

Wyner-Ziv Κωδικοποίηση Στερεοσκοπ

Ελίνα Σουκαλλάρη

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙ



Ιωάννινα, Ιούνιος 2012



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF IOANNINA



ΒΙΒΛΙΟΘΗΚΗ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΙΩΑΝΝΙΝΩΝ



826000345866



ΚΩΔΙΚΟΠΟΙΗΣΗ WYNER-ZIV ΣΤΕΡΕΟΣΚΟΠΙΚΟΥ ΒΙΝΤΕΟ

**Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ**

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύστασης
του Τμήματος Πληροφορικής
Εξεταστική Επιτροπή

Από την

Ελίνα Σουκαλλάρη

ως μέρος των Υποχρεώσεων

για τη λήψη

του

**ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ**

Ιούνιος 2012



ΑΦΙΕΡΩΣΗ

Στην οικογένειά μου και στους φίλους μου!



ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Λυσίμαχο Παύλο Κόντη για την πολύτιμη βοήθειά του κατά τη διάρκεια της εκπόνησης της εργασίας αυτής. Επίσης, ευχαριστώ όλους όσους με βοήθησαν στην περάτωσή της.



ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	vi
ΠΕΡΙΛΗΨΗ	x
EXTENDED ABSTRACT IN ENGLISH	xi
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	1
1.1. Εισαγωγή	1
1.2. Διάρθρωση Εργασίας	2
ΚΕΦΑΛΑΙΟ 2. ΚΩΔΙΚΟΠΟΙΗΣΗ ΕΙΚΟΝΑΣ ΚΑΙ ΒΙΝΤΕΟ	3
2.1. Εισαγωγή	3
2.2. Κωδικοποίηση Εικόνας	3
2.3. Κωδικοποίηση Βίντεο	5
2.3.1. Κωδικοποίηση εικόνων χωρίς εκτίμηση και αντιστάθμιση κίνησης (no motion compensation)	7
2.3.2. Κωδικοποίηση εικόνων με εκτίμηση και αντιστάθμιση κίνησης (motion estimation and compensation)	9
ΚΕΦΑΛΑΙΟ 3. ΚΑΤΑΝΕΜΗΜΕΝΗ ΣΥΜΠΙΕΣΗ ΒΙΝΤΕΟ	15
3.1. Εισαγωγή	15
3.2. Κατανεμημένη Συμπύεση Βίντεο (Distributed Video Coding)	15
3.2.1. Θεώρημα Slepian-Wolf για μη απωλεστική κατανεμημένη κωδικοποίηση	16
3.2.2. Πρακτική κωδικοποίηση Slepian-Wolf	18
3.2.3. Θεωρία ρυθμού-παραμόρφωσης (Rate-Distortion) για απωλεστική συμπύεση με side information στον αποκωδικοποιητή	20
3.2.4. Κωδικοποίηση Wyner-Ziv	21
3.2.5. Κωδικοποίηση βίντεο με χαμηλή πολυπλοκότητα	22
3.2.6. Κωδικοποίηση Wyner-Ziv σε επίπεδο εικονοστοιχείων (pixel domain)	24
3.2.7. Κωδικοποίηση Wyner-Ziv σε επίπεδο μετασχηματισμού (transform-domain)	25
ΚΕΦΑΛΑΙΟ 4. ΣΤΕΡΕΟΣΚΟΠΙΑ	26
4.1. Εισαγωγή	26
4.2. Κατανεμημένη κωδικοποίηση Wyner-Ziv στερεοσκοπικού βίντεο (distributed Wyner-Ziv coding of stereoscopic video)	31
4.2.1. Κατανεμημένη κωδικοποίηση στερεοσκοπικού βίντεο	31
4.2.2. Κωδικοποίηση Wyner-Ziv στερεοσκοπικού βίντεο	32
4.2.3. Joint bitplane LDPC decoding	39
4.3. Αλγόριθμος Μεγιστοποίησης Πιθανότητας	42
4.3.1. Μοντέλο	42
4.3.2. Πρόβλημα	43



4.3.3. Βήμα Ε του αλγορίθμου ΕΜ	44
4.3.4. Βήμα Μ του αλγορίθμου ΕΜ	47
4.3.5. Τερματισμός	49
4.3.6. Τεχνική μάθησης διανυσμάτων κίνησης στον αποκωδικοποιητή σε επίπεδο μετασχηματισμού (transform domain)	49
ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ	51
5.1. Πειραματικά αποτελέσματα	51
5.2. Επίλογος	64
5.3. Μελλοντικές επεκτάσεις	65
ΑΝΑΦΟΡΕΣ	66
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ	73



ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα	Σελ
Σχήμα 2.1: Κωδικοποιητής πηγής και καναλιού.	4
Σχήμα 2.2: Γενικό σύστημα συμπίεσης/κωδικοποίησης εικόνας.	5
Σχήμα 2.3: Κωδικοποίηση βίντεο με πρόβλεψη (prediction).	6
Σχήμα 2.4: (a) τρέχουσα εικόνα (current frame), (b) προηγούμενη εικόνα (previous frame), (c) η διαφορά χωρίς αντιστάθμιση κίνησης (no motion compensation)	7
Σχήμα 2.5: Κωδικοποιητής με βρόχο αποκωδικοποίησης.	8
Σχήμα 2.6: Εκτίμηση κίνησης (motion estimation) με ταίριασμα των μπλοκ της τρέχουσας εικόνας με αυτών της εικόνας αναφοράς.	10
Σχήμα 2.7: Αντιστάθμιση κίνησης (motion compensation) για την εύρεση της πρόβλεψης της τρέχουσας εικόνας (predicted frame) από την ανακατασκευασμένη εικόνα (reference frame).	11
Σχήμα 2.8: Η διαφορά που προκύπτει από την αφαίρεση της πρόβλεψης (που προκύπτει από την αντιστάθμιση κίνησης) από την τρέχουσα εικόνα.	11
Σχήμα 2.9: Κωδικοποιητής βίντεο με εκτίμηση και αντιστάθμιση κίνησης (motion estimation and compensation).	12
Σχήμα 2.10: Κωδικοποιητής βίντεο με εκτίμηση και αντιστάθμιση κίνησης (motion estimation and compensation).	13
Σχήμα 2.11: Αποκωδικοποιητής βίντεο με εκτίμηση και αντιστάθμιση κίνησης (motion estimation and compensation).	13
Σχήμα 2.12: (a) Μέθοδος κωδικοποίησης με πολύπλοκο κωδικοποιητή και απλό αποκωδικοποιητή, (b) Μέθοδος κωδικοποίησης με απλό κωδικοποιητή και πολύπλοκο αποκωδικοποιητή.	14
Σχήμα 3.1: Κατανομημένη συμπίεση δύο πηγών πληροφορίας που εξαρτώνται στατιστικά μεταξύ τους, X και Y. Ο αποκωδικοποιητής αποκωδικοποιεί από κοινού (joint decoder) τα X και Y και έτσι αξιοποιεί την εξάρτησή τους.	16



- Σχήμα 3.2: Θεώρημα Slepian Wolf, 1973: Περιοχή του ρυθμού που μπορεί να επιτευχθεί για κατανεμημένη συμπίεση δύο στατιστικά εξαρτημένων πηγών X και Y . 17
- Σχήμα 3.3: Συμπίεση πηγής X χρησιμοποιώντας side information Y . Το Y είναι διαθέσιμο μόνο στον αποκωδικοποιητή και όχι στον κωδικοποιητή. 17
- Σχήμα 3.4: Απωλεστική συμπίεση πηγής X χρησιμοποιώντας side information Y με το οποίο σχετίζεται στατιστικά. 20
- Σχήμα 3.5: Ένας κωδικοποιητής Wyner-Ziv αποτελείται από έναν κβαντιστή (quantizer) ακολουθούμενο από έναν κωδικοποιητή Slepian-Wolf (Slepian-Wolf Encoder). 21
- Σχήμα 3.6: Κωδικοποίηση – αποκωδικοποίηση βίντεο κινητού τηλεφώνου. 23
- Σχήμα 3.7: Συνδυασμός απλού κωδικοποιητή – πολύπλοκου αποκωδικοποιητή και το αντίστροφο (δηλαδή πολύπλοκος κωδικοποιητής – απλός αποκωδικοποιητής). 23
- Σχήμα 3.8: Wyner-Ziv κωδικοποιητής με απλό κωδικοποιητή και πολύπλοκο αποκωδικοποιητή στο χώρο εικονοστοιχείων (pixel domain). 24
- Σχήμα 3.9: κωδικοποίηση Wyner-Ziv στο χώρο μετασχηματισμού (transform-domain), SWC (Slepian-Wolf Coder). 25
- Σχήμα 4.1: Ένα πείραμα για να καταλάβουμε τη στερεοσκοπία: (α) όταν τα μάτια συγκλίνουν στον αντίχειρα, τότε ο φόντος εμφανίζεται σαν διπλή εικόνα, (β) όταν τα μάτια συγκλίνουν στον φόντο, τότε ο αντίχειρας εμφανίζεται σαν διπλή εικόνα. 27
- Σχήμα 4.2: (α) εικόνα X , (b) φάσμα διαφορών (οριζόντιων μετατοπίσεων), (c) – (f) εικόνες Y σύμφωνα με τις οριζόντιες μετατοπίσεις της διαφοράς Δ της X σε σχέση με την Y . 30
- Σχήμα 4.3: Κατανεμημένη συμπίεση: τα X και Y κωδικοποιούνται ανεξάρτητα (separate encoding) και αποκωδικοποιούνται μαζί (joint decoding). 31
- Σχήμα 4.4: Μοντέλο κωδικοποίησης Wyner-Ziv στερεοσκοπικού βίντεο. 32
- Σχήμα 4.5: Wyner-Ziv κωδικοποίηση του X σε σχέση με το Y . 32
- Σχήμα 4.6: Αποκωδικοποίηση Wyner-Ziv για την εύρεση των κβαντισμένων συντελεστών Q (α) χωρίς εκτίμηση διαφοράς (no disparity compensation), (β) με ιδανική διαφορά (disparity oracle), (γ) εκτίμηση της διαφοράς Δ μέσω EM. 34
- Σχήμα 4.7 αριστερά: υπολογισμός της διαφοράς Δ (disparity), δεξιά: side information Y . 35



- Σχήμα 4.8: Κωδικοποίηση του X σε σχέση με το Y με α) μηδενική κίνηση (zero motion), β) ιδανική εκτίμηση κίνησης (motion oracle), γ) εκτίμηση των διανυσμάτων κίνησης (motion estimation). 37
- Σχήμα 4.9: Κωδικοποίηση του X σε σχέση με το Y με α) μηδενική διαφορά (zero disparity), β) ιδανική εκτίμηση διαφοράς (disparity oracle), γ) εκτίμηση των διανυσμάτων διαφοράς (disparity estimation). 38
- Σχήμα 4.10: Γράφημα διάδοσης πεποιθήσεων του αποκωδικοποιητή LDPC (joint bitplane LDPC decoding). 40
- Σχήμα 4.11: Αριστερά: Ε βήμα, υπολογισμός κατανομής πιθανοτήτων της κίνησης, δεξιά: Μ βήμα, υπολογισμός του soft side information ψ . 45
- Σχήμα 4.12: Αριστερά: Ε βήμα, υπολογισμός κατανομής πιθανοτήτων της διαφοράς Δ , δεξιά: Μ βήμα, υπολογισμός του soft side information ψ . 46
- Σχήμα 4.13: Τεχνική μάθησης διανυσμάτων κίνησης στο χώρο μετασχηματισμού όπου το μη απωλεστικό σύστημα κωδικοποίησης να είναι εντός της διακεκομμένης γραμμής. 50
- Σχήμα 5.1: Αριστερή και δεξιά εικόνα που εφαρμόστηκαν στα πειράματα. 52
- Σχήμα 5.2: Ρυθμός-παραμόρφωση για διαφορετικά GOP, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς. 54
- Σχήμα 5.3: Ρυθμός-παραμόρφωση για τέσσερα συστήματα κωδικοποίησης για GOP = 2, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς. 56
- Σχήμα 5.4: Ο ρυθμός για τέσσερα συστήματα κωδικοποίησης για GOP = 70, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς. 58
- Σχήμα 5.5: Η ποιότητα για τέσσερα συστήματα κωδικοποίησης για GOP = 70, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς. 60
- Σχήμα 5.6: Τα σφάλματα των διανυσμάτων κίνησης και διαφοράς για το σύστημα κωδικοποίησης Wyner-Ziv για GOP = 70, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς. 62



Σχήμα 5.7: Οι πιθανότητες των διανυσμάτων κίνησης και διαφοράς για το σύστημα κωδικοποίησης Wyner-Ziv για $GOP = 70$, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς.



ΠΕΡΙΛΗΨΗ

Ελίνα Σουκαλλάρη του Χαραλάμπους και της Ευαγγελίας. MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Ιούνιος, 2012. Κωδικοποίηση Wyner-Ziv στερεοσκοπικού βίντεο, Επιβλέποντας: Λυσίμαχος Παύλος Κόντης.

Η εργασία ασχολείται με την κωδικοποίηση στερεοσκοπικού βίντεο χρησιμοποιώντας το θεώρημα των Wyner-Ziv. Η κατανεμημένη κωδικοποίηση βίντεο (distributed video coding) μετατοπίζει την πολυπλοκότητα υπολογισμού από τον κωδικοποιητή στον αποκωδικοποιητή. Κωδικοποιεί τις εικόνες του βίντεο ανεξάρτητα και τις αποκωδικοποιεί μαζί αξιοποιώντας το χρονικό πλεόνασμα μέσω της εκτίμησης κίνησης. Στο στερεοσκοπικό βίντεο οι κάμερες δεν επικοινωνούν μεταξύ τους και γι' αυτό βολεύει η κατανεμημένη κωδικοποίηση η οποία υποστηρίζει την ανεξαρτησία των καμερών στον κωδικοποιητή. Οι Slepian-Wolf πρότειναν τα όρια μη απωλεστικής συμπίεσης ενώ οι Wyner-Ziv πρότειναν τα όρια της απωλεστικής συμπίεσης ενός βίντεο. Στον αποκωδικοποιητή η εκτίμηση κίνησης υπολογίζει τα διανύσματα κίνησης (motion vectors) μιας εικόνας με βάση την προηγούμενή της ανακατασκευασμένη εικόνα και με βάση τα δυαδικά ψηφία ισοτιμίας (syndrome) που λαμβάνει από τον κωδικοποιητή για τη συγκεκριμένη εικόνα. Για τα στερεοσκοπικά ζεύγη εικόνων υπολογίζονται τα διανύσματα διαφοράς (disparity vectors). Η τεχνική περιγράφεται στα πλαίσια του αλγορίθμου μεγιστοποίησης πιθανότητας – EM (expectation maximization) και είναι πιο αποδοτική από την κλασική μέθοδο συμπίεσης JPEG ως προς το rate-distortion. Επίσης, είναι πιο αποτελεσματική από τη μέθοδο που δεν κάνει εκτίμηση κίνησης (no motion compensation) και προσεγγίζει κατά πολύ την ιδανική εκτίμηση κίνησης (motion oracle). Τέλος η πολυπλοκότητα υπολογισμού είναι πολύ μικρότερη από αυτή του αλγορίθμου που κάνει πλήρη αναζήτηση των διανυσμάτων κίνησης (full motion search).



EXTENDED ABSTRACT IN ENGLISH

Soukallari, Elinta, MSc. Computer Science Department, University of Ioannina, Greece. June, 2012. Wyner-Ziv coding of stereoscopic video. Thesis Supervisor: Lisimachos Paul Kondi.

Stereoscopy refers to the ability that humans have to see the same scene with both eyes in slightly different ways. Acquisition and display of stereoscopic and more generally multiview sequences have many applications. For medical applications, stereoscopic probes are providing more useful information than a monoscopic one. For scientific exploration and visualization, stereoscopic or multiview display enables the human being to experience 3-D environment over a flat 2-D screen. Another more recent application for stereo/3-D display lies in virtual reality and human-computer interface, where 3-D rendering and display gives viewers an illusion of physically interacting with people or objects in a remote (or non-existent) site.

Efficient compression can be achieved by exploiting source statistics, partially or wholly, at the decoder only. This surprising insight is the consequence of information-theoretic bounds established in the 1970s by Slepian and Wolf for distributed lossless coding, and by Wyner and Ziv for lossy coding with decoder side information who enabled low-complexity video encoding where the bulk of the computation is shifted to the decoder. Examples of such systems include wireless video sensors for surveillance, wireless PC cameras, mobile cameraphones, disposable video cameras, and networked camcorders. In all of these cases, compression must be implemented at the camera where memory and computation are scarce. Schemes that build upon these theorems are generally referred to as distributed coding algorithms.



Stereoscopic video consists of two sequences of frames, one left and one right sequence, captured by two stereo cameras, each of these cameras captures the left sequence and the right sequence, respectively. Distributed coding is preferred for compression of stereoscopic video because it avoids communication between the stereo cameras. Since the interframe dependence of the video sequence is exploited only at the decoder, an intraframe encoder can be combined with an interframe decoder. The difficulty, however, lies in discovering and exploiting the scene-dependent disparity at the decoder, while keeping the transmission rate low.

We consider the problem of distributed compression of stereoscopic video with motion and disparity estimation at the decoder. We develop an iterative EM algorithm that alternates between updating the motion and disparity distribution and the source model during the decoding procedure. Towards this, is used a method for joint bitplane LDPC decoding of grayscale images. The experiments show that motion and disparity estimation can achieve twice as efficient compression compared to a system with no disparity and motion compensation and perform nearly as well as a system which knows the disparity and motion through an oracle. It is also superior to conventional compression using JPEG.



ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή

1.2 Διάρθρωση Εργασίας

1.1. Εισαγωγή

Στην καθημερινότητά μας χειριζόμαστε ψηφιακά δεδομένα ήχου, εικόνας και βίντεο τα οποία έχουν τεράστιο μέγεθος δηλαδή τα εικονοστοιχεία που απαιτούνται για την αναπαράστασή τους είναι πάρα πολλά. Ωστόσο, οι δυνατότητες επικοινωνίας, τα μέσα μετάδοσης και αποθήκευσης είναι περιορισμένα και με αρκετό κόστος. Γι' αυτό απαιτούνται τεχνικές που θα μειώσουν το μέγεθος αυτών των δεδομένων. Παραδείγματα προγραμμάτων συμπίεσης που χρησιμοποιούμε καθημερινά έχουν τις εξής καταλήξεις: .zip, .gz, .tar, .mpeg, .mp3, .avi, .jpg κτλ.

Τα τελευταία χρόνια η κωδικοποίηση εικόνας, ήχου και βίντεο είναι μεταξύ των τεχνολογιών πληροφορίας που έχουν συμβάλει πιο πολύ στην αλλαγή και βελτίωση της καθημερινής ζωής. Δημοφιλείς υπηρεσίες που χρησιμοποιούν τέτοιες τεχνολογίες είναι οι ψηφιακές βιντεοκάμερες, η ψηφιακή τηλεόραση, οι συσκευές dvd, οι συσκευές mp3 κτλ. Ο σκοπός των τεχνολογιών κωδικοποίησης είναι να συμπίεσουν την αρχική πληροφορία σε πολύ μικρότερο αριθμό δυαδικών ψηφίων χωρίς να επηρεαστεί η ποιότητα του αποκωδικοποιημένου σήματος και να ικανοποιήσουν ένα σύνολο απαιτήσεων που εξαρτώνται από την εκάστοτε εφαρμογή.



1.2. Διάρθρωση Εργασίας

Στο κεφάλαιο 2 περιγράφονται διαδικασίες για την κωδικοποίηση/συμπύεση εικόνας και βίντεο. Στην παράγραφο 2.2 αναλύονται τα στάδια και τα βήματα κωδικοποίησης εικόνας. Στην παράγραφο 2.3 περιγράφεται η μέθοδος κωδικοποίησης βίντεο με πρόβλεψη (intra-frame coding). Η πρόβλεψη γίνεται με δύο τρόπους: χωρίς εκτίμηση και αντιστάθμιση κίνησης (παράγραφος 2.3.1) και με εκτίμηση και αντιστάθμιση κίνησης (παράγραφος 2.3.2). Στην παράγραφο 3.2 περιγράφεται η έννοια της κατανεμημένης συμπύεσης σύμφωνα με το θεώρημα Slepian-Wolf για μη απωλεστική κατανεμημένη συμπύεση. Στην παράγραφο 3.2.3 περιγράφεται η κατανεμημένη συμπύεση στα πλαίσια του θεωρήματος Wyner-Ziv για απωλεστική συμπύεση για το pixel domain και για το transform domain. Στην παράγραφο αυτή παρουσιάζεται η έννοια της στερεοσκοπίας και της διαφοράς (disparity), η εκτίμηση της διαφοράς (disparity estimation) και η σχέση της με την εκτίμηση κίνησης (motion estimation). Στην παράγραφο 4.2 περιγράφεται το μοντέλο κωδικοποίησης στερεοσκοπικού βίντεο Wyner-Ziv. Επίσης, παρουσιάζεται ο αλγόριθμος EM για τη διαδικασία αποκωδικοποίησης. Στο κεφάλαιο 5 παρατίθενται τα πειραματικά αποτελέσματα από την υλοποίηση του αλγορίθμου, ο επίλογος και τέλος μελλοντικές επεκτάσεις που θα μπορούσαν να γίνουν.



ΚΕΦΑΛΑΙΟ 2. ΚΩΔΙΚΟΠΟΙΗΣΗ ΕΙΚΟΝΑΣ ΚΑΙ ΒΙΝΤΕΟ

-
- 2.1 Εισαγωγή
 - 2.2 Κωδικοποίηση Εικόνας
 - 2.3 Κωδικοποίηση Βίντεο
-

2.1. Εισαγωγή

Στο κεφάλαιο αυτό περιγράφονται διαδικασίες για την κωδικοποίηση/συμπίεση εικόνας και βίντεο. Στην παράγραφο 2.2 αναλύονται τα στάδια και τα βήματα κωδικοποίησης εικόνας. Στην παράγραφο 2.3 περιγράφεται η μέθοδος κωδικοποίησης βίντεο με πρόβλεψη (intra-frame coding). Η πρόβλεψη γίνεται με δύο τρόπους: χωρίς εκτίμηση και αντιστάθμιση κίνησης (παράγραφος 2.3.1) και με εκτίμηση και αντιστάθμιση κίνησης (παράγραφος 2.3.2).

2.2. Κωδικοποίηση Εικόνας

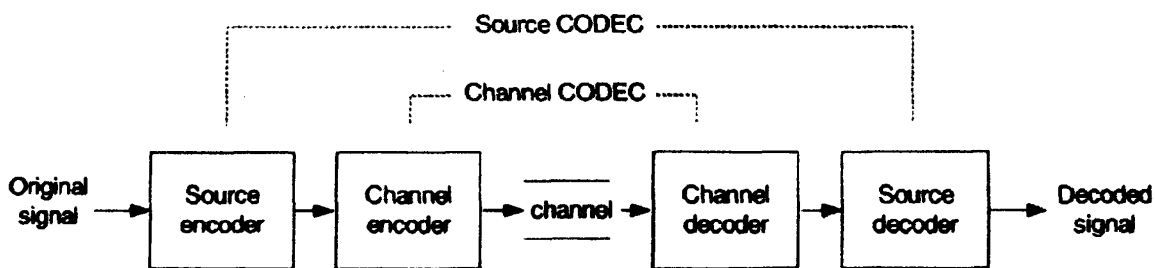
Η εικόνα είναι η προβολή μιας σκηνής τριών διαστάσεων στις δύο διαστάσεις και διαιρείται σε σημεία τα λεγόμενα εικονοστοιχεία (pixels). Με τη συμπίεση αναπαριστούμε την εικόνα σε μορφή που απαιτεί λιγότερα εικονοστοιχεία από την αρχική της μορφή.

Ένας αλγόριθμος συμπίεσης θα έπρεπε να ικανοποιεί τις ακόλουθες απαιτήσεις:



- Να επιτυγχάνει μεγάλη συμπίεση ενώ παράλληλα να διατηρεί αρκετά καλή ποιότητα έτσι ώστε το αποσυμπιεσμένο αρχείο να μην είναι πολύ διαφορετικό από το αρχικό και να περιέχει σχεδόν όλη τη σημαντική πληροφορία.
- Να είναι πολύ απλός με σκοπό την ελάχιστη πολυπλοκότητα που είναι δυνατόν να επιτευχθεί.
- Υπάρχουν δύο είδη τεχνικών συμπίεσης: Συμμετρική και ασύμμετρη. Η πρώτη χρειάζεται τον ίδιο χρόνο για συμπίεση και αποσυμπίεση. Ενώ η δεύτερη απαιτεί περισσότερο χρόνο σε μια από τις δύο τελευταίες διαδικασίες. Η ασύμμετρη είναι προτιμότερη γιατί ενώ το αρχείο συμπιέζεται μια φορά και δεν μας επιβαρύνει ιδιαίτερα, αποσυμπιέζεται αρκετές φορές όπου ο χρόνος είναι πιο σημαντικός.

Η συμπίεση πραγματοποιείται σε δύο στάδια: πρώτα κωδικοποιεί την εικόνα και στη συνέχεια την αποκωδικοποιεί. Αυτό το ζεύγος σταδίων λέγεται κωδικοποιητής.

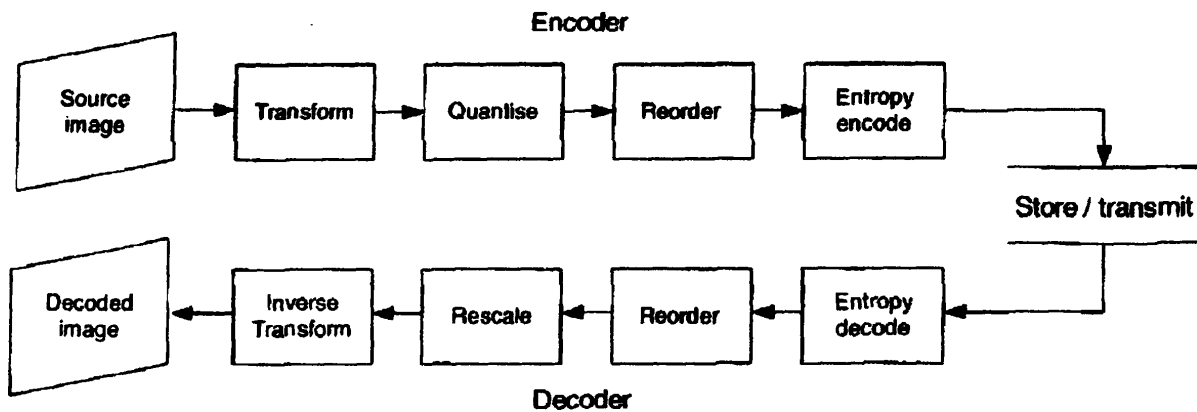


Σχήμα 2.1: Κωδικοποιητής πηγής και καναλιού.

Ένα παράδειγμα κωδικοποιητή φαίνεται στο σχήμα 2.1. Η αρχική ασυμπιεστη εικόνα (original signal) κωδικοποιείται (συμπιέζεται): αυτό είναι η κωδικοποίηση πηγής (source encoder). Η συμπιεσμένη εικόνα που προκύπτει συμπιέζεται περαιτέρω για να προσθέσει προστασία από σφάλματα που μπορεί να συμβούν κατά τη μετάδοσή του στο κανάλι επικοινωνίας (channel encoder). Στον παραλήπτη, ο αποκωδικοποιητής καναλιού (channel decoder) εντοπίζει και διορθώνει τα σφάλματα μετάδοσης και ο αποκωδικοποιητής της πηγής (source decoder) αποσυμπιέζει την εικόνα-σήμα (signal). Η αποσυμπιεσμένη εικόνα μπορεί να είναι ίδια με την αρχική εικόνα (μη απωλεστική συμπίεση) ή μπορεί να παραμορφωθεί (απωλεστική συμπίεση).



Μια εικόνα περιέχει πλεόνασμα που σημαίνει ότι μπορεί να αναπαρασταθεί με πιο συμπαγή τρόπο. Το πλεόνασμα είναι δύο ειδών: χρονικό πλεόνασμα (temporal) και χωρικό (spatial). Χρονικό πλεόνασμα: διαδοχικές εικόνες είναι όμοιες μεταξύ τους. Χωρικό πλεόνασμα: γειτονικά εικονοστοιχεία είναι παρόμοια. Η συμπίεση επιτυγχάνεται αξιοποιώντας το χρονικό και χωρικό πλεόνασμα.



Σχήμα 2.2: Γενικό σύστημα συμπίεσης/κωδικοποίησης εικόνας.

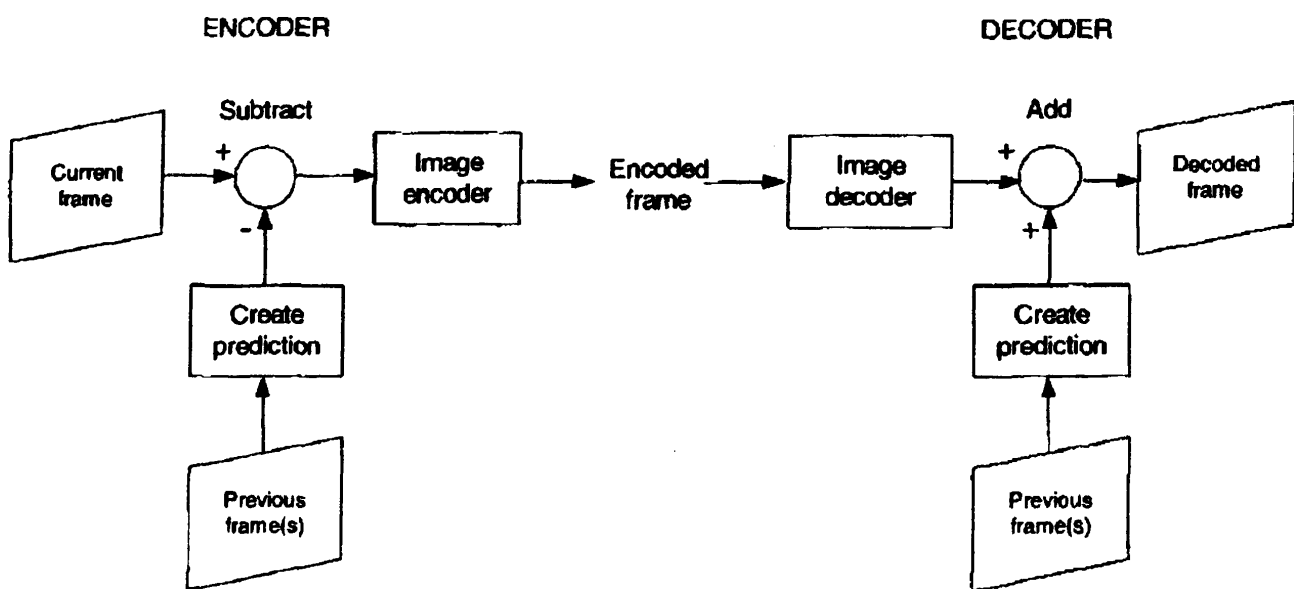
Ένα γενικό σύστημα συμπίεσης/κωδικοποίησης εικόνας φαίνεται στο σχήμα 2.2. Ένας κωδικοποιητής εικόνας κωδικοποιεί και αποκωδικοποιεί κάθε εικόνα ενός βίντεο. Τα στάδια κωδικοποίησης είναι τρία: μετασχηματισμός (transform), κβάντιση (quantise) και κωδικοποίηση εντροπίας (entropy encode). Η αποκωδικοποίηση είναι ίδια με την κωδικοποίηση αλλά με την αντίστροφη σειρά, δηλαδή αποκωδικοποίηση εντροπίας (entropy decode), αντίστροφη κβάντιση (rescale), αντίστροφος μετασχηματισμός (inverse transform).

2.3. Κωδικοποίηση Βίντεο

Ένα βίντεο αποτελείται από μια ακολουθία εικόνων (frame). Κάθε εικόνα μπορεί να συμπεστεί ξεχωριστά χρησιμοποιώντας κωδικοποιητή εικόνας όπως περιγράφηκε στην παράγραφο 2.2. Η συμπίεση κάθε εικόνας ξεχωριστά περιγράφεται ως intra-frame coding όπου κάθε εικόνα κωδικοποιείται χωρίς να αναφέρεται σε κάποια άλλη



εικόνα. Ένας κωδικοποιητής intra-frame είναι ο κωδικοποιητής JPEG. Ωστόσο, καλύτερη συμπίεση μπορεί να επιτευχθεί αξιοποιώντας το χρονικό πλεόνασμα δηλαδή τις ομοιότητες μεταξύ των εικόνων. Αυτό μπορεί να κατορθωθεί προσθέτοντας στον κωδικοποιητή εικόνας μια πρόβλεψη (prediction) της τρέχουσας εικόνας (current frame) που στηρίζεται σε μια ή περισσότερες προηγούμενες εικόνες (previous frames), σχήμα 2.3. Η πρόβλεψη αφαιρείται από την τρέχουσα εικόνα και η διαφορά που προκύπτει (residual frame) κωδικοποιείται χρησιμοποιώντας κωδικοποιητή εικόνας. Το κλειδί σε αυτή την προσέγγιση της πρόβλεψης είναι η εξής: αν η πρόβλεψη είναι ακριβής τότε η διαφορά θα έχει μικρό μέγεθος και άρα η συμπίεσή της (της διαφοράς) θα δώσει πολύ πιο μικρό μέγεθος. Προκειμένου να αποκωδικοποιηθεί η τρέχουσα εικόνα (current frame), ο αποκωδικοποιητής θα πρέπει να αντιστρέψει τη διαδικασία κωδικοποίησης. Στον αποκωδικοποιητή η διαφορά αποκωδικοποιείται και προστίθεται στην πρόβλεψη της τρέχουσας εικόνας που δημιουργεί η προηγούμενη αποκωδικοποιημένη εικόνα (previous frame) και το αποτέλεσμα είναι η αποκωδικοποιημένη ή ανακατασκευασμένη εικόνα (current decoded/reconstructed frame). Αυτό το είδος κωδικοποίησης λέγεται inter-frame coding: οι εικόνες κωδικοποιούνται στηριζόμενες στη σχέση τους με άλλες εικόνες του βίντεο δηλαδή αξιοποιούν την αλληλεξάρτηση μεταξύ τους.

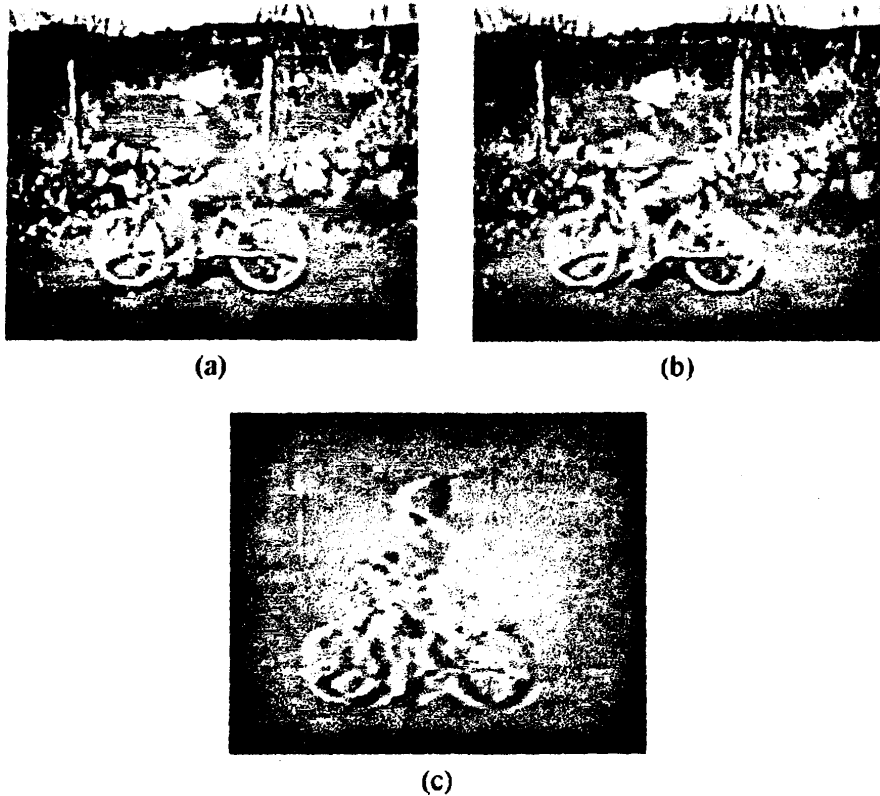


Σχήμα 2.3: Κωδικοποίηση βίντεο με πρόβλεψη (prediction).



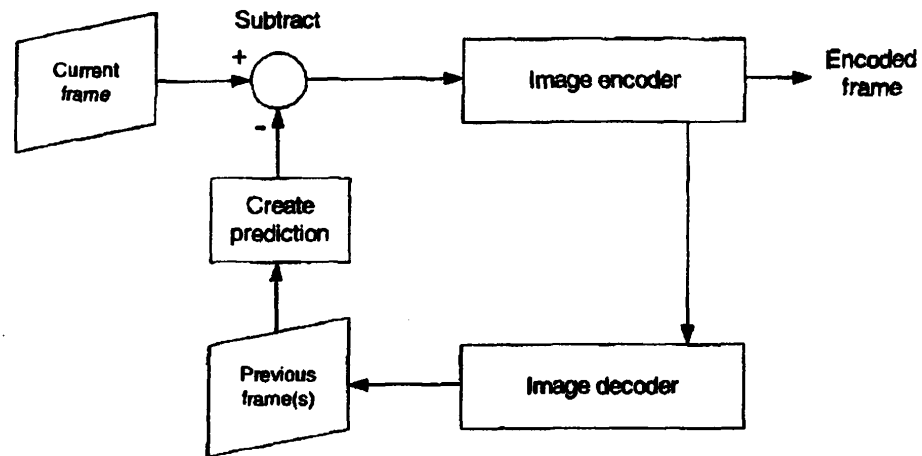
2.3.1. Κωδικοποίηση εικόνων χωρίς εκτίμηση και αντιστάθμιση κίνησης (no motion compensation)

Η πιο απλή μέθοδος πρόβλεψης χρησιμοποιεί μόνο την προηγούμενη εικόνα. Έστω δύο εικόνες σχήμα 2.4(a) και 2.4(b).



Σχήμα 2.4: (a) τρέχουσα εικόνα (current frame), (b) προηγούμενη εικόνα (previous frame), (c) η διαφορά χωρίς αντιστάθμιση κίνησης (no motion compensation)

Η διαφορά τους, σχήμα 2.4(c), προκύπτει αφαιρώντας την προηγούμενη εικόνα από την τρέχουσα. Οι περιοχές της διαφοράς που έχουν γκρι χρώμα έχουν μηδενικές τιμές ενώ αυτές με άσπρο ή μαύρο χρώμα έχουν θετικές ή αρνητικές τιμές αντίστοιχα. Το μεγαλύτερο μέρος της διαφοράς είναι μηδέν που σημαίνει ότι το μέγεθος της είναι πολύ μικρότερο από την τρέχουσα εικόνα και συμφέρει πιο πολύ να συμπιεστεί η διαφορά παρά η ίδια η τρέχουσα.



Σχήμα 2.5: Κωδικοποιητής με βρόχο αποκωδικοποίησης.

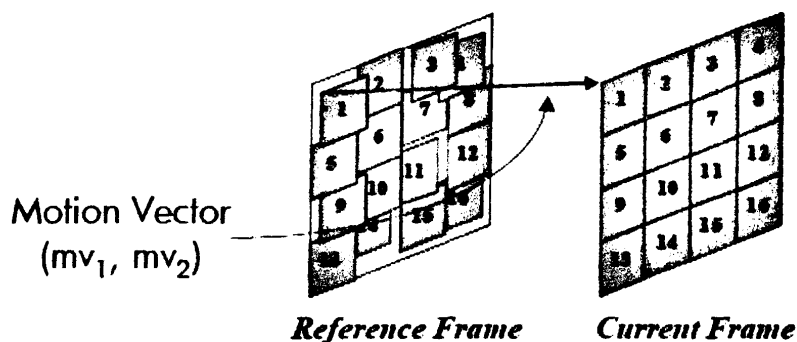
Η διαδικασία με την οποία κωδικοποιείται μια ακολουθία εικόνων με πρόβλεψη είναι η εξής: η πρώτη εικόνα κωδικοποιείται και αποκωδικοποιείται χωρίς να χρησιμοποιεί πρόβλεψη, σχήμα 2.5. Η δεύτερη εικόνα χρησιμοποιεί την πρώτη εικόνα για βρει τη διαφορά τους και να στείλει τη διαφορά στον κωδικοποιητή για συμπίεση. Στον αποκωδικοποιητή η πρώτη εικόνα είναι τώρα πλέον αποκωδικοποιημένη και αν προστεθεί σε αυτήν η αποκωδικοποιημένη διαφορά τότε δεν θα μας δώσει την δεύτερη εικόνα. Το λάθος βρίσκεται στο ότι στον κωδικοποιητή η πρόβλεψη της δεύτερης εικόνας γίνεται από την πρώτη *αρχική* εικόνα ενώ στον αποκωδικοποιητή η πρόβλεψή της (της δεύτερης εικόνας) γίνεται από την πρώτη *αποκωδικοποιημένη* εικόνα. Έτσι οι δύο προβλέψεις είναι διαφορετικές οπότε και η αποκωδικοποίηση της δεύτερης εικόνας θα είναι εσφαλμένη. Για το λόγο αυτό πρέπει να προστεθεί και στον κωδικοποιητή μια διαδικασία αποκωδικοποίησης που θα κάνει πρόβλεψη της δεύτερης εικόνας από την πρώτη εικόνα αλλά αποκωδικοποιημένη τώρα πλέον έτσι ώστε να είναι ίδια με αυτή στον αποκωδικοποιητή. Έτσι στο σχήμα 2.5, η δεύτερη εικόνα χρησιμοποιεί την πρόβλεψη της πρώτης αποκωδικοποιημένης εικόνας και η διαφορά τους πλέον κωδικοποιείται. Η διαφορά εκτός από την κωδικοποίηση και την αποστολή της, επιπλέον περνάει στον βρόχο αποκωδικοποίησης του κωδικοποιητή όπου βρίσκει την πρώτη αποκωδικοποιημένη εικόνα η οποία θα δώσει την πρόβλεψη για την δεύτερη εικόνα κ.ο.κ.

2.3.2. Κωδικοποίηση εικόνων με εκτίμηση και αντιστάθμιση κίνησης (*motion estimation and compensation*)

Η inter-frame κωδικοποίηση χωρίς εκτίμηση κίνησης, δηλαδή η κωδικοποίηση της διαφοράς της τρέχουσας από την προηγούμενη εικόνα, δίνει καλύτερη συμπίεση από την intra-frame κωδικοποίηση όταν δεν υπάρχει κίνηση δηλαδή οι διαδοχικές εικόνες του βίντεο είναι πολύ όμοιες μεταξύ τους. Αν όμως υπάρχει σημαντική διαφορά μεταξύ των διαδοχικών εικόνων δηλαδή υπάρχει κίνηση, τότε η συμπίεση μόνο της διαφοράς δεν είναι τόσο αποδοτική γι' αυτό κάνουμε εκτίμηση αυτής της κίνησης για να βρούμε καλύτερη πρόβλεψη. Η πρόβλεψη με εκτίμηση κίνησης επιτυγχάνεται σε δύο στάδια:

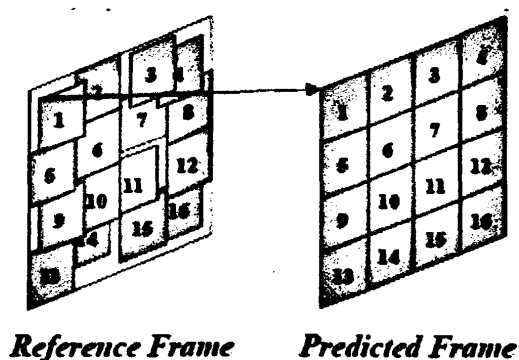
- Εκτίμηση κίνησης (*motion estimation*): σχήμα 2.6, χωρίζουμε την τρέχουσα εικόνα (*current frame*) σε μπλοκ, για κάθε μπλοκ βρίσκουμε το μπλοκ της ανακατασκευασμένης εικόνας (*reference frame*) που ταιριάζει καλύτερα σε αυτό ως εξής: κάθε μπλοκ της τρέχουσας εικόνας το συγκρίνουμε με το αντίστοιχο μπλοκ της προηγούμενης αποκωδικοποιημένης/ανακατασκευασμένης εικόνας καθώς και με άλλα μπλοκ γειτονικά του (του μπλοκ της ανακατασκευασμένης εικόνας) σε μια περιοχή αναζήτησης που το μέγεθός της προσδιορίζεται από την εκάστοτε εφαρμογή. Το κριτήριο για να βρεθεί το καλύτερο μπλοκ είναι το ελάχιστο τετραγωνικό σφάλμα των δύο μπλοκ. Στην περιοχή αναζήτησης της ανακατασκευασμένης εικόνας για το καλύτερο μπλοκ, συγκρίνουμε όλα τα μπλοκ ή ένα υποσύνολό τους. Η θέση του καλύτερου μπλοκ προσδιορίζεται από το διάνυσμα κίνησης που δείχνει την οριζόντια και κάθετη μετατόπισή του σε σχέση με το μπλοκ που έχει μετατόπιση (0,0).



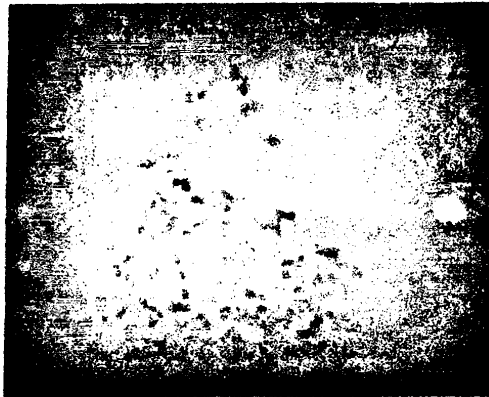


Σχήμα 2.6: Εκτίμηση κίνησης (motion estimation) με ταίριασμα των μπλοκ της τρέχουσας εικόνας με αυτών της εικόνας αναφοράς.

- Αντιστάθμιση κίνησης (motion compensation): Τα διανύσματα κίνησης που βρήκαμε στην εκτίμηση κίνησης, τα χρησιμοποιούμε για να πάρουμε τα κατάλληλα μπλοκ της ανακατασκευασμένης εικόνας και να βρούμε την πρόβλεψη της τρέχουσας εικόνας, σχήμα 2.7. Αφαιρώντας την πρόβλεψη από την τρέχουσα εικόνα προκύπτει η διαφορά η οποία φαίνεται στο σχήμα 2.8. Παρατηρούμε ότι η διαφορά της αντιστάθμισης κίνησης είναι πολύ μικρότερη από την διαφορά του σχήματος 2.4 που δεν χρησιμοποιήθηκε εκτίμηση και αντιστάθμιση κίνησης. Η μέθοδος της πρόβλεψης με εκτίμηση και αντιστάθμιση κίνησης έχει μεγαλύτερο κόστος πολυπλοκότητας από τη μέθοδο της πρόβλεψης χωρίς εκτίμηση και αντιστάθμιση κίνησης λόγω της διαδικασίας εκτίμησης κίνησης.

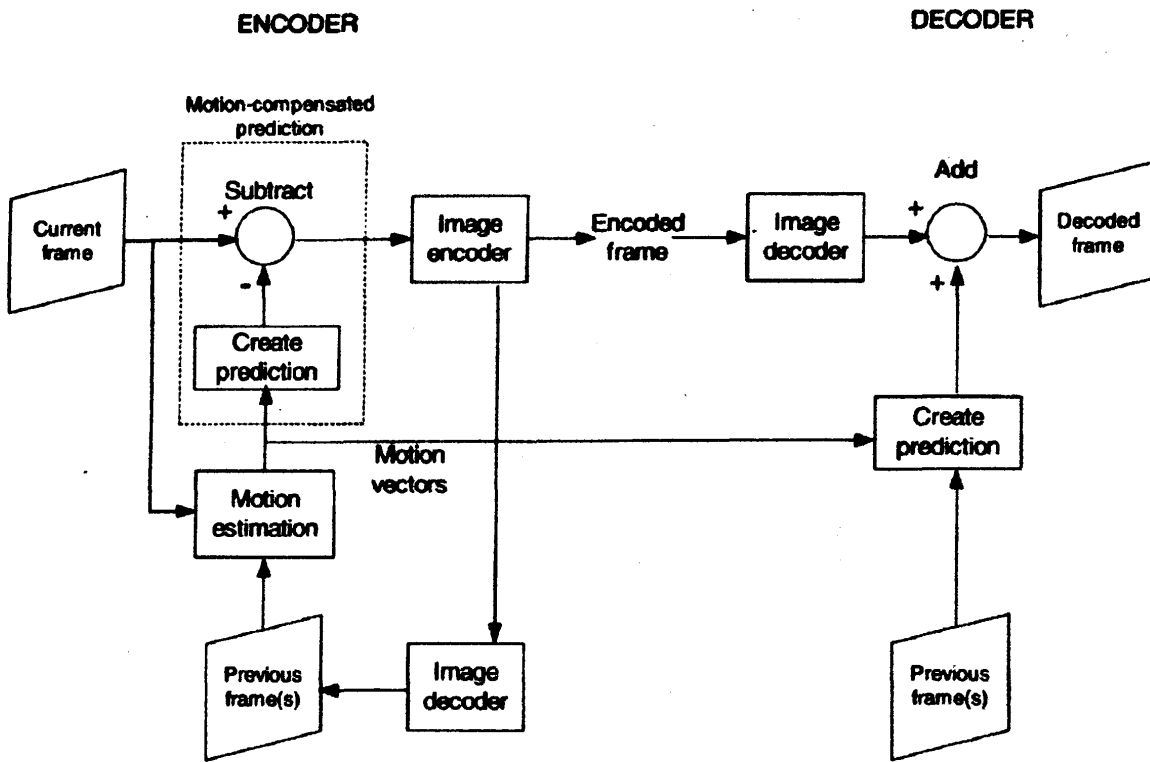


Σχήμα 2.7: Αντιστάθμιση κίνησης (motion compensation) για την εύρεση της πρόβλεψης της τρέχουσας εικόνας (predicted frame) από την ανακατασκευασμένη εικόνα (reference frame).

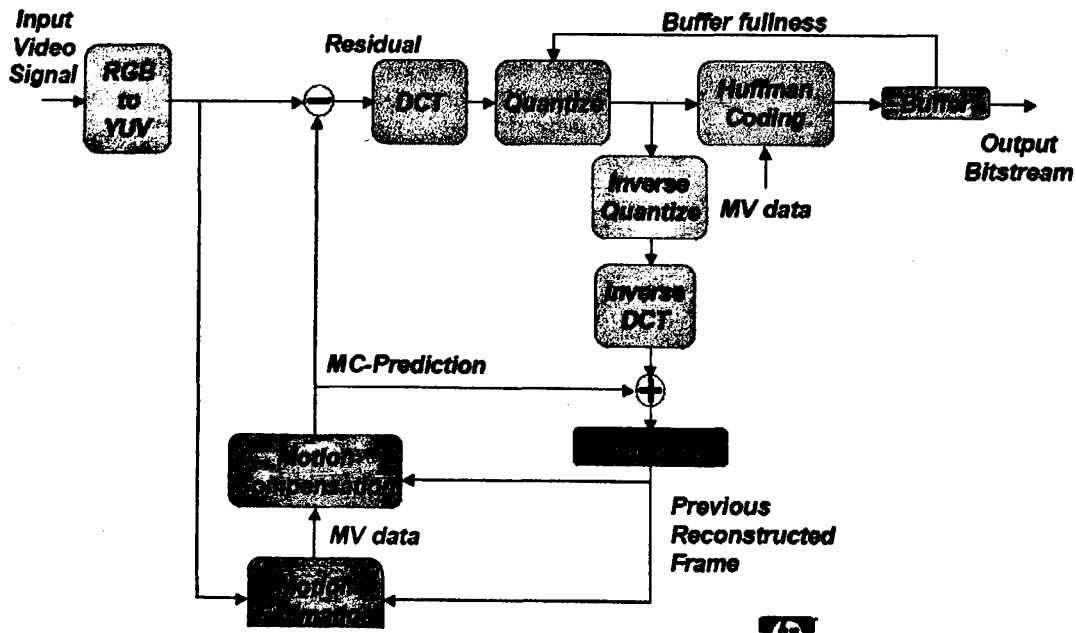


Σχήμα 2.8: Η διαφορά που προκύπτει από την αφαίρεση της πρόβλεψης (που προκύπτει από την αντιστάθμιση κίνησης) από την τρέχουσα εικόνα.

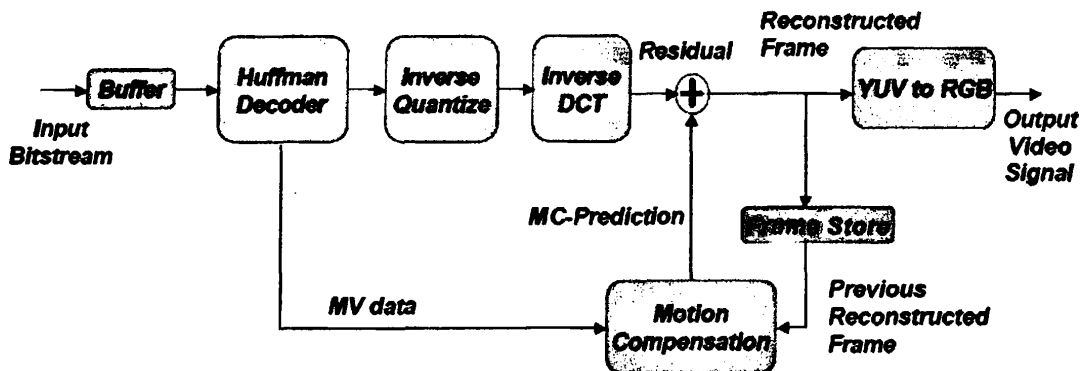
Η όλη διαδικασία inter-frame κωδικοποίησης και αποκωδικοποίησης εικόνας με εκτίμηση και αντιστάθμιση κίνησης φαίνεται στο σχήμα 2.9. Η διαφορά και τα διανύσματα κίνησης κωδικοποιούνται και αποστέλλονται στον αποκωδικοποιητή. Εκεί η διαδικασία αποκωδικοποίησης είναι πιο απλή διότι δεν χρειάζεται να γίνει εκτίμηση κίνησης επειδή τα διανύσματα κίνησης είναι έτοιμα και απλά χρησιμοποιούνται για να βρεθεί η πρόβλεψη (motion compensation). Αν στην πρόβλεψη προσθέσουμε την αποκωδικοποιημένη διαφορά τότε παίρνουμε την αποκωδικοποιημένη τρέχουσα εικόνα (decoded frame). Η διαδικασία κωδικοποίησης/αποκωδικοποίησης φαίνεται πιο αναλυτικά στα σχήματα 2.10, 2.11.



Σχήμα 2.9: Κωδικοποιητής βίντεο με εκτίμηση και αντιστάθμιση κίνησης (motion estimation and compensation).



Σχήμα 2.10: Κωδικοποιητής βίντεο με εκτίμηση και αντιστάθμιση κίνησης (motion estimation and compensation).

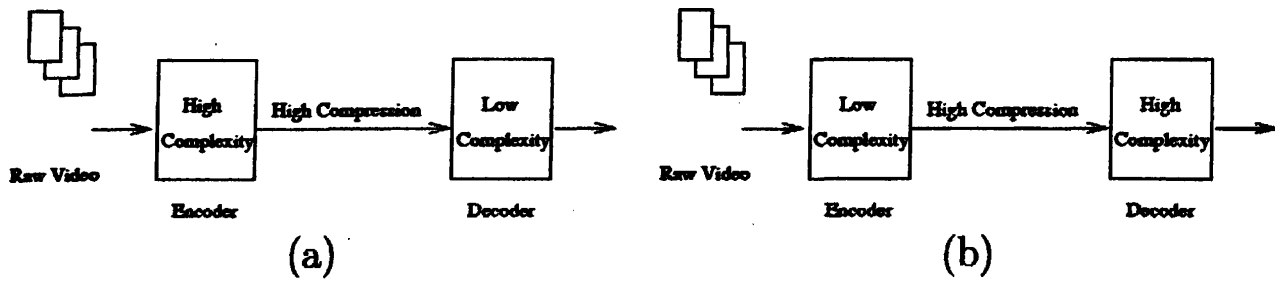


Σχήμα 2.11: Αποκωδικοποιητής βίντεο με εκτίμηση και αντιστάθμιση κίνησης (motion estimation and compensation).

Η μέθοδος κωδικοποίησης ψηφιακού βίντεο που χρησιμοποιείται στις μέρες μας έχει παρουσιαστεί από τα πρότυπα ITU-T και MPEG και στηρίζεται κυρίως σε προσεγγίσεις inter-frame κωδικοποίησης με πρόβλεψη. Σε αυτά τα πλαίσια προσεγγίσεων, ο κωδικοποιητής έχει ως ρόλο να αξιοποιήσει το χρονικό και χωρικό πλεόνασμα της ακολουθίας εικόνων πράγμα το οποίο είναι μια πολύπλοκη διαδικασία. Κατά συνέπεια, οι κωδικοποιητές βίντεο (video encoders) σε όλα τα πρότυπα έχουν κατά πολύ μεγαλύτερη υπολογιστική πολυπλοκότητα από τον αποκωδικοποιητή (decoder) κυρίως λόγω του χρονικού πλεονάσματος που είναι η διαδικασία της εκτίμησης κίνησης (motion estimation), σχήμα 2.12. Αυτό έχει σαν συνέπεια, ο κωδικοποιητής να έχει πολύ μεγαλύτερο κόστος από τον αποκωδικοποιητή. Αυτός ο τύπος αρχιτεκτονικής είναι κατάλληλος για εφαρμογές όπου το βίντεο κωδικοποιείται μια φορά και αποκωδικοποιείται πολλές φορές δηλαδή το κωδικοποιημένο βίντεο εκπέμπεται σε πολλούς δέκτες. Για παράδειγμα, η ραδιοφωνία μεταδίδει σήματα σε ένα ευρύ κοινό, έτσι το κωδικοποιημένο βίντεο αποκωδικοποιείται από τον κάθε έναν από αυτό το κοινό. Ωστόσο υπάρχουν εφαρμογές που απαιτούν απλό κωδικοποιητή και πολύπλοκο αποκωδικοποιητή. Για



την ικανοποίηση αυτής της απαίτησης έχει αναπτυχθεί η κατανεμημένη συμπίεση βίντεο (distributed source coding) που θα αναφερθούμε στο επόμενο κεφάλαιο.



Σχήμα 2.12: (a) Μέθοδος κωδικοποίησης με πολύπλοκο κωδικοποιητή και απλό αποκωδικοποιητή, (b) Μέθοδος κωδικοποίησης με απλό κωδικοποιητή και πολύπλοκο αποκωδικοποιητή.

ΚΕΦΑΛΑΙΟ 3. ΚΑΤΑΝΕΜΗΜΕΝΗ ΣΥΜΠΙΕΣΗ ΒΙΝΤΕΟ

3.1 Εισαγωγή

3.2 Κατανεμημένη Συμπίεση Βίντεο

3.1. Εισαγωγή

Στην παράγραφο 3.2 περιγράφεται η έννοια της κατανεμημένης συμπίεσης σύμφωνα με το θεώρημα Slepian-Wolf για μη απωλεστική κατανεμημένη συμπίεση. Στην παράγραφο 3.2.3 περιγράφεται η κατανεμημένη συμπίεση στα πλαίσια του θεωρήματος Wyner-Ziv για απωλεστική συμπίεση για το pixel domain και για το transform domain.

3.2. Κατανεμημένη Συμπίεση Βίντεο (Distributed Video Coding)

Με την πρόσφατη πρόοδο στα ασύρματα δίκτυα (wireless networks), πολλοί αποστολείς στέλνουν τα δεδομένα τους σε έναν κεντρικό παραλήπτη. Παραδείγματα τέτοιων εφαρμογών είναι η ασύρματη ψηφιακή βιντεοκάμερα, τα δίκτυα αισθητήρων βίντεο και τα συστήματα surveillance. Αυτές οι εφαρμογές απαιτούν ελαφρό κωδικοποιητή και ευέλικτη κατανομή της πολυπλοκότητας, σχήμα 2.12. Επίσης παρατηρείται μια αύξηση στην κωδικοποίηση δεδομένων που βιντεοσκοπούνται από πολλές οπτικές γωνίες μιας σκηνής (multiview video coding). Σε αυτές τις περιπτώσεις οι κάμερες δεν μπορούν να επικοινωνήσουν μεταξύ τους γι' αυτό το

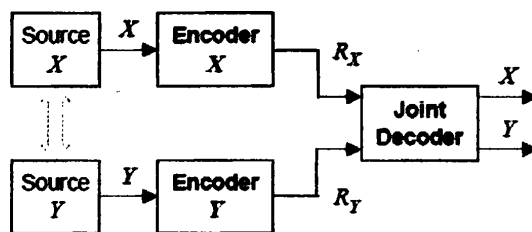


λόγο δεν μπορεί να χρησιμοποιηθεί η κλασική προσέγγιση κωδικοποίησης βίντεο με πρόβλεψη η οποία αξιοποιεί τη συσχέτιση των καμερών στον κωδικοποιητή. Επομένως απαιτείται μια άλλη προσέγγιση κωδικοποίησης βίντεο που ανταποκρίνεται στην παραπάνω απαίτηση.

Το 2002 κάποιες ομάδες ερευνητών επανεξέτασαν το πρόβλημα κωδικοποίησης βίντεο σύμφωνα με δύο θεωρίες του 1970: το θεώρημα Slepian-Wolf [2] και το θεώρημα Wyner-Ziv [3]. Αυτές οι προσπάθειες είχαν ως αποτέλεσμα την κατανεμημένη κωδικοποίηση βίντεο (distributed video coding), και την κωδικοποίηση Wyner-Ziv βίντεο σαν μια ειδική περίπτωση της κατανεμημένης κωδικοποίησης.

3.2.1. Θεώρημα Slepian-Wolf για μη απωλεστική κατανεμημένη κωδικοποίηση

Η κατανεμημένη κωδικοποίηση αναφέρεται στην κωδικοποίηση δύο πηγών πληροφορίας (source) που εξαρτώνται στατιστικά αλλά δεν επικοινωνούν μεταξύ τους και έχουν ξεχωριστό κωδικοποιητή [1], σχήμα 3.1. Κάθε κωδικοποιητής στέλνει μια ξεχωριστή ροή δυαδικών ψηφίων σε έναν αποκωδικοποιητή όπου θα αξιοποιούνται οι στατιστικές εξαρτήσεις όλων των εισερχόμενων ροών και θα γίνεται μια από κοινού αποκωδικοποίηση (joint decoder).

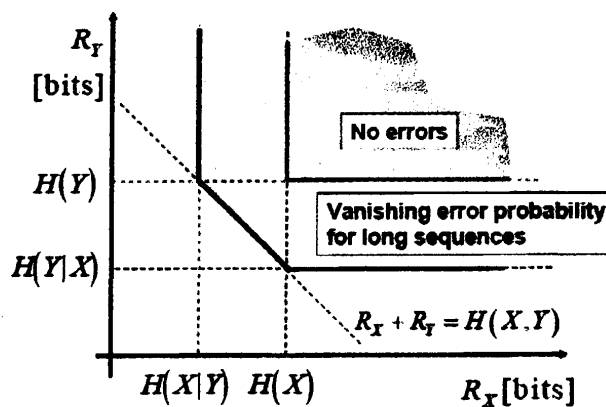


Σχήμα 3.1: Κατανεμημένη συμπίεση δύο πηγών πληροφορίας που εξαρτώνται στατιστικά μεταξύ τους, X και Y . Ο αποκωδικοποιητής αποκωδικοποιεί από κοινού (joint decoder) τα X και Y και έτσι αξιοποιεί την εξάρτησή τους.

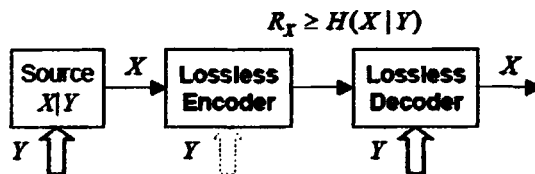
Ας θεωρήσουμε δύο στατιστικά εξαρτώμενες (statistically dependent), ανεξάρτητες και ισόνομα κατανεμημένες (independent and identically distributed), τυχαίες



ακολουθίες X και Y . Με δεδομένους κλασικούς ανεξάρτητους (χωριστούς) κωδικοποιητές (encoder X , encoder Y , σχήμα 3.2) και αποκωδικοποιητές εντροπίας μπορούν να επιτευχθούν τα εξής: $R_X \geq H(X)$, $R_Y \geq H(Y)$, όπου $H(X)$, $H(Y)$ είναι οι εντροπίες των X και Y αντίστοιχα και R_X , R_Y οι ρυθμοί τους (bitrate) [2]. Αν τώρα έχουμε χωριστούς κωδικοποιητές και έναν κοινό αποκωδικοποιητή (joint decoder) τότε: $R_X + R_Y \geq H(X, Y)$, $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$, σχήμα 3.2. Το άθροισμα των ρυθμών, $R_X + R_Y$, μπορεί να φτάσει την από κοινού εντροπία $H(X, Y)$, όπως ακριβώς με τον από κοινού κωδικοποιητή (joint encoder) των X και Y .



Σχήμα 3.2: Θεώρημα Slepian Wolf, 1973: Περιοχή του ρυθμού που μπορεί να επιτευχθεί για καταμετρημένη συμπίεση δύο στατιστικά εξαρτημένων πηγών X και Y .



Σχήμα 3.3: Συμπίεση πηγής X χρησιμοποιώντας side information Y . Το Y είναι διαθέσιμο μόνο στον αποκωδικοποιητή και όχι στον κωδικοποιητή.

Η συμπίεση με side information στον αποκωδικοποιητή, σχήμα 3.3, είναι μια ειδική περίπτωση του προβλήματος της κατανεμημένης κωδικοποίησης, σχήμα 3.1. Η πηγή παράγει μια ακολουθία X με στατιστικά που εξαρτώνται από το side information Y . Μας ενδιαφέρει η περίπτωση στην οποία το side information Y είναι διαθέσιμο στον αποκωδικοποιητή και όχι στον κωδικοποιητή. Αφού $R_Y \geq H(Y)$ είναι εφικτό για κωδικοποίηση Y , η συμπίεση με side information στον αποκωδικοποιητή αντιστοιχεί σε μια από τις γωνίες της περιοχής ρυθμού στο σχήμα 3.2, και επομένως $R_X \geq H(X|Y)$, ανεξάρτητα από την πρόσβαση του κωδικοποιητή στο side information Y .

3.2.2. Πρακτική κωδικοποίηση Slepian-Wolf

Παρόλο που το θεώρημα Slepian-Wolf μας πηγαίνει πίσω στο 1970, μόνο τα τελευταία χρόνια αναδυόμενες εφαρμογές έχουν κινητοποιήσει σοβαρές προσπάθειες για να βρεθούν πρακτικές τεχνικές. Ωστόσο, έχει ήδη κατανοηθεί 30 χρόνια πιο πριν ότι η κωδικοποίηση Slepian-Wolf είναι παρόμοια με την κωδικοποίηση καναλιού (channel coding) [4]. Για να εκτιμήσουμε αυτή την ομοιότητα, ας θεωρήσουμε δύο ανεξάρτητες και ισόνομα κατανεμημένες δυαδικές πηγές X και Y , σχήμα 3.3. Αν X και Y είναι όμοια, τότε έστω μια υποθετική ακολουθία σφαλμάτων $\Delta = X \oplus Y$ αποτελείται από μηδενικά, εκτός από κάποιους άσους που δηλώνουν τις θέσεις στις οποίες τα X και Y διαφέρουν. Προκειμένου να προστατευτεί το X από σφάλματα Δ , θα μπορούσε να εφαρμοστεί μια συστηματική κωδικοποίηση καναλιού και να σταλούν μόνο τα δυαδικά ψηφία ισοτιμίας (parity bits). Στον αποκωδικοποιητή, τα ψηφία ισοτιμίας και το side information Y θα συνενωθούν και θα εκτελεστεί αποκωδικοποίηση για την διόρθωση σφαλμάτων. Αν τα X και Y είναι πολύ όμοια, μόνο λίγα ψηφία ισοτιμίας θα έπρεπε να σταλούν και θα κατορθωνόταν σημαντική συμπίεση. Τονίζεται ότι αυτή η προσέγγιση δεν εκτελεί διόρθωση σφαλμάτων (forward error correction) για την προστασία από σφάλματα του καναλιού μετάδοσης,



αλλά για ένα εικονικό κανάλι συσχέτισης που συλλαμβάνει τη στατιστική συσχέτιση του X και του side information Y .

Χρησιμοποιώντας την ερμηνεία με τα ψηφία ισοτιμίας, στέλνεται ένα δυαδικό διάνυσμα $X_p = XP$, όπου $G = [I|P]$ είναι ο πίνακας (generator matrix) ενός συστηματικού γραμμικού κώδικα μπλοκ C_p (linear block code). Με τον ίδιο τρόπο, στέλνουμε ένα syndrome $S = XH$, όπου H είναι ο πίνακας ελέγχου ισοτιμίας (parity check matrix) ενός γραμμικού κώδικα μπλοκ C_x . Αν $P = H$, τότε οι μεταδιδόμενες ροές δυαδικών ψηφίων είναι παρόμοιες.

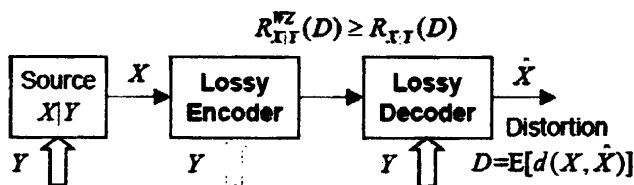
Οι περισσότερες τεχνικές κατανεμημένης κωδικοποίησης στις μέρες μας προέρχονται από αποδεδειγμένες ιδέες κωδικοποίησης καναλιού. Σχετικές πρόσφατες εργασίες έγιναν το 1999 από τους Pradhan και Ramchandran [5]. Αρχικά, ασχολήθηκαν με την ασύμμετρη περίπτωση (asymmetric) της κατανεμημένης κωδικοποίησης με side information στον αποκωδικοποιητή για στατιστικά εξαρτώμενες δυαδικές και γκαουσιανές πηγές χρησιμοποιώντας βαθμωτές (scalar) και trellis coset κατασκευές. Η εργασία τους αργότερα [6]-[9] αφορά την συμμετρική περίπτωση όπου τα X και Y κωδικοποιούνται με τον ίδιο ρυθμό. Wang και Orchard [10] χρησιμοποίησαν μια δομή κώδικα trellis για την ασύμμετρη κωδικοποίηση γκαουσιανών πηγών και έδειξαν βελτιώσεις [5]. Από τότε πιο εξελιγμένες τεχνικές κωδικοποίησης καναλιού έχουν υιοθετηθεί για το πρόβλημα κατανεμημένης κωδικοποίησης πηγής. Αυτά συχνά απαιτούν επαναληπτικούς αποκωδικοποιητές, όπως Bayesian δίκτυα ή Viterbi αποκωδικοποιητές. Ενώ οι κωδικοποιητές τείνουν να είναι πολύ απλοί, ο φόρτος υπολογισμού για τον αποκωδικοποιητή, ο οποίος αξιοποιεί τις στατιστικές των πηγών, είναι πολύ πιο υψηλός. Οι Garcia-Frias και Zhao [9], [10], Bajcsy και Mitran [11], [12], Aaron και Girod [13] ανεξάρτητα πρότειναν τεχνικές συμπίεσης όπου στατιστικά εξαρτώμενες δυαδικές πηγές συμπιέζονται χρησιμοποιώντας turbo κώδικες. Οι τεχνικές που στηρίζονται σε turbo κώδικες χρησιμοποιούνται για τη συμπίεση στατιστικά εξαρτώμενων μη δυαδικών συμβόλων [14], [15] και γκαουσιανών πηγών [13], [16], καθώς επίσης για τη συμπίεση μιας πηγής [10], [17]-[19]. Επαναληπτικοί κώδικες καναλιού μπορούν επίσης να εφαρμοστούν στην



αποκωδικοποίηση πηγής-καναλιού συμπεριλαμβάνοντας τις στατιστικές και της πηγής και του καναλιού στη διαδικασία αποκωδικοποίησης [13], [17]–[21]. Ο Liveris *et al.* [21]–[23], ο Schonberg *et al.* [24]–[26] και άλλοι συγγραφείς [27]–[30] έχουν προτείνει ότι οι κώδικες Low-Density Parity-Check (LDPC) μπορεί να είναι μια δυνατή εναλλακτική των κωδίκων turbo για την κατανεμημένη κωδικοποίηση. Με τους εξελιγμένους κώδικες turbo ή LDPC, η συμπίεση προσεγγίζει τα όρια Slepian-Wolf.

3.2.3. Θεωρία ρυθμού-παραμόρφωσης (Rate-Distortion) για απωλεστική συμπίεση με side information στον αποκωδικοποιητή

Μετά την δημοσίευση των Slepian-Wolf, οι Wyner-Ziv [31]–[33], επέκτειναν τη θεωρία των Slepian-Wolf για να βρουν θεωρητικά όρια για απωλεστική συμπίεση με side information στον αποκωδικοποιητή. Πιο συγκεκριμένα, έστω ότι X και Y είναι δύο ανεξάρτητες και ισόνομα κατανεμημένες τυχαίες ακολουθίες που μοντελοποιούν την πληροφορία πηγής και side information αντίστοιχα. Οι τιμές της πηγής X κωδικοποιούνται χωρίς πρόσβαση στο side information Y , σχήμα 3.4. Ο αποκωδικοποιητής ωστόσο έχει πρόσβαση στο Y , και βρίσκει μια ανακατασκευή \hat{X} της πηγής. Η παραμόρφωση είναι ίση με $D = E[d(X, \hat{X})]$. Το κατώτατο όριο ρυθμού (bit-rate) για την παραμόρφωση D είναι $R_{X|Y}^{WZ}(D)$. Σημειώνουμε με $R_{X|Y}(D)$ το bitrate που απαιτείται αν το side information βρισκόταν στον κωδικοποιητή.

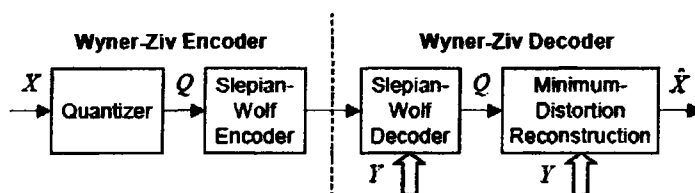


Σχήμα 3.4: Απωλεστική συμπίεση πηγής X χρησιμοποιώντας side information Y με το οποίο σχετίζεται στατιστικά.



Οι Wyner και Ziv απέδειξαν ότι υπάρχει μια απώλεια στο ρυθμό $R_{X|Y}^{WZ}(D)$ - $R_{X|Y}(D) \geq 0$ όταν ο κωδικοποιητής δεν έχει πρόσβαση στο side information. Ωστόσο, επίσης απέδειξαν ότι $R_{X|Y}^{WZ}(D) - R_{X|Y}(D) = 0$ στην περίπτωση των γκαουσιανών πηγών χωρίς μνήμη και τη μέση τετραγωνική παραμόρφωση (mean-squared error distortion) [31], [33]. Για γενικές στατιστικές και μέτρο μέσης τετραγωνικής παραμόρφωσης, ο Zamir [34] απέδειξε ότι η απώλεια ρυθμού είναι λιγότερη από 0.5 bits/pixel.

3.2.4. Κωδικοποίηση Wyner-Ziv



Σχήμα 3.5: Ένας κωδικοποιητής Wyner-Ziv αποτελείται από έναν κβαντιστή (quantizer) ακολουθούμενο από έναν κωδικοποιητή Slepian-Wolf (Slepian-Wolf Encoder).

Γενικά, ένας κωδικοποιητής Wyner-Ziv αποτελείται από έναν κβαντιστή (quantizer) ακολουθούμενο από έναν κωδικοποιητή Slepian-Wolf (Slepian-Wolf Encoder), όπως φαίνεται στο σχήμα 3.5. Οι Cardinal και Van Asche [35] θεώρησαν τη Lloyd κβάντιση για την ιδανική κωδικοποίηση Slepian-Wolf χωρίς side information.

Μια ανεξάρτητη πιο γενική επέκταση του Lloyd αλγορίθμου εμφανίζεται στους D. Rebollo-Monedero, R. Zhang, και B. Girod [36].



3.2.5. Κωδικοποίηση βίντεο με χαμηλή πολυπλοκότητα

Υλοποιήσεις πρόσφατων προτύπων κωδικοποίησης βίντεο, όπως ISO MPEG, ITU-T H.263 και H.264 [37], [38], απαιτούν πολύ περισσότερους υπολογισμούς για τον κωδικοποιητή από ότι για τον αποκωδικοποιητή, τυπικά ο κωδικοποιητής είναι 5 με 10 φορές πιο πολύπλοκος από τον αποκωδικοποιητή. Αυτή η ασυμμετρία ταιριάζει στη ραδιοφωνία, όπου εκπέμπεται σήμα προς πολλούς παραλήπτες, και στα συστήματα video-on-demand όπου πολλοί χρήστες επιλέγουν το βίντεο που θέλουν να δουν. Σε όλα τα παραπάνω συστήματα το βίντεο συμπιέζεται μια φορά και αποσυμπιέζεται πολλές φορές. Ωστόσο, μερικές εφαρμογές μπορεί να απαιτήσουν το ανάποδο σύστημα δηλαδή κωδικοποιητές χαμηλής πολυπλοκότητας σε βάρος αποκωδικοποιητών υψηλής πολυπλοκότητας. Τέτοια παραδείγματα συστημάτων είναι οι κάμερες παρακολούθησης με τους ασύρματους αισθητήρες βίντεο (surveillance), ασύρματες κάμερες υπολογιστών, δικτυωμένες βιντεοκάμερες (camcorders), κάμερες στα κινητά τηλέφωνα κ.α. Σε όλες αυτές τις περιπτώσεις, η συμπίεση πρέπει να γίνει στην κάμερα όπου η μνήμη και οι υπολογιστικές δυνατότητες είναι χαμηλές.

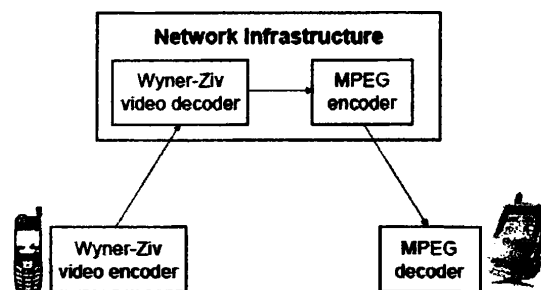
Η θεωρία Wyner-Ziv [31], [37], [38], όπως συζητήθηκε στην παράγραφο 3.3.3, προτείνει ένα σύστημα κωδικοποίησης βίντεο το οποίο κωδικοποιεί τις εικόνες (του βίντεο) ανεξάρτητα και τις αποκωδικοποιεί δεσμευμένα. Ένα τέτοιο σύστημα μπορεί να πετύχει απόδοση που προσεγγίζει κατά πολύ την κλασική inter-frame κωδικοποίηση (π.χ. MPEG). Σε αντίθεση με την βασική κωδικοποίηση βίντεο με πρόβλεψη (hybrid predictive video coding), όπου οι προηγούμενες εικόνες που υπολογίζονται με αντιστάθμιση κίνησης (motion-compensated) χρησιμοποιούνται ως side information, στο παραπάνω σύστημα οι προηγούμενες εικόνες χρησιμοποιούνται μόνο στον αποκωδικοποιητή.

Ένας τέτοιος Wyner-Ziv κωδικοποιητής βίντεο θα έχει πολύ καλή πολυπλοκότητα επειδή κωδικοποιεί κάθε εικόνα ανεξάρτητα από τις άλλες εικόνες και ο αποκωδικοποιητής αξιοποιεί τη στατιστική εξάρτηση μεταξύ των εικόνων με πολύ περισσότερη πολυπλοκότητα λόγω της inter-frame αποκωδικοποίησης. Η επιθυμητή ασυμμετρία είναι σύμφωνη με τους αλγορίθμους Slepian-Wolf και Wyner-Ziv, όπως

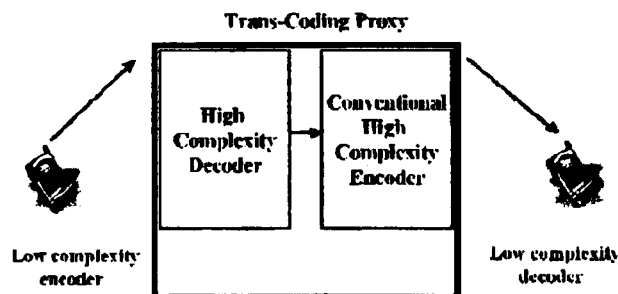


συζητήθηκαν στις παραγράφους §3.3.2, §3.3.4, οι οποίοι τείνουν να έχουν απλούς κωδικοποιητές και πολύ πιο απαιτητικούς αποκωδικοποιητές.

Στην περίπτωση που ο παραλήπτης είναι μια συσκευή με περιορισμένες υπολογιστικές δυνατότητες, δηλαδή και ο κωδικοποιητής και ο αποκωδικοποιητής είναι αναγκαίο να είναι απλοί, όπως συμβαίνει με τα βίντεο κινητών τηλεφώνων, και πάλι είναι ωφέλιμο να χρησιμοποιηθεί η κωδικοποίηση Wyner-Ziv, σχήμα 3.6, σχήμα 3.7. Ένα κινητό τηλέφωνο βιντεοσκοπεί ένα βίντεο και το συμπιέζει χρησιμοποιώντας κωδικοποίηση Wyner-Ziv και το μεταδίδει σε ένα σταθερό σημείο του διαδικτύου (network infrastructure), μετά η δυαδική ροή αποκωδικοποιείται και κωδικοποιείται εκ νέου χρησιμοποιώντας μια κλασική κωδικοποίηση όπως MPEG όπου κωδικοποιείται πολύπλοκα αλλά αποκωδικοποιείται απλά. Έτσι συνδυάζεται απλή κωδικοποίηση και πάλι απλή αποκωδικοποίηση. Στη συνέχεια θα αναφερθούμε στους αλγορίθμους κωδικοποίησης.

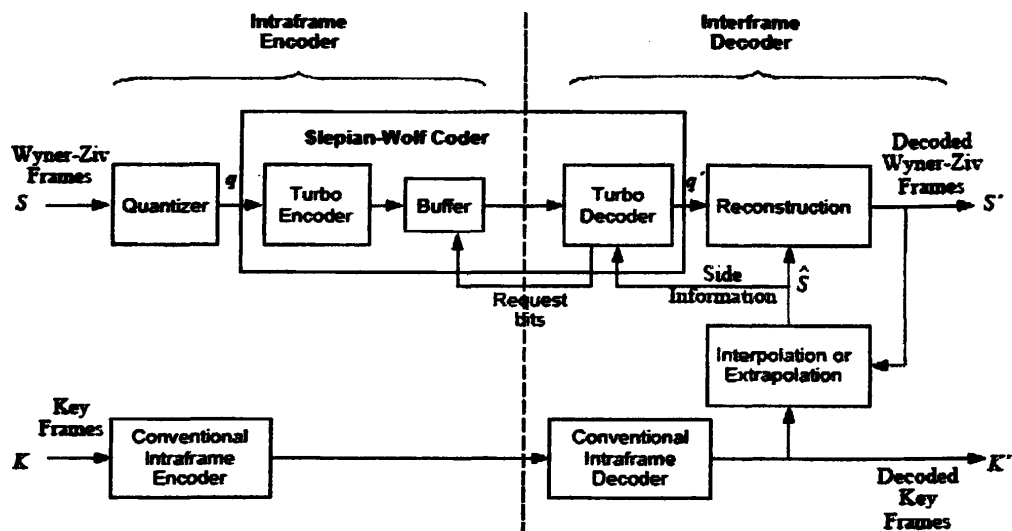


Σχήμα 3.6: Κωδικοποίηση – αποκωδικοποίηση βίντεο κινητού τηλεφώνου.



Σχήμα 3.7: Συνδυασμός απλού κωδικοποιητή – πολύπλοκου αποκωδικοποιητή και το αντίστροφο (δηλαδή πολύπλοκος κωδικοποιητής – απλός αποκωδικοποιητής).

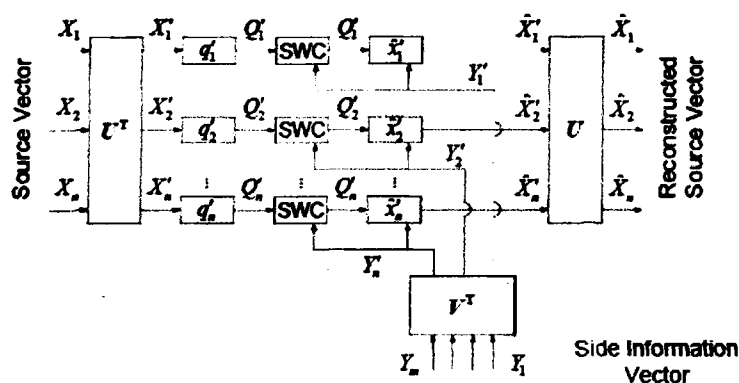
3.2.6. Κωδικοποίηση Wyner-Ziv σε επίπεδο εικονοστοιχείων (pixel domain)



Σχήμα 3.8: Wyner-Ziv κωδικοποιητής με απλό κωδικοποιητή και πολύπλοκο αποκωδικοποιητή στο χώρο εικονοστοιχείων (pixel domain).

Στο σχήμα 3.8 φαίνεται ένας κωδικοποιητής Wyner-Ziv με απλό κωδικοποιητή, όπου η κωδικοποίηση είναι intra-frame, και πολύπλοκο αποκωδικοποιητή, όπου η αποκωδικοποίηση είναι inter-frame, στο χώρο εικονοστοιχείων (pixel domain).

3.2.7. Κωδικοποίηση Wyner-Ziv σε επίπεδο μετασχηματισμού (transform-domain)



Σχήμα 3.9: κωδικοποίηση Wyner-Ziv στο χώρο μετασχηματισμού (transform-domain), SWC (Slepian-Wolf Coder).

Στο σχήμα 3.9 φαίνεται ένας κωδικοποιητής Wyner-Ziv στο χώρο μετασχηματισμού (transform-domain). Η κωδικοποίηση έχει ως εξής: πρώτα εκτελείται μετασχηματισμός σε κάθε εικόνα Wyner-Ziv, οι μετασχηματισμένοι συντελεστές κβαντίζονται και κωδικοποιούνται με έναν Slepian-Wolf turbo κωδικοποιητή [39], [40]. Όπως και στον χώρο εικονοστοιχείων, ο αποκωδικοποιητής παίρνει την προηγούμενη ανακατασκευασμένη εικόνα για να την χρησιμοποιήσει ως side information. Το side information μετασχηματίζεται και σε συνεργασία με τον turbo αποκωδικοποιητή βρίσκει την ανακατασκευασμένη εικόνα. Ο μετασχηματισμός αξιοποιεί επιπλέον και το χωρικό πλεόνασμα για αυτό επιτυγχάνει καλύτερο ζεύγος ρυθμού-παραμόρφωσης (rate-distortion) από το pixel-domain. Παρόλο που το transform-domain έχει μεγαλύτερη πολυπλοκότητα από το pixel-domain, η πολυπλοκότητα του κωδικοποιητή είναι και πάλι μικρή γιατί η εκτίμηση και αντιστάθμιση κίνησης γίνονται στον αποκωδικοποιητή.

ΚΕΦΑΛΑΙΟ 4. ΣΤΕΡΕΟΣΚΟΠΙΑ

4.1 Εισαγωγή

4.2 Κατανεμημένη κωδικοποίηση Wyner-Ziv στερεοσκοπικού βίντεο

4.1. Εισαγωγή

Στην παράγραφο αυτή παρουσιάζεται η έννοια της στερεοσκοπίας και της διαφοράς (disparity), η εκτίμηση της διαφοράς (disparity estimation) και η σχέση της με την εκτίμηση κίνησης (motion estimation). Στην παράγραφο 4.2 περιγράφεται το μοντέλο κωδικοποίησης Wyner-Ziv στερεοσκοπικού βίντεο. Επίσης, παρουσιάζεται ο αλγόριθμος EM για τη διαδικασία αποκωδικοποίησης.

Η τεχνολογία τρισδιάστατων εικόνων θεωρείται ως η επόμενη γενιά της οπτικοποίησης καθώς επιτυγχάνει πιο φυσική οπτική αναπαράσταση. Για να αναπαραχθούν αυτές οι εικόνες, δύο κάμερες τοποθετούνται σε δύο ελαφρά διαφορετικές οπτικές γωνίες και βιντεοσκοπούν δύο εικονοσειρές. Οι στερεοσκοπικές εικονοσειρές και γενικότερα οι ακολουθίες εικόνων που λαμβάνονται από πολλές οπτικές γωνίες (multiview), έχουν πολλές εφαρμογές. Για επιστημονική εξερεύνηση και οπτικοποίηση, η στερεοσκοπική αναπαράσταση δίνει την δυνατότητα να εξερευνάμε τον τρισδιάστατο χώρο σε μια δισδιάστατη οθόνη. Επίσης χρησιμοποιούνται στην τρισδιάστατη τηλεόραση, στις ακολουθίες τρισδιάστατων εικόνων, στη ρομποτική όραση, στις εικονικές μηχανές, στη χειρουργική ιατρική κτλ.



Η στερεοσκοπία προέκυψε από τον τρόπο που ο άνθρωπος αντιλαμβάνεται το βάθος και είναι μια τεχνική που δημιουργεί την ψευδαίσθηση του βάθους σε μια εικόνα. Στηρίζεται στο ότι η τρισδιάστατη φυσική όραση πραγματοποιείται διότι κάθε μάτι βλέπει το ίδιο αντικείμενο από σχετικά μικρή, αλλά διαφορετική οπτική γωνία, με αποτέλεσμα την ίδια στιγμή ο εγκέφαλος να παραλαμβάνει δύο ελαφρά διαφοροποιημένες εικόνες του ίδιου αντικειμένου. Όταν ο εγκέφαλος κάνει υπέρθεση των εικόνων σε μια τότε οι μικρές διαφορές τους στο μάτι ερμηνεύονται σαν η τρίτη διάσταση. Στη συνέχεια παρατίθεται το πώς και το γιατί δουλεύει η στερεοσκοπία: Ας τοποθετήσουμε ένα δάχτυλο μπροστά στο πρόσωπό μας, σχήμα 4.1, [41]. Όταν βλέπουμε το δάχτυλο, τα μάτια συγκλίνουν στο δάχτυλο. Καθώς βλέπουμε τα μάτια να συγκλίνουν στο δάχτυλο, αν προσέξουμε στο φόντο τότε αυτός εμφανίζεται να είναι διπλός. Από την άλλη, αν προσπαθήσουμε να επικεντρώσουμε την προσοχή μας στον φόντο, το δάχτυλο θα εμφανιστεί διπλό.

Αν μπορούσαμε να πάρουμε τις εικόνες που είναι στο αριστερό και στο δεξί μάτι και τις υπερθέσουμε, θα βλέπαμε ότι η μια επικαλύπτει την άλλη. Αν πάρουμε ένα σημείο στην αριστερή εικόνα και το αντίστοιχο στην δεξιά τότε από την υπέρθεση βλέπουμε ότι υπάρχει μια οριζόντια απόσταση αυτών των δύο σημείων η οποία λέγεται διαφορά (disparity). Τα σημεία στα οποία η διαφορά είναι μηδέν είναι αυτά στα οποία τα μάτια συγκλίνουν. Αυτή η αίσθηση βάθους που δίνει η δυνατότητα του ανθρώπου να συνδυάζει δύο διαφορετικές οπτικές γωνίες σε μια εικόνα είναι η στερεοσκοπία.



Σχήμα 4.1: Ένα πείραμα για να καταλάβουμε τη στερεοσκοπία: (α) όταν τα μάτια συγκλίνουν στον αντίχειρα, τότε ο φόντος εμφανίζεται σαν διπλή εικόνα, (β) όταν τα μάτια συγκλίνουν στον φόντο, τότε ο αντίχειρας εμφανίζεται σαν διπλή εικόνα.

Από τη διαφορά (disparity) που συνδέεται με δύο αντίστοιχα σημεία σε ένα στερεοσκοπικό ζεύγος εικόνων, μπορούμε να βρούμε το βάθος του σημείου στον τρισδιάστατο χώρο και επομένως και τις συντεταγμένες του στον πραγματικό χώρο. Η πληροφορία του βάθους μας βοηθά να καταλάβουμε τις συσχετίσεις στον τρισδιάστατο χώρο. Γι' αυτό το λόγο, ο υπολογισμός της διαφοράς (disparity estimation) είναι ένα ουσιώδες βήμα σε οποιοδήποτε σύστημα επεξεργασίας τρισδιάστατης εικόνας/βίντεο.

Το πρόβλημα εκτίμησης διαφοράς (disparity estimation) είναι παρόμοιο με το πρόβλημα εκτίμησης κίνησης (motion estimation) από την άποψη ότι και τα δύο ψάχνουν να βρουν την αντιστοιχία των εικονοστοιχείων των δύο εικόνων. Για την εκτίμηση κίνησης οι δύο εικόνες διαφέρουν στον χρόνο ενώ στην εκτίμηση διαφοράς οι δύο εικόνες λαμβάνονται την ίδια στιγμή αλλά από διαφορετικές οπτικές γωνίες. Η εκτίμηση διαφοράς τείνει να είναι πιο πολύπλοκη από την εκτίμηση κίνησης. Πρώτον, σε δύο χρονικά διαδοχικές εικόνες μόνο ένας περιορισμένος αριθμός αντικειμένων κινούνται στη σκηνή και συνεπώς τα διανύσματα κίνησης είναι μηδέν στα περισσότερα εικονοστοιχεία. Ωστόσο, στις στερεοσκοπικές εικόνες σχεδόν όλα τα εικονοστοιχεία έχουν μη μηδενική διαφορά εκτός από αυτά που είναι πολύ μακριά από το επίπεδο της εικόνας. Δεύτερον, το εύρος των διανυσμάτων κίνησης (δηλαδή η περιοχή τιμών που μπορεί να πάρει ένα διάνυσμα κίνησης) είναι συνήθως αρκετά περιορισμένο, ενώ το εύρος των διανυσμάτων διαφοράς μπορεί να είναι πολύ μεγάλο για αντικείμενα κοντά στην κάμερα. Προφανώς, αν χρειαζόταν να βρούμε την εκτίμηση διαφοράς χρησιμοποιώντας τη μέθοδο ταιριάσματος των μπλοκ τότε η περιοχή αναζήτησης θα ήταν αρκετά μεγάλη.

Η μετάδοση ή η αποθήκευση μιας ακολουθίας στερεοσκοπικών εικόνων απαιτεί το διπλάσιο εύρος καναλιού από μια ακολουθία δισδιάστατων εικόνων. Ο σκοπός ενός συστήματος συμπίεσης είναι να αναπτύξει μια αποτελεσματική μέθοδο κωδικοποίησης η οποία θα αξιοποιεί το πλεόνασμα των δύο ακολουθιών βίντεο (το βίντεο που λαμβάνεται από την αριστερή κάμερα και αυτό που λαμβάνεται από την δεξιά), το χωρικό πλεόνασμα μεταξύ των δύο ακολουθιών βίντεο (δηλαδή η εκτίμηση



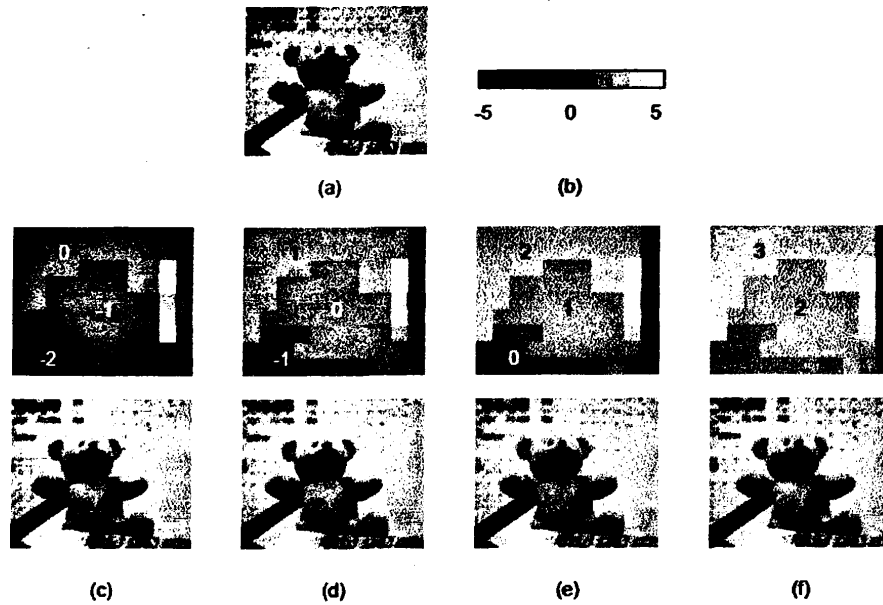
της διαφοράς (disparity estimation) των εικόνων μεταξύ της αριστερής και της δεξιάς ακολουθίας εικόνων), και το χρονικό πλεόνασμα σε κάθε μια ακολουθία (δηλαδή η εκτίμηση της κίνησης (motion estimation)). Αυτό μπορεί να γίνει με το να χρησιμοποιηθεί εκτίμηση και αντιστάθμιση κίνησης και διαφοράς (motion estimation compensation, disparity estimation compensation).

Η διαφορά των στερεοσκοπικών εικόνων μπορεί να χαρακτηριστεί ως μετατοπίσεις μπροστινών αντικειμένων σε σχέση με αντικείμενα του φόντου της εικόνας. Η δυσκολία έγκειται στην εύρεση και στην αξιοποίηση της διαφοράς Δ (disparity) στον αποκωδικοποιητή με την απαίτηση ότι ο ρυθμός (bitrate) είναι χαμηλός. Μια παρόμοια περίπτωση υπάρχει στην κωδικοποίηση Wyner-Ziv χαμηλής πολυπλοκότητας ενός βίντεο βιντεοσκοπημένο από μια μόνο κάμερα [43], [44], [46]. Αυτά τα συστήματα κωδικοποιούν τις εικόνες του βίντεο ανεξάρτητα τη μία από την άλλη και τις αποκωδικοποιούν μαζί, γι' αυτό η εύρεση της κίνησης μεταξύ των εικόνων στον αποκωδικοποιητή είναι πολύ βοηθητική. Ένας πολύπλοκος τρόπος για να εκτιμηθεί η κίνηση είναι να γίνει αποκωδικοποίηση με κάθε πιθανό ταίριασμα των μπλοκ (full motion search) [44]. Μια άλλη προσέγγιση απαιτεί ο κωδικοποιητής να αποστείλει επιπλέον πληροφορία προκειμένου ο αποκωδικοποιητής να κάνει μια καλή εκτίμηση κίνησης πριν την αποκωδικοποίηση [45]. Το πρόβλημα είναι ότι ο ρυθμός της επιπλέον πληροφορίας είναι δύσκολο να καθοριστεί διότι όταν ο κωδικοποιητής στέλνει με σταθερό ρυθμό και η κίνηση είναι μικρή τότε χάνει πληροφορία, ενώ αν η κίνηση είναι αρκετά μεγάλη τότε ο ρυθμός μπορεί να μην επαρκεί για την αποκωδικοποίηση. Λόγω της μεγάλης πολυπλοκότητας και δυσκολίας να προσδιοριστεί ο ρυθμός, δεν χρησιμοποιούνται οι παραπάνω προσεγγίσεις για τη συμπίεση στερεοσκοπικών εικόνων.

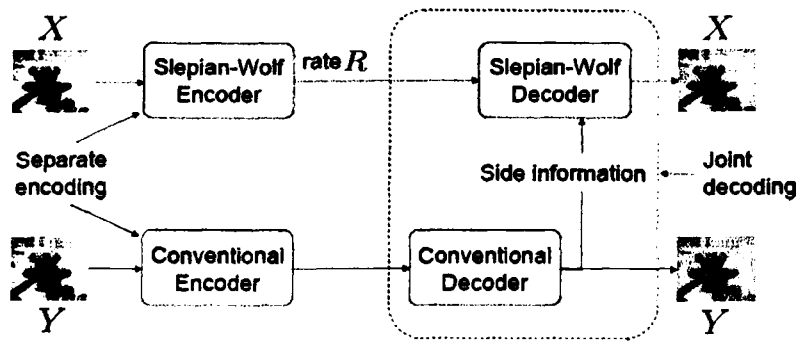
Η σχέση μεταξύ δύο στερεοσκοπικών εικόνων X και Y από άποψη της διαφοράς τους Δ απεικονίζεται στο σχήμα 4.2. Στο (α) απεικονίζεται μια εικόνα X . Ενώ στις εικόνες (c) - (f) δείχνει τέσσερις εικόνες Y που έχουν ληφθεί από τέσσερις διαφορετικές οπτικές γωνίες [47]. Για κάθε ζεύγος δηλαδή της εικόνας X με κάθε μια από τις Y , η οριζόντια μετατόπιση κάθε μπλοκ 8×8 που είναι η διαφορά Δ δείχνει



ποίο μπλοκ του Y (από αυτά που έχουν οριζόντια μετατόπιση το πολύ 5 εικονοστοιχεία) ταιριάζει καλύτερα με κάθε μπλοκ του X (από άποψη ελάχιστου τετραγωνικού σφάλματος). Αυτά τα στερεοσκοπικά ζεύγη όταν τα δούμε στερεοσκοπικά δημιουργούν μια αίσθηση βάθους ανάλογα με τη διαφορά Δ που υπάρχει σε αυτά.



Σχήμα 4.2: (a) εικόνα X , (b) φάσμα διαφορών (οριζόντιων μετατοπίσεων), (c) – (f) εικόνες Y σύμφωνα με τις οριζόντιες μετατοπίσεις της διαφοράς Δ της X σε σχέση με την Y .



Σχήμα 4.3: Κατανεμημένη συμπίεση: τα X και Y κωδικοποιούνται ανεξάρτητα (separate encoding) και αποκωδικοποιούνται μαζί (joint decoding).

Το σχήμα συμπίεσης φαίνεται στο σχήμα 4.3. Οι εικόνες X και Y κωδικοποιούνται ανεξάρτητα (separate encoding) και αποκωδικοποιούνται μαζί (joint decoding). Για λόγους απλότητας θεωρούμε ότι το Y κωδικοποιείται με κλασική μέθοδο και βρίσκεται στον αποκωδικοποιητή. Ο σκοπός είναι να κωδικοποιηθεί το X χωρίς την παρουσία του Y έτσι ώστε να μπορεί να αποκωδικοποιηθεί με την παρουσία (με την βοήθεια) του Y. Το θεώρημα Slepian-Wolf δείχνει ότι ο ρυθμός μετάδοσης του X στον αποκωδικοποιητή θα χρειαστεί $R > H(X|Y)$ δυαδικά ψηφία [2].

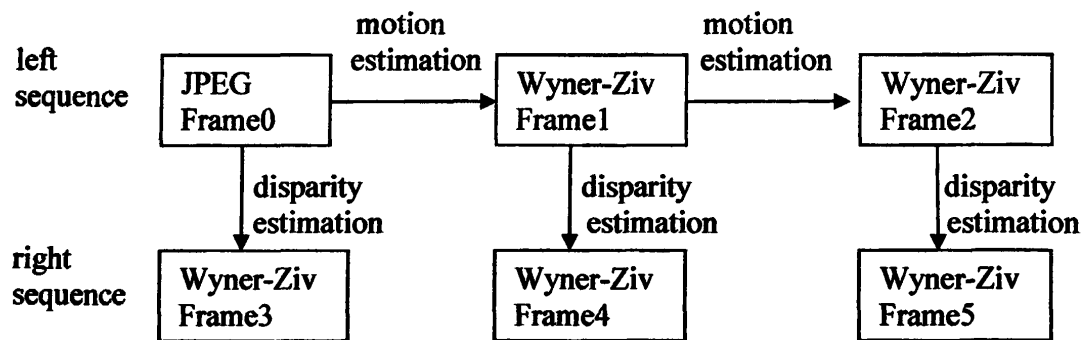
4.2. Κατανεμημένη κωδικοποίηση Wyner-Ziv στερεοσκοπικού βίντεο (distributed Wyner-Ziv coding of stereoscopic video)

4.2.1. Κατανεμημένη κωδικοποίηση στερεοσκοπικού βίντεο

Το σχήμα του μοντέλου κωδικοποίησης Wyner-Ziv στερεοσκοπικού βίντεο φαίνεται στο σχήμα 4.4. Το στερεοσκοπικό βίντεο αποτελείται από δύο ακολουθίες εικόνων, μια αριστερή ακολουθία (left sequence) και μια δεξιά (right sequence). Στην αριστερή ακολουθία, η πρώτη εικόνα κωδικοποιείται με jpeg κωδικοποίηση (intra-frame coding) ενώ όλες οι επόμενες με κωδικοποίηση Wyner-Ziv (inter-frame coding) κάνοντας εκτίμηση κίνησης M (motion estimation). Στη δεξιά ακολουθία, κάθε

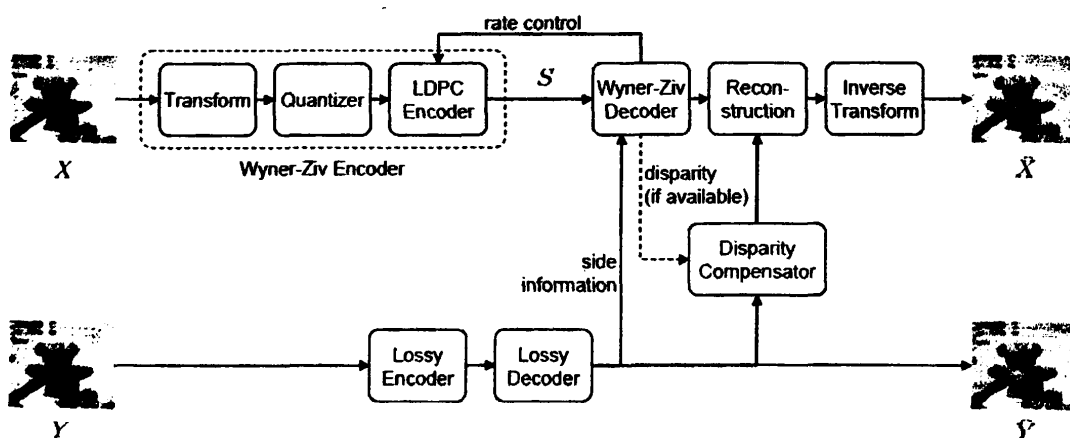


εικόνα κωδικοποιείται από την αντίστοιχη της αριστερή κάνοντας εκτίμηση διαφοράς Δ (disparity estimation). Για τη συμπίεση Wyner-Ziv του στερεοσκοπικού βίντεο χρησιμοποιείται η διαδικασία που θα περιγραφεί στις επόμενες παραγράφους. Η διαδικασία αυτή εφαρμόζεται σε κάθε ζεύγος εικόνων του στερεοσκοπικού βίντεο. Για στερεοσκοπικό βίντεο με τρεις εικόνες για την αριστερή και τρεις εικόνες για την δεξιά ακολουθία, σχήμα 4.4, τέτοια ζεύγη είναι frame0-frame1, frame1-frame2, frame0-frame3, frame1-frame4, frame2-frame5.



Σχήμα 4.4: Μοντέλο κωδικοποίησης Wyner-Ziv στερεοσκοπικού βίντεο.

4.2.2. Κωδικοποίηση Wyner-Ziv στερεοσκοπικού βίντεο



Σχήμα 4.5: Κωδικοποίηση Wyner-Ziv του X σε σχέση με το Y.

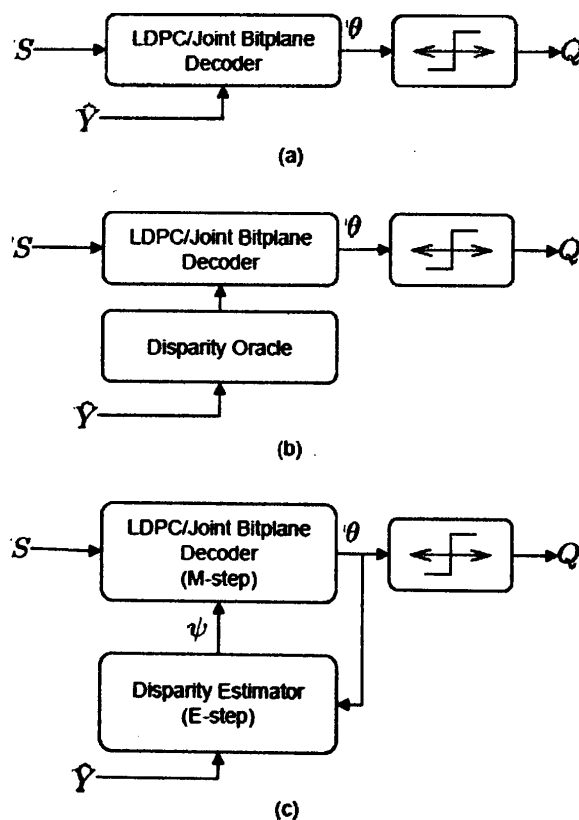


Στο σχήμα 4.5 φαίνεται ο κωδικοποιητής Wyner-Ziv στερεοσκοπικών εικόνων. Το σύστημα εφαρμόζει μετασχηματισμό, για να αξιοποιήσει το χωρικό πλεόνασμα (spatial redundancy), και κβάντιση για απωλεστική κωδικοποίηση. Στη συνέχεια περιγράφονται τα τρία σημαντικά στάδια για την κωδικοποίηση του X : Wyner-Ziv κωδικοποιητής, Wyner-Ziv αποκωδικοποιητής, και ανακατασκευή της αποκωδικοποιημένης εικόνας.

4.2.2.1. Κωδικοποιητής Wyner-Ziv

Όπως στο JPEG [48], η εικόνα X μετασχηματίζεται με DCT και κβαντίζεται. Οι κβαντισμένοι συντελεστές κωδικοποιούνται με Slepian-Wolf όπως στο [49], χρησιμοποιώντας έναν κώδικα LDPC [50] για να βρει το syndrome S .

4.2.2.2. Αποκωδικοποιητής Wyner-Ziv



Σχήμα 4.6: Αποκωδικοποίηση Wyner-Ziv για την εύρεση των κβαντισμένων συντελεστών Q (α) χωρίς εκτίμηση διαφοράς (no disparity compensation), (β) με ιδανική διαφορά (disparity oracle), (γ) εκτίμηση της διαφοράς Δ μέσω EM.

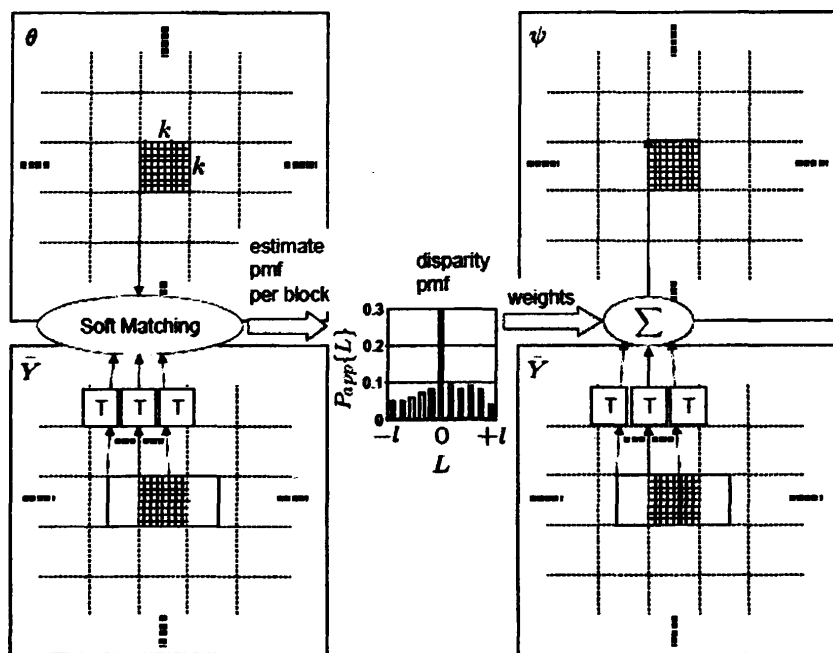
Ο σκοπός του αποκωδικοποιητή Wyner-Ziv είναι να ανακτήσει τους κβαντισμένους συντελεστές Q από το syndrome S και από το side information Y . Το σχήμα 4.6 δείχνει τρεις αποκωδικοποιητές Wyner-Ziv, που διαφέρουν στο πώς χειρίζονται τη διαφορά Δ . Ο βασικός αποκωδικοποιητής στο 4.6(α) δεν κάνει εκτίμηση της διαφοράς (no disparity compensation). Αρχικά υπολογίζει το Q με βάση τους μετασχηματισμένους συντελεστές του side information Y . Οι υπολογισμοί αυτοί αναβαθμίζονται με βάση έναν επαναληπτικό αλγόριθμο μετάδοσης πεποιθήσεων χρησιμοποιώντας το syndrome. Ο αποκωδικοποιητής είναι ένας συνδυασμός ενός LDPC αποκωδικοποιητή και ενός αποκωδικοποιητή joint bitplane [48] που αξιοποιεί την εξάρτηση όλων των δυαδικών ψηφίων του Q . Επειδή υπάρχει διαφορά (disparity) μεταξύ των X και Y , αυτό το σχήμα δεν είναι αποτελεσματικό διότι οι υπολογισμοί του Q από το Y δεν είναι σωστοί όπου υπάρχει κίνηση.

Στο σχήμα 4.6(b) δείχνει ένα μη πρακτικό σχήμα στο οποίο ο αποκωδικοποιητής περιέχει μια ιδανική εκτίμηση της διαφοράς (disparity oracle). Αυτή η μέθοδος κάνει εκτίμηση του X από το Y σύμφωνα με τα ιδανικά διανύσματα διαφοράς και έτσι συγκριτικά με το προηγούμενο σχήμα αποδίδει καλύτερα σε περιοχές της εικόνας όπου υπάρχει διαφορά μεταξύ των X και Y .

Τέλος, στο σχήμα 4.6(c) απεικονίζεται ένας πρακτικός αποκωδικοποιητής που κάνει εκτίμηση της διαφοράς μέσω του αλγορίθμου EM (expectation maximization-μεγιστοποίησης πιθανοφάνειας) δηλαδή ο αλγόριθμος EM δεν τερματίζει όταν βρεθεί το X αλλά όταν ικανοποιηθεί το syndrome. Στη θέση της ιδανικής εκτίμησης διαφοράς, η εκτίμησή της (της διαφοράς) γίνεται διατηρώντας μια εκ των υστέρων κατανομή πιθανοτήτων για τη διαφορά. Κάθε επανάληψη της αποκωδικοποίησης LDPC/joint bitplane στέλνει στον εκτιμητή διαφοράς μια πιθανοτική εκτίμηση του Q (σημειώνεται με θ) προκειμένου να αναβαθμίσει την κατανομή της διαφοράς Δ . Στη



συνέχεια ο εκτιμητής διαφοράς αναβαθμίζει το side information του Y για τον αποκωδικοποιητή LDPC/joint bitplane με το να φέρνει πληροφορία από τα μπλοκ του Y σύμφωνα με την κατανομή της διαφοράς Δ . Το σχήμα 4.7 δείχνει τον υπολογισμό της διαφοράς και του side information Y . Όπως φαίνεται στα αριστερά, κάθε μπλοκ του θ (που είναι η πιθανοτική εκτίμηση του Q) συνεργάζεται με τα υποψήφια μετασχηματισμένα μπλοκ του Y για να παράξει μια εκ των υστέρων κατανομή πιθανότητας της διαφοράς Δ για εκείνο το μπλοκ. Στα δεξιά του σχήματος, τα ίδια υποψήφια μετασχηματισμένα μπλοκ του Y αθροίζονται μαζί σύμφωνα με την κατανομή για να βρεθεί ένα πιο ακριβές side information Y . Αυτά τα βήματα επαναλαμβάνονται με αποκωδικοποίηση LDPC/joint bitplane για να ανακτηθούν οι κβαντισμένοι συντελεστές Q και η διαφορά (δηλαδή τα διανύσματα διαφοράς) Δ . Το Q αφού γίνει αντίστροφος μετασχηματισμός μας δίνει το X .



Σχήμα 4.7 αριστερά: υπολογισμός της διαφοράς Δ (disparity), δεξιά: side information Y .

Μια μονόχρωμη εικόνα αναπαριστάται χρησιμοποιώντας μόνο έναν αριθμό για κάθε εικονοστοιχείο ο οποίος δείχνει την φωτεινότητά του. Αν ένα εικονοστοιχείο



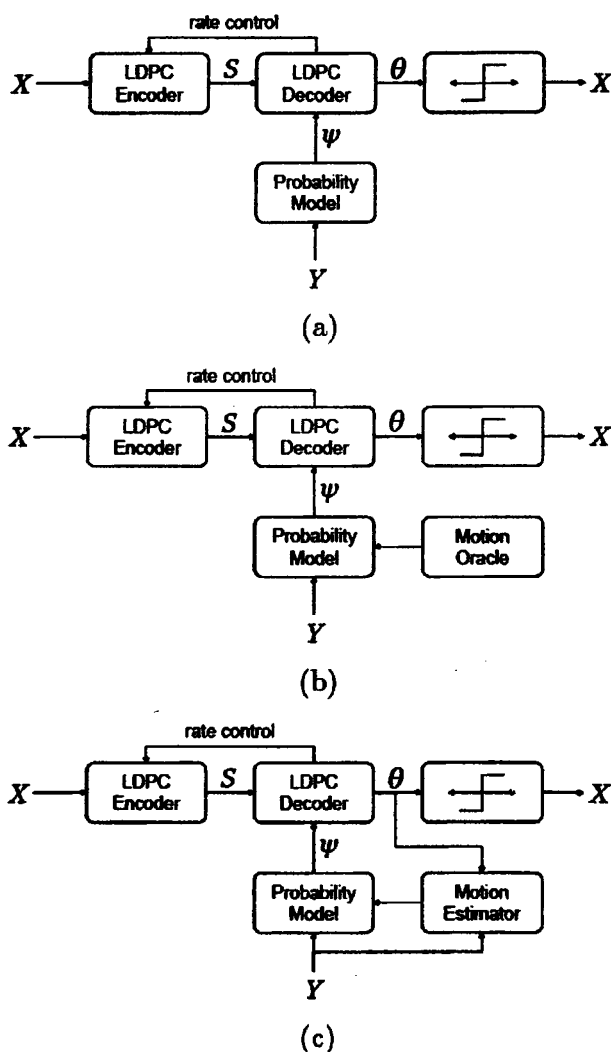
αναπαριστάται χρησιμοποιώντας n δυαδικά ψηφία (bits), τότε ο αριθμός 0 είναι μαύρο, ο $2^n - 1$ είναι άσπρο και οι ενδιάμεσες τιμές είναι αποχρώσεις του γκρι. Η φωτεινότητα συνήθως αναπαριστάται με 8 δυαδικά ψηφία.

Έστω ότι είμαστε σε επίπεδο εικονοστοιχείων, δηλαδή η εικόνα δεν μετασχηματίζεται αλλά κατευθείαν κβαντίζεται. Ας θεωρήσουμε τα συστατικά φωτεινότητας (luminance) δύο διαδοχικών εικόνων (δύο εικόνες από την αριστερή ακολουθία – left sequence, σχήμα 4.4) και στερεοσκοπικών εικόνων βίντεο (μια αριστερή και μια δεξιά εικόνα), που κβαντίζονται με d δυαδικά ψηφία, όπου $d > 1$. Για το πρώτο ζεύγος εικόνων X , Y , έστω ότι το X είναι το κβαντισμένο συστατικό φωτεινότητας της εικόνας που πρόκειται να κωδικοποιηθεί και Y το προηγούμενο κβαντισμένο συστατικό φωτεινότητας της εικόνας που είναι διαθέσιμη στον αποκωδικοποιητή, αντίστοιχα για το στερεοσκοπικό ζεύγος έχουμε το X' και Y' . Συμβολίζουμε με M το πεδίο διανυσμάτων κίνησης (motion vector field) που σχετίζουν το X με το Y , και με Δ τη διαφορά (disparity vector field) που σχετίζει το X' με το Y' . Ο σκοπός είναι να κωδικοποιηθεί το X (αντίστοιχα το X') αποτελεσματικά χωρίς την παρουσία του Y (αντίστοιχα του Y') έτσι ώστε να μπορεί να αποκωδικοποιηθεί αξιόπιστα με την παρουσία του Y (αντίστοιχα του Y'), δηλαδή απαιτείται κατανεμημένη κωδικοποίηση. Θεωρούμε ότι η σχέση του X με το Y (αντίστοιχα του X' με το Y') είναι ένα εικονικό κανάλι, όπου το Y θεωρείται σαν μια θορυβώδης έκδοση του X . Ο αποκωδικοποιητής καλείται να διορθώσει τον θόρυβο εικονικού καναλιού δηλαδή να υπολογίσει το X με δεδομένα το syndrome, δηλαδή τα δυαδικά ψηφία ισοτιμίας (parity bits), και το Y .

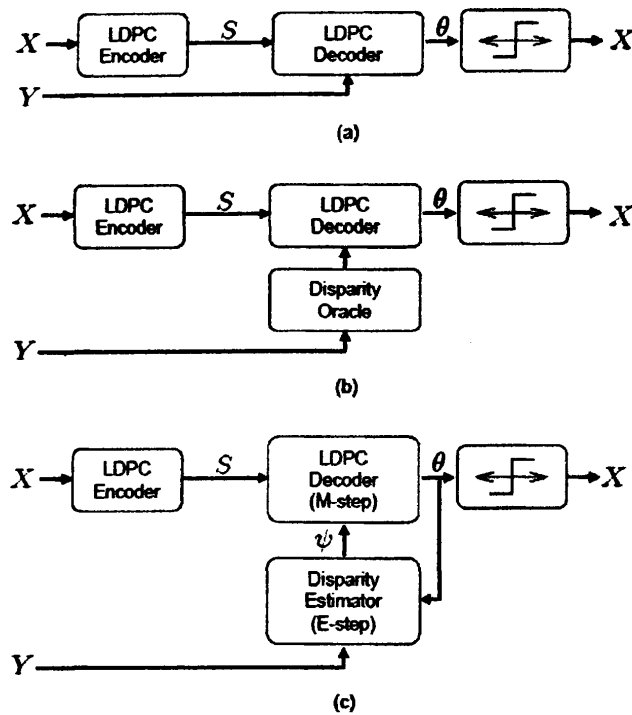
Το σχήμα 4.8 (αντίστοιχα σχήμα 4.9) απεικονίζει τρία συστήματα συμπίεσης που μπορούν να εφαρμοστούν σε αυτό το πρόβλημα. Ο κωδικοποιητής και των τριών συστημάτων είναι παρόμοιος διότι στην κατανεμημένη κωδικοποίηση ο κωδικοποιητής είναι απλός, δηλαδή δεν κάνει εκτίμηση κίνησης και διαφοράς. Ο κωδικοποιητής υπολογίζει το syndrome S του X σύμφωνα με τον κώδικα low-density parity-check (LDPC). Ο κώδικας LDPC είναι κώδικας διόρθωσης σφαλμάτων δηλαδή μέθοδος κωδικοποίησης και μετάδοσης μηνυμάτων σε ένα θορυβώδες κανάλι [51]. Η υλοποίηση της μεθόδου χρησιμοποιεί έναν κώδικα rate-adaptive LDPC



accumulate [52], [53] για να ελέγξει το ρυθμό μετάδοσης δυαδικών ψηφίων (bitrate), δηλαδή το πλήθος δυαδικών ψηφίων που περνάει από ένα σημείο του καναλιού μετάδοσης σε ένα χρονικό διάστημα. Για να πετύχει τον έλεγχο χρησιμοποιεί ένα κανάλι ανάδρασης (rate control). Αυτό επιτρέπει στον κωδικοποιητή να στείλει επιπλέον δυαδικά ψηφία ισοτιμίας (syndrome), μετά από αίτηση του αποκωδικοποιητή, σε περίπτωση που η ανακατασκευή του X δεν είναι σύμφωνη με το S . Τα τρία συστήματα διαφέρουν στην αποκωδικοποίηση, διότι εκεί γίνεται εκτίμηση κίνησης και διαφοράς.



Σχήμα 4.8: Κωδικοποίηση του X σε σχέση με το Y με α) μηδενική κίνηση (zero motion), β) ιδανική εκτίμηση κίνησης (motion oracle), γ) εκτίμηση των διανυσμάτων κίνησης (motion estimation).



Σχήμα 4.9: Κωδικοποίηση του X σε σχέση με το Y με α) μηδενική διαφορά (zero disparity), β) ιδανική εκτίμηση διαφοράς (disparity oracle), γ) εκτίμηση των διανυσμάτων διαφοράς (disparity estimation).

Το βασικό σύστημα στο σχήμα 4.8(α) (αντίστοιχα σχήμα 4.9(α)) εκτελεί την αποκωδικοποίηση του X σε σχέση με τα αντίστοιχα εικονοστοιχεία του Y (side information) (αντίστοιχα του X' σε σχέση με του Y'), δηλαδή κάθε μπλοκ του X χρησιμοποιεί το αντίστοιχο του Y με διάνυσμα κίνησης (αντίστοιχα διάνυσμα διαφοράς) μηδέν. Με άλλα λόγια γίνεται πρόβλεψη του X χωρίς εκτίμηση κίνησης (no motion compensation) (αντίστοιχα διαφοράς) χρησιμοποιώντας το Y και το S . Αρχικά, ο LDPC αποκωδικοποιητής κάνει έναν στατιστικό-πιθανοτικό υπολογισμό (soft estimate) του X χρησιμοποιώντας στατιστική-πιθανοτική πληροφορία του Y (soft side information) η οποία πηγάζει από τα εικονοστοιχεία του Y εφαρμόζοντας ένα μοντέλο πιθανοτήτων. Μετά ξαναυπολογίζει τα παραπάνω χρησιμοποιώντας το S μέσω ενός επαναληπτικού αλγορίθμου διάδοσης πεποιθήσεων. Σε περιοχές που υπάρχει κίνηση (αντίστοιχα διαφορά) μεταξύ X και Y , αυτό το σχήμα έχει μικρή απόδοση.

Για να συγκρίνουμε τα διανύσματα κίνησης και διαφοράς που θα βρεθούν κατά την εκτίμηση κίνησης και διαφοράς του αλγορίθμου υπολογίζονται θεωρητικά, μη πρακτικά διανύσματα κίνησης και διαφοράς τα οποία προκύπτουν μετά από εκτίμηση κίνησης των X και Y (motion oracle), καθώς και εκτίμηση διαφοράς των X' , Y' (disparity estimation). Το σχήμα 4.8(β) (αντίστοιχα σχήμα 4.9(β)) δείχνει έναν μη πρακτικό αποκωδικοποιητή με ενσωματωμένο το motion oracle (αντίστοιχα disparity oracle). Το oracle πληροφορεί το μοντέλο πιθανοτήτων για το ποιά εικονοστοιχεία του Y θα έπρεπε να χρησιμοποιηθούν ως side information κατά την LDPC αποκωδικοποίηση.

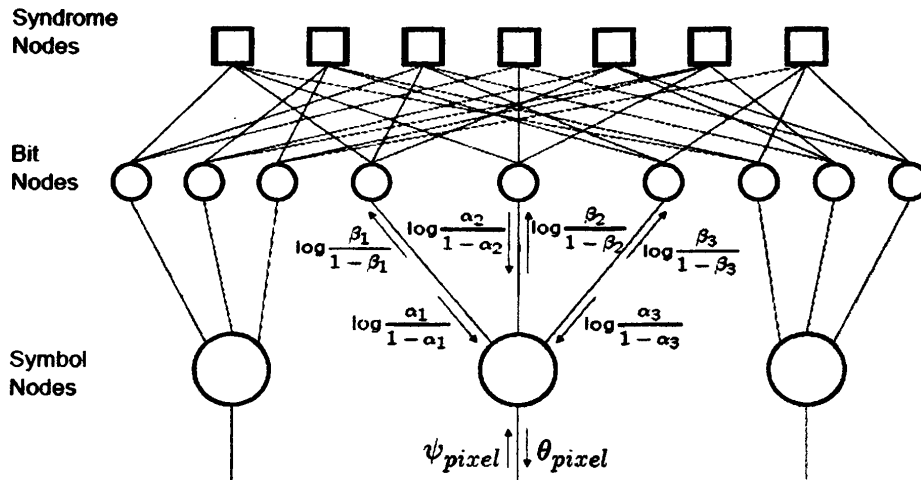
Τέλος, το σχήμα 4.8(γ) (αντίστοιχα σχήμα 4.9(γ)) απεικονίζει έναν πρακτικό αποκωδικοποιητή που εκτιμάει το πεδίο διανυσμάτων κίνησης M (αντίστοιχα πεδίο διανυσμάτων διαφοράς Δ) μέσω του αλγορίθμου μεγιστοποίησης πιθανότητας EM (Expectation Maximization). Στη θέση του motion oracle (αντίστοιχα disparity oracle), μια μέθοδος υπολογισμού κίνησης (motion estimator) (αντίστοιχα διαφοράς) διατηρεί μια εκ των υστέρων κατανομή πιθανοτήτων του M (αντίστοιχα Δ), συγκρίνοντας το Y και τον στατιστικό-πιθανοτικό υπολογισμό (soft estimate) του X από τον LDPC αποκωδικοποιητή. Κάθε επανάληψη του LDPC αποκωδικοποιητή στέλνει στη μέθοδο υπολογισμού κίνησης έναν πρόσφατο στατιστικό υπολογισμό του X προκειμένου να ανανεώσει την κατανομή στο M . Σε ανταπόδοση, το μοντέλο πιθανοτήτων ανανεώνει το side information φέρνοντας πληροφορία από τα εικονοστοιχεία του Y σύμφωνα με την ανανεωμένη κατανομή πιθανοτήτων της κίνησης M . Παρακάτω στην παράγραφο 4.3 θα περιγραφεί ο αλγόριθμος EM.

4.2.3. Joint bitplane LDPC decoding

Η διαδικασία που περιγράφεται σε αυτή την παράγραφο είναι ίδια και για την εκτίμηση κίνησης και για την εκτίμηση διαφοράς. Για τη μέθοδο υπολογισμού κίνησης και διαφοράς (motion estimator and disparity estimator) στο σχήμα 4.8(γ) και στο σχήμα 4.9(γ), προκειμένου να ανανεωθεί/τροποποιηθεί η κατανομή πιθανοτήτων του πεδίου διανυσμάτων κίνησης M και διαφοράς Δ με ακρίβεια, θα έπρεπε να



συγκρίνει το Y με την πιθανοτική εκτίμηση (soft estimate) του X ταυτόχρονα σε όλα τα δυαδικά ψηφία (joint bitplane). Συνεπώς, ο LDPC αποκωδικοποιητής πρέπει να αποκωδικοποιήσει όλα τα δυαδικά ψηφία (bitplanes) του X ταυτόχρονα, σε αντίθεση με τον LDPC αποκωδικοποιητή του [51].



Σχήμα 4.10: Γράφημα διάδοσης πεποιθήσεων του αποκωδικοποιητή LDPC (joint bitplane LDPC decoding).

Η αποκωδικοποίηση LDPC όλων των δυαδικών ψηφίων ταυτόχρονα (joint bitplane LDPC decoding) μοντελοποιεί και αξιοποιεί όλα τα d δυαδικά ψηφία που αναπαριστούν ένα εικονοστοιχείο, χρησιμοποιώντας γράφημα αποκωδικοποίησης για τη διάδοση των πεποιθήσεων, σχήμα 4.10. Όπως ο αποκωδικοποιητής LDPC του [51], αποκωδικοποιεί το X στους κόμβους δυαδικών ψηφίων (bit nodes) σύμφωνα με τους περιορισμούς των syndrome κόμβων (syndrome nodes), διαδίδοντας πεποιθήσεις (log likelihood ratios είναι ο λογάριθμος της πιθανοφάνειας ενός συγκεκριμένου δυαδικού ψηφίου να έχει τιμή 1 δια την πιθανοφάνειά του να έχει τιμή 0) κατά μήκος των ακμών του γραφήματος. Η διαφορά είναι ότι το side information δεν παρέχει log likelihood ratios στα bit nodes απευθείας αλλά αντί αυτού τρέφει νέους κόμβους, τους κόμβους symbol. Όπως φαίνεται στο σχήμα 4.10, κάθε symbol κόμβος (ένας για κάθε εικονοστοιχείο) συλλέγει ψ_{pixel} μαζί με log likelihood ratios

$\log \frac{\alpha_g}{1-\alpha_g}$ από κάθε κόμβο δυαδικού ψηφίου $g \in \{1, \dots, d\}$ που συνδέεται με αυτό το ευκονοστοιχείο. Μετά από τοπικό υπολογισμό, ο κόμβος symbol στέλνει \log likelihood ratios $\log \frac{\beta_g}{1-\beta_g}$ σε κάθε έναν από αυτούς τους κόμβους δυαδικού ψηφίου.

Στη περίπτωση του σχήματος 4.10, το $d = 3$ έτσι η κατανομή τιμών του side information είναι 8 τιμών. Όλοι οι δυνατοί συνδυασμοί των τριών δυαδικών ψηφίων είναι (000, 001, 010, 011, 100, 101, 110, 111). Χρησιμοποιούμε Gray απεικόνιση για λόγους που περιγράφονται στο [42]:

$$\psi_{pixel} = (p_{000}, p_{001}, p_{010}, p_{011}, p_{100}, p_{101}, p_{110}, p_{111}).$$

Τα \log likelihood ratios που στέλνονται στους κόμβους δυαδικών ψηφίων (δυαδικοί κόμβοι) υπολογίζονται μέσω του αλγορίθμου sum-product [54]. Για παράδειγμα, για τον κόμβο δυαδικού ψηφίου 2,

$$\log \frac{\beta_2}{1-\beta_2} = \log \frac{\sum_{c=0,1} p_{a1c} a_1^a (1-a_1)^{1-a} a_3^c (1-a_3)^{1-c}}{\sum_{c=0,1} p_{a0c} a_1^a (1-a_1)^{1-a} a_3^c (1-a_3)^{1-c}}$$

Αυτός ο τύπος ξεκινάει πολλαπλασιάζοντας τα στοιχεία p_{abc} του ψ_{pixel} με τα στοιχεία της δυαδικής κατανομής $(\alpha_1, 1-\alpha_1)$ και $(\alpha_3, 1-\alpha_3)$ σύμφωνα με τις τιμές του πρώτου δυαδικού ψηφίου a και του τρίτου δυαδικού ψηφίου c αντίστοιχα σύμφωνα με τους δείκτες τους. Αυτά τα γινόμενα προστίθενται σε δύο ομάδες σύμφωνα με την τιμή του δεύτερου δυαδικού ψηφίου b . Οι δύο ομάδες διαιρούνται μεταξύ τους και παίρνουμε τον λογάριθμο του αποτελέσματος της διαίρεσης. Για να υπολογίσουμε το $\log \frac{\beta_2}{1-\beta_2}$ αποφεύγουμε να χρησιμοποιούμε το $(\alpha_2, 1-\alpha_2)$ προκειμένου να αποφεύγουμε την ανακύκλωση στον κόμβο δυαδικού ψηφίου 2. Με τον ίδιο τρόπο υπολογίζουμε και τα $\log \frac{\beta_1}{1-\beta_1}$, $\log \frac{\beta_3}{1-\beta_3}$ αλλάζοντας στον



παραπάνω τύπο τους δείκτες του πρώτου, δεύτερου και τρίτου δυαδικού ψηφίου. Οι δυαδικοί κόμβοι προωθούν τα log likelihood ratios $\log \frac{\beta_g}{1-\beta_g}$ στους syndrome κόμβους οι οποίοι εφαρμόζουν τους κανόνες αποκωδικοποίησης LDPC syndrome [51] για να απαντήσουν στους δυαδικούς κόμβους. Οι τελευταίοι τροποποιούν τα log likelihood ratios $\log \frac{a_g}{1-a_g}$ που στέλνονται στους κόμβους symbol.

Τέλος, οι κόμβοι symbol υπολογίζουν την πιθανοτική εκτίμηση (soft estimate) θ_{pixel} πολλαπλασιάζοντας τα στοιχεία του ψ_{pixel} με τα στοιχεία και των τριών ανανεωμένων δυαδικών κατανομών $(\alpha_1, 1-\alpha_1)$, $(\alpha_2, 1-\alpha_2)$ και $(\alpha_3, 1-\alpha_3)$ σύμφωνα με τις τιμές του πρώτου, δεύτερου και τρίτου δυαδικού ψηφίου. Για παράδειγμα για το πρώτο δυαδικό ψηφίο, αν p είναι η πιθανότητα του δυαδικού ψηφίου να είναι 1, τότε $1-p$ είναι η πιθανότητα να είναι 0. Αν ο συνδυασμός των δυαδικών ψηφίων είναι 010 τότε το γινόμενο θα είναι ψ_{pixel} επί $(1-p)^*p^*(1-p)$.

Μετά από Gray απεικόνιση και κανονικοποίηση έχουμε

$$\theta'_{pixel} = (q_{000}, q_{001}, q_{010}, q_{011}, q_{100}, q_{101}, q_{110}, q_{111})$$

Όπου

$$q_{abc} \propto p_{abc} \alpha_1^a (1-\alpha_1)^{1-a} \alpha_2^b (1-\alpha_2)^{1-b} \alpha_3^c (1-\alpha_3)^{1-c}$$

4.3. Αλγόριθμος Μεγιστοποίησης Πιθανότητας

4.3.1. Μοντέλο

Ας θεωρήσουμε τις φωτεινότητες δύο διαδοχικών εικόνων ενός βίντεο X και Y , με την X εικόνα να σχετίζεται με την Y μέσω ενός πεδίου κίνησης M . Κάνοντας



αντιστάθμιση κίνησης του Y για να βρεθεί το X αυτό που προκύπτει αν το αφαιρέσουμε από το X μας δίνει τη διαφορά (residual) την οποία τη χειριζόμαστε ως ανεξάρτητο θόρυβο Laplace Z . Μοντελοποιούμε την εκ των υστέρων κατανομή πιθανότητας της πηγής X που στηρίζεται στην παράμετρο θ ως

$$P_{app}\{X\} \equiv P\{X; \theta\} = \prod_{i,j} \theta(i,j, X(i,j))$$

όπου

$$\theta(i,j, w) = P_{app}\{X(i,j) = w\}$$

ορίζει έναν στατιστικό υπολογισμό (μαλακό soft estimate) του $X(i,j)$ επάνω στις τιμές φωτεινότητας $w \in \{0, \dots, 2^d - 1\}$.

4.3.2. Πρόβλημα

Για το κανονικό ζεύγος εικόνων: ο αποκωδικοποιητής αποσκοπεί στο να υπολογίσει την εκ των υστέρων κατανομή πιθανότητας της κίνησης M , με το δεύτερο βήμα του κανόνα Bayes:

$$P_{app}\{M\} \equiv P\{M|Y, S; \theta\} \propto P\{M\}P\{Y, S|M, \theta\}$$

Το ίδιο ισχύει και για το στερεοσκοπικό ζεύγος εικόνων με τη διαφορά αντί για M να έχουμε D , δηλαδή μόνο την οριζόντια μετατόπιση.

$$P_{app}\{D\} \equiv P\{D|Y, S; \theta\} \propto P\{D\}P\{Y, S|D; \theta\}$$

Πρέπει $D(i,j) \leq n$, όπου n είναι η μέγιστη δυνατή μετατόπιση. Απαιτείται περιορισμός του n για να δουλέψει ο κώδικας και για να επιτευχθεί αυτό, πρέπει οι κάμερες (αριστερή και δεξιά) που βιντεοσκοπούν το ζεύγος να είναι πολύ κοντά μεταξύ τους.



Η μορφή της παραπάνω έκφρασης προτείνει μια επαναληπτική διαδικασία χρησιμοποιώντας τον αλγόριθμο μεγιστοποίησης πιθανότητας EM (Expectation Maximization). Το βήμα E ανανεώνει/τροποποιεί την κατανομή του πεδίου κίνησης σε σχέση με τις παραμέτρους του μοντέλου της πηγής, ενώ το βήμα M ανανεώνει/τροποποιεί τις παραμέτρους του μοντέλου της πηγής σε σχέση με την κατανομή του πεδίου κίνησης. Ας σημειώσουμε ότι η έκφραση $P\{M|Y, S; \theta\}$ είναι η πιθανότητα να παρατηρήσουμε την κίνηση M με δεδομένο ότι σχετίζεται το X (έτσι όπως παραμετροποιείται από το θ) με το Y, και επίσης το S. Στη συνέχεια μελετάται το $P\{Y, S|M; \theta\}$.

4.3.3. Βήμα E του αλγορίθμου EM

Το βήμα E ανανεώνει την κατανομή κίνησης που έχει υπολογιστεί και που πριν την κανονικοποίηση γράφεται ως

$$P_{app}^{(t)}\{M\} := P_{app}^{(t-1)}\{M\} P\{Y, S|M; \theta^{(t-1)}\}$$

Αντίστοιχα για στερεοσκοπικό ζεύγος:

$$P_{app}^{(t+1)}\{D\} := P_{app}^{(t)}\{D\} P\{Y, S|D; \theta^{(t+1)}\}$$

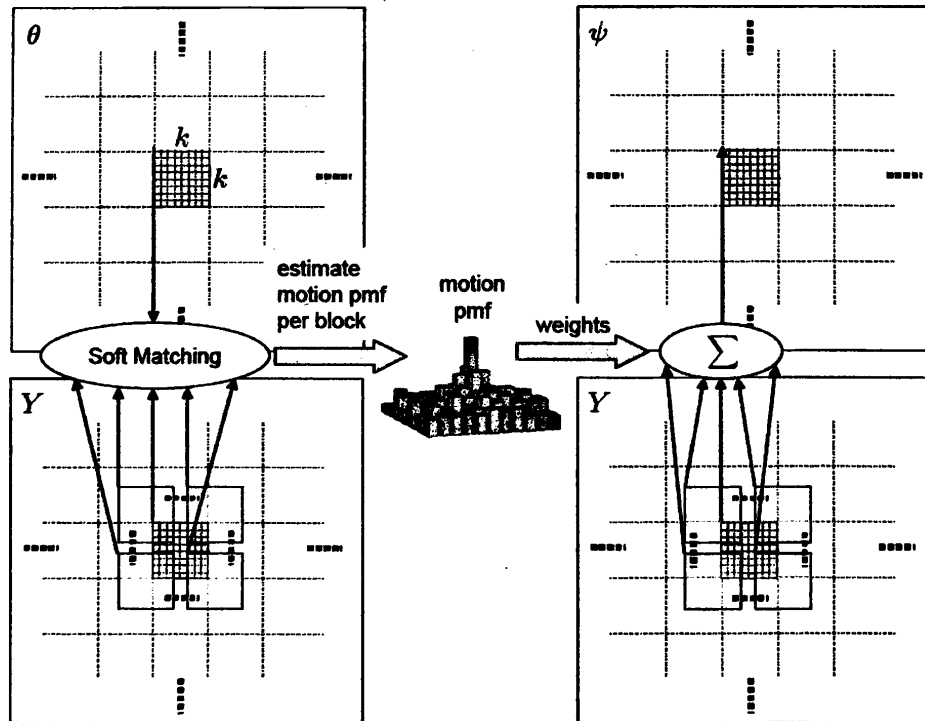
Αλλά αυτή η πράξη έχει κόστος λόγω του μεγάλου αριθμού των πιθανών τιμών της κίνησης. Έτσι απλοποιούμε με δύο τρόπους. Πρώτον, αγνοούμε το syndrome S επειδή αξιοποιείται στο βήμα M (αποκωδικοποίηση LDPC). Δεύτερον, επιτρέπουμε τον υπολογισμό του πεδίου κίνησης M χρησιμοποιώντας διανύσματα κίνησης $M_{u,v}$ για κάθε μπλοκ. Έστω ότι οι γραμμές ενός μπλοκ είναι κ, τότε κάθε μπλοκ του $\theta^{(t-1)}$ συγκρίνεται με το αντίστοιχο μπλοκ του Y καθώς και με αυτά που βρίσκονται σε μια γειτονική περιοχή του που είναι μια καθορισμένη περιοχή αναζήτησης γύρω από αυτό. Για ένα μπλοκ $\theta_{u,v}^{(t-1)}$ που έχει άνω αριστερό εικονοστοιχείο στη θέση (u, v), η



κατανομή αυτού του μπλοκ με μετατόπιση $M_{u,v}$ ανανεώνεται και κανονικοποιείται ως ακολούθως:

$$P_{app}^{(\tau)}\{M_{u,v}\} := P_{app}^{(\tau-1)}\{M_{u,v}\}P\{Y_{(u,v)+M_{u,v}} | M_{u,v}; \theta_{u,v}^{(\tau-1)}\}$$

όπου $Y_{(u,v)+M_{u,v}}$ είναι το μπλοκ του Y διαστάσεων k επί k που έχει άνω αριστερό εικονοστοιχείο στη θέση $((u, v) + M_{u,v})$. Σημειώνουμε με $P\{Y_{(u,v)+M_{u,v}} | M_{u,v}; \theta_{u,v}^{(\tau-1)}\}$ την πιθανότητα το μπλοκ του Y , που είναι στην αντίστοιχη θέση με αυτό του $\theta_{u,v}^{(\tau-1)}$, μετατοπισμένο κατά $M_{u,v}$ και συνδυασμένο με το $\theta_{u,v}^{(\tau-1)}$. Αυτή η διαδικασία γίνεται στην μέθοδο που υπολογίζει την κίνηση δηλαδή την κατανομή πιθανοτήτων των μετατοπίσεων των μπλοκ, αυτό απεικονίζεται στο σχήμα 4.11.

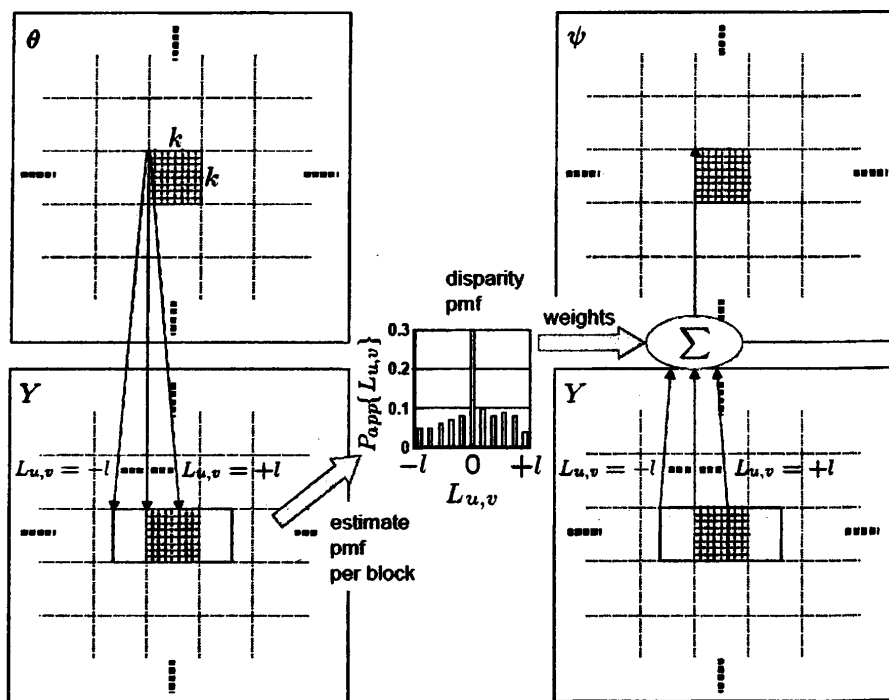


Σχήμα 4.11: Αριστερά: Ε βήμα, υπολογισμός κατανομής πιθανοτήτων της κίνησης, δεξιά: Μ βήμα, υπολογισμός του soft side information ψ .

Για το στερεοσκοπικό ζεύγος: Οι οριζόντιες μετατοπίσεις του πεδίου διαφοράς Δ είναι $L_{u,v}$ και παίρνει τιμές στο διάστημα $-n$ και n . Για ένα μπλοκ $\theta_{u,v}^{(\tau-1)}$ που έχει άνω αριστερό εικονοστοιχείο στη θέση (u, v) , η κατανομή αυτού του μπλοκ με οριζόντια μετατόπιση $L_{u,v}$ ανανεώνεται και κανονικοποιείται ως ακολούθως:

$$P_{app}^{(\tau+1)}\{L_{u,v}\} := P_{app}^{(\tau)}\{L_{u,v}\}P\{Y_{u,v+L_{u,v}} | L_{u,v}; \theta_{u,v}^{(\tau+1)}\}$$

και η συνέχεια είναι όπως για το κανονικό ζεύγος εικόνων και φαίνεται στο σχήμα 4.12.



Σχήμα 4.12: Αριστερά: Ε βήμα, υπολογισμός κατανομής πιθανοτήτων της διαφοράς Δ , δεξιά: Μ βήμα, υπολογισμός του soft side information ψ .

Η διαφορά του κανονικού ζεύγους από το στερεοσκοπικό είναι ότι στην πρώτη η εκτίμηση κίνησης γίνεται σε περιοχή αναζήτησης με οριζόντια και κάθετη μετατόπιση M , ενώ στη δεύτερη γίνεται μόνο με οριζόντια μετατόπιση Δ .

4.3.4. Βήμα M του αλγορίθμου EM

Το βήμα M του αλγορίθμου EM ανανεώνει τον στατιστικό υπολογισμό (soft estimate) μεγιστοποιώντας την πιθανοφάνεια του Y και το syndrome S .

$$\theta^{(t)} := \arg \max_{\theta} P\{Y, S; \theta\} = \arg \max_{\theta} \sum_m P_{app}^{(t)}\{M = m\} P\{Y, S | M = m; \theta\}$$

Αντίστοιχα για το στερεοσκοπικό ζεύγος:

$$\theta^{(t+1)} := \arg \max_{\theta} P\{Y, S; \theta^{(t)}\} = \arg \max_{\theta} \sum_d P_{app}^{(t)}\{D = d\} P\{Y, S | D = d; \theta^{(t)}\}$$

όπου το άθροισμα αναφέρεται σε όλες τις μετατοπίσεις M των μπλοκ του Y . Η ακριβής μεγιστοποίηση της παραπάνω έκφρασης δεν είναι εφικτή, γι' αυτό προσεγγίζεται με το να παραχθεί πρώτα το soft side information $\psi^{(t)}$ και μετά μια επανάληψη του joint bitplane LDPC αποκωδικοποίηση για να παραχθεί το $\theta^{(t)}$.

Η εκ των υστέρων κατανομή πιθανοτήτων της κίνησης M δηλαδή $P_{app}^{(t)}\{M_{u,v}\}$ δίνει βάρος στους υπολογισμούς που γίνονται σε κάθε ένα από τα μπλοκ του Y που έχει μετατόπιση $M_{u,v}$ δηλαδή σε κάθε ένα από τα $Y_{(u,v)+M_{u,v}}$. Μετά αυτοί οι υπολογισμοί προστίθενται για να μας δώσουν το soft side information $\psi_{u,v}^{(t)}$, αυτό φαίνεται στη δεξιά μεριά του σχήματος 4.11. Συγκεκριμένα, η πιθανότητα το side information να έχει τιμή w στο εικονοστοιχείο (i, j) είναι:

$$\theta^{(t+1)} := \arg \max_{\theta} P\{Y, S; \theta^{(t)}\} = \arg \max_{\theta} \sum_d P_{app}^{(t)}\{D = d\} P\{Y, S | D = d; \theta^{(t)}\}$$



Αντίστοιχα για το στερεοσκοπικό ζεύγος:

$$\begin{aligned}\psi^{(t)}(i, j, w) &= \sum_m P_{app}^{(t)}\{M = m\} P\{X(i, j) = w | M = m, Y\} \\ &= \sum_m P_{app}^{(t)}\{M = m\} p_Z(w - Y_m(i, j))\end{aligned}$$

όπου $p_Z(z)$ είναι η συνάρτηση κατανομής πιθανότητας του ανεξάρτητου προσθετικού θορύβου Z , και το Y_m είναι η προηγούμενη ανακατασκευασμένη εικόνα εφαρμόζοντας την μετατόπιση κίνησης m .

Η joint bitplane LDPC αποκωδικοποίηση ξεκινάει στους symbol κόμβους, όπως φαίνεται στο σχήμα 4.10. Εδώ το soft side information $\psi^{(t)}$ συνδυάζεται με log likelihood ratios $\alpha_g^{(t-1)}$ από την προηγούμενη επανάληψη για να παράξει log likelihood ratios $\beta_g^{(t)}$ σύμφωνα με τους υπολογισμούς στην παράγραφο 4.2.4. Οι κόμβοι symbol στέλνουν αυτά τα ratios στους δυαδικούς κόμβους οι οποίοι τα προωθούν στους syndrome κόμβους. Ακολουθώντας τους κανόνες αποκωδικοποίησης LDPC για το syndrome στο [51], οι syndrome κόμβοι απαντούν στους δυαδικούς κόμβους, οι δυαδικοί κόμβοι απαντούν στους symbol κόμβους δίνοντας τον τελευταίο υπολογισμό των log likelihood ratios $\alpha_g^{(t)}$. Τέλος, οι symbol κόμβοι παράγουν το επόμενο soft estimate της πηγής X , το οποίο πριν την κανονικοποίηση στις τιμές $w \in \{0, \dots, 2^d - 1\}$ υπολογίζεται ως:

$$\theta^{(t)}(i, j, w) := \psi^{(t)}(i, j, w) \prod_{g=1}^d (\alpha_g^{(t)})^{1_{[w_g=1]}} (1 - \alpha_g^{(t)})^{1_{[w_g=0]}}$$

όπου w_g είναι το δυαδικό ψηφίο g στην απεικόνιση Gray των τιμών φωτεινότητας w και $1_{[j]}$ είναι η συνάρτηση δεικτών.



4.3.5. Τερματισμός

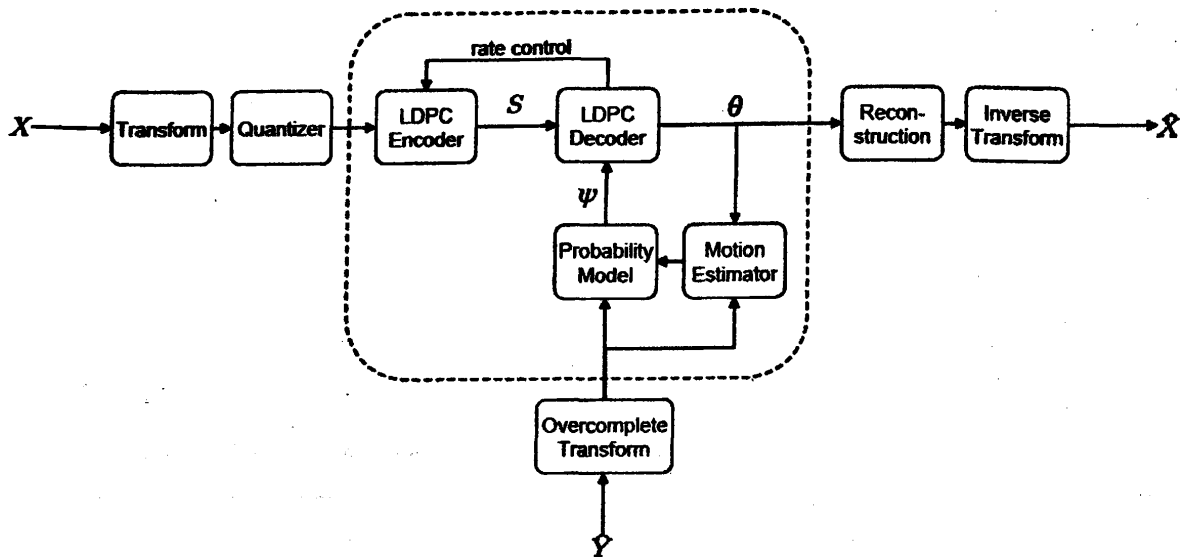
Κάνοντας επαναλήψεις ο αλγόριθμος μεταξύ E-βήματος και M-βήματος μαθαίνει τα διανύσματα κίνησης για κάθε μπλοκ. Ο αλγόριθμος τερματίζει επιτυχώς όταν η έκφραση $\hat{X}(i, j) = \arg \max_w \theta(i, j, w)$ παράγει syndrome ίσο με το S.

4.3.6. Τεχνική μάθησης διανυσμάτων κίνησης στον αποκωδικοποιητή σε επίπεδο μετασχηματισμού (transform domain)

Η τεχνική μάθησης των διανυσμάτων κίνησης αξιοποιεί το πλεόνασμα χρονικά (τμήμα 2), αλλά όχι χωρικά. Επεκτείνοντας την τεχνική στο χώρο μετασχηματισμού επιτυγχάνεται μεγαλύτερη συμπίεση. Αν συμβολίσουμε με X και Y τις εικόνες που δεν έχουν ακόμα κβαντιστεί, \hat{X} και \hat{Y} τις ανακατασκευασμένες εικόνες τους στο χώρο μετασχηματισμού τότε ο σκοπός είναι να κωδικοποιήσουμε το X με σκοπό να το ανακατασκευάσουμε και να βρούμε το \hat{X} χρησιμοποιώντας ως εικόνα αναφοράς (side information) το \hat{Y} . Αυτό φαίνεται στο σχήμα 4.13.

Το μη απωλεστικό κομμάτι που είναι σε διακεκομμένες γραμμές είναι όμοιο με το χώρο εικονοστοιχείων του σχήματος 4.8(c) με τη διαφορά ότι το τελευταίο εργάζεται σε κβαντισμένους συντελεστές. Στον κωδικοποιητή η εικόνα X μετασχηματίζεται χρησιμοποιώντας discrete cosine transform και οι μετασχηματισμένοι συντελεστές κβαντίζονται για να προχωρήσουν στο μη απωλεστικό κομμάτι της συμπίεσης. Στον αποκωδικοποιητή το \hat{Y} δεν χρησιμοποιείται κατευθείαν αλλά πρέπει να μετασχηματιστεί πρώτα.





Σχήμα 4.13: Τεχνική μάθησης διανυσμάτων κίνησης στο χώρο μετασχηματισμού όπου το μη απωλεστικό σύστημα κωδικοποίησης να είναι εντός της διακεκομμένης γραμμής.

ΚΕΦΑΛΑΙΟ 5. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

5.1 Πειραματικά Αποτελέσματα

5.2 Επίλογος

5.3 Μελλοντικές Επεκτάσεις

5.1. Πειραματικά αποτελέσματα

Ο κώδικας διαιρεί την ακολουθία εικόνων του βίντεο σε ομάδες από εικόνες (group of pictures, GOP) με σταθερό το μέγεθος GOP. Η πρώτη εικόνα του GOP κωδικοποιείται χωρίς αναφορά σε κάποια άλλη εικόνα, κωδικοποιείται με jpeg κωδικοποίηση και αποστέλλεται ολόκληρη στον αποκωδικοποιητή. Οι επόμενες εικόνες του GOP κωδικοποιούνται αναφορικά με τις προηγούμενες ανακατασκευασμένες εικόνες στον αποκωδικοποιητή και γι' αυτό αποστέλλονται από τον κωδικοποιητή μόνο τα δυαδικά ψηφία syndrome και γι' αυτό ο ρυθμός μετάδοσης τους είναι κατά πολύ μικρότερος από αυτόν της πρώτης εικόνας. Στα πειράματα χρησιμοποιήθηκε μια εικονοσειρά 70 εικόνων, με όνομα *Leaving_Laptop* όπως φαίνεται στο σχήμα 5.1, με ανάλυση CIF και ρυθμό 15 Hz. Η περιοχή αναζήτησης των οριζόντιων και κάθετων διανυσμάτων κίνησης και διαφοράς είναι 5 εικονοστοιχεία. Μετά από 50 επαναλήψεις του αλγορίθμου EM, αν η ανακατασκευασμένη εικόνα X δεν ικανοποιεί το syndrome, τότε ο αποκωδικοποιητής απαιτεί επιπλέον δυαδικά ψηφία του syndrome από τον κωδικοποιητή μέσω ενός καναλιού ανάδρασης.





Σχήμα 5.1: Αριστερή και δεξιά εικόνα του βίντεο που εφαρμόστηκε στα πειράματα.

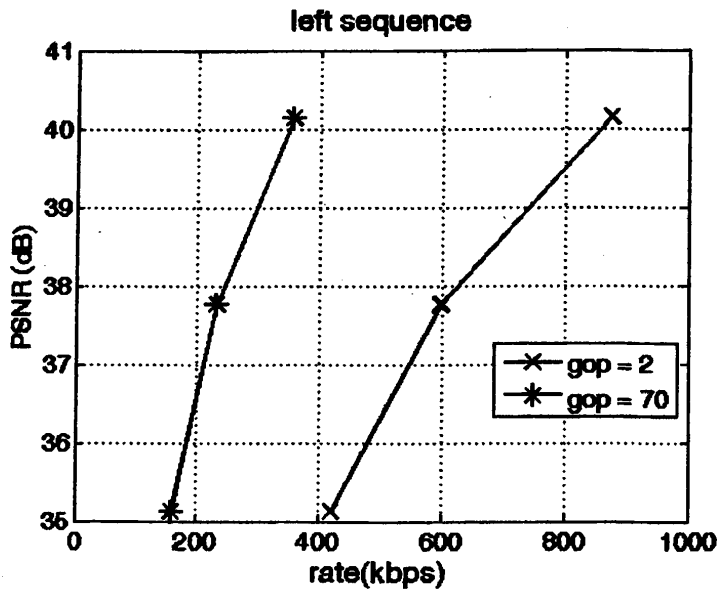
Τα πειράματα έγιναν σε υπολογιστή με επεξεργαστή core i7, 2.67 GHz, 4 GB μνήμη RAM και χρησιμοποιήθηκαν τα προγράμματα visual studio 2005, matlab 7.0.1 και intel ipp βιβλιοθήκη (intel integrated performance primitives library).

Στη συνέχεια εξετάζεται η επίδραση του μεγέθους GOP και των επιπέδων κβάντισης στην απόδοση του κώδικα ως προς το ισοζύγιο ρυθμού-παραμόρφωσης (rate-distortion), σχήμα 5. Τα μεγέθη GOP είναι 2, 70, ενώ τα επίπεδα κβάντισης είναι τρία 0.5, 1, 2. Τα πειραματικά αποτελέσματα μελετώνται με παράμετρο το μέγεθος GOP, ως προς τα τέσσερα συστήματα κωδικοποίησης: ιδανική εκτίμηση κίνησης και διαφοράς (motion oracle), Wyner-Ziv εκτίμηση κίνησης και διαφοράς (motion and disparity learning), κωδικοποίηση εκτίμηση κίνησης και διαφοράς (zero motion and disparity), κωδικοποίηση jpeg (intra coding).

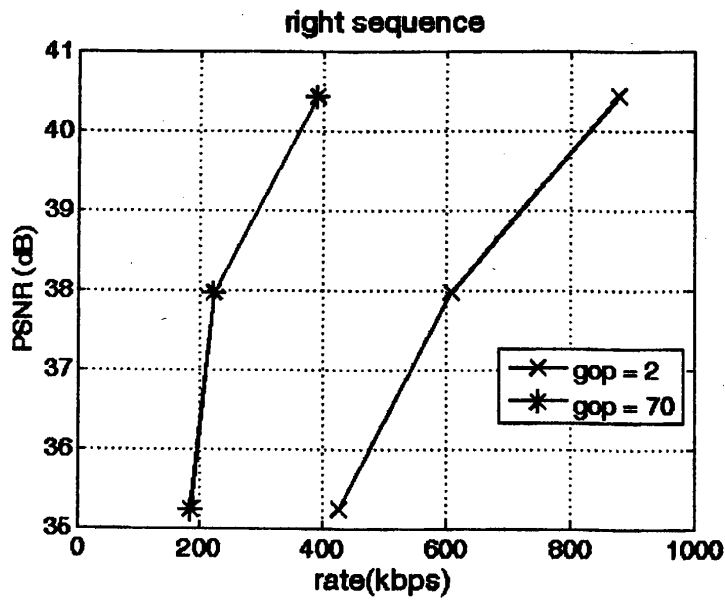
Με παράμετρο το μέγεθος GOP: Και στις τρεις γραφικές παραστάσεις βλέπουμε ότι με δεδομένο ότι η ποιότητα είναι η ίδια, ο ρυθμός είναι καλύτερος για GOP = 70 από ότι για GOP = 2 και αυτό είναι και το αναμενόμενο όπως είπαμε και πιο πάνω. Στο 5(α) κωδικοποιείται η αριστερή ακολουθία του βίντεο κάνοντας οριζόντια και κάθετη εκτίμηση της κίνησης. Στο 5(β) κωδικοποιείται η δεξιά ακολουθία του βίντεο κάνοντας οριζόντια και κάθετη εκτίμηση της διαφοράς. Στο 5(γ) κωδικοποιείται η δεξιά ακολουθία του βίντεο κάνοντας μόνο οριζόντια εκτίμηση της διαφοράς. Το 5(β) είναι καλύτερο ως προς το ρυθμό μετάδοσης από το 5(γ), ωστόσο το 5(γ) λόγω



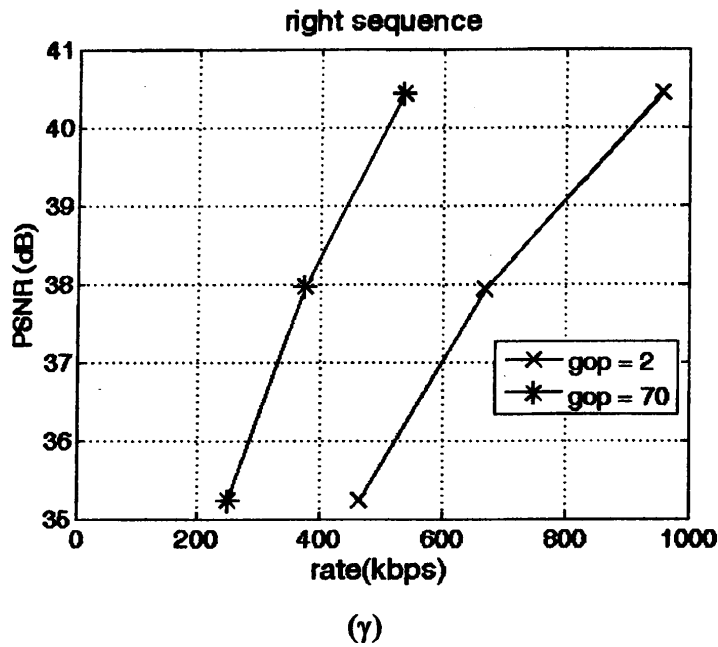
του ότι η αναζήτηση των διανυσμάτων διαφοράς είναι μόνο οριζόντια εξοικονομεί αρκετό χρόνο σε σχέση με το 5(β). Έτσι στο 5(β) κερδίζουμε σε ρυθμό ενώ στο 5(γ) κερδίζουμε σε χρόνο.



(α)

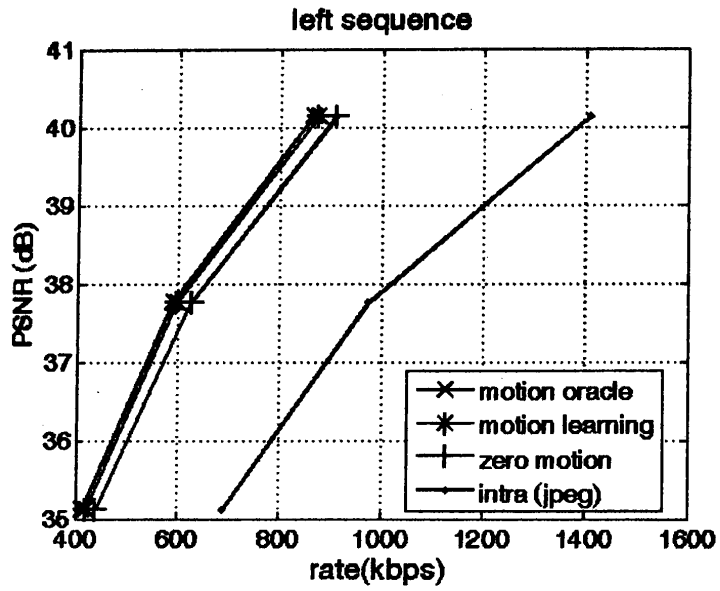


(β)

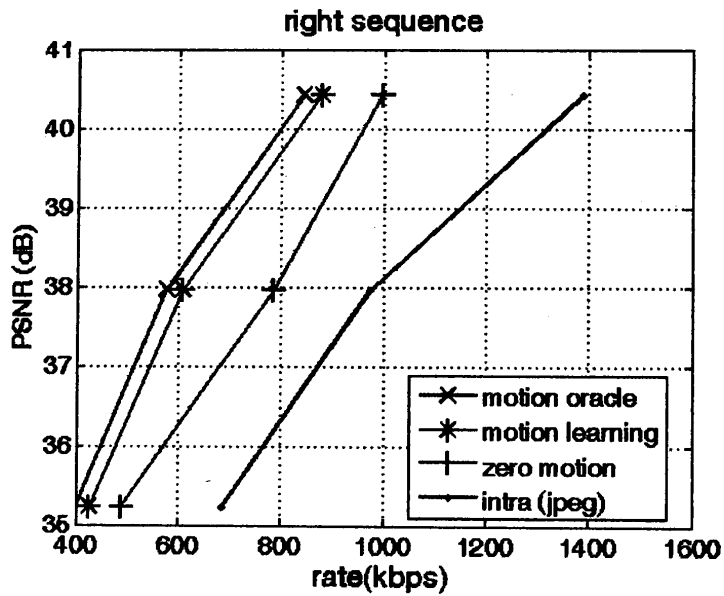


Σχήμα 5.2: Ρυθμός-παραμόρφωση για διαφορετικά GOP, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς.

Με παράμετρο τα τέσσερα συστήματα κωδικοποίησης και μέγεθος GOP 2: Στο 5.2(α) βλέπουμε ότι ο κώδικας Wyner-Ziv, με δεδομένο ότι η ποιότητα είναι η ίδια και για τα τέσσερα συστήματα, απαιτεί κατά πολύ μικρότερο ρυθμό από το jpeg και το zero motion, ενώ προσεγγίζει αρκετά το motion oracle. Ομοίως με τα παραπάνω στο σχήμα 5.1, το 5.2(β) είναι καλύτερο ως προς το ρυθμό μετάδοσης από το 5.2(γ), ωστόσο το 5.2(γ) λόγω του ότι η αναζήτηση των διανυσμάτων διαφοράς είναι μόνο οριζόντια εξοικονομεί αρκετό χρόνο σε σχέση με το 5.2(β). Έτσι στο 5.2(β) κερδίζουμε σε ρυθμό ενώ στο 5.2(γ) κερδίζουμε σε χρόνο.

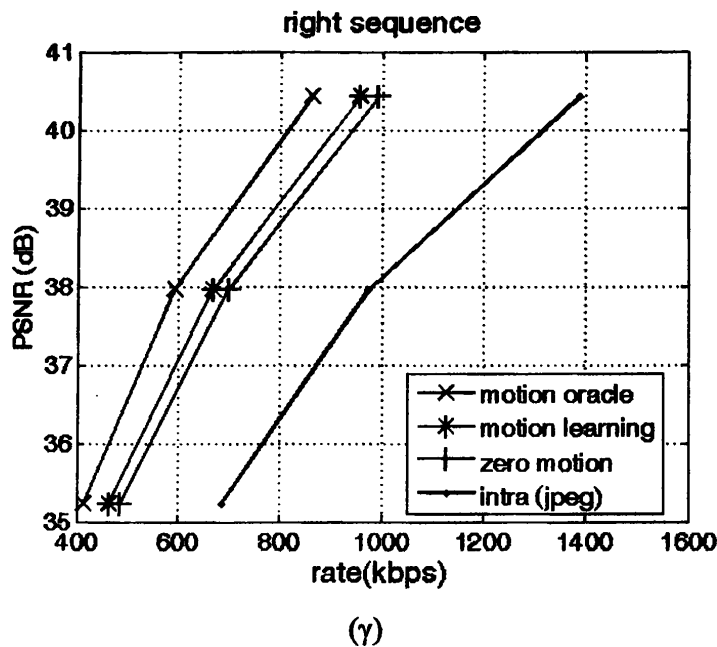


(a)



(b)

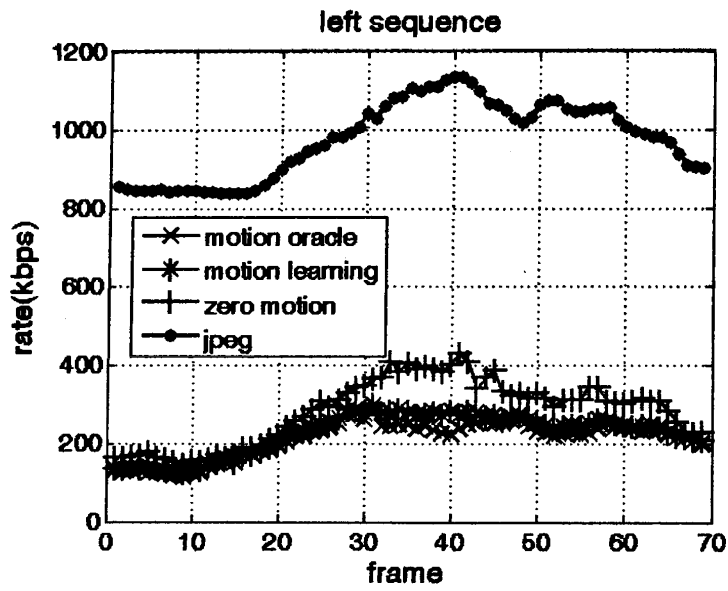




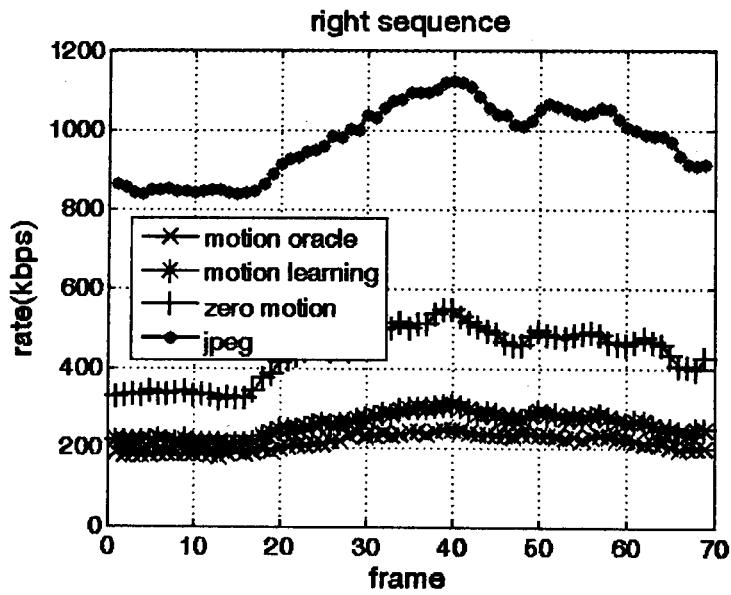
Σχήμα 5.3: Ρυθμός-παραμόρφωση για τέσσερα συστήματα κωδικοποίησης για GOP = 2, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς.

Με παράμετρο τα τέσσερα συστήματα κωδικοποίησης και μέγεθος GOP 70:

Στο σχήμα 5.4 βλέπουμε την ποιότητα των τεσσάρων συστημάτων κωδικοποίησης όπου φαίνεται ότι είναι η ίδια και στα τέσσερα συστήματα. Στο σχήμα 5.3 βλέπουμε το ρυθμό μετάδοσης και για τα τέσσερα συστήματα κωδικοποίησης, με GOP = 70 και για επίπεδο κβάντισης 1. Με δεδομένο ότι η ποιότητα είναι η ίδια και στα τέσσερα συστήματα βλέπουμε ότι στο σχήμα 5.3(α.1) το jpeg απαιτεί κατά πολύ περισσότερο ρυθμό από το Wyner-Ziv (motion learning). Στο 5.3(α.2) φαίνεται πιο ξεκάθαρα η διαφορά ρυθμού, ο Wyner-Ziv έχει μικρότερο ρυθμό σε σχέση με zero motion και πλησιάζει σχεδόν το motion oracle. Ομοίως το ίδιο ισχύει και στα 5.3(β.1), 5.3(β.2) και 5.3(γ.1), 5.3(γ.2). Ομοίως με τα παραπάνω στο σχήμα 5.1, το 5.3(β) είναι καλύτερο ως προς το ρυθμό μετάδοσης από το 5.3(γ), ωστόσο το 5.3(γ) λόγω του ότι η αναζήτηση των διανυσμάτων διαφοράς είναι μόνο οριζόντια εξοικονομεί αρκετό χρόνο σε σχέση με το 7(β). Έτσι στο 7(β) κερδίζουμε σε ρυθμό ενώ στο 7(γ) κερδίζουμε σε χρόνο.

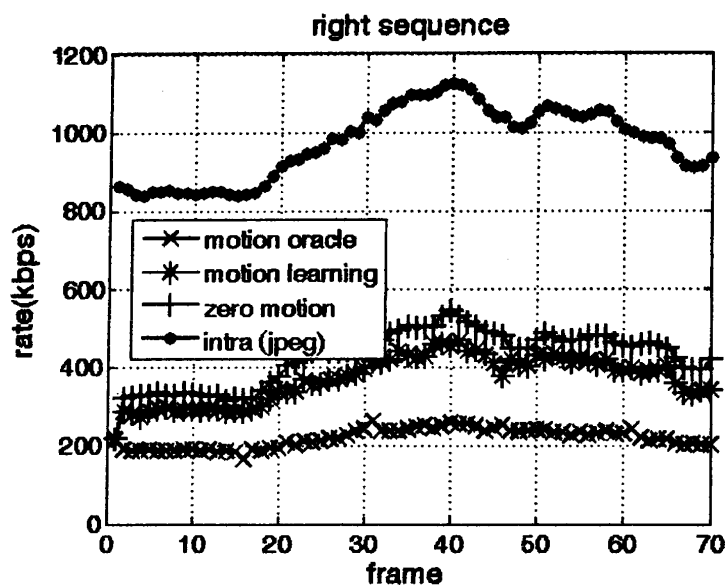


(α)



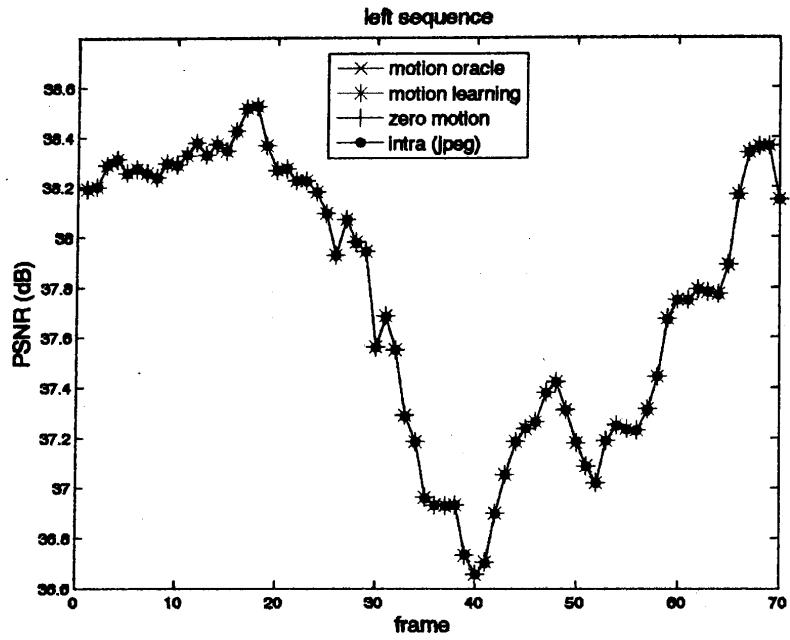
(β)



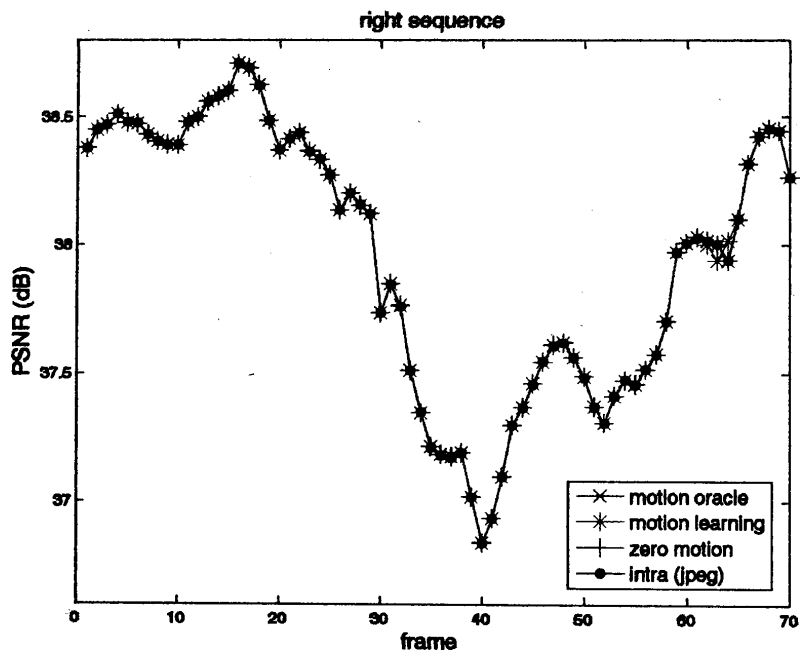


(γ)

Σχήμα 5.4: Ο ρυθμός για τέσσερα συστήματα κωδικοποίησης για $GOP = 70$, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς.

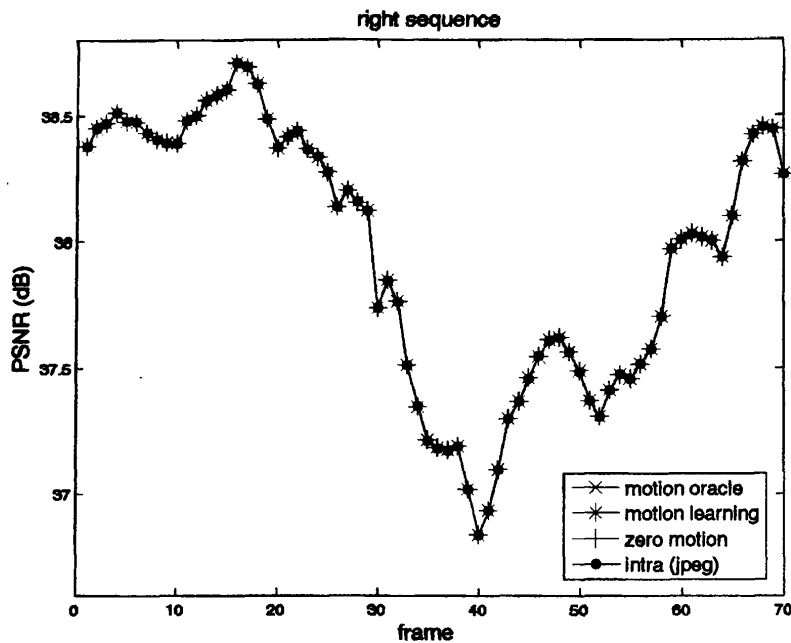


(α)



(β)





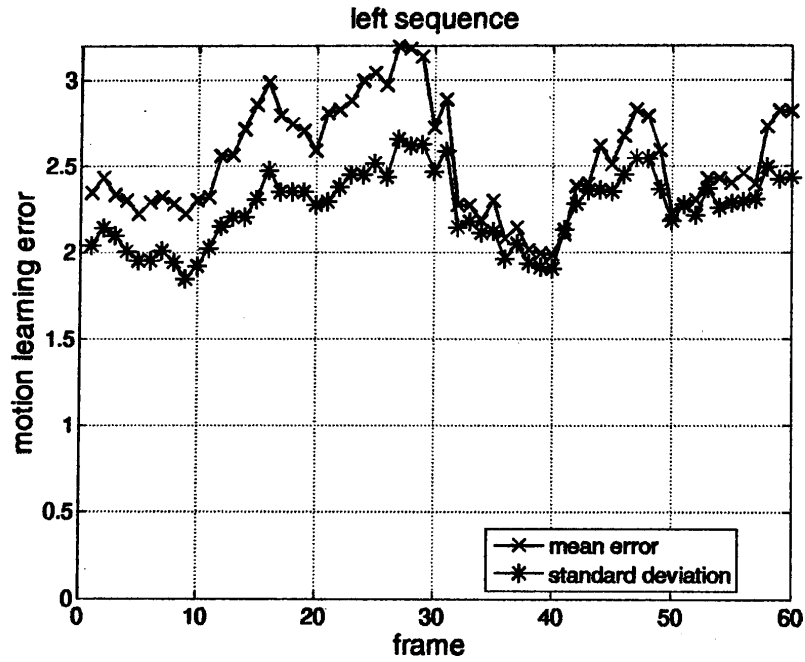
(γ)

Σχήμα 5.5: Η ποιότητα για τέσσερα συστήματα κωδικοποίησης για $\text{GOP} = 70$, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς.

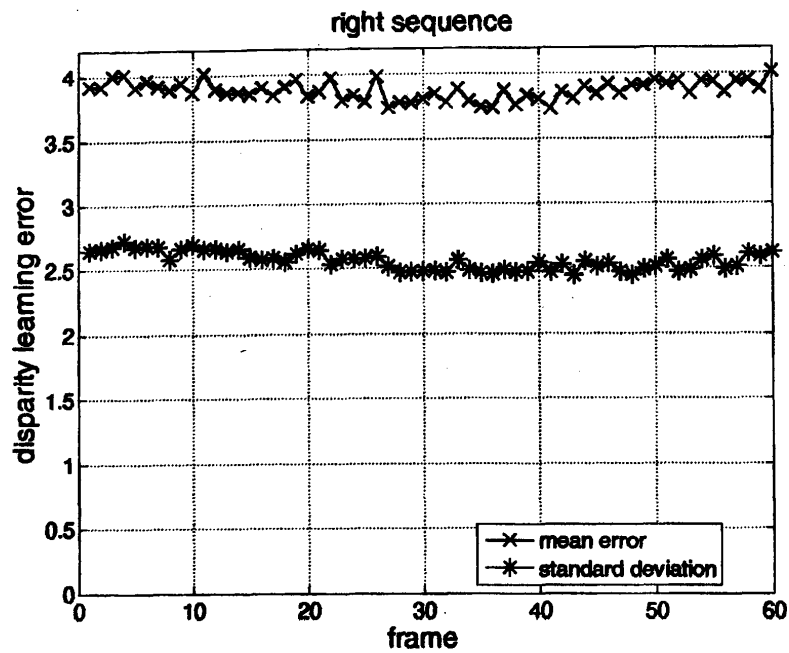
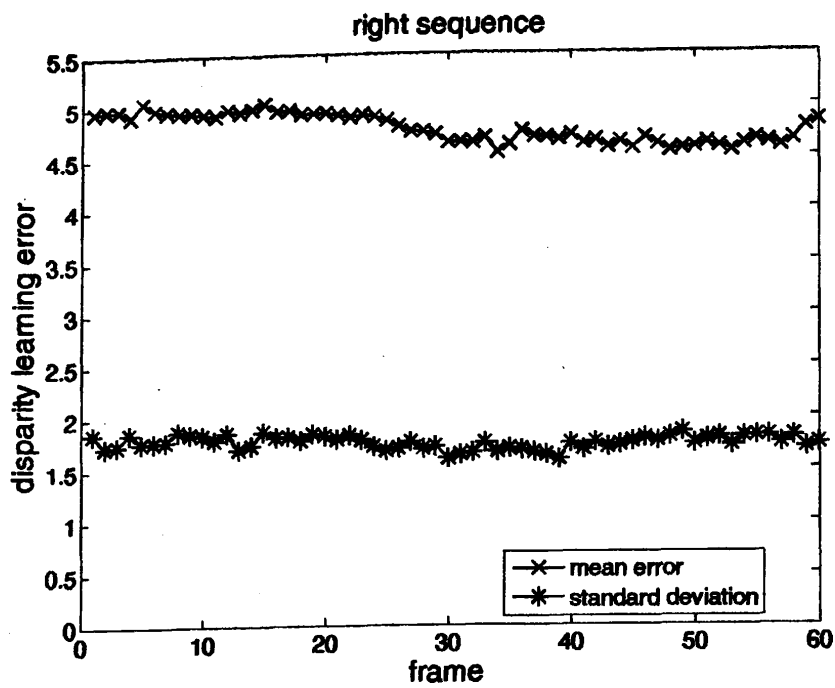
Στο σχήμα 5.5 μετράμε την ορθότητα των διανυσμάτων κίνησης και διαφοράς. Υπολογίζουμε την ευκλείδια απόσταση του διανύσματος κίνησης (ή διαφοράς αντίστοιχα) Wyner-Ziv από το διάνυσμα κίνησης της ιδανικής εκτίμησης κίνησης (motion oracle). Βλέπουμε ότι τα σφάλματα δεν είναι ιδιαίτερα μεγάλα. Η τυπική απόκλιση δείχνει κατά πόσο αποκλίνουν τα σφάλματα σε μια εικόνα και βλέπουμε ότι οι αποκλίσεις δεν υπερβαίνουν το 3. Το σφάλμα μιας εικόνας προκύπτει παίρνοντας τον μέσο όρο των σφαλμάτων όλων των διανυσμάτων κίνησης. Αντίστοιχα προκύπτει η τυπική απόκλιση παίρνοντας τον μέσο όρο των τυπικών αποκλίσεων όλων των διανυσμάτων κίνησης. Η βεβαιότητα των διανυσμάτων κίνησης φαίνεται από τις πιθανότητες στο σχήμα 5.6. Βλέπουμε ότι οι πιθανότητες είναι αρκετά μεγάλες οπότε και τα διανύσματα κίνησης έχουν αρκετά μεγάλη ακρίβεια. Ομοίως με τα παραπάνω στο σχήμα 5.1, το 5.5(β) είναι καλύτερο ως προς το ρυθμό



μετάδοσης από το 5.5(γ), ωστόσο το 5.5(γ) λόγω του ότι η αναζήτηση των διανυσμάτων διαφοράς είναι μόνο οριζόντια εξοικονομεί αρκετό χρόνο σε σχέση με το 5.5(β). Έτσι στο 5.5(β) κερδίζουμε σε ρυθμό ενώ στο 5.5(γ) κερδίζουμε σε χρόνο. Το ίδιο για το σχήμα 5.6.



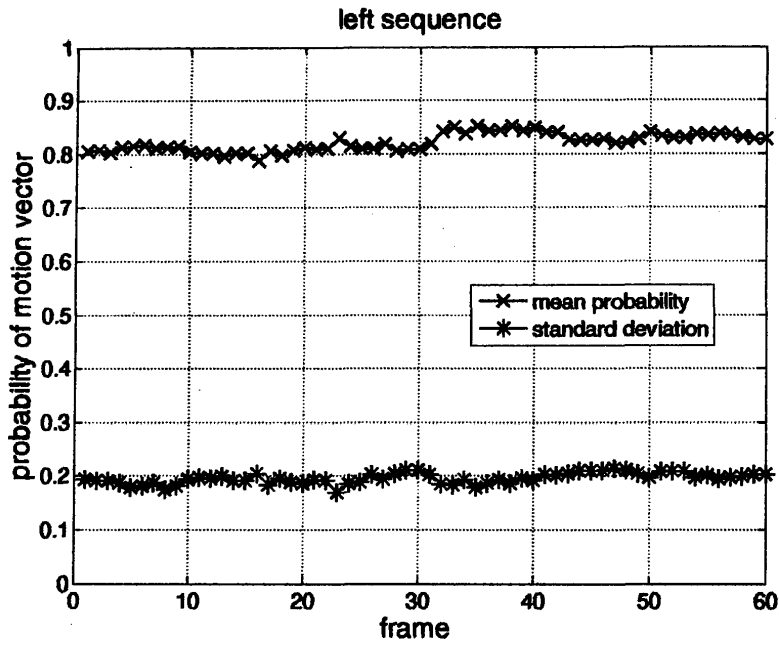
(α)



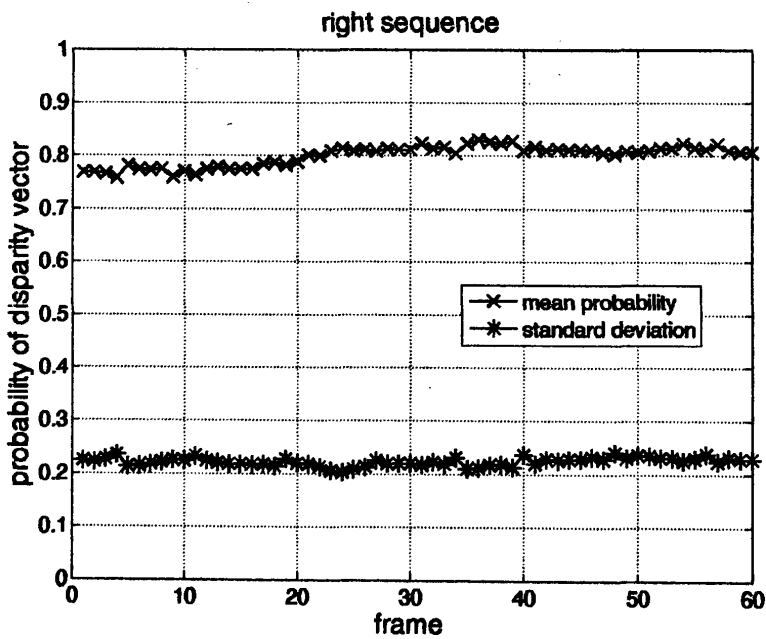
Σχήμα 5.6: Τα σφάλματα των διανυσμάτων κίνησης και διαφοράς για το σύστημα κωδικοποίησης Wyner-Ziv για $GOP = 70$, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence):



οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence):
οριζόντια εκτίμηση διαφοράς.

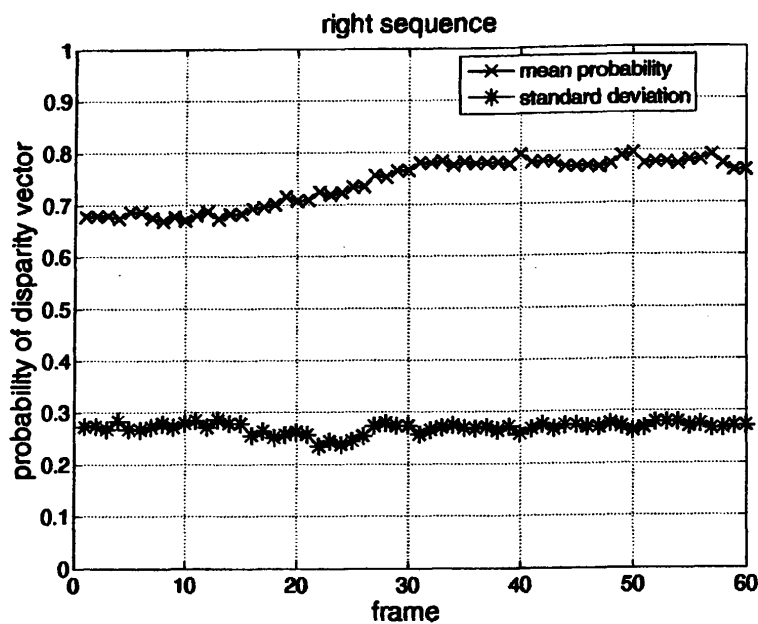


(α)



(β)





(γ)

Σχήμα 5.7: Οι πιθανότητες των διανυσμάτων κίνησης και διαφοράς για το σύστημα κωδικοποίησης Wyner-Ziv για $\text{GOP} = 70$, (α) αριστερή ακολουθία (left sequence): οριζόντια και κάθετη εκτίμηση κίνησης, (β) δεξιά ακολουθία (right sequence): οριζόντια και κάθετη εκτίμηση διαφοράς, (γ) δεξιά ακολουθία (right sequence): οριζόντια εκτίμηση διαφοράς.

5.2. Επίλογος

Στην εργασία αυτή μελετήθηκε το πρόβλημα της κατανεμημένης συμπίεσης στερεοσκοπικού βίντεο. Η τεχνική που λύνει το πρόβλημα αυτό, εκτελεί εκτίμησης κίνησης και διαφοράς (motion and disparity estimation) στον αποκωδικοποιητή Wyner-Ziv. Για το σκοπό αυτό, χρησιμοποιήθηκε ένας επαναληπτικός αλγόριθμος μεγιστοποίησης πιθανότητας - EM (expectation maximization), ο οποίος εναλλάσσεται μεταξύ ανανέωσης της εκτίμησης κίνησης και διαφοράς και της ανανέωσης των πληροφοριών ισοτιμίας που στέλνονται από την πηγή, καθώς και μια μέθοδος joint bitplane LDPC decoding.



Τα πειραματικά αποτελέσματα δείχνουν ότι η μέθοδος συμπίεσης είναι πιο αποδοτική από την κλασική μέθοδο συμπίεσης JPEG ως προς το rate-distortion. Επίσης, είναι πιο αποτελεσματική από τη μέθοδο που δεν κάνει εκτίμηση κίνησης (no motion compensation) και προσεγγίζει κατά πολύ την ιδανική εκτίμηση κίνησης (motion oracle). Τέλος, η πολυπλοκότητα υπολογισμού είναι πολύ μικρότερη από αυτή του αλγορίθμου που κάνει πλήρη αναζήτηση των διανυσμάτων κίνησης (full motion search).

5.3. Μελλοντικές επεκτάσεις

Μια μελλοντική εργασία θα μπορούσε να ασχοληθεί με εικόνες βίντεο που έχουν μεγαλύτερο εύρος διαφοράς (disparity range) καθώς και να μειώσει την υπολογιστική και χρονική πολυπλοκότητα της συμπίεσής τους.



ΑΝΑΦΟΡΕΣ

- [1] B. Girod, A. Aaron, S. Rane, D. Rebollo-Monedero, Distributed video coding, *Proceedings IEEE 93 (1) (2005)* 71–83.
- [2] J. Slepian and J. Wolf, “Noiseless Coding of Correlated Information Sources,” *IEEE Trans. On Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.
- [3] A. Wyner and J. Ziv, “The Rate-Distortion Function for Source Coding with Side Information at the Decoder,” *IEEE Trans. on Information Theory*, vol. 22, no. 1, pp. 1–10, January 1976.
- [4] A. D. Wyner, “Recent Results in the Shannon Theory,” *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 2–10, Jan. 1974.
- [5] S. S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 1999, pp. 158 –167.
- [6] S.S. Pradhan and K. Ramchandran, “Distributed source coding: Symmetric rates and applications to sensor networks,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2000, pp. 363 –372.
- [7] —, “Group-theoretic construction and analysis of generalized coset codes for symmetric/asymmetric distributed source coding,” in *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2000.



- [8] —, “Geometric proof of rate-distortion function of Gaussian sources with side information at the decoder,” in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Sorrento, Italy, June 2000, p. 351.
- [9] S. S. Pradhan, J. Kusuma, and K. Ramchandran, “Distributed compression in a dense microsensor network,” *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 51–60, Mar. 2002.
- [10] X. Wang and M. Orchard, “Design of trellis codes for source coding with side information at the decoder,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2001, pp. 361–370.
- [11] J. Bajcsy and P. Mitran, “Coding for the Slepian-Wolf problem with turbo codes,” in *Proc. IEEE Global Communications Conference*, vol. 2, San Antonio, TX, Nov. 2001, pp. 1400–1404.
- [12] P. Mitran and J. Bajcsy, “Near Shannon-limit coding for the Slepian-Wolf problem,” in *Proc. Biennial Symposium on Communications*, Kingston, Ontario, June 2002.
- [13] A. Aaron and B. Girod, “Compression with side information using turbo codes,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Apr. 2002, pp. 252–261.
- [14] Y. Zhao and J. Garcia-Frias, “Joint estimation and data compression of correlated non-binary sources using punctured turbo codes,” in *Proc. Conference on Information Sciences and Systems*, Princeton, NJ, Mar. 2002.
- [15] —, “Data compression of correlated non-binary sources using punctured turbo codes,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Apr. 2002, pp. 242–251.



- [16] P. Mitran and J. Bajcsy, "Coding for the Wyner-Ziv problem with turbolike codes," in *Proc. IEEE International Symposium on Information Theory*, Lausanne, Switzerland, June 2002, p. 91.
- [17] —, "Turbo source coding: a noise-robust approach to data compression," in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Apr. 2002, p. 465.
- [18] G. Zhu and F. Alajaji, "Turbo codes for non-uniform memoryless sources over noisy channels," *IEEE Communications Letters*, vol. 6, no. 2, pp. 64–66, Feb. 2002.
- [19] J. Garcia-Frias and Y. Zhao, "Compression of binary memoryless sources using punctured turbo codes," *IEEE Communications Letters*, vol. 6, no. 9, pp. 394–396, Sept. 2002.
- [20] J. Garcia-Frias, "Joint source-channel decoding of correlated sources over noisy channels," in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2001, pp. 283–292.
- [21] A. Liveris, Z. Xiong, and C. Georghiades, "Joint source-channel coding of binary sources with side information at the decoder using IRA codes," in *Proc. Multimedia Signal Processing Workshop*, St. Thomas, US Virgin Islands, Dec. 2002.
- [22] —, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Communications Letters*, vol. 6, no. 10, pp. 440–442, Oct. 2002.
- [23] —, "Compression of binary sources with side information at the decoder using LDPC codes," in *Proc. IEEE Global Communications Symposium*, Taipei, Taiwan, Nov. 2002.
- [24] D. Schonberg, S. S. Pradhan, and K. Ramchandran, "LDPC codes can approach the Slepian-Wolf bound for general binary sources," in *Proc. Allerton Conference on Communication, Control, and Computing*, Champaign, IL, Oct. 2002.



- [25] —, “Distributed code constructions for the entire Slepian-Wolf rate region for arbitrarily correlated sources,” in *Proc. Asilomar Conference on Signals and Systems*, Pacific Grove, CA, Nov. 2003, pp. 835–839.
- [26] —, “Distributed code constructions for the entire Slepian-Wolf rate region for arbitrarily correlated sources,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2004, pp. 292–301.
- [27] T. P. Coleman, A. H. Lee, M. Medard, and M. Effros, “On some new approaches to practical Slepian-Wolf compression inspired by channel coding,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2004, pp. 282–291.
- [28] V. Stankovic, A. D. Liveris, Z. Xiong, and C. N. Georghiades, “Design of Slepian-Wolf codes by channel code partitioning,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2004, pp. 302–311.
- [29] C. F. Lan, A. D. Liveris, K. Narayanan, Z. Xiong, and C. Georghiades, “Slepian-Wolf coding of multiple M-ary sources using LDPC codes,” in *Proc. IEEE Data Compression Conference*, Snowbird, UT, Mar. 2004, pp. 549–549.
- [30] N. Gehrig and P. L. Dragotti, “Symmetric and a-symmetric Slepian-Wolf codes with systematic and non-systematic linear codes,” *IEEE Communications Letters*, Jan. 2005, to appear.
- [31] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.
- [32] A. Wyner, “On source coding with side information at the decoder,” *IEEE Transactions on Information Theory*, vol. IT-21, no. 3, pp. 294–300, May 1975.



- [33] —, “The rate-distortion function for source coding with side information at the decoder—II: General sources,” *Information and Control*, vol. 38, no. 1, pp. 60–80, July 1978.
- [34] M. Costa, “Writing on Dirty Paper,” *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [35] J. Cardinal and G. V. Asche, “Joint entropy-constrained multiterminal quantization,” in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Lausanne, Switzerland, June 2002, p. 63.
- [36] D. Rebollo-Monedero, R. Zhang, and B. Girod, “Design of optimal quantizers for distributed source coding,” in *Proc. IEEE Data Compression Conference (DCC)*, Snowbird, UT, Mar. 2003, pp. 13–22.
- [37] Y. Wang, J. Ostermann, and Y. Zhang, *Video Processing and Communications*, 1st ed. New Jersey: Prentice-Hall, 2002.
- [38] T. Wiegand, G. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.
- [39] D. Rebollo-Monedero, A. Aaron, and B. Girod, “Transforms for highrate distributed source coding,” in *Proc. Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2003.
- [40] A. Aaron, S. Rane, E. Setton, and B. Girod, “Transform-domain Wyner-Ziv codec for video,” in *Proc. SPIE Visual Communications and Image Processing*, San Jose, CA, Jan. 2004.
- [41] Stereographics, Inc. Perceiving stereoscopic images.
<http://www.stereographics.com/html/chapter 2.html>.



- [42] D. Chen, D. Varodayan, M. Flierl, B. Girod, Wyner-Ziv coding of multiview images with unsupervised learning of disparity and Gray code, in: Proc. IEEE Internat. Conf. Image Processing, San Diego, CA, 2008, submitted.
- [43] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in Proc. Asilomar Conf. on Signals, Syst., Comput., Pacific Grove, CA, 2002.
- [44] R. Puri and K. Ramchandran, "PRISM: A new robust video coding architecture based on distributed compression principles," in Proc. Allerton Conf. Commun., Contr. and Comput., Allerton, IL, 2002.
- [45] A. Aaron, S. Rane, and B. Girod, "Wyner-Ziv video coding with hashbased motion compensation at the receiver," in Proc. IEEE International Conf. on Image Processing, Singapore, 2004.
- [46] A. Aaron, S. Rane, E. Setton, and B. Girod, "Transform-domain Wyner-Ziv codec for video," in Proc. Visual Commun. and Image Processing, San Jose, CA, 2004.
- [47] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in Proc. Comput. Vision and Pattern Recog., Madison, WI, 2003.
- [48] ITU-T and ISO/IEC JTC1, "Digital compression and coding of continuous-tone still images," ISO/IEC 10918-1 — ITU-T Recommendation T.81 (JPEG), Sept. 1992.
- [49] D. Varodayan, A. Mavlankar, M. Flierl, and B. Girod, "Distributed grayscale stereo image coding with unsupervised learning of disparity," in Proc. IEEE Data Compression Conf., Snowbird, UT, 2007.
- [50] R. G. Gallager, "Low-density parity-check codes," Cambridge MA: MIT Press, 1963.



[51] A. Liveris, Z. Xiong, C. Georghiades, Compression of binary sources with side information at the decoder using LDPC codes, *IEEE Commun. Lett.* 6 (10) (2002) 440–442

[52] D. Varodayan, A. Aaron, B. Girod, Rate-adaptive distributed source coding using low-density parity-check codes, in: *Proc. Asilomar Conf. on Signals, Syst., Comput.*, Pacific Grove, CA, 2005.

[53] D. Varodayan, A. Aaron, B. Girod, Rate-adaptive codes for distributed source coding, *EURASIP Signal Processing J.* 86 (11) (2006) 3123–3130.

[54] F. R. Kschischang, B. J. Frey, H.-A. Loeliger, Factor graphs and the sum-product algorithm, *IEEE Trans. Inform. Theory* 47 (2) (2001) 498–519.



ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Ελίνα Σουκαλλάρη γεννήθηκε στην Πρεμετή Αλβανίας το 1984. Το έτος 2004 εισήχθη στο Τμήμα Πληροφορικής των Θετικών Επιστημών του Πανεπιστημίου Ιωαννίνων και αποφοίτησε το 2008. Συνέχισε για Μεταπτυχιακές σπουδές στο ίδιο τμήμα όπου ασχολήθηκε με συμπύεση βίντεο.



