

# Συστήματα Πολυμέσων και Εικονική Πραγματικότητα

*A. Ντελόπουλος*

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

2014-2015

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>3</b>
1.1	Τί, πώς, πού;	3
1.1.1	Ορισμός	3
1.1.2	Ταξινόμηση	3
1.1.3	Συστήματα πολυμέσων	4
<b>2</b>	<b>Ψηφιοποίηση</b>	<b>6</b>
2.1	Δειγματοληψία	6
2.1.1	Πρακτικές περιπτώσεις δειγματοληψίας	7
2.1.2	Θεώρημα δειγματοληψίας	9
2.2	Κβαντισμός	17
2.2.1	Ορισμοί	18
2.2.2	Σφάλμα κβαντισμού	21
<b>3</b>	<b>Αναπαράσταση και Αναπαραγωγή Πολυμέσων</b>	<b>23</b>
3.1	Εισαγωγή	23
3.2	Ήχος	25
3.2.1	Εσωτερικές αναπαραστάσεις	25
3.2.2	Αναπαραστάσεις μεταφοράς και αποθήκευσης ήχου	26
3.2.3	Αναπαραστάσεις αναπαραγωγής ήχου	26
3.3	Εικόνα	27
3.3.1	Αναπαράσταση του χρώματος	27
3.3.2	Συστήματα Χρωματικών Συντεταγμένων	31
3.3.3	Δεικτοδότηση Χρώματος	31
3.3.4	Εσωτερικές αναπαραστάσεις εικόνων	33
3.3.5	Αναπαραστάσεις μεταφοράς και αποθήκευσης εικόνων	33
3.3.6	Αναπαραστάσεις αναπαραγωγής εικόνων	35

3.4	Βίντεο . . . . .	36
3.4.1	Εσωτερικές Αναπαραστάσεις Βίντεο - Αναπαράσταση της κίνησης . . . . .	36
3.4.2	Αναπαραστάσεις μεταφοράς και αποθήκευσης βίντεο . . . . .	40
3.4.3	Αναπαραστάσεις αναπαραγωγής βίντεο . . . . .	40
<b>4</b>	<b>Συμπίεση / Κωδικοποίηση</b>	<b>41</b>
4.1	Η ανάγκη . . . . .	41
4.2	Βασικές αρχές . . . . .	42
4.2.1	Κωδικοποίηση και Εντροπία . . . . .	43
4.2.2	Αποσυσχέτιση . . . . .	44
4.2.3	DPCM και ADPCM . . . . .	45
4.2.4	Κωδικοποίηση Μετασχηματισμού . . . . .	46
4.2.5	Εκτίμηση Κίνησης . . . . .	48
4.3	Συμπίεση Ήχου . . . . .	49
4.4	Συμπίεση Εικόνων . . . . .	49
4.5	Συμπίεση Βίντεο . . . . .	49
4.5.1	H.26x . . . . .	49
4.5.2	M-JPEG . . . . .	50
4.5.3	MPEG-1 . . . . .	50
4.5.4	MPEG-2 . . . . .	56
4.5.5	MPEG-4 . . . . .	57
<b>5</b>	<b>Αποθήκευση/Αναζήτηση/Ανάκτηση</b>	<b>59</b>

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Τί, πώς, πού;

#### 1.1.1 Ορισμός

Ο όρος *πολυμέσα* ή *πολυμεσικές οντότητες* αναφέρεται σε όλα τα είδη πληροφορίας και δεδομένων που:

1. Είναι αντιληπτά από τις αισθήσεις μας (Προς το παρόν όραση, ακοή, αφή, στο μέλλον τη γεύση και την όσφρηση !)
2. Είναι διαχειρίσιμα από υπολογιστές. Είναι δηλαδή δυνατή η σύλληψη, επεξεργασία και αναπαγωγή τους με τη χρήση υπολογιστών.
3. Μπορούν να αναπαρασταθούν ψηφιακά (ως διαδοχή από bits)

Το συνθετικό *πολυ-* σηματοδοτεί την πρόθεση να χειριζόμαστε, επεξεργαζόμαστε, αποθηκεύουμε, μεταδίδουμε ή παρουσιάζουμε περισσότερα του ενός ταυτόχρονα.

#### 1.1.2 Ταξινόμηση

- *Κείμενο:*

1. Μη δομημένο (π.χ. ASCII text documents)
2. Δομημένο (word processors, postscript, html, κ.λπ.)
3. hypertext

- *Εικόνες:* Φυσικές φωτογραφίες, συνθετικές εικόνες (2D γραφικά), Fax, Interactive στερεοσκοπικές εικόνες, πανοραμικές εικόνες

1. Έγχρωμες

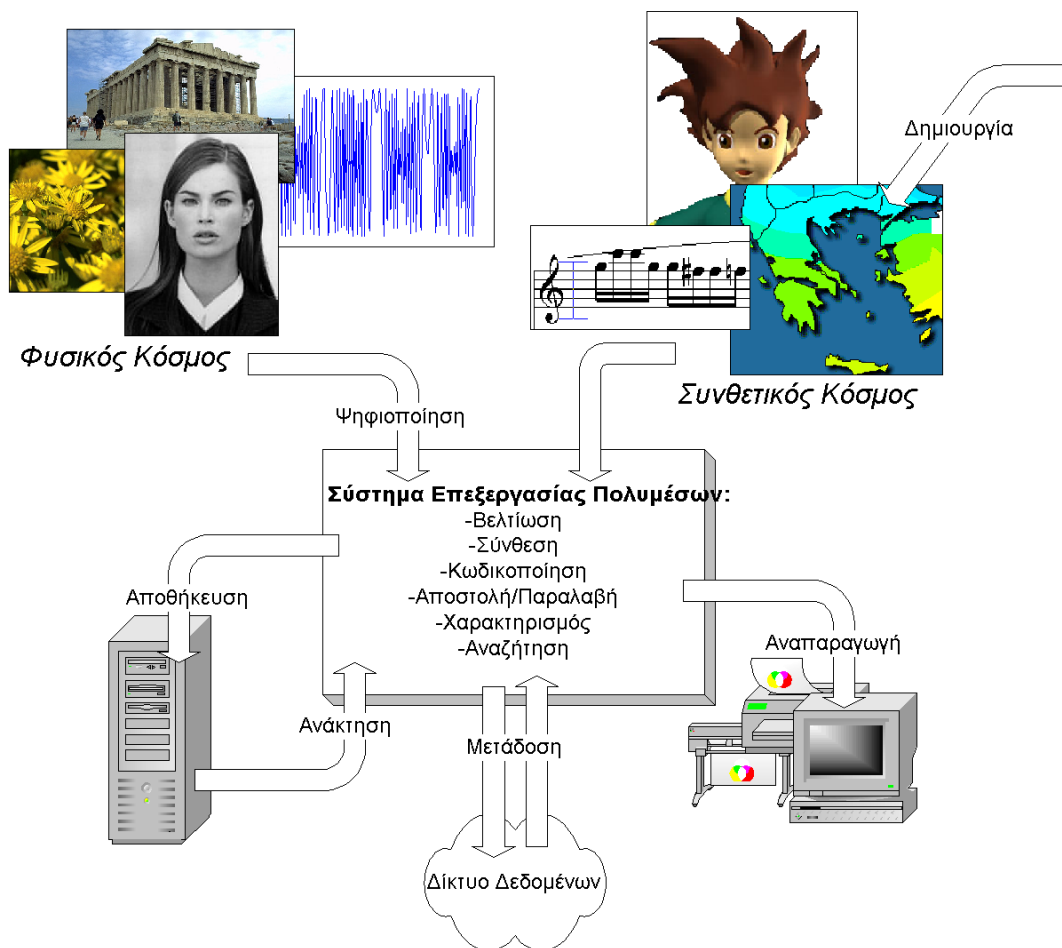
2. Αποχρώσεων του γκρι
  3. Διαδικές
- *Ηχος*: Μουσική, Φωνή, Συνθετικός δομημένος ήχος (π.χ. MIDI)
    1. Μονοκαναλικός - Μονοφωνικός
    2. Πολυκαναλικός (Στερεοφωνικός, surround, κ.λπ.)
  - *Βίντεο*: Φυσικό βίντεο (Ταινίες, ειδήσεις, κ.λπ), 3D γραφικά

### 1.1.3 Συστήματα πολυμέσων

Τα υπολογιστικά συστήματα, είτε υπό τη μορφή κοινών υπολογιστών, είτε υπό τη μορφή εξειδικευμένου hardware και software, που διαχειρίζονται τις πολυμεσικές οντότητες ονομάζονται *συστήματα πολυμέσων*. Όπως προκύπτει από τον εν λόγω ορισμό, τα περισσότερα σύγχρονα υπολογιστικά συστήματα έχουν σε μικρό ή μεγάλο βαθμό το χαρακτήρα του συστήματος πολυμέσων. Μία ενδεικτική αλλά κάθε άλλο παρά εξαντλητική απαρίθμηση τέτοιων συστημάτων περιλαμβάνει:

- τους προσωπικούς υπολογιστές (PCs) και τα περιφερικά τους
- τους web servers
- τα συστήματα μετάδοσης και λήψης ψηφιακής τηλεόρασης
- τα συστήματα παροχής υπηρεσιών βίντεο κατ' απαίτηση (Video on Demand - VoD)
- τα συστήματα τηλεφωνίας (περιλαμβανομένων των κινητών τηλεφώνων)
- τις συσκευές φαξ
- τα συστήματα διαχείρισης δεδομένων ιατρικών εξετάσεων και ειδικότερα απεικονιστικών εξετάσεων (Picture Archiving and Communication Systems - PACS)
- τις παιχνιδιομηχανές (video-games)
- τα συστήματα ψηφιακής επεξεργασίας και έκδοσης εντύπων (Desk-top Publishing DTP και Pre-press)
- τα συστήματα ηλεκτρονικής επεξεργασίας και παραγωγής βίντεο και ήχου (post-production)
- τα συστήματα αρχειοθέτησης οπτικοακουστικού υλικού και ηλεκτρονικών εγγράφων, κ.λπ.

Πάντως παρά την εξαιρετική ποικιλομορφία των συστημάτων πολυμέσων, διακρίνουμε σε όλα τους κάποιες κοινές λειτουργίες που απεικονίζονται στο Σχήμα 1.1



Σχήμα 1.1: Οι λειτουργίες ενός συστήματος πολυμέσων

## Κεφάλαιο 2

# Ψηφιοποίηση: Δειγματοληψία και Κβαντισμός

Το πρώτο στάδιο δημιουργίας οποιουδήποτε είδους πολυμέσων (εκτός αυτών που είναι συνθετικά) είναι η ψηφιοποίηση, η αναπαράσταση δηλαδή ενός φυσικού (αναλογικού) σήματος μέσω μίας ακολουθίας αριθμητικών τιμών. Η εν λόγω ακολουθία είναι μονοδιάστατη, διδιάστατη ή τριδιάστατη κατ'αντιστοιχία με τις διαστάσεις του ψηφιοποιούμενου σήματος (π.χ. ήχου, εικόνας, βίντεο αντίστοιχα). Επιπλέον τα στοιχεία της ακολουθίας μπορεί να είναι βαθμωτά (π.χ. για τις εικόνες τόνων του γκρι) ή διανύσματα (π.χ., για το στερεοφωνικό ήχο).

Η διαδικασία ψηφιοποίησης έχει πάντοτε δύο στάδια: τη δειγματοληψία και τον κβαντισμό.

### 2.1 Δειγματοληψία

Η διαδικασία της δειγματοληψίας συνίσταται στην καταγραφή της τιμής ενός αναλογικού σήματος  $\mathbf{x}_c(\mathbf{v})$ ,  $\mathbf{v} \in R^k$ , σε ένα πλέγμα σημείων  $\mathbf{Tn}$  όπου  $\mathbf{T}$  είναι ένας  $k \times k$  πίνακας και  $\mathbf{n} \in Z^k$ . Αποτέλεσμα της διαδικασίας αυτής είναι η δημιουργία του διακριτού σήματος

$$\mathbf{x}(\mathbf{n}) \triangleq \mathbf{x}_c(\mathbf{Tn}), \mathbf{n} \in Z^k \quad (2.1)$$

Ο πίνακας  $\mathbf{T}$  λέγεται *πίνακας πλέγματος* και καθορίζει πλήρως τις θέσεις των σημείων επί των οποίων θα γίνει η μέτρηση του αρχικού αναλογικού σήματος. Επιπλέον, η απόλυτη τιμή του αντιστρόφου της ορίζουσας του πίνακα πλέγματος,  $\frac{1}{|\det \mathbf{T}|}$ , ισούται με την πυκνότητα δειγματοληψίας (αριθμός δειγμάτων ανά μονάδα μήκους ή χρόνου για  $k = 1$ , εμβαδού για  $k = 2$ , όγκου για  $k = 3$ , κ.λπ.).

### 2.1.1 Πρακτικές περιπτώσεις δειγματοληψίας

*Μονοδιάστατα σήματα* (ήχος):

Στην περίπτωση αυτή είναι  $k = 1$  και ο πίνακας πλέγματος εκφυλίζεται στη βαθμωτή ποσότητα  $T > 0$  που αντιστοιχεί στην περίοδο δειγματοληψίας, ενώ  $f \triangleq \frac{1}{T}$  είναι η συχνότητα δειγματοληψίας.

*Διδιάστατα σήματα* (εικόνες):

Στην περίπτωση αυτή είναι  $k = 2$ . Όταν ο  $2 \times 2$  πίνακας πλέγματος είναι διαγώνιος,

$$\mathbf{T} = \begin{bmatrix} T_h & 0 \\ 0 & T_v \end{bmatrix} \quad (2.2)$$

η δειγματοληψία εφαρμόζεται σε σημεία ορθογώνιου καρτεσιανού πλέγματος και τα στοιχεία της διαγωνίου  $T_h, T_v$  αντιστοιχούν στην περίοδο δειγματοληψίας στην οριζόντια και κατακόρυφη κατεύθυνση αντίστοιχα.

Η εναλλακτική μορφή του (μη διαγώνιου) πίνακα πλέγματος που φαίνεται στην παρακάτω έκφραση,

$$\mathbf{T} = \begin{bmatrix} T & T/2 \\ 0 & T\frac{\sqrt{3}}{2} \end{bmatrix}, T > 0, \quad (2.3)$$

αντιστοιχεί σε δειγματοληψία επί ενός κανονικού εξαγωνικού πλέγματος όπως φαίνεται στο Σχήμα 2.1.

*Τριδιάστατα σήματα* (βίντεο):

Στην περίπτωση αυτή είναι  $k = 3$ . Και στην περίπτωση των τριδιάστατων σημάτων η απλούστερη περίπτωση δειγματοληψίας αντιστοιχεί σε διαγώνιους πίνακες πλέγματος της μορφής,

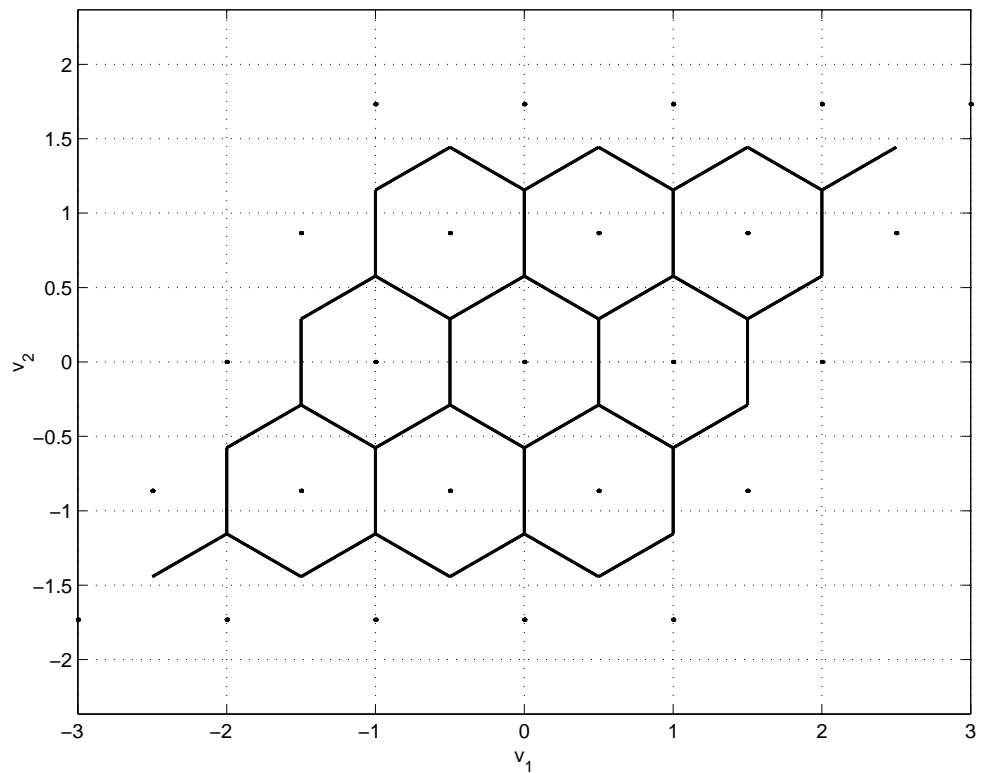
$$\mathbf{T} = \begin{bmatrix} T_t & 0 & 0 \\ 0 & T_h & 0 \\ 0 & 0 & T_v \end{bmatrix} \quad (2.4)$$

Το στοιχείο  $\mathbf{T}_{11} = T_t$  ισούται σ' αυτή την περίπτωση με την περίοδο χρονικής δειγματοληψίας (π.χ.,  $T_t = 1/25$  sec για το ευρωπαϊκό πρότυπο PAL και  $T_t = 1/30$  sec για το αμερικανικό NTSC). Το ψηφιακό βίντεο που δημιουργείται από το συγκεκριμένο είδος δειγματοληψίας αποκαλείται *progressive* ή *non-interlaced* σε αντιδιαστολή προς το *interlaced* βίντεο που παράγεται μέσω του πίνακα πλέγματος της Σχέσης 2.5:

$$\mathbf{T} = \begin{bmatrix} T_t/2 & 0 & 0 \\ 0 & T_h & 0 \\ T_v & 0 & 2T_v \end{bmatrix} \quad (2.5)$$

Στα Σχήματα 2.2 (α) και (β) φαίνονται οι θέσεις των δειγμάτων στον τριδιάστατο χώρο  $(t, h, v)$  όπως προκύπτουν από τις δύο μεθόδους δειγματοληψίας αντίστοιχα. Από τη σύγκριση των σχημάτων





Σχήμα 2.1: Δειγματοληψία επί ενός εξαγωνικού πλέγματος. Οι θέσεις των δειγμάτων αντιστοιχούν στις έντονες στιγμές. Στο ίδιο σχήμα αποτυπώνεται το αντίστοιχο διάγραμμα Voronoi που διαμερίζει το επίπεδο  $(v_1, v_2)$  σε περιοχές των οποίων όλα τα σημεία βρίσκονται πλησιέστερα στο κάθε δείγμα. Στη συγκεκριμένη περίπτωση οι περιοχές αυτές είναι κανονικά εξάγωνα.

προκύπτει ότι τα δείγματα του interlaced βίντεο προκύπτουν από αυτά ενός non-interlaced με τη μισή περίοδο χρονικής δειγματοληψίας ( $T_t/2$ ) αν υποδειγματοληπτήσουμε χωρικά κρατώντας μόνο τις άρτιες οριζόντιες γραμμές για τις άρτιες χρονικές στιγμές και μόνο τις περιττές οριζόντιες γραμμές για τις περιττές χρονικές στιγμές.

Έχει πειραματικά εξακριβωθεί ότι περίοδος χρονικής δειγματοληψίας,  $T_t$ , της τάξης του 1/25 sec ή μικρότερη είναι ικανοποιητική για το ανθρώπινο αισθητήριο. Συγκεκριμένα, όταν τα αντίστοιχα δείγματα παρουσιάζονται στο θεατή με ρυθμό  $1/T_t$  δίνεται στον εγκέφαλο η αίσθηση της *συνεχόμενης, ομαλής κίνησης*. Είναι ενδιαφέρον το γεγονός ότι αυτό ισχύει και στην περίπτωση του interlaced βίντεο όταν κάθε  $T_t/2$  sec *αναανεώνονται* κατά την παρουσίαση του βίντεο στο θεατή εναλλάξ οι άρτιες και οι περιττές γραμμές.

Στην περίπτωση του non-interlaced βίντεο, τα δείγματα που αντιστοιχούν σε μία χρονική στιγμή αποτελούν ένα πλαίσιο (*frame*). Αντίθετα στην περίπτωση του interlaced βίντεο, για τα δείγματα που αντιστοιχούν σε μία χρονική στιγμή χρησιμοποιείται ο όρος *πεδίο (field)*, ενώ ως πλαίσιο θεωρείται το ζεύγος ενός άρτιου κι ενός περιττού πεδίου. (Έτσι και στις δύο περιπτώσεις ένα πλαίσιο περιέχει τον ίδιο αριθμό δειγμάτων).

### 2.1.2 Θεώρημα δειγματοληψίας

Η διαδικασία της δειγματοληψίας όπως ορίστηκε παραπάνω είναι γενικώς *μη αντιστρεπτή*. Τούτο σημαίνει ότι το αρχικό αναλογικό σήμα  $\mathbf{x}_c(\mathbf{v})$  δεν μπορεί να ανασυντεθεί από το διακριτό (προϊόν της δειγματοληψίας) αντίγραφο του  $\mathbf{x}(\mathbf{n})$ . Ωστόσο υπό ορισμένες προϋποθέσεις η εν λόγω ανασύνθεση είναι *εφικτή*. Οι προϋποθέσεις αυτές αφορούν (1) το συχνοτικό περιεχόμενο του αρχικού σήματος και (2) τη συχνότητα δειγματοληψίας.

Για να τεκμηριώσουμε τον παραπάνω ισχυρισμό θα πρέπει να καταφύγουμε στη χρήση του Μετασχηματισμού Fourier (Fourier Transform - FT) και του Διακριτού Μετασχηματισμού Fourier (Discrete Fourier Transform - DFT) των σημάτων  $\mathbf{x}_c(\mathbf{v})$  και  $\mathbf{x}(\mathbf{n})$  αντίστοιχα.

Επειδή οι μετασχηματισμοί αυτοί έχουν 1-1 σχέση με τα αντίστοιχα σήματα, η δυνατότητα ανακατασκευής του  $\mathbf{x}_c(\mathbf{v})$  από το  $\mathbf{x}(\mathbf{n})$  ισοδυναμεί με τη δυνατότητα ανακατασκευής του FT από τον DFT. Η σχέση λοιπόν του FT με τον DFT είναι το ζητούμενο της ακόλουθης ανάλυσης.

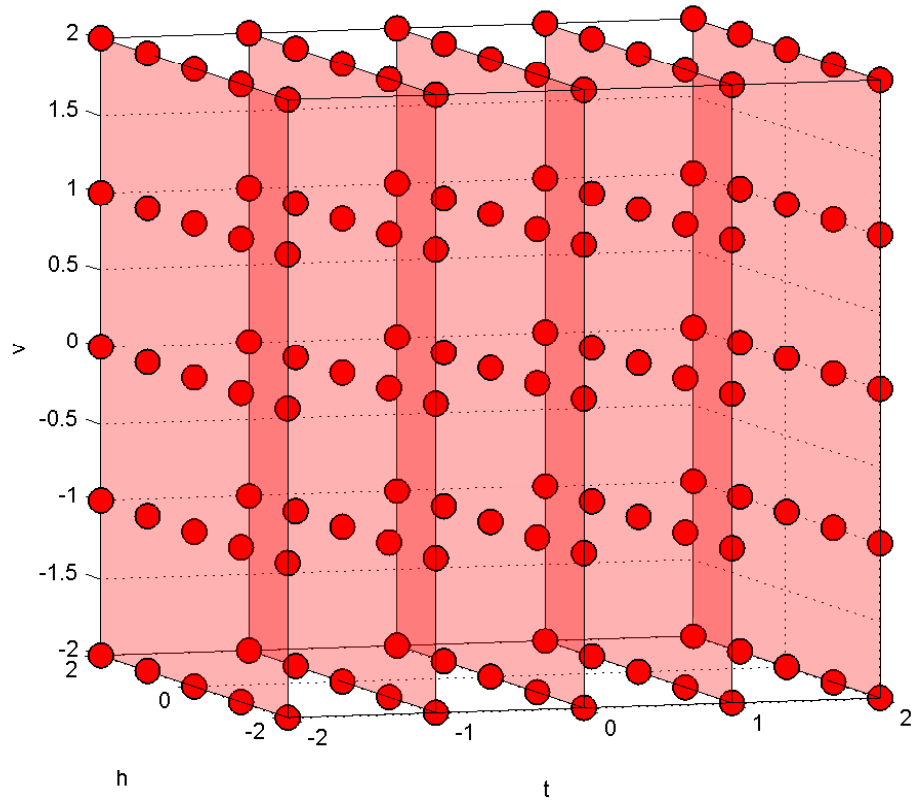
*Ορισμοί:*

Ο *Μετασχηματισμός Fourier* ενός απολύτως ολοκληρώσιμου <sup>1</sup> σήματος  $\mathbf{x}_c(\mathbf{v})$  με  $\mathbf{v} \in R^k$  δίνεται από την έκφραση:

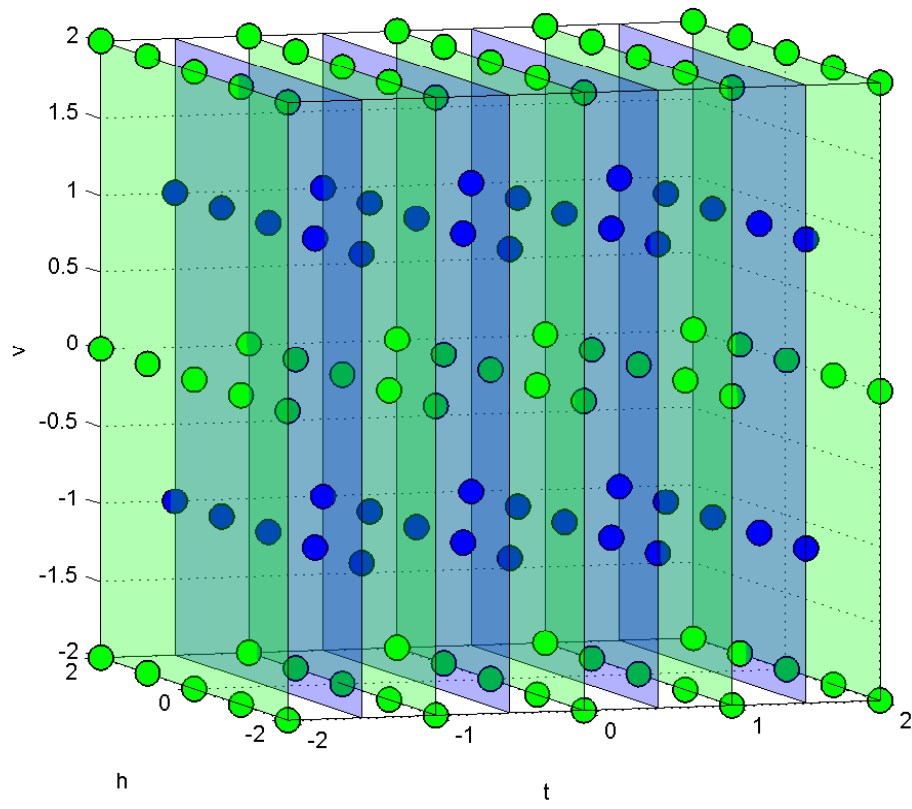
$$\mathbf{X}_c(\mathbf{F}) \triangleq \int_{R^k} \mathbf{x}_c(\mathbf{v}) e^{-j2\pi \mathbf{F}^T \mathbf{v}} d\mathbf{v}, \mathbf{F} \in R^k \quad (2.6)$$

---

<sup>1</sup>  $\int |\mathbf{x}_c(\mathbf{v})| d\mathbf{v} < \infty$



(a)



(b)

Σχήμα 2.2: (a) Non-interlaced και (b) interlaced δειγματοληψία του βίντεο.

Η μεταβλητή  $\mathbf{F}$  αναπαριστά την  $k$ -διάστατη συχνότητα.

Ο *Αντίστροφος Μετασχηματισμός Fourier* επιτρέπει τον υπολογισμό του  $\mathbf{x}_c(\mathbf{v})$  από το  $\mathbf{X}_c(\mathbf{F})$  μέσω της σχέσης:

$$\mathbf{x}_c(\mathbf{v}) = \int_{R^k} \mathbf{X}_c(\mathbf{F}) e^{+j2\pi\mathbf{F}^T\mathbf{v}} d\mathbf{F}, \mathbf{v} \in R^k \quad (2.7)$$

Ο *Διακριτός Μετασχηματισμός Fourier* ενός απολύτως αθροίσιμου <sup>2</sup> σήματος  $\mathbf{x}(\mathbf{n})$  με  $\mathbf{n} \in Z^k$  δίνεται από την έκφραση:

$$\mathbf{X}(\mathbf{f}) \triangleq \sum_{\mathbf{n} \in Z^k} \mathbf{x}(\mathbf{n}) e^{-j2\pi\mathbf{f}^T\mathbf{n}}, \mathbf{f} \in R^k \quad (2.8)$$

Η μεταβλητή  $\mathbf{f}$  αναπαριστά την  $k$ -διάστατη συχνότητα του διακριτού σήματος  $\mathbf{x}(\mathbf{n})$ .

Ο *Αντίστροφος Διακριτός Μετασχηματισμός Fourier* επιτρέπει τον υπολογισμό του  $\mathbf{x}(\mathbf{n})$  από το  $\mathbf{X}(\mathbf{f})$  μέσω της σχέσης:

$$\mathbf{x}(\mathbf{n}) = \int_{c_0} \mathbf{X}(\mathbf{f}) e^{+j2\pi\mathbf{f}^T\mathbf{n}} d\mathbf{f}, \mathbf{n} \in Z^k, \quad (2.9)$$

όπου το χωρίο  $c_0$  ορίζεται ως ακολούθως:

$$c_0 \triangleq \{\xi = [\xi_1, \dots, \xi_k] / |\xi_i| \leq 1/2, i = 1 \dots k\} \quad (2.10)$$

*Ιδιότητες:*

Από τις ιδιότητες των δύο παραπάνω μετασχηματισμών αναφέρουμε εδώ δύο:

1. Γραμμικότητα: Ο (Διακριτός) Μετασχηματισμός Fourier του γραμμικού συνδιασμού δύο ή περισσότερων σημάτων ισούται με τον αντίστοιχο γραμμικό συνδιασμό των (Διακριτών) Μετασχηματισμών Fourier των σημάτων αυτών.
2. Περιοδικότητα του Διακριτού Μετασχηματισμού Fourier: Για κάθε  $\mathbf{m} \in Z^k$  ισχύει ότι

$$\begin{aligned} \mathbf{X}(\mathbf{f} + \mathbf{m}) &= \sum_{\mathbf{n} \in Z^k} \mathbf{x}(\mathbf{n}) e^{-j2\pi(\mathbf{f}+\mathbf{m})^T\mathbf{n}} \\ &= \sum_{\mathbf{n} \in Z^k} \mathbf{x}(\mathbf{n}) e^{-j2\pi\mathbf{f}^T\mathbf{n}} e^{-j2\pi\mathbf{m}^T\mathbf{n}} \\ &= \sum_{\mathbf{n} \in Z^k} \mathbf{x}(\mathbf{n}) e^{-j2\pi\mathbf{f}^T\mathbf{n}} \\ &= \mathbf{X}(\mathbf{f}) \end{aligned}$$

όπου εκμεταλλευτήκαμε την ισότητα  $e^{-j2\pi\mathbf{m}^T\mathbf{n}} = 1$  που ισχύει για  $\mathbf{m}, \mathbf{n} \in Z^k$

*Η Σχέση του Διακριτού Μετασχηματισμού Fourier με το Μετασχηματισμό Fourier:*

---


$$^2 \sum_{\mathbf{n}} |\mathbf{x}(\mathbf{n})| < \infty$$

Έστω διακριτό σήμα  $\mathbf{x}(\mathbf{n}) = \mathbf{x}_c(\mathbf{T}\mathbf{n})$  που προκύπτει από το αντίστοιχο αναλογικό σήμα  $\mathbf{x}_c(\mathbf{n})$  με δειγματοληψία πάνω στο πλέγμα που ορίζεται από τον πίνακα  $\mathbf{T}$ . Με βάση τη Σχέση 2.7 θα έχουμε:

$$\begin{aligned}\mathbf{x}(\mathbf{n}) &= \int_{R^k} \mathbf{X}_c(\mathbf{F}) e^{+j2\pi\mathbf{F}^T\mathbf{T}\mathbf{n}} d\mathbf{F} \\ &= \frac{1}{|\det T|} \int_{R^k} \mathbf{X}_c(\mathbf{T}^{-T}\mathbf{f}) e^{+j2\pi\mathbf{f}^T\mathbf{n}} d\mathbf{f} \\ &= \frac{1}{|\det T|} \int_{R^k} \mathbf{X}_c(\mathbf{S}\mathbf{f}) e^{+j2\pi\mathbf{f}^T\mathbf{n}} d\mathbf{f},\end{aligned}\quad (2.11)$$

όπου χρησιμοποιήσαμε την αλλαγή μεταβλητής  $\mathbf{f} = \mathbf{T}^T\mathbf{F}$  και συνακόλουθα τη σχέση  $d\mathbf{f} = |\det T|d\mathbf{F}$  και για ευκολία τον πίνακα  $\mathbf{S} \triangleq \mathbf{T}^{-T} \equiv (\mathbf{T}^T)^{-1}$ .

Στη συνέχεια θα εκφράσουμε το τελευταίο ολοκλήρωμα της παραπάνω έκφρασης σαν άθροισμα ολοκληρωμάτων της ίδιας μορφής πάνω σε κατάλληλα χωρία του  $R^k$ . Ειδικότερα ορίζουμε τα χωρία  $c_m \triangleq \{\mathbf{f}/\mathbf{f} - \mathbf{m} \in c_0\}$  όπου το  $c_0$  ορίστηκε στη Σχέση 2.10. Μπορεί να διαπιστώσει κανείς ότι τα  $c_m$  με  $\mathbf{m} \in Z^k$  αποτελούν διαμέριση του  $R^k$ . Κατά συνέπεια η τελευταία έκφραση της 2.11 μπορεί να αναπτυχθεί ως εξής:

$$\begin{aligned}\mathbf{x}(\mathbf{n}) &= \frac{1}{|\det T|} \sum_{\mathbf{m} \in Z^k} \int_{c_m} \mathbf{X}_c(\mathbf{S}\mathbf{f}) e^{+j2\pi\mathbf{f}^T\mathbf{n}} d\mathbf{f} \\ &= \frac{1}{|\det T|} \sum_{\mathbf{m} \in Z^k} \int_{c_0} \mathbf{X}_c(\mathbf{S}(\mathbf{f} - \mathbf{m})) e^{+j2\pi(\mathbf{f}-\mathbf{m})^T\mathbf{n}} d\mathbf{f} \\ &= \frac{1}{|\det T|} \int_{c_0} \sum_{\mathbf{m} \in Z^k} \mathbf{X}_c(\mathbf{S}(\mathbf{f} - \mathbf{m})) e^{+j2\pi\mathbf{f}^T\mathbf{n}} d\mathbf{f},\end{aligned}\quad (2.12)$$

όπου, και εδώ, εκμεταλλευτήκαμε την ισότητα  $e^{-j2\pi\mathbf{m}^T\mathbf{n}} = 1$  που ισχύει για  $\mathbf{m}, \mathbf{n} \in Z^k$

Συγκρίνοντας τις Σχέσεις 2.9 και 2.12 καταλήγουμε στο βασικό συμπέρασμα ότι:

$$\mathbf{X}(\mathbf{f}) = \frac{1}{|\det T|} \sum_{\mathbf{m} \in Z^k} \mathbf{X}_c(\mathbf{S}(\mathbf{f} - \mathbf{m})) \quad (2.13)$$

Η Σχέση 2.13 μεταξύ του Μετασχηματισμού Fourier ενός αναλογικού σήματος και του Διακριτού Μετασχηματισμού Fourier σήματος που προκύπτει μετά από δειγματοληψία είναι ενδιαφέρουσα αλλά κάπως απογοητευτική: Ο Διακριτός Μετασχηματισμός ισούται με την επαλληλία άπειρων μετατοπισμένων αντιγράφων του Μετασχηματισμού Fourier. Το κάθε αντίγραφο μάλιστα είναι παραμορφωμένο από την αλλαγή συντεταγμένων  $\mathbf{f} \rightarrow \mathbf{S}\mathbf{f}$ .

Η ακόλουθη Πρόταση προσδιορίζει τις συνθήκες για την ανακατασκευή του  $\mathbf{X}_c(\mathbf{F})$  από το  $\mathbf{X}(\mathbf{f})$

**Πρόταση 2.1.** *Η ανακατασκευή του Μετασχηματισμού Fourier,  $\mathbf{X}_c(\mathbf{F})$ , ενός αναλογικού σήματος από το Διακριτό Μετασχηματισμό Fourier σήματος που προκύπτει μετά από δειγματοληψία*

είναι εφικτή αν και μόνο αν οι περιοχές υποστήριξης (= μη μηδενισμού) των  $\mathbf{X}_c(\mathbf{F} - \mathbf{S}\mathbf{m}_1)$  και  $\mathbf{X}_c(\mathbf{F} - \mathbf{S}\mathbf{m}_2)$  δεν επικαλύπτονται για κάθε  $\mathbf{m}_1 \neq \mathbf{m}_2$  με  $\mathbf{m}_1, \mathbf{m}_2 \in Z^k$

Απόδειξη:

Αν ισχύει η υπόθεση της πρότασης από την 2.13 προκύπτει ότι

$$\mathbf{X}_c(\mathbf{F}) = \begin{cases} |\det T| \mathbf{X}(\mathbf{S}^{-1}\mathbf{F}) \equiv |\det T| \mathbf{X}(\mathbf{T}^T\mathbf{F}) & \text{για } F \text{ εντός της περιοχής υποστήριξης} \\ 0 & \text{αλλιώς} \end{cases}$$

Ο προσεκτικός αναγνώστης θα παρατηρήσει ότι ο πίνακας  $\mathbf{S} = \mathbf{T}^{-T}$  ορίζει ένα άλλο πλέγμα στο πεδίο των συχνοτήτων. Οι συνθήκη της Πρότασης 2.1 ικανοποιείται όταν ολόκληρη η περιοχή υποστήριξης της  $\mathbf{X}_c(\mathbf{F})$  περικλείεται στο χωρίο  $\varphi_0$  που ορίζεται από το διάγραμμα Voronoi των σημείων του νέου αυτού πλέγματος και έχει κέντρο το σημείο  $\mathbf{S}\mathbf{0} = \mathbf{0}$ .

Από την παραπάνω παρουσίαση προκύπτουν οι ακόλουθες τρεις σημαντικές παρατηρήσεις:

**Παρατήρηση 2.1.** Η Πρόταση 2.1 θέτει τις προϋποθέσεις για ανακατασκευή του Μετασχηματισμού Fourier,  $\mathbf{X}_c(\mathbf{F})$ , από το Διακριτό Μετασχηματισμό Fourier,  $\mathbf{X}(\mathbf{f})$ . Επειδή υπάρχει 1-1 αντιστοιχία μεταξύ των συναρτήσεων αυτών και των σημάτων  $\mathbf{x}_c(\mathbf{v})$  και  $\mathbf{x}(\mathbf{n})$  αντίστοιχα, οι προϋποθέσεις της Πρότασης 2.1 διασφαλίζουν τη δυνατότητα τέλει ανάκτησης του αρχικού αναλογικού σήματος  $\mathbf{x}_c(\mathbf{v})$  από τα (αρκετά πυκνά) δείγματά του!

**Παρατήρηση 2.2.** Προϋπόθεση ώστε ολόκληρη η περιοχή υποστήριξης της  $\mathbf{X}_c(\mathbf{F})$  να περικλείεται σε ένα συγκεκριμένο χωρίο είναι το σήμα  $\mathbf{x}_c(\mathbf{v})$  να είναι πεπερασμένου εύρους ζώνης. Στην πραγματικότητα το εύρος ζώνης του αναλογικού σήματος  $\mathbf{x}_c(\mathbf{v})$  ορίζεται ως η διάσταση της περιοχής υποστήριξης της  $\mathbf{X}_c(\mathbf{F})$ . Τούτο θα γίνει πιο εμφανές στην επόμενη ενότητα που θα εξειδικεύσουμε για τα πιο οικεία μονοδιάστατα σήματα ( $k = 1$ ).

**Παρατήρηση 2.3.** Με την υπόθεση πλέον ότι το σήμα  $\mathbf{x}_c(\mathbf{v})$  έχει κάποιο συγκεκριμένο (πεπερασμένο) εύρος ζώνης, η υπόθεση της Πρότασης 2.1 εξαρτάται από το μέγεθος του χωρίου  $\varphi_0$  που όπως προαναφέρθηκε ορίζεται από το διάγραμμα Voronoi των σημείων του πλέγματος  $\mathbf{S}\mathbf{n}$  και περιέχει την αρχή των αξόνων. Είναι φανερό ότι το μέγεθος του εν λόγω χωρίου εξαρτάται αποκλειστικά από τον πίνακα  $\mathbf{S}$  ή ισοδύναμα από τον πίνακα  $\mathbf{T}$ . Ξεφεύγει από τους σκοπούς αυτού του κειμένου η αυστηρή θεμελίωση σχέσεων μεταξύ των δύο πινάκων και της γεωμετρίας του χωρίου. Αξίζει ωστόσο να διατυπώσουμε την ακόλουθη αδρή διαισθητική διαπίστωση: Όσο ο πίνακας  $\mathbf{T}$  μικραίνει τόσο ο πίνακας  $\mathbf{S} = \mathbf{T}^{-T}$  μεγαλώνει, άρα τα σημεία του πλέγματος  $\mathbf{S}\mathbf{n}$  απομακρύνονται μεταξύ τους και συνεπώς η διάσταση του χωρίου  $\varphi_0$  μεγαλώνει. Συνεπώς για να ικανοποιείται η απαίτηση της Πρότασης 2.1 πρέπει, για δεδομένο εύρος ζώνης του  $\mathbf{x}_c(\mathbf{v})$ , ο πίνακας  $\mathbf{T}$  να είναι ‘αρκετά μικρός’ ή ισοδύναμα η δειγματοληψία αρκετά πυκνή. Πόσο πυκνή όμως; Το ακόλουθο Θεώρημα 2.1 δίνει την

απάντηση για την περίπτωση διαγώνιων πινάκων πλέγματος (δηλαδή δειγματοληψίας σε ορθογώνιο πλέγμα σημείων).

**Θεώρημα 2.1.** Έστω σήμα  $\mathbf{x}_c(\mathbf{v})$ , ( $\mathbf{v} \in R^k$ ) με Μετασχηματισμό Fourier  $\mathbf{X}_c(\mathbf{F})$  τέτοιο ώστε:

$$\mathbf{X}_c(\mathbf{F}) = \mathbf{0} \text{ για } \mathbf{F} \notin \mathbf{B}, \quad (2.14)$$

όπου  $\mathbf{B} \triangleq \{\mathbf{F} = [F_1 \cdots F_k]^T / |F_i| \leq b_i, i = 1 \cdots k\}$  στο χώρο  $R^k$  της συχνότητας<sup>3</sup>. Έστω επίσης ότι το  $\mathbf{B}$  είναι το μικρότερο (ορθογώνιο) χωρίο του  $R^k$  που έχει αυτή την ιδιότητα (να περικλείει την περιοχή υποστήριξης του Μετασχηματισμού Fourier  $\mathbf{X}_c(\mathbf{F})$ ). Αν το διακριτό σήμα  $\mathbf{x}(\mathbf{n})$  ( $\mathbf{v} \in Z^k$ ) έχει προέλθει από δειγματοληψία του  $\mathbf{x}_c(\mathbf{v})$  πάνω στο πλέγμα  $\mathbf{Tn}$  όπου ο πίνακας πλέγματος,

$$\mathbf{T} = \begin{bmatrix} T_1 & 0 & \dots & 0 \\ 0 & T_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & & 0 \\ 0 & \dots & 0 & T_k \end{bmatrix}, \quad (2.15)$$

τότε το αναλογικό σήμα  $\mathbf{x}_c(\mathbf{v})$  μπορεί να ανακτηθεί πλήρως από το διακριτό  $\mathbf{x}(\mathbf{n})$  αν και μόνο αν,

$$\frac{1}{T_i} \geq 2b_i, i = 1 \cdots k. \quad (2.16)$$

Απόδειξη:

Ας ονομάσουμε  $\mathbf{B}_H$  το χωρίο που προκύπτει από τη μετατόπιση του  $\mathbf{B}$  κατά  $\mathbf{H} \in \mathbf{R}^k$ , είναι δηλαδή  $\mathbf{B}_H \triangleq \{\mathbf{F} = [F_1 \cdots F_k]^T / |F_i - H_i| \leq b_i, i = 1 \cdots k\}$ . Εύκολα μπορούμε να διαπιστώσουμε ότι δύο οποιαδήποτε μετατοπισμένα αντίγραφα  $\mathbf{X}_c(\mathbf{F} - \mathbf{S}\mathbf{m}_1)$  και  $\mathbf{X}_c(\mathbf{F} - \mathbf{S}\mathbf{m}_2)$  της  $\mathbf{X}_c(\mathbf{F})$  σαν αυτά που αναφέρονται στην υπόθεση της Πρότασης 2.1 έχουν περιοχή υποστήριξης που περικλείεται στα χωρία  $\mathbf{B}_{\mathbf{S}\mathbf{m}_1}$  και  $\mathbf{B}_{\mathbf{S}\mathbf{m}_2}$  αντίστοιχα.

Συνεπώς αν αυτά τα χωρία δεν επικαλύπτονται το ίδιο θα ισχύει και για τις περικλειόμενες περιοχές υποστήριξης. Αντίστροφα, μπορεί να αποδειχτεί ότι αν δύο γειτονικά χωρία επικαλύπτονται το ίδιο ισχύει και για τις περιοχές υποστήριξης (η δεύτερη αυτή συνεπαγωγή αποδεικνύεται με χρήση των συμμετριών του Μετασχηματισμού Fourier αλλά στο παρόν κείμενο θα παραλείψουμε την απόδειξη). Συνοψίζοντας, κριτήριο για την ορθότητα της υπόθεσης της Πρότασης 2.1 είναι το αν (γειτονικά) χωρία  $\mathbf{B}_{\mathbf{S}\mathbf{m}_1}$  και  $\mathbf{B}_{\mathbf{S}\mathbf{m}_2}$  επικαλύπτονται. Θα δείξουμε ότι ικανή και αναγκαία συνθήκη για να μην επικαλύπτονται είναι η συνθήκη της Σχέσης 2.16.

Παρατηρούμε ότι για το διαγώνιο πίνακα πλέγματος  $\mathbf{T}$  του Θεωρήματος ο αντίστοιχος πίνακας

---

<sup>3</sup>Για  $k = 1, 2, 3$  το  $\mathbf{B}$  είναι διάστημα ευθείας, ορθογώνιο παραλληλόγραμμο, και ορθογώνιο παραλληλεπίπεδο αντίστοιχα, πάντα με κέντρο την αρχή των αξόνων.

$\mathbf{S} = \mathbf{T}^{-T}$  είναι επίσης διαγώνιος με διαγώνια στοιχεία  $S_i = 1/T_i$ .

Συνεπώς τα κέντρα των προαναφερθέντων χωρίων έχουν τη μορφή

$$\begin{aligned}\mathbf{S}\mathbf{m}_1 &= \left[ \frac{m_{11}}{T_1}, \frac{m_{12}}{T_2}, \dots, \frac{m_{1k}}{T_k} \right] \text{ και} \\ \mathbf{S}\mathbf{m}_2 &= \left[ \frac{m_{21}}{T_1}, \frac{m_{22}}{T_2}, \dots, \frac{m_{2k}}{T_k} \right],\end{aligned}$$

αντίστοιχα. Εύκολα συμπεραίνουμε ότι δύο οποιαδήποτε τέτοια χωρία δεν επικαλύπτονται αν  $\frac{m_{1i}}{T_i} - \frac{m_{2i}}{T_i} < 2b_i$  για  $m_{1i} \neq m_{2i}$ . Επειδή ή δυσμενέστερη περίπτωση είναι  $|m_{1i} - m_{2i}| = 1$  (γειτονικά χωρία) συμπεραίνουμε ότι αν ισχύει η Σχέση 2.16 διαφορετικά χωρία δεν επικαλύπτονται και συνεπώς ούτε οι περιοχές υποστήριξης των αντιγράφων της  $\mathbf{X}_c(\mathbf{v})$  επικαλύπτονται. Αντιστρόφως αν δεν ισχύει η Σχέση 2.16 δύο γειτονικά χωρία που αντιστοιχούν σε  $\mathbf{m}_1 \neq \mathbf{m}_2$  για τα οποία όλες πλην μιας συντεταγμένης τους είναι ίσες θα επικαλύπτονται. Αυτό αποδεικνύει το ‘μόνο αν’ τμήμα του Θεωρήματος. Η προηγούμενη ανάλυση γίνεται πιο κατανοητή παρατηρώντας τις διαφορές στα Σχήματα 2.3(α) και 2.3(β)

Αξίζει στο σημείο αυτό να κάνουμε μερικές παρατηρήσεις:

**Παρατήρηση 2.4.** Το Θεώρημα 2.1 αποτελεί εξειδίκευση της Πρότασης 2.1 για την περίπτωση δειγματοληψίας πάνω σε ορθογώνια (καρτεσιανά) πλέγματα.

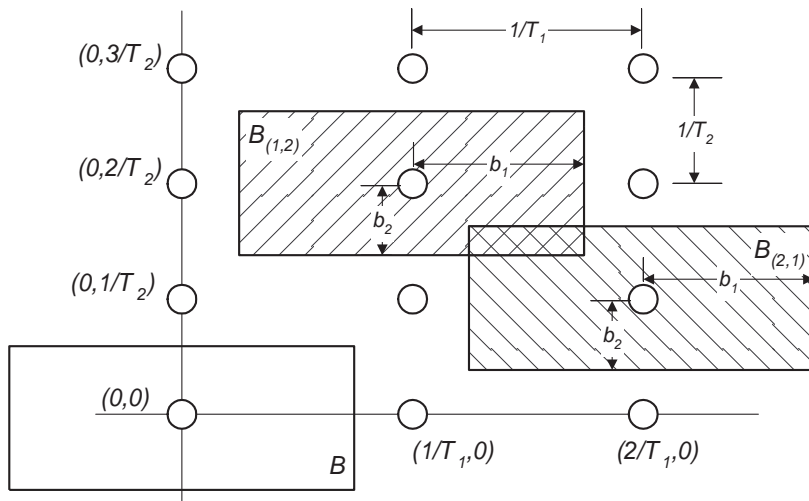
**Παρατήρηση 2.5.** Για την ειδικότερη περίπτωση μονοδιάστατων σημάτων ( $k = 1$ ) το Θεώρημα 2.1 εκφράζει το γνωστό νόμο του Nyquist: ‘Η συχνότητα δειγματοληψίας,  $f_s$ , πρέπει να είναι τουλάχιστον διπλάσια της μέγιστης συχνότητας,  $F_{max}$ , που εμφανίζεται στο δειγματοληπτούμενο αναλογικό σήμα’. Η αντιστοιχία με τις παραμέτρους του θεωρήματος είναι ότι  $f_s \equiv 1/T_1$  και  $F_{max} \equiv b_1$ .

**Παρατήρηση 2.6.** Τόσο η Πρόταση 2.1 όσο και το Θεώρημα 2.1 θέτουν συνθήκες που εξασφαλίζουν τη δυνατότητα ανακατασκευής ενός αναλογικού σήματος από ένα διακριτό (δειγματοληπτιμένο) αντίγραφό του. Δεν υποδεικνύουν ωστόσο κάποια πρακτική μέθοδο για την υλοποίηση της συγκεκριμένης διαδικασίας. Οι μέθοδοι μετατροπής ψηφιακών σημάτων σε αναλογικά (DAC: Digital to Analog Conversion) αποτελούν σημαντικό τομέα της επεξεργασίας σήματος αλλά ξεφεύγουν από το στόχο αυτού του κειμένου.

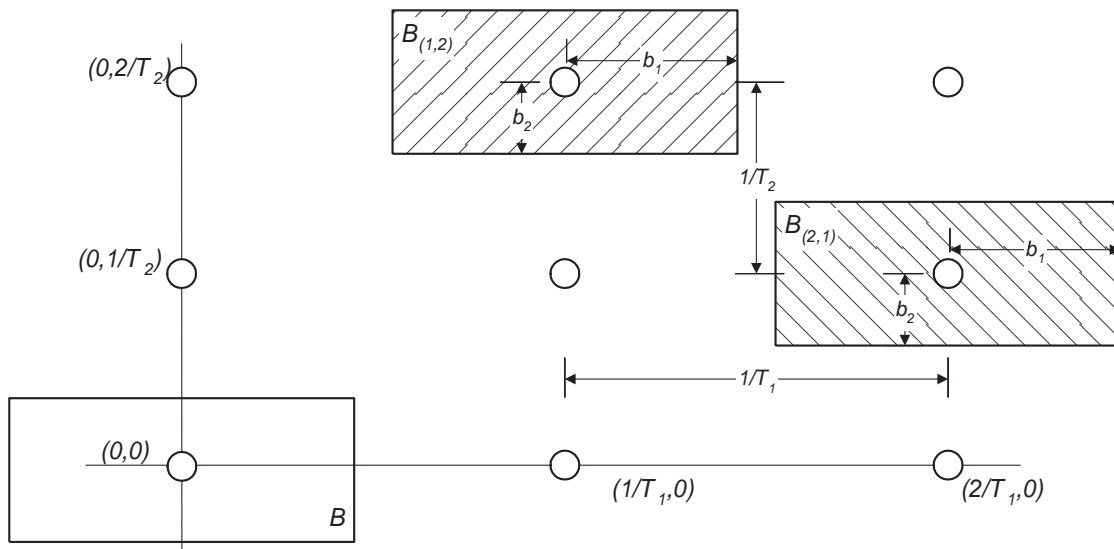
**Παρατήρηση 2.7.** Στην πράξη οι συνθήκες του Θεωρήματος 2.1 επιβάλλονται ‘δια της βίας’. Προ της δειγματοληψίας το αναλογικό σήμα οδηγείται σε βαθυπερατό φίλτρο που αποκόπτει όσες συχνότητες βρίσκονται έξω από το επιλεγμένο χωρίο  $\mathbf{B}$ .

**Παρατήρηση 2.8.** Τί συμβαίνει όμως κατά τη δειγματοληψία αν δεν τηρηθούν οι συνθήκες τέλει ανακατασκευής; Παρουσιάζεται το φαινόμενο της αναδίπλωσης συχνοτήτων (aliasing). Για την ευκολότερη κατανόηση του φαινομένου θα αναφερθούμε σε μονοδιάστατα σήματα ( $k = 1$ ).



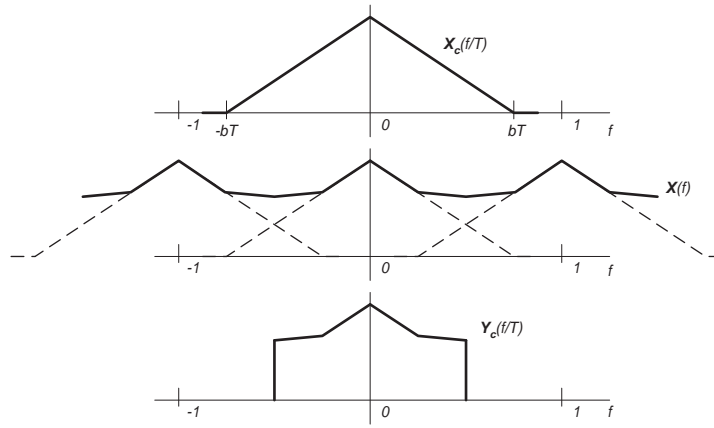


(α)



(β)

Σχήμα 2.3: Δειγματοληψία (α) με και (β) χωρίς alias



Σχήμα 2.4: Σχηματική αναπαράσταση του φαινομένου της αναδίπλωσης συχνοτήτων (aliasing)

Ας θεωρήσουμε λοιπόν το αναλογικό σήμα  $\mathbf{x}_c(v)$  με Μετασχηματισμό Fourier  $\mathbf{X}_c(F)$  και περιοχή υποστήριξης το διάστημα  $[-b, b]$ . Έστω επίσης το διακριτό αντίγραφο του  $\mathbf{x}(n)$  με Διακριτό Μετασχηματισμό Fourier  $\mathbf{X}(f)$  το οποίο προέρχεται από δειγματοληψία με περίοδο  $T$  όπου

$$\frac{1}{T} < 2b.$$

Ανακαλώντας τη Σχέση 2.13 για τη συγκεκριμένη περίπτωση

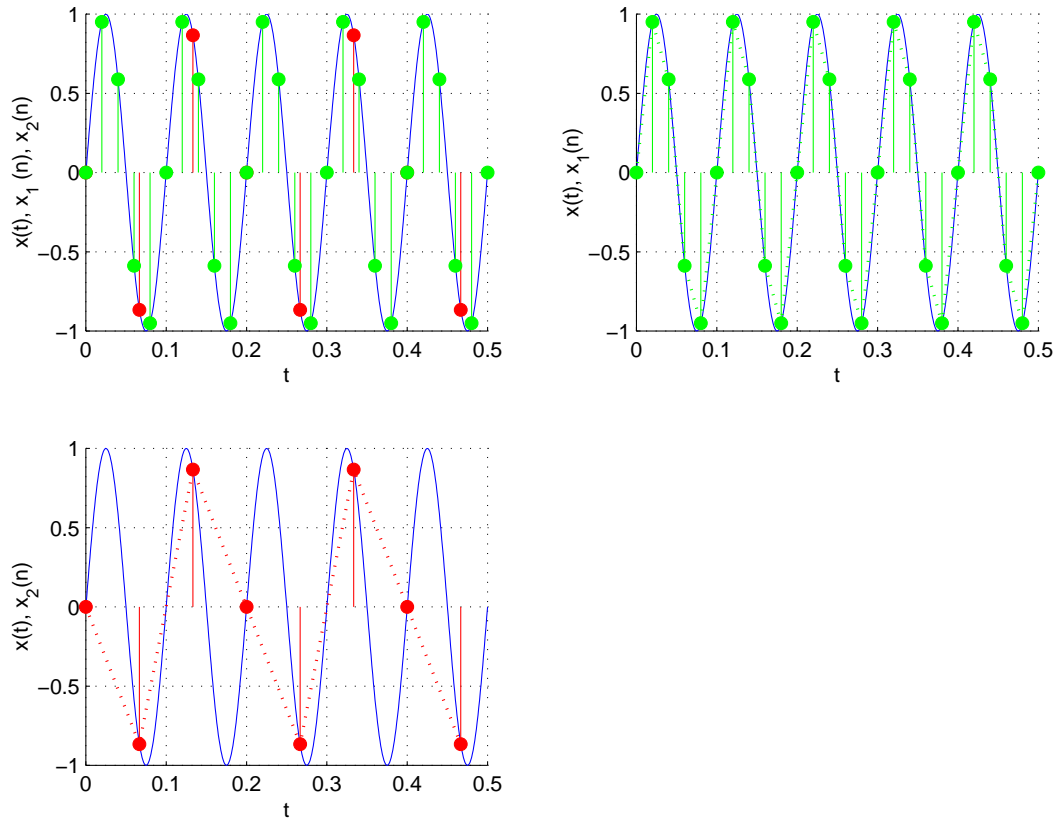
$$\mathbf{X}(f) = \frac{1}{T} \sum_{m \in \mathbb{Z}} \mathbf{X}_c\left(\frac{f}{T} - \frac{m}{T}\right),$$

Στο πάνω μέρος του Σχήματος 2.4 αναπαριστούμε το διάγραμμα μίας αυθαίρετης  $\mathbf{X}_c(F)$  ενώ στο μεσαίο διάγραμμα του ίδιου σχήματος φαίνεται μέρος του προκύπτοντος Διακριτού Μετασχηματισμού Fourier ( $\mathbf{X}(f)$ ). Παρατηρούμε ότι στο ίδιο αποτέλεσμα θα καταλήγαμε αν αντί του  $\mathbf{x}_c(v)$  δειγματοληψτούσαμε το  $\mathbf{y}_c(v)$  του οποίου ο Μετασχηματισμός Fourier,  $\mathbf{Y}_c(F)$ , αναπαρίσται στο κάτω μέρος του σχήματος.

Στα Σχήματα 2.5 και 2.6 φαίνονται τα αποτελέσματα της αναδίπλωσης συχνοτήτων σε μονοδιάστατα και διδιάστατα σήματα.

## 2.2 Κβαντισμός

Η διαδικασία του κβαντισμού είναι το δεύτερο στάδιο στην ψηφιοποίηση ενός αναλογικού σήματος. Σε ελεύθερη απόδοση, με τον όρο *κβαντισμός* (*quantization*) εννοούμε τη στρογγυλοποίηση των τιμών ενός σήματος σε μία από τις πεπερασμένου πλήθους στάθμες που έχουν προκαθοριστεί. Η ανάγκη της διαδικασίας κβαντισμού προκύπτει από την επιλογή μας (η μόνη τεχνικά εφικτή επιλογή) να αναπαριστούμε την τιμή κάθε δείγματος ενός σήματος με πεπερασμένο πλήθος bits. Έτσι αν επιλέξουμε να χρησιμοποιήσουμε  $b$  bits ανά δείγμα οι διαθέσιμες στάθμες κβαντισμού είναι συνολικά το πολύ  $N = 2^b$ .



Σχήμα 2.5: Δειγματοληψία ημιτονικού σήματος χωρίς και με αναδίπλωση.

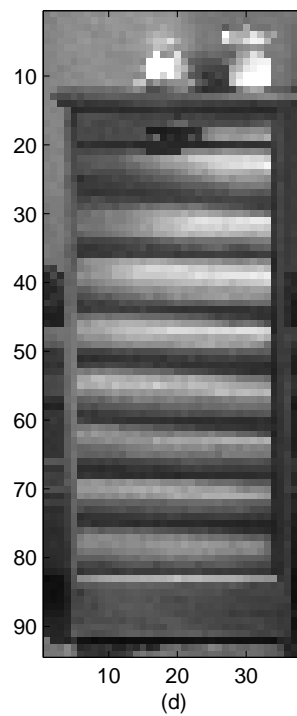
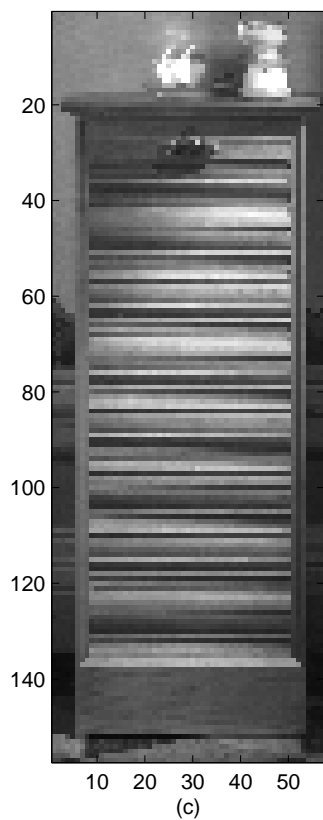
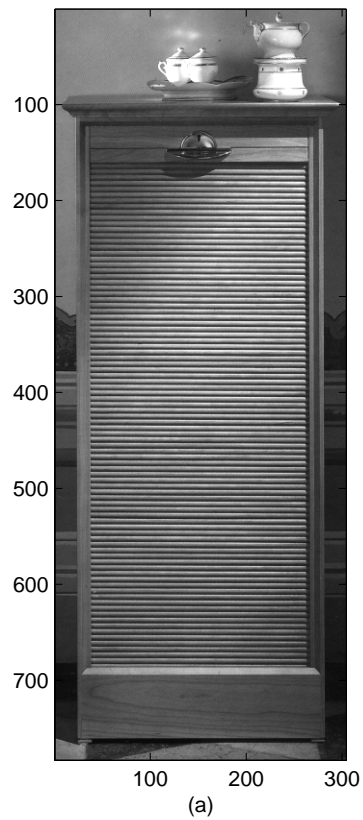
Η ίδια διαδικασία κβαντισμού εφαρμόζεται συχνά όχι σε πρωτογενή σήματα (αμέσως μετά τη δειγματοληψία τους) αλλά και σε παράγωγα μεγέθη όπως σήματα που προκύπτουν από ενδιάμεσους υπολογισμούς (π.χ. μετά από εφαρμογή ψηφιακών φίλτρων) ή και σήματα μετασχηματισμένα στο πεδίο της συχνότητας (π.χ. στους συντελεστές του Διακριτού Μετασχηματισμού Fourier).

Στην απλούστερη περίπτωση ο κβαντισμός είναι μία διαδικασία που εφαρμόζεται σε μεμονωμένα δείγματα ενός σήματος αλλά έχει επίσης ενδιαφέρον (και ευρύ πεδίο εφαρμογών) η διαδικασία του κβαντισμού διανυσμάτων αποτελούμενων από περισσότερα του ενός δείγματα. Αναφερόμαστε στην δεύτερη περίπτωση στον *διανυσματικό κβαντισμό (vector quantization)*

### 2.2.1 Ορισμοί

Στη συνέχεια παρατίθενται οι αυστηρότεροι ορισμοί του απλού (βαθμωτού) και του διανυσματικού κβαντισμού και αναλύεται, συνοπτικά, η αλλοίωση του σήματος εξαιτίας της συγκεκριμένης διαδικασίας.

**Ορισμός 2.1.** Έστω διαμέριση  $\mathbf{P} = [P_0, P_1, \dots, P_{N-1}]^T$  του συνόλου  $R$  των πραγματικών αριθμών όπου τα  $P_i \subseteq R$  ( $i = 0, \dots, N-1$ ) είναι συνεκτικά διαστήματα της πραγματικής ευθείας. Έστω επίσης το σύνολο συμβόλων  $\mathbf{S} = \{s_0, s_1, \dots, s_{N-1}\}$ . Ορίζουμε ως βαθμωτό κβαντιστή ή



Σχήμα 2.6: Διδιάστατη δειγματοληψία φυσικής φωτογραφίας ( $k = 2$ ): Στην πάνω αριστερά δεν υπάρχει αναδίπλωση. Στην πάνω δεξιά η αναδίπλωση αρχίζει να διακρίνεται ενώ η αναδίπλωση στις δύο κάτω φωτογραφίες προκαλεί εξαφάνιση υψηλών χωρικών συχνοτήτων και εμφάνιση πλασματικών χαμηλών.

απλώς κβαντιστή  $N$  επιπέδων τον τελεστή  $Q[ ]$ :

$$s_i = Q[x] \quad \forall x \in P_i, \quad i = 0, \dots, N-1 \quad (2.17)$$

Ορίζουμε αντίστοιχα ως αποκβαντιστή (dequantizer) τη συνάρτηση

$$x_i = \widehat{Q}[s_i] \quad i = 0, \dots, N-1 \quad (2.18)$$

που απεικονίζει τα  $N$  σύμβολα  $s_i \in S$  σε πραγματικούς αριθμούς  $x_i \in L$ . Τα στοιχεία  $(x_i)$  του πεπερασμένου συνόλου  $L \subseteq R$  ονομάζονται στάθμες ή επίπεδα κβαντισμού των  $Q$  και  $\widehat{Q}$ .

Για συντόμευση στη γραφή του συνδιασμού κβαντισμού/αποκβαντισμού θα τον αναπαριστούμε με τον τελεστή  $\overline{Q}[ ]$ , δηλαδή,

$$\overline{Q}[x] \triangleq \widehat{Q}[Q[x]] \quad (2.19)$$

Από τον παραπάνω ορισμό γίνεται φανερό ότι η διαδικασία κβαντισμού είναι μη αντιστρεπτή διαδικασία δεδομένου ότι η διαδοχή κβαντισμού-αποκβαντισμού απεικονίζει οποιοδήποτε τιμή σήματος βρίσκεται εντός του διαστήματος  $P_i$  στο ίδιο σύμβολο  $s_i$  και στη συνέχεια στον ίδιο πραγματικό αριθμό  $x_i$ .

Στη συνηθέστερη περίπτωση τα διαστήματα  $P_i$  της προαναφερθείσας διαμέρισης  $P$  ορίζονται μέσω του συνόλου των άκρων τους για τα οποία χρησιμοποιείται ο όρος *στάθμες αποφάσεων*. Συγκεκριμένα, αν  $\mathbf{d} = [d_1, d_2, \dots, d_{N-1}]^T$  με  $d_i \in R$  ( $i = 1, \dots, N-1$ ) οι  $N-1$  στάθμες αποφάσεων, ορίζουμε την αντίστοιχη διαμέριση μέσω της σχέσης,

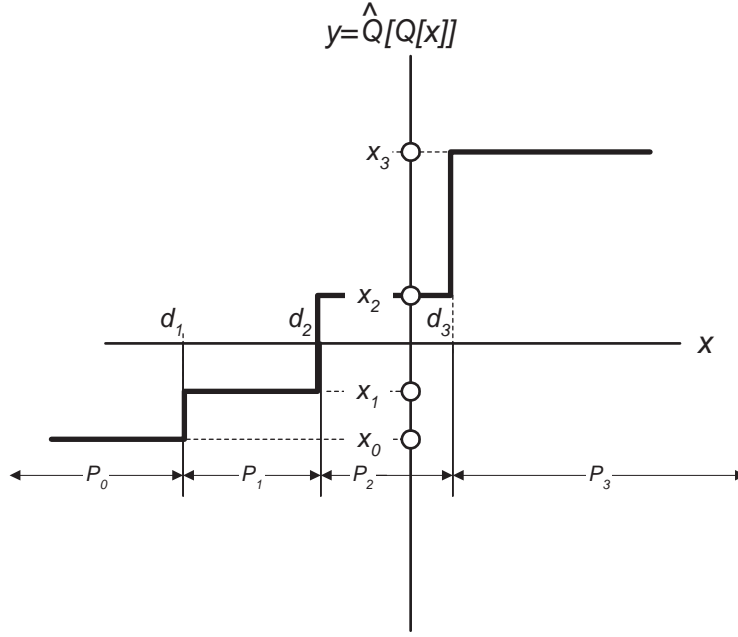
$$P_i = \begin{cases} (-\infty, d_1) & i = 0 \\ [d_i, d_{i+1}) & i = 1, \dots, N-2 \\ [d_{N-1}, +\infty) & i = N-1 \end{cases} \quad (2.20)$$

Είναι επίσης λογικό οι στάθμες κβαντισμού  $x_i$  να ανήκουν στα αντίστοιχα διαστήματα της διαμέρισης ( $x_i \in P_i$ ).

Με βάση τους παραπάνω ορισμούς και παραδοχές ο συνδιασμός των διαδικασιών κβαντισμού αποκβαντισμού παριστάνεται στο Σχήμα 2.7.

**Ορισμός 2.2.** Αν  $x(\mathbf{n}) \quad \mathbf{n} \in R^k$  είναι ένα  $k$ -διάστατο σήμα βαθμωτών τιμών, η ποσότητα  $\widehat{x}(\mathbf{n}) \triangleq \overline{Q}[x(\mathbf{n})] \in L$ , ονομάζεται κβαντισμένο σήμα.

Αξίζει να επισημάνουμε ότι τα κβαντισμένα σήματα είναι οντότητες της ίδιας φύσης με τα σήματα  $x(\mathbf{n})$  (μόνο που το πεδίο τιμών τους περιορίζεται στο πεπερασμένο υποσύνολο  $L$  των πραγματικών αριθμών), και να τα αντιδιαστέλλουμε προς τα σύμβολα  $s_i \in S$  που προκύπτουν στην έξοδο του κβαντιστή  $Q[ ]$ .



Σχήμα 2.7: Βαθμωτός Κβαντισμός/Αποκβαντισμός σε  $N = 4$  στάθμες

Ομοιόμορφος λέγεται ο κβαντιστής/αποκβαντιστής για τον οποίο τα διαστήματα της διαμέρισης  $P$  είναι ισομήκη, δηλαδή  $d_{i+1} - d_i = \Delta$  για κάθε  $i = 1, \dots, N-2$  και επίσης οι στάθμες κβαντισμού ισαπέχουν:  $x_{i+1} - x_i = \Delta$  για κάθε  $i = 0, \dots, N-2$

Κβαντιστής *νεκρής ζώνης* λέγεται αυτός για τον οποίο το διάστημα  $P_i$  που περιέχει το  $x = 0$  απεικονίζεται στη στάθμη  $x_i = 0$ .

Συμμετρικός λέγεται ο κβαντιστής στον οποίο τόσο οι στάθμες απόφασης όσο και οι στάθμες κβαντισμού είναι διατεταγμένες συμμετρικά ως προς το 0 (για άρτιο αριθμό σταθμών κβαντισμού  $N$  θα πρέπει  $d_{N/2} = 0$ ,  $d_i = -d_{N-i}$  και  $x_i = -x_{N-1-i}$  ενώ για περιττό αριθμό  $N$  θα πρέπει  $d_i = -d_{N-i}$ ,  $x_{(N-1)/2} = 0$  και  $x_i = -x_{N-1-i}$ ).

Ο ορισμός του (βαθμωτού) κβαντισμού εύκολα επεκτείνεται σε αυτόν του διανυσματικού κβαντισμού αν η διαμέριση  $P$  καλύπτει το  $n$ -διάστατο χώρο  $R^n$  αντί του  $R$  (δηλαδή  $P_i \subseteq R^n$ ). Βεβαίως σ' αυτή την περίπτωση οι στάθμες κβαντισμού  $x_i \in R^n$ . Αν και στην απλούστερη περίπτωση διανυσματικού κβαντισμού τα χωρία  $P_i$  είναι υπερκύβοι (με κέντρο τα  $x_i$ ) σε πληθώρα εφαρμογών χρησιμοποιούνται διαμερίσεις με μη-κανονικά χωρία ή ακόμη και διαμερίσεις με χωρία  $P_i$  που δεν είναι συνεκτικά.

### 2.2.2 Σφάλμα κβαντισμού

Όπως έχει ήδη αναφερθεί η διαδικασία αποκβαντισμού δεν αντιστρέφει επακριβώς τη διαδικασία κβαντισμού. Ως αποτέλεσμα εμφανίζεται το *σφάλμα κβαντισμού*,

$$\varepsilon \triangleq x - \widehat{Q}[Q[x]] \equiv x - \overline{Q}[x] . \quad (2.21)$$

Εξετάζοντας ειδικότερα τον *ομοιόμορφο κβαντισμό* εύκολα διαπιστώνουμε ότι το σφάλμα κβαντισμού δεν ξεπερνά, κατ' απόλυτη τιμή το μισό του βήματος κβαντισμού  $\Delta$ , και συγκεκριμένα,

$$-\Delta/2 < \varepsilon \leq \Delta/2 .$$

με εξαίρεση τις περιπτώσεις όπου το  $x$  έχει εξαιρετικά μικρές ( $x \in P_0$  με  $x < x_0 - \Delta/2$ ) ή εξαιρετικά μεγάλες τιμές ( $x \in P_{N-1}$  με  $x_{N-1} + \Delta/2 < x$ ).

Για διευκόλυνση της μαθηματικής ανάλυσης θα υποθέσουμε ότι το πεδίο τιμών του  $x$  καλύπτεται εξολοκλήρου από το διάστημα

$$[x_{min}, x_{max}] = [x_0 - \Delta/2, x_{N-1} + \Delta/2] .$$

Με αυτή την παραδοχή και υπό ορισμένες συνθήκες που συχνά ισχύουν σε πρακτικές εφαρμογές, το σφάλμα  $\varepsilon$  μπορεί να θεωρηθεί ως τυχαία μεταβλητή με *ομοιόμορφη κατανομή* στο διάστημα  $[-\Delta/2, \Delta/2]$ . Σε αυτή την περίπτωση η ισχύς του σφάλματος ισούται με τη διασπορά της ομοιόμορφης κατανομής δηλαδή,

$$\sigma_\varepsilon^2 = \frac{\Delta^2}{12}$$

Συνεπώς ο αντίστοιχος σηματοθορυβικός λόγος θα είναι

$$SNR = 10 \log \left( \frac{\sigma_x^2}{\sigma_\varepsilon^2} \right) = 10 \log \left( \frac{12\sigma_x^2}{\Delta^2} \right) \quad (2.22)$$

Οποιαδήποτε αύξηση (μείωση) του βήματος κβαντισμού  $\Delta$  έχει ως αποτέλεσμα την αύξηση (μείωση) της ισχύος του σφάλματος ή ισοδύναμα τη μείωση (αύξηση) του  $SNR$ . Αν θεωρήσουμε δεδομένο το πεδίο τιμών  $[x_{min}, x_{max}]$  του  $x$  γίνεται φανερό ότι

$$\Delta = \frac{x_{max} - x_{min}}{N} = \frac{x_{max} - x_{min}}{2^b} \quad (2.23)$$

όπου υποθέσαμε ότι ο αριθμός  $N$  των κβαντικών επιπέδων ισούται με το μέγιστο αριθμό συμβόλων  $s_i$  που μπορούν να αναπαρασταθούν με  $b$  bits. Αντικαθιστώντας την τελευταία έκφραση στη Σχέση 2.22 προκύπτει η ακόλουθη πολύ ενδιαφέρουσα σχέση μεταξύ  $SNR$  και διαθέσιμου αριθμού bits:

$$SNR = 10 \log (12\sigma_x^2) - 20 \log (x_{max} - x_{min}) + (20 \log (2))b \approx A_x + 6b \quad (\text{σε dB}) \quad (2.24)$$

όπου  $A_x = 10 \log (12\sigma_x^2) - 20 \log (x_{max} - x_{min})$  μία ποσότητα που εξαρτάται αποκλειστικά από τη φύση του σήματος  $x$ . Από την παραπάνω Σχέση 2.24 προκύπτει το συμπέρασμα ότι κάτω από ρεαλιστικές παραδοχές ο σηματοθορυβικός λόγος ενός ομοιόμορφου κβαντιστή βελτιώνεται κατά  $6db$  για κάθε επιπλέον bit που διατίθεται στην αναπαράσταση των συμβόλων  $s_i$  (Σημείωση: κάθε επιπλέον bit αντιστοιχεί σε διπλασιασμό του αριθμού των διαθέσιμων κβαντικών επιπέδων).

## Κεφάλαιο 3

# Αναπαράσταση και Αναπαραγωγή Πολυμέσων

### 3.1 Εισαγωγή: Τρία είδη αναπαράστασης στον κύκλο ζωής των πολυμέσων

Ανεξάρτητα από την προέλευσή τους τα πολυμέσα αποθηκεύονται, υφίστανται επεξεργασία και μεταδίδονται σε ψηφιακή μορφή. Τόσο για λόγους μεθοδολογίας όσο και για πρακτικούς λόγους πρέπει να διακρίνουμε τρία είδη αναπαράστασης των πολυμέσων με την επιπλέον παρατήρηση ότι για την ίδια οντότητα πολυμέσων χρησιμοποιούνται εναλλάξ τα τρία αυτά είδη κατά τη διάρκεια του κύκλου ζωής της:

- *Εσωτερική Αναπαράσταση*: Οι οντότητες πολυμέσων βρίσκονται σ' αυτόν τον τύπο αναπαράστασης αμέσως μετά τη δημιουργία τους (π.χ., δειγματοληψία, σύνθεση) και επανέρχονται σ' αυτή την αναπαράσταση πριν υποστούν οποιαδήποτε επεξεργασία. Η εσωτερική αναπαράσταση είναι ο φυσικός τρόπος αναπαράστασης των πολυμεσικών οντοτήτων. Για παράδειγμα, η εσωτερική αναπαράσταση του ψηφιακού στερεοφωνικού ήχου είναι δύο ακολουθίες από τις τιμές των δειγμάτων έντασης του κάθε καναλιού. Οι Εσωτερικές αναπαραστάσεις των πολυμέσων δεν συναντώνται "ελεύθερες στη φύση". Μπορούν να αναζητηθούν μόνο στις προσωρινές μνήμες των επεξεργαστών και βέβαια στις εξισώσεις βιβλίων και περιοδικών.
- *Αναπαράσταση Μεταφοράς και Αποθήκευσης*: Όταν μία οντότητα πολυμέσων πρόκειται να αποθηκευτεί ή να μεταφερθεί από ένα σύστημα πολυμέσων σε άλλο τότε χρησιμοποιείται ο εν λόγω τύπος αναπαράστασης. Μία πολυμεσική οντότητα επιδέχεται περισσότερες της μιας αναπαραστάσεις μεταφοράς και αποθήκευσης. Όλες, ωστόσο, οι εναλλακτικές αναπαραστάσεις αυτού του είδους αποτελούνται από δύο βασικές συνιστώσες: (α) την εσωτερική αναπαράστα-



ση της οντότητας ή κάποιο μετασχηματισμό της και (β) ένα σύνολο συνοδευτικών δεδομένων που επιτρέπουν την ανασύνθεση και κατανόηση της εσωτερικής αναπαράστασης από ένα σύστημα πολυμέσων που θα κληθεί να διαχειριστεί τη συγκεκριμένη οντότητα. Για παράδειγμα, η αναπαράσταση μεταφοράς και αποθήκευσης του ψηφιακού στερεοφωνικού ήχου μπορεί να αποτελείται από (α) την παράθεση των δειγμάτων (ανά ζεύγη) που προέρχονται εναλλάξ από το αριστερό και δεξί κανάλι και (β) μία επικεφαλίδα που - χρησιμοποιώντας κάποιο τυποποιημένο συντακτικό - περιγράφει την προαναφερθείσα διαδικασία πολύπλεξης, το ρυθμό δειγματοληψίας και τα χαρακτηριστικά του κβαντιστή που χρησιμοποιήθηκαν, τον παραγωγό, τον κάτοχο των πνευματικών δικαιωμάτων της συγκεκριμένης οντότητας κ.λπ. Οι αναπαραστάσεις μεταφοράς και αποθήκευσης πολυμέσων μπορεί να γίνουν ιδιαίτερα περίπλοκες προκειμένου (1) να υποστηρίξουν πρωτόκολλα συμπίεσης των πολυμέσων (βλ. αντίστοιχη ενότητα) και (2) να επιτρέψουν την κατάτμηση των δεδομένων σε μικρότερα πακέτα δεδομένων έτσι ώστε να γίνει εφικτή η σταδιακή τους μετάδοση μέσα από δίκτυα δεδομένων. Η μετάβαση από την εσωτερική αναπαράσταση στην αναπαράσταση μεταφοράς και αποθήκευσης και αντιστρόφως είναι πάντα εφικτή. Όμως οι εν λόγω μεταβάσεις είναι δυνατό να επιφέρουν μερική απώλεια / αλλοίωση των δεδομένων. Για παράδειγμα ο μετασχηματισμός της εσωτερικής αναπαράστασης προκειμένου να ενταχθεί στην αναπαράσταση μεταφοράς και αποθήκευσης μπορεί να εισάγει παραμορφώσεις (τυπική περίπτωση τέτοιων μετασχηματισμών είναι οι αλγόριθμοι συμπίεσης με απώλειες που περιγράφονται στο αντίστοιχο κεφάλαιο). Αντίστροφα κατά την ανασύνθεση της εσωτερικής αναπαράστασης από την αναπαράσταση μεταφοράς και αποθήκευσης είναι δυνατό να χαθούν πληροφορίες όπως αυτές του ρυθμού δειγματοληψίας ή του κατόχου των πνευματικών δικαιωμάτων.

- *Αναπαράσταση Αναπαραγωγής:* Μία πολυμεσική οντότητα μετατρέπεται στην αναπαράσταση αναπαραγωγής της όταν πρόκειται να *παρουσιαστεί στους χρήστες* των πολυμεσικών συστημάτων. Μία πολυμεσική οντότητα είναι δυνατό να επιδέχεται πολλαπλές διαφορετικές αναπαραστάσεις αναπαραγωγής που αντιστοιχούν στα μέσα και τις συνθήκες αναπαραγωγής. Για το παράδειγμα του στερεοφωνικού ήχου, που χρησιμοποιήσαμε και παραπάνω, δύο από τις εναλλακτικές αναπαραστάσεις αναπαραγωγής είναι οι ακόλουθες: (α) Από τις δύο ακολουθίες δειγμάτων που αποτελούν την εσωτερική αναπαράσταση της οντότητας παράγεται μία τρίτη που τα δείγματά της έχουν τιμή ίση προς το μέσο όρο των δειγμάτων αριστερού και δεξιού καναλιού. Η νέα αυτή ακολουθία είναι κατάλληλη για την αναπαραγωγή του ήχου από συστήματα ενός μεγαφώνου (μονοφωνικά). (β) Από τις δύο αρχικές ακολουθίες δειγμάτων παράγονται δύο άλλες μέσω παρεμβολής (ιντερπολατιον) ενός ή περισσοτέρων δειγμάτων μεταξύ διαδοχικών δειγμάτων των αρχικών ακολουθιών. Οι δύο νέες ακολουθίες (που προσομοιώνουν

ρυθμό δειγματοληψίας πολλαπλάσιο του αρχικού) είναι καταλληλότερες για την αναπαραγωγή υψηλής πιστότητας αναλογικού στερεοφωνικού ήχου.

Στις επόμενες ενότητες εξετάζονται λεπτομερέστερα οι χρησιμοποιούμενοι τύποι αναπαράστασης για τις σημαντικότερες μορφές πολυμεσικών οντοτήτων: τον ήχο, την (στατική) εικόνα και το βίντεο.

## 3.2 Ήχος

### 3.2.1 Εσωτερικές αναπαραστάσεις

#### Μονοκαναλικός ήχος

Η απλούστερη εκδοχή ψηφιακού ήχου είναι ο μονοφωνικός (μονοκαναλικός) ήχος, αυτός δηλαδή που αντιστοιχεί στην καταγραφή μίας και μόνο ηχητικής κυματομορφής (ανεξάρτητα του αν αυτή προέρχεται από μία πηγή ή είναι επαλληλία πολλών διαφορετικών συνιστωσών). Η εσωτερική αναπαράσταση του μονοκαναλικού ήχου είναι η ακολουθία των κβαντισμένων δειγμάτων

$$s(n) = \overline{Q}[s_d(n)], \quad n \in Z$$

της ηχητικής καταγραφής  $s_d(n)$ . Ανάλογα με την προέλευσή του το σήμα  $s_d(n)$  προέρχεται είτε από δειγματοληψία αντίστοιχης αναλογικής κυματομορφής ( $s_d(n) = s_c(Tn)$ ) είτε από επεξεργασία ή σύνθεση άλλων ψηφιακών κυματομορφών.

#### Πολυκαναλικός ήχος

Ο πολυκαναλικός ήχος αντιστοιχεί στην παράλληλη καταγραφή  $k \geq 2$  ηχητικών κυματομορφών και αναπαρίσταται ως μία ακολουθία διανυσμάτων (μήκους  $k$  το καθένα) με κβαντισμένες τιμές:

$$\mathbf{s}(n) \equiv \begin{bmatrix} s_1(n) \\ \vdots \\ s_k(n) \end{bmatrix} = \begin{bmatrix} \overline{Q}[s_{d1}(n)] \\ \vdots \\ \overline{Q}[s_{dk}(n)] \end{bmatrix}$$

Είναι ακόμη δυνατό, ως γενίκευση της παραπάνω σχέσης, οι κβαντιστές που χρησιμοποιούνται να είναι διαφορετικοί σε κάθε συνιστώσα (κανάλι). Η εκδοχή του πολυκαναλικού ήχου περιλαμβάνει:

- Τον στερεοφωνικό ήχο
- Τον ήχο surround
- Τον αναδυόμενο ήχο (immersive audio)
- Τον πολυγλωσσικό ήχο

- Την παράλληλη εγγραφή φωνών και μουσικής, κ.λπ.

Στον Πίνακα 3.1 παρουσιάζονται οι πιο τυπικές κατηγορίες ηχητικών σημάτων και οι αντίστοιχες παράμετροι εσωτερικής αναπαράστασης (ρυθμός δειγματοληψίας, αριθμός βιτς του χρησιμοποιούμενου κβαντιστή, αριθμός καναλιών).

Τύπος	Εύρος Αναλογικών Συχνοτήτων	Ρυθμός Δειγματοληψίας	bits/δείγμα	bitrate
Φωνή χαμηλής ποιότητας (π.χ., τηλέφωνο)	200-3200 Hz	8 KHz	16	128 Kbits/sec
Φωνή υψηλής ποιότητας	50-7000 Hz	16 KHz	16	256 Kbits/sec
Ήχος CD	20-20000 Hz	44.1 KHz	16 × 2 κανάλια	1.41 Mbits/sec

Πίνακας 3.1: Είδη

### 3.2.2 Αναπαραστάσεις μεταφοράς και αποθήκευσης ήχου

Οι αναπαραστάσεις μεταφοράς και αποθήκευσης του ήχου εξετάζονται στην ενότητα “Συμπίεση Ηχητικών Σημάτων”.

### 3.2.3 Αναπαραστάσεις αναπαραγωγής ήχου

Η αναπαραγωγή των ηχητικών σημάτων απαιτεί την μετατροπή τους σε αναλογικά σήματα που θα οδηγήσουν τον ενισχυτή (με τη μορφή ηλεκτρικών σημάτων) και στη συνέχεια τα ηχεία που θα μετατρέψουν το ηλεκτρικό σήμα σε ηχητικό. Η περιγραφή της διαδικασίας μετατροπής του ψηφιακού σήματος σε αναλογικό (D/A conversion) ξεφεύγει από τους στόχους του παρόντος κειμένου. Αυτό όμως που ενδιαφέρει να εξετάσουμε είναι η μορφή της αναπαράστασης αναπαραγωγής - του σήματος δηλαδή το οποίο παραδίδεται στο μετατροπέα D/A - και πώς αυτή προκύπτει από την εσωτερική αναπαράσταση του ήχου.

Στην πράξη εμφανίζονται περισσότερες της μιας διαφορετικές αναπαραστάσεις αναπαραγωγής και συνδυασμοί αυτών.

### Υπερδειγματοληψία ηχητικών εγγραφών

Η αναπαράσταση αναπαραγωγής αυτού του τύπου είναι μια υπερδειγματοληπτημένη έκδοση της αρχικής εσωτερικής αναπαράστασης, ήχος δηλαδή που προσομοιώνει δειγματοληψία με πολλαπλάσια συχνότητα δειγματοληψίας σε σχέση με την αρχική. Ποικίλες μέθοδοι μπορούν να χρησιμοποιηθούν για την παρεμβολή, μεταξύ δύο διαδοχικών δειγμάτων του αρχικού σήματος, ενός ή περισσότερων νέων δειγμάτων.

Ένας συνηθισμένος τρόπος είναι η τοποθέτηση π.χ., ενός μηδενικού (0) ανάμεσα στα διαδοχικά δείγματα του αρχικού σήματος και η χρήση στη συνέχεια ενός βαθυπερατού φίλτρου.

Η μέθοδος της υπερδειγματοληψίας αποσκοπεί στη μείωση του θορύβου που εισάγει η μετέπειτα διαδικασία μετατροπής του σήματος σε αναλογικό (D/A).

### Αναπαραγωγή Μονοφωνικού, Στερεοφωνικού, Τετραφωνικού Ήχου

Η αναπαράσταση αναπαραγωγής ταυτίζεται συνήθως με την πολυκαναλική αναπαράσταση και συγκεκριμένα κάθε κανάλι παραδίδεται σε μία συσκευή αναπαραγωγής (‘‘ήχείο’’). Σε περιπτώσεις που το σύστημα αναπαραγωγής διαθέτει λιγότερα ‘‘ήχεία’’ από το πλήθος των καναλιών (π.χ., αναπαραγωγή στερεοφωνικού ήχου από συσκευή ενός ήχου) τότε 2 ή περισσότερα κανάλια μπορούν να συμπτυχθούν σε ένα με δείγματα τιμής ίσης προς το μέσο όρο των επιμέρους καναλιών.

### Ήχος surround και Αναδυόμενος ήχος

(βλ. σχετικό άρθρο (Chris Kyriakakis 1998))

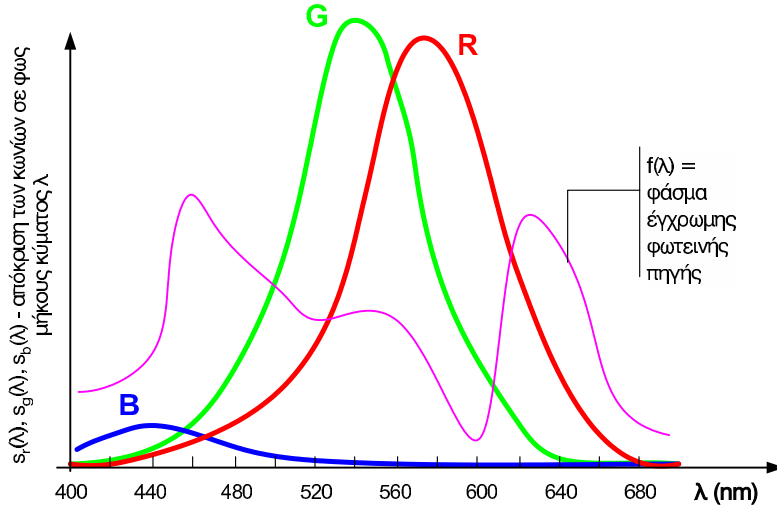
## 3.3 Εικόνα

Για να γίνει κατανοητό το περιεχόμενο της ψηφιακής αναπαράστασης εικόνων χρειάζεται να αντιληφθούμε με ποιο τρόπο τα συστήματα πολυμέσων διαχειρίζονται και αναπαριστούν το χρώμα. Στην επόμενη ενότητα δίνονται οι βασικές αρχές της καταγραφή, αναπαράστασης και αναπαραγωγής της πληροφορίας του χρώματος.

### 3.3.1 Αναπαράσταση του χρώματος

Η χρωματική πληροφορία μίας σημειακής φωτεινής πηγής αναπαρίσταται μέσω ενός διάνυσματος τριών χρωματικών συνιστωσών. Η αναπαράσταση αυτή γίνεται εφικτή χάρη στη φυσιολογία του ανθρώπινου αισθητηρίου της όρασης. Η αίσθηση του χρώματος προκύπτει από το συνδυασμό του ερεθίσματος που δέχονται τρεις διαφορετικοί τύποι *κονίων* (κύτταρα διάσπαρτα στον αμφιβληστροειδή του ματιού). Κάθε τύπος κωνίου έχει διαφορετική συχνοτική απόκριση όταν διεγείρεται από το ορατό φως (Με μήκη κύματος  $\lambda \in [\lambda_{min}, \lambda_{max}]$  όπου  $\lambda_{min} \approx 360nm$  και  $\lambda_{max} \approx 830nm$ ). Αν μία φωτεινή πηγή έχει φασματικό περιεχόμενο  $f(\lambda)$  τότε το ερέθισμα που φτάνει στον εγκέφαλο ισοδυναμεί με το διάνυσμα,

$$\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}, \quad \text{όπου } c_i = \int_{\lambda_{min}}^{\lambda_{max}} s_i(\lambda) f(\lambda) d\lambda, \quad i = 1, 2, 3. \quad (3.1)$$



Σχήμα 3.1: Αποκρίσεις συχνότητας ( $s_i(\lambda)$ ,  $i = 1, 2, 3$ ) στην περιοχή του ορατού φάσματος. Οι συναρτήσεις απόκρισης μεγιστοποιούνται κατά σειρά στις περιοχές του κόκκινου (R), πράσινου (G) και μπλε (B).

Οι συναρτήσεις  $s_i(\lambda)$  μεγιστοποιούνται στις περιοχές του κόκκινου (R), πράσινου (G) και μπλε (B) όπως φαίνεται από τις αντίστοιχες καμπύλες του Σχήματος 3.1.

Αν δύο πηγές με διαφορετικά φασματικά περιεχόμενα  $f_1(\lambda)$  και  $f_2(\lambda)$  αντίστοιχα τυχαίνει να παράγουν το ίδιο διάνυσμα  $\mathbf{c}$  τότε εμφανίζονται να έχουν το ίδιο χρώμα. Το διάνυσμα  $\mathbf{c}$  ονομάζεται *προσλαμβανόμενο χρώμα*.

Για απλούστευση τα ολοκληρώματα στη Σχέση 3.1 θα αντικατασταθούν στη συνέχεια από αθροίσματα της μορφής,

$$c_i \approx \sum_{k=1 \dots N} s_i(\lambda_k) f(\lambda_k)$$

οπότε η Σχέση 3.1 μπορεί να γραφεί

$$\mathbf{c} = \mathbf{S}^T \mathbf{f} \quad (3.2)$$

όπου

$$\mathbf{S} = \begin{bmatrix} s_1(\lambda_1) & s_2(\lambda_1) & s_3(\lambda_1) \\ s_1(\lambda_2) & s_2(\lambda_2) & s_3(\lambda_2) \\ \vdots & \vdots & \vdots \\ s_1(\lambda_N) & s_2(\lambda_N) & s_3(\lambda_N) \end{bmatrix}$$

και

$$\mathbf{f} = \begin{bmatrix} f(\lambda_1) \\ f(\lambda_2) \\ \vdots \\ f(\lambda_N) \end{bmatrix}$$

Αντίστοιχος με τη λειτουργία του ματιού είναι ο μηχανισμός καταγραφής μίας έγχρωμης CCD κάμερας. Όταν η ίδια φωτεινή πηγή καταγράφεται από την κάμερα παράγεται ένα αντίστοιχο διάνυσμα μετρήσεων

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad \text{όπου } a_i = \int_{\lambda_{min}}^{\lambda_{max}} m_i(\lambda) f(\lambda) d\lambda, \quad i = 1, 2, 3. \quad (3.3)$$

όπου  $m_i(\lambda)$  είναι οι αποκρίσεις συχνότητας (φίλτρα) τριών φωτοευαίσθητων αισθητήρων που αντιστοιχούν σε κάθε θέση (pixel) του CCD. Το διάνυσμα  $\mathbf{a}$  ονομάζεται *καταγραφόμενο χρώμα*. Τα τρία φίλτρα που χρησιμοποιούνται έχουν χρωματική επιλεκτικότητα και παρουσιάζουν, συνήθως, μέγιστο στις περιοχές του κόκκινου (R), πράσινου (G) και μπλε (B) κατά μίμηση των συναρτήσεων  $s_i(\lambda)$  του ανθρώπινου ματιού. Από τη φύση του λοιπόν το καταγραφόμενο χρώμα είναι ένα διάνυσμα *RGB*.

Διακριτοποιώντας τη Σχέση 3.3 όπως και προηγουμένως έχουμε:

$$\mathbf{a} = \mathbf{M}^T \mathbf{f} \quad (3.4)$$

όπου ο πίνακας  $\mathbf{M}$  ορίζεται αντίστοιχα προς τον  $\mathbf{S}$ .

Το ζητούμενο είναι πώς με βάση το διάνυσμα  $\mathbf{a}$  μπορεί να αναπαραχθεί φωτεινή ακτινοβολία που να δίνει το ίδιο υποκειμενικό αποτέλεσμα με το φως της αρχικής πηγής, δηλαδή  $\mathbf{c}$ . Η εν λόγω αναπαραγωγή κατά περίπτωση θα γίνει από την οθόνη μιας τηλεόρασης, το μόνιτορ ενός υπολογιστή, την έγχρωμη εκτύπωση σε λευκό χαρτί, κ.λπ. Στη συνέχεια θα εξετάσουμε την απλή, αλλά διαφωτιστική, εκδοχή ενός τριχρωματικού προβολέα (προθεστωρ).

Το φως που παράγεται (για ένα σημείο του πετάσματος) από τον προβολέα είναι η σύνθεση των τριών φωτεινών δεσμών του προβολέα με φάσμα  $p_i(f)$ ,  $i = 1, 2, 3$ . Το φάσμα της κάθε δέσμης είναι συγκεντρωμένο σε διαφορετικές περιοχές φωτεινών συχνοτήτων κατ' αναλογία με τις συναρτήσεις  $s_i(f)$  του ματιού. Ο προβολέας ρυθμίζει το κέρδος κάθε δέσμης με βάση τις συνιστώσες του διανύσματος  $\mathbf{a}$  που έχει παραχθεί από την κάμερα. Συγκεκριμένα, το παραγόμενο φως έχει φάσμα:

$$g(f) = a_1^c p_1(f) + a_2^c p_2(f) + a_3^c p_3(f) \quad (3.5)$$

όπου

$$\mathbf{a}^c \equiv \begin{bmatrix} a_1^c \\ a_2^c \\ a_3^c \end{bmatrix} = \mathbf{T}^c \mathbf{a}, \quad (3.6)$$

όπου ο  $3 \times 3$  πίνακας  $\mathbf{T}^c$  επιτρέπει τη λεγόμενη *διόρθωση χρώματος* του προβολέα (θα δούμε στη συνέχεια το ρόλο του πίνακα αυτού). Αν διακριτοποιήσουμε τη συχνότητα  $f$  όπως και προηγουμένως

η παραπάνω σχέση γράφεται,

$$\mathbf{g} = a_1^c \mathbf{p}_1 + a_2^c \mathbf{p}_2 + a_3^c \mathbf{p}_3 \quad (3.7)$$

$$= [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3] \mathbf{a}^c \quad (3.8)$$

$$= [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3] \mathbf{T}^c \mathbf{a} \quad (3.9)$$

$$= \mathbf{P} \mathbf{a} \quad (3.10)$$

όπου ο  $N \times 3$  πίνακας  $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3] \mathbf{T}^c$  εξαρτάται από το φασματικό περιεχόμενο των τριών δεσμών του προβολέα και τον πίνακα διόρθωσης χρώματος.

Όταν το ανθρώπινο μάτι δεχθεί το φωτεινό ερέθισμα από το συγκεκριμένο σημείο του πετάσματος, σύμφωνα με τη Σχέση 3.2 θα αντιληφθεί το προσλαμβανόμενο χρώμα

$$\hat{\mathbf{c}} = \mathbf{S}^T \mathbf{g} = \mathbf{S}^T \mathbf{P} \mathbf{a} = \mathbf{S}^T \mathbf{P} \mathbf{M}^T \mathbf{f} \quad (3.11)$$

όπου διαδοχικά χρησιμοποιήσαμε τις Σχέσεις 3.7 και 3.4. Για να έχει το προβαλλόμενο σημείο το ίδιο προσλαμβανόμενο χρώμα με αυτό της αρχικής πηγής θα πρέπει  $\hat{\mathbf{c}} = \mathbf{c}$  δηλαδή

$$\mathbf{S}^T \mathbf{f} = \mathbf{S}^T \mathbf{P} \mathbf{M}^T \mathbf{f} \quad (3.12)$$

Μάλιστα η Σχέση 3.12 πρέπει να ισχύει ανεξαρτήτως του χρώματος της αρχικής πηγής  $\mathbf{f}$ . Συνεπώς για να εξασφαλιστεί ότι η διαδικασία καταγραφήσ-αναπαραγωγής δεν αλλοιώνει το χρωματικό περιεχόμενο θα πρέπει:

$$\mathbf{S}^T = \mathbf{S}^T \mathbf{P} \mathbf{M}^T \quad (3.13)$$

όπου ο πίνακας  $\mathbf{S}$  εξαρτάται από την ευαισθησία του ανθρώπινου ματιού στις συχνότητες του ορατού φωτεινού φάσματος, ο πίνακας  $\mathbf{M}$  εξαρτάται αποκλειστικά από την αντίστοιχη ευαισθησία της κάμερας και ο πίνακας  $\mathbf{P}$  καθορίζεται από το φασματικό περιεχόμενο των τριών δεσμών του προβολέα. Η ικανοποίηση της Σχέσης 3.13 είναι εφικτή υπό τις ακόλουθες παραδοχές:

- Αν και ο πίνακας  $\mathbf{S}$  διαφοροποιείται από άνθρωπο σε άνθρωπο, μπορούμε να υποθέσουμε ότι υπάρχει ένας αντιπροσωπευτικός πίνακας που περίπου περιγράφει το μέσο θεατή.
- Για κάθε συνδυασμό κάμερας / προβολέα κάνουμε μία από τις ακόλουθες δύο ρυθμίσεις:

1. Για δεδομένο μοντέλο κάμερας (δηλαδή συγκεκριμένο πίνακα  $\mathbf{M}$  επιλέγεται πειραματικά<sup>1</sup> ο κατάλληλος πίνακας  $\mathbf{P}$  του προβολέα έτσι ώστε να ικανοποιείται η Σχέση 3.13. Τούτο είναι εφικτό με κατάλληλη επιλογή του πίνακα διόρθωσης χρώματος  $\mathbf{T}^c$ .

---

<sup>1</sup> Η πειραματική διαδικασία στηρίζεται στην υποκειμενική κρίση αντιπροσωπευτικού αριθμού ανθρώπων που, για ένα πλήθος μονοχρωματικών πηγών, αναλαμβάνουν να συγκρίνουν το φως της κάθε αρχικής πηγής με αυτό που αναπαράγει ο προβολέας.

2. Επιλέγονται κατάλληλα τα τρία φίλτρα  $m_i(\lambda)$  της κάμερας, ισοδύναμα ο πίνακας  $\mathbf{M}$  έτσι ώστε για δεδομένους πίνακες  $\mathbf{S}$  και  $\mathbf{P}$  να ισχύει η Σχέση 3.13. Και εδώ η επιλογή γίνεται με πειραματικό τρόπο.

Συνοψίζοντας, το προσλαμβανόμενο χρώμα  $\mathbf{c} = \mathbf{S}^T \mathbf{f}$  ενός φωτεινού σημείου αναπαρίσταται από το καταγραφόμενο χρώμα  $\mathbf{a} = \mathbf{M}^T \mathbf{f}$  και από αυτό αναπαράγεται το προσλαμβανόμενο χρώμα  $\hat{\mathbf{c}} = \mathbf{S}^T \mathbf{P} \mathbf{a}$  λαμβάνοντας μέριμνα ώστε  $\hat{\mathbf{c}} \approx \mathbf{c}$ .

### 3.3.2 Συστήματα Χρωματικών Συντεταγμένων

Όπως έχει ήδη αναφερθεί, το καταγραφόμενο χρώμα  $\mathbf{a}$  αναπαριστά κατά προσέγγιση τη συμμετοχή του κόκκινου, πράσινου και μπλε μέρους του ορατού φάσματος στη σύνθεση ενός χρώματος. Για το λόγο αυτό το συγκεκριμένο σύστημα χρωματικών συνιστωσών είναι το γνωστό ως *RGB* (Red-Green-Blue).

Άλλα χρωματικά συστήματα όπως τα *YIQ*, *YUV*, *YCrCb*, *CMY* και *CMYK* είναι επίσης πολύ διαδεδομένα (το τελευταίο χρησιμοποιεί 4 χρωματικές συνιστώσες). Στα συστήματα αυτά, το χρώμα αναπαρίσταται μέσω μίας τριάδας (ή και τετράδας) χρωματικών συντεταγμένων που έχει αμφιμονοσήμαντη σχέση με το διάνυσμα  $\mathbf{a}$  της αναπαράστασης *RGB*. Ο Πίνακας 3.2 περιέχει τους μετασχηματισμούς του *RGB* διανύσματος  $\mathbf{a}$  στις παραπάνω εναλλακτικές αναπαραστάσεις.

### 3.3.3 Δεικτοδότηση Χρώματος

Δεικτοδότηση χρώματος, είναι η εφαρμογή διανυσματικού κβαντισμού στο διανυσματικό χώρο των χρωμάτων  $\mathbf{a} \in R^3$ .

Πειραματικά αποτελέσματα έχουν καταδείξει ότι το ανθρώπινο μάτι μπορεί να διακρίνει μόνο ένα μικρό αριθμό από διαφορετικές χρωματικές αποχρώσεις όταν αυτές συνυπάρχουν σε μία εικόνα. Με βάση αυτή την παρατήρηση η ιδέα του κβαντισμού των χρωμάτων σε μία εικόνα εμφανίστηκε ελκυστική κυρίως για λόγους συμπίεσης του όγκου των αναπαραστάσεων.

Τα σύμβολα του διανυσματικού κβαντιστή που χρησιμοποιείται είναι κατά σύμβαση οι αριθμοί από 1 έως  $N$  όπου  $N$  είναι το πλήθος των κβαντικών σταθμών. Ο αποκβαντισμός γίνεται με εξαιρετικά απλό τρόπο μέσω ενός χρωματικού πίνακα (*color map*)  $\mathbf{A}$  διαστάσεων  $N \times 3$  ο οποίος έχει στην  $i$ -γραμμή του το χρώμα  $\mathbf{a}$  που αντιστοιχεί στην  $i$  κβαντική στάθμη (π.χ. το pixel με δεικτοδοτημένο χρώμα '5' θα έχει πραγματικό χρώμα  $\mathbf{a} = \mathbf{A}(5, 1 : 3)^T$ ).

Τυπικές τιμές του  $N$  είναι  $2^1$ ,  $2^4$ ,  $2^8$  και  $2^{16}$  με πιο συνηθισμένη την  $N = 2^8 = 256$ .



Πίνακας 3.2: Συνηθέστερα συστήματα χρωματικών συντεταγμένων

Σύστημα	Μετασχηματισμός	Χρήση και Σχόλια
YUV	$\mathbf{a}_{YUV} \equiv [Y, U, V]^T = \mathbf{T}\mathbf{a}$ όπου $\mathbf{T} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.147 & 0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix}$	Χρησιμοποιείται από το ευρωπαϊκό πρότυπο τηλεόρασης PAL. Η συνιστώσα $Y$ του διανύσματος αντιπροσωπεύει τη <i>φωτεινότητα</i>
YIQ	$\mathbf{a}_{YIQ} \equiv [Y, I, Q]^T = \mathbf{T}\mathbf{a}$ όπου $\mathbf{T} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix}$	Χρησιμοποιείται από το αμερικανικό πρότυπο τηλεόρασης NTSC
YCrCb	$\mathbf{a}_{YCrCb} \equiv [Y, Cr, Cb]^T = \mathbf{T}\mathbf{a}$ όπου $\mathbf{T} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.5000 & -0.4187 & -0.0813 \\ -0.1687 & -0.3313 & 0.5000 \end{bmatrix}$	Χρησιμοποιείται από τα πρότυπα κωδικοποίησης εικόνων και βίντεο JPEG και MPEG αντίστοιχα
CMY	$\mathbf{a}_{CMY} = [1 \ 1 \ 1]^T - \mathbf{a}$	Το σύστημα (Cyan, Magenta, Yellow) είναι συμπληρωματικό του συστήματος RGB
CMYK	$\mathbf{a}_{CMYK} = [\dot{C}, \dot{M}, \dot{Y}, K]^T$ όπου με αναφορά στις συντεταγμένες του συστήματος CMY, $K = \min(C, M, Y)$ , $\dot{C} = C - K$ , $\dot{M} = M - K$ , $\dot{Y} = Y - K$	Το σύστημα (Cyan, Magenta, Yellow, Black) χρησιμοποιείται σε εφαρμογές έγχρωμης εκτύπωσης σε λευκό χαρτί
HSV και LAB	ορίζονται μέσω μη-γραμμικών μετασχηματισμών του RGB.	Τα HSV και LAB είναι <i>ομοιόμορφα</i> ως προς το ανθρώπινο αισθητήριο. Όταν το μάτι αντιλαμβάνεται πως, μεταξύ δύο χρωμάτων, το ένα πλησιάζει περισσότερο σε ένα χρώμα αναφοράς, η ίδια σχέση ισχύει και για τις Ευκλείδειες αποστάσεις των χρωματικών τους συντεταγμένων.

### 3.3.4 Εσωτερικές αναπαραστάσεις εικόνων

Διακρίνουμε δύο κατηγορίες εικόνων όσον αφορά τη χρωματική τους σύνθεση: τις ψηφιακές εικόνες αποχρώσεων του γκρι και τις έγχρωμες.

Στην πρώτη περίπτωση οι εικόνες αναπαρίστανται ως διδιάστατοι πίνακες κβαντισμένων βαθμωτών τιμών,

$$f(\mathbf{n}) = \overline{Q}[Y(\mathbf{n})], \quad \mathbf{n} = \begin{bmatrix} n_h \\ n_v \end{bmatrix} \in Z^2$$

όπου  $Y(\mathbf{n})$  αντιστοιχεί στη φωτεινότητα (luminance) του δείγματος  $\mathbf{n}$ . Συνεπώς οι εικόνες αποχρώσεων του γκρι με  $N_r$  γραμμές και  $N_c$  στήλες αναπαρίστανται εσωτερικά με πίνακες της μορφής

$$\mathbf{f} = \left[ f \left( \begin{bmatrix} n_h \\ n_v \end{bmatrix} \right) \right]_{\substack{n_h=0 \dots N_c-1 \\ n_v=0 \dots N_r-1}} \quad (3.14)$$

Όσον αφορά στις έγχρωμες εικόνες, αυτές αναπαρίστανται στη συνηθέστερη περίπτωση από τρεις (3) διαφορετικούς πίνακες που αντιστοιχούν στις τρεις χρωματικές συντεταγμένες των διανυσμάτων καταγραφόμενου χρώματος  $\mathbf{a}$  ή κάποιου μετασχηματισμού του (RGB, YUV, YIQ, YCrCb, CMY, HSV, LAB). Ειδικά στην περίπτωση της αναπαράστασης CMYK χρησιμοποιούνται τέσσερις πίνακες. Οι πίνακες αυτοί που από κοινού περιλαμβάνουν την πληροφορία του χρώματος, αποκαλούνται *χρωματικές συνιστώσες* της εικόνας.

Στην απλούστερη περίπτωση οι πίνακες των χρωματικών συνιστωσών μίας εικόνας έχουν τις ίδιες διαστάσεις  $N_r \times N_c$ . Είναι όμως δυνατό να χρησιμοποιηθεί διαφορετικός ρυθμός δειγματοληψίας για κάθε χρωματική συνιστώσα οπότε και οι διαστάσεις των πινάκων θα διαφοροποιούνται αντίστοιχα.

### 3.3.5 Αναπαραστάσεις μεταφοράς και αποθήκευσης εικόνων

Οι συνηθέστερες αναπαραστάσεις μεταφοράς και αποθήκευσης εικόνων εμφανίζονται στον Πίνακα 3.3. Σε όλα τα είδη η αναπαράσταση αποτελείται από μία επικεφαλίδα ακολουθούμενη από την κυρίως αναπαράσταση της εικόνας. Η τελευταία είναι είτε αυτούσια η αντίστοιχη εσωτερική αναπαράσταση είτε ένας μετασχηματισμός της.

Όσον αφορά στην επικεφαλίδα αυτή κατά περίπτωση περιέχει τις εξής ενότητες:

1. Είδος Αναπαράστασης (Format) και Έκδοση της Αναπαράστασης: Η πληροφορία αυτή είναι αναγκαία στο σύστημα που διαβάζει το αρχείο της εικόνας για να ερμηνεύσει με τον ορθό τρόπο το υπόλοιπο μέρος της επικεφαλίδας και την κυρίως αναπαράσταση.
2. Διαστάσεις της Εικόνας
3. Μέθοδος πολύπλεξης των δεδομένων (π.χ. RR...RRGG...GGBB...BBB ή RGBRGBRGB...)

Είδος	Ονομασία		Συμπίεση
TIFF	Tagged Image File Format	Interchange	Κατ' επιλογή (χωρίς απώλειες)
PNM	Portable Map	Any	Όχι
PGM	Portable Map	Gray	Όχι
TGA	Truevision image file	Targa	Όχι
PCX	Paintbrush image file		Όχι
GIF	Graphics Interchange Format		Ναι
BMP	Microsoft Windows bitmap image file		Ναι (χωρίς απώλειες)
JPEG	Joint Photographic Experts Group format		Δύο εκδόσεις: με και χωρίς συμπίεση

Πίνακας 3.3: Αναπαραστάσεις μεταφοράς και αποθήκευσης εικόνων

4. Παράμετροι κβαντισμού: Κατά περίπτωση περιγράφεται το είδος του κβαντιστή, οι στάθμες κβαντισμού, ο αριθμός των bits ανά σύμβολο, κ.λπ.
5. Χρωματικός Πίνακας: Είναι απαραίτητος στην περίπτωση αναπαραστάσεων (όπως η GIF) που χρησιμοποιούν δεικτοδότηση χρώματος (color indexing).
6. Παράμετροι συμπίεσης: Εφόσον η συγκεκριμένη αναπαράσταση χρησιμοποιεί κάποιον αλγόριθμο συμπίεσης, η επικεφαλίδα περιέχει τις παραμέτρους χρήσης του αλγορίθμου (π.χ. τον πίνακα αντιστοίχισης κωδικών λέξεων της κωδικοποίησης Huffman σε σύμβολα). Το είδος των παραμέτρων εξαρτάται από τη συγκεκριμένη μέθοδο συμπίεσης (βλ. Ενότητα: Συμπίεση Ψηφιακών Εικόνων).

Σε όλα τα είδη αναπαράστασης η πληροφορία που περιλαμβάνεται στην επικεφαλίδα, οι συμβάσεις που συνάγονται εμμέσως από την υιοθέτηση ενός συγκεκριμένου τύπου αναπαράστασης και τα κυρίως δεδομένα που ακολουθούν την επικεφαλίδα επιτρέπουν την ανασύνθεση της εσωτερικής αναπαράστασης.

### 3.3.6 Αναπαραστάσεις αναπαραγωγής εικόνων

Οι συσκευές αναπαραγωγής εικόνων περιλαμβάνουν τους προβολείς (projectors), τις οθόνες (υπολογιστή, τηλεόρασης, μικροσυσκευής), και τα διάφορα είδη εκτυπωτών. Οι συσκευές αυτές παραλαμβάνουν την εικόνα στην κατάλληλη ψηφιακή αναπαράσταση αναπαραγωγής από το σύστημα πολυμέσων στο οποίο είναι προσαρτημένες. Το κρίσιμο χαρακτηριστικό των αναπαραστάσεων αναπαραγωγής είναι ο τρόπος αναπαράστασης του χρώματος και των αποχρώσεων του γκρι.

#### Προβολείς

Η αναπαράσταση αναπαραγωγής για τους προβολείς δημιουργείται από την εσωτερική αναπαράσταση των εικόνων με το μηχανισμό που ήδη περιγράφηκε στην Ενότητα 3.3.2. Συγκεκριμένα η αναπαράσταση περιλαμβάνει τρεις  $N_r \times N_c$  πίνακες, έναν για κάθε χρωματική συνιστώσα, με στοιχεία τις συνιστώσες του διανύσματος  $\mathbf{T}^c \mathbf{a}$ .

#### Οθόνες

Η ίδια αναπαράσταση, με διαφορετικούς πίνακες  $\mathbf{T}^c$ , χρησιμοποιείται και για τις σύγχρονες έγχρωμες οθόνες.

## Εκτυπωτές

Σε αντίθεση με τους δύο παραπάνω τύπους συσκευών αναπαραγωγής, οι εκτυπωτές έχουν έναν τελείως διαφορετικό μηχανισμό αναπαραγωγής του χρώματος. Οι δύο βασικές διαφορές του εν λόγω μηχανισμού συνίστανται στο ότι ένας εκτυπωτής:

- Δημιουργεί τα χρώματα *αφαιρετικά*: Τούτο σημαίνει ότι αν σε ένα σημείο δεν τοποθετηθεί καθόλου μελάνι το σημείο αυτό θα είναι λευκό, θα αντιστοιχεί δηλαδή στο μέγιστο της έντασης κάθε χρωματικής συνιστώσας. Για να βαφτεί ένα σημείο κόκκινο πρέπει να χρησιμοποιηθεί μελάνι που θα απορροφά (αφαιρεί) όλες τις άλλες ακτινοβολίες πλην της κόκκινης.
- Έχει *πεπερασμένη παλέτα* τριών ή τεσσάρων μελανιών (ενός για τους ασπρόμαυρους εκτυπωτές) σε αντίθεση με τους προβολείς και τις οθόνες που μπορούν να παράγουν χρώματα με αυθαίρετη ένταση.

Με βάση τα παραπάνω η αναπαράσταση αναπαραγωγής για έναν εκτυπωτή αποτελείται από τόσους πίνακες όσα και τα μελάνια του. Οι πίνακες αυτοί είναι δυαδικοί και οι τιμές τους καθορίζουν το κατά πόσο ένα pixel εκτύπωσης θα μελανωθεί (τιμή 1) ή όχι (τιμή 0). Τα πιξελ εκτύπωσης δεν ταυτίζονται με τα pixel (δείγματα) της προς εκτύπωση εικόνας. Συνήθως αναλογεί ένας μεγάλος αριθμός pixel εκτύπωσης για κάθε δείγμα της εικόνας έτσι ώστε συνδυάζοντας πολλά τέτοια pixel να δίνεται η αίσθηση των διαβαθμίσεων του χρώματος (ή του γκρι). Η μέθοδος προσδιορισμού των τιμών (0 ή 1) που θα λάβουν τα pixel εκτύπωσης προκειμένου να αποδώσουν την αίσθηση του χρώματος ονομάζεται *διττηρινγ* και αποτελεί πολύ ενδιαφέρον κεφάλαιο στην θεωρία της επεξεργασίας εικόνων.

## 3.4 Βίντεο

### 3.4.1 Εσωτερικές Αναπαραστάσεις Βίντεο - Αναπαράσταση της κίνησης

Η ιδέα της αναπαράστασης *κινούμενης εικόνας* είναι τουλάχιστον τόσο παλιά όσο ο παλιός κινηματογράφος. Η παρουσίαση διαδοχικών ακίνητων εικόνων που αναπαριστούν τη χρονική εξέλιξη μίας σκηνής δίνει στον άνθρωπο την αίσθηση της κίνησης. Αρκεί δηλαδή μία σχετικά πυκνή δειγματοληψία στιγμιότυπων για να αποκτήσουμε την *ψευδαίσθηση* της ομαλής κίνησης. Πόσο πυκνή όμως; Η απάντηση βασίζεται στο θεώρημα δειγματοληψίας και στην πειραματική μελέτη της απόκρισης του ανθρώπινου αισθητηρίου σε χρονικές μεταβολές του οπτικού πεδίου. Άρα η προηγούμενη ερώτηση ανάγεται στην ακόλουθη: ποιο το εύρος ζώνης της ανθρώπινης όρασης που σχετίζεται με τις χρονικές (όχι χωρικές) μεταβολές; Συνολικά η απάντηση είναι ότι αντιλαμβανόμαστε ομαλή κίνηση αρκεί να έχουμε μερικές δεκάδες εναλλαγών ανά δευτερόλεπτο. Το σινεμά αρκείται σ' αυτή τη θεώρηση,

έτσι ο κινηματογραφικός φακός ανοιγοκλείνει 29 φορές το δευτερόλεπτο και κάθε φορά προβάλλει το αντίστοιχο πλαίσιο ή *καρέ* (*frame*) .

Η τηλεόραση (στην αναλογική της ακόμα εποχή) επιχειρεί μία κάπως πιο περίπλοκη θεώρηση διερευνώντας το ερώτημα: Υπάρχει σχέση μεταξύ χρονικής και χωρικής δειγματοληψίας όσον αφορά το ανθρώπινο αισθητήριο; Η απάντηση είναι καταφατική. Το ανθρώπινο αισθητήριο λειτουργεί ως βαθυπερατό χωρο-χρονικό φίλτρο και το αντίστοιχο εύρος ζώνης επιτρέπει μια πιο φειδωλή δειγματοληψία όχι χωριστά στο χρόνο και χωριστά στις δύο χωρικές διαστάσεις αλλά από κοινού στις τρεις διαστάσεις. Τηλεοπτικής προέλευσης είναι η ιδέα του πεδίου (*field*) που περιέχει μόνο τις άρτιες ή μόνο τις περιττές γραμμές ενός τηλεοπτικού πλαισίου. Η αίσθηση της κίνησης δημιουργείται με τη διαδοχική εναλλαγή πεδίων με άρτιες και περιττές γραμμές εναλλάξ.

Στην περίπτωση του ψηφιακού βίντεο οι δύο παραπάνω μέθοδοι αναπαράστασης της κίνησης αντιστοιχούν στο *non-interlaced* και *interlaced* βίντεο, αντίστοιχα, που περιγράψαμε στην Ενότητα 2.1.1.

Στην παραπάνω θεώρηση ο όρος δείγμα αναφέρεται στη συνολική χρωματική πληροφορία του κινούμενου οπτικού περιεχομένου. Συνεπώς κάθε δείγμα είναι ένα διάνυσμα αποτελούμενο από τις χρωματικές συνιστώσες (RGB ή YUV ή YIQ ή YCrCb) της συγκεκριμένης χωρο-χρονικής θέσης.

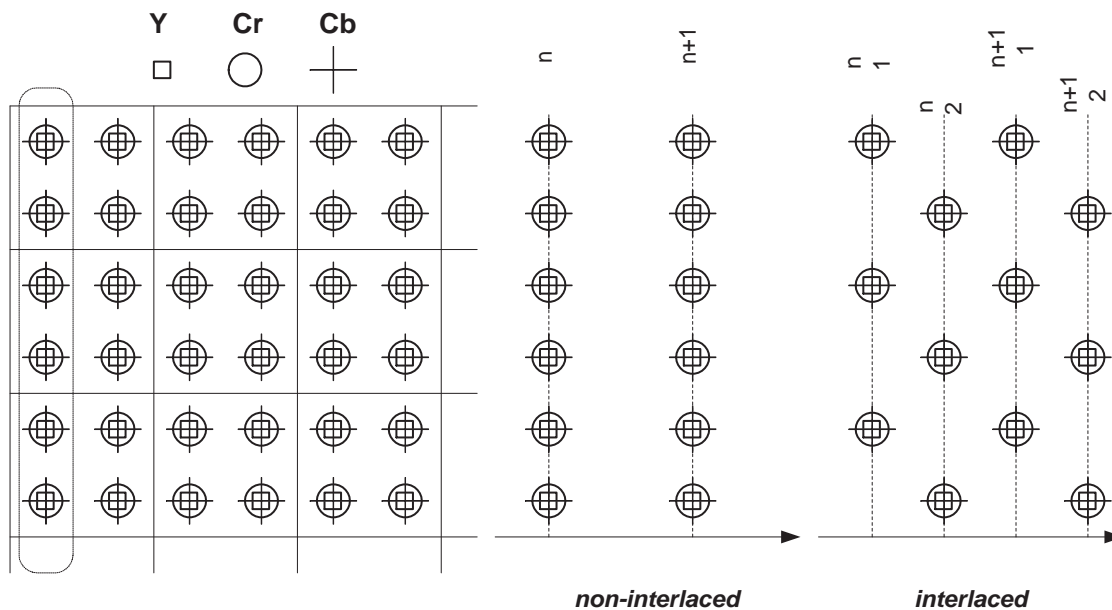
Μια επιπλέον σειρά πειραματικών αποτελεσμάτων μαρτυρούν ότι: Όχι μόνο ο μηχανισμός της όρασης λειτουργεί ως βαθυπερατό χωρο-χρονικό φίλτρο αλλά η ζώνη διέλευσής του είναι μικρότερη για τα σήματα που περιέχουν πληροφορία χρωματικής αναλογίας (συνιστώσες Cr-Cb ή U-V ή I-Q) απ• ότι για τα σήματα φωτεινότητας (Y). Με βάση αυτή την παρατήρηση προκύπτουν δύο συμπεράσματα:

- Είναι επιτρεπτή η δειγματοληψία των συνιστωσών χρωματικής αναλογίας με πίνακες πλέγματος μικρότερης πυκνότητας.
- Οι αναπαραστάσεις χρώματος YCrCb, YUV και YIQ είναι πλεονεκτικότερες στις εφαρμογές ψηφιακού βίντεο απ• ότι η RGB. Στη συνέχεια θα αναφερόμαστε μόνο στην αναπαράσταση χρώματος YCrCb δεδομένου ότι αυτή χρησιμοποιείται στις περισσότερες περιπτώσεις από τις εφαρμογές και τα διεθνή πρότυπα ψηφιακού βίντεο (MPEG και ITU).

Με βάση τις ιδέες της χωρο-χρονικής δειγματοληψίας και της (υπο-)δειγματοληψία των χρωματικών συνιστωσών (Cr και Cb)έχουν υιοθετηθεί τα ακόλουθα ήδη εσωτερικών αναπαραστάσεων του ψηφιακού βίντεο.

#### **Αναπαράσταση 4:4:4**

Στην αναπαράσταση αυτή και οι τρεις χρωματικές συνιστώσες (Y, Cr, Cb) δειγματοληπτούνται με τον ίδιο πίνακα πλέγματος. Το Σχήμα 3.2 (αριστερά) απεικονίζει τις σχετικές θέσεις των δειγμάτων



Σχήμα 3.2: Χωροχρονική δειγματοληψία 4:4:4

των τριών παραπάνω συνιστωσών σε ένα πλαίσιο. Το Σχήμα 3.2 (μεσαίο και δεξιά) αναπαριστά τις χωρο-χρονικές θέσεις των δειγμάτων της πρώτης στήλης του πλαισίου στην περίπτωση του non-interlaced (μεσαίο) και interlaced (δεξιά) βίντεο.

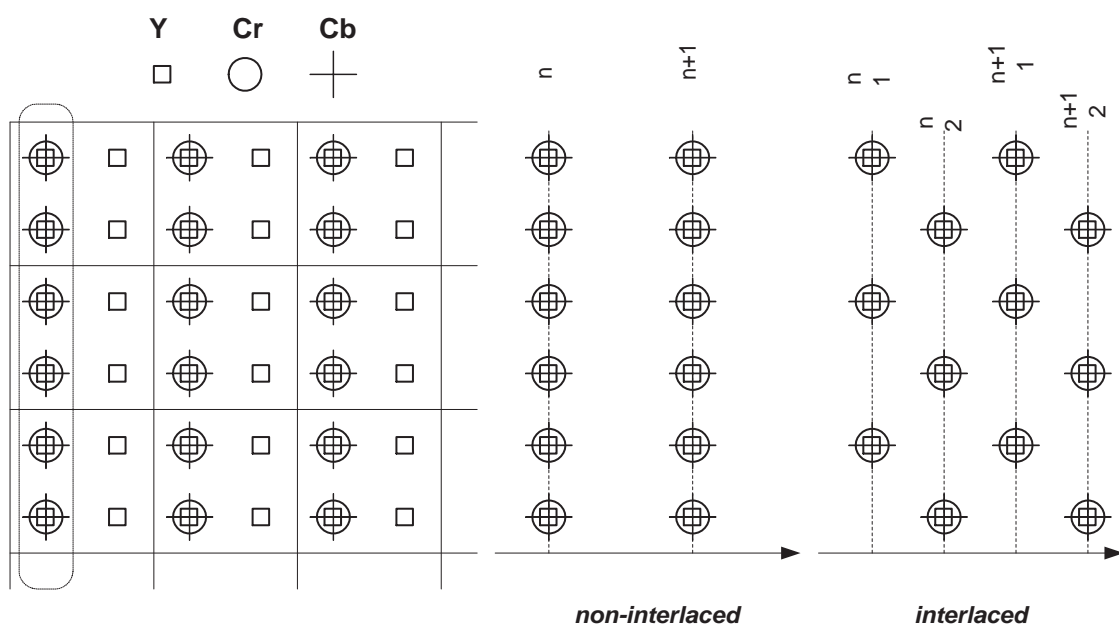
#### Αναπαράσταση 4:2:2

Στην αναπαράσταση αυτή, σε κάθε πλαίσιο, σε 4 δείγματα φωτεινότητας (Y) αντιστοιχούν 2 δείγματα Cr και 2 δείγματα Cb όπως φαίνεται στο Σχήμα 3.3(αριστερά). Όπως και στην προηγούμενη περίπτωση, το Σχήμα 3.3(μεσαίο και δεξιά) αναπαριστά τις χωρο-χρονικές θέσεις των δειγμάτων της πρώτης στήλης του πλαισίου στην περίπτωση του non-interlaced (μεσαίο) και interlaced (δεξιά) βίντεο.

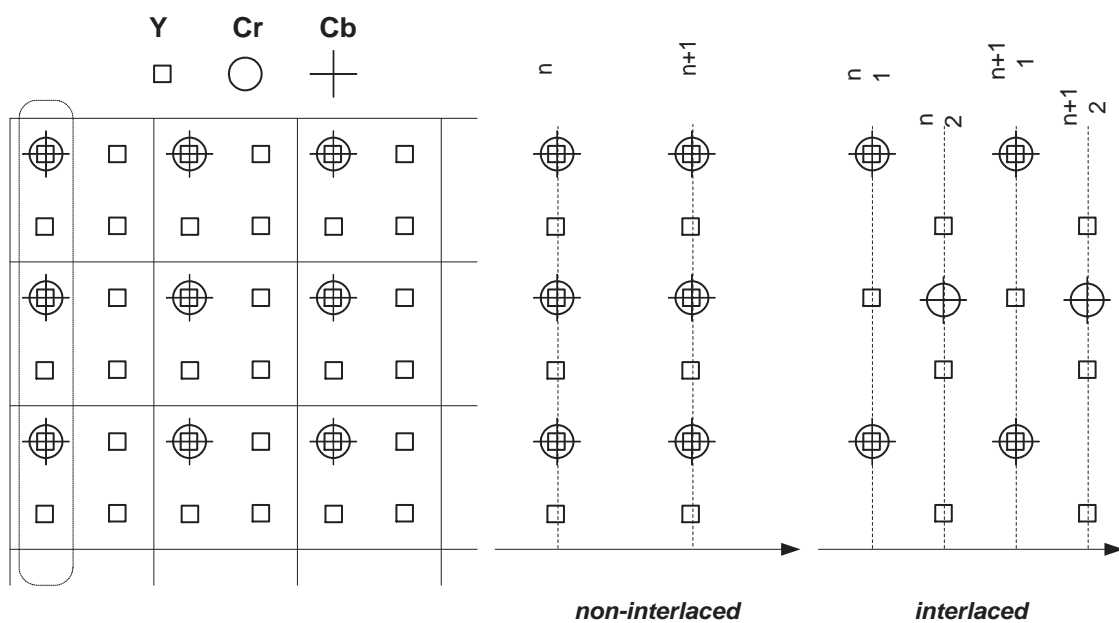
#### Αναπαράσταση 4:2:0

Στην αναπαράσταση αυτή, σε κάθε πλαίσιο, σε 4 δείγματα φωτεινότητας (Y) αντιστοιχούν 1 δείγμα Cr και 1 δείγμα Cb όπως φαίνεται στο Σχήμα 3.4(αριστερά). Αυτό θα δικαιολογούσε την ονομασία 4:1:1 αντί του χρησιμοποιούμενου 4:2:0 το οποίο έχει επικρατήσει. Όπως και στις προηγούμενες περιπτώσεις, το Σχήμα 3.4(μεσαίο και δεξιά) αναπαριστά τις χωρο-χρονικές θέσεις των δειγμάτων της πρώτης στήλης του πλαισίου στην περίπτωση του non-interlaced (μεσαίο) και interlaced (δεξιά) βίντεο.

Όσον αφορά τις διαστάσεις των πλαισίων, αυτές ποικίλουν ανά κατηγορία εφαρμογής.



Σχήμα 3.3: Χωροχρονική δειγματοληψία 4:2:2



Σχήμα 3.4: Χωροχρονική δειγματοληψία 4:2:0



Μορφή	Διαστάσεις	Πλαίσια/sec	Interlaced	Αναπαράσταση
CIF	$288 \times 352$		OXI	4:2:0
QCIF	$144 \times 176$		OXI	4:2:0
SQCIF	$128 \times 96$		OXI	4:2:0
SIF-625	$288 \times 352$	25	OXI	4:2:0
SIF-525	$240 \times 352$	30	OXI	4:2:0
PAL	$576 \times 720$	25	NAI	4:2:2
NTSC	$486 \times 720$	29.97	NAI	4:2:2
HDTV	$720 \times 1280$	59.94	OXI	4:2:0
HDTV	$1080 \times 1920$	29.97	NAI	4:2:0

Πίνακας 3.4: Τα χαρακτηριστικά των συνηθέστερων προτυποποιημένων μορφών βίντεο.

Ο Πίνακας 3.4 παρουσιάζει τις συνηθέστερες προτυποποιημένες μορφές (formats) ψηφιακού βίντεο, τις διαστάσεις πλαισίου, τον αντίστοιχο ρυθμό χρονικής δειγματοληψίας, την ένδειξη για τον αν είναι ντερλασεδ ή όχι και τη χρησιμοποιούμενη αναπαράσταση (4:4:4, 4:2:2 ή 4:2:0).

Τέλος όσον αφορά τον κβαντιστή που χρησιμοποιείται, η συνηθέστερη πολιτική είναι να χρησιμοποιούνται 8 *bits* (256 κβαντικά επίπεδα) για κάθε στοιχείο Y, Cr ή Cb.

### 3.4.2 Αναπαραστάσεις μεταφοράς και αποθήκευσης βίντεο

Οι αναπαραστάσεις αυτές εξετάζονται στις Ενότητες που αναφέρονται στη συμπίεση βίντεο.

### 3.4.3 Αναπαραστάσεις αναπαραγωγής βίντεο

Ισχύει ό,τι και στις ακίνητες εικόνες μόνο που η προβαλλόμενη εικόνα ανανεώνεται ως εξής:

- Στην περίπτωση του non-interlaced βίντεο ανανεώνεται ολόκληρη η εικόνα
- Στην περίπτωση του interlaced βίντεο ανανεώνονται εναλλάξ οι άρτιες και οι περιττές γραμμές.
- Στην περίπτωση των εσωτερικών αναπαραστάσεων 4:2:2 και 4:2:0 χρησιμοποιούνται αλγόριθμοι χωρο-χρονικής παρεμβολής ώστε να υπολογιστούν οι τιμές των δειγμάτων Cr και Cb που λείπουν.

## Κεφάλαιο 4

# Συμπίεση / Κωδικοποίηση

Με τον όρο *συμπίεση* εννοούμε το μετασχηματισμό της εσωτερικής αναπαράστασης των πολυμέσων σε μία διαφορετική μορφή - συμπιεσμένη αναπαράσταση - έτσι ώστε:

- Η συμπιεσμένη αναπαράσταση να χρειάζεται μικρότερο πλήθος bits για την περιγραφή της. Συνεπώς μικρότερο χώρο αποθήκευσης ή/και χρόνο μετάδοσης μέσω δικτύων υπολογιστών.
- Να είναι δυνατή η ανάκτηση μιας εσωτερικής αναπαράστασης από τη συμπιεσμένη αναπαράσταση όπου:
  1. είτε η ανακτηθείσα αναπαράσταση ταυτίζεται με την αρχική εσωτερική αναπαράσταση
  2. είτε η ανακτηθείσα αναπαράσταση μοιάζει αρκετά με την αρχική. Κριτήριο ομοιότητας είναι οι απαιτήσεις της εφαρμογής στην οποία η εν λόγω αναπαράσταση θα χρησιμοποιηθεί.

Στα δύο παραπάνω είδη συμπίεσης αναφερόμαστε με τους όρους 'χωρίς απώλεια' (*lossless*) και 'με απώλεια' (*lossy*) αντίστοιχα.

### 4.1 Η ανάγκη συμπίεσης

Η συμπίεση είναι αναγκαία για την εξοικονόμηση (α) αποθηκευτικού χώρου και (β) εύρους ζώνης κατά τη μεταφορά πολυμεσικών δεδομένων. Πλήθος εφαρμογών πολυμέσων όπως:

- η ψηφιακή τηλεόραση
- η κινητή τηλεφωνία
- η μετάδοση εικόνων και βίντεο μέσω του Internet
- τα DVDs

- η ψηφιακή σύνταξη εφημερίδων και περιοδικών
- η αποστολή fax

και πολλές άλλες θα ήταν πρακτικώς αδύνατες ή οικονομικά ασύμφορες χωρίς τη χρήση αλγορίθμων συμπίεσης. Δύο ενδεικτικά παραδείγματα είναι πειστικά:

#### 1. Αποστολή σελίδας A4 μέσω fax:

Ας υποθέσουμε ότι πρόκειται να στείλουμε μία σελίδα A4 ψηφιοποιημένη με δειγματοληψία 100 dpi (dots per inch) τόσο οριζόντια όσο και κατακόρυφα. Κατά τα συνήθη ο κβαντισμός που χρησιμοποιούμε αναθέτει 1 bit ανά δείγμα ( $= 2^1 = 2$  επίπεδα του γκρι, δηλαδή το σύννηθες άσπρο/μαύρο fax). Η σελίδα A4 έχει διαστάσεις  $8.27in \times 11.69$ . Άρα θα απαιτηθούν  $(8.27 in \times 100 samples/in) \times (11.69 in \times 100 samples/in) \times 1 bit/sample \approx 9.67 \times 10^5 bits$ . Αν ο όγκος αυτός δεδομένων μεταδοθεί μέσω τηλεφωνικής γραμμής με βανδωίδη 9600 bits/sec θα χρειαστούν περίπου  $100 sec \approx 1.5 min$ . Χρησιμοποιώντας το ίδιο βανδωίδη μία απλή συσκευή φαξ στέλνει στον ίδιο χρόνο 6 – 7 σελίδες χάρη στη συμπίεση που υλοποιεί.

#### 2. Ψηφιακή Τηλεόραση:

Το ψηφιακό σήμα βίντεο τύπου PAL (βλ. Πίνακα 3.4) έχει διαστάσεις  $576 \times 720$  και αναπαράσταση 4:2:2. Άρα για την αναπαράσταση ενός πλαισίου απαιτούνται:  $(576 \times 720)$  δείγματα φωτεινότητας (Y) συν  $\frac{1}{2} \times (576 \times 720)$  δείγματα Cr και άλλα τόσα δείγματα Cb. Επομένως απαιτούνται  $(1 + \frac{1}{2} + \frac{1}{2})(576 \times 720) = 829440$  δείγματα ανά πλαίσιο. Επειδή όπως έχει αναφερθεί ο συνηθέστερος κβαντιστής των χρωματικών συνιστωσών χρησιμοποιεί 8 bits/sample, η αναπαράσταση του πλαισίου απαιτεί 6635520 bits και με δεδομένο ότι στο PAL μεταδίδονται 25 πλαίσια/sec, το απαιτούμενο εύρος ζώνης είναι  $6635520 \times 25 bits/sec \approx 158 Mbits/sec$ . Αξίζει να παρατηρήσει κανείς ότι το εύρος ζώνης μιας ψηφιακής δορυφορικής ζεύξης (όπως αυτές που χρησιμοποιούν τα συνδρομητικά δορυφορικά κανάλια) είναι  $\approx 32 Mbits/sec$  και χρησιμοποιούνται για να προσφέρουν στους συνδρομητές μπουκέτα με 10 ως και 20 κανάλια.

## 4.2 Οι βασικές αρχές και μέθοδοι συμπίεσης

Ο στόχος της μείωσης του απαιτούμενου πλήθους bits για την αναπαράσταση μίας πολυμεσικής οντότητας επιτυγχάνεται με το συνδυασμό των παρακάτω τεχνικών:

- Ελεγχόμενη παραμόρφωση
- Μείωση της περίσσειας (redundancy) πληροφορίας

- Κωδικοποίηση εντροπίας

Οι μέθοδοι συμπίεσης ‘χωρίς απώλειες’ αξιοποιούν αποκλειστικά τις δύο τελευταίες τεχνικές. Αντίθετα οι μέθοδοι συμπίεσης ‘με απώλειες’ συνδυάζουν και τις τρεις.

#### 4.2.1 Κωδικοποίηση και Εντροπία

Οποιαδήποτε αναπαράσταση μεταφοράς και αποθήκευσης πολυμεσικών οντοτήτων αποτελείται από μία ακολουθία συμβόλων τα οποία ανήκουν σε ένα αλφάβητο  $W$  (π.χ., βλ. τα σύμβολα στην έξοδο ενός κβαντιστή  $Q[\ ]$  στην Ενότητα 2.2.1). Κάθε σύμβολο παριστάνεται με ένα πλήθος από bits (0 ή 1).

**Ορισμός 4.1.** Ορίζουμε ως κωδική λέξη του συμβόλου  $s_i$ ,  $i = 1 \cdots N$  τη (μοναδική) ακολουθία  $W_i$  πεπερασμένου πλήθους bits που το περιγράφει κατά την ψηφιακή του αναπαράσταση.

Το μήκος  $|W_i|$  των κωδικών λέξεων μπορεί να είναι είτε κοινό για όλα τα σύμβολα - οπότε το σύνολο των κωδικών λέξεων,  $W$ , λέγεται κώδικας σταθερού μήκους - ή να διαφοροποιείται ανάλογα με το σύμβολο - οπότε ο αντίστοιχος κώδικας,  $W$ , λέγεται μεταβλητού μήκους.

**Ορισμός 4.2.** Έστω  $N$  σύμβολα  $s_i$ ,  $i = 1 \cdots N$  και ένας μηχανισμός  $\mathbf{S}$  (πηγή) που παράγει τα εν λόγω σύμβολα με τυχαίο τρόπο αλλά με σταθερή πιθανότητα  $p_i = \text{Prob}(s_i)$   $i = 1 \cdots N$ . Ορίζουμε ως εντροπία της πηγής  $\mathbf{S}$  την ποσότητα,

$$H \triangleq - \sum_{i=1}^N p_i \log_2 \{p_i\} \quad (4.1)$$

Από τη Θεωρία Πληροφορίας δανειζόμαστε την ακόλουθη πολύ σημαντική πρόταση:

**Πρόταση 4.1.** Έστω πηγή  $\mathbf{S}$  που παράγει τα σύμβολα  $s_i$ ,  $i = 1 \cdots N$  με πιθανότητες  $p_i = \text{Prob}(s_i)$   $i = 1 \cdots N$  αντίστοιχα. Αν  $W$  οποιοσδήποτε κώδικας που αναπαριστά τα σύμβολα  $s_i$  με τις κωδικές λέξεις  $W_i$  και  $L_w \triangleq \sum_{i=1}^N p_i |W_i|$  το μέσο μήκος των κωδικών λέξεων, τότε

$$H \leq L_w \quad (4.2)$$

Το νόημα της Πρότασης 4.1 είναι ότι η εντροπία  $H$  αποτελεί το κάτω φράγμα για τον οποιοδήποτε ‘καλό’ κώδικα. Οι κώδικες που επιτυγχάνουν ή προσεγγίζουν αυτό το κάτω φράγμα ονομάζονται κώδικες εντροπίας. Ο γνωστότερος κώδικας εντροπίας είναι ο κώδικας *Huffman*.

Ο κώδικας *Huffman* είναι κώδικας μεταβλητού μήκους και αναθέτει στα σύμβολα κωδικές λέξεις μήκους αντιστρόφως αυξανόμενου με την πιθανότητα εμφάνισής τους.

Ο κώδικας *Huffman* για τα σύμβολα  $s_i$   $i = 1 \cdots N$  με πιθανότητες  $p_i$   $i = 1 \cdots N$ , δημιουργείται με τη βοήθεια ενός δυαδικού δένδρου που κατασκευάζεται ως εξής:

1. Δημιουργούμε ένα σύνολο  $N$  που περιέχει ανά πάσα στιγμή ζεύγη  $(s, p)$  όπου  $s$  ένα ενδιάμεσο σύμβολο και  $p$  η πιθανότητά του.
2. Στην αρχή το σύνολο  $N$  περιέχει τα ζεύγη  $(s_i, p_i)$   $i = 1 \cdots N$ .
3. Το δέντρο έχει ως ακραίους κόμβους (φύλλα) τα σύμβολα  $s_i$ .
4. Σε κάθε βήμα
  - Επιλέγονται δύο στοιχεία του συνόλου  $N$  που έχουν τις μικρότερες πιθανότητες.
  - Τα στοιχεία αυτά διαγράφονται από το  $N$  και αντικαθίστανται από ένα νέο ενδιάμεσο σύμβολο με πιθανότητα ίση προς το άθροισμα των πιθανοτήτων των δύο διαγραφέντων.
  - Στο δέντρο προστίθεται ένας νέος κόμβος, που αντιστοιχεί στο νέο ενδιάμεσο σύμβολο.
  - Προστίθενται δύο σύνδεσμοι μεταξύ του νέου κόμβου και αυτών που αντιστοιχούν στα σύμβολα που διαγράφηκαν.
  - Οι δύο νέοι σύνδεσμοι σηματοδοτούνται με τα ψηφία 0 και 1 (δεν έχει σημασία η σειρά).
5. Η δημιουργία του δένδρου ολοκληρώνεται όταν στο σύνολο  $N$  μείνει ένα μόνο στοιχείο με πιθανότητα 1.
6. Ο κώδικας Huffman του κάθε συμβόλου  $s_i$  προκύπτει από τα ψηφία 0 και 1 που συναντώνται ακολουθώντας το μονοπάτι από τη ρίζα του δένδρου μέχρι το αντίστοιχο φύλλο.

#### 4.2.2 Αποσυσχέτιση

Η εντροπία μίας ακολουθίας κβαντισμένων δειγμάτων ταυτίζεται με την εντροπία πηγής που συναποτελείται (α) από το μηχανισμό γένεσης των δειγμάτων,  $\mathbf{x}(\mathbf{n})$ , (π.χ., διαδικασία δειγματοληψίας) ακολουθούμενη (β) από τον αντίστοιχο κβαντιστή  $N = 2^b$  κβαντικών επιπέδων,  $s(\mathbf{n}) = Q[\mathbf{x}(\mathbf{n})] \in S$ . Αποδεικνύεται ότι αν πριν από τον κβαντιστή εφαρμοστεί ένας (αντιστρεπτός) μετασχηματισμός,  $F[\ ]$  των δειγμάτων τέτοιος ώστε τα νέα δείγματα,  $\mathbf{y}(\mathbf{n}) = F[\mathbf{x}(\mathbf{n})]$ , είναι μεταξύ τους στατιστικώς (πιο) ανεξάρτητα και στη συνέχεια χρησιμοποιήσουμε κβαντιστή  $N$  επιπέδων, η πηγή που προκύπτει έχει μικρότερη εντροπία από την αρχική.

Κατά συνέπεια σύμφωνα με την Πρόταση 4.1 η νέα ακολουθία δειγμάτων,  $\mathbf{y}(\mathbf{n})$ , επιδέχεται μεγαλύτερη συμπίεση. Αυτή η παρατήρηση είναι κομβική για το σύνολο σχεδόν των μεθόδων συμπίεσης. Στην πράξη, το σύνολο των αλγορίθμων συμπίεσης εμπεριέχει ένα στάδιο στο οποίο επιχειρείται η στατιστική αποσυσχέτιση των δειγμάτων τους κωδικοποιούμενου σήματος<sup>1</sup>

---

<sup>1</sup>Οι έννοιες της στατιστικής ανεξαρτησίας και της στατιστικής αποσυσχέτισης δεν ταυτίζονται παρά μόνο στην περίπτωση που τα δείγματα ακολουθούν την κανονική κατανομή. Ωστόσο, υπό προϋποθέσεις που συνήθως είναι αληθείς, η μείωση της στατιστικής συσχέτισης συνεπάγεται μείωση της στατιστικής εξάρτησης.

Τυπικά παραδείγματα εφαρμογής της διαδικασίας αποσυσχέτισης, όπως αυτή χρησιμοποιείται στους αλγορίθμους συμπίεσης είναι:

- Οι αλγόριθμοι Differential Pulse Code Modulation (DPCM) και Adaptive DPCM (ADPCM) στην συμπίεση ψηφιακού ήχου.
- Η χρήση κωδικοποίησης μετασχηματισμού - συνήθως Διακριτού Μετασχηματισμού Συνημιτόνου (Discrete Cosine Transform - DCT) στην κωδικοποίηση ήχου, εικόνας και βίντεο.
- Οι αλγόριθμοι εκτίμησης και αντιστάθμισης κίνησης στην κωδικοποίηση βίντεο.

#### 4.2.3 DPCM και ADPCM

Η αρχή της μεθόδου DPCM βρίσκεται στην παρατήρηση ότι δοθέντος μονοδιάστατου σήματος  $x(n)$  το σήμα

$$d(n) = x(n) - x(n-1), \quad (4.3)$$

παρουσιάζει συνήθως μικρότερη συσχέτιση από το αρχικό. Πράγματι, υποθέτοντας χωρίς βλάβη της γενικότητας ότι  $E\{x(n)\} = 0$ , η συνάρτηση αυτοσυσχέτισης  $r_d(m)$  του σήματος διαφορών  $d(n)$  υπολογίζεται ως εξής:

$$\begin{aligned} r_d(m) &= E\{d(n)d(n+m)\} \\ &= E\{(x(n) - x(n-1))(x(n+m) - x(n+m-1))\} \\ &= E\{x(n)x(n+m)\} + E\{x(n-1)x(n+m-1)\} \\ &\quad - E\{x(n)x(n+m-1)\} - E\{x(n-1)x(n+m)\} \\ &= 2r_x(m) - r_x(m-1) - r_x(m+1) \\ &\approx 0, \end{aligned} \quad (4.4)$$

όπου στην τελευταία γραμμή της 4.4 υποθέσαμε ότι ο συντελεστής  $r_d(m)$  της συνάρτησης αυτοσυσχέτισης του σήματος  $x(n)$  είναι πολύ κοντά στο μέσο όρο των συντελεστών  $r_x(m-1)$  και  $r_x(m+1)$ . Όπως φαίνεται και από την 4.4 υπό ορισμένες προϋποθέσεις (αλλά όχι και πάντα) η συσχέτιση μεταξύ των δειγμάτων του  $d(n)$  είναι μικρή ακόμη κι αν η αντίστοιχη συσχέτιση για το  $x(n)$  είναι σημαντική. Άρα είναι προτιμότερη, συνήθως, η συμπίεση/κωδικοποίηση του  $d(n)$ .

Η αντίστροφη διαδικασία δίνεται από τη Σχέση

$$x(n) = d(n) + x(n-1), \quad (4.5)$$

όπου η αναπαγωγή του  $n$ -οστού δείγματος του αρχικού σήματος γίνεται με βάση το αντίστοιχο δείγμα της ακολουθίας διαφορών και του προηγούμενου (ήδη αναπαχθέντος) δείγματος  $x(n-1)$ .

Στην πράξη η διαδικασία είναι ελαφρώς πιο περίπλοκη λόγω της παρεμβολής του κβαντιστή/αποκβαντιστή. Έτσι αντί του ζεύγους εξισώσεων 4.3, 4.5 χρησιμοποιούνται οι ακόλουθες:

$$d(n) = x(n) - \hat{x}(n-1), \quad (4.6)$$

όπου

$$\begin{aligned} \hat{x}(n) &= \hat{d}(n) + \hat{x}(n-1) \\ \hat{d}(n) &= \bar{Q}[d(n)]. \end{aligned} \quad (4.7)$$

Η διαδικασία DPCM είναι κατά βάση μία διαδικασία πρόβλεψης ενός βήματος του δείγματος  $x(n)$  από το αμέσως προηγούμενό του. Η διαδικασία αυτή γενικεύεται (και βελτιώνεται) αν στην πρόβλεψη συμμετέχουν περισσότερα δείγματα από το παρελθόν και μάλιστα με κατάλληλα βάρη. Στην περίπτωση αυτή αντί των Σχέσεων 4.3 και 4.5 έχουμε τις

$$\begin{aligned} d(n) &= x(n) - \mathbf{a}^T \mathbf{x}(n-1) \\ x(n) &= d(n) + \mathbf{a}^T \mathbf{x}(n-1) \end{aligned} \quad (4.8)$$

όπου το διάνυσμα  $\mathbf{x}(n-1) \triangleq [x(n-1) \ x(n-2) \ \cdots \ x(n-p)]^T$  περιέχει τα  $p$  παρελθόντα δείγματα και  $\mathbf{a} = [a_1 \ a_2 \ \cdots \ a_p]^T$  είναι το αντίστοιχο διάνυσμα βαρών - ή αλλιώς συντελεστών πρόβλεψης.

Και στην περίπτωση αυτή οι σχέσεις που χρησιμοποιούνται στην πράξη τροποποιούνται λόγω της ύπαρξης του κβαντιστή αναλογία με τις 4.7.

Ενδιαφέρον παρουσιάζει τέλος ο υπολογισμός των βαρών και του μήκους,  $p$ , του διανύσματος  $\mathbf{a}$  με βάση τα στατιστικά χαρακτηριστικά του σήματος  $x(n)$ . Συγκεκριμένα, πληθώρα τεχνικών επιτρέπει την επιλογή του καταλληλότερου διανύσματος συντελεστών πρόβλεψης με κριτήριο την όσο το δυνατό καλύτερη αποσυσχέτιση των δειγμάτων του  $d(n)$ . Για τις τεχνικές αυτής της μορφής χρησιμοποιείται ο όρος Adaptive DPCM (ADPCM).

#### 4.2.4 Κωδικοποίηση Μετασχηματισμού

Η μέθοδοι συμπίεσης με χρήση Μετασχηματισμού επιχειρούν την αποσυσχέτιση των δειγμάτων χρησιμοποιώντας επεξεργασίας κατά block (block processing).

##### Μονοδιάστατοι Μετασχηματισμοί

Το αρχικό σήμα  $x(n)$ ,  $n \in Z$  χωρίζεται σε μη επικαλυπτόμενα διαδοχικά blocks δειγμάτων  $\mathbf{x}(k) = [x(kN) \ x(kN-1) \ \cdots \ x((k-1)N+1)]^T$  μήκους  $N$  το καθένα και μετασχηματίζονται με τη χρήση του αντιστρεπτού πίνακα  $\mathbf{F}$  ( $\det \mathbf{F} \neq 0$ ),

$$\mathbf{y}(k) = \mathbf{F}\mathbf{x}(k). \quad (4.9)$$

Τα αρχικά δείγματα, κατά την αποκωδικοποίηση επανυπολογίζονται ως

$$\mathbf{x}(k) = \mathbf{F}^{-1}\mathbf{y}(k). \quad (4.10)$$

χάρη στην αντιστρεψιμότητα του πίνακα  $\mathbf{F}$ . Στόχοι του μετασχηματισμού είναι

1. Τα δείγματα των διανυσμάτων  $\mathbf{y}(k)$  να είναι πιο ασυσχέτιστα από αυτά των  $\mathbf{x}(k)$
2. Εντός κάθε διανύσματος  $\mathbf{y}(k)$  η θέση κάθε δείγματος να προσδιορίζει τη σπουδαιότητά του από πλευράς πληροφορίας που εμπεριέχει και μάλιστα αυτή η σπουδαιότητα να διαφοροποιείται μεταξύ των δειγμάτων. Τούτο θα δώσει τη δυνατότητα για χρήση διαφορετικών κβαντιστών για θέσεις με διαφορετική σπουδαιότητα.

### Διδιάστατοι Μετασχηματισμοί

Όπως και στην περίπτωση των μονοδιάστατων μετασχηματισμών το αρχικό σήμα  $x(\mathbf{n})$ ,  $\mathbf{n} = [n_1 \ n_2]^T \in Z^2$  χωρίζεται σε μη επικαλυπτόμενα διδιάστατα blocks δειγμάτων

$$x(\mathbf{k}) = [x(\mathbf{n})]_{\substack{n_1=k_1N \dots (k_1-1)N+1 \\ n_2=k_2N \dots (k_2-1)N+1}}, \quad \mathbf{k} = [k_1 \ k_2]^T \in Z^2 \quad (4.11)$$

και το κάθε block μετασχηματίζεται με τη χρήση του αντιστρεπτού πίνακα  $\mathbf{F}$  ( $\det \mathbf{F} \neq 0$ ),

$$\mathbf{y}(k) = \mathbf{F}\mathbf{x}(k)\mathbf{F}^T. \quad (4.12)$$

Τα αρχικά δείγματα, κατά την αποκωδικοποίηση επανυπολογίζονται ως

$$\mathbf{x}(k) = \mathbf{F}^{-1}\mathbf{y}(k)\mathbf{F}^{-T}. \quad (4.13)$$

### Μετασχηματισμός Karhunen Loeve - KL

Ο μετασχηματισμός Karhunen Loeve είναι ο ιδανικός μετασχηματισμός από πλευράς αποσυσχέτισης.

Στη μονοδιάστατη εκδοχή του ο πίνακας  $\mathbf{F}$  ορίζεται ως ακολούθως:

Αν  $\mathbf{R}_x \equiv E\{\mathbf{x}(k)\mathbf{x}^T(k)\}$  είναι ο πίνακας αυτοσυσχέτισης του διανύσματος  $\mathbf{x}(k)$  (Για στάσιμα σήματα ο πίνακας αυτός δεν εξαρτάται από το  $k$ ) και

$$\mathbf{R}_x = \mathbf{U}\mathbf{S}\mathbf{U}^T$$

η ανάλυση ιδιοτιμών του (eigenvalue decomposition) όπου  $\mathbf{U}$  ο ορθοκανονικός πίνακας ιδιοδυνατισμάτων ( $\mathbf{U}^T = \mathbf{U}^{-1}$ ) και  $\mathbf{S}$  ο διαγώνιος πίνακας ιδιοτιμών του  $\mathbf{R}_x$  τότε

$$\mathbf{F} \triangleq \mathbf{U}^T \quad (4.14)$$



Ως αποτέλεσμα,

$$\mathbf{R}_y \equiv E\{\mathbf{y}(k)\mathbf{y}^T(k)\} = E\{\mathbf{U}^T \mathbf{x}(k)\mathbf{x}^T(k)\mathbf{U}\} = \mathbf{U}^T \mathbf{R}_x \mathbf{U} = \mathbf{S}$$

Συνεπώς, επειδή ο πίνακας ιδιοτιμών  $\mathbf{S}$  είναι διαγώνιος, τα δείγματα του  $\mathbf{y}(k)$  είναι μεταξύ τους πλήρως ασυσχέτιστα.

Ο ορισμός του μετασχηματισμού KL επεκτείνεται στις δύο διαστάσεις αν χειριστούμε τα διδιάστατα blocks της Σχέσης 4.11 σαν μονοδιάστατα διανύσματα αφού τοποθετήσουμε τη μία στήλη τους κάτω από την άλλη.

Παρά τις εξαιρετικές του ιδιότητες ως προς την αποσυσχέτιση ο μετασχηματισμός KL δεν χρησιμοποιείται στην πράξη για τους εξής λόγους:

- Τις περισσότερες φορές η υπόθεση ότι τα σήματα είναι *στάσιμα* παραβιάζεται.
- Ο ακριβής υπολογισμός (εκτίμηση) του πίνακα αυτοσυσχέτισης δεν είναι πάντα εφικτός.
- Δεν υπάρχει *ταχύς αλγόριθμος* υπολογισμού του μετασχηματισμού.

### Μετασχηματισμός DCT

Ο Διακριτός Μετασχηματισμός Συνημιτόνου (Discrete Cosine Transform - DCT) είναι ο πλέον χρησιμοποιούμενος μετασχηματισμός στην επεξεργασία και ιδίως στη συμπίεση πολυμέσων. Ο πίνακας  $\mathbf{F}$  του μετασχηματισμού ορίζεται από τη σχέση:

$$F_{kl} = \begin{cases} \sqrt{2/N} \cos\left(\frac{\pi}{N}k(l + \frac{1}{2})\right), & k = 1, \dots, N-1 \\ \sqrt{1/N} \cos\left(\frac{\pi}{N}k(l + \frac{1}{2})\right), & k = 0 \end{cases} \quad (4.15)$$

Η δημοτικότητα του DCT οφείλεται στους εξής λόγους:

1. Σε πληθώρα περιπτώσεων επιτυγχάνει αποσυσχέτιση εφάμιλλη του KL
2. Έχει εξαιρετική ικανότητα συγκέντρωσης της πληροφορίας σε ένα μικρό πλήθος δειγμάτων (συντελεστών) του block  $\mathbf{y}(\mathbf{k})$ .
3. Επιδέχεται ταχύτατες υλοποιήσεις για  $N$  ίσα με δυνάμεις του 2. Επιπλέον οι υλοποιήσεις αυτές είναι εύκολα μεταφέρσιμες σε hardware.

### 4.2.5 Εκτίμηση Κίνησης

Οι μέθοδοι εκτίμησης κίνησης εφαρμόζονται σε ακολουθίες κινούμενων εικόνων (βίντεο) με σκοπό τη μείωση της συσχέτισης μεταξύ διαδοχικών εικόνων της ακολουθίας. Υπό αυτή την έννοια προσομοιάζουν στις μεθόδους DPCM. Αναλυτικότερη αναφορά στις μεθόδους αυτές περιέχεται στην περιγραφή του προτύπου κωδικοποίησης βίντεο MPEG-1 στην Ενότητα 4.5.3.

### 4.3 Συμπύεση Ήχου

βλ. Άρθρα (Davis Yen Pan 1993) και (A. Spanias 1994) σχετικά με τα υπάρχοντα πρωτόκολλα συμπίεσης ψηφιακού ήχου.

### 4.4 Συμπύεση Εικόνων

βλ. Άρθρο (Gregory K. Wallace 1991) σχετικά με το πρότυπο JPEG.

### 4.5 Συμπύεση Βίντεο

#### 4.5.1 H.26x

Το πρότυπο H.261 αναπτύχθηκε για να χρησιμοποιηθεί σε εφαρμογές τηλεσυνδιάσκεψης πάνω από δίκτυα ISDN σε ρυθμούς  $p \times 64$  kbps,  $p = 1 \dots 30$ . Το πρότυπο εκτός του ότι αποτέλεσε την βάση για τα πρότυπα MPEG-1/2 προσφέρει δύο σημαντικά χαρακτηριστικά: α) προδιαγράφει την μέγιστη καθυστέρηση κωδικοποίησης στα 150 msec και β) η υλοποίηση του αλγορίθμου είναι εφικτή με VLSI chips. Στην προσπάθεια για ενιαίο τρόπο αντιμετώπισης των τηλεοπτικών σημάτων NTSC και PAL το πρότυπο προδιαγράφει μια τυπική εικόνα εισόδου CIF (Common Interchange Format). Προαιρετικά για εφαρμογές χαμηλού ρυθμού μετάδοσης ορίζεται το QCIF το οποίο έχει τις μισές διαστάσεις του CIF. Χαρακτηριστικά αναφέρουμε ότι για ρυθμούς 30 πλαίσια/sec ο ρυθμός μετάδοσης για ασυμπίεστα δεδομένα εισόδου είναι 37.3 Mbps και 9.35 Mbps για CIF και QCIF αντίστοιχα. Ακόμα και σε ρυθμούς των 10 πλαισίων/sec με εικόνες εισόδου QCIF απαιτείται λόγος συμπίεσης 48 : 1 για μετάδοση σε γραμμή 64 kbps.

Το πρότυπο ορίζει ότι τα δεδομένα έχουν μια ιεραρχημένη δομή. Κάθε εικόνα αποτελείται από 12 GOB (group-of-blocks). Κάθε GOB αποτελείται από 33 MacroBlock (MB) το οποίο αποτελείται από υποεικόνες B (blocks) διαστάσεων  $8 \times 8$ . Το macroblock αποτελεί το μικρότερο κομμάτι εικόνας στο οποίο μπορεί να διαφοροποιηθεί ο τρόπος συμπίεσης. Κάθε MB αποτελείται από τέσσερα Block φωτεινότητας Y και τα αντίστοιχα Block χρωματικότητας U και V, τα οποία έχουν προκύψει από υποδειγματοληψία τεσσάρων Block χρωματικότητας διαστάσεων ίδιων με αυτά της φωτεινότητας (Αναπαράσταση 4:2:0). Κατά την διαδικασία προτυποποίησης έγιναν συγκριτικές μελέτες μεταξύ διάφορων αλγορίθμων συμπίεσης. Το πρότυπο ορίζει ένα υβριδικό αλγόριθμο κωδικοποίησης τον Διακριτό Μετασχηματισμό Συνημίτονου (DCT) με Αντιστάθμιση κίνησης. Προηγούμενα κωδικοποιημένα πλαίσια εικόνων φυλάγονται στην μνήμη (Frame Store) τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή. Η μονάδα εκτίμησης κίνησης ανιχνεύει και υπολογίζει την κίνηση στην εικόνα. Υπολογίζεται ένα διάνυσμα κίνησης από το προηγούμενο πλαίσιο εικόνας με ακρίβεια

1 pixel για κάθε MB. Το πρότυπο δεν προδιαγράφει την μέθοδο, αλλά μπορεί να χρησιμοποιηθεί κάποια από τις γνωστές μεθόδους εκτίμησης κίνησης. Η χρονική πρόβλεψη της τιμής κάθε pixel σε κάθε MB αφαιρείται από την τρέχουσα τιμή και προκύπτει το χρονικά αντισταθμισμένο σφάλμα πρόβλεψης. Ένας  $8 \times 8$  DCT εφαρμόζεται σε κάθε μπλοκ σφαλμάτων εκτίμησης. Στην συνέχεια ακολουθεί ο κβαντιστής (Q) και κατόπιν ο κωδικοποιητής εντροπίας (VLC). Κωδικοποιούνται επιπρόσθετα στην περίπτωση μη μηδενικής κίνησης και τα διανύσματα κίνησης. Ένας απομονωτής (video buffer) βρίσκεται στην έξοδο των δεδομένων για να εγγυάται σταθερή ροή δεδομένων. Οι πίνακες κβαντισμού μπορούν να μεταβάλλονται σε κάθε MB ώστε να αποφεύγεται η υπερχειλίση ή το άδειασμα του απομονωτή. Σε αυτήν την περίπτωση κωδικοποιούνται και οι στάθμες του κβαντιστή. Επιπλέον υπάρχει και ο τρόπος λειτουργίας όπου δεν κωδικοποιείται το σφάλμα πρόβλεψης (Interframe Coding) αλλά η αρχική εικόνα (Intraframe Coding). Η εναλλαγή του τρόπου λειτουργίας δεν είναι προκαθορισμένη και μπορεί να συμβαίνει για λόγους εξοικονόμησης εύρους ζώνης (Interframe Coding) ή για λόγους βελτίωσης της ποιότητας της παραγόμενης ακολουθίας εικόνων.

Ο αποκωδικοποιητής χρησιμοποιεί την αντίστροφη διαδικασία.

#### 4.5.2 M-JPEG

Παρότι η ανάπτυξη του JPEG είχε ως σκοπό την συμπίεση ακίνητων εικόνων υπάρχει η δυνατότητα (ειδικά με τη βοήθεια ειδικού υλικού) κωδικοποίησης μιας σειράς JPEG εικόνων σε πραγματικό χρόνο ώστε να δημιουργηθεί μια ακολουθία βίντεο. Αυτή η χρήση του αλγόριθμου JPEG αναφέρεται ως Motion JPEG ή M-JPEG. Το M-JPEG δεν είναι ένα επίσημο πρότυπο βρίσκει όμως χρήση σε πληθώρα εφαρμογών. Τυπικοί λόγοι συμπίεσης είναι μεταξύ  $10 : 1$  έως  $20 : 1$  καθώς μεγαλύτερη συμπίεση δημιουργεί αισθητές απώλειες. Η κωδικοποίηση M-JPEG δεν εκμεταλλεύεται τη συσχέτιση που υπάρχει μεταξύ διαδοχικών frames και αυτός είναι λόγος για τον οποίο δεν επιτυγχάνεται μεγάλη συμπίεση.

#### 4.5.3 MPEG-1

Το πρότυπο MPEG-1 καλύπτει τις εφαρμογές κωδικοποίησης ήχου και βίντεο με bitrate της τάξης του 1.5 Mbits/sec, κατανεμημένου σε δύο τμήματα

*sim* 1.25 Mbits/sec για την κωδικοποίηση βίντεο τυπικών διαστάσεων  $352 \times 240$  pixels και  $\sim 250$  Kbits/sec για την κωδικοποίηση 2 καναλιών ήχου. Το πρότυπο αναφέρεται σε non-interlaced ακολουθίες βίντεο και είναι εν γένει βέλτιστο για αποθήκευση κωδικοποιημένων ακολουθιών σε CD-ROM. Το πρότυπο MPEG-1 δέχεται σαν είσοδο μία ακολουθία εικόνων σε μορφή CCIR 601, δηλαδή εικόνες των 625 (576 πραγματικών) γραμμών στα 50 Hz, με αριθμό πραγματικών pixels ανά γραμμή 720 για την συνιστώσα της φωτεινότητας Y και 360 για τις δύο χρωματικές συνιστώσες

Cr και Cb. Για να είναι δυνατή η κωδικοποίηση των εικόνων σε τόσο χαμηλούς ρυθμούς, οι εικόνες υποδειγματοληπτούνται αφού φιλτραριστούν με κάποιο κατωδιαβατό φίλτρο και προς τις δύο διαστάσεις, με αποτέλεσμα να προκύπτουν εικόνες με 288 πραγματικές γραμμές (στην πράξη επιλέγονται μόνο οι περιττές γραμμές, ενώ αγνοούνται οι άρτιες) και 360 (180) pixels ανά γραμμή για τα σήματα φωτεινότητας (χρωματικότητας).

Ο αλγόριθμος συμπίεσης του βίντεο περιλαμβάνει τα επόμενα κύρια στάδια<sup>2</sup>:

- Χωρισμό των εικόνων σε blocks των  $8 \times 8$  pixels και εφαρμογή του διακριτού μετασχηματισμού συνημίτονου (DCT) σε κάθε μία από αυτές.
- Εφαρμογή τεχνικών εκτίμησης και αντιστάθμισης κίνησης για την μείωση της χρονικής συσχέτισης
- Κωδικοποίηση των μετασχηματισμένων υποεικόνων είτε ανεξάρτητα για κάθε εικόνα (intraframe), είτε με πρόβλεψη της εικόνας από την προηγούμενή της (predictive).
- Παρεμβολή (interpolation) στον δέκτη, των pixels και των γραμμών που είχαν παραληφθεί κατά την αρχική υποδειγματοληψία, για επαναδημιουργία της αρχικής εικόνας σε μορφή CCIR 601.

Στη συνέχεια γίνεται μία περιγραφή των σημαντικότερων τεχνικών που χρησιμοποιεί το πρότυπο MPEG-1. Η περιγραφή αναφέρεται στα εξής σημεία:

1. Αναπαράσταση των εικόνων στην είσοδο του προτύπου και μέθοδοι υποδειγματοληψίας και υπερδειγματοληψίας.
2. Πρόβλεψη εικόνων και μέθοδοι υπολογισμού διανυσμάτων κίνησης.
3. Μετασχηματισμοί και κβαντισμός των εικόνων.
4. Κωδικοποίηση και έλεγχος του bitrate της πληροφορίας.

### **Μορφή - Μετατροπές Ποιότητας Εικόνων**

Όπως ήδη αναφέρθηκε παραπάνω, οι εικόνες μετά την ψηφιοποίησή τους έχουν την μορφή CCIR 601 ( $Y:720 \times 576$ , Cr, Cb: $360 \times 576$ ). Για να επιτύχει μία πρώτη συμπίεση της πληροφορίας, το πρότυπο MPEG-1 χρησιμοποιεί μόνο τις περιττές γραμμές κάθε εικόνας. Για τον σκοπό αυτό η αρχική εικόνα CCIR 601 περνάει από ένα φίλτρο που απορρίπτει τις άρτιες γραμμές. Στην συνέχεια δε ακολουθεί υποδειγματοληψία κατά την οριζόντια διεύθυνση (horizontal decimation), ώστε η τελική προς κωδικοποίηση εικόνα να έχει διαστάσεις  $360 \times 288$  σημείων (μορφή SIF). Η

---

<sup>2</sup>Για τη συμπίεση του Ήχου βλ. αντίστοιχη ενότητα

υποδειγματοληψία γίνεται με την βοήθεια του κατωδιαβατού φίλτρου με συντελεστές κρουστικής απόκρισης  $[-29 \ 0 \ 88 \ 138 \ 88 \ 0 \ -29]$ .

Με ακριβώς αντίστροφη μέθοδο επιτυγχάνεται η υπερδειγματοληψία (oversampling) της εικόνας στην έξοδο του αποκωδικοποιητή MPEG-1, όπου από ένα πεδίο διαστάσεων  $360 \times 288$  σημείων ανακατασκευάζεται η αρχική εικόνα με την εφαρμογή του παρακάτω φίλτρου  $[-12 \ 0 \ 140 \ 256 \ 140 \ 0 \ -12]$

### Χρονική Πρόβλεψη - Εκτίμηση Κίνησης Εικόνων

Η διαδικασία της εκτίμησης κίνησης αποσκοπεί στο να μειώσει τη χρονική συσχέτιση μεταξύ των κωδικοποιούμενων εικόνων. Πρόκειται για τεχνική αντίστοιχη της πρόβλεψης που χρησιμοποιείται στους αλγορίθμους DPCM και ADPCM.

Η εκτίμηση κίνησης εφαρμόζεται σε επιλεγμένα πλαίσια μίας ακολουθίας βίντεο ως εξής:

#### 1. *macroblocks*:

Τα pixels κάθε πλαισίου ομαδοποιούνται σε macroblocks αποτελούμενα από  $4 \times 8 \times 8$  blocks φωτεινότητας (Y) και από ένα  $8 \times 8$  block για τις χρωματικές συνιστώσες Cr και Cb. Η ομαδοποίηση αυτή γίνεται λόγω του ότι υιοθετείται η εσωτερική αναπαράσταση 4:2:0.

#### 2. *Εκτίμηση κίνησης*:

Για κάθε macroblock των εν λόγω πλαισίων (έστω  $B_{t+1}$ ) αναζητείται εκείνο το macroblock του πλαισίου αναφοράς (π.χ. του προηγούμενου πλαισίου) που του μοιάζει περισσότερο (έστω  $M_t$ ). Ως κριτήριο ομοιότητας υιοθετείται μία συνάρτηση κόστους όπως, για παράδειγμα, το άθροισμα των τετραγώνων των διαφορών των συγκρινόμενων macroblocks. Οι συγκρίσεις ομοιότητας αρχούνται στη σύγκριση των blocks της φωτεινότητας.

#### 3. *motion vectors*:

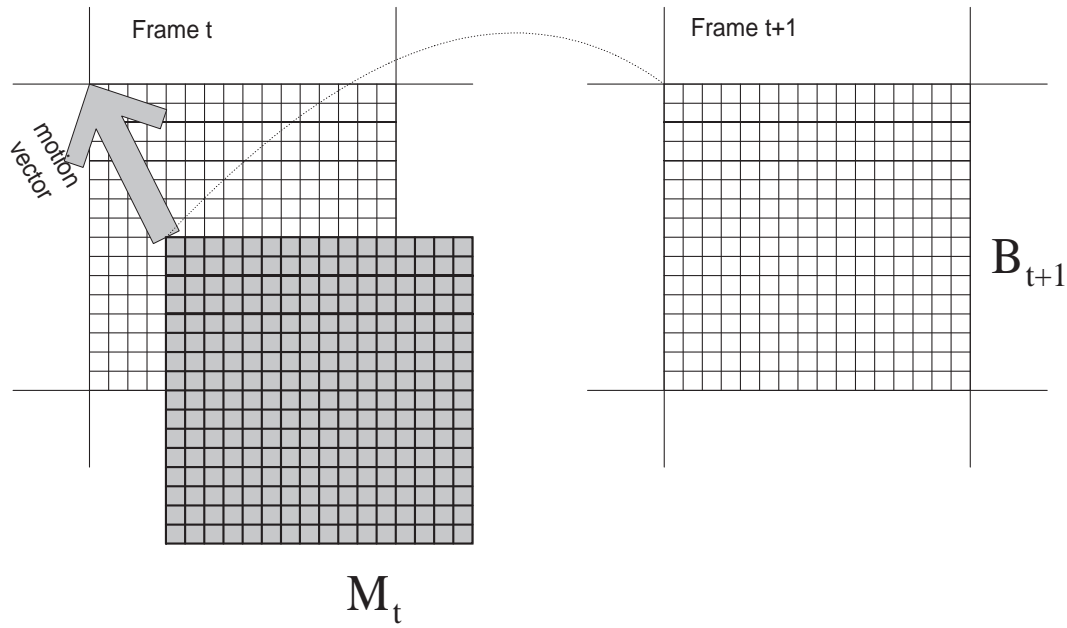
Αποτέλεσμα της εκτίμησης κίνησης είναι τα λεγόμενα *motion vectors*, δηλαδή διανύσματα δύο συντεταγμένων που υποδηλώνουν τη σχετική μετατόπιση του macroblock  $B_{t+1}$  ως προς το macroblock  $M_t$ .

#### 4. *Σφάλμα Πρόβλεψης*:

Στη συνέχεια αντί να κωδικοποιηθεί το περιεχόμενο του macroblock  $B_{t+1}$ , κωδικοποιείται το σφάλμα πρόβλεψης, δηλαδή ένα macroblock διαστάσεων ίδιων με το αρχικό και τιμή:

$$E_{t+1} \triangleq B_{t+1} - M_t$$

Η παραπάνω διαδικασία αναπαρίσταται στο Σχήμα 4.1.



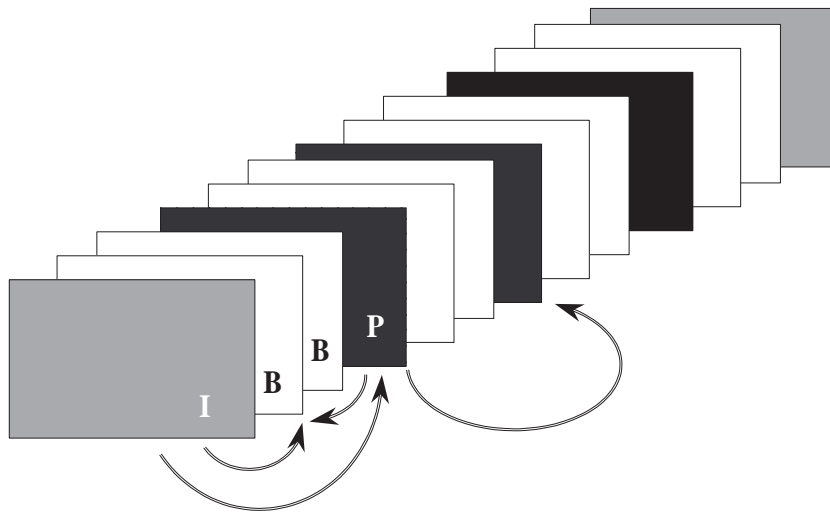
Σχήμα 4.1: Εκτίμηση κίνησης ενός  $16 \times 16$  macroblock,  $B_{t+1}$  του  $t + 1$ -πλασίου με αναφορά το  $t$ -πλαίσιο. Στο εικονιζόμενο παράδειγμα το  $16 \times 16$  macroblock  $M_t$  θεωρήθηκε ότι προσεγγίζει καλύτερα το macroblock  $B_{t+1}$ . Ως αποτέλεσμα θα κωδικοποιηθεί η διαφορά  $E_{t+1} \triangleq B_{t+1} - M_t$  και επιπλέον το αντίστοιχο motion vector  $v_m = (-4, 8)$ .

Ανά τακτά χρονικά διαστήματα η διαδικασία πρόβλεψης διακόπτεται και κωδικοποιείται η πληροφορία ολόκληρης της επόμενης εικόνας προς αποφυγή συσσώρευσης αθροιστικού σφάλματος. Για τον σκοπό αυτό τα πλαίσια διακρίνονται σε:

- Τύπου I (Intraframe) το περιεχόμενο των οποίων κωδικοποιείται αυτόνομα σαν να πρόκειται για ακίνητες εικόνες.
- Τύπου P (Predicted) στα οποία κάθε macroblock προβλέπεται με αναφορά το *προηγούμενο* I ή το *προηγούμενο* P πλαίσιο (ανάλογα με το ποιο είναι το πιο πρόσφατο).
- Τύπου B (Bidirectionnaly Interpolated Frames), για τα οποία ως αναφορά θεωρούνται ταυτόχρονα το *προηγούμενο* I ή P πλαίσιο *καθώς και* το *επόμενο* I ή P πλαίσιο. Η εκτίμηση κίνησης έχει ως αποτέλεσμα την εξεύρεση δύο διαφορετικών macroblocks, το  $M_{t-}$  (από την αναζήτηση στο προηγούμενο I ή P πλαίσιο) και το  $M_{t+}$  (από την αναζήτηση στο επόμενο I ή P πλαίσιο). Το macroblock σφάλματος που κωδικοποιείται είναι το:

$$E_{t+1} \triangleq B_{t+1} - \frac{1}{2}(M_{t-} + M_{t+})$$

Στη συνηθέστερη περίπτωση η ακολουθία εικόνων του υπό κωδικοποίηση βίντεο υποδιαιρείται σε ομάδες πλασίων (Group of Pictures) που αρχίζουν με ένα πλαίσιο τύπου I και περιέχουν πλαίσια



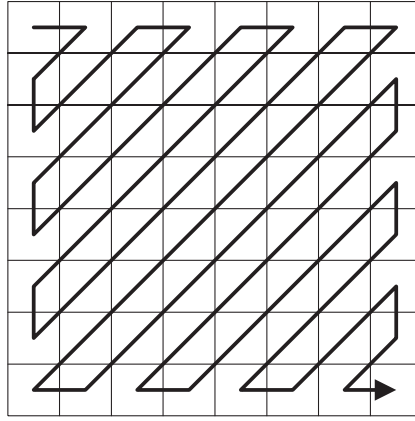
Σχήμα 4.2: Διαδοχή και μέθοδος πρόβλεψης των I, P και B πλαισίων εντός ενός GOP 12 πλαισίων.

τύπου P και B σε συγκεκριμένες θέσεις όπως φαίνεται στο Σχήμα 4.2. Το εικονιζόμενο GOP έχει τη διάταξη IBBPBBPBBPBB. Στην πλευρά του αποκωδικοποιητή η αποκωδικοποίηση αναπαράγει πρώτα την πληροφορία που μπορεί να αναπαραχθεί. Έτσι για το προηγούμενο παράδειγμα ακολουθείται η διαδοχή **IBBPBBPBBPBBIBB** όπου τα δύο πλαίσια τύπου B που σημειώνονται με έντονη γραφή ανήκουν στο προηγούμενο GOP και το τελευταίο I (επίσης με έντονη γραφή) ανήκει στο επόμενο GOP.

### Μετασχηματισμός και Κβαντισμός Εικόνων

Όλα τα πλαίσια που υπολογίστηκαν μετά την εκτίμηση κίνησης και την πρόβλεψη, δηλαδή οι ίδιες οι εικόνες για τα πλαίσια τύπου I και τα σφάλματα πρόβλεψης για τα πλαίσια τύπου P και B κωδικοποιούνται μέσω του διδιάστατου Διακριτού Μετασχηματισμού Συνημίτονου (DCT). Ο μετασχηματισμός εφαρμόζεται σε κάθε block διαστάσεων  $8 \times 8$  της εικόνας (τόσο στα blocks Y όσο και στα blocks Cr και Cb), δημιουργώντας ένα αντίστοιχο πίνακα συντελεστών του μετασχηματισμού των ίδιων διαστάσεων. Τα στοιχεία των πινάκων αυτών περιγράφουν το περιεχόμενο κάθε υποεικόνας στο πεδίο των χωρικών συχνοτήτων. Το στοιχείο (0, 0) (πάνω αριστερά) αντιστοιχεί στον συντελεστή DC του μετασχηματισμού, τα γειτονικά σε αυτό στοιχεία αντιστοιχούν σε χαμηλές συχνότητες κ.ο.κ. Ο μετασχηματισμός DCT έχει επιλεγεί για τους λόγους που αναφέρονται στην αντίστοιχη ενότητα.

Αφού υπολογιστούν οι συντελεστές DCT κάθε block στην συνέχεια υπόκεινται σε κβαντισμό. Συνήθως χρησιμοποιείται κβαντιστής νεκρής ζώνης που εκμηδενίζει τους μικρούς συντελεστές



Σχήμα 4.3: Σάρωση zig-zag των χβαντισμένων συντελεστών DCT ενός block  $8 \times 8$ .

DCT.

### Κωδικοποίηση χωρίς απώλειες

Στη συνέχεια, για κάθε block, τα σύμβολα που προκύπτουν από των χβαντισμό διατάσσονται σε ένα μονοδιάστατο διάνυσμα μήκους 64 με σειρά που υποδεικνύεται από τη λεγόμενη *σάρωση zig-zag* (βλ. Σχήμα 4.3).

Τα σύμβολα των παραπάνω διανυσμάτων κωδικοποιούνται σε μήκη διαδρομής με τη χρήση του αλγορίθμου RLE και στη συνέχεια κατά Huffman. Αναλυτικότερα, αν  $s(0)$  είναι το σύμβολο που αναθέτει ο χβαντιστής στη νεκρή ζώνη και  $s(i)$  οποιοδήποτε άλλο σύμβολο παράγει ο χβαντιστής, η κωδικοποίηση μήκους διαδρομής αντικαθιστά συμβολοσειρές της μορφής

$$\underbrace{s(0) \cdots s(0)}_n s(i), \quad n \geq 0,$$

με τα νέα (συνοπτικά) σύμβολα  $A_{ni}$  που υποδηλώνουν  $n$  μηδενικά ακολουθούμενα από το σύμβολο  $s(i)$ . Τα νέα συνοπτικά σύμβολα είναι αυτά που κωδικοποιούνται κατά Huffman.

Οι δύο παραπάνω αλγόριθμοι συμπίεσης (χωρίς απώλειες) χρησιμοποιούνται και για την κωδικοποίηση των motion vectors τα οποία ομαδοποιούνται σε επίπεδο πλαισίου.

### Έλεγχος του bitrate

Η κωδικοποίηση MPEG-1 επιδέχεται μηχανισμούς ελέγχου του bitrate του κωδικοποιημένου βίντεο. Συνοπτικά οι εν λόγω μηχανισμοί λειτουργούν ως εξής:

- Το προϊόν της τελικής κωδικοποίησης μεταφέρεται σε έναν buffer ελέγχου ο οποίος αδειάζει με ρυθμό ίσο προς το επιθυμητό bitrate (target bitrate).



- Όταν η στάθμη του εν λόγω buffer ξεπερνά κάποιο προκαθορισμένο όριο ειδοποιείται ο κβαντιστής να γίνει αυστηρότερος (π.χ., αυξάνοντας το εύρος της νεκρής ζώνης) επιφέροντας, βέβαια, χειροτέρευση της ποιότητας κωδικοποίησης (μεγαλύτερες απώλειες).
- Αντίθετα, όταν η στάθμη του buffer βρεθεί χαμηλότερα από ένα κάτω όριο ενεργοποιείται η αντίστροφη διαδικασία.

Επιτυγχάνεται έτσι η διατήρηση του μέσου bitrate πολύ κοντά στην τιμή του target bitrate. Το εν λόγω είδος κωδικοποίησης ονομάζεται κωδικοποίηση Σταθερού Bitrate (Constant Bitrate - *CBR*).

Όταν ο παραπάνω μηχανισμός ελέγχου του bitrate δεν υπάρχει ή ενεργοποιείται μόνο σε πολύ ακραίες περιπτώσεις, η ποιότητα κωδικοποίησης παραμένει (σχεδόν) σταθερή καθόλη τη διάρκεια του βίντεο αλλά το bitrate μεταβάλλεται έντονα με το χρόνο. Η κωδικοποίηση του είδους αυτού αποκαλείται Μεταβλητού Bitrate (Variable Bitrate - *VBR*).

#### 4.5.4 MPEG-2

Το πρότυπο MPEG-2 έχει σχεδιαστεί για να αντιμετωπίσει σαφώς υψηλότερους ρυθμούς μετάδοσης, οι οποίοι αγγίζουν τα 60 Mbits/sec. Επίσης, έχει σχεδιαστεί έτσι ώστε να μπορεί να αντιμετωπίσει μια πλειάδα formats για την ακολουθία εισόδου, να μπορεί να περιλάβει μέχρι και 5 ανεξάρτητα κανάλια ήχου (surround sound), είναι δε σχεδιασμένο για την κωδικοποίηση interlaced ακολουθιών. Το πρότυπο MPEG-2 απευθύνεται στις εφαρμογές ψηφιακής τηλεόρασης, τηλεόρασης υψηλής ευκρίνειας (HDTV), στα DVD και το ψηφιακό σινεμά.

Οι προδιαγραφές κωδικοποίησης βίντεο του MPEG-2 έχουν τις ίδιες βασικές αρχές με τις αντίστοιχες του MPEG-1. Ίσως η σημαντικότερη διαφοροποίηση βρίσκεται στην υποστήριξη της λεγόμενης *κλιμακούμενης κωδικοποίησης* (scalable coding). Η εν λόγω κλιμάκωση της πληροφορίας συνίσταται στη δημιουργία ενός stream δεδομένων που αποτελείται από επιμέρους παράλληλα streams έτσι ώστε ένα από αυτά να περιέχει την απαραίτητη (συνήθως χαμηλής ποιότητας) κωδικοποιημένη μορφή του βίντεο και τα υπόλοιπα, εφόσον χρησιμοποιηθούν, να συνεισφέρουν στην αναβάθμιση της ποιότητας του αρχικού. Τέσσερα είδη κλιμάκωσης υποστηρίζονται από το πρότυπο:

- *Διαμέριση Δεδομένων (Data Partitioning)* όπου για παράδειγμα το βασικό stream περιέχει πληροφορία μόνο από τους συντελεστές DCT που αντιστοιχούν σε χαμηλές συχνότητες ενώ συμπληρωματικά streams περιέχουν πληροφορία για συντελεστές υψηλότερων συχνοτήτων.
- *Σηματοθορυβική Κλιμάκωση (SNR Scalability)*: όπου για το βασικό stream περιέχει πληροφορία μόνο για τα πιο σημαντικά bits (most significant bits) των συντελεστών DCT ενώ συμπληρωματικά streams περιέχουν πληροφορία για τα λιγότερο σημαντικά bits (least significant bits) των ίδιων συντελεστών

- *Χωρική Κλιμάκωση (Spatial Scalability)*: όπου στο βασικό stream μεταφέρεται πληροφορία μίας χωρικά υποδειγματοληπτημένης έκδοσης του βίντεο ενώ τα συμπληρωματικά streams μεταφέρουν την πληροφορία που απαιτείται για την αύξηση της ανάλυσης.
- *Χρονική Κλιμάκωση (Temporal Scalability)*: όπου δύο διαφορετικά streams περιέχουν πληροφορία για χρονικά υποδειγματοληπτημένες εκδόσεις του βίντεο αλλά όταν συνδυαστούν αυξάνουν τη χρονική ευκρίνεια.

#### 4.5.5 MPEG-4

Σε αντίθεση, ή καλύτερα σε επέκταση των προτύπων MPEG-1 και MPEG-2 για την κωδικοποίηση βίντεο και ήχου το MPEG-4 εισάγει την ιδέα της κωδικοποίησης οπτικοακουστικών αντικειμένων (audiovisula objects - AVO). Τα εν λόγω αντικείμενα μπορεί να είναι:

- Κινούμενες φυσικές εικόνες (φυσικό βίντεο)
- Ακίνητες εικόνες
- Συνθετικά διδιάστατα ή τρισδιάστατα αντικείμενα
- Ψηφιακός ήχος (μονοκαναλικός ή πολυκαναλικός)
- Γραφικά
- Κείμενο

Η κωδικοποίηση αφορά (α) στο ίδιο το οπτικοακουστικό περιεχόμενο όπου χρησιμοποιούνται αλγόριθμοι όπως αυτοί των προτύπων MPEG-1 και MPEG-2 ή και άλλοι, (β) στην ιεραρχία των αντικειμένων. Για παράδειγμα το αντικείμενο *παρουσιαστής ειδήσεων* αποτελείται από τα επιμέρους αντικείμενα *κινούμενη εικόνα του παρουσιαστή* και *φωνή του παρουσιαστή*, (γ) πληροφορία χρονικής και χωρικής συσχέτισης (alignment) των επιμέρους αντικειμένων μίας ιεραρχίας.

Σημαντική πρωτοτυπία του προτύπου MPEG-4 είναι η ανάθεση στον *τελικό θεατή* της διαδικασίας *σύνθεσης* του τελικού αποτελέσματος. Ο θεατής παραλαμβάνει streams με τις προαναφερθείσες κωδικοποιημένες πληροφορίες για τα αντικείμενα που συμμετέχουν σε μία *σκηνή* και αναλαμβάνει να τα συνθέσει σύμφωνα με τις οδηγίες που του αποστέλλονται σε χωριστό (παράλληλο) stream. Για τις εν λόγω οδηγίες το πρότυπο ορίζει μία γλώσσα περιγραφής των σκηνών αποκαλούμενη *Binary Format for Scenes - BIFS*. Η γλώσσα BIFS έχει δανειστεί σε σημαντικό βαθμό τις ιδέες της *Virtual Reality Modeling Language - VRML* έχει όμως προσθέσει και νέα στοιχεία για την ένταξη στις σκηνές φυσικού βίντεο και εικόνων.

Μία συνοπτική παρουσίαση του προτύπου MPEG-4 περιέχεται στο Άρθρο (Rob Koenen 1999) ενώ μία πιο αναλυτική παρουσίαση της γλώσσας BIFS παρουσιάζεται στο Άρθρο (Julien Signes 1998).

## Κεφάλαιο 5

# Αποθήκευση/Αναζήτηση/Ανάκτηση

Οι κυρίαρχη τεχνολογία για την οργανωμένη αποθήκευση, αναζήτηση (με βάση το περιεχόμενο) και ανάκτηση οπτικοακουστικού υλικού ενσωματώνεται στο πρότυπο *MPEG-7*. Το πρότυπο έχει ως βασικό στόχο την περιγραφή της πληροφορίας που συνοδεύει τις πολυμεσικές οντότητες (metadata) έτσι ώστε να είναι εύκολη

- η αρχειονομική τους ταξινόμηση
- η αναζήτηση σε βάσεις δεδομένων που τις περιέχουν (ή σε τηλεπικοινωνιακά κανάλια που τις μεταδίδουν)
- η πλοήγηση στο εσωτερικό τους (π.χ. μέσω της προεπισκόπησης επιλεγμένων αποσπασμάτων)

Το πρότυπο υποστηρίζει metadata τα οποία είτε δημιουργούνται χειροκίνητα (π.χ., τα ονόματα των συντελεστών μιας ταινίας, ο χρόνος παραγωγής της, τα γεγονότα στα οποία αναφέρεται ένα ντοκιμαντέρ, κ.λπ.) είτε εξάγονται αυτόματα (ή ημιαυτόματα) από το οπτικοακουστικό περιεχόμενο (π.χ., τα χρονικά όρια των κινηματογραφικών πλάνων σε μία ταινία, το αντιπροσωπευτικό χρώμα και η υφή μίας εικόνας, το σχήμα των κύριων αντικειμένων σε μία σκηνή, κ.λπ.).

Αντίθετα με τα προηγούμενα πρότυπα της επιτροπής MPEG, το MPEG-7 δεν ασχολείται καθόλου με τον τρόπο κωδικοποίησης των πολυμεσικών οντοτήτων (data στην ορολογία του προτύπου). Ασχολείται όμως με την οργάνωση και τη συμπίεση των metadata.

Το Άρθρο (Shih-Fu Chang, Thomas Sikora and Atul Puri 2001) δίνει συνοπτικά τη συνολική ιδέα του εν λόγω προτύπου. Ο τρόπος με τον οποίο αναλύεται και περιγράφεται η οπτική πληροφορία περιγράφεται συνοπτικά στο Άρθρο (Thomas Sikora 2001). Ειδικότερα στα Άρθρα (Shih-Fu Chang, Thomas Sikora and Atul Puri 2001), (Mirosław Bober 2001) και (Sylvie Jeannin and Ajay Divakaran 2001) παρουσιάζεται ο τρόπος με τον οποίο το πρότυπο MPEG-7 χειρίζεται την πληροφορία του χρώματος και της υφής, του σχήματος των αντικειμένων και της κίνησης αντίστοιχα.

Τέλος στο Άρθρο (Philippe Salembier and John R. Smith 2001) παρουσιάζεται ο τρόπος οργάνωσης της πληροφορίας ώστε να είναι δυνατή η αναζήτηση πολυμεσικών οντοτήτων και η πλοήγηση στο εσωτερικό τους (π.χ. στις σκηνές ενός βίντεο).

# Βιβλιογραφία

- A. Spanias. 1994. ‘Speech Coding: A Tutorial Review.’ *Proceedings of the IEEE* 82(10):1541–1582.
- Chris Kyriakakis. 1998. ‘Fundamental and Technological Limitations of Immersive Audio Systems.’ *Proceedings of the IEEE* 86(5):941–951.
- Davis Yen Pan. 1993. ‘Digital Audio Compression.’ *Digital Technical Journal* 5(2):1–14.
- Gregory K. Wallace. 1991. ‘The JPEG Still Picture Compression Standard.’ *IEEE Transactions on Consumer Electronics* .
- Julien Signes. 1998. ‘Binary Format for Scene (BIFS): Combining MPEG-4 media to build rich multimedia services.’ *SPIE Proceedings* .
- Mirosław Bober. 2001. ‘MPEG-7 Visual Shape Descriptors.’ *IEEE Transactions on Circuits and Systems for Video Technology* 11(6):716–719.
- Philippe Salembier and John R. Smith. 2001. ‘MPEG-7 Multimedia Description Schemes.’ *IEEE Transactions on Circuits and Systems for Video Technology* 11(6):748–759.
- Rob Koenen. 1999. ‘MPEG-4 multimedia for our time.’ *IEEE Spectrum* 36(2):26–33.
- Shih-Fu Chang, Thomas Sikora and Atul Puri. 2001. ‘Overview of the MPEG-7 Standard.’ *IEEE Transactions on Circuits and Systems for Video Technology* 11(6):688–695.
- Sylvie Jeannin and Ajay Divakaran. 2001. ‘MPEG-7 Visual Motion Descriptors.’ *IEEE Transactions on Circuits and Systems for Video Technology* 11(6):720–724.
- Thomas Sikora. 2001. ‘The MPEG-7 Visual Standard for Content Description • An Overview.’ *IEEE Transactions on Circuits and Systems for Video Technology* 11(6):696–702.