



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Μέθοδος Πρόβλεψης της Αντιληπτής Ποιότητας Ψηφιακού Βίντεο Συναρτήσεως του Ρυθμού Κωδικοποίησης και της Δυναμικής του Περιεχομένου

Χαρίλαος Γ. Κουμαράς

ΑΘΗΝΑ
ΣΕΠΤΕΜΒΡΙΟΣ 2007



ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Μέθοδος Πρόβλεψης της Αντιληπτής Ποιότητας Ψηφιακού Βίντεο Συναρτήσεως του Ρυθμού Κωδικοποίησης και της Δυναμικής του Περιεχομένου

Χαρίλαος Γ. Κουμαράς

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:

Δρακούλης Μαρτάκος, Αν. Καθηγητής ΕΚΠΑ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:

Δρακούλης Μαρτάκος, Αν. Καθηγητής ΕΚΠΑ

Ιωάννης Σταυρακάκης, Καθηγητής ΕΚΠΑ

Αναστάσιος Κούρτης, Κύριος Ερευνητής, ΕΚΕΦΕ Δημόκριτος

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Δρακούλης Μαρτάκος

Αν. Καθηγητής ΕΚΠΑ

Τμ. Πληροφορικής & Τηλεπικοινωνιών

Ηλίας Μανωλάκος

Αν. Καθηγητής ΕΚΠΑ

Τμ. Πληροφορικής & Τηλεπικοινωνιών

Ιωάννης Σταυρακάκης

Καθηγητής ΕΚΠΑ

Τμ. Πληροφορικής & Τηλεπικοινωνιών

Λάζαρος Μεράκος

Καθηγητής ΕΚΠΑ

Τμ. Πληροφορικής & Τηλεπικοινωνιών

Αναστάσιος Κούρτης

Κύριος Ερευνητής, ΕΚΕΦΕ Δημόκριτος

Ινστ. Πληροφορικής & Τηλεπικοινωνιών

Εμμανουήλ Σαγκριώτης

Αν. Καθηγητής ΕΚΠΑ

Τμ. Πληροφορικής & Τηλεπικοινωνιών

Σέργιος Θεοδωρίδης

Καθηγητής ΕΚΠΑ

Τμ. Πληροφορικής & Τηλεπικοινωνιών

Ημερομηνία εξέτασης **13/09/2007**

ΠΕΡΙΛΗΨΗ

Ερευνητικά η διατριβή εστιάζεται στο χώρο της ψηφιακής κωδικοποίησης βίντεο και ειδικότερα στην εκτίμηση του επιπέδου της αντιληπτής ποιότητας κωδικοποιημένου βίντεο σύμφωνα με το πρότυπο MPEG-4/H.264.

Σε αυτά τα πλαίσια διερευνώνται αρχικά οι στατιστικές ιδιότητες ενός σήματος κωδικοποιημένου σύμφωνα με το πρότυπο MPEG-4/H.264, όπου παρουσιάζονται οι υψηλές συσχετίσεις μεταξύ διαδοχικών καρτέ, βάση των οποίων επιτυγχάνεται η συμπίεση των δεδομένων κατά τη διαδικασία κωδικοποίησης.

Το κυρίως μέρος της διατριβής πραγματεύεται την εκτίμηση του επιπέδου της αντιληπτής ποιότητας ενός ψηφιακά κωδικοποιημένου βίντεο, προτείνοντας πρωτότυπη μέθοδο πρόβλεψης του επιπέδου της αντιληπτής ποιότητας συναρτήσεως του ρυθμού κωδικοποίησης και της χωρικής και χρονικής δυναμικής του περιεχομένου. Η προτεινόμενη μέθοδος εφαρμόζεται αρχικά στα πλαίσια βίντεο μικρής διάρκειας με ομοιογενές και ετερογενές περιεχόμενο. Η αξιοπιστία της μεθόδου εξετάζεται εκτενώς και αποδεικνύεται ότι παρέχει αξιόπιστα αποτελέσματα στα βίντεο βραχείας διάρκειας.

Εν συνεχεία, η προτεινόμενη μέθοδος πρόβλεψης επεκτείνεται σε βίντεο μακράς διάρκειας, τα οποία χαρακτηρίζονται εν γένει ως ετερογενή λόγω της έντονης μεταβολής της δυναμικής του περιεχομένου τους με τη πάροδο του χρόνου. Για τις ανάγκες της επέκτασης αναπτύσσεται και προτείνεται πρωτότυπη μέθοδος κατάτμησης ενός βίντεο μεγάλης διάρκειας στις στοιχειώδεις σκηνές του, μέσω ανιχνεύσεως των ορίων των σκηνών του.

Επιπλέον, διερευνάται η σχέση της απόδοσης σε όρους αντιληπτής ποιότητας, που επιτυγχάνεται μέσω κωδικοποίησης σταθερού και μεταβλητού ρυθμού. Σύμφωνα με τη διεθνή βιβλιογραφία, η μέθοδος κωδικοποίησης μεταβλητού ρυθμού παρέχει ποιοτικότερη κωδικοποίηση, εξαιτίας του γεγονότος ότι διατηρεί σταθερούς τους συντελεστές κβάντισης. Η διατριβή, εν προκειμένω, πρωτοτυπεί επειδή διερευνά με ποσοτικό τρόπο την υπεροχή αυτή για ποικίλους ρυθμούς κωδικοποίησης και μεγέθη χωρικής ανάλυσης.

Επίσης, διερευνάται η επίδραση των σφαλμάτων του δίκτυο μεταφοράς και ειδικότερα της απώλειας πακέτων δεδομένων στο επίπεδο της αντιληπτής ποιότητας μιας ροής βίντεο πραγματικού χρόνου. Το προτεινόμενο θεωρητικό μοντέλο πρωτοτυπεί συγκριτικά με τη σχετική βιβλιογραφία, διότι η ισχύς του δεν περιορίζεται από το πρότυπο κωδικοποίησης ούτε από τη δυναμική του περιεχομένου της υπηρεσίας. Αντιθέτως, πρόκειται για ένα γενικευμένο μοντέλο απεικόνισης του ρυθμού απώλειας πακέτων δεδομένων στην αντιληπτή ποιότητα της μεταδιδόμενης υπηρεσίας.

Εν κατακλείδι, η διατριβή θέτει προς συζήτηση τις δυνατότητες συνδυασμού της προτεινόμενης μεθόδου πρόβλεψης και του μοντέλου απεικόνισης του ρυθμού απώλειας πακέτων στο επίπεδο της αντιληπτής ποιότητας, ώστε να προκύψει το θεωρητικό υπόβαθρο για τη διατεμαστική εκτίμηση της ποιότητας ενός ψηφιακά κωδικοποιημένου βίντεο σύμφωνα με το πρότυπο MPEG.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Ψηφιακό Βίντεο

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Αντιληπτή ποιότητα υπηρεσίας, Ποιότητα Βίντεο, MPEG, Πρόβλεψη, Δυναμική περιεχομένου, Ρυθμός κωδικοποίησης

ABSTRACT

The dissertation is focused in the research field of digital video coding and more specifically on the assessment of video quality level according to the encoding standard MPEG-4/H.264.

In this context, it is initially researched the statistical properties of a MPEG-4/H.264 encoded video signal, where it is revealed the high correlation relations between successive frames, on which the high compression ratio of the encoding process is based.

The main part of the dissertation deals with the evaluation of the perceived quality of video encoded signals, proposing a novel method for predicting the perceived quality level in correlation with the encoding bit rate and the spatiotemporal content dynamics. The proposed method is initially examined on short in duration video clips with homogeneous and heterogeneous content. Upon extended validation, it is proved the efficiency of the proposed method on videos with short duration.

Afterwards, the proposed method is extended to videos with long duration, which are generally characterized as heterogeneous because of their continuously variable content dynamics. For the needs of the extension, it is proposed and developed prototype shot boundary detection method, which is exploited for the division of the long video sequence to its elementary scenes.

Moreover, it is researched in terms of perceived quality, the trade off between the constant bit rate and variable bit rate encoding schemes. According to the relative literature, the variable bit rate scheme provides higher encoding efficiency, because of the fixed quantization coefficients. The dissertation goes a step beyond the current situation and presents a quantitative approach to this generally accepted statement for various encoding bit rates and spatial resolutions.

Furthermore, it is researched for real time video streaming services the perceptual impact of the packet loss ratio of the transport network on the transmitting service. The proposed theoretical framework is totally novel in comparison to the already existing ones, because it is generic, without its validity being restricted by the encoding standard, the encoding efficiency and the content dynamics.

Finally, the dissertation discusses and presents the possibilities of combining the proposed video quality prediction method with the novel packet loss mapping model, in order to set the basis for an end-to-end video quality assessment framework for MPEG-based encoded signals.

SUBJECT AREA: Digital Video

KEYWORDS: Perceived Quality of Service (PQoS), Video Quality, MPEG, Prediction, Content dynamics, Encoding bit rate

*Με βάση μια υπόθεση απλή, ξεκίνησα για
ένα ταξίδι επικίνδυνο,
Και βλέπω κιόλας μπρος μου το κατόφλι
που οδηγεί σε νέους τόπους.*

Εμμανουήλ Καντ



Ευχαριστίες

‘Το μυστικό της επιτυχίας είναι η εμμονή στον επιδιωκόμενο στόχο’. Το γνωμικό αυτό του **Benjamin Disraeli** συνοψίζει απλά, λιτά και ταυτοχρόνως ουσιαστικά τον κατευθυντήριο άξονα που ακολούθησα για την ολοκλήρωση της παρούσας διδακτορικής διατριβής. Ενός έργου, το οποίο πέρα από οποιαδήποτε ακαδημαϊκή αξία και επιστημονική συνεισφορά, πρόκειται προπαντός για μια προσωπική κατάθεση ψυχής.

Στο πνευματικό και νοητικό ταξίδι αναζήτησης που μου πρόσφερε η εκπόνηση της διδακτορικής μου διατριβής στο εργαστήριο ψηφιακών τηλεπικοινωνιών του ΕΚΕΦΕ «Δημόκριτος», υπήρξαν πολλοί συμπαράστατες και συνοδοιπόροι που ο καθένας με το δικό του τρόπο συνεισέφερε ένα λίθο στην ανοικοδόμηση του παρόντος έργου.

Καταρχάς οφείλω να ευχαριστήσω τον επιβλέποντά μου Καθηγητή του τμήματος Πληροφορικής και Τηλεπικοινωνιών του Παν. Αθηνών Δρακούλη Μαργτάκο, ο οποίος κυριολεκτικά «αγκάλιασε» εξαρχής την προσπάθεια μου, συμβάλλοντας ουσιαστικά και καταλυτικά, με πολύτιμες συμβουλές στη διεξαχθείσα έρευνα.

Ιδιαίτερη μνεία και ευχαριστίες θα ήθελα να απευθύνω στον κύριο ερευνητή του ΕΚΕΦΕ «Δημόκριτος» Δρα Αναστάσιο Κούρτη, ο οποίος με ενέπνευσε και με ενδυνάμωσε στην ερευνητική προσπάθεια που κατέβαλα, διδάσκοντάς μου ότι οι μικρές εσωτερικές σπίθες της δυνατότητας δύνανται να μετατραπούν σε ζωηρές φλόγες επίτευξης στόχων. Επιπλέον θα ήθελα να τον ευχαριστήσω για την εμπιστοσύνη που μου έδειξε όταν με δέχτηκε στην ομάδα του εργαστηρίου που διευθύνει. Ελπίζω να στάθμια αντάξιος των προσδοκιών του.

Καθοριστική ήταν επίσης και η συνεισφορά του Καθηγητή Ιωάννη Σταυρακάκη, ο οποίος ως μέλος της τριμελούς συμβουλευτικής επιτροπής μέσω τακτικών συναντήσεων προσέφερε εποικοδομητικά σχόλια και παρατηρήσεις στη διεξαχθείσα έρευνα.

Επίσης, ευχαριστώ το φωτισμένο πνεύμα και Καθηγητή Χαράλαμπο Μάντακα, ο οποίος με τον απλό αλλά συνάμα επαναστατικό τρόπο του, μας διδάσκει το ήθος που πρέπει να έχει κάθε νέος επιστήμονας που δραστηριοποιείται στο ερευνητικό γίγνεσθαι.

Θα ήθελα επίσης να ευχαριστήσω τον Επ. Καθηγητή Χαράλαμπο Σιάνη, τον Επ. Καθηγητή Γεώργιο Κορμέντζα και τον Επ. Καθηγητή Ευάγγελο Πάλλη, των οποίων οι παρατηρήσεις τους με βοήθησαν να καθορίσω τις αποσπασμένες προτεραιότητες κάθε πτυχής της παρούσας διδακτορικής διατριβής, προσφέροντάς μου πολύτιμες συμβουλές και καθοδήγηση.

Ευχαριστίες οφείλω επίσης στους συναδέλφους ερευνητές και φίλους του ΕΚΕΦΕ «Δημόκριτος», το Δρα Γεώργιο Γαρδίτη, τον υπ. διδάκτορα Γεώργιο Ξυλούρη, τον υπ. διδάκτορα Θωμά Πλιάκα, τον υπ. διδάκτορα Ιωάννη Γιαννουλάκη, τον υπ. διδάκτορα Μανώλη Καφεντζάκη, το Μιχάλη Μπατιστάτο, το Νικόλα Ζώτο και το Σπύρο Ορφανό με τους οποίους μοιραστήκαμε ιδέες, απόψεις και προτάσεις για ποικίλα θέματα που πολλές φορές οδήγησαν σε πολώρες συζητήσεις και ενίοτε έντονες αντιπαραθέσεις.

Επιπλέον, πολλοί υπήρξαν οι δάσκαλοι και καθοδηγητές μου, που ο καθένας με διαφορετικό τρόπο συνέβαλε στη πνευματική εξέλιξη και νοητική πρόοδό μου. Θέλω ιδιαίτερα να ευχαριστήσω τη Μ. Βαβούτη, το Δ. Μαλλούχο, τη Ε. Τσαβλή, το Θ. Φλωρόπουλο, το Π. Σχοινέζο, την Α. Τσακαλάκη και τον Π. Μηλιό που με ενέπνευσαν και αφύπνισαν τη ψυχή μου. Τους Ν. Σάββα, Θ. Παλαιοθόδωρο, Καψάλη, Α. Φελιμέργια, Κ. Σκουτέλα, Σ. Αντωνιάδου και Α. Αποστολοπούλου-Αρβανιτοπούλου που με δίδαξαν να σκέφτομαι. Επίσης τους συναδέλφους Σ. Φωκά και Σ. Κονίδη, με τους οποίους μοιραζόμαστε κοινές φιλοσοφικές και κοινωνικές αντιλήψεις.

Εν κατακλείδι, θα ήθελα να ευχαριστήσω τους γονείς μου Νότα και Γιώργο Κουμαρά, τον αδερφό μου Βάιο Κουμαρά και τη σύζυγο του Αναστασία Στεργίου, τις φίλες και φίλους μου, αλλά προπαντός τα ανίψια μου, τα οποία μέσω της παιδικής ματιάς τους με διδάσκουν ότι παρόλο που η επιστήμη ίσως σε κάνει σοφότερο, η πραγματική φώτιση βρίσκεται μέσα μας.

Αθήνα, Αύγουστος 2007

Χαρίλαος Κουμαράς

Περιεχόμενα

<i>Ευρετήριο Σχημάτων</i>	17
<i>Ευρετήριο Πινάκων</i>	19
ΚΕΦΑΛΑΙΟ 1	21
<i>Ανασκόπηση της Διδακτορικής Διατριβής</i>	21
<i>Ερευνητικό Αντικείμενο</i>	21
<i>Περιγραφή Παρούσης Κατάστασης– Συνεισφορά της Διδακτορικής Διατριβής</i>	21
<i>Δομή της Διδακτορικής Διατριβής</i>	25
ΚΕΦΑΛΑΙΟ 2	31
<i>Πρότυπα Κωδικοποίησης Ψηφιακού Βίντεο</i>	31
<i>Από το αναλογικό στο ψηφιακό σήμα βίντεο: Η ανάγκη για συμπίεση</i>	31
<i>Πρότυπα Κωδικοποίησης Βίντεο</i>	31
<i>Ομάδες εργασίας στην Ανάπτυξη Προτύπων Ψηφιακού Βίντεο</i>	31
<i>Γένεση και διαδρομή των προτύπων κωδικοποίησης ψηφιακού βίντεο</i>	35
H.261.....	35
MPEG-1.....	36
MPEG-2/H.262.....	37
MPEG-3.....	38
MPEG-4.....	38
MPEG-4 AVC Part 10 / H.264.....	44
<i>Συμπεράσματα του Κεφαλαίου</i>	48
ΚΕΦΑΛΑΙΟ 3	49
<i>Θεμελιώδεις Αρχές Κωδικοποίησης Ψηφιακού Βίντεο</i>	49
<i>Χωρικός και Χρονικός Πλεονασμός</i>	49
<i>Αξιοποίηση Χρονικού Πλεονασμού κατά την Κωδικοποίηση</i>	52
<i>Αξιοποίηση Χωρικού Πλεονασμού κατά την Κωδικοποίηση</i>	56
<i>Διακριτός Μετασχηματισμός Συνημιτόνου</i>	57
<i>Κβάντιση και Σάρωση των Συντελεστών DCT</i>	60
<i>Εντροπική Κωδικοποίηση των Δεδομένων</i>	63
Huffman Coding.....	63
Arithmetic Coding.....	65
<i>Απλοποιημένο Block Διάγραμμα ενός Ψηφιακού Κωδικοποιητή</i>	66
<i>Συμπεράσματα του Κεφαλαίου</i>	67
ΚΕΦΑΛΑΙΟ 4	69
<i>Συνοπτική Βιβλιογραφική Ανασκόπηση των Μεθόδων και Τεχνικών Εκτίμησης της Ποιότητας Βίντεο</i>	69
<i>Ποιοτική Υποβάθμιση Βίντεο Λόγω Συμπίεσης/Κωδικοποίησης</i>	69
<i>Μέθοδοι Εκτίμησης της Αντιληπτής Ποιότητας ψηφιακού Βίντεο</i>	72
<i>Υποκειμενικές Μέθοδοι Εκτίμησης της Αντιληπτής Ποιότητας Βίντεο</i>	73

Υποκειμενικές Μέθοδοι Μονού Οπτικού Ερεθίσματος.....	73
Υποκειμενικές Μέθοδοι Διπλού Οπτικού Ερεθίσματος.....	76
Επιλογή της κατάλληλης μεθόδου διπλού ή μονού ερεθίσματος.....	79
Εναλλακτικές Μέθοδοι Εκτίμησης.....	80
Αντικειμενικές Μέθοδοι Εκτίμησης της Αντιληπτής Ποιότητας Βίντεο.....	81
Η Συνεισφορά της Διατριβής: Από την Εκτίμηση στην Πρόβλεψη της Ποιότητας.....	88
ΚΕΦΑΛΑΙΟ 5.....	89
Στατιστική Ανάλυση και Μοντελοποίηση Βίντεο H.264 Μεταβλητού Ρυθμού.....	89
Εισαγωγή.....	89
Σχετική Βιβλιογραφία στην Στατιστική Ανάλυση Κίνησης Ψηφιακού Βίντεο.....	90
Το Πρότυπο H.264: Βασικές Ιδιότητες και Παρούσα Κατάσταση.....	90
Στατιστική Ανάλυση Κωδικοποιημένου Σήματος Σύμφωνα με το Πρότυπο H.264.....	92
- Ανάλυση σε Επίπεδο Καρέ.....	92
- Ανάλυση σε Επίπεδο I/B/P.....	95
Μοντέλα Παραγωγής Συνθετικής Ροής H.264.....	99
- Συνοπτική Βιβλιογραφική Ανασκόπηση.....	99
- Auto-Regressive Models.....	100
- Μαρκοβιανά Μοντέλα.....	101
- Μοντέλα Μακροπρόθεσμης Εξάρτησης.....	101
- Υβριδικά Μοντέλα Markov και AR.....	101
Προτεινόμενο Τροποποιημένο Μαρκοβιανό Μοντέλο.....	102
Συμπεράσματα του Κεφαλαίου.....	105
ΚΕΦΑΛΑΙΟ 6.....	107
Σύγκριση της Αντιληπτής Ποιότητας Βίντεο Σταθερού και Μεταβλητού Ρυθμού.....	107
Μέθοδοι Κωδικοποίησης Σταθερού και Μεταβλητού Ρυθμού (CBR-VBR).....	107
Μεθοδολογία Σύγκρισης Αντιληπτής Ποιότητας Κωδικοποίησης VBR και CBR.....	108
- PSNR.....	109
- SSIM.....	109
- Delta.....	109
- DVQ.....	110
Πειραματική Διαδικασία Λήψης Μετρήσεων Αντιληπτής Ποιότητας CBR-VBR.....	110
Συμπεράσματα του Κεφαλαίου.....	119
ΚΕΦΑΛΑΙΟ 7.....	121
Μέθοδος Πρόβλεψης της Αντιληπτής Ποιότητας για Βίντεο Βραχείας Διάρκειας.....	121
Η Ανάγκη για Μέθοδο Πρόβλεψης της Αντιληπτής Ποιότητας Συμπεσιμένου Βίντεο.....	121
Λογισμικό Εργαλείο Μέτρησης της Αντιληπτής Ποιότητας.....	123
Διακύμανση του MPQoS Συναρτήσει του Ρυθμού Κωδικοποίησης.....	126
Εκθετική προσέγγιση του MPQoS Συναρτήσει του Ρυθμού Κωδικοποίησης.....	132
Ταχεία Εκτίμηση των Στοιχείων του Διανύσματος Ποιότητας.....	135
Επέκταση της Μεθόδου σε Βίντεο Σύντομης Διάρκειας με Ανομοιογενές Περιεχόμενο.....	139

Συμπεράσματα του Κεφαλαίου.....	142
ΚΕΦΑΛΑΙΟ 8.....	145
<i>Επέκταση της Προτεινόμενης Μεθόδου Πρόβλεψης της Αντιληπτής Ποιότητας σε Βίντεο Μακράς Διαρκείας.....</i>	145
<i>Απαιτήσεις για Επέκταση της Προτεινόμενης Μεθόδου.....</i>	145
<i>Τεχνική Κατάτμησης των Σκηνών σε DCT Βίντεο Χαμηλού Ρυθμού.....</i>	146
<i>Σύντομη Βιβλιογραφική Ανασκόπηση των Τεχνικών Κατάτμησης Σκηνών.....</i>	147
<i>Η Προτεινόμενη Μέθοδος Κατάτμησης.....</i>	149
<i>Ενσωμάτωση της Μεθόδου Κατάτμησης στην Τεχνική Πρόβλεψης της Ποιότητας.....</i>	152
<i>Σχολιασμός στην ακρίβεια και αποδοτικότητα της εκτεταμένης μεθόδου.....</i>	154
<i>Πειραματική εφαρμογή της εκτεταμένης μεθόδου.....</i>	155
<i>Συμπεράσματα του Κεφαλαίου.....</i>	156
ΚΕΦΑΛΑΙΟ 9.....	159
<i>Διερεύνηση και Μοντελοποίηση της Επίδρασης του Ρυθμού Απώλειας Πακέτων του Δικτύου Μεταφοράς στην Υποβάθμιση της Αντιληπτής Ποιότητας.....</i>	159
<i>Υποβάθμιση της Ποιότητας κατά τη Μετάδοση του Βίντεο.....</i>	159
<i>Συνοπτική Βιβλιογραφική Ανασκόπηση της Επίπτωσης Δικτυακών Παραμέτρων στην Ποιότητα του Μεταδιδόμενου Βίντεο.....</i>	160
<i>Επίδραση της Απώλειας Πακέτων στο Αποκωδικοποιημένο Βίντεο.....</i>	161
<i>Μοντελοποίηση της Απώλειας Πακέτων στο Αποκωδικοποιημένο Βίντεο.....</i>	163
<i>Το Αναμενόμενο Πλήθος Αποκωδικοποιημένων Καρέ I (N_{dec-I}).....</i>	164
<i>Το Αναμενόμενο Πλήθος Αποκωδικοποιημένων Καρέ P (N_{dec-P}).....</i>	164
<i>Το Αναμενόμενο Πλήθος Αποκωδικοποιημένων Καρέ B (N_{dec-B}).....</i>	164
<i>Πειραματική Επαλήθευση του Προτεινόμενου Μοντέλου.....</i>	165
<i>Επέκταση του Μοντέλου για Πρόβλεψη της Διατερματικής Αντιληπτής Ποιότητας.....</i>	167
<i>Συμπεράσματα του Κεφαλαίου.....</i>	168
ΚΕΦΑΛΑΙΟ 10.....	169
<i>Συμπεράσματα και Συνεισφορά της Διατριβής.....</i>	169
<i>Σύνοψη των Προτεινόμενων Αναλύσεων, Προτύπων Μεθόδων και Μοντέλων.....</i>	169
<i>Ερευνητικά Ζητήματα που Θέτονται για Περαιτέρω Διερεύνηση.....</i>	171
ΑΝΑΦΟΡΕΣ.....	173
ΒΙΟΓΡΑΦΙΚΟ ΤΟΥ ΣΥΓΓΡΑΦΕΑ.....	183

Ευρετήριο Σχημάτων

Σχήμα 2.1	Τα συστατικά μιας ροής MPEG-4.....	39
Σχήμα 2.2	Παράδειγμα κωδικοποίησης H.264 μεταβλητού block	45
Σχήμα 3.1	Χωρικός και Χρονικός Πλεονασμός.....	50
Σχήμα 3.2	Καρέ που διακρίνονται οι χωρικές περιοχές με συχνοτικό πλεονασμό.....	51
Σχήμα 3.3	Συσχετίσεις μεταξύ των καρέ που δομούν ένα GOP	53
Σχήμα 3.4	Παράδειγμα χρήσης blocks μεταβλητού μεγέθους στο πρότυπο H.264	54
Σχήμα 3.5	Απεικόνιση των διανυσμάτων κίνησης σε ένα καρέ της ταινίας Spiderman II	55
Σχήμα 3.6	Παράδειγμα υπολογισμού καρέ εντροπικών διαφορών.....	55
Σχήμα 3.7	Intra prediction τεχνικές του προτύπου H.264	56
Σχήμα 3.8	Παράδειγμα της αντιληπτής απόδοσης του αντιστρόφου DCT	58
Σχήμα 3.9	Τυπικό παράδειγμα των τιμών φωτεινότητας ενός block.....	59
Σχήμα 3.10	Το block του σχ.3.9 μετά την εφαρμογή του DCT	59
Σχήμα 3.11	Οι συχνότητες βάσης του 8x8 DCT	60
Σχήμα 3.12	Οι τιμές του block μετά τη χρήση κβάντισης.....	62
Σχήμα 3.13	Σάρωση των κβαντισμένων συντελεστών DCT.....	62
Σχήμα 3.14	Παράδειγμα δέντρου συχνοτήτων εμφάνισης των συμβόλων	64
Σχήμα 3.15	Γενικευμένο block διάγραμμα κωδικοποίησης ψηφιακού βίντεο.....	66
Σχήμα 3.16	Οι τιμές φωτεινότητας του αποκωδικοποιημένου block του σχήματος 3.9	67
Σχήμα 4.1	Παράδειγμα εμφάνισης τεχνουργήματος Moiré σε δύο σύνολα παραλλήλων	71
Σχήμα 4.2	Γραφική αναπαράσταση της μεθόδου SSM/ACR (ITU-T Rec. P.910)	74
Σχήμα 4.3	Οδηγίες προς τους θεατές και κλίμακα βαθμολόγησης της μεθόδου ACR.	75
Σχήμα 4.4	Γραφική αναπαράσταση της μεθόδου DSIS (ITU-T Rec P.910).....	77
Σχήμα 4.5	Οδηγίες προς τους θεατές και κλίμακα βαθμολόγησης της μεθόδου DSIS/DCR.	78
Σχήμα 4.6	Γραφική αναπαράσταση της μεθόδου DSIS (ITU-T Rec P.910).....	78
Σχήμα 4.7	Οδηγίες προς τους θεατές και κλίμακα βαθμολόγησης της μεθόδου DSIS/DCR.	79
Σχήμα 4.8	Block διάγραμμα των μεθόδων πλήρους αναφοράς πολλαπλών καναλιών.....	82
Σχήμα 4.9	Block διάγραμμα των μεθόδων πλήρους αναφοράς μονού καναλιού.....	84
Σχήμα 4.10	Block διάγραμμα των μεθόδων μειωμένης αναφοράς.....	86
Σχήμα 5.1	Στατιστική Ανάλυση σε επίπεδο καρέ	93
Σχήμα 5.2	Συνολικό αριθμός Macro-Blocks που κωδικοποιήθηκαν ως Intra.....	93
Σχήμα 5.3	Ιστογράμματα μεγέθους των καρέ (α), (β) και οι αυτοσυσχετίσεις αυτών (c) και (d) 94	
Σχήμα 5.5	Πειραματικά ιστογράμματα μεγέθους καρέ με τις αντίστοιχες Γάμμα κατανομές	97
Σχήμα 5.6	Αντιπροσωπευτικές Αυτοσυσχετίσεις	98
Σχήμα 5.7	Block διάγραμμα του προτεινόμενου μοντέλου.....	104
Σχήμα 5.8	Q-Q Διαγράμματα δειγμάτων διάρκειας 3 λεπτών	104
Σχήμα 5.9	Αυτοσυσχέτιση πραγματικών και συνθετικών I καρέ μακράς/βραχείας εξάρτησης.105	
Σχήμα 6.1	Το προτεινόμενο επίπεδο ταξινόμησης της δυναμικής του περιεχομένου	111
Σχήμα 6.2	Απεικόνιση της χωρικής και χρονικής δυναμικής των πέντε βίντεο αναφοράς.....	113
Σχήμα 6.3	Ο λόγος PQVBR/PQCBR του SNR.....	114
Σχήμα 6.4	Ο λόγος PQVBR/PQCBR του SSIM.....	114
Σχήμα 6.5	Ο λόγος PQVBR/PQCBR του DVQ.....	115
Σχήμα 6.6	Ο λόγος PQVBR/PQCBR του Delta.....	115
Σχήμα 6.7	Σύγκριση του PQVBR/PQCBR σε δύο καρέ του βίντεο Suzie	116
Σχήμα 6.8	Ο λόγος PQVBR/PQCBR για κωδικοποίηση ASP MPEG-4/CIF και οι αντίστοιχες αποκλεισεις μεταξύ πειραματικών και θεωρητικών τιμών.	117
Σχήμα 7.1	Σύγκριση αποτελεσμάτων του QMS και SSCQE.	124

Σχήμα 7.2	Πειραματικά αποτελέσματα του QMS	125
Σχήμα 7.3	MPQoS συναρτήσει του ρυθμού κωδικοποίησης για MPEG-4 ASP CIF.....	128
Σχήμα 7.4	MPQoS συναρτήσει του ρυθμού κωδικοποίησης για MPEG-4 ASP QCIF.....	128
Σχήμα 7.5	MPQoS συναρτήσει του ρυθμού κωδικοποίησης για MPEG-2 CIF	129
Σχήμα 7.6	MPQoS συναρτήσει του ρυθμού κωδικοποίησης για H.264 Baseline CIF.....	129
Σχήμα 7.7	MPQoS για βίντεο χαμηλής δυναμικής (Talk Show) και κυμαινόμενης διάρκειας..	131
Σχήμα 7.8	MPQoS για βίντεο υψηλής δυναμικής (Sports) και κυμαινόμενης διάρκειας.....	131
Σχήμα 7.9	Πειραματικές και θεωρητικές καμπύλες MPQoS για MPEG-4/CIF	133
Σχήμα 7.10	Πειραματικές και θεωρητικές καμπύλες MPQoS για MPEG-2/CIF	134
Σχήμα 7.11	Πειραματικά προσδιορισμένη διακύμανση των QV στοιχείων συναρτήσει της δυναμικής του περιεχομένου για MPEG-4/CIF	136
Σχήμα 7.12	Καμπύλες MPQoS για βραχείας διάρκειας βίντεο ετερογενούς περιεχομένου	140
Σχήμα 7.13	Πειραματικές και θεωρητικές καμπύλες MPQoS επιδεικνύοντας την αποδοτικότητα της προτεινόμενης μεθόδου.....	141
Σχήμα 7.14	Συνάρτηση Πυκνότητας Πιθανότητας του στιγμιαίου PQoS.....	142
Σχήμα 8.1	Διακύμανση του φαινομένου τετραγώνισης για απόσπασμα του Spiderman II.....	151
Σχήμα 8.3	Επέκταση της μεθόδου πρόβλεψης σε βίντεο μακράς διάρκειας.....	153
Σχήμα 8.4	Στοιχειώδεις σκηνές του συνθετικού βίντεο	155
Σχήμα 9.1	Γραφική απεικόνιση του GOP(12,3).....	162
Σχήμα 9.2	Σχηματική αναπαράσταση της ομοιόμορφης απώλειας πακέτων.....	166
Σχήμα 9.3	Πειραματική και Θεωρητική αναπαράσταση του ποσοστού αποκωδικοποιημένων καρέ Q συναρτήσει του ρυθμού απώλειας πακέτων.....	166

Ευρετήριο Πινάκων

Πίνακας 2.1	Εξέλιξη της δημοσίευσης του προτύπου H.264.....	45
Πίνακας 2.2	Σύγκριση των χαρακτηριστικών του MPEG-4 και H.264	46
Πίνακας 2.3	Συνοπτική παρουσίαση των προφίλ του H.264 (Απρίλιος 2007).....	48
Πίνακας 3.1	Παραδείγματα κβάντισης.....	61
Πίνακας 3.2	Παράδειγμα Huffman Coding.....	64
Πίνακας 3.3	Παράδειγμα Arithmetic Coding I	65
Πίνακας 3.4	Παράδειγμα Arithmetic Coding II.....	65
Πίνακας 3.5	Παράδειγμα Αποκωδικοποίησης Arithmetic Coding.....	66
Πίνακας 5.1	Στατιστικά ανά είδος καρτέ για το κωδικοποιημένο βίντεο	95
Πίνακας 5.2	Συντελεστές απόκλισης για ποικίλες παραμέτρους κβάντισης.....	96
Πίνακας 5.3	Στατιστικά της Γάμμα μοντελοποίησης για ποικίλα επίπεδα κβάντισης.....	96
Πίνακας 5.4	Οι παράμετροι του μοντέλου της συνάρτησης αυτοσυσχέτισης	98
Πίνακας 5.5	Συντελεστές συσχέτισης διαφορετικών τύπων καρτέ.....	103
Πίνακας 6.1	Οι πειραματικές τιμές του γενικευμένου λόγου CIF PQVBR/PQCVR	117
Πίνακας 6.2	Οι πειραματικές τιμές του γενικευμένου λόγου QCIF PQVBR/PQCVR	118
Πίνακας 7.1	Τα βίντεο αναφοράς της προτεινόμενης μεθόδου πρόβλεψης της ποιότητας	126
Πίνακας 7.2	Τα στοιχεία των διανυσμάτων ποιότητας για MPEG-4 CIF και QCIF.....	134
Πίνακας 7.3	Παράδειγμα προσδιορισμού των ρυθμών κωδικοποίησης.....	138
Πίνακας 7.4	Βίντεο με συγκεκριμένο λόγο υψηλής και χαμηλής δυναμικής του περιεχομένου	139
Πίνακας 7.5	Μέσο σφάλμα εκτίμησης για τα ανομοιογενή βίντεο του σχήματος 7.12.....	140
Πίνακας 7.6	Χαρακτηριστικά των βραχείας διάρκειας βίντεο ετερογενούς περιεχομένου	140
Πίνακας 8.1	Σύγκριση της απόδοσης της προτεινόμενης μεθόδου για απότομες αλλαγές σκηνών..	152
Πίνακας 8.2	Σύγκριση της απόδοσης της προτεινόμενης μεθόδου για σταδιακές αλλαγές σκηνών	152
Πίνακας 9.1	Επεξήγηση των συμβόλων του προτεινόμενου μοντέλου.....	163
Πίνακας 9.2	Στατιστικά του βίντεο Alladin	166

ΚΕΦΑΛΑΙΟ 1

Ανασκόπηση της Διδακτορικής Διατριβής

Ερευνητικό Αντικείμενο

Η ραγδαία ανάπτυξη των υπηρεσιών πολυμέσων τα τελευταία χρόνια έχει προκαλέσει μια σημαντική κινητικότητα στην ερευνητική κοινότητα για σχεδιασμό, ανάπτυξη και υλοποίηση τεχνικών κωδικοποίησης και συμπίεσης δεδομένων που παράγονται από εφαρμογές πολυμέσων. Πιο συγκεκριμένα, η εκτεταμένη διάδοση και χρήση υπηρεσιών ψηφιακής κινούμενης εικόνας (βίντεο), είχε ως αποτέλεσμα την ανάπτυξη μεθόδων, τεχνικών και προτύπων για τη κωδικοποίηση ασυμπίεστων ψηφιακών σημάτων βίντεο, μέσω της οποίας επιτυγχάνεται μείωση του όγκου δεδομένων με αντίκτυπο την υποβάθμιση της αντιληπτής ποιότητας του τελικού συμπιεσμένου σήματος.

Η παρούσα διδακτορική διατριβή εστιάζεται ερευνητικά στο χώρο της εκτίμησης του επιπέδου της ποιότητας κωδικοποιημένου βίντεο. Πιο συγκεκριμένα, εισάγεται, περιγράφεται, δοκιμάζεται και προτείνεται πρωτότυπη μέθοδος εκτίμησης της αντιληπτής ποιότητας βίντεο, η οποία προβλέπει συναρτήσει της δυναμικής του περιεχομένου και των παραμέτρων κωδικοποίησης την τελική ποιότητα του κωδικοποιημένου βίντεο.

Συνεπώς, η συνεισφορά της παρούσας διατριβής στον τομέα εκτίμησης της αντιληπτής ποιότητας βίντεο δεν περιορίζεται στην απλή παρουσίαση μιας ακόμη μεθόδου εκτίμησης της ποιότητας. Αντιθέτως, η διατριβή καινοτομεί μέσω της προτεινόμενης μεθόδου, εισάγοντας καινούργια κατηγορία τεχνικών εκτίμησης, οι οποίες δεν περιορίζονται απλά σε εκτιμητική δράση του ήδη υπάρχοντος συμπιεσμένου σήματος, αλλά πραγματοποιούν πρόβλεψη της τελικής αντιληπτής ποιότητας ενός ψηφιακού βίντεο, πριν η διαδικασία κωδικοποίησης και συμπίεσης λάβει χώρα.

Περιγραφή Παρούσης Κατάστασης– Συνεισφορά της Διδακτορικής Διατριβής

Οι πολυμεσικές εφαρμογές και υπηρεσίες κατέχουν τη σήμερα πρωτεύοντα ρόλο στα τηλεπικοινωνιακά και υπολογιστικά δίκτυα. Ανάμεσα στις ποικίλες πολυμεσικές εφαρμογές, οι υπηρεσίες ψηφιακού video (μετάδοση διαδοχικών εικόνων με ήχο) έχουν ανακηρυχθεί ως οι πιο δημοφιλείς και διαδεδομένες εφαρμογές στις επιλογές των τελικών χρηστών.

Μέσα σε αυτό το πλαίσιο των παρεχόμενων ψηφιακών οπτικοακουστικών υπηρεσιών, η βέλτιστη αξιοποίηση του διαθέσιμου εύρους ζώνης του εκάστοτε καναλιού μετάδοσης παραμένει πρωτεύον ζήτημα στην επιστήμη αλλά και στη βιομηχανία των τηλεπικοινωνιών. Είναι ευρέως γνωστό ότι οι υπηρεσίες ψηφιακού βίντεο καταναλώνουν μεγάλο μερίδιο της διαθέσιμης χωρητικότητας ενός τηλεπικοινωνιακού καναλιού, εξαιτίας της πολυπλοκότητας της πληροφορίας που μεταφέρουν. Αξίζει να σημειωθεί ότι κανένα σημερινό τηλεπικοινωνιακό δίκτυο δεν μπορεί να υποστηρίξει σε πραγματικό χρόνο το ρυθμό μετάδοσης ασυμπίεστης ψηφιακής οπτικοακουστικής υπηρεσίας (raw video data), ακόμη και σήματος χαμηλής χρονικής και χωρικής ανάλυσης (δηλαδή μικρό μέγεθος καρέ και χαμηλός ρυθμός καρέ ανά δευτερόλεπτο). Το γεγονός αυτό, σε συνδυασμό με τη συνεχώς

αυξανόμενη ζήτηση του καταναλωτικού κοινού για εφαρμογές ψηφιακού βίντεο, έχει καταστήσει αναγκαία την αξιοποίηση τεχνικών κωδικοποίησης της μεταδιδόμενης οπτικοακουστικής πληροφορίας (video coding). [1]

Η κωδικοποίηση ψηφιακού βίντεο αναφέρεται στη διαδικασία συμπίεσης της αρχικής ψηφιακής πληροφορίας με στόχο την μείωση του όγκου των δεδομένων, ώστε να είναι δυνατή η διακίνηση και η μετάδοση του ψηφιακού βίντεο μέσω τηλεπικοινωνιακών δικτύων περιορισμένου εύρους ζώνης, καθώς και η αποθήκευσή του σε οπτικά μέσα περιορισμένης χωρητικότητας. Οι αλγόριθμοι συμπίεσης βίντεο αξιοποιούν το γεγονός ότι ένα οπτικοακουστικό σήμα αποτελείται από διαδοχικές εικόνες/καρέ, οι οποίες παρουσιάζουν σημαντικές ομοιότητες μεταξύ τους στο πεδίο του χρόνου, χώρου και συχνότητας (temporal, spatial and frequency-domain similarities) [2]. Οι ομοιότητες αυτές εμπεριέχουν πλεονάζουσα επαναλαμβανόμενη πληροφορία, η οποία αν αφαιρεθεί μπορεί να επιτευχθεί σημαντική συμπίεση των δεδομένων βίντεο, θυσιάζοντας βεβαίως ένα ποσοστό οπτικής πληροφορίας, η οποία δεν γίνεται άμεσα αντιληπτή από το μηχανισμό όρασης του ανθρώπου, το οποίο εμφανίζει μικρή ευαισθησία και διακριτική ικανότητα σε αυτού του είδους την απολεσθείσα οπτική πληροφορία [3].

Βεβαίως, ανάλογα με τον επιθυμητό λόγο συμπίεσης, το είδος της πληροφορίας που παραλείπεται και την διαδικασία κωδικοποίησης του βίντεο, ενδέχεται να προκληθεί μη αντιστρέψιμη ποιοτική υποβάθμιση του αρχικού οπτικοακουστικού περιεχομένου και για αυτό το λόγο οι ποικίλοι κωδικοποιητές βίντεο διακρίνονται σε απωλεστικούς (lossy codecs) και μη απωλεστικούς (lossless codecs). Η ειδοποιός διαφορά ανάμεσα στις δύο κατηγορίες είναι η χρήση κβαντιστή (quantizer) κατά τη διαδικασία κωδικοποίησης, η οποία μοιραία οδηγεί σε οριστική απώλεια μέρους της αρχικής πληροφορίας, αφού η κβάντιση πρόκειται για μερικώς αντιστρέψιμη διαδικασία [1].

Όσον αναφορά το λόγο συμπίεσης, οι απωλεστικές τεχνικές κωδικοποίησης πετυχαίνουν υψηλότερο λόγο συμπίεσης του αρχικού οπτικοακουστικού σήματος, με κόστος την απώλεια μέρους της αρχικής πληροφορίας και εν τέλει την ποιοτική υποβάθμιση του κωδικοποιημένου/συμπιεσμένου σήματος συγκριτικά με το αρχικό. Αντίστοιχα, οι μη απωλεστικοί αλγόριθμοι διατηρούν την αρχική ποιότητα του σήματος χωρίς όμως να είναι σε θέση να επιτύχουν σημαντική συμπίεση των δεδομένων.

Από τις δύο προαναφερθείσες κατηγορίες κωδικοποιητών, οι απωλεστικοί κωδικοποιητές έχουν ευρέως επικρατήσει στις τηλεπικοινωνιακές εφαρμογές, οι οποίοι αν και υποβαθμίζουν ποιοτικά το αρχικό οπτικοακουστικό σήμα με τρόπο μη αντιστρέψιμο, προτιμούνται γιατί παρέχουν σημαντικά υψηλότερο λόγο συμπίεσης του αρχικού σήματος συγκριτικά με τους μη απωλεστικούς.

Εξαιτίας της ποιοτικής υποβάθμισης που εισάγεται κατά τη διαδικασία απωλεστικής κωδικοποίησης ενός οπτικοακουστικού σήματος, προέκυψε ερευνητικά η έννοια της εκτίμησης της αντιληπτής ποιότητας ψηφιακά κωδικοποιημένου βίντεο συναρτήσει των παραμέτρων κωδικοποίησης. Πιο συγκεκριμένα, στις εφαρμογές οπτικοακουστικού περιεχομένου που έχουν κωδικοποιηθεί/συμπιεστεί σύμφωνα με κάποια απωλεστική μέθοδο είναι πολύ χρήσιμο να καθοριστεί η ποιότητα με την οποία ο τελικός χρήστης λαμβάνει την υπηρεσία (Quality of Service – QoS). Ειδικότερα, η έννοια της ποιότητας σε εφαρμογές ψηφιακού βίντεο αναφέρεται στην ποιότητα του οπτικοακουστικού περιεχομένου, όπως γίνεται αντιληπτή από τον θεατή και συνήθως χαρακτηρίζεται ως αντιληπτή ποιότητα υπηρεσίας (Perceived Quality of Service – PQoS) [4].

Οι κυριότερες μέθοδοι προσδιορισμού της αντιληπτής ποιότητας υπηρεσίας των προσφερόμενων ψηφιακών βίντεο που έχουν προταθεί στην διεθνή βιβλιογραφία ταξινομούνται κυρίως σε δύο κατηγορίες: Τις υποκειμενικές και αντικειμενικές τεχνικές [3].

Στην πρώτη κατηγορία, ένα ακροατήριο καλείται να παρακολουθήσει, σε ελεγχόμενο περιβάλλον, ένα σύνολο ψηφιακών video κωδικοποιημένων με διαφορετικούς τρόπους και να βαθμολογήσει υποκειμενικά την ποιότητα του βίντεο που προβάλλεται. Στη διαδικασία εκτίμησης της ποιότητας, αναλόγως με τη μέθοδο που θα ακολουθηθεί, ενδέχεται να είναι διαθέσιμο το αρχικό ασυμπιεστο σήμα και να προβάλλεται στο κοινό ως σήμα αναφοράς. Η διαδικασία της υποκειμενικής εκτίμησης της ποιότητας ψηφιακού βίντεο μειονεκτεί εξαιτίας του γεγονότος ότι είναι πολύπλοκη, χρονοβόρα και οικονομικά μη συμφέρουσα, αφού απαιτεί ειδικό εξοπλισμό και τη συγκέντρωση μιας ομάδας ανθρώπων που θα συμμετέχουν στη διαδικασία. [3] [5] [6]

Στη δεύτερη περίπτωση χρησιμοποιούνται τεχνικές αντικειμενικής εκτίμησης του video, στις οποίες μαθηματικά μοντέλα προσομοιώνουν τις αντίστοιχες υποκειμενικές μετρήσεις, βασισμένες σε κριτήρια που μπορούν να μετρηθούν με αντικειμενικό τρόπο. Οι αντικειμενικές μέθοδοι μπορούν να ταξινομηθούν ως πλήρης αναφοράς (full reference) [7] [8], μειωμένης αναφοράς (reduced reference) [9] ή μη-αναφοράς (no-reference) [7] [10] ανάλογα με την απαίτηση ύπαρξης του αρχικού ασυμπιεστού οπτικοακουστικού σήματος στη διαδικασία εκτίμησης της ποιότητας.

Κατά συνέπεια, στις μεθόδους πλήρους αναφοράς εκτελείται μια άμεση σύγκριση μεταξύ των διαδοχικών καρέ του συμπιεσμένου και ασυμπιεστού οπτικοακουστικού σήματος, με στόχο την ποσοτικοποίηση της ποιοτικής υποβάθμισης που υπέστη το βίντεο λόγω κωδικοποίησης. Οι πρώτες υλοποιήσεις πλήρους αναφοράς προήλθαν κυρίως από το χώρο της ψηφιακής επεξεργασίας εικόνας (digital image processing), όπου χρησιμοποιήθηκαν απλές μαθηματικές εκφράσεις υπολογισμού σφάλματος μεταξύ των δύο σημάτων, όπως: Ο λόγος σήματος προς θόρυβο (PSNR), το μέσο απόλυτο άθροισμα των διαφορών (MSAD) και το μέσο τετραγωνικό σφάλμα (MSE) [11]:

$$PSNR = 10 \log_{10} \frac{L^2}{MSE} \quad MSAD = \frac{1}{N} \sum_{i=1}^N |x_i - y_i| \quad MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2$$

Όπου το L είναι το δυναμικό εύρος των pixels που απαρτίζουν τα καρέ ενός βίντεο (για 8bits/pixel ασπρόμαυρο σήμα το L ισοδυναμεί με 255) και N είναι ο συνολικός αριθμός των pixels που απαρτίζουν το κάθε καρέ του βίντεο, ενώ x_i και y_i τα καρέ του ασυμπιεστού και του συμπιεσμένου σήματος αντίστοιχα.

Αρχικά, τα απλά προαναφερθέντα μαθηματικά μοντέλα χρησιμοποιήθηκαν ευρέως για την εκτίμηση της αντιληπτής ποιότητας βίντεο και τη σύγκριση της απόδοσης διάφορων τεχνικών και προτύπων συμπίεσης και κωδικοποίησης.

Παρόλα αυτά, η μη ταύτιση των αποτελεσμάτων των μετρικών αυτών με τις αντίστοιχες υποκειμενικές διαδικασίες μέτρησης της ποιότητας βίντεο έγινε αντικείμενο έντονης κριτικής, επιδεικνύοντας ότι δεν παρέχουν αξιόπιστες μετρήσεις της αντιληπτής ποιότητας, αφού ποσοτικοποιούσαν τεχνουργήματα του κωδικοποιημένου σήματος, τα οποία όμως δεν γίνονταν αντιληπτά από το θεατή [5, 12].

Συνεπώς, βασισμένοι στην παρατήρηση ότι η απλή μαθηματική μοντελοποίηση των διαφορών του συμπιεσμένου και ασυμπιεστού σήματος δεν παρέχει αξιόπιστα αποτελέσματα για την αντιληπτή ποιότητα, αναπτύχθηκαν πιο πολύπλοκες μέθοδοι

πλήρους αναφοράς, οι οποίες με τη χρήση κατάλληλων φίλτρων προσομοίωναν τα χαρακτηριστικά του ανθρωπίνου συστήματος όρασης (Human Visual System) [13]. Ένα γενικό μοντέλο αυτών των μεθόδων θα μπορούσε να περιγραφεί από την αξιοποίηση συναρτήσεων ευαισθησίας της αντίθεσης (Contrast Sensitivity Functions) σε κάθε καρέ του συμπιεσμένου και ασυμπιέστου σήματος βίντεο, σε συνδυασμό με αποσύνθεση του αποτελέσματος σε αντίστοιχο πλήθος συχνοτικών καναλιών (channel decomposition). Στη συνέχεια, εφαρμόζοντας βασικές συναρτήσεις υπολογισμού σφάλματος διακριτά σε κάθε ένα από τα προαναφερθέντα κανάλια, προκύπτει ο υπολογισμός ποικίλων επιμέρους σφαλμάτων, τα οποία εν συνεχεία κανονικοποιούνται (error normalization) και ανάλογα με τα χαρακτηριστικά του ανθρωπίνου συστήματος όρασης –όπως το φαινόμενο μάσκας (masking effect), σύμφωνα με το οποίο πολλά ορατά τεχνουργήματα δεν γίνονται αντιληπτά από τους θεατές, διότι «καλύπτονται» από την απεικονιζόμενη χωρική πολυπλοκότητα (spatial complexity)- αντιστοιχίζεται σε κάθε κανονικοποιημένο σφάλμα ένας στατιστικός συντελεστής βάρους, ανάλογος της αντιληπτής σημαντικότητας του κάθε οπτικού καναλιού, από το οποίο έχει προέρθει το συγκεκριμένο σφάλμα. Τελικά, με τη χρήση κάποιων μαθηματικής μεθόδου συνδυαστικής εξαγωγής αποτελέσματος (συνήθως χρησιμοποιείται το Minkowski error pooling), συνυπολογίζονται τα ποικίλα σφάλματα σε μια κοινή εκτίμηση της αντιληπτής ποιότητας. [5]

Αντιστοίχως, οι μέθοδοι μειωμένης αναφοράς για την εκτίμηση της αντιληπτής ποιότητας βίντεο βασίζονται σε ένα παρόμοιο πλαίσιο λειτουργίας, με βασική διαφορά ότι δεν χρησιμοποιούνται τα καρέ του συμπιεσμένου/ασυμπιέστου σήματος στην διαδικασία εκτίμησης, αλλά κάποια στατιστικά στοιχεία, τα οποία έχουν εξαχθεί από αυτά σε προγενέστερο στάδιο [14]. Παρομοίως, οι μέθοδοι μη αναφοράς δεν χρησιμοποιούν καθόλου το αρχικό ασυμπιέστο σήμα στην διαδικασία εκτίμησης της ποιότητας, περιορίζοντας τις απαιτήσεις τους μόνο στο κωδικοποιημένο σήμα [15]. Δηλαδή, οι μέθοδοι μη αναφοράς μπορούν να αποφανθούν για το επίπεδο της αντιληπτής ποιότητας εφαρμόζοντας αλγόριθμους και τεχνικές εκτίμησης απευθείας στα καρέ του κωδικοποιημένου σήματος. Η ανεξαρτησία αυτή από το αρχικό ασυμπιέστο οπτικό σήμα, επιτυγχάνεται επειδή οι μέθοδοι αυτές είναι συνήθως εξειδικευμένες ανάλογα με την τεχνική συμπίεσης που χρησιμοποιείται, εστιάζοντας στην ποσοτικοποίηση συγκεκριμένων αντιληπτών τεχνουργημάτων που προκύπτουν από τα ποικίλα είδη κωδικοποίησης. Συνεπώς, οι μέθοδοι μη αναφοράς δεν μπορούν να εφαρμοστούν καθολικά σε όλα τα συμπιεσμένα σήματα βίντεο, αφού έχουν προταθεί για συγκεκριμένες μεθόδους κωδικοποίησης, για τις οποίες παρέχουν πολύ καλά αποτελέσματα, προσομοιάζοντας επιτυχώς τις αντίστοιχες εκτιμήσεις των υποκειμενικών μεθόδων.

Συνεπώς, η διεθνής βιβλιογραφία σχετικά με την εκτίμηση της αντιληπτής ποιότητας ψηφιακού βίντεο, περιορίζεται σε αντικειμενικές μεθόδους που στοχεύουν στην ποσοτικοποίηση της ποιότητας ψηφιακού βίντεο, το οποίο έχει ήδη υποστεί κωδικοποίηση και συμπίεση, προσομοιάζοντας τα αποτελέσματα των αντιστοίχων υποκειμενικών μεθόδων, στις οποίες γίνεται χρήση ακροατηρίου. Δεδομένου ότι αναμένεται στα μελλοντικά τηλεπικοινωνιακά δίκτυα οι παρεχόμενες πολυμεσικές υπηρεσίες να προσφέρονται σε ποικίλα επίπεδα αντιληπτής ποιότητας, τα οποία θα καθορίζουν το κόστος και τη χρέωση της κάθε υπηρεσίας [16], είναι απαραίτητη η ανάπτυξη μεθόδων και τεχνικών που θα προσδιορίζουν σε μια οπτικοακουστική υπηρεσία εκείνες τις παραμέτρους κωδικοποίησης (δηλαδή το ρυθμό κωδικοποίησης για δεδομένη χωρική και χρονική ανάλυση) που αντιστοιχούν σε συγκεκριμένο επίπεδο αντιληπτής ποιότητας.

Με τις υπάρχουσες μεθόδους αντικειμενικής εκτίμησης της αντιληπτής ποιότητας ψηφιακού βίντεο που έχει υποστεί κωδικοποίηση και συμπίεση, απαιτούνται πολλαπλές επαναλαμβανόμενες δοκιμαστικές κωδικοποιήσεις και εκτιμήσεις ώστε να εντοπιστούν επακριβώς οι παράμετροι κωδικοποίησης που ικανοποιούν ένα συγκεκριμένο επίπεδο ποιότητας. Η διαδικασία όμως αυτή είναι αφενός χρονοβόρα και αφετέρου οικονομικά και εμπορικά μη βιώσιμη.

Εντοπίζοντας το πρόβλημα αυτό, η παρούσα διδακτορική διατριβή περιγράφει, προτείνει και δοκιμάζει πρωτότυπη μέθοδο πρόβλεψης και εκτίμησης της αντιληπτής ποιότητας ψηφιακά κωδικοποιημένου βίντεο συναρτήσει των παραμέτρων κωδικοποίησης και της δυναμικής του περιεχομένου, χωρίς να απαιτείται η χρονοβόρα και επίπονη διαδικασία επαναλαμβανόμενων δοκιμαστικών κωδικοποιήσεων και εκτιμήσεων. Με την προτεινόμενη τεχνική, οι πάροχοι υπηρεσιών βίντεο, ανάλογα με τη δυναμική του προς κωδικοποίηση περιεχομένου, είναι σε θέση να καθορίσουν άμεσα, γρήγορα, οικονομικά και με ικανοποιητική ακρίβεια τις παραμέτρους κωδικοποίησης που αντιστοιχούν σε συγκεκριμένα αντιληπτά επίπεδα ποιότητας, στα οποία θα προσφερθεί η εν λόγω υπηρεσία.

Στα πλαίσια αυτά, διερευνήθηκε και καθορίστηκε ποσοτικά η εξάρτηση της αντιληπτής ποιότητας (PQoS) από το περιεχόμενο, δηλαδή από τη δυναμική (dynamics) του βίντεο, αφού είναι γνωστό ότι ένα βίντεο χαμηλής δυναμικής (αργές κινήσεις, στατική εικόνα, κλπ) χρειάζεται μικρότερο ρυθμό κωδικοποίησης από ένα αντίστοιχο με υψηλή δυναμική (με γρήγορες σκηνές και εναλλαγές εικόνας), ώστε να επιτευχθεί το ίδιο επίπεδο ποιότητας [17]. Σύμφωνα με την ποσοτική εκτίμηση της δυναμικής του περιεχομένου, η προτεινόμενη μέθοδος προβλέπει με ικανοποιητική ακρίβεια τη μεταβολή της αντιληπτής ποιότητας συναρτήσει των παραμέτρων κωδικοποίησης. Από την μεταβολή αυτή είναι δυνατόν να καθοριστεί, με αντικειμενικό τρόπο, η αντιληπτή ποιότητα για ποικίλους ρυθμούς κωδικοποίησης και επίπεδα δυναμικής του περιεχομένου, χωρίς να απαιτούνται επαναλαμβανόμενες δοκιμαστικές κωδικοποιήσεις και εκτιμήσεις της ποιότητας. Πρόκειται δηλαδή για πρωτότυπη μέθοδο πρόβλεψης με αντικειμενικό τρόπο της αντιληπτής ποιότητας που θα έχει το συμπιεσμένο βίντεο ανάλογα με τη δυναμική του περιεχομένου του, εφόσον κωδικοποιηθεί με συγκεκριμένες παραμέτρους.

Συνεπώς, η παρούσα διδακτορική διατριβή συνεισφέρει ερευνητικά στο χώρο της αντιληπτής ποιότητας βίντεο, εισάγοντας για πρώτη φορά την έννοια της πρόβλεψης (και όχι απλά της εκτίμησης) της ποιότητας που αναμένεται να έχει ένα ψηφιακό βίντεο μετά το στάδιο της κωδικοποίησης, συναρτήσει του επιλεγμένου ρυθμού κωδικοποίησης και της δυναμικής του περιεχομένου του. Στη συνέχεια της παρούσης διατριβής, θα παρουσιαστούν αναλυτικά μετρήσεις και δοκιμές της προτεινόμενης μεθόδου, ενώ θα συζητηθούν και πιθανοί περιορισμοί.

Δομή της Διδακτορικής Διατριβής

Η παρούσα διατριβή πραγματεύεται τη διεύρυνση της έννοιας της αντιληπτής ποιότητας υπηρεσίας (PQoS), εστιάζοντας κυρίως στις εφαρμογές ψηφιακού βίντεο του προτύπου MPEG-4 (Part 2 – Visual) και δευτερευόντως του MPEG-4 AVC/H.264. Στόχος της είναι η συνεισφορά στην ανάπτυξη πρωτότυπης αντικειμενικής μεθόδου πρόβλεψης της αντιληπτής ποιότητας κωδικοποιημένου βίντεο συναρτήσει του ρυθμού κωδικοποίησης και της δυναμικής του περιεχομένου του.

Πιο συγκεκριμένα, η παρούσα διατριβή, εξαιρουμένου του παρόντος εισαγωγικού κεφαλαίου, δομείται από εννέα αυτοτελή κεφάλαια, εκ των οποίων τα τρία πρώτα είναι υποστηρικτικά των υπολοίπων έξι.

Στο δεύτερο κεφάλαιο παρουσιάζεται εν συντομία μια ιστορική αναδρομή της εξέλιξης των προτύπων κωδικοποίησης ψηφιακού βίντεο, καθώς και των φορέων που είναι αρμόδιοι για την προτυποποίηση των τεχνικών κωδικοποίησης. Στόχος του κεφαλαίου είναι να γνωστοποιήσει στον αναγνώστη τις βασικές εξελίξεις των προτύπων κωδικοποίησης ψηφιακού βίντεο, καθώς και τις αλληλεπιδράσεις που υπήρξαν μεταξύ των προτύπων και οδήγησαν στην εξέλιξη αυτών μέσα από την εμφάνιση νεότερων προτύπων.

Το τρίτο κεφάλαιο πραγματεύεται τις θεμελιώδεις αρχές της κωδικοποίησης ψηφιακού βίντεο, εξοικειώνοντας τον αναγνώστη με έννοιες που είναι απαραίτητες για την κατανόηση του περιεχομένου των λοιπών κεφαλαίων. Περιγράφονται εν συντομία οι βασικοί μηχανισμοί που συναντώνται στις σύγχρονες μεθόδους κωδικοποίησης βίντεο όπως η τεχνική εκτίμησης και αντιστάθμισης της κίνησης, ο διακριτός μετασχηματισμός συνημιτόνου, η κβάντιση των μετασχηματισμένων τιμών και η εντροπική κωδικοποίηση αυτών.

Εν συνεχεία, το τελευταίο υποστηρικτικό κεφάλαιο της διατριβής, το τέταρτο εν σειρά, παρουσιάζει μια συνοπτική βιβλιογραφική ανασκόπηση της κατηγοριοποίησης των μεθόδων εκτίμησης της αντιληπτής ποιότητας ψηφιακού βίντεο. Ταξινομώντας αρχικά τις τεχνικές εκτίμησης σε υποκειμενικές και αντικειμενικές, εν συνεχεία παρατίθεται μια σύντομη παρουσίαση των σημαντικότερων τεχνικών που είναι απαραίτητη για την κατανόηση των λοιπών κεφαλαίων της διατριβής. Με τη βοήθεια του κεφαλαίου ο αναγνώστης μπορεί να αντιληφθεί την καινοτομία της μεθόδου που προτείνεται από την παρούσα διατριβή, η οποία αποτελεί τομή στις καθιερωμένες τεχνικές, εισάγοντας για πρώτη φορά την έννοια της πρόβλεψης της αντιληπτής ποιότητας, κάνοντας ένα βήμα πέρα από τις συμβατικές τεχνικές εκτίμησης αυτής.

Στα πλαίσια των εισαγωγικών αυτών κεφαλαίων, οι ακόλουθες υποστηρικτικές δημοσιεύσεις της παρούσης διατριβής έχουν πραγματοποιηθεί σε τεχνικά βιβλία συμμετοχών:

- H. Koumaras, F. Liberal, L. Sun, "PQoS Assessment Methods for Multimedia Services", Chapter contribution in "Wireless Multimedia: Quality of Service and Solutions" (Αποδεκτό για δημοσίευση κατόπιν συγκεκριμένων τροποποιήσεων)
- H. Koumaras, E. Pallis, A. Kourtis, D. Martakos, "Principles of Digital Video Coding", Encyclopedia of Information Science & Technology 2nd edition, Idea Group Pub.
- H. Koumaras, E. Pallis, G. Xilouris, A. Kourtis, D. Martakos, "Perceived Quality Evaluation for Multimedia Services", Encyclopedia of Mobile Computing & Commerce, Idea Group Pub.

Με το πέμπτο κατά σειρά κεφάλαιο, ο αναγνώστης εισέρχεται στο κυρίως μέρος της διατριβής, όπου παρουσιάζεται μια στατιστική ανάλυση και μαθηματική μοντελοποίηση των δεδομένων ενός κωδικοποιημένου ψηφιακού βίντεο σύμφωνα με το πρότυπο MPEG-4/H.264. Στόχος αυτού του κεφαλαίου της διατριβής είναι η παρουσίαση, ανάλυση και κατανόηση των στατιστικών ιδιοτήτων των δεδομένων ενός κωδικοποιημένου βίντεο, όπως επί παραδείγματι ο βαθμός συσχέτισης των γειτονικών καρτέ. Η ανάλυση αυτή αποτελεί τη θεμέλια βάση της υπόλοιπης διατριβής.

Τα ερευνητικά αποτελέσματα αυτού του κεφαλαίου έχουν δημοσιευτεί στις εξής επιστημονικές εργασίες:

- H. Koumaras, C. Skianis, G. Xilouris, G. Gardikis, A. Kourtis, "A Markov Modified Model of H.264 VBR Video Traffic", IST Mobile Summit 2006, Mykonos, Greece, 4-8 June 2006
- H. Koumaras, C. Skianis, G. Gardikis, A. Kourtis, "Analysis of H.264 video encoded traffic", INC 2005 Fifth International Network Conference, Samos Island, Greece, July 2005
- H. Koumaras, C. Skianis, D. Martakos, A. Kourtis, "Analysis and Modeling of H.264 Unconstrained VBR Video Traffic", Journal of Communications and Networking (υπό κρίση)

Στο έκτο κεφάλαιο της διατριβής διερευνάται η σύγκριση σε επίπεδο αντιληπτής ποιότητας της μεθόδου κωδικοποίησης μεταβλητού και σταθερού ρυθμού (Variable Bit rate – Constant Bit Rate) για το πρότυπο MPEG-4. Τα αποτελέσματα της σύγκρισης δείχνουν πολύ μικρή αντιληπτή υπεροχή της κωδικοποίησης μεταβλητού έναντι του σταθερού ρυθμού. Σε αυτά τα πλαίσια, η σύγκριση αυτή εξυπηρετεί και τα υπόλοιπα μέρη της διατριβής. Με τη χρήση των αποτελεσμάτων αυτού του μέρους δεν είναι πλέον απαραίτητη η εξαγωγή διπλών αποτελεσμάτων –δηλαδή μεταβλητού και σταθερού ρυθμού- για το πέρας της διατριβής, αφού συμπεράσματα μόνο της μιας μεθόδου κωδικοποίησης μπορούν πολύ εύκολα να επεκταθούν και στην άλλη μέθοδο κωδικοποίησης μέσω της προτεινόμενης σύγκρισης/απεικόνισης.

Η σχετική έρευνα έχει δημοσιευτεί σε διεθνές επιστημονικό περιοδικό στην εργασία:

- H. Koumaras, G. Gardikis, A. Kourtis, D. Martakos, "Quantitative Perceptual Efficiency of VBR over CBR for MPEG-4 Video", Journal of Electronic Imaging (JEI), SPIE, IS&T, July-September 2007.

Στο έβδομο κεφάλαιο της διατριβής, διερευνάται και επιλύεται το πρόβλημα του καθορισμού των παραμέτρων κωδικοποίησης που ικανοποιούν ένα συγκεκριμένο επίπεδο αντιληπτής ποιότητας υπηρεσίας για βίντεο σύντομης διάρκειας. Η προτεινόμενη μέθοδος καινοτομεί σε σύγκριση με τις υπάρχουσες μεθόδους εκτίμησης της ποιότητας κωδικοποιημένου βίντεο, διότι παρέχει πρόβλεψη της τελικής ποιότητας πριν το τελικό στάδιο της κωδικοποίησης. Τα αποτελέσματα της προτεινόμενης μεθόδου δεν περιορίζονται μόνο σε ποιοτικές εκτιμήσεις. Αντιθέτως, η διατριβή περιγράφει με αναλυτικό τρόπο και με ικανοποιητική ακρίβεια την εξάρτηση της ποιότητας του βίντεο συναρτήσει του ρυθμού κωδικοποίησης για ποικίλα επίπεδα χωρικής και χρονικής δυναμικής του περιεχομένου.

Η ανάπτυξη της προαναφερθείσας μεθόδου έχει υποστηριχθεί με τις ακόλουθες επιστημονικές δημοσιεύσεις σε διεθνή περιοδικά και ανακοινώσεις σε διεθνή συνέδρια:

- H. Koumaras, A. Kourtis, D. Martakos, J. Lauterjung, "Quantified PQoS Assessment Based on Fast Estimation of the Spatial and Temporal Activity Level", Journal of Multimedia Tools and Applications, Springer Editions, pp. 355-374, Volume 34, Number 3, September, 2007.

- H. Koumaras, A. Kourtis, D. Martakos, "Evaluation of Video Quality Based on Objectively Estimated Metric", Journal of Communications and Networking, Vol. 7(3), Sep 2005,
- H. Koumaras, D. Martakos, A. Kourtis, "Objective Measurement of Perceived QoS for Homogeneous MPEG-4 Video Content", Encyclopedia of Multimedia Technology and Networking, Idea Group, April 2005.
- H. Koumaras, G. Gardikis, A. Kourtis, "Objective Evaluation of the Perceived Quality of Video Content", IST Mobile Summit 2003, Aveiro, Portugal, June 2003
- H. Koumaras, E. Pallis, G. Xilouris, A. Kourtis, D. Martakos, J. Lauterjung, "Pre-Encoding PQoS Assessment Method for Optimized Resource Utilization", 2nd Inter. Conference on Performance Modelling and Evaluation of Heterogeneous Networks, Het-NeTs04, Ilkley, United Kingdom, July 2004.
- H. Koumaras, E. Pallis, G. Gardikis, A. Kourtis, "Real Time PQoS Enhancement of IP Multimedia Services Over Fading and Noisy DVB-T Channel", TEMU2006, Int. Conf. on Telecommunications and Multimedia, Heraklion, Greece, 5-7 July 2006.
- H. Koumaras, T. Pliakas, A. Kourtis, "A Novel Method for Pre-Encoding Video Quality Prediction", IST Mobile Summit 2007, Budapest, Hungary, 1-5 July 2007.
- H. Koumaras, A. Kourtis, "Video Quality Prediction Based on the Spatial and Temporal Classification of the Uncompressed Content", The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Comm. (PIMRC), Athens, Greece, 3-7 Sep. 2007.

Στο όγδοο κεφάλαιο της διατριβής, η προαναφερθείσα αντικειμενική μέθοδος επεκτείνεται σε βίντεο μεγάλης διάρκειας. Για το λόγο αυτό διερευνάται, αναλύεται και προτείνεται μέθοδος εντοπισμού των ορίων μια σκηνής (shot boundary detection) σε κωδικοποιημένο βίντεο. Η προτεινόμενη μέθοδος βασίζεται στην παρατήρηση ότι ένα κωδικοποιημένο βίντεο παρουσιάζει στιγμιαία υποβάθμιση της αντιληπτής ποιότητας κατά την εναλλαγή μιας σκηνής, εξαιτίας της χαμηλής χρονικής συσχέτισης μεταξύ των αντιστοίχων διαδοχικών καρέ. Επιπλέον, η μέθοδος αυτή καινοτομεί σε σύγκριση με την υπάρχουσα βιβλιογραφία, διότι δεν απαιτεί τη χρήση κάποιου πολύπλοκου μηχανισμό κατωφλίου αναφοράς για τον εντοπισμό των σκηνών. Ενσωματώνοντας τη μέθοδο αυτή στην προαναφερθείσα μέθοδο πρόβλεψης της ποιότητας βίντεο πριν το τελικό στάδιο της κωδικοποίησης, επεκτείνεται η εγκυρότητα της μεθόδου και σε βίντεο μεγάλης διάρκειας.

Οι ακόλουθες δημοσιευμένες επιστημονικές εργασίες σε διεθνή επιστημονικά περιοδικά και βιβλία συμμετοχών υποστηρίζουν τα ερευνητικά αποτελέσματα του ογδού κεφαλαίου:

- H. Koumaras, G. Gardikis, G. Xilouris, E. Pallis, A. Kourtis, "Shot Boundary Detection without Threshold Parameters", Journal of Electronic Imaging (JEI), SPIE, Vol.15(2), April 2006.
- H. Koumaras, E. Pallis, A. Kourtis, D. Martakos, "An Automatic Shot Boundary Detection Method for Low Bit Rate DCT-based Video", Journal of Telecommunication Systems, Springer Ed (υπό κρίση).
- H. Koumaras, G. Xilouris, E. Pallis, G. Gardikis, A. Kourtis, "Shot Boundary Detection Techniques For Video Sequences", Encyclopedia of Mobile Computing & Commerce, Idea Group Pub.

Στο ένατο μέρος της διατριβής διερευνάται η επίδραση του ρυθμού απώλειας πακέτων του δικτύου μεταφοράς στην υποβάθμιση της αντιληπτής ποιότητας ροής βίντεο πραγματικού χρόνου. Η περιγραφείσα έρευνα κρίνεται ιδιαίτερα ενδιαφέρουσα και βρίσκεται υπό εξέλιξη, έχοντας τα πρώτα αποτελέσματα αυτής ανακοινωθεί στην ακόλουθη εργασία, η οποία απέσπασε το βραβείο της καλύτερης εργασίας του συνεδρίου (Best Paper Award):

- H. Koumaras, A. Kourtis, C-H Lin, C-K Shieh, "A Theoretical Framework for End-to-End Video Quality Prediction of MPEG-based Sequences", The Third Inter. Conf. on Networking and Services - ICNS07, Athens, Greece, June 19-25 2007.

Σε αυτή την ερευνητική περιοχή, έχει επίσης υποβληθεί για δημοσίευση και βρίσκεται υπό κρίση το ακόλουθο άρθρο:

- H. Koumaras, C-H Lin, C-K Shieh, A. Kourtis, "End-to-End Video Quality Assessment of MPEG-based Broadcasting Services ", IEEE Transactions on Broadcasting, Special issue on Quality Issues in Multimedia Broadcasting.

Τέλος, το δέκατο κεφάλαιο συζητά πώς τα ερευνητικά αποτελέσματα της παρούσας διατριβής μπορούν να αξιοποιηθούν προς τη διατύπωση και επίλυση πρωτότυπων ερευνητικών ζητημάτων, τα οποία χρήζουν περαιτέρω διερεύνησης. Τα θέματα αυτά θα αποτελέσουν αντικείμενο μελλοντικής εργασίας.

Μέχρι τη στιγμή συγγραφής της διατριβής, έχουν καταμετρηθεί συνολικά 14 ετεροαναφορές στις εξής δημοσιεύσεις:

- Way-Chuang Ang, Chee-Hong The, Tat-Chee Wan and Rahmat Budiarto, "Performance Evaluation of Unidirectional Lightweight Encapsulation using ns-2 and DVB-S2 Test bed", Proceedings of the 2007 IEEE International Conference on Telecommunications and Malaysia International Conference on Communications, 14-17 May 2007, Penang, Malaysia.
- M. Ries, C. Crespi, O. Nemethova, M. Rupp, "Content Based Video Quality Estimation for H.264/AVC Video Streaming", in Proc. Proceedings of IEEE Wireless and Communications & Networking Conference, Hong Kong, March, 2007.
- Eric A. Silva, Karen Panetta, Sos S. Agaian, "Quantifying image similarity using measure of enhancement by entropy", Mobile Multimedia/Image Processing for Military and Security Applications 2007, Sos S. Agaian, Sabah A. Jassim, Editors, 65790U, Proceedings of SPIE -- Volume 6579, May. 2, 2007.
- Ulrich Engelke and Hans-Jürgen Zepernick, "Perceptual Quality Measures for Image and Video Services", Euro-NGI Workshop on Socio-Economic Aspects of Next Generation Internet, Lyngby, Denmark, October 9-10, 2006.
- Astrid Undheim, Yuan Lin, Peder J. Emstad, "Characterization of Slice-Based H.264/AVC Encoded Video Traffic," ecumn, pp. 263-272, Fourth European Conference on Universal Multiservice Networks (ECUMN'07), 2007.
- Michal Ries, Olivia Nemethova, Markus Rupp, "Motion Based Reference-Free Quality Estimation for H.264/AVC Video Streaming", International Symposium on Wireless Pervasive Computing , San Juan, Poerto Rico, 5-7 February 2007.

- M. Fouad, F. Bayoumi, H. Onsi, M. Darwish, "Shot Transition Detection with Minimal Decoding of MPEG Video Streams", Transactions on Engineering, Computing and Technology, Vol.17, Dec. 2006.
- J. Doggen, F. Schueren, "MPEG4 Part 10-H.264 Modeling in OPNET", Masters Thesis, Department of Industrial Sciences and Technology, University College of Antwerpen, Belgium, July 2006.
- F. Liberal et al "PQoS Based Model for Assessing Significance Statistically", 3rd Inter. Conference on Performance Modelling and Evaluation of Heterogeneous Networks, Het -NeTs 05, 18-20 July 2005, Ilkley, U.K., July 2005.
- T. Pliakas, G. Kormentzas, Ch. Skianis "Scalable Video Streaming Traffic Delivery in IP/UMTS Networking Environments", Journal of Mobile Multimedia, 2007.
- Christopher John, "Effect of content on perceived video quality", TLEN 5380 – Video Technology, Colorado, 9 August 2006.
- J. Jiang, Z. Li and G. Xiao, "Real-time shot cut detection in compressed domain", Journal of Electronic Imaging, SPIE, 2007.
- T. Pliakas, G. Kormentzas, Ch. Skianis "End-to-end QoS Issues of MPEG-4 FGS Video Streaming Traffic Delivery in an IP/DVB/UMTS Network", Special Issue of Elsevier journal of Operational Research.
- Ulrich Engelke and Hans-Jürgen Zepernick, "Perceptual-based Quality Metrics for Image and Video Services: A Survey", 3rd EuroNGI Conference on Next Generation Internet Networks, Trondheim, Norway, 21-23 May 2007.

ΚΕΦΑΛΑΙΟ 2

Πρότυπα Κωδικοποίησης Ψηφιακού Βίντεο

Από το αναλογικό στο ψηφιακό σήμα βίντεο: Η ανάγκη για συμπίεση

Είναι ευρέως γνωστό ότι ένα τυπικό αναλογικό τηλεοπτικό σήμα PAL καταλαμβάνει εύρος ζώνης (bandwidth) ίσο με 8MHz. Για τη ψηφιοποίηση του σήματος αυτού, χωρίς απώλεια πληροφορίας, σύμφωνα με το θεώρημα Nyquist, πρέπει να πραγματοποιηθεί δειγματοληψία στη διπλάσια συχνότητα, δηλαδή στα 16MHz (που αντιστοιχεί σε 16.000.000 δείγματα το δευτερόλεπτο). Για να έχουμε επαρκή ευκρίνεια (δηλαδή ικανοποιητικές αποχρώσεις του γκρι σε ασπρόμαυρο σήμα) πρέπει κάθε δείγμα να περιγράφεται τουλάχιστον από 8 bits (συνεπώς $2^8=256$ διακριτά επίπεδα κωδικοποίησης). Συνεπώς ο ψηφιακός ρυθμός μετάδοσης που αντιστοιχεί σε αυτά τα χαρακτηριστικά είναι $16 \cdot 8 = 128 \text{Mbits/s}$.

Είναι προφανές ότι ο ρυθμός των 128 Mbit το δευτερόλεπτο είναι υψηλός, καθιστώντας αδύνατη τη μετάδοση ασυμπίεστου ψηφιακού σήματος βίντεο ανάλυσης PAL. Το πρόβλημα του ρυθμού μετάδοσης είναι ακόμη εντονότερο αν θελήσουμε να έχουμε υψηλότερη ευκρίνεια (περισσότερες αποχρώσεις του γκρι σε ασπρόμαυρο σήμα), οπότε θα χρησιμοποιήσουμε μήκος λέξης των 16, ή 24bits. Οπότε οι αντίστοιχοι ρυθμοί μετάδοσης θα είναι ακόμη υψηλότεροι. Συνεπώς, εκ των πραγμάτων προκύπτει η ανάγκη για συμπίεση του βίντεο με στόχο τη μείωση του απαιτούμενου ρυθμού μετάδοσης, σε συνδυασμό με ταυτόχρονη διατήρηση της αντιληπτής ποιότητας σε ικανοποιητικά επίπεδα.

Εκ του αποτελέσματος λοιπόν, θα μπορούσαμε να πούμε ότι η συμπίεση του βίντεο, καθιστά εφικτή: Πρώτον, τη μετάδοση ψηφιακού βίντεο τηλεοπτικής ανάλυσης (PAL ή NTSC) μέσω τηλεπικοινωνιακών δικτύων που δεν μπορούν να υποστηρίξουν τους ρυθμούς μετάδοσης του ασυμπίεστου ("raw") βίντεο. Δεύτερον, την αποδοτικότερη αξιοποίηση των διαθέσιμων πόρων μετάδοσης (εύρος ζώνης) και αποθήκευσης (χωρητικότητα).

Πρότυπα Κωδικοποίησης Βίντεο

Ομάδες εργασίας στην Ανάπτυξη Προτύπων Ψηφιακού Βίντεο

ITU-T VCEQ

Η ομάδα εργασίας Video Coding Expert Group της ITU του τμήματος προτυποποίησης τηλεπικοινωνιακών θεμάτων (ITU-T) ασχολείται με την ανάπτυξη προτύπων ή συστάσεων για τις τηλεπικοινωνίες. Το τμήμα προτυποποίησης τηλεπικοινωνιακών θεμάτων συντελείται από υπό-ομάδες εργασίας, όπου η κάθε υπό-ομάδα έχει μια συγκεκριμένη θεματική περιοχή δράσης [2]. Η υπό-ομάδα 16 (SubGroup 16 - SG16) είναι υπεύθυνη για θέματα πολυμεσικών υπηρεσιών και εφαρμογών, όπως πολυμεσικά τερματικά, συστήματα, πρωτόκολλα και κωδικοποίηση σήματος (media coding). Η υπό-ομάδα 16, διαιρείται και εκείνη σε μικρότερες εξειδικευμένες ομάδες εργασίας (working party), εκ των οποίων η ομάδα 3 εξειδικεύεται με την κωδικοποίηση σήματος (media coding). Μεταξύ των ποικίλων ενοτήτων εργασίας αυτής της ομάδας, η ενότητα (question) 6 ασχολείται με τη ψηφιακή κωδικοποίηση βίντεο και ουσιαστικά εκείνα τα μέλη που συμμετέχουν ενεργά στις

συναντήσεις που πραγματοποιούνται στα πλαίσια αυτής της ενότητας συνθέτουν την VCEG. Συνεπώς, η επίσημη ονομασία της VCEG είναι ITU-T SG16 WP3 Q6. Η ονομασία VCEG υιοθετήθηκε σχετικά πρόσφατα, ενώ νωρίτερα η συγκεκριμένη ομάδα εργασίας ήταν γνωστή πέραν από την επίσημη ονομασία της και ως Low Bitrate Coding (LBC) Experts Group.

Ο στόχος της ITU-T SG16 WP3 Q6 είναι η προώθηση, διατύπωση και παραγωγή προτάσεων για μεθόδους κωδικοποίησης βίντεο, οι οποίες να είναι κατάλληλες για εφαρμογές τηλεδιάσκεψης (π.χ. βίντεο-τηλεφωνία) και λοιπές εφαρμογές βίντεο (π.χ. βίντεο ροών, τηλεοπτικής εκπομπής, μεταφόρτωσης αρχείου, αποθήκευσης και αναπαραγωγής πολυμεσικού περιεχομένου ή ψηφιακού κινηματογράφου). Επιπλέον, η ITU-T SG16 WP3 Q6 εστιάζει και στην επέκταση των υπαρχόντων προτύπων κωδικοποίησης, ενώ δημιουργεί τις κατάλληλες συνθήκες για νέες προτάσεις κωδικοποίησης, που θα υιοθετούν πρωτοποριακές τεχνικές για τη βελτίωση του trade-off ανάμεσα στο λόγο συμπίεσης, το επίπεδο της αντιληπτής ποιότητας, την καθυστέρηση που εισάγει η διαδικασία της κωδικοποίησης και την πολυπλοκότητα του αλγορίθμου.

ISO/MPEG

Η Moving Pictures Expert Group (MPEG) είναι μια ομάδα εργασίας του Διεθνή Οργανισμού Προτυποποίησης (International Organization of Standardization - ISO) και της Διεθνής Ηλεκτροτεχνικής Επιτροπής (International Electrotechnical Committee – IEC). Επίσημα η MPEG είναι το Working Group (WG) 11 της SubCommittee 29 του Joint Technical Committee 1 και κατά συνέπεια η επίσημη ονομασία της είναι ISO/IEC JTC1/SC29/WG11.

Σκοπός της MPEG είναι η ανάπτυξη διεθνών προτύπων για συμπίεση, επεξεργασία και κωδικοποιημένη αναπαραστάση κινούμενων εικόνων, ήχου ή συνδυασμό αυτών με σκοπό την κάλυψη ενός μεγάλου εύρους εφαρμογών [18]. Οι στόχοι της MPEG είναι [19]:

- § Να δραστηριοποιείται εντός του ISO/IEC ως το υπεύθυνο σώμα για την πρόταση νέων προτύπων σχετικά με το αντικείμενο εργασίας.
- § Να συνεργάζεται με παρόμοια σώματα προτυποποίησης που ασχολούνται με παρεμφερείς εφαρμογές. Για το λόγο αυτό η MPEG δημιουργεί συνδέσμους με συγγενή σώματα προτυποποίησης, καθώς και με σχετικούς οργανισμούς και εταιρείες, ανταλλάζοντας και συναλλάζοντας τεχνογνωσία, τεχνικές απαιτήσεις και προδιαγραφές, ώστε να επιτευχθεί η μέγιστη δυνατή εμπορική αξιοποίηση των προτύπων.
- § Να θέτει τις απαιτήσεις ώστε να είναι δυνατή η εφαρμογή των προτύπων της MPEG σε διαφορετικές εφαρμογές, όπως οι τηλεπικοινωνίες και η τηλεοπτική βιομηχανία. Επιπλέον, πραγματοποιούνται όλες οι απαραίτητες ενέργειες για συνδυασμό των προτύπων της MPEG με σχετικά πρότυπα κωδικοποίησης εικόνας και ήχου άλλων φορέων.
- § Να καθορίσει τις μεθόδους και διαδικασίες για την υποκειμενική εκτίμηση της αντιληπτής ποιότητας βίντεο, ήχου, εικόνας και συνδυασμό αυτών.
- § Να αξιολογήσει τα ποικίλα χαρακτηριστικά των τεχνολογιών υλοποίησης των αλγορίθμων κωδικοποίησης ήχου, κινούμενης εικόνας και συνδυασμών αυτών.
- § Να αξιολογήσει τα χαρακτηριστικά την ψηφιακής αποθήκευσης και παράδοσης πολυμέσων, που έχουν προταθεί από το WG11.

- § Να αναπτύξει πρότυπα για την κωδικοποίηση κινούμενης εικόνας, ήχου και συνδυασμό αυτών, λαμβάνοντας υπόψιν την ποιότητα του κωδικοποιημένου σήματος, την αποδοτική κωδικοποίηση και παραμέτρους για την παράδοση και αποστολή του περιεχομένου.
- § Να προτείνει πρότυπα για την κωδικοποιημένη αναπαράσταση κινούμενων εικόνων
- § Να προτείνει πρότυπα για την κωδικοποιημένη αναπαράσταση ήχου
- § Να προτείνει πρότυπα για την κωδικοποιημένη αναπαράσταση κινούμενων εικόνων σε συνδυασμό με ήχο
- § Να προτείνει πρότυπα για πρωτόκολλα που σχετίζονται με την κωδικοποιημένη αναπαράσταση κινούμενων εικόνων, ήχου και συνδυασμό αυτών.

Η MPEG διαιρείται σε υπό-ομάδες ειδικών εργασιών, όπου η κάθε μία εξειδικεύεται σε ένα συγκεκριμένο θέμα σχετικό με τις διαδικασίες προτυποποίησης. Οι ευθύνες των υπό-ομάδων αυτών είναι τα εξής [18]:

- § Requirements: Να εντοπίζει τις ανάγκες και απαιτήσεις της αγοράς για νέα πρότυπα.
- § Systems: Μελετά τις προδιαγραφές των συστημάτων καταγραφής, αναπαραγωγής και επεξεργασίας ήχου, βίντεο και εικόνων, καθώς και τις διαδικασίες μεταφοράς των δεδομένων αυτών μέσω τηλεπικοινωνιακών δικτύων.
- § Description: Μελετά τη δήλωση και περιγραφή των ψηφιακών πολυμεσικών αντικειμένων.
- § Video: Μελετά την κωδικοποίηση κινούμενης εικόνας.
- § Audio: Μελετά την κωδικοποίηση ήχου.
- § Synthetic Natural Hybrid Coding: Μελετά την κωδικοποίηση συνθετικού βίντεο και ήχου σε συνδυασμό με φυσικό βίντεο και ήχο.
- § Integration: Ελέγχει τη συμβατότητα των υλοποιήσεων με το πρότυπο και αναπτύσσει το λογισμικό αναφοράς.
- § Test: Μελετά υποκειμενικές μεθόδους εκτίμησης της ποιότητας εικόνας βίντεο
- § Implementation: Καθορίζει τις προδιαγραφές ανάπτυξης και υλοποίησης σχετικών εφαρμογών.
- § Liaison: Αναπτύσσει δεσμούς με άλλους σχετικούς φορείς και ομάδες εργασίας.

Τα μέλη της MPEG προέρχονται από ποικίλα ινστιτούτα, πανεπιστήμια, ερευνητικούς και βιομηχανικούς φορείς. Η δυνατότητα να γίνει κάποιος μέλος και να συμμετέχει στις συνεδριάσεις της MPEG περιορίζεται μόνο σε εκπροσώπους εθνικών σωμάτων προτυποποίησης. Μια εταιρεία ή ένα ίδρυμα που επιθυμεί να συμμετέχει στην ομάδα εργασίας MPEG πρέπει πρώτα να υποβάλλει αίτηση για να συμπεριληφθεί στο εκάστοτε εθνικό σώμα προτυποποίησης και μέσω αυτού να συμμετέχει εν συνεχεία ενεργά στις συνεδριάσεις της MPEG.

Οι συνεδριάσεις της MPEG λαμβάνουν χώρα κάθε 2-3 μήνες. Από το 1988, έχουν συνολικά πραγματοποιηθεί 64 συνεδριάσεις. Η παρθενική συνάντηση της ομάδας MPEG πραγματοποιήθηκε το 1988 στην Ottawa του Καναδά. Κάθε συνεδρίαση διαρκεί συνήθως 5 ημέρες, όπου οι ποικίλες υπό-ομάδες εργασίας συναντιούνται πριν την κυρίως συνεδρίαση, όπου ποικίλες προτάσεις καταθέτονται προς έγκριση και σταδιακά δρομολογούνται προς

προτυποποίηση. Οι αποφάσεις της κάθε συνεδρίασης καταγράφονται υπό τη μορφή εισερχόμενων και εξερχόμενων εγγράφων. Τα εισερχόμενα έγγραφα περιλαμβάνουν αναφορές των υπό-ομάδων εργασίας, προτάσεις για αλλαγές ή επεκτάσεις των υπό ανάπτυξη προτύπων και δηλώσεις άλλων φορέων σχετικές με τις εργασίες της MPEG. Τα εξερχόμενα έγγραφα περιλαμβάνουν πρακτικά και αναφορές των συνεδριάσεων καθώς και πρόχειρες εκδοχές των προτύπων πριν λάβουν την τελική τους μορφή.

JVT

Η Joint Video Team (JVT) είναι μια ομάδα εργασίας που ιδρύθηκε το 2001 και αποτελείται από μέλη της ISO/IEC JTC1/SC29/WG11 (MPEG) και της ITU-T SG16 WP3 Q6 (VCEG). Η γενεσιουργός αιτία της JVT ήταν η απαίτηση της MPEG για εργαλεία κωδικοποίησης βίντεο υψηλής απόδοσης. Εκείνη την εποχή ο βέλτιστος μηχανισμός κωδικοποίησης ήταν ο MPEG-4 Part 2 Visual, ο οποίος όμως ήταν βασισμένος σε μια αρκετά ξεπερασμένη τεχνολογία, που περιγραφόταν στο πρότυπο H.263 του 1995. Με τις εξελίξεις που είχαν πραγματοποιηθεί στο χώρο της υπολογιστικής ισχύος και στον ερευνητικό τομέα της κωδικοποίησης κινούμενης εικόνας, ήταν πλέον ξεκάθαρο ότι ήταν γόνιμο το έδαφος για την ανάπτυξη ενός νέου προτύπου κωδικοποίησης. Το νέο πρότυπο θα αποτελούσε σημείο καμπής στην εξέλιξη της απόδοσης των μεθόδων κωδικοποίησης βίντεο.

Έπειτα από εκτίμηση των τότε γνωστών τεχνικών κωδικοποίησης κινούμενης εικόνας, η MPEG κατέληξε στο συμπέρασμα ότι ο υπό-ανάπτυξη αλγόριθμος κωδικοποίησης της VCEG H.26L ήταν ο καταλληλότερος για περαιτέρω βελτίωση και εξέλιξη, ώστε να επιτευχθεί ο στόχος για βέλτιστη κωδικοποίηση. Κατόπιν αυτής της επιλογής, αποφασίστηκε ότι μέλη της MPEG και της VCEG θα συστήσουν μια ομάδα δράσης που από κοινού θα διευθύνει τα τελικά στάδια εξέλιξης του H.26L. Η κοινή αυτή ομάδα δράσης ονομάστηκε JVT και αυτή η από κοινού εξέλιξη του H.26L οδήγησε τελικά στην γέννηση του νεότερου προτύπου κωδικοποίησης κινούμενης εικόνας του H.264 το 2003.

Αξίζει να σημειωθεί ότι ένα από το πλέον προκαταρκτικό μέλημα των μελών της JVT ήταν η αναζήτηση και η εξεύρεση πιθανών υπαρκτών πατεντών ευρεσιτεχνίας σχετικών με τα βασικά προφίλ του υπό-ανάπτυξη προτύπου. Στόχος ήταν η διασφάλιση της υλοποίησης του πλέον βασικού προφίλ του νέου προτύπου από οποιονδήποτε φορέα χωρίς να προκύπτουν ζητήματα πληρωμής για αδειοδότηση και πνευματικά δικαιώματα σε τρίτους φορείς που θα τα διατηρούν μέσω κάποιας σχετικής πατέντας ευρεσιτεχνίας. Η επιτυχία του νέου προτύπου θα κρινόταν σημαντικά από τέτοια ζητήματα ελεύθερης υλοποίησης, μιας και η αντίστοιχη περίπτωση του προτύπου MPEG-4 είχε δημιουργήσει σημαντικά προβλήματα σε οργανισμούς και εταιρείες. Πλέον είχε γίνει κατανοητό ότι αν υπάρχει υψηλό κόστος στην αδειοδότηση ενός προτύπου, τότε οι υποψήφιοι φορείς αν και αρχικά επιθυμούν την υιοθέτηση του νέου προτύπου, τελικά οδηγούνται στην απόρριψή του, λόγω του υψηλού κόστους αγοράς των σχετικών δικαιωμάτων. Μάλιστα, δεν ήταν και λίγοι οι υποστηρικτές της άποψης ότι η αποτυχία του προτύπου MPEG-4 με τις περιορισμένου εύρους εμπορικές εφαρμογές που αναπτύχθηκαν σε σύγκριση με το προγενέστερο MPEG-2, οφειλόταν κυρίως στην πολύπλοκη και οικονομικά μη συμφέρουσα διαδικασία αδειοδότησης. Συνεπώς, για να μπορέσει να επιτύχει το νέο πρότυπο θα έπρεπε να είναι απαλλαγμένο από τέτοιου είδους προβλήματα, γεγονός που και η JVT γνώριζε από την εμπειρία των μελών της και που ουσιαστικά φρόντισε να μη συμβεί.

Γένεση και διαδρομή των προτύπων κωδικοποίησης ψηφιακού βίντεο

Όπως αναφέρθηκε στο προηγούμενο εδάφιο, η ψηφιοποίηση του τηλεοπτικού σήματος αποτέλεσε την απαρχή της κωδικοποίησης ψηφιακού σήματος. Η χρήση κωδικοποίησης PCM των τηλεοπτικών προγραμμάτων σε ρυθμό των 140 Mbps εισήχθη τη δεκαετία του 1970. Η υψηλή τιμή όμως του απαιτούμενου ρυθμού μετάδοσης περιόρισε τη χρήση του σε επαγγελματικό επίπεδο για τη διανομή των τηλεοπτικών προγραμμάτων και την επεξεργασία (π.χ. μοντάζ) του ψηφιακού υλικού. Η χρήση του PCM σε εφαρμογές βιντεοκλήσης και βιντεοδιάσκεψης δε μπορούσε να υλοποιηθεί σε αυτούς τους ρυθμούς μετάδοσης.

Τη δεκαετία του 1980, ο βίντεο-κωδικοποιητής COST211, ο οποίος ήταν βασισμένος στη διαμόρφωση DPCM, κατέληξε στο πρώτο πρότυπο κωδικοποίησης βίντεο από την ITU με την ονομασία H.120. Ο ρυθμός κωδικοποίησης που στόχευε το πρότυπο H.120 [20] ήταν 2Mbps για την Ευρώπη και 1.544 Mbps για τη Β. Αμερική. Το τελικό κωδικοποιημένο σήμα, αν και προσέφερε πολύ καλή χωρική ανάλυση, παρόλα αυτά η ποιότητα στη χρονική ανάλυση ήταν πολύ χαμηλή. Σύντομα έγινε αντιληπτό, ότι για να βελτιωθεί η τελική ποιότητα του κωδικοποιημένου σήματος (σε χωρική και χρονική ανάλυση), χωρίς να επιβαρυνθεί επιπλέον ο ρυθμός κωδικοποίησης, θα έπρεπε να χρησιμοποιείται στη κωδικοποίηση λιγότερο από ένα bit σε κάθε pixel. Αυτό είναι εφικτό μόνο εάν μια ομάδα από pixels κωδικοποιηθούν μαζί, ώστε η μέση τιμή των bits που αντιστοιχούν σε ένα pixel να είναι μικρότερη της μονάδας, δηλαδή κλασματική. Η διαπίστωση αυτή οδήγησε στην αποκαλούμενη κωδικοποίηση τετράγωνων δομών (block-based encoding).

H.261

Στα τέλη της δεκαετίας του 1980, στο τμήμα προτυποποίησης της ITU είχαν κατατεθεί 15 διαφορετικές προτάσεις βασισμένες στην block-based κωδικοποίηση, εκ των οποίων οι 14 χρησιμοποιούσαν το Διακριτό Μετασχηματισμό Συνημίτονου (DCT). Την ίδια ακριβώς περίοδο, το Joint Photographic Experts Group (JPEG) επέλεξε το μετασχηματισμό DCT ως το κύριο μετασχηματισμό στη διαδικασία συμπίεσης στατικών εικόνων. Προφανώς η επιλογή αυτή επηρέασε και την ITU, η οποία παρουσίασε μια εξέλιξη του προτύπου H.120, βασισμένο σε DPCM και DCT. Το αποτέλεσμα ήταν βελτιωμένος λόγος συμπίεσης σε συνδυασμό με καλύτερη ποιότητα κωδικοποίησης τόσο σε χωρικό όσο και σε χρονικό επίπεδο. Πιο συγκεκριμένα, η αντιληπτή ποιότητα που επιτύγχανε σε ρυθμό κωδικοποίησης 384kbps ήταν ικανοποιητική και ο ρυθμός αυτός ήταν αποδεκτός – όντας χαμηλός - για εφαρμογές τηλεδιάσκεψης. Αργότερα, οι υποστηριζόμενοι ρυθμοί αυτού του προτύπου καθορίστηκαν να είναι τιμές πολλαπλάσιες των 64kbps ($p \times 64kbps$, $p \in [1,30]$), οι οποίες συμπεριλήφθηκαν το 1990 στο πρότυπο H.261.

Αρχικά το H.261 προοριζόταν για μετάδοση βίντεο σε γραμμές ISDN, για αυτό το λόγο και στην πρώτη έκδοση του προτύπου οι υποστηριζόμενοι ρυθμοί είναι ακέραια πολλαπλάσια των 64kbps. Αργότερα το εύρος των υποστηριζόμενων ρυθμών κωδικοποίησης επεκτάθηκε ώστε να συμπεριλάβει επιπλέον τιμές ανάμεσα σε 40kbps και 2Mbps. Το πρότυπο υποστηρίζει χωρικές αναλύσεις CIF (352x288) και QCIF(176x144) με χρωματική δειγματοληψία 4:2:0, που σημαίνει ότι οι αντίστοιχες χρωματικές χωρικές αναλύσεις είναι 176x144 και 88x72, αντίστοιχα.

MPEG-1

Παράλληλα με την ITU, το MPEG forum εργαζόταν στην ανάπτυξη ενός άλλου προτύπου ψηφιακής κωδικοποίησης βίντεο: Το MPEG-1. Το MPEG forum είναι μια επιτροπή που δρα στα πλαίσια του διεθνούς οργανισμού τυποποίησης (ISO). Επίσημα είναι γνωστή ως ISO/IEC JTC1/SC29/WG11. Ιδρύθηκε το 1988 και είναι μέλος του **JTC1** (Joint ISO/IEC Technical Committee on Information Technology - Ενωμένη Τεχνική Επιτροπή ISO/IEC στην Τεχνολογία της Πληροφορικής). Ο συντονιστής της επιτροπής MPEG είναι ο Leonardo Chiariglione, γνωστός και ως «πατέρας» του MPEG. Η επιτροπή πραγματοποιεί 3 - 4 συναντήσεις το χρόνο σε διάφορες πόλεις του κόσμου, όπου συζητούνται οι εξελίξεις της ερευνητικής δραστηριότητας που έχει αναπτυχθεί στο διάστημα μεταξύ των συναντήσεων, θέτονται στόχοι και διατυπώνονται οι προδιαγραφές πάνω στις οποίες οι εταιρείες θα αναπτύξουν τα προϊόντα τους, ώστε να είναι συμβατά με τα αντίστοιχα πρότυπα. Τα αρχικά του ακρωνυμίου MPEG προέρχονται από τις λέξεις Moving Picture Experts Group (Ομάδα Ειδικών στην Κινούμενη Εικόνα). Ο όρος MPEG στην επιστημονική κοινότητα έχει επικρατήσει να αναφέρεται στην οικογένεια των προτύπων που δημιουργήθηκαν από την ομάδα ISO/IEC JTC1/SC29/WG11 και χρησιμοποιούνται ευρέως για τη μετάδοση οπτικών και ηχητικών δεδομένων σε ψηφιακή συμπιεσμένη μορφή.

Αρχές της δεκαετίας του 1990, το MPEG forum ξεκίνησε τις διαδικασίες ανάπτυξης ενός νέου προτύπου που θα επιτύγχανε συμπίεση μεγάλου μήκους ταινιών σε οπτικά μέσα όπως το Compact Disc, με ποιότητα εφάμιλλη του αναλογικού οικιακού προτύπου VHS. Έχοντας ως θεμέλιο λίθο το νεότερο πρότυπο του H.261, το MPEG Forum προέβη στην ολοκλήρωση του προτύπου MPEG-1: *Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s*. Σύμφωνα με τις προδιαγραφές, το πρότυπο κωδικοποίησης MPEG-1 ήταν ικανό να συμπίεσει εντός των ορίων χωρητικότητας ενός οπτικού δίσκου (700MB), μια ταινία μεγάλου μήκους σε ρυθμό μέχρι 1.5Mbps και χωρική ανάλυση CIF (352x288 για PAL και 352x240 για NTSC). Στα πλαίσια του προτύπου MPEG-1 έγινε αντιληπτό ότι για ένα κωδικοποιημένο βίντεο που είναι αποθηκευμένο σε ένα οπτικό μέσο για μελλοντική προβολή, δεν παίζει κάποιο ιδιαίτερο ρόλο ο παράγοντας της καθυστέρησης κατά τη διάρκεια της κωδικοποίησης και συμπίεσης αυτού. Δηλαδή σε αντίθεση με το πρότυπο H.261, που στοχεύει στην κωδικοποίηση και αποστολή του κωδικοποιημένου σήματος σε πραγματικό χρόνο στα πλαίσια κάποιας εφαρμογής τηλεδιάσκεψης (όπου τα περιθώρια για χρονική καθυστέρηση κατά την κωδικοποίηση είναι πολύ στενά), το MPEG-1 δεν προοριζόταν για εφαρμογές πραγματικού χρόνου.

Συνεπώς η πολυπλοκότητα της κωδικοποίησης του προτύπου MPEG-1 μπορούσε να αυξηθεί, ώστε να βελτιστοποιηθεί η αποδοτικότητα της κωδικοποίησης, εισάγοντας όμως χρονική καθυστέρηση κατά τη διάρκεια της κωδικοποίησης και αποκωδικοποίησης του σήματος. Για παράδειγμα στο χρονικό πεδίο, θα μπορούσε να χρησιμοποιηθεί αποκλειστικά ο μετασχηματισμός DCT αντί του DPCM, σε συνδυασμό με τη χρήση αντιστάθμισης της κίνησης (motion compensation), η οποία ήταν προαιρετική στο πρότυπο H.261, ώστε να διατηρείται η πολυπλοκότητα σε χαμηλά επίπεδα. Το πρότυπο MPEG-1 υιοθέτησε τη χρήση αντιστάθμισης της κίνησης, η οποία κατά τη διαδικασία κωδικοποίησης συγκρίνει τη κίνηση των αντικειμένων ανάμεσα στα προγενέστερα και μεταγενέστερα γειτονικά καρέ, συσχετίζοντας τα απεικονιζόμενα αντικείμενα και μειώνοντας το χρονικό πλεονασμό του σήματος. Έτσι, εισήχθησαν τα bi-directional καρέ, στα οποία η εκτίμηση της κίνησης και η αντιστάθμιση αυτής λαμβάνει χώρα ανάμεσα στο προηγούμενο και το επόμενο καρέ ως προς το υπό-κωδικοποίηση.

Το πρότυπο MPEG-1 είχε μεγάλη απήχηση στη χρήση ψηφιακών βίντεο στους υπολογιστές, αφού πολύ γρήγορα στην αγορά υπήρχαν διαθέσιμοι κωδικοποιητές υλικού και λογισμικού. Επιπλέον, στην αγορά της Κίνας το CDi (που είχε την εμπορική υποστήριξη της Philips) και το Video CD, που περιείχε μια ταινία μεγάλου μήκους κωδικοποιημένη σύμφωνα με το πρότυπο MPEG-1 σε ένα CD, γνώρισε πολύ μεγάλη επιτυχία. Πληθώρα κινηματογραφικών τίτλων ήταν διαθέσιμοι προς πώληση, ενώ ποικίλες ήταν και οι επιλογές οικιακών συσκευών αναπαραγωγής Video CD.

MPEG-2/H.262

Σταδιακά το πρότυπο MPEG-1 επεκτάθηκε ώστε να συμπεριλάβει υψηλότερους ρυθμούς κωδικοποίησης και μεγαλύτερη χωρική ανάλυση. Επιπλέον η επέκταση του προτύπου συμπεριέλαβε και πεπλεγμένη σάρωση, ώστε να είναι δυνατή η χρήση του σε τηλεοπτικές μεταδόσεις. Αν και αρχικά η τηλεοπτική βιομηχανία ήταν αντίθετη απέναντι σε οποιαδήποτε χρήση κωδικοποίησης/συμπίεσης του τηλεοπτικού σήματος, τελικά το μεταγενέστερο πρότυπο MPEG-2 υιοθετήθηκε στην κωδικοποίηση βίντεο, προσφέροντας ποιότητα εφάμιλλη της αντίστοιχης αναλογικής μετάδοσης.

Το πρότυπο MPEG-2: *Generic coding of moving pictures and associated audio information* αναπτύχθηκε στοχεύοντας στη ψηφιακή τηλεόραση. Η χωρική ανάλυση της εικόνας ακολουθεί το τηλεοπτικό πρότυπο CCIR-601 (broadcast quality - ποιότητα εκπομπής), δηλαδή 704x480 pixels (NTSC) ή 704x576 pixels (PAL) για εικόνα πλεκτής σάρωσης (interlaced). Ο ρυθμός μετάδοσης κυμαίνεται από 3 ως 10 Mbits/sec. Οι εφαρμογές που συναντάμε το MPEG-2 είναι στην καλωδιακή τηλεόραση (Cable TV), στη ψηφιακή δορυφορική τηλεόραση (Digital Video Broadcasting Satellite /DVB-S) αλλά και στην αντίστοιχη επίγεια ψηφιακή (Digital Video Broadcasting Terrestrial /DVB-T), η οποία στις αρχές του 2006 ξεκίνησε να εκπέμπει και στην ελληνική επικράτεια. Μάλιστα, τα κράτη μέλη της Ευρωπαϊκής Ένωσης έχουν δεσμευτεί ότι το έτος 2015 είναι η καταλυτική ημερομηνία για να έχει ολοκληρωθεί πλήρως η μετάβαση από την αναλογική στην επίγεια ψηφιακή τηλεόραση. Επίσης, το MPEG-2 έχει γνωρίσει μεγάλη επιτυχία στην αποθήκευση και διανομή κινηματογραφικών ταινιών μέσω του οπτικού δίσκου DVD (Digital Versatile Disc).

Στο πρότυπο MPEG-2, ο αριθμός των αμφίδρομα προβλεπόμενων καρέ (B frames) μπορεί να μεταβληθεί ώστε να επιτύχουμε την επιθυμητή πολυπλοκότητα και κατά συνέπεια χρονική υστέρηση κατά την κωδικοποίηση. Η τεχνική αυτή μπορεί να βρει εφαρμογή σε τηλεπικοινωνιακά συστήματα, όπου απαιτείται σε πραγματικό χρόνο η ψηφιοποίηση και συμπίεση του οπτικοακουστικού συστήματος. Εξαιτίας αυτής της δυνατότητας προσαρμογής της πολυπλοκότητας του προτύπου MPEG-2, η ITU-T υιοθέτησε το πρότυπο MPEG-2 υπό το γενικότερο όνομα H.262 και για τηλεπικοινωνιακές εφαρμογές.

Επιπλέον το πρότυπο MPEG-2/H.262 εισήγαγε την κλιμακωτή (scalable) κωδικοποίηση, όπου από μια μονή ροή MPEG-2 δύναται να εξαχθούν σήματα βίντεο διαφορετικής αντιληπτής ποιότητας (π.χ. με διαφορετική χωρική, χρονική ή ποιοτική ανάλυση) ανάλογα με το ποσοστό της ροής που χρησιμοποιείται από τον αποκωδικοποιητή. Η ιδιότητα αυτή της κλιμάκωσης έχει πολύ μεγάλη σημασία σε δικτυακές εφαρμογές. Από τη μεριά της υπηρεσίας και του παρόχου, τα ποικίλα επίπεδα κλιμάκωσης μπορούν να τιμολογηθούν διαφορετικά αφού αντιστοιχούν σε διαφορετική ποιότητα υπηρεσίας. Από τη μεριά του διαχειριστή δικτύου, η κλιμακωτή ροή MPEG-2 έχει το πλεονέκτημα ότι ανάλογα με την επικρατούσα δικτυακή κίνηση και τη πιθανή συμφόρηση μπορεί να παραλείψει συγκεκριμένα τμήματα δεδομένων, τα οποία αντιστοιχούν σε υψηλά χωρικά, χρονικά και

ποιοτικά επίπεδα ανάλυσης, με αποτέλεσμα ο τελικός χρήστης να μπορέσει να λάβει την υπηρεσία, έστω και ποιοτικά υποβαθμισμένη. Σε αντίθετη περίπτωση, η αποστολή απλής ροής MPEG-2 θα ισοδυναμούσε με μη επιτυχή λήψη της υπηρεσίας, και κατά συνέπεια την καθολική απώλεια αυτής.

MPEG-3

Μετά το πρότυπο MPEG-2, το οποίο προσέφερε κωδικοποίηση με ανάλυση ίση του τηλεοπτικού σήματος και υψηλό ρυθμό κωδικοποίησης, η προσπάθεια εστιάστηκε στην ανάπτυξη ενός νέου προτύπου, το οποίο θα εστιαζόταν στην κωδικοποίηση βίντεο σήματος υψηλής ευκρίνειας. Οι προσπάθειες για την ανάπτυξη και καθορισμό του νέου προτύπου, το οποίο έφερε το όνομα MPEG-3, τελικά εγκαταλείφθηκαν αφού η εμπορική επιτυχία του MPEG-2 και η επέκταση των δυνατοτήτων αυτού σύμφωνα με τις τροποποιήσεις του αρχικού προτύπου να κωδικοποιεί βίντεο σε αναλύσεις υψηλής ευκρίνειας δεν άφησε κανένα ουσιαστικό περιθώριο για περαιτέρω ενασχόληση με το MPEG-3 και οι εργασίες εξέλιξης και προτυποποίησης αυτού εγκαταλείφθηκαν.

MPEG-4

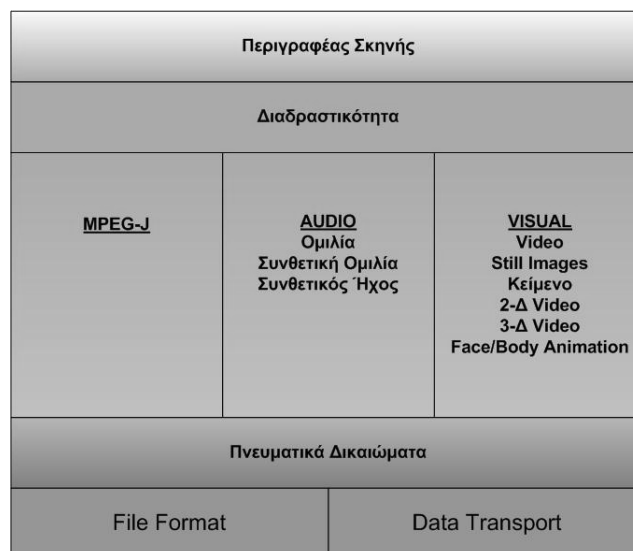
Στη συνέχεια, ο στόχος της MPEG και της ITU ήταν η ανάπτυξη νέων προτύπων, τα οποία θα παρείχαν υψηλής ποιότητας κωδικοποιημένο βίντεο σε πολύ χαμηλό ρυθμό κωδικοποίησης (της τάξης των 64kbps ή και λιγότερο). Επιπλέον άρχισε να διαφαίνεται σταδιακά η ανάγκη για ένα πρότυπο, το οποίο θα μπορούσε να συμπεριλάβει εκτός από τον υψηλό λόγο συμπίεσης φυσικού βίντεο και άλλες πολυμεσικές τεχνικές απεικόνισης, όπως συνδυασμός φυσικού και συνθετικού βίντεο, διαδραστικό βίντεο και κωδικοποίηση βίντεο με δυνατότητα επιλογής γωνίας θέασης.

Η MPEG, προς αυτή την κατεύθυνση ανέπτυξε το πρότυπο MPEG-4, το οποίο εκτός από καλύτερη απόδοση σε χαμηλούς ρυθμούς κωδικοποίησης, συμπεριλάμβανε και μια πληθώρα άλλων εργαλείων για την κωδικοποίηση πολυμεσικών μορφών βίντεο. Το MPEG-4 εισάγει για πρώτη φορά την έννοια της αντικειμενοστραφούς κωδικοποίησης βίντεο. Πιο συγκεκριμένα, ο όρος οπτικοακουστικά αντικείμενα χρησιμοποιείται για να περιγράψει τις διάφορες οντότητες που απαρτίζουν την εικόνα, οι οποίες μπορούν να διαχειριστούν αυτόνομα και ανεξάρτητα από τις υπόλοιπες, τόσο από τον κωδικοποιητή όσο και από τον αποκωδικοποιητή. Όπως φαίνεται και στο Σχήμα 2.1, με τον όρο οντότητες εννοούμε φυσικά ή συνθετικά αντικείμενα και ήχους που όλα μαζί συνθέτουν το οπτικοακουστικό σήμα (πχ. Ήχος, Αφήγηση, Συνθετική Αφήγηση, Συνθετικό Ήχος, Βίντεο, Εικόνες, Δισδιάστατα γραφικά, Τρισδιάστατα Γραφικά, Κείμενο, Συνθετική απεικόνιση προσώπου και σώματος). Το πρότυπο MPEG-4 συνδυάζει όλα αυτά τα φυσικά ή συνθετικά αντικείμενα σύμφωνα με ένα περιγραφέα σκηνής και με χρήση κατάλληλου συγχρονισμού, ώστε να αναπαραστήσει σωστά την αρχική εικόνα που αναλύθηκε και εισήχθη προς κωδικοποίηση. Η χωρική ανάλυση της εικόνας μπορεί να είναι ακόμη και στα χαμηλά επίπεδα των 176x144 pixels (QCIF) σε σχετικά χαμηλούς ρυθμούς κωδικοποίησης που κυμαίνονται ανάμεσα στα 48 και 128kbps, κατάλληλα δηλαδή για μετάδοση σε δίκτυα με περιορισμένο εύρος ζώνης ανά συνδρομητή, όπως τα δίκτυα κινητών επικοινωνιών.

Συνεπώς, το πρότυπο MPEG-4 κάλυψε και μια πληθώρα νέων πολυμεσικών εφαρμογών, όπως η κωδικοποίηση βίντεο με τη δυνατότητα πολλαπλών σκηνών θέασης, συνθετικό βίντεο, συνδυασμός συνθετικού και πραγματικού βίντεο, διαδραστικό βίντεο κλπ. Οι

ποικίλες αυτές εφαρμογές ενσωματώθηκαν στο πρότυπο του MPEG-4 μέσω διαφόρων εργαλείων κωδικοποίησης, τα οποία είναι οργανωμένα σε προφίλ (profile). Το κάθε προφίλ προσδιορίζει και ορίζει ένα προτεινόμενο σύνολο εργαλείων κωδικοποίησης του προτύπου, που είναι κατάλληλα για συγκεκριμένες εφαρμογές. Η ταξινόμηση των προφίλ περιλαμβάνει:

- Απλά προφίλ που περιλαμβάνουν εκείνα τα εργαλεία που απαιτούνται για τη κωδικοποίηση φυσικών τετράγωνων καρτέ
- Αντικειμενοστραφή προφίλ για την κωδικοποίηση οπτικών αντικειμένων τυχαίου σχήματος
- Προφίλ ακίνητης εικόνας για την κωδικοποίηση ακίνητων εικόνων
- Κλιμακωτά προφίλ για τη κωδικοποίηση σε πολλαπλά χωρικά, χρονικά και ποιοτικά επίπεδα και
- Προφίλ στούντιο για τη κωδικοποίηση υψηλής ποιότητας βίντεο για επαγγελματικές εφαρμογές



Σχήμα 2.1 Τα συστατικά μιας ροής MPEG-4

Συνεπώς, κάθε υλοποίηση MPEG-4 είναι συμβατή με ορισμένα προφίλ του προτύπου, που σημαίνει ότι μπορεί να κωδικοποιήσει/αποκωδικοποιήσει μόνο περιεχόμενο MPEG-4 κωδικοποιημένο σύμφωνα με το συγκεκριμένο προφίλ. Επιπλέον το κάθε προφίλ υποδιαιρείται σε διαφορετικά επίπεδα (levels), τα οποία ορίζουν διάφορα επίπεδα πολυπλοκότητας τόσο σε χωρική και χρονική ανάλυση, όσο και σε ρυθμό κωδικοποίησης.

Οπότε, σύμφωνα με το σχήμα 2.1, η παραγόμενη ροή MPEG-4 ανάλογα με το προφίλ και το επίπεδο περιλαμβάνει τα αντίστοιχα οπτικοακουστικά αντικείμενα, τον περιγραφέα σκηνής καθώς, πιθανές διαδραστικές υπηρεσίες και πνευματικά δικαιώματα. Στη συνέχεια τα δεδομένα αυτά οργανώνονται με τέτοιο τρόπο ώστε είτε να αποθηκευτούν τοπικά σύμφωνα με κάποια συγκεκριμένη δομή αρχείου ή να σταλούν σαν ροή δεδομένων πάνω από κάποιο τηλεπικοινωνιακό δίκτυο.

Αυτή τη στιγμή το πρότυπο MPEG-4 περιλαμβάνει 23 μέρη, εκ των οποίων τα μέρη 1-8 και 10 ήταν δημοσιευμένα στο αρχικό πρότυπο 14496, ενώ τα υπόλοιπα μέρη ήταν ακόμη υπό

ανάπτυξη και δεν είχαν καταλήξει στην τελική τους μορφή. Εξαιτίας του γεγονότος ότι το MPEG-4 είναι αυτή τη στιγμή το πιο ευρέως διαδεδομένο, υλοποιήσιμο και εφαρμόσιμο πρότυπο, για το λόγο αυτό θα αναφερθούμε αναλυτικά στα μέρη του MPEG-4 που περιλαμβάνονται στο πρότυπο, καθώς και σε αυτά που βρίσκονται υπό ανάπτυξη (σε παρένθεση σημειώνεται η χρονολογία της τελευταίας ενημερωμένης έκδοσης):

- Μέρος 1, Systems (2004): Περιγραφέας σκηνής, πολυπλεξία ήχου, βίντεο και σχετικής πληροφορίας, συγχρονισμός, διαχείριση του buffer, διαχείριση πνευματικών δικαιωμάτων.
- Μέρος 2, Visual (2004): Κωδικοποίηση φυσικών και συνθετικών οπτικών αντικειμένων.
- Μέρος 3, Audio (2005): Κωδικοποίηση φυσικών και συνθετικών ηχητικών αντικειμένων.
- Μέρος 4, Conformance Testing (2004): Συνθήκες προσαρμογής, διαδικασίες δοκιμών, ροές δοκιμών.
- Μέρος 5, Reference Software (2001): Το δημόσια διαθέσιμο λογισμικό αναφοράς το οποίο υλοποιεί την πληθώρα των περιγραφόμενων εργαλείων κωδικοποίησης του προτύπου.
- Μέρος 6, Delivery Multimedia Integration Framework (2000): Περιγράφεται το πλαίσιο για την παράδοση / ενοποίηση πολυμεσικών ροών.
- Μέρος 7, Optimised Visual Reference Software (2004): Λογισμικό βελτιστοποίησης για την υλοποίηση επιλεγμένων εργαλείων κωδικοποίησης. Το μέρος αυτό είναι τεχνική αναφορά και όχι πρότυπο.
- Μέρος 8, Carriage of MPEG-4 over IP (2004): Καθορίζει το μηχανισμό μεταφοράς MPEG-4 δεδομένων πάνω από δίκτυα IP.
- Μέρος 9, Reference Hardware Description (2004): Περιγραφή VHDL των εργαλείων κωδικοποίησης MPEG-4. Το μέρος αυτό είναι τεχνική αναφορά και βρίσκεται ακόμη σε φάση ανάπτυξης.
- Μέρος 10, Advanced Video Coding (2005): Αποδοτική κωδικοποίηση φυσικού βίντεο με πολύ χαμηλό ρυθμό. Το μέρος αυτό αρχικά βρισκόταν σε διαδικασία ανάπτυξης έως το 2004 που ορίστηκε ως νέο πρότυπο (H.264/MPEG-4 Part 10).
- Μέρος 11, Scene Description and Application Engine (2005): Σε αυτό το μέρος καθορίζονται τα ακόλουθα:
 - Η κωδικοποιημένη αναπαράσταση της χωροχρονικής θέσης των οπτικοακουστικών αντικειμένων καθώς και η διαδραστική τους σχέση (περιγραφή σκηνής)
 - Η κωδικοποιημένη αναπαράσταση δισδιάστατων και τρισδιάστατων συνθετικών αντικειμένων, τα οποία μπορούν να εκδηλωθούν οπτικά ή ηχητικά.
 - Ο καθορισμός της Extensible MPEG-4 Textual (XMT), μιας αναπαράστασης του πολυμεσικού υλικού που περιγράφεται στο πρότυπο και είναι βασισμένη στην Extensible Markup Language (XML)

- Μέρος 12, ISO Base Media File Format (2006): Το μέρος αυτό του προτύπου είναι κοινό με το μέρος 12 του JPEG 2000, ISO/IEC 15444, όπου καθορίζεται η διαμόρφωση και οι χρήσεις της δομής του αρχείου που ορίζεται από το πρότυπο. Η δομή του αρχείου καθορίζεται με τέτοιο τρόπο ώστε να είναι δυνατή η μεταφορά πολυμεσικού περιεχομένου. Επιπλέον, έχει ληφθεί ειδική μέριμνα ώστε η αναπαράσταση του περιεχομένου να γίνεται με έναν ευέλικτο και επεκτάσιμο τρόπο που διευκολύνει την ανταλλαγή, διαχείριση, επεξεργασία και παρουσίασή του. Η αναπαράσταση του περιεχομένου μπορεί να είναι είτε τοπική σε ένα σύστημα που υποστηρίζει αναπαραγωγή πολυμεσικού περιεχομένου, είτε δικτυακή ή με κάποιο άλλο μηχανισμό ροών βίντεο. Η δομή του αρχείου είναι αντικειμενοστραφής και σχεδιασμένη ώστε να είναι ανεξάρτητη από οποιοδήποτε δικτυακό πρωτόκολλο επικοινωνίας, χωρίς να παραβλέπει όμως αποδοτική υποστήριξη αυτών.
- Μέρος 13, Intellectual Property Management and Protection Extensions - IPMP (2004): Το μέρος αυτό περιγράφει τον ορισμό, το συντακτικό και σημασιολογία των επεκτάσεων IPMP.
- Μέρος 14, MPEG-4 File Format (2003): Πρόκειται για τον καθορισμό της μορφής του πολυμεσικού μεταφορέα, ο οποίος περιέχει και διανέμει πολυμεσικό υλικό κωδικοποιημένο σύμφωνα με το πρότυπο MPEG-4. Όπως παρόμοιοι σύγχρονοι πολυμεσικοί μεταφορείς, το MPEG-4 Part 14 προβλέπει και επιτρέπει τη μεταφορά πολυμεσικού περιεχομένου σε πραγματικό χρόνο μέσω κάποιου τηλεπικοινωνιακού δικτύου. Η επίσημη επέκταση του αρχείου μεταφοράς είναι .mp4, ενώ ανεπίσημα ο μεταφορέας πολλές φορές ονοματίζεται ως MP4, ο οποίος έχει τις ρίζες του στον μεταφορέα της Apple Quicktime με συγκεκριμένες τροποποιήσεις.
- Μέρος 15, AVC File Format (2004): Το μέρος αυτό καθορίζει πώς περιεχόμενο που είναι κωδικοποιημένο σύμφωνα με το πρότυπο MPEG-4 Part 10 αποθηκεύεται σε αρχείο, ώστε να αυξηθούν οι δυνατότητες διανομής του περιεχομένου και διαλειτουργικότητας σε διαφορετικές συσκευές. Επιπλέον, το μέρος αυτό προβλέπει πώς ροές AVC μπορούν να συνδυαστούν χρησιμοποιώντας hint tracks με άλλο πολυμεσικό περιεχόμενο, όπως ήχος και πώς μπορεί περιεχόμενο AVC να προετοιμαστεί κατάλληλα ώστε να είναι δυνατή η διανομή του σε πραγματικό χρόνο από εξυπηρετητή ροών.
- Μέρος 16, Animation Framework Extension (2006): Το μέρος αυτό καθορίζει το MPEG-4 Animation Framework extension (AFX) για την αναπαράσταση γραφικών τριών διαστάσεων. Με τη χρήση αυτού του μοντέλου εργασίας, το MPEG-4 επεκτείνεται πέρα από την παραδοσιακή κωδικοποίηση φυσικού βίντεο, με τη χρήση υψηλού επιπέδου συνθετικών αντικειμένων.
- Μέρος 17, Timed Text Subtitle Format (2006): Το μέρος αυτό αναπτύχθηκε ως απάντηση στην ανάγκη που υπήρχε για κωδικοποίηση κειμένου σε εξαιρετικά χαμηλό ρυθμό, όντας συστατικό της οπτικοακουστικής αναπαράστασης. Πιο συγκεκριμένα, το μέρος αυτό του προτύπου προβλέπει πώς υπότιτλοι και Karaoke στίχοι μπορούν να κωδικοποιηθούν και να διανεμηθούν ως ξεχωριστές ροές κειμένου σε πολύ χαμηλό ρυθμό. Βεβαίως, παρά τον πρωταρχικό στόχο των εφαρμογών αυτών που είναι οι κινητοί χρήστες, υπηρεσίες πάνω από δίκτυα IP ευρείας ζώνης επίσης θα ωφεληθούν, αφού δεν θα σπαταλείται διαθέσιμο εύρος ζώνης.

- Μέρος 18, Font Compression and Streaming (2004): Ως φυσική συνέχεια του προηγούμενου μέρους, το παρόν μέρος καλύπτει τη δυνατότητα κωδικοποίησης και διανομής συγκεκριμένων δεδομένων που σχετίζονται με την αναπαράσταση γραμματοσειρών.
- Μέρος 19, Synthesized Texture Stream (2004): Σε αυτό το μέρος καθορίζεται η μετάδοση δεδομένων συνθετικής σύστασης/αναπαράστασης ως τμήμα οπτικοακουστικού περιεχομένου, κωδικοποιημένου σύμφωνα με το πρότυπο MPEG-4. Πιο συγκεκριμένα καθορίζει
 - Τη δομή των δεδομένων της συνθετικής σύστασης/αναπαράστασης
 - Τις μεθόδους κίνησης της συνθετικής σύστασης/αναπαράστασης
 - Την κωδικοποιημένη αναπαράσταση ροών δεδομένων της συνθετικής σύστασης/αναπαράστασης.
- Μέρος 20, Lightweight Scene Representation (LSeR) (2006): Το μέρος αυτό καθορίζει τη μορφή ενός περιγραφέα σκηνών (LSeR) και μια συσσωρευτική μορφή (SAF) κατάλληλη για αναπαράσταση και παράδοση εμπλουτισμένων πολυμεσικών υπηρεσιών σε τερματικές συσκευές περιορισμένων επεξεργαστικών δυνατοτήτων (όπως τα κινητά τηλέφωνα τρίτης γενιάς). Ο όρος εμπλουτισμένη πολυμεσική υπηρεσία αναφέρεται σε ένα εύρος εφαρμογών από ταινίες εμπλουτισμένες με διαδραστικές δυνατότητες έως πολύπλοκες διαδραστικές εφαρμογές πολλαπλών βημάτων, όπου σε κάθε βήμα διαφορετικό πολυμεσικό είδος χρησιμοποιείται.
- Μέρος 21, MPEG-J Graphical Framework eXtension (GFX) (2006): Περιγράφει ένα περιβάλλον προγραμματισμού για διαδραστικές πολυμεσικές εφαρμογές. Έχει σχεδιαστεί για την ανάπτυξη εφαρμογών σε συσκευές περιορισμένων επεξεργαστικών πόρων όπως τα κινητά τηλέφωνα. Πιο συγκεκριμένα το GFX προσφέρει ένα πλαίσιο που συνδυάζει ένα μέρος των εφαρμογών Java της MPEG (MPEG-J) με Java APIs για τρισδιάστατη αναπαράσταση συνθετικών μορφών. Έτσι δίνεται η δυνατότητα στους προγραμματιστές να αναπτύξουν εφαρμογές που συνδυάζουν οπτικοακουστικές MPEG ροές με τρισδιάστατα συνθετικά γραφικά και αλληλεπίδραση του χρήστη.
- Μέρος 22, Open Font Format Specification (OFFS) based on OpenType (υπό ανάπτυξη)
- Μέρος 23, Symbolic Music Representation (SMR) (υπό ανάπτυξη)

Με την προαναφερθείσα σύντομη περιγραφή του προτύπου MPEG-4 ISO 14496, είναι εύκολο να γίνει αντιληπτό ότι το MPEG-4 περιλαμβάνει μια μεγάλη γκάμα εφαρμογών και εργαλείων, όπου το καθένα στοχεύει σε διαφορετικό είδος εφαρμογών και υπηρεσιών. Συνεπώς, η υλοποίηση όλων των μερών του προτύπου ταυτοχρόνως δεν θα εξυπηρετούσε κανένα συγκεκριμένο σκοπό και θα οδηγούσε σε μια πολύπλοκη και μη εφαρμόσιμη υλοποίηση. Για το λόγο αυτό, εισάγεται η έννοια των προφίλ και των επιπέδων, ώστε να διευκρινιστούν τα μέρη του προτύπου που υλοποιούνται από κάθε εφαρμογή. Με τον τρόπο αυτό επιτυγχάνουμε συμβατότητα και διαλειτουργικότητα ανάμεσα σε διαφορετικές υλοποιήσεις, που είναι συμβατές με ένα συγκεκριμένο προφίλ του προτύπου.

Στη συνέχεια παραθέτουμε το σύνολο των προφίλ που ορίζονται από το πρότυπο του MPEG-4, καθώς και μια σύντομη περιγραφή των χαρακτηριστικών που υλοποιούν:

- Simple Profile: Χαμηλής πολυπλοκότητας κωδικοποίηση τετράγωνων καρέ βίντεο
- Advanced Simple Profile: Κωδικοποίηση τετράγωνων καρέ βίντεο με βελτιωμένη αποδοτικότητα και υποστήριξη βίντεο πεπλεγμένης σάρωσης.
- Advanced Real-time Simple Profile: Κωδικοποίηση τετράγωνων καρέ βίντεο για παράδοση σε πραγματικό χρόνο μέσω ροών βίντεο.
- Core Profile: Βασική κωδικοποίηση αντικειμένων βίντεο τυχαίας μορφής.
- Main Profile: Κωδικοποίηση με πληθώρα χαρακτηριστικών αντικειμένων βίντεο.
- Advanced Coding Efficiency Profile: Υψηλής αποδοτικότητας κωδικοποίησης αντικειμένων βίντεο.
- N-Bit Profile: Κωδικοποίηση αντικειμένων βίντεο με ανάλυση δείγματος διάφορη των 8 bits.
- Simple Scalable Profile: Κλιμακωτή κωδικοποίηση τετράγωνων καρέ βίντεο.
- Fine Granular Scalability Profile: Υψηλού επιπέδου κωδικοποίηση τετράγωνων καρέ βίντεο.
- Core Scalable Profile: Κλιμακωτή κωδικοποίηση αντικειμένων βίντεο.
- Scalable Texture Profile: Κλιμακωτή κωδικοποίηση ακίνητης σύστασης.
- Advanced Scalable Texture Profile: Κλιμακωτή κωδικοποίησης ακίνητης σύστασης με βελτιωμένη αποδοτικότητα και αντικειμενοστροφή χαρακτηριστικά.
- Advanced Core Profile: Συνδυάζει χαρακτηριστικά των προφίλ Simple, Core και Advanced Scalable Texture.
- Simple Studio Profile: Αντικειμενοστροφής κωδικοποίηση βίντεο υψηλής ποιότητας.
- Core Studio Profile: Αντικειμενοστροφής κωδικοποίηση βίντεο υψηλής ποιότητας με βελτιωμένη αποδοτικότητα συμπίεσης.
- Basic Animated Texture Profile: Δισδιάστατη κωδικοποίηση πλέγματος με ακίνητη σύσταση.
- Simple Face Animation Profile: Μοντέλα ανθρώπινου προσώπου με κίνηση.
- Simple Face and Body Animation Profile: Μοντέλα ανθρώπινου προσώπου και σώματος με κίνηση.
- Hybrid Profile: Συνδυασμός των προφίλ Simple, Core, Basic Animated Texture και Simple Face Animation.

Οι κατακριτές του προτύπου MPEG-4 υποστηρίζουν ότι η χρήση πολλαπλών προφίλ και τεχνικών ήταν μια έντεχνη παρέκκλιση από τον αρχικό στόχο, που ήταν η ανάπτυξη ενός προτύπου που θα επιτύγχανε ικανοποιητική ποιότητα κωδικοποίησης σε χαμηλό ρυθμό. Επειδή το νέο πρότυπο δεν κατάφερε να βελτιστοποιήσει σημαντικά το λόγο ποιότητας έναντι του ποσοστού συμπίεσης, η εισαγωγή επιπλέον πολυμεσικών δυνατοτήτων διέσωσαν το νέο πρότυπο από την αποτυχία. Παρόλα αυτά τα ενδεχομένως βάσιμα σχόλια, θα ήταν πραγματικά κακοπροαίρετη πράξη να μην αναγνωριστεί η σημαντική συνεισφορά του προτύπου MPEG-4 στην εξέλιξη και διάδοση των προτύπων κωδικοποίησης ψηφιακού βίντεο, κυρίως μέσω των κινητών τερματικών.

MPEG-4 AVC Part 10 / H.264

Τα MPEG-4 Visual και MPEG-4 AVC (H.264) μοιράζονται την ίδια καταγωγή και μερικά κοινά χαρακτηριστικά γνωρίσματα (και τα δύο ακολουθούν τις αποδεδειγμένες τεχνικές των προηγούμενων προτύπων), για αυτό και το H.264 αναφέρεται ως επέκταση του MPEG-4. Παρόλα αυτά, τα δύο πρότυπα έχουν σημαντικά διαφορετικές οπτικές, προσπαθώντας να βελτιώσουν τα παλαιότερα πρότυπα με διαφορετικούς τρόπους.

Ο σκοπός του MPEG-4 Visual, είναι να επεκταθεί πέρα από την κωδικοποίηση ορθογώνιων κινούμενων εικόνων βίντεο και να παράσχει ένα ανοικτό, εύκαμπτο πλαίσιο για τις πολυμεσικές εφαρμογές. Αντίθετα, το H.264 εξακολουθεί να στοχεύει στη συμπίεση των ορθογώνιων εικόνων βίντεο, όπως και τα προηγούμενα πρότυπα, αλλά να το κάνει με έναν αποδοτικότερο, ανθεκτικό σε λάθη και αποτελεσματικό τρόπο.

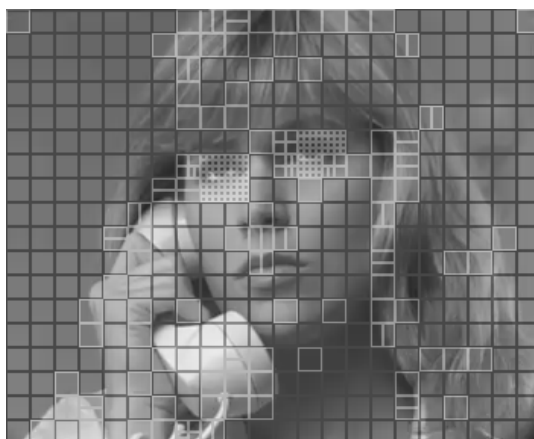
Το MPEG-4 Visual είναι ο ωριμότερος των δύο προτύπων (η πρώτη έκδοσή του δημοσιεύθηκε το 1999, ενώ το H.264 έγινε διεθνές πρότυπο με τη Recommendation του 2003). Δεν υπάρχει καμία αμφιβολία ότι το H.264 μπορεί να ξεπεράσει το MPEG-4 Visual στην αποδοτικότητα συμπίεσης αλλά δεν έχει ακόμη τη διαλειτουργικότητα (interoperability) των παλαιότερων προτύπων. Επιπλέον, ενώ η πολιτική χορήγησης αδειών όσον αφορά το MPEG-4 Visual είναι γνωστή, το κόστος της χορήγησης αδειών του H.264 δεν έχει ακόμη καθοριστεί, παρόλο που γίνονται σαφείς κινήσεις προς την ελεύθερη αδειοδότηση των απλών/βασικών προφίλ του.

Στον πίνακα 2.1 απεικονίζεται μια λεπτομερής περιγραφή της εξέλιξης των προτύπων MPEG-4 και H.263 μέχρι την από κοινού γέννηση του προτύπου MPEG-4 AVC/H.264. Το 1993 ξεκίνησε η ανάπτυξη του προτύπου MPEG-4, ενώ την ίδια χρονιά τα πρώτα αποτελέσματα του προτύπου H.263 άρχισαν να προκύπτουν. Στη συνέχεια το 1995, το MPEG-4 κάνει κάλεσμα για προτάσεις αποδοτικών μεθόδων κωδικοποίησης βίντεο, ενώ το H.263 επιλέγεται ως βασικό εργαλείο κωδικοποίησης. Το 1998, πραγματοποιείται το κάλεσμα για το project H.26L. Το 1999 δημοσιεύεται η πρώτη έκδοση του προτύπου MPEG-4 και ορίζεται το πρώτο δοκιμαστικό μοντέλο του H.26L. Το 2000 η MPEG κάνει κάλεσμα για προτάσεις σχετικές με υψηλής αποδοτικότητας μεθόδους κωδικοποίησης βίντεο. Το 2001, δημοσιεύεται η δεύτερη έκδοση του προτύπου MPEG-4 και το πρότυπο H.26L υιοθετείται ως βάση για το MPEG-4 Part 10. Επίσης την ίδια χρονιά δημιουργείται η JVT, ενώ το 2002 πραγματοποιούνται δύο τροποποιήσεις στη δεύτερη έκδοση του προτύπου του MPEG-4. Το 2003 δημοσιεύεται το H.264/MPEG-4 Part 10.

Έτος	Γεγονός
1993	Ξεκίνησε η ανάπτυξη του προτύπου MPEG-4. Πρώτα αποτελέσματα του προτύπου H.263. το MPEG-4 κάνει κάλεσμα για προτάσεις αποδοτικών μεθόδων κωδικοποίησης βίντεο.
1995	Το H.263 επιλέγεται ως βασικό εργαλείο κωδικοποίησης.
1998	Κάλεσμα για το project H.26L
1999	Δημοσιεύεται η πρώτη έκδοση του προτύπου MPEG-4. Ορίζεται το πρώτο δοκιμαστικό μοντέλο του H.26L
2000	Η MPEG κάνει κάλεσμα για προτάσεις υψηλής αποδοτικότητας βίντεο κωδικοποίησης
2001	Δημοσιεύεται η δεύτερη έκδοση του προτύπου MPEG-4. Το πρότυπο H.26L υιοθετείται ως βάση για το MPEG-4 Part 10. Δημιουργείται η JVT.
2002	Πραγματοποιήθηκαν δύο τροποποιήσεις στη δεύτερη έκδοση του προτύπου του MPEG-4
2003	Δημοσιεύεται το H.264/MPEG-4 Part 10

Πίνακας 2.1 Εξέλιξη της δημοσίευσης του προτύπου H.264

Στον πίνακα 2.2 απεικονίζονται σε συνοπτική μορφή οι βελτιώσεις του προτύπου H.264 έναντι του MPEG-4. Όπως μπορεί να παρατηρηθεί, το H.264 στοχεύει στην αποδοτική συμπίεση φυσικού βίντεο, χωρίς να περιλαμβάνει προφίλ σχετικά με την κωδικοποίηση συνθετικού ή διαδραστικού περιεχομένου. Εξαιτίας αυτής της εξειδίκευσης, το σύνολο των διαθέσιμων προφίλ είναι μόλις τρία σε σύγκριση με τα δέκα εννιά του προτύπου MPEG-4, εστιάζοντας στην υψηλή συμπίεση φυσικού βίντεο. Για να επιτευχθεί υψηλός λόγος συμπίεσης, διατηρώντας ταυτόχρονα σε ικανοποιητικό επίπεδο την αντιληπτή ποιότητα, η κωδικοποίηση του βίντεο όσο και η αντιστάθμιση της κίνησης πραγματοποιείται με τη χρήση μεταβλητού μεγέθους block που το μικρότερο μέγεθος φτάνει τα 4x4 pixels. Στο σχήμα 2.2 απεικονίζεται ένα παράδειγμα κωδικοποίησης του προτύπου H.264 με τη χρήση blocks μεταβλητών μεγεθών ανάλογα με τη χωρική πολυπλοκότητα του περιεχομένου.



Σχήμα 2.2 Παράδειγμα κωδικοποίησης H.264 μεταβλητού block

Επιπλέον το πρότυπο H.264 ενσωματώνει στην αρχιτεκτονική του αποκωδικοποιητή φίλτρο εξομάλυνσης των οπτικών τεχνουργημάτων όπως το φαινόμενο τετραγώνισης. Με τον

τρόπο αυτό επιτυγχάνεται βελτιστοποίηση του αποκωδικοποιημένου οπτικού σήματος και της τελικής αντιληπτής ποιότητας. Τέλος, το βασικό προφίλ του νέου προτύπου θα μπορεί να χρησιμοποιηθεί χωρίς να απαιτείται η αγορά αντίστοιχης άδειας, γεγονός που θα συμβάλει καταλυτικά στην εξάπλωση και στην υιοθέτηση του νέου προτύπου σε ποικίλες εμπορικές εφαρμογές. Σε αυτά τα πλαίσια το νέο πρότυπο έχει ήδη υιοθετηθεί στα DVD επόμενης γενιάς (Blue Ray), τα οποία θα διαδεχθούν τα ευρέως διαδεδομένα DVD συμβατικής ευκρίνειας που χρησιμοποιούν το πρότυπο MPEG-2.

Σύγκριση	MPEG-4 Visual	H.264
Υποστηριζόμενα είδη κωδικοποίησης	Τετράγωνα καρέ και πεδία, αντικείμενα βίντεο τυχαίας μορφής, textures, sprites, συνθετικά αντικείμενα βίντεο, 2-Δ και 3-Δ αντικείμενα mesh	Τετράγωνα καρέ και πεδία
Συνολικό πλήθος προφίλ	19	3
Αποδοτικότητα συμπίεσης	Μεσαία	Υψηλή
Υποστήριξη βίντεο ροής	Scalable coding	Switching slices
Ελάχιστο μέγεθος block για αντιστάθμιση κίνησης	8x8	4x4
Ακρίβεια διανύσματος κίνησης	Μισό ή ένα τέταρτο του Pixel	Τέταρτο του Pixel
Μετασχηματισμός	8x8 DCT	4X4 DCT approximation
Ενσωματωμένο φίλτρο κατά του φαινομένου τετραγώνισης	-	√
Απαιτήση για αγορά άδειας στα πλαίσια εμπορικής εφαρμογής	√	Δεν απαιτείται για το βασικό προφίλ (Baseline Profile)

Πίνακας 2.2 Σύγκριση των χαρακτηριστικών του MPEG-4 και H.264

Η αναθεωρημένη έκδοση του προτύπου το 2007 έχει επεκτείνει τα διαθέσιμα προφίλ από τα τρία που είχαν περιγραφεί αρχικά σε επτά, ώστε να συμπεριληφθούν προφίλ υψηλής ευκρίνειας και επαγγελματικών εφαρμογών. Σύμφωνα με την τελευταία έκδοση του προτύπου, τα προφίλ του H.264 είναι τα εξής:

- **Baseline Profile (BP):** Το προφίλ αυτό προορίζεται για χαμηλού κόστους εφαρμογές, όπου υπάρχουν περιορισμένοι διαθέσιμοι υπολογιστικοί πόροι. Ήδη το προφίλ αυτό χρησιμοποιείται ευρέως σε εφαρμογές τηλεδιάσκεψης καθώς και σε κινητά τερματικά.
- **Main Profile (MP):** Αρχικά το προφίλ αυτό προοριζόταν για εφαρμογές εκπομπής τηλεοπτικών προγραμμάτων. Με τη ευρύτερη τάση όμως για μετάβαση σε σήματα υψηλής ευκρίνειας, σταδιακά το προφίλ αυτό παραμερίστηκε και επικράτησε το μεταγενέστερο High Profile.
- **Extended Profile (XP):** Το προφίλ αυτό προορίζεται για εφαρμογές ροής βίντεο σε δίκτυα υπολογιστικών και κινητών επικοινωνιών, περιλαμβάνοντας υψηλή συμπίεση

και ανοχή στα λάθη, καθώς και εναλλαγή των χαρακτηριστικών της ροής σε πραγματικό χρόνο.

- High Profile (HiP): Πρόκειται για το πρωταρχικό προφίλ για εφαρμογές εκπομπής ψηφιακού τηλεοπτικού σήματος υψηλής ευκρίνειας και αποθήκευσης αυτού σε οπτικά μέσα. Πρόκειται για το προφίλ που έχει ήδη υιοθετηθεί από το Blu-ray Disc.
- High 10 Profile (Hi10P): Το προφίλ αυτό επεκτείνει το υπάρχον High Profile κάνοντας ένα βήμα πέρα από τις παρούσες δυνατότητες των καταναλωτικών προϊόντων. Η επέκταση αυτή προβλέπει την υποστήριξη μέχρι και 10 bits –αντί του συμβατικού μήκους των 8 bits- ανά δείγμα της αποκωδικοποιημένης εικόνας.
- High 4:2:2 Profile (Hi422P): Επεκτείνοντας επιπλέον το High 10 Profile, το προφίλ αυτό στοχεύει σε βίντεο πλεκτής σάρωσης, υποστηρίζοντας χρωματική δειγματοληψία 4:2:2 και μέχρι 10 bits ανά δείγμα της αποκωδικοποιημένης εικόνας.
- High 4:4:4 Predictive Profile (Hi444PP): Το προφίλ αυτό επεκτείνει ακόμη περισσότερο το High 4:2:2 Profile, προσθέτοντας υποστήριξη για χρωματική δειγματοληψία 4:4:4 και μέχρι 14 bits ανά δείγμα. Επιπλέον προβλέπει τη χρήση μη απωλεστικής κωδικοποίησης, καθώς και τη κωδικοποίηση του κάθε καρέ ξεχωριστά στα τρία χρωματικά πεδία.

Επιπροσθέτως, το πρότυπο προβλέπει και τέσσερα εξειδικευμένα προφίλ που περιλαμβάνουν μόνο κωδικοποίηση Intra και ορίζονται ως υποσύνολα των προηγούμενων προφίλ. Τα εξειδικευμένα αυτά υποσύνολα προορίζονται για επαγγελματικές εφαρμογές και είναι τα ακόλουθα:

- High 10 Intra Profile: Πρόκειται για το High 10 Profile με αποκλειστική χρήση Intra κωδικοποίησης.
- High 4:2:2 Intra Profile: Πρόκειται για το High 4:2:2 Profile με αποκλειστική χρήση Intra κωδικοποίησης.
- High 4:4:4 Intra Profile: Πρόκειται για το High 4:4:4 Profile με αποκλειστική χρήση Intra κωδικοποίησης.
- CAVLC 4:4:4 Intra Profile: Πρόκειται για το High 4:4:4 Profile με αποκλειστική χρήση Intra κωδικοποίησης και CAVLC ως μέθοδο εντροπικής κωδικοποίησης.

Στον πίνακα 2.3 παρουσιάζονται με συνοπτικό τρόπο τα χαρακτηριστικά του κάθε ενός από τα επτά προφίλ που περιλαμβάνονται στην τελευταία τροποποίηση του προτύπου H.264 τον Απριλίο 2007.

	Baseline	Extended	Main	High	High 10	High 4:2:2	High 4:4:4
I and P Slices	√	√	√	√	√	√	√
B Slices	-	√	√	√	√	√	√
SI and SP Slices	-	√	-	-	-	-	-
Multiple Reference Frames	√	√	√	√	√	√	√
In-Loop Deblocking Filter	√	√	√	√	√	√	√
CAVLC Entropy Coding	√	√	√	√	√	√	√
CABAC Entropy Coding	-	-	√	√	√	√	√
Flexible Macroblock Ordering (FMO)	√	√	-	-	-	-	-
Arbitrary Slice Ordering (ASO)	√	√	-	-	-	-	-
Redundant Slices (RS)	√	√	-	-	-	-	-
Data Partitioning	-	√	-	-	-	-	-
Interlaced Coding (PicAFF, MBAFF)	-	√	√	√	√	√	√
4:2:0 Chroma Format	√	√	√	√	√	√	√
Mo-chrome Video Format (4:0:0)	-	-	-	√	√	√	√
4:2:2 Chroma Format	-	-	-	-	-	√	√
4:4:4 Chroma Format	-	-	-	-	-	-	√
8 Bit Sample Depth	√	√	√	√	√	√	√
9 and 10 Bit Sample Depth	-	-	-	-	√	√	√
11 to 14 Bit Sample Depth	-	-	-	-	-	-	√
8x8 vs. 4x4 Transform Adaptivity	-	-	-	√	√	√	√
Quantization Scaling Matrices	-	-	-	√	√	√	√
Separate Cb and Cr QP control	-	-	-	√	√	√	√
Separate Color Plane Coding	-	-	-	-	-	-	√
Predictive Lossless Coding	-	-	-	-	-	-	√
	Baseline	Extended	Main	High	High 10	High 4:2:2	High 4:4:4

Πίνακας 2.3 Συνοπτική παρουσίαση των προφίλ του H.264 (Απρίλιος 2007)

Συμπεράσματα του Κεφαλαίου

Στο παρόν κεφάλαιο παρουσιάστηκαν τα ποικίλα πρότυπα κωδικοποίησης ψηφιακού βίντεο, η εξέλιξη αυτών, καθώς και τα χαρακτηριστικά τους. Ακολουθώντας χρονικά την εξέλιξη των προτύπων έγινε σαφής η τεχνολογική εξέλιξη και βελτίωση των προτύπων κωδικοποίησης έως το νεότερο H.264, το οποίο χαρακτηρίζεται επί της παρούσης ως το βέλτιστο πρότυπο κωδικοποίησης, ενσωματώνοντας τα τελευταία επιτεύγματα στον ερευνητικό χώρο της κωδικοποίησης ψηφιακού βίντεο και επιτυγχάνοντας μέγιστο λόγο συμπίεσης.

ΚΕΦΑΛΑΙΟ 3

Θεμελιώδεις Αρχές Κωδικοποίησης Ψηφιακού Βίντεο

Χωρικός και Χρονικός Πλεονασμός

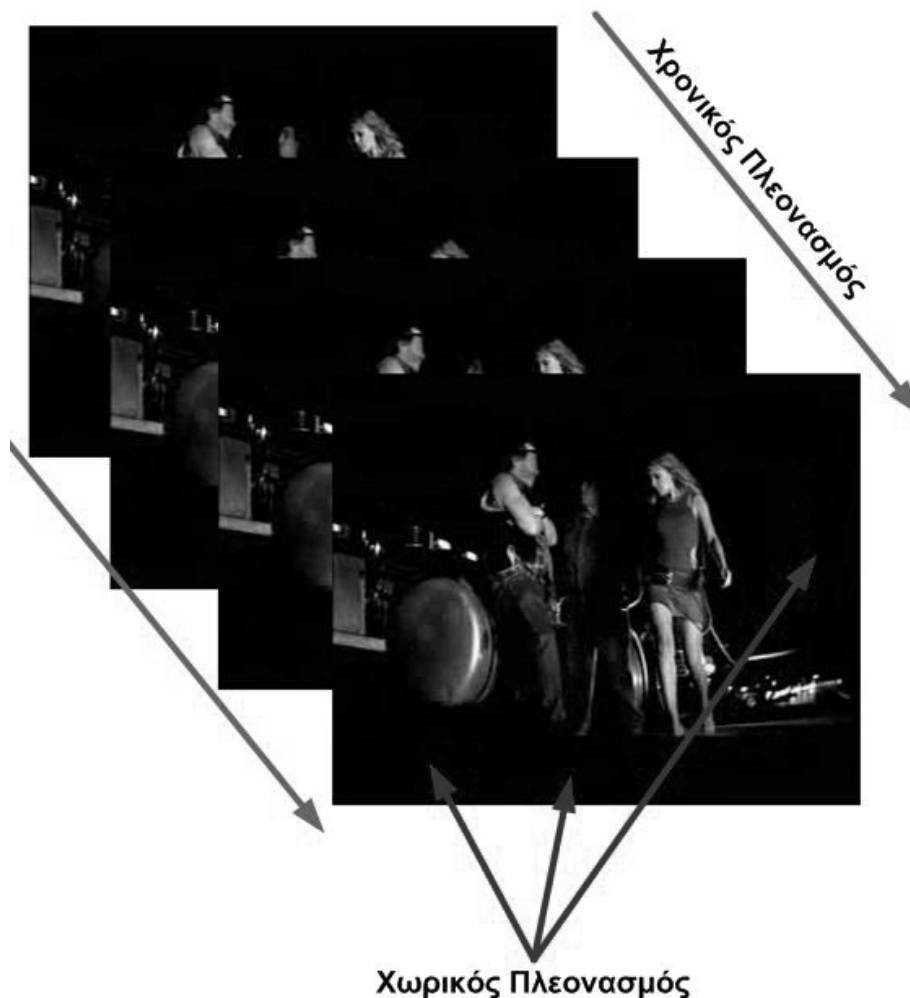
Η στατιστική ανάλυση ενός σήματος βίντεο υποδεικνύει ότι υπάρχει σημαντική οπτική συσχέτιση τόσο διαπλαισιακή (μεταξύ των διαδοχικών καρέ), όσο και ενδοπλαισιακή (της οπτικής πληροφορίας που απεικονίζεται σε ένα καρέ). Η συσχέτιση αυτή μπορεί να χαρακτηριστεί ως πλεονάζουσα πληροφορία και μπορεί επιτυχώς να αξιοποιηθεί για τη συμπίεση του όγκου των δεδομένων του σήματος. Επιπλέον, γνωρίζοντας τους περιορισμούς του ανθρώπινου οπτικού συστήματος, το οποίο παρουσιάζει υψηλή ευαισθησία μόνο στις χαμηλές χωρικές συχνότητες, μπορούμε να επιτύχουμε περαιτέρω συμπίεση των δεδομένων παραλείποντας συγκεκριμένες συχνότητες της πλεονάζουσας πληροφορίας, η αφαίρεση των οποίων θα προκαλέσει μικρής κλίμακας ποιοτικό υποβιβασμό.

Βασισμένοι σε αυτές τις παρατηρήσεις, οι μέθοδοι κωδικοποίησης βίντεο διαιρούνται σε δύο κατηγορίες: Στις μη απωλεστικές και στις απωλεστικές [21]. Σε ένα χωρίς απώλειες σύστημα συμπίεσης, ο στατιστικός πλεονασμός αφαιρείται με τέτοιο τρόπο ώστε το αρχικό σήμα να μπορεί να αναδημιουργηθεί πλήρως στο δέκτη. Μέχρι σήμερα, μη απωλεστικοί μέθοδοι επιτυγχάνουν ένα μέτριο λόγο συμπίεσης του σήματος βίντεο και για το λόγο αυτό δεν έχουν βρει ευρεία εφαρμογή. Αντιθέτων στις τεχνικές απωλεστικής συμπίεσης επιτυγχάνεται υψηλότερος λόγος συμπίεσης με κόστος ότι το αποκωδικοποιημένο σήμα έχει ποιοτικά υποβαθμιστεί σε σύγκριση με το αρχικό. Πρωτεύοντας στόχος κάθε απωλεστικής τεχνικής συμπίεσης βίντεο, είναι να εξισορροπηθεί ο υψηλός λόγος συμπίεσης με ελαχιστοποίηση της αντίστοιχης ποιοτικής υποβάθμισης.

Συνεπώς, η βασική αρχή πάνω στην οποία στηρίζονται όλες οι μέθοδοι ψηφιακής συμπίεσης βίντεο είναι το γεγονός ότι το σήμα εμπεριέχει ένα σημαντικό ποσοστό πλεονασμού (redundancy). Με τον όρο πλεονασμό εννοούμε την πληροφορία που μπορεί να συμπίεστεί ή εν μέρη να παραληφθεί, χωρίς να προκαλέσει σημαντική ποιοτική υποβάθμιση στο τελικό βίντεο. Υπάρχουν δύο είδη πλεονασμού [21], όπως απεικονίζονται και στο σχήμα 3.1:

1. Χωρικός πλεονασμός (Spatial Redundancy) : Η τιμή ενός pixel που εμπεριέχεται σε ένα καρέ του σήματος, μπορεί σε κάθε δεδομένη χρονική στιγμή να εκτιμηθεί με μεγάλη ακρίβεια από τιμές γειτονικών pixels, στηριζόμενοι στο γεγονός ότι οι γειτονικές τιμές δειγμάτων μιας εικόνας σχετίζονται μεταξύ τους σε μεγάλο ποσοστό.

2. Χρονικός Πλεονασμός (Temporal Redundancy) : Ένα σήμα βίντεο δομείται από διαδοχικά καρέ, τα οποία απεικονίζουν την εξελισσόμενη δράση ανά τακτά χρονικά διαστήματα εντός του ενός δευτερολέπτου. Συνεπώς τα διαδοχικά καρέ παρουσιάζουν μεγάλες ομοιότητες, δημιουργώντας έντονες συσχετίσεις μεταξύ τους, οι οποίες χαρακτηρίζονται ως χρονικός πλεονασμός.

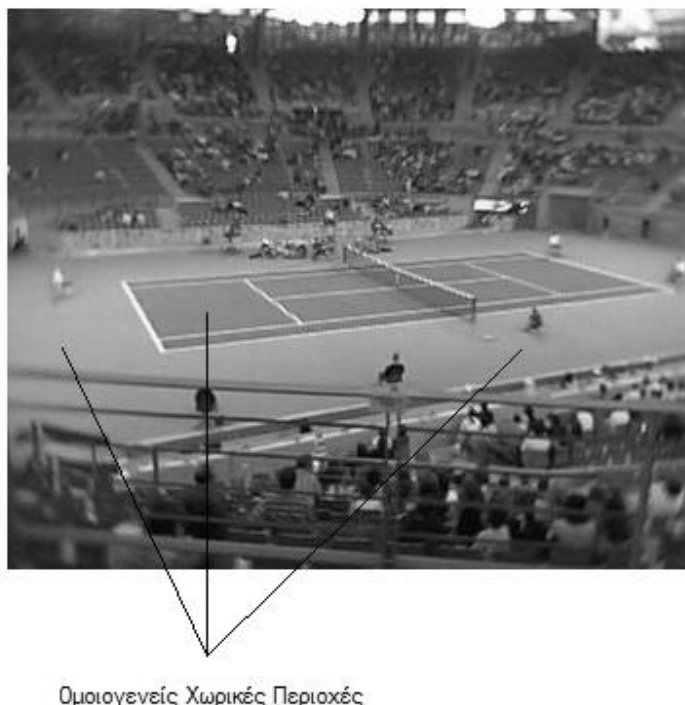


Σχήμα 3.1 Χωρικός και Χρονικός Πλεονασμός

Οπότε η πλειοψηφία των προτύπων κωδικοποίησης και συμπίεσης ψηφιακού βίντεο στηρίζονται στην αξιοποίηση του χωρικού και χρονικού πλεονασμού στοχεύοντας στη μείωση αυτού, γεγονός που θα επιφέρει και συμπίεση των δεδομένων. Πιο συγκεκριμένα, οι τρεις θεμελιώδεις αρχές μείωσης του πλεονασμού κατά τη διάρκεια κωδικοποίησης είναι:

- Μείωση Χωρικού Πλεονασμού: Για να επιτευχθεί μείωση του χωρικού πλεονασμού μεταξύ των pixels που δομούν ένα καρέ, αξιοποιείται η ομοιότητα των τιμών των γειτονικών pixels με τη χρήση κατάλληλων μετασχηματισμών.
- Μείωση Χρονικού Πλεονασμού: Για να επιτευχθεί μείωση του χρονικού πλεονασμού μεταξύ των διαδοχικών καρέ ενός βίντεο, χρησιμοποιούνται καρέ αναφοράς και αξιοποιείται η κωδικοποίηση μόνο των διαφορών μεταξύ των διαδοχικών καρέ. Οπότε είναι εφικτό να προβλέψουμε ένα σημαντικό τμήμα της εικόνας και να προσθέσουμε απλά τις επιμέρους διαφορές, χωρίς να χρειάζεται να κωδικοποιείται σε κάθε καρέ ολόκληρη η πλεονάζουσα πληροφορία.
- Εντροπική Κωδικοποίηση: Για να επιτευχθεί περαιτέρω συμπίεση των δεδομένων, πραγματοποιείται εντροπική συμπίεση με την εφαρμογή κωδικοποίησης μεταβλητού μήκους στα σύμβολα της ροής δεδομένων.

Συνεπώς, οι αλγόριθμοι συμπίεσης βασίζονται στην αφαίρεση του υπάρχοντος πλεονασμού (χρονικός, χωρικός ή/και πλεονασμός συχνότητας) σε ένα σήμα βίντεο. Το σχήμα 3.2 παρουσιάζει ένα παράδειγμα ενός πλαισίου (frame), στο οποίο έχουν τονιστεί οι περιοχές που αν και περιέχουν λεπτομερέστατη οπτική πληροφορία, το ανθρώπινο οπτικό σύστημα τις αντιλαμβάνεται εν γένει ως ομοιογενείς εξαιτίας της μικρής διακριτικής του ικανότητας. Συνεπώς, στις τονισμένες περιοχές, υπάρχει μικρή αντιληπτή διαφοροποίηση στο περιεχόμενο της εικόνας και ως εκ τούτου υπάρχει σημαντικός συχνотικός πλεονασμός. Εξαιτίας του γεγονότος ότι το ανθρώπινο οπτικό σύστημα είναι περισσότερο ευαίσθητο στις χαμηλές συχνότητες, η αφαίρεση των υψηλών συχνοτήτων δεν γίνεται αντιληπτή ως ποιοτική υποβάθμιση από τον ανθρώπινο εγκέφαλο, παρόλο που ένα μεγάλο μέρος του συχνотικού εύρους της έχει αφαιρεθεί. Έτσι επιτυγχάνεται συμπίεση των δεδομένων, διατηρώντας το ποιοτικό επίπεδο της εικόνας σε υψηλό επίπεδο.



Σχήμα 3.2 Καρέ που διακρίνονται οι χωρικές περιοχές με συχνотικό πλεονασμό

Σχετικά με την αξιοποίηση του χρονικού και χωρικού πλεονασμού, όλα τα πρότυπα κωδικοποίησης αξιοποιούν δυο διακριτές φάσεις στην τεχνική συμπίεσης: Την Ενδο-πλαισιακή και τη Δια-πλαισιακή κωδικοποίηση.

- Στην Ενδο-πλαισιακή Κωδικοποίηση (Intra-frame Coding) για την μείωση του χωρικού πλεονασμού, το κάθε καρέ αντιμετωπίζεται σαν αυτόνομη μονάδα και κωδικοποιείται ανεξάρτητα από τα υπόλοιπα καρέ αξιοποιώντας τον εσωτερικό του χωρικό πλεονασμό, οπότε το τελικό σήμα είναι μια σειρά από διακριτές ακίνητες εικόνες.
- Στη Δια-πλαισιακή Κωδικοποίηση (Inter-frame Coding) για τη μείωση του χρονικού πλεονασμού, λαμβάνονται υπόψιν οι ομοιότητες μεταξύ των διαδοχικών καρέ

(motion estimation) και κωδικοποιείται η διαφορά τους με χρήση της τεχνικής αντιστάθμισης της κίνησης (motion compensation). Έτσι, στο τελικό σήμα υπάρχει μία αλληλεξάρτηση μεταξύ των καρέ, αφού για να αποκωδικοποιηθεί και να ανασυγκροτηθεί κάποιο καρέ πιθανώς να πρέπει να ληφθούν πληροφορίες και από γειτονικά (προηγούμενα ή επόμενα). Με αυτόν τον τρόπο, επιτυγχάνεται υψηλός λόγος συμπίεσης, αφού κωδικοποιούνται μόνο οι διαφορές μεταξύ των διαδοχικών καρέ.

Στη συνέχεια παρουσιάζονται αναλυτικά οι τεχνικές ενδο-πλαισιακής και δια-πλαισιακής κωδικοποίησης προς αξιοποίηση του χρονικού και χωρικού πλεονασμού.

Αξιοποίηση Χρονικού Πλεονασμού κατά την Κωδικοποίηση

Η αξιοποίηση του χρονικού πλεονασμού κατά τη διαδικασία κωδικοποίησης στοχεύει στη μείωση της πλεονάζουσας πληροφορίας μεταξύ διαδοχικών καρέ με την ανάπτυξη δια-πλαισιακών (inter-frame coding) συσχετίσεων. Για το λόγο αυτό, σε μια ροή δεδομένων (bit-stream) κωδικοποιημένου βίντεο σύμφωνα με το πρότυπο MPEG τα καρέ διακρίνονται σε τρεις κατηγορίες:

- I (Intra frames) : Αυτό το είδος των καρέ κάνει χρήση του intra frame-coding (ενδο-πλαισιακής κωδικοποίησης). Η αποκωδικοποίηση των καρέ τύπου I δεν απαιτεί τη λήψη δεδομένων από κάποιο καρέ αναφοράς, αφού είναι τα μόνα που η κωδικοποίησή τους είναι ασυσχέτιστη με τα γειτονικά καρέ και κατά συνέπεια η αποκωδικοποίησή τους μπορεί να γίνει χωρίς αναφορά σε κάποιο άλλο. Πρόκειται για τα μεγαλύτερα σε μέγεθος καρέ σε μια ροή δεδομένων MPEG και αποτελούν σημεία αναφοράς για την κωδικοποίηση των μελλοντικών καρέ ώστε να αξιοποιηθεί και να μειωθεί ο χρονικός πλεονασμός. Εξαιτίας αυτής της συσχέτισης, τα I καρέ επαναλαμβάνονται περιοδικά στο ρεύμα ροής ενός βίντεο MPEG για να αποφευχθεί η διάδοση πιθανών σφαλμάτων που δημιουργεί η συσχετισμένη αποκωδικοποίηση των υπολοίπων καρέ από τα I καρέ. Η ομάδα των καρέ που δομείται από ένα I και από όλα τα διαδοχικά καρέ μέχρι το επόμενο I, ονομάζεται Group of Pictures (GOP). Το μήκος του GOP ορίζεται κατά τη κωδικοποίηση και ουσιαστικά αναφέρεται στην περιοδικότητα εμφάνισης των I καρέ, καθώς και στην αντιστάθμιση μιας αποδοτικής κωδικοποίησης και μιας λιγότερο ανθεκτικής σε λάθη μετάδοσης ροής MPEG. Συνεπώς τα I καρέ αποτελούν το θεμέλιο λίθο στη μείωση του χρονικού πλεονασμού κατά την κωδικοποίηση ενός ψηφιακού βίντεο.
- P (Predicted frames) : Τα καρέ τύπου P είναι πλαίσια, των οποίων η κωδικοποίηση/αποκωδικοποίηση είναι συσχετισμένη και εξαρτημένη από ένα προηγούμενο I ή P πλαίσιο (το οποίο με τη σειρά του εξαρτάται από ένα P ή I καρέ). Με τη βοήθεια της εκτίμησης της κίνησης, ο μηχανισμός κωδικοποίησης προβλέπει σε ένα P καρέ τις σχετικές θέσεις των macroblocks ή blocks σε σχέση με το P ή I καρέ αναφοράς. Συνεπώς δεν κωδικοποιείται όλο το καρέ, αλλά μόνο η πληροφορία της συσχέτισης με το καρέ αναφοράς που ονομάζεται διάνυσμα κίνησης. Επιπλέον κωδικοποιείται και η διαφορά του εκτιμώμενου καρέ που δομείται με τις συσχετίσεις σε σχέση με το πραγματικό καρέ, μέσω της τεχνικής αντιστάθμισης της κίνησης, ώστε κατά την αποκωδικοποίηση να έχουμε μια πλήρη αναδόμηση του αρχικού καρέ. Τα P καρέ μπορούν και αυτά με τη σειρά τους να αποτελέσουν καρέ αναφοράς για τα επόμενα πλαίσια, συμβάλλοντας στην πιθανή εισαγωγή και διάδοση σφαλμάτων εντός ενός GOP, εξαιτίας των δομημένων συσχετίσεων στις οποίες

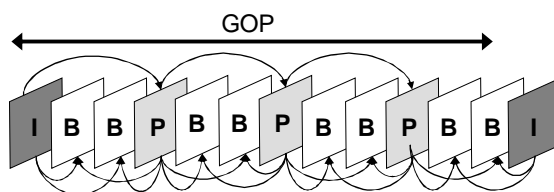
βασίζεται η αποκωδικοποίησή τους. Τα P καρέ είναι εν γένει μικρότερα σε μέγεθος από τα I καρέ, γιατί δεν περιέχουν ολόκληρη την οπτική πληροφορία του κωδικοποιημένου πλαισίου, αξιοποιώντας τον χρονικό πλεονασμό. Συνοψίζοντας, η διαδικασία της κωδικοποίησης ενός P καρέ είναι: Εντοπισμός των συσχετισμένων macroblocks/blocks στο καρέ αναφοράς (που μπορεί να είναι το προγενέστερο I ή P –ανάλογα με τη θέση του κωδικοποιημένου P εντός του GOP) και δημιουργία ενός γραμμικού συνδυασμού αυτών μέσω κατάλληλων διανυσμάτων κίνησης. Στη συνέχεια πραγματοποιείται η αντιστάθμιση της κίνησης με υπολογισμό και κωδικοποίηση των αντίστοιχων διαφορών μεταξύ του πραγματικού και του εκτιμώμενου καρέ. Συνεπώς, ένα P καρέ δομείται από τα διανύσματα κίνησης, τα Intra-coded blocks που δεν συσχετίστηκαν με κάποιο προγενέστερο καρέ P ή I και τη διαφορά του εκτιμώμενου πλαισίου από το πραγματικό καρέ.

- B (Bi-directional frames): Τα καρέ τύπου B είναι πλαίσια που δημιουργούνται συσχετίζοντας σε επίπεδο macroblock/block το περιεχόμενο ενός προηγούμενου και ενός επόμενου πλαισίου P καρέ. Τα B καρέ δε συντελούν στη διάδοση των σφαλμάτων, γιατί δεν χρησιμοποιούνται ως σημεία αναφοράς για περαιτέρω μείωση του χρονικού πλεονασμού και η εκτίμησή τους περιορίζεται μεταξύ δύο διαδοχικών P καρέ. Συνεπώς, ο 'κύκλος ζωής' των B καρέ περιορίζεται μόνο τοπικά και δεν επεκτείνεται μεταβιβάζοντας πληροφορίες σε άλλα πλαίσια. Η διαδικασία δόμησης τους είναι παρόμοια με αυτή των P καρέ, με βασική διαφορά ότι το εκτιμώμενο B πλαίσιο κατασκευάζεται λαμβάνοντας ως καρέ αναφοράς το προηγούμενο και το αμέσως επόμενο P καρέ. Πρέπει να διευκρινιστεί σε αυτό το σημείο, ότι ανάλογα με τη δομή του GOP και τη θέση στην οποία βρίσκεται ένα B καρέ εντός αυτού, τότε ενδέχεται –αν το B βρίσκεται στο τέλος του GOP- να χρησιμοποιηθούν για την εκτίμηση του B καρέ χωρίς βλάβη της γενικότητας, το προγενέστερο P και το επόμενο I καρέ (όπως ακριβώς απεικονίζεται στο σχήμα 3.3).

Τα B καρέ είναι μικρότερα σε μέγεθος από τα υπόλοιπα καρέ και αυτό εξαιτίας της δια-κατευθυντικής εκτίμησης, στην οποία στηρίζεται η δόμησή τους, που εκμεταλλεύεται στο έπακρον τη μείωση του χρονικού πλεονασμού που υπάρχει μεταξύ δύο διαδοχικών P καρέ.

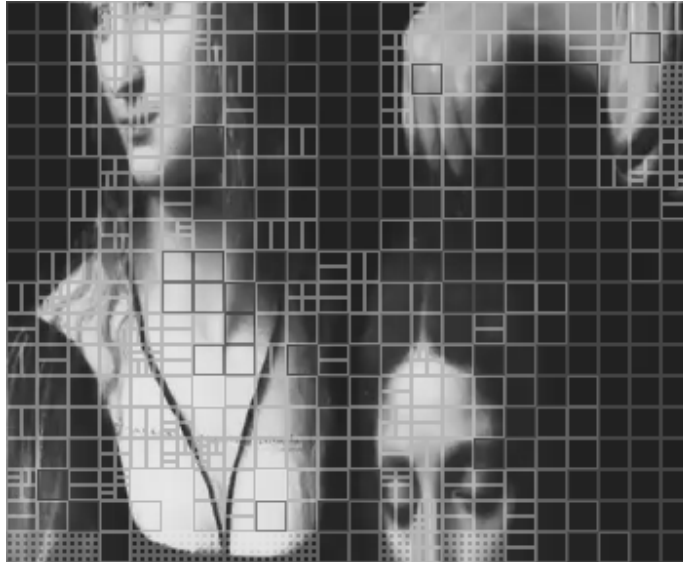
Ένα γεγονός που αξίζει να σημειωθεί είναι ότι η σειρά με την οποία κωδικοποιούνται τα καρέ δεν είναι ίδια με τη σειρά αναπαραγωγής. Τα B καρέ κατασκευάζονται συσχετίζοντας το περιεχόμενο δύο διαδοχικών P καρέ. Συνεπώς είναι προφανές ότι τα δύο P καρέ αναφοράς θα πρέπει να έχουν δημιουργηθεί πριν από το αντίστοιχο B, παρόλο που κατά την αναπαραγωγή το B παρεμβάλλεται ανάμεσά τους.

Το σχήμα 3.3 απεικονίζει μια τυπική περίπτωση της δομής ενός GOP, στο οποίο απεικονίζονται και οι συσχετίσεις που δημιουργούνται μεταξύ των I, B και P καρέ, ώστε να μειωθεί ο χρονικός πλεονασμός του περιεχομένου.



Σχήμα 3.3 Συσχετίσεις μεταξύ των καρέ που δομούν ένα GOP

Κατά συνέπεια, για να πραγματοποιηθεί η μείωση του χρονικού πλεονασμού μέσω των τριών κατηγοριών καρέ που προαναφέραμε, δύο διακριτές τεχνικές κωδικοποίησης πραγματοποιούνται: Η εκτίμηση κίνησης (Motion Estimation) και η αντιστάθμιση της κίνησης (Motion Compensation). Και οι δύο τεχνικές εφαρμόζονται κυρίως σε macroblocks ή blocks ανάλογα με το προφίλ και το πρότυπο κωδικοποίησης. Στα νεότερα πρότυπα, όπως απεικονίζεται στο σχήμα 3.4 για το πρότυπο H.264, blocks μεταβλητού μεγέθους μπορούν να χρησιμοποιηθούν ανάλογα με το περιεχόμενο για να επιτευχθεί βέλτιστη απόδοση των τεχνικών αυτών.



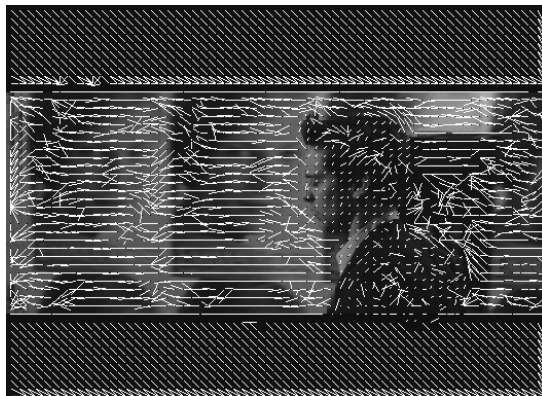
Σχήμα 3.4 Παράδειγμα χρήσης blocks μεταβλητού μεγέθους στο πρότυπο H.264

Κατά την εκτίμηση κίνησης (motion estimation), ο αλγόριθμος κωδικοποίησης αναζητεί στο/στα καρέ αναφοράς, που μπορεί να είναι προγενέστερα ή μεταγενέστερα ανάλογα με το είδος του καρέ που κωδικοποιείται (B ή P), τη βέλτιστη περιοχή macroblock/block που ταιριάζει με το περιεχόμενο που κωδικοποιείται και που θα χρησιμοποιηθεί ως αναφορά για τη δόμησή του. Η διαδικασία αυτή πραγματοποιείται μέσω σύγκρισης συγκεκριμένων τετράγωνων περιοχών (blocks ή macroblocks) μεταξύ του καρέ αναφοράς και του καρέ που κωδικοποιείται, μέχρι η βέλτιστη ταύτιση να εντοπιστεί. Στη διεθνή βιβλιογραφία έχει προταθεί πολύ μεγάλος αριθμός αλγορίθμων για τη διαδικασία αυτή, με στόχο την αύξηση της ταχύτητας εκτέλεσης και απόδοσης του αλγορίθμου και τη μείωση της πολυπλοκότητάς του [22], [23]. Επειδή στόχος της παρούσας διατριβής δεν είναι η διερεύνηση των αλγορίθμων εκτίμησης της κίνησης κατά τη διάρκεια κωδικοποίησης, δεν θα προβούμε σε λεπτομερή περιγραφή και βιβλιογραφική ανασκόπηση. Αξίζει όμως να σημειωθεί ότι στην συντριπτική τους πλειοψηφία οι αλγόριθμοι εκτίμησης της κίνησης χρησιμοποιούν ως κοινό μετρικό για την ποσοτικοποίηση των χωρικών διαφορών μεταξύ των διαδοχικών καρέ το Άθροισμα της Απόλυτης Διαφοράς (SAD – Sum of Absolute Difference) που ορίζεται ως εξής:

$$SAD(d_x, d_y) = \sum_{i=0}^{15} \sum_{j=0}^{15} |f(i, j) - g(i - d_x, j - d_y)|$$

Όπου $f(i, j)$ και $g(i, j)$ είναι οι τιμές της έντασης της φωτεινότητας του pixel ενός block ή macroblock της θέσης (i, j) στο καρέ που κωδικοποιείται και στο καρέ αναφοράς αντίστοιχα.

Η περιοχή αναφοράς καθορίζεται μέσω της χρήσης διανυσμάτων κίνησης (d_x, d_y), που προσδιορίζουν επακριβώς τη θέση της βέλτιστης περιοχής συσχέτισης στο καρέ αναφοράς. Ένα τυπικό παράδειγμα των διανυσμάτων κίνησης παρουσιάζεται στο σχήμα 3.5, όπου απεικονίζεται ένα καρέ από την ταινία Spiderman II, στο οποίο προβάλλονται τα διανύσματα κίνησης. Τα απεικονιζόμενα διανύσματα κίνησης δείχνουν τις συσχετίσεις που έχουν υπολογιστεί μεταξύ των blocks του απεικονιζόμενου καρέ και του/των καρέ αναφοράς.



Σχήμα 3.5 Απεικόνιση των διανυσμάτων κίνησης σε ένα καρέ της ταινίας Spiderman II

Όταν ο αλγόριθμος εκτίμησης της κίνησης εντοπίσει τη βέλτιστη περιοχή συσχέτισης, τότε έπεται η διαδικασία αντιστάθμισης της κίνησης. Κατά τη διάρκεια αυτού του σταδίου, το καρέ που βρίσκεται υπό κωδικοποίηση αναδομείται σύμφωνα με τις συσχετίσεις που έχουν προκύψει από το προγενέστερο στάδιο, αυτό της εκτίμησης της κίνησης, με τη χρήση των διανυσμάτων κίνησης και το καρέ αναφοράς. Με τον τρόπο αυτό δημιουργείται ένα εκτιμώμενο καρέ, το οποίο στη συνέχεια αφαιρείται από το αντίστοιχο πραγματικό καρέ με αποτέλεσμα τη δημιουργία ενός πλαισίου διαφορών, που εμπεριέχει μια απεικόνιση της εντροπικής τους διαφοράς. Το παραγόμενο πλαίσιο διαφορών είναι επί της ουσίας η απεικόνιση των διαφορών μεταξύ του εκτιμώμενου και του πραγματικού καρέ. Στο σχήμα 3.6, παρουσιάζεται ένα τυπικό παράδειγμα αυτής της διαδικασίας αφαίρεσης των δύο καρέ και υπολογισμού της διαφοράς τους, που απεικονίζεται στο κέντρο. Οι στατικές περιοχές (γκρι χρώμα) αντιστοιχούν σε μηδενική διαφορά, ενώ οι σκοτεινές περιοχές υποδηλώνουν αρνητικές διαφορές και οι φωτεινές θετικές διαφορές αντίστοιχα.



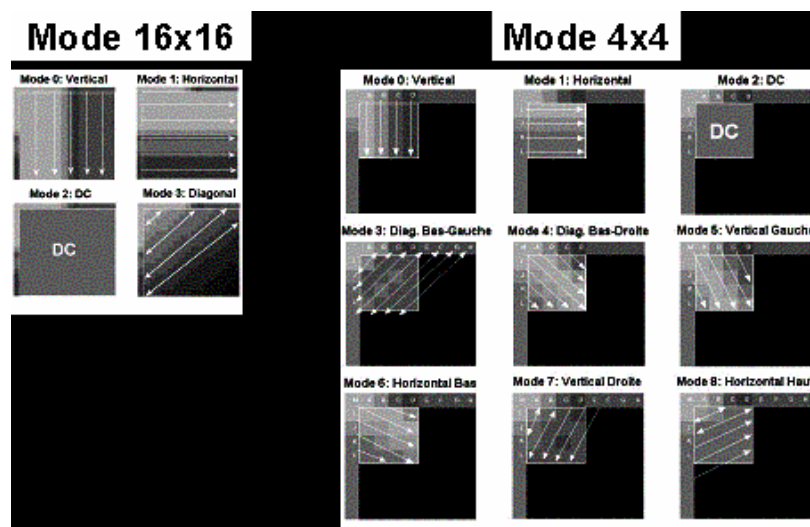
Σχήμα 3.6 Παράδειγμα υπολογισμού καρέ εντροπικών διαφορών

Συνεπώς, το στάδιο της αντιστάθμισης της κίνησης χρησιμοποιείται συμπληρωματικά του μηχανισμού εκτίμησης της κίνησης, διορθώνοντας και αντισταθμίζοντας στο εκτιμώμενο καρέ τις διαφορές που υπάρχουν συγκριτικά με το πραγματικό καρέ. Επομένως, η διαδικασία αντιστάθμισης της κίνησης υπολογίζει τις διαφορές αυτές, οι οποίες ενσωματώνονται στη ροή και υπόκεινται κανονικά σε όλα τα επόμενα στάδια της κωδικοποίησης.

Αξιοποίηση Χωρικού Πλεονασμού κατά την Κωδικοποίηση

Παρομοίως με τις τεχνικές μείωσης του χρονικού πλεονασμού σε ένα βίντεο, όπου πραγματοποιούνται συσχετίσεις μεταξύ του περιεχομένου των διαδοχικών καρέ, συσχετίσεις μπορούν να πραγματοποιηθούν και σε τμήματα της εικόνας που ανήκουν εντός του ίδιου καρέ. Πιο συγκεκριμένα, ανάλογα με την ομοιογένεια του περιεχομένου μπορεί να πραγματοποιηθεί πρόβλεψη των τιμών κάποιων blocks του καρέ, αξιοποιώντας δείγματα που έχουν ήδη κωδικοποιηθεί ως Intra blocks. Δηλαδή ο κωδικοποιητής προβλέπει τις τιμές ενός συνόλου pixels, βασισμένος στα προηγούμενα γειτονικά κωδικοποιημένα pixels που ανήκουν σε ήδη κωδικοποιημένα Intra blocks. Ανάλογα με το ποσοστό ομοιογένειας του περιεχομένου, η μείωση του χωρικού πλεονασμού μπορεί να είναι ιδιαίτερα υψηλή, που με τη σειρά της προκαλεί υψηλό λόγο συμπίεσης. Πρέπει σε αυτό το σημείο να σημειωθεί ότι η πρόβλεψη σε χωρικό επίπεδο πραγματοποιείται στο κωδικοποιημένο και όχι στο οπτικό χώρο. Η κωδικοποίηση –όπως θα δούμε αναλυτικά παρακάτω- πραγματοποιείται κυρίως με τη χρήση διακριτού μετασχηματισμού συνημιτόνου.

Στο τελευταίο πρότυπο H.264 οι διαθέσιμες τεχνικές για Intra Coding Prediction, όπως αποκαλείται αυτή η μέθοδος αξιοποίησης του ενδο-πλαισιακού χωρικού πλεονασμού κατατάσσεται σε δύο κατηγορίες ανάλογα με την επιθυμητή ακρίβεια: Η πρώτη κατηγορία περιλαμβάνει τις τεχνικές που προβλέπουν επιφάνειες εμβαδού 16x16 Pixels και στοχεύουν σε περιεχόμενο με υψηλό δείκτη χωρικής ομοιογένειας. Η δεύτερη κατηγορία περιλαμβάνει τις τεχνικές που προβλέπουν τη χωρική πληροφορία μιας επιφάνειας 4x4 pixels, στοχεύοντας σε περιεχόμενο με χαμηλό δείκτη χωρικής ομοιογένειας. Οι ποικίλες κατευθύνσεις αυτών των τεχνικών, οι οποίες επιλέγονται ανάλογα με τη δυναμική και την πολυπλοκότητα του περιεχομένου, παρουσιάζονται στο σχήμα 3.7.



Σχήμα 3.7 Intra prediction τεχνικές του προτύπου H.264

Πέραν από τις τιμές των κωδικοποιημένων blocks που σε περιπτώσεις ομοιογενούς περιεχομένου παρουσιάζουν μεγάλη συσχέτιση μεταξύ τους, τα διανύσματα κίνησης γειτονικών blocks ή macroblocks, που επιδεικνύουν τη συσχέτιση αναφοράς σε προγενέστερα κωδικοποιημένα καρέ, ενδέχεται επίσης να παρουσιάζουν αντίστοιχη υψηλή συσχέτιση. Επειδή η κίνηση του περιεχομένου εκτείνεται ομοιογενώς σε μεγάλη περιοχή του καρέ, τα διανύσματα κίνησης διατάσσονται κατά μία έννοια παράλληλα. Συνεπώς, η αξιοποίηση αυτού του πλεονασμού μπορεί να γίνει πραγματοποιώντας πρόβλεψη του κάθε διανύσματος κίνησης από προγενέστερα κωδικοποιημένα διανύσματα κίνησης. Το πιο απλό σχήμα πρόβλεψης για ένα διάνυσμα κίνησης ενός macroblock είναι η χρήση του διανύσματος που αντιστοιχεί στο οριζοντίως γειτονικό macroblock, εξασφαλίζοντας με τον τρόπο αυτό αξιοποίηση της πλεονάζουσας πληροφορίας χωρίς να απαιτείται η εφαρμογή κάποιου απαιτητικού αλγορίθμου πρόβλεψης. Εναλλακτικά, μπορούν να χρησιμοποιηθούν πιο πολύπλοκα σχήματα πρόβλεψης όπου λαμβάνουν υπόψιν τους όλες τις τιμές των διανυσμάτων κίνησης των γειτονικών macroblocks ή blocks. Η επιλογή του αλγορίθμου που θα εφαρμοστεί εξαρτάται από το περιεχόμενο του καρέ που κωδικοποιείται, ενώ το πλήθος των διαθέσιμων τεχνικών πρόβλεψης και η ακρίβεια αυτών (macroblock ή block) καθορίζεται από το πρότυπο κωδικοποίησης.

Διακριτός Μετασχηματισμός Συνημιτόνου

Στη συνέχεια τα I, B και P καρέ που έχουν προκύψει μέσω των τεχνικών εκτίμησης της κίνησης, καθώς και τα καρέ διαφορών που έχουν υπολογιστεί από την τεχνική αντιστάθμισης της κίνησης μετασχηματίζονται σε ένα άλλο χώρο, όπου οι συσχετισμένες οπτικές συχνότητες αναλύονται με μη συσχετισμένο τρόπο. Ο λόγος που η οπτική πληροφορία μετασχηματίζεται ώστε να επιτύχουμε αυτή την από-συσχέτιση γίνεται για να εκμεταλλευτούμε στη διαδικασία συμπίεσης το γεγονός ότι το ανθρώπινο οπτικό σύστημα παρουσιάζει διαφορετικά επίπεδα ευαισθησίας στο ευρύ φάσμα των οπτικών συχνοτήτων. Μέσω του μετασχηματισμού οι διάφορες οπτικές συχνότητες μπορούν να προσπελαστούν με διακριτό τρόπο και κατά συνέπεια να αφαιρεθούν εκείνες, στις οποίες οι άνθρωποι οφθαλμοί δεν είναι ευαίσθητοι και κατά συνέπεια η απουσία τους δεν θα γίνει εντόνως αντιληπτή [2].

Ο πιο διαδεδομένος μετασχηματισμός για την κωδικοποίηση ψηφιακού βίντεο είναι ο διακριτός συνημιτονικός μετασχηματισμός (Discrete Cosine Transformation – DCT), ο οποίος πλεονεκτεί έναντι του διακριτού μετασχηματισμού Fourier [24]. Ο DCT εφαρμόζεται εν γένει σε ένα καρέ διαστάσεων $N \times N$ δημιουργώντας το μετασχηματισμένο καρέ Y , το οποίο διατηρεί βεβαίως τις ίδιες διαστάσεις με το αρχικό καρέ περιέχοντας τους αντίστοιχους συντελεστές που αντιστοιχούν στη βάση συχνοτήτων του DCT. Δηλαδή πραγματοποιείται μια ανάλυση της οπτικής πληροφορίας στις οπτικές συχνότητες από τις οποίες δομείται. Η δράση του DCT θα μπορούσε να περιγραφεί χρησιμοποιώντας έναν τετραγωνικό πίνακα μετασχηματισμού A . Κατά συνέπεια ο μετασχηματισμός γράφεται:

$$Y = AXA^T$$

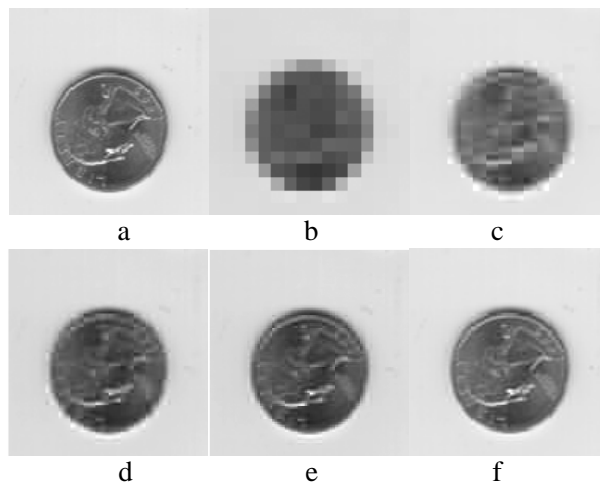
Όπου το X είναι ο πίνακας που περιέχει την πληροφορία του αρχικού καρέ και Y ο αντίστοιχος μετασχηματισμένος πίνακας. Τα στοιχεία του πίνακα μετασχηματισμού A δίνονται από τον εξής τύπο:

$$A_{ij} = C_i \cos \frac{(2j+1)ip}{2N} \quad \text{όπου } C_i = \begin{cases} \sqrt{1/N}, & i=0 \\ \sqrt{2/N}, & i>0 \end{cases}$$

Κατά συνέπεια, συνδυάζοντας τις δύο εκφράσεις ο DCT μπορεί να περιγραφεί από την ακόλουθη σχέση:

$$Y_{xy} = C_x C_y \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} X_{ij} \cos \frac{(2j+1)yp}{2N} \cos \frac{(2i+1)xp}{2N}$$

Το πλεονέκτημα του DCT είναι ότι μπορεί να αναχθεί η αρχική εικόνα με την εφαρμογή του αντίστροφου μετασχηματισμού. Επιπλέον, εάν κατά την εφαρμογή του αντίστροφου αλγορίθμου δεν χρησιμοποιηθούν όλες οι συχνότητες, αλλά μόνο τμήμα των χαμηλών συχνοτήτων, στις οποίες παρουσιάζει και τη μεγαλύτερη ευαισθησία το ανθρώπινο οπτικό σύστημα, τότε η ποιότητα της μετασχηματισμένης εικόνας μπορεί να είναι αρκετά ικανοποιητική. Με τον τρόπο αυτό επιτυγχάνεται συμπίεση των δεδομένων, χωρίς να πραγματοποιείται σημαντική ποιοτική υποβάθμιση, γεγονός ύψιστης σημασίας για τα πρότυπα κωδικοποίησης βίντεο. Σε αυτά τα πλαίσια το σχήμα 3.8 παρουσιάζει ένα παράδειγμα της απόδοσης του αντιστρόφου DCT με τη χρήση ενός μέρους από το σύνολο των 64 συντελεστών συχνοτήτων ανά block. Όπως παρατηρείται, είναι εφικτό να αναχθεί η αρχική εικόνα με πολύ ικανοποιητική ποιότητα με τη χρήση λιγότερων εκ των μισών συντελεστών ποιότητας. Πιο συγκεκριμένα, η εικόνα a απεικονίζει την αρχική μας πηγαία φωτογραφία. Στη συνέχεια, στη b απεικονίζεται το αποτέλεσμα του αντιστρόφου μετασχηματισμού DCT με τη χρήση ενός μόνο συντελεστή συχνοτήτων σε κάθε block. Στην c χρησιμοποιούνται 4 συντελεστές, ενώ στην d αξιοποιούνται 8 συντελεστές. Στην e όπου χρησιμοποιούνται 18 συντελεστές μπορούμε να παρατηρήσουμε ότι έχουμε ανακτήσει σε ικανοποιητικό βαθμό την αρχική μας εικόνα. Τέλος στην f χρησιμοποιούνται ακριβώς οι μισοί συντελεστές συχνοτήτων (δηλαδή 32) και το αποτέλεσμα είναι άμεσα συγκρίσιμο με την αρχική μας εικόνα, χωρίς να είναι αισθητή κάποια ποιοτική υποβάθμιση.



Σχήμα 3.8 Παράδειγμα της αντιληπτής απόδοσης του αντιστρόφου DCT

Συνεπώς, με κόστος την υποβάθμιση της ποιότητας, η αρχική εικόνα μπορεί να αναδομηθεί χρησιμοποιώντας ένα περιορισμένο αριθμό συντελεστών του διακριτού μετασχηματισμού συνημιτόνου.

Για λόγους καλύτερης κατανόησης παραθέτουμε ένα παράδειγμα της χρήσης του μετασχηματισμού, όπως αυτός εφαρμόζεται σε ένα καρέ Ι. Στο σχήμα 3.9 απεικονίζονται τυπικές τιμές της συνιστώσας φωτεινότητας (Υ) ενός block που προέρχεται από το υπό εξέταση Ι καρέ. Πρέπει να σημειωθεί ότι τα νούμερα αυτά έχουν επιλεγεί εντελώς τυχαία και δεν αντιστοιχούν σε κάποια συγκεκριμένη εικόνα.

Εφαρμόζοντας στο block του σχήματος το διακριτό μετασχηματισμό συνημιτόνου (DCT), το συγκεκριμένο block μετασχηματίζεται όπως ακριβώς απεικονίζεται στο σχήμα 3.10.

131	134	139	143	144	144	144	144
134	141	143	146	149	146	146	146
143	145	150	153	148	146	146	146
148	151	152	151	150	148	148	148
150	150	151	152	152	145	145	145
152	151	151	151	150	147	146	145
151	152	153	153	151	148	147	147
151	152	153	153	151	149	149	148

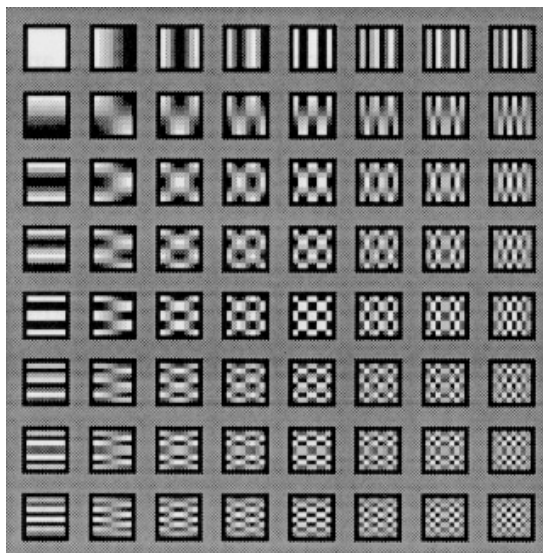
Σχήμα 3.9 Τυπικό παράδειγμα των τιμών φωτεινότητας ενός block

294,94	0,32	-3,08	-1,37	0,44	-0,02	-0,42	0,19
-5,81	-4,07	-1,19	-0,41	-0,15	-0,15	-0,08	-0,11
-2,62	-2,40	-0,34	0,09	-0,29	0,01	0,06	-0,17
-2,06	-0,19	0,10	0,65	0,31	-0,17	0,00	0,21
-0,25	-0,20	0,28	0,11	0,00	-0,01	0,02	0,16
-0,26	0,13	0,32	-0,20	-0,21	0,31	0,33	-0,22
-0,27	0,35	0,12	-0,00	-0,12	0,66	0,28	-0,19
-0,31	0,36	-0,50	-0,32	-0,52	0,37	0,07	-0,01

Σχήμα 3.10 Το block του σχ.3.9 μετά την εφαρμογή του DCT

Οι τιμές που περιέχονται στο σχήμα 3.10 μετά την εφαρμογή του DCT είναι οι συντελεστές του διακριτού μετασχηματισμού, οι οποίοι με τη σειρά τους αντιστοιχούν στις αντίστοιχες χωρικές συχνότητες βάσης που απεικονίζονται στο σχήμα 3.11. Η τιμή 294,94 που βρίσκεται στη θέση (0,0) ονομάζεται συντελεστής DC, ενώ οι υπόλοιποι ονομάζονται συντελεστές AC. Ως συντελεστής DC ορίζεται αυτός που η χωρική του συχνότητα θεωρείται μηδενική και αποτελεί σημείο αναφοράς στις δύο κατευθύνσεις, ενώ AC είναι οι συντελεστές που η συχνότητά τους είναι διάφορη του μηδενός σε τουλάχιστον μία διεύθυνση, όπως μπορεί πολύ εύκολα να παρατηρηθεί και στις αντίστοιχες συχνότητες βάσης στο σχήμα 3.11. Καθώς κινούμαστε προς τα δεξιά και απομακρυνόμαστε από τη DC συνιστώσα παρατηρούμε ότι αυξάνει η οριζόντια χωρική συχνότητα των βάσεων DCT, ενώ καθώς κινούμαστε προς τα κάτω αυξάνεται αντιστοίχως η κάθετη χωρική συχνότητα.

Αξίζει να σημειωθεί ότι το μεγαλύτερο μέρος της μετασχηματισμένης οπτικής πληροφορίας περιέχεται στο συντελεστή DC, η οποία φέρει τιμή που είναι μεγαλύτερη από το διπλάσιο του μέσου όρου των υπολοίπων AC συντελεστών. Ο συντελεστής DC δε μπορεί σε καμία περίπτωση να παραλειφθεί, σε αντίθεση με τους συντελεστές AC της κάτω δεξιά περιοχής που αντιστοιχούν σε υψηλές συχνότητες και όπως επιδείχθηκε στο σχήμα 3.8, η παράλειψή τους δεν προκαλεί αξιοσημείωτη ποιοτική υποβάθμιση.



Σχήμα 3.11 Οι συχνότητες βάσης του 8x8 DCT

Κβάντιση και Σάρωση των Συντελεστών DCT

Στο μετασχηματισμένο block του παραδείγματος, που περιέχει τους συντελεστές DCT εφαρμόζεται η διαδικασία κβάντισης, η οποία χρησιμοποιείται για να περιοριστεί το εύρος των συντελεστών DCT σε ένα μικρό σύνολο διακριτών ακεραίων τιμών. Η πιο δημοφιλής μέθοδος κβάντισης είναι η κλιμακωτή, η οποία αντιστοιχίζει ένα συνεχές δείγμα στην είσοδο σε ένα κβαντισμένο στην έξοδο. Ένας κβαντιστής $Q()$ μπορεί να περιγραφεί με μαθηματικό τρόπο ως εξής:

$$Q(x) = g(\lfloor f(x) \rfloor)$$

Όπου

- x είναι ένας πραγματικός αριθμός
- $\lfloor x \rfloor$ είναι η συνάρτηση στρογγυλοποίησης στο μικρότερο ακέραιο
- $f(x)$ και $g(i)$ είναι πραγματικές συναρτήσεις

Η ακέραια τιμή $i = \lfloor f(x) \rfloor$ είναι η αναπαράσταση που συνήθως αποθηκεύεται ή μεταδίδεται, αλλά η τελική κβαντισμένη τιμή μπορεί επιπλέον να τροποποιηθεί χρησιμοποιώντας τη συνάρτηση $g(i)$. Οπότε, κατά την εφαρμογή ενός κβαντιστή κλίμακας, πραγματοποιείται η στρογγυλοποίηση του πραγματικού αριθμού προς τον κοντινότερο ακέραιο, σύμφωνα με τον ακόλουθο τρόπο:

$$Y = QP \text{ round} \left(\frac{X}{QP} \right)$$

Όπου QP είναι το βήμα κβάντισης, X ο αρχικός πραγματικός αριθμός και Y η αντίστοιχη τελική κβαντισμένη τιμή. Ο πίνακας 3.1 περιέχει μερικά αντιπροσωπευτικά παραδείγματα για ποικίλα βήματα κβάντισης.

X	Y		
	$QP=1$	$QP=2$	$QP=3$
0	0	0	0
1	1	0	0
2	2	2	0
3	3	2	3
4	4	4	3
5	5	4	3
6	6	6	6
7	7	6	6
8	8	8	6
9	9	8	9

Πίνακας 3.1 Παραδείγματα κβάντισης

Εφαρμόζοντας την προαναφερθείσα διαδικασία κβάντισης στους συντελεστές DCT, επιτυγχάνουμε περιορισμό του εύρους των τιμών τους με αποτέλεσμα την αποδοτικότερη συμπίεση αυτών. Όμως, πρέπει να τονιστεί ότι η κβάντιση είναι μια μη αντιστρέψιμη διαδικασία, οπότε η πληροφορία που χάνεται από τη διαδικασία αυτή είναι ο κύριος λόγος που οδηγεί στην ποιοτική υποβάθμιση των ψηφιακά κωδικοποιημένων βίντεο, αφού στη συνέχεια με την εφαρμογή του αντίστροφου μετασχηματισμού DCT δεν αποκαθίστανται επακριβώς οι αρχικές τιμές των pixels. Όσο μεγαλύτερο είναι το βήμα κβάντισης, τόσο αποδοτικότερη θα είναι η συμπίεση των δεδομένων, αλλά και πιο έντονη η ποιοτική υποβάθμιση του περιεχομένου. Μικρό βήμα κβάντισης θα οδηγήσει σε χαμηλού λόγου συμπίεση και απαραίτητη ποιοτική υποβάθμιση.

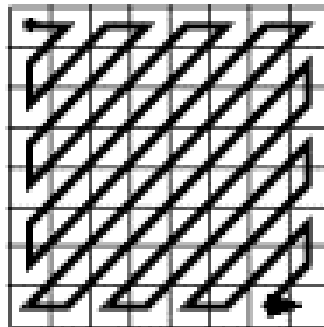
Επιστρέφοντας στο block του παραδείγματος, παρατηρούμε ότι αν κβαντίσουμε τις αντίστοιχες τιμές, τότε οι συντελεστές AC υψηλών χωρικών συχνοτήτων μηδενίζονται, γεγονός που πρακτικά σημαίνει ότι αφαιρούνται από το καρέ οι υψηλές χωρικές συχνότητες τόσο στην οριζόντια όσο και στην κατακόρυφη συνιστώσα. Επιπλέον, ο μηδενισμός των συντελεστών αυτών επιτρέπει, όπως θα δούμε στη συνέχεια, την αποτελεσματική συμπίεσή τους με τη χρήση εντροπικής κωδικοποίησης, όπως Run Length Coding. Το εν λόγω block μετά την κβάντιση έχει τη μορφή που φαίνεται στο σχήμα 3.12.

295	0	-3	-1	0	0	0	0
-6	-4	-1	0	0	0	0	0
-3	-2	0	0	0	0	0	0
-2	0	3	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Σχήμα 3.12 Οι τιμές του block μετά τη χρήση κβάντισης

Πρέπει να σημειωθεί ότι σε κάποιες υλοποιήσεις κωδικοποιητών πριν το στάδιο της κωδικοποίησης πραγματοποιείται η εφαρμογή ενός κατώφλιου, η οποία εκμηδενίζει τους συντελεστές DCT που είναι μικρότεροι ή ίσοι με ένα προκαθορισμένο κατώφλι. Η διαδικασία αυτή συντελεί στην επιπλέον εκμηδένιση κάποιων τιμών που δεν μηδενίζονται από τη διαδικασία της κβάντισης. Οπότε πραγματοποιείται μια περαιτέρω αφαίρεση των υψηλών οριζοντίων και καθέτων χωρικών συχνοτήτων από το βίντεο που κωδικοποιείται.

Εξαιτίας των κβαντισμένων παραμέτρων DCT σε ένα block, οι διαφορές και οι ασυνέχειες της φωτεινότητας μεταξύ των γειτονικών blocks αυξάνονται από τη διαφοροποίηση των συντελεστών. Αυτό συμβαίνει επειδή η διαδικασία κβάντισης είναι απωλεστική, δηλαδή μη αναστρέψιμη, αφού δεν δύναται να προβλεφθεί η αρχική τιμή του κάθε συντελεστή, με αποτέλεσμα την εμφάνιση οπτικών τεχνουργημάτων με κυρίαρχο το φαινόμενο τετραγώνισης (blockiness). Οπότε έχουμε μια ισοδύναμη κατάσταση με αυτή που επιδείχθηκε στο σχήμα 3.8, όπου στην εφαρμογή του αντιστρόφου μετασχηματισμού δεν χρησιμοποιήθηκαν όλοι οι συντελεστές DCT, γεγονός ταυτόσημο με τον εκμηδενισμό ή την παράλειψη κάποιων εξ' αυτών.



Σχήμα 3.13 Σάρωση των κβαντισμένων συντελεστών DCT

Στη συνέχεια ακολουθεί η σάρωση των κβαντισμένων συντελεστών με προκαθορισμένο τρόπο όπως ακριβώς φαίνεται στο σχήμα 3.13. Η διαδικασία αυτή είναι προπαρασκευαστική του επόμενου σταδίου, αυτού της περαιτέρω συμπίεσης των δεδομένων από αλγόριθμους σειριακής εντροπικής κωδικοποίησης. Συνεπώς είναι απαραίτητο τα δεδομένα να επεξεργαστούν σειριακά και για το λόγο αυτό εκτελείται το εν λόγω στάδιο της σάρωσης των δεδομένων και μετατροπής της προβολής τους από

δισδιάστατο πίνακα σε πίνακα $1 \times N$ (όπου N το πλήθος των συντελεστών του δισδιάστατου πίνακα). Ο λόγος που επιλέχθηκε ο συγκεκριμένος τρόπος σάρωσης (κυματοειδής (zig-zag) αρχίζοντας από την τιμή DC) είναι προφανής: Η παραγόμενη σειριακή διάταξη των δεδομένων ακολουθεί μια φθίνουσα πορεία σημαντικότητας στην εφαρμογή του αντιστρόφου διακριτού μετασχηματισμού συνημιτόνου, εναλλασσόμενων οριζοντίων και καθέτων συχνοτήτων, παραθέτοντας αρχικά τις χαμηλές χωρικές συχνότητες και τελευταίες τις υψηλές, οι οποίες έχουν εν γένει εκμηδενιστεί από το στάδιο της κβάντισης.

Εντροπική Κωδικοποίηση των Δεδομένων

Στη συνέχεια η σειριακή ροή των κβαντισμένων συντελεστών DCT συμπιέζονται μέσω της εφαρμογής εντροπικής μεθόδου, η οποία αξιοποιεί τον πλεονασμό της διαδοχικής επανάληψης των ίδιων αριθμητικών τιμών των συντελεστών. Σε αυτά τα πλαίσια οι δύο μέθοδοι που αξιοποιούνται κυρίως στα πρότυπα κωδικοποίησης βίντεο είναι η κωδικοποίηση Huffman [25] και η αριθμητική κωδικοποίηση [26].

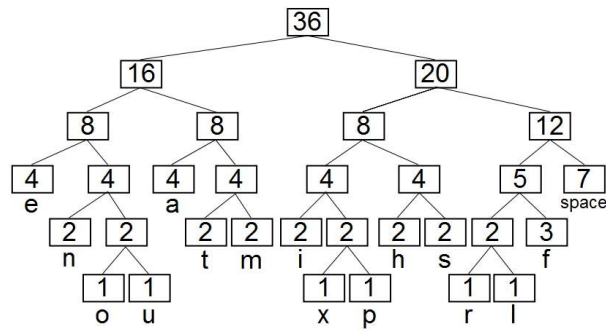
Αυτοί οι μέθοδοι συμπίεσης αντιστοιχίζουν κάθε σύμβολο της ροής βίντεο με ένα κωδικό μεταβλητού μήκους βασισμένοι στην πιθανότητα εμφάνισης του κάθε συμβόλου. Τα σύμβολα που εμφανίζονται τακτικότερα αντιστοιχίζονται με κωδικούς μικρού μήκους, ενώ σε σύμβολα που εμφανίζονται αραιότερα ορίζονται κωδικοί μεγάλου μήκους. Λαμβάνοντας υπόψιν ότι η ροή ενός σήματος βίντεο αποτελείται από μεγάλο πλήθος ομοίων συμβόλων, η αντικατάσταση αυτών με κωδικούς μεταβλητού μήκους σύμφωνα με την παραπάνω τεχνική οδηγεί σε αποδοτικότερη συμπίεση των δεδομένων [27].

Για λόγους κατανόησης και πληρότητας θα αναφερθούμε εν συντομία και στις δύο προαναφερθείσες τεχνικές, παρουσιάζοντας αμφότερα από ένα παράδειγμα.

Huffman Coding

Στην επιστήμη των υπολογιστών και στη θεωρία πληροφορίας, η κωδικοποίηση Huffman είναι ευρέως γνωστή ως ένας αλγόριθμος εντροπικής κωδικοποίησης που χρησιμοποιείται για μη απωλεστική συμπίεση δεδομένων. Ο όρος αναφέρεται στη χρήση ενός πίνακα με κωδικούς μεταβλητού μήκους, οι οποίοι διαμορφώνονται και αντιστοιχίζονται με τα σύμβολα των δεδομένων ανάλογα με τη συχνότητα εμφάνισης αυτών. Ο αλγόριθμος αυτός αναπτύχθηκε από τον D. Huffman κατά τη διάρκεια εκπόνησης της διδακτορικής του διατριβής στο MIT και δημοσιεύτηκε το 1952 [25].

Για να κατανοήσουμε τον τρόπο συμπίεσης του αλγορίθμου Huffman, ας θεωρήσουμε τη φράση "This is an example of Huffman coding". Η φράση αυτή δομείται από σύμβολα, τα οποία είναι τα γράμματα που σχηματίζουν τις λέξεις και το κενό μεταξύ των λέξεων. Συνεπώς, η συχνότητα εμφάνισης των συμβόλων αυτών μπορεί να αναπαρασταθεί με ένα δέντρο συχνοτήτων, όπως ακριβώς απεικονίζεται στο σχήμα 3.12.



Σχήμα 3.14 Παράδειγμα δέντρου συχνοτήτων εμφάνισης των συμβόλων

Σύμβολο	Συχνότητα	Κωδικός
κενό	7	111
a	4	010
e	4	000
f	3	1101
h	2	1010
i	2	1000
m	2	0111
n	2	0010
s	2	1011
t	2	0110
l	1	11001
o	1	00110
p	1	10011
r	1	11000
u	1	00111
x	1	10010

Πίνακας 3.2 Παράδειγμα Huffman Coding

Όπως παρατηρούμε, το σύμβολο με τη μεγαλύτερη συχνότητα εμφάνισης στη φράση “This is an example of Huffman coding” είναι το σύμβολο του κενού, το οποίο εμφανίζεται 7 φορές. Όπως παρατηρούμε στον πίνακα 3.2, αντιστρόφως ανάλογα με τη συχνότητα εμφάνισης των συμβόλων ορίζεται το μήκος του κωδικού που θα αντικαταστήσει το συγκεκριμένο σύμβολο. Συνεπώς παρατηρούμε ότι τα σύμβολα με μεγάλη συχνότητα εμφάνισης αντικαθίστανται από μήκος λέξης 3 bits, ενώ τα σύμβολα μικρής συχνότητας αντικαθίστανται από λέξεις μήκους 5 bits.

Η μέθοδος κωδικοποίησης Huffman χρησιμοποιεί συγκεκριμένο τρόπο για την επιλογή της αναπαράστασης και του κωδικού του κάθε συμβόλου. Ο κωδικός αυτός είναι μια αλληλουχία από bits που αντικαθιστούν ένα μόνο σύμβολο των δεδομένων και καθορίζεται από τη θέση του αντίστοιχου συμβόλου στο δέντρο Huffman. Η αντιστοίχιση μεταξύ των κωδικών και των συμβόλων είναι 1-1, που σημαίνει ότι ο ίδιος κωδικός δεν μπορεί να αντιστοιχεί σε περισσότερα του ενός σύμβολα, όπως και ότι ένα σύμβολο δεν μπορεί να αντικατασταθεί από περισσότερους του ενός κωδικούς. Η βασική έννοια στην οποία βασίζεται η μέθοδος κωδικοποίησης Huffman είναι ότι αντικαθιστά σύμβολα ή χαρακτήρες με λέξεις μικρότερου μήκους σε σχέση με αυτές που απαιτούνται για την αναπαράσταση αυτών.

Arithmetic Coding

Η αριθμητική κωδικοποίηση είναι μια μέθοδος μη απωλεστικής συμπίεσης δεδομένων που κατατάσσεται στην κατηγορία των τεχνικών εντροπικής κωδικοποίησης μεταβλητού μήκους, όπου σε αντίθεση με την κωδικοποίηση Huffman, δεν αντικαθιστά τα δεδομένα με κωδικούς μεταβλητού μήκους αλλά κωδικοποιεί το πλήρες μήνυμα με ένα μόνο αριθμό n όπου παίρνει τιμές στην περιοχή $0.0 \leq n < 1.0$.

Για να γίνει κατανοητός ο αλγόριθμος, στον οποίο βασίζεται η αριθμητική κωδικοποίηση, παραθέτουμε το ακόλουθο παράδειγμα. Η βασική ιδέα είναι ότι από το φάσμα των πιθανοτήτων $[0,1]$ αντιστοιχίζουμε σε κάθε σύμβολο ένα συγκεκριμένο εύρος ανάλογα με τη συχνότητα εμφάνισής του. Όσο μεγαλύτερη είναι η πιθανότητα εμφάνισης του συμβόλου, τόσο μεγαλύτερο εύρος διαθέτουμε στο συγκεκριμένο σύμβολο. Οπότε, πρέπει να καθορίσουμε το εύρος του κάθε συμβόλου ανάλογα με τη συχνότητα εμφάνισης αυτού. Ας θεωρήσουμε ότι θέλουμε να κωδικοποιήσουμε τη λέξη ΚΑΛΑ σύμφωνα με τη μέθοδο αριθμητικής κωδικοποίησης. Στον πίνακα 3.3 εξετάζουμε την απλή περίπτωση όπου το πλήθος των διαθέσιμων συμβόλων περιορίζεται στα εξής τρία: Κ, Λ και Α.

Θεωρώντας ότι η συχνότητα εμφάνισης του συμβόλου Α είναι διπλάσια από αυτή των υπολοίπων συμβόλων, παραθέτουμε στο σύμβολο αυτό διπλάσιο εύρος σε σύγκριση με τα υπόλοιπα δύο.

Σύμβολο	Συχνότητα	Εύρος
A	2	[0.0 , 0.5)
K	1	[0.5 , 0.75)
Λ	1	[0.75 , 1.0)

Πίνακας 3.3 Παράδειγμα Arithmetic Coding I

Τότε ξεκινάμε την εφαρμογή του εξής αλγορίθμου

- Χαμηλή Τιμή = 0
- Υψηλή Τιμή = 1
- Επανάληψη για όλα τα σύμβολα
 - Εύρος= Υψηλή – Χαμηλή τιμή
 - Υψηλή τιμή = Χαμηλή τιμή + εύρος * Υψηλή τιμή του συμβόλου που κωδικοποιείται
 - Χαμηλή τιμή = Χαμηλή τιμή + εύρος * Χαμηλή τιμή του συμβόλου που κωδικοποιείται

Οπότε με τη βοήθεια του πίνακα 3.3, για τη λέξη ΚΑΛΑ προκύπτει:

Σύμβολο	Εύρος	Χαμηλή τιμή	Υψηλή τιμή
		0	1
K	1	0.5	0.75
A	0.25	0.5	0.625
Λ	0.125	0.59375	0.625
A	0.03125	0.59375	0.609375

Πίνακας 3.4 Παράδειγμα Arithmetic Coding II

Συνεπώς το αποτέλεσμα της κωδικοποίησης της λέξης ΚΑΛΑ είναι ο αριθμός 0.59375.

Για την αποκωδικοποίηση αυτής θα χρησιμοποιηθεί ακριβώς η αντίστροφη διαδικασία, η οποία παρατίθεται παρακάτω:

- Εντοπισμός του πρώτου συμβόλου από την τιμή 0.59375 ανάλογα με το διάστημα που ανήκει στον πίνακα 3.3
- Επανάληψη για όλα τα σύμβολα:
 - Εύρος = Υψηλή τιμή – Χαμηλή τιμή
 - $ΕπομενοV\ ΑριθμοV = \frac{Τρεχων\ Αριθμος - Χαμηλη\ τιμη\ του\ συμβολου}{ΕυροV\ του\ symbol\ ou}$

Οπότε αποκωδικοποιώντας την τιμή 0.59375 χρησιμοποιώντας τον πίνακα 3.3, παίρνουμε τα εξής αποτελέσματα

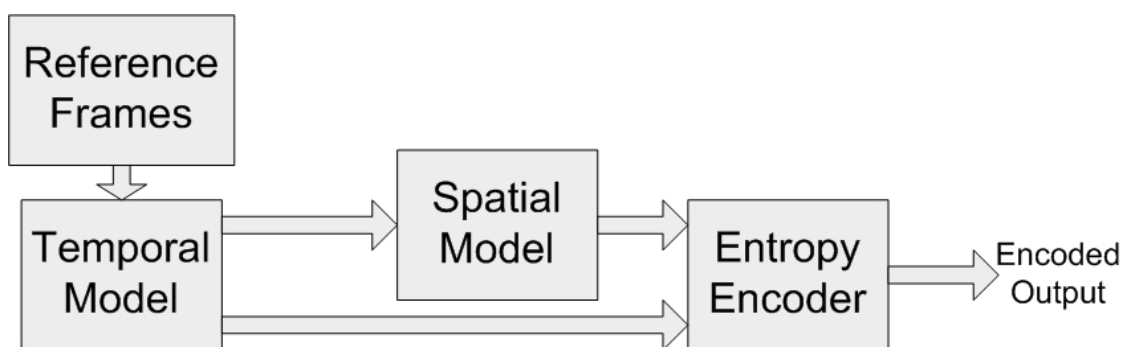
Σύμβολο	Εύρος	Αριθμός
Κ	0.25	0.59375
Α	0.5	0.375
Λ	0.25	0.75
Α	0.5	0

Πίνακας 3.5 Παράδειγμα Αποκωδικοποίησης Arithmetic Coding

Σε μερικές υλοποιήσεις δεσμεύεται ένα ειδικό σύμβολο για να υποδηλώνει το τέλος του αρχείου. Εναλλακτικά δηλώνεται το συνολικό μήκος του αρχείου, οπότε ο αλγόριθμος γνωρίζει ακριβώς σε ποιο σημείο να σταματήσει την αποκωδικοποίηση.

Απλοποιημένο Block Διάγραμμα ενός Ψηφιακού Κωδικοποιητή

Βασισμένοι στην περιγραφή των ποικίλων σταδίων κωδικοποίησης των προηγούμενων εδαφίων της παρούσης διατριβής, το ακόλουθο βασικό block διάγραμμα του σχήματος 3.15 μπορεί να συνοψίσει τη διαδικασία κωδικοποίησης ενός ψηφιακού βίντεο.



Σχήμα 3.15 Γενικευμένο block διάγραμμα κωδικοποίησης ψηφιακού βίντεο

Σύμφωνα με το διάγραμμα αυτό, παρατηρούμε τις εξαρτήσεις που υπάρχουν ανάμεσα στα διάφορα στάδια κωδικοποίησης που στοχεύουν στην μείωση του χρονικού, χωρικού και συχνοτικού πλεονασμού.

Το στάδιο της αποκωδικοποίησης δεν απεικονίζεται στο σχήμα 3.15. Η διαδικασία είναι ακριβώς η ίδια, ακολουθώντας τα βήματα της κωδικοποίησης αντίστροφα. Στο σχήμα 3.16 απεικονίζονται οι τιμές του αποκωδικοποιημένου block, του οποίου οι τιμές χρησιμοποιήθηκαν ως παράδειγμα στα διάφορα στάδια κωδικοποίησης. Συγκρίνοντας το αποκωδικοποιημένο block του σχήματος 3.16 με το αντίστοιχο αρχικό block του σχήματος 3.9, μπορούμε να παρατηρήσουμε ότι υπάρχουν μικρές μεταβολές στις τιμές των rixels από τη διαδικασία κωδικοποίησης. Οι μεταβολές αυτές είναι εντονότερες ανάλογα με το λόγο συμπίεσης που επιλέγεται στον κωδικοποιητή μέσω των δυνατών επιπέδων κβάντισης. Όσο εντονότερες είναι οι διαφορές μεταξύ του κωδικοποιημένου και του αρχικού σχήματος, τόσο εντονότερη είναι και η ποιοτική υποβάθμιση του τελικού βίντεο.

131	134	139	142	144	144	144	144
136	139	143	146	147	146	146	145
143	145	148	150	150	148	147	147
148	150	152	152	151	149	148	147
150	151	152	152	150	148	146	146
150	151	152	152	150	148	146	145
151	152	153	152	150	148	147	147
151	152	153	153	151	149	149	148

Σχήμα 3.16 Οι τιμές φωτεινότητας του αποκωδικοποιημένου block του σχήματος 3.9

Αξίζει να σημειωθεί ότι τα πρότυπα κωδικοποίησης δεν περιγράφουν λεπτομερώς τη διαδικασία του κάθε σταδίου της κωδικοποίησης, αλλά περιγράφουν τη δομή που πρέπει να έχει η κωδικοποιημένη ροή των δεδομένων του βίντεο. Συνεπώς, κάθε κωδικοποιητής μπορεί να αναπτυχθεί αξιοποιώντας διαφορετικούς αλγορίθμους στα ποικίλα στάδια της κωδικοποίησης. Οπότε διαφορετικές υλοποιήσεις του ίδιου προτύπου κωδικοποίησης καταλήγουν στην ανάπτυξη κωδικοποιητών με διαφορετική απόδοση, δημιουργώντας με τον τρόπο αυτό έντονο ανταγωνισμό μεταξύ των κατασκευαστών, καθώς και γόνιμο έδαφος για ερευνητική δραστηριότητα προς τη βελτιστοποίηση των υπαρχόντων αλγορίθμων και τεχνικών κωδικοποίησης και συμπίεσης.

Συμπεράσματα του Κεφαλαίου

Στο παρόν κεφάλαιο παρουσιάστηκαν τα ποικίλα στάδια κωδικοποίησης και συμπίεσης ενός ψηφιακού βίντεο. Οι αλγόριθμοι κωδικοποίησης βίντεο αξιοποιούν το χωρικό, χρονικό, συχνοτικό και εντροπικό πλεονασμό, αφαιρώντας την αντίστοιχη πλεονάζουσα πληροφορία. Με τον τρόπο αυτό επιτυγχάνεται συμπίεση των δεδομένων. Ανάλογα με το ποσοστό αξιοποίησης της πλεονάζουσας πληροφορίας, μεταβάλλεται η επίδραση της συμπίεσης στην ποιοτική υποβάθμιση του κωδικοποιημένου βίντεο. Τα βασικά στάδια που περιγράφηκαν στον παρόν κεφάλαιο είναι επί της ουσίας όμοια σε όλα τα πρότυπα κωδικοποίησης της MPEG και της ITU. Οποιαδήποτε περαιτέρω βελτίωση επιτυγχάνεται από την εξέλιξη των αλγορίθμων στα επιμέρους στάδια της κωδικοποίησης.

ΚΕΦΑΛΑΙΟ 4

Συνοπτική Βιβλιογραφική Ανασκόπηση των Μεθόδων και Τεχνικών Εκτίμησης της Ποιότητας Βίντεο

Ποιοτική Υποβάθμιση Βίντεο Λόγω Συμπίεσης/Κωδικοποίησης

Η πλειονότητα των προτύπων κωδικοποίησης βίντεο στηρίζεται στη συσχέτιση των γειτονικών καρέ (motion estimation/compensation), στο διακριτό μετασχηματισμό συνημίτονου, στην κβάντιση των συντελεστών του μετασχηματισμού και στην εντροπική κωδικοποίηση της σειριακής ροής δεδομένων.

Συνεπώς, η ποιοτική υποβάθμιση του τελικού κωδικοποιημένου ή συμπιεσμένου σήματος προέρχεται κυρίως από τη κβάντιση των συντελεστών του διακριτού μετασχηματισμού συνημίτονου. Η υποβάθμιση αυτή οφείλεται στο γεγονός ότι η κβάντιση είναι μερικώς αντιστρέψιμη διεργασία, με αποτέλεσμα να μην είναι εφικτή η πλήρης ανάκτηση των αρχικών δεδομένων.

Δεδομένου ότι υπάρχουν και άλλοι παράμετροι που επηρεάζουν την τελική ποιότητα του κωδικοποιημένου/συμπιεσμένου σήματος, όπως η απόδοση του αλγορίθμου εκτίμησης και αντιστάθμισης της κίνησης, παρόλα αυτά οι παράμετροι αυτοί δεν εισάγουν άμεσα κάποιο τεχνούργημα στο τελικό προϊόν, αλλά επηρεάζουν την απόδοση της συμπίεσης έμμεσα. Η μέτρια απόδοση ενός αλγορίθμου εκτίμησης της κίνησης, θα οδηγήσει κατά το στάδιο της κβάντισης σε μεγαλύτερη απώλεια δεδομένων, σε σύγκριση με έναν αποδοτικότερο αλγόριθμο. Συνεπώς εάν σε έναν απωλεστικό αλγόριθμο κωδικοποίησης αφαιρεθεί το στάδιο της κβάντισης, τότε ουσιαστικά μετατρέπεται σε μη-απωλεστικό, αφού διατηρείται πλήρως η αρχική ποιότητα στο κωδικοποιημένο/συμπιεσμένο σήμα.

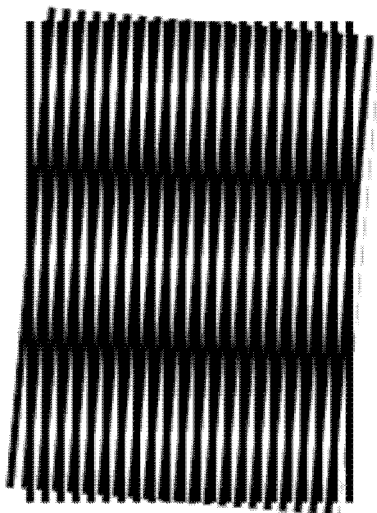
Η ποιοτική υποβάθμιση που προκαλείται εξαιτίας της κωδικοποίησης, μπορεί να περιγραφεί επιτυχώς από την εμφάνιση συγκεκριμένων τεχνουργημάτων στο συμπιεσμένο σήμα. Τα τεχνουργήματα αυτά, επειδή φέρουν συγκεκριμένα χαρακτηριστικά, μπορούν να ομαδοποιηθούν και να ταξινομηθούν ως εξής [28]:

- § Το φαινόμενο τετραγώνισης (blocking effect) ή τετραγωνοποίησης (blockiness effect) αναφέρεται στην εμφάνιση ενός τεχνουργήματος τετράγωνου σχήματος στο κωδικοποιημένο σήμα. Η εμφάνιση αυτού του φαινομένου προέρχεται από την ανεξάρτητη κβάντιση των αυτόνομων blocks (περιοχή 8x8 pixels) που περιέχουν τη μετασχηματισμένη πληροφορία, όπως προέκυψε ύστερα από εφαρμογή του DCT. Το φαινόμενο τετραγώνισης είναι η επικρατέστερη και συνηθέστερη οπτική υποβάθμιση που παρατηρείται στο κωδικοποιημένο σήμα. Εξαιτίας της μη φυσικής μορφής του, η ύπαρξη του σε ομοιογενείς περιοχές της κωδικοποιημένης εικόνας προκαλεί ευκολότερα τον εντοπισμό του από τον θεατή. Αντιθέτως, το φαινόμενο τετραγώνισης σε μη ομοιογενείς περιοχές (περιοχές με έντονη χωρική πληροφορία) γίνεται λιγότερο αντιληπτό εξαιτίας του φαινομένου μάσκας.
- § Θόλωμα της εικόνας, δηλαδή απώλεια υψηλών χωρικών συχνοτήτων και μείωση της οξύτητας των ακμών της προβαλλόμενης εικόνας. Η απώλεια των υψηλών χωρικών

συχνοτήτων οφείλεται στην κβάντιση των συντελεστών του διακριτού μετασχηματισμού συνημίτονου με αποτέλεσμα τον εκμηδενισμό εκείνων που αντιστοιχούν στις υψηλές χωρικές συχνότητες.

- § Χρωματική απόσπαση (color bleeding) είναι το φαινόμενο της εισροής χρώματος μεταξύ συνοριακών περιοχών με έντονα διαφορετικές χρωματικές συνιστώσες. Το τεχνούργημα αυτό προκαλείται από την υπό-δειγματοληψία των χρωματικών δεδομένων. Επιπλέον, η χρωματική απόσπαση εκτείνεται συνήθως εντός των ορίων ενός Macro-Block.
- § Το φαινόμενο της απεικόνισης της βάσης του DCT (DCT basis image effect) είναι κυρίαρχο όταν ένας μόνο συντελεστής DCT είναι κυρίαρχος σε ένα block. Σε επίπεδα κβάντισης χαμηλής ποιότητας, το φαινόμενο αυτό προκαλεί έμφαση της κυρίαρχης εικόνας βάσης του DCT και υποβιβασμό των υπολοίπων.
- § Πλάγιες ευθείες γραμμές στην κωδικοποιημένη εικόνα, συνήθως παρουσιάζουν το φαινόμενο της κλίμακας (staircase effect). Το φαινόμενο αυτό οφείλεται στο γεγονός ότι οι βάσεις DCT μπορούν να αποδώσουν καλύτερα οριζόντια ή κάθετες γραμμές, αφού η αναπαράσταση γραμμών με άλλη διεύθυνση απαιτούν την ύπαρξη υψηλών χωρικών συχνοτήτων, οι οποίες όμως αφαιρούνται κατά την κωδικοποίηση του σήματος.
- § Το φαινόμενο δακτυλιδιού (ringing effect), το οποίο στην ουσία πρόκειται για το φαινόμενο Gibbs στις εφαρμογές του ψηφιακού βίντεο εξαιτίας της χρήσης του διακριτού μετασχηματισμού συνημίτονου (μετασχηματισμός συγγενής του διακριτού μετασχηματισμού Fourier). Στα μαθηματικά το φαινόμενο Gibbs, το οποίο πήρε την ονομασία του από τον αμερικανό ψυχολόγο J. Willard Gibbs, αναφέρεται στην συμπεριφορά μιας σειράς Fourier σε σημείο ασυνέχειας. Πιο συγκεκριμένα παρατηρείται ότι το n άθροισμα της σειράς Fourier στο σημείο ασυνέχειας έχει σημαντικές ταλαντώσεις, οι οποίες οδηγούν σε τιμές μεγαλύτερες από τις αντίστοιχες της συνάρτησης. Επιπλέον, ο πλεονασμός αυτός δεν εξαλείφεται καθώς αυξάνεται η συχνότητα, απλά προσεγγίζει ένα συγκεκριμένο πεπερασμένο όριο. Στις εφαρμογές βίντεο, το φαινόμενο Gibbs είναι πιο εμφανές κατά μήκος των ακμών υψηλής αντίθεσης, οι οποίες κατά μία έννοια είναι οπτικά σημεία ασυνέχειας. Λαμβάνοντας υπόψιν και τη κβάντιση που πραγματοποιείται στη διαδικασία κωδικοποίησης, τότε η ασυνέχεια στις υψηλές συχνότητες ενισχύεται περαιτέρω.
- § Εσφαλμένες ακμές είναι το αποτέλεσμα της μεταφοράς ασυνεχειών στα όρια των blocks (εξαιτίας του φαινομένου τετραγώνισης) μέσω του μηχανισμού των καρτέ αναφοράς και της αξιοποίησης του χρονικού πλεονασμού.
- § Μη ομαλή κίνηση μπορεί να προκληθεί από τη χαμηλή απόδοση των μηχανισμών εκτίμησης της κίνησης, ειδικά σε περιπτώσεις όπου παρατηρείται έντονη ανομοιογένεια στις κινήσεις των αντικειμένων που απεικονίζονται σε διαδοχικά καρτέ.
- § Σε σημαντικό αριθμό των δημοφιλών προφίλ κωδικοποίησης, ο μηχανισμός εκτίμησης της κίνησης αξιοποιεί μόνο τη συνιστώσα φωτεινότητας ενώ για τη χρωματική συνιστώσα χρησιμοποιείται το ίδιο διάνυσμα κίνησης χωρίς περαιτέρω επεξεργασία. Η απλοποίηση αυτή μπορεί σε μερικές περιπτώσεις να οδηγήσει σε χρωματική ασυμφωνία ενός macroblock ή block ανάλογα με τη δυναμική του περιεχομένου.

- § Σε έντονα ομοιογενείς επιφάνειες παρατηρείται το χρονικό τεχνούργημα που ονομάζεται mosquito noise, το οποίο παρουσιάζεται ως χρωματική και φωτεινή διακύμανση γύρω από ακμές ή κινούμενα αντικείμενα. Πρόκειται για χρονικής φύσεως τεχνούργημα που προέρχεται από τη εμφάνιση προαναφερθέντων χωρικών τεχνουργημάτων (π.χ. φαινόμενο Gibbs) σε ακμές και σημεία ασυνέχειας. Εξαιτίας της διαφορετικής έντασης της εμφάνισης των φαινομένων αυτών στα διαδοχικά καρέ, το αποτέλεσμα είναι να προκαλείται αυτό το χρονικά μεταβαλλόμενο τεχνούργημα.
- § Εάν κατά τη ψηφιοποίηση των καρέ η συχνότητα δειγματοληψίας υπερβεί το όριο Nyquist τότε παρατηρείται το φαινόμενο Αναδίπλωσης (Aliasing Effect). Το τεχνούργημα αυτό αναφέρεται στο φαινόμενο που προκαλεί δύο διαφορετικά σήματα να γίνονται δυσδιάκριτα κατά τη δειγματοληψία τους με αποτέλεσμα να μην είναι δυνατή η μοναδική ανακατασκευή τους. Η ονομασία του φαινομένου προέρχεται από το γεγονός ότι στο πεδίο των συχνοτήτων παρατηρούνται περιοχές αναδίπλωσης (επικάλυψης) μεταξύ των δειγμάτων. Ένα τυπικό παράδειγμα χωρικής επικάλυψης είναι η δομή Moiré. Στο σχήμα 4.1 απεικονίζεται ένα τυπικό παράδειγμα εμφάνισης του τεχνουργήματος Moiré σε δύο σύνολα παραλλήλων με σχετική γωνία 5 μοιρών.



Σχήμα 4.1 Παράδειγμα εμφάνισης τεχνουργήματος Moiré σε δύο σύνολα παραλλήλων

Πρέπει να διευκρινιστεί ότι τα τεχνουργήματα αυτά εμφανίζονται κυρίως σε τεχνικές κωδικοποίησης που βασίζονται στην αξιοποίηση των μετασχηματισμών περιοχών blocks. Παρόλα αυτά σε άλλες τεχνικές κωδικοποίησης, όπως αυτές που χρησιμοποιούν τα wavelets, κάποια από τα προαναφερθέντα τεχνουργήματα δεν παρατηρούνται (όπως το φαινόμενο τετραγώνισης).

Μέθοδοι Εκτίμησης της Αντιληπτής Ποιότητας ψηφιακού Βίντεο

Στις εφαρμογές οπτικοακουστικού περιεχομένου είναι πολύ χρήσιμο να καθοριστεί το επίπεδο της ποιότητας υπηρεσίας (Quality of Service – QoS), με το οποίο ο τελικός χρήστης λαμβάνει μια δεδομένη υπηρεσία. Η έννοια της ποιότητας υπηρεσίας σε οπτικοακουστικές και πολυμεσικές εφαρμογές αναφέρεται στο επίπεδο της ποιότητας του οπτικού και ακουστικού περιεχομένου, όπως γίνεται αντιληπτό από τον χρήστη (Perceived Quality of Service – PQoS).

Η απαρχή της ανάπτυξης μεθόδων για την εκτίμηση της ποιότητας υπηρεσίας των εφαρμογών πολυμέσων βασίστηκε κυρίως στην εφαρμογή ενός μαθηματικού μοντέλου σφάλματος ανάμεσα στο ασυμπιέστο και στο κωδικοποιημένο/συμπιεσμένο/ποιοτικά υποβαθμισμένο ψηφιακό βίντεο. Τα πρώιμα αυτά μαθηματικά μοντέλα είχαν αρχικά προταθεί για την αξιολόγηση της απόδοσης των ποικίλων τεχνικών συμπίεσης ψηφιακής εικόνας, οι οποίες αποτέλεσαν τον πρόδρομο των μεθόδων της ψηφιακής κωδικοποίησης βίντεο.

Οι πιο διαδεδομένες πρωταρχικές μέθοδοι και τα αντίστοιχα μετρικά που βασίστηκαν σε ένα αμιγώς μαθηματικό πλαίσιο σφάλματος είναι ο λόγος κορυφής σήματος προς θόρυβο (Peak Signal to Noise Ratio – PSNR) και το μέσο τετραγωνικό σφάλμα (Mean Square Error - MSE):

$$PSNR = 10 \log_{10} \frac{L^2}{MSE}$$

Όπου το L αντιστοιχεί στη μέγιστη δυναμική τιμή που μπορεί να πάρει ένα pixel (στην περίπτωση περιγραφής 8bits/pixel μονοτονικού σήματος, το L παίρνει την τιμή 255)

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2$$

Όπου N είναι ο συνολικός αριθμός των pixels που συγκροτούν ένα πλαίσιο (frame) και x_i / y_i είναι η τιμή του i pixel στο ασυμπιέστο και στο συμπιεσμένο βίντεο αντίστοιχα.

Συνεπώς, αυτές οι πρωταρχικές μέθοδοι, αν και παρείχαν μια ποσοτική προσέγγιση της ποιοτικής υποβάθμισης του κωδικοποιημένου σήματος, παρόλα αυτά δεν παρείχαν αξιόπιστες μετρήσεις της αντιληπτής ποιότητας, αφού παρουσίαζαν σημαντικές αποκλίσεις από τις αντίστοιχες υποκειμενικές εκτιμήσεις. Η αιτία αυτής της απόκλισης ανάμεσα στην αντικειμενική μέτρηση και την υποκειμενική εκτίμηση της ποιοτικής υποβάθμισης οφείλεται κυρίως στο γεγονός ότι δεν λαμβάνονται υπόψιν τα χαρακτηριστικά, τις ιδιότητες και τις ευαισθησίες του ανθρωπίνου οπτικού συστήματος, αλλά ούτε και τα ιδιαίτερα τεχνουργήματα που εμφανίζονται από τη διαδικασία κωδικοποίησης. Σε αυτά ακριβώς τα ιδιαίτερα τεχνουργήματα και τις ευαισθησίες του ανθρωπίνου οπτικού συστήματος είναι βασισμένοι οι σύγχρονοι αλγόριθμοι εκτίμησης της αντιληπτής ποιότητας ψηφιακού βίντεο.

Σήμερα, η εκτίμηση της αντιληπτής ποιότητας βίντεο έχει πλέον εξελιχθεί πέραν των αρχικών αυτών μοντέλων, όντας πλέον αντικείμενο υποκειμενικών και αντικειμενικών διαδικασιών. Οι υποκειμενικές μέθοδοι απαιτούν για την υλοποίησή τους έναν ικανό αριθμό θεατών που θα παρακολουθήσουν και θα εκτιμήσουν υποκειμενικά την ποιότητα ενός ψηφιακού βίντεο, γεγονός που τις καθιστά ιδιαίτερα χρονοβόρες και οικονομικά μη

συμφέρουσες. Αντίστοιχα, οι αντικειμενικές τεχνικές εκτίμησης της αντιληπτής ποιότητας βασίζονται σε μοντέλα και τεχνικές που χρησιμοποιούν ποικίλα μετρικά βασισμένα στα τεχνουργήματα που προκύπτουν λόγω κωδικοποίησης, με σκοπό την επιτυχή εξομοίωση των υποκειμενικών μεθόδων. Στη συνέχεια θα παρουσιαστούν λεπτομερώς οι δύο προαναφερθείσες κατηγορίες μεθόδων εκτίμησης της αντιληπτής ποιότητας.

Υποκειμενικές Μέθοδοι Εκτίμησης της Αντιληπτής Ποιότητας Βίντεο

Οι υποκειμενικές μέθοδοι της αντιληπτής ποιότητας έχουν κυρίως προταθεί από τη διεθνή τηλεπικοινωνιακή ένωση (International Telecommunications Union - ITU) και την εξειδικευμένη ομάδα σε ζητήματα ποιότητας βίντεο (Video Quality Experts Group - VQEG). Οι τεχνικές αυτές στηρίζονται σε ένα γενικό πλαίσιο, το οποίο περιλαμβάνει ένα σύνολο θεατών τυχαία επιλεγμένων ώστε να μη διαθέτουν εξειδικευμένες γνώσεις σχετικά με τη κωδικοποίηση βίντεο. Οι θεατές καλούνται να παρακολουθήσουν και να αξιολογήσουν μια σειρά επεξεργασμένων βίντεο, υπό πλήρως ελεγχόμενες συνθήκες φωτισμού και θέασης. Η ακριβής διαδικασία παρακολούθησης (π.χ. η ταυτόχρονη ή όχι προβολή του ασυμπίεστου σήματος, η διάρκεια θέασης) και αξιολόγησης (π.χ. συνεχής για όλη τη διάρκεια θέασης ή διακριτή εφ' άπαξ στο τέλος της θέασης) διαφέρει ανάλογα με τη μέθοδο. Στη συνέχεια, πραγματοποιείται συλλογή και στατιστική ανάλυση των δεδομένων για την εξαγωγή μιας κοινής εκτίμησης. Η «Μέση Βαθμολόγηση του Κοινού» (Mean Opinion Score – MOS) θεωρείται η πιο διαδεδομένη και αξιόπιστη στατιστική μέθοδος επεξεργασίας των δεδομένων και εφαρμόζεται ευρέως σε όλες τις υποκειμενικές μεθόδους.

Οι προδιαγραφές των μεθόδων υποκειμενικής εκτίμησης περιγράφονται στα ITU-R Rec. T.500-11 και ITU-T Rec. P.910, όπου παραθέτονται συγκεκριμένες συνθήκες θέασης, κριτήρια για τους θεατές, ο τρόπος επιλογής του οπτικοακουστικού υλικού, περιγραφή της μεθόδου εκτίμησης, η βαθμολογική κλίμακα καθώς και η στατιστική μέθοδος ανάλυσης των αποτελεσμάτων. Η διαφορά των δύο προτύπων έγκειται στο γεγονός ότι το ITU-R Rec. T.500-11 αναφέρεται κυρίως σε τηλεοπτικές εφαρμογές, ενώ το ITU-T Rec. P.910 πραγματεύεται πολυμεσικές υλοποιήσεις. Παρ' όλα αυτά οι διαφορές μεταξύ των μεθόδων δεν είναι ουσιώδεις και περιορίζονται σε επιμέρους διαφοροποιήσεις.

Μια γενική κατηγοριοποίηση των υποκειμενικών τεχνικών εκτίμησης είναι ο διαχωρισμός τους σε μεθόδους μονού και διπλού οπτικού ερεθίσματος. Ο διαχωρισμός αυτός πραγματοποιείται με κριτήριο την προβολή του ασυμπίεστου σήματος ταυτόχρονα με το ποιοτικά υποβαθμισμένο. Στη συνέχεια αναλύονται οι σημαντικότερες μέθοδοι μονού και διπλού οπτικού ερεθίσματος.

Υποκειμενικές Μέθοδοι Μονού Οπτικού Ερεθίσματος

§ Single Stimulus Method (SSM) / Absolute Category Rating (ACR)

Στη μέθοδο μονού ερεθίσματος, πολλαπλές διακριτές σκηνές βίντεο προβάλλονται στο ακροατήριο, το οποίο καλείται να τις βαθμολογήσει, χωρίς να υπάρχει ταυτόχρονη προβολή των αντιστοίχων βίντεο αναφοράς. Εν γένει υπάρχουν δύο προσεγγίσεις: Οι SSM τεχνικές όπου προβάλλονται άπαξ οι υποβαθμισμένες σκηνές και οι SSM που η προβολή των υποβαθμισμένων σκηνών επαναλαμβάνεται πολλάκις. Οι βαθμολογικές κλίμακες που χρησιμοποιούνται σε αυτές τις μεθόδους είναι οι ακόλουθες:

- ο Λεκτική

Η ακόλουθη περιγραφική κλίμακα χρησιμοποιείται στην αξιολόγηση και ποσοτικοποίηση της ποιοτικής υποβάθμισης:

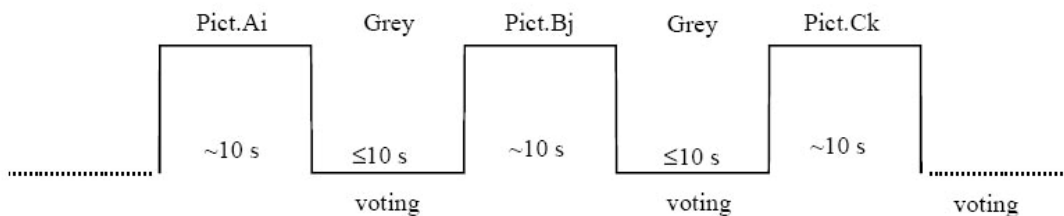
- § Ανεπαίσθητη / μη αντιληπτή (υποβάθμιση)
- § Αισθητή αλλά όχι ενοχλητική (υποβάθμιση)
- § Ελαφρώς ενοχλητική (υποβάθμιση)
- § Ενοχλητική (υποβάθμιση)
- § Πολύ Ενοχλητική (υποβάθμιση)

- ο Αριθμητική

- § Χρησιμοποιείται μια διακριτή αριθμητική κλίμακα [1,10], όπου το 1 αντιστοιχεί στη χειρίστη ποιότητα και το 10 στη βέλτιστη. Παρόλο που η συγκεκριμένη αριθμητική κλίμακα χρησιμοποιεί ακέραιες τιμές, το τελικό αποτέλεσμα x της εκτίμησης που προκύπτει από τη στατιστική επεξεργασία των μεμονωμένων ερωτηματολογίων ενδέχεται να μην είναι ακέραια αλλά πραγματική, με $1 \leq x \leq 10$

- ο Συνεχής

- § Η κλίμακα αυτή είναι συνεχής, χωρίς να προσδίδεται σε αυτή αριθμητική τιμή που να είναι γνωστή στο θεατή. Η λήψη των μετρήσεων γίνεται με τη βοήθεια μηχανικού μέσου, που έχει τη μορφή κυλιόμενης χειρολαβής. Ο θεατής βαθμολογεί με σημείο αναφοράς τη μέγιστη θέση της κυλιόμενης χειρολαβής, που αντιστοιχεί στη βέλτιστη ποιότητα. Κατά τη διάρκεια της θέασης, ο χρήστης μετακινεί την κυλιόμενη χειρολαβή συνεχώς ανάλογα με την αντιλαμβανόμενη ποιότητα, που εκείνος υποκειμενικά εκτιμά. Τη στιγμή της θέασης, κατάλληλο λογισμικό αναλαμβάνει τη δειγματοληπτική λήψη των ποικίλων θέσεων που βρίσκεται η χειρολαβή, με συχνότητα που αντιστοιχεί σε πολλές λήψεις το δευτερόλεπτο. Το ειδικό αυτό λογισμικό, αναλαμβάνει τη γραμμική αντιστοίχιση των θέσεων της χειρολαβής σε αριθμητική κλίμακα μεγάλου εύρους, όπως η εκατοστιαία (1-100). Έτσι προκύπτει μια συνεχής υποκειμενική απεικόνιση της αντιληπτής ποιότητας του κωδικοποιημένου βίντεο.



Ai Sequence A under test condition i
Bj Sequence B under test condition j
Ck Sequence C under test condition k

Σχήμα 4.2 Γραφική αναπαράσταση της μεθόδου SSM/ACR (ITU-T Rec. P.910)

Η ITU ονοματίζει την μέθοδο SSM ως Absolute Category Rating (ACR). Η αναλυτική περιγραφή της μεθόδου, καθώς και ο τρόπος εκτέλεσής της υπάρχουν στο έγγραφο ITU-R P.910 [29]. Εκεί ορίζεται ότι έπειτα από κάθε παρουσίαση ενός βίντεο υπό εκτίμηση, οι θεατές ερωτώνται να αξιολογήσουν την ποιότητά του. Η χρονική ροή πραγματοποίησης της εκτίμησης απεικονίζεται στο σχήμα 4.2. Επιπλέον, στο σχήμα 4.3 απεικονίζονται οι οδηγίες προς τους θεατές σχετικά με τη διαδικασία αξιολόγησης και την επεξήγηση της βαθμολογικής κλίμακας σύμφωνα με το έγγραφο ITU-R P.910.

Good morning and thank you for coming.

In this experiment you will see short video sequences on the screen that is in front of you. Each time a sequence is shown, you should judge its quality by using one of the five levels of the following scale.

- | | |
|---|-----------|
| 5 | Excellent |
| 4 | Good |
| 3 | Fair |
| 2 | Poor |
| 1 | Bad |

Observe carefully the entire video sequence before making your judgement.

Σχήμα 4.3 Οδηγίες προς τους θεατές και κλίμακα βαθμολόγησης της μεθόδου ACR.

Το χρονικό παράθυρο που διατίθεται στους θεατές για τη βαθμολόγηση του βίντεο ορίζεται ως μικρότερο ή ίσο των δέκα δευτερολέπτων ανάλογα με τον μηχανισμό ψηφοφορίας που χρησιμοποιείται. Επιπλέον η διάρκεια του βίντεο μπορεί να τροποποιηθεί ανάλογα με το περιεχόμενο του βίντεο που τίθεται υπό αξιολόγηση.

§ Single Stimulus Continuous Quality Evaluation (SSCQE)

Στη μέθοδο μονού ερεθίσματος με συνεχή εκτίμηση της ποιότητας [30], [31], οι θεατές καλούνται να παρακολουθήσουν ένα πρόγραμμα διάρκειας 20-30 λεπτών χωρίς να υπάρχει ταυτόχρονη προβολή του αρχικού ασυμπιέστου σήματος. Κατά τη διάρκεια της θέασης του ποιοτικά υποβαθμισμένου σήματος, οι συμμετέχοντες στη διαδικασία χρησιμοποιούν ένα μοχλό ώστε να αξιολογούν συνεχώς το επίπεδο της αντιληπτής ποιότητας. Μετακινώντας το μοχλό κάθετα, σύμφωνα με την αναγραφόμενη λεκτική κλίμακα (από «κακό» έως «άριστο»), το σύστημα αντιστοιχεί αυτόματα τις θέσεις του μοχλού σε αριθμητική κλίμακα από 1 έως 100. Ανά τακτά χρονικά διαστήματα, δειγματοληπτικά το σύστημα συλλέγει τα αριθμητικά δεδομένα που αντιστοιχούν στην αξιολόγηση των χρηστών, τα οποία στη συνέχεια με στατιστική επεξεργασία προσδίδουν μια ενιαία εκτίμηση για το υποβαθμισμένο σήμα.

Πιο αναλυτικά, η μέθοδος SSCQE χωρίζεται σε τρία στάδια, εκ των οποίων μόνο το πρώτο έχει συμπεριληφθεί στο ITU-R Recommendation 500-7:

- Στάδιο 1

Το πρώτο στάδιο περιλαμβάνει τη συνεχή εκτίμηση της οπτικοακουστικής πηγής μονού ερεθίσματος και τη συγκέντρωση δεδομένων από τις στιγμιαίες εκτιμήσεις που πραγματοποιεί ο κάθε χρήστης, παρακολουθώντας τα οπτικοακουστικά σήματα της δοκιμής. Ήδη από αυτό το στάδιο, η μέθοδος SSCQE μπορεί να θεωρηθεί πλήρης, αφού είναι δυνατή η εξαγωγή συμπερασμάτων για την αθροιστική κατανομή της ποιοτικής διακύμανσης συναρτήσει του χρόνου. Συνεπώς, το στάδιο 1 της SSCQE μεθόδου είναι κατάλληλο και ικανό για τη διεξαγωγή συγκριτικών δοκιμών.

- Στάδιο 2

Το στάδιο αυτό χρησιμοποιείται για τη διόρθωση των διαβαθμίσεων των μετρήσεων που έχουν ληφθεί κατά το πρώτο στάδιο, σύμφωνα με την κλίμακα που προτείνεται και περιγράφεται από την ITU στο έγγραφο ITU-R Recommendation BT 500-7. Ουσιαστικά πρόκειται για μια απεικόνιση των μετρήσεων σε μια κλίμακα που αποτελείται από πέντε τιμές, όπου το 5 είναι το άριστα, και στον αντίποδα βρίσκεται το 1.

- Στάδιο 3

Στο τρίτο στάδιο της υποκειμενικής εκτίμησης, προβλέπεται το φιλτράρισμα των αποτελεσμάτων με τη χρήση κατάλληλων φίλτρων, όπου θέτουν διαφορετικά βάρη στις ποικίλες εκτιμήσεις της ποιότητας. Ο σκοπός αυτού του σταδίου είναι να απαλλάξει τις μετρήσεις από φαινόμενα που προέρχονται από ψυχολογικά φαινόμενα όπως το φαινόμενο άφεσης (forgiveness effect). Πιο συγκεκριμένα, το φαινόμενο άφεσης προκαλεί στο χρήστη την εντύπωση ότι το βίντεο παραμένει ποιοτικά υποβαθμισμένο έπειτα από τη θέαση ενός οπτικού τεχνουργήματος, χωρίς αυτό να συμβαίνει στην πραγματικότητα. Οπότε, οι αντίστοιχες μετρήσεις που λαμβάνονται είναι σχετικά χαμηλές, χωρίς όμως να υπάρχει αντίστοιχο αντίκτυπο στο προβαλλόμενο σήμα. Οπότε, μέσω του τρίτου σταδίου, οι συγκεντρωμένες μετρήσεις εξομαλύνονται από λανθασμένες εκτιμήσεις που προκαλούν ποικίλα ψυχολογικά φαινόμενα.

Υποκειμενικές Μέθοδοι Διπλού Οπτικού Ερεθίσματος

§ Double Stimulus Continuous Quality Evaluation (DSCQE)

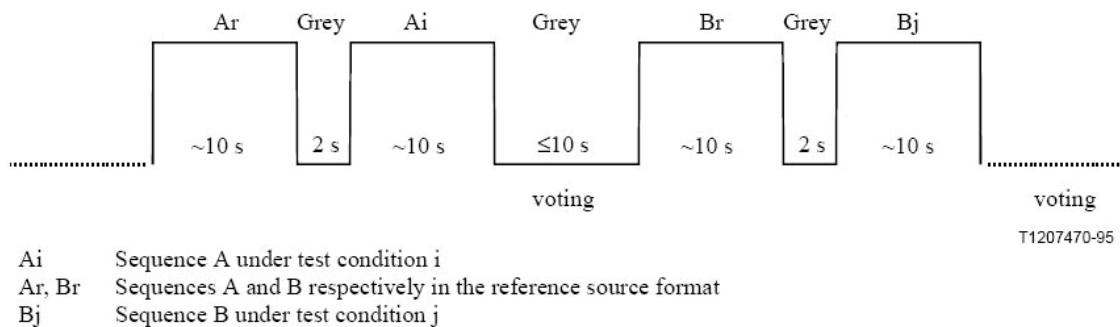
Η DSCQE είναι μια παραλλαγή της SSCQE μεθόδου, όπου η βασική διαφορά είναι η ύπαρξη του αρχικού οπτικοακουστικού περιεχομένου ως πηγή αναφοράς, μέσω της οποίας ο θεατής καλείται να αξιολογήσει το επίπεδο της αντιληπτής ποιότητας του υποβαθμισμένου σήματος συγκριτικά με το αντίστοιχο σήμα αναφοράς. Για αυτό το λόγο, η DSCQE στην υλοποίηση της χρησιμοποιεί δύο οθόνες, όπου στη μία προβάλλεται το συμπίεσμένο/κωδικοποιημένο οπτικοακουστικό σήμα, ενώ στην άλλη οθόνη προβάλλεται το βίντεο αναφοράς. Με τον τρόπο αυτό, οι θεατές εκτιμούν την αντιληπτή ποιότητα του κωδικοποιημένου σήματος, συγκρίνοντας άμεσα την πιστότητα και την ποιότητά του σε σύγκριση με το πρωτότυπο.

Κατά τη διάρκεια της αξιολόγησης, οι θεατές χρησιμοποιούν ένα μοχλό ώστε να βαθμολογούν συνεχώς το επίπεδο της αντιληπτής ποιότητας. Μετακινώντας το μοχλό κάθετα, σύμφωνα με την αναγραφόμενη λεκτική κλίμακα (από «κακό» έως «άριστο»), το σύστημα αντιστοιχεί αυτόματα τις θέσεις του μοχλού σε αριθμητική κλίμακα από 1 έως 100.

Η μέθοδος DSCQE χρησιμοποιείται κυρίως για την αξιολόγηση μικρών διαφορών μεταξύ του βίντεο αναφοράς και του ποιοτικά υποβαθμισμένου. Για αναλυτική περιγραφή τόσο της μεθόδου, όσο και της διαδικασίας εκτέλεσης αυτής ο αναγνώστης μπορεί να ανατρέξει στην αναφορά [32].

§ Double Stimulus Impairment Scale (DSIS) / Degradation Category Rating (DCR)

Η μέθοδος DSIS ή DCR προβάλλει τα βίντεο υπό τη μορφή ζευγαριών, όπου το πρώτο βίντεο είναι πάντοτε το βίντεο αναφοράς, ενώ το δεύτερο απεικονίζει το ίδιο περιεχόμενο με το πρώτο, αλλά ποιοτικά υποβαθμισμένο. Για την ταυτόχρονη προβολή των ζευγαριών μπορούν να χρησιμοποιηθούν είτε δύο ξεχωριστά μόνιτορ ή μόνο ένα, ανάλογα με τη χωρική ανάλυση που έχει επιλεγεί.



Σχήμα 4.4 Γραφική αναπαράσταση της μεθόδου DSIS (ITU-T Rec P.910)

Η χρονική ροή πραγματοποίησης της μεθόδου DSIS / DCR απεικονίζεται στο σχήμα 4.4. Το χρονικό παράθυρο που διατίθεται στους θεατές για τη βαθμολόγηση του βίντεο ορίζεται μικρότερο ή ίσο των δέκα δευτερολέπτων ανάλογα με το μηχανισμό ψηφοφορίας που χρησιμοποιείται. Επιπλέον η διάρκεια του βίντεο μπορεί να τροποποιηθεί ανάλογα με το περιεχόμενο του βίντεο που τίθεται υπό αξιολόγηση.

Κατά τη διάρκεια της αξιολόγησης οι θεατές αξιολογούν την αντιληπτή ποιότητα του δεύτερου κατά σειρά βίντεο, συγκριτικά με το αντίστοιχο βίντεο αναφοράς. Για την αξιολόγηση δίνονται σε κάθε θεατή οι ακόλουθες οδηγίες σχετικά με τη διαδικασία και με την κλίμακα βαθμολόγησης. Οι οδηγίες αυτές απεικονίζονται στο σχήμα 4.5, όπως ακριβώς περιγράφονται στο ITU-R P.910.

Good morning and thank you for coming.

In this experiment you will see short video sequences on the screen that is in front of you. Each sequence will be presented twice in rapid succession: within each pair only the second sequence is processed. At the end of each paired presentation you should evaluate the impairment of the second sequence with respect to the first one. You will express your judgement by using the following scale:

- 5 Imperceptible
- 4 Perceptible but not annoying
- 3 Slightly annoying
- 2 Annoying
- 1 Very annoying

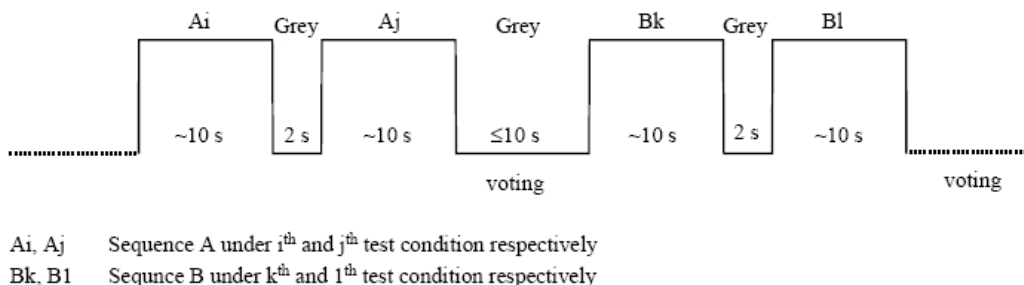
Observe carefully the entire pair of video sequences before making your judgement.

Σχήμα 4.5 Οδηγίες προς τους θεατές και κλίμακα βαθμολόγησης της μεθόδου DSIS/DCR.

Ανάλογα με το ποσοστό ακρίβειας που απαιτούμε, πραγματοποιούνται οι αντίστοιχες επαναλήψεις της πειραματικής διαδικασίας, χρησιμοποιώντας ακριβώς τις ίδιες συνθήκες θέασης και τα ίδια βίντεο.

§ Pair Comparison method (PC)

Η μέθοδος της σύγκρισης σε ζεύγη προϋποθέτει ότι το αρχικά συμπιεσμένο σήμα (π.χ. A) και το ίδιο σήμα κωδικοποιημένο ποικιλοτρόπως (π.χ. B, C) συνδυάζονται με όλους τους δυνατούς τρόπους AB, BA, AC, CA. Οπότε όλοι οι δυνατοί συνδυασμοί των υπό δοκιμή σημάτων θα λάβουν μέρος στην αξιολόγηση (πχ. AB, BA). Έπειτα από την εμφάνιση κάθε ζεύγους, ο θεατής καλείται να αξιολογήσει ποιο σήμα από τα δύο που συνιστούν το ζεύγος είναι ποιοτικά καλύτερο.



Σχήμα 4.6 Γραφική αναπαράσταση της μεθόδου DSIS (ITU-T Rec P.910).

Η χρονική ροή της αξιολόγησης απεικονίζεται στο σχήμα 4.6. Σύμφωνα με το σχήμα, ο χρόνος που έχει ο θεατής για την αξιολόγηση του προβαλλόμενου βίντεο είναι ίση ή μικρότερη των δέκα δευτερολέπτων, ανάλογα με το μηχανισμό ψηφοφορίας που έχει επιλεγεί. Ο χρόνος θέασης πρέπει να είναι περίπου ίσος με δέκα δευτερόλεπτα, ενώ μπορεί να μειωθεί ανάλογα με το περιεχόμενο του βίντεο. Επιπλέον, κατά τη αξιολόγηση μικρών χωρικών αναλύσεων μπορεί η προβολή να πραγματοποιηθεί στο ίδιο μόνιτορ.

Για την αξιολόγηση δίνονται σε κάθε θεατή οι ακόλουθες οδηγίες σχετικά με τη διαδικασία και με την κλίμακα βαθμολόγησης. Οι οδηγίες αυτές απεικονίζονται στο σχήμα 4.7, όπως ακριβώς περιγράφονται στο ITU-R P.910.

Good morning and thank you for coming.

In this experiment you will see short video sequences on the screen that is in front of you. Each sequence will be presented twice in rapid succession: each time through a different codec. The order of the sequences and the combination of codecs in the pairs vary in a random way. At the end of each paired presentation you should express your preference by ticking one of the boxes shown below. You will tick box 1 if you prefer the first sequence or box 2 if you prefer the second sequence of the pair



Observe carefully the entire pair of video sequences before making your judgement.

Σχήμα 4.7 Οδηγίες προς τους θεατές και κλίμακα βαθμολόγησης της μεθόδου DSIS/DCR.

Επιλογή της κατάλληλης μεθόδου διπλού ή μονού ερεθίσματος

Ένα σημαντικό ζήτημα στην επιλογή της μεθόδου που θα χρησιμοποιηθεί στην αξιολόγηση ψηφιακά κωδικοποιημένου βίντεο είναι το γεγονός ότι η μία κατηγορία των μεθόδων απαιτεί την ύπαρξη του αρχικού ασυμπίεστου βίντεο αναφοράς, ενώ η άλλη οικογένεια των μεθόδων χρησιμοποιεί μόνο το κωδικοποιημένο σήμα.

Η οικογένεια των μεθόδων που χρησιμοποιεί μόνο το κωδικοποιημένο σήμα στη διαδικασία αξιολόγησης δεν λαμβάνει υπόψιν την ακρίβεια και πιστότητα της απόδοσης του κωδικοποιημένου σήματος σε σύγκριση με το αντίστοιχο ασυμπίεστο. Οπότε, για αυτό το λόγο η μέθοδος DCR επιλέγεται ευρέως για την αξιολόγηση συστημάτων υψηλής ποιότητας, όπου η πιστότητα στην αναπαράσταση του βίντεο αποτελεί πρωτεύοντα παράγοντα. Επιπλέον η ύπαρξη στην βαθμολογική κλίμακα του DCR των όρων “imperceptible/perceptible” είναι πολύ σημαντική στην αξιολόγηση τέτοιων συστημάτων, διότι αποτυπώνουν εάν τα τεχνουργήματα που εισάγονται από το σύστημα γίνονται αντιληπτά από το θεατή ή όχι.

Από την άλλη μεριά η μέθοδος ACR είναι πιο εύκολα υλοποιήσιμη και γρήγορη στην εφαρμογή της, παρέχοντας αξιολογήσεις σχετικά με την ποιότητα ενός κωδικοποιημένου σήματος και το μέσο επίπεδο ποιότητας που αναμένει ένας θεατής. Συνεπώς, η μέθοδος ACR προσφέρεται για δοκιμές καταλληλότητας ενός συστήματος για τις ανάγκες του μέσου θεατή.

Σε αυτά τα πλαίσια, το μεγάλο πλεονέκτημα της μεθόδου PC είναι η μεγάλη διακριτική ακρίβεια που παρέχει στις αξιολογήσεις μεταξύ των συμπιεσμένων βίντεο. Με τον τρόπο αυτό μπορούν να ποσοτικοποιηθούν μικρές διαφορές στην αντιληπτή ποιότητα μιας πολυμεσικής υπηρεσίας, καθιστώντας κατάλληλη τη μέθοδο αυτή για σύγκριση ποικίλων τεχνικών και προτύπων κωδικοποίησης ψηφιακού βίντεο.

Τέλος, όταν επιλέγεται μεγάλος αριθμός δειγμάτων για να αξιολογηθούν με την ίδια τεχνική εκτίμησης, τότε η μέθοδος PC αποφεύγεται λόγω των πολλών συνδυαστικών παρουσιάσεων των βίντεο που την καθιστούν ιδιαίτερα χρονοβόρα. Σε αυτή την περίπτωση προτιμάται αρχικά η εκτέλεση μιας μεθόδου ACR ή DCR και στη συνέχεια έπεται μια αξιολόγηση PC μόνο για εκείνα τα βίντεο που έχουν ισοβαθμίσει.

Εναλλακτικές Μέθοδοι Εκτίμησης

Οι τεχνικές και μέθοδοι εκτίμησης που έχουν προταθεί στη διεθνή βιβλιογραφία, μπορούν να ταξινομηθούν σε δύο κατηγορίες, ανάλογα με το σκοπό της εκάστοτε μεθόδου.

- Οι μέθοδοι εκτίμησης που στοχεύουν στον καθορισμό εκείνων των παραμέτρων κωδικοποίησης (δηλ. Χωρική ανάλυση, ρυθμός των καρτέ, ρυθμός κωδικοποίησης), που απαιτούνται για την επιτυχή ολοκλήρωση μιας συγκεκριμένης πολυμεσικής εφαρμογής. Με άλλα λόγια, ο σκοπός των μεθόδων αυτών έγκειται στην εκτίμηση και στον καθορισμό εκείνου του επιπέδου της αντιληπτής ποιότητας υπηρεσίας που απαιτείται από μια συγκεκριμένη εφαρμογή.
- Οι μέθοδοι εκτίμησης που στοχεύουν στην ποσοτικοποίηση της αντιληπτής ποιότητας μιας πολυμεσικής υπηρεσίας αξιοποιώντας την εμφάνιση τεχνουργήματων που προκαλούνται από τη διαδικασία κωδικοποίησης. Σε αντίθεση με τις προαναφερθείσες τεχνικές εκτίμησης, ο σκοπός των μεθόδων αυτών δεν είναι ο εντοπισμός του κατάλληλου ποιοτικού επιπέδου ανάλογα με την παρεχόμενη υπηρεσία, αλλά η κατάταξη του κωδικοποιημένου περιεχομένου σε μια κλίμακα αντιληπτής ποιότητας.

Οι μέθοδοι της πρώτης κατηγορίας λαμβάνουν υπόψιν τους ένα σημαντικό πλήθος παραμέτρων και μετρικών για να καθορίσουν το επιθυμητό επίπεδο της αντιληπτής ποιότητας που αντιστοιχεί σε μια συγκεκριμένη υπηρεσία. Τα μετρικά αυτά εξαρτώνται άμεσα από το είδος της υπηρεσίας και από τη συναισθηματική συμπεριφορά του χρήστη [33]. Για παράδειγμα, η ταξινόμηση της υπηρεσίας ως πρωτευούσης ή δευτερευούσης σημασίας ανάλογα με την πολυπλοκότητα αυτής [34] παίζει σημαντικό ρόλο στο καθορισμό των ποιοτικών απαιτήσεών της. Για το λόγο αυτό, πλήθος παραμέτρων μετράται με σκοπό τον καθορισμό του κατάλληλου επιπέδου ποιότητας ανάλογα με το είδος της υπηρεσίας. Οι παράμετροι που λαμβάνονται υπόψιν σε αυτές τις υποκειμενικές μεθόδους περιλαμβάνουν μεταξύ άλλων:

- Τα χαρακτηριστικά του χρήστη (π.χ. γνωστικό υπόβαθρο, εξοικείωση με την εφαρμογή, ηλικία)
- Οι συνθήκες θέασης (π.χ. συνολικός αριθμό θεατών, κατανομή των χαρακτηριστικών των θεατών κλπ)
- Η φυσική κατάσταση του χρήστη κατά τη θέαση της υπηρεσίας (π.χ. καρδιακός παλμός)
- Η συμπεριφορά του χρήστη (π.χ. παρακολούθηση των κινήσεων του κεφαλιού και των ματιών του θεατή)

Παρά το μεγάλο δια-επιστημονικό ενδιαφέρον που παρουσιάζουν οι εναλλακτικοί μέθοδοι, πρέπει να τονιστεί ότι υπάρχουν πολλά άλυτα ζητήματα σε τεχνικό, θεωρητικό και πρακτικό επίπεδο σχετικά με την αποτελεσματικότητα αυτών. Ένας χρήστης που συμμετέχει σε μια τέτοια διαδικασία εκτίμησης είναι υποχρεωμένος να συμπεριφερθεί με εντελώς φυσιολογικό τρόπο ενώ φοράει σε πολλά σημεία του σώματός του τεχνικό εξοπλισμό και καλώδια – με σημαντικότερο τη συσκευή παρακολούθησης των οφθαλμών του χρήστη. Προφανώς, ένας

θεατής σε αυτή την κατάσταση δεν μπορεί να επιφέρει ανεπηρέαστα την εκτίμησή του για την προβαλλόμενη υπηρεσία.

Επίσης και από τεχνικής άποψης υπάρχουν ζητήματα που χρίζουν επίλυσης ώστε να βελτιώσουν την αξιοπιστία των εναλλακτικών αυτών υποκειμενικών μεθόδων εκτίμησης της ποιότητας. Πιο συγκεκριμένα, ζητήματα όπως η απορύθμιση της συσκευής ανίχνευσης των κινήσεων των ματιών, καθώς και η επανατοποθέτηση αυτής από εξειδικευμένο προσωπικό για την επαναλειτουργία της, θέτουν σημαντικά ερωτήματα αξιοπιστίας και εφαρμογής των τεχνικών αυτών σε πραγματικές συνθήκες [33].

Οι τεχνικές της δεύτερης κατηγορίας, αξιοποιούν την εμφάνιση συγκεκριμένων τεχνουργημάτων που προκαλούνται από τη διαδικασία κωδικοποίησης με στόχο την ποιοτική εκτίμηση του συμπιεσμένου σήματος βίντεο. Επειδή ο τρόπος ανίχνευσης και ποσοτικοποίησης της εμφάνισης των τεχνουργημάτων μπορεί να πραγματοποιηθεί ευκολότερα και ακριβέστερα με τη χρήση τεχνικών επεξεργασίας εικόνας, οι τεχνικές αυτές υλοποιούνται από αντικειμενικές μεθόδους εκτίμησης, οι οποίες αναλύονται και περιγράφονται στο ακόλουθο εδάφιο.

Αντικειμενικές Μέθοδοι Εκτίμησης της Αντιληπτής Ποιότητας Βίντεο

Οι υποκειμενικές μέθοδοι εκτίμησης της αντιληπτής ποιότητας βίντεο παρόλο που παρέχουν πολύ ακριβή αποτελέσματα εκτίμησης, είναι ιδιαίτερος πολύπλοκες στην εφαρμογή τους από πρακτική άποψη, λαμβάνοντας υπόψιν την προετοιμασία που απαιτείται για την εκτέλεση αυτών, καθώς και για την επεξεργασία των ληφθέντων μετρήσεων. Πρόκειται δηλαδή για ιδιαίτερος χρονοβόρες διαδικασίες, των οποίων η εφαρμογή και εκτέλεση περιορίζεται για αποκλειστικά επιστημονικούς και πειραματικούς σκοπούς.

Για το λόγο αυτό, μια έντονη προσπάθεια έχει ξεκινήσει πρόσφατα, η οποία έχει εστιάσει στην ανάπτυξη οικονομικότερων, ταχύτερων και ευκολότερα υλοποιήσιμων μεθόδων και τεχνικών εκτίμησης της αντιληπτής ποιότητας. Οι τεχνικές αυτές προσομοιάζουν τα αντίστοιχα αποτελέσματα που προκύπτουν από μεθόδους υποκειμενικής εκτίμησης, βασισμένες σε κριτήρια και μετρικά που μπορούν να μετρηθούν με αντικειμενικό τρόπο. Για το λόγο αυτό οι τεχνικές αυτές καλούνται αντικειμενικές και έχουν ως βασικό χαρακτηριστικό τους ότι δεν απαιτείται η ύπαρξη θεατών για την εκτίμηση της ποιότητας.

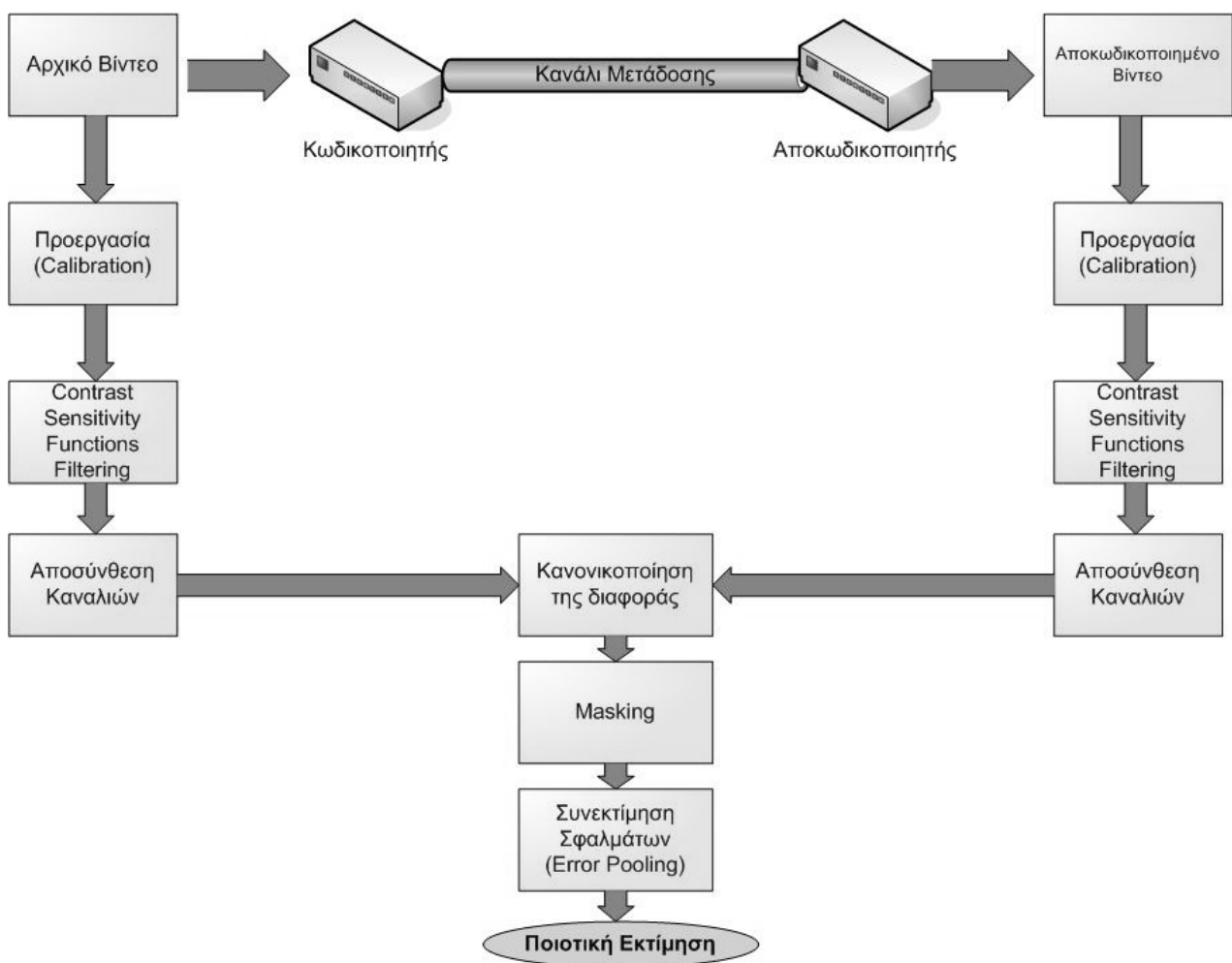
Οι αντικειμενικές μέθοδοι εκτίμησης της ποιότητας βίντεο ταξινομούνται με βασικό κριτήριο την διαθεσιμότητα και απαίτηση ύπαρξης του σαρχικού ασυμπιέστου βίντεο ως οντότητα αναφοράς, με βάση το οποίο θα πραγματοποιηθεί η εκτίμηση της ποιότητας. Επιπλέον το βίντεο αναφοράς θεωρείται ότι είναι άριστης ποιότητας. Αναλυτική βιβλιογραφική έρευνα των τεχνικών εκτίμησης της ποιότητας βίντεο μπορεί να βρεθεί στις αναφορές [3], [35], [36].

§ Αντικειμενικές Μέθοδοι Πλήρους Αναφοράς

Η πλειοψηφία των μεθόδων αντικειμενικής εκτίμησης της αντιληπτής ποιότητας κατατάσσονται στην κατηγορία μεθόδων πλήρους αναφοράς, που σημαίνει ότι απαιτείται για την εκτίμηση της ποιότητας η σύγκριση καρέ προς καρέ του ασυμπιέστου και συμπιεσμένου βίντεο αντίστοιχα. Για το λόγο αυτό, πληθώρα τεχνικών εκτίμησης της

ποιότητας εικόνας έχουν βρει εφαρμογή και στο χώρο του βίντεο, λαμβάνοντας υπόψιν ότι η σύγκριση πραγματοποιείται καρέ-καρέ, δηλαδή μεταξύ εικόνων.

Επί του συνόλου, οι τεχνικές εκτίμησης πλήρους αναφοράς μπορούν να ταξινομηθούν σε δύο κατηγορίες: Τις τεχνικές πολλαπλών καναλιών (multi channel methods) και τις τεχνικές μονού καναλιού (single channel method). Οι μέθοδοι της πρώτης κατηγορίας λαμβάνουν υπόψιν στην διαδικασία εκτίμησης πολλαπλά μετρικά και φίλτρα, τα οποία εφαρμόζονται τμηματικά στην συνολική εικόνα και στη συνέχεια συνεκτιμώνται λαμβάνοντας υπόψιν τις ευαισθησίες του ανθρωπίνου συστήματος όρασης, για την εξαγωγή μίας και μοναδικής εκτίμησης. Αντιθέτως, οι μέθοδοι της δεύτερης κατηγορίας εστιάζουν στην ανίχνευση και ποσοτικοποίηση συγκεκριμένων τεχνουργημάτων σε όλη την εικόνα ή το βίντεο –όπως το φαινόμενο τετραγώνισης- για να εκτιμήσουν το επίπεδο της αντιληπτής ποιότητας.



Σχήμα 4.8 Block διάγραμμα των μεθόδων πλήρους αναφοράς πολλαπλών καναλιών.

Σύμφωνα με το σχήμα 4.8, παρατηρούμε ότι οι τεχνικές πλήρους αναφοράς πολλαπλών καναλιών ακολουθούν εν γένει μια συγκεκριμένη διαδικασία, η οποία διαφοροποιείται σε επιμέρους σημεία. Επί του συνόλου, θα μπορούσαμε να περιγράψουμε τις τεχνικές αυτές σύμφωνα με τα ακόλουθα βασικά στάδια: Αρχικά πραγματοποιείται μια προεργασία των σημάτων βίντεο, η οποία σχετίζεται με το χωρικό και χρονικό συντονισμό των δύο σημάτων.

Στη συνέχεια ακολουθεί η εφαρμογή φίλτρων (π.χ. Sobel) στα συγκρινόμενα σήματα, τα οποία ως επί το πλείστον σχετίζονται με την αύξηση της αντίθεσης ή των ακμών του περιεχομένου. Στη συνέχεια ακολουθεί η αποσύνθεση των σημάτων στα βασικά τους κανάλια και η σύγκριση αυτών, με αποτέλεσμα τον υπολογισμό των αντιστοίχων διαφορών εφαρμόζοντας κάθε φορά το μετρικό που προτείνεται από τη εκάστοτε μέθοδο. Έχοντας υπολογίσει την εκτίμηση που αντιστοιχεί σε κάθε κανάλι, στη συνέχεια εφαρμόζεται φίλτρο-μάσκα, το οποίο στοχεύει να προσομοιάσει τα χαρακτηριστικά του ανθρωπίνου οπτικού συστήματος. Με τον τρόπο αυτό προσδίδονται διαφορετικά στατιστικά βάρη στις εκτιμήσεις του κάθε καναλιού, γεγονός που επηρεάζει τη συμμετοχή αυτών στη διαμόρφωση του τελικού αποτελέσματος που πραγματοποιείται κυρίως με την εφαρμογή του Minkowski Error Pooling.

Παρακάτω παρουσιάζονται εν συντομία μερικές αντιπροσωπευτικές μέθοδοι πολλαπλών καναλιών. Η μέθοδος Visible Difference Predictor (VDP) [37], [38] στοχεύει στον υπολογισμό ενός χάρτη πιθανοτήτων θέασης των τεχνουργημάτων μεταξύ του καρέ αναφοράς και του ποιοτικά υποβαθμισμένου. Δηλαδή κάθε σημείο του παραγόμενου χάρτη απεικονίζει την πιθανότητα ένας θεατής να παρατηρήσει την ποιοτική υποβάθμιση σε συγκεκριμένο σημείο μεταξύ του ασυμπιεστού και συμπιεσμένου σήματος. Παρόμοια τεχνική ακολουθεί και ο αλγόριθμος του Lubin [39], ο οποίος επίσης προσπαθεί να υπολογίσει τη πιθανότητα ανίχνευσης των διαφορών μεταξύ του ασυμπιεστού και του συμπιεσμένου σήματος.

Ο αλγόριθμος του Watson [40] βασίζεται στην εφαρμογή του διακριτού μετασχηματισμού συνημιτόνου και στην εφαρμογή ενός κατωφλίου ορατότητας του κάθε συντελεστή DCT. Με τον τρόπο αυτό παραλείπονται τα τεχνουργήματα που δεν γίνονται αντιληπτά από τον θεατή και ποσοτικοποιούνται μόνο οι ορατές διαφορές.

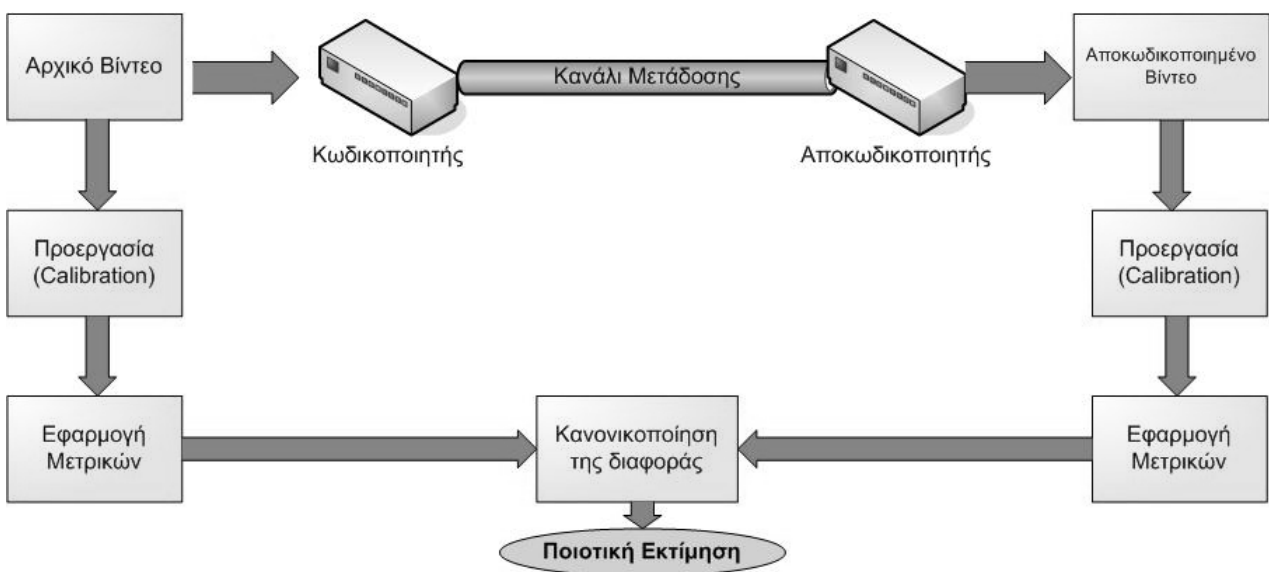
Στην [41], ο Tan et al. υλοποιούν το Video Distortion Meter χρησιμοποιώντας ένα μετρικό της αντιληπτής ποιότητας σε συνδυασμό με έναν εξομοιωτή των χρονικών φαινομένων θέασης, όπως η ομαλοποίηση (smoothing) και η χρονική συγκάλυψη (temporal masking) της υποβάθμισης, που προκαλούν ασύμμετρη ανίχνευση των τεχνουργημάτων. Το φαινόμενο αυτό αναφέρεται στο γεγονός ότι οι θεατές τείνουν να παρατηρούν ως πιο σημαντική τη μετάβαση της ποιότητας από καλή σε κακή σε σύγκριση με την αντίστροφη μετάβαση.

Στην [42] ο Winkler παρουσιάζει μια αντικειμενική μέθοδο για εκτίμηση της ποιότητας έγχρωμου βίντεο. Ο προτεινόμενος αλγόριθμος χρησιμοποιεί ένα χρωματικό μετασχηματισμό και υλοποιεί σύγκριση για την εκτίμηση της ποιότητας στον προτεινόμενο χρωματικό χώρο.

Ο A. B. Watson στην [13] ανέπτυξε το Digital Video Quality (DVQ) Metric, το οποίο εφαρμόζεται απευθείας στο DCT χώρο και για αυτό το λόγο προτιμάται από άποψη υλοποίησης, ειδικά για εφαρμογές κωδικοποίησης ψηφιακού βίντεο [43], τα οποία βασίζονται στο διακριτό μετασχηματισμό συνημιτόνου. Το μετρικό αυτό είναι βασισμένο στο διακριτό μετασχηματισμό συνημιτόνου και αξιοποιεί πτυχές του ανθρωπίνου συστήματος όρασης και επεξεργασίας, όπως προσαρμογή φωτεινότητας, αποσύνθεση σε κανάλια φωτός και χρώματος, χωρικό και χρονικό φιλτράρισμα, κανάλια ανάλυσης χωρικών συχνοτήτων, μάσκα αντίθεσης και στατιστική συνεκτίμηση όλων των προαναφερθέντων εκτιμήσεων [13]. Πιο συγκεκριμένα, το μετρικό DVQ χρησιμοποιεί το διακριτό μετασχηματισμό συνημιτόνου για να πραγματοποιήσει την αποσύνθεση του αρχικού σήματος σε χωρικά κανάλια. Το γεγονός αυτό πλεονεκτεί στην υλοποίηση του συγκεκριμένου μετρικού, δεδομένου ότι υπάρχει πληθώρα αποδοτικών λογισμικών και

υλικών που πραγματοποιούν το μετασχηματισμό αυτό. Επιπλέον, στις περιπτώσεις κωδικοποίησης βίντεο με τη χρήση κάποιου απωλεστικού προτύπου συμπίεσης (MPEG, ITU), ο μετασχηματισμός έχει ήδη πραγματοποιηθεί ως μέρος της διαδικασίας συμπίεσης/κωδικοποίησης του βίντεο.

Ένα άλλο μετρικό προτάθηκε από τον Tan και Ghanbari [44], το οποίο στοχεύει ειδικώς στην εκτίμηση της ποιότητας κωδικοποιημένου βίντεο σύμφωνα με το πρότυπο MPEG και συνδυάζει μια τυπική μέθοδο εκτίμησης σφάλματος με ένα μοντέλο μέτρησης του φαινομένου τετραγωνίσης. Το αντιληπτό μοντέλο, περιλαμβάνει περαιτέρω διόρθωση Γάμμα, υπολογισμό της αντίθεσης, χωρικό φιλτράρισμα, χρονικό φιλτράρισμα, συχνοτική αποσύνθεση σε δύο κανάλια (διαγώνιο και οριζόντιο/κάθετο), απόκριση αντίθεσης, υπολογισμό του μέσου σφάλματος, συνεκτίμηση όλων των μετρήσεων και τέλος εφαρμογή φίλτρου μάσκας της κίνησης.



Σχήμα 4.9 Block διάγραμμα των μεθόδων πλήρους αναφοράς μονού καναλιού.

Στο σχήμα 4.9 απεικονίζεται το block διάγραμμα των μεθόδων πλήρους αναφοράς μονού καναλιού. Όπως γίνεται άμεσα αντιληπτό, η βασική διαφορά μεταξύ των μεθόδων μονού και πολλαπλών καναλιών είναι ότι στις τεχνικές μονού καναλιού δεν πραγματοποιείται κατακερματισμός της εικόνας σε ποικίλα κανάλια, στα οποία εν συνεχεία θα αντιστοιχηθεί διαφορετικό στατιστικό βάρος ανάλογα με την ευαισθησία του ανθρωπίνου οπτικού συστήματος. Αντιθέτως, τα μετρικά αυτά δεν λαμβάνουν υπόψιν τους ψυχολογικούς μηχανισμούς και τις ευαισθησίες της ανθρώπινης όρασης, ακολουθώντας μια πιο τεχνική προσέγγιση. Οπότε το εκάστοτε μετρικό εφαρμόζεται σε όλη την εικόνα χρησιμοποιώντας το ίδιο στατιστικό βάρος σε κάθε σημείο αυτής.

Οι πρώτες προσπάθειες αυτού του χώρου ήταν η εφαρμογή καθαρά τεχνικών μετρικών όπως το μέσο τετραγωνικό σφάλμα ή το PSNR χωρίς να υπάρχει κάποια άμεση σύνδεση με το επίπεδο της αντιληπτής ποιότητας. Παρόλο που έχει επιδειχθεί η αναποτελεσματικότητα του συγκεκριμένου μετρικού για την επιτυχή και ακριβή εκτίμηση της ποιότητας, παρόλα αυτά χρησιμοποιείται ακόμη ευρέως στην βιβλιογραφία για να είναι δυνατή η άμεση σύγκριση με προγενέστερες δημοσιεύσεις και επιστημονικές εργασίες.

Πρόσφατα, έχουν προταθεί αντικειμενικές τεχνικές πλήρους αναφοράς μονού καναλιού, οι οποίες έχουν επιφέρει πολύ αξιόπιστα αποτελέσματα, χωρίς να απαιτούν μεγάλη υπολογιστική ισχύ ή να στηρίζονται στην πολύπλοκη εφαρμογή πλήθους φίλτρων και μασκών. Ενδεικτική τεχνική των μεθόδων πλήρους αναφοράς μονού καναλιού αυτής της κατηγορίας είναι το *SSIM*, το οποίο ποσοτικοποιεί την δομική ομοιότητα ανάμεσα στα δύο υπό εξέταση σήματα. Το μετρικό αυτό είναι βασισμένο στη βασική αρχή ότι ο θεατής εξάγει τη δομική πληροφορία από ένα οπτικό ερέθισμα. Εάν θεωρήσουμε ότι x και y είναι δύο καρέ από το ασυμπιέστο και συμπιεσμένο βίντεο αντίστοιχα, τότε το *SSIM* ορίζεται ως εξής:

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2s_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(s_x^2 + s_y^2 + C_2)}$$

όπου μ_x, μ_y είναι η μέση τιμή του x και y , $\sigma_x, \sigma_y, \sigma_{xy}$ είναι η διασπορά του x, y και συνδιασπορά του x και y , αντίστοιχα. Οι σταθερές C_1 και C_2 ορίζονται ως εξής:

$$C_1 = (K_1 L)^2 \quad C_2 = (K_2 L)^2$$

όπου L είναι δυναμικό εύρος των τιμών ενός pixel και $K_1 = 0.01, K_2 = 0.03$. [45], [46].

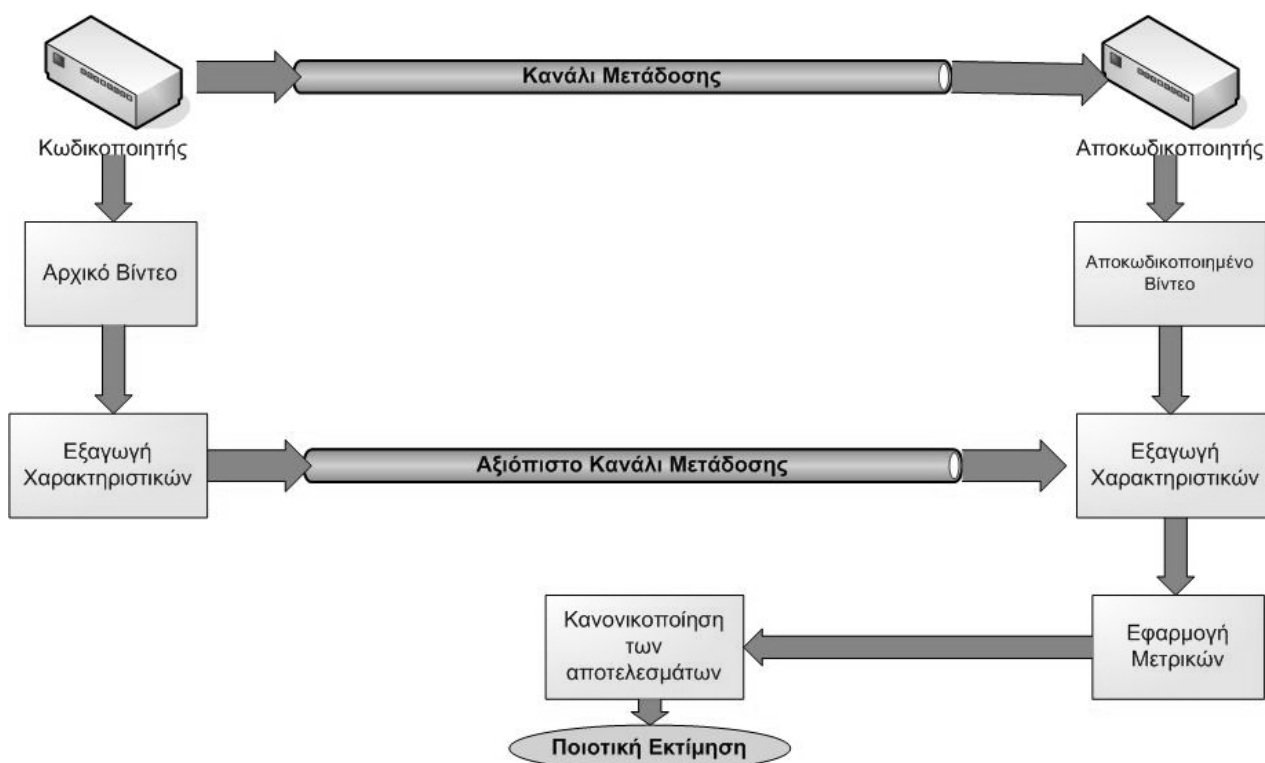
Πρόσφατα παρατηρείται μια έντονη δραστηριότητα στο χώρο των μεθόδων αντικειμενικής εκτίμησης μονού καναλιού, όπου καινούργιες τεχνικές προτείνονται αξιοποιώντας ως μετρικά την εντροπία του περιεχομένου [47], [48].

§ Αντικειμενικές Μέθοδοι Μειωμένης Αναφοράς

Οι αντικειμενικές μέθοδοι μειωμένης αναφοράς εξάγουν πληροφορίες από το ασυμπιέστο και συμπιεσμένο σήμα αντίστοιχα και τις συγκρίνουν με στόχο την εκτίμηση της ποιότητας του κωδικοποιημένου σήματος. Αυτό γίνεται με κόστος τη μετάδοση των εξαγόμενων πληροφοριών ως πλεονάζουσα πληροφορία στο κανάλι μεταφοράς, προκαλώντας αύξηση του συνολικού όγκου των δεδομένων, γεγονός ιδιαίτερος κρίσιμο στις περιπτώσεις περιορισμένου εύρους ζώνης.

Εν γένει, οι αντικειμενικές μέθοδοι μειωμένης αναφοράς –όπως απεικονίζεται και στο block διάγραμμα του σχήματος 4.10- βασίζονται στην εξαγωγή συγκεκριμένων χαρακτηριστικών στο σήμα αναφοράς και στο αποκωδικοποιημένο αντίστοιχα. Στη συνέχεια τα δεδομένα αυτά χρησιμοποιούνται από το εκάστοτε προτεινόμενο μετρικό, το οποίο τα ποσοτικοποιεί και έπειτα από μια διαδικασία κανονικοποίησης εξάγει την εκτίμηση της ποιότητας.

Στην [49] προτείνεται μέθοδος μειωμένης αναφοράς που βασίζεται στην εξαγωγή τεσσάρων χωρικών τεχνουργημάτων για κάθε καρέ του σήματος βίντεο. Στη συνέχεια υπολογίζεται ένα στατιστικό άθροισμα με βάρη των μετρήσεων αυτών, μειώνοντας το συνολικό όγκο των εξαγόμενων δεδομένων σε μία μόνο αριθμητική τιμή ανά καρέ. Στη συνέχεια σύγκριση των εκτιμώμενων τιμών μεταξύ του κωδικοποιημένου και του ασυμπιέστου βίντεο πραγματοποιούνται με σκοπό την εκτίμησης της αντιληπτής ποιότητας.



Σχήμα 4.10 Block διάγραμμα των μεθόδων μειωμένης αναφοράς

Η προσέγγιση αυτή έχει περαιτέρω επεκταθεί λαμβάνοντας υπόψιν την επίδραση διαφόρων παραμέτρων του συστήματος όπως ο ρυθμός κωδικοποίησης και η κωδικοποίηση καναλιού στην αντιληπτή ποιότητα της υπηρεσίας βίντεο [50]. Στην [51] τέσσερα χαρακτηριστικά υπολογίζονται σε ένα εναλλακτικό χρωματικό χώρο, όπου λαμβάνονται υπόψιν τα φασματικά και χρονικά χαρακτηριστικά του περιεχομένου, καθώς και το φαινόμενο τετραγώνισης. Στην [52] χρησιμοποιούνται τέσσερα μετρικά μειωμένης αναφοράς (θόρυβος, φαινόμενο τετραγώνισης, θόλωμα και φαινόμενο δακτυλιδιού) σε συνδυασμό με τρία επιπλέον χρονικά μετρικά (απώλεια πακέτων, καθυστέρηση και διασπορά καθυστέρησης). Παρομοίως, η μέθοδος που προτείνεται στην [53] βασίζεται σε μια κατάτμηση του σήματος μέσω διακριτών κυματιδιακών μετασχηματισμών που επιτρέπει την επιλεκτική μετάδοση συντελεστών και συνεπώς κυμαινόμενη πλεονάζουσα πληροφορία.

Στην [54] προτείνονται πρωτότυπα αναγνωριστικά ίχνη σημάτων βίντεο (video traces), τα οποία περιέχουν τρεις επιπλέον περιγραφείς του περιεχομένου ανά καρέ για την εκτίμηση της ποιότητας σε συνδυασμό με την πληροφορία της κίνησης. Επίσης, ένα μετρικό μειωμένης αναφοράς που βασίζεται στα στατιστικά του μετασχηματισμένου σήματος βίντεο στον κυματιδιακό χώρο προτείνεται στην [14], όπου η μέτρηση μειωμένης αναφοράς περιλαμβάνει δεκαοχτώ παραμέτρους. Επίσης μια προσπάθεια να αποφευχθεί η ανεξάρτητη μετάδοση των παραμέτρων/χαρακτηριστικών της μέτρησης μειωμένης αναφοράς πάνω από το κανάλι μετάδοσης περιγράφεται στην [55], όπου η πλεονάζουσα πληροφορία κωδικοποιείται και ενσωματώνεται μαζί με το σήμα βίντεο, υπό τη μορφή υδατοσήματος. Συνεπώς με τον τρόπο αυτό δεν παρατηρείται αύξηση στον συνολικό όγκο της μεταδιδόμενης πληροφορίας και επιπλέον δεν απαιτείται η ύπαρξη βοηθητικού αξιόπιστου καναλιού για την μεταφορά των παραμέτρων της εκτίμησης μειωμένης αναφοράς.

Επίσης στην [56] περιγράφεται μέθοδος μειωμένης εκτίμησης για τηλεοπτικά σήματα, η οποία έχει εκτενώσ δοκιμαστεί από την VQEG και προσφάτως έγινε πρότυπο από το American National Standards Institute (ANSI) [57]. Η μέθοδος αυτή λαμβάνει υπόψιν τις ποικίλες παραμέτρους του ανθρώπινου οπτικού συστήματος και προτείνει την εξαγωγή επτά διακριτών παραμέτρων από το βίντεο, που απαιτούν για τη μετάδοσή τους το 14% του συνολικού εύρους ζώνης του χρησιμοποιούμενου τηλεοπτικού καναλιού.

§ Αντικειμενικές Μέθοδοι μη-Αναφοράς

Συγκριτικά με τις προαναφερθείσες μεθόδους εκτίμησης του επιπέδου της αντιληπτής ποιότητας βίντεο, η κατηγορία των αντικειμενικών μεθόδων μη αναφοράς αποτελεί μεγαλύτερη πρόκληση από ερευνητική άποψη δεδομένου ότι δεν παρέχονται πληροφορίες για το βίντεο αναφοράς. Για το λόγο αυτό, η πλειοψηφία των τεχνικών μη αναφοράς βασίζονται κυρίως στη μέτρηση και ποσοτικοποίηση της εμφάνισης συγκεκριμένων τεχνουργημάτων που προκύπτουν από τη διαδικασία κωδικοποίησης [58].

Στις αναφορές [15], [59], [60] προτείνονται μετρικά μη αναφοράς που βασίζονται στο φαινόμενο τετραγώνισης. Αντίστοιχα στις αναφορές [61], [62] εισάγονται τεχνικές μη αναφοράς, οι οποίες αξιοποιούν μετρικά που ποσοτικοποιούν τεχνουργήματα θολώματος, ενώ παρομοίως στις αναφορές [63], [64] αξιοποιείται το ποσοστό της ευκρίνειας της εικόνας. Υπάρχουν επίσης τεχνικές οι οποίες χρησιμοποιούν ταυτόχρονα πολλαπλά μετρικά μη αναφοράς, όπως η [65], η οποία εξάγει πληροφορία για το ποσοστό του φαινομένου τετραγώνισης και του θολώματος του υπό εξέταση βίντεο και η [66] που επιπροσθέτως λαμβάνει υπόψιν και το θόρυβο της εικόνας και η [67] το φαινόμενο δακτυλιδιού. Στην αναφορά [68] προτείνεται μέθοδος εκτίμησης μη αναφοράς του φαινομένου τετραγώνισης, του θολώματος και της διακοπτόμενης κίνησης μέσω ενός μοντέλου υψηλού επιπέδου της γνωστικής συμπεριφοράς που στηρίζεται στην επίδραση του κατακερματισμού των απεικονιζόμενων προσώπων. Η μελέτη στηρίζεται στην υπόθεση ότι τα απεικονιζόμενα πρόσωπα σε ένα βίντεο προσελκύουν την προσοχή των θεατών, οπότε η ποιοτική υποβάθμιση αυτών συντελεί τα μέγιστα στην εκτίμηση της αντιληπτής ποιότητας. Στην αναφορά [69] προτείνεται μέθοδος μη αναφοράς όπου το πλήθος των διακοπών κατά τη μετάδοση του βίντεο μέσω του διαδικτύου αξιοποιείται ως μετρικό. Μια άλλη προσέγγιση [70] εκτίμησης της αντιληπτής ποιότητας με τεχνική μη αναφοράς στηρίζεται στην εξαγωγή πληροφοριών από το βίντεο σχετικά με το ρυθμό κωδικοποίησης, τη χρονική ανάλυση του σήματος και τον τύπο του κωδικοποιητή που χρησιμοποιείται. Η τεχνική αυτή εξελίχθηκε πρόσφατα περαιτέρω για να εξειδικευτεί στα χαρακτηριστικά των τελευταίων προτύπων κωδικοποίησης, όπως το H.264 [58].

Άλλες τεχνικές μη αναφοράς [71] αξιοποιούν την υπόθεση ότι η ποιοτική υποβάθμιση ενός κωδικοποιημένου σήματος είναι άμεσα εξαρτημένη από την ακρίβεια των διανυσμάτων κίνησης, από όπου μπορεί να εξαχθεί και η σχετική πληροφορία. Επίσης στη βιβλιογραφία μπορούν να βρεθούν τεχνικές [72], οι οποίες χρησιμοποιούν την ενθυλάκωση πληροφορίας εντός του σήματος μέσω της τεχνικής του υδατοσήματος. Στη συνέχεια, προτείνουν μοντέλο, το οποίο συσχετίζει την ποιοτική υποβάθμιση της αντιληπτής ποιότητας του βίντεο με την αντίστοιχη υποβάθμιση του υδατοσήματος [73], [74]. Τέλος μερικές τεχνικές χρησιμοποιούν τεχνητά νευρωνικά δίκτυα στην εκτίμησης του επιπέδου της αντιληπτής ποιότητας [75], [76], [77].

Η Συνεισφορά της Διατριβής: Από την Εκτίμηση στην Πρόβλεψη της Ποιότητας

Στα προηγούμενα εδάφια παρουσιάστηκε μια συνοπτική βιβλιογραφική ανασκόπηση των υποκειμενικών και αντικειμενικών τεχνικών εκτίμησης του επιπέδου της αντιληπτής ποιότητας βίντεο. Όπως μπορεί να παρατηρηθεί, όλες οι τεχνικές εκτίμησης που παρουσιάστηκαν, ανεξάρτητα από την αρχή λειτουργίας τους, απαιτούν την ύπαρξη του κωδικοποιημένου και ποιοτικά υποβαθμισμένου σήματος για να εξάγουν τα συμπεράσματά τους σχετικά με το επίπεδο της αντιληπτής ποιότητας. Συνεπώς, όλες οι τεχνικές πραγματεύονται την έννοια της εκτίμησης ενός ήδη κωδικοποιημένου σήματος.

Η παρούσα διδακτορική διατριβή επεκτείνει την τρέχουσα έννοια της αντικειμενικής εκτίμησης της ποιότητας βίντεο προτείνοντας πρωτότυπη μέθοδο πρόβλεψης του επιπέδου της ποιότητας ενός βίντεο εάν κωδικοποιηθεί χρησιμοποιώντας συγκεκριμένες παραμέτρους. Προς υλοποίηση της προτεινόμενης μεθόδου αξιοποιούνται ως μετρικά η χωρική και χρονική δυναμική του περιεχομένου που πρόκειται να κωδικοποιηθεί. Ταξινομώντας το προς κωδικοποίηση βίντεο ως στατικό ή έντονα δυναμικό, η προτεινόμενη μέθοδος είναι σε θέση να προβλέψει με ακρίβεια το επίπεδο της αντιληπτής ποιότητας συναρτήσει του ρυθμού κωδικοποίησης που θα χρησιμοποιηθεί κατά τη συμπίεση αυτού.

Με τον τρόπο αυτό οι αντικειμενικές τεχνικές μη αναφοράς επεκτείνονται πέρα από την απλή εκτίμηση ενός κωδικοποιημένου βίντεο, συμπεριλαμβάνοντας πλέον της έννοια της πρόβλεψης. Η προτεινόμενη τεχνική πρόβλεψης της αντιληπτής ποιότητας θα επιτρέψει στους παρόχους υπηρεσιών βίντεο να καθορίσουν σύντομα και οικονομικά για βίντεο με διαφορετική δυναμική τις παραμέτρους κωδικοποίησης που ικανοποιούν τα επιθυμητά επίπεδα αντιληπτής ποιότητας.

ΚΕΦΑΛΑΙΟ 5

Στατιστική Ανάλυση και Μοντελοποίηση Βίντεο H.264 Μεταβλητού Ρυθμού

Εισαγωγή

Οι πολυμεσικές εφαρμογές και υπηρεσίες κατέχουν σήμερα ένα σημαντικό μέρος της συνολικής κίνησης των υπολογιστικών και τηλεπικοινωνιακών δικτύων. Ανάμεσα στα ποικίλα είδη των πολυμέσων, οι υπηρεσίες βίντεο (μετάδοση κινούμενης εικόνας) είναι οι επικρατέστερες και οι πιο διαδεδομένες.

Όπως έχει ήδη αναφερθεί σε προηγούμενα κεφάλαια, οι απαιτήσεις σε χωρητικότητα και εύρος ζώνης για ασυμπιεστο ψηφιακό βίντεο είναι τόσο υψηλές, που καθιστούν πρακτικά αδύνατη τόσο την αποθήκευση, όσο και τη μετάδοσή του. Για το λόγο αυτό κρίνεται αναγκαία η χρήση προτύπων κωδικοποίησης και συμπίεσης του αρχικού σήματος βίντεο.

Από τις απαρχές των τεχνικών κωδικοποίησης/συμπίεσης ψηφιακού βίντεο, δύο βασικές μέθοδοι κωδικοποίησης έχουν προταθεί και χρησιμοποιούνται ευρέως έως και σήμερα: Η κωδικοποίηση σταθερού ρυθμού (Constant Bit Rate – CBR) και μεταβλητού ρυθμού (Variable Bit Rate – VBR). Η κωδικοποίηση μεταβλητού ρυθμού διατηρεί σταθερές τις παραμέτρους κβάντισης για όλη τη διαδικασία κωδικοποίησης σε αντίθεση με την κωδικοποίηση σταθερού ρυθμού, η οποία τις μεταβάλλει δυναμικά αντιστρόφως ανάλογα με τη χωρική δυναμική του περιεχομένου, ώστε ο τελικός ρυθμός των δεδομένων να παραμείνει σταθερός.

Σήμερα, η επιλογή του μεταβλητού ρυθμού κωδικοποίησης προτιμάται στις τηλεπικοινωνιακές εφαρμογές έναντι του σταθερού, εξαιτίας κάποιων πλεονεκτημάτων όπως [78]:

- § Καλύτερη ποιότητα βίντεο για το ίδιο μέσο ρυθμό κωδικοποίησης, χωρίς να απαιτείται η πολυπλοκότητα ενός αλγορίθμου μεταβολής των παραμέτρων κβάντισης.
- § Μικρότερη καθυστέρηση στη μετάδοση του βίντεο, εξαιτίας του γεγονότος ότι το μέγεθος της προσωρινής μνήμης στον κωδικοποιητή μπορεί να μειωθεί σημαντικά.
- § Αυξημένη χωρητικότητα αποδοχής κλήσεων, αφού το εύρος ζώνης που καταναλώνεται ανά κλήση για υπηρεσία βίντεο μεταβλητού ρυθμού είναι χαμηλότερη σε σχέση με βίντεο σταθερού ρυθμού και ισοδύναμη ποιότητα υπηρεσίας.
- § Καλύτερη αξιοποίηση της διαθέσιμης χωρητικότητας του καναλιού μετάδοσης με τη χρήση στατιστικής πολυπλεξίας ροών μεταβλητού ρυθμού σε σύγκριση με τη χρήση ισάριθμων και ισοδύναμων ροών σταθερού ρυθμού.

Παρόλα αυτά, όταν θεωρούμε μετάδοση και στατιστική πολυπλεξία μιας ροής βίντεο μεταβλητού ρυθμού, μέσω ενός κοινού τηλεπικοινωνιακού καναλιού (όπως τα υπολογιστικά δίκτυα), η βελτίωση στην αποδοτικότερη αξιοποίηση των διαθέσιμων δικτυακών πόρων δεν μπορεί μονοσήμαντα να οριστεί από το λόγο συμπίεσης των δεδομένων του βίντεο. Η κωδικοποίηση μεταβλητής ροής εμπεριέχει μεγάλες διακυμάνσεις, οι οποίες είναι συσχετισμένες με τη δυναμική του περιεχομένου του βίντεο [79]. Η ευμετάβλητη φύση των χαρακτηριστικών της κίνησης βίντεο μεταβλητού ρυθμού, δημιουργεί

μια πρόκληση στον αποδοτικό σχεδιασμό των τηλεπικοινωνιακών δικτύων και των μηχανισμών ελέγχου της κίνησης. Κατά συνέπεια, μια λεπτομερής στατιστική μελέτη των ιδιοτήτων της κίνησης είναι απαραίτητη για τη βέλτιστη διαχείριση του δικτύου, καθώς και για την κατανόηση των ενδογενών ιδιοτήτων της κωδικοποίησης βίντεο.

Σε αυτά τα πλαίσια, το παρόν κεφάλαιο παρουσιάζει μια στατιστική μελέτη και ανάλυση βίντεο H.264 κωδικοποιημένο σε μεταβλητό ρυθμό.

Σχετική Βιβλιογραφία στην Στατιστική Ανάλυση Κίνησης Ψηφιακού Βίντεο

Οι πρώτες μελέτες στην ανάλυση και μελέτη των στατιστικών ιδιοτήτων των δεδομένων βίντεο κωδικοποιημένων με τη μέθοδο μεταβλητού ρυθμού, περιελάμβαναν ποικίλες προσεγγίσεις, όπως τις αριθμητικές διαφορές στη συνιστώσα φωτεινότητας μεταξύ διαδοχικών καρέ [80] ή την κατανομή της διάρκειας των σκηνών του περιεχομένου του βίντεο [81]. Στη συνέχεια, κάποιες εργασίες εστίασαν στην ανάλυση της κίνησης ανά καρέ και GOP [82], [83].

Αποτελέσματα από τις εργασίες αυτές έδειξαν ότι η κατανομή του μεγέθους ενός καρέ ακολουθεί την κανονική κατανομή [84],[85],[86]. Επιπλέον, οι αυτοσυσχετίσεις των παραγόμενων ροών βίντεο βρέθηκαν να έχουν μια φθίνουσα εκθετική μορφή [84],[85],[87],[88],[89] ενώ κάποιες άλλες εργασίες [86], [90], [91] που εξέτασαν το φαινόμενο πιο διεξοδικά, κατέληξαν σε ένα πιο αναλυτικό συμπέρασμα προσδιορίζοντας ότι η φθίνουσα κλίση της αυτοσυσχέτισης είναι μεγάλη στα πρώτα βήματα και σταδιακά φθίνει.

Όπως έχει ήδη αναφερθεί, τα πιο διαδεδομένα πρότυπα κωδικοποίησης είναι αυτά που έχουν αναπτυχθεί και προταθεί από την MPEG και την ITU. Το τελευταίο επίτευγμα που προήλθε από τη συνεργασία των δύο αυτών φορέων είναι το πρότυπο H.264/MPEG-4 AVC [92]. Για πληρότητα του παρόντος κεφαλαίου, συνοπτικά αναφέρουμε ότι το νέο πρότυπο επιτυγχάνει 40-50% αποδοτικότερο λόγο κωδικοποίησης σε σύγκριση με το προγενέστερο πρότυπο που έχει γνωρίσει μεγάλη εμπορική επιτυχία: το MPEG-2. Εξαιτίας των προτερημάτων του νέου προτύπου έναντι προγενέστερων του [93], [94], αναμένεται ότι θα επικρατήσει στα μελλοντικά τηλεπικοινωνιακά δίκτυα και στα συστήματα κινητών τηλεπικοινωνιών, καθιστώντας αναγκαία τη διερεύνηση των στατιστικών του ιδιοτήτων και την ανάπτυξη μοντέλων που θα προσομοιάζουν ροές βίντεο H.264.

Σε αυτά τα πλαίσια το παρόν κεφάλαιο παρουσιάζει μια στατιστική ανάλυση ενός βίντεο κωδικοποιημένου σύμφωνα με το πρότυπο H.264. Η διεξαχθείσα ανάλυση βοηθάει στην κατανόηση των συσχετίσεων και ιδιοτήτων που εμφανίζονται στη ροή βίντεο H.264. Επιπλέον, καθίσταται δυνατή η σύγκριση με παλαιότερα πρότυπα, γεγονός που μπορεί να οδηγήσει στη διάκριση εγγενών διαφορών, οι οποίες θα συμβάλλουν στην καλύτερη διαχείριση των τηλεπικοινωνιακών δικτύων που θα χρησιμοποιήσουν το πρότυπο αυτό για τη διανομή πολυμεσικών υπηρεσιών.

Το Πρότυπο H.264: Βασικές Ιδιότητες και Παρούσα Κατάσταση

Για πληρότητα του παρόντος κεφαλαίου, σε αυτή την ενότητα επαναλαμβάνουμε εν συντομία τις βασικές ιδιότητες του προτύπου H.264 και συζητάμε την παρούσα κατάσταση στην εξέλιξη του προτύπου.

Το 1998 η ITU-T VCEG ανακοίνωσε ένα κάλεσμα για προτάσεις (H.26L Project) με κύριο στόχο την ανάπτυξη ενός νέου προτύπου που θα διπλασιάσει την απόδοση της κωδικοποίησης σε σύγκριση με τα υπάρχοντα πρότυπα. Το 2001, η VCEG και η MPEG δημιούργησαν την Joint Video Team (JVT) με στόχο να ολοκληρώσουν από κοινού το υπό ανάπτυξη πρότυπο και να το καταθέσουν για επίσημη έγκριση με την ονομασία H.264/MPEG-4 AVC [92].

Το νέο πρότυπο έχει να επιδείξει σημαντικά βήματα προόδου στον τομέα της αποδοτικότερης κωδικοποίησης, της ευρωστίας του σήματος μετάδοσης και της πληθώρας των πιθανών εφαρμογών του [95]. Ενδεικτικά, μερικά από τα σημαντικότερα νέα χαρακτηριστικά του προτύπου είναι:

- § Το μεταβλητό μέγεθος του block, με μεγέθη που φτάνουν έως και 4x4 σε συνδυασμό με μετασχηματισμούς αντίστοιχου μεγέθους.
- § Υψηλή ακρίβεια $\frac{1}{4}$ των διανυσμάτων κίνησης.
- § Εκτεταμένη επιλογή καρέ αναφοράς.
- § Ενσωματωμένο φίλτρο de-blocking στο μηχανισμό αντιστάθμισης της κίνησης.
- § Ποικίλους μεθόδους εντροπικής συμπίεσης (CAVLC και CABAC).

Ο κύριος στόχος των παραπάνω βελτιώσεων στο νέο πρότυπο είναι η βελτιστοποίηση του λόγου της ποιότητας βίντεο προς το ποσοστό συμπίεσης του αρχικού σήματος. Επιπλέον, τρία βασικά προφίλ έχουν προταθεί:

- § Το Baseline Profile: Σχεδιάστηκε για να ελαχιστοποιήσει την πολυπλοκότητα και να προσφέρει υψηλά επίπεδα ευρωστίας με ευέλικτο τρόπο πάνω από μια πληθώρα δικτυακών τεχνολογιών και συνθηκών μετάδοσης.
- § Το Main Profile: Σχεδιάστηκε με έμφαση στην βελτιστοποίηση της απόδοσης στη διαδικασία κωδικοποίησης.
- § Το Extended Profile: Σχεδιάστηκε για να επεκτείνει την ευρωστία του Baseline Profile με μεγαλύτερου βαθμού ευρωστία και αποδοτικότητα στη διαδικασία κωδικοποίησης.

Αυτή τη στιγμή, το Baseline Profile φαίνεται ότι προσφέρει μια ευρέως ικανοποιητική λύση για τις υπηρεσίες ψηφιακού βίντεο. Πρόσφατα η JVT ολοκλήρωσε τις εργασίες της σε μια τροποποίηση και επέκταση του προτύπου, ώστε να συμπεριλάβει κλιμακωτή κωδικοποίηση (Scalable Video Coding - SVC) στο ήδη υπάρχον πρότυπο. Από άποψη δομής, μια ροή δεδομένων κλιμακωτής κωδικοποίησης περιλαμβάνει μία βασική ροή και μία (ή περισσότερες) ροές βελτίωσης, που δρουν συμπληρωματικά ως προς την αντίστοιχη βασική ροή. Η βασική ροή πρέπει να συμβαδίζει με τις προδιαγραφές ενός εκ των τριών προφίλ του H.264. Συνεπώς με το νέο αυτό προφίλ, ένας χρήστης θα μπορεί να βελτιώνει τη χωρική ή χρονική ανάλυση ή ποιότητα ή συνδυασμό αυτών που λαμβάνεται μέσω της βασικής ροής, με τη λήψη ενός ή περισσότερων ροών βελτίωσης της κλιμακωτής κωδικοποίησης.

Επίσης, πραγματοποιούνται και ενέργειες προς την επέκταση του προτύπου ώστε να συμπεριλάβει πιο απαιτητικές εφαρμογές σε επίπεδο χωρικής ανάλυσης, χρωματικής αναπαράστασης και απόδοσης του τελικού λόγου κωδικοποίησης [96], [97], [98].

Στατιστική Ανάλυση Κωδικοποιημένου Σήματος Σύμφωνα με το Πρότυπο H.264

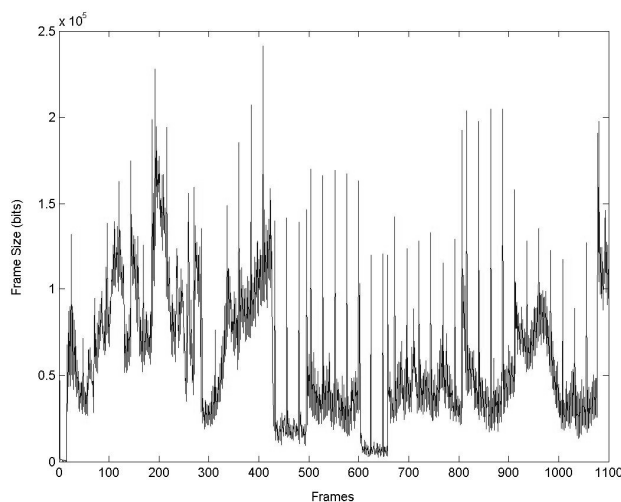
Για τη στατιστική ανάλυση σήματος κωδικοποιημένου σύμφωνα με το πρότυπο H.264, ο κωδικοποιητής αναφοράς JM χρησιμοποιήθηκε για την παραγωγή των κωδικοποιημένων σημάτων. Κατά τη διάρκεια της συμπίεσης επιλέχθηκε κωδικοποίηση μεταβλητού ρυθμού, χωρίς της ύπαρξη μηχανισμού ελέγχου ροής (δηλαδή σταθεροί συντελεστές κβάντισης). Όπως και στα λοιπά πρότυπα κωδικοποίησης, παρομοίως και στο πρότυπο H.264 χρησιμοποιούνται τα τρία είδη καρτέ, δηλαδή τα I, B και P. Ο συνδυασμός διαδοχικών I, P και B καρτέ στο κωδικοποιημένο σήμα, συγκροτούν ένα Group of Pictures (GOP), του οποίου το μήκος ορίζεται από την απόσταση δύο διαδοχικών καρτέ I και από την περιοδικότητα των B και P αντίστοιχα. Για τις ανάγκες του παρόντος κεφαλαίου επιλέχθηκε η ακόλουθη δομή GOP IPBPBPBPBP... με μήκος GOP από 3 έως και 12. Η χρονική ανάλυση καθορίστηκε στα 25 καρτέ το δευτερόλεπτο για όλη την πειραματική διαδικασία. Ως περιεχόμενο υπό δοκιμή επιλέχθηκε ένα απόσπασμα από την ταινία δράσης Spider-Man II, που περιείχε συνολικά 18357 καρτέ σε μορφή YUV 4:2:0 και χωρική ανάλυση 528x384. Για επέκταση των αποτελεσμάτων του παρόντος κεφαλαίου σε διαφορετικές χωρικές αναλύσεις, ο αναγνώστης μπορεί να ανατρέξει στην αναφορά [99].

- Ανάλυση σε Επίπεδο Καρέ

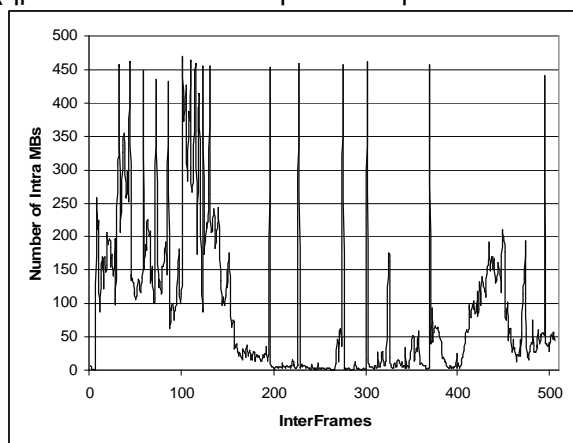
Αρχικά πραγματοποιείται στατιστική ανάλυση σε επίπεδο καρτέ, που σημαίνει ότι θα εξεταστούν όλα τα καρτέ ως ισοδύναμα αδιαφορώντας για τον τύπο τους (δηλ. I, B ή P). Στο σχήμα 5.1 παρουσιάζονται τα αντίστοιχα αποτελέσματα. Σύμφωνα με το γράφημα αυτό, στο οποίο απεικονίζεται ένα απόσπασμα 1100 καρτέ από το συνολικό σήμα που κωδικοποιήθηκε με συντελεστή κβάντισης 20 σε όλα τα καρτέ και μήκος GOP 12 καρτέ, παρατηρούμε τη μεγάλη διακύμανση του μεγέθους κάθε καρτέ. Μελετώντας πιο αναλυτικά το πειραματικό αυτό γράφημα, προκύπτει το συμπέρασμα ότι η διακύμανση στο μέγεθος του κάθε καρτέ έχει άμεση ταύτιση με το είδος του καρτέ, αφού οι υψηλότερες περιοδικές κορυφές αντιστοιχούν σε I καρτέ, ενώ οι μικρότερες αντιστοιχούν σε B καρτέ, τα οποία παρουσιάζουν και το μεγαλύτερο λόγο συμπίεσης.

Επιπλέον, η περιοδικότητα δύο συνεχόμενων κορυφών που αντιστοιχούν σε I καρτέ, στην ουσία καθορίζει το μήκος της δομής GOP. Είναι επίσης προφανές ότι το γράφημα του μεγέθους του κάθε καρτέ ακολουθεί τη δυναμική του περιεχομένου, αφού οι κορυφές που αντιστοιχούν σε καρτέ με περιεχόμενο υψηλής δυναμικής έχουν μεγαλύτερο μέγεθος σε σύγκριση με τα καρτέ που αντιστοιχούν σε στατικό περιεχόμενο, τα οποία απαιτούν λιγότερα bits για την επιτυχή περιγραφή τους, εξαιτίας του υψηλότερου λόγου συμπίεσης.

Επίσης μια άλλη ενδιαφέρουσα παρατήρηση είναι ότι τα inter-frames, και ειδικότερα τα P καρτέ, παρουσιάζουν μια πιο έντονη διακύμανση σε σύγκριση με τα Intra-frames. Αυτό προέρχεται από το γεγονός ότι ανάλογα με τη δυναμική του περιεχομένου, κάποια Macro-Blocks που ανήκουν σε inter-frame και αποτυγχάνουν να συσχετιστούν με προηγούμενα καρτέ, τότε μοιραία κωδικοποιούνται ως intra. Αυτό συνεπάγεται μείωση του λόγου συμπίεσης, με αποτέλεσμα την παραγωγή καρτέ μεγαλύτερου μεγέθους. Το σχήμα 5.2 απεικονίζει τον συνολικό αριθμό των Macro-Blocks που κωδικοποιήθηκαν ως Intra από τα συνολικά 1100 καρτέ του σχήματος 5.1. Όπως μπορεί να παρατηρηθεί από τη μορφή του γραφήματος με τον αριθμό των Intra κωδικοποιημένων Macro-Blocks, η επίδραση των P και I καρτέ είναι σημαντική στην τελική μορφή του γραφήματος του μεγέθους κάθε καρτέ (σχήμα 5.1). Εν ολίγοις, προκύπτει ότι τα inter-frames επηρεάζουν ουσιαστικά τη μορφή της παραγόμενης κίνησης βίντεο.



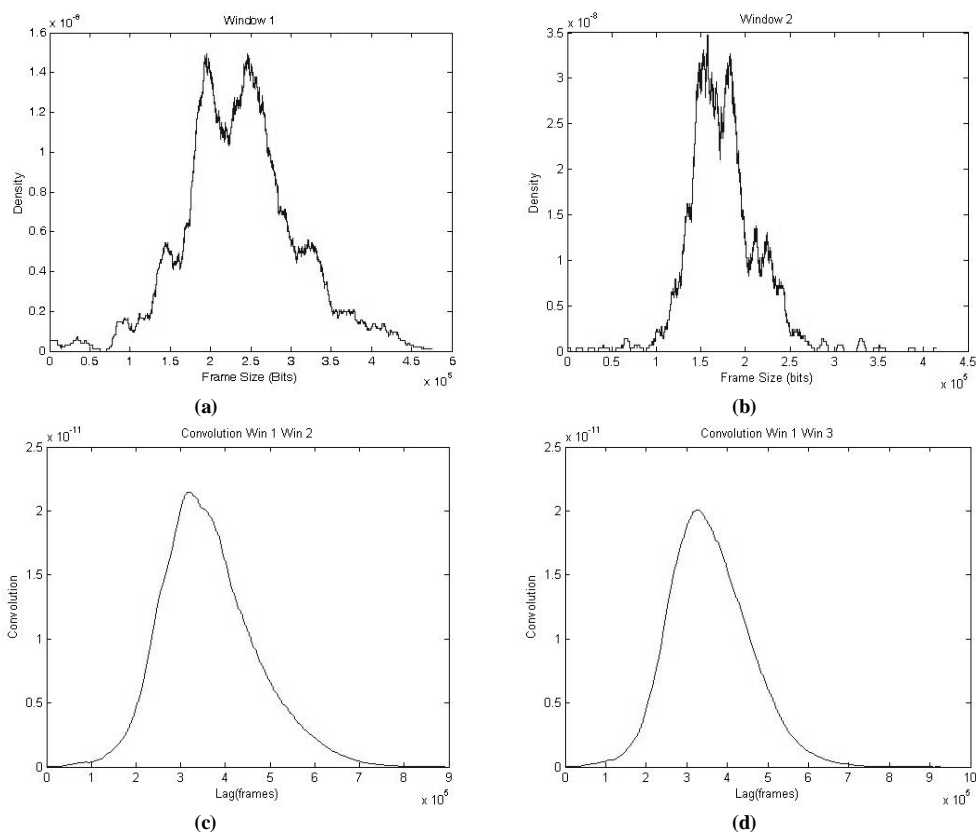
Σχήμα 5.1 Στατιστική Ανάλυση σε επίπεδο καρτέ



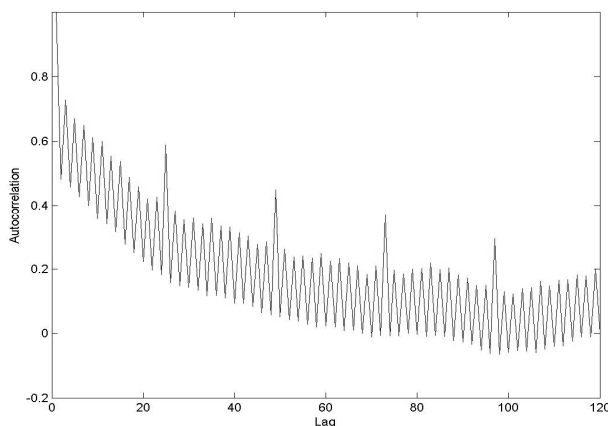
Σχήμα 5.2 Συνολικό αριθμός Macro-Blocks που κωδικοποιήθηκαν ως Intra

Ένα βασικό ζήτημα στη ανάλυση και διερεύνηση των στατιστικών ιδιοτήτων της κίνησης βίντεο μεταβλητού ρυθμού, είναι εάν τηρείται η ιδιότητα της στάσιμης διεργασίας. Δηλαδή εάν η κωδικοποιημένη κίνηση βίντεο διατηρεί σε όλη της την έκταση τις ίδιες στατιστικές ιδιότητες. Για το λόγο αυτό, χρησιμοποιήθηκαν τέσσερα διακριτά χρονικά παράθυρα από την ταινία δράσης “Spider-Man II”. Για κάθε ένα από αυτά τα διακριτά παράθυρα, υπολογίστηκε η συνάρτηση πυκνότητας πιθανότητας (Probability Density Function) του μεγέθους του κάθε καρτέ.

Οι παραγόμενες συναρτήσεις πυκνότητας πιθανότητας απεικονίζονται στο σχήμα 5.3 και παρουσιάζουν μεγάλη ομοιότητα, γεγονός που παραπέμπει ότι τα εξεταζόμενα δείγματα μπορούν από στατιστικής άποψης να θεωρηθούν ως μια στάσιμη διεργασία [82], [100]. Για περαιτέρω εξέταση του δείγματος, επεκτείνεται η πειραματική διαδικασία στην εξέταση στάσιμης διεργασίας δεύτερης τάξης [82], [100]. Για το σκοπό αυτό, υπολογίστηκαν οι αυτοσυσχετίσεις των αντιστοίχων ζευγαριών των συναρτήσεων πυκνότητας πιθανότητας του σχήματος 5.3. Τα αποτελέσματα απεικονίζονται στο σχήμα 5.3 (c) και (d), επιδεικνύοντας σχεδόν πανομοιότυπη μορφή για όλα τα εξεταζόμενα ζευγάρια. Κατά συνέπεια, ενισχύεται περαιτέρω η προαναφερθείσα άποψη ότι το εξεταζόμενο δείγμα μπορεί να θεωρηθεί ως στάσιμη διεργασία.



Σχήμα 5.3 Ιστογράμματα μεγέθους των καρτέ (a), (b) και οι αυτοσυσχετίσεις αυτών (c) και (d)



Σχήμα 5.4 Αυτοσυσχέτιση των μεγεθών 1100 καρτέ

Στη συνέχεια υπολογίστηκε η συνάρτηση αυτοσυσχέτισης για 1100 καρτέ και το αποτέλεσμα απεικονίζεται στο σχήμα 5.4. Όπως μπορεί να παρατηρηθεί η συνάρτηση αυτοσυσχέτισης αποτελείται από μια υπέρθεση περιοδικών κορυφών πάνω σε μια φθίνουσα καμπύλη. Οι υψηλότερες κορυφές αντιστοιχούν στην αυτοσυσχέτιση των Intra καρτέ της ακολουθίας βίντεο, οι οποίες ενδιάμεσα ακολουθούνται από 11 χαμηλότερες κορυφές μέχρι την επόμενη Intra κορυφή. Αυτές οι ενδιάμεσες κορυφές αντιστοιχούν σε P καρτέ, οι οποίες εξ' ορισμού αναμέναμε να είναι μικρότερες από τις αντίστοιχες I κορυφές. Τέλος, οι κοιλάδες που σχηματίζονται ανάμεσα σε δύο διαδοχικές P κορυφές αντιστοιχούν

σε B καρέ του βίντεο, τα οποία όπως έχουμε ήδη σχολιάσει έχουν το μικρότερο μέγεθος από τα δύο.

Βασισμένοι στα έως τώρα στατιστικά αποτελέσματα, προκύπτει το συμπέρασμα ότι το κωδικοποιημένο σήμα H.264 μπορεί επιτυχώς να περιγραφεί ως μια υπέρθεση τριών αυτόνομων κατανομών, οι οποίες προκύπτουν από τα τρία διαφορετικά είδη καρέ (δηλαδή τα I, P και B). Κατά συνέπεια, εστιάζοντας στην κατανομή κάθε τύπου ξεχωριστά και αναλύοντάς την, μπορούμε να εξάγουμε πιο αναλυτικά και λεπτομερή αποτελέσματα για τις στατιστικές ιδιότητες της κίνησης βίντεο H.264. Σύμφωνα με τις παρατηρήσεις αυτές, στην επόμενη ενότητα εξετάζεται η στατιστική μελέτη και ανάλυση σε επίπεδο I/B/P καρέ.

- Ανάλυση σε Επίπεδο I/B/P

Για την ανάλυση σε επίπεδο I/B/P χρησιμοποιήθηκε, όπως και στην προηγούμενη ενότητα, το απόσπασμα των 18357 καρέ από την ταινία δράσης “Spider-Man II” σε μορφή YUV 4:2:0 και χωρική ανάλυση 528x384. Για να μπορέσουμε να εστιάσουμε στις ενδογενείς στατιστικές ιδιότητες του κωδικοποιημένου σήματος, εξετάσαμε κωδικοποιήσεις με διάφορες τιμές των παραμέτρων κωδικοποίησης και μήκη GOP. Σε κάθε κωδικοποίηση, καταγράφηκαν τα δεδομένα που παρέχει ο κωδικοποιητής, όπως το είδος του καρέ και το μέγεθός του. Ο πίνακας 5.1 περιέχει τη μέση τιμή και την απόκλιση για τις κωδικοποιημένες ακολουθίες που χρησιμοποιήθηκαν στην πειραματική διαδικασία. Στον πίνακα αυτό χρησιμοποιείται ο συμβολισμός (x,y,z)-I για να απεικονίσει τον εκάστοτε συντελεστή κβάντισης για τα καρέ I,B,P και το μήκος της δομής GOP αντίστοιχα.

Από τον πίνακα 5.1 μπορεί να εξαχθεί ότι μεγαλύτεροι συντελεστές κβάντισης, οι οποίοι συνεπάγονται χαμηλής ποιότητας κωδικοποίηση, αντιστοιχούν σε μικρότερα μεγέθη καρέ και απόκλισης σε σύγκριση με χαμηλότερες παραμέτρους κβάντισης. Επιπλέον, διαφαίνεται ότι η μεταβολή της περιόδου επανεμφάνισης I καρέ, δηλαδή το μήκος του GOP, δεν επηρεάζει τις αντίστοιχες στατιστικές τιμές, οι οποίες παραμένουν πρακτικά σταθερές. Διευκρινίζεται ότι η μέση τιμή, η τυπική απόκλιση, η μέγιστη και η ελάχιστη τιμή που απεικονίζονται στον πίνακα 5.1 είναι σε Kbits.

Quantization Settings / Frame Types	I Frames (in Kbits)				B Frames (in Kbits)				P Frames (in Kbits)			
	Mean	σ	min	max	Mean	σ	min	max	Mean	σ	min	max
(10,10,10)-12	354.47	87.29	35.7	610.1	227.65	57.67	6.3	576.5	271.34	65.67	24.2	588.7
(20,20,20)-12	148.16	58.25	2.62	325.8	43.02	33.17	0.26	291.5	67.01	41.98	0.62	294.8
(30,30,30)-12	53.91	25.68	1.70	146.7	7.86	8.71	0.24	109.1	16.33	13.93	0.24	116.9
(20,20,20)-3	147.21	57.32	2.62	325.8	42.86	33.25	0.28	290.6	67.22	42.15	0.51	294.8
(20,20,20)-6	148.46	58.03	2.62	325.8	43.10	33.30	0.26	291.5	67.06	42.08	0.50	294.8

Πίνακας 5.1 Στατιστικά ανά είδος καρέ για το κωδικοποιημένο βίντεο

Παρομοίως, ο πίνακας 5.2 περιέχει τους συντελεστές απόκλισης για ποικίλες τιμές των συντελεστών κβάντισης, οι οποίες είναι ενδεικτικές του σχήματος των κατανομών που αντιστοιχούν στο μέγεθος του κάθε τύπου καρέ. Ο συντελεστής απόκλισης ορίζεται ως S_x/\bar{x} , όπου S_x είναι η τυπική απόκλιση και \bar{x} η μέση τιμή. Επιπλέον, οι εξαγόμενες τιμές είναι συνεπείς με τις προαναφερθείσες παρατηρήσεις σχετικά με τον πίνακα 5.1.

Quantization Settings / Frame Types	I Frames	B Frames	P Frames
	Variation Coefficient	Variation Coefficient	Variation Coefficient
(10,10,10)-12	0.2463	0.2533	0.2420
(20,20,20)-12	0.3932	0.7710	0.6265
(30,30,30)-12	0.4763	1.1081	0.8530
(20,20,20)-3	0.3894	0.7758	0.6271
(20,20,20)-6	0.3909	0.7726	0.6275

Πίνακας 5.2 Συντελεστές απόκλισης για ποικίλες παραμέτρους κβάντισης

Για να μελετήσουμε περαιτέρω τις στατιστικές ιδιότητες του κωδικοποιημένου σήματος, υπολογίστηκε η Συνάρτηση Πυκνότητας Πιθανότητας (Probability Density Function –PDF). Τα πειραματικά αποτελέσματα απεικονίζονται στο σχήμα 5.5, όπου μπορούμε να παρατηρήσουμε ότι οι κατανομές ακολουθούν μια καμπανοειδή κατανομή. Αξιοποιώντας την μέθοδο των στιγμών (moments method) [82] υπολογίζουμε τη βέλτιστη κατανομή Γάμμα, που περιγράφει καλύτερα την πειραματική πυκνότητα πιθανότητας. Η Γάμμα κατανομή πιθανότητας δίνεται από την ακόλουθη σχέση

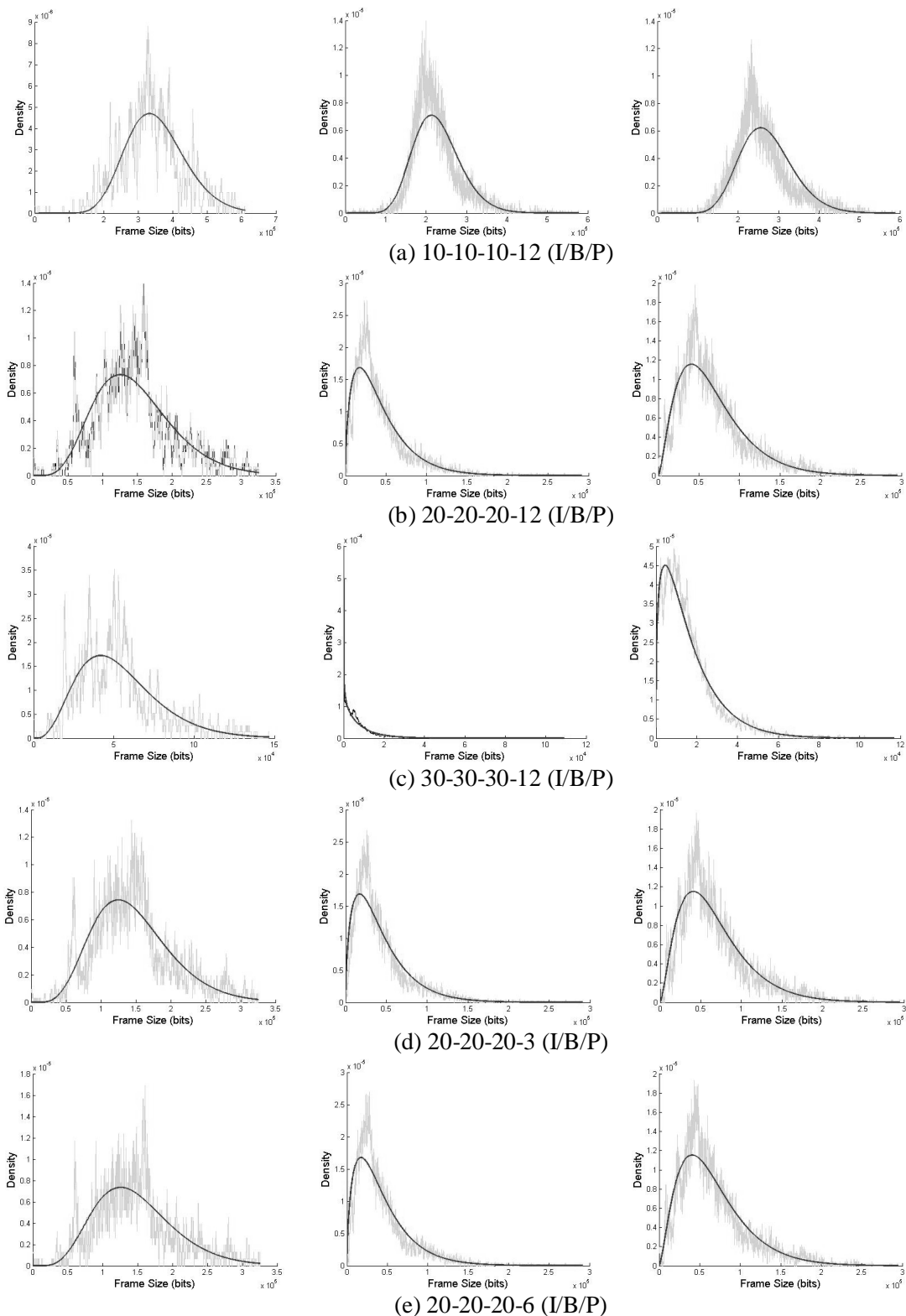
$$f_k = \frac{(x/m)^{p-1}}{m\Gamma(p)} e^{-x/m}, \mu, p > 0, x \geq 0, \Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt$$

Η Γάμμα κατανομή έχει μέση τιμή ίση με μ και απόκλιση μ^2 . Συνεπώς σύμφωνα με τη μέθοδο των στιγμών, εξισώνοντας τις αντίστοιχες πειραματικές τιμές της μέσης τιμής και της απόκλισης, τα οποία τα απεικονίζουμε ως m και v αντίστοιχα, με τις αντίστοιχες της Γάμμα κατανομής, μπορούμε να συμπεράνουμε ότι $\mu=v/m$ και $p=m^2/v$. Ο πίνακας 5.3 απεικονίζει τις αντίστοιχες παραμέτρους της κατανομής Γάμμα για κάθε εξεταζόμενη τιμή κβάντισης. Επιπλέον στις απεικονιζόμενες πειραματικές κατανομές του σχήματος 5.5, με συνεχή μαύρη γραμμή παρατίθεται η αντίστοιχη θεωρητική Γάμμα κατανομή, όπως προκύπτει από την πειραματική διαδικασία που περιγράφηκε.

Quantization Settings / Frame Types	I Frames		B Frames		P Frames	
	μ	p	μ	p	μ	p
(10,10,10)-12	21499	16.487	14608	15.584	15895	17.071
(20,20,20)-12	22905	6.468	25572	1.682	26287	2.549
(30,30,30)-12	12235	4.406	7857	0.813	11874	1.376
(20,20,20)-3	22316	6.597	25808	1.661	26446	2.542
(20,20,20)-6	22683	6.545	25720	1.676	26404	2.540

Πίνακας 5.3 Στατιστικά της Γάμμα μοντελοποίησης για ποικίλα επίπεδα κβάντισης

Στη συνέχεια, η συνάρτηση αυτοσυσχέτισης υπολογίζεται για κάθε τύπο καρέ. Αντιπροσωπευτικά παραθέτουμε την περίπτωση με επίπεδα κβάντισης 10-10-10 και περίοδο Intra ίση με 12.



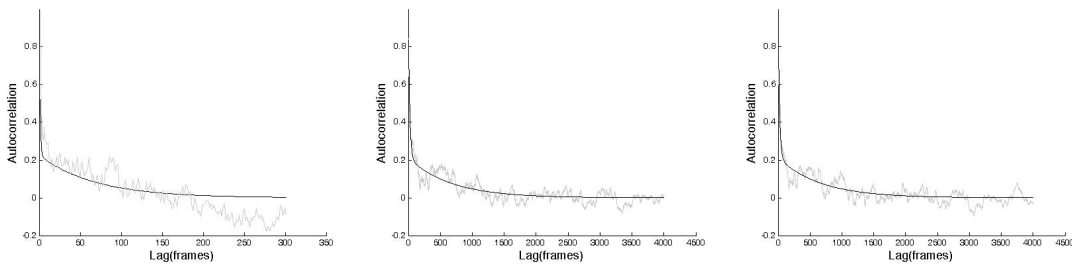
Σχήμα 5.5 Πειραματικά ιστογράμματα μεγέθους καρέ με τις αντίστοιχες Γάμμα κατανομές

Όπως μπορούμε να παρατηρήσουμε στο σχήμα 5.6, η συνάρτηση αυτοσυσχέτισης ακολουθεί μια φθίνουσα μορφή, όπως ακριβώς και σε προηγούμενα πρότυπα κωδικοποίησης, όπως το H.261 [82] και MPEG-1 [83]. Πιο συγκεκριμένα, εξαιτίας της

μορφής που παρουσιάζει η πειραματική συνάρτηση αυτοσυσχέτισης, μπορούμε να την αναλύσουμε σε δύο μέρη: Ένα αρχικό με μεγάλη κλίση και στη συνέχεια ένα σκέλος που φθίνει ασυμπτωτικά προς μια συγκεκριμένη τιμή. Σύμφωνα με την παρατήρηση αυτή, το φαινόμενο αυτό μπορεί επιτυχώς να συλληφθεί και να περιγραφεί από ένα άθροισμα με στατιστικά βάρη δύο γεωμετρικών όρων της ακόλουθης μορφής:

$$r_k = wI_1^k + (1-w)I_2^k \quad \text{με } |\lambda_2| < |\lambda_1| < 1$$

όπου $\lambda_{1,2}$ είναι οι ρυθμοί με την οποία φθίνει η συνάρτηση αυτοσυσχέτισης. Επιπλέον, η εξίσωση αυτή αντιστοιχεί στη γενική μορφή των μοντέλων auto-regressive δεύτερης τάξης AR(2) [84], [101], τα οποία έχουν εκτενώς εξεταστεί και έχει διαπιστωθεί ότι προσομοιάζουν επιτυχώς κάποιες ιδιότητες της πραγματικής κίνησης βίντεο.



Σχήμα 5.6 Αντιπροσωπευτικές Αυτοσυσχετίσεις

Autocorrelation Type /Quantization Scale	I	B	P
10-10-10-12	w = 0.2248 $\lambda_1=0.9855$ $\lambda_2=0.3884$	w = 0.2132 $\lambda_1=0.9985$ $\lambda_2=0.9526$	w = 0.2014 $\lambda_1=0.9985$ $\lambda_2=0.9467$
20-20-20-12	w = 0.2472 $\lambda_1=0.9828$ $\lambda_2=0.7216$	w = 0.1859 $\lambda_1=0.9987$ $\lambda_2=0.9658$	w = 0.1712 $\lambda_1=0.9987$ $\lambda_2=0.9594$
30-30-30-12	w = 0.2011 $\lambda_1=0.9819$ $\lambda_2=0.7417$	w = 0.2650 $\lambda_1=0.9981$ $\lambda_2=0.9457$	w = 0.1923 $\lambda_1=0.9984$ $\lambda_2=0.9465$
20-20-20-3	w = 0.1931 $\lambda_1=0.9973$ $\lambda_2=0.9093$	w = 0.1686 $\lambda_1=0.9980$ $\lambda_2=0.9683$	w = 0.1590 $\lambda_1=0.9982$ $\lambda_2=0.9482$
20-20-20-6	w = 0.2124 $\lambda_1=0.9942$ $\lambda_2=0.8377$	w = 0.1746 $\lambda_1=0.9988$ $\lambda_2=0.9672$	w = 0.1648 $\lambda_1=0.9986$ $\lambda_2=0.9566$

Πίνακας 5.4 Οι παράμετροι του μοντέλου της συνάρτησης αυτοσυσχέτισης.

Για να εξετάσουμε περαιτέρω την εγκυρότητα της εξίσωσης $r_k = wI_1^k + (1-w)I_2^k$, οι απαραίτητες παράμετροι υπολογίστηκαν μέσω της μεθόδου ελαχίστων τετραγώνων. Για την περίπτωση των I καρέ για τις πρώτες 300 ολισθήσεις, ενώ για τα B και P καρέ για τις πρώτες 4000 ολισθήσεις. Τα αποτελέσματα αυτής της θεωρητικής προσέγγισης

απεικονίζονται με συνεχόμενη μαύρη γραμμή στο προηγούμενο ενδεικτικό παράδειγμα (10-10-10)-12 του σχήματος 5.6. Όπως μπορεί εύκολα να παρατηρηθεί το προτεινόμενο θεωρητικό μοντέλο περιγράφει επιτυχώς τόσο τους βραχυπρόθεσμους, όσο και τους μακροπρόθεσμους όρους της πειραματικής συνάρτησης αυτοσυσχέτισης. Στον πίνακα 5.4 εμφανίζονται οι τιμές των παραμέτρων του προτεινόμενου μοντέλου για όλες τις εξεταζόμενες περιπτώσεις κωδικοποίησης.

Έχοντας ολοκληρώσει τη στατιστική ανάλυση σε επίπεδο I/P/B για ποικίλα σχήματα κωδικοποίησης σύμφωνα με το πρότυπο H.264 είναι πλέον εφικτό να μοντελοποιήσουμε τις στατιστικές ιδιότητες της εξεταζόμενης κίνησης βίντεο. Η αξιοποίηση των προτεινόμενων αυτών μοντέλων μπορεί να οδηγήσει σε μια επιτυχή γεννήτρια συνθετικής κίνησης, η οποία θα προσομοιάζει τις ιδιότητες και τα χαρακτηριστικά πραγματικής ροής H.264.

Η κωδικοποίηση μεταβλητού ρυθμού παράγει κίνηση που χαρακτηρίζεται από μεγάλες διακυμάνσεις γύρω από την επιλεγμένη μέση τιμή. Οι διακυμάνσεις αυτές σχετίζονται άμεσα με τη χωρική και χρονική δυναμική του περιεχομένου. Η πολύπλοκη αυτή φύση της κωδικοποίησης μεταβλητού ρυθμού δημιουργεί μια πρόκληση στον αποδοτικό σχεδιασμό των υπολογιστικών και τηλεπικοινωνιακών δικτύων μέσω των οποίων πραγματοποιείται η διακίνηση τέτοιων δεδομένων. Κατά συνέπεια κρίνεται αναγκαία η ακριβής μελέτη της επίδρασης που έχει η κίνηση βίντεο στην απόδοση του δικτύου. Μια μέθοδος για να γίνει η μελέτη αυτή είναι η πραγματοποίηση πειραμάτων με πραγματικές πηγές ροών βίντεο στα υπάρχοντα τηλεπικοινωνιακά ή υπολογιστικά δίκτυα. Όμως η πραγματοποίηση πειραμάτων σε δίκτυα με τη χρήση πραγματικής κίνησης, ουσιαστικά δεν διευκολύνει στη βέλτιστη προσαρμογή των δικτύων αυτών για αποδοτική μεταφορά κίνησης βίντεο, αφού παρέχει περιορισμένα αποτελέσματα. Αντιθέτως, η στατιστική ανάλυση πρωτογενώς των χαρακτηριστικών κίνησης βίντεο και δευτερευόντως η ανάπτυξη ενός μοντέλου παραγωγής συνθετικής κίνησης βίντεο με δυνατότητες μεταβολής των στατιστικών παραμέτρων της, θα προσφέρει πιο γενικευμένα και αξιόπιστα αποτελέσματα.

Σε αυτά τα πλαίσια, το επόμενο εδάφιο παραθέτει μια συνοπτική βιβλιογραφική ανασκόπηση των μοντέλων προσομοίωσης κίνησης βίντεο, ενώ παρουσιάζει, περιγράφει και αξιολογεί πρωτότυπη μέθοδο παραγωγής συνθετικής κίνησης H.264.

Μοντέλα Παραγωγής Συνθετικής Ροής H.264

Το παρόν εδάφιο παρουσιάζει μια συνοπτική βιβλιογραφική ανασκόπηση των μοντέλων παραγωγής συνθετικής ροής βίντεο. Επιπλέον παρουσιάζεται, περιγράφεται και αξιολογείται πρωτότυπη μέθοδος παραγωγής κίνησης H.264.

- Συνοπτική Βιβλιογραφική Ανασκόπηση

Κατά τη διάρκεια των τελευταίων ετών, η διερεύνηση μοντέλων παραγωγής συνθετικής ροής ψηφιακού βίντεο έχει αποτελέσει βασικό άξονα στην επιστημονική κοινότητα του ψηφιακού βίντεο. Στη διεθνή βιβλιογραφία έχουν προταθεί πληθώρα μεθόδων και τεχνικών, οι οποίες εν γένει μπορούν να ταξινομηθούν ως εξής [102]:

- Auto-Regressive Models

Τα διακριτά Auto-Regressive (AR) μοντέλα πρώτης τάξης – DAR(1) υπήρξαν τα πρωταρχικά μοντέλα που προτάθηκαν στην παραγωγή συνθετικής ροής βίντεο. Στην αναφορά [85], ένα μοντέλο DAR(1) αξιολογείται στην παραγωγή συνθετικής κίνησης διάρκειας 10 δευτερολέπτων στην ακόλουθη AR διεργασία:

$$x(n) = a_1 x(n-1) + e(n)$$

όπου $x(n)$ είναι το μέγεθος του n καρέ, $e(n)$ μια Gaussian διεργασία με απόκλιση σ^2 και a_1 είναι ο συντελεστής αυτοσυσχέτισης στην θέση ολίσθησης -1. Η ακολουθία βίντεο θεωρείται στατιστικώς στάσιμη. Επιπλέον, το μοντέλο αυτό μπορεί να θεωρηθεί ως μια διακριτού χρόνου διεργασία Markov, η οποία είναι μια ειδική περίπτωση της γενικής μορφής

$$x(n) = \sum_{i=1}^p a_i x(n-i) + e(n)$$

όπου p είναι ο βαθμός της τάξης της AR διεργασίας. Σε αυτή τη μαθηματική διατύπωση, οι διαφορές $e(n)$ θεωρούνται μη συσχετισμένες και κατανομημένες σύμφωνα με την κανονική κατανομή. Στην αναφορά [103], ο Xu *et al* προτείνουν μια Γάμμα κατανομή των διαφορών αντί για την κανονική κατανομή στην AR διεργασία, αναλύοντας ένα μοντέλο πρώτης τάξης Γάμμα AR (Gamma AR – GAR) της μορφής:

$$f_e(s) = a^3 + 3a^3(1-a) \frac{b}{b+2} + 3a(i-a)^2 \left(\frac{b}{b+s} \right)^2 + (1-a)^3 \left(\frac{b}{b+s} \right)^3$$

όπου $f_e(s)$ είναι ο μετασχηματισμός Laplace των διαφορών, a ο συντελεστής AR και β η παράμετρος κλίμακας. Αν και το GAR μοντέλο ξεπέρασε σε ορισμένες περιπτώσεις την απόδοση παλαιότερων DAR(1) μοντέλων, το γεγονός ότι η διαδικασία γένεσης των διαφορών είναι σημαντικά πολύπλοκη για AR μοντέλα μεγάλης τάξης, αποτρέπει στην πράξη η υλοποίηση και εφαρμογή τέτοιων μοντέλων.

Επιπλέον, ως επέκταση του προηγούμενου μοντέλου, το μοντέλο Gamma Beta AR (GBAR) έχει προταθεί στην βιβλιογραφία [104]. Στο μοντέλο αυτό προτείνεται ότι οι AR συντελεστές A_n ακολουθούν την Beta κατανομή και οι διαφορές B_n ακολουθούν την Gamma κατανομή. Πιο συγκεκριμένα, προτείνεται ότι:

$$X_n = A_n X_{n-1} + B_n$$

Αν και το μοντέλο GBAR είναι πιο αποδοτικό σε σύγκριση με το GAR μοντέλο, παρόλα αυτά δεν είναι κατάλληλο για τη μελέτη αλγορίθμων ελέγχου εισδοχής σε δίκτυα ATM.

Επίσης, ένα γενικευμένο AR μοντέλο περιγράφεται στην αναφορά [105], όπου αρχικά το μοντέλο αναλύει τη Gamma διεργασία σε μια ακολουθία όρων δευτέρου βαθμού. Στη συνέχεια, κάθε στοιχείο της ακολουθίας προσδιορίζεται από την ύψωση στο τετράγωνο μιας Gaussian διεργασίας, η οποία παράγεται από ένα AR μοντέλο με δεδομένο πίνακα συμμεταβλητότητας.

- Μαρκοβιανά Μοντέλα

Επιπλέον, στη διεθνή βιβλιογραφία έχουν προταθεί πληθώρα μοντέλων που βασίζονται στη χρήση αλυσίδων Markov. Τα μοντέλα αυτά βασίζονται στο γεγονός ότι μια παρούσα κατάσταση Markov εξαρτάται μονοσήμαντα από την αμέσως προηγούμενη κατάστασή της και από καμία άλλη προγενέστερή της. Η πρόταση αυτή μπορεί επιτυχώς να περιγραφεί από την ακόλουθη πρόταση: Μια στοχαστική διεργασία X_k με χώρο $S = \{1, 2, 3, \dots\}$ είναι μαρκοβιανή εάν για κάθε n και για όλες τις καταστάσεις i_1, i_2, \dots ικανοποιεί την ακόλουθη ιδιότητα:

$$P[X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_1 = i_1] = P[X_n = i_n | X_{n-1} = i_{n-1}]$$

Ένα μοντέλο παραγωγής συνθετικής ροής βίντεο έχει προταθεί στην αναφορά [106], το οποίο είναι ικανό να αναπαράγει πολλαπλά επίπεδα δυναμικής του περιεχομένου καθώς και τις αλλαγές σκηνών. Ο πίνακας μετάβασης καταστάσεων P της Μαρκοβιανής αλυσίδας δίνεται από τη σχέση

$$P = \rho I + (1 - \rho)Q$$

Όπου ρ είναι ο συντελεστής αυτοσυσχέτισης, I είναι ο μοναδιαίος πίνακας και κάθε γραμμή του πίνακα Q περιέχει τις αντίστοιχες πιθανότητες της αρνητικής διωνυμικής κατανομής [107].

- Μοντέλα Μακροπρόθεσμης Εξάρτησης

Στον ίδιο ερευνητικό χώρο, έχουν παρουσιαστεί και εργασίες, οι οποίες επιχειρούν να προσομοιάσουν πιο πιστά τις πολυπλοκότητες που παρουσιάζονται μακροπρόθεσμα στη φθίνουσα συνάρτηση αυτοσυσχέτισης. Αξιοποιώντας πιο ακριβείς μεθόδους όπως τη Autoregressive Moving-Average Process (ARMA), η οποία προσθέτει κάθε φορά ένα κινητό μέσο όρο, έχουμε το εξής μοντέλο:

$$x(n) = a_0 + \sum_{i=1}^p a_i x(n-i) + \sum_{j=0}^p b_j e(n)$$

όπου p είναι η τάξη της AR διεργασίας και q είναι η τάξη του κινητού μέσου. Μοντέλα αυτής της κατηγορίας προτείνονται και περιγράφονται αναλυτικά στην αναφορά [108].

- Υβριδικά Μοντέλα Markov και AR

Επίσης έχουν προταθεί υβριδικά μοντέλα, τα οποία συνδυάζουν Gaussian AR(1) μοντέλα με παραμέτρους που διαμορφώνονται σύμφωνα με τις μεταβάσεις μιας μαρκοβιανής αλυσίδας. Ενδεικτικά παραθέτουμε τις ακόλουθες βιβλιογραφικές αναφορές [79], [109], [110].

Σε αντίθεση με τα προαναφερθέντα μοντέλα και μεθόδους, το παρόν κεφάλαιο εισάγει πρωτότυπο μοντέλο που συνδυάζει τρεις διακριτές DAR(1) διεργασίες για την παραγωγή συνθετικών ροών I, B και P σε συνδυασμό με ένα μαρκοβιανό μηχανισμό απόφασης για την εξομοίωση των αλλαγών σκηνών. Στο επόμενο εδάφιο, παρουσιάζουμε το προτεινόμενο μοντέλο, το οποίο ονομάζεται Τροποποιημένο Μαρκοβιανό Μοντέλο.

Προτεινόμενο Τροποποιημένο Μαρκοβιανό Μοντέλο

Στα προηγούμενα εδάφια παρουσιάστηκαν οι στατιστικές ιδιότητες των I, P και B καρέ. Αυτή η αναλυτική και ποσοτική διερεύνηση της κίνησης βίντεο κωδικοποιημένου σύμφωνα με το πρότυπο H.264 είναι απαραίτητη τόσο στην κατανόηση των ιδιοτήτων του, όσο και στην παραγωγή συνθετικής κίνησης. Το προτεινόμενο μοντέλο για να είναι αποδοτικό και ακριβές πρέπει να τηρεί ορισμένα κριτήρια [102]:

- Να ικανοποιεί τις στατιστικές ιδιότητες της πραγματικής κίνησης βίντεο, όπως η συνάρτηση πυκνότητας πιθανότητας και αυτοσυσχέτισης.
- Τα χαρακτηριστικά του συνθετικού βίντεο πρέπει να είναι όμοια με αυτά του πραγματικού βίντεο ώστε να είναι δυνατή η χρησιμοποίηση της συνθετικής κίνησης αντί της πραγματικής, στη μελέτη ενός δικτύου. Αυτό σημαίνει ότι το προτεινόμενο μοντέλο πρέπει να προσομοιάζει επιτυχώς τις εξαρτήσεις μακράς και βραχείας κλίμακας του πραγματικού βίντεο.
- Η πολυπλοκότητα του προτεινόμενου μοντέλου να διατηρηθεί σε όσο το δυνατόν χαμηλότερα επίπεδα.
- Να παράγει με αποδοτικό τρόπο συνθετική κίνηση βίντεο, προσομοιάζοντας επιτυχώς περιεχόμενο τόσο στατικό όσο και έντονα δυναμικό.

Σύμφωνα με τα προηγούμενα κριτήρια και βασισμένοι στην στατιστική και ποσοτική ανάλυση των προηγούμενων εδαφίων, είναι δυνατή η παραγωγή συνθετικής ροής βίντεο, με την αξιοποίηση DAR(1) μοντέλων.

Οι μαρκοβιανές ιδιότητες συνοψίζονται στο γεγονός ότι η κατανομή του μεγέθους του καρέ F_n εξαρτάται μονοσήμαντα από το μέγεθος του προγενέστερου καρέ F_{n-1} . Τέτοια διεργασία καθορίζεται από τη στατική οριακή κατανομή του F_n και άλλων παραμέτρων, οι οποίες – ανεξάρτητα από την οριακή κατανομή- καθορίζουν τις συσχετίσεις της παραγόμενης ροής.

Σύμφωνα με το DAR(1) μοντέλο, η πρώτη τάξης φόρμα δίνεται από τη σχέση που ακολουθεί

$$F_n = V_n F_{n-1} + (1 - V_n) Y_n \text{ για } n=1,2,\dots$$

Όπου V_n είναι ανεξάρτητα κατανομημένες δυαδικές τυχαίες μεταβλητές με $P(V_n=1) = 1 - P(V_n=0) = a$ με $0 \leq a < 1$ και Y_n είναι ανεξάρτητα κατανομημένες τυχαίες μεταβλητές με οριακή κατανομή π.

Δηλαδή, το μοντέλο ορίζει ότι η παρούσα κατάσταση είναι ένα μίγμα δύο ανεξάρτητων τυχαίων μεταβλητών: Καθορίζεται είτε από την τελευταία παρατήρηση με πιθανότητα a , ή από ένα άλλο ανεξάρτητο δείγμα από την ίδια κατανομή. Συνεπώς, πρόκειται για ένα απλό και γενικό μοντέλο αφού το π είναι η κατανομή κάθε τυχαίας μεταβλητής και η δομή της συσχέτισης είναι ανεξάρτητη του π. Πιο συγκεκριμένα, η συνάρτηση αυτοσυσχέτισης του F_n δίνεται από τη σχέση $r_F(k) = a^k, k = 0, 1, \dots$ και F_n είναι μια μαρκοβιανή αλυσίδα με πίνακα πιθανοτήτων μετάβασης να ορίζεται από τη σχέση $aI + (1-a)Q$, όπου I είναι ο μοναδιαίος πίνακας και Q είναι ένας τετραγωνικός πίνακας, του οποίου οι σειρές παίρνουν τιμές από την κατανομή π.

Η προσέγγιση αυτή στο προτεινόμενο μοντέλο χρησιμοποιείται ανεξάρτητα για την παραγωγή της κίνησης των ροών I, B, P προσπαθώντας να συλλάβει επιτυχώς το γεγονός ότι ένα βίντεο αποτελείται από σκηνές διαφορετικής χωρικής και χρονικής δραστηριότητας. Συνεπώς ένα υποψήφιο μοντέλο παραγωγής συνθετικής κίνησης βίντεο θα πρέπει να είναι ικανό να προσομοιάσει επιτυχώς τις προαναφερθείσες ένδο-πλαισιακές και διά-πλαισιακές ιδιότητες των σκηνών. Πιο συγκεκριμένα, κατά τη διάρκεια μιας σκηνής, τα μεγέθη του καρτέ ίδιου τύπου παραμένουν πρακτικώς σταθερά, ενώ κατά την αλλαγή μιας σκηνής παρατηρείται σημαντική μεταβολή.

Συνεπώς, στο προτεινόμενο μοντέλο θεωρούνται δύο διακριτές διεργασίες για την παραγωγή καρτέ:

- Η πρώτη διεργασία προσομοιάζει την ένδο-πλαισιακή κατάσταση, που σημαίνει ότι το μέγεθος του τρέχοντος καρτέ διατηρεί τα χαρακτηριστικά του προγενέστερου του.
- Η δεύτερη διεργασία προσομοιάζει τη δια-πλαισιακή κατάσταση, όπου το μέγεθος καρτέ παράγεται από μια AR(1) διεργασία της μορφής $x(n)=\alpha_1x(n-1)+e(n)$ βασισμένοι στο μέγεθος του προηγούμενου καρτέ, με α_1 να συμβολίζει τον συντελεστή αυτοσυσχέτισης με ολίσθηση -1 και $e(n)$ ένα υπόλοιπο, που ακολουθεί την κανονική κατανομή.

Ειδικά για την περίπτωση των I καρτέ, τα οποία ουσιαστικά καθορίζουν σε σημαντικό βαθμό τη συσχέτιση των καρτέ ανάμεσα στα διαδοχικά GOP μιας ροής βίντεο, όπως απεικονίζεται αυτή στη συνάρτηση αυτοσυσχέτισης των I καρτέ, η προαναφερθείσα πρώτη διεργασία εξειδικεύεται ως εξής: $I(n)=\alpha_1I(n-1)$ όπου α_1 είναι ο συντελεστής αυτοσυσχέτισης για ολίσθηση -1.

Παρόλο που η συσχέτιση ανάμεσα στα GOP (*inter-GOP correlation*), που περιγράφεται ικανοποιητικά από τη συνάρτηση αυτοσυσχέτισης των I καρτέ, είναι ένα σημαντικό μέγεθος για την επιτυχή παραγωγή συνθετικής κίνησης βίντεο, μια εξίσου σημαντική παράμετρος είναι η συσχέτιση ανάμεσα στα I/P/B εντός του ίδιου GOP (*intra-GOP correlation*). Εάν θεωρήσουμε μη συσχετισμένη παραγωγή ροών I, P και B, τότε η συνολική παραγόμενη κίνηση βίντεο δε θα αναπαριστά επιτυχώς τις στατιστικές ιδιότητες της πραγματικής κίνησης, με αποτέλεσμα η συνθετική αυτή κίνηση να μη μπορεί να αξιοποιηθεί στη βελτιστοποίηση των παραμέτρων ενός δικτύου.

Για το λόγο αυτό, στο προτεινόμενο μοντέλο οι συντελεστές συσχέτισης υπολογίζονται ανάμεσα στους πρώτους γείτονες I-P, I-B και P-B κάθε δομής GOP. Ο πίνακας 5.5 αναπαριστά τους συντελεστές συσχέτισης για ποικίλα σχήματα κωδικοποίησης, δείχνοντας ότι υπάρχει μεγάλη συσχέτιση ανάμεσα στους πρώτους γείτονες I-P κάθε GOP, μικρότερη ανάμεσα στα P-B και ακόμη μικρότερη για την περίπτωση των I-B καρτέ [107].

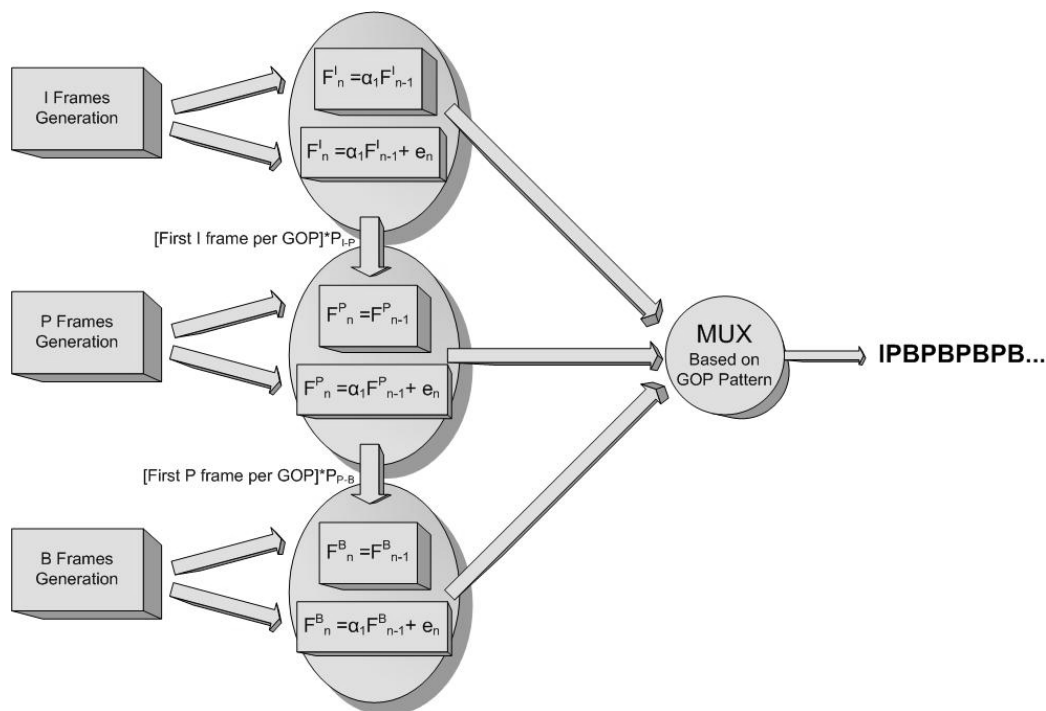
Quantization Scale	I-P Correlation Coefficient	I-B Correlation Coefficient	P-B Correlation Coefficient
(10-10-10)-12	0.6412	0.1750	0.1721
(20-20-20)-12	0.6250	0.0653	0.0948
(30-30-30)-12	0.5180	-0.0381	0.0916

Πίνακας 5.5 Συντελεστές συσχέτισης διαφορετικών τύπων καρτέ.

Τα αποτελέσματα του πίνακα 5.5 είναι σύμφωνα με τα αντίστοιχα στατιστικά που έχουν προκύψει από την εργασία του M. Dai [111], επικυρώνοντας την εγκυρότητάς τους.

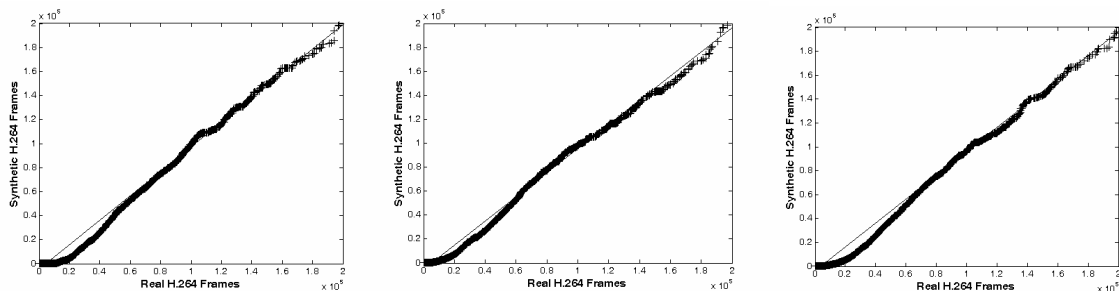
Σύμφωνα με τα προαναφερθέντα, στο σχήμα 5.7 απεικονίζεται σχηματικά ο αλγόριθμος του προτεινόμενου μοντέλου παραγωγής συνθετικής ροής H.264. Όπως παρατηρείται, εκτός από τη χρήση τριών διακριτών DAR(1) μοντέλων για την παραγωγή των ροών I/P/B, το προτεινόμενο μοντέλο αξιοποιεί και τους συντελεστές συσχέτισης μεταξύ I-P και P-B. Στη διαδικασία παραγωγής συνθετικής κίνησης, ο συντελεστής συσχέτισης μεταξύ των καρέ I-B δεν λαμβάνεται υπόψιν, λόγω της χαμηλής τιμής του και για αυτό στο προτεινόμενο μοντέλο η παραγωγή I και B καρέ θεωρείται ασυσχέτιστη.

Πιο συγκεκριμένα, βασισμένοι στο μήκος και στη δομή του GOP της παραγόμενης συνθετικής κίνησης, ο προτεινόμενος αλγόριθμος κατά τη διάρκεια αρχικοποίησης ενός συνθετικού GOP πραγματοποιεί την εξής διαδικασία: Αρχικά διαδίδεται το μέγεθος ενός I και P καρέ, τα οποία αφού πολλαπλασιαστούν με τους αντίστοιχους συντελεστές συσχέτισης των I-P και P-B καρέ, χρησιμοποιούνται ως είσοδοι στις αντίστοιχες P και B διεργασίες. Με τον τρόπο αυτό, επιτυγχάνεται η προσομοίωση της *intra*-GOP εξαρτήσεως της πραγματικής κίνησης, το οποίο είναι ένα από τα σημαντικότερα ζητήματα στην μοντελοποίηση κίνησης βίντεο.



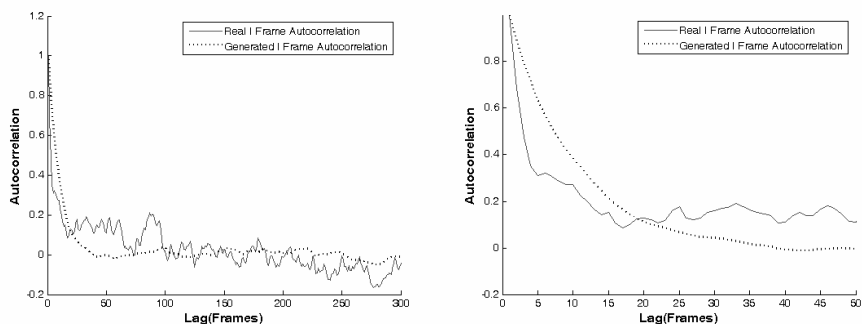
Σχήμα 5.7 Block διάγραμμα του προτεινόμενου μοντέλου

Στο προτεινόμενο μοντέλο, η συνιστώσα MUX στο block διάγραμμα είναι υπεύθυνη για τη σωστή πολυπλεξία των ροών I, B και P σύμφωνα με τη επιθυμητή δομή GOP.



Σχήμα 5.8 Q-Q Διαγράμματα δειγμάτων διάρκειας 3 λεπτών

Για την περίπτωση του σχήματος κωδικοποίησης (20-20-20)-12, παρήχθησαν από το προτεινόμενο μοντέλο ποικίλα συνθετικά δείγματα διάρκειας 3 λεπτών και συγκρίθηκαν με αντίστοιχα δείγματα πραγματικής κίνησης. Από τα αντίστοιχα Q-Q διαγράμματα του σχήματος 5.8, μπορούμε να παρατηρήσουμε την καλή συμπεριφορά του προτεινόμενου μοντέλου και την ορθότητα της παραγόμενης συνθετικής κίνησης.



Σχήμα 5.9 Αυτοσυσχέτιση πραγματικών και συνθετικών I καρέ μακράς/βραχείας εξάρτησης

Επίσης, για το προτεινόμενο μοντέλο ελέγχθηκε και η ακρίβεια των μακρών και βραχέων εξαρτήσεων αυτοσυσχέτισης. Στο σχήμα 5.9 απεικονίζονται οι αυτοσυσχετίσεις των πραγματικών και συνθετικών I καρέ. Όπως μπορεί να παρατηρηθεί, το προτεινόμενο μοντέλο προσομοιάζει ικανοποιητικώς τη μακρά και βραχεία δια-GOP συσχέτιση της πραγματικής H.264 κίνησης.

Συμπεράσματα του Κεφαλαίου

Στο παρόν κεφάλαιο διερευνήθηκε και παρουσιάστηκε ότι το πιο πρόσφατο πρότυπο κωδικοποίησης βίντεο H.264 μπορεί επιτυχώς να μοντελοποιηθεί από την προσαρμογή ήδη γνωστών τεχνικών παραγωγής συνθετικών ροών βίντεο. Το συμπέρασμα αυτό έγκειται στο γεγονός ότι το νέο πρότυπο διατηρεί τις βασικές στατιστικές ιδιότητες των προηγούμενων προτύπων. Πιο συγκεκριμένα, η κίνηση βίντεο H.264 παραμένει μια υπέρθεση τριών διακριτών κατανομών, οι οποίες προέρχονται από τρεις διαφορετικές κατηγορίες καρέ (I/B/P), όπως ακριβώς και στα προηγούμενα πρότυπα. Κατά συνέπεια, προσεγγίζοντας την παραγωγή συνθετικής κίνησης H.264 με βάση το τριπλό αυτό μοντέλο είναι δυνατή η χρήση υπάρχοντων τεχνικών και μεθόδων, όπως το DAR(1) και τα μαρκοβιανά μοντέλα. Όλα τα νέα χαρακτηριστικά του προτύπου H.264, όπως η χρήση μεταβλητού μεγέθους blocks, μεγαλύτερη ακρίβεια στα ανύσματα κίνησης, εκτεταμένο εύρος επιλογής των καρέ αναφοράς, de-blocking φίλτρο και νέες τεχνικές εντροπικής κωδικοποίησης, πετυχαίνουν καλύτερη απόδοση συμπίεσης σε σύγκριση με τα προγενέστερα πρότυπα, χωρίς να τροποποιούνται ουσιαστικά οι στατιστικές ιδιότητες της παραγόμενης κίνησης βίντεο σε σύγκριση με προγενέστερα πρότυπα.

ΚΕΦΑΛΑΙΟ 6

Σύγκριση της Αντιληπτής Ποιότητας Βίντεο Σταθερού και Μεταβλητού Ρυθμού

Μέθοδοι Κωδικοποίησης Σταθερού και Μεταβλητού Ρυθμού (CBR-VBR)

Από τις απαρχές των τεχνικών συμπίεσης ψηφιακού βίντεο, δύο βασικοί τύποι κωδικοποίησης έχουν προταθεί και χρησιμοποιούνται ευρέως έως και σήμερα: Η κωδικοποίηση σταθερού ρυθμού (Constant Bit Rate – CBR) και μεταβλητού ρυθμού (Variable Bit Rate – VBR). Η κωδικοποίηση μεταβλητού ρυθμού διατηρεί τις ίδιες παραμέτρους κβάντισης σε όλη τη διάρκεια κωδικοποίησης, σε αντίθεση με την κωδικοποίηση σταθερού ρυθμού, η οποία μεταβάλλει δυναμικά τους συντελεστές κβάντισης. Η δυναμική αυτή μεταβολή πραγματοποιείται σύμφωνα με κάποιο αλγόριθμο ελέγχου του ρυθμού κωδικοποίησης, ο οποίος ως επί το πλείστον μεταβάλλει τις παραμέτρους κβάντισης αντιστρόφως ανάλογα με τη πολυπλοκότητα του περιεχομένου, ώστε να διατηρήσει σταθερό τον τελικό ρυθμό της παραγόμενης ροής.

Σήμερα, η επιλογή του μεταβλητού ρυθμού κωδικοποίησης προτιμάται στις τηλεπικοινωνιακές εφαρμογές έναντι του σταθερού ρυθμού κωδικοποίησης, εξαιτίας κάποιων πλεονεκτημάτων όπως [78]:

- § Καλύτερη ποιότητα βίντεο για τον ίδιο μέσο ρυθμό κωδικοποίησης, χωρίς να απαιτείται η πολυπλοκότητα ενός αλγορίθμου μεταβολής του ρυθμού.
- § Μικρότερη καθυστέρηση στη μετάδοση του βίντεο, εξαιτίας του γεγονότος ότι το μέγεθος της προσωρινής μνήμης στον κωδικοποιητή μπορεί να μειωθεί σημαντικά, το οποίο προκαλεί αντίστοιχα και μικρότερη καθυστέρηση προώθησης της ροής στο δίκτυο.
- § Αυξημένη χωρητικότητα αποδοχής κλήσεων, αφού το εύρος ζώνης που καταναλώνεται ανά κλήση για υπηρεσία βίντεο μεταβλητού ρυθμού είναι χαμηλότερη σε σχέση με την ισοδύναμη ποιότητα υπηρεσίας που αντιστοιχεί σε βίντεο σταθερού ρυθμού.
- § Καλύτερη αξιοποίηση της διαθέσιμης χωρητικότητας του καναλιού μετάδοσης με τη χρήση στατιστικής πολυπλεξίας ροών μεταβλητού ρυθμού σε σύγκριση με τη χρήση ισάριθμων και ισοδύναμων ροών σταθερού ρυθμού.

Βάση των προαναφερθέντων πλεονεκτημάτων, είναι σήμερα γενικά αποδεκτό ότι η χρήση βίντεο μεταβλητού ρυθμού παρέχει αποδοτικότερη αξιοποίηση των διαθέσιμων δικτυακών πόρων.

Επιπλέον, εκτός από τα δικτυακά πλεονεκτήματα που έχει η χρήση κωδικοποίησης μεταβλητού ρυθμού, το γεγονός ότι η κωδικοποίηση/συμπίεση ενός βίντεο προκαλεί ποιοτική υποβάθμιση αυτού, προκαλεί περαιτέρω ενδιαφέρον για διερεύνηση της συσχέτισης των παραμέτρων κωδικοποίησης και της αντιληπτής ποιότητας που λαμβάνει ο τελικός χρήστης [112], αντικείμενο που αποτελεί και τον κεντρικό ερευνητικό άξονα της παρούσας διδακτορικής διατριβής.

Μεταξύ των ποικίλων παραμέτρων κωδικοποίησης (όπως ο ρυθμός κωδικοποίησης, η χωρική και χρονική ανάλυση) που επηρεάζουν την τελική αντιληπτή ποιότητα υπηρεσίας που λαμβάνει ο τελικός χρήστης, η επιλογή σταθερού ή μεταβλητού ρυθμού κωδικοποίησης

θεωρείται η πλέον σημαντικότερη παράμετρος για το τελικό αποτέλεσμα. Όπως περιγράφηκε αναλυτικά στο τέταρτο κεφάλαιο, πληθώρα υποκειμενικών και αντικειμενικών μεθόδων έχουν αναπτυχθεί για να ποσοτικοποιήσουν την αντιληπτή ποιότητα βίντεο που έχει υποστεί διαδικασία κωδικοποίησης και συμπίεσης [13], [15], [32], [37], [44], [45], [46], [113], [114], [115], [116], [117]. Παρόλο που έχουν προταθεί πολυάριθμες ερευνητικές μέθοδοι προς αυτή την κατεύθυνση, η ποσοτική διερεύνηση της αντιληπτής υπεροχής της ποιότητας που παρέχει η μέθοδος μεταβλητού ρυθμού κωδικοποίησης σε σύγκριση με τη μέθοδο σταθερού ρυθμού κωδικοποίησης δεν έχει πραγματοποιηθεί. Σήμερα, οι διαχειριστές τηλεπικοινωνιακών συστημάτων επιλέγουν εν γένει την κωδικοποίηση μεταβλητού ρυθμού χωρίς να γνωρίζουν την ουσιαστική αντιληπτή υπεροχή στην ποιότητα σε σύγκριση με το μηχανισμό του σταθερού ρυθμού.

Το παρόν κεφάλαιο πραγματεύεται με αναλυτικό και ποσοτικό τρόπο την υπεροχή της αντιληπτής ποιότητας που παρέχει η κωδικοποίηση μεταβλητού ρυθμού έναντι του σταθερού ρυθμού. Πιο συγκεκριμένα, θεωρώντας σταθερές παραμέτρους κωδικοποίησης (δηλ. χωρική και χρονική ανάλυση, προφίλ κωδικοποίησης, GOP δομή κλπ), παρουσιάζεται με αναλυτικό τρόπο η ποσοτική υπεροχή του μεταβλητού ρυθμού κωδικοποίησης έναντι του σταθερού για ποικίλες τιμές ρυθμού κωδικοποίησης (bit rate). Για λόγους συντόμευσης και ευκολότερης κατανόησης του κειμένου από τον αναγνώστη, για το πέρας του παρόντος κεφαλαίου θα χρησιμοποιηθούν οι αγγλικές συντομεύσεις CBR και VBR για την κωδικοποίηση σταθερού και μεταβλητού ρυθμού αντίστοιχα.

Μεθοδολογία Σύγκρισης Αντιληπτής Ποιότητας Κωδικοποίησης VBR και CBR

Για την ποσοτικοποίηση της αντιληπτής διαφοράς ανάμεσα στις μεθόδους κωδικοποίησης CBR και VBR, προτιμήθηκε η χρήση αντικειμενικών μεθόδων εκτίμησης της ποιότητας εικόνας βίντεο έναντι των αντιστοιχών υποκειμενικών. Ο λόγος για αυτή την επιλογή ήταν ότι οι διαφορές ανάμεσα στις δύο μεθόδους κωδικοποίησης αναμένεται να είναι μικρές, χωρίς την εμφάνιση τεχνουργημάτων, οπότε μια υποκειμενική μέθοδος δεν θα ήταν ικανή να παρέχει αξιόπιστα αποτελέσματα, δεδομένου ότι το στατιστικό λάθος που θα εισέλθει στις μετρήσεις από τη στατιστική επεξεργασία των αποτελεσμάτων δεν θα επιτρέψει την εξαγωγή κάποιου συγκεκριμένου συμπεράσματος, αφού θα υπερκαλύψει τα αποτελέσματα της εκτιμητικής μεθόδου. Κατόπιν αυτού του περιορισμού, αποφασίστηκε η χρήση τεσσάρων αντικειμενικών μεθόδων πλήρους αναφοράς:

- § PSNR
- § SSIM Metric [45], [46]
- § DVQ Metric [13]
- § Delta Metric [118]

Η χρήση των τεσσάρων αντικειμενικών μετρικών πλήρους αναφοράς παρέχει ακριβή αποτελέσματα για την αντιληπτή υπεροχή των υπό-εξέταση μεθόδων κωδικοποίησης. Επιπλέον περιορίζει στο μέγιστο δυνατό την εισαγωγή σφαλμάτων από την εκτιμητική διαδικασία, γεγονός που περιορίζει την περαιτέρω διάδοση αυτών μέσω της επακόλουθης επεξεργασίας των συλλεγόμενων πειραματικών μετρήσεων.

Στα πλαίσια της πληρότητας του παρόντος κεφαλαίου, παραθέτουμε μια συνοπτική περιγραφή για τα μετρικά που θα χρησιμοποιηθούν στην πειραματική διαδικασία.

- PSNR

Το PSNR είναι ένα ευρέως διαδεδομένο μετρικό που χρησιμοποιείται για τη σύγκριση της απόδοσης των ποικίλων τεχνικών κωδικοποίησης βίντεο. Παρόλο που έχει ευρέως επιδειχθεί η αναποτελεσματικότητα του συγκεκριμένου μετρικού για την επιτυχή και ακριβή εκτίμηση της αντιληπτής ποιότητας, παρόλα αυτά χρησιμοποιείται ακόμη και σήμερα στη βιβλιογραφία για να είναι δυνατή η άμεση σύγκριση με προγενέστερες δημοσιεύσεις και επιστημονικές εργασίες. Ο ορισμός του PSNR είναι ο ακόλουθος:

$$PSNR = 10 \log_{10} \frac{L^2}{MSE}$$

Όπου το L είναι το δυναμικό εύρος των pixels που απαρτίζουν τα καρέ του βίντεο και MSE το μέσο τετραγωνικό σφάλμα που δίνεται από την ακόλουθη σχέση:

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2$$

όπου το N είναι ο αριθμός των pixels ανά καρέ για δεδομένη ανάλυση εικόνας και x_i / y_i η τιμή του pixel i στο ασυμπιεστο και συμπιεσμένο βίντεο αντίστοιχα.

- SSIM

Το SSIM είναι μετρικό πλήρους αναφοράς, το οποίο ποσοτικοποιεί τη δομική ομοιότητα ανάμεσα σε δύο καρέ. Το μετρικό αυτό είναι βασισμένο στη βασική αρχή ότι το ανθρώπινο οπτικό σύστημα εξάγει τη δομική πληροφορία από ένα οπτικό ερέθισμα. Εάν θεωρήσουμε ότι x και y είναι δύο καρέ του ασυμπιεστού και συμπιεσμένου βίντεο αντίστοιχα, τότε το SSIM ορίζεται ως εξής:

$$SSIM(x, y) = \frac{(2m_x m_y + C_1)(2s_{xy} + C_2)}{(m_x^2 + m_y^2 + C_1)(s_x^2 + s_y^2 + C_2)}$$

όπου μ_x, μ_y είναι η μέση τιμή του x και y , $\sigma_x, \sigma_y, \sigma_{xy}$ είναι η διασπορά του x, y και συνδιασπορά του x και y , αντίστοιχα. Οι σταθερές C_1 και C_2 ορίζονται ως εξής:

$$C_1 = (K_1 L)^2 \quad C_2 = (K_2 L)^2$$

όπου L είναι δυναμικό εύρος των τιμών ενός pixel και $K_1 = 0.01, K_2 = 0.03$. [45], [46]

- Delta

Πρόκειται για μετρικό αντικειμενικής εκτίμησης πλήρους αναφοράς που αξιοποιεί τα δεδομένα της συνιστώσας της φωτεινότητας. Στο μετρικό Delta υπολογίζονται οι αποστάσεις μεταξύ των καρέ του ασυμπιεστού και συμπιεσμένου βίντεο, δηλαδή η μαθηματική απόσταση των αριθμητικών τιμών της φωτεινότητας στα υπό εξέταση καρέ. Θεωρώντας ότι f και g είναι δύο καρέ που ορίζονται στο εύρος τιμών ενός pixel X , τα οποία

χρησιμοποιούν τον ίδιο αριθμό φωτεινών επιπέδων, με Γ_f και Γ_g να είναι οι υπό-γράφοι των καρτέ f και g αντίστοιχα, τότε το μετρικό Delta Δ_g ορίζεται ως εξής:

$$\Delta_g(\Gamma_f, \Gamma_g) = \left[\frac{1}{n(X)n(Y)} \sum_{x \in X} \sum_{y \in Y} |d((x, y), \Gamma_f) - d((x, y), \Gamma_g)|^p \right]^{1/p} \text{ for } 1 \leq p \leq \infty$$

- DVQ

Το μετρικό DVQ είναι βασισμένο στο διακριτό μετασχηματισμό συνημίτονου και αξιοποιεί πτυχές του ανθρώπινου συστήματος όρασης, όπως προσαρμογή φωτεινότητας, αποσύνθεση σε κανάλια φωτός και χρώματος, χωρικό και χρονικό φιλτράρισμα, κανάλια ανάλυσης χωρικών συχνοτήτων, μάσκα αντίθεσης και στατιστική συνεκτίμηση όλων των προαναφερθέντων εκτιμήσεων [13]. Πιο συγκεκριμένα, το μετρικό DVQ χρησιμοποιεί το διακριτό μετασχηματισμό συνημίτονου για να πραγματοποιήσει την αποσύνθεση του αρχικού σήματος σε χωρικά κανάλια. Το γεγονός αυτό πλεονεκτεί στην υλοποίηση του συγκεκριμένου μετρικού, δεδομένου ότι υπάρχει πληθώρα αποδοτικών λογισμικών και υλικών που πραγματοποιούν το μετασχηματισμό αυτό. Επιπλέον, στις περιπτώσεις κωδικοποίησης βίντεο με τη χρήση απωλεστικού προτύπου συμπίεσης (MPEG, ITU), ο μετασχηματισμός έχει ήδη πραγματοποιηθεί ως μέρος της διαδικασίας συμπίεσης και κωδικοποίησης του βίντεο.

Η χρήση των τεσσάρων μετρικών πλήρους αναφοράς παρέχει υψηλή ακρίβεια στα εξαγόμενα αποτελέσματα της πειραματικής σύγκρισης της αντιληπτής ποιότητας μεταξύ κωδικοποίησης σταθερού και μεταβλητού ρυθμού. Εξαιτίας της έντονης ετερογενούς φύσης της μεθόδου που στηρίζεται η υλοποίηση του κάθε μετρικού, πιθανές αδυναμίες στην εκτίμηση της αντιληπτής ποιότητας βίντεο με τη χρήση του ενός μετρικού, αντισταθμίζεται από τη χρήση κάποιου άλλου εκ των υπολοίπων τριών. Οπότε με την προτεινόμενη μέθοδο θα οδηγηθούμε στην εξαγωγή ενός γενικευμένου και αξιόπιστου αποτελέσματος.

Πειραματική Διαδικασία Λήψης Μετρήσεων Αντιληπτής Ποιότητας CBR-VBR

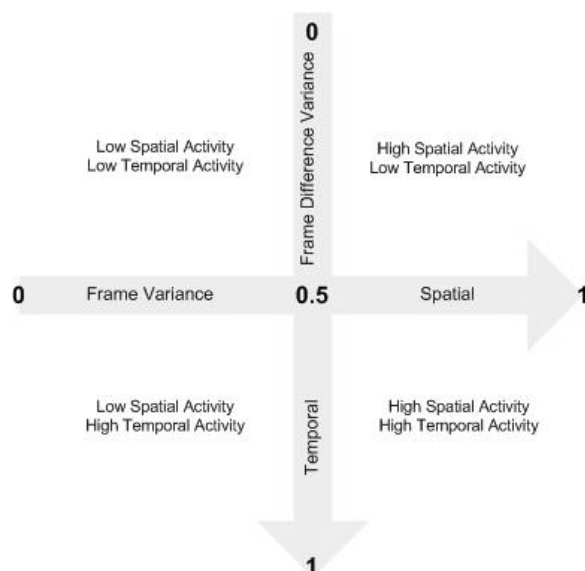
Για το πειραματικό μέρος του κεφαλαίου χρησιμοποιήθηκε η υλοποίηση του προτύπου MPEG-4 που έχει αναπτυχθεί από τη DICAS, η οποία είναι επίσημο μέλος του MPEG Industry Forum και υποστηρικτής του AVC Alliance. Η συμμετοχή της DICAS σε αυτές τις ομάδες εργασίας υποδηλώνει στενή σχέση ανάμεσα στην υλοποίηση του κωδικοποιητή και στην περιγραφή του προτύπου.

Σχετικά με το προφίλ της κωδικοποίησης, επιλέχθηκε το Advanced Simple Profile (ASP), επειδή είναι το πιο διαδεδομένο και ευρέως εφαρμόσιμο προφίλ στις εμπορικές υλοποιήσεις του προτύπου MPEG-4, όπως ψηφιακές κάμερες, κινητές τηλεφωνικές συσκευές κλπ, σε αντίθεση με τα υπόλοιπα προφίλ που προορίζονται για εξειδικευμένες υλοποιήσεις.

Επιπλέον, τα προφίλ Simple, Core και Main είναι υποσύνολα του ASP, που σημαίνει ότι οποιαδήποτε ροή βίντεο συμβατή με τα προφίλ αυτά είναι επίσης συμβατή και με το Advanced Simple Profile. Δηλαδή, η επιλογή του ASP είναι μια ευρύτερη περίπτωση των λοιπών προαναφερθέντων προφίλ.

Επίσης, οι προδιαγραφές του Advanced Simple Profile περιλαμβάνουν εκτός από τις πλέον δημοφιλείς χωρικές αναλύσεις (CIF και QCIF), τα κυριότερα χαρακτηριστικά κωδικοποίησης φυσικού βίντεο σύμφωνα με το πρότυπο MPEG-4, όπως: Τα βασικά οπτικά εργαλεία, ανοχή σε λάθη, μικρή επικεφαλίδα, B-VOP, μέθοδος κβάντισης $\frac{1}{2}$ και εκτίμηση κίνησης με ακρίβεια $\frac{1}{2}$ του pixel. Συνεπώς η επιλογή του Advanced Simple Profile για την πραγματοποίηση του πειραματικού μέρους του παρόντος κεφαλαίου, είναι πολυπλεύρως ενδιαφέρουσα.

Στη διαδικασία εκτίμησης της αντιληπτής ποιότητας βίντεο, χρησιμοποιήθηκαν πέντε βίντεο αναφοράς (Suzie, Cactus, Table Tennis, Flower Garden και Mobile & Calendar). Αυτά τα βίντεο αναφοράς αντιπροσωπεύουν ένα ευρύ φάσμα όσον αφορά το επίπεδο της δυναμικής του περιεχομένου. Προς επαλήθευση αυτής της παρατήρησης, ταξινομήσαμε τα βίντεο της πειραματικής διαδικασίας σύμφωνα με τη χωρική και χρονική δυναμική του περιεχομένου. Για το λόγο αυτό θεωρήσαμε το καρτεσιανό επίπεδο [119] του σχήματος 6.1, όπου ο οριζόντιος άξονας αναφέρεται στη χωρική δυναμική του βίντεο και ο κατακόρυφος στη χρονική αντίστοιχα.



Σχήμα 6.1 Το προτεινόμενο επίπεδο ταξινόμησης της δυναμικής του περιεχομένου

Σύμφωνα με την προσέγγιση αυτή, το φάσμα της δυναμικής του περιεχομένου ενός βίντεο αναλύεται σε τέσσερα τεταρτημόρια, τα οποία είναι τα εξής:

- Πάνω αριστερό τεταρτημόριο: Χαμηλή Χωρική – Χαμηλή Χρονική Συνιστώσα
- Πάνω δεξιό τεταρτημόριο: Υψηλή Χωρική – Χαμηλή Χρονική Συνιστώσα
- Κάτω αριστερό τεταρτημόριο: Χαμηλή Χωρική – Υψηλή Χρονική Συνιστώσα
- Κάτω δεξιό τεταρτημόριο: Υψηλή Χωρική – Υψηλή Χρονική Συνιστώσα

Για την ποσοτική μέτρηση της χωρικής συνιστώσας ενός βίντεο χρησιμοποιήθηκε ο μέσος όρος της απόκλισης των τιμών φωτεινότητας των pixels ενός καρτέ. Στη διαδικασία αυτή δεν λαμβάνουμε υπόψιν τη χρωματική συνιστώσα, διότι δεν επηρεάζει την πολυπλοκότητα της κωδικοποίησης και συνεπώς δεν επιφέρει καμία μεταβολή στον αντίστοιχο λόγο απόδοσης.

Θεωρώντας ότι ένα καρέ περιλαμβάνει N pixels x_i , τότε ανά καρέ η μέση απόκλιση των τιμών ορίζεται ως:

$$S^2_{frame,y} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

Συνεπώς σύμφωνα με της εξίσωση αυτή, η μέση απόκλιση ανά καρέ για όλη της διάρκεια ενός βίντεο, δεδομένου ότι αποτελείται από K καρέ, ορίζεται ως:

$$\frac{1}{K} \sum_{k=1}^K S^2_{frame,y} = \frac{1}{K} \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^N (x_{k,i} - \bar{x}_k)^2$$

Για τη μέτρηση της χρονικής συνιστώσας της δυναμικής του περιεχομένου ενός βίντεο χρησιμοποιήθηκε η μέση απόκλιση της διαφοράς φωτεινότητας μεταξύ των διαδοχικών καρέ. Το προτεινόμενο μετρικό παράγει χαμηλές τιμές για στατικά βίντεο, όπου το περιεχόμενο των διαδοχικών καρέ έχει μεταβληθεί ελάχιστα, και υψηλές τιμές για βίντεο υψηλής δυναμικής, όπου τα διαδοχικά καρέ διαφέρουν σημαντικά. Θεωρώντας ότι ένα καρέ περιέχει N pixels x_i και ότι το βίντεο υπό εξέταση περιέχει συνολικά K καρέ, τότε η μέση διαφορά μεταξύ των διαδοχικών καρέ για όλη τη διάρκεια του βίντεο, ορίζεται ως εξής:

$$\frac{1}{K-1} \sum_{k=2}^K \frac{1}{N} \sum_{i=1}^N (x_{k,i} - x_{k-1,i})$$

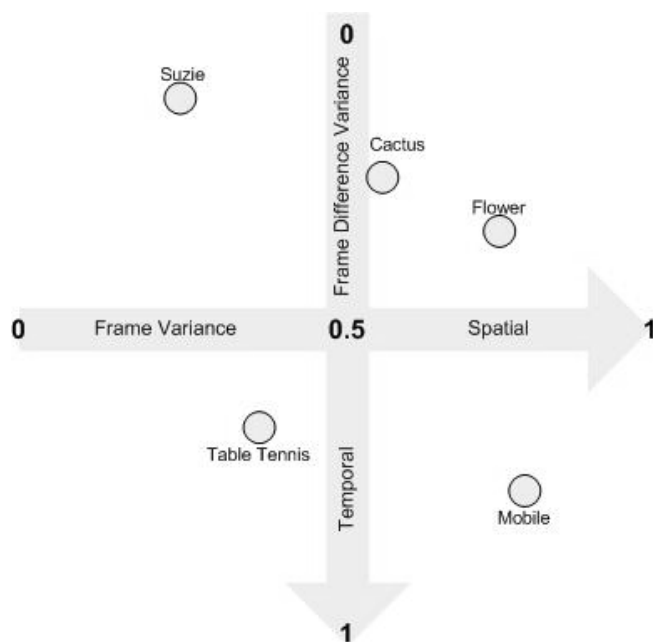
Συνεπώς, η μέση απόκλιση της διαφοράς φωτεινότητας μεταξύ διαδοχικών καρέ για όλη τη διάρκεια του βίντεο υπό εξέταση δίνεται από τη σχέση:

$$\frac{1}{K-1} \sum_{k=2}^K \left(\frac{1}{N} \sum_{i=1}^N (x_{k,i} - x_{k-1,i}) - \frac{1}{K-1} \sum_{k=2}^K \frac{1}{N} \sum_{i=1}^N (x_{k,i} - x_{k-1,i}) \right)^2$$

Η κλίμακα που χρησιμοποιείται στους δύο άξονες του καρτεσιανού χώρου που απεικονίζεται στο σχήμα 6.1 αναφέρεται στις κανονικοποιημένες μετρήσεις της χωρικής και χρονικής συνιστώσας (δηλαδή θεωρούμε μια κλίμακα από το 0 έως το 1) των προαναφερθέντων μετρικών. Η διαδικασία κανονικοποίησης που ακολουθήθηκε, θέτει ως αρχή το βίντεο με τα υψηλότερα αποτελέσματα στο χωρικό και χρονικό άξονα, τοποθετώντας το στις καρτεσιανές τιμές (0.75, 0.75), δηλαδή στο κέντρο του τέταρτου τεταρτημορίου. Η υπόθεση αυτή, χωρίς βλάβη της γενικότητας, επιτρέπει στη διαδικασία ταξινόμησης που πραγματοποιούμε να θεωρήσουμε βίντεο, τα οποία ενδεχομένως να εμφανίζουν υψηλότερες χωροχρονικές συνιστώσες συγκριτικά με τα βίντεο υπό δοκιμή.

Κανονικοποιώντας τις μετρήσεις των πέντε βίντεο σύμφωνα με τις σταθερές που προκύπτουν ώστε το βίντεο 'Mobile & Calendar', το οποίο παρήγαγε τα υψηλότερα αποτελέσματα, να τοποθετηθεί εξ ορισμού στο μέσο του τέταρτου τεταρτημορίου, προκύπτουν τα αποτελέσματα που απεικονίζονται στο σχήμα 6.2

Όπως, μπορεί να παρατηρηθεί τα χρονοχωρικά χαρακτηριστικά του περιεχομένου των βίντεο αναφοράς που χρησιμοποιούνται στο παρόν κεφάλαιο, είναι αντιπροσωπευτικά όλου του εύρους της δυναμικής συνιστώσας, αφού κατανέμονται και στα τέσσερα τεταρτημόρια του προτεινόμενου καρτεσιανού χώρου.



Σχήμα 6.2 Απεικόνιση της χωρικής και χρονικής δυναμικής των πέντε βίντεο αναφοράς

Τα πέντε βίντεο αναφοράς του παρόντος κεφαλαίου κωδικοποιήθηκαν και συμπίεστηκαν από την αρχική τους ασυμπίεστη ψηφιακή μορφή σύμφωνα με το πρότυπο MPEG-4 Advanced Simple Profile με χωρική ανάλυση CIF. Επιπλέον, επιλέχθηκαν οι ακόλουθες τιμές ρυθμού κωδικοποίησης 100, 150, 200, 250, 300, 400, 500 kbps, χρησιμοποιώντας μέθοδο κωδικοποίησης σταθερού και μεταβλητού ρυθμού.

Στον κωδικοποιητή, οι παράμετροι κωδικοποίησης επιλέχθηκαν ως εξής: Η δομή MPEG GOV ορίστηκε με Key Frame Period 100 καρέ και B Frame Period 2 καρέ. Η σχέση ποιότητας/ταχύτητα κωδικοποίησης ορίστηκε στην επιλογή “Υψηλή Ποιότητα”. Σχετικά με την κωδικοποίηση σταθερού ρυθμού, χρησιμοποιήθηκε υλοποίηση VBV Buffer, με μέγεθος προσωρινής μνήμης 1 δευτερόλεπτο.

Στη συνέχεια, από κάθε κωδικοποιημένο βίντεο εξήχθησαν με τυχαίο τρόπο 20 αποκωδικοποιημένα καρέ, τα οποία χρησιμοποιήθηκαν για τη διαδικασία εκτίμησης της ποιότητας ανάμεσα στη μέθοδο VBR και CBR σε όλους τους προαναφερθέντες ρυθμούς (δηλ. 100-500kbps). Τα επιλεγμένα αποκωδικοποιημένα καρέ περιλαμβάνουν καρέ I, B ή P, πιστοποιώντας με τον τρόπο αυτό ότι η πειραματική μας διαδικασία θέτει υπό εξέταση την αποδοτικότητα της Intra αλλά και της Inter κωδικοποίησης.

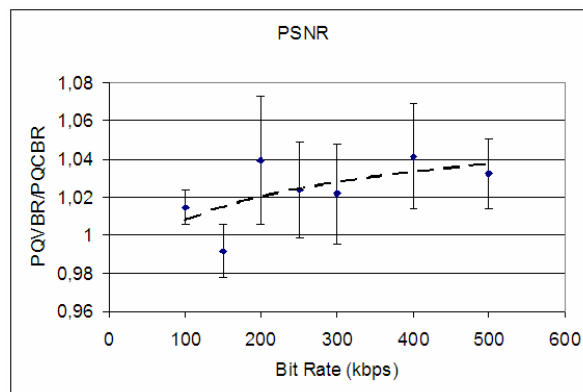
Για να μπορέσουμε να κανονικοποιήσουμε την κλίμακα αξιολόγησης της ποιότητας βίντεο που χρησιμοποιεί το κάθε μετρικό, η ακόλουθη διαδικασία πραγματοποιήθηκε: Για κάθε ένα από τα πέντε βίντεο που χρησιμοποιήθηκαν στα πειράματα, υπολογίστηκε ο λόγος της αντιληπτής ποιότητας του VBR (Perceived Quality of VBR – PQVBR) προς την αντιληπτή ποιότητα του CBR (Perceived Quality of CBR – PQCBR) σε κάθε εξεταζόμενο ρυθμό κωδικοποίησης. Οπότε, για κάθε μετρικό εξάγεται ο λόγος PQVBR/PQCBR, ο οποίος είναι κανονικοποιημένος και ανεξάρτητος από την κλίμακα που χρησιμοποιεί το κάθε μετρικό. Επιπλέον, για εκείνα τα μετρικά (VQM και Delta), που δεν ακολουθούν μια αύξουσα κλίμακα αξιολόγησης (δηλαδή όσο βελτιώνεται η ποιότητα βίντεο, τόσο οι τιμές που παρέχει το μετρικό μικραίνουν), ο αντίστροφος λόγος των τιμών χρησιμοποιήθηκε στον υπολογισμό του PQVBR/PQCBR. Με αυτό τον τρόπο επιτυγχάνουμε εναρμόνιση των αποτελεσμάτων με τα υπόλοιπα μετρικά, στα οποία υψηλότερες τιμές αντιστοιχούν σε καλύτερη ποιότητα.

Για κάθε x_i των εξαγόμενων λόγων PQVBR/PQCBR κάθε μετρικού στους ρυθμούς κωδικοποίησης i , υπολογίζεται η μέση τιμή και το απόλυτο σφάλμα της μέσης τιμής, σύμφωνα με τις ακόλουθες εξισώσεις.

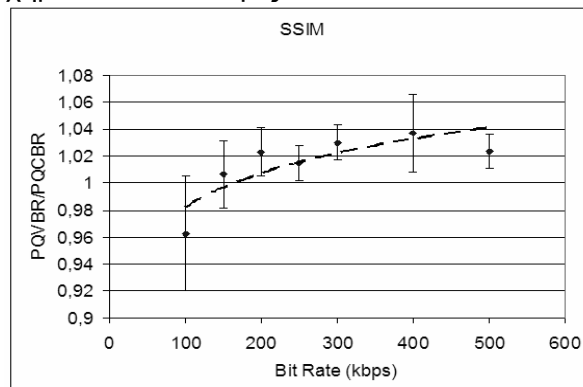
$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \delta x = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N(N-1)}}$$

Με τον τρόπο αυτό, κάθε μετρικό παρέχει μια μοναδική εκτίμηση του λόγου PQVBR/PQCBR σε κάθε υπό εξέταση ρυθμό κωδικοποίησης i για όλα τα υπό εξέταση βίντεο. Συνεπώς, η εκτίμηση αυτή του λόγου PQVBR προς PQCBR είναι της μορφής $x = \bar{x} \pm \delta x$. Επιπλέον, για να διατηρήσουμε την ακρίβεια των εξαγόμενων αποτελεσμάτων, χρησιμοποιήθηκε το κριτήριο Chauvenet στις πειραματικές τιμές ώστε να απορριφθούν οι ανακριβείς εκτιμήσεις.

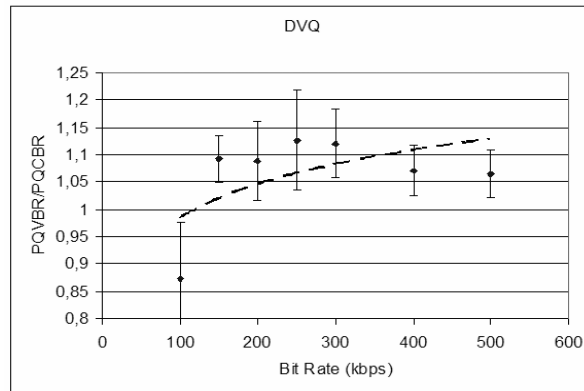
Σε αυτά τα πειραματικά πλαίσια, τα σχήματα 6.3-6 απεικονίζουν τους πειραματικούς λόγους PQVBR/PQCBR για κάθε ένα από τα χρησιμοποιούμενα μετρικά επί του συνόλου των πειραματικών σημάτων.



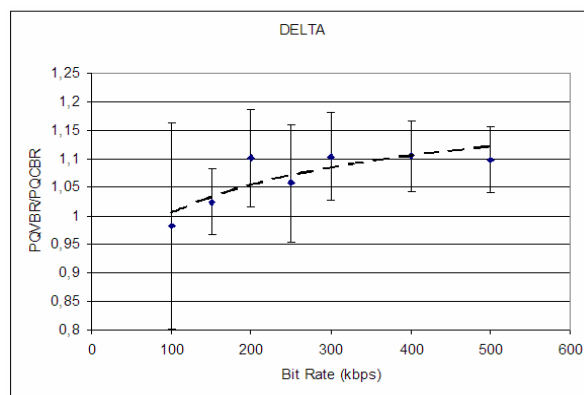
Σχήμα 6.3 Ο λόγος PQVBR/PQCBR του SNR



Σχήμα 6.4 Ο λόγος PQVBR/PQCBR του SSIM



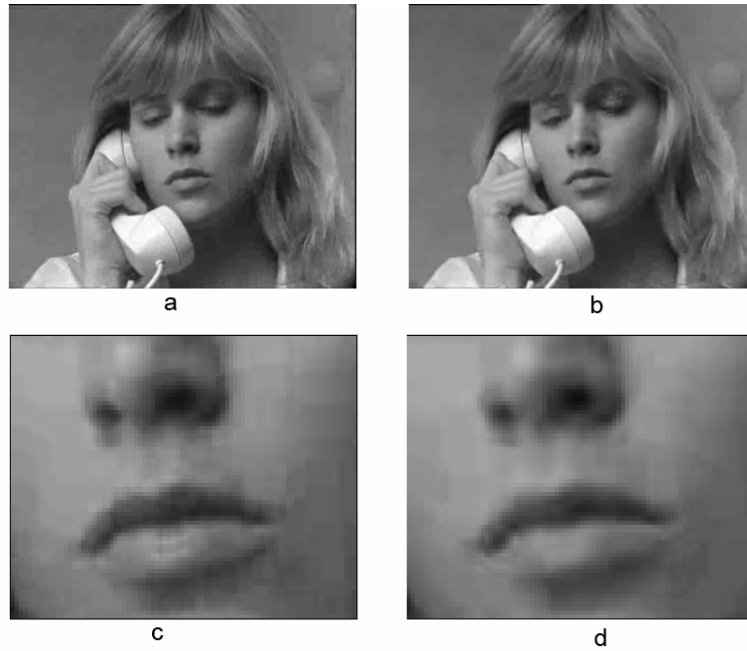
Σχήμα 6.5 Ο λόγος PQVBR/PQCBR του DVQ



Σχήμα 6.6 Ο λόγος PQVBR/PQCBR του Delta

Σε κάθε σχήμα απεικονίζεται με γραφικό τρόπο η αντίστοιχη τιμή του απόλυτου σφάλματος της μέσης τιμής. Με βάση τα πειραματικώς εξαγόμενα σημεία, χαράζουμε τη βέλτιστη εκθετική καμπύλη, ώστε να δοθεί μια εποπτική εικόνα της τάσης του λόγου PQVBR/PQCBR. Όπως είναι εμφανές η αντιληπτή υπεροχή του VBR έναντι του CBR είναι αύξουσα σε σχέση με το ρυθμό κωδικοποίησης. Πιο συγκεκριμένα, το VBR φαίνεται να είναι πιο αποδοτικό σε σχέση με το CBR, εκτός από την περιοχή μικρών τιμών του ρυθμού κωδικοποίησης (<125kbps). Επιπλέον, για υψηλές τιμές του ρυθμού κωδικοποίησης ο λόγος PQVBR/PQCBR παραμένει πρακτικά σταθερός.

Για να μπορέσουμε να επιδείξουμε τη διαφορά στην αντιληπτή ποιότητα ανάμεσα στη μέθοδο κωδικοποίησης σταθερού και μεταβλητού ρυθμού, το σχήμα 6.7 απεικονίζει δύο αντιπροσωπευτικά καρέ του σήματος αναφοράς Suzie, το οποίο κωδικοποιήθηκε με ASP CIF/200kbps με CBR –a- και VBR –b- αντίστοιχα. Οι υπόλοιπες παράμετροι κωδικοποίησης παρέμειναν σταθεροί. Σύμφωνα με το σχήμα, αν και αρχικά καμιά σημαντική διαφορά δεν παρατηρείται ανάμεσα στα δύο καρέ, μια πιο προσεκτική εξέταση (-c- και -d-) αποκαλύπτει την υπεροχή του VBR έναντι του CBR.



Σχήμα 6.7 Σύγκριση του PQVBR/PQCBR σε δύο καρέ του βίντεο Suzie

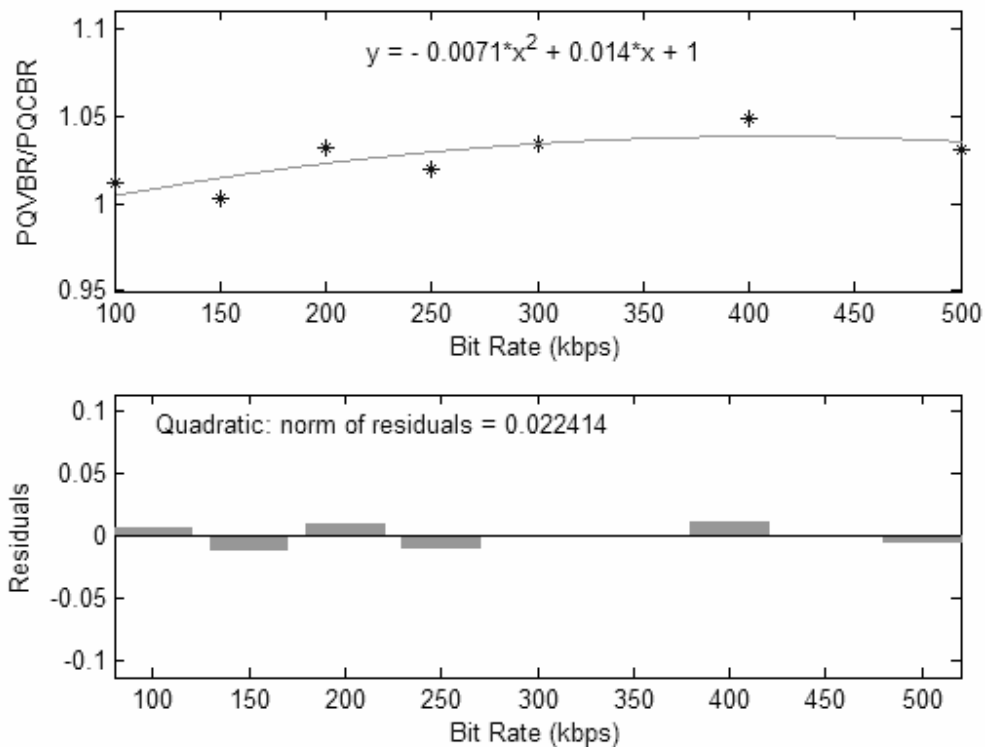
Για να οδηγηθούμε σε μια γενικευμένη εκτίμηση της αντιληπτής υπεροχής του VBR έναντι του CBR για την περίπτωση CIF χωρικής ανάλυσης, ένα συγκεντρωτικό γράφημα συνεκτιμά τα προηγούμενα γραφήματα που απεικονίζονται στα Σχήματα 6.3-6. Δηλαδή η γενικευμένη αυτή καμπύλη συνδυάζει τις επιμέρους εκτιμήσεις των χρησιμοποιούμενων μετρικών σε μία μοναδική εκτίμηση.

Για να συνεκτιμήσουμε τους διακριτούς λόγους PQVBR/PQCBR των ποικίλων μετρικών σε μια κοινή συνεκτίμηση, χρησιμοποιούμε την ακόλουθη διαδικασία από τη θεωρία σφαλμάτων:

Θεωρώντας τους λόγους x_1, x_2, \dots, x_4 του PQVBR/PQCBR για τα τέσσερα μετρικά της πειραματικής διαδικασίας, όπου $x_k = \bar{x}_k \pm \delta x_k$ με $K = \{1, 2, 3, 4\}$, υπολογίζουμε ένα γενικευμένο $X_i + dX_i$ λόγο PQVBR/PQCBR για κάθε ρυθμό κωδικοποίησης i , βασισμένοι στις ακόλουθες εξισώσεις:

$$X_i = \frac{\sum_{k=1}^N w_k x_k}{\sum_{k=1}^N w_k}, \text{ με } w_k = 1/(dx_k)^2 \text{ και } dX_i = \sqrt{1/\sum_{k=1}^N w_k}$$

Εφαρμόζοντας τη μέθοδο αυτή στα πειραματικά δεδομένα που απεικονίζονται στις γραφικές παραστάσεις των σχημάτων 6.3-6, θα οδηγηθούμε σε ένα γενικευμένο λόγο του PQVBR/PQCBR, ο οποίος απεικονίζεται στο σχήμα 6.8.



Σχήμα 6.8 Ο λόγος PQVBR/PQCBR για κωδικοποίηση ASP MPEG-4/CIF και οι αντίστοιχες αποκλίσεις μεταξύ πειραματικών και θεωρητικών τιμών.

Οι αντίστοιχες πειραματικές τιμές του γενικευμένου λόγου PQVBR/PQCBR, απεικονίζονται στον πίνακα 6.1 που ακολουθεί:

Ρυθμός Κωδικοποίησης (kbps)	Γενικευμένος Λόγος PQVBR/PQCBR
100	1.01124±0.00891
150	1.00307±0.01146
200	1.03195±0.01538
250	1.01919±0.01158
300	1.03337±0.01159
400	1.04887±0.01747
500	1.03026±0.00997

Πίνακας 6.1 Οι πειραματικές τιμές του γενικευμένου λόγου CIF PQVBR/PQCBR

Όπως παρατηρείται στη γραφική απεικόνιση των πειραματικών αποτελεσμάτων, ο γενικευμένος λόγος του PQVBR/PQCBR για ρυθμούς κωδικοποίησης από 100kbps ως 500kbps, μπορεί επιτυχώς να προσεγγιστεί από την ακόλουθη εξίσωση δευτέρου βαθμού:

$$y = -0.0071x^2 + 0.014x + 1, 100 \leq x \leq 500$$

Όπου γ είναι ο γενικευμένος λόγος του PQVBR/PQCBR και x ο ρυθμός κωδικοποίησης. Σύμφωνα με την εξίσωση αυτή, είναι δυνατό να υπολογιστεί πριν πραγματοποιηθεί η κωδικοποίηση με αναλυτικό τρόπο η αντιληπτή υπεροχή του μεταβλητού ρυθμού έναντι του σταθερού ρυθμού για την περίπτωση κωδικοποίησης MPEG-4 ASP/CIF. Η ακρίβεια του προτεινόμενου μοντέλου μπορεί να παρατηρηθεί από τις αντίστοιχες διαφορές μεταξύ πειραματικών και θεωρητικών τιμών, όπως απεικονίζονται στο σχήμα 6.8. Η νόρμα των διαφορών υπολογίστηκε ίση με 0,022414 υποδεικνύοντας ικανοποιητική προσέγγιση των πειραματικών τιμών από το θεωρητικό μοντέλο.

Αναλύοντας περισσότερο τα εξαγόμενα αποτελέσματα του γενικευμένου λόγου PQVBR/PQCBR, μπορούμε να παρατηρήσουμε ότι αν και το PQVBR υπερέχει του PQCBR για όλη την έκταση του χρησιμοποιούμενου ρυθμού, παρόλα αυτά η αντίστοιχη αντιληπτή υπεροχή του PQVBR είναι της τάξης του 4-5% έναντι του PQCBR. Επιπλέον, η υπεροχή αυτή παραμένει πρακτικά σταθερή και αμετάβλητη για ρυθμούς μεγαλύτερους των 200kbps, δείχνοντας με τον τρόπο αυτό την περιορισμένα υψηλότερη απόδοση του μεταβλητού ρυθμού κωδικοποίησης.

Για πληρότητα της εξέτασης της απόδοσης του PQVBR έναντι του PQCBR, η πειραματική διαδικασία επεκτάθηκε ώστε να συμπεριλάβει βίντεο ανάλυσης QCIF σε ρυθμούς 32 και 64 kbps. Επαναλαμβάνοντας την ίδια ακριβώς πειραματική διαδικασία, όπως περιγράφηκε προηγουμένως για την περίπτωση της ανάλυσης CIF και διατηρώντας σταθερές τις υπόλοιπες παραμέτρους κωδικοποίησης στα πέντε βίντεο αναφοράς (Suzie, Cactus, Table Tennis, Flower Garden και Mobile & Calendar), τότε τα εξαγόμενα κωδικοποιημένα βίντεο χρησιμοποιήθηκαν ως δεδομένα στα τέσσερα μετρικά. Στη συνέχεια, υπολογίστηκε ο γενικευμένος λόγος PQVBR/PQCBR, όπως ακριβώς και στην περίπτωση της χωρικής CIF ανάλυσης. Τα αποτελέσματα παρουσιάζονται στον πίνακα 6.2, όπου απεικονίζονται τόσο οι επί μέρους, όσο και οι γενικευμένοι λόγοι QCIF PQVBR/PQCBR.

Μετρικό	PQVBR/PQCBR 32kbps	PQVBR/PQCBR 64kbps
PSNR	1.0218±0.0069	1.0314±0.0037
SSIM	1.0224±0.0044	1.0153±0.0041
Delta	1.1026±0.0573	1.1224±0.0317
VQM	1.0720±0.0351	1.0832±0.0201
Μέσος Όρος	1.0231±0.0037	1.0259±0.0027

Πίνακας 6.2 Οι πειραματικές τιμές του γενικευμένου λόγου QCIF PQVBR/PQCBR

Τα πειραματικά αποτελέσματα δείχνουν ότι για την περίπτωση του QCIF τα επίπεδα του λόγου PQVBR/PQCBR κινούνται σε παρόμοια και χαμηλότερα επίπεδα από αυτά του CIF. Πιο συγκεκριμένα η υπεροχή του VBR έναντι του CBR κυμαίνεται στο 2.5%, γεγονός που υποδεικνύει την περιορισμένα υψηλότερη απόδοση της αντιληπτής ποιότητας του VBR και στην περίπτωση χαμηλών ρυθμών κωδικοποίησης με μικρές χωρικές αναλύσεις.

Συμπεράσματα του Κεφαλαίου

Το κεφάλαιο αυτό παρουσίασε μια αναλυτική και ποσοτική διερεύνηση της σύγκρισης της μεθόδου κωδικοποίησης μεταβλητού ρυθμού έναντι της κωδικοποίησης σταθερού ρυθμού, σχετικά με την επιτεύξιμη ποιότητα βίντεο του κωδικοποιημένου σήματος. Η ερευνητική διαδικασία εξέτασε διαφορετικές τιμές ρυθμού κωδικοποίησης και χωρικής ανάλυσης. Τα αποτελέσματα έδειξαν ότι για την περίπτωση κωδικοποίησης CIF MPEG-4 ASP, η αντιληπτή απόδοση του VBR έναντι του CBR είναι της τάξης του 4-5%, ενώ για την περίπτωση QCIF MPEG-4 ASP ο αντίστοιχος λόγος υπολογίσθηκε να είναι της τάξης του 2.5%. Συνεπώς, με την ερευνητική αυτή διαδικασία πιστοποιήθηκε ότι η μέθοδος κωδικοποίησης μεταβλητού ρυθμού –αν και από τεχνικής άποψης- προσφέρει κάποια πλεονεκτήματα έναντι της μεθόδου σταθερού ρυθμού, από την οπτική της αντιληπτής ποιότητας βίντεο οι οποιεσδήποτε διαφορές είναι πρακτικά ασήμαντες.

ΚΕΦΑΛΑΙΟ 7

Μέθοδος Πρόβλεψης της Αντιληπτής Ποιότητας για Βίντεο Βραχείας Διαρκείας

Η Ανάγκη για Μέθοδο Πρόβλεψης της Αντιληπτής Ποιότητας Συμπιεσμένου Βίντεο

Οι πολυμεσικές εφαρμογές που διανέμουν οπτικοακουστικό περιεχόμενο σε τηλεπικοινωνιακά δίκτυα (όπως υπηρεσίες Video on Demand και ροές βίντεο πραγματικού χρόνου) βασίζονται στις τεχνικές ψηφιακής κωδικοποίησης, όπως τα πρότυπα MPEG και ITU που παρουσιάστηκαν στο δεύτερο κεφάλαιο. Όπως έχει ήδη αναφερθεί, η ψηφιακή κωδικοποίηση προκαλεί την εμφάνιση τεχνουργημάτων στην τελική ποιότητα του βίντεο με αποτέλεσμα την ποιοτική υποβάθμιση αυτής. Εξαιτίας του γεγονότος ότι οι παράμετροι με ισχυρή επίδραση στο επίπεδο της ποιότητας του συμπιεσμένου βίντεο είναι κυρίως ο ρυθμός κωδικοποίησης, η χωρική και χρονική ανάλυση, έχει προκύψει το ερευνητικό θέμα της συσχέτισης των παραμέτρων κωδικοποίησης και της χωρικής/χρονικής δυναμικής του περιεχομένου με την αντιληπτή ποιότητα που επιτυγχάνει η διαδικασία κωδικοποίησης.

Ένας από τους μελλοντικούς στόχους στη διαχείριση και διανομή οπτικοακουστικών υπηρεσιών στα τηλεπικοινωνιακά δίκτυα είναι η παροχή πολυμεσικών εφαρμογών σε ποικίλα επίπεδα αντιληπτής ποιότητας και κόστους [16]. Στη βιβλιογραφία υπάρχουν ποικίλες προσεγγίσεις, οι οποίες εν γένει εισάγουν την έννοια της αντιληπτής ποιότητας υπηρεσίας (Perceived Quality of Service) για εφαρμογές βίντεο, καλύπτοντας τις περιπτώσεις χαμηλής, μεσαίας και υψηλής ποιότητας. Επιπλέον, η αντιληπτή ποιότητα υπηρεσίας δίνει τη δυνατότητα στον διαχειριστή του δικτύου και τον πάροχο της υπηρεσίας να ελαχιστοποιήσει τους απαιτούμενους δικτυακούς και αποθηκευτικούς πόρους που χρειάζονται για να διατηρηθεί η αντιληπτή ποιότητα σε ένα συγκεκριμένο επίπεδο, το οποίο αντιστοιχεί σε δεδομένο ποσοστό ευχαρίστησης και ικανοποίησης του χρήστη.

Η εκτίμηση της αντιληπτής ποιότητας υπηρεσίας, που για την περίπτωση του ψηφιακού βίντεο εξειδικεύεται στην εκτίμηση της ποιότητας βίντεο, πραγματοποιείται κυρίως από διαδικασίες υποκειμενικής ή αντικειμενικής εκτίμησης, οι οποίες έχουν εκτενώς αναλυθεί στο τέταρτο κεφάλαιο. Το σημείο που πρέπει εντόνως να σημειωθεί είναι ότι στις τεχνικές και των δύο κατηγοριών, η εκτίμηση της ποιότητας πραγματοποιείται έπειτα από τη διαδικασία της κωδικοποίησης. Συνεπώς, οποιαδήποτε εκτίμηση της ποιότητας πραγματοποιείται με σκοπό να αξιολογηθεί ένα ήδη υπάρχον κωδικοποιημένο σήμα βίντεο και ουχί να εντοπιστούν εκείνες οι παράμετροι κωδικοποίησης που ικανοποιούν και αντιστοιχούν σε ένα συγκεκριμένο επίπεδο ποιότητας και ευχαρίστησης του χρήστη. Δηλαδή, οι υπάρχουσες μέθοδοι πραγματοποιούν μια εκ των υστέρων διαδικασία εκτίμησης της ποιότητας του ήδη κωδικοποιημένου βίντεο. Συνεπώς, δεν δύνανται να χρησιμοποιηθούν από τον πάροχο στον καθορισμό των παραμέτρων κωδικοποίησης που ικανοποιούν συγκεκριμένα επίπεδα ποιότητας. Μια τέτοια προετοιμασία με τις παρούσες τεχνικές εκτίμησης, θα απαιτούσε πολλαπλές επαναληπτικές κωδικοποιήσεις και αξιολογήσεις της ποιότητας βίντεο μέχρι να εντοπιστούν οι παράμετροι που ικανοποιούν το επιθυμητό επίπεδο ποιότητας του βίντεο. Συνεπώς, σήμερα δεν υπάρχει κάποια εμπορικά υλοποιήσιμη τεχνική ή μέθοδος που να προσφέρει με αποδοτικό τρόπο την πρόβλεψη ή εκτίμηση των παραμέτρων κωδικοποίησης που ικανοποιούν –για δεδομένο περιεχόμενο– ένα επιθυμητό επίπεδο ποιότητας.

Συνεπώς, έχοντας εντοπίσει την ανάγκη για τεχνικές και μεθόδους που καθορίζουν και προβλέπουν τις παραμέτρους κωδικοποίησης που ικανοποιούν συγκεκριμένο επίπεδο αντιληπτής ποιότητας, το παρόν κεφάλαιο εισάγει, περιγράφει και δοκιμάζει πρωτότυπη μέθοδο πρόβλεψης της αντιληπτής ποιότητας βίντεο συναρτήσει του ρυθμού κωδικοποίησης και της δυναμικής του περιεχομένου.

Πιο συγκεκριμένα το παρόν κεφάλαιο εισάγει μέθοδο ποσοτικής εκτίμησης του επιπέδου της αντιληπτής ποιότητας ενός βίντεο MPEG-4 σύντομης διάρκειας, παρέχοντας πρόβλεψη του επιπέδου της τελικής ποιότητας πριν το στάδιο της κωδικοποίησης. Η μέθοδος στηρίζεται στην αξιοποίηση αντικειμενικής μετρικής, η οποία εξάγεται πειραματικά και είναι ενδεικτική της χωρικής και χρονικής δυναμικής του περιεχομένου του βίντεο. Το γεγονός ότι η προτεινόμενη μέθοδος παρέχει πρόβλεψη της ποιότητας πριν το στάδιο της κωδικοποίησης, έχει ως άμεσα αποτελέσματα αφενός την αποδοτικότερη διαχείριση των διαθέσιμων πόρων του συστήματος και αφετέρου τη γρήγορη και οικονομικά ανεκτή εκτίμηση της αντιληπτής ποιότητας εφαρμογών βίντεο.

Προς την κατεύθυνση αυτή, ένα λογισμικό μετρικό σύστημα [15] αντικειμενικής μέτρησης μη αναφοράς χρησιμοποιήθηκε για την ποσοτική μέτρηση του επιπέδου της ποιότητας των βίντεο αναφοράς που χρησιμοποιήθηκαν στο παρόν κεφάλαιο. Το μετρικό σύστημα παρέχει αντικειμενική εκτίμηση της ποιότητας βίντεο για κάθε καρέ χωρίς να απαιτείται η χρήση του βίντεο αναφοράς. Αρχικά, η λήψη των μετρήσεων της ποιότητας βίντεο υπολογίστηκαν για μικρής διάρκειας MPEG-4 βίντεο, τα οποία φέρουν ομοιόμορφα χωροχρονικά χαρακτηριστικά της δυναμικής του περιεχομένου. Η γραφική αναπαράσταση αυτών των αποτελεσμάτων συναρτήσει του χρόνου παρουσιάζει τη στιγμιαία διακύμανση της αντιληπτής ποιότητας υπηρεσίας (Perceived Quality of Service – PQoS) για κάθε καρέ που συνιστά το υπό δοκιμή βίντεο. Στη συνέχεια, από τη στιγμιαία διακύμανση του PQoS εξάγεται η μέση διακύμανση του PQoS (Mean PQoS –MPQoS) για όλη τη διάρκεια του βίντεο.

Η προαναφερθείσα πειραματική διαδικασία επαναλήφθηκε για το ίδιο περιεχόμενο, μεταβάλλοντας κάθε φορά το ρυθμό κωδικοποίησης και τη χωρική ανάλυση. Η γραφική απεικόνιση αυτής της εξάρτησης καθορίζει τον πειραματικό προσδιορισμό της αντίστοιχης συνάρτησης του MPQoS συναρτήσει του ρυθμού και των λοιπών παραμέτρων κωδικοποίησης. Στη συνέχεια πραγματοποιείται σύγκριση των πειραματικά εξαγόμενων γραφικών καμπυλών με τις αντίστοιχες θεωρητικές αλγεβρικές συναρτήσεις ωφελείας [120], δείχνοντας την ποιοτική τους ομοιότητα, γεγονός που ενισχύει την εγκυρότητα του MPQoS ως μετρική οντότητα για την αντικειμενική πρόβλεψη της ποιότητας κωδικοποιημένου βίντεο.

Μια γενικευμένη προσέγγιση του θεωρητικού μοντέλου των συναρτήσεων ωφελείας παρουσιάζεται στην εργασία [120], όπου οι αλγεβρικές μορφές τους χρησιμοποιούνται για την αναπαράσταση και απεικόνιση του μεγέθους της ευχαρίστησης του χρήστη συναρτήσει των σχετικών πόρων του συστήματος που έχουν διατεθεί σε ανταγωνιστικές μεταξύ τους πολυμεσικές υπηρεσίες. Ο όρος συνάρτηση ωφελείας και η μαθηματική αναπαράσταση του βαθμού ευχαρίστησης του χρήστη που προέρχεται από την επιλογή συγκεκριμένων παραμέτρων ποιότητας υπηρεσίας εισήχθησαν για πρώτη φορά στην εργασία [121].

Η πειραματική διαδικασία για τη θεμελίωση της προτεινόμενης μεθόδου του παρόντος κεφαλαίου, πραγματοποιήθηκε στα βίντεο αναφοράς (Suzie, Cactus, Flower Garden, Mobile and Calendar), τα οποία όπως αποδείχτηκε στο έκτο κεφάλαιο με τη χρήση κατάλληλων μετρικών, είναι αντιπροσωπευτικά όλου του εύρους της χωρικής και χρονικής δυναμικής του περιεχομένου.

Επιπλέον, το παρόν κεφάλαιο αποδεικνύει ότι πειραματικές καμπύλες του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης μπορούν επιτυχώς να προσεγγιστούν από μια οικογένεια εκθετικών συναρτήσεων, η χρήση των οποίων περιορίζει τα QoS χαρακτηριστικά του κάθε βίντεο σε τρεις αλληλεξαρτημένες παραμέτρους. Οι τρεις αυτές παράμετροι ορίζουν μοναδικά για κάθε επίπεδο δυναμικής του περιεχομένου, το αντίστοιχο άνυσμα ποιότητας (Quality Vector – QV). Δείχνοντας ότι οι παράμετροι αυτοί είναι συσχετισμένοι, αποδεικνύεται ότι ο πειραματικός εντοπισμός της μία μόνο παραμέτρου ενός δεδομένου περιεχομένου, επαρκεί για τον αναλυτικό προσδιορισμό των υπολοίπων δύο. Προς αυτή την κατεύθυνση, το παρόν κεφάλαιο δείχνει ότι μία μόνο δοκιμαστική μέτρηση του MPQoS είναι ικανή για τον αναλυτικό προσδιορισμό της εκθετικής σχέσης του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης. Σαν αποτέλεσμα, το προτεινόμενο μετρικό μπορεί να αξιοποιηθεί ως κριτήριο για ποιοτικές εκτιμήσεις πριν την τελική διαδικασία της κωδικοποίησης, σχετικά με τον εντοπισμό εκείνων των παραμέτρων κωδικοποίησης που ικανοποιούν ένα συγκεκριμένο επίπεδο ποιότητας, λαμβάνοντας υπόψη τη δυναμική του περιεχομένου.

Σε αυτό το σημείο πρέπει να τονιστεί ότι η προτεινόμενη μέθοδος αναπτύχθηκε βασισμένη σε βίντεο μικρής διάρκειας, των οποίων το περιεχόμενο χαρακτηρίζεται ως ομοιογενές όσον αφορά τη χωρική και χρονική δυναμική του. Στις τελευταίες ενότητες του παρόντος κεφαλαίου επεκτείνεται η εγκυρότητα της προτεινόμενης μεθόδου και βίντεο σύντομης διάρκειας με ανομοιογενές περιεχόμενο. Όπως θα παρουσιαστεί, η εγκυρότητα της προτεινόμενης μεθόδου διατηρείται και σε αυτές τις περιπτώσεις, δείχνοντας ότι η ακρίβεια των προβλέψεων είναι ικανοποιητική. Για βίντεο μεγαλύτερης διάρκειας, όπου η δυναμική του περιεχομένου μεταβάλλεται σημαντικά, η προτεινόμενη μέθοδος θα επεκταθεί με τη χρήση μεθόδου ανίχνευσης των ορίων των σκηνών, ώστε να εφαρμόζεται τοπικά και εντός της κάθε σκηνής. Η επέκταση της μεθόδου σε βίντεο μεγαλύτερης διάρκειας παρουσιάζεται στο επόμενο κεφάλαιο.

Το υπόλοιπο του παρόντος κεφαλαίου που έπεται της εισαγωγικής αυτής ενότητας, παρουσιάζει, αναλύει, υλοποιεί και δοκιμάζει την προαναφερθείσα μέθοδο.

Λογισμικό Εργαλείο Μέτρησης της Αντιληπτής Ποιότητας

Για τις ανάγκες του παρόντος κεφαλαίου χρησιμοποιήθηκε μια υλοποίηση λογισμικής προσέγγισης για τη μέτρηση και την ποσοτικοποίηση της αντιληπτής ποιότητας: Το Quality Meter Software (QMS). Το QMS στηρίζεται σε έναν από τους πλέον διαδεδομένους αλγορίθμους αντικειμενικής μέτρησης της ποιότητας χωρίς την απαίτηση βίντεο αναφοράς [15]. Δεδομένου ότι ανήκει στην κλάση των αντικειμενικών μεθόδων μη αναφοράς, η χρήση του μπορεί να χαρακτηριστεί ως εύκολη και σύντομη, χωρίς να απαιτείται κάποια προεργασία.

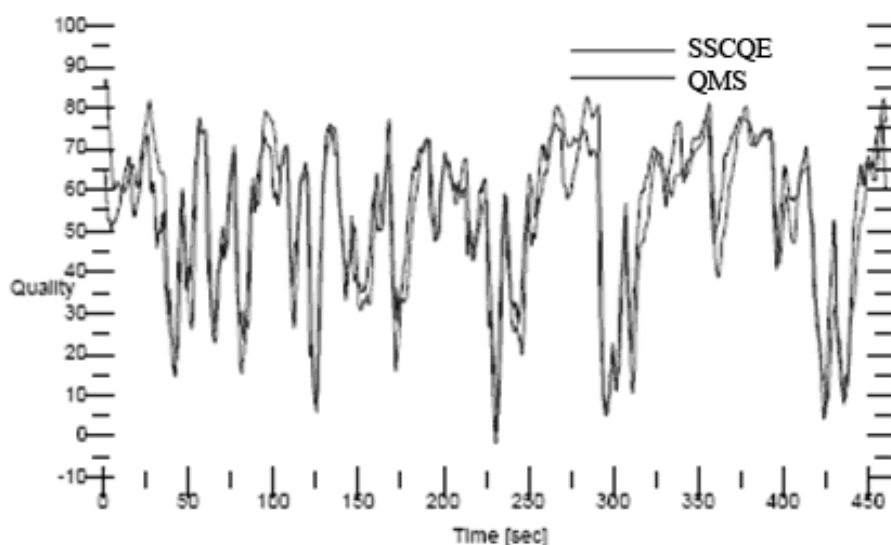
Το QMS μετράει με αντικειμενικό τρόπο τη στιγμιαία μεταβολή του PQoS στη κλίμακα από το 1 έως το 100. Ο αλγόριθμος που υλοποιείται στηρίζεται σε ανύσματα αναφοράς, τα οποία περιέχουν τη διαφορά της φωτεινότητας γειτονικών pixels που βρίσκονται κατά μήκος των blocks και macroblocks του κωδικοποιημένου σήματος. Η υψηλή συμπίεση που επιτυγχάνεται κατά τη διαδικασία κωδικοποίησης MPEG-4, όπως έχει ήδη αναφερθεί σε προγενέστερο κεφάλαιο, προκαλεί την απώλεια των υψηλών συχνοτήτων στους αντίστοιχους συντελεστές του διακριτού συνημιτονικού μετασχηματισμού (DCT). Εντός ενός block MPEG-4 (8x8 pixels), οι διαφορές στη φωτεινότητα εξαλείφονται και ομαλοποιούνται από τη διαδικασία κωδικοποίησης. Αντιθέτως για όλα εκείνα τα ζευγάρια των γειτονικών

pixels που βρίσκονται κατά μήκος και παραπλεύρως των πλευρών των γειτονικών DCT blocks, οι διαφορές στη συνιστώσα φωτεινότητας οξύνονται. Η διαφορά της μέσης φωτεινότητας των προαναφερθέντων ζευγαριών pixels είναι άρρηκτα συνδεδεμένη με τις παραμέτρους κωδικοποίησης και κυρίως με το επιλεγμένο ρυθμό αυτής [44].

Συνεπώς, χαμηλός ρυθμός κωδικοποίησης προκαλεί όξυνση των διαφορών φωτεινότητας μεταξύ των γειτονικών blocks, γεγονός που συνεπάγεται την ενίσχυση του φαινομένου τετραγώνισης, το οποίο όπως έχει αναφερθεί σε προηγούμενο κεφάλαιο είναι ένα από τα κυρίαρχα τεχνουργήματα που προκύπτουν από τα πρότυπα κωδικοποίησης MPEG. Η ενίσχυση του φαινομένου τετραγώνισης με τη χρήση χαμηλού ρυθμού κωδικοποίησης συνεπάγεται την υποβάθμιση της αντιληπτής ποιότητας του κωδικοποιημένου βίντεο. Βασισμένοι στο γεγονός αυτό, το εργαλείο QMS αξιοποιεί τις διακυμάνσεις των διαφορών της φωτεινής συνιστώσας ως αντικειμενική μετρική στη διαδικασία ποσοτικοποίησης και εκτίμησης της ποιότητας βίντεο.

Η αξιοπιστία του QMS ως μετρικό αντικειμενικής μέτρησης έχει επιβεβαιωθεί με τη σύγκριση αποτελεσμάτων υποκειμενικής εκτίμησης της ποιότητας με τα αντίστοιχα που προήλθαν από τη χρήση του QMS [15]. Τα αποτελέσματα της υποκειμενικής εκτίμησης υπολογίστηκαν με την εφαρμογή της μεθόδου SSCQE, η οποία έχει αναλυθεί στο τέταρτο κεφάλαιο.

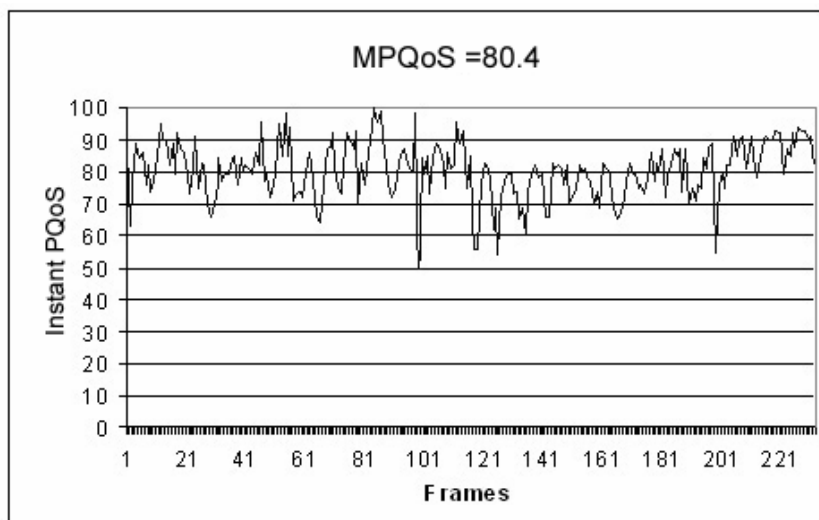
Η σύγκριση αυτή, που απεικονίζεται στο σχήμα 7.1, απέδειξε ότι το QMS, παρά το γεγονός ότι βασίζεται σε ένα σχετικά απλό αλγόριθμο, προσομοιάζει επιτυχώς τα αντίστοιχα αποτελέσματα των υποκειμενικών μεθόδων εκτίμησης.



Σχήμα 7.1 Σύγκριση αποτελεσμάτων του QMS και SSCQE.

Επί παραδείγματι, στην εικόνα 7.2 απεικονίζεται ένα τυπικό γράφημα μέτρησης της αντιληπτής ποιότητας με τη χρήση του λογισμικού QMS για το βίντεο αναφοράς Mobile & Calendar, το οποίο κωδικοποιήθηκε σύμφωνα με το πρότυπο MPEG-4 σε Simple Profile με ρυθμό 800kbps και ανάλυση CIF. Το μήκος του GOP ορίστηκε στα 100 καρέ και ο ρυθμός αυτών στα 25 καρέ το δευτερόλεπτο. Η στιγμιαία μεταβολή του PQoS συναρτήσεως του χρόνου (όπου ο χρόνος απεικονίζεται μέσω της εναλλαγής των καρέ) κυμαίνεται ανάλογα με τη χωρική και χρονική δυναμική του περιεχομένου. Για διαδοχικά καρέ με έντονη

απεικονιζόμενη δραστηριότητα, το επίπεδο του PQoS πέφτει. Αντιθέτως για καρέ με χαμηλή δυναμική οι τιμές του PQoS είναι μεγαλύτερες.



Σχήμα 7.2 Πειραματικά αποτελέσματα του QMS

Πρέπει ιδιαίτερος να τονιστεί ότι στο παρόν κεφάλαιο, όλα τα πειραματικά δεδομένα πραγματοποιήθηκαν με τη μέθοδο Σταθερού Ρυθμού Κωδικοποίησης (CBR), η οποία παράγει σταθερή ροή δεδομένων με κόστος την διακύμανση της ποιότητας (οι συντελεστές κβάντισης μεταβάλλονται αντιστρόφως ανάλογα της δυναμικής του περιεχομένου). Η περίπτωση της μεταβλητής ροής δεν θα εξεταστεί σε αυτό το κεφάλαιο, αφού μπορεί εύκολα να προκύψει χρησιμοποιώντας το μοντέλο απεικόνισης μεταξύ σταθερού και μεταβλητού ρυθμού του προηγούμενου κεφαλαίου. Συνεπώς, όλα τα πειραματικά δεδομένα προέρχονται από τη χρήση σταθερού ρυθμού, η οποία από άποψη αντιληπτής ποιότητας προκαλεί μεγαλύτερη διακύμανση και κατά συνέπεια μεγαλύτερο στατιστικό σφάλμα.

Συνεπώς η χρήση πειραματικών καμπυλών του PQoS συναρτήσεϊ του χρόνου για δεδομένο ρυθμό κωδικοποίησης, μπορούν να χρησιμοποιηθούν ως μετρικό για τη κατηγοριοποίηση της δυναμικής ενός σύντομου σε διάρκεια βίντεο. Το αίτημα σύντομης διάρκειας εισάγεται ώστε να έχει νόημα η κατάταξη του βίντεο ανάλογα με τη δυναμική του περιεχομένου, διότι σε βίντεο μεγάλης διάρκειας παρατηρείται έντονη διαφοροποίηση του περιεχομένου. Πρόκειται δηλαδή για μια απαίτηση, η οποία –όπως θα επιδειχθεί παρακάτω- διατηρεί σε υψηλά επίπεδα την ακρίβεια της προτεινόμενης μεθόδου.

Αξιοποιώντας την έννοια του Μέσου PQoS (MPQoS) ως μετρικό, ο μέσος όρος του στιγμιαίου PQoS (Instant PQoS) για όλη τη διάρκεια του βίντεο, μπορεί να αξιοποιηθεί για την κατάταξη ενός βίντεο στην κλίμακα της αντιληπτής ποιότητας.





$$MPQoS = \frac{\sum_{i=1}^N Instant PQoS_i}{N}, \text{ όπου } N \text{ είναι ο συνολικός αριθμός των καρέ του σήματος}$$

Για παράδειγμα, θεωρώντας τρία επίπεδα ποιότητας, τα οποία ορίζονται ως χαμηλό, μεσαίο και υψηλό (οι αντίστοιχες αριθμητικές κλίμακες μπορούν να οριστούν ως 70-80, 80-90 και 90-100), το βίντεο που απεικονίζεται στο σχήμα 7.2, έχει MPQoS ίσο με 80,4 και μπορεί να ταξινομηθεί και να κατηγοριοποιηθεί ως μεσαίας ποιότητας βίντεο. Τα ακριβή αριθμητικά όρια των λεκτικών ποιοτικών κατηγοριών μπορούν να οριστούν επακριβώς ανάλογα με τις ανάγκες του παρόχου της υπηρεσίας.

Διακύμανση του MPQoS Συναρτήσεως του Ρυθμού Κωδικοποίησης

Για τον επακριβή καθορισμό της διακύμανσης του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης καθώς και του επιπέδου της χωρικής και χρονικής δραστηριότητας του περιεχομένου (όπως αυτή απεικονίζεται από τη γραφική αναπαράσταση του στιγμιαίου PQoS συναρτήσεως του χρόνου από το QMS εργαλείο), χρησιμοποιήθηκαν ως πειραματικό υπόβαθρο τα τέσσερα βίντεο αναφοράς σύντομης διάρκειας του προηγούμενου κεφαλαίου. Τα τέσσερα βίντεο αναφοράς απεικονίζονται στον πίνακα 7.1 και επιλέχθηκαν έτσι ώστε να είναι αντιπροσωπευτικά, όπως επιδείχθηκε ήδη στο προηγούμενο κεφάλαιο, όλου του εύρους της χωρικής και χρονικής δραστηριότητας του περιεχομένου.

Το επίπεδο της χωρικής και χρονικής δραστηριότητας κάθε βίντεο είναι καθοριστικό στην απόδοση της κωδικοποίησης και της τελικής αντιληπτής ποιότητας, εξαιτίας του γεγονότος ότι οι μέθοδοι κωδικοποίησης αξιοποιούν το χωρικό και χρονικό πλεονασμό για να μπορέσουν να επιτύχουν υψηλή συμπίεση των δεδομένων. Συνεπώς, όσο πιο στατικό χρονικά και ομοιογενές χωρικά είναι το περιεχόμενο ενός βίντεο, τότε ο αντίστοιχος πλεονασμός είναι μεγάλος και ο λόγος της συμπίεσης θα είναι υψηλός. Όπως παρουσιάστηκε στο τρίτο κεφάλαιο, εξαιτίας του γεγονότος ότι τα διαδοχικά καρέ φέρουν σημαντικές ομοιότητες και είναι ισχυρώς συσχετισμένα (χρονικός συσχετισμός), ο κωδικοποιητής προσπαθεί να συμπίεσει τα δεδομένα βίντεο αξιοποιώντας αυτό τον πλεονασμό. Στο χωρικό επίπεδο, ο κωδικοποιητής αξιοποιεί την υψηλή συσχέτιση ανάμεσα στα γειτονικά pixels του ίδιου καρέ (χωρικός συσχετισμός), με αποτέλεσμα τη δημιουργία προβλέψεων των τιμών αυτών σύμφωνα με γειτονικά δείγματα.

Clip 1	Χαμηλό Χωρικό & Χρονικό Περιεχόμενο	Suzie	
Clip 2	Μεσαίο Χωρικό & Χρονικό Περιεχόμενο	Cactus	
Clip 3		Flower Garden	
Clip 4		Mobile & Calendar	
		Υψηλό Χωρικό & Χρονικό Περιεχόμενο	

Πίνακας 7.1 Τα βίντεο αναφοράς της προτεινόμενης μεθόδου πρόβλεψης της ποιότητας

Κατά συνέπεια, θα αξιοποιήσουμε το επίπεδο της χωρικής και χρονικής δραστηριότητας του περιεχομένου ενός βίντεο ως κριτήριο και μετρική για την ποσοτικοποίηση της

δυναμικής του, η οποία επηρεάζει άμεσα τη χωρική και χρονική συσχέτιση αυτού. Βίντεο με στατικό περιεχόμενο (π.χ. συζητήσεις, ομιλίες κλπ) έχουν μικρή χωρική και χρονική δραστηριότητα σε αντίθεση με βίντεο που περιέχουν έντονο περιεχόμενο, με γρήγορες κινήσεις και πολύπλοκες σκηνές (π.χ. αθλητικά, σκηνές δράσεις κλπ), τα οποία αντιστοιχούν σε υψηλού επιπέδου χωρική και χρονική δραστηριότητα. Τα βίντεο αναφοράς που επιλέχθηκαν προς μελέτη καλύπτουν ευρέως το φάσμα της χωρικής και χρονικής δραστηριότητας.

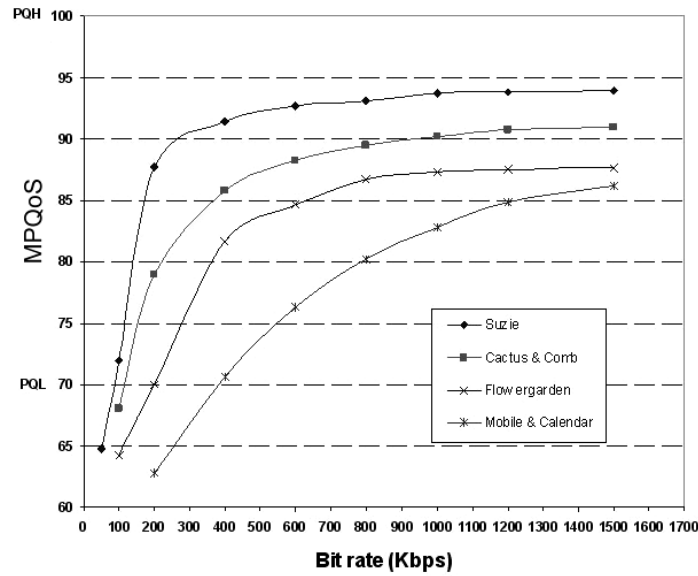
Για τις ανάγκες του παρόντος κεφαλαίου της διατριβής, κάθε βίντεο του πίνακα 7.1, κωδικοποιήθηκε από την αρχική ασυμπίεστη μορφή σε ISO MPEG-4, Simple Profile με τη χρήση διαφορετικών ρυθμών κωδικοποίησης (από 50kbps έως 1.5Mbps για CIF ανάλυση και 20kbps έως 800kbps για QCIF ανάλυση, με περίοδο 1 καρέ τα 100 καρέ και στις δύο περιπτώσεις). Για κάθε ρυθμό κωδικοποίησης, δημιουργήθηκε ένα διαφορετικό αρχείο MPEG-4 με ανάλυση 352x288 (CIF) και 176x144 (QCIF) αντίστοιχα. Ο ρυθμός εναλλαγής των καρέ ορίστηκε στα 25 καρέ/δευτερόλεπτο για όλα τα βίντεο της πειραματικής διαδικασίας.

Κάθε ένα από τα κωδικοποιημένα MPEG-4 βίντεο χρησιμοποιήθηκαν ως σήμα εισόδου στο λογισμικό εκτίμησης της ποιότητας QMS. Από τις παραγόμενες καμπύλες του στιγμιαίου PQoS συναρτήσεως του χρόνου –που εκφράζεται με τη διαδοχική αλληλουχία των καρέ– το MPQoS υπολογίστηκε για κάθε βίντεο. Η πειραματική αυτή διαδικασία επαναλήφθηκε για κάθε βίντεο σε ανάλυση CIF και QCIF αντίστοιχα σε όλους τους υπό εξέταση ρυθμούς.

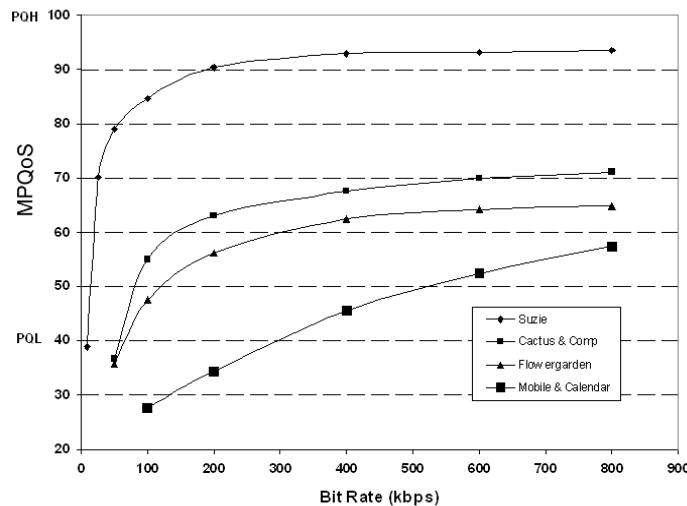
Τα αποτελέσματα αυτών των πειραματικών διεργασιών για τα βίντεο με χωρική ανάλυση CIF απεικονίζονται στο γράφημα 7.3, όπου το PQ_L αντιστοιχεί στην χαμηλότερη αποδεκτή ποιότητα MPQoS, η οποία για βίντεο ανάλυσης CIF και θέαση σε LCD/TFT οθόνη αντιστοιχεί στην ποιότητα 70 της εκατοστιαίας κλίμακας [122]. Το PQ_H υποδηλώνει την καλύτερη εφικτή ποιότητα που κάθε βίντεο μπορεί να επιτύχει. Το γράφημα 7.4 απεικονίζει τα αντίστοιχα πειραματικά αποτελέσματα για τα ίδια με βίντεο, αλλά με χωρική ανάλυση QCIF αυτή τη φορά, όπου το PQ_L αντιστοιχεί με το 40 στην εκατοστιαία κλίμακα, εξαιτίας της χαμηλότερης χωρικής ανάλυσης και της απαίτησης για απεικόνιση των αποτελεσμάτων στην ίδια βαθμολογική κλίμακα. Συνεπώς η χωρική ανάλυση QCIF αντιστοιχεί κατά προσέγγιση σε 40% ποιοτική υποβάθμιση σε σύγκριση με την αντίστοιχη PQ_L τιμή για την περίπτωση της CIF χωρικής ανάλυσης. Συγκρίνοντας τις καμπύλες των δύο γραφημάτων, μπορεί να εξαχθεί ότι χαμηλότερη χωρική ανάλυση (QCIF) οδηγεί σε καμπύλες του MPQoS που μπορούν να συγκλίνουν ταχύτερα και σε μικρότερους ρυθμούς στις μέγιστες PQ_H τιμές τους, οι οποίες όμως συγκλίνουν σε υποβαθμισμένες PQ_H τιμές συγκριτικά με τις αντίστοιχες των υψηλότερων αναλύσεων (CIF).

Σχολιάζοντας τις καμπύλες του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης, όπως απεικονίζονται στα γραφήματα 7.3 και 7.4, τα ακόλουθα συμπεράσματα μπορούν να εξαχθούν:

- Ο ελάχιστος ρυθμός κωδικοποίησης (BR_L) που αντιστοιχεί στο μικρότερο αποδεκτό επίπεδο MPQoS, δηλαδή το PQ_L , εξαρτάται από τη χωρική και χρονική δυναμική του περιεχομένου.
- Η διακύμανση του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης απεικονίζεται μέσω μιας αύξουσας συνάρτησης, η οποία δεν είναι γραμμική.
- Η ποιοτική βελτίωση δεν είναι σημαντική πέραν ενός καθορισμένου κατωφλίου, η τιμή του οποίου είναι άμεσα εξαρτημένη από τη χωρική και χρονική δυναμική του περιεχομένου του βίντεο.



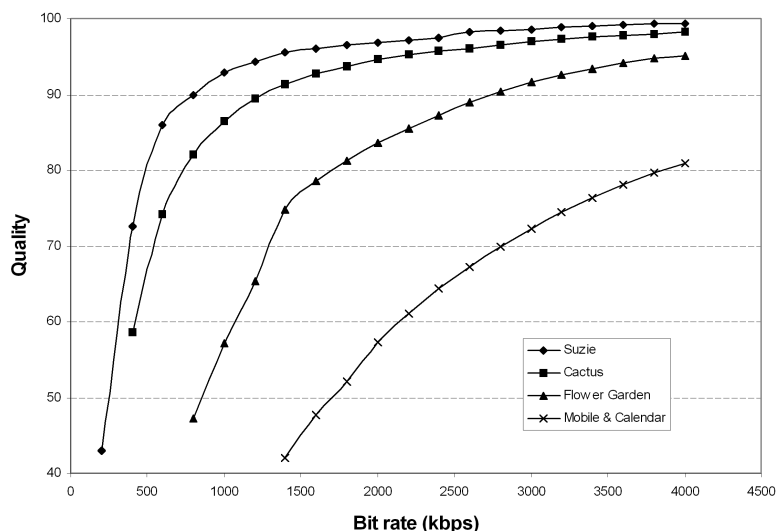
Σχήμα 7.3 MPQoS συναρτήσεσι του ρυθμού κωδικοποίησης για MPEG-4 ASP CIF



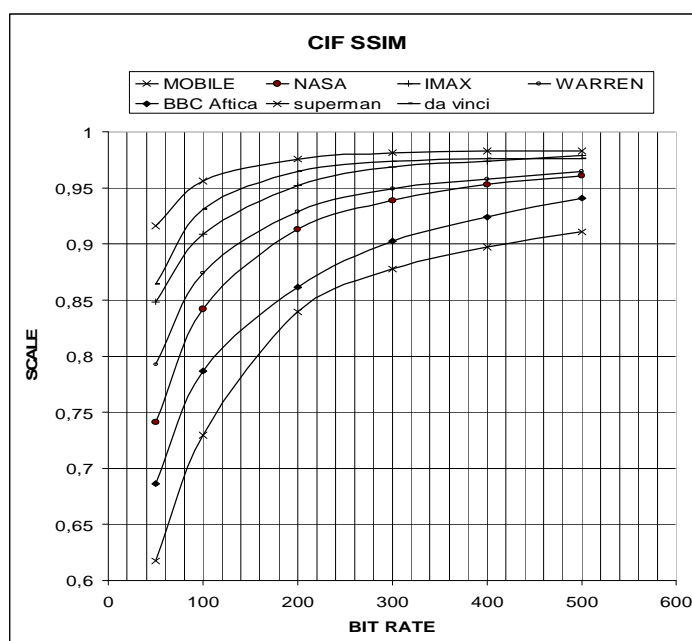
Σχήμα 7.4 MPQoS συναρτήσεσι του ρυθμού κωδικοποίησης για MPEG-4 ASP QCIF

Συγκρίνοντας τις πειραματικές καμπύλες των δύο προαναφερθέντων γραφημάτων με τις αντίστοιχες καμπύλες που προκύπτουν από τις θεωρητικές συναρτήσεις ωφέλειας, όπως αυτές περιγράφονται στην αναφορά [120], ποιοτικές ομοιότητες μεταξύ αυτών μπορούν να παρατηρηθούν. Δηλαδή παρατηρείται εμφανής ποιοτική ομοιότητα μεταξύ των πειραματικών καμπύλων, που προήλθαν από τη χρήση του λογισμικού QMS, και των αντιστοίχων θεωρητικά αναμενόμενων, αποδεικνύοντας την ορθότητα και την εγκυρότητά τους. Απευθείας ποσοτική σύγκριση των πειραματικών και των θεωρητικών καμπυλών δεν είναι εφικτή, διότι οι συναρτήσεις ωφέλειας είναι γενικευμένες προσεγγίσεις της ευχαρίστησης του χρήστη, λαμβάνοντας υπόψιν μια πληθώρα παραμέτρων τόσο στον οριζόντιο, όσο και στον κατακόρυφο άξονα. Απεικονίζοντας την ευχαρίστηση του χρήστη και των πόρων που έχουν αναλωθεί στην υπό εξέταση πολυμεσική εφαρμογή της γενικευμένης συνάρτησης ωφέλειας στο MPQoS και στο ρυθμό κωδικοποίησης αντίστοιχα, οδηγούμαστε στις προτεινόμενες πειραματικές καμπύλες. Συνεπώς, οι πειραματικές

καμπύλες των γραφημάτων 7.3 και 7.4 προσφέρουν μια ποσοτική προσέγγιση και αναπαράσταση, που καθιστούν τις αντίστοιχες θεωρητικές συναρτήσεις ωφελείας πρακτικά χρήσιμες και εμπορικά αξιοποιήσιμες.



Σχήμα 7.5 MPQoS συναρτήσε του ρυθμού κωδικοποίησης για MPEG-2 CIF



Σχήμα 7.6 MPQoS συναρτήσε του ρυθμού κωδικοποίησης για H.264 Baseline CIF

Για περαιτέρω πειραματική επιβεβαίωση της μορφής των καμπύλων της αντιληπτής ποιότητας συναρτήσε του ρυθμού κωδικοποίησης, στο σχήμα 7.5 έχουν υπολογιστεί οι ίδιες καμπύλες για κωδικοποίηση CIF MPEG-2 με χρήση αντικειμενικού μετρικού πλήρους αναφοράς μεταξύ των κωδικοποιημένων βίντεο και των αντιστοιχων βίντεο αναφοράς. Επεκτείνοντας ακόμη περισσότερο τη διαδικασία επιβεβαίωσης της ορθότητας της μορφή των καμπυλών και της ανεξαρτησίας τους από τη μέθοδο κωδικοποίησης και το χρησιμοποιούμενο μετρικό, στο σχήμα 7.6 παρουσιάζονται οι αντίστοιχες καμπύλες

χρησιμοποιώντας βίντεο μη αναφοράς και κωδικοποίηση H.264. Ως μετρικό για τη μέτρηση του επιπέδου τη αντιληπτής ποιότητας χρησιμοποιήθηκε το SSIM [45] που παρέχει αντικειμενική εκτίμηση πλήρους αναφοράς.

Οπότε, συνοψίζοντας τα πειραματικά αποτελέσματα που απεικονίζονται στα σχήματα 7.3, 7.4, 7.5 και 7.6 μπορούμε να αποφανθούμε ότι η γενικευμένη μορφή της καμπύλης της αντιληπτής ποιότητας συναρτήσεως του ρυθμού κωδικοποίησης είναι ανεξάρτητη από το χρησιμοποιούμενο μετρικό, το πρότυπο κωδικοποίησης και τα βίντεο που χρησιμοποιούνται. Όλες οι πειραματικές μετρήσεις οδήγησαν στην ίδια μη γραμμική μορφή, η οποία μετατοπίζεται και μεταβάλλει την κλίση της ανάλογα με τη χωρική και χρονική δυναμική του περιεχομένου. Αξίζει να σημειωθεί, ότι παρά την αλλαγή της πειραματικής διαδικασίας, η σειρά με την οποία μετατοπίζονται η καμπύλες ανάλογα με το περιεχόμενο είναι ακριβώς η ίδια και σύμφωνη με την κατάταξη της χωρικής και χρονικής δυναμικής του περιεχομένου των βίντεο αναφοράς, όπως απεικονίζεται στον καρτεσιανό χώρο του σχήματος 6.2. Οπότε και με αυτή την παρατήρηση, επαληθεύεται ότι η μετατόπιση των καμπυλών μπορεί να χρησιμοποιηθεί ως αντικειμενικό κριτήριο για την κατάταξη του περιεχομένου ενός βίντεο ανάλογα με τη χωρική και χρονική δυναμική του περιεχομένου του.

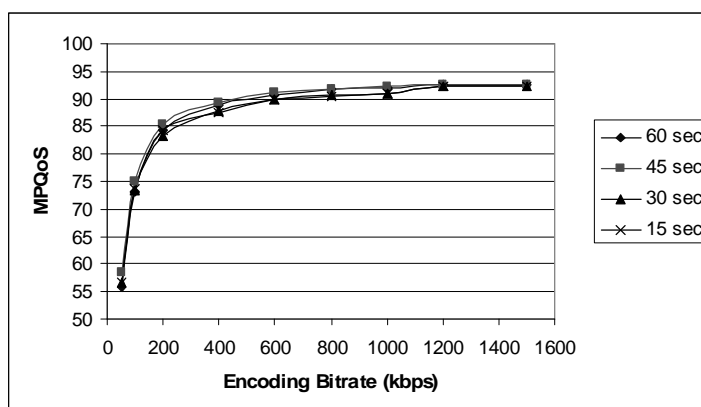
Επιπλέον, είναι ύψιστης σημασίας (με βάση την προαναφερθείσα απεικόνιση) το γεγονός ότι οι αλγεβρικές συναρτήσεις ωφέλειας δεν είναι πανομοιότυπες για όλα τα είδη του περιεχομένου ενός βίντεο, αλλά αποτελούνται από ένα σύνολο καμπυλών, οι οποίες σαν οικογένεια συναρτήσεων ακολουθούν την ίδια γενική μορφή. Η παρατήρηση αυτή προσδίδει μια πολυδιάστατη χροιά στη φύση της συνάρτησης ωφέλειας. Η διαφοροποίηση μεταξύ των καμπυλών –τόσο στην κλίση τους, όσο και στη θέση τους στο επίπεδο ωφέλειας-πόρων, είναι ισχυρά εξαρτημένη από τη χωρική και χρονική δυναμική του περιεχομένου του βίντεο. Συνεπώς, η καμπύλη έχει μικρή κλίση και μετατοπίζεται στην κάτω δεξιά περιοχή του επιπέδου ωφέλειας-πόρων, για περιεχόμενο υψηλής δυναμικής. Αντιθέτως, για περιεχόμενο χαμηλής δυναμικής, η αντίστοιχη καμπύλη μετατοπίζεται στην επάνω αριστερή περιοχή και η κλίση παίρνει υψηλότερες τιμές.

Πρακτικά, η μετατόπιση της καμπύλης στην επάνω αριστερή περιοχή σημαίνει ότι το περιεχόμενο με χαμηλή δυναμική –επί παραδείγματι ένα talk show- αντιστοιχεί για χαμηλότερο ρυθμό κωδικοποίησης σε καλύτερο επίπεδο ποιότητας (MPQoS) σε σύγκριση με περιεχόμενο υψηλής χωρικής και χρονικής δυναμικής. Επί προσθέτως, όταν ο ρυθμός κωδικοποίησης μειωθεί κάτω από ένα συγκεκριμένο κατώφλι, το οποίο είναι άμεσα εξαρτημένο από τη δυναμική του περιεχομένου, το MPQoS (δηλαδή το επίπεδο ωφέλειας) ουσιαστικά καταρρέει.

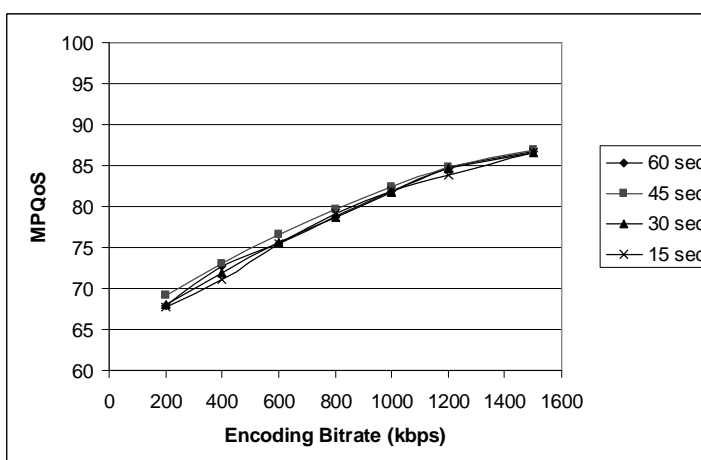
Αντίστοιχα, η μετατόπιση της καμπύλης στην κάτω δεξιά περιοχή του επιπέδου ωφέλειας-πόρων, σημαίνει ότι βίντεο με περιεχόμενο έντονης δυναμικής απαιτεί υψηλότερο ρυθμό κωδικοποίησης για να φτάσει επιτυχώς ένα ικανοποιητικό επίπεδο ποιότητας (MPQoS). Παρόλα αυτά, η αντίστοιχη μέγιστη τιμή MPQoS προσεγγίζεται ομαλότερα αλλά μόνο ασυμπτωτικά σε σύγκριση με την αντίστοιχη περίπτωση περιεχομένου χαμηλής δυναμικής. Επιπλέον, οι υψηλές τιμές της αντιληπτής ποιότητας, ενδέχεται να μην επιτευχθούν από βίντεο με έντονα δυναμικό περιεχόμενο, ακόμη και σε πολύ υψηλούς ρυθμούς κωδικοποίησης.

Σε αυτά τα πλαίσια, πειραματικές καμπύλες του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης υπολογίστηκαν και για ένα σύνολο πραγματικών βίντεο, τα οποία εγράφησαν από τηλεοπτικό σήμα επίγειας ψηφιακής τηλεόρασης, που εκπέμπεται πανελλαδικά, σε μορφή DV (Digital Video)/PAL. Στη συνέχεια τα ψηφιακά αυτά βίντεο

κωδικοποιήθηκαν και συμπίεστηκαν σύμφωνα με το πρότυπο MPEG-4 σε ανάλυση CIF ακολουθώντας την ίδια πειραματική διαδικασία με προηγουμένως. Τα τηλεοπτικά προγράμματα επιλέχθηκαν ώστε να περιέχουν σχετικά ομοιογενές περιεχόμενο, όπως ομιλία, ποδόσφαιρο κλπ., με διάρκεια από 15 έως 60 δευτερόλεπτα. Εκτελώντας πολυάριθμα πειράματα με περιεχόμενο ίδιας κατηγορίας αλλά διαφορετικής διάρκειας, παρατηρήθηκε ότι η μορφή της καμπύλης του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης διατηρεί κοινά χαρακτηριστικά ανεξάρτητα από τη διάρκεια του βίντεο, με μέγιστο σφάλμα απόκλισης 3%. Τα πειραματικά αποτελέσματα αυτής της σύγκρισης απεικονίζονται στα σχήματα 7.7-8, όπου απεικονίζονται τα αντίστοιχα αποτελέσματα των δύο πιο αντιπροσωπευτικών κατηγοριών (Talk Shows και Sports).



Σχήμα 7.7 MPQoS για βίντεο χαμηλής δυναμικής (Talk Show) και κυμαινόμενης διάρκειας.



Σχήμα 7.8 MPQoS για βίντεο υψηλής δυναμικής (Sports) και κυμαινόμενης διάρκειας.

Επιπλέον, σύμφωνα με το επίπεδο της χωρικής και χρονικής δυναμικής του κάθε τηλεοπτικού δείγματος, οι παραγόμενες καμπύλες του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης ακολούθησαν την αντίστοιχη μορφή και κλίση με τις αντίστοιχες καμπύλες αναφοράς του σχήματος 7.3-4. Δηλαδή, τα τηλεοπτικά δείγματα με περιεχόμενο χαμηλής χωρικής και χρονικής δυναμικής (π.χ. Talk Show, Ομιλίες) έδωσαν καμπύλες που έχουν χαρακτηριστικά όμοια με του σήματος αναφοράς Suzie. Αντίστοιχα, τα τηλεοπτικά δείγματα με περιεχόμενο υψηλής δυναμικής (π.χ. ποδόσφαιρο, κολύμβηση) παρήγαγαν καμπύλες όμοιες με την αντίστοιχη του σήματος αναφοράς Mobile and Calendar.

Συνεπώς μπορεί να εξαχθεί το συμπέρασμα ότι το MPQoS ως μετρική παρέχει ικανοποιητικά διακριτές καμπύλες τόσο για βίντεο αναφοράς όσο και για πραγματικά τηλεοπτικά δείγματα, ανεξάρτητα από τη διάρκειά τους. Βεβαίως, τα δείγματα που χρησιμοποιήθηκαν επιλέχθηκαν ώστε να περιέχουν σχετικά ομοιογενές περιεχόμενο. Την περίπτωση βίντεο μεγάλης διάρκειας, του οποίου η δυναμική του περιεχομένου μεταβάλλεται με τη πάροδο του χρόνου, θα την εξετάσουμε ως ξεχωριστή περίπτωση στο επόμενο κεφάλαιο, όπου η προτεινόμενη μέθοδος θα συνδυαστεί με τη χρήση πρωτότυπης μεθόδου ανίχνευσης των ορίων των σκηνών.

Εκθετική προσέγγιση του MPQoS Συναρτήσεως του Ρυθμού Κωδικοποίησης

Παρατηρώντας τις καμπύλες των σχημάτων 7.3 και 7.4, συμπεραίνουμε ότι κάθε καμπύλη μπορεί επιτυχώς να περιγραφεί από τις ακόλουθες τρεις παραμέτρους:

- Την ελαχίστη τιμή του ρυθμού κωδικοποίησης (BR_L) που αντιστοιχεί στο χαμηλότερο αποδεκτό επίπεδο ποιότητας (δηλαδή η τιμή 70 στην εκατοστιαία κλίμακα για ανάλυση CIF).
- Τη μέγιστη τιμή του επιπέδου ποιότητας (PQ_H) που φτάνει ένα βίντεο
- Την παράμετρο α που ορίζει τη μορφή και κατά συνέπεια την κλίση της καμπύλης.

Στην αναφορά [120], προτείνεται ότι τα QoS χαρακτηριστικά μιας συγκεκριμένης πολυμεσικής υπηρεσίας μπορούν επιτυχώς να περιγραφούν από ένα Διάνυσμα Ποιότητας Υπηρεσίας. Συνεπώς τα ποιοτικά χαρακτηριστικά κάθε υπηρεσίας ορίζονται μονοσήμαντα από το διάνυσμα ποιότητας (QoS Vector), το οποίο φέρει τη μορφή $QoS\ Vector = (q_1, q_2, \dots, q_n)$ και μπορεί να χρησιμοποιηθεί για τον καθορισμό των απαραίτητων πόρων που χρειάζονται για την επίτευξη ενός συγκεκριμένου επιπέδου ποιότητας. Προσαρμόζοντας αυτή τη γενική προσέγγιση του διανύσματος ποιότητας στην ανάγκης της παρούσας διατριβής, για υπηρεσίες βίντεο το διάνυσμα ποιότητας (Quality Vector – QV) ορίζεται ως:

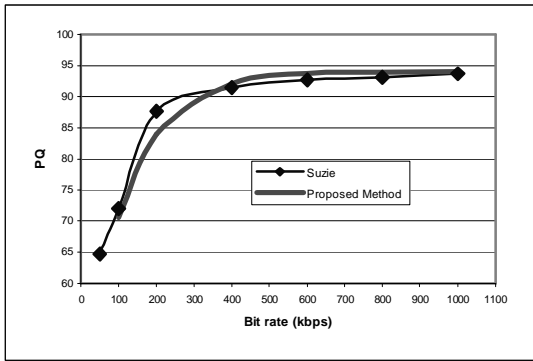
$$QV = (\alpha, BR_L, PQ_H)$$

Οι πειραματικές καμπύλες των σχημάτων 7.3 έως 7.6 μπορούν να προσεγγιστούν επιτυχώς από μια οικογένεια εκθετικών συναρτήσεων. Σύμφωνα με αυτή την πρόταση, το MPQoS ενός βίντεο MPEG-4 κωδικοποιημένο στο ρυθμό BR, μπορεί αναλυτικά να περιγραφεί από την ακόλουθη εξίσωση:

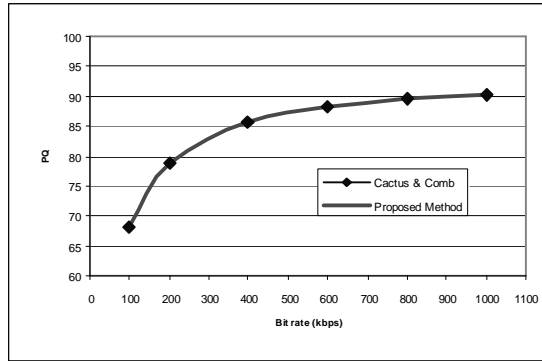
$$MPQoS = [PQ_H - PQ_L] (1 - e^{-\alpha [BR - BR_L]}) + PQ_L, \alpha > 0 \text{ και } BR > BR_L$$

Όπου η παράμετρος α είναι η σταθερά χρόνου της εκθετικής και καθορίζει τη μορφή της καμπύλης.

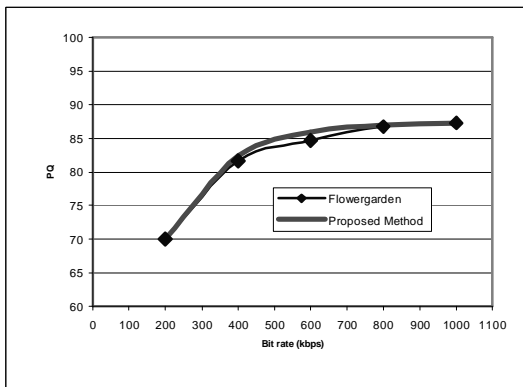
Από τη στιγμή που για όλα τα υπό εξέταση βίντεο το μέγιστο σφάλμα απόκλισης μεταξύ των πειραματικών και εκθετικών MPQoS καμπυλών μετρήθηκε στη χειρότερη περίπτωση να είναι μικρότερο του 4%, μπορεί τότε να εξαχθεί ότι το προτεινόμενο εκθετικό μοντέλο προσεγγίζει ικανοποιητικά τις αντίστοιχες πειραματικές καμπύλες. Τα αντίστοιχα γραφήματα 7.9 και 7.10 απεικονίζονται παρακάτω για την περίπτωση κωδικοποίησης MPEG-4 και MPEG-2 αντίστοιχα.



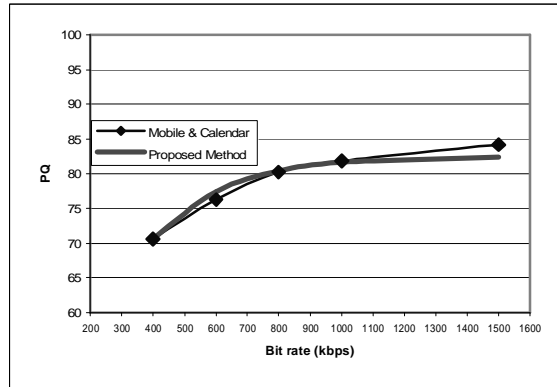
(a) Suzie ($\alpha=0.0083$, $BR_L=95$, $PQ_H^{-1}=93.91^{-1}$)



(b) Cactus ($\alpha=0.0056$, $BR_L=110$, $PQ_H^{-1}=90.89^{-1}$)

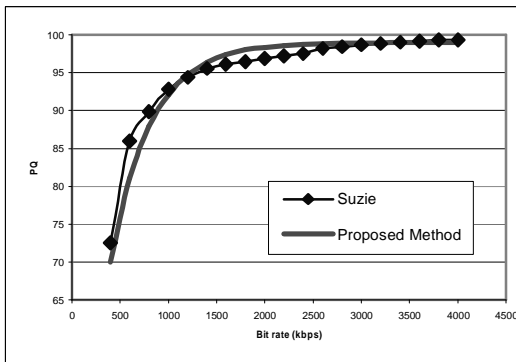


(c) Flower ($\alpha=0.0063$, $BR_L=200$, $PQ_H^{-1}=87.62^{-1}$)

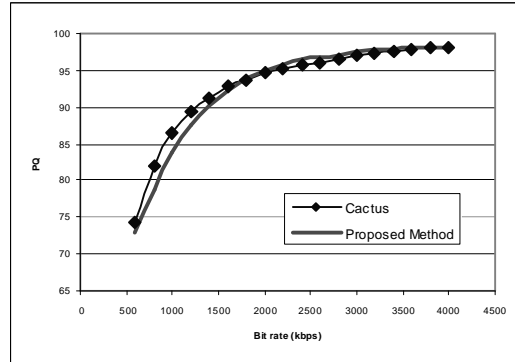


(d) Mobile ($\alpha=0.0045$, $BR_L=400$, $PQ_H^{-1}=86.20^{-1}$)

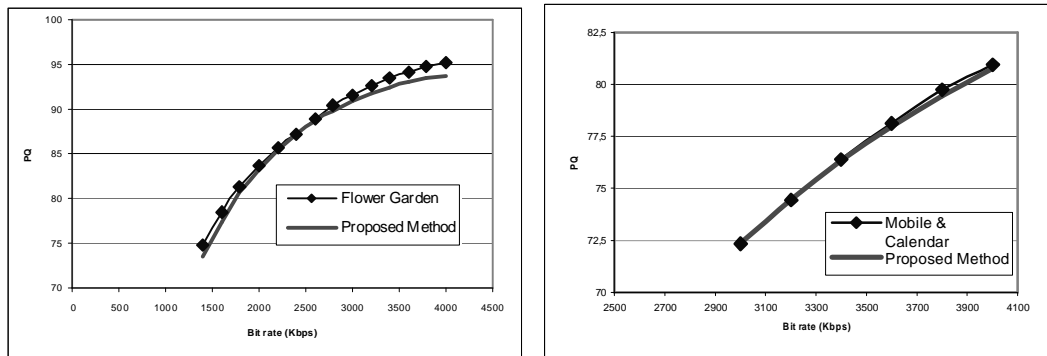
Σχήμα 7.9 Πειραματικές και θεωρητικές καμπύλες MPQoS για MPEG-4/CIF



(a) Suzie ($\alpha=0.024$, $BR_L=400$, $PQ_H^{-1}=99.35^{-1}$)



(b) Cactus ($\alpha=0.015$, $BR_L=550$, $PQ_H^{-1}=98.19^{-1}$)



(c) Flower Garden ($\alpha=0.001$, $BR_L=1250$, $PQ_H^{-1}=95.18^{-1}$) (d) Mobile ($\alpha=0.0006$, $BR_L=2800$, $PQ_H^{-1}=80.95^{-1}$)

Σχήμα 7.10 Πειραματικές και θεωρητικές καμπύλες MPQoS για MPEG-2/CIF

Κατά συνέπεια, κάθε διάνυσμα ποιότητας QV περιέχει τις παραμέτρους που είναι απαραίτητες για την αναλυτική περιγραφή της εξάρτησης του MPQoS από το ρυθμό κωδικοποίησης και δευτερευόντως από τη χωρική ανάλυση, σύμφωνα με το προτεινόμενο εκθετικό μοντέλο.

Οπότε, κάθε πειραματική καμπύλη των γραφημάτων 7.3 έως και 7.6 μπορεί επιτυχώς να προσεγγιστεί από ένα μοναδικό διάνυσμα ποιότητας QV , το οποίο χαρακτηρίζει μοναδικά το συγκεκριμένο βίντεο.

Τα αντίστοιχα διανύσματα ποιότητας απεικονίζονται στον πίνακα 7.2, όπου έχει ληφθεί ως $PQ_L=70$ για χωρική ανάλυση CIF και $PQ_L=40$ για QCIF.

Test Sequence	α	BR_L (Kbps)	PQ_H (Quality Units)
Suzie (MPEG-4 CIF)	0.0083	95	93.91
Cactus (MPEG-4 CIF)	0.0063	110	90.89
Flower (MPEG-4 CIF)	0.0056	200	87.62
Mobile (MPEG-4 CIF)	0.0045	400	86.20
Suzie (MPEG-4 QCIF)	0.013	22	93.50
Cactus (MPEG-4 QCIF)	0.007	55	71.04
Flower (MPEG-4 QCIF)	0.006	65	64.79
Mobile (MPEG-4 QCIF)	0.005	300	57.32

Πίνακας 7.2 Τα στοιχεία των διανυσμάτων ποιότητας για MPEG-4 CIF και QCIF

Πέραν από τα απεικονιζόμενα βίντεο αναφοράς, πειραματικές καμπύλες του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης με τις αντίστοιχες εκθετικές προσεγγίσεις τους συγκρίθηκαν και για τηλεοπτικά δείγματα, διάρκειας 30 δευτερολέπτων, τα οποία κατεγράφησαν σε μορφή DV PAL από τηλεοπτικό πρόγραμμα επίγειας ψηφιακής τηλεόρασης και κωδικοποιήθηκαν ακολουθώντας την ίδια πειραματική διαδικασία που περιγράφηκε στο προηγούμενο εδάφιο. Το περιεχόμενο των τηλεοπτικών δειγμάτων συμπεριλαμβάνει ένα μεγάλο εύρος χωρικής και χρονικής δυναμικής δραστηριότητας (π.χ. ομιλία, αθλητικά). Τα αποτελέσματα έδειξαν ότι το προτεινόμενο εκθετικό μοντέλο προσέγγισε ικανοποιητικά τις αντίστοιχες πειραματικές καμπύλες, με σφάλμα απόκλισης μικρότερο του 4%. Επιπλέον, οι τιμές των στοιχείων των διανυσμάτων ποιότητας ήταν εντός του εύρους των διανυσμάτων αναφοράς που απεικονίζονται στον πίνακα 7.2.

Ταχεία Εκτίμηση των Στοιχείων του Διανύσματος Ποιότητας

Ο ακριβής καθορισμός του ρυθμού κωδικοποίησης που αντιστοιχεί σε ένα καθορισμένο και επιθυμητό επίπεδο MPQoS προσφέρει βέλτιστη αξιοποίηση των αποθηκευτικών μέσων καθώς και του διαθέσιμου εύρους ζώνης κατά τη μετάδοση κωδικοποιημένου βίντεο. Εξαιτίας του γεγονότος ότι ο εκτιμώμενος ρυθμός κωδικοποίησης είναι ακριβώς εκείνος που απαιτείται για την επίτευξη ενός συγκεκριμένου επιπέδου ποιότητας, τότε υπάρχει βέλτιστη διαχείριση των διαθέσιμων πόρων.

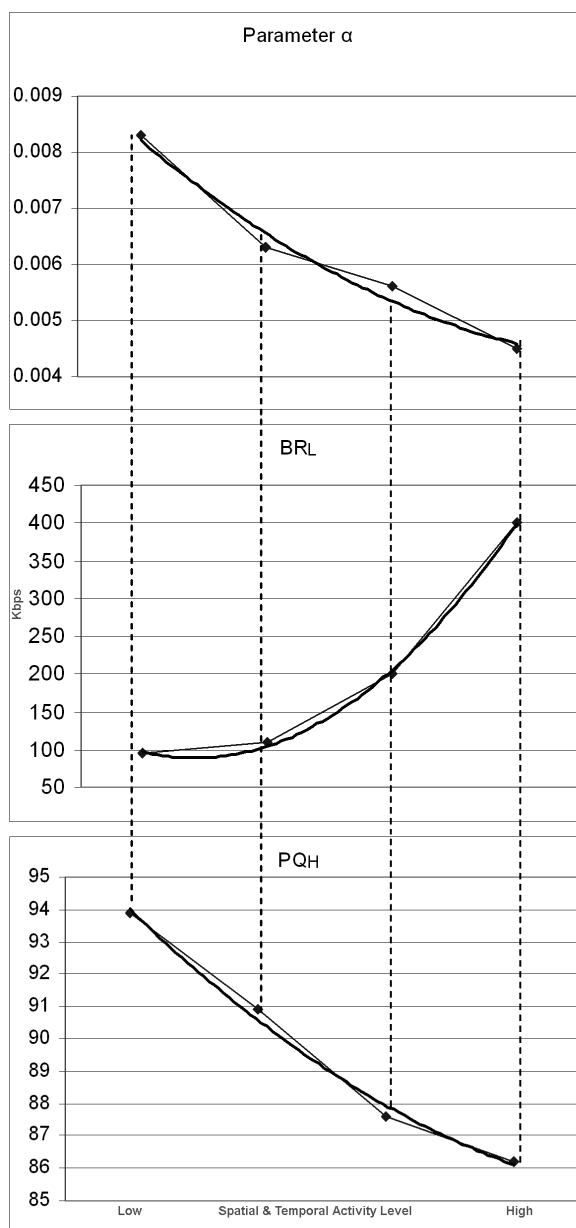
Επιπλέον, η εφαρμογή των μεθόδων για εκτίμηση της διακύμανσης του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης είναι πολύ σημαντικοί για τα δίκτυα κινητών επικοινωνιών, διότι βοηθούν προς την εξάπλωση και εξέλιξη μιας μαζικής αγοράς καταναλωτών πολυμεσικού υλικού, όπου παρέχεται η δυνατότητα στο χρήστη να επιλέξει το επίπεδο της ποιότητας που επιθυμεί να παρακολουθήσει.

Επί της παρούσης για να επιτευχθεί μια τέτοια ποιοτική ταξινόμηση, δεδομένου ενός βίντεο σύντομης διάρκειας, πρέπει πρώτα να κατηγοριοποιηθεί ανάλογα με τη δυναμική του περιεχομένου του. Στη συνέχεια, πρέπει να κωδικοποιηθεί/συμπιεστεί σε συγκεκριμένους ρυθμούς που αντιστοιχούν σε διαφορετικά επίπεδα ποιότητας. Σήμερα, ο καθορισμός των ρυθμών μπορεί να πραγματοποιηθεί μόνο με επαναληπτικές και διαδοχικές κωδικοποιήσεις και εκτιμήσεις της ποιότητας έως ότου εντοπιστούν με ικανοποιητική ακρίβεια οι αντίστοιχοι ρυθμοί. Προφανώς, η διαδικασία αυτή είναι πολύπλοκη, χρονοβόρα, οικονομικά μη συμφέρουσα και πρακτικά μη υλοποιήσιμη. Η παρούσα διατριβή, με την προτεινόμενη εκθετική μοντελοποίηση και το διάνυσμα ποιότητας με τα στοιχεία BR_L , PQ_H και α , παραθέτει μια μέθοδο πρόβλεψης του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης, χωρίς να απαιτούνται επαναλαμβανόμενες κωδικοποιήσεις.

Επιπλέον, δείχνοντας ότι τα στοιχεία του διανύσματος ποιότητας είναι συσχετισμένα τότε η πειραματική εκτίμηση ενός μόνο είναι ικανή για τον καθορισμό των υπολοίπων δύο, που τελικά επιτρέπουν την πρόβλεψη της αντίστοιχης θεωρητικής εκθετικής καμπύλης του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης.

Η συσχέτιση μεταξύ των τριών στοιχείων του διανύσματος ποιότητας μπορεί να εξαχθεί πειραματικά. Θεωρώντας τα τέσσερα βίντεο αναφοράς του πίνακα 7.1, τα οποία καλύπτουν ένα ευρύ και αντιπροσωπευτικό φάσμα χωρικής και χρονικής δυναμικής, η διακύμανσή τους συναρτήσεως της χωρικής και χρονικής δυναμικής απεικονίζεται στον γράφημα 7.11. Η αντιπροσωπευτική περίπτωση που εξετάζεται είναι για κωδικοποίηση σύμφωνα με το

πρότυπο MPEG-4 και χωρική ανάλυση CIF. Προφανώς παρόμοιες καμπύλες μπορούν να εξαχθούν και για άλλες αναλύσεις ή κωδικοποιητές άλλων προτύπων.



Σχήμα 7.11 Πειραματικά προσδιορισμένη διακύμανση των QV στοιχείων συναρτήσει της δυναμικής του περιεχομένου για MPEG-4/CIF

Σύμφωνα με το γράφημα 7.11 είναι προφανές ότι υπάρχει έντονη συσχέτιση μεταξύ των στοιχείων του διανύσματος ποιότητας. Οπότε, εάν ένα από τα τρία στοιχεία υπολογιστεί πειραματικά, τότε τα άλλα δύο μπορούν επακριβώς να καθοριστούν. Μεταξύ των τριών στοιχείων, το PQ_H είναι το καταλληλότερο για να εκτιμηθεί πειραματικά, δεδομένου της εκθετικής μορφής που ακολουθεί η διακύμανση του $MPQoS$ συναρτήσει του ρυθμού κωδικοποίησης, γεγονός που σημαίνει ότι σε υψηλούς ρυθμούς κωδικοποίησης το $MPQoS$ έχει φτάσει στον κόρο. Χρησιμοποιώντας το εργαλείο εκτίμησης της ποιότητας, που για την περίπτωση του MPEG-4 είναι το QMS (αναλύθηκε και εξηγήθηκε σε προηγούμενο εδάφιο),

μία μοναδική εκτίμηση του MPQoS σε υψηλό ρυθμό κωδικοποίησης είναι ικανή για τον ακριβή καθορισμό του PQ_H ενός δεδομένου βίντεο.

Τα βήματα που έπονται είναι απλά: Χρησιμοποιώντας την πειραματικά εκτιμώμενη τιμή PQ_H και τις καμπύλες αναφοράς του σχήματος 7.11, οι αντίστοιχες τιμές του BR_L και α μπορούν να εξαχθούν με γραφικό τρόπο. Συνεπώς, έχοντας εκτιμήσει τα τρία στοιχεία του διανύσματος ποιότητας ενός βίντεο, τότε η αναλυτική εκθετική έκφραση του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης μπορεί να εξαχθεί χρησιμοποιώντας την εξίσωση

$$MPQoS = [PQ_H - PQ_L] (1 - e^{-\alpha [BR - BR_L]}) + PQ_L, \text{ όπου } \alpha > 0 \text{ και } BR > BR_L$$

Για να μπορέσουμε να ακολουθήσουμε μια πιο αναλυτική υπολογιστική μέθοδο, η πειραματική εξάρτηση της παραμέτρου α , του BR_L και του PQ_H από τη χωρική και χρονική δυναμική του περιεχομένου μπορεί επιτυχώς να περιγραφεί με τη χρήση δυναμοσειρών της

πολυωνυμικής μορφής $\sum_{k=0}^{\infty} b_k x^k = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \dots$, όπου τα b_0, b_1, \dots είναι

πραγματικοί συντελεστές. Στα πλαίσια της παρούσης διατριβής, η ανάπτυξη των δυναμοσειρών μέχρι όρους δεύτερης τάξης παρέχει ικανοποιητική ακρίβεια για την θεωρητική προσέγγιση των αντίστοιχων πειραματικών δεδομένων, όπως παρουσιάζεται στο σχήμα 7.11. Επιπλέον, με τον τρόπο αυτό η πολυπλοκότητα των προτεινόμενων εκφράσεων διατηρείται σε χαμηλό επίπεδο, επιτρέποντας την περαιτέρω χρήση και αξιοποίηση αυτών.

$$\left\{ \begin{array}{l} \text{Parameter_}\alpha(x) = \sum_{k=0}^{\infty} b_k x^k = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \dots \approx 0.0103 - 0.0023x + 0.0002x^2 \\ \\ BR_L(x) = \sum_{k=0}^{\infty} c_k x^k = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \dots \approx 181.25 - 130.75x + 46.25x^2 \\ \\ PQ_H(x) = \sum_{k=0}^{\infty} d_k x^k = d_0 + d_1 x + d_2 x^2 + d_3 x^3 + \dots \approx 98.255 - 4.64x + 0.4x^2 \end{array} \right.$$

Όπου η μεταβλητή x σχετίζεται με το επίπεδο της δυναμικής του περιεχομένου

Όπως περιγράφηκε και νωρίτερα, μια μόνο μέτρηση/εκτίμηση του MPQoS με τη χρήση του εργαλείου εκτίμησης της ποιότητας σε υψηλό ρυθμό κωδικοποίησης είναι ικανή για τον ακριβή καθορισμό του PQ_H για ένα συγκεκριμένο βίντεο. Αντικαθιστώντας την εκτιμώμενη τιμή PQ_H στην εξίσωση $98.255 - 4.64x + 0.4x^2$, η αντίστοιχη τιμή x (που είναι ενδεικτική της χωρικής και χρονικής δυναμικής του περιεχομένου) μπορεί να υπολογιστεί επακριβώς με την επίλυση αυτής. Από τις δύο παραγόμενες ρίζες, η μικρότερη θετική περιγράφει επιτυχώς τη συμπεριφορά της δυναμικής του βίντεο και χρησιμοποιείται στις επόμενες δύο εξισώσεις για τον υπολογισμό της παραμέτρου α και του BR_L . Συνεπώς, έχοντας καθορίσει επακριβώς τα τρία στοιχεία (α, BR_L, PQ_H), η αναλυτική εκθετική έκφραση της συνάρτησης του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης μπορεί αναλυτικά πλέον να περιγραφεί χρησιμοποιώντας την εξίσωση $MPQoS = [PQ_H - PQ_L] (1 - e^{-\alpha [BR - BR_L]}) + PQ_L$. Με τον τρόπο αυτό επιτυγχάνεται η πρόβλεψη της καμπύλης του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης χωρίς να απαιτείται η χρήση επαναλαμβανόμενων κωδικοποιήσεων και εκτιμήσεων.

Η μεταβλητή x είναι ισχυρά συσχετισμένη με τη χωρική και χρονική δυναμική του περιεχομένου. Επιπλέον, έχει πειραματικά επαληθευτεί ότι η μεταβλητή x είναι ανάλογη της δραστηριότητας: Από τη διπλή ρίζα της δευτεροβάθμιας εξίσωσης που επιλύουμε για την εύρεση της μεταβλητής x , η μικρότερη από τις δύο είναι ανάλογη ενώ η δεύτερη είναι αντιστρόφως ανάλογη της δυναμικής του περιεχομένου. Οπότε, μόνο η μικρότερη από τις δύο έχει φυσική σημασία και χρησιμοποιείται περαιτέρω στη διαδικασία εκτίμησης της ποιότητας, ώστε να διατηρηθεί συμφωνία μεταξύ της θεωρητικής και πειραματικής συνάρτησης ωφέλειας.

Επί παραδείγματι, ένα τηλεοπτικό βίντεο διάρκειας 25 δευτερολέπτων κωδικοποιείται σύμφωνα με το πρότυπο MPEG-4 (Simple Profile)/CIF στα 2Mbps CBR. Το κωδικοποιημένο βίντεο χρησιμοποιείται ως πηγή στο εργαλείο εκτίμησης της ποιότητας QMS. Η εξαγόμενη καμπύλη του στιγμιαίου PQoS ανά καρέ χρησιμοποιείται για τον υπολογισμό του MPQoS, που εκφράζει τη μέση ποιότητα για όλη τη διάρκεια του βίντεο. Για το βίντεο του παραδείγματος το MPQoS υπολογίστηκε ίσο με 90 στην εκατοστιαία κλίμακα. Δεδομένου ότι τα 2 Mbps είναι υψηλός ρυθμός για κωδικοποίηση MPEG-4/CIF, η τιμή 90 του MPQoS μπορεί να θεωρηθεί ότι προσεγγίζει ικανοποιητικά το PQ_H αφού σε αυτό το ρυθμό η αντιληπτή ποιότητα έχει φτάσει στον κόρο. Στη συνέχεια χρησιμοποιώντας την τιμή αυτή στο σύστημα εξισώσεων των παραμέτρων ποιότητας, υπολογίζουμε τη μεταβλητή x , η οποία είναι ενδεικτική της δυναμικής του περιεχομένου και χρησιμοποιείται για περαιτέρω υπολογισμό του BR_L και της παραμέτρου α . Ακολουθώντας αυτή τη διαδικασία, τα δύο αυτά στοιχεία υπολογίζονται πολύ εύκολα να είναι $BR_L=117$ και $\alpha=0.0062$. Οπότε η εξίσωση που περιγράφει με αναλυτικό τρόπο την εξάρτηση του MPQoS συναρτήσει του ρυθμού κωδικοποίησης για το συγκεκριμένο βίντεο είναι:

$$MPQoS = [90 - 70] (1 - e^{-0.0062 [BR-117]}) + 70, \quad BR > 117$$

Οπότε, εάν ο πάροχος της υπηρεσίας επιθυμεί να προσφέρει το συγκεκριμένο βίντεο στις ποιότητες 70, 80 και 85, τότε χρησιμοποιώντας την παραπάνω εξίσωση είναι σε θέση να υπολογίσει τους ρυθμούς κωδικοποίησης που ικανοποιούν τα συγκεκριμένα επίπεδα ποιότητας. Ο πίνακας 7.3 περιέχει τους αντίστοιχους ρυθμούς κωδικοποίησης για το συγκεκριμένο βίντεο.

MPQoS Level	BR (Kbps)
70	117
80	229
85	341

Πίνακας 7.3 Παράδειγμα προσδιορισμού των ρυθμών κωδικοποίησης

Οπότε είναι προφανές ότι η προτεινόμενη μέθοδος πρόβλεψης των ρυθμών που αντιστοιχούν σε συγκεκριμένα επίπεδα της αντιληπτής ποιότητας ενός κωδικοποιημένου βίντεο είναι γρήγορη, ακριβής, εμπορικά υλοποιήσιμη και αξιοποιήσιμη, με μοναδική απαίτηση τη χρήση ενός μετρικού μη αναφοράς και μια δοκιμαστική κωδικοποίηση σε έναν υψηλό ρυθμό.

Επέκταση της Μεθόδου σε Βίντεο Σύντομης Διάρκειας με Ανομοιογενές Περιεχόμενο

Η μέθοδος πρόβλεψης της αντιληπτής ποιότητας συναρτήσει του ρυθμού κωδικοποίησης και της δυναμικής του περιεχομένου που περιγράφηκε στο προηγούμενο εδάφιο, αξιολογήθηκε σε βίντεο σύντομης διάρκειας ομοιογενούς περιεχομένου, επιδεικνύοντας πολύ ικανοποιητικά επίπεδα ακρίβειας (χείριστο σφάλμα εκτίμησης 4%). Το παρόν εδάφιο θα εξετάσει την επέκταση της μεθόδου σε βίντεο σύντομης διάρκειας με ανομοιογενές περιεχόμενο ως προς τη χρονική και χωρική δυναμική του.

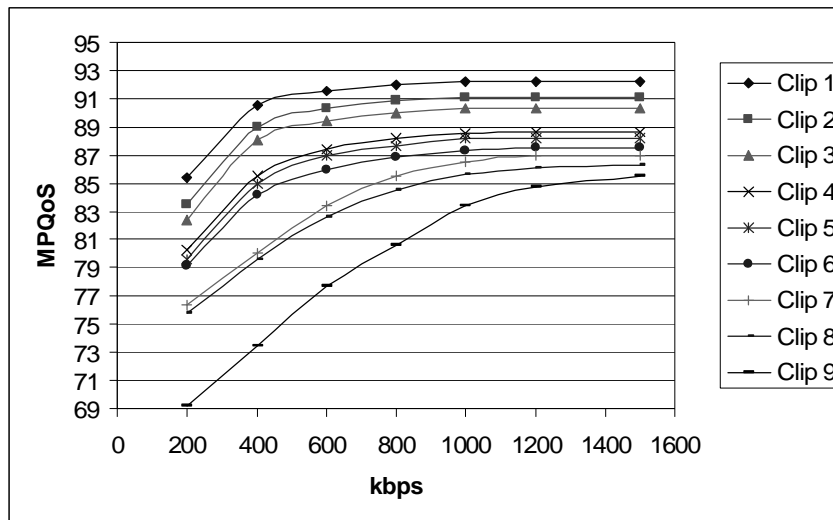
Εξαιτίας του γεγονότος ότι δεν είναι εφικτό να συλλέξουμε δείγματα βίντεο που να καλύπτουν όλα τα επίπεδα της ανομοιογένειας του περιεχομένου, για τις ανάγκες της παρούσας διατριβής χρησιμοποιήθηκαν επεξεργασμένα βίντεο με ελεγχόμενο επίπεδο ανομοιογένειας. Πιο συγκεκριμένα, τα βίντεο αυτά κατασκευάστηκαν χρησιμοποιώντας αποσπάσματα από δύο τηλεοπτικά δείγματα αντίθετης δυναμικής: Ένα talk show και μια έντονη σκηνή από ποδοσφαιρικό αγώνα. Τα ανομοιογενή βίντεο κατασκευάστηκαν χρησιμοποιώντας διαδοχικές σκηνές των δύο βίντεο (talk show και ποδόσφαιρο), δημιουργώντας με τον τρόπο αυτό βίντεο με συγκεκριμένο λόγο υψηλής και χαμηλής δυναμικής του περιεχομένου. Ο πίνακας 7.4 συνοψίζει τα χαρακτηριστικά των βίντεο που παρήχθησαν με αυτή τη διαδικασία:

Όνομα Σήματος	Συνολική Διάρκεια Talk Show (sec)	Συνολική Διάρκεια Football (sec)	Λόγος Talk/Football
Clip 1	120	0	∞
Clip 2	105	15	7.00
Clip 3	90	30	3.00
Clip 4	75	45	1.67
Clip 5	60	60	1.00
Clip 6	45	75	0.60
Clip 7	30	90	0.33
Clip 8	15	105	0.14
Clip 9	0	120	0.00

Πίνακας 7.4 Βίντεο με συγκεκριμένο λόγο υψηλής και χαμηλής δυναμικής του περιεχομένου

Τα δύο βίντεο που χρησιμοποιήθηκαν στη σύνθεση των σημάτων του πίνακα 7.4 ήταν σε μορφή DV PAL, στην οποία μάλιστα πραγματοποιήθηκε και η επεξεργασία τους για τη σύνθεση των ανομοιογενών σημάτων. Στη συνέχεια, τα παραγόμενα ανομοιογενή βίντεο κωδικοποιήθηκαν σύμφωνα με το πρότυπο MPEG-4 Simple Profile σε ποικίλους ρυθμούς, που κυμαίνονται από 200kbps έως 1.5Mbps και μήκος GOP τα 100 καρέ. Για κάθε ρυθμό κωδικοποίησης δημιουργήθηκε ένα αρχείο MPEG-4, ανάλυσης CIF με ρυθμό 25 καρέ το δευτερόλεπτο.

Σε κάθε κωδικοποιημένο βίντεο εκτιμήθηκε η αντιληπτή ποιότητα ανά καρέ με τη χρήση του εργαλείου QMS. Από τα αποτελέσματα της εκτίμησης της ποιότητας ανά καρέ, υπολογίστηκε η μέση ποιότητα για όλη τη διάρκεια του βίντεο, όπως ακριβώς και σε προγενέστερο εδάφιο. Τα αποτελέσματα αυτής της διαδικασίας απεικονίζονται στις γραφικές παραστάσεις του MPQoS συναρτήσει του ρυθμού κωδικοποίησης που φαίνονται στο σχήμα 7.12. Παρατηρείται ότι οι γραφικές αυτές είναι όμοιες με τις αντίστοιχες καμπύλες αναφοράς του σχήματος 7.3, αν κρατήσουμε τις αναλογίες αναφορικά με τη χωρική και χρονική δυναμική του περιεχομένου.



Σχήμα 7.12 Καμπύλες MPQoS για βραχείας διάρκειας βίντεο ετερογενούς περιεχομένου

Στη συνέχεια στο κάθε ένα από τα βίντεο με διαφορετικό επίπεδο δυναμικής του πίνακα 7.5 εφαρμόστηκε η προτεινόμενη τεχνική πρόβλεψης της εκθετικής συνάρτησης του MPQoS συναρτήσε του ρυθμού κωδικοποίησης. Οι θεωρητικές εκθετικές εκτιμήσεις συγκρίθηκαν με τις αντίστοιχες πειραματικές του σχήματος 7.12 και υπολογίστηκε το αντίστοιχο μέσο σφάλμα. Τα αποτελέσματα της σύγκρισης αυτής απεικονίζονται στον πίνακα 7.5 όπου παρατηρούμε ότι το μέγιστο μέσο σφάλμα είναι 4,364%.

Όνομα Σήματος	Μέσο Σφάλμα %
Clip 1	0.434
Clip 2	0.708
Clip 3	0.942
Clip 4	1.884
Clip 5	0.836
Clip 6	1.062
Clip 7	1.934
Clip 8	4.364
Clip 9	0.718

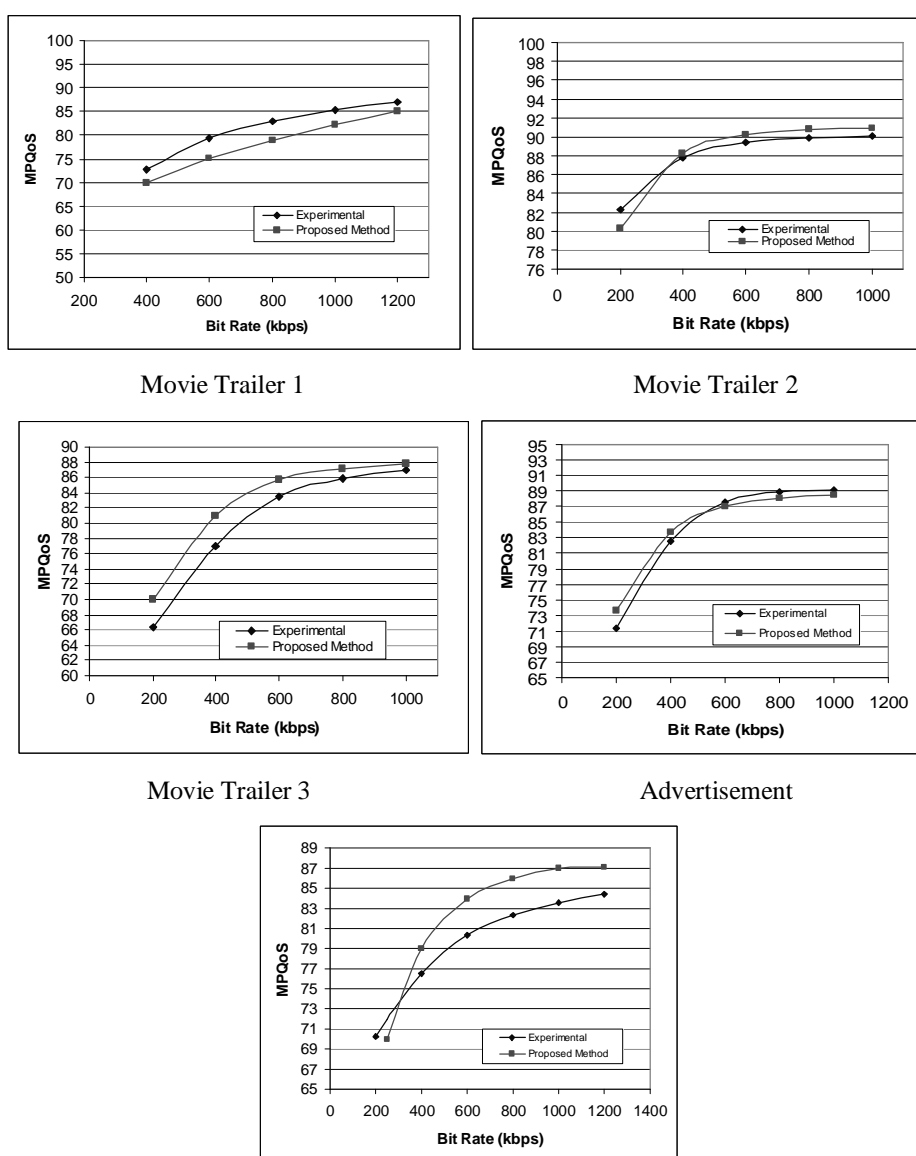
Πίνακας 7.5 Μέσο σφάλμα εκτίμησης για τα ανομοιογενή βίντεο του σχήματος 7.12

Επιπλέον η προτεινόμενη τεχνική δοκιμάστηκε και σε βίντεο μη αναφοράς, αλλά σύντομης διάρκειας, όπως trailer ταινιών, μουσικά βίντεο και διαφημίσεις. Τα χαρακτηριστικά των βίντεο που χρησιμοποιήθηκαν για αυτή τη διαδικασία φαίνονται στον πίνακα 7.6 που ακολουθεί.

Όνομα Σήματος	Διάρκεια
Movie Trailer 1 (action)	1 min 42 sec
Movie Trailer 2 (comic)	1 min 40 sec
Movie Trailer 3 (comedy)	1 min 37 sec
Advertisement	15 sec
Jazz video clip	2 min 10 sec

Πίνακας 7.6 Χαρακτηριστικά των βραχείας διάρκειας βίντεο ετερογενούς περιεχομένου

Τα βίντεο αυτά καταγράφηκαν σε μορφή DV PAL μέσω του προγράμματος της επίγειας ψηφιακής τηλεόρασης που εκπέμπεται στην επικράτεια. Στη συνέχεια κωδικοποιήθηκαν σύμφωνα με το πρότυπο MPEG4 Simple Profile/CIF σε διαφορετικούς ρυθμούς που ποικίλουν από 200kbps μέχρι 1,2Mbps και μήκος GOP ίσο με 100 καρέ. Για κάθε ρυθμό κωδικοποίησης ένα αρχείο MPEG-4 με CIF ανάλυση δημιουργήθηκε. Σε όλες τις περιπτώσεις κατά τη διαδικασία κωδικοποίησης επιλέχθηκε ο ρυθμός των καρέ ανά δευτερόλεπτο να είναι 25 για όλα τα βίντεο. Ακολουθώντας την ίδια ακριβώς διαδικασία, όπως και στο προηγούμενο εδάφιο, υπολογίστηκαν οι πειραματικές καμπύλες του MPQoS συναρτήσεως του ρυθμού κωδικοποίησης. Στη συνέχεια, εφαρμόζοντας την προτεινόμενη μέθοδο οι αντίστοιχες θεωρητικές εκθετικές καμπύλες υπολογίστηκαν και συγκρίθηκαν με τις αντίστοιχες πειραματικές, με σφάλμα ίσο και μικρότερο του 4,364%. Οι αντίστοιχες πειραματικές και θεωρητικές καμπύλες ανά βίντεο φαίνονται στο σχήμα 7.13.



Movie Trailer 1

Movie Trailer 2

Movie Trailer 3

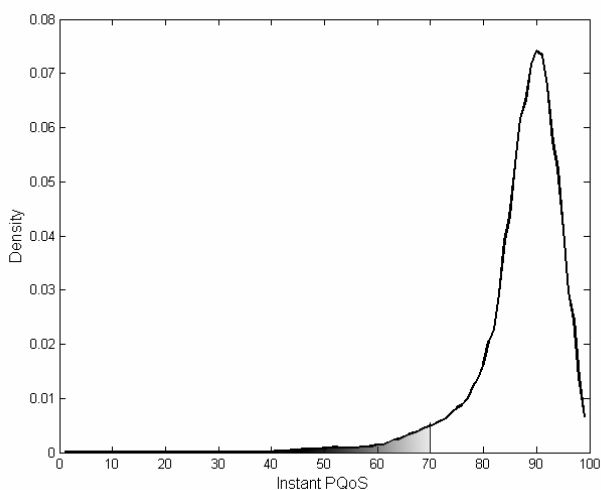
Advertisement

Jazz Video Clip

Σχήμα 7.13 Πειραματικές και θεωρητικές καμπύλες MPQoS επιδεικνύοντας την αποδοτικότητα της προτεινόμενης μεθόδου

Για να εξετάσουμε τη διακύμανση του στιγμιαίου PQoS στα βίντεο ετερογενούς περιεχομένου, η συνάρτηση πυκνότητας πιθανότητας (Probability Density Function –PDF) υπολογίστηκε για όλα τα βίντεο του πίνακα 7.5 και 7.6. Η παραγόμενη συνάρτηση πυκνότητας πιθανότητας απεικονίζεται στο σχήμα 7.14, όπου παρατηρείται ότι οι τιμές του στιγμιαίου PQoS παρουσιάζουν μεγάλη συγκέντρωση γύρω από την τιμή του MPQoS που είναι το 86,93 με σφάλμα τυπικής απόκλισης ίσο με 8,7. Συνεπώς, η πιθανότητα να έχουμε μη αποδεκτό επίπεδο ποιότητας (δηλαδή τιμές στιγμιαίου PQoS κάτω από 70 – σκιασμένη περιοχή στο γράφημα, είναι κατά προσέγγιση ίση με 0,0375. Οπότε, η διακύμανση του στιγμιαίου PQoS γύρω από το MPQoS δεν επηρεάζει σημαντικά την ακρίβεια της μετρικής ικανότητας του MPQoS.

Παρόλα αυτά πρέπει εντόνως να τονιστεί ότι η προτεινόμενη μέθοδος χρησιμοποιείται για την πρόβλεψη ποιοτικών επιπέδων που είναι εν γένει αποδεκτά από το μέσο θεατή. Η εξάρτηση του MPQoS για πολύ χαμηλούς ρυθμούς κωδικοποίησης δεν εξετάζεται διότι δεν έχει πρακτικό ενδιαφέρον, αφού το παραγόμενο βίντεο έχει έντονα προβλήματα κατά την αναπαραγωγή του και κατά συνέπεια μη αποδεκτή ποιότητα.



Σχήμα 7.14 Συνάρτηση Πυκνότητας Πιθανότητας του στιγμιαίου PQoS

Συμπεράσματα του Κεφαλαίου

Τα υπάρχοντα λογισμικά ή υλικά συστήματα εκτίμησης της ποιότητας κωδικοποιημένου βίντεο παρέχουν μέτρηση της στιγμιαίας ποιότητας συναρτήσει του χρόνου. Σε αυτό το κεφάλαιο, ο μέσος όρος της αντιληπτής ποιότητας (το MPQoS) εισάγεται ως μετρικό για τη μέτρηση της μέσης αντιληπτής ποιότητας όλης τη διάρκεια του βίντεο. Πειραματικές καμπύλες του MPQoS συναρτήσει του ρυθμού κωδικοποίησης συγκρίθηκαν με τις αντίστοιχες θεωρητικές συναρτήσεις ωφελείας, δείχνοντας έντονη ομοιότητα στη μορφή τους και επιβεβαιώνοντας την ορθότητα των πειραματικών μετρήσεων. Επιπλέον, η αντιστοίχιση της ευχαρίστησης του χρήστη και των χρησιμοποιούμενων πόρων, της θεωρητικής συνάρτησης ωφέλειας με την μέση αντιληπτή ποιότητα (MPQoS) και το ρυθμό κωδικοποίησης αντίστοιχα, αποκαλύπτει την πολυδιάστατη φύση της συνάρτησης ωφέλειας, αφού δεν είναι μοναδική αλλά αντιθέτως ευμετάβλητη και έντονα εξαρτημένη από τη δυναμική του περιεχομένου του βίντεο.

Επιπλέον, παρουσιάστηκε ότι οι πειραματικές καμπύλες του MPQoS συναρτήσει του ρυθμού κωδικοποίησης μπορούν επιτυχώς να προσεγγιστούν από μια οικογένεια εκθετικών συναρτήσεων, με μέγιστο σφάλμα ταύτισης μικρότερο του 4%. Με την προσέγγιση αυτή καθίσταται πλέον δυνατή η αναλυτική περιγραφή του MPQoS συναρτήσει του ρυθμού κωδικοποίησης. Βασισμένη σε αυτή την προσέγγιση, αναπτύσσεται και προτείνεται πρωτότυπη μέθοδος για ταχεία πρόβλεψη της συνάρτησης του MPQoS ως προς τον ρυθμό κωδικοποίησης, ανάλογα με τη χωρική και χρονική δυναμική του περιεχομένου. Η πρωτότυπη αυτή μέθοδος επιτρέπει τη βελτιστοποίηση των διαθέσιμων πόρων κωδικοποίησης, αφού καθορίζεται επακριβώς ανάλογα με το περιεχόμενο του βίντεο, ο ρυθμός που ικανοποιεί ένα συγκεκριμένο επίπεδο ποιότητας.

Στη συνέχεια πραγματοποιήθηκαν πειράματα τόσο με βίντεο αναφοράς, όσο και με πραγματικά βίντεο που καταγράφηκαν από τηλεοπτικά προγράμματα ώστε να επιδειχθεί η ακρίβεια της μεθόδου για βίντεο σύντομης διάρκειας με ομοιογενές και ετερογενές περιεχόμενο. Συνεπώς, στο παρόν κεφάλαιο έχει καλυφθεί πλήρως η εξέταση της ορθότητας της προτεινόμενης μεθόδου για βίντεο σύντομης διάρκειας. Στο επόμενο κεφάλαιο θα εξεταστεί η επέκταση της μεθόδου για βίντεο μεγάλης διάρκειας.

ΚΕΦΑΛΑΙΟ 8

Επέκταση της Προτεινόμενης Μεθόδου Πρόβλεψης της Αντιληπτής Ποιότητας σε Βίντεο Μακράς Διάρκειας

Απαιτήσεις για Επέκταση της Προτεινόμενης Μεθόδου

Στο προηγούμενο κεφάλαιο περιγράφηκε, παρουσιάστηκε και δοκιμάστηκε πρωτότυπη μέθοδος για την πρόβλεψη της ποιότητας ψηφιακού βίντεο κωδικοποιημένου σύμφωνα με το πρότυπο MPEG-4, συναρτήσει του ρυθμού κωδικοποίησης και της δυναμικής του περιεχομένου. Η ακρίβεια της μεθόδου δοκιμάστηκε επιτυχώς σε βίντεο σύντομης διάρκειας ομοιογενούς και ετερογενούς περιεχομένου. Στο παρόν κεφάλαιο θα μελετηθεί η επέκταση της μεθόδου σε βίντεο μεγάλης διάρκειας, όπου το έντονα ετερογενές περιεχόμενο δεν επιτρέπει τη χρήση της μέσης ποιότητας (MPQoS) για όλη τη διάρκεια του βίντεο. Δεδομένου ότι η συνάρτηση του MPQoS συναρτήσει της δυναμικής του περιεχομένου διαφοροποιείται σημαντικά, τότε η χρήση του MPQoS θα οδηγήσει σε ανακριβή αποτελέσματα.

Για να επεκταθεί η τεχνική πρόβλεψης σε βίντεο μεγάλης διάρκειας θα πρέπει η προτεινόμενη μέθοδος να συνδυαστεί με τεχνική ανίχνευσης των ορίων των σκηνών. Με τον τρόπο αυτό, το βίντεο μεγάλης διάρκειας θα αναλυθεί σε επιμέρους ομοιογενή αποσπάσματα μικρής διάρκειας, στα οποία η προτεινόμενη μέθοδος θα εφαρμοστεί μεμονωμένα και ανεξάρτητα με ικανοποιητική ακρίβεια, όπως επιδείχθηκε στο προηγούμενο κεφάλαιο.

Προτού προβούμε σε λεπτομερή παρουσίαση της επεκταμένης μεθόδου, για τις ανάγκες της παρούσας διατριβής διερευνήθηκε και αναπτύχθηκε μέθοδος ανίχνευσης των ορίων των σκηνών, η οποία μπορεί να συνδυαστεί με την υπάρχουσα μέθοδο πρόβλεψης της ποιότητας, χωρίς να απαιτείται κάποια επιπλέον διαδικασία. Συνεπώς, η μέθοδος θα πρέπει να εφαρμόζεται στο αποκωδικοποιημένο βίντεο και να αξιοποιεί τα τεχνουργήματα που προέρχονται από την κωδικοποίηση MPEG. Λαμβάνοντας υπόψιν το γεγονός ότι για την εκτίμηση της ποιότητας αξιοποιήθηκε η ένταση του φαινομένου τετραγώνισης στο αποκωδικοποιημένο σήμα MPEG-4, προφανώς η προτεινόμενη τεχνική ανίχνευσης των ορίων των σκηνών θα πρέπει να ενσωματώνει στους μηχανισμούς της, την ανίχνευση κάποιου τεχνουργήματος που προκύπτει από τη διαδικασία κωδικοποίησης.

Σε αυτά τα πλαίσια, στο επόμενο εδάφιο παρουσιάζεται πρωτότυπη τεχνική ανίχνευσης των ορίων των σκηνών ενός κωδικοποιημένου σήματος MPEG, η οποία αναπτύχθηκε για τις ανάγκες της παρούσας διατριβής. Η κατάτμηση ενός βίντεο μεγάλης διάρκειας στις σκηνές από τις οποίες δομείται είναι ένα από τα κυρίως θέματα στην ερευνητική περιοχή της επεξεργασίας ψηφιακού βίντεο, όπου πολυάριθμες μέθοδοι έχουν προταθεί, αξιοποιώντας μετρικά που εφαρμόζονται είτε στο συμπιεσμένο ή απευθείας στο ασυμπιεστο σήμα, συγκρίνοντας τα διαδοχικά καρέ. Στην πλειοψηφία τους οι υπάρχουσες τεχνικές για κωδικοποιημένο βίντεο είναι αποτελεσματικές σε υψηλούς ρυθμούς κωδικοποίησης (>1Mbps) [123], ενώ η ακρίβεια τους ελαττώνεται για μικρότερους ρυθμούς (π.χ. @256kbps), οι οποίοι όμως είναι ευρέως αξιοποιήσιμοι στο πρότυπο MPEG-4 [124].

Για το λόγο αυτό παρουσιάζεται στον παρόν κεφάλαιο πρωτότυπη μέθοδος για την ανίχνευση των σκηνών σε βίντεο κωδικοποιημένο σύμφωνα με πρότυπο βασισμένο στο

μετασχηματισμό διακριτού συνημίτονου (DCT) σε χαμηλούς ρυθμούς. Πιο συγκεκριμένα αξιοποιείται η μετρική της έντασης του αντιληπτού τεχνουργήματος του φαινομένου τετραγώνισης σε κάθε καρέ του αποκωδικοποιημένου βίντεο. Με πειραματικές δοκιμές που θα παρουσιαστούν στον παρόν κεφάλαιο, αξιολογείται η προτεινόμενη τεχνική και η αποδοτικότητα αυτής έναντι των υπαρχόντων. Τέλος περιγράφεται η αξιοποίηση αυτής στα πλαίσια πρόβλεψης της αντιληπτής ποιότητας για βίντεο μεγάλης διάρκειας.

Τεχνική Κατάτμησης των Σκηνών σε DCT Βίντεο Χαμηλού Ρυθμού

Η εννοιολογική και οντολογική επεξεργασία οποιουδήποτε σήματος βίντεο απαιτεί την οργάνωση και καταχώρηση του βίντεο σε στοιχειώδεις σκηνές, όπου η κάθε μία αποτελείται από ένα πλήθος διαδοχικών καρέ παρόμοιου περιεχομένου (που έχει καταγραφεί συνεχώς με χρήση της ίδιας κάμερας). Επιπλέον, η δυναμική του περιεχομένου μιας σκηνής τόσο στη χωρική όσο και στη χρονική συνιστώσα είναι ομοιογενής αφού αποτελείται από καρέ παρόμοιου περιεχομένου.

Συνεπώς, η εφαρμογή της μεθόδου πρόβλεψης της μέσης ποιότητας σε μια σκηνή θα μας δώσει άριστα αποτελέσματα, εξαιτίας της χωρικής και χρονικής ομοιογένειας του περιεχομένου εντός μιας σκηνής. Αυτή η διεργασία κατάτμησης σε στοιχειώδεις σκηνές μπορεί να εφαρμοστεί είτε σε απότομες αλλαγές σκηνών, όπου η μετάπτωση από το τέλος μιας σκηνής στην αρχή της επόμενης συμβαίνει απότομα μεταξύ δύο διαδοχικών καρέ, ή σε προοδευτικές αλλαγές σκηνών, όπου η μετάπτωση πραγματοποιείται σταδιακά σε ένα πλήθος καρέ, στα οποία οι δύο γειτονικές σκηνές βαθμιαία αλλάζουν.

Λαμβάνοντας υπόψιν ότι οι υπάρχουσες τεχνικές ανίχνευσης απαιτούν

- Πολύπλοκους μηχανισμούς ανίχνευσης, στους οποίους πρέπει να προηγηθεί διαδικασία προετοιμασίας του συστήματος ανάλογα με το περιεχόμενο που θα κατανεμηθεί (γεγονός που δεν επιτρέπει τη χρήση ετερογενών περιεχομένων μεγάλης διάρκειας) [125]
- Τη χρήση υψηλών ρυθμών κωδικοποίησης για να επιτύχουν υψηλά ποσοστά επιτυχούς ανίχνευσης και αποδοτικότητας (γεγονός που αυξάνει σημαντικά την πολυπλοκότητα της τεχνικής) [126]

είναι προφανής η ανάγκη για ανάπτυξη πρωτότυπης τεχνικής κατάτμησης των σκηνών, η οποία να δύναται να εφαρμοστεί σε βίντεο μεγάλης διάρκειας χωρίς να απαιτεί υψηλή κατανάλωση πόρων από το σύστημα τόσο για την κωδικοποίηση όσο και για τη προσαρμογή του κατωφλίου αναφοράς στη δυναμική του περιεχομένου που τίθεται υπό επεξεργασία. Επιπλέον, η προτεινόμενη τεχνική πρέπει να έχει υψηλή απόδοση και ακρίβεια στα βίντεο χαμηλού ρυθμού, γεγονός που αποτελεί την αχίλλειο φτέρνα της πλειοψηφίας των υπαρχόντων τεχνικών. Επίσης, στα πλαίσια επέκτασης της μεθόδου πρόβλεψης, η τεχνική κατάτμησης που θα αξιοποιηθεί για το σκοπό αυτό, θα πρέπει να χρησιμοποιεί τεχνικές ποσοτικοποίησης, οι οποίες ήδη χρησιμοποιούνται στον αλγόριθμο πρόβλεψης της ποιότητας (π.χ. Εκτιμητής του φαινομένου τετραγώνισης). Έτσι θα είναι δυνατή η αξιοποίηση της τεχνικής κατάτμησης συνδυαστικά με την μέθοδο πρόβλεψης της ποιότητας που περιγράφηκε στο προηγούμενο κεφάλαιο, χωρίς να απαιτείται η ανάπτυξη επιπλέον διεργασιών ή διεπαφών.

Σε αυτά τα πλαίσια, στο επόμενο εδάφιο εισάγεται πρωτότυπη τεχνική κατάτμησης και ανίχνευσης των στοιχειωδών σκηνών ενός DCT-βίντεο χαμηλού ρυθμού, η οποία αξιοποιεί

το αντιληπτό μετρικό της έντασης του φαινομένου τετραγώνισης (blockiness) στο αποκωδικοποιημένο σήμα. Πιο συγκεκριμένα, ενώ η γενικότερη τάση για βελτίωση της απόδοσης των υπάρχοντων μεθόδων είναι η ενσωμάτωση στον μηχανισμό εκτίμησης περισσότερων γειτονικών καρέ, η προτεινόμενη τεχνική εστιάζει στη μέτρηση της αντιληπτής ποιότητας μέσω της έντασης του φαινομένου τετραγώνισης εντός κάθε καρέ, τεχνοτροπία που ήδη παρουσιάστηκε στην προτεινόμενη μέθοδο πρόβλεψης της ποιότητας του προηγούμενου κεφαλαίου. Συνεπώς, είναι προφανής η συμβατότητα της προτεινόμενης τεχνικής με τις απαιτήσεις που τέθηκαν, γεγονός που επιτρέπει την ενσωμάτωσή της εντός του μηχανισμού της τεχνικής πρόβλεψης της ποιότητας.

Το υπόλοιπο του παρόντος κεφαλαίου δομείται ως εξής: Στο επόμενο εδάφιο πραγματοποιείται μια σύντομη βιβλιογραφική ανασκόπηση στις τεχνικές κατάτμησης και ανίχνευσης των στοιχειωδών σκηνών, μέσω της οποίας διαφαίνεται η πρωτοτυπία της προτεινόμενης τεχνικής. Στη συνέχεια η προτεινόμενη τεχνική παρουσιάζεται και η αξιοπιστία της δοκιμάζεται και συγκρίνεται με ήδη υπάρχουσες τεχνικές. Έπειτα, η προτεινόμενη τεχνική συνδυάζεται με την μέθοδο πρόβλεψης της ποιότητας του προηγούμενου κεφαλαίου, για την επέκταση αυτής σε ετερογενή βίντεο μεγάλης διάρκειας. Εν κατακλείδι, στο τελευταίο εδάφιο παρουσιάζονται και σχολιάζονται τα συμπεράσματα του παρόντος κεφαλαίου.

Σύντομη Βιβλιογραφική Ανασκόπηση των Τεχνικών Κατάτμησης Σκηνών

Οι ποικίλες τεχνικές ανίχνευσης και κατάτμησης των στοιχειωδών σκηνών ενός βίντεο μπορούν να ταξινομηθούν ανάλογα με το εάν η εφαρμογή τους περιορίζεται σε κωδικοποιημένο σήμα ή σε ασυμπιεστο [123], [127], [128], [129]. Παρόλα αυτά, ανεξάρτητα από τη φύση του βίντεο (συμπιεσμένο ή ασυμπιεστο) που θα χρησιμοποιηθεί, κάθε μέθοδος κατάτμησης απαιτεί δύο συναρτήσεις για να παρέχει αξιόπιστα αποτελέσματα: Τη συνάρτηση απόφασης και τη συνάρτηση κατωφλίου. Η πρώτη καθορίζει τον αλγόριθμο σύμφωνα με τον οποίο η τεχνική ορίζει ως ανιχνεύσιμη αλλαγή σκηνής ένα χρονικό σημείο του βίντεο, ενώ η δεύτερη καθορίζει –συνήθως με δυναμικό τρόπο- την παράμετρο κατωφλίου που θα αξιοποιηθεί από τη συνάρτηση απόφασης.

Η συντριπτική πλειοψηφία των τεχνικών κατάτμησης, που εφαρμόζονται σε ασυμπιεστα βίντεο, είναι βασισμένες στη σύγκριση της μορφής των ιστογραμμάτων χρώματος και φωτεινότητας μεταξύ διαδοχικών καρέ. Βασισμένοι στην πειραματική διαπίστωση ότι όλα τα καρέ που ανήκουν σε μια κοινή σκηνή χαρακτηρίζονται από παρόμοια ιστογράμματα, τότε ανίχνευση μεταβολής σε ιστόγραμμα συνεπάγεται πιθανή αλλαγή σκηνής [130]. Μια άλλη ομάδα τεχνικών κατάτμησης για ασυμπιεστα βίντεο εστιάζουν στην ταξινόμηση των καρέ, αξιοποιώντας μεταβολές στις τιμές των pixels μεταξύ γειτονικών καρέ [131]. Παρομοίως, άλλες τεχνικές στηρίζονται στη χρήση φίλτρων ανίχνευσης ακμών και σύγκρισης αυτών μεταξύ των γειτονικών καρέ [132].

Σχετικά με τις τεχνικές κατάτμησης που εξειδικεύονται σε συμπιεσμένα/κωδικοποιημένα βίντεο υπάρχουν ποικίλες προσεγγίσεις, οι οποίες αξιοποιούν τα ειδικά χαρακτηριστικά του κάθε προτύπου κωδικοποίησης. Συνεπώς, συναντάμε τεχνικές που βασίζονται είτε μεμονωμένα ή συνδυαστικά στο διακριτό μετασχηματισμό συνημίτονου, στους DC όρους των διανυσμάτων κίνησης και στο πλήθος των Intra-κωδικοποιημένων MacroBlocks [133], [134].

Επιπλέον, κάποιες τεχνικές κατάτμησης εφαρμόζονται απευθείας στη ροή μεταβλητού ρυθμού (variable bit rate coding stream) των κωδικοποιημένων βίντεο, αξιοποιώντας τη μεταβολή του ρυθμού κωδικοποίησης συναρτήσει των σκηνών. Ο Krunz στο [135] και [136] χρησιμοποίησε τη μεταβολή στο μέγεθος των διαδοχικών I καρέ για να κατατμήσει τις σκηνές. Ο Melamed και ο Pendakis στο [137] εισήγαγαν μια συνάρτηση ένδειξης λαμβάνοντας υπόψιν την ελάχιστη διάρκεια σκηνής. Επίσης, ο Yoo στο [138] πρότεινε μια μέθοδο αναγνώρισης των ορίων των σκηνών.

Ωστόσο, όλες οι προαναφερθείσες τεχνικές για να μπορέσουν να εντοπίσουν τις μεταβολές στο μετρικό μέγεθος που χρησιμοποιούν είναι απαραίτητο να χρησιμοποιήσουν ένα δυναμικό κατώφλι αναφοράς. Συνεπώς, ένα κοινό πρόβλημα στις υπάρχουσες τεχνικές κατάτμησης είναι η επιλογή του κατάλληλου βέλτιστου κατωφλίου, ώστε να ανιχνευτούν επιτυχώς τα όρια των σκηνών [139]. Εάν επιλεγεί η χρήση ενός καθολικού κατωφλίου για την κατάτμηση των σκηνών ενός βίντεο μεγάλης διάρκειας τότε το ποσοστό επιτυχούς ανίχνευσης κυμαίνεται έως και 20% ανάλογα με τη δυναμική του περιεχομένου και το είδος της αλλαγής της σκηνής (απότομη ή βαθμιαία) [140]. Για να βελτιωθεί η απόδοση της τεχνικής και για να περιοριστεί το ποσοστό διακύμανσης της απόδοσης, κάποιες από τις τελευταίες προτεινόμενες τεχνικές χρησιμοποιούν δυναμικό κατώφλι, το οποίο αυτό-προσαρμόζεται ανάλογα με τη δυναμική του περιεχομένου [123], [129], με κόστος την υψηλή πολυπλοκότητα της τεχνικής.

Επειδή οι προαναφερθέντες τεχνικές κατάτμησης με δυναμικό κατώφλι απαιτούν μεγάλη υπολογιστική ισχύ για την ταυτόχρονη εκτέλεση της συνάρτησης απόφασης και της συνάρτησης δυναμικού κατωφλίου, η εφαρμογή τους σε συνδυασμό με την προτεινόμενη τεχνική πρόβλεψης της ποιότητας δεν αποτελεί βέλτιστη λύση. Βεβαίως θα μπορούσε να χρησιμοποιηθεί οποιαδήποτε εξ'αυτών των τεχνικών κατάτμησης, αλλά στα πλαίσια συνοχής και πληρότητας της παρούσης διατριβής η τεχνική ανίχνευσης των ορίων των σκηνών που θα χρησιμοποιηθεί για την επέκταση της μεθόδου σε βίντεο μεγάλης διάρκειας θα πρέπει να χρησιμοποιεί διεργασίες, οι οποίες ήδη υλοποιούνται στον αλγόριθμο πρόβλεψης της ποιότητας (π.χ. Εκτιμητής του φαινομένου τετραγώνισης). Με τον τρόπο αυτό επιτυγχάνεται η ενδογενής επέκταση της τεχνικής, χωρίς επιπλέον προσθήκες ή καινοτόμες διεργασίες στον μηχανισμό πρόβλεψης.

Για το λόγο αυτό στα πλαίσια του παρόντος εδαφίου, προτείνεται πρωτότυπη τεχνική κατάτμησης σκηνών, της οποίας η συνάρτηση απόφασης στηρίζεται σε αντιληπτό μετρικό, το οποίο υπολογίζει τη μέση ένταση του φαινομένου τετραγώνισης για κάθε καρέ, όπως ακριβώς και η μέθοδος πρόβλεψης της ποιότητας του προηγούμενου κεφαλαίου. Συνεπώς, η προτεινόμενη συνάρτηση απόφασης μπορεί να συνδυαστεί με την χρήση κατάλληλης συνάρτησης απεικόνισης με την εκτίμηση της αντιληπτής ποιότητας. Επιπλέον, στην προτεινόμενη μέθοδο δεν χρησιμοποιείται πολύπλοκη συνάρτηση κατωφλίου για τη διαδικασία ανίχνευσης. Αντιθέτως, η προτεινόμενη μέθοδος χρησιμοποιεί ένα εσωτερικό στατικό κατώφλι στη διαδικασία απόφασης, το οποίο καθορίζει εάν η μετρήσιμη διαφορά της έντασης του φαινομένου τετραγώνισης αντιστοιχεί σε κατάτμηση σκηνής. Δεν απαιτείται κανένας υπολογισμός δυναμικού κατωφλίου, επιτρέποντας την ενσωμάτωση της μεθόδου στην πρωτότυπη μέθοδο πρόβλεψης της αντιληπτής ποιότητας για βίντεο μεγάλης διάρκειας.

Στο επόμενο εδάφιο παραθέτουμε μια συνοπτική παρουσίαση της προτεινόμενης τεχνικής κατάτμησης σκηνών.

Η Προτεινόμενη Μέθοδος Κατάτμησης

Η προτεινόμενη τεχνική κατάτμησης αξιοποιεί το γεγονός ότι κατά τη διάρκεια μιας αλλαγής σκηνής υπάρχει έντονα ασυσχέτιστο περιεχόμενο και κατά συνέπεια η αντίστοιχη αποδοτικότητα των αλγορίθμων εκτίμησης και αντιστάθμισης της κίνησης μειώνεται σημαντικά. Αυτό έχει ως άμεση συνέπεια την αύξηση του πλήθους των blocks που κωδικοποιούνται ως intra, τα οποία άμεσα και έμμεσα ενισχύουν το φαινόμενο τετραγώνισης.

Το φαινόμενο τετραγώνισης, όπως έχει ήδη εκτενώς αναφερθεί σε προηγούμενο εδάφιο, αναφέρεται στο τεχνούργημα τετράγωνης μορφή που εμφανίζεται κατά την αναπαραγωγή του ψηφιακά κωδικοποιημένου βίντεο. Το φαινόμενο τετραγώνισης είναι αποτέλεσμα της ανεξάρτητης κβάντισης των DCT blocks που στοιχειοθετούν κάθε καρέ. Εξαιτίας του διακριτού συνημιτονικού μετασχηματισμού σε block του καρέ μεγέθους 8x8 pixels, οι διακυμάνσεις της φωτεινότητας στα όρια των block ενισχύονται, γεγονός που γίνεται αντιληπτό ως φαινόμενο τετραγώνισης. Ειδικά για εφαρμογές βίντεο στα πλαίσια κινητών επικοινωνιών, όπου ο ρυθμός κωδικοποίησης είναι χαμηλός, το φαινόμενο τετραγώνισης είναι το κυρίαρχο τεχνούργημα κατά την αναπαραγωγή του κωδικοποιημένου βίντεο. Ειδικότερα, κατά τη διάρκεια εναλλαγής μιας σκηνής παρατηρείται σημαντική αύξηση του φαινομένου τετραγώνισης, η οποία μπορεί να μην είναι άμεσα αντιληπτή από το θεατή [141] αλλά είναι ικανού μεγέθους για να μετρηθεί και να ποσοτικοποιηθεί.

Πιο συγκεκριμένα στην προτεινόμενη μέθοδο υπολογίζεται η μέση διακύμανση της φωτεινότητας στα όρια των blocks και η τιμή αυτή περιγράφεται ως $\langle offset \rangle$ για όλα τα ζεύγη των pixels ενός καρέ με πλάτος w και ύψος h :

$$\langle offset \rangle_{Frame} = \frac{\langle offset \rangle_v + \langle offset \rangle_H}{2}$$

Οι τιμές του $\langle offset \rangle$ για τα κάθετα ζεύγη ενός καρέ μπορούν να οριστούν ως

$$\langle offset \rangle_v = \sum_{i=1}^{w-8/8} \sum_{j=1}^h \frac{|Pixel_{8ij} - Pixel_{8i+1j}|}{h(w-8)/8}$$

Παρομοίως για τα οριζόντια ζεύγη το $\langle offset \rangle$ είναι

$$\langle offset \rangle_H = \sum_{i=1}^w \sum_{j=1}^{h-8/8} \frac{|Pixel_{i8j} - Pixel_{i8j+1}|}{(h-8)w/8}$$

Έπειτα υπολογίζουμε το μέσο όρο του offset ανά καρέ (δηλαδή $\langle offset \rangle_H + \langle offset \rangle_v / 2$) και το αποτέλεσμα κανονικοποιείται μεταξύ των τιμών 0 και 100, όπου το 0 απεικονίζει τη μέγιστη τιμή του φαινομένου τετραγώνισης και 100 τη μικρότερη, όντας με τον τρόπο αυτό σε συμφωνία με την αντιληπτή ποιότητα του βίντεο.

Τότε, η διαφορά των $\langle offset \rangle$ μεταξύ διαδοχικών καρέ μπορεί να απεικονιστεί ως εξής:

$$OD_n = \langle offset \rangle_n - \langle offset \rangle_{n-1}$$

Οπότε, εάν ο ακόλουθος αλγόριθμος απόφασης

$$D_n = \begin{cases} 100, & \text{Av } OD_n < 0 \text{ kai } |OD_n| > 15 \\ 1, & \text{Av } OD_n > 0 \end{cases}$$

εφαρμοστεί στην κανονικοποιημένη από 1 έως 100 μέτρηση της έντασης του φαινομένου τετραγώνισης, τότε οι καταγεγραμμένες διαφορές OD_n ελέγχονται για την ανίχνευση των ορίων κάθε σκηνής, επιτυγχάνοντας με τον τρόπο αυτό μια τεχνική κατάτμησης του βίντεο στις στοιχειώδεις σκηνές από τις οποίες δομείται. Η επιλογή του στατικού ορίου 15, επιλέχθηκε έπειτα από ποικίλες πειραματικές δοκιμές, όπου διαπιστώθηκε ότι πρόκειται για τη τιμή που δίδει βέλτιστα αποτελέσματα για κωδικοποίηση MPEG-4 ASP CIF και για ρυθμούς έως 500kbps (όπου είναι ουσιαστικά και το εύρος των ρυθμών κωδικοποίησης που εστιάζει η προτεινόμενη μέθοδος πρόβλεψης της ποιότητας).

Για την εκτίμηση της απόδοσης που προσφέρει η προτεινόμενη μέθοδος, τα ευρέως διαδεδομένα μετρικά της ακρίβειας (precision) και ανάκλησης (recall) χρησιμοποιήθηκαν. Ο ορισμός των μετρικών αυτών εστιάζει στην ποσοτικοποίηση της ικανότητας της μεθόδου να μην ανιχνεύει εσφαλμένα όρια σκηνών και πραγματικά/υπαρκτά όρια αντίστοιχα.

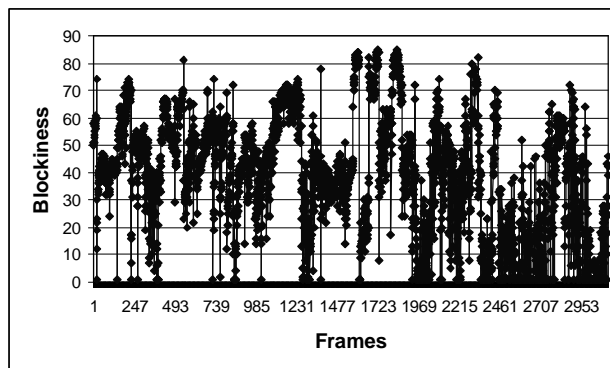
Συνεπώς, ο αναλυτικός ορισμός των μετρικών αυτών είναι ο ακόλουθος:

$$Anakl\ hsh = \frac{AnicneuseiV}{AnicneuseiV + Mh\ anafergenteV\ AnicneuseiV}$$

$$Akribeia = \frac{AnicneuseiV}{AnicneuseiV + Esfal\ meneV\ AnicneuseiV}$$

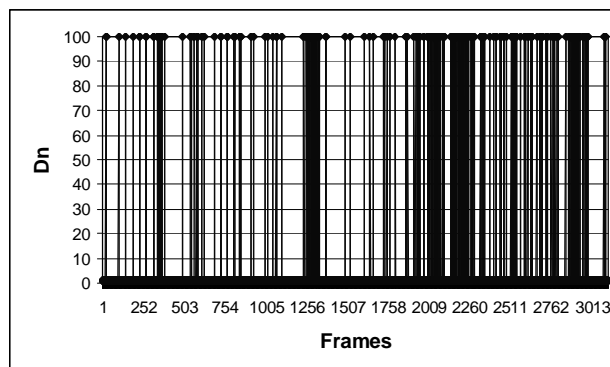
Αξίζει να σημειωθεί ότι στην αναφορά [142] προτείνεται ότι οι εσφαλμένες ανιχνεύσεις θα πρέπει να αγνοούνται και να μην λαμβάνονται υπόψιν στη διαδικασία αξιολόγησης της αποδοτικότητας μιας μεθόδου κατάτμησης σκηνών. Παρόλα αυτά η παρούσα διατριβή δε συντάσσεται υπέρ αυτής της άποψης διότι η συχνή εσφαλμένη ανίχνευση στοιχειωδών σκηνών θα έχει ως συνέπεια την περαιτέρω εσφαλμένη εφαρμογή της προτεινόμενης μεθόδου πρόβλεψης της ποιότητας στις στοιχειώδεις αυτές σκηνές. Συνεπώς, λαμβάνοντας υπόψιν ότι η τεχνική κατάτμησης σκηνών που προτείνεται στο παρόν εδάφιο στοχεύει στην αξιοποίησή της από το μηχανισμό πρόβλεψης της ποιότητας του προηγούμενου κεφαλαίου, είναι ύψιστης σημασίας η αξιολόγηση του ρυθμού εσφαλμένων ανιχνεύσεων και της ακρίβειας της προτεινόμενης τεχνικής.

Για τις ανάγκες της αξιολόγησης της προτεινόμενης τεχνικής κατάτμησης σκηνών, χρησιμοποιήθηκε ένα βίντεο μήκους 3113 καρέ, το οποίο περιέχει απόσπασμα από την ταινία δράσης Spiderman II. Το αρχικό βίντεο εξήχθη από DVD και κωδικοποιήθηκε σε ASP MPEG-4 CIF στα 256kbps με 25 καρέ το δευτερόλεπτο. Στο κωδικοποιημένο σήμα εφαρμόστηκε η προαναφερθείσα μέθοδος ποσοτικοποίησης της έντασης του φαινομένου τετραγώνισης, ώστε στη συνέχεια να κανονικοποιηθούν τα αποτελέσματα και να εφαρμοστεί η προτεινόμενη τεχνική κατάτμησης σκηνών. Τα αποτελέσματα αυτής της διαδικασίας απεικονίζονται στο σχήμα 8.1, όπου παρουσιάζονται τα πειραματικά αποτελέσματα της διακύμανσης της έντασης του φαινομένου τετραγώνισης ανάλογα με τη δυναμική του περιεχομένου.



Σχήμα 8.1 Διακύμανση του φαινομένου τετραγώνισης για απόσπασμα του Spiderman II

Στη συνέχεια εφαρμόζεται ο αλγόριθμος απόφασης D_n που περιγράφηκε πρωτύτερα, παρέχοντας εκείνα τα καρέ στα οποία –σύμφωνα με την εκτίμηση της τεχνικής-πραγματοποιείται αλλαγή σκηνής. Τα καρέ αυτά σηματοδοτούνται στην έξοδο της συνάρτησης απόφασης με την τιμή 100, ενώ τα υπόλοιπα παίρνουν την τιμή 1.



Σχήμα 8.2 Εφαρμογή του αλγορίθμου απόφασης στο απόσπασμα του Spiderman II

Το σχήμα 8.2 απεικονίζει το πειραματικό αποτέλεσμα της εφαρμογής της συνάρτησης απόφασης στο βίντεο της αξιολόγησης, όπου με την πειραματική τιμή εκατό κατατάσσονται τα καρέ που ενδέχεται να καθορίζουν τα όρια αλλαγής σκηνής.

Προτού ολοκληρωθεί η ανίχνευση των καρέ, θα ακολουθήσει μια διαδικασία φιλτραρίσματος των πρωτογενών ανιχνεύσιμων καρέ του σχήματος 8.2. Όπως μπορεί να παρατηρηθεί από τα πειραματικά αποτελέσματα, όταν πραγματοποιείται μια αλλαγή σκηνής, τότε λόγω της αύξησης του φαινομένου τετραγώνισης και της εσωτερικής εξάρτησης των καρέ σε μια κωδικοποίηση MPEG, παρατηρείται το φαινόμενο της διάδοσης του φαινομένου τετραγώνισης και στα αμέσως επόμενα γειτονικά καρέ, με αποτέλεσμα την εσφαλμένη ανίχνευση αυτών ως πιθανών αλλαγών σκηνών. Το φαινόμενο αυτό της διάδοσης της έντασης του φαινομένου τετραγώνισης, έπειτα από την εφαρμογή του αλγορίθμου απόφασης, παρατηρείται στα πειραματικά αποτελέσματα του σχήματος 8.2, όπου εμφανίζονται έντονες πυκνές περιοχές που απαρτίζονται από διαδοχικά υποψήφια καρέ.

Για να ελαχιστοποιηθεί η ένταση του φαινομένου, το οποίο θα οδηγούσε σε πληθώρα εσφαλμένων αναφορών για αλλαγή σκηνής, εισάγεται ένα διάστημα αδράνειας έπειτα από

κάθε επιτυχής ανίχνευση πιθανής αλλαγής σκηνής, κατά τη διάρκεια του οποίου δεν λαμβάνεται υπόψιν οποιαδήποτε ανιχνεύσιμη αλλαγή σκηνής, ακόμη και αν ο αλγόριθμος απόφασης αναφέρει την ύπαρξη κάποιων. Αξιοποιώντας το γεγονός ότι η μέση διάρκεια μιας σκηνής είναι αισθητά μεγαλύτερη του ενός δευτερολέπτου [143], τότε το διάστημα αδράνειας μπορεί με ασφάλεια να θεωρηθεί ίσο με 25 καρέ με μικρή πιθανότητα απόρριψης καρέ που αντιστοιχεί όντως σε αλλαγή σκηνής.

Εφαρμόζοντας το διάστημα αδράνειας των 25 καρέ στα πειραματικά αποτελέσματα του σχήματος 8.2, τα εξαγόμενα δεδομένα είναι απαλλαγμένα από τη διάδοση του φαινομένου τετραγώνισης και κατά συνέπεια θεωρούμε ότι αντιπροσωπεύουν επιτυχώς και επακριβώς τις αλλαγές των σκηνών του υπό εξέταση σήματος. Για το λόγο αυτό αξιολογούνται για την ακρίβεια και αξιοπιστία τους, πραγματοποιώντας άμεση σύγκριση με το περιεχόμενο του εξεταζόμενου σήματος και εφαρμόζοντας τα μετρικά ανάκλησης και ακρίβειας.

Μέθοδος Κατάτμησης	Ανάκληση	Ακρίβεια
Προτεινόμενη Μέθοδος	0.81	0.72
Μέθοδος 1	0.52	0.79
Μέθοδος 2	0.74	0.58

Πίνακας 8.1 Σύγκριση της απόδοσης της προτεινόμενης μεθόδου για απότομες αλλαγές σκηνών

Τα αντίστοιχα αποτελέσματα απεικονίζονται στον πίνακα 8.1 για το σήμα Spiderman II, το οποίο περιέχει μόνο απότομες αλλαγές σκηνών, μαζί με τις αποδόσεις για απότομες αλλαγές σκηνών άλλων δύο τεχνικών ανίχνευσης σκηνών κωδικοποιημένου MPEG βίντεο για τον ίδιο ρυθμό κωδικοποίησης (256kbps) [144]. Η μέθοδος 1 αναλύεται στην αναφορά [145] και η μέθοδος 2 στην αναφορά [146].

Μέθοδος Κατάτμησης	Ανάκληση	Ακρίβεια
Προτεινόμενη Μέθοδος	0.99	0.33
Μέθοδος 1	0.21	0.06
Μέθοδος 2	0.24	0.20

Πίνακας 8.2 Σύγκριση της απόδοσης της προτεινόμενης μεθόδου για σταδιακές αλλαγές σκηνών

Η απόδοση της προτεινόμενης τεχνικής δοκιμάστηκε και συγκρίθηκε επίσης και για την περίπτωση περιεχομένου με σταδιακή αλλαγή σκηνών. Τα αποτελέσματα απεικονίζονται στον πίνακα 8.2, όπου φαίνεται η υπεροχή της προτεινόμενης τεχνικής έναντι των δύο άλλων μεθόδων.

Ενσωμάτωση της Μεθόδου Κατάτμησης στην Τεχνική Πρόβλεψης της Ποιότητας

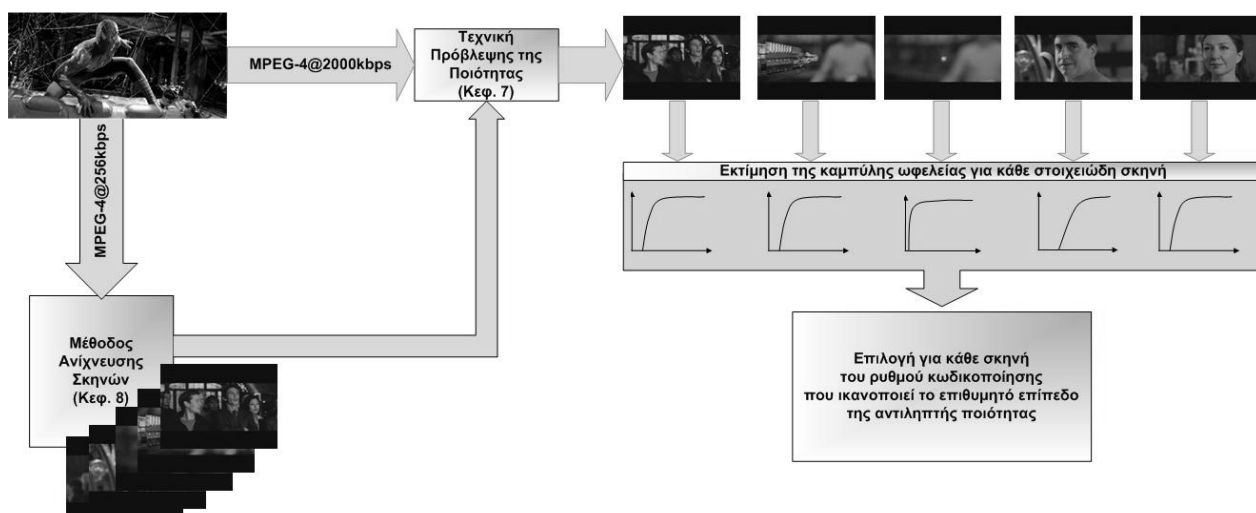
Το παρόν εδάφιο θα παρουσιάσει την ενσωμάτωση της μεθόδου κατάτμησης ενός βίντεο στις στοιχειώδεις σκηνές από τις οποίες δομείται με σκοπό την επέκταση της τεχνικής πρόβλεψης της ποιότητας σε βίντεο μεγάλης διάρκειας με έντονα ετερογενές περιεχόμενο.

Η τεχνική κατάτμησης των σκηνών που αναπτύχθηκε στο προηγούμενο εδάφιο δύναται να κατατμήσει με υψηλή ακρίβεια και αποδοτικότητα ένα βίντεο που κωδικοποιείται σύμφωνα

με το πρότυπο MPEG-4 στις στοιχειώδεις σκηνές από τις οποίες αποτελείται. Μια σκηνή ορίζεται ως στοιχειώδης όταν η καταγραφή της έχει πραγματοποιηθεί αδιάκοπα με τη χρήση της ίδιας κάμερας. Η εναλλαγή μεταξύ δύο στοιχειωδών σκηνών μπορεί να είναι είτε απότομη, δηλαδή τα όρια της εναλλαγής καθορίζονται απευθείας σε ένα καρέ, ή σταδιακή, όπου τα όρια των δύο σκηνών ορίζονται σε ένα πλήθος διαδοχικών καρέ.

Κάθε στοιχειώδης σκηνή, επειδή έχει καταγραφεί με τη χρήση της ίδιας κάμερας, ως επί το πλείστον έχει περιεχόμενο το οποίο μπορεί να ικανοποιεί τις απαιτήσεις για εφαρμογή της τεχνικής πρόβλεψης της ποιότητας του προηγούμενου κεφαλαίου. Η συντριπτική πλειοψηφία των στοιχειωδών σκηνών που προέρχονται από βίντεο επαγγελματικού επιπέδου μπορεί ασφαλώς να χαρακτηριστεί ως ομοιογενής αφού περιλαμβάνει μικρής διάρκειας ομοιογενές περιεχόμενο με συγκεκριμένη θεματολογία. Στις περιπτώσεις ερασιτεχνικού βίντεο, λόγω της απουσίας του μοντάζ, ενδέχεται κάποια στοιχειώδη σκηνή να περιέχει ετερογενές περιεχόμενο, αλλά λόγω περιορισμένης διάρκειας οι ιδιότητές της συγκαταλέγονται εντός το φάσμα των βίντεο βραχείας διάρκειας ετερογενούς περιεχομένου που εξετάστηκαν στο προηγούμενο κεφάλαιο και ικανοποιούν με ακρίβεια την προτεινόμενη τεχνική. Η περίπτωση σκηνής, η οποία είναι αυτοτελές βίντεο μεγάλης διάρκειας και περιέχει έντονα ετερογενές περιεχόμενο δεν εξετάζεται στα πλαίσια της παρούσας διατριβής, δεδομένου ότι πρόκειται για περίπτωση που δεν συναντάται ευρέως.

Η βασική αρχή επέκτασης της μεθόδου είναι ότι εάν η τεχνική πρόβλεψης της μεθόδου εφαρμοστεί ανεξάρτητα σε κάθε στοιχειώδη σκηνή που αποτελεί το βίντεο, τότε λόγω της υψηλής ομοιογένειας της κάθε σκηνής, η απόδοση και η ακρίβεια της μεθόδου θα είναι ικανοποιητική. Δηλαδή θα καθοριστεί για κάθε στοιχειώδη σκηνή του βίντεο μεγάλης διάρκειας εκείνος ο ρυθμός κωδικοποίησης που ικανοποιεί το επιθυμητό επίπεδο ποιότητας ανάλογα με τη δυναμική του περιεχομένου. Προσδιορίζοντας για κάθε στοιχειώδη σκηνή την εξάρτηση της ποιότητας και του ρυθμού κωδικοποίησης, τότε καθίσταται δυνατόν να καθοριστεί για κάθε σκηνή ο ρυθμός που αντιστοιχεί στο επίπεδο της ποιότητας που ο πάροχος επιθυμεί να προσφέρει το συγκεκριμένο βίντεο. Με τον τρόπο αυτό το σύνολο του βίντεο μπορεί να κωδικοποιηθεί με τέτοιο τρόπο ώστε να παρέχεται σταθερή αντιληπτή ποιότητα για όλη τη διάρκειά του.



Σχήμα 8.3 Επέκταση της μεθόδου πρόβλεψης σε βίντεο μακράς διάρκειας

Δεδομένου ενός βίντεο μεγάλης διάρκειας, στο οποίο θέλουμε να εφαρμοστεί η τεχνική πρόβλεψης της ποιότητας, ο μηχανισμός που πρέπει να ακολουθηθεί περιγράφεται στο σχήμα 8.3. Το βίντεο μεγάλης διάρκειας κωδικοποιείται αρχικά σύμφωνα με το πρότυπο MPEG-4 σε ρυθμό κατάλληλο για την εφαρμογή της τεχνικής κατάτμησης των σκηνών. Στη συνέχεια το κωδικοποιημένο σήμα χρησιμοποιείται ως είσοδο στην προτεινόμενη τεχνική κατάτμησης, η οποία θα εντοπίσει εκείνα τα καρέ που σηματοδοτούν την έναρξη στοιχειωδών σκηνών. Στη συνέχεια η προτεινόμενη μέθοδος πρόβλεψης αξιοποιεί τα σηματοδοτημένα καρέ στο αρχικό ασυμπιεστο βίντεο και διαχειρίζεται την κάθε ανιχνεύσιμη στοιχειώδη σκηνή ως αυτόνομο βίντεο. Με τον τρόπο αυτό εκτιμάται η συνάρτηση ωφελείας για κάθε στοιχειώδη σκηνή του βίντεο, επιτρέποντας την επιλογή του κατάλληλου ρυθμού κωδικοποίησης για κάθε στοιχειώδη σκηνή, ώστε να ικανοποιήσει το επιθυμητό επίπεδο αντιληπτής ποιότητας.

Σχολιασμός στην ακρίβεια και αποδοτικότητα της εκτεταμένης μεθόδου

Η ακρίβεια και η αποδοτικότητα της μεθόδου πρόβλεψης της ποιότητας στις στοιχειώδεις σκηνές ενός βίντεο μεγάλης διάρκειας δεν θα εξεταστεί περαιτέρω στο παρόν κεφάλαιο. Η περίπτωση αυτή έχει ήδη εξεταστεί στο κεφάλαιο 7, μέσω των βίντεο σύντομης διάρκειας με ομοιογενές και ετερογενές περιεχόμενο, επιδεικνύοντας την υψηλή απόδοση και ακρίβεια της προτεινόμενης τεχνικής. Συνεπώς, η ακρίβεια της προτεινόμενης τεχνικής ενδέχεται να περιοριστεί περαιτέρω μόνο από την ανεπιτυχή απόδοση της τεχνικής κατάτμησης των στοιχειωδών σκηνών.

Δεδομένου όμως των τιμών ακρίβειας και επιτυχούς ανίχνευσης της προτεινόμενης τεχνικής, όπως μετρήθηκε και αποδίδεται στον πίνακα 8.1, τα ακόλουθα σχόλια προκύπτουν:

- Η εσφαλμένη ανίχνευση μιας σκηνής δεν περιορίζει την απόδοση της μεθόδου πρόβλεψης της ποιότητας. Αντιθέτως, διασπά μια ενιαία σκηνή σε περισσότερα μέρη, στα οποία εφαρμόζεται ο αλγόριθμος πρόβλεψης με μεγαλύτερη ακρίβεια. Συνεπώς, εσφαλμένη ανίχνευση προκαλεί αύξηση της ακρίβειας της μεθόδου πρόβλεψης.
- Η εσφαλμένη ανίχνευση μιας σκηνής αυξάνει την πολυπλοκότητα της εκτεταμένης μεθόδου, αφού απαιτείται η παράλληλη εφαρμογή του αλγορίθμου εκτίμησης της συνάρτησης ωφελείας σε περισσότερα αποσπάσματα του βίντεο, γεγονός που καταναλώνει περισσότερους πόρους του υπολογιστικού συστήματος.
- Η ανεπιτυχής ανίχνευση μια σκηνής περιορίζει εν μέρει την απόδοση της μεθόδου πρόβλεψης, αφού σκηνή μεγαλύτερης διάρκειας συνεπάγεται μεγαλύτερη διασπορά στις τιμές πρόβλεψης της ποιότητας, αφού δημιουργείται ανομοιογένεια στη δυναμική του περιεχομένου.

Αξίζει όμως να σημειωθεί ότι η άστοχη ανίχνευση μια σκηνής, έχει συνήθως ως αποτέλεσμα ένα απόσπασμα του βίντεο, το οποίο είναι εντός των προδιαγραφών των ετερογενών βίντεο βραχείας διάρκειας που εξετάστηκαν επιτυχώς στο κεφάλαιο 7. Οπότε, πρακτικά μια ανεπιτυχής ανίχνευση του ορίου μια σκηνής, δεν προκαλεί σημαντική υποβάθμιση στην απόδοση και την ακρίβεια εφαρμογής της μεθόδου πρόβλεψης. Λαμβάνοντας υπόψιν και τις υψηλές τιμές επιτυχούς ανίχνευσης του μηχανισμού κατάτμησης σκηνών που παρουσιάζονται στους πίνακες 8.1 και 8.2, μπορεί ασφαλώς να εξαχθεί το συμπέρασμα ότι ακόμη και στη χειρότερη περίπτωση δεν θα προκύψουν στοιχειώδεις σκηνές όπου η εκτεταμένη μέθοδος

πρόβλεψης δεν θα μπορεί να εφαρμόσει την διαδικασία εκτίμησης της συνάρτησης ωφελείας με ικανοποιητική ακρίβεια.

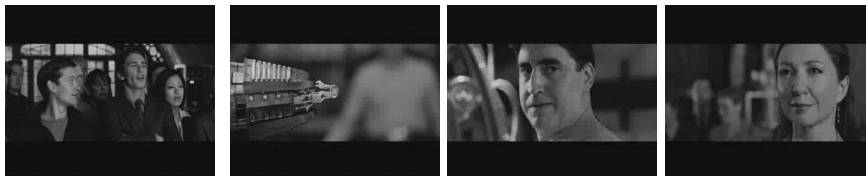
- Η ανεπιτυχής ανίχνευση μιας σκηνής μειώνει την πολυπλοκότητα εφαρμογής της εκτεταμένης μεθόδου πρόβλεψης της ποιότητας, αφού περιορίζεται ο αριθμός των παραλλήλων εφαρμογών εκτίμησης της συνάρτησης ωφελείας και κατά συνέπεια των υπολογιστικών πόρων που απαιτούνται.

Συνοψίζοντας, σχετικά με την απόδοση και την ακρίβεια της εκτεταμένης μεθόδου πρόβλεψης της ποιότητας, σύμφωνα με τα παραπάνω σχόλια μπορούμε ασφαλώς να συμπεράνουμε ότι εξαρτάται και διαμορφώνεται πλήρως από την επιμέρους απόδοση της μεθόδου στις στοιχειώδεις σκηνές. Οπότε ουσιαστικά το όλο ζήτημα απόδοσης της εκτεταμένης τεχνικής ανάγεται στη μελέτη βίντεο ετερογενούς και ομοιογενούς περιεχομένου βραχείας διάρκειας, το οποίο έχει ήδη καλυφθεί στο κεφάλαιο 7.

Πειραματική εφαρμογή της εκτεταμένης μεθόδου

Για λόγους συνοχής και σύγκρισης του παρόντος κεφαλαίου που παρουσίασε την εκτεταμένη μέθοδο και του προηγούμενου κεφαλαίου αξιολόγησης της μεθόδου πρόβλεψης σε βίντεο βραχείας διάρκειας, το εν λόγω εδάφιο παρουσιάζει πειραματική εφαρμογή της εκτεταμένης μεθόδου.

Για να επιδειχθεί η βελτίωση στην ακρίβεια και αποδοτικότητα πρόβλεψης της συνάρτησης ωφελείας με την εφαρμογή της εκτεταμένης μεθόδου, ως πηγή θα χρησιμοποιηθεί ένα απόσπασμα από την ταινία *Spiderman II* με έντονα ετερογενές περιεχόμενο. Στόχος της πειραματικής εφαρμογής είναι να επιδειχθεί η βελτίωση της απόδοσης και ακρίβειας που προσφέρει η εκτεταμένη μέθοδος με τη συνδυαστική χρήση της τεχνικής κατάτμησης των σκηνών.

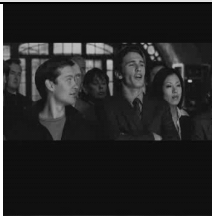





Σχήμα 8.4 Στοιχειώδεις σκηνές του σήματος

Στο σχήμα 8.4, παρουσιάζονται τα πρώτα καρέ των στοιχειωδών σκηνών που ανιχνεύονται από την εφαρμογή του μηχανισμού κατάτμησης σκηνών στο βίντεο της δοκιμής, το οποίο για το λόγο αυτό κωδικοποιήθηκε σύμφωνα με το πρότυπο MPEG-4 σε ανάλυση CIF στα 265 kbps. Στη συνέχεια, η τεχνική εκτίμησης εφαρμόζεται στη κάθε στοιχειώδη σκηνή με αποτέλεσμα την εκτίμηση των θεωρητικών συναρτήσεων ωφελείας και των διανυσμάτων ποιότητας. Συνεπώς, με τη χρήση της εκθετικής προσέγγισης της συνάρτησης ωφελείας, εάν θεωρήσουμε ότι το επιθυμητό επίπεδο αντιληπτής ποιότητας είναι το 80, τότε μπορεί να προσδιοριστεί επακριβώς ο ρυθμός κωδικοποίησης που απαιτείται για κάθε σκηνή.

Επαναλαμβάνοντας τη διαδικασία χωρίς τη χρήση του μηχανισμού κατάτμησης των στοιχειωδών σκηνών, τότε το MPQoS που προκύπτει από την ενιαία κωδικοποίηση και εκτίμηση της ποιότητας του κωδικοποιημένου βίντεο MPEG-4/CIF στα 1200kbps είναι 87. Με τη χρήση των εξισώσεων του κεφαλαίου 7 και προσδιορισμού της εκθετικής προσέγγισης της συνάρτησης ωφελείας προκύπτει ότι για να επιτευχθεί αντιληπτή ποιότητα 80 για το σύνολο του βίντεο, τότε πρέπει να κωδικοποιηθεί στα 280kbps. Το αποτέλεσμα

αυτό έρχεται σε αντιπαράθεση στον πίνακα 8.3 με τα αντίστοιχα αποτελέσματα που προκύπτουν από την εφαρμογή της εκτεταμένης μεθόδου.

Στοιχειώδης σκηνή	Απλή Μέθοδος	Εκτεταμένη Μέθοδος
	280 kbps	200 kbps
	280 kbps	220 kbps
	280 kbps	350 kbps
	280 kbps	150 kbps

Πίνακας 8.3 Σύγκριση των αποτελεσμάτων της απλής και επεκταμένης μεθόδου

Όπως μπορεί να παρατηρηθεί από τα αντίστοιχα αποτελέσματα, η εφαρμογή της απλής μεθόδου πρόβλεψης σε όλη τη διάρκεια του βίντεο, οδηγεί σε κωδικοποίηση που επί του συνόλου μπορεί να προσφέρει μέση ποιότητα ίση με την επιθυμητή, αλλά σε επίπεδο στοιχειώδης σκηνής, η αντίστοιχη ποιότητα που επιτυγχάνεται ποικίλει. Στην περίπτωση εφαρμογής της επεκταμένης μεθόδου με τη χρήση της τεχνικής κατάτμησης σκηνών, υπολογίζεται διαφορετικός ρυθμός κωδικοποίησης για κάθε στοιχειώδη σκηνή, ο οποίος ικανοποιεί ενδοσκοπικά το επιθυμητό επίπεδο αντιληπτής ποιότητας. Με την εκτεταμένη μέθοδο επιτυγχάνεται σταθερό επίπεδο αντιληπτής ποιότητας για όλη τη διάρκεια του βίντεο, δεδομένου ότι κάθε στοιχειώδης σκηνή έχει κωδικοποιηθεί στο ρυθμό που καθορίζεται από τη χωρική και χρονική δυναμική του περιεχομένου, ώστε να ικανοποιηθεί το επιθυμητό επίπεδο της ποιότητας.

Συμπεράσματα του Κεφαλαίου

Το παρόν κεφάλαιο περιγράφει, μελετάει και παρουσιάζει την επέκταση της μεθόδου πρόβλεψης της αντιληπτής ποιότητας του κεφαλαίου 7 σε βίντεο μεγάλης διάρκειας και ετερογενούς περιεχομένου. Η επέκταση της μεθόδου πραγματοποιείται με τη συνδυαστική χρήση πρωτότυπης τεχνικής ανίχνευσης των ορίων των σκηνών, η οποία πραγματοποιεί κατάτμηση ενός βίντεο μεγάλης διάρκειας στις στοιχειώδεις σκηνές από τις οποίες αποτελείται.

Η προτεινόμενη τεχνική κατάτμησης σκηνών εξειδικεύεται σε βίντεο κωδικοποιημένα σύμφωνα με το πρότυπο MPEG-4 και χρησιμοποιεί ως μετρικό το γεγονός ότι η ένταση του φαινομένου τετραγώνισης ενισχύεται στην εναλλαγή μιας σκηνής, εξαιτίας της μειωμένης συσχέτισης του περιεχομένου των συνοριακών καρέ των δύο σκηνών και κατά συνέπεια της περιορισμένης απόδοσης των αλγορίθμων εκτίμησης και αντιστάθμισης της κίνησης.

Επιπλέον, παρουσιάζεται με πειραματικό τρόπο η βελτίωση που επιφέρει η επέκταση της μεθόδου πρόβλεψης σε βίντεο με έντονα ετερογενές περιεχόμενο, όπου η εφαρμογή της απλής τεχνικής προκαλεί αποκλίσεις, οι οποίες για βίντεο μεγάλης διάρκειας θα οδηγούσαν σε μη αποδεκτά αποτελέσματα.

Τέλος, σχολιάζεται πώς η επέκταση της μεθόδου πρόβλεψης με τη χρήση της τεχνικής κατάτμησης των σκηνών δε μεταβάλλει την απόδοση και την ακρίβεια της μεθόδου, αφού ουσιαστικά ανάγει το πρόβλημα ενός βίντεο μεγάλης διάρκειας με ετερογενές περιεχόμενο στην περίπτωση ομοιογενών βίντεο μικρής διάρκειας.

ΚΕΦΑΛΑΙΟ 9

Διερεύνηση και Μοντελοποίηση της Επίδρασης του Ρυθμού Απώλειας Πακέτων του Δικτύου Μεταφοράς στην Υποβάθμιση της Αντιληπτής Ποιότητας

Υποβάθμιση της Ποιότητας κατά τη Μετάδοση του Βίντεο

Τα προηγούμενα κεφάλαια πραγματεύτηκαν την υποβάθμιση που υφίσταται ένα ψηφιακό βίντεο όταν συμπιέζεται σύμφωνα με ένα πρότυπο κωδικοποίησης. Όπως έχει ήδη αναφερθεί εξαιτίας της χρήσης διακριτού συνημιτονικού μετασχηματισμού, της κβάντισης των αντιστοιχών παραμέτρων και των μηχανισμών εκτίμησης και αντιστάθμισης της κίνησης, οι τεχνικές κωδικοποίησης υποβαθμίζουν ποιοτικά το ψηφιακό βίντεο. Η συνεισφορά της παρούσας διατριβής έως το προηγούμενο κεφάλαιο εστιάστηκε στην περιγραφή και δοκιμή πρωτότυπης μεθόδου πρόβλεψης του επιπέδου της αντιληπτής ποιότητας πριν πραγματοποιηθεί το στάδιο της κωδικοποίησης.

Από κει και πέρα, σε εφαρμογές πραγματικού χρόνου, ένα ψηφιακά κωδικοποιημένο βίντεο, ενδέχεται να υποστεί περαιτέρω υποβάθμιση της ποιότητας κατά τη μετάδοση του περιεχομένου μέσω του δικτύου μεταφοράς. Ένα μείζον ζήτημα στα υπολογιστικά και τηλεπικοινωνιακά δίκτυα δεδομένων πακετομεταγωγής είναι ότι τα χαρακτηριστικά του καναλιού μεταφοράς μεταβάλλονται με την πάροδο του χρόνου, εξαιτίας του γεγονός ότι οι διαθέσιμοι δικτυακοί πόροι διαμοιράζονται ταυτόχρονα με άλλους χρήστες. Εξαιτίας αυτής της ευμετάβλητης κατάστασης, το κανάλι μεταφοράς, μέσω του οποίου θα μεταφερθεί η υπηρεσία βίντεο, ενδέχεται να εμφανίζει σφάλματα, με κυρίαρχο φαινόμενο –ιδιαιτέρως στα ασύρματα δίκτυα- η απώλεια πακέτων δεδομένων. Λαμβάνοντας υπόψιν ότι κατά τη μετάδοση μιας υπηρεσίας βίντεο πραγματικού χρόνου μέσω ενός δικτύου πακετομεταγωγής τα δεδομένα του βίντεο κατακερματίζονται και ομαδοποιούνται σε πακέτα δεδομένων, τα οποία με τη σειρά τους μεταφέρονται μέσω του δικτύου, τότε γίνεται άμεσα αντιληπτό ότι απώλεια των πακέτων συνεπάγεται και ποιοτική υποβάθμιση της ποιότητας του μεταδιδόμενου βίντεο.

Ειδικότερα, οι εφαρμογές βίντεο πραγματικού χρόνου που χρησιμοποιούν κωδικοποιημένο βίντεο είναι ιδιαίτερα ευαίσθητες στα σφάλματα μετάδοσης εξαιτίας της αλληλεξάρτησης των δεδομένων του συμπιεσμένου βίντεο. Για αυτό το λόγο είναι πολύ αυστηρές οι απαιτούμενες προδιαγραφές σχετικά με την ανεκτικότητα στα λάθη, ώστε να μην παρατηρηθεί ποιοτική υποβάθμιση στη εκπεμπόμενη υπηρεσία βίντεο. Ειδικότερα κατά τη μετάδοση βίντεο μέσω ασύρματων δικτύων πακετομεταγωγής, κάθε μεταδιδόμενο πακέτο από άκρη σε άκρη ενδέχεται να ληφθεί στην άλλη μεριά είτε σωστά, ή με λάθη ή να μην παραληφθεί ποτέ λόγω απώλειας. Στις δύο τελευταίες περιπτώσεις η αντίστοιχη αντιληπτή ποιοτική υποβάθμιση είναι όμοια, αφού ο αποκωδικοποιητής απορρίπτει τα πακέτα που περιέχουν λάθη, προκαλώντας ποιοτική υποβάθμιση.

Στόχος του παρόντος κεφαλαίου είναι η διερεύνηση της συσχέτισης μεταξύ του ρυθμού απώλειας πακέτων σε δίκτυα πακετομεταγωγής και της ποιοτικής υποβάθμισης που προκαλείται στο μεταδιδόμενο βίντεο.

Συνοπτική Βιβλιογραφική Ανασκόπηση της Επίπτωσης Δικτυακών Παραμέτρων στην Ποιότητα του Μεταδιδόμενου Βίντεο

Η σχετική βιβλιογραφία στη θεματική ενότητα της διερεύνησης της επίδρασης των δικτυακών παραμέτρων στο επίπεδο της αντιληπτής ποιότητας είναι σχετικά περιορισμένη, αφού μικρός αριθμός εργασιών έχει παρουσιαστεί. Η πλειοψηφία των δημοσιευμένων εργασιών εστιάζει στη μελέτη της αντιληπτής επίπτωσης της απώλειας πακέτων δεδομένων κατά τη μετάδοση της οπτικοακουστικής υπηρεσίας, θεωρώντας είτε ασύρματη ζεύξη με κακή ποιότητα λήψης ή ενσύρματο κανάλι μετάδοσης με έντονη συμφόρηση δεδομένων.

Σε αυτά τα πλαίσια, ο S. Kanumuri [147] παρουσίασε ένα αναλυτικό στατιστικό μοντέλο της αντιληπτής επίδρασης του ρυθμού απώλειας πακέτων στην ποιότητα αποκωδικοποιημένου MPEG-2 βίντεο, διερευνώντας ποικίλες παραμέτρους (όπως ο ρυθμός των καρτέ, ο τύπος του καρτέ που συμβαίνει η απώλεια κλπ.) που επηρεάζουν την ένταση της επίπτωσης στο μεταδιδόμενο βίντεο.

Παρομοίως, στην [148] παρουσιάζεται ένα θεωρητικό πλαίσιο μοντελοποίησης της επίδρασης στην ποιότητα υπηρεσίας ροής βίντεο πραγματικού χρόνου ενός ασύρματου δικτύου επιρρεπές σε σφάλματα. Πιο συγκεκριμένα, παρουσιάζεται μια διατεμαχική μελέτη της ποιοτικής υποβάθμισης του βίντεο, αξιοποιώντας τη μοντελοποίηση της αντιληπτής απόσβεσης των σφαλμάτων μετάδοσης. Το προτεινόμενο μοντέλο, παρόλο που παρέχει πολύ ακριβή αποτελέσματα και μπορεί να αξιοποιηθεί από τον πάροχο της υπηρεσίας για την εκτίμηση της ενδεχόμενης αντιληπτής υποβάθμισης κατά τη μετάδοση, μειονεκτεί στο γεγονός ότι δεν είναι γενικευμένο, αλλά αντιθέτως παρουσιάζει ισχυρή εξάρτηση από τη δυναμική του περιεχομένου και από το είδος της κωδικοποίησης. Πιο συγκεκριμένα, για να είναι ο πάροχος της υπηρεσίας σε θέση να εφαρμόσει το προτεινόμενο μοντέλο πρέπει αρχικώς να ποσοτικοποιήσει τη χωρική και χρονική δυναμική του περιεχομένου, ώστε να προσαρμόσει τις αντίστοιχες παραμέτρους του μοντέλου ανάλογα. Αυτή η προαπαιτούμενη διαδικασία καθιστά τη μέθοδο χρονοβόρα και εμπορικά μη υλοποιήσιμη. Επιπλέον, η ισχυρή εξάρτηση του προτεινόμενου μοντέλου από τη χωρική και χρονική δυναμική του περιεχομένου, περιορίζει την εφαρμογή της σε βίντεο βραχείας διάρκειας και ομοιογενούς περιεχομένου.

Στην [149] ο Gastaldo προτείνει μια μέθοδο για την εκτίμηση της αντιληπτής ποιότητας ροών βίντεο MPEG-2 που βασίζεται στη θεωρία νευρωνικών δικτύων χωρίς να απαιτείται αποκωδικοποίηση του βίντεο. Παρόλα αυτά το προτεινόμενο θεωρητικό πλαίσιο περιορίζεται στα τεχνουργήματα που προκαλούνται από τη διαδικασία συμπίεσης και δεν ανιχνεύει την υποβάθμιση που οφείλεται σε δικτυακές παραμέτρους.

Η ταυτόχρονη επίδραση του ρυθμού κωδικοποίησης και της απώλειας κελιών ATM στην ποιότητα βίντεο MPEG-2 διερευνάται στις αναφορές [150] και [151], όπου για δεδομένο ρυθμό απώλειας πακέτων καθορίζεται ο βέλτιστος ρυθμός κωδικοποίησης. Μια παρόμοια μελέτη [152] δείχνει ότι αύξηση του ρυθμού κωδικοποίησης μπορεί να επιφέρει βελτίωση της αντίστοιχης αντιληπτής ποιότητας μόνο όταν η αντίστοιχη απώλεια κελιών ATM είναι μικρότερη από ένα προκαθορισμένο όριο. Ένα θεωρητικό πλαίσιο που αξιοποιεί μεθόδους αντικειμενικής εκτίμησης για τη μοντελοποίηση της δικτυακής απόδοσης παρουσιάζεται στην [153].

Μια πληθώρα εργασιών [77], [154], [155], [156], εστιάζουν στη διερεύνηση της μέσης υποβάθμισης της αντιληπτής ποιότητας υπό τυπικές δικτυακές συνθήκες. Πιο συγκεκριμένα, εφαρμογές τηλεδιάσκεψης εξετάζονται στην [154] ώστε να καθοριστούν οι

προδιαγραφές που απαιτούνται στο διαθέσιμο εύρος ζώνης, στην καθυστέρηση και στην απώλεια πακέτων. Στην [155] διερευνάται η επίδραση του ρυθμού απώλειας πακέτων στην ποιότητα εφαρμογών βίντεο ροής πραγματικού χρόνου. Στην [77] μια υποδομή νευρωνικού δικτύου αξιολογείται στις υποκειμενικές εκτιμήσεις ενός βίντεο διάρκειας δέκα δευτερολέπτων σε συνθήκες μεταβλητού εύρους ζώνης και ρυθμού απώλειας πακέτων. Στην [156] χρησιμοποιείται το MOS για την εκτίμηση της υποκειμενικής ποιότητας βίντεο υπό συνθήκες απώλειας κελιών ATM, όπου αποδείχθηκε ότι η δυναμική του περιεχομένου επιφέρει σημαντική μεταβολή στα αντίστοιχα αποτελέσματα.

Αντίστοιχα, στην [157] χρησιμοποιούνται δύο υποκειμενικές μέθοδοι εκτίμησης της αντιληπτής ποιότητας για την ποσοτικοποίηση της επίδρασης του ρυθμού απώλειας πακέτων στην αντιληπτή ποιότητα που παρακολουθούν οι θεατές. Μια παρόμοια προσέγγιση με χρήση υποκειμενικών μεθόδων παρουσιάζεται στην [158].

Αξίζει να σημειωθεί ότι στις εργασίες που χρησιμοποιούνται υποκειμενικές μέθοδοι για τη διερεύνηση της απώλειας πακέτων στο επίπεδο της αντιληπτής ποιότητας, η έννοια της αντιληπτής ποιότητας διαφέρει σημαντικά σε κάθε εργασία. Επί παραδείγματι, στις [159] και [160] οι θεατές ερωτώνται να εκτιμήσουν τη συχνότητα εμφάνισης τεχνουργημάτων, ενώ στις [161] και [162] ζητείται από τους θεατές να εκτιμήσουν την ένταση των τεχνουργημάτων. Συνεπώς οποιαδήποτε σύγκριση μεταξύ των αποτελεσμάτων καθίσταται ιδιαίτερος δύσκολη.

Σε αυτά τα πλαίσια, το παρόν κεφάλαιο περιγράφει, προτείνει και δοκιμάζει πρωτότυπο γενικευμένο μοντέλο της επίδρασης του ρυθμού απώλειας πακέτων στην αντιληπτή ποιότητα υπηρεσίας βίντεο-ροής πραγματικού χρόνου. Η γενικευμένη φύση του προτεινόμενου μοντέλου βασίζεται στο γεγονός ότι η απόδοσή του δεν εξαρτάται από τη δυναμική του περιεχομένου, αλλά μόνο από τη δομή του GOP του κωδικοποιημένου σήματος.

Επίδραση της Απώλειας Πακέτων στο Αποκωδικοποιημένο Βίντεο

Ένα βίντεο μπορεί να διαιρεθεί και να οργανωθεί σε πακέτα με έναν από τους ακόλουθους τρόπους:

- Ο κατακερματισμός των δεδομένων να πραγματοποιηθεί σε πακέτα σταθερού μεγέθους (όπως τα κελιά ATM ή τα πακέτα μεταφοράς MPEG-2 που έχουν σταθερό μέγεθος 188 bytes).
- Ο κατακερματισμός των δεδομένων να πραγματοποιηθεί σε πακέτα μεταβλητού μεγέθους, των οποίων η συνάρτηση πυκνότητας πιθανότητας του μεγέθους να χαρακτηρίζεται από μια μέση τιμή.

Και στις δυο περιπτώσεις, η απώλεια ενός πακέτου ισοδυναμεί συνήθως με απώλεια τμήματος ή ολόκληρου του καρτέ, ανάλογα με τον τύπο του (δηλ. I, B ή P). Τυπικές περιπτώσεις απώλειας πακέτων είναι :

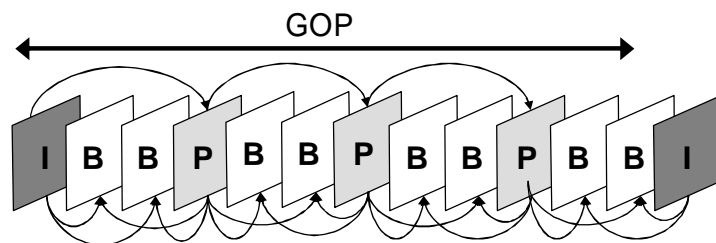
- Το απολεσθέν πακέτο να περιέχει την επικεφαλίδα του καρτέ και δεδομένα
- Το απολεσθέν πακέτο να περιέχει δεδομένα του καρτέ
- Το απολεσθέν πακέτο να περιέχει δεδομένα και το τέλος του καρτέ.

Ανεξάρτητα από το είδος απώλειας, στο παρόν κεφάλαιο για λόγους απλότητας θα θεωρήσουμε ότι μια απώλεια ενός πακέτου –ανεξάρτητα από το περιεχόμενό του- προκαλεί ανεπιτυχή αποκωδικοποίηση του αντιστοίχου καρέ. Το ακριβές σφάλμα που προκαλεί η απώλεια ενός πακέτου στο αποκωδικοποιημένο σήμα εξαρτάται από το είδος του καρέ που συνέβη η απώλεια. Λόγω της αλληλεξάρτησης των καρέ ενός σήματος MPEG, ανάλογα με το είδος του καρέ, στο οποίο ανήκει το απολεσθέν πακέτο, η αντίστοιχη αντιληπτή επίπτωση μπορεί να περιοριστεί μόνο τοπικά (περίπτωση B καρέ), να διαδοθεί σε μετέπειτα καρέ του GOP (περίπτωση P καρέ) ή να διαδοθεί σε όλα τα καρέ του GOP (περίπτωση I καρέ).

Συνεπώς, το σφάλμα που εισάγεται από την απώλεια ενός πακέτου ενδέχεται να διαδοθεί ανάλογα με τη θέση του συγκεκριμένου πακέτου στη ροή του κωδικοποιημένου βίντεο. Το μέγεθος της διάδοσης σφάλματος ενδέχεται να περιοριστεί από μηχανισμούς απόκρυψης σφάλματος, οι οποίοι προσπαθούν να αντικαταστήσουν την απολεσθείσα πληροφορία ώστε να περιορίσουν η διάδοση του σφάλματος στα μεταγενέστερα καρέ.

Στο παρόν κεφάλαιο, το μοντέλο που θα παρουσιαστεί στηρίζεται στην παραδοχή ότι η μέθοδος αποκωδικοποίησης δεν υλοποιεί κάποια μέθοδο απόκρυψης της απώλειας του πακέτου. Συνεπώς οποιαδήποτε διάδοση σφάλματος πραγματοποιείται πλήρως εντός της δομής GOP.

Όπως έχει αναφερθεί στο κεφάλαιο τρία σχετικά με τη δομή ενός σήματος MPEG, κάθε κωδικοποιημένο σήμα περιλαμβάνει τριών ειδών καρέ: I, P και B. Το σύνολο των καρέ που εμπεριέχεται μεταξύ δύο διαδοχικών I καρέ ορίζει ένα GOP και συνήθως περιγράφεται με δύο παραμέτρους $GOP(N,M)$, όπου N είναι το μήκος του, δηλαδή ο αριθμός των καρέ που δομούν το GOP και M-1 το πλήθος των B καρέ μεταξύ των I-P και P-P. Σύμφωνα με αυτό το συμβολισμό, η δομή GOP που απεικονίζεται στο σχήμα 9.1 περιγράφεται ως $GOP(12,3)$.



Σχήμα 9.1 Γραφική απεικόνιση του $GOP(12,3)$

Δηλαδή ο συμβολισμός $GOP(12,3)$ σημαίνει ότι ένα GOP περιλαμβάνει ένα I καρέ, τρία P καρέ και 8 B καρέ. Όπως φαίνεται και στο σχήμα, το δεύτερο I καρέ σηματοδοτεί την έναρξη του επόμενου GOP. Τα απεικονιζόμενα βέλη παρουσιάζουν τις εξαρτήσεις των καρέ, τα οποία παίζουν καθοριστικό ρόλο για τη διάδοση του σφάλματος κατά τη διαδικασία αποκωδικοποίησης.

Λαμβάνοντας υπόψιν την ιεραρχική δομή ενός κωδικοποιημένου σήματος MPEG, ένα καρέ μπορεί να θεωρηθεί ως άμεσα ή έμμεσα μη αποκωδικοποιήσιμο. Άμεσα μη αποκωδικοποιήσιμο είναι ένα καρέ, για το οποίο έχουν απολεσθεί τα πακέτα που μεταφέρουν τα δεδομένα του. Αντίστοιχα, έμμεσα μη αποκωδικοποιήσιμο χαρακτηρίζεται ένα καρέ επειδή τα καρέ από τα οποία εξαρτάται η αποκωδικοποίησή του είναι άμεσα ή έμμεσα μη αποκωδικοποιήσιμα.

Συνοψίζοντας το παρόν εδάφιο, το προτεινόμενο μοντέλο της επίδρασης του ρυθμού απώλειας πακέτων του δικτύου μεταφοράς στην υποβάθμιση της αντιληπτής ποιότητας στηρίζεται στις εξής παραδοχές:

- Η μέθοδος αποκωδικοποίησης δεν υλοποιεί κάποια τεχνική απόκρυψης της απώλειας του πακέτου.
- Το κατώφλι αποκωδικοποίησης λαμβάνεται ίσο με τη μονάδα που σημαίνει ότι απώλεια ενός πακέτου –ανεξάρτητα από το περιεχόμενό του- προκαλεί ανεπιτυχή αποκωδικοποίηση ολόκληρου του καρέ.
- Η οργάνωση των δεδομένων πραγματοποιείται σε πακέτα μεταβλητού μεγέθους με μέγιστη τιμή, που σημαίνει ότι ορίζεται ένα μέγιστο κατώφλι στο μέγεθος που μπορούν να έχουν τα πακέτα.
- Η πιθανότητα απώλειας ενός πακέτου θεωρείται ανεξάρτητη από το είδος του καρέ, του οποίου τα δεδομένα περιέχονται στο πακέτο.
- Θεωρείται σταθερός ρυθμός p απώλειας πακέτων για όλη τη διάρκεια του βίντεο.
- Κάθε πακέτο μεταφέρει μόνο έναν τύπο καρέ.

Το ακόλουθο εδάφιο παρουσιάζει το προτεινόμενο μοντέλο.

Μοντελοποίηση της Απώλειας Πακέτων στο Αποκωδικοποιημένο Βίντεο

Για τη μέτρηση της υποβάθμισης της αντιληπτής ποιότητας στο αποκωδικοποιημένο βίντεο, προς αποφυγήν χρήσης εξαισιωματικών υποκειμενικών τεχνικών, που όπως συζητήθηκαν στο προηγούμενο εδάφιο δεν μπορούν να αξιοποιηθούν για περαιτέρω σύγκριση η χρήση των αποτελεσμάτων, στο προτεινόμενο μοντέλο επιλέχθηκε η χρήση αντικειμενικού μετρικού. Πιο συγκεκριμένα, το ποσοστό αποκωδικοποιημένων καρέ Q (Decodable Frame Rate) [163] επιλέχθηκε για την ποσοτικοποίηση της υποβάθμισης της ποιότητας. Το μετρικό αυτό ορίζεται ως

$$Q = \frac{N_{dec}}{(N_{total-I} + N_{total-P} + N_{total-B})}$$

όπου N_{dec} είναι το πλήθος των επιτυχώς αποκωδικοποιημένων καρέ (δηλαδή N_{dec-I} , N_{dec-P} , και N_{dec-B}).

Σύμβολο	Επεξήγηση
$N_{total-I}$ $N_{total-P}$ $N_{total-B}$	Ο αριθμός των καρέ κάθε τύπου
N_{dec-I} N_{dec-P} N_{dec-B}	Ο αριθμός των επιτυχώς αποκωδικοποιημένων καρέ κάθε τύπου
N_{dec}	Ο συνολικός αριθμός των επιτυχώς αποκωδικοποιημένων καρέ
N_{GOP}	Ο συνολικός αριθμός GOP σε ολόκληρο το βίντεο
C_I C_P C_B	Ο μέσος αριθμός πακέτων που μεταφέρουν τα δεδομένα κάθε τύπο καρέ
p	Ρυθμός απώλειας πακέτων

Πίνακας 9.1 Επεξήγηση των συμβόλων του προτεινόμενου μοντέλου

Στη συνέχεια υπολογίζουμε με στατιστικό τρόπο τα μεγέθη N_{dec-I} , N_{dec-P} , και N_{dec-B} για δομή GOP ίδια με αυτή που απεικονίζεται στο σχήμα 9.1. Για λόγους ευκολότερης παρουσίασης

και κατανόησης των εξισώσεων του προτεινόμενου μοντέλου, ο πίνακας 9.1 συνοψίζει και επεξηγεί όλα τα σύμβολα που θα εμφανίζονται στις εξισώσεις του επόμενου εδαφίου.

Το Αναμενόμενο Πλήθος Αποκωδικοποιημένων Καρέ I (N_{dec-I})

Ένα καρέ I, σύμφωνα με τις παραδοχές του προτεινόμενου μοντέλου, αποκωδικοποιείται επιτυχώς μόνο όταν ληφθούν όλα τα πακέτα που ανήκουν σε αυτό. Κατά συνέπεια η πιθανότητα επιτυχούς αποκωδικοποίησης ενός I καρέ ορίζεται ως

$$s(I) = (1 - p)^{C_I}$$

Άρα, ο αναμενόμενος αριθμός επιτυχώς αποκωδικοποιημένων I καρέ για όλο το βίντεο υπολογίζεται ίσος με

$$N_{dec-I} = (1 - p)^{C_I} * N_{GOP}$$

Το Αναμενόμενο Πλήθος Αποκωδικοποιημένων Καρέ P (N_{dec-P})

Εντός ενός GOP ένα καρέ P είναι αποκωδικοποιήσιμο μόνο εφόσον τα προγενέστερα I και P καρέ έχουν επιτυχώς αποκωδικοποιηθεί και τα πακέτα που το δομούν έχουν παραληφθεί χωρίς απώλειες. Εντός ενός GOP υπάρχουν N_p P καρέ και η πιθανότητα ένα P καρέ να αποκωδικοποιηθεί επιτυχώς είναι

$$\begin{aligned} s(P_1) &= (1 - p)^{C_I} * (1 - p)^{C_P} = (1 - p)^{C_I + C_P} \\ s(P_2) &= (1 - p)^{C_I} * (1 - p)^{C_P} * (1 - p)^{C_P} = (1 - p)^{C_I + 2C_P} \\ &\dots \dots \dots \\ s(P_{N_p}) &= (1 - p)^{C_I} * (1 - p)^{N_p * C_P} = (1 - p)^{C_I + N_p * C_P} \end{aligned}$$

Όθεν, ο αναμενόμενος αριθμός επιτυχώς αποκωδικοποιημένων P καρέ για όλο το βίντεο υπολογίζεται ίσος με

$$N_{dec-P} = (1 - p)^{C_I} * \sum_{j=1}^{N_p} (1 - p)^{jC_P} * N_{GOP}$$

Το Αναμενόμενο Πλήθος Αποκωδικοποιημένων Καρέ B (N_{dec-B})

Εντός ενός GOP ένα καρέ B είναι αποκωδικοποιήσιμο μόνο εφόσον τα προγενέστερα I και P καρέ έχουν επιτυχώς αποκωδικοποιηθεί και τα πακέτα που το δομούν έχουν παραληφθεί χωρίς απώλειες. Εξαιτίας του γεγονότος ότι τα διαδοχικά B καρέ –στη δομή GOP του σχήματος 9.1 που εξετάζεται- έχουν ακριβώς τις ίδιες εξαρτήσεις από τα γειτονικά καρέ, τα εξετάζουμε ως ομάδα B και όχι ανεξάρτητα.

Μόνο η τελευταία B ομάδα έχει διαφορετική εξάρτηση, αφού η επιτυχής αποκωδικοποίησή της εξαρτάται από το προγενέστερο P και το μεταγενέστερο I καρέ, που σημαίνει ότι εμμέσως επηρεάζεται από την επιτυχή αποκωδικοποίηση δύο διαδοχικών I καρέ.

Οπότε, η πιθανότητα ένα B καρέ να αποκωδικοποιηθεί επιτυχώς είναι

$$\begin{aligned}
S(B_1) &= (1-p)^{C_1} * (1-p)^{C_P} * (1-p)^{C_B} \\
S(B_2) &= (1-p)^{C_1} * (1-p)^{2C_P} * (1-p)^{C_B} \\
&\dots\dots\dots \\
S\left(B_{\frac{N}{M}-1}\right) &= (1-p)^{C_1} * (1-p)^{\left(\frac{N}{M}-1\right)*C_P} * (1-p)^{C_B} \\
S\left(B_{\frac{N}{M}}\right) &= (1-p)^{2C_1} * (1-p)^{\left(\frac{N}{M}-1\right)*C_P} * (1-p)^{C_B}
\end{aligned}$$

Επομένως, ο αναμενόμενος αριθμός επιτυχώς αποκωδικοποιημένων Β καρέ για όλο το βίντεο υπολογίζεται ίσος με

$$\begin{aligned}
N_{dec-B} &= (M-1) * \sum_{j=1}^{\frac{N}{M}} S(B_j) * N_{GOP} \\
&= \left[(M-1) * (1-p)^{C_1} * \sum_{j=1}^{N_p} (1-p)^{jC_P} * (1-p)^{C_B} + (M-1) * (1-p)^{2C_1} * (1-p)^{N_p C_P} * (1-p)^{C_B} \right] * N_{GOP} \\
&= \left[(1-p)^{C_1 + N_p C_P} + \sum_{j=1}^{N_p} (1-p)^{jC_P} \right] * (M-1) * (1-p)^{C_1 + C_B} * N_{GOP}
\end{aligned}$$

Σύμφωνα με τις παραπάνω εξισώσεις το μετρικό Q (Ποσοστό Αποκωδικοποιημένων Καρέ) για δίκτυο με σταθερό ρυθμό p απώλειας πακέτων για όλη τη διάρκεια της υπηρεσίας, διαμορφώνεται ως εξής:

$$Q = \frac{N_{dec}}{(N_{total-I} + N_{total-P} + N_{total-B})} = \frac{N_{dec-I} + N_{dec-P} + N_{dec-B}}{(N_{total-I} + N_{total-P} + N_{total-B})} \Rightarrow$$

$$Q = \frac{(1-p)^{C_1} * N_{GOP} + (1-p)^{C_1} * \sum_{j=1}^{N_p} (1-p)^{jC_P} * N_{GOP} + \left[(1-p)^{C_1 + N_p C_P} + \sum_{j=1}^{N_p} (1-p)^{jC_P} \right] * (M-1) * (1-p)^{C_1 + C_B} * N_{GOP}}{(N_{total-I} + N_{total-P} + N_{total-B})}$$

Από τη μαθηματική μορφή της έκφρασης του Q και του ρυθμού απώλειας p εντός αυτής, γίνεται άμεσα αντιληπτό ότι η επίδραση του ποσοστού απώλειας πακέτων στην ποιότητα του μεταδιδόμενου βίντεο είναι μη γραμμική.

Πειραματική Επαλήθευση του Προτεινόμενου Μοντέλου

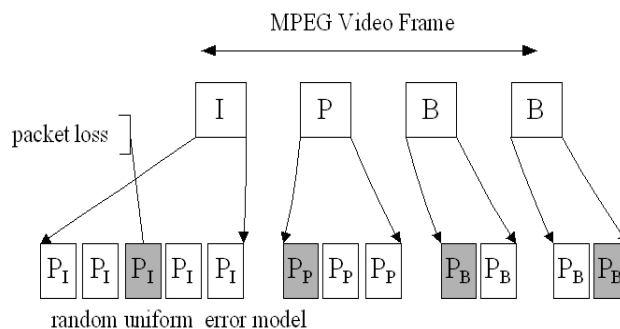
Για την πειραματική επαλήθευση του προτεινόμενου μοντέλου πραγματοποιήθηκε εξομοίωση μιας υπηρεσίας βίντεο ροής σε περιβάλλον επιρρεπές σε λάθη στο λογισμικό NS-2, χρησιμοποιώντας ομοιόμορφη κατανομή ως μοντέλο απώλειας πακέτων [164]. Το βίντεο που χρησιμοποιήθηκε στην προσομοίωση ήταν η παιδική ταινία “Alladin”, η οποία αποτελείται από 89998 καρέ και κωδικοποιήθηκε σύμφωνα με το πρότυπο MPEG-4 με τη δομή GOP IBBPBBPBBPBB (N=12, M=3).

Στατιστικά του βίντεο Alladin	
$N_{total-I}$	7500
$N_{total-P}$	22500
$N_{total-B}$	559998
N_{GOP}	7500
C_I	5.295733333
C_B	2.192506417
C_P	3.087066667

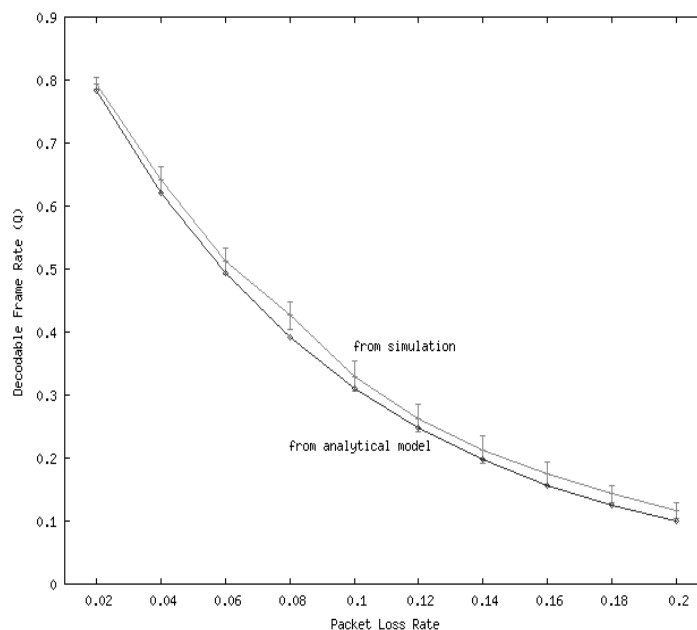
Πίνακας 9.2 Στατιστικά του βίντεο Alladin

Τα στατιστικά χαρακτηριστικά του κωδικοποιημένου σήματος συνοψίζονται στον πίνακα 9.2, ενώ το αντίστοιχο ίχνος βίντεο (video trace) είναι διαθέσιμο στη διεύθυνση [165]. Επιπλέον, καθορίστηκε ότι το μέγιστο μήκος μεταδιδόμενου πακέτου είναι 1000 bytes, τιμή που οδήγησε σε κατακερματισμό των καρτέ σε περισσότερα του ενός πακέτου μεταφοράς.

Για το μοντέλο ομοιόμορφης απώλειας πακέτων, που απεικονίζεται στο σχήμα 9.2, επιλέχθηκε να εξεταστεί το εύρος από 2% έως 20% με βήμα 2%.



Σχήμα 9.2 Σχηματική αναπαράσταση της ομοιόμορφης απώλειας πακέτων



Σχήμα 9.3 Πειραματική και Θεωρητική αναπαράσταση του ποσοστού αποκωδικοποιημένων καρτέ Q συναρτήσει του ρυθμού απώλειας πακέτων

Τα αποτελέσματα της προσομοίωσης απεικονίζονται στο σχήμα 9.3 μαζί με την αντίστοιχη θεωρητική καμπύλη που προκύπτει από τη στατιστική μοντελοποίηση που παρουσιάστηκε στο προηγούμενο εδάφιο. Από τη σύγκριση των δύο καμπυλών προκύπτει το επιτυχές ταίριασμα μεταξύ της θεωρητικά παραγόμενης καμπύλης και της αντίστοιχης πειραματικής, επιβεβαιώνοντας με τον τρόπο αυτό την ορθότητα του αντιστοιχίου πειραματικού μοντέλου. Με τη μεταβολή του ρυθμού απώλειας πακέτων, παρατηρούμε ότι το ποσοστό των επιτυχώς αποκωδικοποιημένων καρτέ μειώνεται, γεγονός που αντιστοιχεί σε ανάλογη ποιοτική υποβάθμιση της αντίστοιχης αντιληπτής ποιότητας.

Το ομοιογενές μοντέλο απώλειας πακέτων που χρησιμοποιήθηκε για την παραγωγή της αντίστοιχης πειραματικής καμπύλης αντιστοιχεί στην εξέταση της χειρότερης περίπτωσης που μπορεί να προκύψει για δεδομένο ρυθμό απώλειας. Το συμπέρασμα αυτό είναι βάσιμο εξαιτίας της υπόθεσης που έχει τεθεί στα πλαίσια του προτεινόμενου μοντέλου, σύμφωνα με την οποία το κατώφλι αποκωδικοποίησης λαμβάνεται ίσο με τη μονάδα που σημαίνει ότι απώλεια ενός πακέτου –ανεξάρτητα από το περιεχόμενό του– προκαλεί ανεπιτυχή αποκωδικοποίηση ολόκληρου του καρτέ. Στην περίπτωση της ομοιόμορφης κατανομής των απωλειών των πακέτων μεταφοράς, θεωρώντας συγκεκριμένο ποσοστό απώλειας, παρατηρείται διάχυση των απωλειών στο σύνολο των καρτέ με περιορισμένη πιθανότητα να επαναληφθεί απώλεια πακέτου εντός του ίδιου καρτέ. Συνεπώς, το σχήμα ομοιογενούς απώλειας πακέτων προκαλεί τη χειρίστη περίπτωση ποιοτικής υποβάθμισης που ισοδυναμεί με το μέγιστο αριθμό ανεπιτυχώς αποκωδικοποιημένων καρτέ.

Επέκταση του Μοντέλου για Πρόβλεψη της Διατερματικής Αντιληπτής Ποιότητας

Το προτεινόμενο στατιστικό μοντέλο εκτίμησης του ποσοστού αποκωδικοποιημένων καρτέ μιας ροής MPEG υπό συνθήκες απώλειας πακέτων μπορεί να αξιοποιηθεί συνδυαστικά με την προτεινόμενη τεχνική πρόβλεψης της αντιληπτής ποιότητας των κεφαλαίων 7 και 8.

Πιο συγκεκριμένα, εξαιτίας του γεγονότος ότι το προτεινόμενο θεωρητικό μοντέλο απεικόνισης του ρυθμού απώλειας πακέτων σε ανεπιτυχώς αποκωδικοποιημένα καρτέ είναι γενικευμένο, δηλαδή δεν εξαρτάται από τη χωρική και χρονική δυναμική του περιεχομένου, η συνδυαστική του χρήση με τη προτεινόμενη μέθοδο πρόβλεψης καθίσταται πρακτικά δυνατή, χωρίς να απαιτείται περαιτέρω προσαρμογή ή αναθεώρηση των προτεινόμενων μοντέλων και μεθόδων.

Έστω ότι ένας υποθετικός πάροχος υπηρεσιών επιθυμεί να διαθέσει μέσω του τηλεπικοινωνιακού του δικτύου ένα μουσικό βίντεο σύντομης διάρκειας σε ποικίλες ποιότητες. Εφαρμόζοντας τη μέθοδο πρόβλεψης του κεφαλαίου 7, μπορεί να καθορίσει τους ρυθμούς κωδικοποίησης που ικανοποιούν τα επιθυμητά επίπεδα ποιότητας.

Στη συνέχεια, ανάλογα με τις συνθήκες που επικρατούν στο δίκτυο μεταφοράς και διανομής (π.χ. ρυθμός απώλειας πακέτων), μπορεί να αξιοποιήσει το προτεινόμενο θεωρητικό μοντέλο του παρόντος κεφαλαίου για να υπολογίσει το μέγεθος της ποιοτικής υποβάθμισης που προκαλείται από τις απώλειες πακέτων. Επί παραδείγματι, εάν η επικρατούσα συχνότητα απώλειας πακέτων είναι 2% στο δίκτυο μεταφοράς, τότε σύμφωνα με το προτεινόμενο μοντέλο, μπορεί να προβλεφθεί ότι η χειρίστη περίπτωση υποβάθμισης είναι ότι ο τελικός χρήστης θα υποστεί ποιοτική υποβάθμιση (εξαιτίας της κατάστασης του δικτύου) για το 20% της συνολικής διάρκειας του βίντεο. Στο προτεινόμενο μοντέλο, υπολογίζεται ότι το αντίστοιχο ποσοστό δεν δύναται να αποκωδικοποιηθεί, που σημαίνει ότι ο χρήστης θα βιώσει διακοπές της υπηρεσίας. Πρακτικώς, το ποσοστό αυτό αντιστοιχεί σε

προβλήματα κατά την αναπαραγωγή και ποιοτική υποβάθμιση. Βεβαίως, πρέπει ιδιαίτερα να τονιστεί ότι ανάλογα με τις ποιοτικές ανοχές του αποκωδικοποιητή και του εξυπηρετητή ροών βίντεο, ενδέχεται πέρα κάποιων ρυθμών απώλειας πακέτων να παρατηρείται διακοπή της υπηρεσίας. Τα όρια αυτά παρουσιάζουν ισχυρή εξάρτηση από τις προδιαγραφές του εκάστοτε συστήματος ροών βίντεο πραγματικού χρόνου και δεν αποτελεί αντικείμενο μελέτης της παρούσης διατριβής.

Δεδομένου της πρόβλεψης για το ποσοστό της ποιοτικής υποβάθμισης της υπηρεσίας βίντεο εξαιτίας των συνθηκών που επικρατούν στο κανάλι μετάδοσης, ο πάροχος δύναται να εκτιμήσει τη μέση τελική ποιότητα που θα παραδοθεί στον τελικό χρήστη, εφαρμόζοντας την ακόλουθη εξίσωση:

$$Ektimwmenh\ Tel\ ikh\ Poiiothta = (Arcikh_Poiiothta) * (Pososto_EpitucwW_Apokwdikopoihmenwn_Kare)$$

Στην προτεινόμενη εξίσωση, το ποσοστό επιτυχώς αποκωδικοποιημένων καρτέ χρησιμοποιείται ως συντελεστής υποβάθμισης της αρχικής ποιότητας. Τονίζεται ότι η παραπάνω σχέση παρέχει ενδεικτική εκτίμηση της τελικής ποιότητας που θα παραλάβει ο χρήστης, χωρίς να λαμβάνονται υπόψη με ποσοτικό τρόπο τα τεχνουργήματα που προκαλούνται στην πράξη από τα ανεπιτυχώς αποκωδικοποιημένα καρτέ.

Συμπεράσματα του Κεφαλαίου

Το παρόν κεφάλαιο περιγράφει, μελετάει και παρουσιάζει θεωρητικό μοντέλο απεικόνισης της επίδρασης του ρυθμού απώλειας πακέτων του δικτύου μεταφοράς στην υποβάθμιση της αντιληπτής ποιότητας βίντεο κωδικοποιημένου σύμφωνα με το πρότυπο MPEG. Πιο αναλυτικά, το προτεινόμενο θεωρητικό μοντέλο παρέχει με αντικειμενικό τρόπο εκτίμηση για το πλήθος των καρτέ που δεν δύνανται να αποκωδικοποιηθούν επιτυχώς στον τελικό χρήστη λόγω απώλειας πακέτων.

Τα πλεονεκτήματα του προτεινόμενου μοντέλου εστιάζονται στη γενικευμένη μορφή του, γεγονός που το καθιστά ανεξάρτητο από τη μέθοδο κωδικοποίησης, τα ακριβή χαρακτηριστικά αυτής, αλλά και από τη χωροχρονική δυναμική του περιεχομένου. Επιπλέον, η χρήση του αντικειμενικού μετρικού του ποσοστού των επιτυχώς αποκωδικοποιημένων καρτέ παρέχει ντετερμινιστικές εκτιμήσεις του ποσοστού υποβάθμισης για συγκεκριμένους ρυθμούς απώλειας πακέτων δικτύου.

Τέλος, παρουσιάζονται οι προοπτικές αξιοποίησης του προτεινόμενου μοντέλου στα πλαίσια παροχής διαθεματικής εκτίμησης της αντιληπτής ποιότητας, λαμβάνοντας υπόψη όλα τα στάδια γένεσης και διαδρομής ενός βίντεο, από το αρχικό στάδιο της κωδικοποίησης, τη μετάδοσή του μέσω του δικτύου μεταφοράς, έως την άφιξή του στον τελικό χρήστη.

ΚΕΦΑΛΑΙΟ 10

Συμπεράσματα και Συνεισφορά της Διατριβής

Σύνοψη των Προτεινόμενων Αναλύσεων, Πρωτότυπων Μεθόδων και Μοντέλων

Ερευνητικά η διατριβή εστιάζεται στο χώρο της ψηφιακής κωδικοποίησης βίντεο σύμφωνα με τα πρότυπα κωδικοποίησης MPEG. Σε αυτά τα πλαίσια διερευνώνται αρχικά οι στατιστικές ιδιότητες ενός σήματος κωδικοποιημένου σύμφωνα με το πρότυπο MPEG-4/H.264, όπου παρουσιάζονται οι υψηλές συσχετίσεις στις οποίες βασίζεται η επίτευξη της συμπίεσης των δεδομένων. Σε αυτά τα πλαίσια προτείνεται επίσης γεννήτρια συνθετικής κίνησης H.264, η οποία προσομοιάζει επιτυχώς τις στατιστικές ιδιότητες του πραγματικού σήματος.

Το κυρίως μέρος της διατριβής εστιάζεται στην εκτίμηση του επιπέδου της αντιληπτής ποιότητας ενός ψηφιακά κωδικοποιημένου βίντεο, προτείνοντας πρωτότυπη μέθοδο πρόβλεψης του επιπέδου της αντιληπτής ποιότητας συναρτήσεως του ρυθμού κωδικοποίησης και της χωρικής και χρονικής δυναμικής του περιεχομένου. Επίσης, προτείνεται θεωρητική μοντελοποίηση της επίδρασης του ποσοστού απώλειας πακέτων του δικτύου μεταφοράς στην ποιοτική υποβάθμιση υπηρεσίας ροής βίντεο πραγματικού χρόνου, η οποία είναι γενικευμένης μορφής και δεν επηρεάζεται από τη δυναμική του περιεχομένου και το είδος της κωδικοποίησης.

Η προτεινόμενη μέθοδος πρωτοτυπεί συγκριτικά με την υπάρχουσα βιβλιογραφία τεχνικών εκτίμησης της ποιότητας βίντεο, διότι επεκτείνεται πέρα από την απλή εκτίμηση της ποιότητας ενός κωδικοποιημένου σήματος, στοχεύοντας στην πρόβλεψη του επιπέδου της ποιότητας που επιτυγχάνεται για δεδομένο περιεχόμενο και συγκεκριμένο ρυθμό κωδικοποίησης. Συνεπώς, πέραν από τις υπάρχουσες τεχνικές εκτίμησης της ποιότητας, η διατριβή πρωτοτυπεί εισάγοντας καινούργια κατηγορία στην υπάρχουσα βιβλιογραφία, αυτή της πρόβλεψης της ποιότητας ενός σήματος πριν πραγματοποιηθεί το τελικό στάδιο κωδικοποίησης.

Πιο αναλυτικά, διερευνάται με αντικειμενικό τρόπο η εξάρτηση του επιπέδου της αντιληπτής ποιότητας συναρτήσεως του ρυθμού κωδικοποίησης και της χωρικής ανάλυσης για ποικίλα επίπεδα της χωρικής και χρονικής δυναμικής συνιστώσας του περιεχομένου. Από τα αντίστοιχα πειραματικά αποτελέσματα, προκύπτει ότι η εξάρτηση μεταξύ της ποιότητας και του ρυθμού κωδικοποίησης είναι αύξουσα και μη γραμμική. Επιπλέον, χαρακτηρίζεται από την ύπαρξη κόρου για κωδικοποιήσεις πέραν ενός καθορισμένου κατωφλίου του ρυθμού κωδικοποίησης, το οποίο εξαρτάται ισχυρά από τη δυναμική του περιεχομένου.

Σύμφωνα με την πειραματική μορφή των καμπύλων, προτείνεται με αναλυτικό τρόπο η χρήση εκθετικής προσέγγισης, η οποία επιτυγχάνει με ικανοποιητική ακρίβεια να περιγράψει επιτυχώς τις ιδιότητες της εξάρτησης της αντιληπτής ποιότητας συναρτήσεως του ρυθμού κωδικοποίησης για ποικίλα επίπεδα δυναμικής του περιεχομένου. Η σύγκριση μεταξύ πειραματικών και θεωρητικών καμπύλων έδειξε ότι το χειρίστο μέσο σφάλμα της προτεινόμενης εκθετικής προσέγγισης είναι μικρότερο του 4%, γεγονός που υποδηλώνει την αξιοπιστία της προτεινόμενης τεχνικής.

Αξιοποιώντας την προτεινόμενη εκθετική προσέγγιση, κάθε πειραματική καμπύλη της ποιότητας συναρτήσει του ρυθμού κωδικοποίησης μπορεί επιτυχώς να περιγραφεί με τη χρήση τριών παραμέτρων που καθορίσουν τη μορφή της αντίστοιχης εκθετικής προσέγγισης. Διερευνώντας την εξάρτηση των παραμέτρων από την πολυπλοκότητα της δυναμικής του περιεχομένου, προτείνεται αναλυτική προσέγγιση των αντιστοιχών πειραματικών αποτελεσμάτων.

Στη συνέχεια, σύμφωνα με τις προαναφερθείσες προτεινόμενες προσεγγίσεις προτείνεται, περιγράφεται και δοκιμάζεται πρωτότυπη μέθοδος πρόβλεψης και εκτίμησης των ρυθμών κωδικοποίησης που ικανοποιούν συγκεκριμένα επίπεδα ποιότητας. Η προτεινόμενη μέθοδος προτείνεται αρχικά στα πλαίσια βίντεο μικρής διάρκειας με ομοιογενές και ετερογενές περιεχόμενο. Η αξιοπιστία της μεθόδου εξετάζεται εκτενώς και αποδεικνύεται ότι παρέχει αξιόπιστα αποτελέσματα για βίντεο βραχείας διάρκειας.

Έπειτα, η προτεινόμενη μέθοδος πρόβλεψης επεκτείνεται σε βίντεο μεγάλης διάρκειας, τα οποία χαρακτηρίζονται εν γένει ως ετερογενή λόγω της έντονης μεταβολής της δυναμικής του περιεχομένου τους με τη πάροδο του χρόνου. Για τις ανάγκες της επέκτασης αναπτύσσεται πρωτότυπη μέθοδος κατάτμησης ενός βίντεο μεγάλης διάρκειας στις στοιχειώδεις σκηνές του, μέσω της ανίχνευσης των ορίων των σκηνών του. Η προτεινόμενη πρωτότυπη μέθοδος στηρίζεται στη χρήση του φαινομένου τετραγώνισης ως μετρικό ανίχνευσης αλλαγής των σκηνών σε βίντεο κωδικοποιημένο σύμφωνα με το πρότυπο MPEG. Οι επιδόσεις της προτεινόμενης μεθόδου συγκρίθηκαν με αντίστοιχες μεθόδους της βιβλιογραφίας, όπου επιδεικνύεται η αντίστοιχη βελτίωση που επιτυγχάνεται.

Η ενσωμάτωση της πρωτότυπης τεχνικής κατάτμησης στη προτεινόμενη μέθοδο πρόβλεψης της ποιότητας αναλύει την περίπτωση ενός βίντεο μακράς διάρκειας σε πολλαπλά ομοιογενή βίντεο βραχείας διάρκειας, τα οποία αντιστοιχούν στις στοιχειώδεις σκηνές από τις οποίες αποτελείται το αρχικό σήμα. Στη συνέχεια, πραγματοποιείται εφαρμογή της μεθόδου ξεχωριστά σε κάθε ανιχνεύσιμη στοιχειώδη σκηνή, γεγονός που ανάγει την όλη διαδικασία στην ήδη γνωστή περίπτωση των ομοιογενών βίντεο βραχείας διάρκειας, στα οποία η προτεινόμενη μέθοδος έχει εξεταστεί και εφαρμοστεί με ικανοποιητική ακρίβεια.

Επιπλέον, η προτεινόμενη επέκταση της μεθόδου μπορεί να αξιοποιηθεί πέρα από τον καθορισμό των ρυθμών κωδικοποίησης που ικανοποιούν τα ποικίλα επίπεδα ποιότητας για τις στοιχειώδεις σκηνές του βίντεο μακράς διάρκειας και ως τεχνική κωδικοποίησης που παρέχει σταθερό αντιληπτό επίπεδο ποιότητας. Όπως είναι ευρέως γνωστό, η μέθοδος κωδικοποίησης παρέχει από τεχνικής μόνο απόψεως κωδικοποίηση σταθερής ποιότητας, χωρίς να λαμβάνεται υπόψη η έννοια της αντιληπτής ποιότητας από τον χρήστη. Αντιθέτως, η προτεινόμενη επεκταμένη μέθοδος πρόβλεψης, μπορεί να αξιοποιηθεί προς αυτή την κατεύθυνση παρέχοντας κωδικοποίηση μεταβλητού ρυθμού και σταθερής αντιληπτής ποιότητας.

Επιπλέον, διερευνάται η σχέση της απόδοσης, σε όρους αντιληπτής ποιότητας, που επιτυγχάνεται μέσω κωδικοποίησης σταθερού και μεταβλητού ρυθμού αντίστοιχα. Σύμφωνα με τη διεθνή βιβλιογραφία, η μέθοδος κωδικοποίησης μεταβλητού ρυθμού παρέχει ποιοτικότερη κωδικοποίηση, εξαιτίας του γεγονότος ότι διατηρούνται σταθεροί οι συντελεστές κβάντισης. Η διατριβή πρωτοτυπεί και διερευνά με ποσοτικό τρόπο την υπεροχή αυτή, αποδεικνύοντας ότι περιορίζεται σε ένα ποσοστό της τάξης του 4%. Επίσης, εξετάζεται η μεταβολή αυτή της υπεροχής για ποικίλους ρυθμούς κωδικοποίησης και χωρικές αναλύσεις.

Η προτεινόμενη αντιστοίχιση του επιπέδου της αντιληπτής ποιότητας που επιτυγχάνεται με τη μέθοδο κωδικοποίησης μεταβλητού ρυθμού έναντι της μεθόδου σταθερού ρυθμού, παρέχει επιπλέον στη διατριβή τη δυνατότητα να μην απαιτούνται πειραματικά αποτελέσματα και για τα δύο είδη των μεθόδων κωδικοποίησης, δεδομένου ότι το επίπεδο της αντιληπτής ποιότητας της μιας μπορεί να αντιστοιχηθεί στο αντίστοιχο επίπεδο της άλλης, αξιοποιώντας την προτεινόμενη σχέση απόδοσης των δύο μεθόδων.

Επίσης, διερευνάται η επίδραση των σφαλμάτων του δικτύου μεταφοράς μέσω της απώλειας πακέτων δεδομένων στο επίπεδο της αντιληπτής ποιότητας υπηρεσίας ροής βίντεο πραγματικού χρόνου. Το προτεινόμενο θεωρητικό μοντέλο πρωτοτυπεί συγκριτικά με τη σχετική βιβλιογραφία διότι η ισχύς του δεν περιορίζεται από το πρότυπο κωδικοποίησης ή τη δυναμική του περιεχομένου της υπηρεσίας που μελετάται. Αντιθέτως, πρόκειται για ένα γενικευμένο μοντέλο απεικόνισης του ρυθμού απώλειας πακέτων στην αντιληπτή ποιότητα της μεταδιδόμενης υπηρεσίας.

Πιο συγκεκριμένα, αναλύοντας τις εξαρτήσεις των καρτέ εντός της δομής GOP ενός βίντεο κωδικοποιημένου σύμφωνα με το πρότυπο MPEG, υπολογίζονται με αναλυτικό τρόπο οι αναμενόμενες απώλειες ανά τύπο καρτέ. Στη συνέχεια εισάγεται το ποσοστό των επιτυχώς αποκωδικοποιημένων καρτέ ως αντικειμενικό μετρικό εκτίμησης της ποιοτικής υποβάθμισης της υπηρεσίας στον τελικό χρήστη. Έπειτα, υπολογίζεται η αντίστοιχη θεωρητική εξάρτηση του ποσοστού επιτυχώς αποκωδικοποιημένων καρτέ συναρτήσει του ρυθμού κωδικοποίησης, η οποία προκύπτει μη γραμμική και φθίνουσα.

Στα πλαίσια της πειραματικής επαλήθευσης των αντιστοιχών θεωρητικών αποτελεσμάτων, πραγματοποιείται εξομοίωση υπηρεσίας ροής βίντεο πραγματικού χρόνου σε περιβάλλον επιρρεπές σε λάθη, μέσω ενός σχήματος απώλειας πακέτων ομοιογενούς κατανομής. Τα εξαγόμενα πειραματικά αποτελέσματα του ποσοστού ποιοτικής υποβάθμισης συναρτήσει του ρυθμού απώλειας πακέτων συγκρίθηκαν με τα αντίστοιχα θεωρητικά, υποδεικνύοντας την αξιοπιστία του προτεινόμενου θεωρητικού πλαισίου.

Αξίζει να σημειωθεί ότι το εξεταζόμενο σχήμα απώλειας πακέτων ομοιογενούς κατανομής αντιστοιχεί στη χειρίστη περίπτωση εισαγωγής σφαλμάτων στο δίκτυο μετάδοσης για μια υπηρεσία βίντεο MPEG, δεδομένου της υπόθεσης του μοντέλου ότι η απώλεια ενός πακέτου που αποτελεί τμήμα ενός καρτέ συνεπάγεται πλήρη απώλεια του καρτέ στον αποκωδικοποιητή.

Τέλος, παρουσιάζεται η δυνατότητα αξιοποίησης του μοντέλου ποιοτικής απεικόνισης του ρυθμού απώλειας πακέτων συνδυαστικά με την προτεινόμενη μέθοδο πρόβλεψης της ποιότητας, στα πλαίσια παροχής διατεματικής εκτίμησης της αντιληπτής ποιότητας υπηρεσίας μιας ροής βίντεο από τη γέννησή της έως την παράδοση αυτής στον τελικό χρήστη.

Ερευνητικά Ζητήματα που Θέτονται για Περαιτέρω Διερεύνηση

Η συνεισφορά της διατριβής στο θεματικό τομέα της εκτίμησης της αντιληπτής ποιότητας ψηφιακά κωδικοποιημένου βίντεο μέσω της προτεινόμενης πρωτότυπης μεθόδου πρόβλεψης της ποιότητας και του μοντέλου ποιοτικής απεικόνισης του ρυθμού απώλειας πακέτων στο επίπεδο της ποιότητας μιας υπηρεσίας ροής βίντεο, δημιουργεί νέα ερευνητικά ερωτήματα και θέτει τις βάσεις για περαιτέρω διερεύνηση.

Πιο συγκεκριμένα, στα πλαίσια μελλοντικής εργασίας, θα διερευνηθεί η περαιτέρω βελτίωση της προτεινόμενης μεθόδου πρόβλεψης της αντιληπτής ποιότητας, ώστε να πραγματοποιηθεί άρση της απαίτησης για πραγματοποίηση μιας δοκιμαστικής κωδικοποίησης σε υψηλό ρυθμό και να δύναται η πρόβλεψη της ποιότητας να πραγματοποιείται απευθείας μέσω της ποσοτικοποίησης της χωρικής και χρονικής δυναμικής του περιεχομένου του ασυμπιέστου βίντεο.

Προς αυτή την κατεύθυνση θα χρειαστεί να διερευνηθεί η εξάρτηση της χωρικής και χρονικής δυναμικής του περιεχομένου του ασυμπιέστου βίντεο και των αντιστοιχών συναρτήσεων ωφέλειας. Το προσδοκώμενο αποτέλεσμα είναι ένας πειραματικός προσδιορισμός της ένα προς ένα αντιστοίχησης της δυναμικής του περιεχομένου στις αντίστοιχες συναρτήσεις ωφελείας.

Στα πλαίσια της περαιτέρω αξιοποίησης της επεκταμένης μεθόδου πρόβλεψης για ετερογενές βίντεο μεγάλης διάρκειας θα διερευνηθεί μέθοδος κωδικοποίησης μεταβλητού ρυθμού, η οποία θα βελτιώνει την αμιγώς από τεχνικής απόψεως επιτυχούσα σταθερή ποιότητα με όρους αντιληπτής ποιότητας από τον τελικό χρήστη. Με τον τρόπο αυτό θα προταθεί πρωτότυπη μέθοδος κωδικοποίησης μεταβλητού ρυθμού που παρέχει κωδικοποίηση σταθερής αντιληπτής ποιότητας.

Από τη διεξαχθείσα έρευνα για τη μοντελοποίηση της επίδρασης του ρυθμού απώλειας πακέτων δεδομένων στο επίπεδο της αντιληπτής ποιότητας σε υπηρεσίας ροής βίντεο πραγματικού χρόνου, περαιτέρω διερεύνηση μπορεί να πραγματοποιηθεί με την εξέταση ποικίλων κατανομών απώλειας και διαφορετικών μεγεθών πακέτων.

Επιπλέον, αξιοποιώντας το προτεινόμενο θεωρητικό μοντέλο απεικόνισης της επίδρασης του ρυθμού απώλειας πακέτων στην αντιληπτή ποιότητα σε συνδυασμό με την προτεινόμενη μέθοδο πρόβλεψης αυτής, δύναται να προταθεί ένα θεωρητικό πλαίσιο πρόβλεψης της διατεμαστικής ποιότητας για υπηρεσίες βίντεο ροής πραγματικού χρόνου.

ΑΝΑΦΟΡΕΣ

1. Ghanbari, M., *Standard Codecs: Image Compression to Advance Video Coding*. IEE Telecommunication series. 2003.
2. Richardson, Iain E. G., *H.264 and MPEG-4 Video Compression*, ed. Wiley. 2004.
3. Winkler, Stefan, *Digital Video Quality*, ed. Wiley. 2005.
4. H. Koumaras, F. Liberal, L. Sun, *PQoS Assessment Methods for Multimedia Services*. Chapter contribution in "Wireless Multimedia: Quality of Service and Solutions", 2007.
5. Z. Wang, H. R. Sheikh and A. C. Bovik *Objective Video Quality Assessment*, in *The Handbook of Video Databases: Design and Applications*, B.F.a.O. Marqure, Editor. 2003, CRC Press. p. 1041-1078.
6. VQEG, *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment*. 2000.
7. L. Lu, Z. Wang, A. C. Bovik and J. Kouloheris *Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of MPEG video*, in *IEEE International Conference on Multimedia*. 2002.
8. Stefan Winkler, Animesh Sharma, David McNally, *Perceptual Video Quality and Blockiness Metrics for Multimedia Streaming Applications*.
9. Ghanbari, I. Pr. Guawan and M. *Reduced-Reference Picture Quality Estimation by Using Local Harmonic Amplitude Information*. in *Communications Symposium*. 2003. London.
10. Zunino, Paolo Gastaldo and Rodolfo, *Neural networks for the no-reference assessment of perceived quality*. Journal of Electronic Imaging, 2005. 14(3): p. 033004
11. H. Koumaras, E. Pallis, G. Xilouris, A. Kourtis, D. Martakos, *Perceived Quality Evaluation for Multimedia Services*, in *Encyclopedia of Mobile Computing & Services*. 2007, Idea Group Pub.
12. Z. Wang, A. C. Bovik and L. Lu, *Why is Image Quality Assessment so Difficult*. International Conference in Acoustics, Speech and Signal Processing, 2002. 4: p. 3313-3316.
13. A. B. Watson, J. Hu and J. F. McGowan, *DVQ: A Digital Video Quality Metric Based on Human Vision*. Journal of Electronic Imaging, 2001. 10(1): p. 20-29.
14. Z. Wang, E. P. Simoncelli, *Reduced-reference image quality assessment using a wavelet-domain natural image statistic model*. Proc. of SPIE Human Vision and Electronic Imaging, 2005. 5666: p. 149-159.
15. Lauterjung, J. *Picture Quality Measurement*. in *Proceedings of the International Broadcasting Convention (IBC)*. 1998. Amsterdam.
16. P. Seeling, M. Reisslein and B. Kulapala *Network Performance Evaluation Using Frame Size and Quality Traces of Single Layer and Two Layer Video: A Tutorial*. IEEE Communications Surveys, Third Quarter 2004. 6(3).
17. H. Koumaras, A. Kourtis, D. Martakos, J. Lauterjung, *Quantified PQoS Assessment Based on Fast Estimation of the Spatial and Temporal Activity Level*. Multimedia Tools and Applications, 2007. 34(3): p. 355-374.
18. MPEG-Converer, *Terms of Reference. Doc. ISO/MPEG N4500*. 2001: Pattaya MPEG Meeting.
19. Pereira, F., *The MPEG-4 Book*. 2002: Prentice Hall.
20. Sullivan, Gary J. *Overview of International Video Coding Standards*. in *ITU-T VICA Workshop*. 2005. ITU Headquarter, Geneva.
21. H. Koumaras, E. Pallis, A. Kourtis, D. Martakos, *Principles of Digital Video Coding*. Encyclopedia of Information Science & Technology 2nd edition, IGI Press, 2007.

22. Borko Furht, Joshua Greenberg, Raymond Westwater, *Motion Estimation Algorithms for Video Compression*. The International Series in Engineering and Computer Science, Springer, 1996.
23. E. Chan, S. Panchanathan, *Review of block matching based motion estimation algorithms for video compression*. Canadian Conference on Electrical and Computer Engineering, 1993. **1**: p. 151 - 154.
24. Blinn, Jim, *What's the deal with the DCT*. IEEE Computer Graphics and Applications, July 1993: p. 78-83.
25. Huffman, D. , *A method for the construction of minimum redundancy codes*. Proceedings of the IRE, 1952. 40: p. pp. 1098–1101.
26. H. Witten, M. Neal and G. Cleary *Arithmetic coding for data compression*. Communications of the ACM, 1987. 30(6): p. pp.520-540.
27. G. Held, T. Marshall. , *Data Compression*. Wiley, 1991. ISBN 0 471 92941 7.
28. M. Yuen, H. R. Wu *A survey of hybrid MC/DPCM/DCT video coding distortions*. Signal Processing. 70(3): p. 247-278.
29. ITU, *Subjective video quality assessment methods for multimedia applications*. Recommendation ITU-T P.910, 1999.
30. MOSAIC, *A new single stimulus quality assessment methodology*. RAC R2111, 1996.
31. J. Alpert, P. Evain, *Subjective Quality Evaluation – The SSCQE and DSCQE methodologies*. EBU Technical Review Spring 1997, 1997.
32. ITU, *Methodology for the Subjective Assessment of the Quality of Television Pictures*. Recommendation ITU-R BT.500-10, 2000.
33. J. Mullin, L. Smallwood, A. Watson and G. Wilson, *New techniques for assessing audio and video quality in real-time interactive communications*. Third International Workshop on Human Computer Interaction with Mobile Devices, Lille, France, 2001.
34. Buxton, W., *Integrating the periphery and context: A new taxonomy of telematics*, in *Graphics Interface 1995*.
35. Winkler, Stefan, *Issues in vision modeling for perceptual video quality assessment*. Signal Processing, 1999. 70: p. 247-248.
36. Ulrich Engelke, Hans-Jürgen Zepemick, *Perceptual Quality Measures for Image and Video Services*. Euro-NGI Workshop on Socio-Economic Aspects of Next Generation Internet, Lyngby, Denmark, October 9-10, 2006.
37. Daly, S., *The Visible Difference Predictor: An algorithm for the Assessment of Image Fidelity*. in Proceedings SPIE, 1992. 1616: p. 2-15.
38. Daly, S., *The visible difference predictor: An algorithm for the assessment of image fidelity*. Digital Images and Human Vision, 1993. The MIT Press: p. 179-206.
39. Lubin, J., *The use of psychological data and models in the analysis of display system performance*. Digital Images and Human Vision, 1993. The MIT Press: p. 163-178.
40. Watson, A. B., *DCTune: A technique for visual optimization of DCT matrices for individual images* Society for information display digest of technocal papers, 1993. XXIV: p. 946-949.
41. K. T. Tan, M. Ghanbari and D.E. Pearson, *An objective measurement tool for MPEG video quality*. Signal Processing, 1998. **70**: p. 279-294.
42. Winkler, Stefan, *A perceptual distortion metric for digital color video*. Proceedings of SPIE, 1999: p. 175-184.
43. Watson, A. B., *Toward a perceptual video quality metric*. Proceedings of SPIE Human Vision and Electronic Imaging, 1998. **3299**: p. 19-147.
44. K. T. Tan, M. Ghanbari, *A Multi-Metric Objective Picture Quality Measurement Model for MPEG Video*. IEEE TRANSACTIONS ON VIDEO TECHNOLOGY, 2000. **10**(7): p. 1208-1213.

45. Z. Wang, L. Lu, A. C. Bovik, *Video Quality Assessment Based on Structural Distortion Measurement*. Signal Processing: Image Communication, special issue on Objective video quality metrics, 2004. 19(2): p. 121-132.
46. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*. IEEE TRANSACTIONS ON IMAGE PROCESSING, 2004. 13(4): p. 1-14.
47. M. Ries, C. Crespi, O. Nemethova, M. Rupp, *Content Based Video Quality Estimation for H.264/AVC Video Streaming*. Proc. Proceedings of IEEE Wireless and Communications & Networking Conference, Hong Kong, 2007.
48. Eric A. Silva, Karen Panetta, Sos S. Agaian, *Quantifying image similarity using measure of enhancement by entropy*. Mobile Multimedia/Image Processing for Military and Security Applications 2007, Sos S. Agaian, Sabah A. Jassim, Editors, 65790U, Proceedings of SPIE, 2007. 6579.
49. T. M. Kusuma, H.-J. Zepernick, *A reduced-reference perceptual quality metric for in-service image quality assessment*. Joint First Workshop on Mobile Future and Symposium on Trends in Communications, 2003: p. 71-74.
50. U. Engelke, H. J. Zepernick, T. M. Kusuma, *Perceptual evaluation of motion JPEG2000 quality over wireless channels*. Proc. of Symp. on Trends in Communications, 2006: p. 92-96.
51. P. Le Callet, C. Viard-Gaudin, D. Barba, *Continuous quality assessment of MPEG2 video with reduced reference*. Proc. of 1st Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, 1995.
52. M. Montenovo, A. Perot, M. Carli, P. Cicchetti, A. Neri, *Objective evaluation of video services*. Proc. of 2nd Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2006.
53. S. S. Hemami, M. A. Masry, *A scalable video quality metric and applications*. Proc. of 1st Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2005.
54. O. A. Lotfallah, M. Reisslein, S. Panchanathan, *A framework for advanced video traces: Evaluating visual quality for video transmission over lossy networks*. EURASIP Journal on Applied Signal Processing, 2006. 2006(Article ID 42083).
55. Zhou Wang, Guixing Wu, Hamid R. Sheikh, Eero P. Simoncelli, En-Hui Yang and Alan C. Bovik, *Quality-Aware Images*. IEEE TRANSACTIONS ON IMAGE PROCESSING.
56. M. H. Pinson, S. Wolf, *A new standardized method for objectively measuring video quality*. IEEE Trans. on Broadcasting, 2004. 50(3): p. 312-322.
57. American National Standards Institute, *American national standard for telecommunications - digital transport of one-way video signals - parameters for objective performance assessment*. ANSI T1.801.03, 2003.
58. Michal Ries, Olivia Nemethova, Markus Rupp, *Motion Based Reference-Free Quality Estimation for H.264/AVC Video Streaming*. International Symposium on Wireless Pervasive Computing, San Juan, Poerto Rico, 2007.
59. H. R. Wu, M. Yuen, *A generalized block-edge impairment metric for video coding*. IEEE Signal Processing Letters, 1997. 4(11): p. 317-320.
60. S. Liu, A. C. Bovik, *Efficient dct-domain blind measurement and reduction of blocking artifacts*. IEEE Transactions on Circuits and Systems for Video Technology, 2002. 12(12): p. 1139-1149.
61. P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahim, *A no-reference perceptual blur metric*. in Proc. of IEEE Int. Conf. on Image Processing, 2002. 3: p. 57-60.
62. X. Marichal, W. Y. Ma, H. J. Zhang, *Blur determination in the compressed domain using DCT information*. in Proc. of IEEE Int. Conf. on Image Processing, 2002. 2: p. 386-390.

63. R. Ferzli, L. J. Karam, J. Caviedes, *A robust image sharpness metric based on kurtosis measurement of wavelet coefficients*,. Proc. of 1st Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2005.
64. J. Caviedes, S. Gurbuz, *No-reference sharpness metric based on local edge kurtosis*. in Proc. of IEEE Int. Conf. on Image Processing, 2002. 3: p. 53-56.
65. Z. Wang, H. R. Sheikh, A. C. Bovik, *No-reference perceptual quality assessment of JPEG compressed images*,. in Proc. of IEEE Int. Conf. on Image Processing, 2002. 1: p. 477-480.
66. M. C. Q. Farias, S. K. Mitra, *No-reference video quality metric based on artifact measurements*. in Proc. of IEEE Int. Conf. on Image Processing, 2002. 3: p. 141-144.
67. R. Barland, A. Saadane, *A reference free quality metric for compressed images*. in Proc. of 2nd Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2006.
68. A. Cavallaro, S. Winkler, *Segmentation-driven perceptual quality metrics*. in Proc. of IEEE Int. Conf. on Image Processing, 2004. 5: p. 3543-3546.
69. R. R. Pastrana-Vidal, J. C. Gicquel, *Automatic quality assessment of video fluidity impairments using a no-reference metric*. in Proc. of 2nd Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2006.
70. M. Ries, O. Nemethova, M. Rupp, *Reference-free video quality metric for mobile streaming applications*. in Proc. of 8th Int. Symp. on DSP and Communication Systems & 4th Workshop on the Internet, Telecommunications and Signal Processing, 2005: p. 98-103.
71. M. Caramma, R. Lancini, M. Marconi, *Subjective quality evaluation of video sequences by using motion information*. in Proc. of IEEE Int. Conf. on Image Processing, 1999. 2: p. 313-316.
72. M. C. Q. Farias, S. K. Mitra, M. Carli, *Video quality objective metric using data hiding*. in Proc. of IEEE Workshop on Multimedia Signal Processing, 2002: p. 464-467.
73. Y. Fu-zheng, W. Xin-dai, C. Yi-lin, W. Shuai, *A no-reference video quality assessment method based on digital watermark*. in Proc. of 14th IEEE Conf. on Personal, Indoor and Mobile Radio Communications, 2003. 3: p. 2707-2710.
74. S. Wang, D. Zheng, J. Zhao, W. J. Tam, F. Speranza, *An Image Quality Evaluation Method Based on Digital Watermarking*. IEEE transactions on Circuits Systems Video Technology, 2007. **17**(1).
75. P. Fleury, J. Reichel, T. Ebrahimi, *Image quality prediction for bitrate allocation*. in Proc. of IEEE Int. Conf. on Image Processing, 1996. 3: p. 339-342.
76. P. Gastaldo, S. Rovetta, R. Zunino, *Objective quality assessment of MPEG2 video streams by using CBP neural networks*. IEEE Trans. on Neural Networks, 2002. 13(4): p. 939-947.
77. S. Mohamed, G. Rubino, *A study of real-time packet video quality using random neural networks*. IEEE Trans. on Circuits and Systems for Video Technology, 2002. 12(12): p. 1071-1083.
78. T. Lacshman, A. Ortega, A. Reibman, *VBR Video: Tradeoffs and Potentials*. Invited Paper, Proceedings of the IEEE, 1998. 86(5).
79. F. Yegenoglu, B. Jabbari and Y.-Q. Zhang, *Motion-classified autoregressive modeling of variable bit rate video*. IEEE transactions on Circuits Systems Video Technology, 1993. **3**(1): p. pp42-53.
80. H.S. Chin, J.W. Goodge, R. Griffiths and D.J. Parish, *Statistics of video signals for viewphone-type pictures*. IEEE Journal on Selected Areas in Communications, 1989. **7**(5): p. 826-832.
81. W. Verbiest, L. Pinnoo and B. Voeten, *The impact of the ATM concept on video coding*. IEEE Journal on Selected Areas in Communications, 1988. **6**(9): p. 1623-1632.

82. C. Skianis, K. Kontovasilis, A. Drigas and M. Moatsos, *Measurement and Statistical Analysis of Asymmetric Multipoint Videoconference Traffic in IP Networks*. Telecommunication Systems, 2003. 23: p. 95-122.
83. N.D. Doulamis, A.D. Doulamis, G.E. Konstantoulakis and G.I. Stassinopoulos, *Efficient Modeling of VBR MPEG-1 Coded Video Sources*. IEEE Transactions on Circuits Systems Video Technology, 2000. 10(1): p. 93–112.
84. D.P. Heyman, A. Tabatabai and T.V. Lakshman, *Statistical analysis and simulation study of video teleconference traffic in ATM networks*. IEEE Transactions on Circuits Systems Video Technology, 1992. 2(1): p. 49–59.
85. B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J.D. Robbins, *Performance models of statistical multiplexing in packet video communications*. IEEE Transactions on Communications, 1988. 36(7): p. 834–843.
86. M. Nomura, T. Fujii and N. Ohta, *Basic characteristics of variable rate video coding in ATM environment*. IEEE Journal on Selected Areas in Communications, 1989. 7(5): p. 752–760.
87. Heyman, D.M. Cohen and D.P., *Performance modeling of video teleconferencing in ATM networks*. IEEE Transactions on Circuits Systems Video Technology, 1993. 3(6): p. 408–422.
88. Haskell, B.G., *Buffer and channel sharing by several interframe picturephone coders*. Bell Systems Technical Journal, 1972. 51(1): p. 261-289.
89. D.M. Lucantoni, M.F. Neuts and A.R. Reibman, *Methods for performance evaluation of VBR video traffic models*. IEEE/ACM Transactions on Networking, 1994. 2(2): p. 176–180.
90. Sengupta, G. Ramamurthy and B. *Modeling and analysis of a variable bit rate video multiplexer*. in *Proc. of the 7th Internat. Teletraffic Congress Seminar*. 1990. Morristown, NJ.
91. R.M. Rodriguez-Dagnino, M.R.K. Khansari and A. Leon-Garcia, *Prediction of bit rate sequences of encoded video signals*. IEEE Journal on Selected Areas in Communications, 1991. 9(3): p. 305–314.
92. T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, *Overview of the H.264/AVC Video Coding Standard*. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue in H.264, 2003.
93. B. Girod, E. Steinbach, N. Faerber, *Performance of the H.263 Video Compression Standard*. Journal of VLSI Signal Processing” Systems for Signal, Image and Video Technnology. Special issue on recent development in video: Algorithms, Implementation and Applications, 1997. 17: p. 101-111.
94. E. Steinbach, N. Faerber, B. Girod, *Standard Compatible Extension of H.263 for robust video transmission in Mobile environments*. IEEE Transactions on Circuits and Systems for Video Technology, 1997. 7(6): p. 872-881.
95. G. J. Sullivan, P. Topiwala, and A. Luthra. *The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions*. in *SPIE Annual Conference on Applications of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard H.264/AVC*. 2004. China.
96. Y. Su, M. T. Sun. *Encoder Optimization for H.264/AVC Fidelity Range Extensions*. in *Proceedings of VCIP 2005*. 2005. China.
97. Sullivan, G., *The H.264/MPEG-4 AVC video coding standard and its deployment status*, in *Proceedings of VCIP 2005*. 2005: China.
98. Y.L. Lai, Y.Y. Tseng, C. W. Lin, Z. Zhou, M.T. Sun. *H.264 Encoder Speed-UP via Joint Algorithm/Code-Level Optimization*. in *in Proceedings of VCIP 2005*. 2005. China.
99. Doggen Jeroen, Van der Schueren Filip, *MPEG4 Part 10 - H.264 Modeling in OPNET*. Masters Thesis, Department of Industrial Sciences and Technology, University College of Antwerpen, Belgium, 2006.

100. Haag, M. *Stationary and Nonstationary Random Processes*, online available at <http://cnx.rice.edu/content/m10684/latest>. 2002 [cited 2002].
101. Miller, D.R. Cox and H.D., *The Theory of Stochastic Processes*, ed. C. Hall. 1965.
102. Alheraish, A., *Autoregressive video conference models*. International Journal of Network Management, 2004. 14: p. 329-337.
103. S. Xu, Z. Hung, *A gamma autoregressive video model on ATM networks*. IEEE Transactions on Circuits and Systems for Video Technology, 1996. 8(4): p. 138-142.
104. McKenzie, E., *Autoregressive moving average processes with negative binomial and geometri mariginal distributions*. Advances in Applied Probability, 1986. 18: p. 679-705.
105. Zhang, Q., *A general AR-based technique for the generation of arbitrary gamma VBR video traffic in ATM networks*. IEEE Transactions on Circuits and Systems for Video Technology, 1999. 9(10): p. 1130-1137.
106. P. Sen, B. Maglaris, N. E. Rikli and D. Anastasiou, *Models for packet switching of variable bit rate video in ATM networks*. IEEE Journal on Selected Areas in Communications, 1989. 7(5): p. 865-869.
107. H. Koumaras, C. Skianis, G. Gardikis, A. Kourtis, *Analysis of H.264 video encoded traffic*. INC 2005 Fifth International Network Conference, Samos Island, Greece, 2005.
108. R. Grunenfelder, J. P. Cosmas, S. Manthorpe, A. Odinma-Okafor, *Characterization of Video Codecs as Autoregressive Moving Average processes and Related Queuing System Performance*. IEEE Journal on Selected Areas in Communications, 1991. 9(3): p. 284-293.
109. Sengupta, G. Ramamurthy and B., *Modeling and analysis of a variable bit rate video multiplexer*. Proc. of the 7th Internat. Teletraffic Congress Seminar, 1990.
110. C. Shim, I. Ryoo, J. Lee and S. Lee, *Modeling and call admission control algorithm of variable bit rate video in ATM Networks*. IEEE Journal on Selected Areas in Communications, 1994. 12(2): p. 332-344.
111. M. Dai, D. Loguinov *Analysis and modeling of MPEG-4 and H.264 multi-layer video traffic*. INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, 2005. 4: p. 2257-2267.
112. F. Liberal, A. Perro, J. Fajardo, *PQoS Based Model for Assessing Significance Statistically*, in *Third International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs 05)*. 2005: Ilkley U.K.
113. F. Pereira, T. Alpert, *MPEG-4 Video Subjective Test Procedures and Results*. IEEE TRANSACTIONS ON VIDEO TECHNOLOGY, 1997. 7(1): p. 32-51.
114. T. Alpert, L. Contin, *DSCQE Experiment for the Evaluation of the MPEG-4 VM on Error Robustness Functionality ISO/IEC - JTC1/SC29/WG11, MPEG97/M1604*, 1997, 1997.
115. Stephen Wolf, Margaret H. Pinson. *Spatial - Temporal Distortion Metrics for in0service Quality Monitoring of any Digital Video System*. in *SPIE International Symposium on Voice, Video and Data Communications*. 1999. Boston.
116. Bradley, A. P., *A Wavelet Difference Predictor*. IEEE TRANSACTIONS ON IMAGE PROCESSING, 1999. 5: p. 717-730.
117. Y. K. Lai, J. Kuo, *A Haar Wavelet Approach to Compressed Image Quality Measurement*. Journal of Visual Communication and Image Understanding, 2000. 11: p. 81-84.
118. Baddeley, A J, *An error metric for binary images*, in *Inernational Workshop on Robust Computer Vision 1992*: Bonn.
119. N. Cranley, L. Murphy, *Incorporating User Perception in Adaptive Video Streaming Systems*. in *Digital Multimedia Perception and Design* (Eds. G. Ghinea and S. Chen), 2006.
120. W. Lee, J. Srivastava, *An Algebraic QoS-Based Resource Allocation Model for Competitive Multimedia Applications*. International Journal of Multimedia Tools and Applications, 2001. 13(197-212).

121. B. Sabata, S. Chatterjee, J. Sydir, *Dynamic Adaptation of Video for Transmission under Resource Constraints*. International Conference of Image Processing, Chicago, 1998.
122. M. H. Pinson, S. Wolf, *The Impact of Monitor Resolution and Type on Subjective Video Quality Testing*. NTIA Technical Memorandum TM-04-412, U.S. DEPARTMENT OF COMMERCE, 2004.
123. R. Lienhart, *Comparison of Automatic Shot Boundary Detection Algorithms*. In Image and Video Processing VII 1999, Proc. SPIE, 1999: p. 3656-29.
124. R. lienhart, *Methods for Content Analysis, Indexing and Comparison on Digital Video Sequences*. Ph.D. thesis, Shaker Verlag GmbH, ISBN 3-8265-4125-1, 1998.
125. Lu, Y-P. Tan, *An Effective Post-Refinement Method for Shot Boundary Detection*. IEEE Transactions on Circuits and Systems for Video Technology, 2005. 15(11): p. 1407-1421.
126. A. Hampapur, R. Jain, T. E. Weymouth, *Production Model Based Digital Video Segmentation*. Multimedia Tools and Applications, 1995. 1(1): p. 9-46.
127. H. Li, G. Liu, Z. Zhang, Y. Li, *Adaptive Scene-Detection Algorithm for VBR Video Stream*. IEEE Transactions on Multimedia, 2004. 6(4).
128. I. Koprinska, S. Carrato, *Temporal Video Segmentation: A survey*. Signal Processing: Image Communication, 2001. 16: p. 477-500.
129. A. Dailianas, R. B. Allen, P. England, *Comparison of automatic video segmentation algorithms*. Proceedings of SPIE, 1995. 2615: p. 2-16.
130. H. Ueda, T. Miyatake, S. Yoshizawa, *IMPACT: An interactive Natural-motion-picture Dedicated Multimedia Authoring System*. in proceedings of CHI, ACM, 1991: p. 343-350.
131. H.J. Zhang, A. Kankanhalli, S.W. Smoliar, *Automatic Partitioning of Full-motion Video*. Multimedia Systems, 1993. 1(1): p. 10-28.
132. R. Zabih, J. Miller, K. Mai, *A Feature-Based Algorithm for Detecting and Classifying Scene Breaks*. Proc. ACM Multimedia, San Francisco, CA, 1993: p. 189-200.
133. B. Yeo, B. Liu, *Rapid Scene Analysis on Compressed Video*. IEEE Transactions on Circuits and Systems for Video Technology, 1995. 5(6).
134. J. Meng, Y. Juan, S-F Chang, *Scene Change Detection in a MPEG compressed Video Sequence*. Proceedings of SPIE Int. Symp. Electronic Imaging 2417, San Jose, 1995: p. 14-25.
135. M. Krunz, S. K. Tripathi, *Scene-based characterization of VBR MPEG-compressed video traffic*. Sigmetrics 97, Seattle, WA, 1997.
136. M. Krunz, H. Hughes, *A traffic model for MPEG-coded VBR streams*. Sigmetrics 95, Ottawa, Canada, 1995.
137. B. Melamed, D. E. Pendarakis, *Modeling full-length VBR video using markov-renewal-modulated TES models*. IEEE Journal in Selected Areas of Communications, 1998. 16: p. 600-611.
138. S. J. Yoo, S. D. Kim, *Scene-based traffic modeling queueing analysis of MPEG video sequences*. Electronics Letters, 1998. 34(15): p. 1484-1485.
139. H. Lu, Y. Tan, *An effective Post-Refinement Method for Shot boundary Detection*. IEEE Transactions on Circuits and Systems for Video Technology, 2005. 15(11).
140. C. O'Toole, A. Smeaton, N. Murphy, S. Marlow, *Evaluation of automatic shot boundary detection on a large video suite*. 2nd U.K. Conf. Image Retrieval: The challenge of Image Retrieval, Newcastle, U.K., 1999.
141. W. J. Tam, L. Stelmach, L. Wang, D. Lauzon, P. Gray, *Visual masking at video scene cuts*. Proceedings of SPIE, 1995. 2411: p. 111-119.
142. I. K. Sethi, N. V. Patel, *A statistical approach to scene change detection*. Proceedings of SPIE, 1995. 2420: p. 329-339.

143. E. Casilari, A. Reyes, A. Díaz-Estrella, F. Sandoval, *Heavy-tailed distribution of scene duration in VBR video*. Electronics Letters, 1999 35(2).
144. Ullas Gargi, Rangachar Kasturi, Susan H. Strayer, *Performance characterization of Video-Shot-Change Detection Methods*. IEEE transactions on Circuits Systems Video Technology, 2000. 10(1): p. 1-11.
145. B.-L. Yeo, B. Liu, *A unified approach to temporal segmentation of motion JPEG and MPEG compressed video*. in Proc. 2nd Int. Conf. Multimedia Computing and Systems, 1995: p. 81-83.
146. K. Shen, E. J. Delp, *A fast algorithm for video parsing using MPEG compressed sequences*. IEEE Int. Conf. Image Processing, 1995: p. 252-255.
147. S. Kanumuri, P. C. Cosman, A.R. Reibman, V.A. Vaishampayan, *Modeling Packet-Loss Visibility in MPEG-2 Video*. IEEE Transactions on Multimedia, 2006. 8(2): p. 341-355.
148. Z. He, H. Xong, *Transmission Distortion Analysis for Real-Time Video Encoding and Streaming over Wireless Networks*. IEEE Transactions on Circuits and Systems for Video Technology, 2006. 16(9): p. 1051-1062.
149. P. Gastaldo, S. Rovetta, R. Zunino, *Objective quality assessment of MPEG-2 video streams by using CBP neural networks*. IEEE Trans. Neural Networks, 2002. 13: p. 939-947.
150. O. Verscheure, P. Frossard, M. Hamdi, *Joint impact of MPEG-2 encoding rate and ATM cell losses on video quality*. in Global Telecommunications Conf. (GLOBECOM), 1998. 1: p. 71-76.
151. O. Verscheure, P. Frossard, M. Hamdi, *User-oriented QoS analysis in MPEG-2 video delivery*. Journal of Real-Time Imaging, 1999. 5: p. 305-314.
152. J. Lu, M. Chatterjee, M. D. Schwartz, M. K. Ravel, W. M. Osberger, *Measuring ATM video quality of service using an objective picture quality model*. in Proc. SPIE, Multimedia Systems and Applications II, 1999. 3845: p. 290-297.
153. A. E. Conway, Y. Zhu, *Applying objective perceptual quality assessment methods in network performance modeling*. in Proc. Eleventh Int Conf. on Computer Communications and Networks, 2002: p. 116-123.
154. G. W. Cermak, *Videoconferencing Service Quality as a Function of Bandwidth, Latency, and Packet Loss*. Verizon Laboratories, T1A1.3/2003-026, 2003.
155. B. Chen, J. Francis, *Multimedia performance evaluation*. AT&T Tech. Mem, 2003.
156. C. J. Hughes, M. Ghanbari, D. E. Pearson, V. Seferidis, J. Xiong, *Modeling and subjective assessment of cell discard in ATM video*. IEEE Trans. Image Processing, 1993. 2: p. 212-222.
157. S. Winkler, R. Campos, *Video quality evaluation for internet streaming applications*. in Proc. SPIE, Human Vision and Electronic Imaging VIII, 2003. 5007: p. 104-115.
158. K. Brunnstrom, B. N. Schenkman, *Quality of video affected by packet loss distortion, compared to the predictions of a spatio-temporal model*. in Proc. SPIE, Human Vision and Electronic Imaging VIII, 2002. 4662: p. 149-158.
159. M. S. Moore, J. M. Foley, S. K. Mitra, *Detectability and annoyance value of MPEG-2 artifacts inserted into uncompressed video sequences*. in Proc. SPIE, Human Vision and Electronic Imaging V, 2000. 3959: p. 99-110.
160. M. S. Moore, S. K. Mitra, J. M. Foley, *Defect visibility and content importance implications for the design of an objective video fidelity metric*. in Proc. IEEE ICIP, 2002. 3: p. 45-48.
161. M. G. Ramos, S. S. Hemami, *Suprathreshold wavelet coefficient quantization in complex stimuli: Psychophysical evaluation and analysis*. Journal of Optics Society Amer., 2001. 18(10): p. 2385-2397.

162. D. Chandler, S. S. Hemami, *Effects of natural images on the detectability of simple and compound wavelet subband quantization distortion*. Journal of Optics Society Amer., 2003. 20(7): p. 1164–1180.
163. A. Ziviani, B.E. Wolfinger, J. F. Rezende, O. C. Duarte, and S. Fdida, *Joint Adoption of QoS schemes for MPEG streams*. Multimedia Tools and Applications. to appear.
164. Cheng-Han Lin, Chih-Heng Ke, Ce-Kuen Shieh, Naveen Chilamkurti, *The Packet Loss Effect on MPEG Video Transmission in Wireless Networks*. The IEEE 20th International Conference on Advanced Information Networking and Applications (AINA'06), Vienna, Austria, 2006.
165. Video Trace Files, available online at http://www.tkn.tu-berlin.de/research/trace/pics/FrameTrace/mp4/Verbose_Aladdin.dat.

ΒΙΟΓΡΑΦΙΚΟ ΤΟΥ ΣΥΓΓΡΑΦΕΑ

Full Name **Harilaos G. Koumaras**
Position **Associate Researcher at the NCSR “Demokritos”**
Address **37 Simmahidon Str., Athens, 11854, Greece**
Phone Number **210 3470417, 6944 595 494**
Date of Birth **20 June 1980**
Nationality **Greek**
Web Page **<http://aias.iit.demokritos.gr/~koumaras>**
e-mail **koumaras@iit.demokritos.gr**



Harilaos Koumaras was born in Athens, Greece in 1980. He received his BSc degree in Physics in 2002 from the University of Athens, Physics Department and his MSc in Electronic Automation and Information Systems in 2004, being scholar of the non-profit organization Alexander S Onassis, from the University of Athens, Physics and Informatics Department. He has received twice the Greek State Foundations (IKY) scholarship for excellent performance during the academic years 2000-01 and 2003-04. He has also granted with honors the classical piano and harmony degrees from the classical music department of Attiko Conservatory. He joined the Digital Telecommunications Lab at the National Centre of Scientific Research "Demokritos" in 2003 and since then he has participated in the EU funded projects SOQUET, ENTHRONE and national funded project PYTHAGORAS with presentations and publications at international conferences, scientific journals and book chapters. Currently, he completes his PhD dissertation, having granted the scholarship of NCSR "Demokritos". At the same time, he is an associate lecturer at the Business College of Athens (BCA), teaching modules related to Information Technology. His research interests include objective/subjective evaluation of the perceived quality of multimedia services, video quality and picture quality evaluation, video traffic modeling, digital terrestrial television and video compression techniques. Currently, he is the author or co-author of more than 20 scientific papers in international journals and technical books, numbering 14 non-self citations. He is a member of IEEE, SPIE and National Geographic Society.

EDUCATIONAL BACKGROUND

2004-2007: University of Athens

PhD degree. Informatics and Telecommunications Department
Scholar of NCSR “Demokritos”

2002-2004: University of Athens

Physics Department – Informatics and Telecommunications Department
Postgraduate Studies (MSc): Electronic Automation & Information Systems 8.82/10.00
Master Thesis: Objective Evaluation of Perceived Quality of Service for MPEG-4
Multimedia Applications. Scholar of non profit organization “Alexander S. Onassis”

1998-2002: University of Athens

Physics Department, University Degree in Physics: 7.53/10.00

1988-2002: Attiko Conservatory, Department of Classical Piano

Piano Degree (Excellent with Honors) - Piano teacher: D. Mallouhos

1996-1999: Attiko Conservatory, Department of Music Theory

Harmony Degree (Excellent with Honors) - Major Teacher: J. Christofilos

1995-1998: 63rd State School of Athens

High school Degree - Apolytirio (Excellent 19^{2/10}/20)

SCIENTIFIC PUBLICATIONS

Journals

- J8** : **“A theoretical framework for end-to-end video quality assessment”**
Harilaos Koumaras, Cheng-Han Lin, Ce-Kuen Shieh, Anastasios Kourtis
Working paper.
- J6** : **“Quantitative Perceptual Comparison of VBR over CBR for MPEG-4 Video”**
H. Koumaras, G. Gardikis, A. Kourtis, D. Martakos
Journal of Electronic Imaging (*JEI*), SPIE, IS&T, Vol. 16, No.3, July 2007
- J5** : **“Automatic Scene Change Detection for DCT-based Compressed Video at Low Bit Rates”**
H. Koumaras, E. Pallis, A. Kourtis, D. Martakos
Journal of Telecommunication Systems, Springer Ed., (*Pending Review results*)
- J4** : **“Boundary Shot Detection without Threshold Parameters”**
H. Koumaras, G. Gardikis, G. Xilouris, E. Pallis, A. Kourtis
Journal of Electronic Imaging (*JEI*), SPIE, IS&T, Vol.15, No.2, April 2006 (IF 0.710)
- J3** : **“Unidirectional Lightweight Encapsulation (ULA) – Performance Evaluation and Application Perspectives”**
G. Xilouris, G. Gardikis, H. Koumaras, A. Kourtis
IEEE Transactions on Broadcasting, Vol. 52(3), Sept. 2006 pp.374 – 380 (IF 0.560)
- J2** : **“Quantified PQoS Assessment Based on Fast Estimation of the Spatial and Temporal Activity Level”**
H. Koumaras, A. Kourtis, D. Martakos, J. Lauterjung
Journal of Multimedia Tools and Applications, DOI.10.1007/s11042-007-0111-1
Springer Editions, March 05, 2007 (IF 0.416)
- J1** : **“Evaluation of Video Quality Based on Objectively Estimated Metric”**
H. Koumaras, A. Kourtis, D. Martakos
Journal of Communications & Networking, Korean Inst. of Commun. Sciences, Vol.7(3), 235-242, Sep. 2005 (IF 0.479)

Book Chapters

- BC2** : **“PQoS Assessment Methods for Multimedia Services”**
H. Koumaras, F. Liberal, L. Sun
In *“Wireless Multimedia: Quality of Service and Solutions”*
- BC1** : **“Customer Acceptance of Internet banking services in Greece”**
M. Makris, H. Koumaras, A. Konstantopoulou, S. Konidis, S. Kostakis
In *“Advances in Banking Technology and Management: Impacts on ICT and CRM”*

Chapter Contributions in Encyclopedias and Handbooks

- EN5** : **“Principles of Digital Video Coding”**
H. Koumaras, E. Pallis, A. Kourtis, D. Martakos
Encyclopedia of Information Technology, 2nd Edition, Idea Group Pub. (**Under publication**)
- EN4** : **“Digital Television for Handheld”**
G. Gardikis, H. Koumaras, A. Kourtis
Encyclopedia of Information Technology, 2nd Edition, Idea Group Pub. (**Under publication**)
- EN3** : **“Indexing of Noteworthy Frames From Video Services”**
H. Koumaras, G. Gardikis, G. Xilouris, E. Pallis, A. Kourtis
Encyclopedia of Mobile Computing & Services, Information Science Reference Pub., ISBN: 978-1-59904-675-4, October 2007
- EN2** : **“Perceived Quality Evaluation for Multimedia Services”**
H. Koumaras, E. Pallis, G. Xilouris, A. Kourtis, D. Martakos
Encyclopedia of Mobile Computing & Services, Information Science Reference Pub., ISBN: 978-1-59904-675-4, October 2007
- EN1** : **“Objective Measurement of Perceived QoS for Homogeneous MPEG-4 Video Content”**
H. Koumaras, D. Martakos, A. Kourtis
Encyclopedia of Multimedia Technology and Networking, pp.770-776, Idea Group Pub. ISBN 1-59140-561-0 (hard cover) -- ISBN 1-59140-796-6 (ebook), May 2005.

Conferences

- C11** : **"Video Quality Prediction based on the Spatial and Temporal Classifier of the Uncompressed Content"**
H. Koumaras, A. Kourtis
The 18th Annual IEEE Inter. Symposium on Personal, Indoor and Mobile Radio Comm. - IEEE PIMRC 2007, Athens, Greece, 3-7 September 2007
- C10** : **"A Theoretical Framework for End-to-End Video Quality Prediction of MPEG-based Sequences"**
H. Koumaras, A. Kourtis, C-H Lin, C-K Shieh (**Best Paper Award**)
The 3rd Int. Conf. on Networking & Services - ICNS07, Athens, Greece, June 19-25 2007
- C9** : **"A Novel Method for Pre-Encoding Video Quality Prediction"**
H. Koumaras, T. Pliakas, A. Kourtis
IST Mobile Summit 2007, Budapest, Hungary, 1-5 July 2007
- C8** : **"Real-time, Dynamic Resource Allocation in DVB-S.2/RCS Networks"**
G. Gardikis, H. Koumaras, G. Xilouris, E. Pallis, A. Kourtis
Eighth International Symposium on Interworking, Santiago, Chile, January 15 - 19, 2007

- C7 : “Real Time PQoS Enhancement of IP Multimedia Services Over Fading & Noisy DVB-T Channel”**
 H. Koumaras, E. Pallis, G. Gardikis, A. Kourtis
 TEMU2006, Int. Conf. on Telecommunications and Multimedia, Heraklion, Crete, Greece, 5-7 July 2006
- C6 : “Multi-layer Resource Management in DVB-S.2 Networks: The IMOSAN approach”**
 G. Gardikis, G. Xilouris, H. Koumaras, A. Kourtis
 TEMU2006, Int. Conf. on Telecom and Multimedia, Heraklion, Greece, 5-7 July 2006
- C5 : “A Markov Modified Model of H.264 VBR Video Traffic”**
 H. Koumaras, C. Skianis, A. Kourtis
 IST Mobile & Wireless Communications Summit 2006, Mykonos, Greece.
- C4 : “Analysis of H.264 Video Coded Traffic”**
 H. Koumaras, C. Skianis, G. Gardikis, A. Kourtis
 2nd Int. Network Conference – INC 2005, 4-7 July 2005, pp.441-448 Samos, Greece.
- C3 : “Broadband Data Access over Hybrid DVB-T Networks”**
 G. Gardikis, G. Kormentzas, G. Xilouris, H. Koumaras, A. Kourtis
 3rd Int. Conference on Performance Modelling & Evaluation of Heterogeneous Networks, Het-NeTs 05, 18-20 July 2005, pp. 53/1-53/9, Ilkley, U.K., July 2005.
- C2 : “Pre-Encoding PQoS Assessment Method for Optimized Resource Utilization”**
 H. Koumaras, E. Pallis, G. Xilouris, A. Kourtis, D. Martakos, J. Lauterjung
 2nd Int. Conference on Performance Modelling & Evaluation of Heterogeneous Networks Het-NeTs 04, 26-28 July 2004, pp. 14/1-14/10, ISBN : 0-9540151-6-9 Ilkley, UK.
- C1 : “Objective Evaluation of the Perceived Quality of Video Content”**
 H. Koumaras, G. Gardikis, A. Kourtis
 IST Mobile Summit 2003, June 15-18 2003, Vol. I, pp.237-240 Aveiro, Portugal

NON-SELF CITATIONS

Non-Self Citations: 14

- Way-Chuang Ang, Chee-Hong The, Tat-Chee Wan and Rahmat Budiarto, “Performance Evaluation of Unidirectional Lightweight Encapsulation using ns-2 and DVB-S2 Test bed”, Proceedings of the 2007 IEEE International Conference on Telecommunications and Malaysia International Conference on Communications, 14-17 May 2007, Penang, Malaysia.
- M. Ries, C. Crespi, O. Nemethova, M. Rupp, “Content Based Video Quality Estimation for H.264/AVC Video Streaming”, in Proc. Proceedings of IEEE Wireless and Communications & Networking Conference, Hong Kong, March, 2007.
- Eric A. Silva, Karen Panetta, Sos S Agaian, “Quantifying image similarity using measure of enhancement by entropy”, Mobile Multimedia/Image Processing for Military and Security Applications 2007, Sos S. Agaian, Sabah A. Jassim, Editors, 65790U, Proceedings of SPIE -- Volume 6579, May. 2, 2007.

- Ulrich Engelke and Hans-Jürgen Zepernick, "Perceptual Quality Measures for Image and Video Services", Euro-NGI Workshop on Socio-Economic Aspects of Next Generation Internet, Lyngby, Denmark, October 9-10, 2006.
- Astrid Undheim, Yuan Lin, Peder J. Emstad, "Characterization of Slice-Based H.264/AVC Encoded Video Traffic," *ecumn*, pp. 263-272, Fourth European Conference on Universal Multiservice Networks (ECUMN'07), 2007.
- Michal Ries, Olivia Nemethova, Markus Rupp, "Motion Based Reference-Free Quality Estimation for H.264/AVC Video Streaming", International Symposium on Wireless Pervasive Computing, San Juan, Poerto Rico, 5-7 February 2007.
- M. Fouad, F. Bayoumi, H. Onsi, M. Darwish, "Shot Transition Detection with Minimal Decoding of MPEG Video Streams", *Transactions on Engineering, Computing and Technology*, Vol.17, Dec. 2006.
- J. Doggen, F. Schueren, "MPEG4 Part 10-H.264 Modeling in OPNET", Masters Thesis, Department of Industrial Sciences and Technology, University College of Antwerpen, Belgium, July 2006.
- F. Liberal et al "PQoS Based Model for Assessing Significance Statistically", 3rd Inter. Conference on Performance Modelling and Evaluation of Heterogeneous Networks, Het - NeTs 05, 18-20 July 2005, Ilkley, U.K., July 2005.
- T. Pliakas, G. Kormentzas, Ch. Skianis "Scalable Video Streaming Traffic Delivery in IP/UMTS Networking Environments", *Journal of Mobile Multimedia*, 2007.
- Christopher John, "Effect of content on perceived video quality", TLEN 5380 – Video Technology, Colorado, 9 August 2006.
- J. Jiang, Z. Li and G. Xiao, "Real-time shot cut detection in compressed domain", *Journal of Electronic Imaging*, SPIE, 2007.
- T. Pliakas, G. Kormentzas, Ch. Skianis "End-to-end QoS Issues of MPEG-4 FGS Video Streaming Traffic Delivery in an IP/DVB/UMTS Network", Special Issue of Elsevier journal of Operational Research.
- Ulrich Engelke and Hans-Jürgen Zepernick, "Perceptual-based Quality Metrics for Image and Video Services: A Survey", 3rd EuroNGI Conference on Next Generation Internet Networks, Trondheim, Norway, 21-23 May 2007.

EDITORIAL and CHAIR SERVICES

- Associate Editor of Telecommunication Systems Journal, published by Springer Science & Business Publishers, ISSN: 1018-4864
- Session Chair at the 18th Annual IEEE Inter. Symposium on Personal, Indoor and Mobile Radio Comm. - IEEE PIMRC 2007, Athens, Greece, 3-7 September 2007
- Session Chair at the Third International Conference on Networking and Services - ICNS07, June 19-25, 2007 - Athens, Greece

RESEARCH and ACADEMIC INTERESTS

- Subjective Methods for Picture/Video Quality Evaluation
- Full Reference Methods for Objective Picture/Video Quality Evaluation
- Reduced Reference Methods for Objective Picture/Video Quality Evaluation
- No Reference Methods for Objective Picture/Video Quality Evaluation
- Digital Encoding Standards (i.e. MPEG-1/2/4, H.264)

- Modeling of video traffic (Frame Layer, GOP layer)
- Shot Boundary Detection
- Real Time Streaming Protocols / Applications
- Mapping of Network Parameters (i.e. delay, jitter) to Perceived Video Quality
- Network QoS issues (i.e. DiffServ implementations, Traffic Marking/Classification)
- IP over DVB issues (i.e. High Definition Video over DVB)

HONORS - DISTINCTIONS

- 2007** : Best Paper Award at ICNS07
- 2003-2004:** Distinction by the Foundation of National Scholarships (IKY) for primacy and excellent performance during the second year of postgraduate studies at the Physics/Informatics Department
- 1999-2000:** Distinction by the Foundation of National Scholarships (IKY) for excellent performance during the second year of graduate studies at the Physics Department
- 1997-1998:** Diploma and Metal by the Mayor of Athens Mr. Avramopoulos for excellent performance during the two last classes of High School
- 1997** : Distinction of Excellence by TEAPAE
- 1992-1998:** Distinctions and Honors by the Ministry of Education and Religion for all classes

SCHOLARSHIPS

- 2004-2007:** Scholar of NCSR «Demokritos» for Doctorate dissertation
- 2003-2004:** Scholar of non profit organization “Alexander S. Onassis” for postgraduate studies
- 2003** : Scholarship of excellence by IKY for highest performance in the postgraduate studies
- 1999-2000:** Scholarship of excellence by IKY for highest performance as a sophomore

RESEARCH PROJECTS

- 2006- . . .** : Participation in EC project “ENTHRONE II”
- 2006- . . .** : Participation in EC project “IMOSAN”
- 2005- 06** : Participation in national funded project “PYTHAGORAS II”
- 2003- 06** : Participation in EC project “ATHENA”
- 2003- 06** : Participation in EC project “ENTHRONE”
- 2002- 03** : Participation in EC project “SOQUET”

MEMBERSHIPS

- Member of IEEE
- Member of SPIE
- Member of National Geographic Society
- Member of Onassis Foundation Scholars Association