



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ  
ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ  
ΣΕ  
«ΣΥΣΤΗΜΑΤΑ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ ΚΑΙ ΕΙΚΟΝΩΝ:  
ΘΕΩΡΙΑ, ΥΛΟΠΟΙΗΣΕΙΣ, ΕΦΑΡΜΟΓΕΣ»

## **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

“MOTION COMPENSTATION – SCALABLE VIDEO  
CODING”

ΑΘΑΝΑΣΟΠΟΥΛΟΣ ΔΙΟΝΥΣΙΟΣ  
ΔΙΠΛΩΜΑΤΟΥΧΟΣ ΜΗΧΑΝΙΚΟΣ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: κ. ΣΤΟΥΡΑΪΤΗΣ  
ΤΡΙΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:  
κ. ΣΤΟΥΡΑΪΤΗΣ, κ. ΜΠΕΡΜΠΕΡΙΔΗΣ, κ. ΠΑΛΙΟΥΡΑΣ

*Πάτρα, Ιούνιος 2006*

## *Πρόλογος*

Η εργασία αυτή πραγματοποιήθηκε στα πλαίσια της εκπόνησης διπλωματικής εργασίας του Διατμηματικού Μεταπτυχιακού Προγράμματος «Συστήματα Επεξεργασίας Σημάτων και Εικόνων: Θεωρία, Υλοποιήσεις, Εφαρμογές». Στο μεταπτυχιακό πρόγραμμα εισήχθηκα το 2004, αφού αποφοίτησα από την Πολυτεχνική Σχολή Μηχανικών Η/Υ και Πληροφορικής, του Πανεπιστημίου Πατρών.

Η παρακολούθηση του μεταπτυχιακού προγράμματος είχε ως αποτέλεσμα τη διεύρυνση των γνώσεών μου στο χώρο των τηλεπικοινωνιών και της επεξεργασίας σημάτων. Οι μεταπτυχιακές μου σπουδές ολοκληρώνονται με την ακόλουθη διπλωματική εργασία όπου μελετάται η κλιμακοθετήσιμη κωδικοποίηση βίντεο (scalable video coding), ένα θέμα με έντονο ερευνητικό ενδιαφέρον και ένα από τα αναδυόμενα στο χώρο της κωδικοποίησης βίντεο. Στην εργασία αυτή θα μελετήσουμε το χρονικό και το χωρικό μετασχηματισμό wavelet στους οποίους βασίζεται η κλιμακοθετήσιμη κωδικοποίηση βίντεο και θα προτείνουμε τεχνικές για βελτίωση της απόδοσης ενός κλιμακοθετήσιμου συστήματος κωδικοποίησης βίντεο.

Κλείνοντας, θέλω να ευχαριστήσω θερμά τον Καθηγητή κ. Στουραϊτή Αθανάσιο, για την ανάθεση και την επίβλεψη της εργασίας αυτής, για τον χρόνο που αφιέρωσε, την απαραίτητη καθοδήγηση καθώς και για την στηριξή του στο προσωπό μου. Τέλος, ευχαριστώ όλους όσους συνέβαλλαν στον εμπλουτισμό των γνώσεών μου στη διάρκεια της φοιτητικής μου πορείας και όσους στάθηκαν δίπλα μου τα χρόνια αυτά.

Πάτρα, Ιούνιος 2006

ΑΘΑΝΑΣΟΠΟΥΛΟΣ ΔΙΟΝΥΣΙΟΣ

## Περιεχόμενα

Λίστα Εικόνων .....	- 4 -
Κεφάλαιο 1 : Εισαγωγή.....	- 6 -
1.1 Η Ανάγκη για Κλιμακοθετήσιμη Κωδικοποίηση Βίντεο .....	- 6 -
1.1.1 Χωρική Κλιμακοθετησιμότητα (Spatial Scalability).....	- 8 -
1.1.2 Ποιοτική Κλιμακοθετησιμότητα (Quality Scalability).....	- 9 -
1.1.3 Χρονική Κλιμακοθετησιμότητα (Temporal Scalability).....	- 9 -
1.1.4 Κλιμακοθετησιμότητα στην Πολυπλοκότητα (Complexity Scalability).....	- 10 -
Κεφάλαιο 2 : Ο Μετασχηματισμός Wavelet .....	- 11 -
2.1 Μαθηματικό Υπόβαθρο .....	- 11 -
2.1.1 Επέκταση Διακριτών Σημάτων.....	- 11 -
2.1.2 Τράπεζα Φίλτρων (Filter Bank).....	- 15 -
2.2 Ο Μετασχηματισμός Wavelet .....	- 18 -
2.2.1 Συνεχής Μετασχηματισμός Wavelet (CWT).....	- 19 -
2.2.2 Ανάλυση Πολλαπλής Διακριτικής Ικανότητας (Multiresolution analysis) .....	- 22 -
2.2.3 The Lifting Scheme .....	- 30 -
Κεφάλαιο 3 : Κλιμακοθετήσιμη Κωδικοποίηση Βίντεο (Scalable Video Coding) .....	- 34 -
3.1 Κλιμακοθετήσιμη Κωδικοποίηση Εικόνας με χρήση των Wavelets .....	- 34 -
3.2 Κινησιακά Αντισταθμισμένη Πρόβλεψη κατά Μπλοκ (Block-based Motion Compensated Prediction) .....	- 36 -
3.2.1 Εκτίμηση Κίνησης κατά Μπλοκ (Block-based Motion Estimation) .....	- 37 -
3.2.2 Κινησιακά Αντισταθμισμένη Πρόβλεψη και Παρεμβολή Εικονοστοιχείου .....	- 39 -
3.3 Κινησιακά Αντισταθμισμένη Ανανέωση (Motion Compensated Update) ..	- 41 -
3.4 Δομές Χρονικής Πρόβλεψης για Κωδικοποίηση Βίντεο .....	- 52 -
3.5 Προχωρημένη Κινησιακά Αντισταθμισμένη Πρόβλεψη και Ανανέωση ..	- 55 -
3.6 Χωρικός Μετασχηματισμός wavelet .....	- 56 -
3.7 Ο Αλγόριθμος SPIHT – 3D SPIHT .....	- 58 -
3.8 Ανακεφαλαίωση.....	- 63 -
Κεφάλαιο 4 : Πειραματικές Μετρήσεις.....	- 65 -
4.1 Σύγκριση full-pixel, sub-pixel .....	- 65 -
4.2 Σύγκριση Φίλτρων Παρεμβολής .....	- 71 -
4.3 Επίδραση του Αριθμού Χρονικών Επιπέδων .....	- 76 -
4.4 Μελέτη της Επίδρασης της Παραμέτρου GOP .....	- 81 -

4.5 Ανακεφαλαίωση.....	- 85 -
Βιβλιογραφία .....	- 87 -

## Λίστα Εικόνων

Εικόνα 1.1 Αντιπροσωπευτικό σενάριο για video streaming μέσω του Internet, όπου πολλοί πελάτες συνδέονται μέσω ετερογενών ασύρματων πρωτοκόλλων .....	7 -
Εικόνα 1.2 Προσαρμογή σε ετερογενή συσκευές και υποδομές δικτύων με το πρότυπο MPEG-21 DIA .....	9 -
Εικόνα 2.1 Παραδείγματα διανυσμάτων για επέκταση του χώρου $R^2$ . (a) Ορθοκανονική περίπτωση (b) Biorthogonal περίπτωση .....	13 -
Εικόνα 2.2 Τράπεζα φίλτρων με φίλτρα ανάλυσης $h_0[n]$ , $h_1[n]$ και φίλτρα σύνθεσης $g_0[n]$ , $g_1[n]$ .....	16 -
Εικόνα 2.3 Τεμαχισμός του συχνοτικοχρονικού επιπέδου (a) STFT (b) Wavelet μετασχηματισμός .....	18 -
Εικόνα 2.4 Σύγκριση μετασχηματισμού Fourier και μετασχηματισμού Wavelet.....	20 -
Εικόνα 2.5 Διακριτοποίηση των παραμέτρων κλίμακας και χρονικής μετατόπισης $a = a_0^m$ , $\tau = n\tau_0 a_0^m$ . (α) Το πλέγμα που προκύπτει (β) Τα αντίστοιχα σύνολα συναρτήσεων.....	22 -
Εικόνα 2.6 (a) Haar συναρτήσεις βάσης (b) Ορθογωνιότητα των συναρτήσεων. ....	23 -
Εικόνα 2.7 Haar wavelet (α) Μητέρα συνάρτηση (β) Συνάρτηση κλιμακοθέτησης (γ) Συχνотικό περιεχόμενο της μητέρας συνάρτησης (δ) Συχνотικό περιεχόμενο της συνάρτησης κλιμακοθέτησης.....	24 -
Εικόνα 2.8 Υλοποίηση μετασχηματισμού Wavelet με χρήση τράπεζας φίλτρων.....	28 -
Εικόνα 2.9 Μετασχηματισμός wavelet με χρήση του πολυφασικού (polyphase) πίνακα... -	31 -
Εικόνα 2.10 Αναπαράσταση της wavelet ανάλυσης βασισμένη στο lifting scheme ..	33 -
Εικόνα 3.1 Αρχιτεκτονική κλιμακοθετήσιμης κωδικοποίησης εικόνας βασισμένη στο μετασχηματισμό wavelet .....	35 -
Εικόνα 3.2 Motion Compensated Prediction .....	37 -
Εικόνα 3.3 Εκτίμηση κίνησης κατά μπλοκ.....	38 -
Εικόνα 3.4 Εκτίμηση κίνησης με παρεμβολή μισού εικονοστοιχείου. Το τρέχων μπλοκ έχει μέγεθος $2 \times 2$ εικονοστοιχεία και βρίσκεται στη θέση $(m,n)$ . Παρουσιάζεται μέρος της περιοχής αναζήτησης, με διαφορετικά σύμβολα να υποδηλώνουν οριζόντιες, κάθετες και διαγώνιες θέσεις παρεμβολής. Οι θέσεις του καλύτερου ταιριάσματος στο πλήρες πλέγμα και στο πλέγμα μισών εικονοστοιχείων συμβολίζονται με γκρι τετράγωνα και απεικονίζονται και τα διανύσματα μετατόπισης. ....	39 -
Εικόνα 3.5 Wavelet ανάλυση στο πεδίο του χρόνου με εφαρμογή των βημάτων πρόβλεψης και ανανέωσης.....	42 -
Εικόνα 3.6 Διάγραμμα εφαρμογής του μετασχηματισμού Haar .....	43 -
Εικόνα 3.7 Διάγραμμα εφαρμογής μετασχηματισμού Haar στην κατεύθυνση της κίνησης .....	44 -
Εικόνα 3.8 Εφαρμογή του βήματος MCU με αντιστροφή των διανυσμάτων κίνησης -	45 -

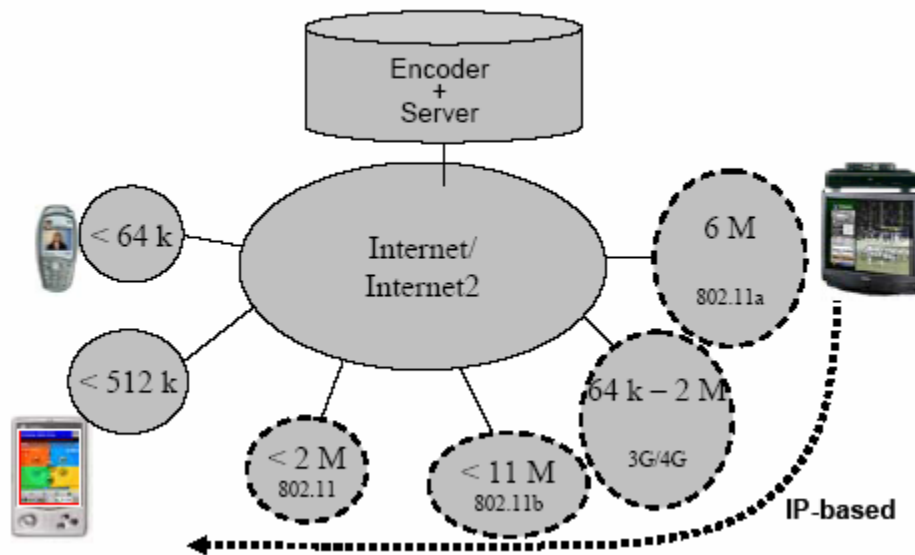
Εικόνα 3.9 Εφαρμογή του βήματος ανανέωσης στην κατεύθυνση της κίνησης. Το συγκεκριμένο παράδειγμα είναι σε μια διάσταση, με μέγεθος μπλοκ δύο και ακρίβεια μισού εικονοστοιχείου .....	- 46 -
Εικόνα 3.10 Καταστάσεις σύνδεσης ενός εικονοστοιχείου.....	- 49 -
Εικόνα 3.11 Πρόταση των Girod και Han για το χειρισμό των πολυσύνδετων εικονοστοιχείων .....	- 51 -
Εικόνα 3.12 Σύστημα κωδικοποίησης βίντεο κλειστού βρόχου.....	- 52 -
Εικόνα 3.13 Σύστημα κωδικοποίησης βίντεο ανοιχτού βρόχου.....	- 53 -
Εικόνα 3.14 Παράδειγμα MH-MCP .....	- 56 -
Εικόνα 3.15 Εφαρμογή του μετασχηματισμού wavelet στο πρώτο πλαίσιο της ακολουθίας Mobile με χρήση των 9/7 biorthogonal Daubechies wavelets. ....	- 57 -
Εικόνα 3.16 Παράδειγμα χωροχρονικού μετασχηματισμού wavelet, 2 επιπέδων στον χρόνο και 3 επιπέδων στο χώρο.....	- 58 -
Εικόνα 3.17 Αλγόριθμος SPIHT - Χωρική σχέση των συντελεστών.....	- 59 -
Εικόνα 3.18 Δένδρα προσανατολισμού στον 3D-SPIHT .....	- 62 -
Εικόνα 4.1 Σύγκριση της ακρίβειας ενός και μισού εικονοστοιχείου .....	- 67 -
Εικόνα 4.2 Αποκωδικοποιημένο πλαίσιο της ακολουθίας βίντεο Mobile (α) Αρχικό πλαίσιο (β) Πλαίσιο ακρίβειας μισού εικονοστοιχείου (γ) Πλαίσιο ακρίβειας ενός εικονοστοιχείου.....	- 69 -
Εικόνα 4.3 Πλαίσιο σφάλματος (α) Ακρίβεια μισού εικονοστοιχείου (β) Ακρίβεια ενός εικονοστοιχείου.....	- 70 -
Εικόνα 4.4 Σύγκριση φίλτρων παρεμβολής.....	- 73 -
Εικόνα 4.5 Αποκωδικοποιημένο πλαίσιο της ακολουθίας βίντεο Mobile (α) Bilinear φίλτρο (β) 8-tap φίλτρο (γ) 6-tap φίλτρο. ....	- 74 -
Εικόνα 4.6 Αποκωδικοποιημένο πλαίσιο της ακολουθίας Stefan χρησιμοποιώντας (α) 3 χρονικά επίπεδα (β) 4 χρονικά επίπεδα. ....	- 79 -
Εικόνα 4.7 Αποκωδικοποιημένο πλαίσιο της ακολουθίας Mobile χρησιμοποιώντας (α) 3 χρονικά επίπεδα (β) 4 χρονικά επίπεδα. ....	- 80 -
Εικόνα 4.8 Αποκωδικοποιημένο πλαίσιο της ακολουθίας Foreman χρησιμοποιώντας (α) GOP = 16 (β) GOP =32. ....	- 83 -
Εικόνα 4.9 Αποκωδικοποιημένο πλαίσιο της ακολουθίας Mobile χρησιμοποιώντας (α) GOP = 16 .....	- 84 -

## Κεφάλαιο 1 : Εισαγωγή

### 1.1 Η Ανάγκη για Κλιμακοθετήσιμη Κωδικοποίηση Βίντεο

Στις μέρες μας ολοένα και περισσότερο αυξάνεται ο όγκος των δεδομένων βίντεο που χειριζόμαστε καθημερινά. Μια σημαντική πρόκληση είναι η διανομή του περιεχομένου βίντεο σε διαφορετικούς χρήστες. Το ψηφιακό βίντεο στη δεκαετία 1980 διανεμόταν μεταξύ εταιριών που διέθεταν ακριβά και αξιόπιστα κανάλια με μεγάλο εύρος ζώνης και δυνατότητα παροχής ποιότητας εξυπηρέτησης (QoS). Όμως, από το 1995 και έπειτα η εξάπλωση του Παγκόσμιου Ιστού άλλαξε τη φιλοσοφία της εμπορικής διανομής, με αποτέλεσμα σήμερα εκατομμύρια χρήστες να χρησιμοποιούν υπηρεσίες βίντεο ροοθήκευσης (video-streaming) για να δούνε ταινίες, μουσικά βίντεο και άλλα γεγονότα μέσω του Παγκόσμιου Ιστού. Επιπλέον, η ανάπτυξη της τεχνολογίας σύλληψης εικόνας προκάλεσε τη διεύρυνση καμερών εικόνας και βίντεο στις κινητές συσκευές, όπως τα κινητά τηλέφωνα και τα PDAs. Αυτό είχε αποτέλεσμα στις μέρες μας, ακόμα και οι φορητές συσκευές να λειτουργούν ως καταναλωτές ή και παραγωγοί ψηφιακού βίντεο.

Στις περισσότερες περιπτώσεις η μετάδοση βίντεο πραγματοποιείται μέσω μη αξιόπιστων δικτύων, χρησιμοποιώντας πρωτόκολλα βασιζόμενα στα πακέτα όπως για παράδειγμα το TCP/IP το οποίο δεν υποστηρίζει λειτουργίες ποιότητας εξυπηρέτησης για ανθεκτική και χρονικά ορθή διανομή περιεχομένου πολυμέσων. Το πρόβλημα διογκώνεται με την χρήση ασύρματων υπηρεσιών στο περιβάλλον του χρήστη, αφού τα ασύρματα δίκτυα εμφανίζουν απώλειες πακέτων και μεγάλες διακυμάνσεις στο εύρος ζώνης. Για την τυπική περίπτωση της βίντεο ροοθήκευσης η σημερινή κατάσταση απεικονίζεται στην Εικόνα 1.1, όπου παρουσιάζονται διάφορες συσκευές και δίκτυα μαζί με το αντίστοιχο απαιτούμενο εύρος επικοινωνίας [1].



**Εικόνα 1.1** Αντιπροσωπευτικό σενάριο για video streaming μέσω του Internet, όπου πολλοί πελάτες συνδέονται μέσω ετερογενών ασύρματων πρωτοκόλλων

Για να αντιμετωπιστούν τα προβλήματα που δημιουργεί η ποικιλία των μέσων μετάδοσης βίντεο, η κοινότητα του MPEG αποφάσισε να διεξάγει έρευνα στις τεχνολογίες κωδικοποίησης βίντεο που μπορούν να χειριστούν τις απαιτήσεις μετάδοσης. Η έρευνα αυτή άρχισε το 2002 και συνεχίζεται ακόμα. Όσον αφορά το τεχνολογικό κομμάτι είναι γενικώς αποδεκτό ότι η κλιμακοθετήσιμη κωδικοποίηση (scalable coding) είναι η πιο βιώσιμη τεχνικά λύση στα προβλήματα της μετάδοσης βίντεο σε ετερογενή δίκτυα και τερματικά χρηστών με διαφορετικές δυνατότητες. Η κλιμακοθετήσιμη κωδικοποίηση βίντεο αποτελεί ένα πλαίσιο εργασίας όπου από μια ενιαία συμπιεσμένη ακολουθία βίντεο μπορούν να προκύψουν αναπαραστάσεις του βίντεο με διαφορετική ποιότητα, ανάλυση (resolution) και χρονικό ρυθμό πλαισίων.

Ένα σημαντικό κομμάτι της επιτυχίας ενός συγκεκριμένου κλιμακοθετήσιμου αλγόριθμου κωδικοποίησης είναι το πόσο αποδοτική είναι η συμπίεση που δημιούργησε. Συνεπώς, μια βασική απαίτηση κάθε αλγόριθμου κλιμακοθετήσιμης κωδικοποίησης είναι ότι πρέπει να αποδίδει το ίδιο καλά με τα υπάρχοντα μη κλιμακοθετήσιμα συστήματα κωδικοποίησης βίντεο όσον αφορά την οπτική ποιότητα. Αυτό θα πρέπει να συμβαίνει για την πλειοψηφία των απεικονίσεων, των επιπέδων ποιότητας και των χρονικών ρυθμών πλαισίων. Μια άλλη απαίτηση αφορά τον χρόνο αποκωδικοποίησης. Γενικά για βίντεο ροοθήκευση (video-streaming) ο χρόνος αποκωδικοποίησης είναι της τάξης δευτερολέπτων ενώ για εφαρμογές όπως παρακολούθηση βίντεο (video monitoring) απαιτείται χρόνος της τάξης των 100 msec. Όσον αφορά τις απαιτήσεις για κλιμακοθετησιμότητα (scalability) του περιεχομένου βίντεο, το MPEG ορίζει κάποιον



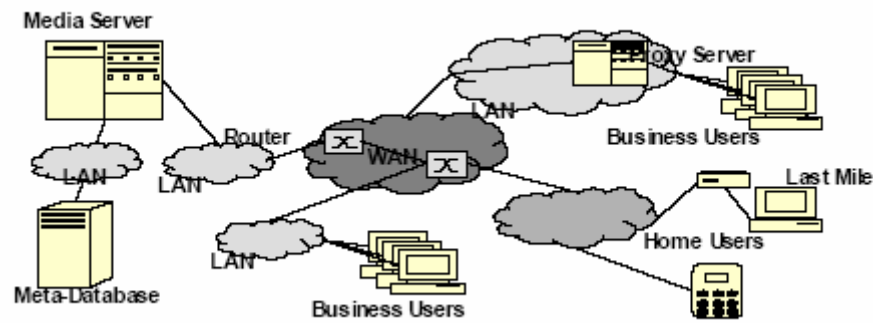
ελάχιστο αριθμό χωρικών, χρονικών και ποιοτικών επιπέδων που πρέπει να διατίθενται. Τέλος, θα πρέπει να λάβουμε υπόψη και την πολυπλοκότητα της κωδικοποίησης και αποκωδικοποίησης. Οι απαιτήσεις αυτές αναλύονται περισσότερο στις ακόλουθες υποενότητες.

### 1.1.1 Χωρική Κλιμακοθετησιμότητα (Spatial Scalability)

Η κλιμακοθετήσιμη κωδικοποίηση βίντεο πρέπει να υποστηρίζει το λιγότερο τέσσερα επίπεδα χωρικής ανάλυσης (ευκρίνειας) τα οποία κυμαίνονται από 1152 x 1408 έως 144 x 176 εικονοστοιχεία. Στο εύρος αυτό περιέχονται και τα δύο πολύ γνωστά πρότυπα για αναπαράσταση βίντεο: CIF (288 x 352) και QCIF (144 x 176). Η ικανότητα να παράγουμε πολλαπλές αναλύσεις από μία κοινή ακολουθία ψηφίων είναι σημαντική για αρκετές εφαρμογές. Μια από αυτές είναι η επιτήρηση χώρου μέσω βίντεο και τα συστήματα παρακολούθησης, όπου η λειτουργία χωρικής κλιμακοθετησιμότητας έχει δύο λόγους ύπαρξης [1].

- Στο πρώτο σενάριο πολλές λήψεις από διαφορετικές θέσεις καταφτάνουν στον σταθμό επιτήρησης. Βασιζόμενοι στη δυνατότητα χωρικής κλιμακοθετησιμότητας κάθε λήψη μπορεί να μεγενθυθεί. Έτσι μειώνεται το εύρος επικοινωνίας μεταξύ του δικτύου καμερών και του σταθμού παρακολούθησης και το υψηλής ευκρίνειας περιεχόμενο μεταδίδεται μόνο έπειτα από συναγερμό ή από επιλογή του χρήστη.
- Υποστήριξη αποθήκευσης βίντεο με λειτουργία διαγραφής. Στα συστήματα παρακολούθησης η σημαντικότητα των καταγεγραμμένων δεδομένων και κατ' επέκταση της απαιτούμενης ανάλυσης των ακολουθιών βίντεο, μειώνεται με το πέρασμα του χρόνου. Για παράδειγμα, το πλήρες ευκρίνειας βίντεο χρειάζεται να αποθηκευτεί μόνο για τρεις ημέρες. Οι ακολουθίες που είναι για χρονικό διάστημα μεγαλύτερο από μια εβδομάδα χρειάζονται μια ελάχιστη ευκρίνεια για την αρχειοθέτησή τους.

Τέλος, μια μεγάλη κατηγορία συστημάτων που απαιτούν κλιμακοθετησιμότητα στην ανάλυση είναι οι αρχιτεκτονικές επικοινωνίας που χρησιμοποιούν το πλαίσιο εργασίας MPEG-21 Digital Item Adaptation (DIA), όπου ετερογενείς υποδομές επικοινωνίας χρησιμοποιούνται για video streaming προς ένα πλήθος συσκευών όπως απεικονίζεται στην Εικόνα 1.2. Σε αυτήν την περίπτωση, η προσαρμογή της ευκρίνειας βίντεο είναι σημαντική για να ικανοποιηθεί το εύρος ζώνης του καναλιού και η ευκρίνεια του αποκωδικοποιητή βίντεο.



Εικόνα 1.2 Προσαρμογή σε ετερογενή συσκευές και υποδομές δικτύων με το πρότυπο MPEG-21 DIA

### 1.1.2 Ποιοτική Κλιμακοθετησιμότητα (Quality Scalability)

Πρακτικά η ιδιότητα η ακολουθία που προκύπτει να είναι κλιμακοθετήσιμη στην ποιότητα σημαίνει να έχουμε κλιμακοθετησιμότητα στο ρυθμό δυαδικών ψηφίων (bit rate). Σύμφωνα με τις απαιτήσεις του MPEG τα κλιμακοθετήσιμα συστήματα κωδικοποίησης πρέπει να παρέχουν μια μεγάλη ποικιλία ρυθμών μετάδοσης δυαδικών ψηφίων που αντιστοιχούν στους ρυθμούς μετάδοσης που υποστηρίζουν τα σημερινά δίκτυα. Ένα πρακτικό παράδειγμα εύρους ρυθμών δυαδικών ψηφίων φαίνεται στην Εικόνα 1.1. Πρέπει επίσης να αναφερθεί ότι για συγκεκριμένες εφαρμογές αρχειοθέτησης βίντεο, όπως η βιομηχανία ψηφιακού σινεμά, η εξερεύνηση του διαστήματος ή κάποιες ιατρικές εφαρμογές, η δυνατότητα για κωδικοποίηση χωρίς απώλειες (lossless) πρέπει να υποστηρίζεται.

Η λειτουργία κλιμακοθετησιμότητας στην παρεχόμενη ποιότητα είναι πολύ χρήσιμη όταν η μετάδοση λαμβάνει χώρα σε μη αξιόπιστα δίκτυα, όπου οι αλλαγές στο εύρος ζώνης μπορούν να αντιμετωπιστούν αποτελεσματικά αλλάζοντας κατά τη μετάδοση (on-the-fly) την ποιότητα του μεταδιδόμενου βίντεο. Μία άλλη εφαρμογή έχουμε στη διανομή περιεχομένου, όπου ο χρήστης χρεώνεται ανάλογα με την ποιότητα που επιθυμεί και έτσι με την κλιμακοθετήσιμη κωδικοποίηση από μια ακολουθία μπορούν να προκύψουν τα κατάλληλα επίπεδα ποιότητας που θα μεταδοθούν στους αντίστοιχους χρήστες [1].

### 1.1.3 Χρονική Κλιμακοθετησιμότητα (Temporal Scalability)

Η χρονική κλιμακοθετησιμότητα της ακολουθίας βίντεο ισοδυναμεί με τη δυνατότητα κλιμακοθετησιμότητας του ρυθμού πλαισίων (frame rate). Ο ρυθμός πλαισίων του βίντεο αντιστοιχεί στον αριθμό των πλαισίων που προβάλλονται στη διάρκεια ενός

δευτερολέπτου. Έτσι, αλλάζοντας το ρυθμό των πλαισίων στην πραγματικότητα αλλάζουμε την ποιότητα του βίντεο που αντιλαμβάνεται το ανθρώπινο μάτι. Ο πιο συνηθισμένος ρυθμός πλαισίων είναι 30 Hz. Η κλιμακοθετήσιμη κωδικοποίηση βίντεο πρέπει να παρέχει τουλάχιστον τέσσερα επίπεδα χρονικής κλιμακοθετησιμότητας με ρυθμό αποκωδικοποίησης πλαισίων έως και 60 Hz [1].

Η κύρια εφαρμογή της χρονικής κλιμακοθετησιμότητας είναι η παραγωγή και διανομή πολυκαναλικού περιεχομένου, όπου η ίδια ακολουθία βίντεο προορίζεται για ένα πλήθος συσκευών που υποστηρίζουν διαφορετικές χρονικές αναλύσεις. Για παράδειγμα, 7.5 Hz, 15 Hz, 30 Hz, και 60 Hz πρέπει να υποστηρίζονται ώστε να εξυπηρετηθούν ποικίλοι χρήστες, από συσκευές स्टούντιο μέχρι βίντεο κατά απαίτηση (video-on-demand) σε κινητά τηλέφωνα με μικρές επεξεργαστικές ικανότητες. Επιπλέον, η χρονική κλιμακοθετησιμότητα μπορεί να χρησιμοποιηθεί για να εξομοιώσει λειτουργίες fast forward/backward που συναντούμε στις αναλογικές συσκευές, όπως τα βίντεο. Τέλος, η χρονική κλιμακοθετησιμότητα βρίσκει εφαρμογή σε εφαρμογές παρακολούθησης, όπου οι κάμερες καταγράφουν στατικές εικόνες και έτσι βίντεο υψηλού χρονικού ρυθμού πλαισίων είναι απαραίτητο μόνο έπειτα από κάποιο συναγερμό.

#### **1.1.4 Κλιμακοθετησιμότητα στην Πολυπλοκότητα (Complexity Scalability)**

Μια άλλη σημαντική λειτουργία είναι η κλιμακοθετησιμότητας στην πολυπλοκότητα. Είναι επιθυμητό η πολυπλοκότητα αποκωδικοποίησης να μεταβάλλεται ανάλογα με την χρονική και χωρική ανάλυση. Επίσης, ανάλογα με τους διαθέσιμους πόρους στην πλατφόρμα κωδικοποίησης θα πρέπει να συμβαίνει κατά τη μετάδοση (on-the-fly) προσαρμογή της ακολουθίας δυαδικών ψηφίων, ώστε να επιλέγεται η υποακολουθία που οδηγεί σε μικρής πολυπλοκότητας αποκωδικοποίηση. Η πολυπλοκότητα σχετίζεται με αλγοριθμικά χαρακτηριστικά και με την πλατφόρμα υλοποίησης, οπότε για να έχουμε αξιόπιστα αποτελέσματα θα πρέπει να χρησιμοποιηθούν σωστά μοντέλα πολυπλοκότητας.

Σημαντικές εφαρμογές της κλιμακοθετησιμότητας στην πολυπλοκότητα έχουμε στην περιοχή της ασύρματης βίντεο ροοθήκευσης (video streaming) σε κινητές συσκευές, όπως κινητά τηλέφωνα και PDAs. Επιπλέον, ανάλογα με την εφαρμογή, μπορεί να είναι απαραίτητο η κωδικοποίηση να πληρεί ορισμένα όρια πολυπλοκότητας. Αυτό είναι σημαντικό σε σενάρια όπου οι συσκευές κωδικοποίησης διαμοιράζονται στο δίκτυο, όπως στις εφαρμογές παρακολούθησης.

## Κεφάλαιο 2 : Ο Μετασχηματισμός wavelet

Στο κεφάλαιο αυτό θα μελετήσουμε το κυματιδιακό μετασχηματισμό (wavelet transform), ο οποίος αποτελεί το βασικό εργαλείο για την κλιμακοθετήσιμη κωδικοποίηση τόσο εικόνων όσο και ακολουθιών βίντεο. Θα μελετήσουμε το συνεχή μετασχηματισμό wavelet, την ιδέα της ανάλυσης πολλαπλής διακριτικής ικανότητας (multiresolution analysis) καθώς και την υλοποίηση του μετασχηματισμού wavelet με χρήση του σχήματος ανόρθωσης (lifting scheme), το οποίο προκάλεσε νέο ενδιαφέρον στο χώρο της κλιμακοθετήσιμης κωδικοποίησης βίντεο. Για να γίνει όμως κατανοητός ο μετασχηματισμός wavelet, θα πρέπει πρώτα να παρουσιάσουμε το μαθηματικό υπόβαθρο στο οποίο στηρίζεται κάθε επέκταση διακριτού σήματος και να δούμε πως η επέκταση αυτή μπορεί να υλοποιηθεί με τη χρήση μιας τράπεζα φίλτρων κάνοντας ευκολότερα υλοποιήσιμη υπολογιστικά την επέκταση.

### 2.1 Μαθηματικό Υπόβαθρο

#### 2.1.1 Επέκταση Διακριτών Σημάτων

Στην παράγραφο αυτή θα ασχοληθούμε με την γραμμική επέκταση σημάτων ή συναρτήσεων. Δοσμένου δηλαδή ενός σήματος  $x$  από κάποιο χώρο  $S$ , όπου ο  $S$  μπορεί να είναι είτε πεπερασμένης διάστασης (π.χ  $\mathbb{R}^n$ ,  $\mathbb{C}^n$ ) είτε όχι ( $l_2(\mathbb{Z})$ ,  $L_2(\mathbb{R})$ ), θέλουμε να βρούμε ένα σύνολο βασικών σημάτων (ή αλλιώς συναρτήσεις βάσης)  $\{\varphi_i\}_{i \in \mathbb{Z}}$  για το χώρο αυτό ώστε να γράψουμε το  $x$  ως γραμμικό συνδυασμό

$$x = \sum_i a_i \varphi_i \quad (2.1)$$

Επιπλέον, επιθυμούμε το σύνολο  $\{\varphi_i\}_{i \in \mathbb{Z}}$  να είναι πλήρες για το χώρο  $S$ , δηλαδή κάθε σήμα  $x$  που ανήκει στο χώρο αυτό να μπορεί να αναλυθεί όπως στην παραπάνω σχέση. Στην περίπτωση αυτή θα υπάρχει επίσης και ένα δυειδές (dual) σύνολο  $\{\tilde{\varphi}_i\}_{i \in \mathbb{Z}}$  έτσι ώστε οι συντελεστές επέκτασης της σχέσης (2.1) να μπορούν να υπολογιστούν ως

$$a_i = \sum_n \tilde{\varphi}_i[n] x[n],$$

όταν τα  $x$  και  $\tilde{\varphi}_i$  είναι διακριτές ακολουθίες και ως

$$a_i = \int \tilde{\varphi}_i(t)x(t)dt ,$$

όταν είναι συνεχείς χρόνου ακολουθίες. Παρατηρούμε ότι οι παραπάνω σχέσεις είναι στην πραγματικότητα το εσωτερικό γινόμενο του σήματος  $x$  με τα  $\tilde{\varphi}_i$ , το οποίο συμβολίζεται  $\langle \tilde{\varphi}_i, x \rangle$ . Δύο σημαντικές περιπτώσεις έχουμε όταν το σύνολο  $\varphi_i$  είναι ορθοκανονικό ή biorthogonal [2] [1].

Όταν έχουμε μια ορθοκανονική βάση για το σύνολο  $S$  τότε το σύνολο  $\varphi_i$  και το δυειδές του ταυτίζονται, δηλαδή  $\varphi_i = \tilde{\varphi}_i$ . Έτσι ισχύει,

$$\langle \varphi_i, \varphi_j \rangle = \delta[i - j] ,$$

όπου  $\delta[i] = 1$  αν  $i = 0$  αλλιώς είναι 0 παντού, (η γνωστή συνάρτηση Δέλτα). Μία άλλη σημαντική ιδιότητα στην περίπτωση αυτή είναι η διατήρηση της ενέργειας, δηλαδή ισχύει  $\|x\|^2 = \|X\|^2$ . Με  $X$  συμβολίζουμε την ακολουθία των συντελεστών επέκτασης (μετασχηματισμού), δηλαδή  $X[i] = a_i$ .

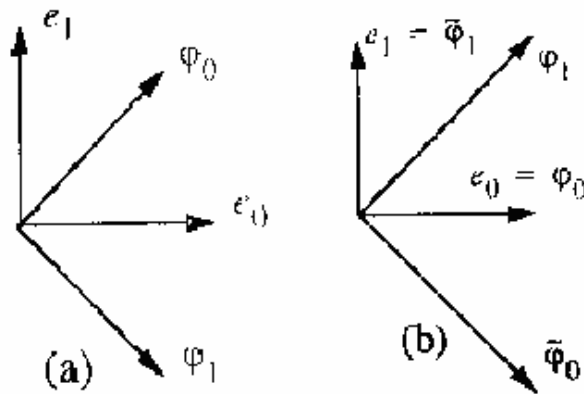
Αν τώρα έχουμε biorthogonal επέκταση, τότε οι συναρτήσεις  $\varphi_i$  και  $\tilde{\varphi}_i$  ικανοποιούν την σχέση

$$\langle \varphi_i, \tilde{\varphi}_j \rangle = \delta[i - j].$$

Στην περίπτωση αυτή δεν ισχύει η διατήρηση της ενέργειας. Για να έχουμε όμως μια ευσταθή επέκταση, οι συντελεστές της επέκτασης πρέπει να ικανοποιούν την σχέση

$$A \sum_i |a_i|^2 \leq \|x\|^2 \leq B \sum_i |a_i|^2 ,$$

για κάποιες θετικές σταθερές  $A, B$ . Οι δύο περιπτώσεις απεικονίζονται στην Εικόνα 2.1.



**Εικόνα 2.1** Παραδείγματα διανυσμάτων για επέκταση του χώρου  $R^2$ . (a) Ορθοκανονική περίπτωση  
(b) Biorthogonal περίπτωση

Όπως αναφέραμε προηγουμένως το σήμα  $x$  το οποίο και επιθυμούμε να αναλύσουμε σύμφωνα με την σχέση (2.1) ανήκει σε κάποιο χώρο  $S$ . Ο χώρος αυτός στην περίπτωση που έχουμε διακριτές ακολουθίες  $x[n]$  είναι σχεδόν πάντα ο χώρος  $l_2(Z)$ . Ο χώρος αυτός έχει άπειρη διάσταση και περιλαμβάνει εκείνες τις ακολουθίες για τις οποίες  $\|x\| < \infty$ . Ένας άλλος χώρος που θα αναφερθεί συχνά σε αυτή τη διπλωματική εργασία είναι ο  $L_2(R)$ . Μια συνάρτηση  $f(t)$  ορισμένη στο  $R$  ανήκει στο χώρο  $L_2(R)$  αν ισχύει  $\sqrt{\int_{t \in R} |f(t)|^2 dt} < \infty$ .

Ο πιο γνωστός ορθογώνιος μετασχηματισμός που χρησιμοποιείται σε δεκάδες εφαρμογές είναι ο γνωστός μας μετασχηματισμός Fourier. Οι συναρτήσεις βάσης του είναι τριγωνομετρικές συναρτήσεις και συγκεκριμένα οι  $\cos\left(\frac{2\pi nx}{l}\right), \{n | n \geq 0, n \in Z\}$  και

$\sin\left(\frac{2\pi nx}{l}\right), \{n | n \geq 0, n \in Z\}$ , όπου  $l$  είναι η περίοδος της συνάρτησης και  $x$  η μεταβλητή

της συνάρτησης. Παρατηρούμε ότι κάθε συνάρτηση βάσης είναι ορθογώνια σε κάθε άλλη συνάρτηση βάσης, οπότε έχουμε ένα ορθογώνιο σύνολο βάσεων όχι όμως και ορθοκανονικό, αφού  $\int_0^l \varphi^2(i) dx \neq 0$  και συγκεκριμένα  $\int_0^l \sin^2\left(\frac{2\pi nx}{l}\right) dx = \frac{l}{2}$ . Έτσι, αν

θέλουμε να περιγράψουμε μία συνάρτηση  $f(x)$  στο χώρο που δημιουργούν οι βάσεις του μετασχηματισμού Fourier, θα έχουμε

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left( a_n \cos\left(\frac{2\pi nx}{l}\right) + b_n \sin\left(\frac{2\pi nx}{l}\right) \right),$$

$$\text{όπου } a_0 = 2, a_n = \frac{2}{l} \int_0^l f(x) \cos\left(\frac{2\pi nx}{l}\right) dx \text{ και } b_n = \frac{2}{l} \int_0^l f(x) \sin\left(\frac{2\pi nx}{l}\right) dx$$

Οι όροι  $2/l$  χρησιμοποιούνται για την κανονικοποίηση των τιμών. Οι συναρτήσεις που δημιουργούν το μετασχηματισμό Fourier μπορούν προσεγγίσουν πολύ καλά κάθε συνάρτηση που ανήκει στο χώρο  $L_2(\mathbb{R})$ . Επιπλέον, χειρίζονται και τη μη συνέχεια των συναρτήσεων, αλλά όχι ιδανικά, ενώ δυσκολεύονται με τις πολύ fractal και ασυνεχείς συναρτήσεις. Αυτό το μειονέκτημα του μετασχηματισμού Fourier ήταν ο λόγος που οδηγήθηκαμε στην ανάπτυξη των wavelets. Το πρώτο wavelet ανακαλύφθηκε από τον Alfrad Haar το 1909 και αποτέλεσε την έμπνευση των μαθηματικών για τη δημιουργία και άλλων wavelets.

## Haar

Για να γίνει πιο κατανοητή η παρακάτω ανάλυση στην περίπτωση των wavelets, θα παρουσιάσουμε μία άλλη σημαντική περίπτωση επέκτασης διακριτών σημάτων, την Haar επέκταση [2]. Οι συναρτήσεις βάσης στην περίπτωση αυτή είναι οι

$$\varphi_{2k}[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = 2k, 2k+1, \\ 0 & \text{αλλιώς} \end{cases}, \quad \varphi_{2k+1}[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = 2k \\ -\frac{1}{\sqrt{2}} & n = 2k+1 \\ 0 & \text{αλλιώς} \end{cases}$$

Όπως είδαμε πριν οι συντελεστές επέκτασης θα δίνονται από τις σχέσεις

$$\begin{aligned} X[2k] &= \langle \varphi_{2k}, x \rangle = \frac{1}{\sqrt{2}} (x[2k] + x[2k+1]), \\ X[2k+1] &= \langle \varphi_{2k+1}, x \rangle = \frac{1}{\sqrt{2}} (x[2k] - x[2k+1]). \end{aligned} \quad (2.2)$$

Έτσι το σήμα  $x$  αναλύεται ως

$$x[n] = \sum_{k \in \mathbb{Z}} X[k] \varphi_k[n] \quad (2.3)$$

Τα  $\phi_k[n]$  στην Haar ανάλυση αποτελούν μια ορθοκανονική βάση για το χώρο  $l_2(\mathbb{Z})$ .

### 2.1.2 Τράπεζα Φίλτρων (Filter Bank)

Η επέκταση ενός σήματος χρησιμοποιώντας τις συναρτήσεις βάσης σχετίζεται άμεσα με την υλοποίηση μιας τράπεζας φίλτρων. Δηλαδή, μπορούμε να υλοποιήσουμε το μετασχηματισμό που δημιουργούν οι συναρτήσεις βάσης με φίλτρα [2]. Ας εξετάσουμε αρχικά την περίπτωση της Haar επέκτασης που είδαμε στην προηγούμενη παράγραφο. Ας θεωρήσουμε το φίλτρο με κρουστική απόκριση

$$h_0[n] = \begin{cases} \frac{1}{\sqrt{2}} & n = -1, 0 \\ 0 & \text{αλλού} \end{cases} .$$

Τότε οι συντελεστές επέκτασης  $X[2k]$  της σχέσης (2.2) είναι το αποτέλεσμα της συνέλιξης του  $h_0[n]$  με το  $x[n]$  τις χρονικές στιγμές  $2k$  αφού

$$h_0[n] * x[n] |_{n=2k} = \sum_{l \in \mathbb{Z}} h_0[2k-l]x[l] = \frac{1}{\sqrt{2}}x[2k] + \frac{1}{\sqrt{2}}x[2k+1] = X[2k].$$

Όμοια αν ορίσουμε το φίλτρο  $h_1[n]$  με κρουστική απόκριση

$$h_1[n] = \begin{cases} \frac{1}{\sqrt{2}}, & n = 0 \\ -\frac{1}{\sqrt{2}}, & n = -1 \\ 0, & \text{αλλού} \end{cases}$$

τότε προκύπτουν οι περιττοί συντελεστές  $X[2k+1]$  από την ακόλουθη συνέλιξη:

$$h_1[n] * x[n] |_{n=2k} = \sum_{l \in \mathbb{Z}} h_1[2k-l]x[l] = \frac{1}{\sqrt{2}}x[2k] - \frac{1}{\sqrt{2}}x[2k+1] = X[2k+1].$$

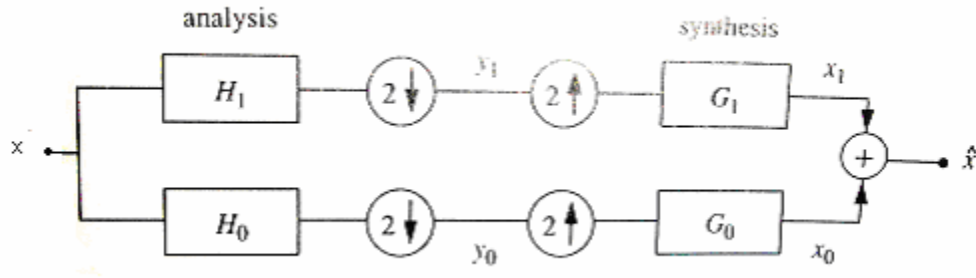
Γνωρίζουμε όμως ότι το να υπολογίσουμε μια συνέλιξη σε ζυγά στιγμιότυπα αντιστοιχεί στο να περάσουμε την ακολουθία από το φίλτρο και έπειτα να κάνουμε υποδειγματοληψία κατά δύο. Έτσι οι ακολουθίες  $X[2k]$  και  $X[2k+1]$  προκύπτουν από μία τράπεζα φίλτρων με φίλτρα  $h_0[n]$ ,  $h_1[n]$  ακολουθούμενα από υποδειγματοληψία κατά



δύο, όπως φαίνεται στο αριστερό μέρος της Εικόνα 2.2. Ονομάζουμε το μέρος αυτό ανάλυση τράπεζας φίλτρων . Είναι σημαντικό να αναφέρουμε ότι οι κρουστικές αποκρίσεις των φίλτρων ανάλυσης είναι οι χρονικά αντίθετες εκδοχές των συναρτήσεων βάσης, δηλαδή

$$h_0[n] = \varphi_0[-n] \quad h_1[n] = \varphi_1[-n]$$

και τα φίλτρα είναι μη αιτιατά. Επιπλέον, η ανάλυση είναι αμετάβλητη στις μετατοπίσεις (shift invariant) μόνο για ζυγό αριθμό μετατόπισης.



Εικόνα 2.2 Τράπεζα φίλτρων με φίλτρα ανάλυσης  $h_0[n]$ ,  $h_1[n]$  και φίλτρα σύνθεσης  $g_0[n]$ ,  $g_1[n]$

Αφού υλοποιήσαμε την ανάλυση θέλουμε τώρα να επανασηματίσουμε το αρχικό σήμα. Ορίζουμε δύο φίλτρα  $g_0$ ,  $g_1$  με κρουστικές αποκρίσεις  $\varphi_0$  και  $\varphi_1$  αντίστοιχα. Οπότε,

$$\begin{aligned} g_0[n] &= \varphi_0[n], & g_1[n] &= \varphi_1[n] \\ \varphi_{2k}[n] &= g_0[n-2k], & \varphi_{2k+1}[n] &= g_1[n-2k] \end{aligned}$$

Από την σχέση ανακατασκευής (2.3) έχουμε

$$x[n] = \sum_{k \in \mathbb{Z}} X[2k] \varphi_{2k}[n] + \sum_{k \in \mathbb{Z}} X[2k+1] \varphi_{2k+1}[n] = \sum_{k \in \mathbb{Z}} X[2k] g_0[n-2k] + \sum_{k \in \mathbb{Z}} X[2k+1] g_1[n-2k] \quad (2.4)$$

Παρατηρούμε ότι κάθε συντελεστής  $X[2k]$  ( $X[2k+1]$ ) προσθέτει ένα αντίγραφο της κρουστικής απόκρισης  $g_0[n]$  ( $g_1[n]$ ) μετατοπισμένο κατά  $2k$ . Αυτό μπορεί να υλοποιηθεί με μια υπερδειγματοληψία κατά δύο, ακολουθούμενη από συνέλιξη με το φίλτρο  $g_0[n]$  ( $g_1[n]$ ). Η διαδικασία αυτή φαίνεται στο δεξιό μέρος της Εικόνας 2.2 και ονομάζεται σύνθεση τράπεζας φίλτρων.

Αν μελετήσουμε την παραπάνω ανάλυση (2.2) από την πλευρά της επεξεργασίας σήματος, βλέπουμε ότι το ένα κανάλι υπολογίζει τη μέση τιμή και το άλλο τη διαφορά δύο διαδοχικών δειγμάτων. Τα φίλτρα αυτά δεν είναι τα καλύτερα δυνατά

χαμηλοδιαβατά και υψηλοδιαβατά φίλτρα, έχουν όμως καλό χρονικό εντοπισμό. Από την σχέση (2.4) βλέπουμε ότι το πρώτο άθροισμα είναι η ορθογώνια προβολή της εισόδου στον υπόχωρο που δημιουργείται από τα  $\varphi_{2k}[n]$ , ή διαφορετικά μία μέση ή κατώτερης ποιότητας εκδοχή του  $x[n]$ . Αν ονομάσουμε το σήμα που προκύπτει  $x_0[n]$  τότε  $x_0[2k] = x_0[2k+1] = \frac{1}{2}(x[2k] + x[2k+1])$ . Αντίστοιχα, το άλλο άθροισμα είναι η ορθογώνια προβολή στο χώρο που δημιουργείται από τα  $\varphi_{2k+1}[n]$ . Ονομάζοντας το σήμα που προκύπτει  $x_1[n]$ , έχουμε  $x_1[2k] = \frac{1}{2}(x[2k] - x[2k+1])$ ,  $x_1[2k+1] = -x_1[2k]$ . Το σήμα αυτό αντιστοιχεί στη διαφορά ή λεπτομέρεια που είναι απαραίτητη για να κατασκευάσουμε το  $x[n]$  από την κατώτερη ποιότητας εκδοχή  $x_0[n]$ . Το άθροισμα τους ανακατασκευάζει ορθά το  $x[n]$  αφού το άθροισμα  $x_0[2k] + x_1[2k]$  μας δίνει το  $x[2k]$  και το άθροισμα  $x_0[2k+1] + x_1[2k+1]$  το  $x[2k+1]$ .

Είδαμε λοιπόν πως μπορούμε να υλοποιήσουμε την Haar επέκταση με μια τράπεζα φίλτρων. Επαναλαμβάνουμε δύο χαρακτηριστικά της, τα οποία και ισχύουν στην περίπτωση που έχουμε οποιαδήποτε άλλη ορθογώνια επέκταση.

1. Οι κρουστικές αποκρίσεις των φίλτρων σύνθεσης είναι ίσες με το πρώτο σύνολο των συναρτήσεων βάσης

$$g_i[n] = \varphi_i[n], \quad i = 0, 1.$$

2. οι κρουστικές αποκρίσεις των φίλτρων ανάλυσης είναι οι χρονικά αντίθετες εκδοχές των φίλτρων σύνθεσης

$$h_i[n] = g_i[-n], \quad i = 0, 1.$$

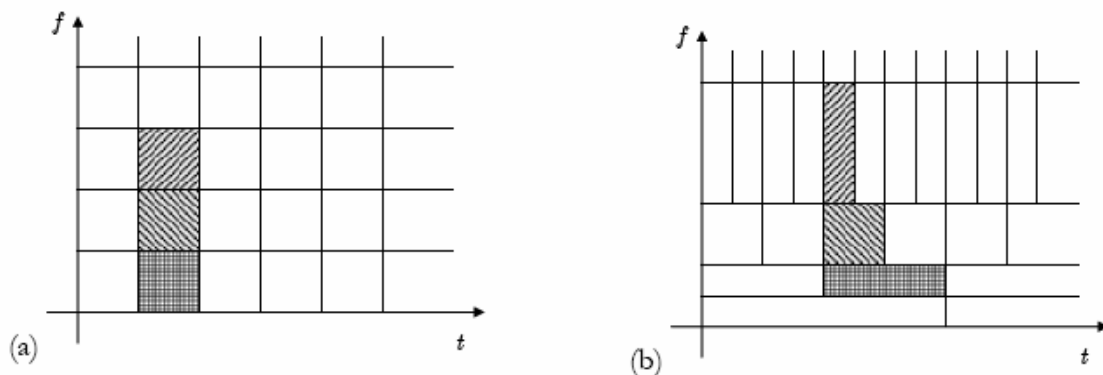
Έστω, τώρα ότι δεν έχουμε ορθοκανονική επέκταση όπως στην προηγούμενη περίπτωση αλλά biorthogonal και θέλουμε να δημιουργήσουμε μια τράπεζα φίλτρων που να υλοποιεί την επέκταση. Ονομάζουμε  $h_0[n]$ ,  $h_1[n]$  τα φίλτρα ανάλυσης και  $g_0[n]$ ,  $g_1[n]$  τα φίλτρα σύνθεσης που υλοποιούν την τέλεια ανακατασκευή. Αν κάνουμε την ανάλυση με το μετασχηματισμό  $z$  καταλήγουμε ότι οι συνθήκες που πρέπει να ισχύουν για τέλεια ανακατασκευή είναι

$$G_0(z)H_0(z) + G_1(z)H_1(z) = 2$$

$$G_0(z)H_0(-z) + G_1(z)H_1(-z) = 0$$

## 2.2 Ο Μετασχηματισμός Wavelet

Όπως αναφέραμε στην προηγούμενη παράγραφο οι μετασχηματισμοί wavelets ανακαλύφθηκαν για να αντιμετωπιστούν τα μειονεκτήματα του μετασχηματισμού Fourier. Το ένα μειονέκτημα είναι ότι δεν μπορούν οι τριγωνομετρικές συναρτήσεις βάσης να προσεγγίσουν επιτυχώς κάποιες ασυνεχείς ή fractal συναρτήσεις. Το δεύτερο μειονέκτημα είναι ότι με το μετασχηματισμό Fourier μπορούμε να πάρουμε χρήσιμες πληροφορίες για το συχνοτικό περιεχόμενο του σήματος, όχι όμως και για την χρονική στιγμή που εμφανίζονται οι συχνότητες. Για παράδειγμα, έστω ένα σήμα που αποτελείται από  $k$  ταλαντώσεις διαφορετικών συχνοτήτων στα χρονικά διαστήματα  $[i_0 \ i_1]$ , ...,  $[i_{k-1} \ i_k]$ . Ο μετασχηματισμός Fourier θα μας πληροφορήσει για την ύπαρξη των  $k$  συχνοτήτων, όχι όμως και για την χρονική στιγμή που υπήρχαν στο σήμα. Ένας τρόπος για να αντιμετωπίσουμε το πρόβλημα αυτό είναι να παραθυροποιήσουμε τις τριγωνομετρικές συναρτήσεις βάσης για να αποκτήσουμε έτσι δυνατότητα χρονικού εντοπισμού. Η μέθοδος αυτή αποτελεί τον Short-Time Fourier Transform (STFT), ο οποίος προτάθηκε αρχικά από τον Gabor. Με την χρήση του STFT το συχνοτικοχρονικό επίπεδο τεμαχίζεται ομοιόμορφα σε ένα πλέγμα τετραγώνων, Εικόνα 2.3(a). Η ανακάλυψη των wavelets επέφερε λύση τόσο στην προσέγγιση fractal συναρτήσεων όσο και στη δυνατότητα χρονικού εντοπισμού (localization). Το συχνοτικοχρονικό πλέγμα που δημιουργείται από τα wavelets είναι τελείως διαφορετικό και απεικονίζεται στην Εικόνα 2.3(b). Παρατηρούμε ότι η δυνατότητα χρονικού εντοπισμού είναι αντιστρόφως ανάλογη με το συχνοτικό εντοπισμό. Στις επόμενες παραγράφους θα αναλύσουμε τα wavelets και το μαθηματικό υπόβαθρο στο οποίο στηρίζονται και θα γίνει κατανοητός ο τρόπος με τον οποίο λαμβάνουμε την απεικόνιση της Εικόνας 2.3(b).



Εικόνα 2.3 Τεμαχισμός του συχνοτικοχρονικού επιπέδου (a) STFT (b) Wavelet μετασχηματισμός

### 2.2.1 Συνεχής Μετασχηματισμός Wavelet (CWT)

Αρχικά θα μελετήσουμε το συνεχή μετασχηματισμό wavelet. Ένας τρόπος για να πάρουμε αναπαράσταση όπως αυτή της Εικόνας 2.3 είναι να ορίσουμε για βάση την οικογένεια συναρτήσεων

$$\psi_{a,\tau} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right), \quad a > 0, \tau \in R \quad (2.5)$$

όπου  $\psi$  είναι μια δοσμένη συνάρτηση (η οποία ονομάζεται “mother wavelet”) που έχει καλή δυνατότητα εντοπισμού στον χρόνο και στην συχνότητα [1], [2], [16]. Με  $\tau$  συμβολίζουμε την μετατόπιση της συνάρτησης στον χρόνο και  $a$  είναι η παράμετρος κλίμακας (scale). Αυτές οι παράμετροι πρέπει να έχουν τέτοιες τιμές ώστε οι συναρτήσεις  $\psi_{a,\tau}(t)$  να έχουν πεπερασμένη νόρμα, δηλαδή

$$\|\psi_{a,\tau}(t)\|^2 = \int_{-\infty}^{+\infty} |\psi_{a,\tau}(t)|^2 dt = Q$$

Ο παράγοντας  $1/\sqrt{a}$  στην σχέση (2.5) χρησιμοποιείται για να διατηρηθεί η νόρμα όλων των συναρτήσεων  $\psi_{a,\tau}$  ίδια και ίση με τη νόρμα της  $\psi(t)$ . Επιπλέον, ο εντοπισμός (localization) της μητέρας συνάρτησης μπορεί να εξασφαλιστεί με το να την περιορίσουμε στον χρόνο και στην συχνότητα [2], δηλαδή

$$|\psi(t)| \leq c(1+|t|)^{-1-\varepsilon}, \quad |\Psi(\omega)| \leq c(1+|\omega|)^{-1-\varepsilon}$$

για κάποια  $c, \varepsilon > 0$ , όπου με  $\Psi(\omega)$  συμβολίζουμε το μετασχηματισμό Fourier της  $\psi(t)$ .

Ο Συνεχής Μετασχηματισμός Wavelet (Continuous Wavelet Transform) του σήματος  $f(t)$  ορίζεται ως

$$CWT_f(a, \tau) = \langle f, \psi_{a,\tau} \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(t) \overline{\psi\left(\frac{t-\tau}{a}\right)} dt \quad (2.6)$$

όπου με  $\bar{g}$  συμβολίζουμε τη συζυγή της συνάρτησης  $g$  και με  $\langle f, g \rangle$  το εσωτερικό γινόμενο των σημάτων ή συναρτήσεων  $f, g$ . Βλέπουμε ότι η σχέση (2.6) στην πραγματικότητα υπολογίζει την ομοιότητα μεταξύ του σήματος  $f(t)$  και ενός μέλους της

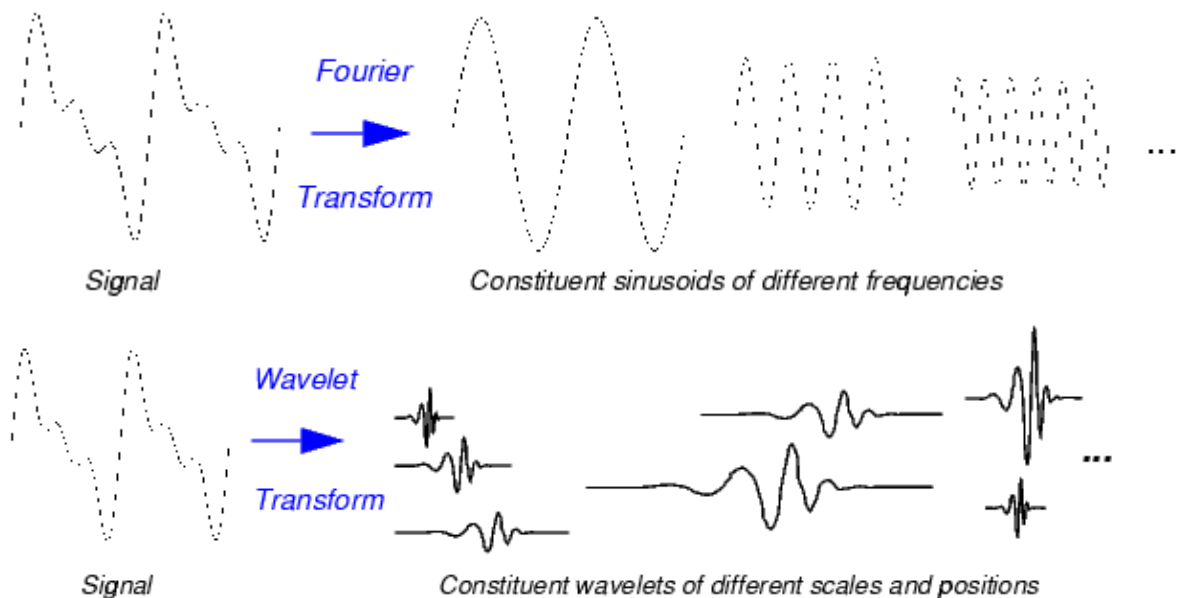
οικογένειας συναρτήσεων που ορίζεται από την σχέση (2.5). Για να εξασφαλίσουμε την ύπαρξη του αντίστροφου συνεχή μετασχηματισμού wavelet, θα πρέπει η συνάρτηση  $\psi$  να ικανοποιεί την ιδιότητα της αποδεκτικότητας (admissibility), η οποία εκφράζεται από την σχέση

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$$

Τότε η  $f(t)$  μπορεί να ανακατασκευαστεί από τον αντίστροφο συνεχή μετασχηματισμό wavelet (ICWT) ως

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} CWT_f(a, \tau) \psi_{a, \tau}(t) d\tau$$

Ερμηνεύοντας την παραπάνω σχέση βλέπουμε ότι κάθε σήμα  $f(t)$  μπορεί να γραφτεί ως υπέρθεση μετατοπισμένων και διαστελλομένων wavelets. Η ανακατασκευή δίνεται από το διπλό ολοκλήρωμα με το  $\tau \in R$  και  $a > 0$ . Ο μετασχηματισμός wavelet απεικονίζεται στην Εικόνα 2.4 και για να γίνει πιο κατανοητός παρουσιάζεται σε σύγκριση με το μετασχηματισμό Fourier.



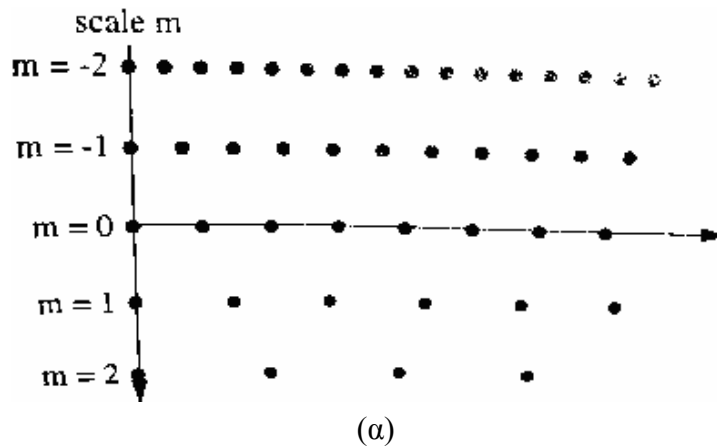
**Εικόνα 2.4** Σύγκριση μετασχηματισμού Fourier και μετασχηματισμού Wavelet

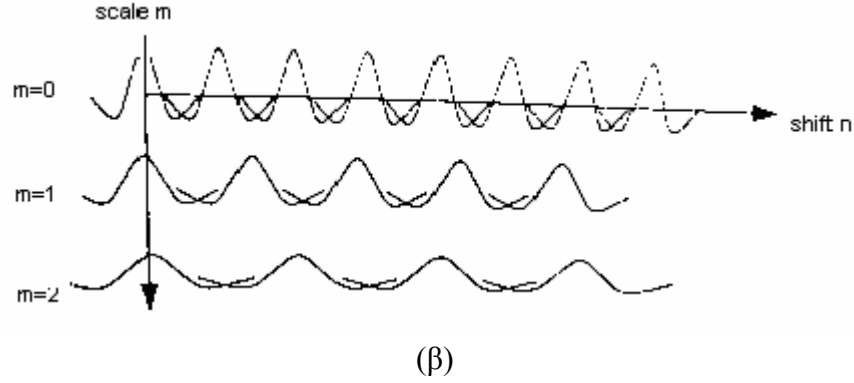
Παρόλο που ο ύπαρξη του CWT είναι πολύ σημαντική, στις σημερινές εφαρμογές έχουμε να αντιμετωπίσουμε κυρίως διακριτά σήματα και επομένως θέλουμε να

υλοποιήσουμε διακριτούς μετασχηματισμούς που να υλοποιούν ορθοκανονικές επεκτάσεις. Η Daubechies απέδειξε ότι είναι δυνατές αριθμητικά σταθερές ανακατασκευές [7] αν χρησιμοποιήσουμε για διακριτοποίηση των παραμέτρων κλίμακας και χρονικής μετατόπισης τις εξής τιμές:

$$a = a_0^m \quad \tau = n\tau_0 a_0^m, \quad m, n \in \mathbb{Z}, \quad \alpha_0 > 1, \tau_0 > 0$$

Αντικαθιστώντας στην σχέση (2.5) τις παραπάνω τιμές προκύπτει η διακριτοποιημένη οικογένεια των wavelets  $\psi_{m,n}(t) = a_0^{-m/2} \psi(a_0^{-m}t - n\tau_0)$  η οποία και αποτελεί τη βάση επέκτασης. Διαφορετικές τιμές του  $m$  αντιστοιχούν σε wavelets διαφορετικής κλίμακας, οπότε έχουν διαφορετικό πλάτος. Στενά, υψηλής συχνότητας wavelets έχουν μικρό χρονικό βήμα για να καλύψουν όλη την χρονική περιοχή, ενώ πιο πλατιά και χαμηλής συχνότητας wavelets έχουν μεγαλύτερα χρονικά βήματα.. Έτσι, προκύπτει και η ανάλυση που παρουσιάζεται στην Εικόνα 2.5. Αυτό είναι και το σημαντικό πλεονέκτημα των wavelets, ότι μπορούμε να λάβουμε ανάλυση πολλαπλής διακριτικής ικανότητας στο συχνοτικοχρονικό πλέγμα, σε αντίθεση με το μετασχηματισμό STFT όπου το συχνοτικοχρονικό πλέγμα τεμαχίζεται ομοιόμορφα. Για  $\alpha_0 = 2$ ,  $\tau_0 = 1$  προκύπτει η δυαδική περίπτωση για την οποία γνωρίζουμε ότι υπάρχουν ορθοκανονικές βάσεις και είναι δυνατή η τέλεια ανακατασκευή από τους συντελεστές μετασχηματισμού.





Εικόνα 2.5 Διακριτοποίηση των παραμέτρων κλίμακας και χρονικής μετατόπισης

$a = a_0^m$ ,  $\tau = n\tau_0 a_0^m$ . (α) Το πλέγμα που προκύπτει (β) Τα αντίστοιχα σύνολα συναρτήσεων

## 2.2.2 Ανάλυση Πολλαπλής Διακριτικής Ικανότητας (Multiresolution analysis)

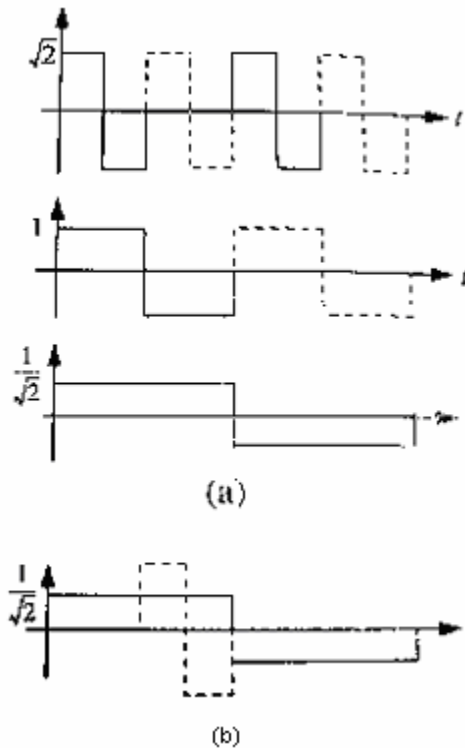
Όπως είδαμε στην προηγούμενη παράγραφο στόχος μας είναι να αναλύσουμε ένα σήμα  $f(t)$  με wavelet συναρτήσεις βάσης, όπου οι παράμετροι κλίμακας και χρονικής μετατόπισης έχουν διακριτοποιηθεί. Στην περίπτωση όμως που έχουμε μια διακριτή οικογένεια συναρτήσεων wavelet δεν είμαστε σίγουροι πάντα ότι μπορούμε από τους συντελεστές του μετασχηματισμού να ανακατασκευάσουμε το αρχικό σήμα με έναν αριθμητικά σταθερό τρόπο. Για να έχουμε σταθερότητα ανακατασκευής στο χώρο  $L_2(\mathbb{R})$  θα πρέπει για το σύνολο των συναρτήσεων βάσης, έστω  $\{g_n\}_{n \in \mathbb{N}}$ , να ισχύει η σχέση

$$c \sum_n |a[n]|^2 \leq \int_{-\infty}^{+\infty} |a[n]g_n(t)|^2 dt \leq C \sum_n |a[n]|^2 \quad (2.7)$$

όπου  $a[n]$  οι συντελεστές μετασχηματισμού [1], [2]. Στην περίπτωση που  $c = C = 1$  έχουμε μια ορθοκανονική βάση. Όπως αναφέρθηκε στην προηγούμενη παράγραφο, για  $a_0 = 2$ ,  $\tau_0 = 1$  και μια κατάλληλη μητέρα συνάρτηση wavelet προκύπτει ορθοκανονική επέκταση και μπορούμε να έχουμε τέλεια ανακατασκευή από τους συντελεστές μετασχηματισμού. Για παράδειγμα μια κατάλληλη μητέρα συνάρτηση αποτελεί το Haar wavelet. Στο σημείο αυτό θα πρέπει να αναφέρουμε ότι κάθε wavelet αποτελείται από δύο κομμάτια, από την συνάρτηση κλιμακοθέτησης (scaling function) που θα συμβολίζουμε με  $\varphi(t)$ , και τη «μητέρα» συνάρτηση (mother wavelet) που θα συμβολίζουμε με  $\psi(t)$ . Ο ρόλος κάθε μιας από τις συναρτήσεις αυτές θα γίνει κατανοητός παρακάτω. Στην περίπτωση του Haar wavelet η μητέρα συνάρτηση δίνεται από τον τύπο

$$\psi(t) = \begin{cases} 1, & t \in [0, 0.5) \\ -1, & t \in [0.5, 1) \\ 0, & \text{αλλού} \end{cases}$$

οπότε έχουμε την οικογένεια συναρτήσεων  $\psi_{m,n}(t) = 2^{-m/2} \psi(2^{-m}t - n\tau_0)$ ,  $m, n \in \mathbb{Z}$ . Έτσι, προκύπτουν οι συναρτήσεις βάσης της Εικόνας 2.6(α). Στην Εικόνα 2.6(β) βλέπουμε σχηματικά ότι οι συναρτήσεις διαφορετικής κλίμακας είναι ορθογώνιες μεταξύ τους αφού το εσωτερικό τους γινόμενο ισούται με τη μέση τιμή της μικρότερης συνάρτησης.

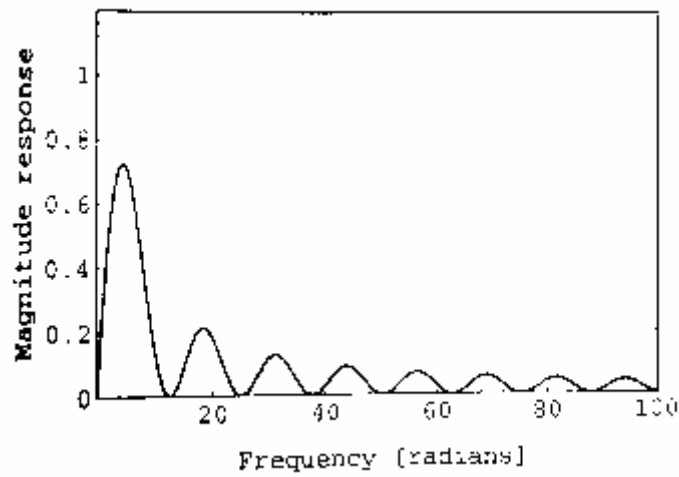
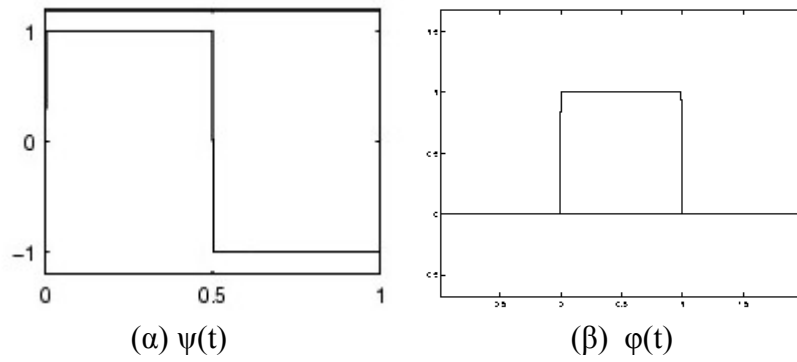


Εικόνα 2.6 (α) Haar συναρτήσεις βάσης (β) Ορθογωνιότητα των συναρτήσεων.

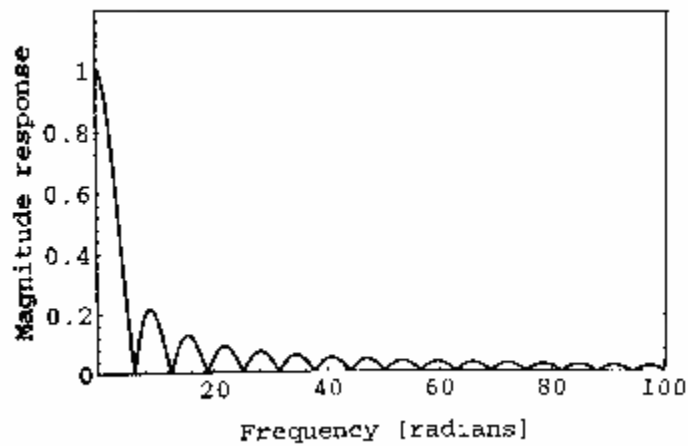
Η αντίστοιχη συνάρτηση κλιμακοθέτησης στο Haar wavelet είναι η  $\varphi(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{αλλού} \end{cases}$ .

Στην Εικόνα 2.7 απεικονίζονται οι συναρτήσεις  $\psi(t)$ ,  $\varphi(t)$  μαζί με το πλάτος του συχνοτικού τους περιεχομένου  $|\Psi(\omega)|$  και  $|\Phi(\omega)|$  αντίστοιχα.





(γ)  $|\Psi(\omega)|$



(δ)  $|\Phi(\omega)|$

**Εικόνα 2.7** Haar wavelet (α) Μητέρα συνάρτηση (β) Συνάρτηση κλιμακοθέτησης (γ) Συχνοτικό περιεχόμενο της μητέρας συνάρτησης (δ) Συχνοτικό περιεχόμενο της συνάρτησης κλιμακοθέτησης.

Όμως οι wavelet συναρτήσεις βάσης που προκύπτουν από την Haar συνάρτηση έχουν το εξής μειονέκτημα για πολλές εφαρμογές. Λόγω του ότι η  $\psi(t)$  δεν είναι συνεχής, οι wavelet συναρτήσεις που προκύπτουν δεν δίνουν καλές προσεγγίσεις για συναρτήσεις που είναι ομαλές (“smooth”). Αυτό το πρόβλημα απασχόλησε πολλούς ερευνητές στις αρχές ανάπτυξης της θεωρίας wavelet. Σκοπός ήταν να βρεθούν συστήματα που να έχουν και την πολυκλιμακωτή (multiscale) δομή της Haar βάσης και η μητέρα wavelet  $\psi(t)$  να είναι πιο ομαλή. Μια στρατηγική που επιτυγχάνει αυτόν το στόχο είναι η ανάλυση πολλαπλής διακριτικής ικανότητας (multiresolution analysis) που αναπτύχθηκε από τους Mallat και Meyer και αποτελεί όχι μόνο εργαλείο εύρεσης συναρτήσεων wavelet αλλά και ένα σημαντικό μαθηματικό πλαίσιο εργασίας. Επιπλέον, η ανάλυση πολλαπλής διακριτικής ικανότητας οδηγεί στη δημιουργία ορθοκανονικών βάσεων.

Η κύρια ιδέα της ανάλυσης πολλαπλής διακριτικής ικανότητας (multiresolution analysis) είναι να αναλύσουμε το σήμα σε μια σειρά διαδοχικών προσεγγίσεων. Το αρχικό σήμα αναπαρίσταται από μια κατώτερης ποιότητας προσέγγιση συν ένα σήμα λεπτομέρειας. Οι δυο υπόχωροι που ανήκουν τα δύο αυτά σήματα είναι κάθετοι μεταξύ τους. Με άλλα λόγια το σήμα της λεπτομέρειας είναι η διαφορά του αρχικού σήματος από την εκδοχή κατώτερης ποιότητας. Έπειτα, επαναλαμβάνουμε τη διαδικασία στο κατώτερης ποιότητας σήμα και παίρνουμε ένα σήμα ακόμα κατώτερης ποιότητας συν το σήμα λεπτομέρειας. Τελικά, προκύπτει ότι κάθε σήμα στο χώρο  $L_2(\mathbb{R})$  μπορεί να παραχθεί από τους χώρους των διαδοχικών λεπτομερειών. Η ακολουθία των υποχώρων προσέγγισης  $\{V_m\}_{m \in \mathbb{Z}} \subset L_2(\mathbb{R})$  που προκύπτει από την ανάλυση πολλαπλής διακριτικής ικανότητας ικανοποιεί τις ακόλουθες τρεις απαιτήσεις [1], [2], [16]. Αναφέρουμε ότι όσο το  $m$  αυξάνεται προκύπτουν υπόχωροι κατώτερης ποιότητας.

1. Οι χώροι  $V_m$  παράγονται από μια συνάρτηση κλιμακοθέτησης (scaling function)  $\varphi \in L_2(\mathbb{R})$ . Δηλαδή, για κάθε  $m$  η οικογένεια συναρτήσεων  $\varphi_{m,n}(t) = 2^{-m/2} \varphi\left(\frac{t-n}{2^m}\right), n \in \mathbb{Z}$  παράγει το χώρο  $V_m$  και ικανοποιεί την συνθήκη της σχέσης (2.7). Δηλαδή, οι συναρτήσεις  $\varphi_{m,n}(t)$  αποτελούν μια ορθοκανονική βάση στον χώρο  $V_m$ .
2. Οι υπόχωροι είναι εμφωλιασμένοι, δηλαδή  $\forall m \in \mathbb{Z}, V_{m+1} \subset V_m$ .
3. Η ορθογώνια προβολή  $P_{V_m}$  στο χώρο  $V_m$  ικανοποιεί τις σχέσεις  $\lim_{m \rightarrow +\infty} P_{V_m} f = 0$  και  $\lim_{m \rightarrow -\infty} P_{V_m} f = f$  για κάθε  $f \in L_2(\mathbb{R})$ .

Οι συναρτήσεις  $\varphi_{m,n}$  είναι αυτές που παράγουν το χώρο  $V_m$ , ενώ το σήμα λεπτομέρειας παράγεται από τις συναρτήσεις  $\psi_{m,n}$ . Από τις παραπάνω απαιτήσεις προκύπτει ότι αν  $f(t) \in V_m$  τότε  $f(\frac{t}{2}) \in V_{m+1}$  και ότι ο υπόχωρος  $V_m$  είναι αμετάβλητος για μετατοπίσεις  $2^m$ . Όπως είδαμε στην πρώτη απαίτηση, η προσέγγιση της  $f(t)$  στον υπόχωρο  $V_m$  υπολογίζεται ως η ορθογώνια προβολή του σήματος στη βάση που δημιουργούν οι συναρτήσεις  $\varphi_{m,n}$ . Δηλαδή,

$$\forall f(t) \in L^2(R), P_{V_m} f(t) = \sum_{n=-\infty}^{+\infty} \langle f, \varphi_{m,n} \rangle \varphi_{m,n} = \sum_{n=-\infty}^{+\infty} a_m[n] \varphi_{m,n}$$

Οι συντελεστές  $a_m[n]$  αποτελούν και τη διακριτή προσέγγιση της  $f(t)$  στον υπόχωρο  $V_m$ . Έστω  $W_m$  το ορθογώνιο συμπλήρωμα του  $V_m$  στον  $V_{m-1}$ , το οποίο ικανοποιεί την σχέση  $W_m \oplus V_m = V_{m-1}$ . Η διαφορά στην πληροφορία μεταξύ της προσέγγισης στον υπόχωρο  $V_{m-1}$  και τον  $V_m$  καθορίζεται με το να ορίσουμε μια ορθοκανονική βάση στον  $W_m$  και να προβάλλουμε το σήμα  $f(t)$  στη βάση αυτή. Η βάση αυτή δημιουργείται από τις μετατοπίσεις και διαστολές μίας συνάρτησης  $\psi(t)$ , την οποία και ονομάσαμε προηγουμένως «μητέρα» wavelet. Έτσι, η οικογένεια συναρτήσεων  $\psi_{m,n}(t) = 2^{-m/2} \psi(\frac{t-n}{2^m}), n \in Z$  αποτελεί ορθοκανονική βάση για το χώρο  $W_m$ . Αν προβάλλουμε την  $f(t)$  στη βάση αυτή παίρνουμε

$$\forall f(t) \in L^2(R), P_{W_m} f(t) = \sum_{n=-\infty}^{+\infty} \langle f, \psi_{m,n} \rangle \psi_{m,n} = \sum_{n=-\infty}^{+\infty} d_m[n] \psi_{m,n}$$

Σύμφωνα με ένα θεώρημα [2] όποτε ισχύουν οι τρεις απαιτήσεις που αναφέραμε, τότε οι συναρτήσεις  $\psi_{m,n}$  αποτελούν ορθοκανονική βάση για το χώρο  $L_2(R)$ . Εδώ έγκειται και η σημαντικότητα της ανάλυσης πολλαπλής διακριτικής ικανότητας (multiresolution analysis). Στο ότι μπορούμε να αναλύσουμε το αρχικό σήμα σε ένα σήμα κατώτερης ποιότητας συν ένα σήμα λεπτομέρειας και να επαναλάβουμε τη διαδικασία στο σήμα κατώτερης ποιότητας κ.ο.κ. Τελικά όμως, προκύπτει ότι όλα τα σήματα των λεπτομερειών είναι ικανά να παράγουν το αρχικό μας σήμα.

### Τράπεζα φίλτρων και wavelets

Στην παράγραφο 2.1.2 είδαμε πως μία επέκταση μπορεί να υλοποιηθεί με την χρήση τράπεζας φίλτρων. Αυτό μπορεί να γίνει και στην περίπτωση του μετασχηματισμού wavelet. Έτσι, οι συντελεστές  $a_m[n]$  και  $d_m[n]$  μπορούν να υπολογιστούν και με έναν διαφορετικό τρόπο, σύμφωνα με τον πυραμιδικό (pyramidal) αλγόριθμο. Ο αλγόριθμος

αυτός χρησιμοποιεί φίλτρα και υπολογίζει τα  $\alpha_m[n]$  αναδρομικά βασιζόμενος στο ότι οι χώροι είναι εμφωλιασμένοι, αφού  $\forall m \in Z, V_{m+1} \subset V_m$ . Άρα το  $\varphi(t/2)$  του υποχώρου  $V_{m+1}$  μπορεί να γραφτεί ως γραμμικός συνδυασμός της βάσης  $\varphi(t-n)$  του  $V_m$ . Αν ονομάσουμε  $h[n]$  τους συντελεστές αυτού του γραμμικού συνδυασμού τότε θα δίνονται από την σχέση

$$h[n] = \left\langle \frac{1}{\sqrt{2}} \varphi\left(\frac{t}{2}\right), \varphi(t-n) \right\rangle$$

Ισοδύναμα, αν μελετήσουμε την σχέση στο πεδίο συχνότητας τότε θα πρέπει να ισχύει  $\Phi(\omega) = \frac{1}{\sqrt{2}} H(e^{j\omega/2}) \Phi(\omega/2)$ , όπου  $H, \Phi$  οι μετασχηματισμοί Fourier των  $h[n]$  και  $\varphi(t)$  αντίστοιχα. Αν λοιπόν έχουμε ένα φίλτρο με κρουστική απόκριση  $h[n]$ , τότε η προσέγγιση της συνάρτησης  $f(t)$  στην κλίμακα  $m$  μπορεί να υπολογιστεί από την συνέλιξη της  $\alpha_{m-1}[n]$  με το φίλτρο  $h[-n]$  ακολουθούμενο από υποδειγματοληψία κατά δύο:

$$a_m[n] = \sum_{k=-\infty}^{\infty} h[k-2n] a_{m-1}[k] = a_{m-1}[2n] * h[-2n]$$

Οι συντελεστές  $d_m[n]$  αποτελούν τη λεπτομέρεια της  $f(t)$  στην ανάλυση που αντιστοιχεί στον υπόχωρο  $W_m$ . Όμοια το wavelet  $\psi \in W_{m+1} \subset V_m$ , οπότε γράφεται ως γραμμικός συνδυασμός της βάσης  $\varphi(t-n)$  του  $V_m$ . Βλέπουμε λοιπόν την σχέση που συνδέει την συνάρτηση κλιμακοθέτησης ενός wavelet με τη «μητέρα» συνάρτηση wavelet [2]. Αν ονομάσουμε  $g[n]$  τους συντελεστές αυτού του γραμμικού συνδυασμού τότε θα δίνονται από την σχέση  $g[n] = \left\langle \frac{1}{\sqrt{2}} \psi\left(\frac{t}{2}\right), \varphi(t-n) \right\rangle$ . Ισοδύναμα, στο πεδίο συχνότητας θα ισχύει

$\Psi(\omega) = \frac{1}{\sqrt{2}} G(e^{j\omega/2}) \Phi(\omega/2)$ . Έτσι αποδεικνύεται ότι για τα φίλτρα  $G, H$  όταν έχουμε ορθογώνια επέκταση ισχύει

$$G(e^{j\omega}) = -e^{-j\omega} H^*(e^{j(\omega+\pi)})$$

ή ισοδύναμα στο πεδίο του χρόνου

$$g[n] = (-1)^n h[-n+1] \quad (2.8)$$

Όμοια με προηγουμένως, το  $d_m[n]$  μπορεί να υπολογιστεί επαναληπτικά συνελλίσοντας την ακολουθία  $a_{m-1}[n]$  με το φίλτρο  $g[-n]$  ακολουθούμενο από υποδειγματοληψία κατά δύο:

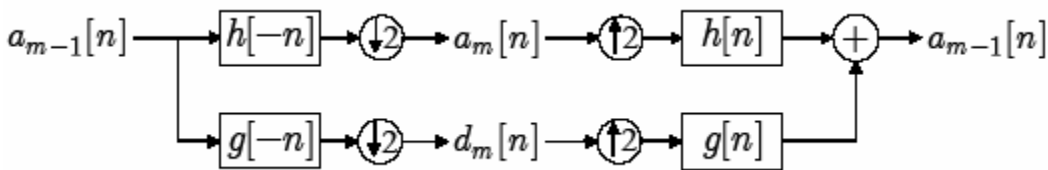
$$d_m[n] = \sum_{k=-\infty}^{\infty} g[k-2n]a_{m-1}[k] = a_{m-1}[2n] * g[-2n]$$

Τα  $h, g$  φίλτρα είναι χαμηλοδιαβατά και ζωνοδιαβατά αντίστοιχα με τις κρουστικές αποκρίσεις να συνδέονται σύμφωνα με την σχέση 2.8. Τους αρχικούς συντελεστές  $a_0[n]$  αποτελεί το ίδιο το διακριτό σήμα που θέλουμε να μετασχηματίσουμε.

Συνοψίζοντας, ο αλγόριθμος από την προσέγγιση  $a_{m-1}[n]$  της συνάρτησης  $f(t)$  υπολογίζει τις ακολουθίες  $a_m[n]$  και  $d_m[n]$  [1], [2]. Η αντίστροφη διαδικασία αποτελείται από την κατασκευή του ποιοτικότερου σήματος  $a_{m-1}[n]$  από τα  $a_m[n]$  και  $d_m[n]$ . Η ανακατασκευή γίνεται σύμφωνα με τον τύπο

$$a_{m-1}[n] = \sum_{k=-\infty}^{\infty} h[n-2k]a_m[k] + \sum_{k=-\infty}^{\infty} g[n-2k]d_m[k]$$

Από την παραπάνω σχέση φαίνεται ότι πριν εφαρμόσουμε τα φίλτρα  $h[n], g[n]$  στις ακολουθίες  $a_m[n], d_m[n]$  αντίστοιχα πρέπει να υπερδειγματοληπτήσουμε τις τελευταίες κατά δύο. Επαναλαμβάνοντας διαδοχικά τον πυραμιδικό αλγόριθμο στο αρχικό διακριτό σήμα  $a_0[n]$  για  $M$  επίπεδα λαμβάνουμε τον Διακριτό (Δυαδικό) Μετασχηματισμό Wavelet (DWT):  $(a_M[n], d_M[n], d_{M-1}[n], \dots, d_1[n])$ . Το μπλοκ διάγραμμα της wavelet ανάλυσης και σύνθεσης για σήματα μίας διάστασης φαίνεται στην Εικόνα 2.8. Το αρχικό σήμα κατασκευάζεται αναδρομικά από τη wavelet αναπαράσταση (τα  $a_M[n], d_M[n]$  δίνουν το  $a_{M-1}[n]$ , το οποίο μαζί με το  $d_{M-1}[n]$  δίνει το  $a_{M-2}[n]$  κ.ο.κ).



Εικόνα 2.8 Υλοποίηση μετασχηματισμού Wavelet με χρήση τράπεζας φίλτρων

### Δισδιάστατη περίπτωση

Στην περίπτωση που έχουμε δισδιάστατο σήμα, όπως για παράδειγμα εικόνες, η ανάλυση πολλαπλής διακριτικής ικανότητας  $M$  επιπέδων μας δίνει ένα σύνολο  $3M + 1$  υποζωνών.

Αυτό συμβαίνει, γιατί η αρχική εικόνα αναλύεται σε 4 υποζώνες, αφού εφαρμόζουμε τα χαμηλοδιαβατά και υψιπερατά φίλτρα κατά γραμμές και κατά στήλες. Έτσι, διακρίνουμε τις υποζώνες: χαμηλοπερατή-χαμηλοπερατή, χαμηλοπερατή-υψιπερατή, υψιπερατή-χαμηλοπερατή, υψιπερατή-υψιπερατή. Η διαδικασία επαναλαμβάνεται στην χαμηλοπερατή-χαμηλοπερατή υποζώνη και έτσι η DWT αναπαράσταση  $M$  επιπέδων ενός δισδιάστατου σήματος  $f$  είναι:  $\{aa_M[i,j], (ad_m[i,j], da_m[i,j], dd_m[i,j])\}_{1 \leq m \leq M}$ . Όπως αναφέραμε προηγουμένως, η δισδιάστατη wavelet ανάλυση αποτελείται από την εφαρμογή του μονοδιάστατου DWT κατά γραμμές και στήλες. Πρώτα συνελλίσουμε τις γραμμές του  $aa_{m-1}[i,j]$  με ένα μονοδιάστατο φίλτρο και υποδειγματοληπτούμε κατά δύο και έπειτα συνελλίσουμε τις στήλες των σημάτων που προέκυψαν με το μονοδιάστατο φίλτρο και υποδειγματοληπτούμε ξανά κατά δύο.

Είδαμε λοιπόν πώς μπορούμε να υλοποιήσουμε wavelet επεκτάσεις διακριτών σημάτων με χρήση μιας τράπεζας φίλτρων. Στην περίπτωση που έχουμε ορθοκανονική επέκταση τα φίλτρα ανάλυσης  $h[-n]$ ,  $g[-n]$  είναι απλά χρονικά αντίθετες εκδοχές των φίλτρων σύνθεσης  $h[n]$ ,  $g[n]$  και ικανοποιούν την συνθήκη τέλει ανακατασκευής [2] που είναι η εξής:

$$|H(\omega)|^2 + |H(\omega+\pi)|^2 = 2, \quad |G(\omega)|^2 + |G(\omega+\pi)|^2 = 2$$

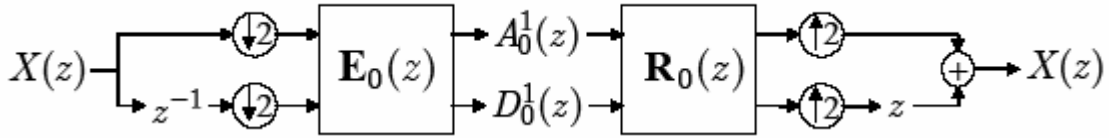
Οι ορθοκανονικές επεκτάσεις έχουν όμως και ένα σοβαρό μειονέκτημα, ότι δεν υπάρχουν ορθοκανονικά γραμμικής φάσης FIR φίλτρα, εκτός από την περίπτωση της Haar επέκτασης. Για να αποκτήσουμε τη γραμμική φάση μπορούμε να αφήσουμε τον περιορισμό της ορθοκανονικότητας και να χρησιμοποιήσουμε biorthogonal βάσεις. Με αυτόν τον τρόπο πετυχαίνουμε περισσότερη ελευθερία στο σχεδιασμό των συναρτήσεων βάσης. Υπενθυμίζουμε ότι στην περίπτωση biorthogonal επεκτάσεων [1], [2] υπάρχουν δύο σύνολα βάσεων  $\{\tilde{\varphi}_{m,n}\}$ ,  $\{\varphi_{m,n}\}$  που αντιστοιχούν στους χώρους  $\tilde{V}_m, V_m$  και τα οποία ικανοποιούν την σχέση  $\langle \tilde{\varphi}_{m,n-l}, \varphi_{m,n-k} \rangle = \delta[k-l]$ . Η ανάλυση γίνεται χρησιμοποιώντας τη βάση  $\{\tilde{\varphi}_{m,n}\}$  ενώ η σύνθεση χρησιμοποιώντας τη βάση  $\{\varphi_{m,n}\}$ . Αντίστοιχα, ορίζουμε και τα σύνολα βάσεων  $\{\tilde{\psi}_{m,n}\}$ ,  $\{\psi_{m,n}\}$  που αντιστοιχούν στους χώρους  $\tilde{W}_m, W_m$ . Στην biorthogonal επέκταση οι υπόχωροι  $\tilde{V}_m$  και  $\tilde{W}_m$  όπως και οι  $V_m$  και  $W_m$  δεν είναι ορθογώνια συμπληρώματα των υποχώρων  $\tilde{V}_{m-1}$  και  $V_{m-1}$  αντίστοιχα. Οι υπόχωροι  $V_m$  και  $W_m$  είναι ορθογώνιοι μεταξύ τους όπως και οι  $\tilde{V}_m$  και  $\tilde{W}_m$ . Για να υπάρχει τέλεια ανακατασκευή από την τράπεζα φίλτρων θα πρέπει να ισχύει η συνθήκη:

$$\begin{cases} \tilde{H}^*(\omega)H(\omega) + \tilde{G}^*(\omega)G(\omega) = 2 \\ \tilde{H}^*(\omega + \pi)H(\omega) + \tilde{G}^*(\omega + \pi)G(\omega) = 0 \end{cases}$$

Με  $\tilde{H}(\omega), \tilde{G}(\omega)$  και  $H(\omega), G(\omega)$  συμβολίζουμε τους μετασχηματισμούς Fourier των φίλτρων ανάλυσης και σύνθεσης αντίστοιχα. Η κατασκευή ορθοκανονικών και biorthogonal wavelet βάσεων υπήρξε αντικείμενο μεγάλης ερευνητικής προσπάθειας την περασμένη δεκαετία, οδηγώντας στην εύρεση ποικίλων wavelets χρήσιμων για εφαρμογές συμπίεσης σημάτων.

### 2.2.3 The Lifting Scheme

Στην προηγούμενη παράγραφο είδαμε πως μπορούμε να υλοποιήσουμε το μετασχηματισμό wavelet με την χρήση μιας τράπεζας φίλτρων, όπως φαίνεται στην Εικόνα 2.8. Όπως εξηγήσαμε πρώτα λαμβάνει χώρα η συνέλιξη και έπειτα η υποδειγματοληψία. Αυτό σημαίνει ότι και άσκοπα υπολογίζουμε τα μισά δείγματα και χρόνο ξοδεύουμε. Το ίδιο συμβαίνει και κατά την ανακατασκευή όπου προσθέτουμε μηδενικά και κάνουμε πολλαπλασιασμούς με αυτά. Η μέθοδος με την οποία αντιμετωπίζεται το πρόβλημα αυτό ονομάζεται σχήμα ανόρθωσης (lifting scheme) και στηρίζεται σε δύο βασικές ιδέες [1], [2], [16], [19]. Η πρώτη είναι η δειγματοληψία να γίνεται πριν το φιλτράρισμα ώστε να μειώσουμε στο μισό τις πράξεις και η δεύτερη είναι αντί να χρησιμοποιούμε συνέλιξεις να χρησιμοποιήσουμε στοιχεία από τις δύο ακολουθίες  $a[n]$  και  $d[n]$  και να προσθέσουμε τη μία στην άλλη. Σύμφωνα με τη μέθοδο, αρχικά χωρίζουμε την συνάρτηση σε δύο υποακολουθίες, η μία περιέχει τα ζυγά στοιχεία και η άλλη τα περιττά. Η πρώτη ιδέα υλοποιείται με την χρήση του πολυφασικού (polyphase) πίνακα των φίλτρων ανάλυσης και σύνθεσης, όπως φαίνεται στην Εικόνα 2.9. Παρατηρούμε ότι αντί να φιλτράρουμε με τα φίλτρα  $H(z)$ ,  $G(z)$  φιλτράρουμε με τον πίνακα  $E_0(z)$ . Με  $A_0^{-1}(z)$  συμβολίζουμε την χαμηλοδιαβατή προσέγγιση του σήματος στο πεδίο  $Z$  και με  $D_0^{-1}(z)$  αντίστοιχα το σήμα λεπτομέρειας. Όπως είπαμε προηγουμένως, εδώ η διαδικασία της υποδειγματοληψίας εκτελείται πριν από το φιλτράρισμα και έτσι γλιτώνουμε τους περιττούς υπολογισμούς που θα έπρεπε έπειτα να πετάξουμε κατά τη διαδικασία υποδειγματοληψίας, όπως αυτή γίνεται στην Εικόνα 2.8. Κατά τη διαδικασία της σύνθεσης έχουμε τον πολυφασικό πίνακα  $R_0(z)$  και αφού φιλτράρουμε τα σήματα με αυτόν τον πίνακα ακολουθεί η υπερδειγματοληψία κατά δύο και η πρόσθεση των δύο σημάτων που προέκυψαν.



Εικόνα 2.9 Μετασχηματισμός wavelet με χρήση του πολυφασικού (polyphase) πίνακα

Ο πολυφασικός πίνακας της ανάλυσης  $E_0(z)$  ορίζεται ως

$$\mathbf{E}_0(z) = \begin{bmatrix} H_0(z) & H_1(z) \\ G_0(z) & G_1(z) \end{bmatrix}$$

όπου με  $F_{\{0,1\}}(z)$  συμβολίζουμε τις πολυφασικού τύπου-I συνιστώσες του φίλτρου  $F(z)$  που δίνονται από τις σχέσεις:

$$F_0(z) = \frac{1}{2} \left( F(z^{\frac{1}{2}}) + F(-z^{\frac{1}{2}}) \right)$$

$$F_1(z) = \frac{1}{2} z^{\frac{1}{2}} \left( F(z^{\frac{1}{2}}) - F(-z^{\frac{1}{2}}) \right)$$

Με τις παραπάνω σχέσεις οι ζυγοί και μονοί συντελεστές του φίλτρου  $F(z)$  διαχωρίζονται στα φίλτρα  $F_0(z)$ ,  $F_1(z)$  αντίστοιχα. Το σήμα  $X(z)$  επίσης υφίσταται διαχωρισμό στη ζυγή και μονή ακολουθία. Έτσι προκύπτουν οι ακολουθίες:

$$X_0(z) = \frac{1}{2} \left( X(z^{\frac{1}{2}}) + X(-z^{\frac{1}{2}}) \right)$$

$$X_1(z) = \frac{1}{2} z^{-\frac{1}{2}} \left( X(z^{\frac{1}{2}}) - X(-z^{\frac{1}{2}}) \right)$$

Η wavelet ανάλυση γράφεται τώρα ως

$$\begin{bmatrix} A_0^1(z) \\ D_0^1(z) \end{bmatrix} = \mathbf{E}_0(z) \begin{bmatrix} X_0(z) \\ X_1(z) \end{bmatrix}$$

και αντίστοιχα η διαδικασία της wavelet σύνθεσης ως



$$\begin{bmatrix} X_0(z) \\ X_1(z) \end{bmatrix} = \mathbf{R}_0(z) \begin{bmatrix} A_0^1(z) \\ D_0^1(z) \end{bmatrix}$$

Για να έχουμε τέλεια ανακατασκευή θα πρέπει να ισχύει η σχέση

$$\mathbf{E}_0(z^{-1})\mathbf{R}_0(z) = \mathbf{I}$$

Έτσι, ο πίνακας  $\mathbf{R}_0(z)$  συνδέεται με τον πίνακα σύνθεσης  $\mathbf{E}_0(z)$  με τον τύπο

$$\mathbf{R}_0(z) = \frac{1}{\det \mathbf{E}_0(z)} \begin{bmatrix} G_1(z) & -H_1(z) \\ -G_0(z) & H_0(z) \end{bmatrix}$$

Αφού προκύψουν οι πολυφασικές συνιστώσες  $X_0(z)$ ,  $X_1(z)$  τις προσθέτουμε και ανακατασκευάζεται έτσι το αρχικό σήμα  $X(z)$ . Δηλαδή, έχουμε

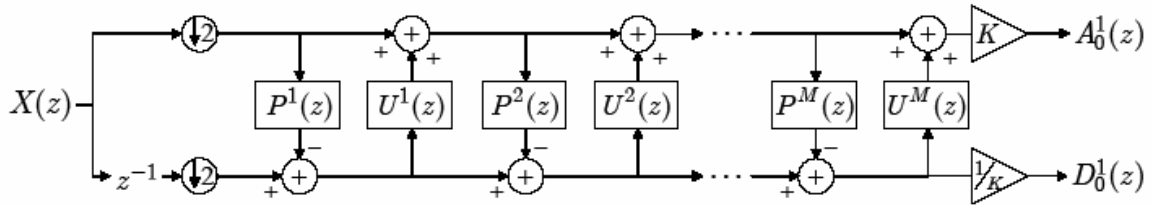
$$X(Z) = X_0(z^2) + zX_1(z^2)$$

### Παραγοντοποίηση του $\mathbf{E}_0(z)$ – Lifting Scheme

Οι Daubechies και Sweldens [7] έδειξαν ότι μια τράπεζα φίλτρων η οποία υλοποιεί biorthogonal ανάλυση και σύνθεση μπορεί να υλοποιηθεί από μια ακολουθία βημάτων ανανέωσης (liftings) και προβλέψεων (dual-liftings). Αντί λοιπόν να εφαρμόσουμε τον πολυφασικό πίνακα σύνθεσης στις ακολουθίες εισόδου  $X_0(z)$ ,  $X_1(z)$  εφαρμόζουμε αυτήν την ακολουθία βημάτων, όπως φαίνεται στην Εικόνα 2.10. Η απόδειξη του παραπάνω θεωρήματος στηρίζεται στον αλγόριθμο του Ευκλείδη για την πολυωνυμική διαίρεση. Έτσι υπάρχουν πολλές διαφορετικές λύσεις για την υλοποίηση μιας τράπεζας φίλτρων με το σχήμα ανόρθωσης (lifting scheme). Γενικά οι παραγοντοποιήσεις αυτές εκφράζονται από την σχέση

$$\mathbf{E}(z) = \begin{bmatrix} K & 0 \\ 0 & 1/K \end{bmatrix} \prod_{i=1}^M \begin{bmatrix} 1 & U^i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -P^i(z) & 1 \end{bmatrix}$$

Τα φίλτρα  $P^i(z)$  και  $U^i(z)$  υποδηλώνουν τελεστές πρόβλεψης και ανανέωσης αντίστοιχα και οι  $K$ ,  $1/K$  είναι παράγοντες κανονικοποίησης ώστε ο μετασχηματισμός να είναι ορθοκανονικός (κατά προσέγγιση). Αξίζει να αναφέρουμε ότι αν και για κάθε βήμα  $i$   $1 \leq i \leq M$  το βήμα πρόβλεψης εμφανίζεται να προηγείται του βήματος ανανέωσης, οι παραγοντοποιήσεις του lifting scheme μπορούν να υλοποιηθούν με την αντίθετη σειρά των βημάτων.



Εικόνα 2.10 Αναπαράσταση της wavelet ανάλυσης βασισμένη στο lifting scheme

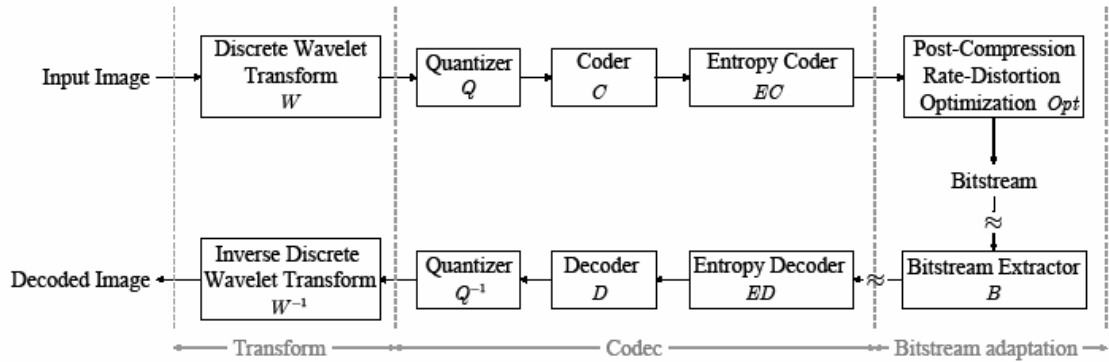
Ο αντίστροφος μετασχηματισμός, δηλαδή η διαδικασία κατά την οποία από τα σήματα  $A_0^1(z)$ ,  $D_0^1(z)$  θα προκύψει το σήμα  $X(z)$ , υλοποιείται αν ακολουθήσουμε το σχήμα από δεξιά προς τα αριστερά, αντιστρέψουμε τους παράγοντες κανονικοποίησης, την κατεύθυνση ροής και τα πρόσημα στα σημεία άθροισης. Τελικά, το ανασχηματισμένο σήμα προκύπτει με υπερδειγματοληψία και ένωση των πολυφασικών συνιστωσών. Το σχήμα ανόρθωσης (lifting scheme) συνδυάζει πολλά πλεονεκτήματα όπως χαμηλή πολυπλοκότητα υλοποίησης, αφού μειώνεται ο αριθμός των απαιτούμενων πολλαπλασιασμών και αθροίσεων, μετασχηματισμούς από ακέραιους σε ακέραιους και πλήρη αντιστρεψιμότητα ακόμα και αν χρησιμοποιηθούν μη γραμμικά ή προσαρμοζόμενα φίλτρα για τα βήματα ανανέωσης και πρόβλεψης. Τα πλεονεκτήματα αυτά έχουν οδηγήσει σε πολλές ενδιαφέρουσες εφαρμογές, όπως η προσαρμοζόμενη wavelet ανάλυση για κωδικοποίηση εικόνας. Μια άλλη εφαρμογή, την οποία και θα αναλύσουμε στο επόμενο κεφάλαιο είναι η κλιμακοθετήσιμη κωδικοποίηση βίντεο με χρήση wavelets.

## Κεφάλαιο 3 : Κλιμακοθετήσιμη Κωδικοποίηση Βίντεο (Scalable Video Coding)

Στο προηγούμενο κεφάλαιο μελετήσαμε το μαθηματικό υπόβαθρο στο οποίο στηρίζεται ο μετασχηματισμός wavelet και αναπτύξαμε τις ιδέες στις οποίες βασίζεται. Στο κεφάλαιο αυτό θα παρουσιάσουμε την εφαρμογή του μετασχηματισμού wavelet στην κλιμακοθετήσιμη κωδικοποίηση εικόνας και βίντεο. Η κλιμακοθετήσιμη κωδικοποίηση βίντεο με χρήση των wavelets αποτελεί το βασικό κομμάτι της διπλωματικής εργασίας και αναλύεται εκτενώς τόσο θεωρητικά στις παρακάτω παραγράφους όσο και πειραματικά στο Κεφάλαιο 4. Στην παράγραφο 3.1 παρουσιάζεται η διαδικασία της κλιμακοθετήσιμης κωδικοποίησης εικόνας και στις παραγράφους 3.2 - 3.4 τα βασικά συστατικά ενός συστήματος κωδικοποίησης βίντεο. Στην συνέχεια στην παράγραφο 3.5 παρουσιάζουμε προχωρημένες τεχνικές κωδικοποίησης βίντεο και στην παράγραφο 3.6 αναλύουμε το χωρικό μετασχηματισμό wavelet. Τέλος, κλείνουμε το κεφάλαιο με τη μελέτη ενός από τους πιο γνωστούς αλγόριθμους κωδικοποίησης συντελεστών του μετασχηματισμού wavelet, τον αλγόριθμο SPIHT.

### 3.1 Κλιμακοθετήσιμη Κωδικοποίηση Εικόνας με χρήση των Wavelets

Το βασικό πλαίσιο εργασίας για ένα σύστημα scalable κωδικοποίησης εικόνας παρουσιάζεται στην Εικόνα 3.1. Όπως βλέπουμε ο κωδικοποιητής αποτελείται από το κομμάτι του μετασχηματισμού wavelet ( $W$ ), το οποίο μπορεί να είναι οποιοσδήποτε δισδιάστατος διακριτός μετασχηματισμός wavelet, το κομμάτι του κβαντιστή ( $Q$ ), τον κωδικοποιητή ( $C$ ) και τον κωδικοποιητή εντροπίας ( $EC$ ). Πριν από τον αποκωδικοποιητή έχουμε τα δομικά στοιχεία (modules)  $OpT$  και  $B$ , τα οποία αναλαμβάνουν την παραγωγή της κατάλληλης υποακολουθίας από την συμπιεσμένη ακολουθία που έχει δημιουργηθεί. Η δημιουργία της υποακολουθίας γίνεται με το κατάλληλο ταίριασμα της χωρητικότητας του καναλιού με την ποιότητα και ανάλυση της αποκωδικοποιημένης εικόνας που έχει επιλέξει ο χρήστης. Στην πραγματικότητα τα δομικά στοιχεία  $B$  και  $OpT$  αλληλεπιδρούν όμοια με μία εφαρμογή πελάτη-εξυπηρετητή (client-server). Ο χρήστης ορίζει το επιθυμητό ρυθμό δυαδικών ψηφίων και το επίπεδο ανάλυσης στο δομικό στοιχείο  $B$ , το οποίο ζητά τα κατάλληλα σημεία της κωδικοποιημένης ακολουθίας που έχουν δημιουργηθεί από το  $OpT$  δομικό στοιχείο. Στον αποκωδικοποιητή εκτελείται η αντίστροφη διαδικασία. Πρώτα η ακολουθία που έχουμε λάβει περνάει από τον αποκωδικοποιητή εντροπίας ( $ED$ ), τον αποκωδικοποιητή ( $D$ ), τον αντίστροφο κβαντιστή ( $Q^{-1}$ ) και έπειτα από τον αντίστροφο δισδιάστατο μετασχηματισμό wavelet ( $W^{-1}$ ).



**Εικόνα 3.1** Αρχιτεκτονική κλιμακοθετήσιμης κωδικοποίησης εικόνας βασισμένη στο μετασχηματισμό wavelet

Ο ρόλος του κβαντιστή του συστήματος της Εικόνας 3.1 είναι να αντιστοιχίσει τους συντελεστές του μετασχηματισμού wavelet σε ένα πεπερασμένο σύνολο τιμών. Συνήθως, ο κβαντιστής που χρησιμοποιείται ανήκει στην κατηγορία εμφωλιασμένων κβαντιστών, όπως για παράδειγμα ο Successive Approximation Quantization (SAQ).

Η διαδικασία κωδικοποίησης (C) αναλαμβάνει να κωδικοποιήσει τις κβαντισμένες τιμές, εκμεταλλευόμενη την σχέση που συνδέει τους συντελεστές που έχουν προκύψει από το δισδιάστατο μετασχηματισμό wavelet. Οι πιο γνωστοί αλγόριθμοι κωδικοποίησης είναι ο Set Partitioning In Hierarchical Trees (SPIHT) [14] [15], ο QuadTree-Limited (QT-L) και ο Embedded Block Coding with Optimal Truncation (EBCOT). Στην εργασία αυτή ως αλγόριθμος κωδικοποίησης επιλέχτηκε ο SPIHT, ο οποίος αναλύεται στην παράγραφο 3.7.

Στην περίπτωση ενός κλιμακοθετήσιμου συστήματος κωδικοποίησης βίντεο, το σύστημα μοιάζει με αυτό της Εικόνας 3.1. Η διαφορά είναι ότι πριν από το δισδιάστατο μετασχηματισμό, πραγματοποιείται μετασχηματισμός στον άξονα του χρόνου. Δηλαδή, τα πλαίσια που αποτελούν την ακολουθία του βίντεο μετασχηματίζονται στον άξονα του χρόνου με κάποιον από τους μετασχηματισμούς wavelet. Συνήθως για το χρονικό μετασχηματισμό επιλέγεται ο μετασχηματισμός Haar ή ο 5/3 wavelet μετασχηματισμός. Όμοια, στον αποκωδικοποιητή προστίθεται στο τέλος το κομμάτι του αντίστροφου χρονικού μετασχηματισμού wavelet. Επιπλέον, για τον κωδικοποιητή (C) επιλέγεται ένας τρισδιάστατος κωδικοποιητής συντελεστών, όπως για παράδειγμα ο 3D-SPIHT [15]. Η διαδικασία της κλιμακοθετήσιμης κωδικοποίησης βίντεο με χρήση του μετασχηματισμού wavelet αποτελεί το κύριο αντικείμενο της διπλωματικής εργασίας και αναλύεται στις επόμενες παραγράφους.

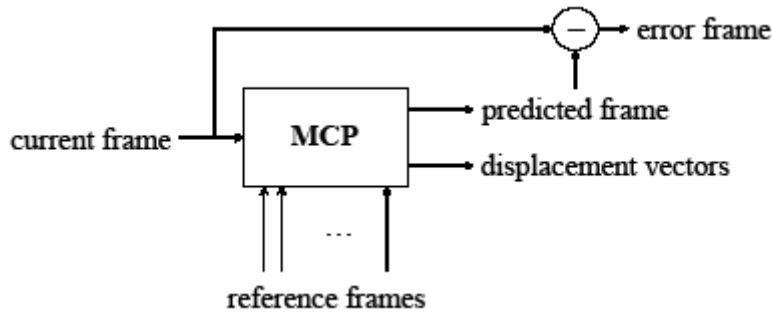
### 3.2 Κινησιακά Αντισταθμισμένη Πρόβλεψη κατά Μπλοκ (*Block-based Motion Compensated Prediction*)

Όπως αναφέραμε στην προηγούμενη παράγραφο το πρώτο βήμα στην κωδικοποίηση ακολουθιών βίντεο είναι ο χρονικός μετασχηματισμός της ακολουθίας, δηλαδή η μείωση της πλεονάζουσας πληροφορίας μεταξύ διαδοχικών πλαισίων. Η κινησιακά αντισταθμισμένη πρόβλεψη (Motion Compensated Prediction (MCP)) αποτελεί ένα ισχυρό εργαλείο για τη μείωση της χρονικής συσχέτισης μεταξύ διαδοχικών πλαισίων βίντεο και αποτελεί βασικό κομμάτι σε κάθε κωδικοποιητή βίντεο είτε επιθυμούμε κλιμακοθετήσιμη είτε όχι κωδικοποίηση. Στην παράγραφο αυτή θα μελετήσουμε τον τρόπο λειτουργίας της κινησιακά αντισταθμισμένης πρόβλεψης [1], [5], [6].

Στην πιο γενική περίπτωση, η διαδικασία MCP δημιουργεί συσχέτιση μεταξύ περιοχών του τρέχοντος πλαισίου και προηγούμενων ή μελλοντικών πλαισίων τα οποία είναι διαθέσιμα τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή. Τα πλαίσια αυτά τα αναφέρουμε συχνά ως πλαίσια αναφοράς. Δοσμένων των πλαισίων αναφοράς και των παραμέτρων αντιστοίχισης η διαδικασία MCP δημιουργεί το προβλεπόμενο πλαίσιο και τα διανύσματα μετατόπισης τα οποία είναι οι παράμετροι της αντιστοίχισης σε σχέση με το πλαίσιο αναφοράς. Από τη διαφορά του τωρινού πλαισίου και του πλαισίου που προέβλεψε η διαδικασία MCP προκύπτει ένα πλαίσιο σφάλματος. Στα περισσότερα συστήματα κωδικοποίησης, έπειτα από τη διαδικασία MCP, το πλαίσιο σφάλματος και τα διανύσματα μετατόπισης μεταδίδονται στον αποκωδικοποιητή. Η δομή της διαδικασίας MCP παρουσιάζεται στην Εικόνα 3.2. Όπως είπαμε, για τη δημιουργία του προβλεπόμενου πλαισίου χρειάζονται οι παράμετροι αντιστοίχισης μεταξύ του τωρινού πλαισίου και του πλαισίου αναφοράς. Για να βρούμε τις παραμέτρους αυτές εφαρμόζουμε κάποιον από τους πολλούς διαθέσιμους αλγόριθμους εκτίμησης κίνησης. Ενδεικτικά αναφέρουμε τον αλγόριθμο πλήρους αναζήτησης, τον ιεραρχικό αλγόριθμο και τον PHODS.

Διατυπώνοντας τα παραπάνω μαθηματικά αν  $A_t[m, n]$  το τωρινό πλαίσιο,  $A_{t-1}[m, n]$  το πλαίσιο αναφοράς και  $d_m, d_n$  τα διανύσματα μετατόπισης που έχουν προκύψει από τον αλγόριθμο εκτίμησης κίνησης τότε η έξοδος της διαδικασίας MCP θα είναι το πλαίσιο σφάλματος  $H_t[m, n]$ , που δίνεται από την σχέση:

$$H_t[m, n] = A_t[m, n] - A_{t-1}[m - d_m, n - d_n]$$



Εικόνα 3.2 Motion Compensated Prediction

Αν το τωρινό πλαίσιο και το πλαίσιο αναφοράς έχουν μεγάλη συσχέτιση μεταξύ τους, τότε το πλαίσιο σφάλματος θα περιέχει έναν μεγάλο αριθμό εικονοστοιχείων που έχουν τιμή κοντά στο μηδέν. Έτσι, το πλαίσιο σφάλματος θα έχει μικρότερη ενέργεια από το τωρινό πλαίσιο, οπότε θα είναι ευκολότερο να υποστεί συμπίεση. Αυτή είναι και η μεγάλη συμβολή της διαδικασίας MCP. Στην περίπτωση που το τωρινό πλαίσιο και το πλαίσιο αναφοράς είναι αρκετά διαφορετικά μεταξύ τους, η διαδικασία MCP δε θα επιφέρει πλεονέκτημα όσον αφορά τη συμπίεση. Συνήθως αυτό συμβαίνει όταν έχουμε εμφάνιση νέων αντικειμένων, αλλαγές στη φωτεινότητα της σκηνής, αλλαγές στο ζουμ της κάμερας κτλ. Η διαδικασία MCP μπορεί να λύσει το πρόβλημα αυτό αλλά λαμβάνοντας υπόψη κάθε φορά ένα από τα παραπάνω φαινόμενα και όχι όλα. Επίσης, όσο πιο προχωρημένος είναι ο αλγόριθμος εκτίμησης κίνησης που χρησιμοποιείται για τη διαδικασία MCP, τόσο πιο πολύπλοκο είναι να υλοποιήσουμε ένα πρακτικό σύστημα κωδικοποίησης βίντεο. Ως αποτέλεσμα οι περισσότεροι αλγόριθμοι εκτίμησης κίνησης που χρησιμοποιούνται είναι αλγόριθμοι εκτίμησης κίνησης κατά μπλοκ, οπότε και η διαδικασία MCP εφαρμόζεται κατά μπλοκ.

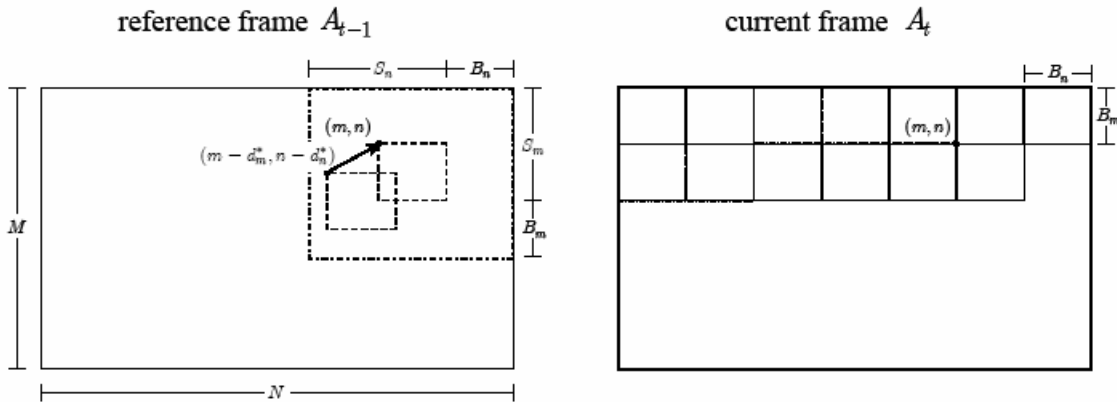
### 3.2.1 Εκτίμηση Κίνησης κατά Μπλοκ (Block-based Motion Estimation)

Στην ενότητα αυτή θα αναλύσουμε τον αλγόριθμο εκτίμησης κίνησης κατά μπλοκ (block-based). Το τρέχων πλαίσιο  $A_i[m,n]$ , που αποτελείται από  $M \times N$  εικονοστοιχεία (pixels), χωρίζεται σε μη επικαλυπτόμενα μπλοκ διαστάσεων  $B_m \times B_n$  εικονοστοιχείων. Κάθε μπλοκ προβλέπεται, έπειτα από μια διαδικασία ταιριάσματος με κάποιο από τα μπλοκ που περιέχονται σε περιοχή  $S_m \times S_n$  εικονοστοιχείων γύρω από τη θέση του, όπως φαίνεται στην Εικόνα 3.3. Το κριτήριο για το ταιρίασμα που χρησιμοποιείται περισσότερο είναι το άθροισμα απόλυτων διαφορών (SAD), αφού παρέχει ακριβή ταιριάσματα χωρίς την ανάγκη για εκτέλεση πολλαπλασιασμών. Έτσι, επιλέγουμε εκείνα

τα διανύσματα κίνησης για τα οποία το SAD ελαχιστοποιείται. Διατυπώνοντας τα παραπάνω μαθηματικά έχουμε:

$$(d_m^*, d_n^*) : \arg \min_{(d_m, d_n)} \sum_{i=0}^{B_m-1} \sum_{j=0}^{B_n-1} C(A_t[m+i, n+j] - A_{t-1}[m+i-d_m, n+j-d_n]) \quad (3.1)$$

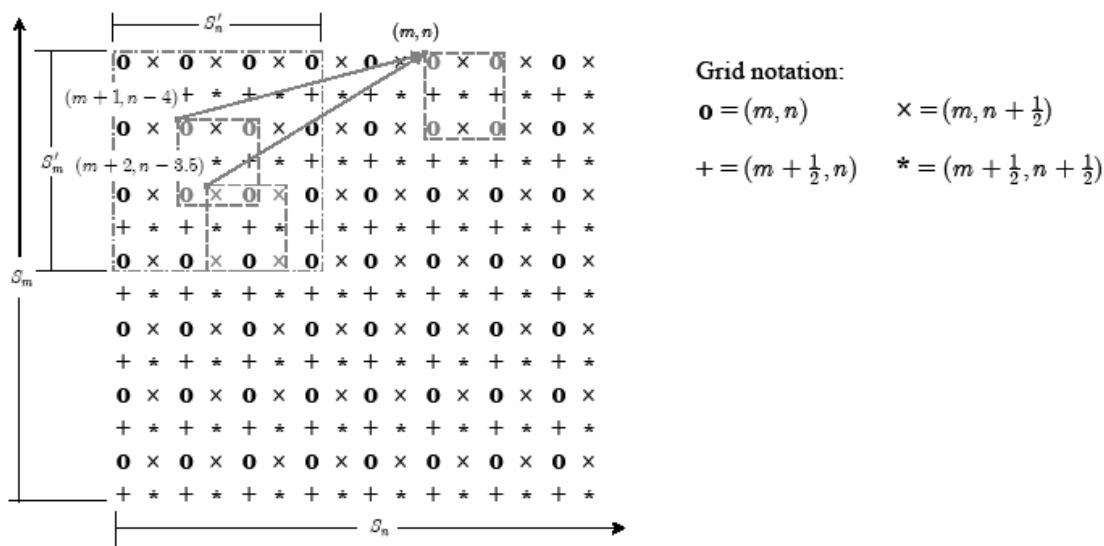
Οι τιμές  $d_m, d_n$  βρίσκονται στα όρια  $[-S_m/2, S_m/2]$  και  $[-S_n/2, S_n/2]$  αντίστοιχα και  $C(a) = |a|$  όταν έχουμε το SAD κριτήριο. Εναλλακτικά μπορούμε να χρησιμοποιήσουμε και το κριτήριο αθροίσματος των τετραγωνικών διαφορών (SSD), δηλαδή  $C(a) = a^2$ . Στους πρακτικούς αλγόριθμους εκτίμησης κίνησης κατά μπλοκ, η επιθυμητή λύση  $(d_m^*, d_n^*)$  είναι ίδια για όλα τα εικονοστοιχεία  $A_t[m+i, n+j]$ , δηλαδή όλα τα εικονοστοιχεία του μπλοκ της θέσης  $(m, n)$  συσχετίζονται με το πρότυπο αναφοράς με την ίδια λύση  $(d_m^*, d_n^*)$ . Για να λύσουμε την σχέση (3.1) ένας κλασικός τρόπος είναι η χρήση ενός πλήρους αναζήτησης (brute-force) αλγόριθμου.



Εικόνα 3.3 Εκτίμηση κίνησης κατά μπλοκ

Για να αυξήσουμε την ακρίβεια της εκτίμησης κίνησης κατά μπλοκ εφαρμόζουμε τεχνικές παρεμβολής [17]. Ανάλογα με την εφαρμογή και το μέγεθος του μπλοκ απαιτείται και διαφορετικός βαθμός παρεμβολής. Για παράδειγμα, από πειραματικές μετρήσεις προέκυψε ότι για μέγεθος μπλοκ  $16 \times 16$ , τα μεταδιδόμενα σήματα τηλεόρασης απαιτούν παρεμβολή με ακρίβεια ενός τετάρτου του εικονοστοιχείου ενώ βιντεοφωνικά σήματα συνήθως απαιτούν ακρίβεια μισού εικονοστοιχείου. Ένα υποθετικό παράδειγμα εκτίμησης κίνησης με παρεμβολή μισού εικονοστοιχείου, παρουσιάζεται στην Εικόνα 3.4. Σε αυτό το παράδειγμα υποθέτουμε ότι  $B_m = B_n = 2$  και  $S_m = S_n = 10$ . Όπως φαίνεται τα καλύτερα ταιριάσματα στο πλήρες (full-pixel) πλέγμα και στο πλέγμα που υποστηρίζει ακρίβεια μισού εικονοστοιχείου (half-pixel), βρίσκονται σε γειτονικές θέσεις. Γενικά, όταν επιθυμούμε ακρίβεια υποεικονοστοιχείου (sub-pixel)

accuracy), οι πρακτικοί αλγόριθμοι εκτίμησης κίνησης κατά μπλοκ εφαρμόζουν την υπόθεση της ομοιογενούς κίνησης. Δηλαδή, εκτελούν ιεραρχική αναζήτηση γύρω από το καλύτερο ταίριασμα που βρήκαν στην προηγούμενη ακρίβεια παρεμβολής. Για παράδειγμα, στην Εικόνα 3.4 ένα πρακτικό σύστημα εκτίμησης κίνησης αρχικά βρίσκει το καλύτερο ταίριασμα στο πλήρες πλέγμα εκτελώντας πλήρη αναζήτηση στα  $S_m \times S_n$  γειτονικά εικονοστοιχεία. Η διαδικασία αυτή μας έδωσε το διάνυσμα μετατόπισης  $(-1, 4)$ . Έπειτα, γίνεται αναζήτηση σε μια επιφάνεια  $S'_m \times S'_n$  θέσεων μισών εικονοστοιχείων (half-pixel) γύρω από την καλύτερη θέση που βρέθηκε προηγουμένως, δηλαδή την  $(-1, 4)$ . Έτσι, βρέθηκε εκείνο το διάνυσμα το οποίο ελαχιστοποιεί το SAD και το οποίο είναι το  $(-2, 3.5)$ . Η διαδικασία αυτή μπορεί να συνεχιστεί και να προκύψει έτσι ακρίβεια ενός τετάρτου εικονοστοιχείου.



**Εικόνα 3.4** Εκτίμηση κίνησης με παρεμβολή μισού εικονοστοιχείου. Το τρέχων μπλοκ έχει μέγεθος 2x2 εικονοστοιχεία και βρίσκεται στη θέση  $(m, n)$ . Παρουσιάζεται μέρος της περιοχής αναζήτησης, με διαφορετικά σύμβολα να υποδηλώνουν οριζόντιες, κάθετες και διαγώνιες θέσεις παρεμβολής. Οι θέσεις του καλύτερου ταίριασματος στο πλήρες πλέγμα και στο πλέγμα μισών εικονοστοιχείων συμβολίζονται με γκρι τετράγωνα και απεικονίζονται και τα διανύσματα μετατόπισης.

### 3.2.2 Κινησιακά Αντισταθμισμένη Πρόβλεψη και Παρεμβολή Εικονοστοιχείου

Έπειτα από το βήμα της εκτίμησης κίνησης κάθε εικονοστοιχείο στο πλαίσιο  $A_t[m, n]$  αντιστοιχίζεται μέσω του διανύσματος μετατόπισης  $(d_m^{F_t(t-1)}, d_n^{F_t(t-1)})$  με ένα εικονοστοιχείο στο πλαίσιο αναφοράς, όπου με  $F_{\Delta t}(\tau)$  συμβολίζουμε την προς τα εμπρός (forward) πρόβλεψη εκτίμησης κίνησης που αντιστοιχεί το πλαίσιο αναφοράς της



χρονικής στιγμής  $\tau$  με το πλαίσιο της χρονικής στιγμής  $\tau + \Delta\tau$ . Έτσι, έπειτα από την πρόβλεψη της κίνησης το πλαίσιο σφάλματος θα δίνεται από την σχέση

$$H_t[m, n] = A_t[m, n] - I_{(i_m^{F_1(t-1)}, i_n^{F_1(t-1)})} A_{t-1}[m - \lfloor d_m^{F_1(t-1)} \rfloor, n - \lfloor d_n^{F_1(t-1)} \rfloor]$$

όπου  $i_m^{F_1(t)} = \lfloor d_m^{F_1(t)} \rfloor - d_m^{F_1(t)}$ ,  $i_n^{F_1(t)} = \lfloor d_n^{F_1(t)} \rfloor - d_n^{F_1(t)}$  είναι η κλασματική θέση της μετατόπισης,  $I_{(k,l)} A[m, n]$  είναι το εικονοστοιχείο έπειτα από την παρεμβολή στη κλασματική θέση  $(k, l)$  από το εικονοστοιχείο  $A[m, n]$ , όπου  $k, l = \{0, \frac{1}{R}, \dots, \frac{R-2}{R}, \frac{R-1}{R}\}$  και  $R$  η ακρίβεια παρεμβολής (π.χ  $R = 2$  για παρεμβολή μισού εικονοστοιχείου). Στην περίπτωση που έχουμε προς τα πίσω (backward) εκτίμηση κίνησης, τότε θα την συμβολίζουμε με  $B_{\Delta\tau}(\tau)$  και θα αντιστοιχούμε το πλαίσιο αναφοράς της χρονικής στιγμής  $\tau$  με το πλαίσιο της χρονικής στιγμής  $\tau - \Delta\tau$ .

Ο τελεστής παρεμβολής  $I(k, l)$  μπορεί να σχεδιαστεί με πολλούς τρόπους [1], [3]. Ένας από αυτούς είναι να παραθυροποιήσουμε την κρουστική απόκριση της συνάρτησης sinc γύρω από τη θέση  $(k, l)$ , υπό την υπόθεση ότι το διάνυσμα κίνησης είναι σταθερό στην περιοχή υποστήριξης του φίλτρου παρεμβολής. Σαν παράθυρο χρησιμοποιείται το Hamming παράθυρο. Έτσι οι συντελεστές του φίλτρου παρεμβολής δίνονται από τον τύπο

$$w_{m+k} = \sum_a w(a) \frac{\sin \pi(m+k-a)}{\pi(m+k-a)}$$

Το φίλτρο αυτό εφαρμόζεται στην εικόνα πρώτα κατά γραμμές και έπειτα κατά στήλες. Στην περίπτωση που το φίλτρο παρεμβολής έχει μήκος οκτώ και επιθυμούμε παρεμβολή 1/8 εικονοστοιχείου ( $R = 8$ ) οι συντελεστές του φίλτρου παρουσιάζονται στον ακόλουθο πίνακα. Να τονίσουμε ότι ο τρόπος με τον οποίο θα γίνει η παρεμβολή επηρεάζει σε πολλές περιπτώσεις την απόδοση του συστήματος και την ποιότητα της λαμβανόμενης ακολουθίας βίντεο. Στην εργασία αυτή μελετήσαμε διάφορα φίλτρα παρεμβολής και τα αποτελέσματα παρουσιάζονται στο Κεφάλαιο 4.

1/8-pixel	1/4-pixel	3/8-pixel	1/2-pixel	5/8-pixel	3/4-pixel	7/8-pixel
-0.0072	-0.0110	-0.0117	-0.0105	-0.0081	-0.0053	-0.0026
0.0284	0.0452	0.0505	0.0465	0.0363	0.0233	0.0105
-0.0902	-0.1437	-0.1624	-0.1525	-0.1224	-0.0812	-0.0380
0.9742	0.8950	0.7713	0.6165	0.4465	0.2777	0.1249

0.1249	0.2777	0.4465	0.6165	0.7713	0.8950	0.9742
-0.0380	-0.0812	-0.1224	-0.1525	-0.1624	-0.1437	-0.0902
0.0105	0.0233	0.0363	0.0465	0.0505	0.0452	0.0284
-0.0026	-0.0053	-0.0081	-0.0105	-0.0117	-0.0110	-0.0072

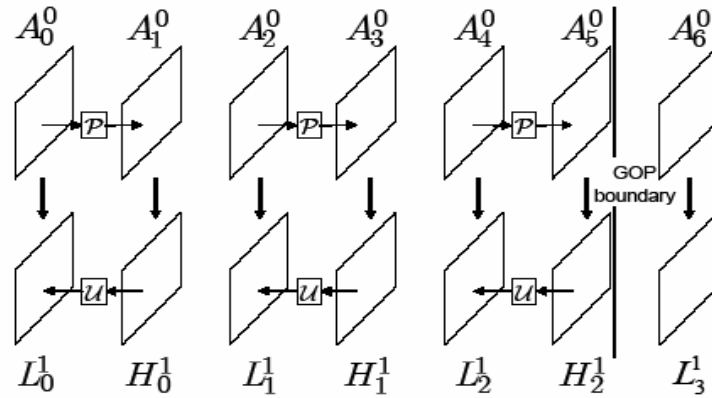
### 3.3 Κινησιακά Αντισταθμισμένη Ανανέωση (Motion Compensated Update)

Στην προηγούμενη παράγραφο μελετήσαμε το πλαίσιο εργασίας για την κινησιακά αντισταθμισμένη πρόβλεψη, όπου προβλέπουμε ένα πλαίσιο έχοντας ένα άλλο ως πρότυπο αναφοράς. Τώρα, θα δούμε πως αυτό το βήμα μπορεί να συμπληρωθεί από μια άλλη διαδικασία, που στόχος της είναι να ανανεώσει το πλαίσιο αναφοράς βασισόμενη στο πλαίσιο σφάλματος που παράχθηκε από το βήμα της πρόβλεψης. Η διαδικασία αυτή ονομάζεται Κινησιακά Αντισταθμισμένη Ανανέωση (Motion Compensated Update (MCU)). Για να κατανοήσουμε καλύτερα τις ιδέες που βρίσκονται πίσω από αυτή τη διαδικασία θα περιγράψουμε πρώτα την κλασσική wavelet ανάλυση που βασίζεται στο σχήμα ανόρθωσης (lifting scheme) και μετά θα μελετήσουμε την επέκταση που ενσωματώνει στα βήματα και την εκτίμηση κίνησης.

Όπως έδειξε η Daubechies [7] κάθε wavelet ανάλυση μπορεί να παραγοντοποιηθεί σε μια σειρά βημάτων πρόβλεψης και ανανέωσης. Την πρόβλεψη και την ανανέωση μπορούμε να τις δούμε ως συμπληρωματικά βήματα που χρησιμοποιούνται για την κατασκευή συστημάτων τέλει ανακατασκευής [5], [6]. Η πρόβλεψη χρησιμοποιεί το κομμάτι αναφοράς του σήματος για να προβλέψει το τρέχων κομμάτι και παράγει ένα σήμα σφάλματος. Το βήμα ανανέωσης παίρνει ως είσοδο το σήμα σφάλματος που παράχθηκε από την πρόβλεψη και τροποποιεί το σήμα αναφοράς. Έτσι, δημιουργείται ένα νέο πλαίσιο αναφοράς με βελτιωμένες ιδιότητες πρόβλεψης. Επιπλέον, η διαδικασία ανανέωσης κανονικοποιεί το πλάτος του σήματος αναφοράς έτσι ώστε να έχουμε ορθοκανονική ανάλυση. Η παραπάνω δομή που βασίζεται στο lifting scheme παρουσιάστηκε αναλυτικότερα στην παράγραφο 2.2.3. Εκεί είδαμε πως μπορούμε να εκτελέσουμε την wavelet ανάλυση με μία σειρά (μη μοναδική) βημάτων πρόβλεψης και ανανέωσης.

Ας πάρουμε για παράδειγμα ένα σύνολο εικόνων (GOP) αποτελούμενο από 6 πλαίσια βίντεο, όπως φαίνεται στην Εικόνα 3.5, στο οποίο θα εφαρμόσουμε wavelet ανάλυση στο πεδίο του χρόνου. Η πρόβλεψη εφαρμόζεται στα πλαίσια  $A_{2t+1}$ ,  $t = \{0,1,2\}$

χρησιμοποιώντας ως αναφορά τα πλαίσια  $A_{2t}$ . Το αποτέλεσμα της πρόβλεψης είναι η παραγωγή των πλαισίων σφάλματος  $H_t$ . Έπειτα σύμφωνα με τη δομή ανόρθωσης (lifting scheme), η πληροφορία από τα πλαίσια  $H_t$  μεταδίδεται πίσω στα πλαίσια αναφοράς δημιουργώντας τα ανανεωμένα πλαίσια αναφοράς  $L_t$ . Αυτό το βήμα αποτελεί το βήμα της ανανέωσης. Η παραπάνω διαδικασία μπορεί να επαναληφθεί χρησιμοποιώντας τα πλαίσια  $L_t$  ως πλαίσια εισόδου, δημιουργώντας έτσι το επόμενο χρονικό επίπεδο ανάλυσης. Παρατηρούμε ότι κάθε φορά τα πλαίσια εισόδου μειώνονται στον μισό αριθμό.



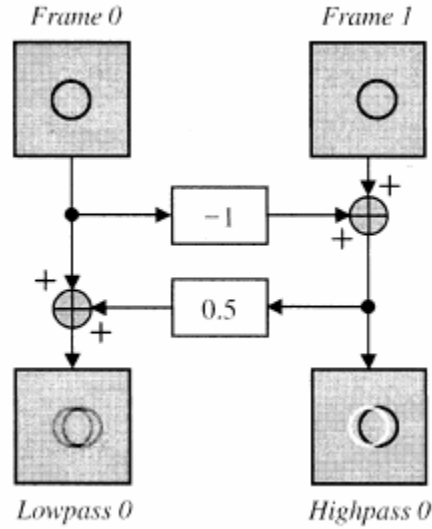
Εικόνα 3.5 Wavelet ανάλυση στο πεδίο του χρόνου με εφαρμογή των βημάτων πρόβλεψης και ανανέωσης

Αν χρησιμοποιήσουμε τη γνωστή μας Haar wavelet ανάλυση που είδαμε στο Κεφάλαιο 2, τότε τα πλαίσια  $H_t$ ,  $L_t$  θα δίνονται από την σχέση 3.2. Τα  $L_t$  αντιστοιχούν στα πλαίσια κατώτερης ποιότητας, ενώ τα  $H_t$  στα πλαίσια της λεπτομέρειας δηλαδή στην πραγματικότητα έχουμε ένα χαμηλοπερατό και ένα υψιπερατό φιλτράρισμα.

$$H_t[m, n] = \frac{1}{\sqrt{2}}(A_{2t+1}[m, n] - A_{2t}[m, n]) \quad (3.2)$$

$$L_t[m, n] = \sqrt{2}A_{2t}[m, n] + H_t[m, n]$$

Μια σχηματική αναπαράσταση της Haar ανάλυσης [4], όπου βλέπουμε και τα βήματα πρόβλεψης και ανανέωσης φαίνεται στο διάγραμμα της Εικόνας 3.6. Το διάγραμμα αυτό είναι ισοδύναμο με τις παραπάνω εξισώσεις αρκεί να πολλαπλασιάσουμε με κάποιο παράγοντα κανονικοποίησης τα πλαίσια που προκύπτουν (με  $\sqrt{2}$  το χαμηλοπερατό πλαίσιο και με  $\frac{1}{\sqrt{2}}$  το υψιπερατό πλαίσιο). Επιπλέον, στο διάγραμμα απεικονίζονται τα πλαίσια  $L_t$ ,  $H_t$  που προέκυψαν.



Εικόνα 3.6 Διάγραμμα εφαρμογής του μετασχηματισμού Haar

Αντί για την Haar ανάλυση θα μπορούσαμε να έχουμε οποιοδήποτε άλλο μετασχηματισμό wavelet χρησιμοποιώντας τα αντίστοιχα βήματα πρόβλεψης και ανανέωσης. Ένας άλλος πολύ ενδιαφέρων και διαδεδομένος wavelet μετασχηματισμός είναι ο 5/3. Τα βήματα του σχήματος ανόρθωσης είναι τα εξής:

$$H_t[m, n] = A_{2t+1}[m, n] - \frac{1}{2}(A_{2t}[m, n] + A_{2t+2}[m, n])$$

$$L_t[m, n] = A_{2t}[m, n] + \frac{1}{4}(H_{t-1}[m, n] + H_t[m, n])$$

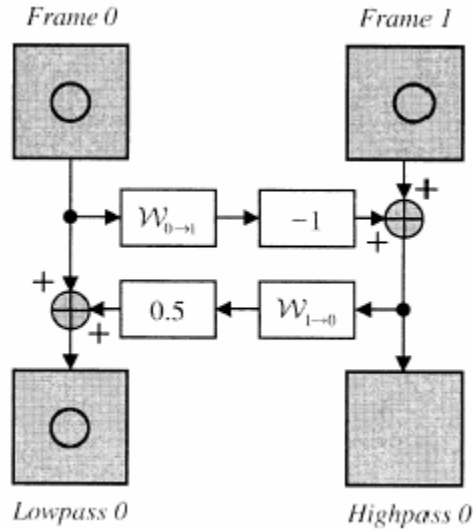
Όπως είπαμε στην αρχή της ενότητας μπορούμε να επεκτείνουμε το παραπάνω πλαίσιο εργασίας ώστε τα βήματα πρόβλεψης και ανανέωσης να περιλαμβάνουν και την εκτίμηση κίνησης [5], [6]. Η ενσωμάτωση της πληροφορίας κίνησης στις wavelet αναλύσεις ήταν για πολύ καιρό αντικείμενο έρευνας. Πρόσφατα, προτάθηκε πως μπορούμε να έχουμε τέλεια ανακατασκευή με το να εφαρμόσουμε στην ακολουθία βίντεο το σχήμα ανόρθωσης (lifting scheme) στην κατεύθυνση της κίνησης. Δηλαδή, αρκεί να αντικαταστήσουμε τα κλασικά βήματα πρόβλεψης και ανανέωσης με την κινησιακά αντισταθμισμένη πρόβλεψη και την κινησιακά αντισταθμισμένη ανανέωση αντίστοιχα.

Έστω  $W_{k_1 \rightarrow k_2}$  η αντιστοίχιση του πλαισίου  $k_1$  στο σύστημα συντεταγμένων του πλαισίου  $k_2$ , έτσι ώστε  $W_{k_1 \rightarrow k_2}(A_{k_1})[m, n] \approx A_{k_2}[m, n]$ . Στην περίπτωση της Haar ανάλυσης τα βήματα της σχέσης (3.2) τροποποιούνται ως εξής:

$$H_t[m, n] = \frac{1}{\sqrt{2}} (A_{2t+1}[m, n] - W_{2t \rightarrow 2t+1}(A_{2t})[m, n])$$

$$L_t[m, n] = \sqrt{2} A_{2t}[m, n] + W_{2t+1 \rightarrow 2t}(H_t)[m, n]$$

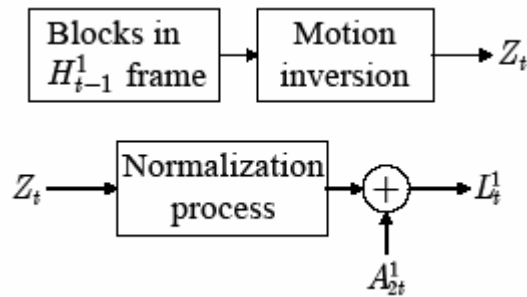
Το ακόλουθο διάγραμμα τώρα για δύο συνεχόμενα πλαίσια είναι αυτό της Εικόνας 3.7. Παρατηρούμε ότι χωρίς την εκτίμηση κίνησης η Haar ανάλυση δημιουργεί τεχνήματα (τα λεγόμενα ghosting artifacts) στη χαμηλοδιαβατή υποζώνη και επιπλέον στην υψηλοδιαβατή ζώνη υπάρχει αρκετή ενέργεια. Αντίθετα, στην περίπτωση που ενσωματώσουμε την εκτίμηση κίνησης τα υψηλοδιαβατά πλαίσια θα έχουν μηδενική ενέργεια με την προϋπόθεση ότι η κίνηση έχει μοντελοποιηθεί σωστά. Επίσης, το χαμηλοδιαβατό πλαίσιο αποτελεί τέλεια αναπαράσταση του αρχικού πλαισίου 0 και δεν περιέχει τεχνήματα.



Εικόνα 3.7 Διάγραμμα εφαρμογής μετασχηματισμού Haar στην κατεύθυνση της κίνησης

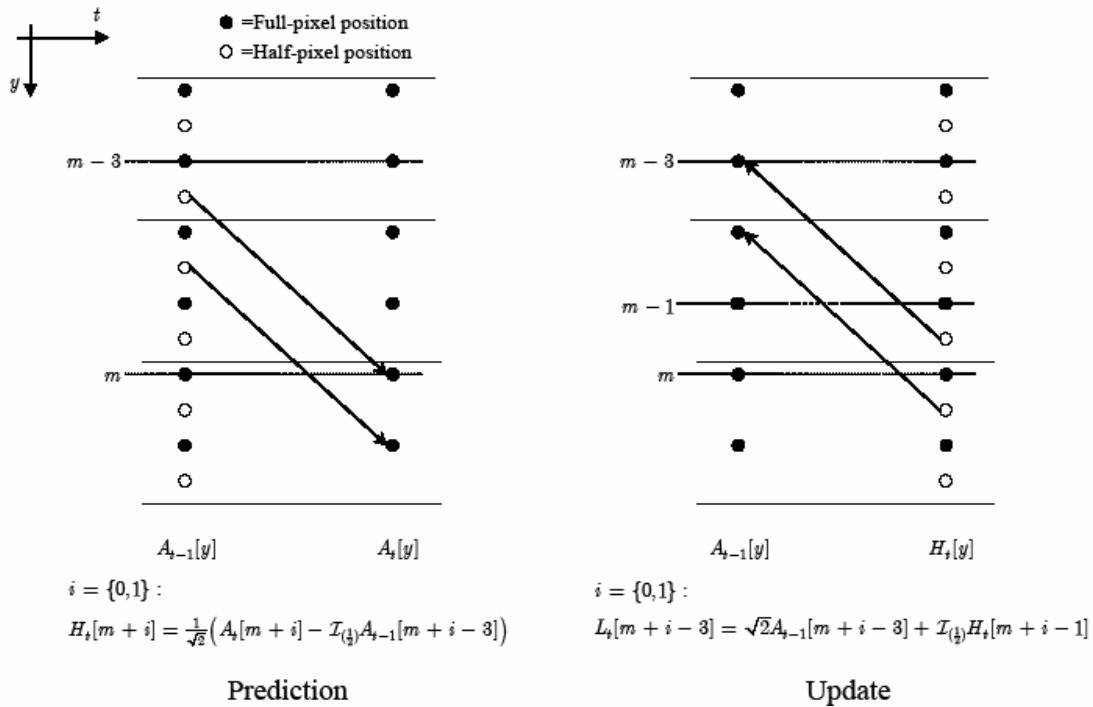
Όπως φαίνεται και παραπάνω χρειαζόμαστε δύο εκτιμήσεις κίνησης την  $W_{2t \rightarrow 2t+1}$  και την  $W_{2t+1 \rightarrow 2t}$ . Μπορούμε όμως να εκτιμήσουμε τις παραμέτρους που περιγράφουν την αντίστροφη κίνηση  $W_{2t+1 \rightarrow 2t}$ , αντιστρέφοντας τις παραμέτρους που έχουν προκύψει από την προς τα εμπρός εκτίμηση  $W_{2t \rightarrow 2t+1}$  [1]. Έτσι η διαδικασία MCU μπορεί να υλοποιηθεί όπως στην Εικόνα 3.8. Αρχικά, δημιουργείται ένα προσωρινό πλαίσιο  $Z_t$  αντιστρέφοντας το πλαίσιο σφάλματος  $H_t$  με την χρήση των αντίστροφων παραμέτρων κίνησης. Το πλαίσιο που προκύπτει κανονικοποιείται και προστίθεται στο πλαίσιο αναφοράς για να προκύψει το ανανεωμένο πλαίσιο  $L_t$ . Αφού τα διανύσματα κίνησης μπορούν να προέρχονται από οποιοδήποτε μπλοκ μέσα στην περιοχή αναζήτησης του πλαισίου

αναφοράς, η αντιστροφή του πλαισίου σφάλματος θα δημιουργήσει εικονοστοιχεία στο πλαίσιο  $L_t$  τα οποία θα είναι ασύνδετα, συνδεδεμένα ή πολυσύνδετα σε εικονοστοιχεία του αντίστοιχου πλαισίου σφάλματος. Η κατάσταση σύνδεσης καθενός εικονοστοιχείου του πλαισίου  $L_t$  εξαρτάται από το αν υπήρχαν ένα, πολλά ή κανένα διανύσματα μετατόπισης προς εκείνη τη θέση του αρχικού πλαισίου αναφοράς. Η διαδικασία κανονικοποίησης διαιρεί την τιμή των δειγμάτων με τον αριθμό των συνδέσεων κάθε εικονοστοιχείου, έτσι ώστε να αποφύγει τον σχηματισμό τεχνημάτων κίνησης (artifacts) στο πλαίσιο  $L_t$ .



**Εικόνα 3.8** Εφαρμογή του βήματος MCU με αντιστροφή των διανυσμάτων κίνησης

Για να γίνει πιο κατανοητή η διαδικασία αντιστροφής της κίνησης θα παρουσιάσουμε ένα απλό παράδειγμα μίας διάστασης ενός συνδεδεμένου μπλοκ (Εικόνα 3.9). Κατά τη διάρκεια του βήματος πρόβλεψης το μπλοκ στη θέση  $m$  προβλέπεται από το μπλοκ στη θέση  $m - 2.5$  του παρεμβαλλόμενου πλαισίου αναφοράς  $IA_{t-1}$ . Δηλαδή το διάνυσμα μετατόπισης για το συγκεκριμένο μπλοκ έχει την τιμή  $-2.5$ . Έπειτα δημιουργείται το πλαίσιο σφάλματος  $H_t$ . Το βήμα ανανέωσης αντιστρέφει την πληροφορία του πλαισίου σφάλματος στο μπλοκ της θέσης  $m - 3$  του πλαισίου αναφοράς  $A_{t-1}$ , δηλαδή στην αμέσως προηγούμενη ακέραια θέση. Έτσι το μπλοκ στη θέση  $m - 3$  του πλαισίου  $A_{t-1}$  θα ανανεωθεί από τις τιμές του μπλοκ του πλαισίου σφάλματος που βρίσκεται στη θέση  $m - 3 + 2.5 = m - 0.5$ . Με αυτόν τον τρόπο δημιουργείται το ανανεωμένο μπλοκ του πλαισίου  $L_t$  στη θέση  $m - 3$ . Ωστόσο, όπως αναφέρθηκε προηγουμένως επιθυμούμε να διαιρέσουμε με τον αριθμό των συνδέσεων, γι' αυτό εφαρμόζουμε την αντιστροφή σε ένα προσωρινό πλαίσιο  $Z_t$  και μετά το κανονικοποιούμε και το προσθέτουμε στο πλαίσιο αναφοράς  $A_{t-1}$ , όπως είδαμε στην Εικόνα 3.8.



**Εικόνα 3.9** Εφαρμογή του βήματος ανανέωσης στην κατεύθυνση της κίνησης. Το συγκεκριμένο παράδειγμα είναι σε μια διάσταση, με μέγεθος μπλοκ δύο και ακρίβεια μισού εικονοστοιχείου

Συνοψίζοντας, ο αλγόριθμος για την κινησιακά αντισταθμισμένη wavelet ανάλυση ή αλλιώς κινησιακά αντισταθμισμένο χρονικό φίλτράρισμα (Motion Compensated Temporal Filtering (MCTF)), στην περίπτωση που έχουμε Haar wavelet ανάλυση αποτελείται από τα ακόλουθα βήματα:

$$c_u[m, n] = 0, Z_t[m, n] = 0$$

$$H_t[m, n] = \frac{1}{\sqrt{2}} (A_{2t+1}[m, n] - \mathcal{I}_{(i_m^{\mathcal{F}_1(2t)}, i_n^{\mathcal{F}_1(2t)})} A_{2t}[m - \lfloor d_m^{\mathcal{F}_1(2t)} \rfloor, n - \lfloor d_n^{\mathcal{F}_1(2t)} \rfloor])$$

$$i_m^{\text{res}} = \begin{cases} 0, & \text{if } i_m^{\mathcal{F}_1(2t)} = 0 \\ 1 - i_m^{\mathcal{F}_1(2t)}, & \text{otherwise} \end{cases}, i_n^{\text{res}} = \begin{cases} 0, & \text{if } i_n^{\mathcal{F}_1(2t)} = 0 \\ 1 - i_n^{\mathcal{F}_1(2t)}, & \text{otherwise} \end{cases}$$

$$d_m^{\text{res}} = \begin{cases} 0, & \text{if } i_m^{\mathcal{F}_1(2t)} = 0 \\ 1, & \text{otherwise} \end{cases}, d_n^{\text{res}} = \begin{cases} 0, & \text{if } i_n^{\mathcal{F}_1(2t)} = 0 \\ 1, & \text{otherwise} \end{cases}$$

$$Z_t[m - \lfloor d_m^{\mathcal{F}_1(2t)} \rfloor, n - \lfloor d_n^{\mathcal{F}_1(2t)} \rfloor] \leftarrow Z_t[m - \lfloor d_m^{\mathcal{F}_1(2t)} \rfloor, n - \lfloor d_n^{\mathcal{F}_1(2t)} \rfloor] \\ + \mathcal{I}_{(i_m^{\text{res}}, i_n^{\text{res}})} H_t[m - d_m^{\text{res}}, n - d_n^{\text{res}}]$$

$$c_u[m - \lfloor d_m^{\mathcal{F}_1(2t)} \rfloor, n - \lfloor d_n^{\mathcal{F}_1(2t)} \rfloor] \leftarrow c_u[m - \lfloor d_m^{\mathcal{F}_1(2t)} \rfloor, n - \lfloor d_n^{\mathcal{F}_1(2t)} \rfloor] + 1$$

$$L_t[m, n] = \sqrt{2} A_{2t}[m, n] + \frac{1}{\max\{c_u[m, n], 1\}} Z_t[m, n]$$

Ο  $c_u[m, n]$  είναι ένας πίνακας  $M \times N$  που περιέχει τον αριθμό των συνδέσεων σε κάθε εικονοστοιχείο  $(m, n)$  του πλαισίου  $Z_t$ . Αφού πολλές περιοχές στο τρέχων πλαίσιο μπορούν να προβλεφθούν από την ίδια περιοχή του πλαισίου αναφοράς, ο πίνακας αυτός μπορεί να πάρει οποιαδήποτε μη αρνητική τιμή. Από τα παραπάνω βήματα του αλγόριθμου παρατηρούμε ότι αν το εικονοστοιχείο  $A_{2t}[m, n]$  είναι ασύνδετο τότε ισχύει  $Z_t[m, n] = 0$ , οπότε είναι σαν να εφαρμόζουμε τη διαδικασία MCTF χωρίς το βήμα της ανανέωσης. Για να υπολογίσουμε τώρα τον αντίστροφο μετασχηματισμό, αρκεί να αντιστρέψουμε την σειρά των εξισώσεων και να λύσουμε ως προς  $A_{2t}[m, n]$  και  $A_{2t+1}[m, n]$ . Τα βήματα για τον υπολογισμό των  $A_{2t}[m, n]$  και  $A_{2t+1}[m, n]$  είναι τα εξής:



$$c_u[m, n] = 0, Z_t[m, n] = 0$$

$$i_m^{res} = \begin{cases} 0, & \text{if } i_m^{F_1(2t)} = 0 \\ 1 - i_m^{F_1(2t)}, & \text{otherwise} \end{cases}, i_n^{res} = \begin{cases} 0, & \text{if } i_n^{F_1(2t)} = 0 \\ 1 - i_n^{F_1(2t)}, & \text{otherwise} \end{cases}$$

$$d_m^{res} = \begin{cases} 0, & \text{if } i_m^{F_1(2t)} = 0 \\ 1, & \text{otherwise} \end{cases}, d_n^{res} = \begin{cases} 0, & \text{if } i_n^{F_1(2t)} = 0 \\ 1, & \text{otherwise} \end{cases}$$

$$Z_t[m - \lfloor d_m^{F_1(2t)} \rfloor, n - \lfloor d_n^{F_1(2t)} \rfloor] \leftarrow Z_t[m - \lfloor d_m^{F_1(2t)} \rfloor, n - \lfloor d_n^{F_1(2t)} \rfloor] \\ + \mathcal{I}_{(i_m^{res}, i_n^{res})} H_t[m - d_m^{res}, n - d_n^{res}]$$

$$c_u[m - \lfloor d_m^{F_1(2t)} \rfloor, n - \lfloor d_n^{F_1(2t)} \rfloor] \leftarrow c_u[m - \lfloor d_m^{F_1(2t)} \rfloor, n - \lfloor d_n^{F_1(2t)} \rfloor] + 1$$

$$A_{2t}[m, n] = \frac{1}{\sqrt{2}} \left( L_t[m, n] - \frac{1}{\max\{c_u[m, n], 1\}} Z_t[m, n] \right)$$

$$A_{2t+1}[m, n] = \sqrt{2} H_t[m, n] + \mathcal{I}_{(i_m^{F_1(2t)}, i_n^{F_1(2t)})} A_{2t}[m - \lfloor d_m^{F_1(2t)} \rfloor, n - \lfloor d_n^{F_1(2t)} \rfloor].$$

Η ύπαρξη του αντίστροφου μετασχηματισμού και της τέλει ανακατασκευής εξασφαλίζεται πάντα από το γεγονός ότι χρησιμοποιούμε την lifting δομή, η οποία είναι πάντα αντιστρέψιμη. Αυτή είναι και η μεγάλη συνεισφορά της lifting δομής, ότι δηλαδή το σχήμα είναι πάντα αντιστρέψιμο ανεξάρτητα από τον τρόπο με τον οποίο θα γίνει η παρεμβολή [4]. Όπως φαίνεται και από τις παραπάνω εξισώσεις για την ανακατασκευή του  $A_{2t}$  χρησιμοποιούμε το πλαίσιο  $L_t$  και το πλαίσιο  $H_t$  το οποίο περιέχει την πληροφορία για τα παρεμβαλλόμενα εικονοστοιχεία στο  $A_{2t}$ . Επειδή όμως η ίδια παρεμβολή υπάρχει και στο  $L_t$  στην πραγματικότητα ακυρώνεται. Αν δεν εφαρμόζαμε την lifting δομή τότε για να έχουμε τέλεια ανακατασκευή θα έπρεπε η διαδικασία της παρεμβολής να είναι αντιστρέψιμη από μόνη της. Για παράδειγμα στην κλασική περίπτωση της Haar ανάλυσης έχουμε

$$H_t[m, n] = \frac{1}{\sqrt{2}} (A_{2t+1}[m, n] - \mathcal{I}_{(i_m^{F_1(t-1)}, i_n^{F_1(t-1)})} A_{2t}[m - \lfloor d_m^{F_1(t-1)} \rfloor, n - \lfloor d_n^{F_1(t-1)} \rfloor])$$

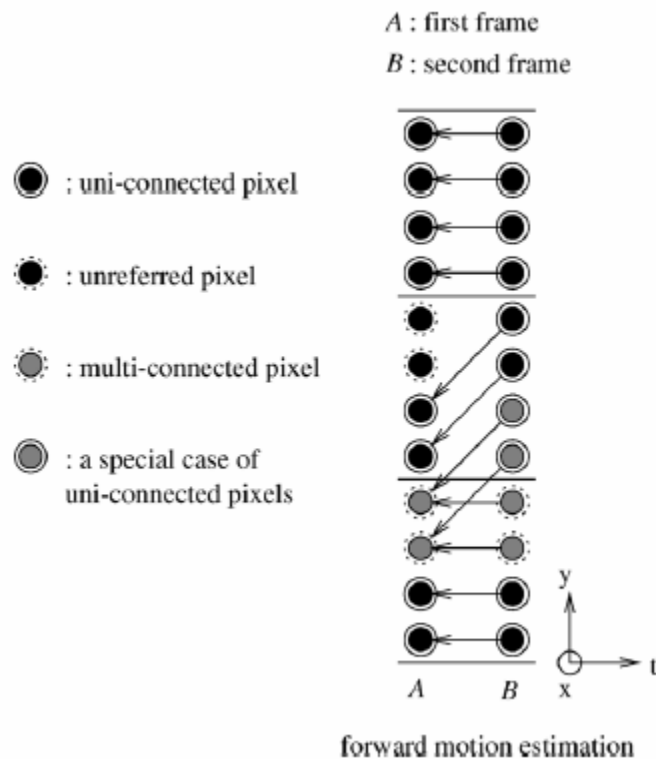
$$L_t[m - \lfloor d_m^{F_1(t-1)} \rfloor, n - \lfloor d_n^{F_1(t-1)} \rfloor] = \frac{1}{\sqrt{2}} (\mathcal{I}_{(i_m^{res}, i_n^{res})} A_{2t+1}[m - d_m^{res}, n - d_n^{res}] + A_{2t}[m - \lfloor d_m^{F_1(t-1)} \rfloor, n - \lfloor d_n^{F_1(t-1)} \rfloor])$$

Για να ανακατασκευάσουμε το  $A_{2t}$  χρειαζόμαστε το  $L_t$  και το  $H_t$  που περιέχει την πληροφορία για τα παρεμβαλλόμενα εικονοστοιχεία του πλαισίου  $A_{2t}$ . Η πληροφορία

αυτή όμως δεν υπάρχει τώρα στο πλαίσιο  $L_t$  και έτσι απαιτείται για τέλεια ανακατασκευή η αντιστρεψιμότητα του αλγόριθμου παρεμβολής.

### Κατάσταση Σύνδεσης Εικονοστοιχείου

Όπως είδαμε προηγουμένως, η αντιστροφή των διανυσμάτων μετατόπισης θα δημιουργήσει εικονοστοιχεία στο πλαίσιο  $L_t$  τα οποία είναι ασύνδετα, συνδεδεμένα ή πολυσύνδετα σε εικονοστοιχεία του αντίστοιχου πλαισίου σφάλματος  $H_t$  [1]. Η κατάσταση σύνδεσης καθενός εικονοστοιχείου του πλαισίου  $L_t$  εξαρτάται από το αν υπήρχαν ένα, πολλά ή κανένα διανύσματα προς εκείνη τη θέση του αρχικού πλαισίου αναφοράς  $A_{2t}$  (πλαίσιο A) από το πλαίσιο  $A_{2t+1}$  (πλαίσιο B). Αν πολλά εικονοστοιχεία από το πλαίσιο B συνδέονται στο ίδιο σημείο του πλαισίου A, τότε εκείνο το σημείο στο πλαίσιο A ονομάζεται πολυσύνδετο. Από τα εικονοστοιχεία του πλαισίου B που συνδέονται στο ίδιο εικονοστοιχείο του πλαισίου A, μόνο ένα από αυτά κατηγοριοποιείται ως συνδεδεμένο ενώ τα άλλα ως ασύνδετα. Στην Εικόνα 3.10 παρουσιάζονται όλες οι δυνατές καταστάσεις ενός εικονοστοιχείου.



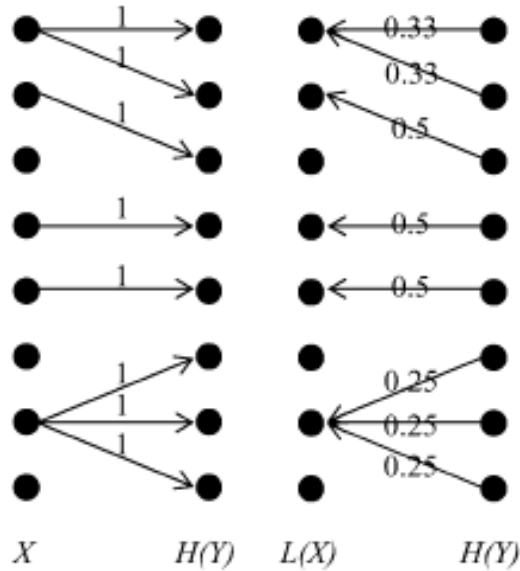
Εικόνα 3.10 Καταστάσεις σύνδεσης ενός εικονοστοιχείου

Το πως θα χειριστούμε αυτά τα ασύνδετα και πολυσύνδετα εικονοστοιχεία έχει μεγάλη σημασία για την ποιότητα του πλαισίου  $L_t$  και στο κέρδος που θα έχουμε από την κωδικοποίηση. Αρχικά ο Ohm και έπειτα οι Choi και Woods [6] πρότειναν για τα

ασύνδετα εικονοστοιχεία στο πλαίσιο αναφοράς A, έστω  $a$  η τιμή τους, να βάζουμε τη τιμή  $\sqrt{2}a$  στο αντίστοιχο εικονοστοιχείο του πλαισίου  $L_t$ . Για τα ασύνδετα εικονοστοιχεία στο πλαίσιο B, βάζουμε την τιμή  $\frac{b-a}{\sqrt{2}}$  στο αντίστοιχο σημείο του πλαισίου  $H_t$ , όπου  $a$  και  $b$  οι τιμές στο σημείο των πλαισίων A και B αντίστοιχα. Στα συνδεδεμένα εικονοστοιχεία εφαρμόζουμε κανονικά το κινησιακά αντισταθμισμένο χρονικό φίλτράρισμα (Motion Compensated Temporal Filtering).

Για τα ασύνδετα εικονοστοιχεία στο πλαίσιο A όλες οι προτεινόμενες μέθοδοι βάζουν την τιμή  $\sqrt{2}a$  στο αντίστοιχο εικονοστοιχείο του πλαισίου  $L_t$  και διαφοροποιούνται στον χειρισμό των ασύνδετων εικονοστοιχείων του πλαισίου B. Στις συνηθισμένες μεθόδους MCTF από τα πολλά εικονοστοιχεία του πλαισίου B που συνδέονταν στο ίδιο σημείο του πλαισίου A, το ποιο θα επιλεγόταν ως συνδεδεμένο εξαρτιόταν από την σειρά ανάλυσης (scan order). Δηλαδή, ως συνδεδεμένο θεωρούταν το πρώτο από τα σημεία του πλαισίου B που συναντούσαμε. Η μέθοδος αυτή έχει το βασικό μειονέκτημα ότι μπορεί να επιλέξει ως συνδεδεμένο ένα «κακό» εικονοστοιχείο. Έτσι, και η διαφορά  $\frac{b-a}{\sqrt{2}}$  θα έχει μεγάλη τιμή και οι χρονικές χαμηλοδιαβατές υποζώνες θα έχουν κακή οπτική ποιότητα. Η καλή ποιότητα των  $L_t$  μας ενδιαφέρει πολύ, γιατί συχνά τα πλαίσια αυτά χρησιμοποιούνται ως δεδομένα στην περίπτωση που ο δέκτης επιθυμεί πλαίσια χαμηλού χρονικού ρυθμού (low-frame-rate). Πρόσφατα, ο Chen και Woods πρότειναν στην περίπτωση που πολλά σημεία του πλαισίου B δείχνουν στο ίδιο σημείο του πλαισίου A, να επιλέγουμε ως συνδεδεμένο αυτό με την μικρότερη τιμή  $\frac{b-a}{\sqrt{2}}$ .

Μια πιο πρόσφατη πρόταση από τους Girod και Han [11], ακολουθεί την εξής διαδικασία. Αν κάποιο σημείο στο πλαίσιο A είναι πολυσύνδετο τότε για τη δημιουργία του  $L_t$  στο σημείο εκείνο, προσθέτουμε στην τιμή του εικονοστοιχείου όλα τα σημεία του πλαισίου  $H_t$  που δείχνουν προς αυτό με βάρος  $1/(M+1)$ . Τα συνδεδεμένα και ασύνδετα σημεία τα χειριζόμαστε όπως και προηγουμένως. Όλες οι δυνατές καταστάσεις απεικονίζονται στην Εικόνα 3.11.



Εικόνα 3.11 Πρόταση των Girod και Han για το χειρισμό των πολυσύνδετων εικονοστοιχείων

Ο τρόπος χειρισμού των πολυσύνδετων εικονοστοιχείων, μελετήθηκε και από τους Tillier, Pesquet-Popescu και Mihaela van der Schaar [10]. Στην εργασία εκείνη έδειξαν τους διαφορετικούς τρόπους χειρισμού των πολυσύνδετων εικονοστοιχείων. Μελέτησαν και το μη γραμμικό τρόπο συνδυασμού των  $M$  σημείων του πλαισίου  $B$  που σχετίζονται με το ίδιο σημείο του πλαισίου  $A$ . Μία μη γραμμική μέθοδος είναι η εφαρμογή του μέσου φίλτρου (median filter) στις τιμές των  $M$  σημείων. Από τις πειραματικές μελέτες που έγιναν φάνηκε ότι η καλύτερη αντιμετώπιση είναι το κανονικοποιημένο άθροισμα, δηλαδή να διαιρούμε την τιμή καθενός από τα  $M$  σημεία του πλαισίου  $H_t$  με  $1/M$  και να τα προσθέτουμε μεταξύ τους. Διατυπώνοντας τα παραπάνω μαθηματικά [10] αν  $p$  το πολυσύνδετο σημείο και  $m_i$  τα σημεία από το πλαίσιο  $B$  που συνδέονται σε αυτό τότε

$$L_t(p) = \sqrt{2}p + \sum_{i=1}^M a_i H_t(m_i) \quad \text{με } a_i = 1/M. \quad \text{Για τιμές στο πλαίσιο } H_t \text{ στα σημεία } m_i,$$

ισχύει  $H_t(m_i) = \frac{1}{\sqrt{2}}(m_i - p)$ . Το κανονικοποιημένο άθροισμα είναι και ο τρόπος που

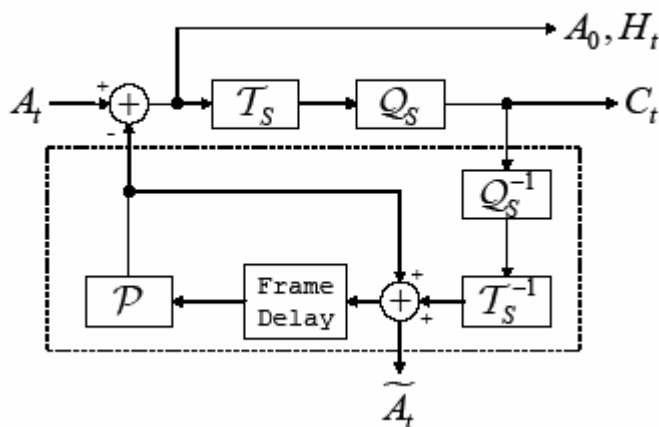
επιλέχτηκε σε αυτήν τη διπλωματική εργασία να χειριζόμαστε τα πολυσύνδετα σημεία. Στον αλγόριθμο που δώσαμε προηγουμένως για την wavelet Haar ανάλυση είναι φανερό ότι χρησιμοποιούμε την παραπάνω μέθοδο για την ανανέωση του πλαισίου  $L_t$ .

### 3.4 Δομές Χρονικής Πρόβλεψης για Κωδικοποίηση Βίντεο

Στην ενότητα αυτή θα παρουσιάσουμε τις δύο δομές [1] που χρησιμοποιούνται στα συστήματα κωδικοποίησης βίντεο με κλιμακοθετήσιμες ιδιότητες: την κλειστού βρόχου (closed-loop) δομή και τη πιο πρόσφατη ανοιχτού βρόχου (open-loop) δομή που υλοποιεί την χρονική ανάλυση χρησιμοποιώντας το κινησιακά αντισταθμισμένο χρονικό φίλτράρισμα (MCTF), όπως περιγράφηκε στην προηγούμενη παράγραφο.

Αρχικά θα μελετήσουμε τον τρόπο λειτουργίας των συστημάτων κωδικοποίησης κλειστού βρόχου. Όλα τα σημερινά προτυποποιημένα σχήματα κωδικοποίησης βίντεο βασίζονται σε μια δομή στην οποία ο δισδιάστατος χωρικός μετασχηματισμός και η κβαντοποίηση εφαρμόζονται στο πλαίσιο σφάλματος που προκύπτει από τον κλειστό βρόχο στον οποίο γίνεται η πρόβλεψη. Μια απλή δομή που περιγράφει τέτοιες αρχιτεκτονικές απεικονίζεται στην Εικόνα 3.12. Η λειτουργία της χρονικής πρόβλεψης  $P$  συνήθως περιέχει την βασιζόμενη σε μπλοκ κινησιακά αντισταθμισμένη πρόβλεψη (MCP). Ο αποκωδικοποιητής λαμβάνει το συμπιεσμένο πλαίσιο σφάλματος  $C_t$  και την πληροφορία για τα διανύσματα κίνησης και εφαρμόζει τον ίδιο βρόχο χρησιμοποιώντας την λαμβάνουσα πληροφορία ώστε να ξαναγίνει η κινησιακά αντισταθμισμένη πρόβλεψη μέσα στον τελεστή  $P$ . Έτσι, το επαναδημιουργημένο πλαίσιο την χρονική στιγμή  $t$  μπορεί να γραφτεί ως:

$$\tilde{A}_t = P\tilde{A}_{t-1} + T_s^{-1}Q_s^{-1}C_t, \quad \tilde{A}_0 = T_s^{-1}Q_s^{-1}C_0 \quad (3.3)$$

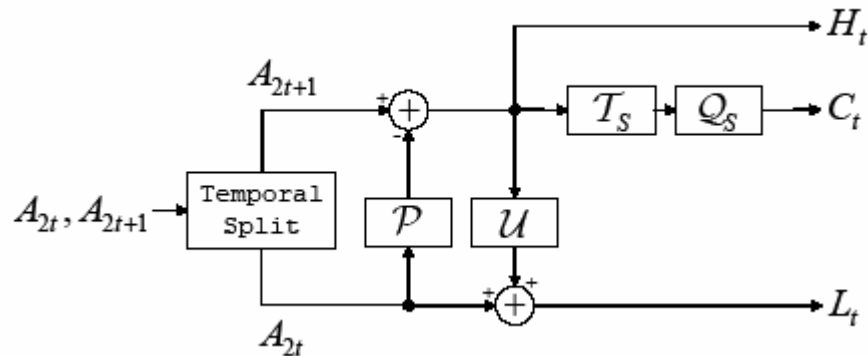


Εικόνα 3.12 Σύστημα κωδικοποίησης βίντεο κλειστού βρόχου

Βλέπουμε επίσης ότι το πρώτο πλαίσιο  $A_0$  στέλνεται απευθείας στον αποκωδικοποιητή και ότι τα επόμενα πλαίσια υπολογίζονται στον αποκωδικοποιητή αναδρομικά, δηλαδή

για τον υπολογισμό του  $A_t$  χρειαζόμαστε το πλαίσιο  $A_{t-1}$ , όπως φαίνεται και στην σχέση 3.3. Η αναδρομή αυτή δημιουργεί το γνωστό πρόβλημα ολίσθησης (drift) μεταξύ του κωδικοποιητή και του αποκωδικοποιητή αν η πληροφορία που χρησιμοποιούν είναι διαφορετική, για παράδειγμα αν κάποια χρονική στιγμή το  $C_t$  που λαμβάνει ο αποκωδικοποιητής είναι διαφορετικό του  $Q_S T_S H_t$  που έχει στείλει ο κωδικοποιητής. Αυτό συμβαίνει συχνά σε πρακτικά συστήματα, αφού είτε υπάρχουν λάθη στη μετάδοση είτε υπάρχει απώλεια στα συμπιεσμένα δεδομένα λόγω της περιορισμένης χωρητικότητας του καναλιού. Τέτοια προβλήματα συναντάμε στα ασύρματα ή στα βασισμένα σε IP δίκτυα, όπου ένας αριθμός πελατών ανταγωνίζονται για τους διαθέσιμους πόρους του δικτύου. Κάποιες προτάσεις για τη λύση του προβλήματος προσπαθούν να απαλείψουν την ολίσθηση μειώνοντας τεχνητά στην πλευρά του κωδικοποιητή το ρυθμό δεδομένων της πληροφορίας  $C_t$  σε ένα βασικό στρώμα στο οποίο το δίκτυο μπορεί να εγγυηθεί τη σωστή μετάδοση. Ένα παράδειγμα τέτοιου αποκωδικοποιητή είναι ο MPEG-4 FGS. Το πρόβλημα με αυτόν τον τρόπο αντιμετώπισης είναι η μείωση της αποδοτικότητας της κωδικοποίησης.

Έτσι, πολλές άλλες προτάσεις οδήγησαν τελικά σε ένα ανοιχτού βρόχου σύστημα, όπως βλέπουμε στην Εικόνα 3.13. Το σύστημα αυτό εφαρμόζει το μετασχηματισμό wavelet στο χρόνο με χρήση και της αντιστάθμισης κίνησης, εφαρμόζει δηλαδή τη διαδικασία (MCTF) που μελετήσαμε εκτενώς στην προηγούμενη ενότητα.



Εικόνα 3.13 Σύστημα κωδικοποίησης βίντεο ανοιχτού βρόχου

Το σύστημα αυτό αρχικά χωρίζει τα πλαίσια σε ζυγές και μονές ακολουθίες. Έπειτα με την χρήση χρονικής πρόβλεψης η πληροφορία στο πλαίσιο  $A_{2t+1}$  αντιστοιχίζεται με την πληροφορία στο πλαίσιο  $A_{2t}$ . Έπειτα, ο τελεστής  $U$  αντιστρέφει την πληροφορία του πλαισίου σφάλματος που έχει δημιουργηθεί από την πρόβλεψη και την προσθέτει στο πλαίσιο  $A_{2t}$ . Έτσι για κάθε ζεύγος εισόδου  $A_{2t}$ ,  $A_{2t+1}$  παράγεται στην έξοδο ένα πλαίσιο σφάλματος  $H_t$  και το ανανεωμένο πλαίσιο  $L_t$ . Το πλαίσιο σφάλματος  $H_t$  αντιστοιχεί σε

ένα υψηλοδιαβατό φιλτράρισμα της πληροφορίας των πλαισίων  $A_{2t}$ ,  $A_{2t+1}$ , ενώ το πλαίσιο  $L_t$  σε ένα χαμηλοδιαβατό. Ο τελεστής  $U$  εκτελεί την κινησιακά αντισταθμισμένη ανανέωση είτε χρησιμοποιώντας το αντίστροφο διάνυσμα που δημιούργησε ο τελεστής πρόβλεψης  $P$ , είτε δημιουργεί ένα νέο διάνυσμα κάνοντας προς τα πίσω (backward) εκτίμηση κίνησης. Η παραπάνω διαδικασία μπορεί να επαναληφθεί θεωρώντας ως είσοδο τώρα τα πλαίσια  $L_t$ , τα οποία είναι τα μισά από τα αρχικά πλαίσια  $A$ . Έτσι, μπορούμε να κάνουμε χρονική ανάλυση σε όσα επίπεδα επιθυμούμε. Ο αποκωδικοποιητής της ανοιχτού βρόχου δομής εκτελεί την αντίστροφη διαδικασία. Αρκεί να εφαρμόσουμε ότι στην Εικόνα 3.18, αλλά από δεξιά προς τα αριστερά και αφού έχουμε αντιστρέψει τα πρόσημα των τελεστών  $P$ ,  $U$ . Τέλος, ενώνουμε σε μια ακολουθία τα πλαίσια  $A_{2t}$  και  $A_{2t+1}$  που έχουν προκύψει. Επομένως, στον αποκωδικοποιητή θα έχουμε:

$$\tilde{A}_{2t} = \tilde{L}_t - UT_s^{-1}Q_s^{-1}C_t, \quad \tilde{A}_{2t+1} = P\tilde{A}_{2t} + T_s^{-1}Q_s^{-1}C_t$$

όπου με  $\tilde{A}_{2t}$ ,  $\tilde{A}_{2t+1}$  συμβολίζουμε τα αναδημιουργημένα πλαίσια τις χρονικές στιγμές  $2t$ ,  $2t + 1$ . Όπως βλέπουμε, ακόμα και αν σε κάποια χρονική στιγμή  $C_t \neq Q_s T_s H_t$  στον αποκωδικοποιητή, το σφάλμα αυτό επηρεάζει μόνο τη δημιουργία των πλαισίων  $\tilde{A}_{2t}$ ,  $\tilde{A}_{2t+1}$  και δε διαδίδεται γραμμικά στις επόμενες ακολουθίες του βίντεο, όπως συνέβαινε στην κλειστού βρόχου δομή. Η διάδοση του σφάλματος μπορεί να συμβεί μόνο κατά τα διαδοχικά χρονικά επίπεδα από τα επαναδημιουργημένα πλαίσια  $\tilde{L}_t$ . Όμως, έπειτα από τη δημιουργία της χρονικής ανάλυσης, μπορεί να εφαρμοστεί σε κάθε ομάδα εικόνων (Group Of Pictures) εμφωλιασμένη κωδικοποίηση βάζοντας έτσι σε προτεραιότητα την πληροφορία των υψηλότερων χρονικών επιπέδων. Έτσι, το αποτέλεσμα της διάδοσης του σφάλματος στην χρονική πυραμίδα είναι περιορισμένο και δεν έχει σημαντικές επιπτώσεις στην ποιότητα του βίντεο. Πειραματικές μετρήσεις που έγιναν επιβεβαίωσαν ότι οι κλιμακοθετήσιμοι MCTF κωδικοποιητές βίντεο που βασίζονται στην αρχιτεκτονική ανοιχτού βρόχου μπορούν να συγκριθούν στον ρυθμό παραμόρφωσης με ισοδύναμους μη κλιμακοθετήσιμους κωδικοποιητές που χρησιμοποιούν τη δομή κλειστού βρόχου.

### 3.5 Προχωρημένη Κινησιακά Αντισταθμισμένη Πρόβλεψη και Ανανέωση

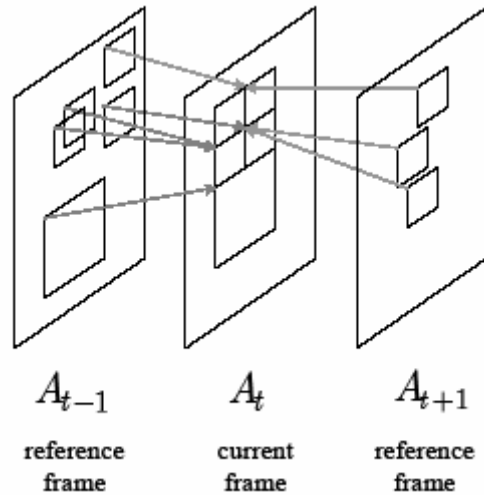
Τα σχήματα MCP και MCU που παρουσιάστηκαν στις προηγούμενες παραγράφους, αν και εφαρμόζονται στην πλειοψηφία των σημερινών πλατφόρμων, έχουν κάποια μειονεκτήματα. Περιορίζοντας τον αλγόριθμο σε ένα σταθερό μέγεθος μπλοκ, δεν είναι δυνατή η προσαρμογή στο συνεχώς μεταβαλλόμενο περιεχόμενο της σκηνής. Επιπλέον, με το να περιορίζουμε το εύρος αναζήτησης σε ένα ή δύο πλαίσια αναφοράς δεν μπορούμε να χειριστούμε αποδοτικά κάποια φαινόμενα που σχετίζονται με τη λήψη του βίντεο. Τέλος, η ανεξάρτητη βελτιστοποίηση της πρόβλεψης για κάθε μπλοκ μπορεί να προκαλέσει ασυνέχεια των εικονοστοιχείων στα σύνορα των μπλοκ, το οποίο και δημιουργεί τεχνήματα στα όρια των μπλοκ (blocking artifacts).

Για να μειωθούν τα προαναφερθέντα προβλήματα δημιουργήθηκαν κάποιες επεκτάσεις του βασικού MCP σχήματος [18] [13], όπως η multihypothesis motion-compensated prediction (MH-MCP) ή αλλιώς motion-compensated prediction with superimposed signals. Ένα χαρακτηριστικό παράδειγμα MH-MCP [1] φαίνεται στην Εικόνα 3.14. Σε αυτήν την περίπτωση, μη επικαλυπτόμενα μπλοκ διαφόρων μεγεθών προβλέπονται από ένα μελλοντικό και ένα προηγούμενο πλαίσιο αναφοράς. Κάθε πρόβλεψη είναι υπέρθεση μπλοκ από τα πλαίσια αναφοράς. Στην γενική περίπτωση που έχουμε MH-MCP χρησιμοποιώντας  $T$  πλαίσια αναφοράς των χρονικών στιγμών  $q = \{t - t^{init}, \dots, t - 1, t + 1, \dots, t + t^{end}\}$ , το πλαίσιο σφάλματος δίνεται από την σχέση:

$$H_t[m,n] = A_t[m,n] - \sum_{q=t-t^{init}}^{t-1} (w_q[m,n] A_q[m - d_m^{F_{t-q}(q)}, n - d_n^{F_{t-q}(q)}]) - \sum_{q=t+1}^{t+t^{end}} (w_q[m,n] A_q[m - d_m^{B_{q-t}(q)}, n - d_n^{B_{q-t}(q)}])$$

Στην παραπάνω σχέση με  $w_q[m,n]$  συμβολίζουμε τα χωρικά μεταβαλλόμενα βάρη του πλαισίου αναφοράς  $A_q$ . Με  $d_m^{F_{t-q}(q)}$ ,  $d_n^{F_{t-q}(q)}$  και  $d_m^{B_{q-t}(q)}$ ,  $d_n^{B_{q-t}(q)}$  συμβολίζουμε τα διανύσματα προς τα εμπρός (forward) και προς τα πίσω (backward) πρόβλεψης αντίστοιχα, που αντιστοιχούν στο σημείο  $A_q[m,n]$ .





Εικόνα 3.14 Παράδειγμα MH-MCP

Για MH-MCP κατά μπλοκ, τα βάρη  $w_q[m,n]$  είναι σταθερά για τα σημεία που αντιστοιχούν στο μπλοκ του πλαισίου αναφοράς  $q$ . Ένας κωδικοποιητής που υλοποιεί τη διαδικασία MH-MCP βασιζόμενος στην παραπάνω σχέση θα πρέπει να υπολογίσει τα βέλτιστα  $w_q[m,n]$  και  $d_m^{F_{t-q}(q)}, d_n^{F_{t-q}(q)}$  ή  $d_m^{B_{q-t}(q)}, d_n^{B_{q-t}(q)}$  των  $T$  πλαισίων αναφοράς, ώστε να ελαχιστοποιήσει τη διασπορά του  $H_i[m,n]$ . Στην πράξη, όλοι οι προτεινόμενοι αλγόριθμοι περιορίζουν την πολυπλοκότητα με το να επιτρέπουν μικρές τιμές για το  $T$  και συγκεκριμένα μεγέθη μπλοκ. Για παράδειγμα, στο πρότυπο H.264, τα επιτρεπτά μεγέθη μπλοκ περιορίζονται σε  $B_m, B_n = \{4,8,16\}$  ενώ η μεγαλύτερη τιμή για το  $T$  που έχει χρησιμοποιηθεί σε πειραματικές μετρήσεις είναι  $T = 5$ .

### 3.6 Χωρικός Μετασχηματισμός Wavelet

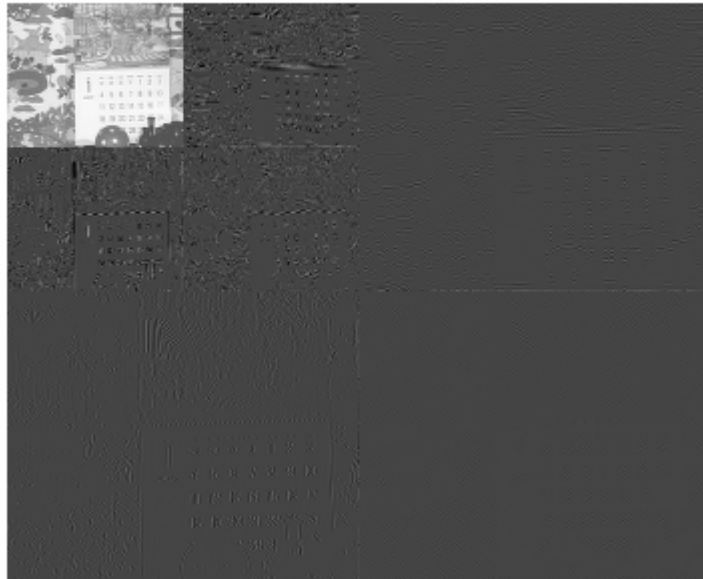
Έπειτα από το χρονικό μετασχηματισμό wavelet που εφαρμόζεται στην ακολουθία βίντεο ακολουθεί ο χωρικός μετασχηματισμός wavelet των πλαισίων, ώστε να εκμεταλλευτούμε την χωρική συσχέτιση των συντελεστών και να μειώσουμε την πλεονάζουσα πληροφορία. Ο χωρικός μετασχηματισμός εφαρμόζεται σε όλα τα χαμηλοπερατά πλαίσια (πλαίσια  $L_t$ ) και υψιπερατά πλαίσια (πλαίσια  $H_t$ ) που έχουν προκύψει από το χρονικό μετασχηματισμό. Ο αριθμός των επιπέδων του μετασχηματισμού είναι επιλογή του χρήστη. Έτσι, με την χρήση των wavelets προκύπτει μια ακολουθία βίντεο κλιμακοθετήσιμη τόσο στον χρόνο (πλαίσια ανά δευτερόλεπτο) όσο και στην αναλύσή της (διαστάσεις βίντεο). Συνήθως, για το δισδιάστατο χωρικό μετασχηματισμό wavelet χρησιμοποιούνται τα biorthogonal 9/7 Daubechies wavelets. Όπως είναι λογικό, τα φίλτρα με τα οποία γίνεται ο μετασχηματισμός επηρεάζουν την

απόδοση του συστήματος. Στην Εικόνα 3.15 παρουσιάζεται το αποτέλεσμα από εφαρμογή δύο επιπέδων των biorthogonal 9/7 Daubechies wavelets στο πρώτο πλαίσιο της γνωστής ακολουθίας βίντεο Mobile.

Mobile, Frame 1

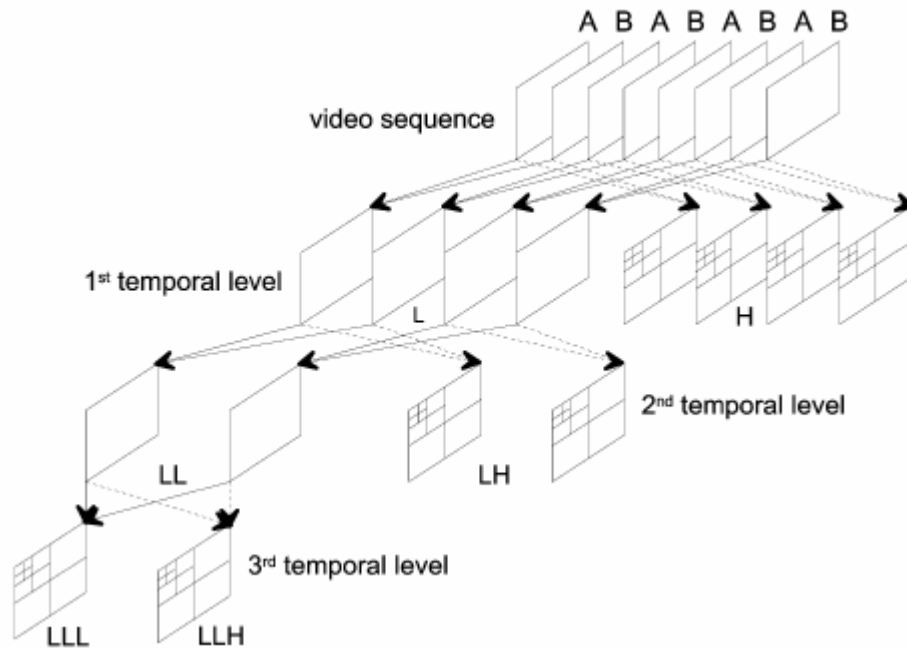


2-level spatial wavelet transform with 9/7 biorthogonal Daubechies



**Εικόνα 3.15** Εφαρμογή του μετασχηματισμού wavelet στο πρώτο πλαίσιο της ακολουθίας Mobile με χρήση των 9/7 biorthogonal Daubechies wavelets.

Συνοψίζοντας, ο μετασχηματισμός wavelet στην κλιμακοθετήσιμη κωδικοποίηση βίντεο εφαρμόζεται τόσο στον χρόνο όσο και στο χώρο. Στο μετασχηματισμό στον χρόνο συνήθως χρησιμοποιείται το Haar wavelet ή τα 5/3 wavelets ενώ για το μετασχηματισμό στο χώρο τα πλέον χρησιμοποιούμενα wavelets είναι τα biorthogonal 9/7. Ένα παράδειγμα της διαδικασίας του χωροχρονικού μετασχηματισμού wavelet παρουσιάζεται την Εικόνα 3.16, όπου η ακολουθία βίντεο αποτελείται από 8 πλαίσια και εφαρμόζουμε χρονικό μετασχηματισμό δύο επιπέδων και χωρικό μετασχηματισμό τριών επιπέδων.



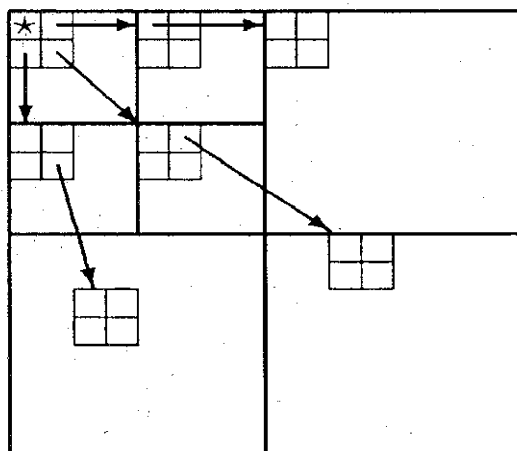
Εικόνα 3.16 Παράδειγμα χωροχρονικού μετασχηματισμού wavelet, 2 επιπέδων στον χρόνο και 3 επιπέδων στο χώρο

### 3.7 Ο Αλγόριθμος SPIHT – 3D SPIHT

Στις προηγούμενες παραγράφους αναλύσαμε το χωροχρονικό μετασχηματισμό wavelet της ακολουθίας βίντεο. Οι συντελεστές που προέκυψαν από το μετασχηματισμό πριν μεταδοθούν από το κανάλι θα πρέπει να κωδικοποιηθούν με κάποιον κατάλληλο αλγόριθμο κωδικοποίησης. Στην παράγραφο αυτή θα περιγράψουμε τον αλγόριθμο κωδικοποίησης SPIHT [14], καθώς και την επέκτασή του στον τρισδιάστατο χώρο τον 3D-SPIHT [15]. Ο αλγόριθμος SPIHT είναι ένας αλγόριθμος που αποθηκεύει τους πιο σημαντικούς συντελεστές με έναν εμφωλιασμένο τρόπο, μαζί με την πληροφορία για τις συντεταγμένες τους. Έτσι, η λήψη των ψηφίων κωδικοποίησης μπορεί να σταματήσει σε

οποιοδήποτε σημείο επιθυμούμε και να ανασχηματιστεί η αρχική εικόνα με την αντίστοιχη ποιότητα. Επιπλέον, το στάδιο της κβαντοποίησης γίνεται έμμεσα, ξεκινώντας με ένα υψηλό κατώφλι το οποίο στη διάρκεια του αλγορίθμου συνεχώς μειώνεται. Ο αλγόριθμος βασίζεται στην τεχνική κωδικοποίησης εικόνας Embedded Zerotree Wavelet (EZW) την οποία εισήγαγε ο Sharipo. Η τεχνική EZW στηρίζεται στα ακόλουθα τρία σημεία: 1) ταξινόμηση των στοιχείων της εικόνας κατά απόλυτη τιμή. Αυτό συμβαίνει γιατί όσο μεγαλύτερη τιμή έχουν τα στοιχεία της εικόνας, τόσο μεγαλύτερη πληροφορία περιέχουν, οπότε πρέπει να μεταδοθούν πρώτα. 2) μετάδοση κατά επίπεδο των δυαδικών ψηφίων (bit plane) 3) εκμετάλλευση της ομοιότητας του wavelet μετασχηματισμού εικόνας στις διαφορετικές κλίμακες. Για παράδειγμα, περιοχές με μεγάλες τιμές στην χαμηλοδιαβατή υποζώνη θα εμφανίζουν μεγάλες τιμές στις αντίστοιχες θέσεις του επόμενου επιπέδου. Αυτό που θέλουμε να κάνουμε είναι να αποθηκεύσουμε όλες τις τιμές των συντελεστών που είναι πάνω από το κατώφλι μαζί με την πληροφορία των συντεταγμένων τους.

Η καινοτομία του αλγορίθμου προέρχεται από τον τρόπο με τον οποίον γίνεται η σάρωση των στοιχείων της εικόνας. Η σάρωση γίνεται με δυναμικό τρόπο ξεκινώντας από τους συντελεστές της χαμηλοδιαβατής υποζώνης. Ο τρόπος λειτουργίας εμπνέεται από το γεγονός ότι στο πεδίο wavelet οι μη σημαντικοί συντελεστές συγκεντρώνονται σε δέντρα, τα οποία ονομάζουμε zerotrees. Έτσι, μπορούμε να αναπαραστήσουμε όλες τις συντεταγμένες ενός υποδέντρου που δεν έχει σημαντικούς συντελεστές σε σχέση με το τρέχον κατώφλι με μόνο ένα ψηφίο. Η χωρική σχέση των συντελεστών απεικονίζεται με μία δομή δέντρου, η οποία ονομάζεται δέντρο χωρικού προσανατολισμού. Για παράδειγμα, ένα τέτοιο δέντρο στην περίπτωση που έχουμε μετασχηματισμό wavelet τεσσάρων επιπέδων απεικονίζεται στην Εικόνα 3.17.



Εικόνα 3.17 Αλγόριθμος SPIHT - Χωρική σχέση των συντελεστών

Κάθε κόμβος στο δέντρο αντιστοιχεί σε ένα εικονοστοιχείο και αναγνωρίζεται από τις συντεταγμένες του. Οι άμεσοι απόγονοι του αντιστοιχούν στα εικονοστοιχεία του ίδιου χωρικού προσανατολισμού του επόμενου επιπέδου της πυραμίδας. Η σχέση αυτή απεικονίζεται με τα αντίστοιχα βέλη. Οι έμμεσοι απόγονοι ενός εικονοστοιχείου είναι όλοι οι απόγονοι των απογόνων του. Από τον ορισμό του δέντρου χωρικού προσανατολισμού παρατηρούμε ότι κάθε κόμβος είτε έχει τέσσερις απογόνους είτε κανέναν. Τα εικονοστοιχεία του υψηλότερου επιπέδου της πυραμίδας είναι οι ρίζες των δέντρων. Επιπλέον, παρατηρούμε ότι το εικονοστοιχείο στην χαμηλοδιαβατή υποζώνη του τέταρτου επιπέδου είναι μαρκαρισμένο με αστερίσκο, δηλαδή το στοιχείο αυτό δεν είναι ρίζα κανενός δέντρου. Οι συμβολισμοί που χρησιμοποιούνται κατά τη διάρκεια της κωδικοποίησης είναι οι ακόλουθοι.

- $O(i, j)$ : το σύνολο των συντεταγμένων όλων των άμεσων απογόνων του κόμβου  $(i, j)$ .
- $D(i, j)$ : το σύνολο των συντεταγμένων όλων των απογόνων του κόμβου  $(i, j)$ .
- $H$ : το σύνολο των συντεταγμένων όλων των κόμβων στο υψηλότερο επίπεδο.
- $L(i, j) = D(i, j) - O(i, j)$ .

Για παράδειγμα εκτός από το υψηλότερο και το χαμηλότερο επίπεδο της πυραμίδας οι άμεσοι απόγονοι του κόμβου  $(i, j)$  είναι:  $O(i, j) = \{(2i, 2j), (2i, 2j+1), (2i+1, 2j), (2i+1, 2j+1)\}$ .

Κατά την κωδικοποίηση είναι σημαντικό να γνωρίζουμε την σειρά με την οποία τα υποσύνολα ελέγχονται για την σημαντικότητα τους. Για αυτό το λόγο ο αλγόριθμος κωδικοποίησης χρησιμοποιεί τρεις διατεταγμένες λίστες: τη λίστα που περιέχει ασήμαντα σύνολα (LIS), τη λίστα που περιέχει ασήμαντα εικονοστοιχεία (LIP) και τη λίστα που περιέχει τα σημαντικά εικονοστοιχεία. Σε όλες τις λίστες κάθε καταχώρηση αναγνωρίζεται από τις συντεταγμένες της  $(i, j)$ , που στην περίπτωση των LIP και LSP αναπαριστούν τα αντίστοιχα εικονοστοιχεία, ενώ στην λίστα LIS είτε το σύνολο  $D(i, j)$  είτε το  $L(i, j)$ . Για να διαχωρίζουμε τις δύο περιπτώσεις της λίστας LIS διακρίνουμε την καταχώρηση τύπου A αν αναπαριστά το σύνολο  $D(i, j)$  και την καταχώρηση τύπου B αν αναπαριστά το σύνολο  $L(i, j)$ . Ο ψευδοκώδικας του αλγόριθμου είναι ο ακόλουθος:

#### Ψευδοκώδικας

1. Αρχικοποίηση: στείλε στην έξοδο το  $n = \lfloor \log_2(\max_{(i,j)} \{c_{i,j}\}) \rfloor$ . Θέσε την LSP λίστα ως κενή λίστα και πρόσθεσε τις συντεταγμένες  $(i, j) \in H$  στην LIP και μόνο όσους έχουν απογόνους στη λίστα LIS ως τύπου A εγγραφές.
2. Βήμα διατάξης (Sorting Pass):
  - 2.1 για κάθε εγγραφή  $(i, j)$  στη λίστα LIP:

- 2.1.1 στείλε στην έξοδο το  $S_n(i, j)$ .
- 2.1.2 αν  $S_n(i, j) = 1$  τότε μετακίνησε το  $(i, j)$  στη λίστα LSP και στείλε στην έξοδο το πρόσημο του  $c_{i,j}$ .
- 2.2 για κάθε εγγραφή  $(i, j)$  στη λίστα LIS:
- 2.2.1 αν η εγγραφή είναι τύπου A τότε
- στείλε στην έξοδο το  $S_n(D(i, j))$ .
  - αν  $S_n(D(i, j)) = 1$  τότε
    - για κάθε  $(k, l) \in O(i, j)$ 
      - στείλε στην έξοδο το  $S_n(k, l)$ .
      - αν  $S_n(k, l) = 1$  τότε πρόσθεσε το  $(k, l)$  στη λίστα LSP και στείλε στην έξοδο το πρόσημο του  $c_{k,l}$ .
      - αν  $S_n(k, l) = 0$  τότε πρόσθεσε το  $(k, l)$  στο τέλος της λίστας LIP.
  - Αν  $L(i, j) \neq 0$  τότε μετακίνησε το  $(i, j)$  στο τέλος της λίστας LIS, ως εγγραφή τύπου B και πήγαινε στο Βήμα 2.2.2 αλλιώς διέγραψε την εγγραφή  $(i, j)$  από τη λίστα LIS.
- 2.2.2 αν η εγγραφή είναι τύπου B τότε
- στείλε στην έξοδο το  $S_n(L(i, j))$ .
  - αν  $S_n(L(i, j)) = 1$  τότε
    - πρόσθεσε κάθε  $(k, l) \in O(i, j)$  στο τέλος της λίστας LIS ως εγγραφή τύπου A.
    - Διέγραψε το  $(i, j)$  από τη λίστα LIS.
3. Βήμα Εκλέπτυνσης (Refinement Pass): για κάθε εγγραφή  $(i, j)$  στη λίστα LSP εκτός από αυτές που συμπεριλήφθηκαν στο πιο πρόσφατο βήμα διάταξης στείλε στην έξοδο το  $n$ -οστό πιο σημαντικό ψηφίο του  $|c_{i,j}|$ .
4. Ανανέωση βήματος κβαντισμού: μείωσε το  $n$  κατά 1 και πήγαινε στο Βήμα 2.

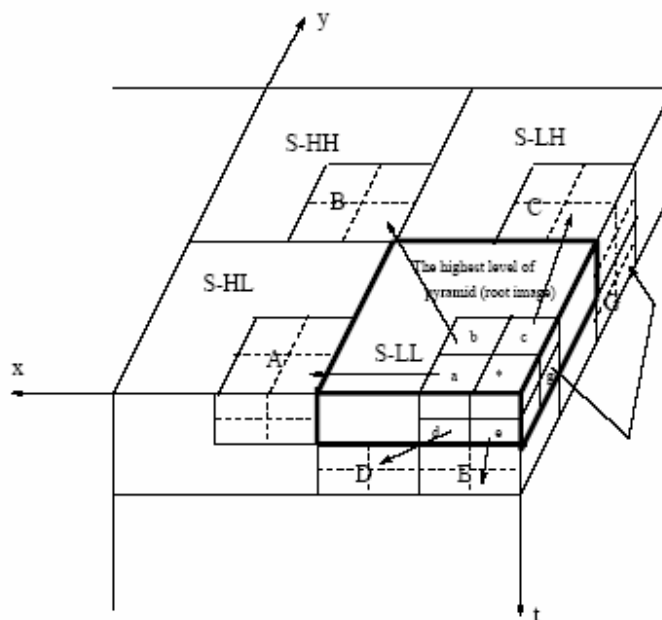
Ως αρχικό κατώφλι επιλέγεται η τιμή του μεγαλύτερου συντελεστή και μετά το πέρας των βημάτων διάταξης και ανανέωσης η τιμή του κατωφλίου μειώνεται στο μισό. Κατά τη διάρκεια του βήματος διάταξης (sorting pass) τα εικονοστοιχεία της λίστας LIP, τα οποία δεν ήταν σημαντικά στο προηγούμενο βήμα διάταξης, ελέγχονται για την σημαντικότητα τους ως προς το τρέχων κατώφλι και όσα είναι σημαντικά μετακινούνται στην λίστα LSP. Όμοια τα σύνολα που περιέχονται στην LIS λίστα ελέγχονται και όσα είναι σημαντικά απομακρύνονται από τη λίστα και διαμελίζονται. Από τα νέα υποσύνολα όσα περιέχουν παραπάνω από ένα στοιχεία προστίθενται στη λίστα LIS, ενώ αυτά με ένα στοιχείο στη λίστα LIP ή τη LSP αναλόγως με το αν είναι ασήμαντα ή σημαντικά αντίστοιχα. Το βήμα εκλέπτυνσης (refinement) εξετάζει όλα τα στοιχεία της LSP λίστας εκτός από αυτά που προστέθηκαν στο πιο πρόσφατο βήμα και αποστέλλει στην έξοδο το αντίστοιχο ψηφίο. Αυτό το βήμα στην πραγματικότητα ανανεώνει της τιμές των συντελεστών. Ο αλγόριθμος εκτελείται μέχρι να επιτευχθεί ο επιθυμητός λόγος συμπίεσης.

Η αποκωδικοποίηση γίνεται με όμοιο τρόπο. Όλα τα κομμάτια εξόδου μετατρέπονται σε είσοδο και όλες οι συγκρίσεις με το κατώφλι μετατρέπονται σε ανανεώσεις στους συντελεστές. Όταν ανανεώνονται οι συντελεστές, η αρχική τους τιμή ανατίθεται σε 1.5 φορές το τρέχων κατώφλι, γιατί το κατώφλι μας λέει ότι η τιμή είναι το λιγότερο όσο το

κατώφλι αλλά όχι περισσότερο από δύο φορές το κατώφλι. Έτσι, υποθέτουμε ότι η τιμή είναι κάπου κοντά στο ενδιάμεσο. Επιπλέον, ο κωδικοποιητής και ο αποκωδικοποιητής έχουν την ίδια ακριβώς πολυπλοκότητα. Όπως και με οποιαδήποτε μέθοδο κωδικοποίησης η αποδοτικότητα του αλγόριθμου μπορεί να βελτιωθεί περνώντας την έξοδο του από έναν κωδικοποιητή εντροπίας, αλλά εις βάρος του χρόνου κωδικοποίησης/αποκωδικοποίησης.

### 3D SPIHT

Ο αλγόριθμος που περιγράψαμε προηγουμένως είναι ο δισδιάστατος SPIHT, που χρησιμοποιείται για κωδικοποίηση εικόνας. Η επέκταση του στον τρισδιάστατο χώρο, η οποία βρίσκει εφαρμογή σε ακολουθίες βίντεο που έχουν υποστεί τρισδιάστατο μετασχηματισμό wavelet, μελετάται σε αυτήν την παράγραφο. Αφού η κωδικοποίηση λαμβάνει χώρα στο τρισδιάστατο χώρο, η ακολουθία βίντεο είναι κλιμακοθετήσιμη τόσο στον χρόνο όσο και στο χώρο. Κάθε κόμβος τώρα θα έχει οκτώ απογόνους αντί για τέσσερις και τα δέντρα θα είναι τρισδιάστατα. Έτσι οι απόγονοι των συντεταγμένων  $(i, j)$  του πλαισίου χρονικής στιγμής  $k$  (εκτός από το υψηλότερο και το χαμηλότερο επίπεδο) θα είναι οι:  $O(i, j, k) = \{(2i, 2j, 2k), (2i, 2j+1, 2k), (2i+1, 2j, 2k), (2i+1, 2j+1, 2k), (2i, 2j, 2k+1), (2i, 2j+1, 2k+1), (2i+1, 2j, 2k+1), (2i+1, 2j+1, 2k+1)\}$ . Για να γίνει πιο κατανοητή η μορφή των δέντρων προσανατολισμού στην τρισδιάστατη περίπτωση απεικονίζουμε γραφικά τους απογόνους του υψηλότερου επιπέδου υποθέτοντας ότι αυτό έχει διαστάσεις  $4 \times 4 \times 2$ .



Εικόνα 3.18 Δένδρα προσανατολισμού στον 3D-SPIHT

Βλέπουμε λοιπόν ότι η διαδικασία κωδικοποίησης και αποκωδικοποίησης παραμένει ίδια. Αυτό που αλλάζει είναι η μορφή των δέντρων προσανατολισμού που είναι τρισδιάστατη.

### 3.8 Ανακεφαλαίωση

Το κεφάλαιο αυτό αποτελεί το βασικό κομμάτι της διπλωματικής εργασίας. Σε αυτό το κεφάλαιο αναλύθηκε η διαδικασία της κλιμακοθετήσιμης κωδικοποίησης βίντεο και τα συστατικά από τα οποία αποτελείται. Είδαμε τον χρονικό και το χωρικό μετασχηματισμό wavelet, καθώς και τον αλγόριθμο κωδικοποίησης των συντελεστών wavelet, τον 3D-SPIHT.

Θα πρέπει να αναφέρουμε ότι σε αυτήν την διπλωματική εργασία μελετήσαμε τη διαδικασία κλιμακοθετήσιμης κωδικοποίησης βίντεο, κατά την οποία εφαρμόζουμε πρώτα το μετασχηματισμό wavelet στο πεδίο του χρόνου και έπειτα στα πλαίσια που προκύπτουν εφαρμόζουμε το χωρικό δισδιάστατο μετασχηματισμό wavelet. Η μέθοδος αυτή ονομάζεται κινησιακά αντισταθμισμένο χρονικά φιλτράρισμα (Motion Compensated Temporal Filtering) ή αλλιώς αρχιτεκτονική  $t + 2D$ . Υπάρχει όμως και μία άλλη κατηγορία κλιμακοθετήσιμης κωδικοποίησης βίντεο, κατά την οποία πρώτα εφαρμόζουμε τον δισδιάστατο μετασχηματισμό wavelet στα πλαίσια του βίντεο και έπειτα εφαρμόζουμε το μετασχηματισμό wavelet στο πεδίο του χρόνου. Έτσι, η εκτίμηση της κίνησης μεταξύ δύο πλαισίων λαμβάνει χώρο στο πεδίο του μετασχηματισμού wavelet. Γι' αυτό και η αρχιτεκτονική αυτή, ονομάζεται in-band αντιστάθμιση κίνησης ή και αρχιτεκτονική  $2D + t$ .

Ένα σημαντικό εμπόδιο στην εκτίμηση της κίνησης στο πεδίο του μετασχηματισμού wavelet είναι ότι ο κλασικός δυαδικός μετασχηματισμός wavelet (DWT) είναι μόνο περιοδικά αμετάβλητος στις μετατοπίσεις (shift-invariant). Για παράδειγμα, αν σε ένα μονοδιάστατο σήμα έχουμε μετατόπιση περιττού αριθμού, τότε δεν μπορούμε να έχουμε εκτίμηση κίνησης με μηδενικό σφάλμα. Αντίθετα, αν έχουμε μετατόπιση ζυγού αριθμού τότε η εκτίμηση κίνησης στο πεδίο wavelet είναι ίδια με την εκτίμηση που θα είχαμε αν εφαρμόζαμε τη διαδικασία στο πεδίο του χρόνου. Ο τρόπος με τον οποίο αντιμετωπίζεται το πρόβλημα αυτό είναι να κατασκευάσουμε για κάθε επίπεδο ανάλυσης από τη δυαδική αναπαράσταση wavelet (DWT) του πλαισίου αναφοράς, την υπερπλήρη δυαδική αναπαράσταση wavelet (Overcomplete DWT), η οποία είναι αμετάβλητη σε μετατοπίσεις. Ο Andreopoulos [1], πρότεινε ένα νέο πλαίσιο εργασίας για τη δημιουργία του ODWT, το οποίο έχει μειωμένο υπολογιστικό κόστος. Με την in-band αρχιτεκτονική [1], [8] οι διαδικασίες MCP και MCU είναι ίδιες, αλλά λαμβάνουν χώρο στο πεδίο του



DWT. Επιπλέον, όπως και στην περίπτωση των  $t + 2D$  συστημάτων, έχουμε αρχιτεκτονικές κλειστού και ανοιχτού βρόχου.

Το βασικό πλεονέκτημα της in-band αρχιτεκτονικής είναι ότι παρέχει καλύτερες δυνατότητες για κλιμακοθετησιμότητα στην ανάλυση (resolution) του βίντεο. Στα συστήματα  $t + 2D$  όταν επιθυμούμε να κάνουμε αποκωδικοποίηση του βίντεο σε χαμηλή ανάλυση, θα πρέπει να υποδειγματοληπτήσουμε τα διανύσματα μετατόπισης. Έτσι, η διαδικασία αντιστάθμισης κίνησης θα είναι διαφορετική στο αποκωδικοποιητή απ' ότι στον κωδικοποιητή. Το πρόβλημα αυτό δεν εμφανίζεται στις in-band αρχιτεκτονικές, γιατί η εκτίμηση κίνησης γίνεται στις υποζώνες. Από πλευράς PSNR, πειραματικές μετρήσεις έδειξαν [1], ότι οι δύο αρχιτεκτονικές εμφανίζουν παρόμοια αποτελέσματα στις περισσότερες ακολουθίες βίντεο, ενώ στις υπόλοιπες η αρχιτεκτονική  $t + 2D$  εμφανίζει καλύτερη απόδοση. Τέλος, ένα μειονέκτημα των in-band αρχιτεκτονικών είναι ότι έχουμε περισσότερο υπολογιστικό κόστος, αφού πρέπει να υπολογίσουμε τον ODWT.

## Κεφάλαιο 4 : Πειραματικές Μετρήσεις

Στο Κεφάλαιο αυτό θα μελετήσουμε πειραματικά τον τρόπο με τον οποίο διάφορες παράμετροι επηρεάζουν την ποιότητα ή την πολυπλοκότητα ενός συστήματος κλιμακοθετήσιμης κωδικοποίησης βίντεο. Οι μέθοδοι για μέτρηση της ποιότητας χωρίζονται σε δύο κατηγορίες: α) αντικειμενικές β) υποκειμενικές. Οι αντικειμενικές μέθοδοι σκοπεύουν στο να υπολογίσουν μαθηματικά την αλλοίωση που υπεισέρχεται στην ακολουθία βίντεο λόγω της συμπίεσης, ενώ οι υποκειμενικές αξιολογούν το πως αντιλαμβάνεται το ανθρώπινο μάτι την ποιότητα του βίντεο. Στις πειραματικές μετρήσεις που ακολουθούν λάβαμε υπόψη και τις δύο κατηγορίες μεθόδων. Συγκεκριμένα ως μέτρο για την αντικειμενική ποιότητα χρησιμοποιήθηκε το Peak Signal-to-Noise Ratio (PSNR) το οποίο δίνεται από τον τύπο

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (A[m,n] - \tilde{A}[m,n])^2}, \text{ όπου } A[m,n] \text{ το αρχικό πλαίσιο και}$$

$\tilde{A}[m,n]$  το ανασχηματισμένο. Η μέτρηση του PSNR έγινε για ρυθμούς μετάδοσης ψηφίων που κυμαίνονται από 250-1250 Kbps, ώστε να καλύπτεται ένα ευρύ φάσμα πρακτικών εφαρμογών.

Για τις πειραματικές μετρήσεις χρησιμοποιήθηκαν δύο CIF (352x288) ακολουθίες βίντεο με ρυθμό ανανέωσης 30 Hz. Οι ακολουθίες βίντεο είναι η Mobile και η Foreman και επιλέχτηκαν έτσι ώστε να αναπαριστούν διαφορετικές κατηγορίες περιεχομένου βίντεο. Συγκεκριμένα, η ακολουθία Mobile περιλαμβάνει πανοραμική λήψη και ποικίλες κινήσεις αντικειμένων, ενώ η ακολουθία Foreman περιέχει ομιλία ανθρώπου χωρίς πολύπλοκες σκηνές. Στις μετρήσεις της παραγράφου 4.3 χρησιμοποιήθηκε και η ακολουθία βίντεο Stefan. Τέλος, πρέπει να αναφέρουμε ότι η διαδικασία κωδικοποίησης και μετάδοσης των ακολουθιών βίντεο έγινε για την συνιστώσα φωτεινότητας (luminance) Y, χωρίς να λαμβάνουμε υπόψη τις χρωματικές συνιστώσες των ακολουθιών.

### 4.1 Σύγκριση full-pixel, sub-pixel

Στην παράγραφο αυτή θα συγκρίνουμε την απόδοση της ακρίβειας μισού εικονοστοιχείου με την απόδοση που έχουμε όταν χρησιμοποιούμε ακρίβεια ενός εικονοστοιχείου. Στην περίπτωση της ακρίβειας μισού εικονοστοιχείου η παρεμβολή έγινε χρησιμοποιώντας το 8-tap φίλτρο που αναφέρεται στο Κεφάλαιο 3.

**Σενάριο**

Αριθμός πλαισίων: 16

Ανάλυση: 352 x 288 (CIF resolution, 30 Hz)

Αριθμός χρονικών επιπέδων μετασχηματισμού: 4

Αριθμός χωρικών επιπέδων μετασχηματισμού: 3

Wavelet χρονικού μετασχηματισμού: Haar wavelet

Wavelet χωρικού μετασχηματισμού: 9/7 biorthogonal wavelets

Μέγεθος μπλοκ: 16 x 16

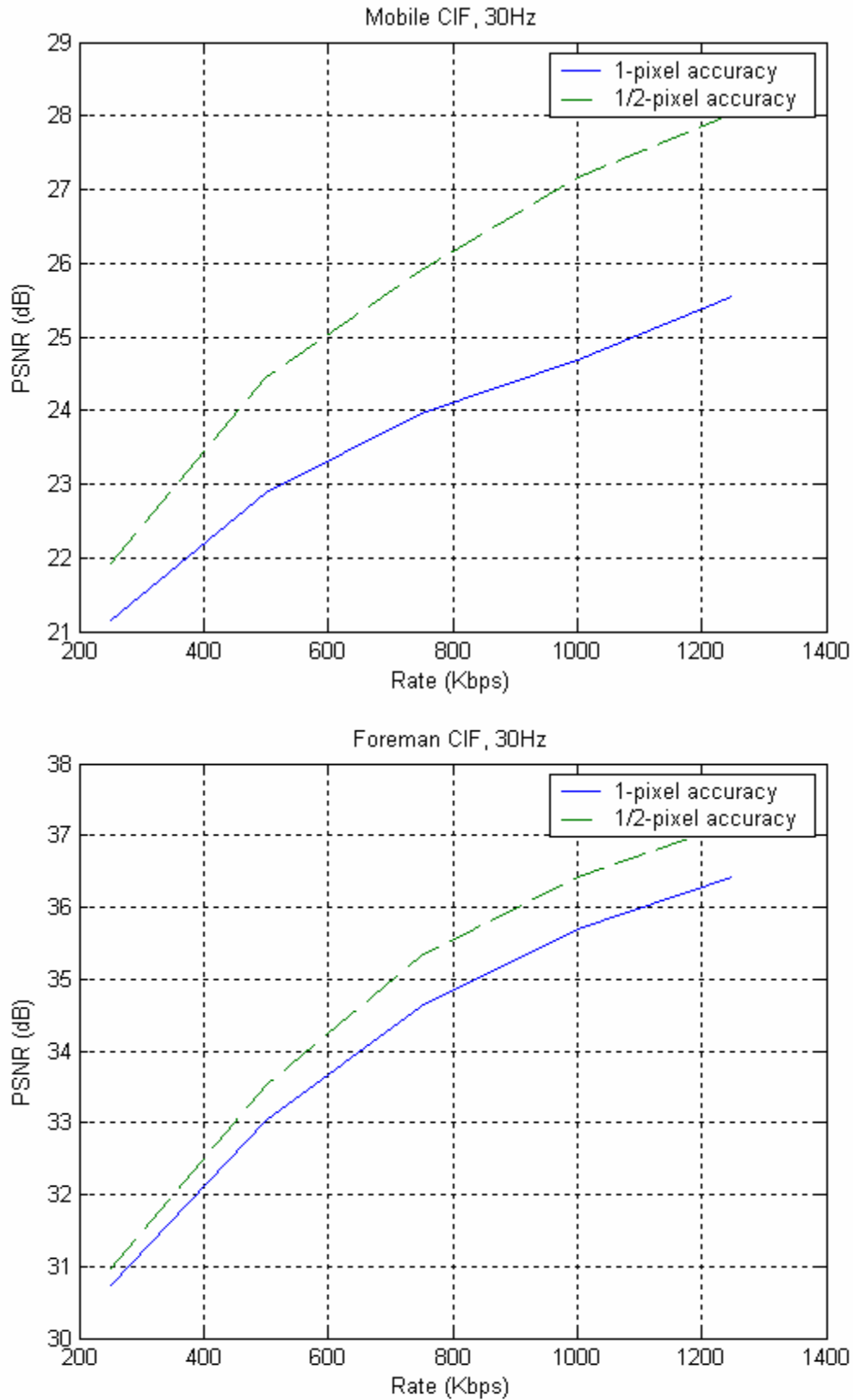
Περιοχή αναζήτησης για εκτίμηση κίνησης:  $\pm 8$  εικονοστοιχεία

***Mobile***

Rate (Kbps)	250	500	750	1000	1250
Full-pixel (PSNR)	21.16	22.89	23.96	24.69	25.56
Half-pixel(PSNR)	21.92	24.46	25.90	27.17	28.04

***Foreman***

Rate (Kbps)	250	500	750	1000	1250
Full-pixel (PSNR)	30.73	33.02	34.62	35.69	36.41
Half-pixel(PSNR)	30.98	33.52	35.31	36.43	37.17



Εικόνα 4.1 Σύγκριση της ακρίβειας ενός και μισού εικονοστοιχείου

Mobile Original Frame 1



( $\alpha$ )

Mobile Frame 1, 1/2-pixel accuracy



( $\beta$ )

Mobile Frame 1, 1-pixel accuracy



(γ)

**Εικόνα 4.2** Αποκωδικοποιημένο πλαίσιο της ακολουθίας βίντεο Mobile (α) Αρχικό πλαίσιο (β) Πλαίσιο ακρίβειας μισού εικονοστοιχείου (γ) Πλαίσιο ακρίβειας ενός εικονοστοιχείου

Πλαίσια σφάλματος  $H_t$

$H_t$ , 1/2-pixel accuracy



(α)

Ht, 1-pixel accuracy



(β)

**Εικόνα 4.3** Πλαίσιο σφάλματος (α) Ακρίβεια μισού εικονοστοιχείου (β) Ακρίβεια ενός εικονοστοιχείου

### Συμπεράσματα

Παρατηρούμε ότι και στις δύο ακολουθίες βίντεο η απόδοση βελτιώνεται χρησιμοποιώντας ακρίβεια μισού εικονοστοιχείου, Εικόνα 4.1. Η βελτίωση αυτή οφείλεται στο γεγονός ότι η μετατόπιση κίνησης του κάθε πλαισίου σε σχέση με το επόμενο πλαίσιο δεν είναι πάντα ακέραιου αριθμού εικονοστοιχείων. Έτσι, όταν χρησιμοποιούμε πλέγμα ακρίβειας μισού εικονοστοιχείου έχουμε καλύτερη ανάλυση και είναι αναμενόμενο η πρόβλεψη να είναι καλύτερη. Αυτό φαίνεται και στην Εικόνα 4.3, όπου βλέπουμε το πλαίσιο σφάλματος  $H_t$  που προέκυψε από την εκτίμηση κίνησης μεταξύ των πλαισίων 1 και 2 της ακολουθίας βίντεο Mobile στην περίπτωση που έχουμε ακρίβεια μισού και ενός εικονοστοιχείου. Όπως παρατηρούμε, το πλαίσιο σφάλματος  $H_t$  έχει αρκετά μικρότερη ενέργεια στην περίπτωση της ακρίβειας μισού εικονοστοιχείου, το οποίο σημαίνει ότι η κίνηση έχει προβλεφθεί αποτελεσματικότερα.

Η βελτίωση της μετρικής PSNR στην ακολουθία βίντεο Mobile είναι αρκετά μεγαλύτερη από ότι στην ακολουθία βίντεο Foreman. Συγκεκριμένα στη Mobile ακολουθία έχουμε βελτίωση από 0.76 dB έως 2.48 dB, ενώ στην ακολουθία Foreman από 0.25 dB έως 0.76 dB ανάλογα με το ρυθμό δυαδικών ψηφίων ανά δευτερόλεπτο. Επίσης, παρατηρούμε ότι

και στις δύο ακολουθίες η βελτίωση είναι μεγαλύτερη στους υψηλούς ρυθμούς δυαδικών ψηφίων απ' ότι στους χαμηλότερους ρυθμούς. Εκτός από τη βελτίωση της μετρικής PSNR, η ακρίβεια μισού εικονοστοιχείου δίνει και οπτικά καλύτερη εικόνα, όπως φαίνεται στην Εικόνα 4.2, όπου απεικονίζεται το πλαίσιο 1 για τις δύο περιπτώσεις και για ρυθμό 1000 Kbps.

## 4.2 Σύγκριση Φίλτρων Παρεμβολής

Όπως αναφέραμε στο Κεφάλαιο 3, το φίλτρο με το οποίο θα γίνει η παρεμβολή της εικόνας στην περίπτωση που έχουμε ακρίβεια μισού εικονοστοιχείου, επηρεάζει την απόδοση του συστήματος. Στην παράγραφο αυτή θα μελετήσουμε και θα συγκρίνουμε τα εξής τρία φίλτρα παρεμβολής:

1. Το κλασσικό φίλτρο παρεμβολής δύο συντελεστών  $[1 \ 1] / 2$ , όπου το παρεμβαλλόμενο εικονοστοιχείο είναι η μέση τιμή των δύο γειτονικών εικονοστοιχείων. Το φίλτρο αυτό αναφέρεται συχνά στη βιβλιογραφία ως Bilinear.
2. Ένα φίλτρο οκτώ συντελεστών, το οποίο προτάθηκε στην εργασία [3] και χρησιμοποιήθηκε και στην εργασία [1]. Οι συντελεστές του φίλτρου είναι:  $[-0.0105 \ 0.0465 \ -0.1525 \ 0.6165 \ 0.6165 \ -0.1525 \ 0.0465 \ -0.0105]$ .
3. Το φίλτρο παρεμβολής έξι συντελεστών, το οποίο χρησιμοποιείται στο πρότυπο H.264 και είναι το  $[1 \ -5 \ 20 \ 20 \ -5 \ 1] / 32$ .

Τα παραπάνω φίλτρα θα αναφέρονται για λόγους συντομίας ως Bilinear, 8-tap, 6-tap αντίστοιχα. Ακολουθούν οι πίνακες και οι γραφικές παραστάσεις για τα τρία φίλτρα παρεμβολής. Οι τιμές για το 8-tap φίλτρο είναι όμοιες με αυτές τις παραγράφου 4.1.

### Σενάριο

Αριθμός πλαισίων: 16

Ανάλυση: 352 x 288 (CIF resolution, 30 Hz)

Αριθμός χρονικών επιπέδων μετασχηματισμού: 4

Αριθμός χωρικών επιπέδων μετασχηματισμού: 3

Wavelet χρονικού μετασχηματισμού: Haar wavelet

Wavelet χωρικού μετασχηματισμού: 9/7 biorthogonal wavelets

Μέγεθος μπλοκ: 16 x 16

Περιοχή αναζήτησης για εκτίμηση κίνησης:  $\pm 8$  εικονοστοιχεία

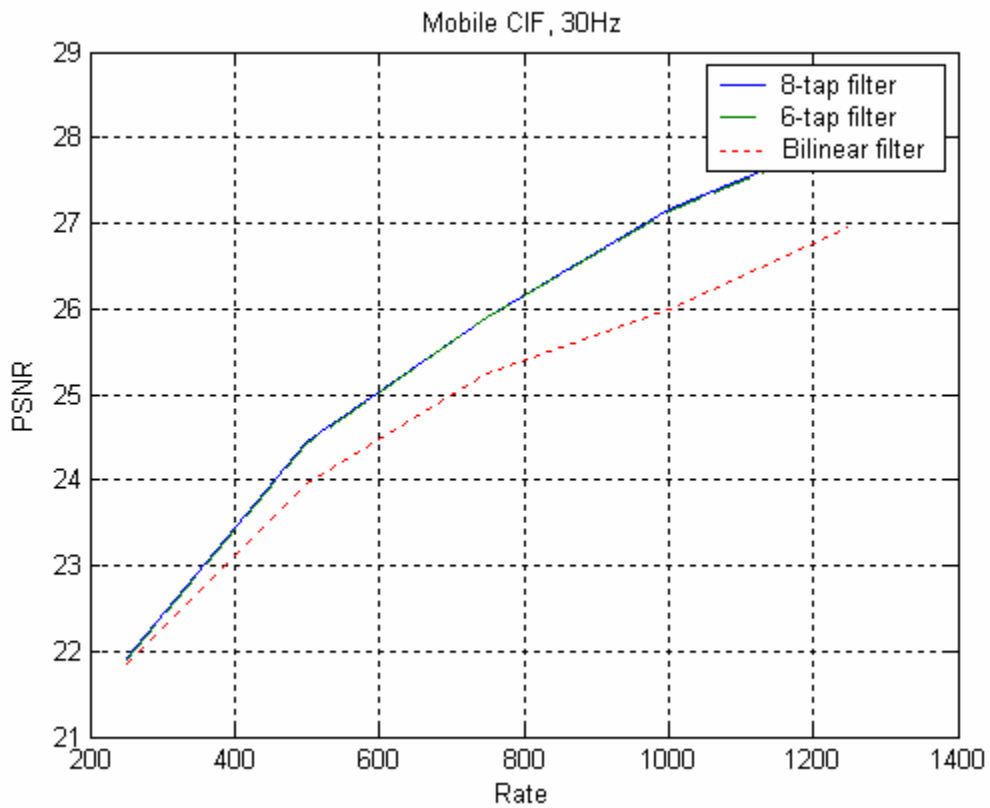


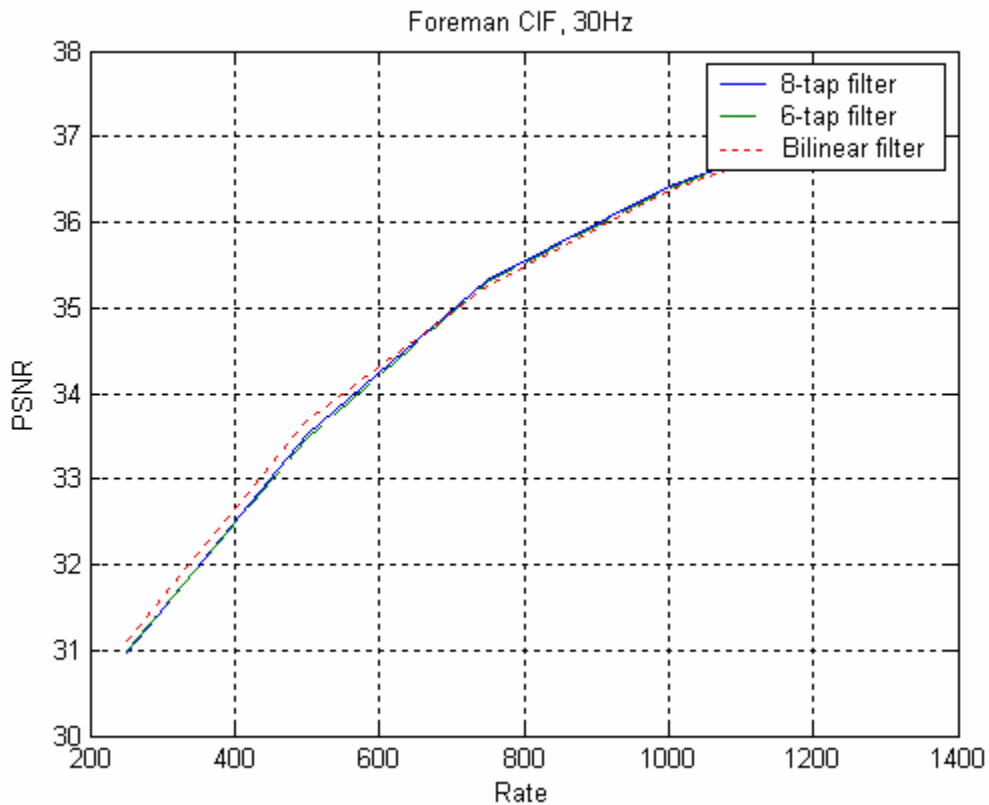
**Mobile**

Rate \ PSNR	250	500	750	1000	1250
8-tap	21.92	24.46	25.90	27.17	28.04
6-tap	21.89	24.42	25.89	27.13	28.01
Bilinear	21.85	23.97	25.24	26.00	26.96

**Foreman**

Rate \ PSNR	250	500	750	1000	1250
8-tap	30.98	33.52	35.31	36.43	37.17
6-tap	30.99	33.48	35.30	36.40	37.17
Bilinear	31.12	33.68	35.25	36.37	37.10





Εικόνα 4.4 Σύγκριση φίλτρων παρεμβολής



(α)

Mobile Frame 1, 8-tap Interpolation



(β)

Mobile Frame 1, 6-tap Interpolation



(γ)

**Εικόνα 4.5** Αποκωδικοποιημένο πλαίσιο της ακολουθίας βίντεο Mobile (α) Bilinear φίλτρο (β) 8-tap φίλτρο (γ) 6-tap φίλτρο.

## Συμπεράσματα

Παρατηρούμε ότι η επιλογή του φίλτρου παρεμβολής έχει επίδραση στη μετρική ποιότητας PSNR, Εικόνα 4.4. Στην ακολουθία βίντεο Mobile, η επιλογή του 8-tap φίλτρου παρεμβολής επιφέρει βελτίωση στο PSNR μέχρι και 1.17 dB (ρυθμός 1000 Kbps) σε σχέση με το Bilinear φίλτρο. Σε σύγκριση με το 6-tap φίλτρο παρεμβολής, το 8-tap παρουσιάζει ίδια αποτελέσματα με μηδαμινές διαφορές. Όμως στην ακολουθία βίντεο Foreman, η χρήση του 8-tap φίλτρου αντί για το κλασικό Bilinear φίλτρο δεν παρουσιάζει βελτίωση όπως συνέβη στην ακολουθία Mobile. Αντίθετα, για μικρούς ρυθμούς δεδομένων εμφανίζει μικρότερες τιμές PSNR. Για παράδειγμα για ρυθμό 500 Kbps έχει μικρότερη τιμή PSNR κατά 0.16 dB. Σε σχέση με το 6-tap φίλτρο οι διαφορές είναι και πάλι μηδαμινές. Όσον αφορά την οπτική ποιότητα βλέπουμε στην Εικόνα 4.5, το πλαίσιο 1 για ρυθμό δυαδικών ψηφίων 1000 Kbps και για τα τρία διαφορετικά φίλτρα παρεμβολής. Επίσης, θα πρέπει να αναφέρουμε ότι η πολυπλοκότητα για την πράξη της παρεμβολής είναι  $O(L)$ , αφού για κάθε δείγμα που δημιουργείται χρειάζονται  $L$  πολλαπλασιασμοί και  $L - 1$  προσθέσεις, όπου  $L$  το μήκος του φίλτρου παρεμβολής. Συνολικά, αν υποθέσουμε ότι έχουμε ακρίβεια μισού εικονοστοιχείου δημιουργούνται  $3(N_1+N_2)$  νέα δείγματα, όπου  $N_1, N_2$  οι διαστάσεις της εικόνας.

Από τις παραπάνω παρατηρήσεις καταλήγουμε στα εξής συμπεράσματα:

- Το 6-tap φίλτρο παρεμβολής εμφανίζει όμοιες επιδόσεις με το 8-tap φίλτρο παρεμβολής. Οπότε, για να έχουμε μικρότερη πολυπλοκότητα υπολογισμού συμφέρει πάντα να χρησιμοποιούμε το 6-tap φίλτρο αντί για το 8-tap.
- Το κλασικό Bilinear φίλτρο έχει την μικρότερη πολυπλοκότητα υπολογισμού, αφού έχει τον μικρότερο αριθμό συντελεστών. Είδαμε όμως ότι σε κάποιες περιπτώσεις, όπως στην ακολουθία βίντεο Mobile, η χρήση του 8-tap φίλτρου παρεμβολής επιφέρει σημαντική βελτίωση στο PSNR. Θα πρέπει λοιπόν να βρούμε έναν προσαρμοστικό τρόπο με τον οποίο να επιλέγουμε κάθε φορά, ανάλογα με την ακολουθία βίντεο, το κατάλληλο φίλτρο παρεμβολής. Έτσι, θα επιτυγχάνουμε τη μικρότερη δυνατή πολυπλοκότητα με την χρήση του Bilinear φίλτρου και στις περιπτώσεις που το 8-tap επιφέρει βελτίωση στο PSNR θα χρησιμοποιούμε αυτό.

Ο προσαρμοστικός τρόπος με τον οποίο θα επιλέγουμε το κατάλληλο φίλτρο παρεμβολής είναι η χρήση της μετρικής JND (Just Noticeable Difference) [9]. Η μετρική αυτή χρησιμοποιείται συχνά στην υδατογράφιση εικόνας και ορίζεται ως:

$$JND_x(i, j) = 1 - \frac{1}{1 + \theta \sigma_x^2(i, j)},$$

όπου το  $\sigma_x^2(i, j)$  συμβολίζει την τοπική διασπορά της εικόνας  $x$  στο παράθυρο που έχει κέντρο τις συντεταγμένες  $(i, j)$  και  $\theta$  είναι μία παράμετρος που εξαρτάται από την συγκεκριμένη εικόνα. Η παράμετρος  $\theta$  ορίζεται ως  $\theta = \frac{D}{\sigma_{x \max}^2}$ , όπου  $\sigma_{x \max}^2$  είναι η μέγιστη τοπική διασπορά για την συγκεκριμένη εικόνα και το  $D$  είναι μια παράμετρος πειραματικά καθορισμένη που παίρνει την τιμή 50 ή 100.

Τα φίλτρα παρεμβολής έχουν διαφορετικό συχνοτικό περιεχόμενο, αφού έχουν διαφορετικό αριθμό συντελεστών. Έτσι, η δημιουργία του παρεμβαλλόμενου εικονοστοιχείου διαφέρει. Συγκεκριμένα τα φίλτρα παρεμβολής 8-tap και 6-tap είναι πιο υψηλοδιαβατά απ' ό τι το Bilinear φίλτρο, δηλαδή περνάνε και πιο υψηλές συχνότητες. Έτσι, σε εικόνες όπου εμφανίζουν ομοιόμορφες περιοχές χωρίς έντονες διακυμάνσεις το κλασσικό Bilinear φίλτρο είναι το ιδανικότερο. Σε εικόνες όμως με έντονες διακυμάνσεις και υψηλές συχνότητες, είναι προτιμότερο κατά την παρεμβολή να χρησιμοποιούμε φίλτρα με μεγαλύτερο αριθμό συντελεστών, όπως το 8-tap, 6-tap που χρησιμοποιήσαμε στις πειραματικές μας μετρήσεις. Η μετρική JND εξαρτάται από τη διασπορά με τρόπο ανάλογο. Αν η διασπορά είναι μικρή τότε και η μετρική JND θα έχει μικρή τιμή, ενώ για μεγάλες διασπορές η τιμή JND πλησιάζει τη μονάδα. Βλέπουμε λοιπόν ότι η μετρική JND είναι κατάλληλη για να αποτυπώσει το περιεχόμενο της κάθε εικόνας, άρα και το φίλτρο με το οποίο θα γίνει η παρεμβολή. Στις πειραματικές μετρήσεις που έγιναν χρησιμοποιήθηκε παράθυρο 3 x 3 και ως τιμή κατωφλίου η τιμή 0.5. Για εικόνες που εμφανίζουν μέση τιμή JND πάνω από 0.5 χρησιμοποιούμε το 6-tap φίλτρο ενώ για μέση τιμή JND κάτω από το κατώφλι το Bilinear φίλτρο. Η ορθή λειτουργία της μετρικής JND σε σχέση με την επιλογή του κατάλληλου φίλτρου παρεμβολής επιβεβαιώθηκε πειραματικά, όχι μόνο με τις ακολουθίες βίντεο Mobile, Foreman αλλά και με άλλες γνωστές ακολουθίες βίντεο όπως οι Football, Stefan, Silent, Coastguard.

### 4.3 Επίδραση του Αριθμού Χρονικών Επιπέδων

Στην παράγραφο αυτή θα επιχειρήσουμε να μελετήσουμε την επίδραση του αριθμού χρονικών επιπέδων μετασχηματισμού στην απόδοση του συστήματος κωδικοποίησης βίντεο. Υπενθυμίζουμε ότι κάθε φορά που εφαρμόζουμε χρονικό μετασχηματισμό ο αριθμός των πλαισίων για το επόμενο επίπεδο μειώνεται στο μισό. Θεωρητικά, για να έχουμε όσο το δυνατό καλύτερη συμπίεση και αναπαράσταση από την οικογένεια

συναρτήσεων wavelet, θα πρέπει να εφαρμόσουμε χρονικό μετασχηματισμό τόσων επιπέδων μέχρι ο αριθμός των πλαισίων που προκύπτει να είναι μικρότερος από το πλήθος των συντελεστών του χρονικού μετασχηματισμού. Για παράδειγμα, αν έχουμε  $GOP = 16$  και εφαρμόζουμε Haar χρονικό μετασχηματισμός wavelet τότε θα πρέπει να εφαρμόσουμε χρονικό μετασχηματισμό τεσσάρων επιπέδων, έτσι ώστε να μείνει μόνο ένα πλαίσιο οπότε δεν είναι εφικτός περαιτέρω χρονικός μετασχηματισμός. Στην πράξη όμως δεν συμφέρει πάντα να εφαρμόζουμε χρονικό μετασχηματισμό. Αυτό εξαρτάται από τον αν η εκτίμηση της κίνησης μεταξύ δύο πλαισίων είναι σωστά εκτιμώμενη [3]. Αν δηλαδή σε δύο πλαίσια δεν υπάρχει σωστή εκτίμηση κίνησης, τότε δεν συμφέρει να εκτελέσουμε χρονικό μετασχηματισμό γιατί όχι μόνο τα πλαίσια σφάλματος  $H_t$  θα έχουν μεγάλες τιμές αλλά και τα πλαίσια ανανέωσης  $L_t$  θα εμφανίζουν θάμπωμα (λόγω της πληροφορίας του πλαισίου  $H_t$  που θα προστεθεί στο αρχικό πλαίσιο). Στους παρακάτω πίνακες παρουσιάζονται τα αποτελέσματα από εφαρμογή τριών επιπέδων χρονικού μετασχηματισμού και συγκρίνονται με αυτά της παραγράφου 4.1 που λάβαμε για τέσσερα επίπεδα χρονικού μετασχηματισμού. Εκτός από τις ακολουθίες βίντεο Foreman, Mobile παρουσιάζονται αποτελέσματα και για την ακολουθία βίντεο Stefan. Επιπλέον, παρουσιάζονται και οι εικόνες για τις ακολουθίες Stefan, Mobile στα 500 Kbps.

### Σενάριο

Αριθμός πλαισίων: 16

Ανάλυση: 352 x 288 (CIF resolution, 30 Hz)

Αριθμός χρονικών επιπέδων μετασχηματισμού: 3

Αριθμός χωρικών επιπέδων μετασχηματισμού: 3

Wavelet χρονικού μετασχηματισμού: Haar wavelet

Wavelet χωρικού μετασχηματισμού: 9/7 biorthogonal wavelets

Μέγεθος μπλοκ: 16 x 16

Περιοχή αναζήτησης για εκτίμηση κίνησης:  $\pm 8$  εικονοστοιχεία

Φίλτρο παρεμβολής: το 8-tap της παραγράφου 4.2.

### *Mobile*

Rate \ PSNR	250	500	750	1000	1250
temp_level = 3	21.22	23.72	25.45	26.72	27.80
temp_level = 4	21.92	24.46	25.90	27.17	28.04

**Foreman**

PSNR \ Rate	250	500	750	1000	1250
temp_level = 3	30.82	33.56	35.40	36.52	37.39
temp_level = 4	30.98	33.52	35.31	36.43	37.17

**Stefan**

PSNR \ Rate	250	500	750	1000	1250
temp_level = 3	21.46	23.27	24.61	25.80	27.01
temp_level = 4	21.32	23.04	24.37	25.52	26.76

Stefan Frame 1, temp-levels = 3



(a)

Stefan Frame 1, temp-levels = 4



(β)

**Εικόνα 4.6** Αποκωδικοποιημένο πλαίσιο της ακολουθίας Stefan χρησιμοποιώντας (α) 3 χρονικά επίπεδα (β) 4 χρονικά επίπεδα.

Mobile Frame 1, temp-levels = 3



(α)



Mobile Frame 1, temp-levels = 4



(β)

**Εικόνα 4.7** Αποκωδικοποιημένο πλαίσιο της ακολουθίας Mobile χρησιμοποιώντας (α) 3 χρονικά επίπεδα (β) 4 χρονικά επίπεδα.

### Συμπεράσματα

Από τους παραπάνω πίνακες βλέπουμε ότι για την ακολουθία βίντεο Mobile ο περαιτέρω χρονικός μετασχηματισμός βελτιώνει την μετρική PSNR κατά 0.24 – 0.7 dB, για την ακολουθία Foreman έχουμε βελτίωση 0.16 dB στα 250 Kbps και μείωση σε όλους τους άλλους ρυθμούς από 0.04 έως 0.22 dB, και τέλος στην ακολουθία Stefan έχουμε μείωση για όλους τους ρυθμούς δυαδικών ψηφίων από 0.14 έως 0.28 dB. Παρατηρούμε δηλαδή ότι η επίδραση του αριθμού χρονικών επιπέδων διαφέρει με το είδος της ακολουθίας βίντεο. Αυτό είναι φανερό και από την Εικόνα 4.6, όπου βλέπουμε ότι στην ακολουθία Stefan η εικόνα για αριθμό χρονικών επιπέδων τέσσερα είναι χειρότερη και παρουσιάζει περισσότερο θάμπωμα. Αντίθετα στην ακολουθία Mobile η εικόνα που προέκυψε για αριθμό χρονικών επιπέδων τέσσερα είναι οπτικά καλύτερη, Εικόνα 4.7. Από άποψη πολυπλοκότητας για GOP = 16 και 4 επίπεδα χρονικού μετασχηματισμού χρειάζεται να εκτελέσουμε τη διαδικασία χρονικού φιλτραρίσματος με αντιστάθμιση κίνησης (MCTF) για 15 ζεύγη πλαισίων ενώ αν επιλέξουμε 3 επίπεδα χρονικού μετασχηματισμού για 14 ζεύγη πλαισίων. Έτσι μας συμφέρει και από άποψη πολυπλοκότητας να χρησιμοποιούμε λιγότερα επίπεδα χρονικού μετασχηματισμού.

Στηριζόμενοι στα παραπάνω συμπεράσματα προτείνουμε έναν προσαρμοστικό τρόπο εκτίμησης του αριθμού των χρονικών επιπέδων, ώστε να επιτυγχάνουμε καλύτερη απόδοση όταν είναι εφικτό, αλλιώς να εκτελούμε μετασχηματισμό μικρότερου αριθμού χρονικών επιπέδων και να έχουμε κέρδος και από πλευράς πολυπλοκότητας. Η διαδικασία αυτή εκτελείται ως εξής: μόλις τελειώσει η διαδικασία χρονικού φιλτραρίσματος με αντιστάθμιση κίνησης ενός επιπέδου ελέγχουμε τις τιμές των πλαισίων  $H_i$ . Συγκεκριμένα σε κάθε μπλοκ υπολογίζουμε τη απόλυτη μέση τιμή του σφάλματος και αν το ποσοστό των μπλοκ που έχουν μεγαλύτερη μέση τιμή πάνω από ένα κατώφλι  $C_1$  είναι μεγαλύτερο από μία συγκεκριμένη τιμή  $C_2$ , δεν προχωράμε σε εφαρμογή του χρονικού μετασχηματισμού σε επόμενο επίπεδο. Στις μετρήσεις οι τιμές  $C_1 = 10$ ,  $C_2 = 10$  λειτούργησαν ορθά για τις ακολουθίες βίντεο που εφαρμόστηκαν. Αν δηλαδή περισσότερο από 10% των μπλοκ έχουν μέση τιμή σφάλματος μεγαλύτερη από 10, επιλέγουμε να σταματήσουμε αλλιώς εκτελούμε το χρονικό μετασχηματισμό στο επόμενο επίπεδο. Ο λόγος που επιλέχτηκε η μελέτη των πλαισίων  $H_i$  για τον καθορισμό του αριθμού χρονικών επιπέδων είναι ότι στα πλαίσια  $H_i$  αντικατοπτρίζεται η ποιότητα της εκτίμησης κίνησης. Οπότε αν δεν έχουμε καλή εκτίμηση κίνησης είναι αναμενόμενο στο επόμενο χρονικό επίπεδο η εκτίμηση να είναι ακόμα χειρότερη.

Από άποψη πολυπλοκότητας η παραπάνω διαδικασία ελέγχου δεν επιβαρύνει το σύστημα. Χρειάζεται μόνο ο υπολογισμός της μέσης τιμής σε κάθε μπλοκ του πλαισίου  $H_i$  και ο υπολογισμός του ποσοστού των τιμών που υπερβαίνουν την τιμή κατωφλίου  $C_1$ . Το κέρδος όμως που έχουμε από μη εφαρμογή του μετασχηματισμού σε επόμενο χρονικό επίπεδο είναι πολύ μεγάλο, αφού η διαδικασία MCTF κοστίζει υπολογιστικά.

#### **4.4 Μελέτη της Επίδρασης της Παραμέτρου GOP**

Στην παράγραφο αυτή θα μελετήσουμε τον τρόπο με τον οποίο η παράμετρος GOP (Group Of Pictures) επηρεάζει την απόδοση του συστήματος. Η τιμή της GOP εκφράζει τον αριθμό πλαισίων τα οποία όλα μαζί αποτελούν μια ομάδα και στα οποία θα εφαρμοστεί η κωδικοποίηση βίντεο (συντά στη βιβλιογραφία όρος συναντάται και ως GOF (Group Of Frames)). Έτσι ο αριθμός GOP εκφράζει το πόσα πλαίσια επεξεργάζονται μαζί. Στο σενάριο των προηγούμενων παραγραφών χρησιμοποιήθηκε  $GOP = 16$  και εφαρμόσαμε μετασχηματισμό τεσσάρων (ή τριών) χρονικών επιπέδων. Το ερώτημα που θα προσπαθήσουμε να απαντήσουμε είναι το εξής: συμφέρει στην ακολουθία βίντεο που έχουμε να εφαρμόσουμε χρονικό μετασχηματισμό τεσσάρων επιπέδων για  $GOP = 16$  ή είναι προτιμότερο να έχουμε  $GOP = 32$  και να εφαρμόσουμε χρονικό μετασχηματισμό πέντε επιπέδων; Από άποψη πολυπλοκότητας πλεονεκτεί η πρώτη περίπτωση. Αν υποθέσουμε ότι η ακολουθία βίντεο αποτελείται από 32 πλαίσια

(για χάρη ευκολίας) στην πρώτη περίπτωση εφαρμόζουμε τη διαδικασία MCTF σε  $2*(8+4+2+1) = 30$  ζεύγη πλαισίων, ενώ στην δεύτερη περίπτωση η διαδικασία MCTF εφαρμόζεται σε  $(16+8+4+2+1) = 31$  ζεύγη πλαισίων. Έτσι για κάθε 32 πλαίσια η δεύτερη περίπτωση εκτελεί τη διαδικασία MCTF μία φορά παραπάνω. Από άποψη απόδοσης όμως πλεονεκτεί η δεύτερη περίπτωση όπως φαίνεται στα αποτελέσματα που παρουσιάζονται στους παρακάτω πίνακες. Τα αποτελέσματα για GOP = 16 είναι όμοια με αυτά των προηγούμενων παραγραφών.

### Σενάριο

Αριθμός πλαισίων: 32

Ανάλυση: 352 x 288 (CIF resolution, 30 Hz)

Αριθμός χρονικών επιπέδων μετασχηματισμού: 5

Αριθμός χωρικών επιπέδων μετασχηματισμού: 3

Wavelet χρονικού μετασχηματισμού: Haar wavelet

Wavelet χωρικού μετασχηματισμού: 9/7 biorthogonal wavelets

Μέγεθος μπλοκ: 16 x 16

Περιοχή αναζήτησης για εκτίμηση κίνησης:  $\pm 8$  εικονοστοιχεία

Φίλτρο παρεμβολής: το 8-tap της παραγράφου 4.2.

### Σημείωση

Στους πίνακες που ακολουθούν στην περίπτωση του GOP = 32, η τιμή του PSNR που αναγράφεται είναι η μέση τιμή για τα πρώτα 16 πλαίσια της ακολουθίας, ώστε να υπάρχει αντιστοιχία με την περίπτωση που GOP = 16 και η σύγκριση να είναι ορθή. Επιπλέον, για τον ίδιο λόγο στην περίπτωση του GOP = 32 η τιμή του PSNR που παρουσιάζεται για Rate = 250, στην πραγματικότητα είναι η μέτρηση που προέκυψε για Rate = 500. Αντίστοιχα και για τον άλλο ρυθμό δυαδικών ψηφίων.

### **Mobile**

Rate	250	500
PSNR		
GOP = 16	21.92	24.46
GOP = 32	22.83	25.17

### **Foreman**

Rate	250	500
PSNR		
GOP = 16	30.98	33.52

Το πλαίσιο 1 των ακολουθιών που λάβαμε για τις δύο τιμές GOP παρουσιάζεται στην Εικόνα 4.8 και αντιστοιχεί σε ρυθμό δυαδικών ψηφίων 250 Kbps.

Foreman Frame 1, GOP = 16



(α)

Foreman Frame 1, GOP = 32



(β)

**Εικόνα 4.8** Αποκωδικοποιημένο πλαίσιο της ακολουθίας Foreman χρησιμοποιώντας (α) GOP = 16 (β) GOP = 32.

Mobile Frame 1, GOP = 16



(α)

Mobile Frame 1, GOP = 32



(β)

**Εικόνα 4.9** Αποκωδικοποιημένο πλαίσιο της ακολουθίας Mobile χρησιμοποιώντας (α) GOP = 16  
(β) GOP = 32.

## Συμπεράσματα

Παρατηρούμε ότι στην περίπτωση που έχουμε  $GOP = 32$  έχουμε αύξηση της μετρικής PSNR. Στην περίπτωση της ακολουθίας Mobile έχουμε αύξηση 0.91 dB (Rate = 250 Kbps) και 0.71 dB (Rate = 500 Kbps), ενώ για την ακολουθία βίντεο Foreman αύξηση 0.60 dB και 0.56 dB αντίστοιχα. Η βελτίωση της ποιότητας φαίνεται και οπτικά στις παραπάνω εικόνες, όπου για  $GOP = 32$  έχουμε λιγότερο θάμπωμα. Καταλήγουμε στο συμπέρασμα ότι η χρησιμοποίηση  $GOP = 32$  και πέντε χρονικών επιπέδων μετασχηματισμού πλεονεκτεί από την χρησιμοποίηση  $GOP = 16$  και τεσσάρων χρονικών επιπέδων. Όμως, όπως αναφέραμε στην αρχή της παραγράφου το τμήμα που πληρώνουμε είναι η αύξηση του υπολογιστικού κόστους. Σε κάθε περίπτωση, ανάλογα με τις ανάγκες του συστήματος και τον παράγοντα που θεωρούμε κάθε φορά σημαντικότερο, μπορούμε να επιλέξουμε την κατάλληλη περίπτωση.

## **4.5 Ανακεφαλαίωση**

Το κεφάλαιο αυτό αποτελείται από τις πειραματικές μετρήσεις και τη μελέτη της επίδρασης ορισμένων παραμέτρων στην απόδοση του συστήματος κωδικοποίησης βίντεο. Η συνεισφορά της διπλωματικής εργασίας στο χώρο της κλιμακοθετήσιμης κωδικοποίησης βίντεο είναι η εξής:

1. Προτείνουμε έναν προσαρμοστικό τρόπο επιλογής του φίλτρου παρεμβολής. Οι πειραματικές μετρήσεις έδειξαν ότι το φίλτρο παρεμβολής επηρεάζει την απόδοση του συστήματος αλλά και το υπολογιστικό κόστος της διαδικασίας παρεμβολής. Σύμφωνα με τον προσαρμοστικό τρόπο επιλογής, σε ακολουθίες βίντεο όπου η χρησιμοποίηση του 6-tap φίλτρου θα επιφέρει βελτίωση επιλέγουμε αυτό το φίλτρο παρεμβολής, αλλιώς χρησιμοποιούμε το κλασικό διγραμμικό (Bilinear) φίλτρο.
2. Προτείνουμε έναν προσαρμοστικό τρόπο επιλογής του αριθμού των επιπέδων του χρονικού μετασχηματισμού. Έτσι, όταν περαιτέρω χρονικός μετασχηματισμός δεν πρόκειται να επιφέρει βελτίωση στην απόδοση του συστήματος, σταματάμε σε εκείνο το επίπεδο τον χρονικό μετασχηματισμό wavelet.
3. Με τις κατάλληλες πειραματικές μετρήσεις δείχνουμε ότι ο παράγοντας  $GOP$  επηρεάζει την απόδοση του συστήματος. Συγκεκριμένα, είναι προτιμότερο να εφαρμόσουμε  $GOP = 32$  και 5 επίπεδα χρονικού μετασχηματισμού από την επιλογή  $GOP = 16$  και 4 χρονικά επίπεδα. Το τμήμα στην περίπτωση αυτή είναι αύξηση στο υπολογιστικό κόστος.

## **Μελλοντική εργασία – Επεκτάσεις**

Ως μελλοντική εργασία, θα μπορούσαμε να μελετήσουμε την επίδραση ενός διαφορετικού χρονικού μετασχηματισμού wavelet, από τον κλασσικό μετασχηματισμό Haar και να προτείνουμε έναν προσαρμοστικό τρόπο επιλογής του κατάλληλου χρονικού μετασχηματισμού, ώστε να έχουμε καλύτερη ποιότητα με όσο το δυνατόν μικρότερη πολυπλοκότητα. Επιπλέον, ενδιαφέρον παρουσιάζει και η υλοποίηση της άλλης κατηγορίας κλιμακοθετήσιμης κωδικοποίησης βίντεο, δηλαδή της in-band αρχιτεκτονικής και η σύγκριση της ως προς την απόδοση και την πολυπλοκότητα με την MCTF αρχιτεκτονική. Τέλος, αξίζει να μελετήσουμε και άλλους μετασχηματισμούς που σχετίζονται με τα wavelets, όπως για παράδειγμα τα bandelets. Στα bandelets εφαρμόζουμε κανονικά το δισδιάστατο μετασχηματισμό wavelet. Στους συντελεστές όμως που προκύπτουν από το μετασχηματισμό, εφαρμόζουμε και έναν επιπλέον γεωμετρικό ορθογώνιο μετασχηματισμό για να εκμεταλλευτούμε την πιθανή πλεονάζουσα γεωμετρική πληροφορία. Τα bandelets εμφανίζουν καλύτερες επιδόσεις, υστερούν όμως σε πολυπλοκότητα και χρόνο υλοποίησης του μετασχηματισμού.

## Βιβλιογραφία

- [1] Andreopoulos Ioannis, “Wavelet-based Scalable Video Coding”, *PhD Thesis Brussels 2005*.
- [2] Martin Vetterli, Jelena Kovacevic, “Wavelets and Subband Coding”, *Prentice Hall*.
- [3] John W.Woods, Peisong Chen, “Bidirectional MC-EZBC With Lifting Implementation”, *IEEE Transactions on Circuit and Systems for Video Technology, Vol. 14, NO.10, October 2004*.
- [4] Andrew Secker, David Taubman, “Lifting-Based Invertible Motion Adaptive Transform (LIMAT) Framework for Highly Scalable Video Compression”, *IEEE Transactions on Image Processing, Vol. 12, NO.12, December 2003*.
- [5] Jens-Rainer Ohm, “Advances in Scalable Video Coding”, *Proceedings of the IEEE, Vol. 93, NO.1, January 2005*.
- [6] John W.Woods, Seung-Jong Choi, “Motion-Compensated 3-D Subband Coding of Video”, *IEEE Transactions on Image Processing, Vol. 8, NO.2, February 1999*.
- [7] Ingrid Daubechies, Wim Sweldens, “Factoring Wavelet Transforms Into Lifting Steps”, *Journal of Fourier Analysis and Applications, vol. 4, pp. 247-269, 1998*.
- [8] D.S.Turaga, Michaela van der Schaar, B. Pesquet-Popescu, “Complexity Scalable Motion Compensated Wavelet Video Encoding”, *IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, NO. 8, August 2005*.
- [9] L.Song, J. Xu, H. Xiong, F. Wu, “Content Adaptive Update Steps for Lifting-Based Motion Compensated Temporal Filtering”.
- [10] C. Tillier, B. Pesquet-Popescu, “Improved Update Operators for Lifting-Based Motion-Compensated Temporal Filtering”, *IEEE Signal Processing Letters, Vol. 12, No. 2, February 2005*.
- [11] B. Girod, S. Han, “Optimum Update for Motion-Compensated Lifting”, *IEEE Signal Processing Letters, Vol. 12, No. 2, February 2005*.
- [12] D. Taubman, N. Mehrseresht, “Adaptively Weighted Update Steps In Motion Compensated Lifting Based Scalable Video Compression”.
- [13] R. Xiong, F. Wu, S. Li, Z. Xiong, Y. Zhang, “Exploiting Temporal Correlation With Adaptive Block-Size Motion Alignment for 3D Wavelet Coding”.



- [14] A. Said, W. Pearlman, “A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 6, NO. 3, June 1996.
- [15] B. Kim, W. Pearlman, “An Embedded Wavelet Video Coder Using Three-Dimensional Set Partitioning in Hierarchical Trees (SPIHT)”.
- [16] K. Taavo, H. Andren, “Scalable Video Using Wavelets”, *Master Thesis, February 2002*.
- [17] P. Thevenaz, T. Blu, M. Unser, “Image Interpolation and Resampling”
- [18] D.S.Turaga, M.v.d. Schaar, Y. Andreopoulos, A. Munteanu, P. Schelkens, “Unconstrained Motion Compensated Temporal Filtering (UMCTF) for Efficient and Flexible Interframe Wavelet Video Coding”
- [19] C. Valens, “The Fast Lifting Wavelet Transform”

-----