



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ & ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ταχείες τεχνικές αναζήτησης βίντεο

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΤΟΥ

ΣΕΡΑΦΕΙΜ Σ. ΠΑΠΑΣΤΕΦΑΝΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π. (2005)

Αθήνα, Μάρτιος 2010



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ & ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ταχείες τεχνικές αναζήτησης βίντεο

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΣΕΡΑΦΕΙΜ Σ. ΠΑΠΑΣΤΕΦΑΝΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π. (2005)

Συμβουλευτική Επιτροπή: Γεώργιος Στασινόπουλος
Νικόλαος Μήτρου
Βασίλειος Λούμος

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την ...

...
Γ. Στασινόπουλος
Καθηγητής Ε.Μ.Π.

...
Ν. Μήτρου
Καθηγητής Ε.Μ.Π.

...
Β. Λούμος
Καθηγητής Ε.Μ.Π.

...
Μ. Θεολόγου
Καθηγητής Ε.Μ.Π.

...
Ε. Καγιάφας
Καθηγητής Ε.Μ.Π.

...
Α. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

...
Δ. Ρεΐσης
Αναπλ. καθηγητής
Παν. Αθηνών

Αθήνα, Μάρτιος 2010

.....

ΣΕΡΑΦΕΙΜ Σ. ΠΑΠΑΣΤΕΦΑΝΟΣ

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

© 2010 - All rights reserved

Copyright © **ΣΕΡΑΦΕΙΜ Σ. ΠΑΠΑΣΤΕΦΑΝΟΣ 2010**

Με επιφύλαξη παντός δικαιώματος. All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

ΠΕΡΙΛΗΨΗ

Η διατριβή πραγματεύεται το πρόβλημα της ευθείας αναζήτησης βίντεο, δηλαδή τη χρήση ενός αποσπάσματος για να βρεθεί το βίντεο από το οποίο αυτό προέρχεται. Η ευθεία αναζήτηση βίντεο έχει διάφορες εφαρμογές, για παράδειγμα για προστασία πνευματικών δικαιωμάτων, αναγνώριση όμοιων βίντεο, έρευνα αγοράς κοκ.

Τα σημαντικότερα προβλήματα της ευθείας αναζήτησης βίντεο είναι: Η ταχύτητα δεικτοδότησης και αναζήτησης του βίντεο και η ευελιξία της αναζήτησης. Η ταχύτητα δεικτοδότησης και αναζήτησης του βίντεο είναι ένα πολύ σημαντικό πρόβλημα, το οποίο και επιδεινώνεται όταν το βίντεο έχει υποστεί συμπίεση με κάποιο νεώτερο πρωτόκολλο συμπίεσης βίντεο οπότε είναι αρκετά χρονοβόρα και δύσκολη η εξαγωγή των συνηθισμένων χαρακτηριστικών του βίντεο που χρησιμοποιούνται στην αναζήτηση. Ομοίως, είναι σημαντικό η αναζήτηση να είναι ευέλικτη σε τυχόν αλλαγές που μπορεί να υποστεί κάποιο βίντεο, για παράδειγμα αν αλλάξει η ποιότητα του ή η ανάλυση του.

Προτείνονται τρεις τεχνικές αναζήτησης βίντεο: Η αναζήτηση με πιθανότητες, η αναζήτηση με βάση δεδομένων και η αναζήτηση με *intra macroblock*. Η αναζήτηση με πιθανότητες βασίζεται σε ένα πιθανοτικό μοντέλο που περιγράφει τον αριθμό των διαφορών στις αλλαγές πλάνου που θα αναγνωριστούν σε δύο παραλλαγές του ίδιου βίντεο ή σε δύο διαφορετικά βίντεο. Η αναζήτηση με βάση δεδομένων βασίζεται στην αυτοματοποιημένη δημιουργία ερωτημάτων προς μια βάση δεδομένων για να βρεθεί το υποψήφιο βίντεο με βάση τις αλλαγές πλάνου ενός αποσπάσματος. Τέλος, η αναζήτηση με *intra macroblock* εξετάζει τα χαρακτηριστικά των *macroblock* του βίντεο για να κάνει την αναζήτηση.

Κύριο σημείο αυτών των τεχνικών είναι ότι χρησιμοποιούν απλά χαρακτηριστικά για την αναζήτηση, κάτι που τις κάνει πολύ γρήγορες τόσο κατά τη δεικτοδότηση όσο και κατά την αναζήτηση αλλά και αρκετά ευέλικτες στις συνήθεις αλλαγές που μπορεί να υποστεί κάποιο βίντεο.

Οι τεχνικές αυτές υλοποιούνται σε πραγματικά συστήματα και, μέσω αρκετών πειραμάτων εξετάζεται τόσο η ταχύτητα τους όσο και η ευελιξία τους στην εύρεση βίντεο. Τα πειραματικά αποτελέσματα αποδεικνύουν ότι οι τεχνικές που προτείνονται επιλύουν τα σημαντικότερα προβλήματα της ευθείας αναζήτησης βίντεο.

ABSTRACT

The thesis examines the direct video search problem, which is about the usage of a video clip in order to find the original video from which it has been taken. Direct video search has several applications, as for copyright protection, recognition of same videos, market research, etc.

The most important problems of direct video search are the indexing and search speed, especially when the video has been compressed with some state of the art video compression protocols. In that particular case, extracting the usual indexing characteristics is difficult and time consuming. Another problem is the flexibility of search in case the video has undergone changes (in quality or in analysis).

Three direct video search techniques are proposed: probability search, database search and intra macroblock search. The probability search is based upon a probability model that describes the number of differences in recognised shot cuts between two variations of the same video or in two different videos. The database search, is based on the automated query creation on a database based on the shot cut positions of the video clip. Finally, the intra macroblock search examines the characteristics of the video macroblocks for the search.

The main point of these techniques is that they use simple characteristics for search, something that makes them very fast both in indexing and search and flexible in common video changes.

These techniques are implemented in real systems and the speed and the flexibility of the video proposed search techniques is examined through several experiments. The results that are derived by these experiments prove that the proposed techniques resolve the most important problems of the field.

Ευχαριστίες

Με αυτό το μικρό σημείωμα θα ήθελα να εκφράσω τις ευχαριστίες μου σε κάποια άτομα τα οποία με βοήθησαν στη συγγραφή της διδακτορικής μου διατριβής: Αρχικά θέλω να εκφράσω την ευγνωμοσύνη μου στον καθηγητή κ. Γ. Στασινόπουλο, ο οποίος τα τελευταία 6 χρόνια με στήριξε και με βοήθησε αφάνταστα. Επίσης, ευχαριστώ τα άλλα δύο μέλη της τριμελούς συμβουλευτικής επιτροπής μου, δηλαδή τον καθηγητή κ. Ν. Μήτρου και τον καθηγητή κ. Β. Λούμο οι οποίοι ήταν πάντοτε διαθέσιμοι όταν τους χρειάστηκα. Κατά τη διάρκεια των τελευταίων ετών συνεργάστηκα με πολλά άτομα από το εργαστήριο τηλεπικοινωνιών του ΕΜΠ στα πλαίσια μιας σειράς προγραμμάτων και δημοσιεύσεων τους οποίους επίσης θέλω να ευχαριστήσω για τη βοήθεια τους αλλά και την άριστη συνεργασία που είχαμε. Έτσι λοιπόν, ευχαριστώ πολύ τους Δρ. Φ. Ανδριτσόπουλο, Δρ. Α. Νικολαΐδη, Δρ. Γ. Δουμένη, Δρ. Ι. Σιφναίο, Δρ. Ν. Παπαουλάκη, Β. Ματεεβίτσι, Μ. Δράκο και Β. Μπιλιλή. Τέλος, οφείλω ένα μεγάλο ευχαριστώ στην οικογένεια μου για όλα όσα μου έχει προσφέρει.

Σεραφείμ Παπαστέφανος
Αθήνα, Μάρτιος 2010

Περιεχόμενα

1	Εισαγωγή	1
1.1	Γενικά για την αναζήτηση βίντεο	2
1.2	Εφαρμογές αναζήτησης βίντεο	4
1.2.1	Προστασία πνευματικών δικαιωμάτων	4
1.2.2	Αφαίρεση αντιγράφων	5
1.2.3	Εύρεση αποσπασμάτων	5
1.2.4	Καταγραφή πληροφοριών	6
1.2.5	Αναζήτηση διαφημίσεων	6
1.2.6	Διόρθωση βίντεο	7
1.2.7	Εφαρμογές σημασιολογικής αναζήτησης	7
1.3	Ορισμοί	7
1.4	Μικρή εισαγωγή στη συμπίεση βίντεο	8
1.4.1	Γενικά	9
1.4.2	Χαρακτηριστικά Ψηφιακού βίντεο	9
1.4.3	Πρόβλεψη με βάση την κίνηση	10
1.4.4	Μετασχηματισμός διαφορών	11
1.4.5	Είδη εικόνων	12
1.4.6	Άλλες έννοιες	13
1.4.6.1	Ανάλυση	13
1.4.6.2	Interlace	14
1.4.6.3	Ρυθμός μετάδοσης εικόνων	16
1.4.6.4	Ρυθμός μετάδοσης bit	17
1.5	Τι περιλαμβάνεται στα επόμενα	17
2	Ανίχνευση αλλαγών πλάνου	19
2.1	Υπάρχουσες μέθοδοι	19
2.2	Μέτρηση επίδοσης	21
2.3	Νέες προκλήσεις	21
2.4	Μέθοδος intra macroblock	22
2.4.1	Λειτουργία	23
2.4.2	Υλοποίηση	24
2.4.3	Παρουσίαση της μεθόδου	25
2.5	Μέθοδος ιστογραμμάτων φωτεινότητας	26

2.5.1	Περιγραφή	28
2.5.2	Υλοποίηση	30
2.5.3	Παρουσίαση της μεθόδου	30
2.6	Μέτρηση επίδοσης τεχνικών	31
2.6.1	Επίδοση intra	34
2.6.2	Επίδοση ιστογραμμάτων	35
2.7	Συμπεράσματα	37
3	Αναζήτηση βίντεο	39
3.1	Υπάρχουσες μέθοδοι	39
3.2	Προβλήματα της αναζήτησης	44
3.3	Μετρικές απόδοσης αναζήτησης βίντεο	46
3.4	Οι λύσεις που προτείνονται	47
3.5	Λάθη κατά την ανίχνευση αλλαγών πλάνου	49
3.5.1	Χαμένες και εσφαλμένες αναγνωρίσεις αλλαγών πλάνου	49
3.5.2	Ανάλυση Jitter	50
3.5.2.1	Deinterlace	51
3.5.2.2	Αλλαγή στο ρυθμό εικόνων	52
3.5.2.3	Λάθος αναγνώριση βαθμιαίων αλλαγών πλάνου	54
3.6	Πιθανοτική μέθοδος	55
3.6.1	Ορισμός της (m_q, f_q) -ισότητας	56
3.6.2	Αντίληψη συμβάντων με την (m_q, f_q) -ισότητα	57
3.6.3	Σύγκριση ακολουθιών με συμβάντα	60
3.6.4	Μέτρηση διαφορών	62
3.6.4.1	Αλγόριθμος μέτρησης διαφορών (συμμετρικός)	62
3.6.4.2	Αλγόριθμος μέτρησης διαφορών (μη συμμετρικός)	63
3.6.4.3	Σύγκριση μεταξύ του συμμετρικού και του μη συμμετρικού αλγορίθμου	65
3.6.4.4	Εξαρτήσεις του συμμετρικού και του QEO αλγορίθμου	65
3.6.4.5	Αλγόριθμος μέτρησης διαφορών με σύγκριση	66
3.6.5	Χρήση αλλαγών πλάνου	68
3.6.6	Έλεγχος ορθότητας ανάλυσης	69
3.6.6.1	Δημιουργία των R, P και Q	69
3.6.6.2	Αναλυτικά και πειραματικά αποτελέσματα	71
3.6.6.3	Παρουσία πρόσθετων συμβάντων	74
3.6.7	Μετρήσεις για την e	76
3.6.8	Χειρισμός Jitter	79
3.6.9	Γενικά για την πιθανοτική μέθοδο	79
3.7	Μέθοδος βάσης δεδομένων	80
3.7.1	Η δημιουργία της ερώτησης στη βάση	81
3.7.1.1	Απλή περίπτωση	81
3.7.1.2	Ανοχή σε Jitter	82

3.7.1.3	Ανοχή σε σφάλματα	83
3.7.2	Εύρεση των δυνατών περιπτώσεων	84
3.7.2.1	Ανοχή σε εσφαλμένες αναγνωρίσεις	84
3.7.2.2	Ανοχή σε χαμένες αναγνωρίσεις	85
3.7.2.3	Αριθμός περιπτώσεων	87
3.7.3	Γενικά για τη μέθοδο βάσης δεδομένων	87
3.8	Μέθοδος intra macroblock	88
3.8.1	Περιγραφή	88
3.8.2	Δυσκολίες που παρουσιάζονται σε αλλαγμένα βίντεο	90
3.8.2.1	Αλλαγή συμπίεσης/ποιότητας	91
3.8.2.2	Αλλαγή ανάλυσης	93
3.8.2.3	Οπτικές αλλαγές	95
3.8.3	Αναζήτηση με τη μέθοδο intra macroblock	95
3.8.3.1	Απλή αναζήτηση	95
3.8.3.2	Απλή αναζήτηση με έλεγχο διακοπής	98
3.8.3.3	Αναζήτηση μεγάλων τιμών	98
3.8.3.4	Αναζήτηση διαστημάτων	99
3.8.3.5	Ανοχή σε jitter	100
3.8.4	Γενικά για τη μέθοδο intra macroblock	102
3.9	Συμπεράσματα	103
4	Συστήματα αναζήτησης	105
4.1	Εξαγωγή στοιχείων από βίντεο	106
4.1.1	FFMpeg	107
4.1.2	DirectShow	107
4.1.3	Σύγκριση μεταξύ Ffmpeg και DirectShow	108
4.2	Σύστημα ιστού	109
4.2.1	Απαιτήσεις	109
4.2.2	Αρχιτεκτονική	110
4.2.2.1	Βάση δεδομένων	110
4.2.2.2	Δεικτοδότηση και εισαγωγή	112
4.2.2.3	Αναζήτηση	113
4.2.2.4	Διεπαφή χρήστη	114
4.2.2.4.1	Συνολική παρουσίαση	114
4.2.2.4.2	Αναλυτική παρουσίαση	115
4.2.2.4.3	Ανέβασμα	115
4.2.2.4.4	Αναζήτηση	117
4.2.2.4.5	Αποτελέσματα	117
4.2.3	Υλοποίηση	119
4.3	Σύστημα κονσόλας	119
4.4	Συνεχής αναζήτηση	120
4.4.1	Απαιτήσεις	120
4.4.2	Αρχιτεκτονική	121
4.4.2.1	Διαχωρισμός μεταξύ πηγών και αναζήτησης	121

4.4.2.2	Η συνεχής αναζήτηση	121
4.4.2.3	Η αρχιτεκτονική της εφαρμογής	122
4.4.2.4	Η βάση δεδομένων	124
4.4.3	Υλοποίηση	126
5	Αποτελέσματα	129
5.1	Ταχύτητα δεικτοδότησης	129
5.2	Ταχύτητα αναζήτησης	131
5.2.1	Μέθοδος βάσης δεδομένων	131
5.2.2	Πιθανοτική και intra macroblock μέθοδοι	133
5.3	Απόδοση αναζήτησης	137
5.3.1	Μέθοδος intra και πιθανοτήτων	138
5.3.2	Μέθοδος βάσης δεδομένων	148
6	Συμπεράσματα	151
A'	Τι υλοποιήθηκε	155
A'.1	Αρχεία ctxt	155
A'.2	GetHistYUV	156
A'.3	HistShotDet	156
A'.4	H264GetCtxt	156
A'.5	CtxtShotDet	156
A'.6	ShotCutDetPerf	156
A'.7	TextVideoAdd	156
A'.8	TextVideoSearch	157
A'.9	TextVideoSceneChangeGUI	157
B'	Δημοσιεύσεις του συγγραφέα	159

Κατάλογος Σχημάτων

1.1	Επίδραση ανάλυσης σε μια εικόνα	14
1.2	Παράδειγμα interlaced βίντεο	15
1.3	Περιπτώσεις deinterlace	16
2.1	Ποσοστά intra macroblock ανα εικόνα	26
2.2	Σφάλματα στην ανίχνευση	27
2.3	Παράδειγμα ιστογράμματος	29
2.4	Διαφορές ιστογραμμάτων	31
2.5	Η περίπτωση της εικόνας 2.2	32
2.6	Παράδειγμα εισαγωγής λογότυπου	33
2.7	Τα <i>Recall</i> και <i>Precision</i> σε σχέση με το κατώφλι για τη μέθοδο των intra macroblock	36
3.1	Είδη σφαλμάτων κατά την αναγνώριση	50
3.2	Αλλαγή στο ρυθμό μετάδοσης	52
3.3	Jitter λόγω αλλαγή στο ρυθμό εικόνων	54
3.4	Ορισμός της αντίληψης κατά (m_q, f_q) της r από την q	56
3.5	Ορισμός της ισοδυναμίας κατά (m_q, f_q)	59
3.6	Παράδειγμα μορφής κατανομών διαφορών	61
3.7	Συμβάντα ως άγκυρες	63
3.8	Αλγόριθμος μέτρησης διαφορών	64
3.9	Οι πιθανότερες τιμές των διαφορών	66
3.10	Μια εικόνα του προγράμματος TextVideoSceneChangeGUI	70
3.11	Αναλυτικά αποτελέσματα: (α) Συμμετρικός αλγόριθμος και (β) Μη συμμετρικός αλγόριθμος	72
3.12	Πειραματικά αποτελέσματα: (α) Συμμετρικός αλγόριθμος και (β) Μη συμμετρικός αλγόριθμος	73
3.13	Θεωρητικά αποτελέσματα με την παρουσία πρόσθετων εικόνων I : (α) Συμμετρικός αλγόριθμος και (β) Μη συμμετρικός αλγόριθμος	75
3.14	Πειραματικά αποτελέσματα με την παρουσία πρόσθετων εικόνων I : (α) Συμμετρικός αλγόριθμος και (β) Μη συμμετρικός αλγόριθμος	77
3.15	Γραφική αναπαράσταση των αποτελεσμάτων του πίνακα 3.4	79

3.16	Δημιουργία συνδυασμών	85
3.17	Αριθμός ερωτημάτων που δημιουργούνται	87
3.18	Εξαγωγή υπογραφής εικόνας με intra mb	89
3.19	Εξαγωγή υπογραφής βίντεο με intra mb	89
3.20	Σύγκριση των ποσοστών intra macroblock τριών ίδιων αποσπασμάτων όταν επανασυμπιεστούν με διαφορετική ποιότητα Q	92
3.21	Σύγκριση των ποσοστών intra macroblock δύο ίδιων αποσπασμάτων όταν το ένα έχει τη μισή ανάλυση	94
3.22	Προσθήκη κανονικών και μεγαλύτερων υπότιτλων σε ένα βίντεο	96
3.23	Σύγκριση των ποσοστών intra macroblock όταν προστεθούν υπότιτλοι σε κάποιο απόσπασμα	97
3.24	Απλή αναζήτηση με ολίσθηση	97
3.25	Αναζήτηση με μεγάλες τιμές	100
3.26	Ανοχή σε jitter	101
4.1	Επίπεδο πρόσβασης στη βάση δεδομένων	112
4.2	Δεικτοδότηση και εισαγωγή στη βάση	112
4.3	Διάγραμμα τάξεων για αναζήτηση	114
4.4	Συνολική παρουσίαση των βίντεο	115
4.5	Ανέβασμα βίντεο στη βάση δεδομένων	116
4.6	Γραφικό περιβάλλον για αναζήτηση	117
4.7	Τα αποτελέσματα της αναζήτησης	118
4.8	Εξαγωγή υπογραφής από διαφορετικές πηγές	122
4.9	Εξήγηση λειτουργίας συνεχούς αναζήτησης	123
4.10	Διάγραμμα τάξεων για τη συνεχή αναζήτηση	125
4.11	Σχεσιακό μοντέλο της βάσης δεδομένων για τη συνεχή αναζήτηση	126
4.12	Το γραφικό περιβάλλον του προγράμματος της συνεχούς αναζήτησης	127
5.1	Πολυπλοκότητα ερωτημάτων προς βάση δεδομένων	132
5.2	Χρόνοι αναζήτησης για όλες τις μεθόδους για διάφορα μήκη αποσπασματος για τη μεγαλύτερη βάση δεδομένων	135
5.3	Χρόνοι αναζήτησης για όλες τις μεθόδους για διάφορα μήκη αποσπασματος για τη μικρότερη βάση δεδομένων	137
5.4	Οι επεξηγήσεις των διαγραμμάτων που ακολουθούν	139
5.5	Επίδοση πιθανοτήτων για απόσπασμα μήκους 1000 εικόνων	140
5.6	Επίδοση intra macroblock για απόσπασμα μήκους 1000 εικόνων	141
5.7	Επίδοση πιθανοτήτων για απόσπασμα μήκους 500 εικόνων	142
5.8	Επίδοση intra macroblock για απόσπασμα μήκους 500 εικόνων	142
5.9	Επίδοση πιθανοτήτων για απόσπασμα μήκους 1500 εικόνων	143
5.10	Επίδοση intra macroblock για απόσπασμα μήκους 1500 εικόνων	143
5.11	Επίδοση πιθανοτήτων για απόσπασμα μήκους 1000 εικόνων με μισή ανάλυση από το αρχικό βίντεο	144

5.12	Επίδοση intra macroblock για απόσπασμα μήκους 1000 εικόνων με μισή ανάλυση από το αρχικό βίντεο	145
5.13	Επίδοση πιθανοτήτων για απόσπασμα μήκους 1000 εικόνων με $Q = 30$	146
5.14	Επίδοση intra macroblock για απόσπασμα μήκους 1000 εικόνων με $Q = 30$	146
5.15	Επίδοση πιθανοτήτων για απόσπασμα μήκους 1000 εικόνων με $Q = 40$	147
5.16	Επίδοση intra macroblock για απόσπασμα μήκους 1000 εικόνων με $Q = 40$	147

Κατάλογος Πινάκων

2.1	Χαρακτηριστικά βίντεο	33
2.2	Απόδοση ανίχνευσης αλλαγών πλάνου με intra	35
2.3	Απόδοση ανίχνευσης αλλαγών πλάνου με ιστογράμματα φωτεινότητας	36
3.1	Χαρακτηριστικά που δε χρειάζονται αποσυμπίεση ([WDV ⁺ 03])	46
3.2	Αποτέλεσμα του α XOR β	67
3.3	Επίδραση στην e του κατωφλιού διαφορών ιστογραμμάτων κατά την ανίχνευση αλλαγών πλάνου	78
3.4	Επίδραση στην e του κατωφλιού intra macroblock και του Q κατά την ανίχνευση αλλαγών πλάνου	78
3.5	Παράδειγμα του πίνακα ΑΛΛΑΓΗ_ΠΛΑΝΟΥ	81
3.6	Παράδειγμα τιμών	99
3.7	Σύγκριση μεθόδων πρώτου επιπέδου	104
5.1	Ταχύτητα εξαγωγής υπογραφής	130
5.2	Ταχύτητα αναζήτησης για τη μέθοδο της βάσης δεδομένων	133
5.3	Απόδοση για τη μέθοδο της βάσης δεδομένων	149

Κεφάλαιο 1

Εισαγωγή

ΤΑ τελευταία χρόνια, λόγω της αύξησης της ταχύτητας των συνδέσεων με το διαδίκτυο, της ανάπτυξης της τεχνολογίας κωδικοποίησης βίντεο αλλά και της ευρείας διάδοσης συσκευών που μπορούν να καταγράψουν άμεσα βίντεο όπως κινητά τηλέφωνα και ψηφιακές κάμερες, έχει γίνει πια κοινός τόπος η παρακολούθηση, αποθήκευση αλλά και δημοσίευση βίντεο στο διαδίκτυο.

Η νέα αυτή τάση δημιουργεί μεγάλες δυνατότητες τόσο για τους επαγγελματίες όσο και για τους απλούς χρήστες. Οι επαγγελματίες μπορούν να μοιράσουν δωρεάν ή επί πληρωμή τα βίντεο τους χωρίς να χρειάζεται να χρησιμοποιήσουν ενδιαμέσους όπως παλαιότερα, έχοντας έτσι πλήρη έλεγχο τόσο στο περιεχόμενο όσο και στην ποιότητα αυτού που διαμοιράζεται. Οι απλοί χρήστες μπορούν με τη βοήθεια υπηρεσιών όπως το Youtube [YOU] να καταγράψουν κάτι που θεωρούν ενδιαφέρον και να το δείξουν σε όλο τον κόσμο. Στο τελευταία βοηθά και η τάση των χρηστών να θέλουν να εμπλουτίσουν το blog τους ή το προφίλ τους στο Facebook [FAC] με με πολυμεσικό περιεχόμενο.

Το βίντεο που βρίσκεται στο διαδίκτυο είναι άναρχα δομημένο: Χρήστες ανταλλάσσουν μεταξύ τους βίντεο, τα δημοσιεύουν στις ιστοσελίδες και τα blog τους, χιλιάδες διαφορετικά βίντεο αποθηκεύονται κάθε μέρα στο Youtube και τις αντίστοιχες υπηρεσίες, οι επαγγελματίες μοιράζουν το βίντεο μέσα από τα δικά τους κανάλια και γενικά επικρατεί η ίδια κατάσταση με κάθε περιεχόμενο που βρίσκεται στο διαδίκτυο.

Ο μόνος τρόπος για να μπει κάποια τάξη σε αυτή την κατάσταση είναι με τη βοήθεια μηχανών αναζήτησης βίντεο, αντίστοιχων με αυτές που υπάρχουν για το κείμενο. Ο τομέας της αναζήτησης βίντεο, αντίθετα με το κείμενο, είναι ακόμα σε πρώιμο στάδιο μιας και υπάρχουν πολλά προβλήματα τα οποία θα αναλυθούν στα επόμενα κεφάλαια. Αντίστοιχη είναι η κατάσταση και στον ήχο και μουσική, όμως στην περίπτωση αυτή η τεχνολογία έχει προχωρήσει περισσότερο και υπάρχουν ακόμη και εμπορικές εφαρμογές με τις οποίες γίνεται αναζήτηση κάποιου μουσικού κομματιού, για παράδειγμα το TrackID της Sony Ericsson [TRA]. Φυσικά αυτή η αναζήτηση δε γίνεται γενικά στο διαδίκτυο, αλλά σε συγκεκριμένες βιβλιοθήκες αναζήτησης.

Στις επόμενες ενότητες της εισαγωγής θα παρουσιαστούν κάποια γενικά στοιχεία για την αναζήτηση βίντεο, οι διάφορες εφαρμογές της αναζήτησης, κάποιοι χρήσιμοι ορισμοί, μια μικρή εισαγωγή στη συμπίεση βίντεο και αχρονιγώς τα περιεχόμενα της ανά χειράς εργασίας.

1.1 Γενικά για την αναζήτηση βίντεο

Ο κλάδος της αναζήτησης βίντεο μπορεί να χωριστεί σε δύο μεγάλες κατηγορίες: Την ευθεία (direct) και τη σημασιολογική (semantic). Στην ευθεία αναζήτηση ως ερώτηση για την αναζήτηση χρησιμοποιείται ένα απόσπασμα του αρχικού βίντεο (πιθανότατα με αλλαγμένα χαρακτηριστικά) ή ακόμα και μια εικόνα αυτού. Αντίθετα στη σημασιολογική αναζήτηση η ερώτηση γίνεται είτε με κάποιες λέξεις κλειδιά, είτε με τη βοήθεια ερώτησης με παράδειγμα (Query by Example). Δηλαδή, στην ευθεία αναζήτηση χρειάζεται κάποιο απόσπασμα για να γίνει η αναζήτηση, ενώ στην σημασιολογική όχι. Παρότι η εργασία αυτή αφορά αποκλειστικά την ευθεία αναζήτηση, για λόγους πληρότητας, στις επόμενες παραγράφους γίνεται μια συνοπτική ανασκόπηση της βιβλιογραφίας για τη σημασιολογική αναζήτηση. Η βιβλιογραφική ανασκόπηση της ευθείας αναζήτησης θα γίνει στο κεφάλαιο 3.

Κατά τη διαδικασία της αναζήτησης βίντεο μπορεί να εξεταστεί ξεχωριστά η δεικτοδότηση (indexing) του βίντεο και ξεχωριστά η αναζήτηση (searching). Κατά τη διαδικασία της δεικτοδότησης εξάγονται από το βίντεο μια σειρά από πληροφορίες οι οποίες είτε αποθηκεύονται σε μια βάση δεδομένων για να χρησιμοποιηθούν αργότερα είτε χρησιμοποιούνται απευθείας για την αναζήτηση.

Στη σημασιολογική αναζήτηση, κατά τη δεικτοδότηση εξάγονται διάφορα εννοιολογικά χαρακτηριστικά από κάθε βίντεο είτε αυτόματα από υπολογιστές, είτε πολλές φορές με τη βοήθεια ανθρώπων οι οποίοι παρακολουθούν τα βίντεο και σημειώνουν τα χαρακτηριστικά αυτά ([YLZ02, LEJ98]). Τέτοια χαρακτηριστικά μπορεί να είναι πάρα πολλά και εξαρτώνται από την εφαρμογή. Μερικά παραδείγματα είναι τα ονόματα ή τα χαρακτηριστικά των ηθοποιών που εμφανίζονται σε κάθε σκηνή, το πότε και αν υπάρχουν συγκεκριμένα συμβάντα (πχ πότε σημειώνεται καλάθι σε έναν αγώνα μπάσκετ), το ποιά ζώα δείχνει ένα ντοκιμαντέρ και ένα σωρό άλλα. Επίσης, μπορούν να εξαχθούν λέξεις κλειδιά από το ηχητικό περιεχόμενο του βίντεο, ή από οπτική αναγνώριση των υποτίτλων αυτού ([HJN02]).

Το εύρος των χαρακτηριστικών που θα μπορούσαν να εξαχθούν ή να αναζητηθούν κατά τη σημασιολογική αναζήτηση είναι τεράστιο, και έτσι τα αυτοματοποιημένα συστήματα σημασιολογικής αναζήτησης αρκετές φορές είναι εξειδικευμένα στη δεικτοδότηση και εύρεση βίντεο προκαθορισμένου είδους. Για παράδειγμα, το [LXY⁺06] εξάγει πληροφορίες από αγώνες μπάσκετ ενώ το [FMR02] έχει να κάνει με την αυτόματη αναγνώριση γεγονότων σε βίντεο ασφαλείας, ενώ άλλες εργασίες βρίσκουν μόνο κάποιο συγκεκριμένο χαρακτηριστικό, για παράδειγμα διαχωρισμός μεταξύ εξωτερικών και εσωτερικών εικόνων

1.1. ΓΕΝΙΚΑ ΓΙΑ ΤΗΝ ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

([SP98]) ή αναγνώριση τοπίων από την πόλη ή από τη φύση ([VFJZ01]).

Ένα αρκετά γνωστό σύστημα σημασιολογικής αναζήτησης είναι το Media-Mill [SWKS07, WSdR⁺07], το οποίο έχει τη δυνατότητα για αναζήτηση με λέξη, με παράδειγμα, αλλά και με βάση την έννοια. Η πρώτη είναι η αναζήτηση που ήδη έχει αναφερθεί, στη δεύτερη ο χρήστης μπορεί να δώσει ένα παράδειγμα αυτού που ζητά, για παράδειγμα να ζωγραφίσει ένα σχήμα και τέλος στην τρίτη γίνεται μια κατάταξη με βάση την έννοια δηλαδή κάποιος μπορεί να ψάξει με βάση τη λέξη «ζώο» και να του επιστρέψει βίντεο με κάθε είδους ζώα, κάτι που δε θα ήταν δυνατό με την αναζήτηση με λέξεις. Παρόμοια χαρακτηριστικά έχει και το Informedia του Carnegie Mellon University ([WCGH99, CHMP04]) που επίσης υποστηρίζει τόσο την αναζήτηση με λέξη όσο και την αναζήτηση με έννοιες. Στην [NH01] παρουσιάζεται ένα σύστημα για την αναγνώριση μιας σειράς από χαρακτηριστικών του βίντεο που βασίζεται στην αντιστοιχία των χαρακτηριστικών χαμηλού επιπέδου με χαρακτηριστικά υψηλού επιπέδου με βάση κάποια πιθανότητα. Τέτοια χαρακτηριστικά υψηλού επιπέδου που αναγνωρίζονται περιλαμβάνουν έννοιες όπως: έκρηξεις, βουνά, παραλίες, εξωτερικός χώρος, μουσική κοκ. Η προηγούμενη εργασία εξελίχθηκε σε ένα πλήρες σύστημα σε συνεργασία και με την IBM ([ABC⁺03]).

Στην [ACGW05] παρουσιάζεται ένα σύστημα το οποίο έχει ως σκοπό να βοηθήσει στην αναζήτηση βίντεο τους ανθρώπους. Προς τούτο, το σύστημα αυτό διαχωρίζει τα βίντεο σε τμήματα με βάση το «θέμα» που αυτά περιέχουν (αντί για διαχωρισμό σε σκηνές ή πλάνα που είναι το σύνηθες) και προσφέρει μια εξελιγμένη διεπαφή χρήστη για γρήγορο χειρισμό αυτών. Παρόμοιος είναι και ο σκοπός του συστήματος ViBE ([TCA⁺04]) για την οργάνωση βάσης δεδομένων με βίντεο. Εδώ, το βίντεο χωρίζεται σε πλάνα και αυτά ομαδοποιούνται με βάση το περιεχόμενο τους και παρουσιάζονται στο χρήστη με μορφή πυραμίδας.

Οι περισσότερες εμπορικές μηχανές αναζήτησης βίντεο γενικού ενδιαφέροντος που υπάρχουν βασίζονται σε σημασιολογική αναζήτηση: Η αναζήτηση γίνεται με βάση κάποιες λέξεις κλειδιά που εισάγει ο χρήστης. Παραδείγματα τέτοιων μηχανών είναι τα AltaVista [ALT], Blinkx [BLI], TruVeo [TRU] και CastTv [CAS]. Παρομοίως φυσικά γίνεται και η αναζήτηση στο Youtube, με βάση λέξεις κλειδιά δηλαδή που έχει εισάγει ο χρήστης για να περιγράψει το βίντεο που ανέβασε, αλλά και χρήση κάποιων άλλων, κυρίως στατιστικών χαρακτηριστικών πχ πόσοι χρήστες τα έχουν δει, πότε ανέβηκαν κοκ.

Όπως φαίνεται από τα παραπάνω, η σημασιολογική αναζήτηση βασίζεται κυρίως στην περιγραφή με λέξεις κλειδιά των βίντεο από τους χρήστες είτε από εξειδικευμένο προσωπικό. Αυτοματοποιημένη σημασιολογική αναζήτηση μπορεί να γίνει μόνο σε συγκεκριμένες κατηγορίες βίντεο όπου μπορούν να αναγνωριστούν συγκεκριμένα συμβάντα (πχ αγώνες μπάσκετ) ή σε συγκεκριμένες κατηγορίες χαρακτηριστικών οι οποίες μπορούν να αναγνωριστούν με διάφορες τεχνικές από τον υπολογιστή (για παράδειγμα εκρήξεις σε ένα βίντεο).

1.2 Εφαρμογές αναζήτησης βίντεο

Η αναζήτηση βίντεο έχει αρκετές εμπορικές εφαρμογές. Παρακάτω παρουσιάζονται αρκετά αναλυτικά οι εφαρμογές της ευθείας αναζήτησης βίντεο και περιληπτικά κάποιες εφαρμογές της σημασιολογικής αναζήτησης:

1.2.1 Προστασία πνευματικών δικαιωμάτων

Η πειρατεία υλικού είναι ένα από τα σημαντικότερα προβλήματα που υπάρχουν σήμερα στο διαδίκτυο. Χιλιάδες άνθρωποι ανταλλάσσουν εύκολα και χωρίς κίνδυνο υλικό για το οποίο κανονικά θα έπρεπε να είχαν πληρώσει. Παρά τις προσπάθειες που γίνονται για προστασία του υλικού, μπορεί κάποιος πολύ εύκολα να βρει όλες τις νέες κυκλοφορίες σε μουσική, βίντεο, βιβλία και λογισμικό ακόμα και πριν αυτά φτάσουν στα καταστήματα, ενώ οι σελίδες που μοιράζουν παράνομο υλικό συγκαταλέγονται στις δημοφιλέστερες του διαδικτύου.

Για την προστασία των πνευματικών δικαιωμάτων, εκτός από τη χρησιμοποίηση ανθρώπων για να ψάχνουν το διαδίκτυο για τέτοιο υλικό, υπάρχουν δύο σημαντικές τεχνικές, είτε προστασία μέσα από το ίδιο το υλικό, είτε με έλεγχο του υλικού που μοιράζεται.

Αναφορικά με την πρώτη περίπτωση υπάρχουν μια σειρά από μεθόδους οι οποίες είτε επιτρέπουν το υλικό να αναπαραχθεί μόνο σε ένα σύστημα ή να αναπαραχθεί μόνο N φορές (DRM - Digital Rights Management [LELD05, Pei02]). Η μέθοδος αυτής προστασίας έχει μια σειρά από προβλήματα τα οποία αναφέρονται στην [MHB03] και έγιναν ευρέως γνωστά με το γνωστό επεισόδιο που έγινε με το σύστημα προστασίας αντιγραφής της εταιρίας Sony [FH06, HF06]. Ακόμα μπορεί να γίνει συσχέτιση του υλικού με έναν συγκεκριμένο χρήστη - έτσι αν αυτός θελήσει να το μοιράσει θα μπορεί να εντοπιστεί ή για παράδειγμα να φαίνεται ότι το υλικό αυτό είναι παράνομο και δε θα έπρεπε να το βλέπουν μη εξουσιοδοτημένοι χρήστες (watermarking [CHO⁺98, FMS99, CX06]). Οι τεχνική αυτή έχει μια σειρά από προβλήματα, αφού υποβαθμίζει την ποιότητα του βίντεο και είναι αρκετά εύκολη η αφαίρεση του watermark ([HSG99, LLB98]).

Για τη δεύτερη περίπτωση γίνεται σύγκριση του υλικού που υπάρχει στο διαδίκτυο με το προστατευμένο υλικό ούτως ώστε να αποφασιστεί αν είναι παράνομο ή όχι. Ένα πλήρες σύστημα για την προστασία των πνευματικών δικαιωμάτων στο διαδίκτυο τόσο για βίντεο όσο και για ήχο και κείμενο προτείνεται στο [Shi99]. Το σύστημα αυτό επισκέπτεται ιστοσελίδες (web crawler) και ελέγχει αν έχουν παράνομο υλικό. Στη σημερινή εποχή φυσικά η διακίνηση παράνομου υλικού δε γίνεται πλέον από το WWW αλλά από συγκεκριμένα peer to peer δίκτυα στα οποία δε μπορούν να συμμετέχουν όποιοι θέλουν, παρά μόνο προκαθορισμένα άτομα. Τα περισσότερα τέτοια peer to peer δίκτυα βασίζονται στο πρωτόκολλο BitTorrent ([BIT]). Συστήματα για προστασία σε αυτή την περίπτωση υπάρχουν στα [LHZ07, Esk03].

Αυτό που προτείνεται εδώ ως εφαρμογή της αναζήτησης βίντεο στην προ-

1.2. ΕΦΑΡΜΟΓΕΣ ΑΝΑΖΗΤΗΣΗΣ ΒΙΝΤΕΟ

στασία των πνευματικών δικαιωμάτων είναι να καταγράφεται ένα μικρό απόσπασμα από τα βίντεο που μοιράζονται μέσω του WWW ή ανταλλάσσονται από τα διάφορα peer to peer προγράμματα και να γίνεται αναζήτηση στη βάση με τα προστατευμένα βίντεο με αυτό το απόσπασμα ως ερώτημα. Επειδή δε χρειάζεται να εξεταστεί ολόκληρο το βίντεο παρά μόνο το μικρό απόσπασμα η διαδικασία αυτή είναι πολύ γρήγορη κάτι που επιτρέπει την εξέταση δεκάδων αρχείων ταυτόχρονα. Η καταγραφή αυτή μπορεί να γίνεται από τον ίδιο τον ISP, μη επιτρέποντας έτσι στους χρήστες να πάρουν παράνομο υλικό.

1.2.2 Αφαίρεση αντιγράφων

Το γεγονός ότι σε υπηρεσίες όπως του YouTube ανεβαίνουν παραπάνω από 65 000 βίντεο ημερησίως (στοιχεία 2006 [YOU06]) σε συνάρτηση με το ότι η αναζήτηση γίνεται με σημασιολογικά κριτήρια, δηλαδή με τη χρήση κάποιων λέξεων κλειδιών έχει ως αποτέλεσμα την επιστροφή ίδιων βίντεο, αντιγράφων το ένα του άλλου όταν γίνεται αναζήτηση με κάποια λέξη κλειδί.

Το πρόβλημα αυτό δε μπορεί να λυθεί με άλλον τρόπο εκτός από την αναζήτηση βίντεο, αφού ούτε από τους χρήστες μπορεί να ζητηθεί να εξετάζουν τα ήδη υπάρχοντα βίντεο πριν ανεβάσουν ένα νέο, ούτε φυσικά να υπάρχουν άνθρωποι οι οποίοι να ελέγχουν όλα αυτά τα βίντεο!

Η εργασία [WHN07] πραγματεύεται ακριβώς αυτό το πρόβλημα: Χρησιμοποιεί μια ιεραρχική προσέγγιση, αρχικά ψάχνει με γρήγορο τρόπο, έχοντας ως βάση την υπογραφή ενός βίντεο που δημιουργείται από το ιστόγραμμα χρωμάτων και αν υπάρχει αμφιβολία τότε χρησιμοποιούνται πιο ακριβές υπολογιστικά μέθοδοι. Στην [LTCJ⁺07] προτείνεται ένα σύστημα για τη σύγκριση μεθόδων απόφασης για αντίγραφα και με τη βοήθεια του εξετάζεται η απόδοση μιας σειράς τέτοιων μεθόδων. Στην [cSCZ00] εξάγεται η υπογραφή ενός βίντεο χρησιμοποιώντας έναν μικρό αριθμό από εικόνες αυτού και γίνεται έλεγχος για ομοιότητα. Ενδιαφέρον αποτέλεσμα αυτού είναι ότι σε πειράματα που έγιναν σε 1800 ώρες βίντεο από το διαδίκτυο που αντιπροσωπεύουν 45000 ξεχωριστά αποσπάσματα βρέθηκε ότι για κάθε βίντεο της συλλογής υπάρχουν 5 αντίγραφα!

Πρέπει εδώ να σημειωθεί ότι πλέον το Youtube έχει ενσωματώσει μια τέτοια τεχνολογία αφαίρεσης αντιγράφων, αφού σε κάποιες αναζητήσεις μπορεί να φανεί ένα μήνυμα που εξηγεί στο χρήστη ότι «κάποια από τα αποτελέσματα ήταν αντίγραφα άλλων αποτελεσμάτων και αφαιρέθηκαν, αν θέλετε μπορείτε να επαναλάβετε την αναζήτηση χωρίς την αφαίρεση των διπλών βίντεο».

1.2.3 Εύρεση αποσπασμάτων

Η εφαρμογή αυτή έχει να κάνει με ερώτηση της μορφής «τι είναι αυτό που παρακολουθώ τώρα;». Ένα χαρακτηριστικό παράδειγμα είναι το ακόλουθο: Κάποιος χρήστης βλέπει ότι η τηλεόραση προβάλλει μια εκπομπή που τον ενδιαφέρει, όμως έχει χάσει την αρχή της και δε γνωρίζει ποιά είναι η εκπομπή αυτή. Χρησιμοποιώντας το κινητό τηλέφωνο του, καταγράφει ένα απόσπασμα από

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

αυτή την εκπομπή και ερωτά κάποια διαδικτυακή υπηρεσία για τον τίτλο της εκπομπής. Η υπηρεσία του απαντά, μαζί με άλλα στοιχεία και επιλογές όπως να αγοράσει την εκπομπή ούτως ώστε να την παρακολουθήσει ολόκληρη.

Η παραπάνω εφαρμογή είναι παρόμοια με το ήδη υπάρχον TrackID [TRA] για τη μουσική. Φυσικά συνήθως το πρόγραμμα της τηλεόρασης είναι καθορισμένο και δημοσιευμένο από πριν, αντίθετα στο ραδιόφωνο δεν δημοσιεύονται οι λίστες με τα τραγούδια που ακούγονται σε κάθε εκπομπή, κάτι που ίσως αλλάζει με την έλευση της δικτυακής τηλεόρασης. Φυσικά, εκτός από την απλή περίπτωση που περιγράφηκε, η εύρεση αποσπασμάτων έχει μεγάλη χρησιμότητα σε εξειδικευμένες εφαρμογές, για παράδειγμα σε τηλεοπτικούς σταθμούς, εταιρίες παραγωγής βίντεο κοκ.

1.2.4 Καταγραφή πληροφοριών

Για τις ανάγκες της έρευνας αγοράς πολύ συχνά χρειάζεται να καταγράφεται το πότε και πού παίζεται κάποιο βίντεο. Για παράδειγμα, πόσες φορές παίχτηκε κάποια διαφήμιση ή πόσες φορές εμφανίστηκαν οι δηλώσεις του πρωθυπουργού. Αυτό μέχρι τώρα γινόταν με ανθρώπους οι οποίοι απλώς παρακολουθούσαν όλα τα κανάλια και κατέγραφαν τις πληροφορίες που χρειαζόνταν. Πλέον, με την ανάπτυξη της δικτυακής τηλεόρασης αλλά και όλων των άλλων υπηρεσιών μέσω των οποίων διαμοιράζεται βίντεο, η χειροκίνητη καταγραφή αυτών των πληροφοριών είναι αδύνατη.

Η λύση σε αυτό το πρόβλημα είναι η χρήση αναζήτησης βίντεο ούτως ώστε να ελέγχονται τα προγράμματα των «καναλιών» της δικτυακής τηλεόρασης για τα αποσπάσματα τα οποία ζητούνται και να καταγράφονται αυτόματα χωρίς ανθρώπινη παρέμβαση. Μια τέτοια εμπορική εφαρμογή είναι το [IPH] το οποίο παρακολουθεί έναν αριθμό από «κανάλια» και καταγράφει το πόσες φορές παίχτηκαν σε αυτά κάποια αποσπάσματα που του έχουν δοθεί σαν είσοδος. Παρόμοιο είναι και το αντικείμενο της διδακτορικής εργασίας [Mia99] όπου προτείνεται ένας αλγόριθμος για παρακολούθηση ειδήσεων. Ο αλγόριθμος αυτός βασίζεται στο γεγονός ότι κάποια συγκεκριμένα βίντεο εμφανίζονται πολλές φορές όταν υπάρχουν νέα για ένα συγκεκριμένο θέμα. Έτσι, το σύστημα που προτείνεται καταγράφει σε πραγματικό χρόνο τις επαναλήψεις αυτές και τις χρησιμοποιεί για παρακολούθηση της είδησης.

1.2.5 Αναζήτηση διαφημίσεων

Παρότι αποτελεί παρακλάδι της καταγραφής πληροφοριών για έρευνα αγοράς, υπάρχουν αρκετές εργασίες σχετικά με την αναγνώριση διαφημίσεων σε βίντεο.

Μια από τις πρώτες εργασίες που ασχολήθηκαν με το θέμα είναι η [LKE97] στην οποία υπολογίζονται κάποια διανύσματα συσχέτισης χρώματος για κάθε εικόνα έτσι ώστε να δημιουργηθεί μια υπογραφή και υπολογίζεται ο αριθμός των πράξεων που χρειάζεται για τη μετατροπή από την υπογραφή της διαφήμισης σε

εκείνη του βίντεο στο οποίο γινόταν η αναζήτηση. Η [SMOM02] αναγνωρίζει τα διαφημιστικά κενά βασιζόμενη στην παρατήρηση ότι πριν και ανάμεσα στις διαφημίσεις υπάρχουν κάποιες μαύρες εικόνες (οι οποίες φυσικά δε γίνονται αντιληπτές από τον άνθρωπο) καθώς και μια πτώση στα επίπεδα του ήχου. Παρόμοιες τεχνικές χρησιμοποιούνται και στην [MC05]. Στην [DCH04] συγκρίνεται η επίδοση δύο τεχνικών αναγνώρισης διαφημίσεων. Η μεν πρώτη βασίζεται στο γεγονός ότι οι διαφημίσεις επαναλαμβάνονται για να τις αναγνωρίσει, ενώ η δεύτερη στην παρατήρηση ότι οι διαφημίσεις έχουν διαφορετικά οπτικοακουστικά χαρακτηριστικά από το υπόλοιπο πρόγραμμα. Τέλος, η [GA05] χρησιμοποιεί ένα συνδυασμό της επαναληψιμότητας των διαφημίσεων και κάποιων χαρακτηριστικών, όπως το πόσες αλλαγές πλάνων έχουν, τις μαύρες εικόνες πριν και μετά, τη μέση τιμή και την τυπική απόκλιση των χρωμάτων, ούτως ώστε να κατατάζει κάποια τμήματα του βίντεο ως διαφημίσεις.

1.2.6 Διόρθωση βίντεο

Παρόμοια με την παραπάνω εφαρμογή είναι και η χρήση της αναζήτησης για διόρθωση βίντεο: Για παράδειγμα μπορεί να εξεταστεί κάποιο βίντεο που έχει καταγραφεί από την τηλεόραση για ύπαρξη διαφημίσεων και στη συνέχεια αυτές να αφαιρεθούν για καλύτερη παρακολούθηση. Φυσικά για να γίνει κάτι τέτοιο με την ευθεία αναζήτηση θα πρέπει να υπάρχουν *a priori* τα βίντεο με τις διαφημίσεις και να εξετάζονται ένα ένα αν εμφανίζονται την κάθε χρονική στιγμή.

1.2.7 Εφαρμογές σημασιολογικής αναζήτησης

Παρότι η ευθεία αναζήτηση είναι καταλληλότερη για τις παραπάνω εφαρμογές, για κάποιες εξ αυτών μπορεί να χρησιμοποιηθεί και η σημασιολογική αναζήτηση. Η πραγματική δύναμη όμως της σημασιολογικής αναζήτησης φαίνεται σε εφαρμογές στις οποίες ο χρήστης αναζητά κάτι που ψάχνει και για το οποίο δεν έχει συγκεκριμένο απόσπασμα, για παράδειγμα ερωτήσεις της μορφής: «βίντεο τα οποία δείχνουν ένα ηλιοβασίλεμα», «βίντεο στο οποίο να εμφανίζεται κάποιος άνθρωπος να χορεύει», «βίντεο με τον τάδε ηθοποιό» κλπ.

1.3 Ορισμοί

Μερικοί χρήσιμοι για τη συνέχεια ορισμοί:

Ορισμός 1.3.1 Εικόνα (frame) είναι μια απλή, στατική εικόνα. Μια ακολουθία από διαδοχικές εικόνες αποτελούν ένα βίντεο.

Ορισμός 1.3.2 Πλάνο (shot) είναι μια ακολουθία από εικόνες που έχουν καταγραφεί με την ίδια κάμερα σε διαδοχικές χρονικές στιγμές, χωρίς καμία διακοπή μεταξύ τους. Αλλαγή πλάνου (shot cut) συμβαίνει όταν υπάρχει διακοπή

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

σε αυτή την ακολουθία και τα επόμενα φραμες είτε προέρχονται από διαφορετική σκηνή είτε από την ίδια.

Ορισμός 1.3.3 Σκηνή (scene) είναι μια ακολουθία από πλάνα τα οποία έχουν τραβηχτεί σε διαδοχικές χρονικές στιγμές, από την ίδια η διαφορετικές κάμερες, στον ίδιο περιβάλλοντα χώρο. Αλλαγή σκηνής (scene change) συμβαίνει όταν μια αλλαγή πλάνου συνοδεύεται και από διαφορετικό περιβάλλον στο νέο πλάνο.

Ορισμός 1.3.4 Απόσπασμα (clip) είναι ένα κομμάτι βίντεο το οποίο μπορεί να περιέχει μία ή περισσότερες σκηνές και χρησιμοποιείται σε μια ερώτηση αναζήτησης.

Ορισμός 1.3.5 Βάση δεδομένων (database) είναι ο χώρος στον οποίο αποθηκεύονται τα βίντεο με τα οποία θα συγκριθεί το απόσπασμα ούτως ώστε να προκύψει απάντηση στην ερώτηση αναζήτησης. Η βάση δεδομένων δεν είναι απαραίτητο να είναι βάση δεδομένων με τη στενή έννοια του όρου, μπορεί να είναι κάποια απλά αρχεία στο δίσκο ή ακόμα και ένα κομμάτι της μνήμης ενός υπολογιστή.

Ορισμός 1.3.6 Εικόνα κλειδί (key frame) ονομάζεται μια εικόνα ενός βίντεο η οποία υποτίθεται ότι έχει κάποιο συγκεκριμένο χαρακτηριστικό. Συνήθως οι εικόνες κλειδιά είναι εκείνες οι εικόνες οι οποίες προσφέρουν μια γρήγορη ανασκόπηση του βίντεο, δηλαδή υπάρχει τουλάχιστον μια εικόνα από κάθε σκηνή του βίντεο. Ένας εύκολος τρόπος για τη δημιουργία εικόνων κλειδιά είναι να θεωρείται εικόνα κλειδί κάθε αλλαγή πλάνου.

Ορισμός 1.3.7 Συντελεστής κβαντισμού Q (quantization factor) στη συμπίεση κατά MPEG είναι ένας αριθμός που καθορίζει το πόσο θα συμπιεστεί το βίντεο. Συγκεκριμένα, όσο μεγαλώνει ο συντελεστής κβαντισμού τόσο η ποιότητα του συμπιεσμένου βίντεο χειροτερεύει, αλλά μειώνονται τα bits που απαιτούνται για κάθε εικόνα. Περισσότερα για τη συμπίεση βίντεο κατά MPEG και το συντελεστή κβαντισμού θα παρουσιαστούν στην επόμενη ενότητα.

1.4 Μικρή εισαγωγή στη συμπίεση βίντεο

Επειδή στα επόμενα κεφάλαια θα αναφερθούν κάποια θέματα σχετικά με τη συμπίεση βίντεο, θα γίνει εδώ μια πολύ μικρή εισαγωγή στο αντικείμενο αυτό. Περισσότερες πληροφορίες μπορούν να βρεθούν στη βιβλιογραφία, για παράδειγμα στο [Ric03].

Η αίσθηση της κίνησης στο βίντεο, όπως είναι ήδη γνωστό, δημιουργείται με τη διαδοχική προβολή στατικών εικόνων οι οποίες αποτυπώνουν διαδοχικές χρονικές στιγμές. Αν τα χρονικά διαστήματα μεταξύ των εικόνων είναι αρκετά μικρά, δημιουργείται στο μάτι η αίσθηση της κίνησης. Οι ίδιες αρχές που ισχύσαν από τη γέννηση του κινηματογράφου, στα τέλη του 19ου αιώνα,

1.4. ΜΙΚΡΗ ΕΙΣΑΓΩΓΗ ΣΤΗ ΣΥΜΠΙΕΣΗ ΒΙΝΤΕΟ

ισχύουν μέχρι και σήμερα, το ψηφιακό βίντεο είναι προβολή διαδοχικών εικόνων, οι οποίες, επειδή αναπαριστούν διαδοχικές χρονικές στιγμές μοιάζουν πολύ μεταξύ τους. Σε αυτήν ακριβώς την ομοιότητα μεταξύ των εικόνων στηρίζεται και η βασικότερη αρχή στη συμπίεση βίντεο, η πρόβλεψη με βάση την κίνηση (motion prediction). Στα επόμενα, αφού αναφερθούν κάποιες λεπτομέρειες για τα πρωτόκολλα που χρησιμοποιούνται για τη συμπίεση αλλά και στο πώς αναπαρίστανται οι εικόνες ψηφιακά, θα γίνει περιγραφή των μεθόδων που χρησιμοποιούνται για τη συμπίεση, καθώς και κάποιων άλλων σημαντικών εννοιών, όπως ο ρυθμός μετάδοσης εικόνων και το interlace (μπορεί να μεταφραστεί σα διάπλεξη, αλλά θα χρησιμοποιηθεί ο αγγλικός όρος).

1.4.1 Γενικά

Οι δημοφιλέστερες τεχνικές συμπίεσης βίντεο βασίζονται σε εργασίες που έχουν γίνει από το MPEG - Motion Pictures Experts Group του ISO. Περιλαμβάνουν μια οικογένεια πρωτοκόλλων, μεταξύ των οποίων τα MPEG-1, MPEG-2, MPEG-4/2, H.264 (MPEG-4/10). Το MPEG-1 δημοσιεύθηκε στα 1991 και είναι το πρωτόκολλο που χρησιμοποιείται στην τεχνολογία Video CD η οποία είχε ως αρχικό σκοπό την αντικατάσταση του βίντεο, παρέχοντας καλύτερη ποιότητα. Το MPEG-1 δεν έγινε ποτέ πολύ δημοφιλές στην οικιακή ψυχαγωγία, όμως χρησιμοποιήθηκε αρκετά από τους χρήστες προσωπικών υπολογιστών. Το MPEG-2 κυκλοφόρησε στα μέσα της δεκαετίας του '90 και έγινε πολύ δημοφιλές λόγω της χρήσης του ως πρωτόκολλο συμπίεσης των DVD τα οποία τελικά αντικατέστησαν το βίντεο. Επίσης χρησιμοποιείται σε κάποιες τεχνολογίες επίγειας ψηφιακής μετάδοσης τηλεοπτικού σήματος (DTV). Το MPEG-4 ήταν μια οικογένεια από πρωτόκολλα που κυκλοφόρησε στα τέλη του 1998. Το δεύτερο μέρος του αφορούσε τη συμπίεση βίντεο. Αρχικά χρησιμοποιήθηκε αποκλειστικά από προσωπικούς υπολογιστές, όμως τα τελευταία χρόνια έχουν κυκλοφορήσει αρκετές συσκευές οικιακής ψυχαγωγίας που υποστηρίζουν το πρωτόκολλο αυτό, ενώ και αυτό χρησιμοποιείται στην επίγεια ψηφιακή τηλεόραση. Το δέκατο μέρος του MPEG-4, το οποίο είναι γνωστότερο ως H.264 ή AVC, στην προτυποποίηση του συνεργάστηκε και το Video Coding Experts Group της ITU-T εκτός από το MPEG, κυκλοφόρησε το 2004 και θεωρείται ότι είναι το μέλλον λόγω της πολύ καλής συμπίεσης προς μέγεθος αρχείου που παράγει. Υποστηρίζεται από τις συσκευές Blue-Ray και HD-DVD και πλέον είναι ένα από τα δημοφιλέστερα πρωτόκολλα βίντεο, τόσο στην επιστημονική έρευνα όσο και σε εμπορικές εφαρμογές.

1.4.2 Χαρακτηριστικά Ψηφιακού βίντεο

Το βίντεο παράγεται από τη συνεχή εναλλαγή εικόνων. Χρειάζονται τουλάχιστον 24 εικόνες ανά δευτερόλεπτο (frames per second - fps) για να δημιουργηθεί η αίσθηση της κίνησης. Εκτός από τις εικόνες, το βίντεο συνήθως συνοδεύεται και από ήχο, ο οποίος, είναι ένα ξεχωριστό σήμα από το βίντεο,

απλώς πρέπει να υπάρχει συγχρονισμός μεταξύ εικόνας και ήχου.

Κάθε εικόνα ενός βίντεο, αποτελείται από $h \times w$ σημεία (pixels), με h και w το ύψος και πλάτος της. Τα σημεία αυτά είναι τα σημεία της αναλογικής εικόνας στα οποία πραγματοποιήθηκε διδιάστατη δειγματοληψία και αναπαρίστανται με τη βοήθεια ενός χρώματος. Η αναπαράσταση των πραγματικών, αναλογικών χρωμάτων με ψηφιακό τρόπο γίνεται με τη βοήθεια των χρωματικών μοντέλων.

Ένα χρωματικό μοντέλο αντιστοιχεί πραγματικά χρώματα σε πλειάδες αριθμών, για παράδειγμα, στο RGB, κάθε χρώμα αντιστοιχίζεται σε τρεις αριθμούς από 0 ως 256 που συμβολίζουν το πόσο κόκκινο, πράσινο και μπλε περιέχεται σε αυτό το χρώμα. Έτσι, το RGB χρειάζεται 3 bytes για την αναπαράσταση ενός pixel και μπορεί να αναπαραστήσει 2^{24} δηλαδή περίπου 16,7 εκατομμύρια χρώματα.

Υπάρχουν διάφορα άλλα χρωματικά μοντέλα εκτός του RGB, όπως το CMY - Cyan Magenda Yellow που χρησιμοποιείται στους εκτυπωτές και το YUV που χρησιμοποιείται ευρύτατα στο ψηφιακό βίντεο. Σε αυτό, κάθε σημείο αποτελείται από μια ακολουθία τριών bytes για τη φωτεινότητα (luminance - Y) και τη χρωματικότητα (chrominance - U και V). Η φωτεινότητα έχει περισσότερη πληροφορία την οποία μπορεί να διακρίνει το ανθρώπινο μάτι, για αυτό και συνήθως η ανάλυση των χρωματικών συνιστωσών είναι μικρότερη. Συγκεκριμένα, στο YUV420 που είναι ο δημοφιλέστερος τύπος YUV, η ανάλυση των U και V είναι η μισή τόσο σε ύψος όσο και σε πλάτος από ότι της Y, δηλαδή είναι το 1/4 της ανάλυσης της Y.

Έτσι, κάθε εικόνα ενός ασυμπίεστου βίντεο με ανάλυση $h \times w$ θα έχει μέγεθος $3 \times h \times w$ στο RGB αλλά μόνο $1,5 \times h \times w$ στο YUV. Στην κωδικοποίηση κατά MPEG χρησιμοποιείται πάντοτε ο χώρος το YUV για την αναπαράσταση του βίντεο. Είναι σύνηθες οι αποφάσεις που είναι απαραίτητες κατά τη συμπίεση να λαμβάνονται μόνο με βάση τη φωτεινότητα του βίντεο μιας και έχει περισσότερη βαρύτητα από τη χρωματικότητα και φυσικά είναι λιγότερη η υπολογιστική επεξεργασία που απαιτείται αν πρέπει να υπολογιστούν οι περιπτώσεις για ένα συντελεστή αντί τριών.

1.4.3 Πρόβλεψη με βάση την κίνηση

Το κύριο χαρακτηριστικό των περισσότερων πρωτοκόλλων συμπίεσης βίντεο είναι η πρόβλεψη με βάση την κίνηση (motion prediction): Οι εικόνα i συγκρίνεται με κάποια προηγούμενη ή επόμενη εικόνα, ανάλογα με τις δυνατότητες κάθε πρωτοκόλλου και αν η διαφορά τους είναι μικρή, τότε αντί να αποσταλεί ολόκληρη η πληροφορία για την εικόνα i αποστέλλονται μόνο οι διαφορές της από την εικόνα $i - 1$. Συγκεκριμένα, στην απλούστερη περίπτωση, κάθε μια εικόνα χωρίζεται σε τμήματα μεγέθους συνήθως 16×16 σημείων τα οποία ονομάζονται macroblocks. Για κάθε ένα macroblock της εικόνας i βρίσκεται με βάση διάφορους αλγόριθμους αναζήτησης το τμήμα της εικόνας $i - 1$ με το οποίο το συγκεκριμένο macroblock έχει τη μικρότερη διαφορά, και τελικά, η πληροφορία που αποστέλλεται για αυτό το macroblock είναι ένας πίνακας

1.4. ΜΙΚΡΗ ΕΙΣΑΓΩΓΗ ΣΤΗ ΣΥΜΠΙΕΣΗ ΒΙΝΤΕΟ

με τις διαφορές του από το σημείο στο οποίο βρέθηκε η ελάχιστη διαφορά και ένα διάνυσμα με τη θέση του σημείου ελάχιστης διαφοράς που ονομάζεται διάνυσμα κίνησης (motion vector). Ένα macroblock στο οποίο γίνεται πρόβλεψη με βάση την κίνηση ονομάζεται inter-coded ή απλά inter.

Εκτός από την πρόβλεψη με βάση την κίνηση, δηλαδή τη χρονική πρόβλεψη (temporal prediction), ένα macroblock μπορεί να συμπιεστεί και με βάση τις τιμές των γειτονικών του macroblock ή ακόμα και να αποσταλούν οι τιμές του ίδιου του macroblock, δηλαδή με χωρική πρόβλεψη (spatial prediction). Ένα macroblock στο οποίο γίνεται χωρική πρόβλεψη ονομάζεται intra-coded ή απλά intra.

Τέλος, μπορεί να υπάρχουν macroblock τα οποία είναι ακριβώς ίδια με κάποια άλλα, οπότε δε χρειάζεται να αποσταλεί πίνακας με τις διαφορές τους από κάποιο άλλο macroblock. Αυτά ονομάζονται skip macroblocks και η μόνη πληροφορία που χρειάζεται για την περιγραφή τους είναι ένα διάνυσμα που αναφέρει το σημείο από το οποίο αυτά προέρχονται. Τα skip macroblock είναι τα πιο οικονομικά από άποψη συμπίεσης για αυτό και τις περισσότερες φορές προτιμώνται έναντι των άλλων τύπων.

Ο σκοπός αυτής της πρόβλεψης, είναι ουσιαστικά να μειωθεί η ποσότητα της πληροφορίας που χρειάζεται για την αναπαράσταση αυτού. Για παράδειγμα, όταν δε γίνει καμία πρόβλεψη για ένα macroblock τότε θα χρειάζονται 256 τυχαίες τιμές για την αναπαράσταση του. Αν όμως αυτό γίνει πρόβλεψη αυτού με βάση κάποιο άλλο macroblock, τότε μπορεί τελικά από τις 256 τιμές, οι 200 να μηδενίζονται αφού θα είναι ίδιες με τις τιμές του άλλου, άρα μόνο οι 56 θα είναι τελικά τυχαίες, κάτι που έχει ως αποτέλεσμα τη μείωση της εντροπίας και τελικά της πληροφορίας που χρειάζεται για την αναπαράσταση αυτού του macroblock.

1.4.4 Μετασχηματισμός διαφορών

Όπως ήδη αναφέρθηκε, εκτός από τα skip macroblocks, θα πρέπει να αποσταλεί μια σειρά από τιμές που αφορούν τη διαφορά της τιμής κάθε ενός σημείου του macroblock από το σημείο της άλλης εικόνας με το οποίο βρέθηκε να έχει τη μικρότερη διαφορά, ή τις τιμές κάθε ενός σημείου με βάση τη χωρική πρόβλεψη. Σε κάθε περίπτωση, ο πίνακας αυτός ο οποίος γενικά περιέχει $16 \times 16 = 256$ τιμές, μετασχηματίζεται στο πεδίο των συχνοτήτων με βάση έναν διακριτό μετασχηματισμό συνημιτόνου (Discrete Cosine Transform). Ενώ στην αρχή οι τιμές των διαφορών είναι ένας πίνακας 256 στοιχείων ο οποίος έχει τυχαίες τιμές σε όλα του τα σημεία, μετά το μετασχηματισμό τα πρώτα στοιχεία του πίνακα παίρνουν μεγαλύτερες τιμές. Μέχρι εδώ, μπορεί να γίνει κατανοητό ότι δεν γίνεται κάποια συμπίεση, αφού για κάθε macroblock υπάρχουν 256 τιμές με ίδιο εύρος με το αρχικό.

Στη συνέχεια όμως, οι τιμές του πίνακα χβαντίζονται, δηλαδή διαιρούνται με κάποιον αριθμό (το συντελεστή χβαντοποίησης) ανάλογα την ποιότητα συμπίεσης και κάποιες εξ αυτών θα πάρουν την ίδια τιμή, ενώ άλλες θα μηδενιστούν

τελείως. Η ποιότητα συμπίεσης ορίζεται από το συντελεστή κβαντισμού Q ο οποίος όσο μεγαλώνει χειροτερεύει μεν η ποιότητα του βίντεο (αφού χάνονται περισσότερες τιμές), αλλά χρειάζονται λιγότερα bits για την αναπαράσταση κάθε εικόνας. Ο συντελεστής κβαντισμού ουσιαστικά δηλαδή καθορίζει το πόσο πολύ θα κβαντιστούν οι τιμές του πίνακα συντελεστών κάθε macroblock.

Ο κβαντισμός των συντελεστών είναι και ο λόγος που πραγματοποιείται ο μετασχηματισμός των τιμών στο πεδίο των συχνοτήτων, αφού έτσι υπάρχει μια ταξινόμηση των τιμών με βάση τη σημαντικότητα τους και πολλές τιμές που δε προσφέρουν σημαντική πληροφορία στο macroblock, λόγω του κβαντισμού, μηδενίζονται και δε καταλαμβάνουν καθόλου πληροφορία.

Τέλος, οι κβαντισμένες τιμές κάθε macroblock συμπιέζονται με μη απωλεστικό τρόπο, για παράδειγμα με Variable Length Coding - VLC ή Context-adaptive Binary Arithmetic Coding - CABAC στο H.264 και πλέον γίνονται μέρος της συμπίεσμένης εικόνας.

1.4.5 Είδη εικόνων

Κάθε μια εικόνα σε ένα κατά βίντεο συμπίεσμένο κατά MPEG μπορεί να είναι τριών διαφορετικών τύπων: I , P και B . Οι εικόνες I περιέχουν μόνο intra macroblocks και δε χρειάζονται άλλες γειτονικές εικόνες για να υπολογιστούν οι τιμές τους. Στις εικόνες αυτές δηλαδή δε γίνεται καμία χρονική πρόβλεψη, κάτι που έχει ως αποτέλεσμα το να έχουν μεγαλύτερο μέγεθος από τα άλλα είδη εικόνων, όμως είναι απαραίτητη η ύπαρξη τους στα βίντεο για δύο λόγους: Η γρήγορη αναζήτηση (fast seek) είναι δυνατή μόνο με την ύπαρξη I εικόνων αφού σε άλλη περίπτωση θα πρέπει να αποσυμπίεστούν όλες οι εικόνες (μιας και η μία θα βασίζεται στην άλλη). Επίσης, σε πρωτόκολλα πριν το H.264 γίνονταν διάφορα αριθμητικά σφάλματα κατά τους μετασχηματισμούς τα οποία συσσωρεύονταν και έπρεπε σε τακτά χρονικά διαστήματα να εισάγεται μια εικόνα I έτσι ώστε τα σφάλματα αυτά να «καθαριστούν». Αντίθετα, στο H.264 γίνεται ακόμα και να μη χρησιμοποιούνται καθόλου εικόνες I αφού εκεί οι μετασχηματισμοί γίνονται με ακεραίους και δεν έχουν σφάλματα ([WSBL03]).

Στις εικόνες P και B μπορεί να υπάρχουν τόσο inter, intra όσο και skip macroblocks. Έτσι, για την αποσυμπίεση τους χρειάζεται να έχουν ήδη αποσυμπίεστεί πρώτα άλλες εικόνες. Η διαφορά των εικόνων P από των B είναι ότι στις μεν πρώτες μπορεί να έχει γίνει πρόβλεψη μόνο με τις προηγούμενες εικόνες (backward prediction), δηλαδή η εικόνα i χρειάζεται μόνο κάποια από τις εικόνες j , με $j < i$ για την πρόβλεψη της, ενώ αντίθετα στις εικόνες B , η πρόβλεψη γίνεται με βάση τόσο τις προηγούμενες όσο και τις επόμενες (forward prediction), ανάλογα με το που βρίσκεται η μικρότερη διαφορά.

Οι εικόνες που περιλαμβάνονται μεταξύ δύο εικόνων I ονομάζονται ομάδα εικόνων (Group Of Pictures - GOP), και έχουν την ιδιότητα ότι αποτελούν μια ανεξάρτητη οντότητα και μπορούν να αποσυμπίεστούν χωρίς τη χρήση άλλων εικόνων. Ο τύπος των εικόνων μιας ομάδας εικόνων συνήθως ακολουθεί κάποια συγκεκριμένη μορφή, για παράδειγμα $IPPPP...$ (χωρίς πρόβλεψη με επόμενες

1.4. ΜΙΚΡΗ ΕΙΣΑΓΩΓΗ ΣΤΗ ΣΥΜΠΙΕΣΗ ΒΙΝΤΕΟ

εικόνες), ή *IBBPBBPBBP...* (αρκετά συνηθισμένη περίπτωση) ή *IIIII...* (χωρίς χρονική πρόβλεψη) κοκ.

1.4.6 Άλλες έννοιες

Παρουσιάζονται στη συνέχεια τρεις έννοιες που δεν αφορούν άμεσα τη συμπίεση βίντεο αλλά είναι πολύ σημαντικές. Η ανάλυση, το interlace, ο ρυθμός μετάδοσης εικόνων και ο ρυθμός μετάδοσης bit.

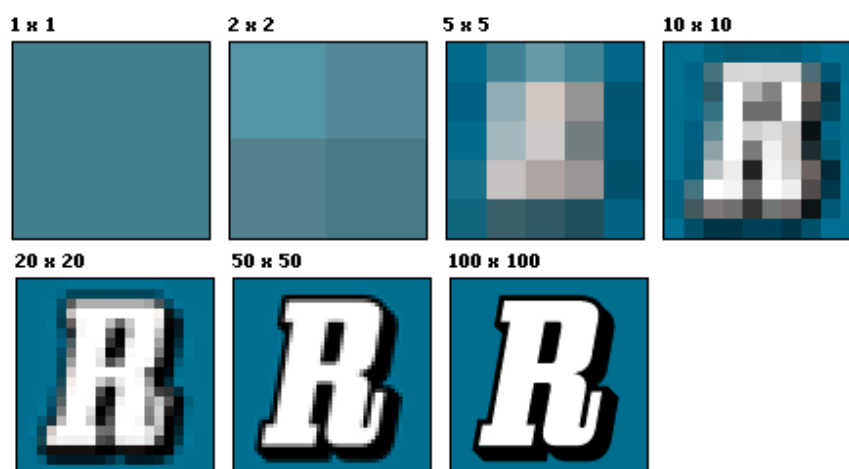
1.4.6.1 Ανάλυση

Όπως ήδη αναφέρθηκε, οι εικόνες που αποτελούν ένα βίντεο και κατ'έπαιση το ίδιο το βίντεο αποτελούνται από $h \times w$ σημεία (pixels), με h και w το ύψος και πλάτος της εικόνας. Υπάρχουν μια σειρά από προκαθορισμένες αναλύσεις οι οποίες έχουν συγκεκριμένη ονομασία στο χώρο της ψηφιακής εικόνας και βίντεο. Έτσι, QCIF είναι η ανάλυση 176×144 , CIF είναι η 352×288 , η VGA είναι η 640×480 και η QVGA η 320×240 .

Το aspect ratio καθορίζει το λόγο πλάτος προς ύψος εικόνας σε ένα βίντεο. Στην τηλεόραση χρησιμοποιείται το $4 : 3$ ενώ στο σινεμά το $16 : 9$. Στους υπολογιστές, αρχικά οι οθόνες ήταν $4 : 3$, όμως τα τελευταία χρόνια αρχίζει να διαδίδεται πολύ το $16 : 9$. Παρότι στο ψηφιακό βίντεο μπορεί να επιλεγούν οποιαδήποτε αυθαίρετα aspect ratios, προτιμώνται τα κλασσικά $4 : 3$ και $16 : 9$ για λόγους συμβατότητας με τα υπάρχοντα βίντεο.

Στις νέες LCD ή Plasma τηλεοράσεις, υπάρχει η ένδειξη Full HD (High Definition), μαζί με κάποιον αριθμό και γράμμα, πχ $1080i$. Ο αριθμός έχει να κάνει με το πόσες «οριζόντιες γραμμές» υπάρχουν στην οθόνη (δηλαδή το ύψος σε pixel). Για παράδειγμα, στο $1080i$ υπάρχουν 1080 οριζόντιες γραμμές, έτσι η ανάλυση που υποστηρίζεται είναι 1920×1080 στα $16 : 9$. Το i σημαίνει interlace σε αντίθεση με το progressive (p). Ένα άλλο παράδειγμα: $720p$ (1280×720 progressive). Οι έννοιες του interlace και του progressive θα εξηγηθούν στη συνέχεια.

Η επίδραση που έχει η ανάλυση σε μια εικόνα, μπορεί να φανεί στην εικόνα 1.1: Αρχικά, η δειγματοληψία έγινε σε ένα μόνο σημείο, οπότε η εικόνα θα έχει ένα μόνο χρώμα. Στη συνέχεια, σε 2×2 έχει 4 χρώματα, 5×5 25 χρώματα κοκ. Όσο η ανάλυση αυξάνει, τόσο θα βελτιώνεται και η ποιότητα του απεικονιζόμενου αντικειμένου. Σε ένα ψηφιακό βίντεο μπορεί ουσιαστικά μόνο να μειωθεί η ανάλυση του, πχ για αναπαραγωγή σε συσκευές με μικρή οθόνη ή εξοικονόμηση χώρου αποθήκευσης και εύρους ζώνης μετάδοσης. Η μείωση αυτή της ανάλυσης θα μπορούσε να γίνει είτε με πολύ απλό τρόπο, απλώς δηλαδή κόβοντας γραμμές και στήλες από pixel, είτε με πιο περίπλοκο τρόπο, συνδυάζοντας δηλαδή τις γραμμές που πρέπει να αποκοπούν σε μια νέα με τα pixel της νέας να έχουν το μέσο όρο των χρωμάτων των δύο γραμμών που ενώθηκαν.



Σχήμα 1.1: Επίδραση ανάλυσης σε μια εικόνα

Αύξηση της ανάλυσης θα μπορούσε θεωρητικά να γίνει με παρόμοιους τρόπους (διπλασιασμό γραμμών και στηλών) αλλά δεν έχει σχεδόν καμία ουσιαστική εφαρμογή μιας και από τη στιγμή που γίνεται η δειγματοληψία μιας εικόνας δε μπορεί να βρεθούν με κανένα τρόπο τα χρώματα που είχε ο χώρος ανάμεσα στα pixels, παρά μόνο να προσεγγιστούν με π.χ παρεμβολή.

1.4.6.2 Interlace

Μια τεχνική που χρησιμοποιείται κατά κόρον κατά την ψηφιοποίηση και μετάδοση τηλεοπτικού σήματος για τη βελτίωση της ποιότητας αυτού χωρίς όμως τη σπατάλη περισσότερου εύρους ζώνης μετάδοσης είναι η τεχνική του **interlace**. Για την απεικόνιση των εικόνων του βίντεο, ουσιαστικά υπάρχουν δύο τρόποι: Ο progressive και ο interlaced. Στο μεν progressive, κάθε εικόνα του βίντεο καταγράφεται, μεταδίδεται και απεικονίζεται ολόκληρη, γραμμή - γραμμή, από πάνω προς τα κάτω και αριστερά προς τα δεξιά. Αντίθετα, στο interlaced, η απεικόνιση γίνεται γραμμή παρά γραμμή, ξεκινώντας τη μία φορά από την πρώτη γραμμή και την άλλη από τη δεύτερη γραμμή, έτσι ώστε τη μια να απεικονίζονται οι περιττές γραμμές της εικόνας και την άλλη οι άρτιες. Κάθε μια από τις επιμέρους εικόνες που έχουν μόνον τις άρτιες ή μόνον τις περιττές γραμμές ονομάζεται πεδίο. Τα πεδία καταγράφονται σε διαδοχικές χρονικές στιγμές. Λόγω του τρόπου που λειτουργούν οι παλαιότερες τηλεοράσεις καθοδικού σωλήνα (CRT) τα δύο πεδία φαίνονταν σα μια ενιαία εικόνα στο ανθρώπινο μάτι, και έτσι μπορούσε να γίνει εξοικονόμηση εύρους ζώνης αφού μεταδίδονταν μόνο οι μισές γραμμές κάθε φορά.

Το πώς απεικονίζεται ένα interlaced βίντεο φαίνεται στο σχήμα 1.2: Στην εικόνα φαίνεται το πρόβλημα που δημιουργείται με το interlace (ονομάζεται δόντια ποντικιού): Ο άντρας κινείται, οπότε οι άρτιες και οι περιττές γραμμές

1.4. ΜΙΚΡΗ ΕΙΣΑΓΩΓΗ ΣΤΗ ΣΥΜΠΙΕΣΗ ΒΙΝΤΕΟ

δείχνουν διαφορετικά πράγματα (αφού η δειγματοληψία τους έγινε σε διαδοχικές χρονικές στιγμές). Η υπόλοιπη εικόνα είναι ακίνητη, άρα δεν δημιουργείται το πρόβλημα εκεί.

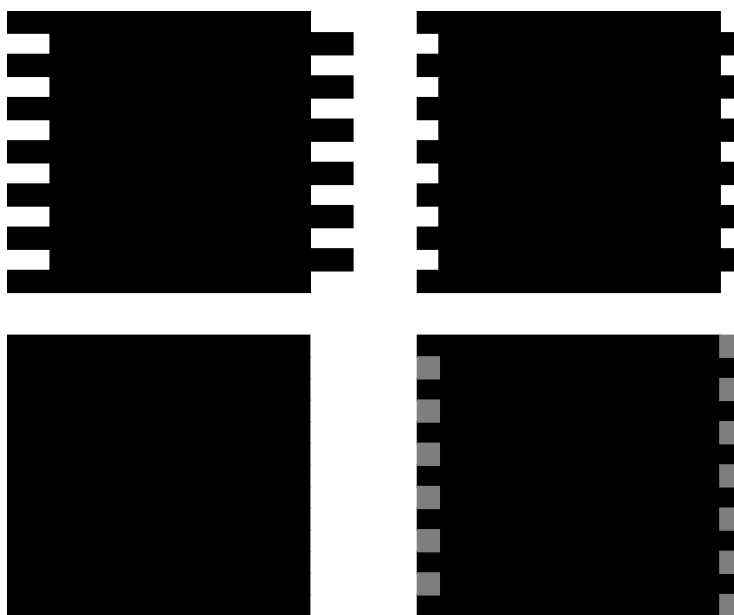


Σχήμα 1.2: Παράδειγμα *interlaced* βίντεο

Επειδή μόνο οι οθόνες καθοδικού σωλήνα μπορούν να απεικονίσουν σωστά το *interlaced* βίντεο, σε όλες τις άλλες περιπτώσεις γίνεται *deinterlacing*, δηλαδή συνδυάζονται τα δύο πεδία σε μια ενιαία εικόνα έτσι ώστε το βίντεο να γίνει *progressive*. Αυτό συμβαίνει συνήθως και κατά την ψηφιοποίηση του τηλεοπτικού σήματος προς συμπίεση, παρότι κάποια πρωτόκολλα συμπίεσης βίντεο μπορούν να συμπίεσουν απευθείας το *interlaced* βίντεο χωρίς να χρειαστεί μετατροπή του σε *progressive*.

Υπάρχουν γενικά τρεις διαφορετικές κατηγορίες μεθόδων για το *deinterlacing*:

- Οι άρτιες και οι περιπτές υποεικόνες συνδυάζονται σε μία ενιαία εικόνα. Είναι μια πολύ κλασική μέθοδος η οποία όμως σε εικόνες με γρήγορη κίνηση δημιουργεί προβλήματα. Για παράδειγμα, στο σχήμα 1.3 φαίνεται το ίδιο σχήμα (ένα τετράγωνο) όπου πάνω αριστερά κινείται οριζόντια με γρήγορη ταχύτητα, πάνω δεξιά κινείται οριζόντια με πιο αργή ταχύτητα ενώ κάτω αριστερά το σχήμα είναι ακίνητο (ή κινείται κάθετα). Επειδή τα πεδία είναι από διαφορετικές χρονικές στιγμές, τα άκρα του τετραγώνου δεν είναι πλέον ευθείες, και μάλιστα όσο γρηγορότερη είναι η κίνηση τόσο περισσότερο δημιουργείται το οδοντωτό σχήμα στα άκρα (δόντια ποντικιού). Για να μην δημιουργούνται τέτοια προβλήματα, υπάρχουν πιο εξελιγμένες τεχνικές στις οποίες τα δύο πεδία αναμιγνύονται και όπου υπάρχουν διαφορές βρίσκεται ο μέσος όρος αυτών, για παράδειγμα το τετράγωνο πάνω δεξιά θα φαίνεται όπως αυτό που είναι κάτω δεξιά.
- Κάθε υποεικόνα μετατρέπεται σε μια ξεχωριστή εικόνα (διπλασιάζοντας



Σχήμα 1.3: Περιπτώσεις deinterlace

τις γραμμές της έτσι ώστε να έχει τη σωστή ανάλυση). Εδώ δεν υπάρχουν τα προβλήματα της προηγούμενης μεθόδου, όμως οι εικόνες έχουν ουσιαστικά τη μισή ανάλυση, αφού για κάθε μια χρησιμοποιούνται μόνο οι μισές γραμμές.

Πιο εξελιγμένες τεχνικές χρησιμοποιούν ανίχνευση κίνησης έτσι ώστε να καλύψουν τις διαφορές μεταξύ των πεδίων. Για παράδειγμα, στο τετράγωνο του σχήματος 1.3 θα ανιχνευόταν η οριζόντια του κίνηση και θα ενώνονταν κατάλληλα οι εικόνες από τα δύο πεδία έτσι ώστε να μη δημιουργηθούν τα προβλήματα της απλής παράθεσης πεδίων, ούτε να μειωθεί η ανάλυση.

1.4.6.3 Ρυθμός μετάδοσης εικόνων

Ο ρυθμός μετάδοσης εικόνων (**frame rate**) έχει να κάνει με το πόσο γρήγορα πρέπει να προβάλλονται οι εικόνες του βίντεο έτσι ώστε αυτό να εμφανίζεται στη σωστή ταχύτητα και να είναι συγχρονισμένο με τον ήχο και τυχόν υποτίτλους. Μετράται σε εικόνες ανά δευτερόλεπτο (fps). Υπάρχουν διάφορες προκαθορισμένες τιμές για τα fps:

- Στο σινεμά, προβάλλονται παραδοσιακά 24 εικόνες ανά δευτερόλεπτο.
- Στην τηλεόραση PAL (ευρωπαϊκή), προβάλλονται 25 εικόνες ανά δευτερόλεπτο. Ουσιαστικά, επειδή χρησιμοποιείται interlaced μετάδοση, προβάλλονται 50 πεδία το δευτερόλεπτο, και έτσι καταλήγουν στις 25 εικόνες.

1.5. ΤΙ ΠΕΡΙΛΑΜΒΑΝΕΤΑΙ ΣΤΑ ΕΠΙΟΜΕΝΑ

- Στην τηλεόραση NTSC (αμερικάνικη), προβάλλονται $60 \times 1000/1001 = 59,94$ πεδία το δευτερόλεπτο, δηλαδή 29,97 εικόνες. Η εισαγωγή αυτού του παράξενου κλάσματος έγινε έτσι ώστε να μην υπάρχει παρεμβολή μεταξύ του χρώματος και του ήχου.
- Στα κινητά τηλέφωνα, χρησιμοποιούνται αρκετά μικρότερες αναλύσεις, της τάξης των 10 με 15 εικόνων ανά δευτερόλεπτο.
- Στο βίντεο υψηλής ποιότητας, υπάρχουν ρυθμοί της τάξης των 50 η 60 εικόνων ανά δευτερόλεπτο, δηλαδή διπλάσιες από ότι στην απλή τηλεόραση.

1.4.6.4 Ρυθμός μετάδοσης bit

Ο ρυθμός μετάδοσης bit (bitrate) είναι το πόσο εύρος ζώνης χρειάζεται για τη μετάδοση του βίντεο, δηλαδή το μέγεθος του βίντεο ανά δευτερόλεπτο. Εξαρτάται από την ανάλυση του βίντεο, το πόσο έχει αυτό συμπιεστεί, αλλά και από το περιεχόμενο του βίντεο. Δηλαδή ένα βίντεο με πολλά χρώματα θα χρησιμοποιεί περισσότερο bitrate από ένα βίντεο γεμάτο με μαύρες εικόνες, ομοίως και ένα βίντεο υψηλότερης ανάλυσης ή με λιγότερη συμπίεση. Συνήθως μετράται σε bit / s και πάντοτε εξαρτάται από το fps, αφού υπολογίζεται από το μέγεθος κάθε συμπιεσμένης εικόνας του βίντεο. Για παράδειγμα, για την μετάδοση του ίδιου ακριβώς βίντεο σε διπλάσιο fps θα χρειάζεται το διπλάσιο bitrate.

Η συμπίεση βίντεο τις περισσότερες φορές στοχεύει στη διατήρηση ενός καθορισμένου και σταθερού ρυθμού μετάδοσης bit για κάθε βίντεο έτσι ώστε να είναι εύκολη η μετάδοση αυτού από κανάλια συγκεκριμένου εύρους ζώνης ή η αποθήκευση αυτού σε μέσα συγκεκριμένης χωρητικότητας όπως τα DVD.

1.5 Τι περιλαμβάνεται στα επόμενα

Στο δεύτερο κεφάλαιο γίνεται μια εισαγωγή στις τεχνικές ανίχνευσης αλλαγών πλάνου και προτείνονται δύο συγκεκριμένες τεχνικές για χρήση στα επόμενα κεφάλαια, μιας και η ανίχνευση αλλαγής πλάνου χρησιμοποιείται σε μεγάλο βαθμό στις τεχνικές αναζήτησης βίντεο που παρουσιάζονται στην εργασία αυτή. Οι τεχνικές ανίχνευσης πλάνου που παρουσιάζονται είναι η ανίχνευση με βάση τον αριθμό των intra macroblock κάθε εικόνας και η ανίχνευση με βάση το ιστόγραμμα φωτεινότητας. Και για τις δύο πραγματοποιείται αρχικά μία θεωρητική ανάλυση του τρόπου λειτουργίας τους και ακολουθούν μια περιγραφή των υλοποιήσεων και κάποια πειραματικά αποτελέσματα

Στο τρίτο κεφάλαιο παρουσιάζονται θεωρητικά οι διάφορες τεχνικές αναζήτησης βίντεο που προτείνει η εργασία. Αφού γίνει μια εισαγωγή στην αναζήτηση βίντεο και τα προβλήματα της, παρουσιάζονται αναλυτικά οι τεχνικές αναζήτησης καθώς και μια σειρά από βελτιώσεις και παραλλαγές αυτών. Οι

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

τεχνικές αυτές είναι η πιθανοτική μέθοδος, η μέθοδος των intra macroblock και η μέθοδος της βάσης δεδομένων.

Στο τέταρτο κεφάλαιο παρουσιάζονται τρία πλήρη συστήματα αναζήτησης βίντεο τα οποία χρησιμοποιούν τις τεχνικές του τρίτου κεφαλαίου. Τα δύο πρώτα είναι παρόμοια με τη διαφορά ότι το μεν πρώτο χρησιμοποιεί μια γραφική διεπαφή χρήστη ενώ το δεύτερο χρησιμοποιεί διεπαφή κειμένου για ευκολότερη εξαγωγή πειραματικών αποτελεσμάτων. Το τρίτο σύστημα ονομάζεται συνεχής αναζήτηση και είναι ένα σύστημα το οποίο δέχεται μια σειρά από αποσπάσματα και αναζητά αυτά τα αποσπάσματα σε κάποιο ζωντανό και συνεχές πρόγραμμα. Η χαρακτηριστικότερη του εφαρμογή είναι προφανώς η αναζήτηση συγκεκριμένων διαφημίσεων σε κάποιο τηλεοπτικό πρόγραμμα.

Γίνεται αναφορά στην αρχιτεκτονική και υλοποίηση των συστημάτων αυτών καθώς και σε διάφορα κοινά χαρακτηριστικά που τυχόν έχουν. Έτσι μπορεί να διαφανούν οι απαιτήσεις ενός πλήρους συστήματος αναζήτησης βίντεο αλλά και τρόποι για την ικανοποίηση αυτών, ενώ γίνονται κατανοητοί οι λόγοι για μια σειρά από αποφάσεις που πάρθηκαν αλλά και ενισχύονται τα θεωρητικά και πειραματικά αποτελέσματα.

Τέλος, στο πέμπτο κεφάλαιο παρουσιάζονται μια σειρά από αναλυτικά πειραματικά αποτελέσματα για τη λειτουργία όλων των μεθόδων αναζήτησης που παρουσιάστηκαν, τόσο για την ταχύτητα λειτουργίας τους όσο και για το πόσο καλά τελικά δουλεύουν, ενώ στο έκτο κεφάλαιο αναφέρονται τα γενικά συμπεράσματα από την εργασία αυτή καθώς και η μελλοντική εργασία που θα μπορούσε να γίνει στον κλάδο της ευθείας αναζήτησης βίντεο.

Κεφάλαιο 2

Ανίχνευση αλλαγών πλάνου

ΕΠΙΕΙΔΗ η ανίχνευση των αλλαγών πλάνου είναι ένα απαραίτητο εργαλείο για τις τεχνικές αναζήτησης βίντεο που θα περιγραφούν στη συνέχεια, θα γίνει εδώ μια σύντομη ανασκόπηση των υπάρχοντων τεχνικών ανίχνευσης αλλαγής πλάνου και κατόπιν θα προταθούν δύο απλές μέθοδοι ανίχνευσης οι οποίες κατόπιν θα χρησιμοποιηθούν για την εξαγωγή πειραματικών αποτελεσμάτων για τις μεθόδους αναζήτησης.

Είναι πολύ σημαντικό εδώ να σημειωθεί ότι η εργασία αυτή δεν έχει ως σκοπό την αντικειμενική εύρεση των αλλαγών πλάνου ή σκηνης ενός βίντεο. Οι περισσότερες εργασίες στο αντικείμενο της εύρεσης αλλαγών πλάνου έχουν ως σκοπό τους το να βρουν τις αλλαγές πλάνου ενός βίντεο, όπως θα τις εύρισκε κάποιος άνθρωπος ο οποίος παρακολουθούσε το βίντεο. Η εύρεση των πραγματικών αλλαγών πλάνου είναι πολύ σημαντική για τη γρήγορη προεπισκόπηση ενός βίντεο, την κατάταξη του βίντεο σε κατηγορίες, την αναζήτηση βίντεο κ.α.

Αντίθετα, σε αυτή την εργασία ορίζονται κάποιες εικόνες του βίντεο ως «αλλαγές πλάνου», ακόμα και αν δεν είναι πραγματικές αλλαγές πλάνου όπως θα τις επέλεγε κάποιος άνθρωπος που παρακολουθεί το βίντεο. Οι αλλαγές πλάνου όμως που χρησιμοποιούνται στην εργασία

έχουν μια πολύ σημαντική ιδιότητα: Πρέπει να ανιχνεύονται οι ίδιες αλλαγές πλάνου στο ίδιο βίντεο, έστω και αν αυτό έχει υποστεί επεξεργασία, πχ να έχει αλλάξει η ανάλυση του, να έχει μειωθεί η ποιότητα του, να έχουν προστεθεί λογότυπα, να έχει αφαιρεθεί ο ήχος του κοκ. Αυτό είναι απαραίτητο, αφού, εφόσον οι αλλαγές πλάνου θα χρησιμοποιηθούν ως χαρακτηριστικό αναζήτησης ενός βίντεο, θα πρέπει να είναι οι ίδιες στο απόσπασμα και στο βίντεο από το οποίο το απόσπασμα προέρχεται.

2.1 Υπάρχουσες μέθοδοι

Η ανίχνευση αλλαγής πλάνου έχει ερευνηθεί εκτενώς εδώ και κάμποσα χρόνια μιας και με την εξαγωγή των εικόνων που έχουν αλλαγές πλάνου μπορεί να γίνει αναζήτηση βίντεο, μπορεί να δημιουργηθεί μια περίληψη του βίντεο,

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

μπορεί το βίντεο να χωριστεί σε τμήματα κλπ [HL01]. Μια αλλαγή πλάνου μπορεί να χωριστεί σε δύο κατηγορίες: Στην απότομη αλλαγή πλάνου (abrupt shot cut) όπου η εικόνα $i - 1$ ανήκει σε διαφορετικό πλάνο από ότι η εικόνα i και στην βαθμιαία αλλαγή πλάνου (gradual shot cut), όπου η εικόνα $i - M$ ανήκει σε διαφορετικό πλάνο από ότι η εικόνα i , αλλά στις εικόνες $i + 1$ ως $i - M - 1$ υπάρχει μια ανάμιξη των δύο αυτών πλάνων με βάση κάποιο εφφέ, όπως fade-in, fade-out, dissolve κοκ. Συνήθως οι βαθμιαίες αλλαγές πλάνου είναι και αλλαγές σκηνής αφού σπάνια οι σκηνοθέτες χρησιμοποιούν τέτοια εφφέ για να περάσουν σε άλλο πλάνο της ίδιας σκηνής.

Στις αναφορές [GKS00] και [Lie99] υπάρχουν εργασίες στις οποίες γίνεται σύγκριση διαφόρων τεχνικών εύρεσης τόσο απότομων όσο και βαθμιαίων αλλαγών σκηνής. Επίσης το TRECVID [KOS06] είναι ένας διαγωνισμός για συστήματα ανάλυσης βίντεο του οποίου ένα τμήμα έχει να κάνει με την αναγνώριση αλλαγών πλάνου.

Ένας λόγος για τον οποίο η εύρεση αλλαγών πλάνου είναι τόσο δημοφιλής είναι ότι είναι δυνατή ακόμα και χωρίς αποσυμπίεση του βίντεο ([WDV⁺03]). Έχουν προταθεί τρόποι για εύρεση αλλαγής πλάνου που χρησιμοποιούν τους συντελεστές DC του συμπιεσμένου κατά MPEG ([Bes04]), πληροφορίες που αφορούν την κίνηση στο βίντεο (όταν αλλάζει το πλάνο θα υπάρχει απότομη αλλαγή στην κατεύθυνση της κίνησης διαφόρων αντικειμένων του βίντεο [BGG99]), στο μέγεθος που έχει κάθε εικόνα όταν συμπιεστεί (αφού για μια εικόνα που είναι αλλαγή πλάνου δε θα μπορεί να χρησιμοποιηθεί η προηγούμενη για πρόβλεψη [LLZL04]) αλλά και τον τρόπο με τον οποίο έχει συμπιεστεί, δηλαδή τον τύπο κάθε ενός macroblock της εικόνας ([PC99] τα περισσότερα macroblock μιας εικόνας που δεν είναι αλλαγή πλάνου θα έχουν προβλεφθεί από αυτά της προηγούμενης εικόνας).

Κάποιες εργασίες αφορούν μόνο την εύρεση βαθμιαίων αλλαγών πλάνου, και μάλιστα με συγκεκριμένο τρόπο ([LZ01]) ενώ οι περισσότερες, όπως η [JBC05] αφορούν τόσο της απότομες όσο και τις βαθμιαίες αλλαγές πλάνου.

Οι περισσότερες από τις παραπάνω εργασίες χρησιμοποιούν σε γενικές γραμμές την ακόλουθη, σχετικά απλή στην κατανόηση της μέθοδο για την εύρεση απότομων αλλαγών πλάνου:

Από κάθε εικόνα i ενός βίντεο εξάγεται με βάση τα χαρακτηριστικά της κάποια μετρική M , έτσι ώστε να δημιουργείται η ακολουθία $M(i)$. Ορίζεται επίσης μια συνάρτηση $D(i, i - 1) = |M(i) - M(i - 1)|$ η οποία επιστρέφει τη διαφορά δύο γειτονικών εικόνων. Τέλος, υπάρχει κάποιο κατώφλι απόφασης για αλλαγή πλάνου ϵ . Τότε, αποφασίζεται ότι υπάρχει αλλαγή σκηνής στην εικόνα i αν $D(i, i - 1) > \epsilon$.

Ο βασικός αυτός αλγόριθμος επιδέχεται μια σειρά από παραλλαγές, οι οποίες αφορούν τα χαρακτηριστικά που εξάγονται από κάθε βίντεο, τη μετρική M που χρησιμοποιείται, το αν θα χρησιμοποιηθούν όλες οι εικόνες του βίντεο ή όχι, το D το οποίο συνήθως δεν έχει την απλή μορφή της προηγούμενης παραγράφου αλλά λαμβάνει υπ' όψιν του διάφορες επιπλέον παραμέτρους και τέλος το ϵ του οποίου επίσης η τιμή μπορεί να αλλάζει [DNR05, Bes04, LLZL04].

2.2 Μέτρηση επίδοσης

Η επίδοση των μεθόδων ανίχνευσης αλλαγής πλάνου έχει να κάνει με την ταχύτητα τους αλλά και με το πόσο σωστά γίνεται η ανίχνευση. Σχετικά με το δεύτερο σκέλος, είναι σημαντικό να οριστούν δύο μεγέθη που χρησιμοποιούνται κατά κόρον στην βιβλιογραφία ([GKS00]). Σε αυτά, ελέγχεται σε κάποιο δοκιμαστικό βίντεο το πόσες αλλαγές πλάνου ανιχνεύθηκαν σωστά N_c , το πόσες χάθηκαν N_m και το πόσες βρέθηκαν χωρίς να υπάρχουν N_f .

Ορισμός 2.2.1 Ως Ανάκληση (recall) ορίζεται το μέγεθος

$$R = \frac{N_c}{N_c + N_m},$$

δηλαδή είναι ένα μέτρο του πόσες αλλαγές πλάνου δεν ανιχνεύονται και όσο πιο κοντά είναι στο 1 τόσο λιγότερες αλλαγές πλάνου θα χάνονται.

Ορισμός 2.2.2 Ως Ακρίβεια (precision) ορίζεται το μέγεθος

$$P = \frac{N_c}{N_c + N_f},$$

δηλαδή είναι ένα μέτρο του πόσες αλλαγές πλάνου βρίσκονται χωρίς να υπάρχουν. Όσο πιο κοντά είναι στη μονάδα τόσο λιγότερες είναι οι εσφαλμένες αλλαγές πλάνου.

Όπως μπορεί να γίνει κατανοητό, υπάρχει μια αντίστροφη σχέση μεταξύ της ανάκλησης και της ακρίβειας, αφού όσο χαλαρότερα γίνονται τα κριτήρια για να μη χαθεί κάποια αλλαγή πλάνου τόσο περισσότερες θα ανιχνεύονται ενώ δεν υπάρχουν, και αντίστροφα, όσο αυστηρότερα γίνονται τα κριτήρια για να μην ανιχνεύονται αλλαγές πλάνου τόσο περισσότερες θα χάνονται.

Η απόδοση των σύγχρονων τεχνικών ανίχνευσης αλλαγής πλάνου είναι αρκετά καλή. Για παράδειγμα, η [JBC05] επιτυγχάνει ανάκληση ίση με 0,99 και ακρίβεια ίση με 0,98 στις απότομες αλλαγές πλάνου, ενώ στις βαθμιαίες πετυχαίνει 0,92 και 0,81 αντίστοιχα. Επίσης, η ταχύτητα εξαγωγής των αλλαγών πλάνου είναι ίση η ακόμα και ξεπερνάει την ταχύτητα πραγματικού χρόνου ([Bes04]). Αντίστοιχα είναι και τα αποτελέσματα των μεθοδολογιών του TRECVID [KOS06].

2.3 Νέες προκλήσεις

Παρά τη βιβλιογραφία που υπάρχει ήδη στο αντικείμενο της ανίχνευσης αλλαγών πλάνου, παρουσιάζονται συνεχώς νέες προκλήσεις στη συγκεκριμένη περιοχή. Μια σημαντική πρόκληση έχει να κάνει με τη γρήγορη ανίχνευση αλλαγών πλάνου σε βίντεο συμπιεσμένο με το πρωτόκολλο H.264.

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

Το H.264 [WSBL03] είναι ένα νέο και πολύ δημοφιλές πρωτόκολλο συμπίεσης βίντεο. Έχει μια σειρά από νέα χαρακτηριστικά συμπίεσης με τη βοήθεια των οποίων, όπως φαίνεται και από την [OBL⁺04] μπορεί να πετύχει ίδια ποιότητα (σηματοθορυβικό λόγο) έχοντας το μισό bit rate από τα παλαιότερα πρωτόκολλα. Το H.264 είναι ένα εξαιρετικά πολύπλοκο πρωτόκολλο και οι λειτουργίες της κωδικοποίησης και αποκωδικοποίησης απαιτούν αρκετό χρόνο [HJKH03, OBL⁺04], για αυτό και η χρήση τεχνικών που απαιτούν πλήρη αποσυμπίεση του βίντεο θα είναι καλό να αποφεύγεται.

Οι περισσότερες υπάρχουσες τεχνικές που υπάρχουν για εύρεση αλλαγών πλάνου χωρίς αποσυμπίεση δε δουλεύουν στο H.264 λόγω της δομής του αλλά και των νέων εργαλείων συμπίεσης που χρησιμοποιούνται. Για παράδειγμα, λόγω του ότι στο H.264 γίνεται και χωρική πρόβλεψη (και όχι μόνο χρονική), δεν είναι εύκολη η εξαγωγή των DC συντελεστών κάθε macroblock όπως είναι στο MPEG-2. Αυτό συμβαίνει επειδή οι τιμές του κάθε macroblock υπολογίζονται με βάση τις τιμές άλλων macroblock. Έτσι, οι DC συνιστώσες δεν περιέχουν το μέσο όρο της ενέργειας στο συγκεκριμένο macroblock αλλά τη διαφορά από κάποιο άλλο.

Οι μόνες πληροφορίες που μπορούν εύκολα να εξαχθούν από το συμπιεσμένο βίντεο είναι ο τύπος του κάθε macroblock και οι πληροφορίες για την κίνηση του κάθε macroblock (αν έχει γίνει πρόβλεψη αυτού με βάση την κίνηση). Έτσι, οι εργασίες ανίχνευσης αλλαγής πλάνου στο H.264 χρησιμοποιούν παρόμοιες τεχνικές με αυτές της [PC99], δηλαδή εξετάζουν τον τύπο κάθε macroblock καθώς και το αν κάθε εικόνα είναι I, P ή B ούτως ώστε να καθοριστεί το αν μια εικόνα είναι αλλαγή πλάνου. Παραδείγματα τέτοιων εργασιών είναι τα [SB08] και [YLWZ04]. Στην εργασία [BDC⁺] εκτός του τύπου κάθε macroblock εξετάζονται και τα χαρακτηριστικά των intra macroblock για να μειωθούν οι εσφαλμένες αναγνωρίσεις αλλαγών πλάνου λόγω εικόνων I που εισάγονται από τον κωδικοποιητή¹. Ένα macroblock που κωδικοποιείται κατά intra μπορεί είτε να κωδικοποιηθεί με χωρική πρόβλεψη ως έχει, είτε να χωριστεί σε 16 επιμέρους τμήματα και σε κάθε ένα να γίνει ξεχωριστή χωριστή πρόβλεψη. Έτσι, συγκρίνοντας τον τύπο των intra macroblock με αυτά των προηγούμενων εικόνων μπορεί να βρεθεί αν υπάρχει πραγματική αλλαγή πλάνου ή απλώς έγινε εισαγωγή μιας εικόνας I. Τέλος, στην [KL08] πραγματοποιείται επίσης εξέταση των intra macroblock έτσι ώστε να μπορεί να γίνει ανίχνευση αλλαγής πλάνου χωρίς αποσυμπίεση.

2.4 Μέθοδος intra macroblock

Σε αυτή την εργασία χρησιμοποιείται μια απλή μέθοδος για την ανίχνευση αλλαγών πλάνου, παρόμοια με αυτές της παραπάνω ενότητας. Το σημαντικό στοιχείο της μεθόδου είναι ότι λειτουργεί με βίντεο συμπιεσμένο κατά H.264

¹Κυρίως για υποστήριξη seek, fast forward, error cancelation κοκ

χωρίς να χρειάζεται η αποσυμπίεση του βίντεο αυτού και έτσι επιλύει ένα από τα προβλήματα που αναφέρθηκαν για την ανίχνευση πλάνου.

2.4.1 Λειτουργία

Η μέθοδος που προτείνεται δουλεύει σε ένα βίντεο που είναι συμπιεσμένο κατά H.264 και οι εικόνες του έχουν τη μορφή *IPPPPP...* Αυτό σημαίνει ότι η πρώτη εικόνα ενός βίντεο είναι τύπου *I* και όλες οι επόμενες είναι τύπου *P*. Αυτή η μορφή επιτρέπεται στο H.264 ενώ δεν επιτρεπόταν σε παλαιότερα πρωτόκολλα ([WSBL03]) αφού στο H.264 γίνονται πράξεις μόνο μεταξύ ακεραίων και έτσι δε συσσωρεύονται αριθμητικά σφάλματα ώστε να είναι απαραίτητη μια εικόνα *I* για να «καθαριστούν» τα σφάλματα αυτά. Υπάρχουν παραλλαγές της μεθόδου που αναφέρθηκαν στην προηγούμενη ενότητα οι οποίες μπορούν να λειτουργήσουν ακόμα και με βίντεο αυθαίρετης μορφής.

Έτσι, μετά την πρώτη εικόνα *I*, υπολογίζεται το ποσοστό των κωδικοποιημένων κατά intra macroblock της εικόνας *i* και αν αυτό το ποσοστό είναι μεγαλύτερο από κάποιο κατώφλι τότε σημαίνει ότι υπάρχουν πολλές αλλαγές σε σχέση με την εικόνα *i - 1* άρα η εικόνα *i* είναι αλλαγή πλάνου. Το κατώφλι αυτό μπορεί να είναι είτε σταθερό είτε να προσαρμόζεται ανάλογα με την ποιότητα συμπίεσης του βίντεο, μιας και βίντεο καλύτερης ποιότητας θα έχει περισσότερα κωδικοποιημένα κατά intra macroblocks. Το σημαντικό πλεονέκτημα της μεθόδου είναι ότι η πληροφορία που χρησιμοποιείται, δηλαδή το ποσοστό των intra macroblock σε κάθε εικόνα, μπορεί να εξαχθεί χωρίς να χρειάζεται αποσυμπίεση του βίντεο, κάτι που την καθιστά πολύ γρήγορη όπως θα φανεί και αργότερα.

Το μειονέκτημα αυτής της μεθόδου έχει να κάνει με το γεγονός ότι το βίντεο πρέπει να είναι της μορφής *IPPPPP...* η οποία δε χρησιμοποιείται ευρέως αφού η μη ύπαρξη εικόνων *I* δεν επιτρέπει τη γρήγορη αναζήτηση και η μη ύπαρξη εικόνων *B* δεν επιτρέπει την πρόβλεψη με βάση επόμενες εικόνες. Ο συνήθης τύπος δηλαδή είναι *IBPBP...* με εισαγωγή ενός *I* καθ' 15 ή 25 εικόνες.

Παρόλα αυτά, η προτεινόμενη μέθοδος είναι εφαρμόσιμη με μικρές αλλαγές όταν υπάρχουν διάσπαρτα *I* frames στο βίντεο. Οι λύσεις είναι δύο:

- Όλες οι εικόνες *I* θεωρούνται αλλαγές πλάνου, κάτι που θα δημιουργήσει πολλές αλλαγές πλάνου ενώ αυτές δεν υπάρχουν.
- Οι πρόσθετες εικόνες *I* αγνοούνται, κάτι που έχει ως αποτέλεσμα το χάσιμο κάποιων αλλαγών πλάνου, αφού δε θα ανιχνεύονται οι αλλαγές πλάνου που τυχαίνουν μαζί με μια εικόνα *I*.

Οι αλλαγές πλάνου που θα χαθούν με τη δεύτερη λύση είναι πολύ λιγότερες από αυτές που θα δημιουργηθούν με την πρώτη λύση. Όμως η επιλογή της δεύτερης λύσης δεν είναι πάντοτε εύκολη, μιας και πολλοί κωδικοποιητές, όταν αντιληφθούν ότι η επόμενη εικόνα αποτελεί αλλαγή πλάνου, την κωδικοποιούν κατά *I* και έτσι δεν υπάρχει κάποιος ασφαλής τρόπος για τον εύκολο διαχωρισμό

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

των πραγματικών αλλαγών πλάνου από τις εικόνες I . Σημειώνεται ότι υπάρχουν στη βιβλιογραφία τεχνικές με τη βοήθεια των οποίων να μπορεί να καθοριστεί με μεγάλη πιθανότητα το αν κάποια εικόνα I είναι ή όχι αλλαγή πλάνου ([BDC⁺]).

Αναφορικά με τις εικόνες B τώρα, μπορούν να χρησιμοποιηθεί η μέθοδος της [PC99]: Εκτός από τον αριθμό των intra macroblock κάθε εικόνας, χρησιμοποιούνται και τα αποτελέσματα της πρόβλεψης κίνησης των εικόνων B . Έστω το παράδειγμα PBP . Αν η αλλαγή πλάνου είναι στην πρώτη εικόνα, τότε η πρώτη εικόνα P θα έχει πολλά intra macroblock και η B θα προβλεφθεί κυρίως από την πρώτη εικόνα P . Αν είναι στη δεύτερη εικόνα (δηλαδή στο B), τότε η δεύτερη εικόνα P θα έχει κυρίως intra macroblock και η B θα προβλεφθεί με βάση αυτήν. Αν τέλος η αλλαγή πλάνου είναι στην τρίτη εικόνα, τότε το B θα προβλεφθεί με βάση την πρώτη εικόνα P , ενώ το δεύτερο P θα έχει περισσότερα intra. Έτσι, ακόμα και στην περίπτωση ενός βίντεο που περιέχει και εικόνες B είναι εύκολη η εξαγωγή των αλλαγών πλάνου χωρίς να χρειάζεται αποσυμπίεση του βίντεο.

2.4.2 Υλοποίηση

Η μέθοδος ανίχνευσης αλλαγών πλάνου που προτείνεται υλοποιήθηκε με τη βοήθεια της βιβλιοθήκης ανοιχτού κώδικα FFmpeg [FFM]. Η FFmpeg είναι μια βιβλιοθήκη που υλοποιεί μια σειρά από πρωτόκολλα συμπίεσης και αποσυμπίεσης βίντεο. Ο κώδικας της βιβλιοθήκης είναι ανοιχτός και είναι σχετικά εύκολο να τροποποιηθούν τα υπάρχοντα πρωτόκολλα ούτως ώστε να προστεθεί νέα λειτουργικότητα. Μια γενική προσέγγιση της εξαγωγής στοιχείων από βίντεο παρουσιάζεται στην ενότητα 4.1.

Αυτό λοιπόν που έγινε στα πλαίσια της εργασίας αυτής ήταν η τροποποίηση της βιβλιοθήκης FFmpeg και ειδικότερα του αρχείου που περιείχε τον κώδικα για την αποσυμπίεση του H.264 και η εισαγωγή εκεί κάποιων μεταβλητών οι οποίες κρατούν τον τύπο κάθε ενός macroblock, δηλαδή ένας μετρητής για τα intra, ένας για τα inter και ένας για τα skip macroblock.

Έτσι, δημιουργήθηκε ένα πρόγραμμα το οποίο το οποίο χρησιμοποιεί την τροποποιημένη βιβλιοθήκη FFmpeg και με το που τελειώνει η αποσυμπίεση κάθε εικόνας να υπολόγιζε το ποσοστό των intra macroblock κάθε ενός frame και τα γράφει σε ένα αρχείο το ένα κάτω από το άλλο. Το είδος αυτό των αρχείων που περιέχει την ακολουθία των ποσοστών ονομάζεται αρχείο ctxt και το πρόγραμμα που τη δημιουργεί H264GetCtxt.

Τέλος, δημιουργήθηκε ένα πρόγραμμα ανίχνευσης αλλαγής πλάνου με όνομα CtxtShotDet το οποίο εξετάζει μία μία τις τιμές ενός αρχείου ctxt και αν κάποια ήταν πάνω από ένα κατώφλι να επέλεγε την εικόνα ως αλλαγή σκηνής. Η τιμή για το κατώφλι επιλέχτηκε με βάση τα πειραματικά αποτελέσματα και μπορούσε να δοθεί ως παράμετρος στο πρόγραμμα. Το γιατί χρησιμοποιήθηκαν δύο ξεχωριστά προγράμματα και ένα ενδιάμεσο αρχείο ctxt θα φανεί στα επόμενα κεφάλαια, αφού τα ίδια εργαλεία θα χρησιμοποιηθούν και για την αναζήτηση βίντεο.

2.4. ΜΕΘΟΔΟΣ INTRA MACROBLOCK

Σε περίπτωση που το βίντεο δεν ήταν συμπιεσμένο με το H.264 ή ήταν άλλου τύπου, για παράδειγμα *IBPBPBP...* τότε γινόταν χρήση του προγράμματος MPlayer [MPL] ούτως ώστε να γίνει η μετατροπή του σε H.264 κατάλληλης μορφής και στη συνέχεια γινόταν εξαγωγή των αλλαγών πλάνου αν έπρεπε να χρησιμοποιηθεί η μέθοδος των intra macroblock (υπάρχουν αντίστοιχα μέθοδοι που ελέγχουν τα macroblock βίντεο συμπιεσμένων με άλλα πρωτόκολλα, η υλοποίηση όμως αυτών των μεθόδων δεν είχε αξία στα πλαίσια της εργασίας αυτής). Φυσικά, για ένα βίντεο αυθαίρετης μορφής μπορεί να χρησιμοποιηθεί και η μέθοδος των ιστογραμμάτων φωτεινότητας που θα παρουσιαστεί στην ενότητα 2.5.

2.4.3 Παρουσίαση της μεθόδου

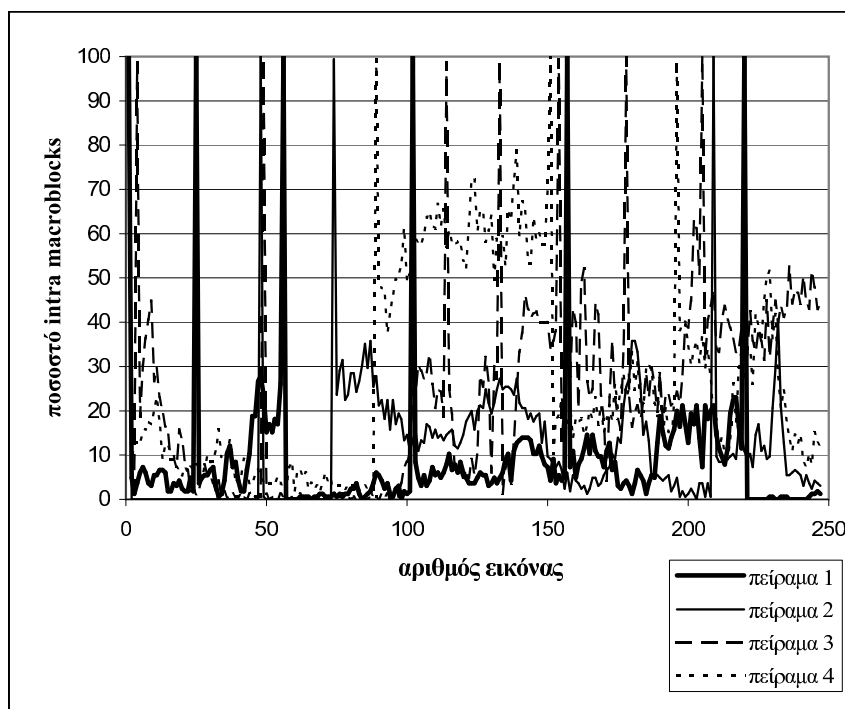
Παρότι στην ενότητα 2.6 γίνεται αναλυτική παρουσίαση της επίδοσης των μεθόδων, εδώ παρουσιάζονται κάποια αρχικά πειράματα που δείχνουν τον τρόπο λειτουργίας της μεθόδου. Έτσι, σχετικά με την ανάκληση και ακρίβεια της αντίχρευσης αλλαγής πλάνου με τη βοήθεια των intra macroblock, στο σχήμα 2.1 φαίνονται τα ποσοστά intra macroblock για κάθε εικόνα για τέσσερα βίντεο μεγέθους 250 εικόνων το καθένα. Τα βίντεο αυτά περιείχαν αποσπάσματα από τέσσερις διαφορετικές κινηματογραφικές ταινίες.

Τα ποσοστά παίρνουν τιμές πάνω από 90% στις περιπτώσεις όπου υπάρχει αλλαγή πλάνου στο βίντεο, έτσι αν το κατώφλι τεθεί ίσο με 90% η απόφαση λοιπόν για την ύπαρξη αλλαγής πλάνου είναι αρκετά εύκολη. Σαν παράδειγμα, στο πείραμα 1, υπάρχουν αλλαγές πλάνου στις εικόνες 25, 56, 102, 157 και 220, ακριβώς οι ίδιες που φαίνονται και από το σχήμα.

Στα συγκεκριμένα παραδείγματα υπάρχει απόλυτη επιτυχία, κάτι όμως που δε συμβαίνει στη γενική περίπτωση όπως θα φανεί και στην ενότητα 2.6 μιας και σε περιπτώσεις που κατά την αλλαγή πλάνου το νέο πλάνο μοιάζει με το πρώτο τα ποσοστά των intra macroblock θα είναι μειωμένα άρα θα χαθεί μια σκηνή, ενώ σε περιπτώσεις που δεν υπάρχει αλλαγή πλάνου αλλά υπάρχει πολύ γρήγορη κίνηση μπορεί να μη μπορεί να γίνει πρόβλεψη κάποιας εικόνας με βάση την προηγούμενη και έτσι εκεί να είναι πολύ μεγάλο το ποσοστό των intra macroblock και να γίνει εσφαλμένη αντίληψη κάποιων σκηνών. Παράδειγμα αυτής της περίπτωσης είναι στις εικόνες 100 ως 150 περίπου του πειράματος 3, όπου το ποσοστό των intra macroblock φτάνει ως και το 80% χωρίς να υπάρχει αλλαγή πλάνου.

Έτσι, για αυτά τα δοκιμαστικά βίντεο και η ανάκληση και η ακρίβεια των αντικειμενικών αλλαγών πλάνου φτάνουν το 100%, κάτι που φυσικά δε θα ισχύει στη γενική περίπτωση που περιλαμβάνει βίντεο που έχουν κακή ποιότητα ή βίντεο με πολύ γρήγορη κίνηση όπως αναφέρθηκε στην παραπάνω παράγραφο.

Στο σχήμα 2.2 φαίνεται ένα χαρακτηριστικό παράδειγμα αυτού του προβλήματος: Στο πείραμα 1, ένα βίντεο έχει κωδικοποιηθεί με συντελεστή κβαντισμού Q_p ίσο με 20, δηλαδή με πολύ καλή ποιότητα, ενώ στο πείραμα 2 με συντελεστή κβαντισμού ίσο με 40, δηλαδή με χειρότερη ποιότητα. Οι αλλαγές πλάνου είναι



Σχήμα 2.1: Ποσοστά *intra macroblock* ανα εικόνα

στις εικόνες 89, 151 και 195.

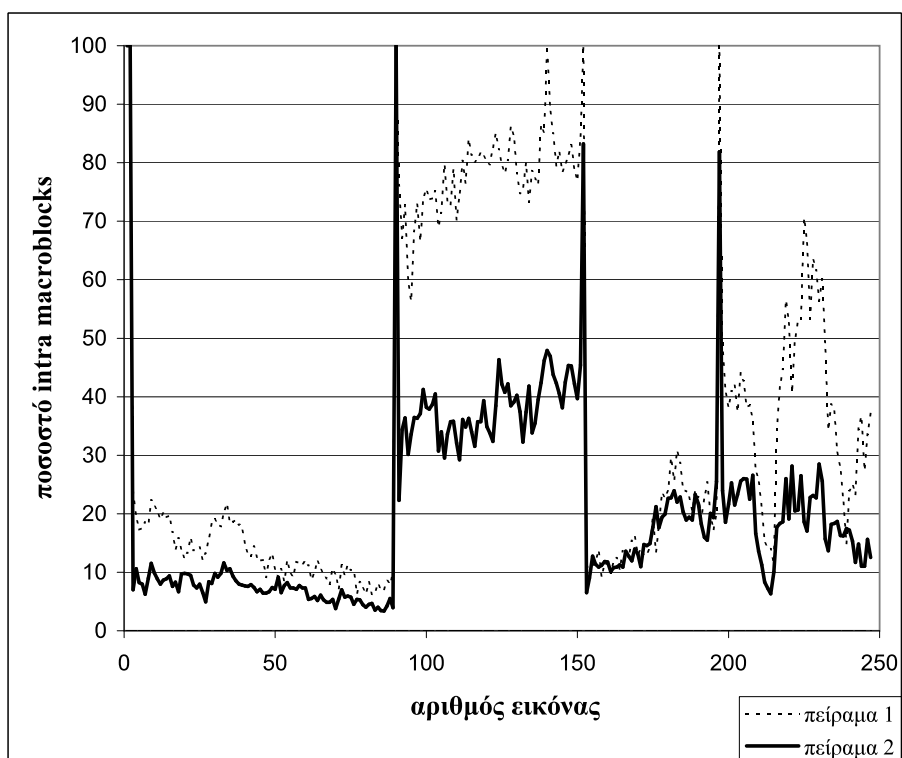
Αν το κατώφλι τεθεί ίσο με 90, στην περίπτωση της συμπίεσης με $Q_p = 20$ έχει ανιχνευθεί μια επιπλέον αλλαγή πλάνου στην εικόνα 141, ενώ στην περίπτωση του $Q_p = 40$ δεν έχουν ανιχνευθεί οι αλλαγές πλάνου στις εικόνες 151 και 195. Αν τώρα το κατώφλι τεθεί ίσο με 80, θα ανιχνευθούν μεν όλες οι αλλαγές πλάνου για $Q_p = 40$, όμως θα ανιχνευθούν και πολλές αλλαγές πλάνου ενώ δεν υπάρχουν στην περίπτωση του $Q_p = 20$ στις εικόνες 120 ως 150. Φαίνεται λοιπόν ότι ανάλογα με τα χαρακτηριστικά που έχει κάποιο βίντεο, μπορεί στις ίδιες εικόνες να ανιχνεύονται και να μην ανιχνεύονται αλλαγές πλάνου.

2.5 Μέθοδος ιστογραμμάτων φωτεινότητας

Η μέθοδος των *intra macroblock* έχει το μειονέκτημα ότι πρέπει το βίντεο να είναι H.264 και μάλιστα κατάλληλης μορφής για να γίνει η εξαγωγή των αλλαγών πλάνου. Μια διαφορετική μέθοδος που χρησιμοποιεί το ιστογράμμα φωτεινότητας κάθε εικόνας θα προταθεί σε αυτή την ενότητα.

Η βασική ιδέα της μεθόδου έχει να κάνει με την εξαγωγή του ιστογράμματος φωτεινότητας για κάθε εικόνα ενός βίντεο και σύγκριση του ιστογράμματος της εικόνας i με αυτό της εικόνας $i - 1$. Αν η διαφορά τους ξεπεράσει κάποιο κατώφλι t τότε η εικόνα i αποτελεί αλλαγή πλάνου. Το ιστογράμμα φωτεινότητας μπορεί εύκολα να εξαχθεί από την DC-συνιστώσα κάθε *macroblock* όταν

2.5. ΜΕΘΟΔΟΣ ΙΣΤΟΓΡΑΜΜΑΤΩΝ ΦΩΤΕΙΝΟΤΗΤΑΣ



Σχήμα 2.2: Σφάλματα στην ανίχνευση

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

η συμπίεση έχει γίνει από κάποιο πρωτόκολλο της οικογένειας των MPEG, ή, αν το βίντεο είναι συμπιεσμένο με κάποιο τελείως αυθαίρετο πρωτόκολλο, μπορεί να γίνει πλήρης αποσυμπίεση αυτού και να εξαχθεί το ιστόγραμμα από το αποσυμπιεσμένο βίντεο. Σε κάθε περίπτωση, δε χρειάζεται η μετατροπή του σε H.264.

2.5.1 Περιγραφή

Για τον υπολογισμό του ιστογράμματος φωτεινότητας χρησιμοποιείται μόνο η Y συνιστώσα του YUV χρώματος κάθε σημείου, και επειδή αυτή είναι ένα byte με τιμές από 0 ως 255 το ιστόγραμμα μπορεί εύκολα να υπολογιστεί με βάση τον τύπο

$$\forall_y \forall_x \text{ιστόγραμμα[εικόνα}[x][y]] ++,$$

με το y να παίρνει τιμές από 0 ως το ύψος και το x από 0 ως το πλάτος του βίντεο. Έτσι, το ιστόγραμμα μιας εικόνας είναι ένας πίνακας 256 στοιχείων, και ανάλογα με την τιμή κάθε σημείου της εικόνας αυξάνεται κατά ένα το αντίστοιχο σημείο του πίνακα αυτού. Όπως αναφέρθηκε και στην εισαγωγή στους χρωματικούς χώρους, η συνιστώσα Y έχει περισσότερη πληροφορία για αυτό και χρησιμοποιείται μόνον αυτή. Οι συνιστώσες U και V χρησιμοποιούνται σε ιστογράμματα χρωμάτων.

Ακολουθεί η ομαδοποίηση των τιμών ώστε να μειωθεί ο υπολογιστικός φόρτος κατά τη σύγκριση των τιμών αλλά και να γίνει πιο ανεκτικός ο αλγόριθμος σύγκρισης σε μικρές διαφορές στη φωτεινότητα. Αντί για 256 τιμές, το ιστόγραμμα πλέον θα περιλαμβάνει 64 τιμές, όπου

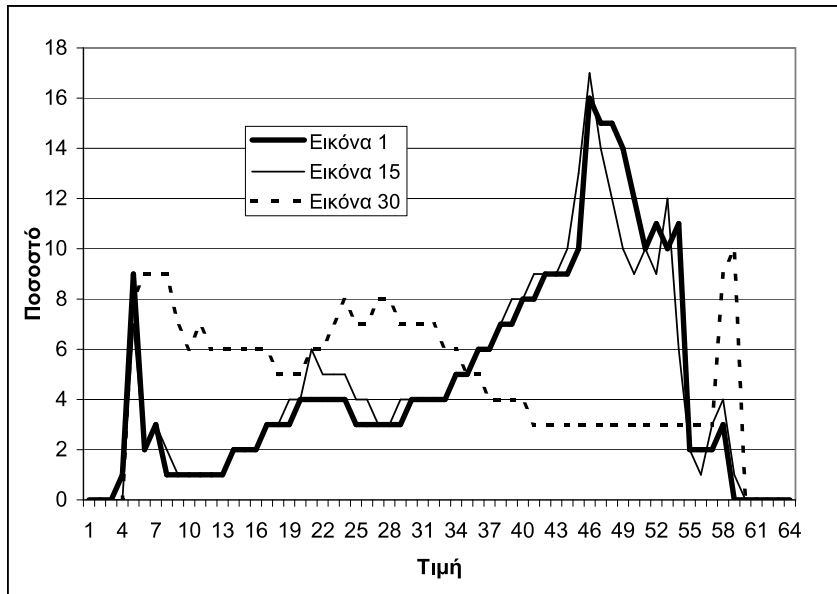
$$\text{νέο}[i] = \sum_{j=0}^3 \text{παλιό}[i \times 4 + j].$$

Δηλαδή, οι τιμές αθροίζονται και ομαδοποιούνται σε τετράδες, οπότε το ιστόγραμμα πλέον θα έχει 64 αντί για 256 κουβάδες. Στη συνέχεια, αντί για απόλυτους αριθμούς, θα γίνει η μετατροπή του σε ποσοστά, έτσι ώστε να υπάρχει ανοχή σε τυχόν αλλαγή ανάλυσης της εικόνας. Έτσι, κάθε τιμή του νέου ιστογράμματος θα είναι ίση με την αντίστοιχη τιμή του παλαιού διαιρεμένη με το άθροισμα όλων των τιμών, δηλαδή με την ανάλυση της εικόνας

$$\text{νέο}[i] = \frac{\text{παλιό}[i]}{\text{ύψος} \times \text{πλάτος}}.$$

Παραδείγματα ιστογραμμάτων φαίνονται στο σχήμα 2.3. Εκεί, υπάρχουν τα ιστογράμματα φωτεινότητας για την εικόνα 1, την εικόνα 15 και την εικόνα 30 ενός βίντεο. Οι εικόνες 1 και 15 ανήκουν στο ίδιο πλάνο ενώ η εικόνα 30 σε διαφορετικό, κάτι που φαίνεται και από τη μορφή των τριών ιστογραμμάτων.

2.5. ΜΕΘΟΔΟΣ ΙΣΤΟΓΡΑΜΜΑΤΩΝ ΦΩΤΕΙΝΟΤΗΤΑΣ



Σχήμα 2.3: Παράδειγμα ιστογράμματος

Τέλος, για τη σύγκριση των τιμών δύο ιστογραμμάτων, χρησιμοποιείται ο τύπος

$$D = \sum_{j=0}^{64} |\text{ιστόγραμμα}_i[j] - \text{ιστόγραμμα}_{i-1}[j]|.$$

Έτσι, δημιουργείται μια ακολουθία από διαφορές του ιστογράμματος της εικόνας i από αυτό της προηγούμενης της, και αν αυτή η διαφορά ξεπεράσει κάποιο καθορισμένο κατώφλι τότε η εικόνα i επιλέγεται ως αλλαγή πλάνου. Η μέγιστη τιμή της διαφοράς δύο ιστογραμμάτων θα είναι ίση με 200, και θα επιτυγχάνεται για παράδειγμα όταν και τα δύο ιστογράμματα έχουν το 100 % των τιμών της φωτεινότητας σε διαφορετικούς κουβάδες.

Όπως έχει ήδη αναφερθεί και πιο πάνω, αντί για το πλήρες ιστόγραμμα φωτεινότητας κάθε εικόνας, θα μπορούσε να δημιουργηθεί και το DC ιστόγραμμα φωτεινότητας, δηλαδή από κάθε intra macroblock που ανήκει σε μια εικόνα I θα χρησιμοποιούνταν μόνο ο μέσος όρος των τιμών του. Αυτή η μέθοδος ουσιαστικά μειώνει 64 φορές την ανάλυση του ιστογράμματος, αφού αντί για $8 \times 8 = 64$ τιμές για κάθε macroblock θα μετρηθεί μόνο μία, ο μέσος όρος αυτών. Το DC ιστόγραμμα φωτεινότητας είναι πολύ σημαντικό σε πρωτόκολλα όπως το MPEG2 στα οποία η DC συνιστώσα κάθε macroblock είναι εύκολο να βρεθεί χωρίς αποκωδικοποίηση του βίντεο, κάτι που αυξάνει πολύ την ταχύτητα ανίχνευσης ([YL95, PBG01]).

Για να μπορέσει να δουλέψει μια τέτοια μέθοδος θα πρέπει να δημιουργούνται εικόνες I σε τακτά χρονικά διαστήματα και να είναι εύκολη η εξαγωγή της DC συνιστώσας κάθε macroblock. Το H.264, είναι αρκετά εξελιγμένο έτσι ώστε

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

να είναι δυνατή η δημιουργία βίντεο που να έχει μόνο μία εικόνα I (την πρώτη) ενώ και ο τρόπος συμπίεσης των *intra macroblock* έχει αλλάξει και δεν είναι εύκολη η εξαγωγή της DC συνιστώσας, οπότε δε μπορεί να γίνει χρήση DC ιστογράμματος φωτεινότητας σε H.264([WSBL03]) .

2.5.2 Υλοποίηση

Για την υλοποίηση αυτής της μεθόδου χρησιμοποιήθηκε η βιβλιοθήκη FFMpeg [FFM]. Δεν έγινε τροποποίηση της βιβλιοθήκης, αλλά υλοποιήθηκε ένα απλό πρόγραμμα το οποίο με χρήση των μεθόδων της βιβλιοθήκης έκανε αποσυμπίεση σε κάποιο βίντεο. Στη συνέχεια, για κάθε μια αποσυμπίεσμένη εικόνα, και επειδή τα MPEG βίντεο χρησιμοποιούν το χώρο YUV, γινόταν απλώς η καταμέτρηση των τιμών φωτεινότητας και η κατασκευή του ιστογράμματος, όπως ακριβώς αναφέρθηκε στην προηγούμενη ενότητα. Το πρόγραμμα αυτό ονομάζεται GetHistYUV.

Το πρόγραμμα που εξετάζει αν κάποια διαφορά είναι αρκετά μεγάλη έτσι ώστε να καθορίσει ότι κάποιο frame είναι αλλαγή πλάνου ονομάζεται HistShot-Det. Το πρόγραμμα αυτό λαμβάνει ένα κατώφλι ως είσοδο και αν κάποια διαφορά το ξεπεράσει τότε το frame θεωρείται αλλαγή πλάνου.

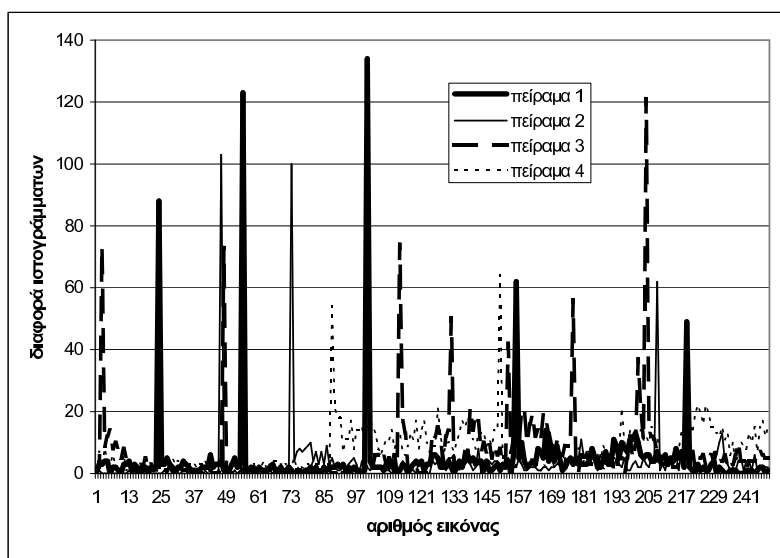
2.5.3 Παρουσίαση της μεθόδου

Οι ακολουθίες των διαφορών για τα ίδια βίντεο με αυτά του σχήματος 2.1 φαίνονται στο σχήμα 2.4. Μπορεί να γίνει κατανοητό ότι και εδώ, όπως στην περίπτωση της ενότητας 2.4.3 είναι αρκετά εύκολη η αναγνώριση των εικόνων με αλλαγή πλάνου. Για παράδειγμα, στο πείραμα 1 μπορούν εύκολα να αναγνωριστούν οι αλλαγές πλάνου που υπάρχουν στις εικόνες 25, 56, 102, 157 και 220.

Τέλος, για το παράδειγμα της εικόνας 2.5, με τη μέθοδο των ιστογραμμάτων δεν υπάρχουν πλέον τα σφάλματα λόγω της διαφορετικής ποιότητας συμπίεσης που υπήρχαν με την μέθοδο των *intra macroblock*, όπως φαίνεται από την εικόνα, όπου οι δύο αλλαγές πλάνου ξεχωρίζουν εμφανώς. Φυσικά, σε άλλες περιπτώσεις ίσως δημιουργηθούν προβλήματα, για παράδειγμα, η τρίτη αλλαγή πλάνου, στην εικόνα 195 δεν ξεχωρίζει τόσο πολύ και ανάλογα με την επιλογή του κατωφλίου απόφασης είναι πιθανό να χανόταν, ή να βρίσκονταν επιπλέον αλλαγές πλάνου στις εικόνες 220 – 230, όμως, το σημαντικό είναι ότι πρέπει να χανόταν ή να βρισκόταν και στις δύο περιπτώσεις, μιας και, όπως έχει ήδη αναφερθεί, δεν ενδιαφέρει να βρίσκονται με αντικειμενικό τρόπο οι αλλαγές πλάνου τις οποίες θα έβλεπε κάποιος άνθρωπος, αλλά να βρίσκονται οι ίδιες αλλαγές πλάνου ακόμα και αν το βίντεο τροποποιηθεί.

Συνεχίζοντας αυτή τη συζήτηση, παρότι η μέθοδος με τα ιστογράμματα φωτεινότητας λειτουργεί καλύτερα από τη μέθοδο των *intra macroblock* σε κάποιες περιπτώσεις και μπορεί να λειτουργήσει σε βίντεο αυθαίρετης μορφής, έχει μια σειρά από άλλα μειονεκτήματα:

2.6. ΜΕΤΡΗΣΗ ΕΠΙΔΟΣΗΣ ΤΕΧΝΙΚΩΝ



Σχήμα 2.4: Διαφορές ιστογραμμάτων

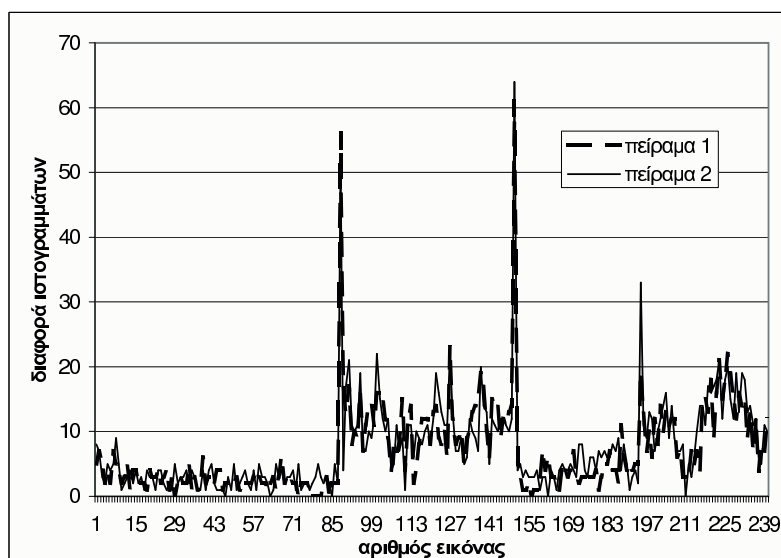
- Η εξαγωγή των ιστογραμμάτων προϋποθέτει αποσυμπίεση, πλήρη στο H.264, τουλάχιστον των συντελεστών DC στο MPEG-2, κάτι που έχει σημαντική επίπτωση στην απόδοση της μεθόδου.
- Τυχόν αλλαγές στη φωτεινότητα του βίντεο, ακόμα και δεν είναι τοπικές δηλαδή πραγματοποιηθούν σε όλες τις εικόνες αυτού, επηρεάζουν πολύ τη μορφή του ιστογράμματος φωτεινότητας και τις διαφορές αυτών. Παρόμοια προβλήματα θα υπάρχουν και αν για παράδειγμα κοπούν τα άκρα από το βίντεο ή προστεθεί κάποιο λογότυπο.
- Όπως μπορεί να φανεί και από το 2.4, ανάλογα με το βίντεο, μπορεί να υπάρξει μεγάλη διακύμανση στις τιμές των διαφορών στα σημεία με αλλαγή πλάνου.

2.6 Μέτρηση επίδοσης τεχνικών

Η επίδοση της προτεινόμενης τεχνικής ανίχνευσης αλλαγών πλάνου έχει να κάνει τόσο με το χρόνο που χρειάζεται για την εξαγωγή των πλάνων για ένα βίντεο συγκεκριμένου μήκους όσο και με το πόσο σωστά βρίσκονται οι αλλαγές πλάνου, δηλαδή το πόσο κοντά στη μονάδα είναι η ανάκληση και η ακρίβεια.

Αναφορικά με την ταχύτητα ανίχνευσης αλλαγών πλάνου, για τη μέθοδο των intra macroblock επειδή η μέθοδος που προτείνεται ουσιαστικά βασίζεται στη μέθοδο δεικτοδότησης για αναζήτηση με βάση τα intra macroblock που θα παρουσιαστεί στο επόμενο κεφάλαιο, τα αποτελέσματα σχετικά με την ταχύτητα εξαγωγής αλλαγών είναι ουσιαστικά τα ίδια με τα αποτελέσματα της ταχύτητας

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ



Σχήμα 2.5: Η περίπτωση της εικόνας 2.2

της μεθόδου αυτής που θα παρουσιαστούν στο κεφάλαιο 5. Για τη μέθοδο των ιστογραμμάτων φωτεινότητας δε θα παρουσιαστούν εδώ αποτελέσματα για την ταχύτητα της, αφού αυτά εξαρτώνται πολύ από τη μορφή του βίντεο από το οποίο θα εξαχθεί το ιστόγραμμα και αρκετά εξ αυτών έχουν ήδη παρουσιαστεί στη βιβλιογραφία ([Bes04]).

Για τον έλεγχο της επίδοσης των τεχνικών ανίχνευσης πλάνου που παρουσιάστηκαν στα προηγούμενα, αρχικά εφαρμόστηκαν μια σειρά από μετατροπές σε ένα αρχικό βίντεο και για κάθε μια εξ αυτών μετρήθηκαν μεγέθη της ανάκλησης (recall) και ακρίβειας (precision) στην ανίχνευση αλλαγών πλάνου σε σχέση με το αρχικό βίντεο. Χρησιμοποιήθηκαν βίντεο διαφόρων κατηγοριών ενώ η επίδοση κάθε τεχνικής εξετάστηκε ξεχωριστά.

Οι μετρικές που θα χρησιμοποιηθούν για τη μέτρηση της επίδοσης είναι οι ανάκληση και ακρίβεια, ελαφρώς παραλλαγμένες. Έτσι, αρχικά γίνεται η ανίχνευση των αλλαγών πλάνου στο αρχικό βίντεο και στη συνέχεια, ελέγχεται αν έχουν βρεθεί οι ίδιες αλλαγές πλάνου στις παραλλαγές του. Δηλαδή, το N_c της ενότητας 2.2 είναι ο αριθμός των αλλαγών πλάνου που ανιχνεύθηκαν στην ίδια εικόνα και στο αρχικό βίντεο και στην παραλλαγή του, το N_m είναι ο αριθμός των αλλαγών πλάνου που ενώ βρέθηκαν στο αρχικό βίντεο δε βρέθηκαν στην παραλλαγή του, ενώ το N_f είναι ο αριθμός των αλλαγών πλάνου που ανιχνεύθηκαν στην παραλλαγή του αρχικού βίντεο ενώ αυτές δεν υπήρχαν στο αρχικό βίντεο.

Δεν έγινε σύγκριση με τις πραγματικές αλλαγές πλάνου που θα ανίχνευε κάποιος άνθρωπος που βλέπει το βίντεο, αφού αυτές ουσιαστικά δεν ενδιαφέρουν την ανάλυση. Αυτό που ενδιαφέρει είναι η σύγκριση των αλλαγών πλάνου στις παραλλαγές του ίδιου βίντεο.

2.6. ΜΕΤΡΗΣΗ ΕΠΙΔΟΣΗΣ ΤΕΧΝΙΚΩΝ

Τα προγράμματα ανίχνευσης αλλαγών πλάνου (CtxtShotDet, HistShot-Det) κατέγραφαν σε ένα αρχείο το 1 για κάθε frame που ήταν αλλαγή πλάνου και το 0 αν δεν ήταν, τη μια τιμή μετά την άλλη. Στη συνέχεια, υλοποιήθηκε σε γλώσσα Perl ένα απλό πρόγραμμα το οποίο δέχεται σαν είσοδο δύο αρχεία κειμένου με αλλαγές πλάνου και υπολογίζει ανάκληση και ακρίβεια. Το πρόγραμμα αυτό απλώς ελέγχει τις ακολουθίες με 0 και 1 των δύο αρχείων και αυξάνει κατάλληλα τους μετρητές για τα N_c , N_f και N_m . Το πρόγραμμα αυτό ονομάζεται ShotCutDetPerf.

Για τις δοκιμές, επιλέχτηκαν βίντεο διαφορετικών κατηγοριών, έτσι ώστε να εξεταστεί σε όσο το δυνατόν περισσότερες περιπτώσεις η απόδοση των μεθόδων. Χρησιμοποιήθηκαν τα εξής βίντεο:

- Ένα βίντεο με αθλητικό περιεχόμενο, συγκεκριμένα, το πρώτο ημίχρονο ενός ποδοσφαιρικού αγώνα με όνομα βίντεο 1.
- Μια κινηματογραφική ταινία δράσης με όνομα βίντεο 2.
- Ένα μουσικό βίντεο κλιπ με όνομα βίντεο 3.
- Ένα ντοκιμαντέρ με όνομα βίντεο 4.

Τα χαρακτηριστικά τους φαίνονται αναλυτικά στον πίνακα 2.1.

Όνομα	Ανάλυση	Διάρκεια (σε εικόνες)	Μέγεθος (σε MB)
Βίντεο 1	576 × 432	76589	404
Βίντεο 2	704 × 384	186399	408
Βίντεο 3	640 × 272	4973	23
Βίντεο 4	640 × 352	108768	280

Πίνακας 2.1: Χαρακτηριστικά βίντεο



Σχήμα 2.6: Παράδειγμα εισαγωγής λογότυπου

Η αρχική μορφή αυτών των βίντεο είχε συμπιεστεί κατά H.264 με συντελεστή χβαντισμού $Q = 26$. Οι αλλαγές που υπέστησαν τα βίντεο είναι:

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

- Κωδικοποίηση των με διαφορετική ποιότητα. Συγκεκριμένα, έγινε κωδικοποίηση σε H.264 με συντελεστή κβαντοποίησης $Q = 30$ και 34 . Με αυτές τις αλλαγές στο Q η ποιότητα αλλάζει πολύ, κάτι που εύκολα μπορεί να φανεί από το γεγονός ότι το βίντεο με τον ποδοσφαιρικό αγώνα για $Q = 26$ ήταν 404 MB, για $Q = 30$ ήταν 228 MB και για $Q = 34$ έπεσε στα 128 MB.
- Μείωση της ανάλυσης στο μισό τόσο οριζόντια όσο και κάθετα. Το μέγεθος του ίδιου βίντεο από 404 πλέον γίνεται 122 MB.
- Προσθήκη ενός λογότυπου που να καλύπτει μεγάλο μέρος του βίντεο (παράδειγμα στην εικόνα 2.6). Το μέγεθος αλλάζει πολύ λίγο.

Τα αποτελέσματα φαίνονται στους πίνακες 2.2 και 2.3 και θα αναλυθούν στα επόμενα.

2.6.1 Επίδοση intra

Στον πίνακα 2.2 φαίνονται τα recall και precision που επιτυγχάνεται με τη μέθοδο των intra macroblock, όταν στο αρχικό βίντεο η ανίχνευση των αλλαγών πλάνου έχει γίνει με κατώφλι 0,90, δηλαδή μια εικόνα αναγνωρίζεται ως αλλαγή πλάνου όταν πάνω από το 90% των macroblock της έχουν κωδικοποιηθεί κατά intra, ενώ για τα τροποποιημένα αποσπάσματα, έχουν επιλεγεί κατώφλια ίσα με 0,75, 0,80, 0,85, 0,90 και 0,95. Τα αποτελέσματα είναι τα συνολικά και για τα 4 βίντεο του πίνακα. 2.1.

Είναι ενδιαφέρον να εξεταστεί περισσότερο η συμπεριφορά της απόδοσης των *Recall* και *Precision* όταν αλλάζει η τιμή του κατωφλιού αναγνώρισης αλλαγής πλάνου για το απόσπασμα. Στο σχήμα 2.7 έχουν σχεδιαστεί οι τιμές των *recall* και *precision* όταν αλλάζει το κατώφλι του αποσπάσματος για τις τέσσερις διαφορετικές αλλαγές. Μια γενική παρατήρηση είναι ότι το *recall* πάντοτε μειώνεται όταν αυξάνει το κατώφλι, ενώ το *precision* πάντοτε αυξάνεται. Αυτή η συμπεριφορά είναι αναμενόμενη αφού όσο αυξάνει το κατώφλι τόσο λιγότερες αλλαγές πλάνου θα αναγνωρίζονται, άρα θα είναι μεγαλύτερη η πιθανότητα να χαθεί κάποια αλλαγή πλάνου που υπάρχει στο αρχικό βίντεο (οπότε χειροτερεύει το *recall*), ενώ, αφού θα αναγνωρίζονται λιγότερες αλλαγές πλάνου θα είναι μικρότερη η πιθανότητα να βρεθεί μια αλλαγή πλάνου στο απόσπασμα που δεν υπάρχει (άρα το *precision* καλυτερεύει).

Το σημείο στο οποίο οι καμπύλες του *recall* και του *precision* τέμνονται είναι εκείνο που η συνολική επίδοση είναι η καλύτερη, δηλαδή εκείνο το σημείο στο οποίο το αρχικό βίντεο και το απόσπασμα έχουν την ελάχιστη διαφορά στον αριθμό των intra macroblock κάθε μιας εικόνας τους. Όπως φαίνεται εύκολα από το σχήμα 2.7, για $Q = 30$ καλύτερη επίδοση επιτυγχάνεται για κατώφλι ίσο με 0,85, για $Q = 34$ για κατώφλι ίσο με 0,80, για τη μισή ανάλυση για κατώφλι περίπου ίσο με 0,87, το ίδιο και για για την εισαγωγή του λογότυπου. Η συμπεριφορά για τα τη χειρότερη ποιότητα ($Q = 30$ και $Q = 34$)

2.6. ΜΕΤΡΗΣΗ ΕΠΙΔΟΣΗΣ ΤΕΧΝΙΚΩΝ

είναι αναμενόμενη, αφού, λόγω της χειρότερης ποιότητας ο αριθμός των intra macroblock θα είναι μικρότερος από το $Q = 26$, μιας και τα intra macroblock έχουν την καλύτερη ποιότητα σε σχέση με τα άλλα είδη των macroblock (με το κόστος της επιπλέον πληροφορίας για την αναπαράστασή τους). Αντίθετα, στις περιπτώσεις της μισής ανάλυσης και του λόγτύπου, η ποιότητα κωδικοποίησης παραμένει η ίδια ($Q = 26$), οπότε ο αριθμός των intra macroblock δεν αλλάζει σχεδόν καθόλου.

T	$Q = 30$		$Q = 34$		Μισή ανάλυση		Logo	
	<i>Rec</i>	<i>Prec</i>	<i>Rec</i>	<i>Prec</i>	<i>Rec</i>	<i>Prec</i>	<i>Rec</i>	<i>Prec</i>
0,75	1	0,85	0,98	0,93	1	0,84	1	0,90
0,80	1	0,92	0,97	0,96	0,99	0,89	1	0,94
0,85	0,98	0,97	0,93	0,99	0,97	0,94	1	0,97
0,90	0,93	1	0,87	1	0,94	0,97	0,97	1
0,95	0,82	1	0,72	1	0,89	0,99	0,82	1

Πίνακας 2.2: Απόδοση ανίχνευσης αλλαγών πλάνου με intra

Πρέπει εδώ να σημειωθεί ότι η επιλογή ενός κατωφλιού ίσο με 0,90 για το αρχικό βίντεο βασίστηκε κυρίως σε πειραματικά αποτελέσματα και για ένα διαφορετικό κατώφλι θα άλλαζε αρκετά η επίδοση των *recall* και *precision* του πίνακα 2.2. Επίσης, υπάρχει εξάρτηση και από το περιεχόμενο του βίντεο. Σε κάθε περίπτωση, για ένα κατώφλι 0,90 στο αρχικό βίντεο, η επιλογή ενός κατωφλιού ίσο με 0,80 ως 0,85 είναι αρκετά καλή για να επιτευχθούν πολύ καλά αποτελέσματα στην αναγνώριση αλλαγών πλάνου.

2.6.2 Επίδοση ιστογραμμάτων

Στον πίνακα 2.3 φαίνονται τα *recall* και *precision* που επιτυγχάνεται με τη μέθοδο των ιστογραμμάτων, όταν για κατώφλι διαφορών μεταξύ διαδοχικών εικόνων έχει επιλεγεί διαφορά ιστογραμμάτων ίση με 50, 75, 100 και στο αρχικό βίντεο και στο τροποποιημένο. Στην περίπτωση των ιστογραμμάτων, αν επιλεγόταν διαφορετικό κατώφλι στο αρχικό και το τροποποιημένο βίντεο η απόδοση της μεθόδου υποβαθμιζόταν πάρα πολύ, οπότε επιλέχθηκαν τα τρία αυτά κατώφλια. Δηλαδή εδώ δεν υπάρχει συσχέτιση μεταξύ του κατωφλιού αναγνώρισης αλλαγής πλάνου και των χαρακτηριστικών συμπίεσης του βίντεο όπως υπήρχε στη μέθοδο intra macroblock. Ως παράδειγμα, στο 50, αλλαγές πλάνου θεωρήθηκε κάθε εικόνα της οποίας η διαφορά του ιστογράμματος φωτεινότητας από την προηγούμενη της ξεπερνούσε το 50, τόσο στο αρχικό όσο και στο τροποποιημένο βίντεο. Τα αποτελέσματα είναι τα συνολικά και για τα 4 βίντεο του πίνακα 2.1.

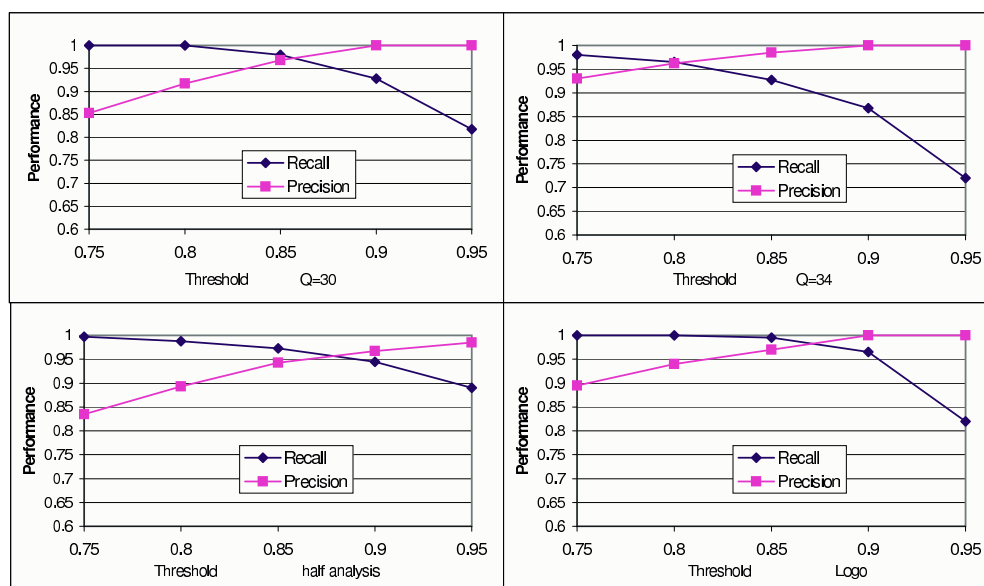
Από τον πίνακα μπορεί να φανεί ότι τα αποτελέσματα είναι πολύ κοντά στη μονάδα σε όλες τις περιπτώσεις. Δε φαίνεται να υπάρχει κάποια εξάρτηση μεταξύ του κατωφλιού και της απόδοσης, κάτι που ίσως ξενίσει αρχικά, μιας

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

	$Q = 30$		$Q = 34$		Μισή ανάλυση		Logo	
T	Rec	$Prec$	Rec	$Prec$	Rec	$Prec$	Rec	$Prec$
50	0,98	0,98	0,96	0,97	0,98	0,94	0,97	0,98
75	0,99	0,98	0,98	0,97	0,99	0,94	0,95	0,99
100	0,98	0,97	0,97	0,96	0,98	0,93	0,90	0,99

Πίνακας 2.3: Απόδοση ανίχνευσης αλλαγών πλάνου με ιστογράμματα φωτεινότητας

και στην περίπτωση του κατωφλιού 50 θα ανιχνευθούν περισσότερες αλλαγές πλάνου από ότι για κατώφλι 100, όπως μπορεί να φανεί από το σχήμα 2.4, άρα θα υπάρχει μεγαλύτερη πιθανότητα διαφορών μεταξύ του αρχικού και του τροποποιημένου βίντεο. Παρότι λοιπόν υπάρχουν περισσότερες διαφορές στο 50 από ότι στο 100, στο 50 υπάρχουν και περισσότερες αλλαγές πλάνου από ότι στο 100, έτσι, το ποσοστό των σφαλμάτων στην ανίχνευση θα είναι αντίστοιχο τόσο στο 50 όσο και στο 100.



Σχήμα 2.7: Τα Recall και Precision σε σχέση με το κατώφλι για τη μέθοδο των *intra macroblock*

Σχετικά με την επιλογή του κατωφλιού, ένα κατώφλι ίσο με 50 συνολικά παρουσιάζει τις καλύτερες επιδόσεις και επίσης, επειδή είναι μικρότερο από τα άλλα θα ανιχνεύει περισσότερες αλλαγές πλάνου, κάτι που, όπως θα φανεί από τις επόμενες ενότητες είναι σημαντικό μιας και έτσι αυξάνεται η ακρίβεια της αναζήτησης. Το σημαντικό στοιχείο εδώ είναι ότι θα πρέπει το κατώφλι πάνω από το οποίο θα πρέπει να είναι οι διαφορές των ιστογραμμάτων για να αναγνωριστεί μία αλλαγή πλάνου θα πρέπει να είναι το ίδιο και για το βίντεο

της βάσης δεδομένων και για το απόσπασμα.

2.7 Συμπεράσματα

Τα αποτελέσματα της δείχνουν ότι και οι δύο τεχνικές ανίχνευσης αλλαγής πλάνου μπορούν να επιτύχουν σχεδόν άριστα αποτελέσματα στην ανίχνευση αλλαγών πλάνου όταν κάποιο βίντεο τροποποιηθεί με διάφορους τρόπους και συγκεκριμένα, αλλάζει η ποιότητα του, η ανάλυση του και το περιεχόμενο του (με την προσθήκη ενός λογότυπου), αφού σχεδόν σε όλες τις περιπτώσεις που παρουσιάστηκαν στην ενότητα 2.6 η επίδοση ήταν πάνω από 90% τόσο στην ανάκληση όσο και στην ακρίβεια.

Η μεν μέθοδος των *intra macroblock* βασίζεται στην αναγνώριση αλλαγών πλάνου απ' ευθείας, χωρίς αποκωδικοποίηση, βίντεο που έχουν συμπίεστεί κατά H.264 με συγκεκριμένα χαρακτηριστικά, ενώ κατά τη μέθοδο των ιστογραμμάτων φωτεινότητας η συμπεριφορά εξαρτάται από το είδος του βίντεο, για παράδειγμα σε βίντεο που βασίζονται στα πρωτόκολλα MPEG μπορεί να δημιουργηθεί το DC ιστόγραμμα με τη βοήθεια των DC συντελεστών των *intra macroblock* ενώ σε βίντεο αυθαίρετης μορφής υπάρχει πάντα η λύση της πλήρους αποσυμπίεσης.

Οι επιδόσεις της μεθόδου των ιστογραμμάτων φωτεινότητας είναι ελαφρώς καλύτερες από αυτές της μεθόδου των *intra macroblock*, κάτι όμως που δεν επηρεάζει την ανάλυση των επόμενων παραγράφων, αφού θα εξεταστούν τρόποι έτσι ώστε αυτές οι μικρές διαφορές στην αναγνώριση των αλλαγών πλάνου να επηρεάζουν την αναζήτηση όσο λιγότερο γίνεται. Όπως θα φανεί στο κεφάλαιο 5 η ταχύτητα εξαγωγής του ποσοστού *intra macroblock* από κάθε εικόνα ενός H.264 είναι πολύ γρηγορότερη από την ταχύτητα πραγματικού χρόνου, και έτσι, με τη μέθοδο των *intra macroblock* επιτυγχάνεται άριστη ταχύτητα εξαγωγής αλλαγών πλάνου σε σχέση με τις υπάρχουσες μεθόδους.

ΚΕΦΑΛΑΙΟ 2. ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

Κεφάλαιο 3

Αναζήτηση βίντεο

ΜΕΤΑ την ανάλυση των μεθόδων εύρεσης αλλαγής πλάνου, το κεφάλαιο αυτό είναι αφιερωμένο στην ευθεία αναζήτηση βίντεο, δηλαδή αναζήτηση χωρίς τη χρήση λέξεων κλειδιά, με χρήση ενός αποσπάσματος βίντεο ως ερώτηση. Εδώ θα γίνει αρχικά μια αναφορά στις υπάρχουσες μεθόδους ευθείας αναζήτησης με εκτενή ανασκόπηση της βιβλιογραφίας και στα προβλήματα που υπάρχουν στην ευθεία αναζήτηση. Στη συνέχεια θα εξεταστούν λεπτομερώς οι μέθοδοι ευθείας αναζήτησης που προτείνει η εργασία.

3.1 Υπάρχουσες μέθοδοι

Παρότι υπάρχουν συστήματα ευθείας αναζήτησης βίντεο εδώ και πάνω από μια δεκαετία [Moh98], η έρευνα στο συγκεκριμένο πεδίο συνεχίζεται με αυξημένο ρυθμό λόγω των συνεχών προκλήσεων που παρουσιάζονται από τη μεγάλη αύξηση της κλίμακας των υπαρχόντων εφαρμογών καθώς και με τις νέες εφαρμογές που συνεχώς παρουσιάζονται.

Για να εξεταστεί αν κάποιο απόσπασμα βίντεο αποτελεί τμήμα ενός άλλου βίντεο που υπάρχει στη βάση δεδομένων είναι απαραίτητη αρχικά η εξαγωγή της υπογραφής τόσο του αποσπάσματος όσο και των βίντεο της βάσης δεδομένων. Στη συνέχεια, γίνεται χρήση αυτών των υπογραφών στην αναζήτηση του αποσπάσματος μεταξύ των βίντεο της βάσης δεδομένων.

Πιο συγκεκριμένα, η εξαγωγή υπογραφής για κάποιο βίντεο, που αλλιώς ονομάζεται δεικτοδότηση, δημιουργεί μια συμπαγή αναπαράσταση ενός βίντεο έτσι ώστε να διευκολύνει την αναζήτηση. Η υπογραφή ενός βίντεο περιέχει μια σειρά από χαρακτηριστικά αυτού του βίντεο τα οποία επιλέγονται από τη μέθοδο της αναζήτησης και θα χρησιμοποιηθούν στη συνέχεια για την αναζήτηση. Τέτοια χαρακτηριστικά μπορεί να είναι η ακολουθία των χρωμάτων των εικόνων του βίντεο, η κίνηση του, η μουσική του και διάφορα άλλα. Η εξαγωγή της υπογραφής είναι χρονοβόρος διαδικασία για αυτό και σχεδόν πάντα η υπογραφή εξάγεται από τα βίντεο της βάσης δεδομένων σε χρονικό διάστημα που προηγείται της αναζήτησης και δεν ενδιαφέρει.

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

Η αναζήτηση σε κάποιες περιπτώσεις μπορεί να διαχωριστεί περαιτέρω, ανάλογα με τους δείκτες που υπάρχουν στη βάση δεδομένων, σε μια αρχική επιλογή κάποιων τμημάτων των βίντεο της βάσης δεδομένων που πιθανόν ταιριάζουν με το απόσπασμα και στη συνέχεια με τη σύγκριση των τμημάτων της βάσης που επελέγησαν με το υπό αναζήτηση απόσπασμα. Το πρώτο μέρος θα ονομαστεί αναζήτηση υποψηφίων και το δεύτερο ταίριασμα αποσπασμάτων ή ακολουθίας (sequence matching), αφού τελικά συγκρίνονται δύο αποσπάσματα ίδιου μεγέθους για ομοιότητα.

Όπως μπορεί να γίνει κατανοητό, η αναζήτηση των υποψηφίων είναι μια προαιρετική διαδικασία και μπορεί να παραλειφθεί - σε μια τέτοια περίπτωση όλες οι πιθανές θέσεις ελέγχονται για ταίριασμα αποσπασμάτων. Το χαρακτηριστικό της αναζήτησης υποψηφίων είναι ότι μειώνει το χώρο αναζήτησης για το ταίριασμα αποσπασμάτων, αφού αντί για όλες τις θέσεις όλων των βίντεο της βάσης δεδομένων, θα επιλεγούν μόνο κάποιες συγκεκριμένες (άρα μειώνεται και ο χρόνος αναζήτησης). Τα σημαντικά χαρακτηριστικά της αναζήτησης υποψηφίων είναι δύο:

- Πρέπει να είναι γρηγορότερη της αναζήτησης αποσπασμάτων σε όλες τις δυνατές θέσεις, αλλιώς δεν έχει νόημα η εισαγωγή ενός επιπλέον σταδίου, αρκεί να γίνει η αναζήτηση αποσπασμάτων.
- Πρέπει να μη χανθεί το σωστό αποτέλεσμα: Η αναζήτηση υποψηφίων πρέπει να λειτουργεί με τέτοιο τρόπο έτσι ώστε να απορρίπτει όσο το δυνατόν περισσότερες θέσεις βίντεο στις οποίες δεν υπάρχει το υπό αναζήτηση απόσπασμα, όμως να μην απορριφθεί και κάποιο αποτέλεσμα το οποίο θα ταίριαζε στο ταίριασμα αποσπασμάτων. Η συγκεκριμένη απαίτηση είναι σημαντική, οπότε είναι καλύτερο να χαλαρώσουν τα κριτήρια της αναζήτησης υποψηφίων έτσι ώστε να «περνάνε» περισσότεροι υποψήφιοι (ακόμα και όλοι) παρά να είναι τόσο αυστηρά που να μην επιτρέπουν στις θέσεις που ταιριάζουν να συγκριθούν με ταίριασμα αποσπάσματος.

Οι υπάρχουσες εργασίες που αφορούν την ευθεία αναζήτηση βίντεο μπορούν να διαχωριστούν σε δύο κύριες κατηγορίες:

- Έλεγχος για ταίριασμα με χρήση εικόνων κλειδιά, όπου κάποιες εικόνες κλειδιά εξάγονται από το απόσπασμα και ελέγχεται το αν ταιριάζουν με τις εικόνες κλειδιά της βάσης δεδομένων και
- έλεγχος για ταίριασμα με βάση τα χαρακτηριστικά της ακολουθίας όλων των εικόνων του βίντεο.

Στην πρώτη κατηγορία περιλαμβάνονται εργασίες όπως η [CSL99] όπου παρουσιάζεται μια μέθοδος εξαγωγής διαβαθμισμένων εικόνων κλειδιά. Κάθε επίπεδο της διαβάθμισης περιέχει διαφορετικό αριθμό εικόνων κλειδιά, με χαρακτηριστικό ότι οι εικόνες κλειδιά των χαμηλότερων επιπέδων περιέχουν αυτές των πρώτων, δηλαδή δημιουργείται μια ιεραρχία η οποία επιτρέπει τη γρήγορη

3.1. ΥΠΑΡΧΟΥΣΕΣ ΜΕΘΟΔΟΙ

αναζήτηση, αφού συγκρίνονται οι εικόνες κλειδιά των υψηλότερων επιπέδων και αν φαίνεται ότι υπάρχει ταίριασμα τότε η σύγκριση γίνεται και με χαμηλότερα επίπεδα ούτως ώστε να ληφθεί η απόφαση. Μια παρόμοια τεχνική παρουσιάζεται στην [NKL01] όπου γίνεται μια δενδρική αναπαράσταση του βίντεο με βάση τα εξής επίπεδα: Πλάνο, εικόνα κλειδί, σκηνή, ομάδα σκηνών και συνολικό βίντεο. Η εργασία αυτή προτείνει τρόπους για ταίριασμα των δενδρικών αυτών αναπαραστάσεων.

Στην [LK07], γίνεται χρήση δύο ουρών για την εξαγωγή των εικόνων κλειδιά οι οποίες μπορούν να χρησιμοποιηθούν τόσο για τη γρήγορη ανασκόπηση του βίντεο όσο και για αναζήτηση. Η εργασία [KP02] εξάγει εικόνες κλειδιά χρησιμοποιώντας τη σωρευμένη κατευθυνόμενη απόκλιση η οποία είναι μια μετρική των διαφορών των εικόνων, οπότε όταν αυτή ξεπεράσει ένα κατώφλι τότε εξάγεται μια εικόνα κλειδί. Στη συνέχεια, αυτές οι εικόνες συγκρίνονται με τις εικόνες της βάσης χρησιμοποιώντας την κατευθυνόμενη απόκλιση τους. Η [LZP99] εξάγει τις εικόνες κλειδιά και συγκρίνει τα αποσπάσματα με βάση την οπτική τους ομοιότητα, για παράδειγμα με χρήση του χρώματος και της υψής. Εκτός από αυτόν τον τρόπο, τα βίντεο συγκρίνονται και με βάση την ομοιότητα της χρονικής σειράς των εικόνων κλειδιών, το πλήθος αυτών, αλλά και με τη διάρκεια των διαστημάτων μεταξύ των εικόνων κλειδιών. Η εργασία αυτή έχει το μειονέκτημα ότι μπορεί να συγκρίνει μόνο ολόκληρα βίντεο και όχι αποσπάσματα. Η εργασία [ZQL⁺01] βρίσκει μια εικόνα μέσα από ένα βίντεο. Χρησιμοποιούνται οι εικόνες κλειδιά του βίντεο σαν χαρακτηριστικά σημεία στο χώρο αναζήτησης και υπολογίζονται «γραμμές» οι οποίες ενώνουν αυτά τα σημεία. Για την αναζήτηση βρίσκεται τελικά η απόσταση της εικόνας από αυτές τις γραμμές.

Αναφορικά με τη δεύτερη κατηγορία των εργασιών ευθείας αναζήτησης βίντεο, η [KP02] εξάγει το ιστόγραμμα χαρακτηριστικών τόσο από το απόσπασμα όσο και από ένα τμήμα ίδιου μήκους από τη βάση δεδομένων και τα συγκρίνει. Στη συνέχεια το παράθυρο αναζήτησης προχωρά στο επόμενο τμήμα της βάσης. Η τεχνική αυτή μπορεί να χρησιμοποιηθεί για ευθεία αναζήτηση τόσο βίντεο όσο και ήχου. Μπορούν να χρησιμοποιηθούν διάφορων ειδών χαρακτηριστικά για το ιστόγραμμα, για παράδειγμα η κατανομή των χρωμάτων. Στην εργασία [NYY00] εξάγεται το ιστόγραμμα από κάθε εικόνα και η μορφή του προσεγγίζεται με τη βοήθεια πολυωνύμων. Το ιστόγραμμα χρησιμοποιεί 32 τιμές (κουβάδες) για την συνιστώσα Y και από 16 για τις χρωματικές. Η πολυωνυμική προσέγγιση αυτή χρησιμοποιείται τελικά για τη σύγκριση μεταξύ του αποσπάσματος και της βάσης δεδομένων. Ομοίως, η [YTR04] εξάγονται οι μέσες τιμές των χρωμάτων, μόνο όμως των εικόνων I , για ταχύτητα και δημιουργείται μια ακολουθία η οποία και χρησιμοποιείται για την αναζήτηση. Στη συγκεκριμένη εργασία χρησιμοποιείται και ο ήχος για ενίσχυση των αποτελεσμάτων. Στην KIM2005 προτείνεται μια μέθοδος που είναι εύρωστη τόσο σε αλλαγές του βίντεο λόγω της συμπίεσης όσο και σε αλλαγές λόγω διαφορετικής ανάλυσης εικόνας. Προς τούτο, οι εικόνες χωρίζονται σε N τμήματα στα οποία ανατίθεται ένας αριθμός από 0 ως $N - 1$, σειριακά, με βάση τη φωτει-

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

νότητα κάθε τμήματος. Στη συνέχεια, η σύγκριση μεταξύ δύο αποσπασμάτων γίνεται τόσο χωρικά, δηλαδή με τη διαφορά κάθε μιας εικόνας όσο όμως και χρονικά δηλαδή ελέγχεται το αν τα επιμέρους τμήματα των αντίστοιχων εικόνων των δύο βίντεο αλλάζουν τιμές με τον ίδιο τρόπο, για παράδειγμα αν το πάνω αριστερό μέρος και των δύο βίντεο έχει την τιμή 2 στην εικόνα i και την τιμή 1 στο ένα βίντεο και την τιμή 3 στο άλλο, τότε σημαίνει ότι η εικόνας δεν αλλάζουν με τον ίδιο τρόπο άρα εκτός από τη χωρική διαφορά θα προστεθεί και η χρονική διαφορά. Στην [CSI06] βρίσκονται βίντεο τα οποία περιέχουν το ίδιο πλάνο, αλλά από διαφορετικές κάμερες, για παράδειγμα με διαφορετικό ζουμ. Αυτό γίνεται εξετάζοντας την κίνηση των διαφόρων αντικειμένων κάθε πλάνου. Στην εργασία [HWL03] δημιουργούνται δύο υπογραφές κάθε βίντεο: Μια λιγότερο λεπτομερής σε έναν χώρο χαρακτηριστικών λίγων διαστάσεων με την οποία γίνεται μια πρώτη διαλογή των βίντεο και μια περισσότερο λεπτομερής με περισσότερες διαστάσεις στο χώρο χαρακτηριστικών με την οποία βρίσκεται το τελικό αποτέλεσμα. Με αυτό τον τρόπο η αναζήτηση πραγματοποιείται γρηγορότερα.

Κάποιες εργασίες μπορεί να περιέχουν τεχνικές που ανήκουν τόσο στην πρώτη όσο και στη δεύτερη κατηγορία όπως η [JWV99]. Σε αυτήν προτείνονται δύο τεχνικές: Μια η οποία εξάγει τις εικόνες κλειδιά και αναζητά το απόσπασμα με βάση αυτές, και μια η οποία εξάγει χαρακτηριστικά για το χρώμα ή την υφή από όλες τις εικόνες (στις οποίες μειώνεται η ανάλυση) και ψάχνει με αυτή την ακολουθία. Η εργασία [cSCZ03] υπολογίζει μια μετρική ομοιότητας δύο βίντεο με όνομα IVS - Ideal Video Similarity. Για τον υπολογισμό της μετρικής αυτής, βρίσκεται το ποσοστό των συστάδων των δύο αποσπασμάτων που περιέχουν εικόνες και από τα δύο αποσπάσματα. Μια συστάδα ορίζεται στην εργασία ως μια ακολουθία από «παρόμοιες» εικόνες σε ένα βίντεο, για παράδειγμα οι εικόνες ενός πλάνου. Ο υπολογισμός της μετρικής IVS είναι αρκετά δύσκολος μιας και πρέπει να βρεθεί η απόσταση μεταξύ όλων των εικόνων έτσι ώστε να δημιουργηθούν οι συστάδες για αυτό και στη συνέχεια της εργασίας προτείνεται η προσέγγιση του IVS μέσω του VVS - Voronoi Video Similarity έτσι ώστε να μειωθεί το κόστος υπολογισμού του IVS. Το VVS μειώνει τον αριθμό των εικόνων που θα χρησιμοποιηθούν για τις συστάδες με το να επιλέγει συγκεκριμένες εικόνες από τα δύο αποσπάσματα με βάση κάποιον κοινό για όλα τα βίντεο τυχαίο αλγόριθμο. Έτσι, η εργασία αυτή κατατάχθηκε σε αυτές που χρησιμοποιούν τεχνικές και από τις δύο κατηγορίες. Συνέχεια της προηγούμενης αποτελεί η εργασία [cSCZ05] στην οποία προτείνονται τεχνικές για γρήγορη αναζήτηση αλλά και για ομαδοποίηση των αποτελεσμάτων της αναζήτησης με βάση την ομοιότητα με χρήση της προηγούμενης μεθόδου. Για τη γρήγορη αναζήτηση χρησιμοποιείται μια βελτίωση της μεθόδου Triangle-Inequality Pruning της [BS99], ενώ για την ομαδοποίηση, κάθε αποτέλεσμα θεωρείται ως ο κόμβος ενός γράφου, και αν η απόσταση μεταξύ δύο αποτελεσμάτων είναι μικρότερη από ένα κατώφλι, θα δημιουργείται ακμή μεταξύ των κόμβων. Χρησιμοποιούνται στη συνέχεια διάφοροι αλγόριθμοι για τη δημιουργία ομάδων, ανάλογα με τη συνεκτικότητα των τμημάτων και για διάφορα κατώφλια.

3.1. ΥΠΑΡΧΟΥΣΕΣ ΜΕΘΟΔΟΙ

Αρκετά σημαντική είναι η [HZ06] η οποία χρησιμοποιεί δυναμικό προγραμματισμό για την εύρεση μικρών τμημάτων βίντεο μέσα σε μεγαλύτερα ακόμα και αν αυτά δεν είναι συνεχόμενα. Επίσης, προτείνονται τέσσερα χαρακτηριστικά για χρήση ως υπογραφή του βίντεο τα οποία δεν είναι ευαίσθητα στις συνήθεις αλλαγές που υφίσταται το βίντεο κατά την επεξεργασία του. Οι τέσσερις τεχνικές που προτείνονται είναι: η διάρκεια των πλάνων του βίντεο, οι αλλαγές του χρωματικού περιεχομένου ως προς το χρόνο, οι αλλαγές της κίνησης ως προς το χρόνο και ο συνδυασμός των δύο τελευταίων. Τέλος περιγράφεται και μια τεχνική για αναζήτηση σε αυτές τις υπογραφές, η οποία βασίζεται σε προσέγγιση αυτών με κείμενο. Αναφορικά με τη διάρκεια των πλάνων, χρησιμοποιείται μια μέθοδος βασισμένη σε ιστόγραμμα των YUV συνιστωσών για την αναγνώριση των αλλαγών πλάνου, και κατόπιν οι διάρκειες αυτών χρησιμοποιούνται για τη δημιουργία της υπογραφής. Υπάρχει πρόβλεψη για το πρόβλημα του jitter αλλά όχι για τυχόν χαμένες ή εσφαλμένες αλλαγές πλάνου που μπορεί να υπάρξουν. Η υπογραφή των χρωμάτων απλώς δημιουργεί ιστόγραμμα με τις τρεις συνιστώσες και ορίζει μια μετρική για την απόσταση των ιστογραμμμάτων. Για την κίνηση, βρίσκεται το κέντρο των πιο φωτεινών και πιο σκοτεινών σημείων κάθε εικόνας και υπολογίζεται το πώς έχουν αυτά μετακινηθεί σε σχέση με την προηγούμενη εικόνα, δημιουργώντας έτσι μια υπογραφή. Τέλος, προτείνεται και ο συνδυασμός των δύο παραπάνω αφού χρησιμοποιούν διαφορετικές ιδιότητες του βίντεο, έτσι ώστε κάθε εικόνα να χαρακτηρίζεται και από τις δύο ιδιότητες. Για την αναζήτηση, οι υπογραφές που περιγράφηκαν μετατρέπονται σε ακολουθίες από ακεραίους και τελικά χρησιμοποιείται δυναμικός προγραμματισμός έτσι ώστε να γίνει η ευθυγράμμιση του αποσπάσματος με το βίντεο της βάσης δεδομένων.

Η εργασία [HB01] παρουσιάζει μια σύγκριση των μεθόδων που μετράνε τη διαφορά μεταξύ δύο βίντεο. Οι μέθοδοι που συγκρίνονται είναι:

- Η διαφορά των εικόνων όπου είναι πολύ απλή και απλώς μετράται η διαφορά των εικόνων κάθε βίντεο. Για παράδειγμα, αν χρησιμοποιηθεί μόνο η Y συνιστώσα, η διαφορά δύο βίντεο μεγέθους L εικόνων θα δίνεται από τον τύπο

$$D = \frac{\sum_{i=1}^L \sum_{x=1}^W \sum_{y=1}^H Y_i(x, y)}{L \times W \times H}.$$

- Η διαφορά των ιστογραμμμάτων, όπου συγκρίνονται τα ιστογράμματα των δύο βίντεο, με τρόπο αντίστοιχο με αυτόν που παρουσιάστηκε ήδη στην ενότητα 2.5.
- Η απόσταση Hausdorff [HKR93]. Έστω δύο πεπερασμένα σύνολα σημείων $A = a_1, a_2, \dots, a_p$ και $B = b_1, b_2, \dots, b_q$. Η απόσταση Hausdorff μεταξύ τους ορίζεται ως

$$H(A, B) = \max(h(A, B), h(B, A)),$$

όπου

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|.$$

Η $\|\cdot\|$ είναι μια νόρμα στα σημεία των A και B . Διαισθητικά, αν $h(A, B) = d$ τότε κάθε σημείο του συνόλου A απέχει το πολύ κατά d από κάποιο σημείο του B και υπάρχει κάποιο σημείο του A που απέχει ακριβώς κατά d από το κοντινότερο σημείο του B . Η $H(A, B)$ μετράει τη διαφορά των δύο συνόλων μετρώντας την απόσταση του σημείου του A που διαφέρει περισσότερο από κάθε σημείο του B και το αντίστροφο.

- Εξαγωγή των ακμών της εικόνας και στη συνέχεια διαχωρισμός της εικόνας σε $n_1 \times n_2$ παράθυρα. Το κάθε παράθυρο έχει μια τιμή που είναι το κεντρικό σημείο των ακμών του.
- Χρήση αμετάβλητων ροπών όπως προτείνονται στην [Hu62]. Χρησιμοποιούνται οι κεντρικές ροπές δεύτερης τάξης για τη μέτρηση της απόστασης.

Οι ίδιοι συγγραφείς, στην [HHB02] παρουσιάζουν μια σύγκριση διάφορων τεχνικών για εξαγωγή υπογραφής από βίντεο οι οποίες στη συνέχεια μπορούν να χρησιμοποιηθούν για να ελεγχθεί το αν τα βίντεο ταιριάζουν. Οι τρεις μέθοδοι που παρουσιάζονται είναι:

- Κατεύθυνση κίνησης: Οι εικόνες χωρίζονται σε κομμάτια και για κάθε ένα εξ αυτών βρίσκεται ο μέσος όρος των διανυσμάτων κίνησης και κβαντίζεται σε Q καθορισμένες κατευθύνσεις. Η υπογραφή κάθε εικόνας είναι ο αριθμός των διανυσμάτων κίνησης που περιέχεται σε κάθε μία από τις προκαθορισμένες κατευθύνσεις.
- Ένταση εικόνων: Η εικόνες χωρίζονται σε N κομμάτια και για κάθε ένα υπολογίζεται ο μέσος όρος του γκριζου. Στα κομμάτια αυτά αντιστοιχίζονται αριθμοί από το 0 ως το $N - 1$, ανάλογα πόσο μεγάλη είναι ένταση του γκριζου σε κάθε μία. Η μέθοδος αυτή παρουσιάστηκε αρχικά στο [BN98] ενώ μια παραλλαγή της χρησιμοποιήθηκε και από το [Moh98]
- Ιστόγραμμα χρωμάτων: Χρησιμοποιεί το ιστόγραμμα χρωμάτων κάθε εικόνας για να δημιουργήσει την υπογραφή του βίντεο. Προτάθηκε αρχικά από το [NYY00].

3.2 Προβλήματα της αναζήτησης

Όπως μπορεί να γίνει κατανοητό από τα παραπάνω, κατά την ευθεία αναζήτηση βίντεο προκύπτουν κάποια δύσκολα προβλήματα:

- Ταχύτητα εξαγωγής χαρακτηριστικών: Επειδή εξάγονται κάποια αρκετά πολύπλοκα χαρακτηριστικά από κάθε εικόνα, όπως για παράδειγμα το

3.2. ΠΡΟΒΛΗΜΑΤΑ ΤΗΣ ΑΝΑΖΗΤΗΣΗΣ

ιστόγραμμα των χρωμάτων, η διαδικασία της δεικτοδότησης ενός βίντεο μπορεί να χρειάζεται πολύ χρόνο. Φυσικά η δεικτοδότηση της βάσης δεδομένων υποτίθεται ότι έχει γίνει από πριν και για κάθε αναζήτηση χρειάζεται μόνο η δεικτοδότηση του αποσπάσματος. Όμως, πρέπει να σημειωθεί ότι, με βάση τα στοιχεία που υπάρχουν για την ποσότητα βίντεο που υπάρχει στο WWW αν η δεικτοδότηση είναι πολύ αργή τότε θα χρειαζόταν πάρα πολύ χρόνο (ή πάρα πολύ υπολογιστική ισχύ) σε σημείο που να είναι ανέφικτη. Ένας τρόπος για να αυξηθεί πολύ η ταχύτητα δεικτοδότησης είναι η χρήση συγκεκριμένων χαρακτηριστικών τα οποία μπορούν να εξαχθούν χωρίς να χρειάζεται η αποσυμπίεση, η έστω όταν δε χρειάζεται η πλήρης αποσυμπίεση του βίντεο, για παράδειγμα κάποια από τα χαρακτηριστικά που πραγματεύεται η [WDV⁺03]. Στη συγκεκριμένη εργασία, υπάρχει ένας σημαντικός πίνακας ο οποίος αναφέρει τα διάφορα χαρακτηριστικά τα οποία μπορούν να εξαχθούν χωρίς πλήρη αποσυμπίεση του κατά MPEG συμπίεσμένου βίντεο, μέρος του οποίου αναπαράγεται στον πίνακα της ανά χείρας εργασίας. Οι όροι αποδοτικότητα και αποτελεσματικότητα του πίνακα έχουν να κάνουν με το πόσο γρήγορη είναι η εξαγωγή του χαρακτηριστικού και πόσο καλή είναι η δεικτοδότηση που γίνεται με αυτό αντίστοιχα.

Στο πρωτόκολλο H.264 όπως ήδη αναφέρθηκε και στην ενότητα 2.3, είναι δύσκολη ως αδύνατη η εξαγωγή των περισσότερων χαρακτηριστικών του πίνακα 3.1 χωρίς να πρέπει να γίνει πλήρης αποσυμπίεση.

Ένας άλλος τομέας όπου η ταχύτητα εξαγωγής των χαρακτηριστικών είναι σημαντική είναι τα συστήματα που παρακολουθούν συνεχώς κάποιους σταθμούς εκπομπής βίντεο με σκοπό να καταγράψουν διάφορες πληροφορίες. Σε αυτά, πρέπει η δεικτοδότηση να μπορεί να γίνεται γρηγορότερα από ότι σε πραγματικό χρόνο, αφού χρειάζεται κάποιος χρόνος και για την αναζήτηση.

- Ταχύτητα αναζήτησης: Οι περισσότερες από τις παραπάνω τεχνικές, χρησιμοποιούν απευθείας ταίριασμα αποσπασμάτων, χωρίς πρώτα να κάνουν κάποια επιλογή των υποψηφίων στους οποίους τελικά θα γίνει το ταίριασμα. Η ύπαρξη πολύπλοκων χαρακτηριστικών σε συνάρτηση με την ποσότητα του βίντεο απαιτεί την όσο γίνεται πιο γρήγορη σύγκριση των χαρακτηριστικών για κάθε εικόνα. Ομοίως και εδώ, για τα συστήματα πραγματικού χρόνου πρέπει και η αναζήτηση να γίνεται γρηγορότερα από τον πραγματικό χρόνο αφού συνήθως η αναζήτηση και η δεικτοδότηση θα γίνονται από το ίδιο σύστημα, ενώ επίσης, ένα σύστημα μπορεί να παρακολουθεί περισσότερες από μια πηγές.
- «Αλλαγμένα» βίντεο: Πολύ σπάνια στην αναζήτηση βίντεο το απόσπασμα περιέχεται ακριβώς το ίδιο στη βάση. Μπορεί να υπάρχουν διαφορές στη φωτεινότητα, την ανάλυση, το fps, την ποιότητα συμπίεσης, να υπάρχει πρόσθετο κείμενο, πρόσθετα λογότυπα, κομμένα αλλά και πρόσθετα κομ-

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

μάτια τόσο χρονικά όσο και χωρικά. Υπάρχουν πολλά πρότυπα εκπομπής, όπως PAL, NTSC, SECAM, τα οποία έχουν διαφορετική ανάλυση αλλά και διαφορετικό frame rate. Επίσης, τα βίντεο μπορεί να είναι σε διαφορετικούς χρωματικούς χώρους πχ RGB, YUV, HSV, οπότε κατά τη μετατροπή υφίσταται αλλαγές. Η ποιότητα συμπίεσης μπορεί όχι μόνο να υποβαθμίσει την ποιότητα του βίντεο αλλά ακόμα και να εισάγει artifacts στο βίντεο.

Οι μέθοδοι ευθείας αναζήτησης θα πρέπει να είναι εύρωστες (robust) ούτως ώστε να βρίσκουν τα αλλαγμένα κομμάτια αλλά όχι τόσο ώστε να μην είναι πλέον αποδοτικές (efficient) δηλαδή εκτός από τη σωστή απάντηση να επιστρέφουν όσο γίνεται λιγότερα λανθασμένα αποτελέσματα.

Χαρακτηριστικό	αποδοτικότητα	αποτελεσματικότητα
DC ιστόγραμμα χρωμάτων	καλή	καλή
DC πίνακας YCbCr	καλή	καλή
Υπολογισμός υφής με DCT AC συντελεστές	καλή	μέτρια
Υπολογισμός υφής με χάρτη DCT άκρων	μέτρια	καλή
Τριδιάστατη κίνηση κάμερας με 6 παραμέτρους	κακή	καλή
Κίνηση κάμερας με 3 παραμέτρους	μέτρια	καλή
Δραστηριότητα κίνησης	καλή	μέτρια
Ιστόγραμμα κίνησης	καλή	μέτρια
Τύπος macroblock	καλή	μέτρια

Πίνακας 3.1: Χαρακτηριστικά που δε χρειάζονται αποσυμπίεση ([WDV⁺03])

3.3 Μετρικές απόδοσης αναζήτησης βίντεο

Για τη μέτρηση της απόδοσης των συστημάτων αναζήτησης βίντεο, ορίζονται οι έννοιες της ανάκλησης και ακρίβειας αναζήτησης βίντεο σε αντιστοιχεία με τους ορισμούς της 2.2 για την ανάκληση και ακρίβεια εύρεσης αλλαγών πλάνου. Έστω ένα σύνολο από αποσπάσματα τα οποία προέρχονται από ένα σύνολο βίντεο που ανήκουν σε μια βάση δεδομένων. Για κάθε ένα από τα αποσπάσματα γίνεται αναζήτηση, και για το σύνολο των αναζητήσεων ορίζονται τα μεγέθη:

Ορισμός 3.3.1 Ως Ανάκληση (recall) στην αναζήτηση βίντεο ορίζεται το μέγεθος

$$R = \frac{N_c}{N_c + N_m},$$

3.4. ΟΙ ΛΥΣΕΙΣ ΠΟΥ ΠΡΟΤΕΙΝΟΝΤΑΙ

όπου N_c είναι ο αριθμός των αποσπασμάτων που βρέθηκαν στο σωστό σημείο και N_m ο αριθμός των αποσπασμάτων που δε βρέθηκαν (*missed detects*). Όσο πιο κοντά στη μονάδα είναι η ανάκληση τόσο πιο σίγουρο είναι ότι με μια αναζήτηση θα βρεθεί, εκτός των άλλων, και η σωστή απάντηση.

Ορισμός 3.3.2 Ακρίβεια (precision) στην αναζήτηση βίντεο ορίζεται το μέγεθος

$$P = \frac{N_c}{N_c + N_f},$$

όπου N_c είναι ο αριθμός των αποσπασμάτων που βρέθηκαν στο σωστό σημείο και N_f είναι ο αριθμός των αποσπασμάτων που βρέθηκαν κάπου, αλλά όχι στο σωστό σημείο (*false alarms*). Όσο πιο κοντά στη μονάδα είναι η ανάκληση τόσο λιγότερες λάθος απαντήσεις θα επιστραφούν με κάθε αναζήτηση.

Μπορεί να γίνει κατανοητό ότι το μέγεθος της ανάκλησης συσχετίζεται με την ευρωστία της αναζήτησης, αφού και τα δύο έχουν να κάνουν με τη σωστή εύρεση των αποσπασμάτων ακόμα και αν αυτά έχουν υποστεί αλλαγές, ενώ το μέγεθος της ακρίβειας έχει να κάνει με την αποδοτικότητα της αναζήτησης, αφού εξαρτώνται από το πόσα αποτελέσματα επιστρέφονται χωρίς να είναι σωστά.

Τα μεγέθη της ανάκλησης και της ακρίβειας είναι συνήθως αντιστρόφως ανάλογα αφού όσο πιο πολύ χαλαρώνουν τα διάφορα κριτήρια αναζήτησης τόσο πιο σίγουρο είναι ότι η σωστή απάντηση θα βρεθεί, αλλά και επίσης, αφού έχουν χαλαρώσει τα κριτήρια, μαζί με τη σωστή απάντηση θα επιστραφούν και μια σειρά από λάθος απαντήσεις. Αντίστροφα, όσο τα κριτήρια γίνονται πιο αυστηρά, τόσο πιο λίγες λάθος απαντήσεις θα επιστραφούν, αλλά θα αυξάνεται η πιθανότητα να κοπεί και η σωστή απάντηση.

Έτσι λοιπόν, για την αναζήτηση υποψηφίων, είναι σημαντικό η ανάκληση να είναι όσο γίνεται πιο κοντά στη μονάδα, έτσι ώστε να μη χαθεί η σωστή απάντηση. Η επιλογή για αυξημένη ανάκληση είναι σημαντική, έστω και αν με αυτή την επιλογή η ακρίβεια της μεθόδου μειωθεί πολύ αφού οι επιπλέον εσφαλμένες απαντήσεις θα κοπούν στο ταίριασμα αποσπασμάτων και τελικά το μόνο κόστος θα είναι η εξέταση περισσότερων υποψηφίων κατά το ταίριασμα.

3.4 Οι λύσεις που προτείνονται

Στην ανά χείρας εργασία προτείνονται τρεις τεχνικές αναζήτησης βίντεο, οι οποίες έχουν ως κύριο χαρακτηριστικό τους την ταχύτητα τόσο στη δεικτοδότηση, αλλά κυρίως στην αναζήτηση. Επίσης, έγινε προσπάθεια έτσι ώστε οι τεχνικές να είναι όσο το δυνατόν περισσότερο ευέλικτες στις διάφορες αλλαγές που μπορεί να υποστεί ένα βίντεο, δηλαδή αλλαγή ανάλυσης, αλλαγή ποιότητας, εισαγωγή υποτίτλων, logos κοκ. Έτσι, οι τεχνικές αυτές βοηθούν στην επιτυχή επίλυση των προβλημάτων αναζήτησης που παρουσιάστηκαν στην ενότητα

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

3.2. Οι δύο από τις τεχνικές αυτές αφορούν την αναζήτηση υποψηφίων, αφού, όπως φάνηκε και από την ανασκόπηση της βιβλιογραφίας, πολύ λίγες εργασίες πραγματοποιούν μία γρήγορη επιλογή υποψηφίων έτσι ώστε να μειώσουν το χώρο αναζήτησης, ενώ η τρίτη είναι συνδυασμός και της αναζήτησης υποψηφίων και του ταιριάσματος αποσπασμάτων. Αφού μπορούν να χρησιμοποιηθούν σαν αναζήτηση υποψηφίων, είναι σημαντικόι οι μέθοδοι που θα περιγραφούν να έχουν όσο γίνεται καλύτερη επίδοση στην ανάληψη, ακόμα και με το κόστος χαμηλότερης επίδοσης στην ακρίβεια, κάτι που είναι δυνατό όπως θα φανεί τόσο στην παρουσίαση των μεθόδων στις επόμενες ενότητες όσο και στα πειραματικά αποτελέσματα του κεφαλαίου 5.

Οι τεχνικές αναζήτησης που παρουσιάζονται είναι οι ακόλουθες:

- Αναζήτηση πιθανοτήτων [SP09, DPS⁺08] , στην οποία με τη βοήθεια της θεωρίας πιθανοτήτων περιγράφεται ένα θεωρητικό μοντέλο για την αναζήτηση το οποίο μπορεί να χρησιμοποιηθεί για την εξαγωγή διαφόρων συμπερασμάτων χρήσιμων για την πραγματοποίηση αναζήτησης με απλό τρόπο. Κύριο στοιχείο της μεθόδου είναι κάποια τυχαία συμβάντα με συγκεκριμένα χαρακτηριστικά που μπορεί να παρουσιαστούν στο βίντεο - ένα τέτοιο συμβάν είναι η αλλαγή πλάνου. Η μέθοδος αυτή αφορά την αναζήτηση υποψηφίων, αφού τα αποτελέσματα της επιστρέφουν ένα σύνολο με πιθανές υποψήφιες θέσεις, οι οποίες όμως θα πρέπει να εξεταστούν με κάποια μέθοδο ταιριάσματος έτσι ώστε να επιλεγούν τα πραγματικά αποτελέσματα.
- Μέθοδος βάσης δεδομένων [ATAP08, SPA08, PAM⁺08, APMT08] , στην οποία εξετάζονται οι ακολουθίες των αλλαγών πλάνου του αποσπάσματος και των βίντεο της βάσης δεδομένων για την αναζήτηση των υποψηφίων και ακολούθως συγκρίνονται μια σειρά από χαρακτηριστικά των εικόνων αλλαγής πλάνου του αποσπάσματος και των υποψηφίων που βρέθηκαν. Και η μέθοδος αυτή είναι τεχνική αναζήτησης υποψηφίων. Το κύριο χαρακτηριστικό της μεθόδου αυτής είναι το ότι κατά την αναζήτηση δημιουργούνται αυτόματα μια σειρά από ερωτήματα προς μια σχεσιακή βάση δεδομένων (SQL queries) τα οποία και χρησιμοποιούνται για την αναζήτηση. Σημειώνεται εδώ, ότι αν το απόσπασμα είναι αρκετά μεγάλο τότε και η μέθοδος της βάσης δεδομένων και της αναζήτησης των πιθανοτήτων μπορεί να επιστρέψουν τόσο λίγα αποτελέσματα έτσι ώστε να μην έχει νόημα η πραγματοποίηση ταιριάσματος ακολουθίας στους υποψηφίους.
- Μέθοδος intra macroblock [DSAP], στην οποία εξετάζεται η ακολουθία των ποσοστών των macroblock που έχουν κωδικοποιηθεί κατά intra για κάθε εικόνα ενός κατάλληλα συμπιεσμένου H.264 βίντεο. Προτείνονται μια σειρά από επιμέρους μεθοδολογίες για την εξέταση, κάποιες εκ των οποίων είναι για ταίριασμα ακολουθιών χωρίς αναζήτηση υποψηφίων, και κάποιες άλλες για αναζήτηση υποψηφίων και ταίριασμα ακολουθιών.

3.5. ΛΑΘΗ ΚΑΤΑ ΤΗΝ ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

Από τα παραπάνω, φαίνεται ότι η αλλαγή πλάνου είναι πολύ σημαντική στις μεθόδους των πιθανοτήτων και της βάσης δεδομένων. Επίσης, οι αλλαγές πλάνου σε ένα H.264 έχουν κάποια πολύ συγκεκριμένα χαρακτηριστικά, που τις κάνουν σημαντικές και σε κάποιες υποπεριπτώσεις της μεθόδου των intra macroblock.

Ήδη από το κεφάλαιο 2 αναλύθηκαν κάποιες τεχνικές ανίχνευσης αλλαγής πλάνου. Εκεί, ορίστηκε ως σημαντικότερο χαρακτηριστικό όχι η αντικειμενική ανίχνευση των αλλαγών πλάνου, αλλά η ανίχνευση των ίδιων αλλαγών τόσο στο αρχικό βίντεο όσο και σε τυχόν παραλλαγές του. Στην επόμενη ενότητα, θα εξεταστούν πιο αναλυτικά τα προβλήματα που παρουσιάζονται κατά τη συγκριτική ανίχνευση αλλαγών πλάνου σε ένα αρχικό βίντεο και στις παραλλαγές αυτού.

3.5 Λάθη κατά την ανίχνευση αλλαγών πλάνου

Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο κάποιες αλλαγές πλάνου μπορεί να βρεθούν στο αρχικό βίντεο και όχι κάποιο απόσπασμα του που έχει υποστεί επεξεργασία και το αντίθετο. Επίσης, ένα άλλο πρόβλημα που δημιουργείται είναι, λόγω του διαφορετικού χρονισμού που υπάρχει στα δύο βίντεο, κάποιες αλλαγές πλάνου να εμφανιστούν λίγο νωρίτερα στο ένα βίντεο από ότι στο άλλο. Μπορεί να εμφανιστούν δηλαδή τα εξής τρία προβλήματα:

Ορισμός 3.5.1 Εσφαλμένη αναγνώριση αλλαγής πλάνου (false alarm) υπάρχει όταν μια αλλαγή πλάνου αναγνωριστεί στο απόσπασμα ενώ δεν έχει αναγνωριστεί στο αρχικό βίντεο.

Ορισμός 3.5.2 Χαμένη αναγνώριση αλλαγής πλάνου (missed detect) υπάρχει όταν μια αλλαγή πλάνου δεν αναγνωριστεί στο απόσπασμα ενώ έχει αναγνωριστεί στο αρχικό βίντεο.

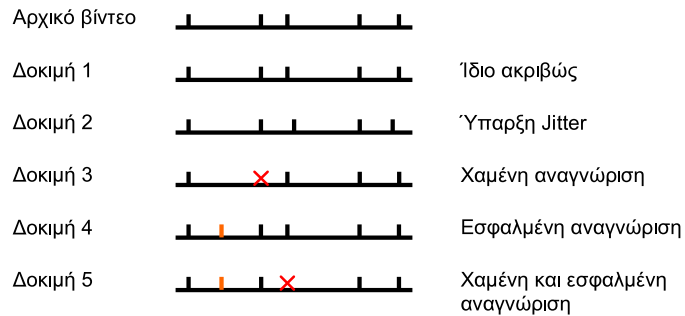
Ορισμός 3.5.3 Jitter αλλαγής πλάνου υπάρχει όταν μια αλλαγή πλάνου εμφανιστεί στο απόσπασμα λίγο νωρίτερα ή αργότερα από ότι στο αρχικό βίντεο.

Στο σχήμα 3.1 υπάρχουν οι αλλαγές πλάνου σε ένα αρχικό βίντεο καθώς και οι περιπτώσεις που μπορούν να υπάρξουν στο απόσπασμα: να υπάρχει ακριβές ταίριασμα, να υπάρχει jitter, χαμένες αναγνωρίσεις και εσφαλμένες αναγνωρίσεις.

3.5.1 Χαμένες και εσφαλμένες αναγνωρίσεις αλλαγών πλάνου

Όπως μπορεί να φανεί στη βιβλιογραφία ([GKS00]) ανάλογα με την ποιότητα κωδικοποίησης οι τιμές ανάκλησης και ακρίβειας μεταβάλλονται ραγδαία. Στο ίδιο συμπέρασμα μπορεί να καταλήξει κανείς και από το παράδειγμα της ενότητας 2.4.3. Δηλαδή, όταν το απόσπασμα υποστεί κάποια επεξεργασία τότε, μετά την

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ



Σχήμα 3.1: Είδη σφαλμάτων κατά την αναγνώριση

επεξεργασία κάποιες αλλαγές πλάνου είτε θα βρεθούν στο απόσπασμα χωρίς να υπάρχουν στο αρχικό βίντεο, είτε δε θα βρεθούν στο απόσπασμα ενώ υπάρχουν στο αρχικό βίντεο.

Επίσης, τα συμπεράσματα αυτά φαίνονται στις επιδόσεις των τεχνικών ανίχνευσης αλλαγής πλάνου που παρουσιάστηκαν στην ενότητα 2.7, αφού σε καμία περίπτωση τα recall και precision δεν είχαν και τα δύο την τιμή 1, και ακόμα και αν ίσχυε κάτι τέτοιο, θα μπορούσαν να βρεθούν κάποια βίντεο στα οποία αυτό να μην ήταν δυνατό.

3.5.2 Ανάλυση Jitter

Στις τηλεπικοινωνίες, ο όρος jitter¹ χρησιμοποιείται για να περιγράψει τη μεταβολή της καθυστέρησης λήψης ενός σήματος. Στην εργασία αυτή, χρησιμοποιείται ο ίδιος όρος για να περιγράψει το φαινόμενο της αναγνώρισης αλλαγών πλάνου σε γειτονικές εικόνες μεταξύ ενός αρχικού βίντεο και του ίδιου βίντεο μετά από επεξεργασία. Δηλαδή, αν στο αρχικό βίντεο αναγνωριστούν αλλαγές πλάνου στην εικόνα i και στην εικόνα j , ενώ αν το ίδιο βίντεο υποστεί επεξεργασία αναγνωριστούν οι αλλαγές πλάνου στην εικόνα i και την εικόνα $j + k$, τότε υπάρχει θιττερ k εικόνων, το οποίο δημιουργεί προβλήματα στην αναγνώριση. Είναι σημαντικό εδώ να τονιστεί ότι η αναγνώριση αλλαγών πλάνου στις εικόνες $i + k$ και $j + k$, δηλαδή μια σταθερή μετατόπιση κατά k δε δημιουργεί προβλήματα στην αναζήτηση, αφού κυρίως ενδιαφέρουν τα διαστήματα μεταξύ των αλλαγών και όχι οι απόλυτες τιμές των εικόνων που αυτές παρουσιάζονται. Προφανώς το jitter μπορεί να αφορά τόσο μια μετατόπιση σε εικόνες, όσο και σε χρόνο.

Οι σημαντικότεροι λόγοι δημιουργίας jitter είναι οι εξής: interlaced βίντεο, αλλαγμένος ρυθμός εικόνων (frame rate), αλλά και λάθη στην αναγνώριση των βαθμιαίων αλλαγών πλάνου.

¹ Παρότι στα ελληνικά θα μπορούσε να αποδοθεί ως «τρεμούλιασμα» συνήθως χρησιμοποιείται ο αγγλικός όρος

3.5.2.1 Deinterlace

Όπως αναφέρθηκε στο πρώτο κεφάλαιο, το interlaced βίντεο έχει πλεονεκτήματα μόνο όταν απεικονίζεται σε οθόνες καθοδικού σωλήνα CRT, οι οποίες έχουν σχεδόν εκλείψει. Έτσι, πλέον όλα τα βίντεο που υπάρχουν είναι σε progressive μορφή, είτε επειδή έτσι καταγράφηκαν είτε μετά από μετατροπή.

Το πρόβλημα είναι ότι ανάλογα με τον τρόπο που γίνεται η μετατροπή σε progressive, μπορεί να δημιουργηθεί τελικά jitter στο βίντεο. Αυτό συμβαίνει επειδή μια αλλαγή πλάνου μπορεί να βρίσκεται είτε στο πρώτο είτε στο δεύτερο πεδίο μιας εικόνας. Όπως μπορεί να γίνει κατανοητό, αν χρησιμοποιείται η πρώτη, απλή μέθοδος της ενότητας 1.4.6.2 για το deinterlace και η αλλαγή πλάνου βρίσκεται στο δεύτερο πεδίο της εικόνας i , τότε η εικόνα αυτή θα αποτελείται από ένα πεδίο από το προηγούμενο πλάνο και ένα πεδίο από το επόμενο πλάνο. Αντίθετα σε περίπτωση που η αλλαγή πλάνου βρίσκεται στο πρώτο πεδίο της εικόνας i τότε αυτή θα αποτελείται από δύο πεδία του ίδιου πλάνου.

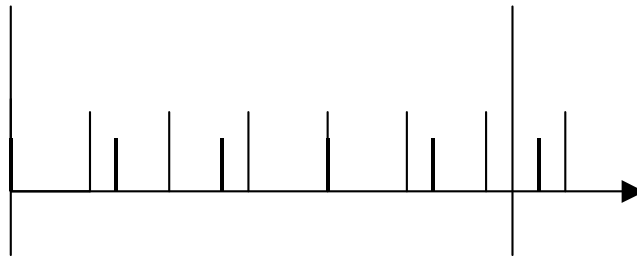
Έτσι, θα ανιχνευθούν τελικά δύο αλλαγές πλάνου αντί για μια αλλαγή πλάνου στην εικόνα i (που ουσιαστικά θα είναι μίξη από δύο διαφορετικά πλάνα αφού οι περιττές γραμμές θα προέρχονται από το προηγούμενο πλάνο και οι άρτιες από το επόμενο), και μια αλλαγή πλάνου στην εικόνα $i + 1$. Επίσης, ανάλογα με τον τρόπο που θα γίνει το deinterlace με την ανίχνευση κίνησης, μπορεί η αλλαγή πλάνου να ανιχνευθεί είτε στην εικόνα i , είτε στην εικόνα $i + 1$. Κάτι αντίστοιχο συμβαίνει και στη μέθοδο με το διπλασιασμό των γραμμών, αφού εκεί ο ρυθμός εικόνων θα είναι διπλάσιος, άρα κάποιες εικόνες θα πρέπει να αγνοηθούν και έτσι, ανάλογα του ποια θα αγνοηθεί η αλλαγή πλάνου θα ανιχνευθεί είτε στην i είτε στην $i + 1$.

Φαίνεται λοιπόν ότι λόγω της διαδικασίας του deinterlace, μπορεί να παρουσιαστούν δύο διαφορετικά φαινόμενα: Το πρώτο είναι η ανίχνευση δύο διαδοχικών σκηνών, στις εικόνες i και $i + 1$, και το δεύτερο είναι η εισαγωγή τυχαίου jitter μιας εικόνας. Το πρώτο πρόβλημα είναι σχετικά εύκολο να επιλυθεί, αφού μπορούν απλά να αγνοηθούν οι αλλαγές πλάνου που ανιχνεύονται σε δύο διαδοχικές εικόνες, αφού δεν χρησιμοποιούνται ποτέ τόσο συχνές αλλαγές πλάνου (αντί για τις δύο συνεχόμενες αλλαγές πλάνου, θα κρατηθεί μόνο η μία, δεν έχει σημασία ποια αφού τελικά το πρόβλημα αυτό θα μετατραπεί στο δεύτερο, δηλαδή εισαγωγή jitter μιας εικόνας). Επίσης, η χρήση της μεθόδου της απλής ένωσης των πεδίων, παρουσιάζει άσχημα αποτελέσματα και, σιγά σιγά, αφού βελτιώνεται και η ταχύτητα των υπολογιστών, εγκαταλείπεται για την μέθοδο deinterlacing με ανίχνευση κίνησης.

Με τη συζήτηση της παραπάνω παραγράφου, εξάγεται το συμπέρασμα ότι λόγω του deinterlacing, μπορεί τελικά να δημιουργηθεί στα βίντεο jitter μιας εικόνας, δηλαδή στο ένα βίντεο η απόσταση μεταξύ δύο αλλαγών πλάνου να είναι ίση με k εικόνες και στο άλλο να είναι $k - 1, k$, ή $k + 1$. Το πώς αυτό το πρόβλημα του jitter μιας εικόνας επιλύεται, αναφέρεται ξεχωριστά σε κάθε μια από τις τεχνικές αναζήτησης που αναλύονται στα επόμενα.

3.5.2.2 Αλλαγή στο ρυθμό εικόνων

Ένα βίντεο τα οποίο μεταδίδεται στο διαδίκτυο, μπορεί να αλλάξει πολλές φορές το ρυθμό μετάδοσης των εικόνων του, ανάλογα με το μέσο στο οποίο πρέπει να προβληθεί. Είναι σημαντικό να αναφερθεί ότι υπάρχουν πολλές περιπτώσεις, μιας και μπορεί μετά από επεξεργασία, κάποιο αρχικό βίντεο να αλλάξει πολλές φορές ρυθμό εικόνων, ή μπορεί το ίδιο βίντεο να βιντεοσκοπηθεί με διαφορετικό ρυθμό εικόνων. Αυτό έχει ως αποτέλεσμα την εισαγωγή jitter. Το παρακάτω παράδειγμα θα ξεκαθαρίσει την κατάσταση: Έστω ότι καταγράφεται ένα πλάνο το οποίο διαρκεί 0,19 δευτερόλεπτα, το οποίο φαίνεται με τις δύο μεγάλες κάθετες γραμμές του σχήματος 3.2. Οι μεσαίες γραμμές έχουν απόσταση ίση με 33,3 ms, δηλαδή είναι δειγματοληψία ίση με 30 fps ($1/30 = 0,033s$), ενώ οι κοντές γραμμές έχουν απόσταση ίση με 40 ms, δηλαδή είναι από δειγματοληψία ίση με 25 fps. Στα 30 fps το νέο πλάνο θα ανιχνευθεί στην 8η εικόνα, ενώ στα 25 fps το νέο πλάνο θα ανιχνευθεί στην 6η εικόνα.



Σχήμα 3.2: Αλλαγή στο ρυθμό μετάδοσης

Έτσι, λόγω του διαφορετικού ρυθμού εικόνων, δημιουργείται ένα jitter δύο εικόνων σε αυτό το απλό παράδειγμα. Το σημαντικό πρόβλημα εδώ είναι ότι το jitter αυτό θα μεγαλώνει σε μεγαλύτερα πλάνα. Για παράδειγμα, σε ένα πλάνο 10 δευτερολέπτων, θα διαρκεί 250 εικόνες στα 25 fps και 300 εικόνες στα 30 fps, δηλαδή jitter 50 εικόνων. Το παραπάνω αποτελεί σημαντικό πρόβλημα, μιας και η σύγκριση των διάρκειών των πλάνων είναι αδύνατη σε αυτή την περίπτωση. Για να λυθεί αυτό υπάρχουν ουσιαστικά δύο λύσεις: Είτε η χρησιμοποίηση του απόλυτου χρόνου σε δευτερόλεπτα για τη μέτρηση των διαστημάτων μεταξύ των πλάνων, είτε η μετατροπή σε έναν κοινό ρυθμό εικόνων που θα κρατείται σταθερός για όλα τα βίντεο που γίνεται η αναζήτηση.

Αναφορικά με την πρώτη μέθοδο, επειδή η διάρκεια του πλάνου είναι περίπου (το περίπου είναι σημαντικό) σταθερή, άσχετα με το ρυθμό εικόνων αυτού, ο λόγος αριθμός εικόνων / ρυθμός εικόνων είναι πάντοτε περίπου σταθερός και ίσος με τη διάρκεια του πλάνου σε δευτερόλεπτα. Οι απόλυτες διάρκειες δε μπορούν να είναι ακριβώς ίσες, εκτός και αν οι ρυθμοί εικόνων είναι πολλαπλάσιοι ο ένας του άλλου. Στο σχήμα 3.2 αυτό μπορεί να φανεί καλύτερα. Αν υποθεθεί ότι το πλάνο είναι στα 30 fps, θα έχει μήκος $1/30 \times 7 = 0,23s$. Αν είναι στα 25 fps, θα έχει μήκος $1/25 \times 5 = 0.2,s$. Έτσι πρέπει να λαμβάνεται υπ' όψιν αυτό το jitter όταν χρησιμοποιείται απόλυτος χρόνος.

3.5. ΛΑΘΗ ΚΑΤΑ ΤΗΝ ΑΝΙΧΝΕΥΣΗ ΑΛΛΑΓΩΝ ΠΛΑΝΟΥ

Για τη μετατροπή σε διαφορετικό ρυθμό εικόνων υπάρχουν διάφοροι τρόποι. Γενικά υπάρχουν διαφορετικές τεχνικές όταν ο ρυθμός εικόνων αυξάνεται και όταν μειώνεται. Ουσιαστικά όλες οι τεχνικές μπορούν να χωριστούν σε απλές και σύνθετες. Στις απλές, για την αύξηση του ρυθμού εικόνων απλώς διπλασιάζονται κάποιες εικόνες, ενώ για τη μείωση απλώς αφαιρούνται κάποιες εικόνες. Για παράδειγμα, για μετατροπή από τα 25 στα 30 fps, αρκεί να διπλασιάζεται κάθε 5η εικόνα ενώ για μετατροπή από τα 30 στα 25 αρκεί να αφαιρείται κάθε 6η εικόνα. Παρά την απλότητα τους και τη μηδαμινή επεξεργασία που χρειάζονται, οι απλές τεχνικές παρουσιάζουν προβλήματα, ειδικά όταν οι διαφορές μεταξύ των ρυθμών εικόνων που πρέπει να μετατραπούν είναι μεγάλες (για παράδειγμα από τα 25 στα 60 fps) οπότε και η κίνηση φαίνεται λιγότερο ομαλή. Αντίθετα, στις πιο σύνθετες τεχνικές, γίνεται ειδική επεξεργασία στις υπάρχουσες εικόνες έτσι ώστε η εικόνα που θα προστεθεί να μην είναι απλώς ίδια με κάποια υπάρχουσα, αλλά, με πρόβλεψη κίνησης να είναι η ενδιάμεση των δύο εικόνων που την περιβάλλουν. Ομοίως, όταν αφαιρεθεί μια εικόνα, η πληροφορία της θα διαμοιραστεί σε αυτές που την περιβάλλουν, έτσι ώστε τελικά η αλλαγή να μην είναι τόσο μεγάλη. Οι τεχνικές αυτές προφανώς κοστίζουν πολύ περισσότερο υπολογιστικά, αλλά το αποτέλεσμα είναι καλύτερο.

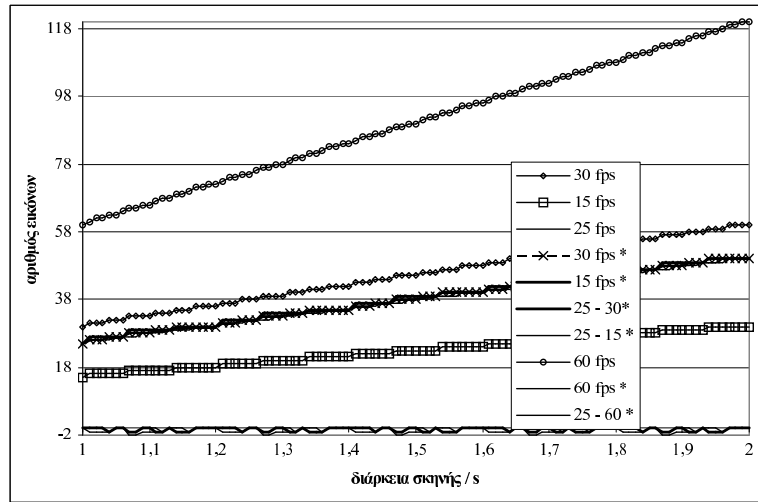
Εκτός της μετατροπής σε άλλο ρυθμό εικόνων η οποία ακόμα και με τον απλό τρόπο έχει υπολογιστικό κόστος, μπορεί σχετικά εύκολα να βρεθεί υπολογιστικά ο αριθμός των εικόνων ενός πλάνου όταν αυτό το πλάνο καταγραφεί σε διαφορετικό ρυθμό εικόνων έτσι ώστε τελικά να γίνει η σύγκριση του αριθμού των εικόνων του πλάνου.

Έστω λοιπόν ένα πλάνο διάρκειας t δευτερολέπτων, το οποίο καταγράφεται στα x fps και έχει i_x εικόνες, δηλαδή είναι $[x \cdot t]$ και $t = t/x - t_x$ όπου με t_x είναι ο χρόνος που περισσεύει από την τελευταία εικόνα μέχρι την αρχή του επόμενου πλάνου. Ο χρόνος αυτός είναι άγνωστος και δεν ενδιαφέρει την ανάλυση παραπάνω αφού πλέον η διάρκεια του πλάνου θα είναι ίση με $t' = i_x/x$. Για παράδειγμα, σε ένα πλάνο διάρκειας $100ms$, με $25fps$ θα είναι $i_x = 3$ εικόνες και $t_x = 20ms$. Όμως πλέον η διάρκεια του πλάνου θα είναι ίση με $3/25 = 0,12s$, αφού η επόμενη εικόνα θα είναι από το επόμενο πλάνο. Για την εύρεση του αριθμού των εικόνων στα y fps, είναι ομοίως $i_y = [y \cdot t'] = [y \cdot i_x/x]$ και $t' = i_x/x = i_y/y - t'_y$. Και το t'_y μπορεί να αγνοηθεί, αφού είναι το υπόλοιπο που περισσεύει από τη μετατροπή και χάνεται.

Ακόμα και με αυτή τη μετατροπή, όμως μπορεί να δημιουργηθεί jitter, λόγω του t'_y .

Για να εξηγηθεί καλύτερα το φαινόμενο αυτό, έγινε το εξής πείραμα: Θεωρήθηκαν διάρκειες πλάνων από 1 μέχρι 2 δευτερόλεπτα, με αύξηση $10ms$ σε κάθε πείραμα. Για κάθε ένα, μετρήθηκε ο αριθμός των εικόνων αν η δειγματοληψία γινόταν με 25, 30, 60 και 15 εικόνες το δευτερόλεπτο, όπως φαίνεται στο σχήμα 3.3.

Όπως είναι φανερό, για τα 25 fps ο αριθμός των εικόνων κυμαίνεται από 25 ως 50, για τα 30 fps από 30 ως 60, για τα 15 fps από 15 ως 30 και για τα 60 fps από 60 ως 120.



Σχήμα 3.3: Jitter λόγω αλλαγή στο ρυθμό εικόνων

Επίσης, έγινε η μετατροπή του αριθμού των εικόνων κάθε πλάνου στα 25 fps με βάση την παρατήρηση για το σταθερό λόγο αριθμού εικόνων προς ρυθμού εικόνων, παίρνοντας τις τιμές με τα * που φαίνονται στο σχήμα. Όπως μπορεί να φανεί, σε κάθε περίπτωση ο αριθμός των εικόνων μετά τη μετατροπή θα είναι είτε ίσος με τον αριθμό των εικόνων για τα 25 fps, με μέγιστη διαφορά ± 2 . Οπότε, μπορεί να θεωρηθεί ότι στις συνήθεις περιπτώσεις των διαφορετικών ρυθμών εικόνων, το jitter που θα δημιουργείται θα είναι μέχρι δύο εικόνες. Αν υπάρχουν εξαιρετικές περιπτώσεις, για παράδειγμα βίντεο με ρυθμό μετάδοσης ίσο με 1 fps, το jitter θα είναι μεγαλύτερο, όμως αυτές μπορούν να αγνοηθούν στη συνέχεια. Έτσι, το πρόβλημα του διαφορετικού fps μπορεί τελικά να λυθεί στις συνήθεις περιπτώσεις, αν μετατραπεί ο ρυθμός εικόνων στα 25 fps και θεωρηθεί ένα jitter μέχρι 2 εικόνες.

3.5.2.3 Λάθος αναγνώριση βαθμιαίων αλλαγών πλάνου

Στις βαθμιαίες αλλαγές πλάνου, οι οποίες συνήθως αποτελούν και αλλαγή σκηνής, το πλάνο δεν αλλάζει από την εικόνα i στην εικόνα $i + 1$, αλλά διαρκεί περισσότερο, για παράδειγμα k εικόνες. Ανάλογα με τον αλγόριθμο εύρεσης αλλαγών πλάνου, μπορεί να βρεθεί αλλαγή πλάνου σε οποιαδήποτε από τις k εικόνες τις βαθμιαίας αλλαγής, ή ακόμα και σε περισσότερες από μια (ή πουθενά). Οι ευκολότερη λύση είναι να μη γίνεται κάποια ιδιαίτερη επεξεργασία όταν υπάρχουν αλλαγές πλάνου και απλώς να θεωρούνται και αυτές εσφαλμένες ή χαμένες αναγνώρισεις, μιας και το jitter που θα δημιουργείται από αναγνώρισεις βαθμιαίων αλλαγών πλάνου σε διαφορετικές εικόνες δεν είναι μόνο 1 ή 2 εικόνες αλλά μπορεί να είναι και πολύ περισσότερο. Επίσης, οι βαθμιαίες αλλαγές πλάνου είναι αρκετά σπάνιες και χρησιμοποιούνται συνήθως σε ειδικά βίντεο, για παράδειγμα trailer κινηματογραφικών ταινιών.

3.6 Πιθανοτική μέθοδος

Η πιθανοτική μέθοδος χρησιμοποιεί τεχνικές από τη θεωρία πιθανοτήτων για να προβλέψει το αν ένα απόσπασμα προέρχεται από κάποιο τμήμα ενός βίντεο της βάσης δεδομένων. Κεντρικό στοιχείο της μεθόδου είναι η έννοια του «συμβάντος», αφού η απόφαση για το αν το απόσπασμα ταιριάζει ή όχι με το συγκεκριμένο τμήμα της βάσης λαμβάνεται με βάση τις διαφορές που έχουν οι ακολουθίες των συμβάντων του αποσπάσματος και του βίντεο της βάσης δεδομένων (όσο λιγότερες οι διαφορές τόσο πιθανότερο το ταίριασμα).

Το συμβάν είναι μια εικόνα ενός βίντεο με ένα συγκεκριμένο χαρακτηριστικό. Το χαρακτηριστικό που θα επιλεγεί για συμβάν θα πρέπει να έχει τις εξής ιδιότητες:

- Τα συμβάντα πρέπει να πραγματοποιούνται τυχαία μέσα στο βίντεο
- Πρέπει να είναι διάσπαρτα σε όλο το μήκος του βίντεο και όχι μαζεμένα κάπου
- Πρέπει να πραγματοποιούνται αρκετά συχνά έτσι ώστε να υπάρχουν αρκετά συμβάντα σε ένα μικρό απόσπασμα
- Πρέπει να είναι εύκολη η αναγνώριση τους.

Εκείνο το χαρακτηριστικό του βίντεο που έχει αυτές τις ιδιότητες είναι η αλλαγή πλάνου, για αυτό και επιλέγεται ως συμβάν.

Ένα πραγματοποιηθέν συμβάν μπορεί να αναγνωριστεί ή όχι. Παρομοίως, ένα συμβάν μπορεί να αναγνωριστεί χωρίς να έχει πραγματοποιηθεί. Ο λόγος είναι οι ατέλειες του μηχανισμού αναγνώρισης συμβάντων. Ειδικότερα για τις αλλαγές πλάνου, όπως έχει ήδη αναφερθεί στο προηγούμενο κεφάλαιο, η πιθανότητα για τέτοιου είδους σφάλματα είναι σημαντική.

Έτσι, ορίζονται τρεις παράμετροι:

Ορισμός 3.6.1 Η πιθανότητα συμβάντος e είναι η πιθανότητα να υπάρχει συμβάν σε μια εικόνα ενός βίντεο.

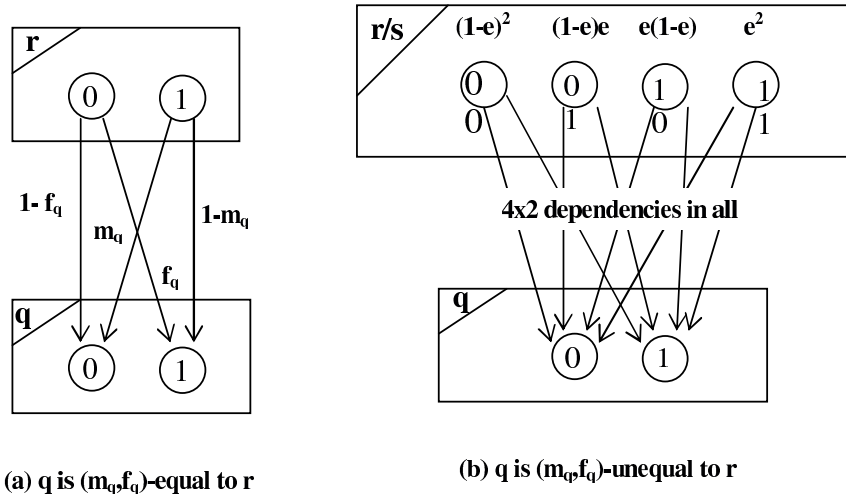
Ορισμός 3.6.2 Η πιθανότητα m για χαμένη αναγνώριση συμβάντος (*missed detection*) είναι η πιθανότητα να μην έχει αναγνωριστεί συμβάν στην εικόνα αλλά στην πραγματικότητα να υπήρχε

Ορισμός 3.6.3 Η πιθανότητα m για εσφαλμένη αναγνώριση συμβάντος (*false alarm*) είναι η πιθανότητα να έχει αναγνωριστεί συμβάν σε μια εικόνα ενώ στην πραγματικότητα δεν υπήρχε.

3.6.1 Ορισμός της (m_q, f_q) -ισότητας

Έστω r μια τυχαία μεταβλητή με τιμές $\{0, 1\}$. Η τιμή 1 (συμβάν) έχει πιθανότητα e και την τιμή 0 (μη-συμβάν) με πιθανότητα $1 - e$. Έστω q μια άλλη μεταβλητή με τις ίδιες τιμές. Η r υποτίθεται ότι αναφέρεται στο πραγματικό γεγονός και η q στο πώς αντιλαμβάνεται το σύστημα το γεγονός αυτό. Έτσι, ορίζεται η ισότητα των δύο αυτών τυχαίων μεταβλητών όσον αφορά τις πιθανότητες της q να έχει χαμένο ή εσφαλμένο συμβάν (σχήμα 3.4a):

Ορισμός 3.6.4 Η q είναι (m_q, f_q) -ίση με την r αν το q είναι 1 με πιθανότητα $e \cdot (1 - m_q) + (1 - e) \cdot f_q$ και 0 με πιθανότητα $(1 - e) \cdot (1 - f_q) + e \cdot m_q$.



(a) q is (m_q, f_q) -equal to r

(b) q is (m_q, f_q) -unequal to r

Σχήμα 3.4: Ορισμός της αντίληψης κατά (m_q, f_q) της r από την q

Όταν η q είναι 1 τότε είτε υπάρχει συμβάν στην r και η q το εντόπισε ($e \cdot (1 - m_q)$), είτε δεν υπάρχει συμβάν στην r αλλά η q εντόπισε εσφαλμένα ένα συμβάν ($(1 - e) \cdot f_q$). Όταν η q είναι 0 τότε είτε υπάρχει συμβάν στην r αλλά η q δεν το εντόπισε ($e \cdot m_q$), είτε δεν υπάρχει συμβάν στην r και η q δεν εντόπισε κάποιο συμβάν εκεί εσφαλμένα ($(1 - e) \cdot (1 - f_q)$). Ο παραπάνω ορισμός προφανώς δεν είναι συμμετρικός. Αν λοιπόν η q είναι (m_q, f_q) -ίση με την r η πιθανότητα της q να ταιριάζει με την r θα είναι το άθροισμα των περιπτώσεων να υπάρχει συμβάν στην r και η q να μη το χάνει ($e(1 - m_q)$) και να μην υπάρχει συμβάν στην r και η q να μην αντιλαμβάνεται συμβάν εσφαλμένα ($(1 - e)(1 - f_q)$), δηλαδή

$$P_r\{q = r\} = 1 - f_q + e(f_q + m_q). \tag{3.1}$$

Με παρόμοιο τρόπο ορίζεται και η ανισότητα:

Ορισμός 3.6.5 Η q είναι (m_q, f_q) -άνιση με την r αν η q είναι (m_q, f_q) -ίση με την s , για κάθε τυχαία μεταβλητή s που παίρνει τις τιμές $\{0, 1\}$ με πιθανότητες $1 - e$ και e αντίστοιχα και η s είναι ανεξάρτητη της r .

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ

Η τυχαία μεταβλητή s δεν παραμετροποιεί ή περιορίζει τον ορισμό της (m_q, f_q) -άνισοτητας. Κάθε s ανεξάρτητη της r μπορεί να χρησιμοποιηθεί και επιπλέον, η s έχει το ίδιο e με το r .

Αν η q είναι (m_q, f_q) -άνιση με την r τότε η πιθανότητα της q να ταιριάζει με την r είναι το άθροισμα (σχήμα 3.4(b))

$$P_r\{q = r\} = e^2(1 - m_q) + e(1 - e)m_q + e(1 - e)f_q + (1 - e)^2(1 - f_q), \quad (3.2)$$

όπου οι όροι του αθροίσματος αναφέρονται αντίστοιχα στις περιπτώσεις:

Το r είναι 1, το s είναι 1, το q δεν χάνει το συμβάν στο s

Το r είναι 0, το s είναι 1, το q χάνει το συμβάν στο s

Το r είναι 1, το s είναι 0, το q αντιλαμβάνεται εσφαλμένο συμβάν στο s

Το r είναι 0, το s είναι 0, το q δεν αντιλαμβάνεται εσφαλμένο συμβάν στο s .

Η (3.2) μετά από πράξεις γίνεται

$$P_r\{q = r\} = 1 - f_q + e \cdot (3f_q + m_q - 2) + 2e^2 + o(h^3). \quad (3.3)$$

Η ισότητα και η ανισότητα κατά (m_q, f_q) μπορούν να γενικευτούν με την έννοια της αντίληψης κατά (m_q, f_q) . Με τη βοήθεια μιας πιθανοτικής θεωρίας καθορίζεται το αν η τυχαία μεταβλητή q είναι αντίγραφο της τυχαίας μεταβλητής r ή μιας άλλης τυχαίας μεταβλητής s , ανεξάρτητης της r . Η τελευταία περίπτωση σημαίνει ουσιαστικά ότι η q είναι διαφορετική από την r .

3.6.2 Αντίληψη συμβάντων με την (m_q, f_q) -ισότητα

Έστω ότι οι τυχαίες μεταβλητές q και p είναι (m_q, f_q) -ίσες και αντίστοιχα (m_p, f_p) -ίσες με την τυχαία μεταβλητή r . Οι πιθανότητες για τα τις τιμές των q και p θα είναι:

(i) Η q και η p έχουν συμβάν	$e(1 - m_q)(1 - m_p) + (1 - e)f_q f_p$
(ii) Η q και η p έχουν μη-συμβάν	$(1 - e)(1 - f_q)(1 - f_p) + em_q m_p$
(iii) Η q έχει συμβάν, η p μη συμβάν	$e(1 - m_q)m_p + (1 - e)f_q(1 - f_p)$
(iv) Η q έχει μη συμβάν, η p συμβάν	$e(1 - f_q)f_p + em_q(1 - m_p)$.

(3.4)

Άθροιση των (i) και (ii) δίνει την πιθανότητα τα p και q να συμπίπτουν:

$$P_r\{q = p\} = 1 - f + e(f - m), \quad (3.5)$$

όπου τέθηκε

$$\begin{aligned} m &= m_p + m_q - 2m_p m_q \\ f &= f_p + f_q - 2f_p f_q. \end{aligned} \quad (3.6)$$

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

Έτσι η ισότητα $P_r\{q = p\}$ της εξίσωσης (3.5) είναι η ίδια με της (3.1) όταν τα m_q και f_q αντικατασταθούν από τα m και f .

Η περίπτωση που η q είναι (m_q, f_q) -ίση με την r ενώ η p είναι (m_q, f_q) -άνιση με την r είναι ίδια με την περίπτωση η q να είναι (m_q, f_q) -άνιση με την r ενώ η p να είναι (m_q, f_q) -ίση με την r ή γενικότερα, και η q και η p να είναι (m_q, f_q) -άνιση και αντίστοιχα (m_p, f_p) -άνιση με την r . Φυσικά στην τρίτη περίπτωση το s το οποίο χρησιμοποιείται για την ανισότητα της p πρέπει να μην είναι το ίδιο με το s για την ανισότητα της q , δηλαδή τα δύο αυτά s να είναι ανεξάρτητα μεταξύ τους. Όλες αυτές οι τυχαίες μεταβλητές χαρακτηρίζονται από μια κοινή πιθανότητα συμβάντος e . Οι περιπτώσεις των τιμών των q και p είναι:

- (i) Τα q και p δε χάνουν ένα συμβάν / τα q και p χάνουν ένα συμβάν.
- (ii) Το q δε χάνει συμβάν, το p αντιλαμβάνεται εσφαλμένο συμβάν / Το q χάνει συμβάν, το p δεν αντιλαμβάνεται εσφαλμένο συμβάν.
- (iii) Τα q δεν αντιλαμβάνεται εσφαλμένο συμβάν, το p χάνει ένα συμβάν / Το q αντιλαμβάνεται εσφαλμένο συμβάν, το p δε χάνει συμβάν.
- (iv) Τα q και p δεν αντιλαμβάνονται εσφαλμένο συμβάν / τα q και p αντιλαμβάνονται εσφαλμένο συμβάν.

Σύμφωνα με τα παραπάνω, η πιθανότητα η τιμή των q και p να συμπίπτει θα δίδεται από τον τύπο

$$P_r\{q = p\} = e^2(1 - m_q)(1 - m_p) + e^2m_qm_p + e(1 - e)(1 - m_q)f_p + e(1 - e)m_q(1 - f_p) + (1 - e)e(1 - f_q)m_p + (1 - e)ef_q(1 - m_p) + (1 - e)^2(1 - f_q)(1 - f_p) + (1 - e)^2f_qf_p. \quad (3.7)$$

Μετά από πράξεις η (3.7) γράφεται

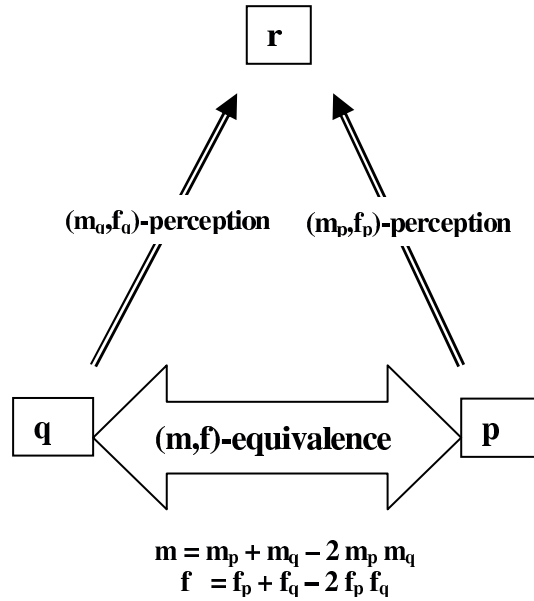
$$P_r\{q = p\} = 1 - f_q - f_p + 2f_qf_p + e(3f_p + 3f_q + m_q + m_p - 2) + 2e^2 + o(h^3), \quad (3.8)$$

με $o(h^3)$ όρους μεγαλύτερης τάξης από τετράγωνο. Όταν το e και οι πιθανότητες για χαμένη και εσφαλμένη αναγνώριση είναι αρκετά μικρές τότε μπορούν να αφαιρεθούν οι όροι 3ης τάξης $6ef_p f_q$ και $eem_p m_q$ και έτσι η (3.7) γίνεται με χρήση της (3.6)

$$P_r\{q = p\} = 1 - f + e(3f + m - 2), \quad (3.9)$$

όπου και στην περίπτωση αυτή η $P_r\{q = p\}$ είναι (περίπου) ίση με την $P_r\{q = p\}$ της (3.3). Τα ζεύγη αβεβαιότητας m_q, f_q και m_p, f_p εισάγονται στην αντίληψη της r από την q μέσω της p . Η περίπτωση όπου η p αντιλαμβάνεται την r μέσω της q θα οδηγούσε στην ίδια έκφραση μιας και τα m και f είναι συμμετρικά στις p και q . Αυτό φαίνεται στο σχήμα 3.5. Έτσι, η q και η p αντιλαμβάνονται το r ισοδύναμα κατά (m, f) . Τα παραπάνω γενικεύουν την περίπτωση να είναι τα q και p «ακριβή αντίγραφα» το ένα του άλλου, όντας και τα δύο «ακριβή αντίγραφα» του r .

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ



Σχήμα 3.5: Ορισμός της ισοδυναμίας κατά (m_q, f_q)

Όταν δεν υπάρχουν χαμένες ή εσφαλμένες αντιλήψεις πλάνων, δηλαδή για $m_q = m_p = f_q = f_p = 0$ η (3.5) δίνει αποτέλεσμα 1, δηλαδή όταν η p και q αντιλαμβάνονται το ίδιο r τότε ταιριάζουν πάντα. Αν όμως αντιλαμβάνονται τα r και s αντίστοιχα, όπου το r είναι ανεξάρτητο του s τότε η (3.8) δίνει πιθανότητα να ταιριάζουν ίση με $1 - 2e + 2e^2$. Αυτή είναι η πιθανότητα των τυχαίων μεταβλητών r και s να ταιριάζουν, δηλαδή είτε και οι δύο να έχουν ένα συμβάν (e^2) είτε και οι δύο να μην έχουν συμβάν $((1-e)^2)$. Πρέπει να σημειωθεί ότι το να περιλαμβάνεται το s στον ορισμό της αντίληψης του r από το q δεν είναι παράδοξο, αφού η ανεξάρτητη τυχαία μεταβλητή s είναι απαραίτητη για έναν πιθανοτικό τρόπο έκφρασης της ανισότητας με το r κρατώντας συμμετρικούς τους ορισμούς της ισότητας και ανισότητας.

Με βάση τα παραπάνω, η r μοντελοποιεί τον πραγματικό κόσμο τον οποίο το σύστημα αντιλαμβάνεται μέσω του μοντέλου p που είναι ένα μη ακριβές αντίγραφο του r (λόγω των εσφαλμένων και χαμένων πλάνων). Η αναζήτηση στο p χρησιμοποιεί άλλο ένα μοντέλο, το q (ερώτημα - query), το οποίο επίσης είναι μη ακριβές αντίγραφο του r και πρέπει να ταιριάζει με τα στοιχεία του p . Αφού δε γίνεται να συγκριθεί το q με το r πρέπει να συγκριθεί το q με το p όπως φαίνεται και από το σχήμα 3.5. Με την παραπάνω ανάλυση αποδείχθηκε ότι είναι δυνατόν ένα συμβάν να θεωρηθεί ότι πραγματοποιήθηκε στο p με πιθανότητα e με αυξημένη όμως αβεβαιότητα όπως φαίνεται από την (3.8).

Έτσι, παρότι οι ρόλοι των p και q είναι συμμετρικοί σύμφωνα με το σχήμα 3.5, αναφέρονται σε τελείως διαφορετικές πρακτικές οντότητες: Το μεν q στο απόσπασμα ενός βίντεο και το p στη βάση δεδομένων μέσα στην οποία θα γίνει η αναζήτηση του αποσπάσματος. Και τα δύο, είναι, λόγω των χαμένων και

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

εσφαλμένων πλάνων, μη ακριβή αντίγραφα του r .

3.6.3 Σύγκριση ακολουθιών με συμβάντα

Έστω οι ακολουθίες τυχαίων μεταβλητών

$$R = \{r_i\}_{i=1}^{|R|}, P = \{p_i\}_{i=1}^{|P|} \text{ και } Q = \{q_i\}_{i=k}^{k+|Q|-1}$$

όπου οι r_i, p_i και q_i έχουν τα χαρακτηριστικά της προηγούμενης ενότητας. Δηλαδή, το R είναι μια πραγματική ακολουθία πειραμάτων που παίρνουν τιμές 0 (μη-συμβάν) ή 1 (συμβάν) της οποίας τα P και Q είναι μη ακριβές αντίγραφα. Η πιθανότητα συμβάντος e είναι αυθαίρετη αλλά σταθερή για όλα τα στοιχεία του R , ενώ το ίδιο ισχύει και για τις πιθανότητες για εσφαλμένη η χαμένη αντίληψη συμβάντος από τα P και Q m_p, f_p και m_q, f_q . Το μήκος $|Q|$ της Q είναι γνωστό με $|Q| \ll |P| = |R|$. Η πραγματική ακολουθία R δεν είναι γνωστή και αυτό που πρέπει να βρεθεί τελικά είναι το σημείο k του R από το οποίο το Q αρχίζει.

Η σχέση των παραπάνω με το πρόβλημα της αναζήτησης βίντεο είναι φανερή. Οι ακολουθίες R, P και Q είναι ακολουθίες συμβάντων σε εικόνες. Η παρουσία μιας αλλαγής πλάνου (ή κάποιου άλλου χαρακτηριστικού) σηματοδοτεί ένα συμβάν (τιμή 1) ενώ η μη αλλαγή πλάνου ένα μη συμβάν (τιμή 0). Το P υποτίθεται ότι είναι μια μεγάλη ακολουθία με όλα τα βίντεο τα οποία βρίσκονται στη βάση δεδομένων, το ένα μετά το άλλο.

Είναι σημαντικό εδώ να διακριθεί ο τρόπος με τον οποίο παράγονται τα P και Q . Το P είναι ένας μεγάλος αριθμός από βίντεο που προέρχονται από διαφορετικές πηγές και με συγκεκριμένα χαρακτηριστικά. Το σύστημα δεν μπορεί να αλλάξει τα χαρακτηριστικά αυτών των βίντεο, αλλά απλώς ανιχνεύει τις αλλαγές πλάνου αυτών. Αντίθετα, το Q είναι το απόσπασμα, το οποίο έχει μικρό μέγεθος και έτσι το σύστημα θα μπορούσε να αλλάξει τα χαρακτηριστικά του και να εφαρμόσει κάποιους πιο εξεζητημένους αλγόριθμους εύρεσης των αλλαγών πλάνου έτσι ώστε να βελτιώσει την ανάκληση και την ακρίβεια ή ακόμα και να βελτιώσει τη μία σε βάρος της άλλης, ανάλογα με τις απαιτήσεις του σχεδιαστή του συστήματος. Έτσι, υπάρχει ένας τρόπος να καθοριστούν σε κάποιο βαθμό οι τιμές των m_q και f_q , δηλαδή των πιθανοτήτων να χαθεί μια αλλαγή πλάνου από το απόσπασμα, ή μια αλλαγή πλάνου που ανιχνεύθηκε στο απόσπασμα αν μην είναι πραγματική. Αυτό δίδει τη δυνατότητα για μερικό έλεγχο των τιμών της αβεβαιότητας που εισάγεται από τα m και f της εξίσωσης (3.6). Αντίθετα, οι αντίστοιχες τιμές για τα βίντεο της βάσης m_p, f_p δεν είναι ούτε ελέγξιμες αλλά ούτε και μπορούν να μετρηθούν με κάποιο τρόπο.

Για την αναζήτηση θα πρέπει να βρεθούν οι θέσεις του P από τις οποίες είναι πιθανό να προέρχεται το Q . Αυτό μπορεί να πραγματοποιηθεί με απλή σύγκριση του πού έχει αλλαγή πλάνου το P και πού το Q , για όλες τις δυνατές θέσεις του P . Συγκεκριμένα:

Ορισμός 3.6.6 Η ολίσθηση του Q μέσα στο P ούτως ώστε να βρεθεί η θέση με τον ελάχιστο αριθμό διαφορών στις εικόνες που θεωρούνται αλλαγές πλάνου

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ

στα P και Q καλείται συγχρονισμός του Q με το P . Όταν το Q βρεθεί στη σωστή θέση k^* ονομάζεται συγχρονισμένο με το R , ενώ σε όλες τις άλλες περιπτώσεις ονομάζεται μη-συγχρονισμένο.

Αν το Q βρεθεί σε μια τυχαία θέση k του P τότε για την πιθανότητα να υπάρχει διαφορά μεταξύ των q_i και p_i θα ισχύει ότι

- εάν το q_i και το p_i αντιλαμβάνονται το ίδιο συμβάν r_i , δηλαδή το Q είναι συγχρονισμένο με το P , η πιθανότητα να διαφέρουν θα δίνεται από την εξίσωση 3.5

$$\epsilon_s = Pr\{q_i \neq p_i\} = (f_q + f_p) - 2f_q f_p + e(m_q + m_p - f_q - f_p), \quad (3.10)$$

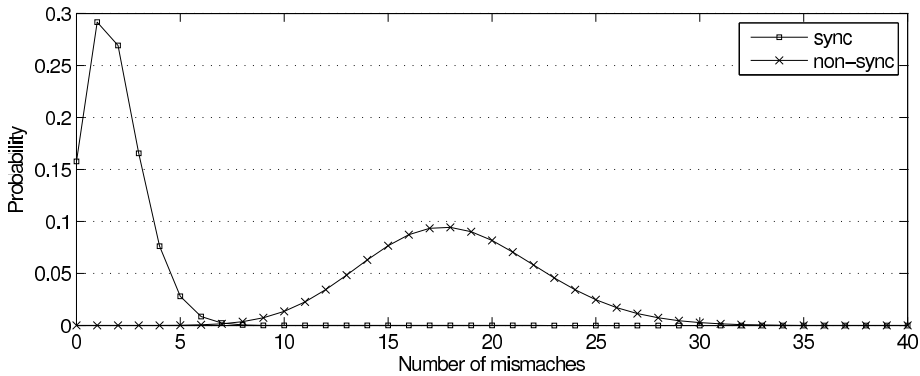
- ενώ αν αντίθετα τα q_i και p_i αντιλαμβάνονται διαφορετικά συμβάντα r_i, r_j δηλαδή δεν είναι συγχρονισμένα τότε η πιθανότητα διαφοράς θα δίνεται από την εξίσωση 3.5

$$\epsilon_n = Pr\{q_i \neq p_i\} = f_q + f_p - 2f_q f_p - e(3f_p + 3f_q + m_q + m_p - 2) - 2e^2, \quad (3.11)$$

όπου το i παίρνει τιμές από το k ως το $k + |Q| + 1$. Έτσι, μπορεί να θεωρηθεί ότι ο έλεγχος για διαφορές για κάθε ένα από τα στοιχεία q_i, p_i μιας ολίσθησης είναι μια σειρά από $|Q|$ πειράματα Bernoulli. Για το σωστό $k = k^*$ τα πειράματα έχουν πιθανότητα «επιτυχίας» ϵ_s ενώ σε κάθε άλλη περίπτωση έχουν πιθανότητα επιτυχίας ϵ_n . «Επιτυχία» στην προκειμένη περίπτωση σημαίνει ότι τα p_i και q_i διαφέρουν. Έτσι, ο αριθμός των διαφορών b θα ακολουθεί τη διωνυμική κατανομή

$$B(b, |Q|, \epsilon) = \binom{|Q|}{b} \epsilon^b (1 - \epsilon)^{|Q| - b} \quad (3.12)$$

με $\epsilon = \epsilon_s$ ή $\epsilon = \epsilon_n$ ανάλογα με την περίπτωση.



Σχήμα 3.6: Παράδειγμα μορφής κατανομών διαφορών

Ένα χαρακτηριστικό παράδειγμα της μορφής των κατανομών φαίνεται στο σχήμα 3.6. Η περίπτωση όπου $\epsilon = \epsilon_n$ βρίσκεται πάντοτε δεξιά αυτής με $\epsilon =$

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

ϵ_s , αφού όταν τα P και Q είναι συγχρονισμένα ο αριθμός των διαφορών θα είναι μικρότερος. Μεταξύ των κορυφών των δύο αυτών κατανομών μπορεί να προστεθεί ένα κατώφλι τ στον αριθμό των διαφορών έτσι ώστε να αποφασίζεται αν το Q είναι συγχρονισμένο ή όχι με το P . Τότε, η συνολική πιθανότητα χαμένης ανίχνευσης μ (το Q είναι συγχρονισμένο με το P αλλά ο συγχρονισμός δεν ανιχνεύεται) και η συνολική πιθανότητα εσφαλμένης ανίχνευσης ϕ (το Q δεν είναι συγχρονισμένο με το P αλλά ανιχνεύεται συγχρονισμός) είναι οι περιοχές (ακριβέστερα τα αθροίσματα αφού οι κατανομές πυκνότητας πιθανότητας είναι διακριτές) που περικλείονται από τις ουρές των κατανομών για τα ϵ_s και ϵ_n , δηλαδή

$$\mu = \sum_{i=\tau}^{|Q|} B(i, |Q|, \epsilon_s) \text{ και } \phi = \sum_{i=0}^{\tau} B(i, |Q|, \epsilon_n) \quad (3.13)$$

Τα μ, ϕ έχουν να κάνουν και με τα $|Q|$ πειράματα και τα οφείλονται στην αβεβαιότητα που εισάγεται από τα m_q, m_p, f_q, f_p . Η επιλογή του τ αλλά και των m_q, f_q (όσο αυτό είναι δυνατόν) εξαρτώνται από την επιλογή του σχεδιαστή για περισσότερες χαμένες ανιχνεύσεις με το κόστος περισσότερων εσφαλμένων ή το αντίθετο, καθώς και ένα όσο είναι δυνατό μεγαλύτερο διάστημα μεταξύ των κατανομών των ϵ_s και ϵ_n .

3.6.4 Μέτρηση διαφορών

Για τη σύγκριση των στοιχείων του Q και του P πρέπει να πραγματοποιηθούν $|Q|$ συγκρίσεις για κάθε ολισθήση k και να μετρηθούν οι διαφορές. Ο αριθμός των ολισθήσεων είναι $|P| - |Q|$ άρα θα χρειαστούν $(|P| - |Q|)|Q|$ συγκρίσεις συνολικά. Σε επίπεδο υλοποίησης, αυτό το αποτέλεσμα μπορεί να επιτευχθεί κάνοντας ένα άλμα στο επόμενο σημείο όπου υπάρχει συμβάν είτε στο Q είτε στο P και μετρώντας τις διαφορές. Επίσης, το ίδιο αποτέλεσμα μπορεί να επιτευχθεί με τη βοήθεια των δυνατοτήτων που δίδονται από το υλικό του υπολογιστή και ειδικότερα από τη σύγκριση αριθμών με τον τελεστή XOR.

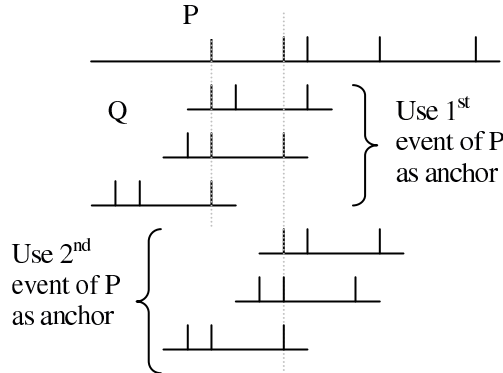
Έστω λοιπόν οι ακολουθίες $\Pi = \{\pi_i\}$ και $K = \{\kappa_i\}$ όπου τα στοιχεία π_i και κ_i περιέχουν θέσεις (σε αριθμό εικόνων) των συμβάντων των P και Q αντίστοιχα. Με βάση αυτές τις ακολουθίες προτείνεται ο αλγόριθμος της επόμενης ενότητας για τη μέτρηση των διαφορών.

3.6.4.1 Αλγόριθμος μέτρησης διαφορών (συμμετρικός)

Χρησιμοποιούνται οι θέσεις συμβάντων των P και Q ως αρχικές θέσεις (άγκυρες) για τη σύγκριση. Αν το Q είναι συγχρονισμένο με το P τότε τουλάχιστον ένα από τα συμβάντα τους πρέπει να ταιριάζει, αλλιώς θα έχουν έναν αριθμό διαφορών ίσο με το άθροισμα των συμβάντων τους. Η χρήση των συμβάντων σαν άγκυρες φαίνεται στο σχήμα 3.7.

Για κάθε ζεύγος π_i/κ_i που αντιστοιχεί σε μια άγκυρα μετράται ο αριθμός των διαφορών με χρήση της συνάρτησης *mismatch_count* του σχήματος 3.8.

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ



Σχήμα 3.7: Συμβάντα ως άγκυρες

Ο αλγόριθμος αυτός καταμετρά αρχικά τις διαφορές δεξιά και κατόπιν αριστερά από την τρέχουσα θέση i, j (η καταμέτρηση προς τα αριστερά παραλείπεται για συντομία). Ο μετρητής των διαφορών αυξάνει όταν υπάρχει συμβάν στο P και όχι στο Q και το αντίθετο. Ο τρόπος με τον οποίο δουλεύει ο αλγόριθμος είναι σχετικά απλός: Για τη θέση i του P και τη θέση j του Q ορίζονται οι δείκτες ii που δείχνει στο i και jj που δείχνει στο j . Όσο οι διαφορές (δηλαδή οι διάρκειες των πλάνων) των P και Q είναι ίδιες, οι ii και jj αυξάνονται. Αν η διάρκεια ενός πλάνου στο Q είναι μικρότερη από το αντίστοιχο πλάνο του P τότε θα αυξηθεί το ii μόνο και θα αυξηθεί ο μετρητής των διαφορών, ομοίως, αν η διάρκεια ενός πλάνου στο Q είναι μεγαλύτερη από το πλάνο στο P τότε θα αυξηθεί το jj και ομοίως θα αυξηθεί ο μετρητής των διαφορών. Αντίστοιχα θα μετρηθούν οι διαφορές και προς τα αριστερά.

Για κάθε μια από τις $e^2|P||Q|$ άγκυρες, η συνάρτηση *mismatch_count* κάνει το πολύ $2e|Q|$ πράξεις, το οποίο είναι ο αριθμός των αναμενόμενων συμβάντων στο P ή το Q για ένα μήκος ίσο με $|Q|$. Το k για το οποίο η *mismatch_count* επιστρέφει την ελάχιστη τιμή είναι και το k^* για το οποίο η Q συγχρονίζεται με την P . Το $e|Q|$, δηλαδή ο αριθμός των συμβάντων σε ένα απόσπασμα μπορεί να θεωρηθεί ως ένας μικρός, σταθερός αριθμός. Σε αυτή την περίπτωση η πολυπλοκότητα της αναζήτησης με τον αλγόριθμο είναι γραμμική ως προς το $e|P|$, δηλαδή τον αριθμό των συμβάντων που περιέχουν τα βίντεο της βάσης δεδομένων.

Η ανάλυση του αλγορίθμου αυτού καλύπτεται από τις ενότητες 3.6.1 και 3.6.3, αφού η θέση συγχρονισμού k^* θα έχει έναν αριθμό διαφορών που ακολουθούν την κατανομή του ϵ_s , ενώ ο αριθμός των διαφορών για τις άλλες θέσεις θα ακολουθεί την κατανομή του ϵ_n .

3.6.4.2 Αλγόριθμος μέτρησης διαφορών (μη συμμετρικός)

Πέραν του συμμετρικού αλγορίθμου μέτρησης διαφορών με τον οποίο μετρώνται οι διαφορές στα συμβάντα τόσο στο P όσο και στο Q , προτείνεται

```

1: Function missmach_count( $K, \Pi$ ) returns  $c$ 
2: for  $i \leftarrow 1, |\Pi|$  do
3:   for  $j \leftarrow 1, |K|$  do
4:     // Count right mismatches
5:      $c \leftarrow 0, ii \leftarrow i, jj \leftarrow j$ 
6:     while  $jj < |K|$  do
7:       while  $\pi_{ii} - \pi_i = \kappa_{jj} - \kappa_j$  do
8:          $ii ++, jj ++$ 
9:       end while
10:      while  $\pi_{ii} - \pi_i < \kappa_{jj} - \kappa_j$  do
11:         $ii ++$ 
12:         $c ++$ 
13:      end while
14:      while  $\pi_{ii} - \pi_i > \kappa_{jj} - \kappa_j$  do
15:         $jj ++$ 
16:         $c ++$ 
17:      end while
18:    end while
19:    // same for left mismatches (omitted)
20:  end for
21: end for
22: return  $c$ 

```

Σχήμα 3.8: Αλγόριθμος μέτρησης διαφορών

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ

εδώ η τροποποίηση του αλγορίθμου του σχήματος 3.8 έτσι ώστε να μετρώνται οι διαφορές μόνο στα σημεία όπου το Q έχει συμβάν. Αυτό μπορεί εύκολα να γίνει αν στη συνάρτηση $mismatch_count$ αφαιρεθεί η γραμμή 10 για τη μέτρηση διαφορών προς τα μπροστά και η αντίστοιχη γραμμή για τις διαφορές προς τα πίσω. Στο εξής θα ονομάζεται QEO - Q Events Only. Ο αλγόριθμος αυτός δεν προσφέρει κάτι αναφορικά με την ταχύτητα, αφού ο αριθμός των πράξεων είναι λίγο πολύ ο ίδιος και στις δύο περιπτώσεις, όμως είναι χρήσιμος σε κάποια συγκεκριμένη μορφή του βίντεο της βάσης δεδομένων P που θα περιγραφεί σε επόμενη ενότητα. Λόγω της μη συμμετρικότητας του αλγορίθμου, δεν καλύπτεται πλήρως από την προηγούμενη ανάλυση, όμως, μια μερική ανάλυση θα παρουσιαστεί στην επόμενη ενότητα.

3.6.4.3 Σύγκριση μεταξύ του συμμετρικού και του μη συμμετρικού αλγορίθμου

Με τη βοήθεια των εξισώσεων που έχουν χρησιμοποιηθεί παραπάνω θα εξαχθούν οι κατανομές πιθανότητας για τον QEO αλγόριθμο.

Για τη συγχρονισμένη περίπτωση, η πιθανότητα για μη ταίριασμα θα είναι:

$$\begin{aligned} Pr\{p = 0|q = 1\} &= \frac{Pr\{p = 0 \wedge q = 1\}}{Pr\{q = 1\}} = \\ &= \frac{e(1 - m_q)m_p + (1 - e)f_q(1 - fp)}{e(1 - m_q) + (1 - e)f_q} \approx \frac{f_q}{f_q + e}, \end{aligned} \quad (3.14)$$

με τους όρους πρώτης τάξεως. Ομοίως, για τη μη συγχρονισμένη περίπτωση είναι

$$\begin{aligned} Pr\{p = 0|q = 1\} &= \frac{e^2(1 - m_q)m_p + (1 - e)ef_q(1 - fp)}{e(1 - m_q) + (1 - e)f_q} + \\ &+ \frac{e(1 - e)(1 - m_q)(1 - fp) + (1 - e)^2f_q(1 - fp)}{e(1 - m_q) + (1 - e)f_q} \approx 1, \end{aligned} \quad (3.15)$$

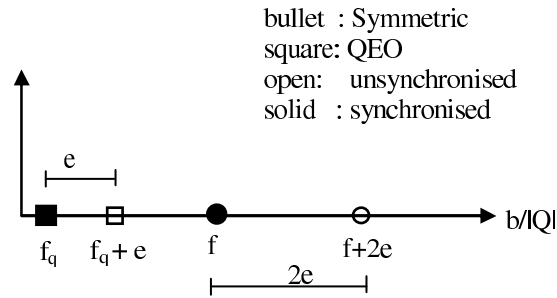
ξανά πάλι με τους όρους πρώτης τάξεως. Ο αριθμητής της παραπάνω έχει μόνο τα 4 από τα συνολικά 16 πιθανά p σε συνάρτηση με το q , δηλαδή μόνο εκείνα όπου το $q = 1$ και $q \neq p$, δηλαδή να μην υπάρχει ταίριασμα αλλά να υπάρχει συμβάν μόνο στο q (και όχι η περίπτωση που δεν υπάρχει ταίριασμα αλλά υπάρχει συμβάν στο p). Έτσι, για το QEO θα χρησιμοποιηθεί η (3.14) σαν ϵ_s και η (3.15) σαν ϵ_n στον τύπο της (3.12). Πλέον, οι δοκιμές θα είναι $(e(1 - m_q) + (1 - e)f_q)|Q| \approx (f_q + e)|Q|$ αντί για $|Q|$.

3.6.4.4 Εξαρτήσεις του συμμετρικού και του QEO αλγορίθμου

Οι κατανομές πιθανότητας των διαφορών εξαρτώνται από την τιμή της πιθανότητας συμβάντος e αλλά και τις πιθανότητες για χαμένη αναγνώριση πλάνου

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

m και εσφαλμένη αναγνώριση πλάνου f . Για ευκολία, στο σχήμα 3.9 φαίνονται οι πιθανότερες τιμές του αριθμού διαφορών για κάθε περίπτωση. Αυτές ουσιαστικά είναι οι κορυφές των διωνυμικών κατανομών των εξισώσεων (3.10), (3.11) και (3.14), (3.15).



Σχήμα 3.9: Οι πιθανότερες τιμές των διαφορών

Στην περίπτωση του συμμετρικού αλγορίθμου η απόσταση μεταξύ της συγχρονισμένης (αριστερά) και της μη συγχρονισμένης (δεξιά) περίπτωσης είναι $2e$ και οι κατανομές ολισθαίνουν προς τα δεξιά (δηλαδή περισσότερες διαφορές) γραμμικά ως προς το f . Αφού του f είναι περίπου ίσο με το άθροισμα των f_p και f_q , είναι σημαντικό να υπάρχουν όσο γίνεται λιγότερες εσφαλμένες αναγνώρισεις αλλαγών πλάνου στο Q , έστω και αν υπάρχουν περισσότερες χαμένες αναγνώρισεις. Στην περίπτωση του QEO αλγορίθμου, ενδιαφέρουν οι εσφαλμένες αναγνώρισεις μόνο στο Q (δηλαδή το f_q), αφού μόνο οι εικόνες που έχουν συμβάντα στο Q λαμβάνονται υπ' όψιν, και η ύπαρξη εσφαλμένων αναγνώρισεων μεταξύ αυτών υποβαθμίζει τη σημασία της αναγνώρισης ενός πραγματικού (και σπάνιου) συμβάντος που προκλήθηκε από το e . Σημειώνεται ότι η προσέγγιση ($f_q + e$) στον παρονομαστή των (3.14) και (3.15) προέρχεται από το γεγονός ότι για το QEO γίνονται $(f_q + e)|Q|$ αντί για $|Q|$ πειράματα, έτσι το μειονέκτημα του μειωμένου αριθμού δοκιμών της QEO περιλαμβάνεται στο σχήμα 3.9. Είναι ενδιαφέρον ότι ο αριθμός των χαμένων αλλαγών πλάνου δεν έχει τόσο μεγάλη σημασία σε καμία από τις δύο περιπτώσεις.

Όπως φαίνεται μια πρώτη ματιά, ο συμμετρικός αλγόριθμος πλεονεκτεί έναντι του QEO αφού η απόσταση είναι $2e$ στη μια περίπτωση και e στην άλλη, ενώ επίσης υπάρχει πολύ μεγαλύτερος αριθμός πειραμάτων. Όμως, η μορφή των διωνυμικών κατανομών εξαρτάται κυρίως από τις τιμές των ϵ_s και ϵ_n και εκεί φαίνεται και η χρησιμότητα του αλγορίθμου QEO, αφού για συγκεκριμένη περίπτωση δίνει «καλύτερες» τιμές στα ϵ_s και ϵ_n , δηλαδή πιο κατάλληλες για το διαχωρισμό της συγχρονισμένης περίπτωσης από τη μη.

3.6.4.5 Αλγόριθμος μέτρησης διαφορών με σύγκριση

Εκτός από τους παραπάνω αλγορίθμους μέτρησης διαφορών, μπορεί να χρησιμοποιηθούν οι δυνατότητες που δίδονται από το υλικό του επεξεργαστή. Ο επεξεργαστής συνήθως παρέχει κατάλληλα λογικά κυκλώματα τα οποία μπο-

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ

ρούν να χρησιμοποιηθούν έτσι ώστε να κάνουν μια σειρά από πράξεις σε έναν μόνο κύκλο ρολογιού.

Μία τέτοια εντολή είναι η XOR - eXclusive OR (αποκλειστικό ή στα ελληνικά). Αυτή, λειτουργεί σε επίπεδο bit και ο τελεστής XOR ανάμεσα στα bit α και β έχει το αποτέλεσμα του πίνακα 3.2, δηλαδή το αποτέλεσμα του α XOR β είναι 1 μόνο αν τα α και β διαφέρουν.

α	β	α XOR β
0	0	0
0	1	1
1	0	1
1	1	0

Πίνακας 3.2: Αποτέλεσμα του α XOR β

Το υλικό του επεξεργαστή προσφέρει δυνατότητες έτσι ώστε αυτή η λειτουργία να μπορεί να γίνει σε ένα byte, word ή double word. Στη συνέχεια θα γίνεται αναφορά σε bytes, αλλά ομοίως μπορούν να επιλεγούν και οι άλλες περιπτώσεις. Τελικά, το αποτέλεσμα του XOR μεταξύ δύο bytes θα είναι 1 σε όσα σημεία τα δύο bytes διαφέρουν. Έτσι, μετρώντας απλώς τους άσσους μπορεί εύκολα να βρεθεί ο αριθμός των διαφορών μεταξύ δύο bytes.

Αυτή η δυνατότητα μπορεί να χρησιμοποιηθεί για πολύ γρήγορη σύγκριση του αποσπάσματος με τα βίντεο της βάσης δεδομένων. Όλα τα βίντεο της βάσης δεδομένων αποθηκεύονται σε μεγάλες ακολουθίες από bytes με μέγεθος ίσο με τον αριθμό των εικόνων τους δια 8. Ομοίως, και το απόσπασμα αποθηκεύεται σε μια ακολουθία από bytes. Το απόσπασμα ολισθαίνει σε όλη τη βάση δεδομένων και για κάθε θέση πραγματοποιείται ένα XOR οπότε το αποτέλεσμα είναι μια τρίτη ακολουθία από bytes. Ο αριθμός των άσπων αυτής της ακολουθίας για κάθε θέση είναι ο αριθμός των διαφορών.

Για την ολίσθηση, θα πρέπει τα bits του αποσπάσματος, να ολισθαίνουν συνέχεια προς τα δεξιά, και να συγκρίνονται με τα bits του βίντεο της βάσης δεδομένων. Υπάρχουν εδώ δύο ολισθήσεις, η ολίσθηση bit και η ολίσθηση byte. Τα bits ολισθαίνουν προς τα δεξιά 8 φορές και στη συνέχεια γίνεται ολίσθηση byte και τα bits ξεκινάνε πάλι από την αρχή.

Για την καταμέτρηση των άσπων, εκτός από κλασσικές μεθόδους, όπως ολίσθηση, εφαρμογή μάσκας και έλεγχος για 1 ή 0, προτιμάται η χρήση lookup table, όπου ένας πίνακας με μέγεθος 256 θέσεις για κάθε θέση από 0 ως 256 θα αποθηκεύει τον αριθμό των bits που είναι 1. Για παράδειγμα, οι θέσεις 0 ως 10 του πίνακα θα είναι: 0, 1, 1, 2, 1, 2, 2, 3, 1, 2, 2. Έτσι, για κάθε byte του αποτελέσματος θα αρκεί η ανάκληση της αντίστοιχής τιμής του πίνακα και η προσθήκη αυτής στο υπάρχον άθροισμα των διαφορών. Ο πίνακας δημιουργείται μόνο μια φορά στην αρχή και έτσι δεν καθυστερεί καθόλου την αναζήτηση.

Σημειώνεται ότι αν αντί για byte χρησιμοποιηθούν words ή double words η χρήση πίνακα δεν είναι τόσο καλή ιδέα λόγω του μεγάλου μεγέθους που θα

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

χρειαζόταν (ειδικά για το double word θα χρειαζόνταν περίπου 4 GB μνήμης). Έτσι, στην περίπτωση αυτή γίνεται χρήση ολίσθησης και μάσκας ενώ επίσης μπορεί να γίνει καταμέτρηση με άθροιση των άσων των επιμέρους 4 bytes του double word.

3.6.5 Χρήση αλλαγών πλάνου

Οι αλλαγές πλάνου είναι το είδος του συμβάντος που χρησιμοποιείται παρακάτω λόγω των χαρακτηριστικών τους: Η εμφάνιση τους είναι τυχαία και η πιθανότητα εμφάνισης e είναι στα ζητούμενα όρια, ενώ υπάρχουν πολλοί τρόποι εύρεσης τους, όπως αυτοί που προτάθηκαν στο προηγούμενο κεφάλαιο. Τα προβλήματα που δημιουργούνται κατά την ανίχνευση των αλλαγών πλάνου έχουν ήδη αναφερθεί σε προηγούμενες ενότητες αυτού του κεφαλαίου, και σε κάθε περίπτωση, εκτός του jitter μπορεί να θεωρηθούν χαμένες ή εσφαλμένες ανιχνεύσεις. Για το jitter γίνεται αναφορά στην ενότητα 3.6.8.

Η βασισμένη στα intra macroblock μέθοδος εξαγωγής αλλαγών πλάνου της ενότητας 2.4 βασίζεται στο γεγονός ότι το βίντεο είναι κωδικοποιημένο με συγκεκριμένο τρόπο, και ειδικά στο ότι η μορφή κάθε ομάδας εικόνων είναι *IPPPP*.. ή έστω *IBBPBP*... δηλαδή δεν έχει πρόσθετες εικόνες *I*. Αυτό συνήθως δε συμβαίνει αφού οι εικόνες *I* αναγκαστικά υπάρχουν για να βοηθήσουν σε λειτουργίες όπως fast forward ή seek, μιας και χωρίς αυτές για να γίνει η μετακίνηση σε άλλο σημείο του βίντεο θα έπρεπε να αποκωδικοποιηθεί ολόκληρο το βίντεο μέχρι εκείνο το σημείο, αφού η μια εικόνα βασίζεται στην προηγούμενη της, και μόνο με την εισαγωγή μιας εικόνας *I* γίνεται απευπλοκή από τις προηγούμενες εικόνες. Αυτές οι πρόσθετες εικόνες *I* μπαίνουν σε τακτά χρονικά διαστήματα τα οποία καθορίζονται από τον κωδικοποιητή.

Για την ακολουθία *P* υπάρχουν δύο περιπτώσεις:

1. Ανελαστική περίπτωση: Το *P* παράγεται από το *R* εκτός ελέγχου του συστήματος. Έτσι, το *P* θα έχει εικόνες που θα είναι πραγματικές αλλαγές πλάνου με την προσθήκη όμως εικόνων *I* που δεν είναι πραγματικές αλλαγές πλάνου. Αυτά ονομάζονται «πρόσθετα» συμβάντα και υποτίθεται ότι δε μπορεί να γίνει προσπάθεια για να αναγνωριστούν ή να εξαλειφθούν.

2. Ελαστική περίπτωση: Το *P* παράγεται από το *R* κάτω από τον έλεγχο του συστήματος. Εδώ, δεν υπάρχουν εικόνες *I* ή ακόμα και αν υπάρχουν υπάρχει τρόπος να αποφασιστεί αν αυτές είναι ή όχι αλλαγές πλάνου, δηλαδή τα πρόσθετα συμβάντα λείπουν ή είναι αναγνωρίσιμα. Ένα πιθανό πρόβλημα θα μπορούσε να υπάρξει αν τύχει κάποιο τέτοιο αναγνωρίσιμο πρόσθετο συμβάν να πέσει στην ίδια εικόνα με ένα πραγματικό συμβάν έτσι ώστε να το καλύψει. Υποτίθεται ότι αυτή η περίπτωση καλύπτεται από τη γενική πιθανότητα απώλειας συμβάντος m_p .

Η ακολουθία *Q* (το απόσπασμα) μπορεί να υποτεθεί ότι πάντοτε παράγεται κάτω από τον έλεγχο του συστήματος λόγω του μικρού μεγέθους του, οπότε το σύστημα φροντίζει να μην υπάρχουν πρόσθετα συμβάντα.

Η παρουσία των πρόσθετων συμβάντων στο *P* είναι ο λόγος που διαχωρί-

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ

στηχε καθαρά ο ρόλος των P και Q και προτάθηκε ο μη συμμετρικός αλγόριθμος αναζήτησης. Οι ρόλοι των P και Q έχουν να κάνουν με τη σχέση τους με το R και δεν είναι συμμετρικοί. Εκτός από το ότι το Q είναι πολύ μικρότερο του P τμήμα αυτής της εργασίας βασίζεται και στο ότι το Q δε θα έχει πρόσθετα συμβάντα ενώ το P μπορεί να έχει, ανάλογα με τα 1, 2 που παρουσιάστηκαν πριν.

Πλέον φαίνεται η χρησιμότητα της μη συμμετρικής αναζήτησης: Μπορεί να μην είναι τόσο καλή όσο η συμμετρική αναζήτηση στη γενική περίπτωση, όμως είναι ένας τρόπος αναζήτησης στην ανελαστική περίπτωση που αναφέρθηκε πιο πάνω. Πραγματικά, με τη μη συμμετρική αναζήτηση χρησιμοποιούνται μόνο τα συμβάντα του Q , τα οποία είναι πραγματικά, και εξετάζεται αν το P έχει συμβάν στην αντίστοιχη εικόνα. Τα συμβάντα του P από μόνα τους αγνοούνται τελείως.

Έτσι, το συμπέρασμα είναι ότι στην ανελαστική περίπτωση θα πρέπει να χρησιμοποιηθεί η μη συμμετρική αναζήτηση, ενώ στην ελαστική περίπτωση η συμμετρική αναζήτηση. Το πλεονέκτημα της μη συμμετρικής αναζήτησης είναι εμφανές: Παρότι το αποτέλεσμα της στο διαχωρισμό του συγχρονισμένου από το μη συγχρονισμένο απόσπασμα δεν είναι τόσο καλό όσο στη συμμετρική, μπορεί να γίνει η αναζήτηση σε υπάρχοντα H.264 βίντεο χωρίς να χρειάζεται η κατάλληλη μετατροπή τους.

3.6.6 Έλεγχος ορθότητας ανάλυσης

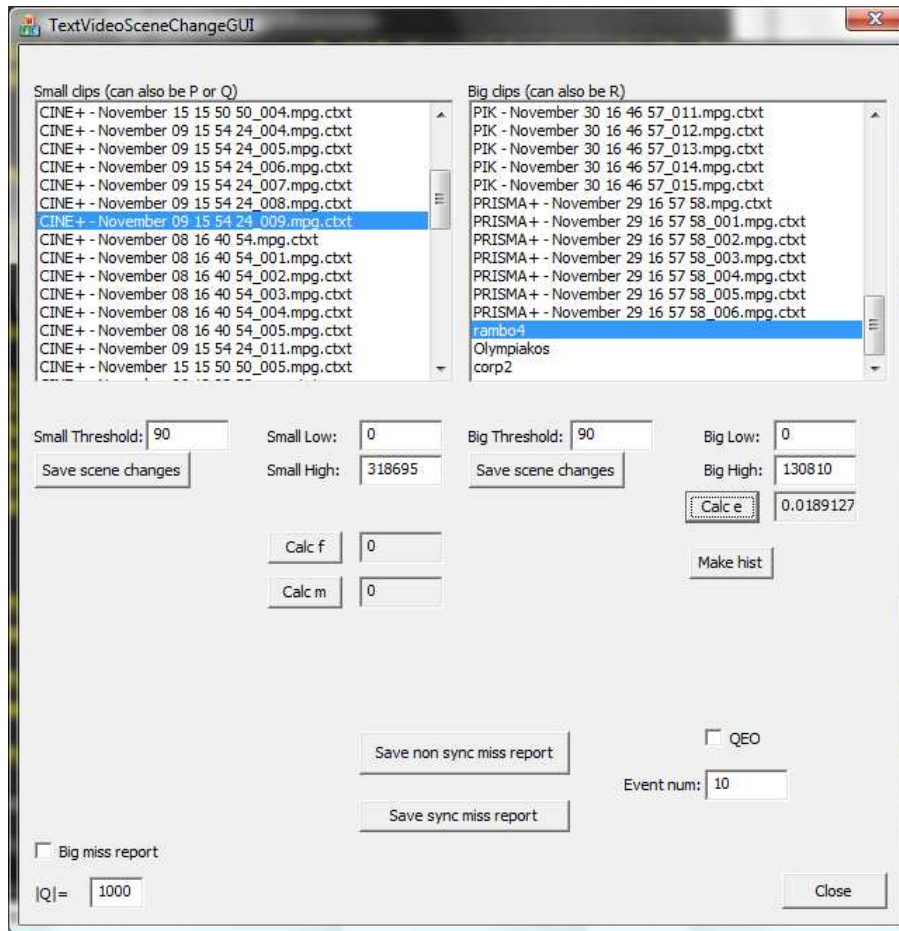
Για την εξαγωγή πειραματικών αποτελεσμάτων προς έλεγχο της ορθότητας της παραπάνω ανάλυσης, δημιουργήθηκε ένα πρόγραμμα με Windows MFC με όνομα TextVideoSceneChangeGUI. Το πρόγραμμα αυτό είχε εκτεταμένες δυνατότητες για ανάλυση βίντεο και θα περιγραφεί πιο αναλυτικά στο παράρτημα Α'. Μια εικόνα του φαίνεται στο σχήμα 3.10. Για τις δοκιμές χρησιμοποιήθηκε η κινηματογραφική ταινία «12 Monkeys» διάρκειας 129 λεπτών.

3.6.6.1 Δημιουργία των R, P και Q

Το βίντεο κωδικοποιήθηκε με το H.264 με διαφορετικό παράγοντα συμπίεσης κάθε φορά (Q_{factor} [WSBL03]) έτσι ώστε να δημιουργηθούν οι περιπτώσεις των R, P και Q . Στη συνέχεια, χρησιμοποιήθηκε ο αλγόριθμος της ενότητας 2.4 με διαφορετικό κατώφλι για κάθε περίπτωση ούτως ώστε να εξαχθούν οι εικόνες στις οποίες υπάρχει αλλαγή πλάνου.

- Για τη δημιουργία του R το βίντεο κωδικοποιήθηκε με ένα παράγοντα συμπίεσης ίσο με 20 και ως συμβάντα θεωρήθηκαν οι εικόνες στις οποίες το ποσοστό των intra coded macroblocks ήταν πάνω από 90%. Αυτός ο ορισμός των συμβάντων, δίνει μια πιθανότητα για συμβάν ίση με $e = 0,0089$.
- Για την περίπτωση του P το βίντεο κωδικοποιήθηκε με παράγοντα συμπίεσης ίσο με 30 και ως κατώφλι απόφασης αλλαγής πλάνου χρησιμοποιή-

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ



Σχήμα 3.10: Μια εικόνα του προγράμματος TextVideoSceneChangeGUI

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ

θηκε το 80%. Έτσι το P είναι ένα ατελές αντίγραφο του R με χειρότερη ποιότητα λόγω της οποίας κάποιες αλλαγές πλάνου πλέον δεν ανιχνεύονται. Αυτό συμβαίνει επειδή με τη μεγαλύτερη συμπίεση ο κωδικοποιητής βίντεο θα προτιμήσει τη συμπίεση με βάση την κίνηση και έτσι ο αριθμός των intra coded macroblocks θα μειωθεί. Σημειώνεται ότι αυτές οι τιμές του παράγοντα συμπίεσης δεν είναι κομμάτι της μεθόδου που προτείνεται εδώ, απλώς βοηθούν στο να δημιουργηθούν διαφορετικές παραλλαγές του ίδιου βίντεο.

Το 80% που επιλέχτηκε ως κατώφλι ούτως ώστε να αντισταθμίσει κάπως τις χαμένες αλλαγές πλάνου, εισάγει την ανίχνευση κάποιων αλλαγών πλάνου οι οποίες δεν υπήρχαν στην πρώτη περίπτωση. Οι πιθανότητες m_p και f_p μπορούν εύκολα να βρεθούν πειραματικά, αν μετρηθούν οι αλλαγές πλάνου που υπάρχουν στο R και όχι στο P (για το m_p) και οι αλλαγές πλάνου που υπάρχουν στο P αλλά όχι στο R (για το f_p). Έτσι, η περίπτωση αυτή έχει $m_p = 0,033$ και $f_p = 0,00033$.

- Τέλος, για το Q το βίντεο κωδικοποιήθηκε με παράγοντα συμπίεσης ίσο με 40, αφού μπορεί να υποθεθεί ότι το απόσπασμα θα είναι κακής ποιότητας και ως κατώφλι απόφασης αλλαγής πλάνου χρησιμοποιήθηκε το 60%. Παρομοίως με την προηγούμενη περίπτωση κάποιες αλλαγές πλάνου θα χαθούν ενώ κάποιες άλλες θα ανιχνευθούν χωρίς να υπάρχουν στο R και μάλιστα τώρα η συχνότητα αυτών θα είναι μεγαλύτερη. Έτσι, μετρήθηκαν τα $m_q = 0,0468$ και $f_q = 0,00084$.

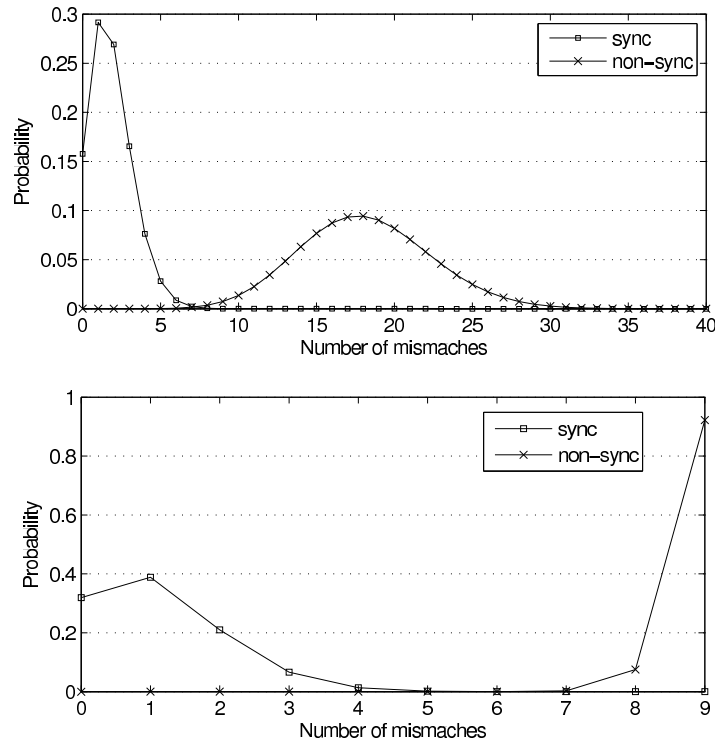
Οι μετρήσεις για το Q αφορούσαν ολόκληρο το βίντεο, όμως για τις δοκιμές που έγιναν παρακάτω χρησιμοποιήθηκε μόνο ένα απόσπασμα, για παράδειγμα με μέγεθος $|Q| = 1000$ εικόνες (40 δευτερόλεπτα στα 25 fps). Για να εξαχθούν τα στατιστικά συμπεράσματα, χρησιμοποιήθηκαν πολλά τέτοια αποσπάσματα έτσι ώστε να αναπαραστήσουν έναν αριθμό από διαφορετικά αποσπάσματα. Συγκεκριμένα, για την περίπτωση που το Q είναι συγχρονισμένο με το P το P «κόπηκε» σε αποσπάσματα μήκους 1000 εικόνων και για κάθε ένα και κάθε ένα εξ αυτών μετρήθηκαν, με βάση τον αλγόριθμο της ενότητας 3.6.4.1 οι διαφορές που έχει από το αντίστοιχο κομμάτι του Q . Αντίστοιχα, για τη μη συγχρονισμένη περίπτωση, για κάθε ένα από τα αποσπάσματα μετρήθηκαν οι διαφορές του με όλες τις θέσεις του P εκτός από τη θέση από την οποία αυτό προήλθε. Όπως μπορεί να γίνει κατανοητό τα πειράματα που αφορούσαν τη μη συγχρονισμένη περίπτωση ήταν πολύ περισσότερα από τη συγχρονισμένη, για αυτό και η μορφή της καμπύλης σε αυτή την περίπτωση ήταν πιο ομαλή.

3.6.6.2 Αναλυτικά και πειραματικά αποτελέσματα

Οι θεωρητικές κατανομές πιθανότητας των διαφορών όταν το Q είναι συγχρονισμένο και όταν το Q δεν είναι συγχρονισμένο, για τα Q, P και R όπως προηγουμένως, φαίνονται στο σχήμα 3.11, τόσο για το συμμετρικό όσο και για

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

το μη συμμετρικό αλγόριθμο. Για το συμμετρικό αλγόριθμο η τιμή $|Q| = 1000$ αντιστοιχεί σε 1000 πειράματα Bernoulli για τη διωνυμική κατανομή για κάθε απόσπασμα που αναζητείται. Για το μη συμμετρικό αλγόριθμο, το μήκος του Q έχει να κάνει με τον αναμενόμενο αριθμό συμβάντων στο Q , δηλαδή $(f_q + e)|Q| \approx f_q|Q| = 1000 \times 0.0089 \approx 9$. Η αριστερή καμπύλη αφορά τη συγχρονισμένη περίπτωση και η δεξιά τη μη συγχρονισμένη. Στη θέση k^* λοιπόν, το πιθανότερο είναι να υπάρχουν το πολύ 4 ή 5 διαφορές μεταξύ του Q και του P , ενώ σε κάθε άλλη θέση το πιθανότερο είναι να υπάρχουν παραπάνω από 10 διαφορές. Η κορυφή της κατανομής στη μη συγχρονισμένη περίπτωση είναι περίπου 18, αφού είναι 9 συμβάντα στο Q και 9 στο P σε τυχαία (άρα και διαφορετικά) σημεία.

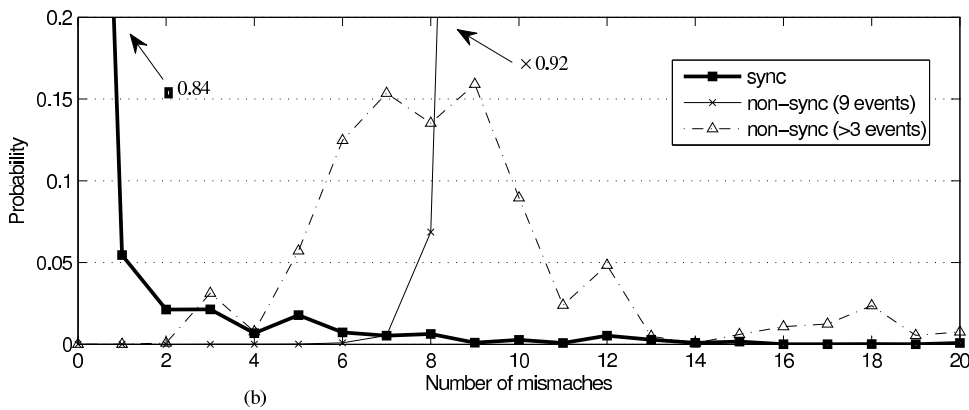
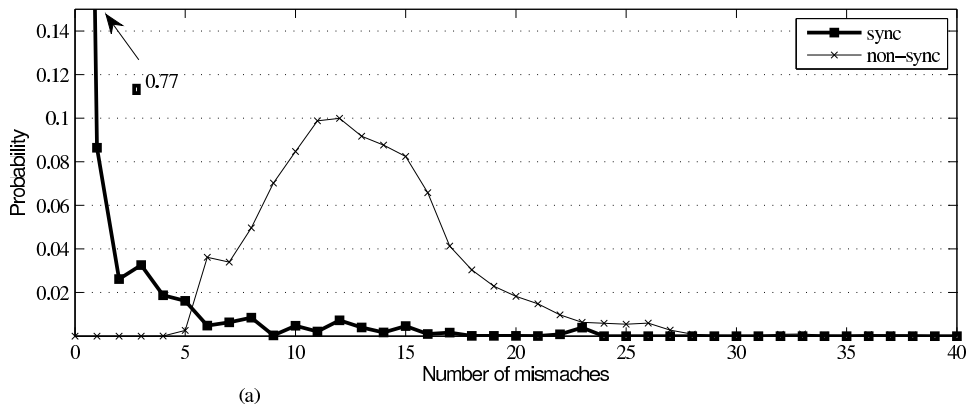


Σχήμα 3.11: Αναλυτικά αποτελέσματα: (a) Συμμετρικός αλγόριθμος και (b) Μη συμμετρικός αλγόριθμος

Τα πειραματικά αποτελέσματα δίδονται στο σχήμα 3.12, όπου φαίνεται η κατανομή του αριθμού των διαφορών που ανιχνεύθηκαν στην πραγματικότητα από τα πειράματα. Με ένα απόσπασμα μήκους 1000 εικόνων, ο συμμετρικός αλγόριθμος μπορεί να εφαρμοστεί απευθείας, όπως φαίνεται στην εικόνα 3.12 (a). Όπως φαίνεται από τη σύγκριση των σχημάτων 3.11 και 3.12 η απόσταση μεταξύ των κατανομών των διαφορών στη συγχρονισμένη και τη μη συγχρονισμένη περίπτωση είναι ελαφρώς καλύτερη στα πειραματικά από ότι στα θεωρητικά αποτελέσματα. Επίσης, και οι δύο καμπύλες είναι μετατοπισμένες προς τα αριστερά και στις δύο περιπτώσεις. Οι διαφορές αυτές οφείλονται στο ότι

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ

στην πειραματική περίπτωση το πότε γίνονται αλλαγές σκηνών αλλά και το πότε αυτές ανιχνεύονται σωστά ή λάθος δεν είναι ομοιόμορφα κατανομημένα σε όλο το βίντεο όπως υποτίθεται στη θεωρητική περίπτωση αλλά υπάρχουν διάφορες εξαρτήσεις. Για παράδειγμα, σε μια σκηνή με πολύ δράση, όπου τα intra-coded macroblocks για μια σειρά από εικόνες ξεπερνούν το κατώφλι που τέθηκε θα μετρηθούν πολλές εσφαλμένες ανιχνεύσεις, οι οποίες δε θα μετρηθούν στις υπόλοιπες σκηνές, οπότε θα υπάρχουν πολλές θέσεις (αποσπάσματα των 1000 εικόνων) στις οποίες ο αριθμός των συμβάντων είναι πολύ μικρότερος από τα περίπου 9 που υπολογίστηκε αρχικά, για αυτό και οι καμπύλες μετατοπίζονται προς τα αριστερά.



Σχήμα 3.12: Πειραματικά αποτελέσματα: (α) Συμμετρικός αλγόριθμος και (β) Μη συμμετρικός αλγόριθμος

Εννοιολογικά αυτό μπορεί να γίνει κατανοητό από το γεγονός ότι αποσπάσματα που έχουν λιγότερα συμβάντα (πχ 0 ή 1) θα ταιριάζουν με πολύ μεγαλύτερη πιθανότητα με τις περισσότερες θέσεις του P , αφού οι αλλαγές πλάνου δεν είναι ομοιόμορφα κατανομημένες άρα αφήνουν μεγάλα κενά. Το παραπάνω είναι αναμενόμενο: Ένα απόσπασμα που δεν έχει συμβάντα έχει πολύ λίγη πληροφορία και τελικά δεν είναι δυνατή η αναζήτηση αυτού με την προτεινόμενη μέθοδο,

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

αντίθετα, για ένα απόσπασμα με πολλές αλλαγές σκηνών η απόφαση για το αν είναι συγχρονισμένο ή όχι με μια θέση μπορεί να εξαχθεί με πολύ μεγαλύτερη βαρύτητα. Για βελτίωση αυτών των πειραματικών αποτελεσμάτων μπορεί να χρησιμοποιηθεί ένα κριτήριο μη αποδοχής αποσπασμάτων με λιγότερα από, για παράδειγμα 5 συμβάντα. Έτσι, η κατανομή διαφορών για τη μη συγχρονισμένη περίπτωση δε θα είχε όρους κάτω από το 5.

Στην περίπτωση του μη συμμετρικού αλγορίθμου, πρέπει να γίνει ένας έλεγχος σχετικά με την καταλληλότητα του αποσπάσματος. Στο παράδειγμα αυτό, είναι γνωστό ότι αναμένονται 9 συμβάντα, όμως, το απόσπασμα μπορεί να έχει έναν αυθαίρετο αριθμό συμβάντων, ακόμα και κανένα. Προφανώς, ένα τέτοιο απόσπασμα, ή ένα απόσπασμα με πολύ λίγα συμβάντα θα είναι ακατάλληλο για την αναζήτηση, οπότε πρέπει να υπάρχει κάποιο κριτήριο που να καθορίζει αν το απόσπασμα έχει επαρκή αριθμό συμβάντων ώστε να έχει νόημα η αναζήτηση. Στα πειραματικά αποτελέσματα, φαίνονται δύο περιπτώσεις στην εικόνα 3.12 (b): Μία που υποστηρίζει αποσπάσματα με ακριβώς 9 συμβάντα, κάτι που δεν είναι δυνατό να εφαρμοστεί στην πραγματικότητα, και μία που υποστηρίζει αποσπάσματα που έχουν περισσότερα από τρία συμβάντα, κάτι που έχει νόημα σε ένα πραγματικό σύστημα. Η καμπύλη με τα τρίγωνα είναι ουσιαστικά ο μέσος όρος των καμπυλών με τα x για ένα συγκεκριμένο αριθμό συμβάντων, δηλαδή 4,5 κοκ. Ο αριθμός των συμβάντων του Q αφορά μόνο τη μη συγχρονισμένη περίπτωση και δεν επηρεάζει τη συγχρονισμένη. Η αβεβαιότητα στον αριθμό των πειραμάτων Bernoulli που θα διεξαχθούν εξηγεί τη διαφορά των αναλυτικών και των πειραματικών αποτελεσμάτων του μη συμμετρικού αλγορίθμου. Η εικόνα 3.11 (b) (αναλυτικά αποτελέσματα) βασίζεται σε έναν καθορισμένο αριθμό 9 συμβάντων και πρέπει να συγκριθεί με την καμπύλη με τα x στην εικόνα 3.12 (b). Στην ίδια εικόνα, τα τρίγωνα αναπαριστούν την περίπτωση να υπάρχουν πάνω από τρία συμβάντα στο απόσπασμα. Είναι ενθαρρυντικό ότι ακόμα και αυτός ο μικρός αριθμός συμβάντων εξασφαλίζει ένα περιθώριο μεταξύ της συγχρονισμένης και της μη συγχρονισμένης περίπτωσης.

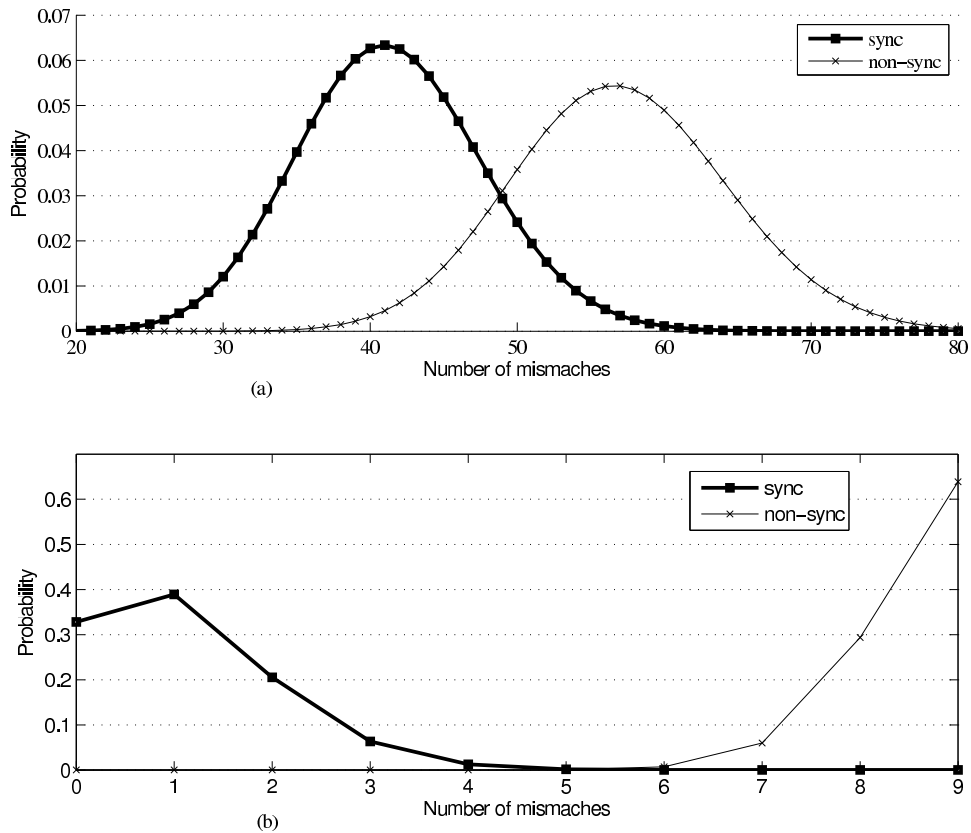
Συμπερασματικά, τα αποτελέσματα του μη συμμετρικού αλγορίθμου είναι πολύ ενθαρρυντικά, έστω και αν δεν υποστηρίζονται πλήρως από το μαθηματικό μοντέλο που προτάθηκε. Ο μη συμμετρικός αλγόριθμος όμως δε φαίνεται να έχει κάποιο πλεονέκτημα έναντι του συμμετρικού, κάτι που θα αλλάξει όταν τα πειράματα γίνουν με βάση την ανελαστική περίπτωση, δηλαδή υπάρχουν πρόσθετα συμβάντα (εικόνες I) στο βίντεο.

3.6.6.3 Παρουσία πρόσθετων συμβάντων

Στα προηγούμενα αναλύθηκε η ελαστική περίπτωση της ενότητας 3.6.5, δηλαδή το P δημιουργήθηκε από το R υπό τον έλεγχο του συστήματος και τυχόν πρόσθετα συμβάντα εξαλείφθηκαν. Για την ανελαστική περίπτωση, μια εικόνα κάθε 25 κωδικοποιείται ως I , δηλαδή έχει 100% intra macroblock. Με βάση τις παραμέτρους της 3.6.6.1 βρέθηκαν τα $f_p = 0,0403$ και $m_p = 0,0336$. Πλέον τα πρόσθετα συμβάντα δε μπορούν να αφαιρεθούν από το P , άρα, οι

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ

μεν χαμένες αλλαγές πλάνου είναι περίπου οι ίδιες (0,0336 τώρα σε σχέση με 0,0330 στην ελαστική περίπτωση), οι δε αλλαγές πλάνου που ανιχνεύθηκαν εσφαλμένα έχουν αυξηθεί περίπου κατά 0,04 (0,0403 σε σχέση με 0,00033 στην ελαστική περίπτωση). Η αύξηση είναι αναμενόμενη αφού πλέον 1 στα 25 (δηλαδή το 4 %) συμβάντα δεν υπάρχουν στην πραγματικότητα. Στο Q δεν υπάρχουν πρόσθετα συμβάντα, αφού το Q είναι πολύ μικρό και μπορεί να υποστεί επεξεργασία (για παράδειγμα ανακωδικοποίηση του σε H.264) για την αφαίρεση τυχών πρόσθετων εικόνων I .



Σχήμα 3.13: Θεωρητικά αποτελέσματα με την παρουσία πρόσθετων εικόνων I : (α) Συμμετρικός αλγόριθμος και (β) Μη συμμετρικός αλγόριθμος

Αυτή η περιοδική εισαγωγή των εικόνων I στο P αλλάζει δραματικά τη θεωρία (σχήμα 3.13) αλλά όχι τόσο την πράξη (σχήμα 3.14). Στην εικόνα 3.13, για το συμμετρικό αλγόριθμο πλέον δεν υπάρχει κανένα περιθώριο διαχωρισμού μεταξύ της συγχρονισμένης και μη περίπτωσης αφού ένα μεγάλο τμήμα τους επικαλύπτεται. Αντίθετα, η εικόνα για το μη συμμετρικό αλγόριθμο δεν αλλάζει σχεδόν καθόλου, κάτι που, στη θεωρία καθιστά το μη συμμετρικό αλγόριθμο αξιόπιστη λύση ακόμα και στην παρουσία πρόσθετων εικόνων I .

Για τα πειραματικά αποτελέσματα, ο συμμετρικός αλγόριθμος διαφέρει αρκετά από τα θεωρητικά, κάτι που οφείλεται στη συσχέτιση που υπάρχει μεταξύ

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

των συμβάντων αντίθετα με τις θεωρητικές υποθέσεις. Φαίνεται ότι στη συγχρονισμένη περίπτωση αναμένεται να υπάρχουν κατά μέσο όρο 40 σφάλματα (αφού στις 1000 εικόνες θα εισαχθούν $1000/25 = 40$ πρόσθετες εικόνες I). Είναι ενδιαφέρον ότι υπάρχουν και λίγες θέσεις με λιγότερα από 40 σφάλματα (για παράδειγμα υπάρχει περίπου 3% πιθανότητα για 38 σφάλματα), στις οποίες τυχαίνει ένα από πρόσθετα συμβάντα του P να πέφτει πάνω σε μια εσφαλμένη αναγνώριση πλάνου του Q .

Ο μη συμμετρικός αλγόριθμος δεν διαφέρει πολύ από την ανελαστική περίπτωση αφού οι διαφορές μετρώνται μόνο στα συμβάντα του Q που είναι όλα πραγματικά. Το μόνο ρίσκο αυτής της διαδικασίας είναι η ύπαρξη ενός εσφαλμένου συμβάντος στο Q και η σύγκριση με ένα πρόσθετο συμβάν στο P , δηλαδή θα φανεί ότι σε εκείνο το σημείο οι ακολουθίες των συμβάντων ταιριάζουν ενώ κανονικά θα διέφεραν. Η πιθανότητα για κάτι τέτοιο είναι ίση με $1/25$. Έτσι, τα αποτελέσματα της εικόνας 3.14 δείχνουν ότι ο μη συμμετρικός αλγόριθμος επιτυγχάνει αρκετά καλά αποτελέσματα στην ανελαστική περίπτωση. Επιπρόσθετα, τα αποτελέσματα βελτιώνονται πολύ όταν εμφανίζεται ένα απόσπασμα που να έχει μεγάλο αριθμό συμβάντων. Σημειώνεται ότι όλα τα αποσπάσματα που έχουν παραπάνω από 9 συμβάντα θα βρίσκονταν στα δεξιά και κάτω από την καμπύλη με τα 9 συμβάντα του σχήματος 3.14.

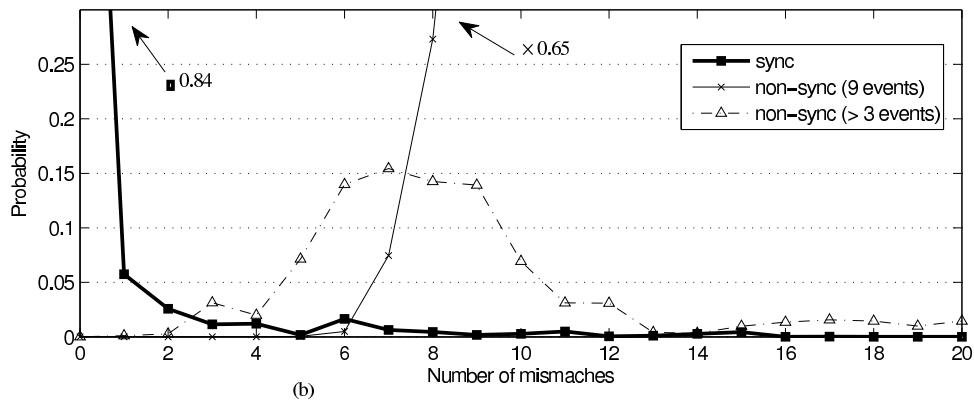
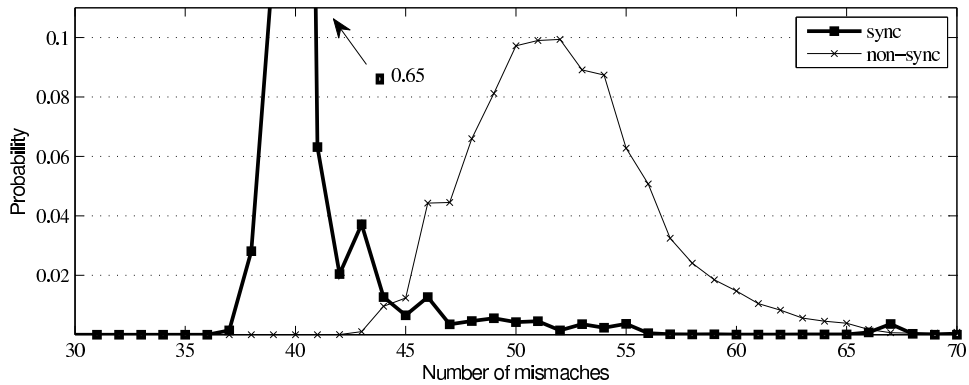
3.6.7 Μετρήσεις για την e

Όπως μπορεί να γίνει κατανοητό, ο αριθμός των αλλαγών πλάνου που αναγνωρίζεται είναι πολύ σημαντικός για την πιθανοτική μέθοδο. Όπως θα φανεί και στα επόμενα, ο αριθμός των συμβάντων είναι σημαντικός και για τις υπόλοιπες μεθόδους. Γενικά, είναι προφανές ότι αν δεν υπάρχουν αρκετά συμβάντα σε κάποιο απόσπασμα τότε αυτό θα ταιριάζει με όλα τα σημεία οπότε η αναζήτηση αυτού του αποσπάσματος θα είναι αδύνατη!

Ο αριθμός των συμβάντων εξαρτάται, εκτός των άλλων, από το είδος του κάθε βίντεο, αφού είναι αναμενόμενο ένα μουσικό βίντεο κλιπ να έχει περισσότερες αλλαγές πλάνου από π.χ. μια τηλεοπτική εκπομπή λόγου. Επίσης, εξαρτάται και από τα κατώφλια για την ανίχνευση που χρησιμοποιήθηκαν στο κεφάλαιο 2. Έτσι, για τα βίντεο του πίνακα 2.1 του κεφαλαίου 2, μετρήθηκαν οι αλλαγές πλάνου που αναγνωρίστηκαν σε διάφορες περιπτώσεις τόσο για τη μέθοδο των ιστογραμμάτων όσο και για τη μέθοδο *intra macroblock*.

Στον πίνακα 3.3, για τη μέθοδο των ιστογραμμάτων, φαίνεται η εξάρτηση του e από το κατώφλι των διαφορών ιστογραμμάτων. Όπως φαίνεται, με ένα κατώφλι ίσο με 50, το e παίρνει τιμή 0,011. Αυτό σημαίνει ότι στις 1000 εικόνες, οι 11 θα αναγνωριστούν ως αλλαγές πλάνου. Αυτές οι 1000 εικόνες, στα 25 fps αντιστοιχούν σε 40 δευτερόλεπτα, δηλαδή ένα βίντεο 40 δευτερολέπτων, θα έχει περίπου 11 συμβάντα όταν το κατώφλι είναι ίσο με 50. Τα αποτελέσματα είναι συνολικά και για τα 4 αρχικά βίντεο. Σημειώνεται ότι για τα τροποποιημένα βίντεο είχαν αντίστοιχες επιδόσεις για αυτό και δεν περιλαμβάνονται στον πίνακα. Επίσης, το βίντεο μουσικής είχε σχεδόν 5 φορές περισσότερα συμβάντα

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ



Σχήμα 3.14: Πειραματικά αποτελέσματα με την παρουσία πρόσθετων εικόνων I: (α) Συμμετρικός αλγόριθμος και (β) Μη συμμετρικός αλγόριθμος

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

από τα άλλα βίντεο!

Κατώφλι	e
50	0,011
75	0,008
100	0,005

Πίνακας 3.3: Επίδραση στην e του κατωφλιού διαφορών ιστογραμμμάτων κατά την ανίχνευση αλλαγών πλάνου

Στον πίνακα 3.4, για τη μέθοδο των intra macroblock φαίνεται η εξάρτηση του e τόσο από το κατώφλι του αριθμού intra macroblock ανά εικόνα όσο όμως και από το συντελεστή κβαντισμού Q . Στην περίπτωση δηλαδή της μεθόδου των intra macroblock το πόσα συμβάντα θα ανιχνευθούν εξαρτάται σημαντικά και από το συντελεστή κβαντοποίησης με τον οποίο έχει συμπιεστεί το H.264 βίντεο. Αυτό συμβαίνει επειδή για μεγαλύτερο συντελεστή κβαντοποίησης, άρα και χειρότερη ποιότητα, ο H.264 κωδικοποιητής προτιμά να χρησιμοποιεί λιγότερα intra macroblock σε κάθε εικόνα. Για περισσότερη εποπτικότητα, τα αποτελέσματα του πίνακα 3.4 σχεδιάζονται στο σχήμα 3.15.

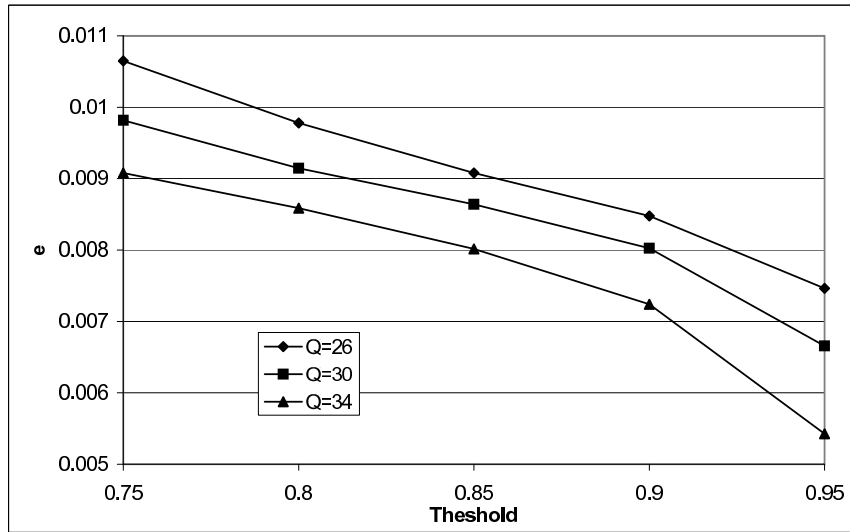
Κατώφλι	$e(Q = 26)$	$e(Q = 30)$	$e(Q = 34)$
0,75	0,011	0,0098	0,0090
0,80	0,0098	0,0091	0,0086
0,85	0,091	0,0086	0,0080
0,90	0,0085	0,0080	0,0072
0,95	0,0075	0,0067	0,0054

Πίνακας 3.4: Επίδραση στην e του κατωφλιού intra macroblock και του Q κατά την ανίχνευση αλλαγών πλάνου

Από το σχήμα 3.15 φαίνεται λοιπόν ότι σε κάθε περίπτωση, όταν αυξάνεται το κατώφλι των intra macroblock για αναγνώριση ως αλλαγή πλάνου, το e μειώνεται. Επίσης, για μεγαλύτερο Q , το e αυξάνεται. Με βάση αυτό το διάγραμμα, μπορεί να εξαχθεί το συμπέρασμα ότι όταν το Q του απόσπασματος είναι μεγαλύτερο από το Q του αρχικού βίντεο (δηλαδή όταν το απόσπασμα έχει χειρότερη ποιότητα από ότι το αρχικό βίντεο), θα πρέπει το κατώφλι που θα χρησιμοποιηθεί στην αναγνώριση αλλαγών πλάνου για το απόσπασμα να είναι μικρότερο από το κατώφλι για την αναγνώριση αλλαγών πλάνου για το αρχικό βίντεο. Σαν παράδειγμα, αν υποθεθεί ότι το αρχικό βίντεο έχει $Q = 26$ και το απόσπασμα $Q = 30$ τότε, αν χρησιμοποιηθεί ένα κατώφλι ίσο με 0,9 στο αρχικό βίντεο, στο απόσπασμα θα πρέπει να χρησιμοποιηθεί ένα κατώφλι ίσο με 0,85, έτσι ώστε και στις δύο περιπτώσεις το e να είναι περίπου το ίδιο (δηλαδή περίπου ίσο με 0,085).

Δυστυχώς, το συμπέρασμα αυτό δεν είναι γενικό, μιας και υπάρχει μεγάλη εξάρτηση τόσο από το περιεχόμενο κάθε ενός βίντεο, όσο και από τους

3.6. ΠΙΘΑΝΟΤΙΚΗ ΜΕΘΟΔΟΣ



Σχήμα 3.15: Γραφική αναπαράσταση των αποτελεσμάτων του πίνακα 3.4

συντελεστές κβαντοποίησης, μπορεί όμως να χρησιμοποιηθεί σαν μια οδηγία, ότι δηλαδή αν το αρχικό βίντεο έχει καλύτερη ποιότητα από το απόσπασμα, το κατώφλι που θα χρησιμοποιηθεί στη μέθοδο αναγνώρισης αλλαγών πλάνου με τα intra macroblock για το αρχικό βίντεο θα πρέπει να είναι μεγαλύτερο από ότι στο απόσπασμα.

3.6.8 Χειρισμός Jitter

Στην ενότητα 3.5.2 παρουσιάστηκαν οι λόγοι για τους οποίους μπορεί να δημιουργηθεί jitter κατά την εύρεση αλλαγών πλάνου και αποδείχθηκε ότι στις συνήθεις περιπτώσεις αυτό θα είναι ίσο με 1 ή 2 εικόνες. Όπως μπορεί να γίνει αντιληπτό, ένα τέτοιο jitter θα δημιουργήσει μεγάλα προβλήματα στον αλγόριθμο μέτρησης διαφορών του σχήματος 3.8.

Μια μικρή τροποποίηση στον αλγόριθμο αυτόν λύνει το πρόβλημα. Συγκεκριμένα, στη γραμμή 7 του 3.5.2 αντί για ισότητα των διαστημάτων θα πρέπει αυτά να διαφέρουν λιγότερο από το jitter, δηλαδή θα γίνει $|(π_{ii} - π_i) - (k_{jj} - k_j)| \leq j$, όπου j είναι η τιμή του jitter. Παρομοίως, η γραμμή 10 θα γίνει $(π_{ii} - π_i) + j < (k_{jj} - k_j)$ και η 14 $(π_{ii} - π_i) + j > (k_{jj} - k_j)$.

Δηλαδή, με το παραπάνω εξετάζεται αν τα πλάνα είναι ίσα ή διαφέρουν το πολύ όσο το jitter που μπορεί να γίνει ανεκτό, οπότε βρίσκεται αν οι αλλαγές πλάνου εμφανίζονται και στα δύο βίντεο στην ίδια εικόνα με μια ανοχή ίση με το jitter.

3.6.9 Γενικά για την πιθανοτική μέθοδο

Τόσο τα θεωρητικά όσο και τα πειραματικά αποτελέσματα δείχνουν ότι ο αριθμός των διαφορών στην ανίχνευση των αλλαγών πλάνου για ένα απόσπα-

σμα Q που προέρχεται από το R στην περίπτωση που το Q είναι συγχρονισμένο με το R θα έχει μια ασφαλή διαφορά από τις μη συγχρονισμένες περιπτώσεις. Φυσικά, στην περίπτωση που γίνεται σύγκριση του Q με κάποιο άλλο R δε θα υπάρχει καν συγχρονισμένη περίπτωση. Παρόλα αυτά, μπορεί να γίνει καταννητό ότι η εύρεση μιας σειρά αποτελεσμάτων επιλέγοντας τη θέση με τα λιγότερα σφάλματα ή τις θέσεις οι οποίες έχουν αριθμό σφαλμάτων κάτω από κάποιο επιλεχθέν κατώφλι δεν είναι αρκετή για να αποφασιστεί ότι πράγματι το απόσπασμα προέρχεται από εκείνο το σημείο της βάσης δεδομένων, ειδικά για αποσπάσματα με μικρό αριθμό από αλλαγές πλάνου. Έτσι, η μέθοδος αυτή, στην περίπτωση μικρών αποσπασμάτων (ειδικά αν αυτά έχουν πολύ λίγες αλλαγές πλάνου) μπορεί να χρησιμοποιηθεί σαν ένα φίλτρο το οποίο θα μειώνει πολύ τον αριθμό των υποψήφιων θέσεων από τις οποίες μπορεί να προήλθε το απόσπασμα με αποδοτικό τρόπο τόσο όσον αφορά την ταχύτητα όσο και την ευελιξία σε αλλαγές χαρακτηριστικών, δηλαδή για αναζήτηση υποψηφίων.

3.7 Μέθοδος βάσης δεδομένων

Η βασική ιδέα στην αναζήτηση βάσης δεδομένων, βασίζεται και εδώ, όπως στην πιθανοτική μέθοδο της ενότητας 3.6, στις αλλαγές πλάνου. Η κύρια διαφορά είναι ότι αντί να ελέγχονται οι διαφορές που έχουν ακολουθίες απο συμβάντα και μη συμβάντα, η αναζήτηση γίνεται με το να ελέγχονται για ταίριασμα ακολουθίες από διάρκειες πλάνου. Οι δύο αυτοί τρόποι έχουν αρκετές ομοιότητες μεταξύ τους, έχοντας όμως ως κύρια διαφορά τον τρόπο που γίνεται τελικά η αναζήτηση: Στο μέθοδο βάσης δεδομένων, οι εικόνες που έχουν αλλαγή πλάνου και η διάρκεια κάθε πλάνου βρίσκονται αποθηκευμένες σε μια βάση δεδομένων, οπότε για την αναζήτηση γίνεται μια ερώτηση (ένα SELECT σε SQL) στη βάση.

Η υλοποίηση με τη βοήθεια της βάσης δεδομένων παρείχε πολλά σημαντικά στοιχεία μιας και μπορούσε να χρησιμοποιηθεί το πλήρες περιβάλλον του RDBMS, κάτι που βοηθά πολύ την υλοποίηση. Παρακάτω θα αναλυθεί λεπτομερώς το γιατί η μέθοδος που αναπτύχθηκε στην ενότητα 3.6 με τη σύγκριση ακολουθιών από συμβάντα είναι πολύ δυσκολότερο να υλοποιηθεί σε βάση δεδομένων από ότι η σύγκριση ακολουθιών από διάρκειες. Παρότι έχουν προταθεί συγκεκριμένες δομές δεδομένων οι οποίες αποθηκεύουν γεωμετρικά στοιχεία και πιθανότατα θα μπορούσαν να χρησιμοποιηθούν για τέτοια αναζήτηση (Spatial Data Sets [Sam89, Fal96]), η κλιμάκωση τους δεν είναι καλή σε χώρους μεγάλων διαστάσεων ([RSB98]) όπως τα ερωτήματα που θα δημιουργηθούν στην περίπτωση της πιθανοτικής μεθόδου, για αυτό και στη συνέχεια θα χρησιμοποιηθούν κλασσικοί σχεσιακοί πίνακες και ερωτήσεις σχεσιακής άλγεβρας για την αναπαράσταση και αναζήτηση των δεδομένων.

Ένα σημαντικό στοιχείο αναφορικά με την υλοποίηση που πρέπει να ξεκαθαριστεί εδώ είναι το τι συμβαίνει με τα άκρα του αποσπάσματος: Επειδή δε μπορεί να γνωρίζει το σύστημα αν στο σημείο που το απόσπασμα αρχίζει και

3.7. ΜΕΘΟΔΟΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

τελειώνει υπάρχει πράγματι αλλαγή πλάνου (το πιθανότερο είναι φυσικά να μην υπάρχει) οι διάρκειες των άκρων αγνοούνται. Έτσι, η πρώτη διάρκεια πλάνου που χρησιμοποιείται είναι από την πρώτη αλλαγή πλάνου μέχρι τη δεύτερη, και η τελευταία διάρκεια είναι από την προτελευταία αλλαγή πλάνου μέχρι την τελευταία.

3.7.1 Η δημιουργία της ερώτησης στη βάση

Η σημαντικότερη διαδικασία της μεθόδου αυτής είναι η δημιουργία του ερωτήματος προς τη βάση δεδομένων (database query). Για να εξηγηθεί το πώς δημιουργείται το ερώτημα, αρχικά θα εξηγηθεί η απλή περίπτωση, στη συνέχεια θα προστεθούν κάποια στοιχεία έτσι ώστε να υπάρχει ανοχή σε καθορισμένο jitter και τέλος η προσθήκη ανοχής σε εσφαλμένες και χαμένες σκηές.

Για την αποθήκευση των αλλαγών πλάνου στη βάση δεδομένων χρησιμοποιήθηκε ένας πίνακας με όνομα ΑΛΛΑΓΗ_ΠΛΑΝΟΥ έχοντας τα εξής πεδία:

- ΚΛΕΙΔΙ_ΒΙΝΤΕΟ το οποίο χρησιμοποιήθηκε για να ξεχωρίζουν οι αλλαγές πλάνου από διαφορετικά μεταξύ τους βίντεο
- ΧΡΟΝΟΣ_ΑΡΧΗΣ για την αποθήκευση του χρόνου ms μέσα στο βίντεο όπου γίνεται η αλλαγή πλάνου
- ΔΙΑΡΚΕΙΑ είναι η διάρκεια του προηγούμενου πλάνου σε ms

Ένα παράδειγμα της μορφής αυτού φαίνεται στον πίνακα 3.5.

ΚΛΕΙΔΙ_ΒΙΝΤΕΟ	ΧΡΟΝΟΣ_ΑΡΧΗΣ	ΔΙΑΡΚΕΙΑ
0	1100	1100
0	4000	2900
0	5540	1540
...
1	560	560
1	1020	460
1	1960	940
1	2760	800
...

Πίνακας 3.5: Παράδειγμα του πίνακα ΑΛΛΑΓΗ_ΠΛΑΝΟΥ

3.7.1.1 Απλή περίπτωση

Για ένα απόσπασμα με k διάρκειες πλάνου, οι οποίες θα έχουν διάρκεια q_1, q_2, \dots, q_k , το ερώτημα σε SQL έχει την εξής γενική μορφή στην απλή περίπτωση:
`SELECT d_1 .ΚΛΕΙΔΙ_ΒΙΝΤΕΟ, d_1 .ΧΡΟΝΟΣ_ΑΡΧΗΣ FROM ΑΛΛΑΓΗ_ΠΛΑΝΟΥ d_1 , ΑΛΛΑΓΗ_ΠΛΑΝΟΥ d_2 , ..., ΑΛΛΑΓΗ_ΠΛΑΝΟΥ d_k`

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

WHERE

```
d1.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ = d2.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ AND
d1.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ = d3.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ AND ...
AND d1.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ = dk.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ AND
d1.ΔΙΑΡΚΕΙΑ = q1 AND
d2.ΔΙΑΡΚΕΙΑ = q2 AND ...
AND dk.ΔΙΑΡΚΕΙΑ = qk AND
d1.ΧΡΟΝΟΣ_ΑΡΧΗΣ+d2.ΔΙΑΡΚΕΙΑ = d2.ΧΡΟΝΟΣ_ΑΡΧΗΣ AND
d2.ΧΡΟΝΟΣ_ΑΡΧΗΣ+d3.ΔΙΑΡΚΕΙΑ = d3.ΧΡΟΝΟΣ_ΑΡΧΗΣ AND ...
dk-1.ΧΡΟΝΟΣ_ΑΡΧΗΣ+dk.ΔΙΑΡΚΕΙΑ = dk.ΧΡΟΝΟΣ_ΑΡΧΗΣ
```

Στο ερώτημα αυτό ορίζονται μια σειρά από «μεταβλητές», οι d_1, d_2, \dots, d_k οι οποίες διατρέχουν τον πίνακα ΑΛΛΑΓΗ_ΠΛΑΝΟΥ. Το να οριστούν τόσες μεταβλητές που αναφέρονται στον ίδιο πίνακα είναι απαραίτητο έτσι ώστε να μπορεί να ελεγχθεί η ακολουθία των διαρκειών των πλάνων, ότι δηλαδή η μία είναι μετά την επόμενη. Φαίνεται λοιπόν ότι στην SQL ο έλεγχος για το αν σε έναν πίνακα υπάρχουν συνεχόμενες τιμές δεν είναι τόσο εύκολος, μιας και θα πρέπει να γίνουν join μεταξύ του ίδιου πίνακα.

Λόγω αυτού του γεγονότος μπορεί να γίνει κατανοητό και το γιατί η πιθανοτική μέθοδος της 3.6 είναι πολύ δύσκολο να υλοποιηθεί σε βάση δεδομένων με αυτόν τον τρόπο. Μιας και υπάρχουν αποσπάσματα που ίσως έχουν και παραπάνω από 1000 εικόνες, θα έπρεπε να οριστούν 1000 μεταβλητές σε έναν πίνακα παρόμοιο με τον ΑΛΛΑΓΗ_ΠΛΑΝΟΥ. Αντίθετα, στην περίπτωση αυτή το k δύσκολα θα πάρει τιμές πάνω από 10, και ακόμα και αν συμβεί κάτι τέτοιο μπορεί εύκολα ο αριθμός του, δηλαδή ο αριθμός των συνεχόμενων διαρκειών πλάνων που αναζητούνται, να περιοριστεί από κάποιο κατώφλι.

Το WHERE μπορεί να χωριστεί σε τρία μέρη: Το πρώτο μέρος καθορίζει ότι πρέπει όλες οι διάρκειες πλάνου που θα βρεθούν να είναι από το ίδιο βίντεο, το δεύτερο μέρος ορίζει ότι η διάρκεια του πρώτου πλάνου θα είναι q_1 , του δεύτερου q_2 κοκ, ενώ το τρίτο μέρος καθορίζει τη σειρά, ότι δηλαδή το δεύτερο πλάνο πρέπει να αρχίζει αμέσως μόλις τελειώσει το πρώτο (αφού το πεδίο ΔΙΑΡΚΕΙΑ του d_i αποθηκεύει την διάρκεια του $i - 1$ πλάνου και όχι του i λόγω του τρόπου που γράφονται τα δεδομένα στη βάση)².

3.7.1.2 Ανοχή σε Jitter

Για την εισαγωγή ανοχής σε jitter ίση με j ms θα πρέπει να αλλάξει το μεσαίο μέρος του απλού ερωτήματος ως εξής:

```
abs(d1.ΔΙΑΡΚΕΙΑ - q1 < j) AND
abs(d2.ΔΙΑΡΚΕΙΑ - q2 < j) AND ...
AND abs(dk.ΔΙΑΡΚΕΙΑ - qk < j) AND
```

²Όταν βρεθεί μια αλλαγή πλάνου είναι γνωστή η διάρκεια του προηγούμενου πλάνου και όχι του τρέχοντος

3.7. ΜΕΘΟΔΟΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

Απλώς λοιπόν θα πρέπει αντί να αναζητούνται διάρκειες πλάνων ακριβώς ίσες με q_1, q_2, \dots, q_k , να αναζητούνται διάρκειες πλάνων που θα διαφέρουν από τα q_1, q_2, \dots, q_k λιγότερο από το κατώφλι j .

3.7.1.3 Ανοχή σε σφάλματα

Δυστυχώς η υλοποίηση της ανοχής σε σφάλματα όπως εσφαλμένες αναγνωρίσεις και χαμένες αναγνωρίσεις αλλαγών πλάνου είναι αρκετά δυσκολότερη από ότι η ανοχή στο jitter.

Έστω μια σειρά από διάρκειες πλάνων q_1, q_2, \dots, q_k από το απόσπασμα. Αυτό που μπορεί να έχει συμβεί σε σχέση με το βίντεο της βάσης δεδομένων από το οποίο αυτές προέρχονται είναι είτε στο αρχικό βίντεο ένας αριθμός από αυτά τα πλάνα να ήταν ένα, αλλά λόγω εσφαλμένων αναγνωρίσεων να θεωρήθηκαν περισσότερα στο απόσπασμα, είτε κάποια από αυτά τα πλάνα να αποτελούνταν από παραπάνω από ένα πλάνα στο αρχικό βίντεο, αλλά λόγω χαμένων αναγνωρίσεων να εμφανίζονται στο απόσπασμα ως ένα.

Ένα παράδειγμα θα ξεκαθαρίσει περισσότερο τα πράγματα: Έστω ότι στο απόσπασμα έχουν βρεθεί τρία πλάνα, με διάρκειες a, b, c . Στο αρχικό βίντεο, μπορεί, λόγω των χαμένων αναγνωρίσεων αντί για τα a, b, c να υπήρχαν τα a_1, a_2, b, c , με $a_1 + a_2 = a$ ή τα a, b, c_1, c_2, c_3 με $c_1 + c_2 + c_3 = c$, ή λόγω εσφαλμένων αναγνωρίσεων το αρχικό βίντεο να ήταν μόνο ένα πλάνο με διάρκεια $a + b + c$, ή δύο πλάνα με διάρκειες $a + b, c$, ή, λόγω ενός συνδυασμού των περιπτώσεων, το αρχικό βίντεο να είχε δύο πλάνα με διάρκειες a_1 και $a_2 + b + c$, με $a_1 + a_2 = a$!

Από τα προηγούμενα μπορεί να γίνει κατανοητό ότι οι δυνατές περιπτώσεις είναι πάρα πολλές, οπότε θα πρέπει να μπουνε κάποια όρια στην ανοχή εσφαλμένων και χαμένων αναγνωρίσεων. Έτσι, μπορούν να χρησιμοποιηθούν δύο μεγέθη:

- Η ανοχή σε χαμένες αναγνωρίσεις στο απόσπασμα md_t που είναι ο αριθμός των αλλαγών πλάνου που μπορεί να έχουν χαθεί σε κάθε ένα πλάνο του αποσπάσματος. Δηλαδή, για κάθε ένα πλάνο του αποσπάσματος, μπορεί να υπάρχουν μέχρι md_t αλλαγές πλάνου στο αρχικό βίντεο οι οποίες δεν έχουν ανιχνευθεί.
- Η ανοχή σε εσφαλμένες αναγνωρίσεις fa_t που είναι ο αριθμός των συνηχόμενων αλλαγών πλάνου που μπορεί να αναγνωριστούν στο απόσπασμα χωρίς να υπάρχουν. Δηλαδή, μέχρι $fa_t + 1$ πλάνα του αποσπάσματος μπορεί στο αρχικό βίντεο να ήταν ένα ενιαίο πλάνο.

Το πώς βρίσκονται όλες οι δυνατές περιπτώσεις για αυθαίρετες τιμές του md_t και του fa_t είναι σχετικά περίπλοκο και θα εξηγηθεί στην επόμενη ενότητα.

3.7.2 Εύρεση των δυνατών περιπτώσεων

Όπως αναφέρθηκε και προηγουμένως, υπάρχουν δύο πιθανότητες: Η μια είναι να υπάρχουν μέχρι md_t συνεχόμενες αλλαγές πλάνου στο αρχικό βίντεο οι οποίες να μην έχουν αναγνωρισθεί στο απόσπασμα, και η άλλη να υπάρχουν μέχρι fa_t συνεχόμενες αλλαγές πλάνου στο απόσπασμα οι οποίες να έχουν αναγνωρισθεί χωρίς όμως να υπάρχουν στο αρχικό βίντεο. Η είσοδος για το πρόβλημα είναι μια σειρά από διάρκειες πλάνων $Q = q_1, q_2, \dots, q_k$.

Στο SELECT της ενότητας 3.7.1.1 ζητώνται από τη βάση όλες οι ακολουθίες διαρκειών που είναι ίσες με q_1, q_2, \dots, q_k .

- Για να υπάρχει ανοχή σε μέχρι fa_t εσφαλμένες αναγνωρίσεις, πρέπει να δημιουργηθούν όλα τα δυνατά αθροίσματα μέχρι fa_t συνεχόμενων q_i και κάθε ένα εξ αυτών να θεωρηθεί ως ξεχωριστό ερώτημα.
- Για να υπάρχει ανοχή σε μέχρι md_t χαμένες αναγνωρίσεις, θα πρέπει, στο ερώτημα προς τη βάση αντί να ελέγχεται μόνο η ισότητα $d_1 = q_1$, να ελέγχονται και τα αθροίσματα των πλάνων της βάσης, δηλαδή οι $d_1 + d_2 = q_1, \dots, \sum_{i=1}^{fa_t} d_i = q_1$, και αντίστοιχα για τα q_2, q_3, \dots, q_k .
- Τέλος, επειδή ενδιαφέρει και ο συνδυασμός των παραπάνω περιπτώσεων, για κάθε μια από τις περιπτώσεις με τα αθροίσματα συνεχόμενων πλάνων των εσφαλμένων αναγνωρίσεων, θα πρέπει να δημιουργηθούν όλοι οι συνδυασμοί με τα αθροίσματα στη βάση για τις τυχόν χαμένες αναγνωρίσεις.

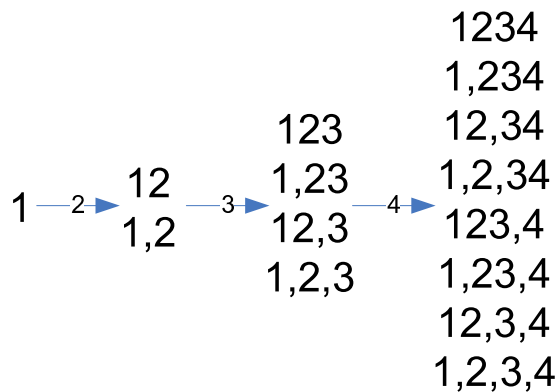
Έτσι λοιπόν, από την αρχική ακολουθία q_1, q_2, \dots, q_k αρχικά θα πρέπει να βρεθούν όλες οι δυνατές περιπτώσεις για την ανοχή σε fa_t εσφαλμένες αναγνωρίσεις, οι οποίες θα δημιουργήσουν μια σειρά από νέες ακολουθίες. Κάθε μια από αυτές τις νέες ακολουθίες, μαζί φυσικά με την αρχική θα δημιουργήσει μια σειρά από ερωτήματα προς τη βάση, από τα οποία θα δημιουργηθούν άλλα ερωτήματα, με ανοχή σε μέχρι md_t χαμένες αναγνωρίσεις στο απόσπασμα.

3.7.2.1 Ανοχή σε εσφαλμένες αναγνωρίσεις

Για την ακολουθία διαρκειών q_1, q_2, \dots, q_k θα πρέπει λοιπόν να δημιουργηθούν όλα τα δυνατά αθροίσματα μέχρι fa_t συνεχόμενων διαρκειών. Για την εξήγηση θα χρησιμοποιηθούν δύο βοηθητικές έννοιες: Η συνταγή (recipe) και ο συνδυασμός permutation. Μια συνταγή περιέχει όλους τους δυνατούς συνδυασμούς για ένα απόσπασμα με συγκεκριμένο αριθμό από διάρκειες. Ο συνδυασμός είναι μια λίστα από λίστες από ακεραίους, πχ $((1, 2), (3), (4, 5, 6))$. Το προηγούμενο σημαίνει ότι σε ένα απόσπασμα με 6 διάρκειες, ένας από τους δυνατούς συνδυασμούς συνεχόμενων διαρκειών είναι ένα νέο απόσπασμα με τρεις μόνο διάρκειες, όπου η πρώτη διάρκεια του νέου είναι ίση με το άθροισμα των δύο πρώτων του παλαιού, η δεύτερη διάρκεια του νέου είναι ίση με την τρίτη διάρκεια του παλαιού και η τρίτη διάρκεια του νέου είναι ίση με το άθροισμα των τριών τελευταίων του παλαιού. Ο τρόπος με τον οποίο δημιουργούνται

3.7. ΜΕΘΟΔΟΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

όλοι οι δυνατοί συνδυασμοί κάθε συνταγής φαίνεται στο σχήμα 3.16: Αρχικά υπάρχει ένας συνδυασμός μόνο με ένα πλάνο. Στη συνέχεια, προστίθεται ένα δεύτερο πλάνο, το οποίο μπορεί είτε να είναι ξεχωριστό από το πρώτο, είτε ενωμένο (στο σχήμα η ύπαρξη κόμματος σημαίνει ότι τα πλάνα είναι ξεχωριστά). Ομοίως, το τρίτο πλάνο μπορεί να ενωθεί με τα άλλα ή όχι. Έτσι, βαθμιαία δημιουργούνται όλοι οι δυνατοί συνδυασμοί κάθε συνταγής, μέχρι τον αριθμό διαρκειών του αποσπάσματος. Η συνταγή για ένα απόσπασμα με k διάρκειες, θα περιέχει ακριβώς 2^{k-1} διαφορετικούς συνδυασμούς.



Σχήμα 3.16: Δημιουργία συνδυασμών

Στη συνέχεια, πρέπει να γίνει ένα φιλτράρισμα αυτών των 2^{k-1} συνδυασμών, αφού κάποιοι συνδυασμοί μπορεί να περιέχουν πολύ μικρό αριθμό διαρκειών (μία ή δύο) και κάποιοι άλλοι μπορεί να έχουν αθροίσματα με παραπάνω από fa_t όρους. Έτσι, αυτοί οι συνδυασμοί θα κοπούν. Οι υπόλοιποι συνδυασμοί θα χρησιμοποιηθούν έτσι ώστε να δημιουργηθούν όλα τα δυνατά αθροίσματα.

Ως παράδειγμα, με ένα απόσπασμα που έχει τις διάρκειες (3000, 5000, 4000), χρησιμοποιώντας την τη τρίτη στήλη του σχήματος 3.16 το αποτέλεσμα θα είναι: (12000), (3000, 9000), (8000, 4000), (3000, 5000, 4000). Η περίπτωση με τα (12000) θα κοπεί από το φιλτράρισμα μιας και έχει πολύ λίγες διάρκειες.

3.7.2.2 Ανοχή σε χαμένες αναγνωρίσεις

Κάθε μια από τις παραπάνω περιπτώσεις που τελικά περνάει το φιλτράρισμα, θα χρησιμοποιηθεί ως ξεχωριστό ερώτημα από τον αλγόριθμο που δημιουργεί τα ερωτήματα για την ανοχή σε μέχρι md_t χαμένες αναγνωρίσεις. Η λύση στο πρόβλημα της δημιουργίας όλων των δυνατών περιπτώσεων όπου μέχρι md_t συνεχόμενα πλάνα από κάποιο βίντεο της βάσης δεδομένων θεωρούνται ως ένα ενιαίο πλάνο από το απόσπασμα μεγέθους q σκηνών μπορεί να βρεθεί σχετικά εύκολα: Είναι όλοι οι αριθμοί με q ψηφία σε αριθμητικό σύστημα με βάση το $fa_t + 1$. Κάθε αριθμός είναι μια διαφορετική ερώτηση στη βάση, όπου η τιμή του κάθε ψηφίου του καθορίζει το πόσες συνεχόμενα πλάνα πρέπει να υπάρχουν στη βάση των οποίων το άθροισμα να είναι ίσο με την αντίστοιχη διάρκεια του

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

αποσπάσματος.

Το παραπάνω θα γίνει κατανοητό με μερικά παραδείγματα. Για ένα απόσπασμα με 3 διάρκειες q_1, q_2, q_3 . και ανοχή σε μέχρι 1 εσφαλμένη αναγνώριση, η λύση είναι όλοι οι αριθμοί με 3 ψηφία σε αριθμητικό σύστημα με βάση το 2, δηλαδή οι αριθμοί 000, 001, 010, 011, 100, 101, 110, 111. Το 000 σημαίνει ότι θα αναζητηθούν στη βάση δεδομένων οι διάρκειες $d_1 = q_1, d_2 = q_2, d_3 = q_3$. Το 001 σημαίνει ότι θα αναζητηθούν στη βάση δεδομένων οι διάρκειες $d_1 = q_1, d_2 = q_2, d_{31} + d_{32} = q_3$, παρομοίως για τα υπόλοιπα μέχρι το 111 που σημαίνει ότι θα αναζητηθούν οι $d_{11} + d_{12} = q_1, d_{21} + d_{22} = q_2, d_{31} + d_{32} = q_3$. Για ένα απόσπασμα με 2 διάρκειες q_1, q_2 και ανοχή σε μέχρι 3 εσφαλμένες αναγνώρισεις η λύση είναι οι αριθμοί με 2 ψηφία σε αριθμητικό σύστημα με βάση το 4, δηλαδή οι αριθμοί 00, 01, 02, 03, 10, 11, 12, 13, 20, 21, 22, 23, 30, 31, 32, 33 όπου το 13 σημαίνει ότι στη βάση δεδομένων θα υπάρχουν δύο διάρκειες d_{11}, d_{12} ίσες με q_1 ακολουθούμενες από άλλες τέσσερις διάρκειες $d_{21}, d_{22}, d_{23}, d_{24}$ ίσες με q_2 και παρομοίως για τα υπόλοιπα.

Προφανώς, ο αριθμός των διαφορετικών περιπτώσεων που θα δημιουργηθούν για ένα απόσπασμα με q σκηνές και ανοχή σε md_t χαμένες αναγνώρισεις είναι ίσος με md_t^q .

Συνεχίζοντας λοιπόν το παράδειγμα της προηγούμενης ενότητας, έχουν μείνει οι διάρκειες (3000, 9000), (8000, 4000) και (3000, 5000, 4000). Για ανοχή σε μια το πολύ χαμένη αναγνώριση, οι αριθμοί είναι 00, 01, 10, 11, για τα δύο πρώτα ερωτήματα και 000, 001, 010, 011, 100, 101, 110, 111 για το τρίτο.

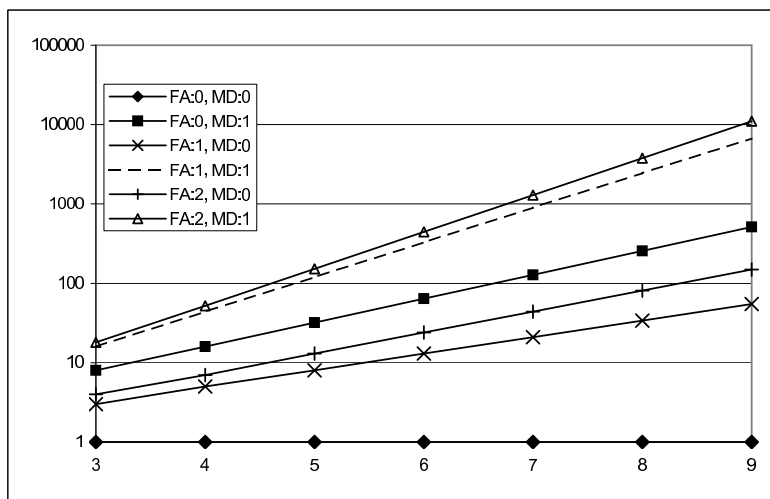
Έτσι, για τα δύο πρώτα ερωτήματα, θα δημιουργηθούν από τέσσερις ερωτήσεις στη βάση, ενώ για το τρίτο θα δημιουργηθούν οκτώ, συνολικά δηλαδή θα δημιουργηθούν 16 ερωτήματα. Κάθε μια εξ αυτών θα είναι παρόμοια με το ερώτημα της 3.7.1.1, απλώς θα υπάρχουν τα αθροίσματα. Ως παράδειγμα, θα γραφεί εδώ το ερώτημα για τα (3000, 9000) και για τον αριθμό 11:

```
SELECT d00.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ, d00.ΧΡΟΝΟΣ_ΑΡΧΗΣ FROM
ΑΛΛΑΓΗ_ΠΛΑΝΟΥ d00, ΑΛΛΑΓΗ_ΠΛΑΝΟΥ d01,
ΑΛΛΑΓΗ_ΠΛΑΝΟΥ d10, ΑΛΛΑΓΗ_ΠΛΑΝΟΥ d11,
WHERE
d00.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ = d01.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ AND
d00.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ = d10.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ AND
d00.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ = d11.ΚΛΕΙΔΙ_ΒΙΝΤΕΟ AND
d00.ΔΙΑΡΚΕΙΑ + d01.ΔΙΑΡΚΕΙΑ = 3000 AND
d10.ΔΙΑΡΚΕΙΑ + d11.ΔΙΑΡΚΕΙΑ = 9000 AND
d00.ΧΡΟΝΟΣ_ΑΡΧΗΣ+d01.ΔΙΑΡΚΕΙΑ = d01.ΧΡΟΝΟΣ_ΑΡΧΗΣ AND
d01.ΧΡΟΝΟΣ_ΑΡΧΗΣ+d10.ΔΙΑΡΚΕΙΑ = d10.ΧΡΟΝΟΣ_ΑΡΧΗΣ AND
d10.ΧΡΟΝΟΣ_ΑΡΧΗΣ+d11.ΔΙΑΡΚΕΙΑ = d11.ΧΡΟΝΟΣ_ΑΡΧΗΣ
```

3.7. ΜΕΘΟΔΟΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

3.7.2.3 Αριθμός περιπτώσεων

Όπως μπορεί να γίνει κατανοητό από τα παραπάνω, ιδιαίτερα όταν υπάρχει ανοχή σε χαμένες αναγνωρίσεις πλάνων, θα δημιουργηθεί ένας τεράστιος αριθμός από διαφορετικά ερωτήματα προς τη βάση. Στο σχήμα 3.17 φαίνεται ο αριθμός των ερωτημάτων που θα δημιουργηθούν για ένα απόσπασμα με 3 ως 9 πλάνα (δηλαδή 5 ως 11 αφού το πρώτο και το τελευταίο πλάνο δε λαμβάνονται υπό όψιν μιας και δεν είναι γνωστή η διάρκεια τους, άρα 4 ως 10 αλλαγές πλάνου) και για διάφορες τιμές της ανοχής σε χαμένες αναγνωρίσεις (MD) και ανοχής σε εσφαλμένες αναγνωρίσεις (FA).



Σχήμα 3.17: Αριθμός ερωτημάτων που δημιουργούνται

Ο αριθμός των ερωτημάτων που δημιουργείται όταν υπάρχει ανοχή σε χαμένες αναγνωρίσεις (MD) αυξάνει πάρα πολύ, ειδικά για απόσπασμα με πολλές αλλαγές πλάνου. Έτσι, στην περίπτωση που ένα απόσπασμα έχει πολλές αλλαγές πλάνου και πρέπει να υπάρχει ανοχή σε χαμένες αναγνωρίσεις, είναι σημαντικό να δημιουργηθούν ερωτήματα μόνο με 3 ή 4 από τα πλάνα που έχει το απόσπασμα, έτσι ώστε να μη δημιουργηθεί τεράστιος αριθμός ερωτημάτων και η αναζήτηση κρατήσει υπερβολικό χρόνο. Επίσης, μπορεί να υπάρχει πρόβλεψη έτσι ώστε να μειωθεί όσο γίνεται περισσότερο ο αριθμός των χαμένων πλάνων στο απόσπασμα, για παράδειγμα, αν χρησιμοποιηθεί η μέθοδος της ενότητας 2.4, μπορεί να επιλεγεί το κατώφλι των *intra macroblock* για την απόφαση για αλλαγή πλάνου πολύ χαμηλό στο απόσπασμα, έτσι ώστε να υπάρχει η διασφάλιση ότι θα βρεθούν όλες οι αλλαγές πλάνου στο απόσπασμα, έστω και αν βρεθούν και κάποιες που δεν υπάρχουν στην πραγματικότητα.

3.7.3 Γενικά για τη μέθοδο βάσης δεδομένων

Η μέθοδος που παρουσιάστηκε στα προηγούμενα είναι συμπληρωματική της πιθανοτικής μεθόδου της ενότητας 3.6. Και οι δύο βασίζονται σε αλλαγές

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

πλάνων (ή αλλιώς συμβάντα), σε ακολουθίες από αλλαγές πλάνων και μη η πιθανοτική μέθοδος, σε ακολουθίες από διάρκειες πλάνων η μέθοδος βάσης δεδομένων. Ουσιαστικά, η μέθοδος αυτή, όπως ήδη αναφέρθηκε, είναι η υλοποίηση της πιθανοτικής μεθόδου σε ένα σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων.

Με τη μέθοδο βάσης δεδομένων το πρόβλημα του jitter επιλύεται αρκετά εύκολα, αντίθετα όμως, για να υπάρχει ανοχή σε εσφαλμένες αναγνώρισεις, και πολύ περισσότερο σε χαμένες αναγνώρισεις αλλαγών πλάνου στο απόσπασμα είναι απαραίτητη η δημιουργία πολλών διαφορετικών ερωτημάτων προς τη βάση δεδομένων, κάτι που θα υποβαθμίζει την ταχύτητα αναζήτησης της μεθόδου αυτής. Το σημαντικότερο χαρακτηριστικό της μεθόδου της βάσης δεδομένων είναι ότι δημιουργούνται μια σειρά από ερωτήματα SQL τα οποία μπορεί να υποβληθούν σε μια βάση δεδομένων και έτσι η αναζήτηση να γίνει αποκλειστικά από τη βάση δεδομένων.

3.8 Μέθοδος intra macroblock

Η τρίτη μέθοδος αναζήτησης πρώτου επιπέδου που θα αναλυθεί είναι η αναζήτηση με βάση των αριθμό των intra macroblock κάθε εικόνας, η οποία χρησιμοποιεί το ίδιο χαρακτηριστικό με τη μέθοδο ανίχνευσης αλλαγών πλάνου της ενότητας 2.4. Στα επόμενα, αρχικά θα γίνει μια γενική περιγραφή του τρόπου λειτουργίας αυτής της μεθόδου, στη συνέχεια μια αναφορά στη λειτουργία της όταν υπάρχουν αλλαγές στο βίντεο και τέλος θα αναφερθούν μια σειρά από υποπεριπτώσεις αυτής της μεθόδου που βελτιώνουν ακόμα περισσότερο την αναζήτηση.

3.8.1 Περιγραφή

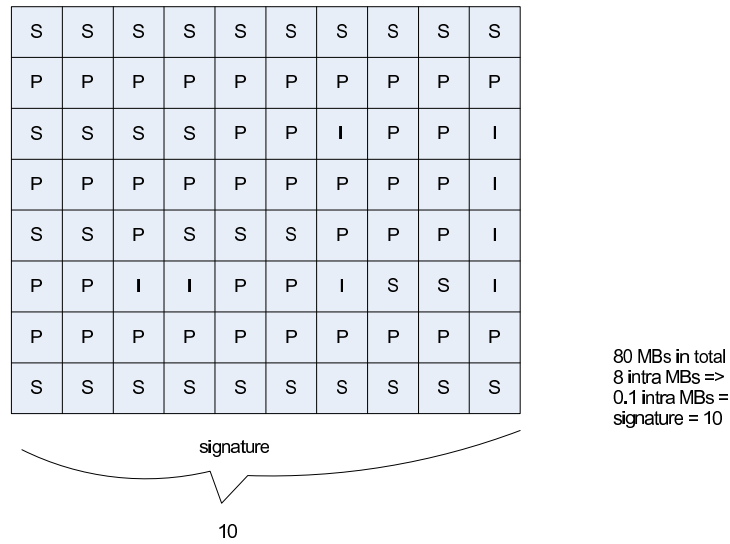
Η λογική αυτής της μεθόδου είναι αρκετά απλή: Από κάθε εικόνα ενός κατάλληλα κωδικοποιημένου H.264 βίντεο εξάγεται το ποσοστό των macroblock που είναι κωδικοποιημένα κατά intra και στρογγυλοποιείται ώστε να γίνει ακέραιος. Αυτός ο ακέραιος αριθμός από 0 ως 100 είναι και η υπογραφή της συγκεκριμένης εικόνας, όπως μπορεί να φανεί στο σχήμα 3.18 ενώ μια ακολουθία τέτοιων αριθμών είναι η υπογραφή του βίντεο. Ο τρόπος εξαγωγής της υπογραφής του βίντεο παρουσιάζεται στο σχήμα 3.19.

Η σύγκριση μεταξύ των υπογραφών δύο βίντεο γίνεται με τη βοήθεια της intra-mb απόστασης τους, η οποία δίνεται από τον ακόλουθο ορισμό:

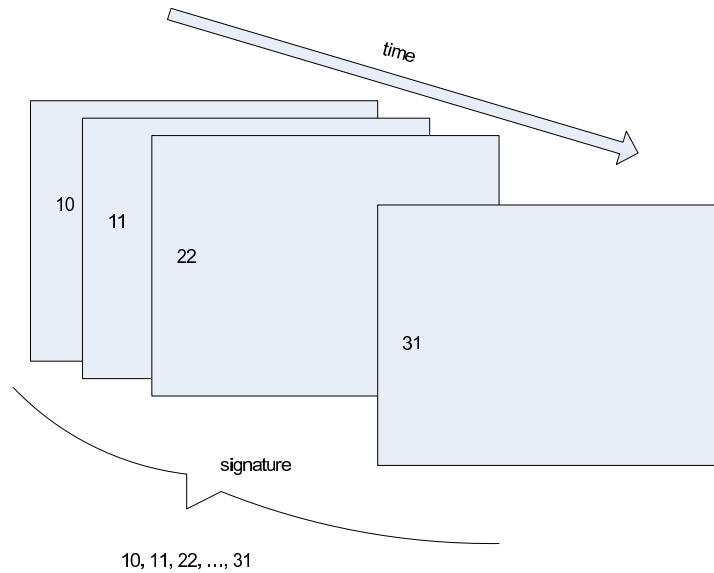
Ορισμός 3.8.1 Έστω δύο αποσπάσματα P και Q ίδιου μήκους k , όπου το P έχει ως υπογραφή την ακολουθία $\{p_1, p_2, \dots, p_k\}$ και το Q την $\{q_1, q_2, \dots, q_k\}$. Η intra-mb απόσταση αυτών D_{intra} ορίζεται ως

$$D_{intra} = \frac{\sum_{i=1}^k (p_i - q_i)^2}{k}.$$

3.8. ΜΕΘΟΔΟΣ INTRA MACROBLOCK



Σχήμα 3.18: Εξαγωγή υπογραφής εικόνας με intra mb



Σχήμα 3.19: Εξαγωγή υπογραφής βίντεο με intra mb

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

Όπως μπορεί να γίνει κατανοητό, το άθροισμα της D_{intra} διαιρείται με τον αριθμό των εικόνων των αποσπασμάτων προς κανονικοποίηση.

Εκτός αυτής της απλής μεθόδου, προτείνεται και μια παραλλαγή της, στην οποία κάθε εικόνα ενός βίντεο χωρίζεται σε n_x οριζόντια και n_y κάθετα τμήματα, δηλαδή συνολικά η εικόνα χωρίζεται σε $N_{xy} = n_x \times n_y$ παράθυρα. Από κάθε ένα από αυτά τα παράθυρα εξάγεται ομοίως το ποσοστό των intra macroblock και έτσι η υπογραφή κάθε εικόνας αντί για έναν ακέραιο αριθμό είναι ένα διάνυσμα από $N_{xy} = n_x \times n_y$ ακέραιους αριθμούς. Η σύγκριση δύο τέτοιων υπογραφών γίνεται με το μέγεθος που ορίζεται ακολούθως:

Ορισμός 3.8.2 Έστω δύο αποσπάσματα P και Q ίδιου μήκους k , όπου το P έχει ως υπογραφή την ακολουθία $\{p_1, p_2, \dots, p_k\}$ και το Q την $\{q_1, q_2, \dots, q_k\}$. Η πολλαπλή intra-mb απόσταση αυτών $D_{intra, N_{xy}}$ ορίζεται ως

$$D_{intra, N_{xy}} = \frac{\sum_{i=1}^k \sum_{y=1}^{n_y} \sum_{x=1}^{n_x} (p_{i,x,y} - q_{i,x,y})^2}{N_{xy}k}.$$

Ομοίως και εδώ το άθροισμα διαιρείται με το $N_{xy}k$ έτσι ώστε το $D_{intra, N_{xy}}$ να είναι κανονικοποιημένο και ανεξάρτητο από τον αριθμό των εικόνων του αποσπάσματος και τον αριθμό των παραθύρων στα οποία χωρίστηκε η εικόνα.

Και στις δύο περιπτώσεις, για τα βίντεο της βάσης δεδομένων δημιουργείται μια μεγάλη ακολουθία από αριθμούς. Αυτή η ακολουθία μπορεί είτε να αποθηκευθεί σε βάση δεδομένων είτε σε κανονικά αρχεία στο σκληρό δίσκο του υπολογιστή. Αναφορικά με την περίπτωση της βάσης δεδομένων, παρουσιάζεται και εδώ το πρόβλημα που αναφέρθηκε στην ενότητα 3.7.1.1, ότι δηλαδή είναι δύσκολο σε μια βάση να ελεγχθεί μια ακολουθία από πολλές τιμές, όπως δηλαδή συμβαίνει και με την πιθανοτική μέθοδο. Έτσι, ακόμα και αν χρησιμοποιηθεί βάση δεδομένων, θα πρέπει, σειριακά, οι υπογραφές όλων των βίντεο να φορτώνονται στη μνήμη και να γίνεται ο έλεγχος, παρόμοια με το αν γινόταν χρήση απλών αρχείων, οπότε, οι τεχνικές αναζήτησης που θα προταθούν στα επόμενα θα βασιστούν σε αποθήκευση των βίντεο της βάσης δεδομένων σε απλά αρχεία στο δίσκο.

3.8.2 Δυσκολίες που παρουσιάζονται σε αλλαγμένα βίντεο

Ένα από τα πολύ σημαντικά χαρακτηριστικά που οφείλουν να έχουν οι μέθοδοι αναζήτησης που προτείνονται στην εργασία αυτή είναι το ότι πρέπει να είναι εύρωστες (robust) σε τυχόν αλλαγές που μπορεί να υποστεί ένα βίντεο. Οι αλλαγές αυτές είναι συνήθως είναι διαφορετική συμπίεση άρα και ποιότητα³, αλλαγή της ανάλυσης καθώς και διάφορες οπτικές αλλαγές, για παράδειγμα προσθήκη υποτίτλων, λογοτύπων κοκ. Παρακάτω θα γίνει μια αναφορά στη συμπεριφορά της μεθόδου των intra macroblock σε αυτές τις περιπτώσεις και

³Συνήθως η ποιότητα υποβαθμίζεται για να βελτιωθεί το ποσοστό συμπίεσης

3.8. ΜΕΘΟΔΟΣ INTRA MACROBLOCK

λύσεις για τα τυχόν προβλήματα που μπορεί να δημιουργηθούν κατά την αναζήτηση. Οι μέθοδοι που παρουσιάστηκαν στις προηγούμενες ενότητες βασίζονται στις αλλαγές πλάνου, και έχει ήδη αποδειχθεί στο κεφάλαιο 2 ότι η επίδοση των τεχνικών ανίχνευσης αλλαγής πλάνου είναι αρκετά εύρωστες.

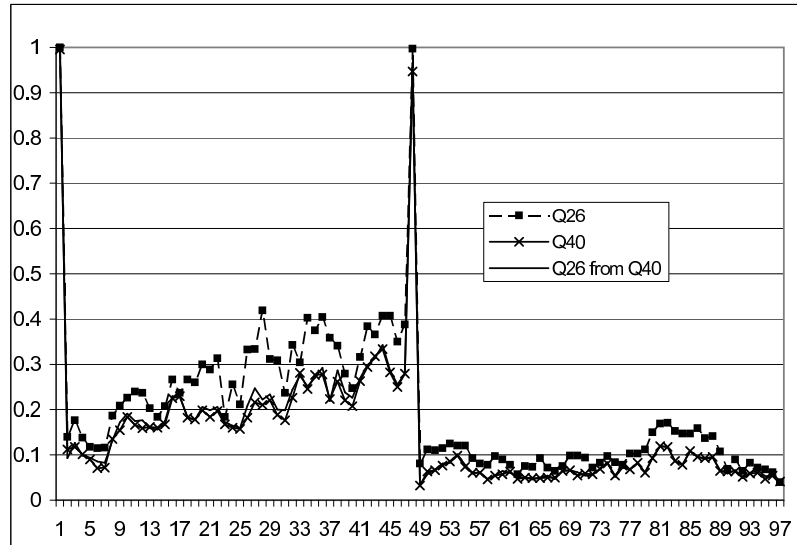
3.8.2.1 Αλλαγή συμπίεσης/ποιότητας

Μια πολύ συχνή αλλαγή σε ένα βίντεο είναι η επανασυμπίεση του με διαφορετικό πρωτόκολλο ή η συμπίεση του με το ίδιο πρωτόκολλο αλλά διαφορετικές παραμέτρους συμπίεσης. Αυτό έχει ως αποτέλεσμα ουσιαστικά να αλλάξει η ποιότητα και τα χαρακτηριστικά του βίντεο. Προφανώς, η ποιότητα ενός βίντεο μπορεί μόνο να υποβαθμιστεί όταν αυτό επανακωδικοποιηθεί, αφού οι αλγόριθμοι συμπίεσης είναι απωλεστικοί και δε μπορεί να δημιουργηθεί πληροφορία που έχει χαθεί. Παρακάτω θα γίνει μια απλή αναφορά στο θέμα αναφορικά κυρίως με το συντελεστή χβαντισμού Q ενός MPEG βίντεο. Προφανώς υπάρχουν και δεκάδες άλλες παράμετροι οι οποίες αλλάζουν την ποιότητα ενός βίντεο οι οποίες όμως δεν είναι δυνατόν να αναπτυχθούν στα πλαίσια αυτής της εργασίας. Σε γενικές γραμμές η συμπεριφορά ενός βίντεο όταν αλλάζουν αυτές οι παράμετροι αναμένεται να είναι παρόμοια με την αλλαγή του Q , αφού η τελική ποιότητα θα είναι είτε καλύτερη, είτε χειρότερη. Το κύριο ενδιαφέρον για την ανάλυση της ενότητας έχει να κάνει με το πόσο αλλάζει το ποσοστό των intra macroblocks κάθε εικόνας όταν αλλάζει η συμπίεση.

Δηλαδή, μετά την επανασυμπίεση ενός βίντεο υπάρχουν δύο περιπτώσεις: Ένα απόσπασμα συμπιέζεται με παραμέτρους καλύτερης ποιότητας / μικρότερης συμπίεσης από την αρχική είτε με παραμέτρους χειρότερης ποιότητας / μεγαλύτερης συμπίεσης από την αρχική. Στην πρώτη περίπτωση, το βίντεο ουσιαστικά θα παραμείνει το ίδιο, αφού δε μπορεί να δημιουργηθεί παραπάνω πληροφορία. Μπορεί να αλλάξει φυσικά ο τρόπος κωδικοποίησής του, δηλαδή κάποια macroblock να κωδικοποιηθούν με διαφορετικό τρόπο (αφού λόγω των διαφορετικών παραμέτρων συμπίεσης αλλάζουν και τα εργαλεία που έχει ο κωδικοποιητής στη διάθεση του για να κάνει τη συμπίεση). Έτσι, σε αυτή την περίπτωση μπορεί να θεωρηθεί ότι το ποσοστό των intra macroblock θα αλλάξει λίγο ή καθόλου, αφού παρότι θα υπάρχει καλύτερη ποιότητα, ο κωδικοποιητής αναγκαστικά θα επιλέξει στα περισσότερα macroblock τον ίδιο τρόπο συμπίεσης με τη χειρότερη ποιότητα αφού αν υπήρχε κάπου καλύτερο ταίριασμα θα το είχε ήδη επιλέξει η πρώτη περίπτωση (αρκεί φυσικά να είναι ο ίδιος κωδικοποιητής). Στη δεύτερη περίπτωση όμως, λόγω της χειρότερης ποιότητας / μεγαλύτερης συμπίεσης, το βίντεο θα αλλάξει, πολύ ή λίγο, ανάλογα με το πόσο αλλάζουν οι παράμετροι. Έτσι, είναι αναμενόμενο εδώ, λόγω της καλύτερης ποιότητας, ο κωδικοποιητής να επιλέξει να κωδικοποιήσει κάποια macroblock διαφορετικά, και συνήθως κάποια inter θα γίνουν intra αφού τα intra έχουν καλύτερη ποιότητα από τα inter και προτιμώνται σε περιοχές μεγάλης πολυπλοκότητας.

Για να εξεταστεί καλύτερη αυτή η συμπεριφορά, χρησιμοποιήθηκε ένα απόσπασμα το οποίο αρχικά είχε συμπειστεί με $Q = 26$ στη συνέχεια μετατραπεί σε

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ



Σχήμα 3.20: Σύγκριση των ποσοστών intra macroblock τριών ίδιων αποσπασμάτων όταν επανασυμπίεστούν με διαφορετική ποιότητα Q

$Q = 40$ και τέλος μετατράπηκε ξανά σε $Q = 26$. Τα αποτελέσματα φαίνονται στην εικόνα 3.20, όπου με τη διακεκομμένη γραμμή με τα τετράγωνα είναι τα ποσοστά για το αρχικό βίντεο με $Q = 26$, με τη γραμμή με τα \times για $Q = 40$ και με την απλή γραμμή τα ποσοστά για $Q = 26$ όταν όμως έγινε επανακωδικοποίηση από το $Q = 40$. Η εικόνα επιβεβαιώνει λίγο πολύ τη συζήτηση της παραπάνω παραγράφου: Όταν η ποιότητα χειροτερεύει τότε το ποσοστό των intra macroblock μειώνεται (διαφορές $Q26$ με $Q40$), ενώ όταν η ποιότητα καλυτερεύει, το ποσοστό των intra macroblock δεν αλλάζει σχεδόν καθόλου (διαφορές $Q26$ from $Q40$ με $Q40$ και $Q26$). Πρέπει να σημειωθεί ότι η διαφορά του $Q26$ με το $Q40$ είναι πολύ μεγάλη⁴ και στη συνήθη περίπτωση δε θα υπάρχουν τόσο μεγάλες διαφορές. Ακόμα και σε αυτή την περίπτωση όμως, μπορεί να φανεί ότι το ποσοστό των intra macroblock έχει τις ίδιες διακυμάνσεις και φυσικά οι αλλαγές πλάνου ξεχωρίζουν πολύ. Παρόλα αυτά, αυτή η διαφορά στα ποσοστά των intra macroblock ίσως δυσκολέψει την εύρεση ενός αποσπάσματος όταν αυτό έχει αρκετά άσχημη ποιότητα. Δύο λύσεις μπορούν να χρησιμοποιηθούν για τη βελτίωση της αναζήτησης σε αυτή την περίπτωση:

- Η πρώτη αφορά τη μεγαλύτερη ελαστικότητα κατά την αναζήτηση όταν ανιχνευθεί ότι συγκρίνονται δύο βίντεο που διαφέρουν πολύ στην ποιότητα τους. Αναμένεται ότι παρότι η διαφορά των δύο βίντεο θα είναι σχετικά μεγάλη με το αν είχαν την ίδια ποιότητα, θα είναι σίγουρα μικρότερη από τη διαφορά δύο άσχετων μεταξύ τους βίντεο, οπότε εφόσον τα κριτήρια επιλογής γίνουν πιο ελαστικά το σωστό βίντεο θα μπορεί να βρεθεί (ίσως με τη μικρή μείωση της ακρίβειας αφού θα υπάρχει η περίπτωση να

⁴Το μέγεθος του αρχείου δοκιμής εξαπλασιάστηκε!

3.8. ΜΕΘΟΔΟΣ INTRA MACROBLOCK

επιλέγονται περισσότερα βίντεο λόγω της χαλάρωσης των κριτηρίων).

- Η δεύτερη λύση έχει να κάνει με την επιλογή ενός αποδεκτού Q για τα βίντεο που θα εισάγονται στη βάση, το οποίο και θα πρέπει να είναι αρκετά μεγάλο - δηλαδή χειρότερη ποιότητα (περισσότερο από 40). Έτσι, αν κάποιο απόσπασμα έχει καλύτερη ποιότητα, θα μπορεί να επανασυμπίεστεί με υποβάθμιση της ποιότητας ώστε να ταιριάζει με την ποιότητα βάσης και να μπορεί να γίνει η σύγκριση. Εδώ φυσικά θα πρέπει να γίνεται επανακωδικοποίηση των βίντεο πριν εισαχθούν στη βάση, ακόμα και αν είναι H.264 κατάλληλης μορφής κάτι που μπορεί να είναι ή όχι αποδεκτό, ανάλογα με τη λειτουργία του συστήματος.
- Μπορεί να προταθούν μια σειρά από βελτιώσεις στον τρόπο που βρίσκεται η intra-mb απόσταση (ορισμός 3.8.1) οι οποίοι θα εξετάζουν για παράδειγμα τις μεταβολές των ποσοστών αντί για τις διαφορές του, ή θα βρίσκουν κάποιο μέσο όρο διαφορών ανά τμήμα μεταξύ δύο αλλαγών πλάνου τον οποίο να αφαιρούν η προσθέτουν στα ποσοστά έτσι ώστε να τα φέρουν πιο κοντά.

Είναι ενδιαφέρον να εξηγηθεί εδώ γιατί θα πρέπει να βρεθεί ο μέσος όρος της διαφοράς για κάθε τμήμα μεταξύ δύο αλλαγών πλάνου του βίντεο και όχι γενικά για ολόκληρο το απόσπασμα: Αυτό έχει να κάνει με το ότι η πολυπλοκότητα ενός βίντεο συνήθως αλλάζει μεταξύ αρκετά μεταξύ των αλλαγών πλάνου. Μπορεί να αλλάξει και εντός του πλάνου αλλά κάτι τέτοιο δε συμβαίνει πολύ συχνά και επίσης η εύρεση αυτών των περιοχών υψηλής και χαμηλότερης πολυπλοκότητας στο ίδιο πλάνο είναι αρκετά δύσκολη. Αντίθετα, μεταξύ των πλάνων υπάρχουν οι αλλαγές πλάνου να βοηθήσουν. Είναι προφανές ότι σε περιοχές υψηλής πολυπλοκότητας ο αριθμός των intra macroblock θα είναι μεγαλύτερος από ότι σε περιοχές χαμηλής πολυπλοκότητας και θα αυξάνει ακόμα περισσότερο για καλύτερη ποιότητα.

Αυτό το ζήτημα φαίνεται και από το παράδειγμα του σχήματος 3.20, όπου στην πρώτη περιοχή ο μέσος όρος των διαφορών για $Q = 26$ από το $Q = 40$ είναι αρκετά μεγαλύτερος από τη δεύτερη περιοχή. Γενικά αυτή η λύση δε θα εξεταστεί εκτενώς στα πλαίσια αυτής της εργασίας, μιας και αυξάνει αρκετά την πολυπλοκότητα αναζήτησης και υλοποίησης των μεθόδων, αλλά θα ελεγχθεί η αποτελεσματικότητά της στο μέλλον.

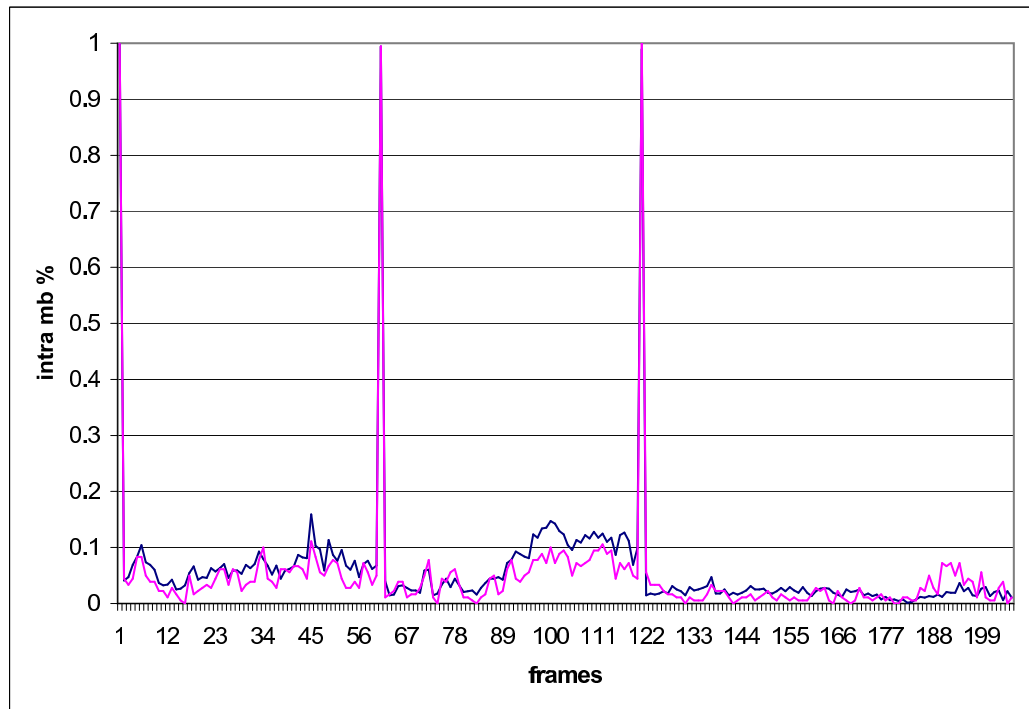
Το πόσο καλά δουλεύουν οι δύο πρώτες λύσεις θα εξεταστεί αναλυτικά στο μεθεπόμενο κεφάλαιο όπου και παρουσιάζονται τα πειραματικά αποτελέσματα.

3.8.2.2 Αλλαγή ανάλυσης

Η αλλαγή ανάλυσης παρότι αλλάζει αρκετά τον αριθμό των intra macroblock δεν αλλάζει τόσο πολύ το ποσοστό τους έτσι ώστε να μη δουλεύει πλέον

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

η μέθοδος των intra macroblock. Αυτός άλλωστε είναι και ο λόγος που η υπογραφή των βίντεο έχει να κάνει με τα ποσοστά των intra macroblock σε κάθε εικόνα και όχι με τον απόλυτο αριθμό τους, αφού ο αριθμός των macroblock αλλάζει με την ανάλυση⁵ Αυτό έχει ήδη φανεί στην ενότητα 2.6, όπου παρότι άλλαξε η ανάλυση πάρα πολύ τα αποτελέσματα ανίχνευσης αλλαγής πλάνου παρέμειναν αρκετά καλά. Αυτό το συμπέρασμα μπορεί να επιβεβαιωθεί και με το παράδειγμα της εικόνας 3.21, όπου παρότι η ανάλυση έχει μειωθεί στο μισό στο ένα βίντεο, τα ποσοστά των intra macroblock παραμένουν πολύ κοντά το ένα στο άλλο και ακολουθούν τις διακυμάνσεις. Όπως ήταν αναμενόμενο, στις αλλαγές πλάνου και στις δύο περιπτώσεις υπάρχει άνοδος στο 100%, ενώ σε κάποια σημεία παίρνει μεγαλύτερες τιμές το απόσπασμα με την κανονική ανάλυση και σε κάποιες άλλες το απόσπασμα με τη μισή ανάλυση. Αυτό είναι τελείως τυχαίο και εξαρτάται από τον τρόπο που λειτουργεί ο κωδικοποιητής και το πώς καταφέρνει να ταιριάζει τα macroblock της τρέχουσας εικόνας με αυτά της προηγούμενης.



Σχήμα 3.21: Σύγκριση των ποσοστών intra macroblock δύο ίδιων αποσπασμάτων όταν το ένα έχει τη μισή ανάλυση

⁵Ως γνωστόν ένα βίντεο ανάλυσης $x \times y$ θα αποτελείται από $\lceil x/16 \rceil \times \lceil y/16 \rceil$ macroblock

3.8.2.3 Οπτικές αλλαγές

Οι πιο συνηθισμένες οπτικές αλλαγές που μπορεί να γίνουν στα βίντεο είναι η προσθήκη λογοτύπων ή υποτίτλων. Μπορεί εύκολα να γίνει κατανοητό ότι ούτε τα λογότυπα ούτε και οι υπότιτλοι επηρεάζουν ουσιαστικά τη μέθοδο:

Τα λογότυπα είναι συνήθως στατικά και καταλαμβάνουν ένα πολύ μικρό τμήμα της οθόνης. Δεν αλλάζουν από τη μία εικόνα στην άλλη, άρα δε θα επηρεάζουν τον αριθμό των intra macroblock. Ακόμα όμως και αν υπάρχει κίνηση σε αυτά ή αν κάποιο λογότυπο καλύπτει μια κίνηση, η διαφορά αυτή στα intra macroblock μεταξύ των δύο βίντεο θα είναι πολύ μικρή και τελικά δε θα επηρεάσει την αναζήτηση. Σε παρόμοια αποτελέσματα κατέληξαν και οι δοκιμές της ενότητας 2.6, όπου η εισαγωγή ενός πολύ μεγάλου λογότυπου δεν επηρέασε σχεδόν καθόλου τα αποτελέσματα της μεθόδου ανίχνευσης αλλαγών πλάνου που βασίστηκε στο ποσοστό των intra macroblock ανά εικόνα.

Παρόμοια πράγματα ισχύουν και για τους υποτίτλους. Εδώ, ανάλογα φυσικά και με το μέγεθος των υποτίτλων, επηρεάζονται περισσότερο τα ποσοστά των intra macroblock αφού όταν εμφανιστούν οι νέοι υπότιτλοι είναι αναπόφευκτο κάποια macroblock να αλλάξουν αρκετά έτσι ώστε να θεωρηθούν ως intra. Για να δοκιμαστεί η ευρωστία κατά την προσθήκη υποτίτλων, σε ένα βίντεο προστέθηκαν αρχικά υπότιτλοι κανονικού μεγέθους και στη συνέχεια μεγαλύτερου από το συνηθισμένο μέγεθος (διπλάσιοι του κανονικού) όπως φαίνεται και στην εικόνα 3.22 και εξετάστηκε το ποσοστό των intra macroblock για κάθε μια από τις τρεις παραλλαγές. Τα ποσοστά των σχήμα των intra macroblock για την περίπτωση αυτή φαίνονται στο σχήμα 3.23. Εύκολα φαίνεται ότι δεν υπάρχει σχεδόν καμία διαφορά και στις τρεις περιπτώσεις, ακόμα και στην περίπτωση που υπάρχουν οι αρκετά μεγάλοι υπότιτλοι.

Έτσι, από τα παραπάνω μπορεί να εξαχθεί ότι η μέθοδος των intra macroblock είναι αρκετά εύρωστη στις συνηθισμένες οπτικές αλλαγές που μπορεί να υποστεί ένα βίντεο, τόσο δηλαδή στην προσθήκη λογοτύπων όσο και στην προσθήκη υποτίτλων.

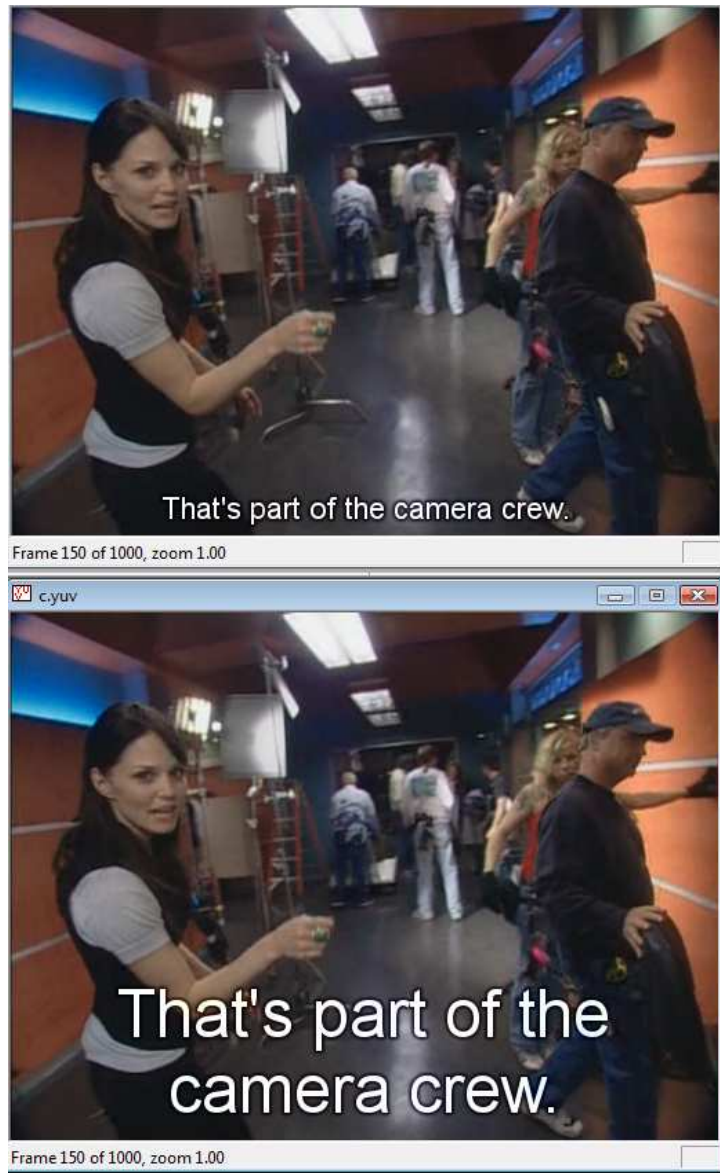
3.8.3 Αναζήτηση με τη μέθοδο intra macroblock

Προτείνονται τέσσερις τρόποι αναζήτησης με βάση την υπογραφή που προτάθηκε παραπάνω: Η απλή αναζήτηση και η απλή αναζήτηση με έλεγχο διακοπής, στις οποίες δε χρειάζεται να δημιουργηθούν κάποιοι παραπάνω δείκτες από την υπογραφή κάθε μιας εικόνας και οι μέθοδοι αναζήτησης μεγάλων τιμών και διαστημάτων, στις οποίες όμως πρέπει να δημιουργηθούν επιπλέον δείκτες εκτός της υπογραφής με τα intra macroblock.

3.8.3.1 Απλή αναζήτηση

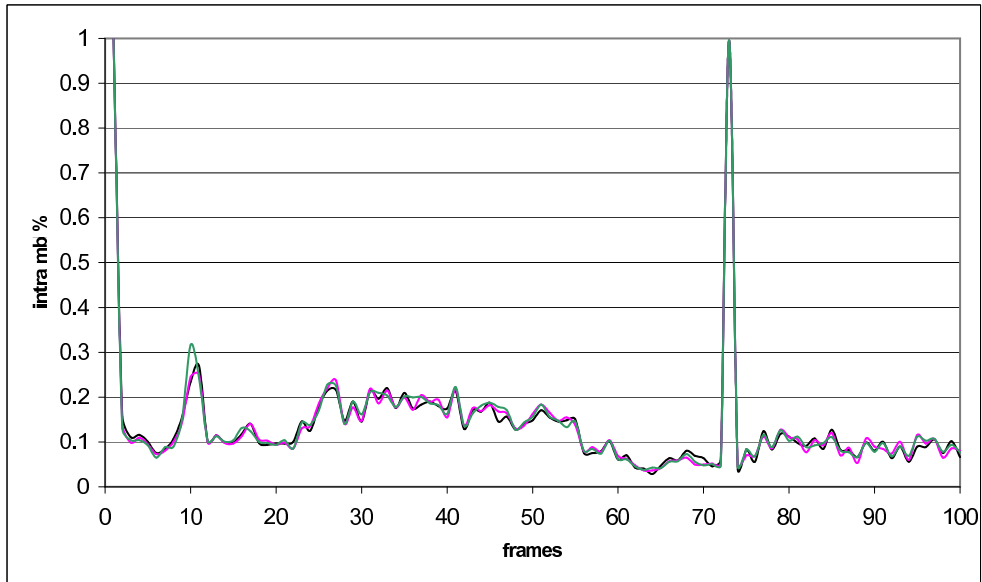
Για την απλή αναζήτηση το απόσπασμα ολισθαίνει σε όλες τις δυνατές θέσεις των βίντεο της βάσης δεδομένων με τον τρόπο που φαίνεται και στην εικόνα 3.24. Για κάθε δυνατή θέση εξάγεται η intra-mb απόσταση D_{intra} .

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

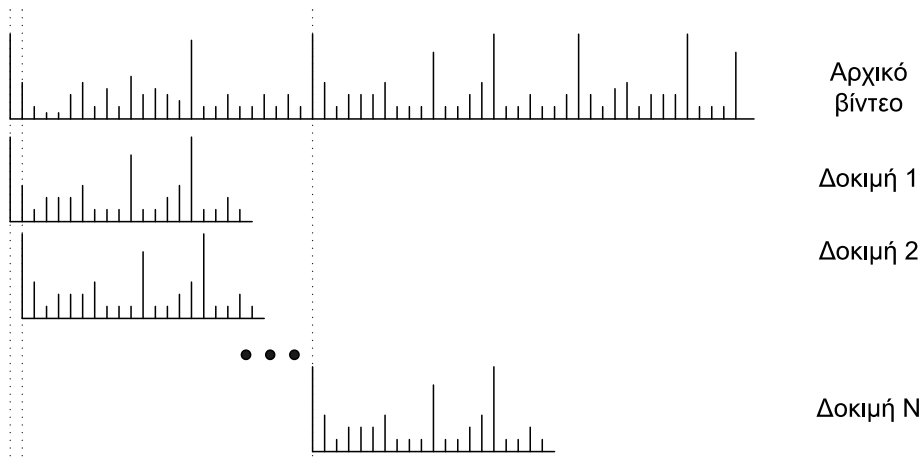


Σχήμα 3.22: Προσθήκη κανονικών και μεγαλύτερων υπότιτλων σε ένα βίντεο

3.8. ΜΕΘΟΔΟΣ INTRA MACROBLOCK



Σχήμα 3.23: Σύγκριση των ποσοστών *intra macroblock* όταν προστεθούν υπότιτλοι σε κάποιο απόσπασμα



Σχήμα 3.24: Απλή αναζήτηση με ολίσθηση

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

Συγκεκριμένα, έστω το βίντεο $P = p_1, p_2, \dots, p_m$ της βάσης δεδομένων και το απόσπασμα $Q = q_1, q_2, \dots, q_n$. Για την απλή αναζήτηση υπολογίζονται τα

$$D_{intra,j} = \frac{\sum_{i=2}^n (p_{i+j-1} - q_i)^2}{n-1}, \forall j \in \{2, m-n\}.$$

Γίνεται διαίρεση με το $(n-1)$ που είναι το μήκος του αποσπάσματος για να υπάρχει κανονικοποίηση με βάση αυτό το μήκος. Η υπογραφή της πρώτης εικόνας δε λαμβάνεται υπό όψιν αφού στα αποσπάσματα η πρώτη εικόνα θα κωδικοποιηθεί με όλα της τα macroblock κατά intra, κάτι που δεν είναι απαραίτητο ότι θα συμβαίνει και στο βίντεο από το οποίο προέρχεται το απόσπασμα.

Οι intra-mb απόστασεις κάθε θέσης μπορούν να ταξινομηθούν και να εμφανιστούν οι καλύτερες ως υποψήφιες θέσεις, ή ακόμα και να εμφανιστούν εκείνες οι θέσεις που έχουν απόσταση κάτω από κάποιο επιλεγμένο κατώφλι.

3.8.3.2 Απλή αναζήτηση με έλεγχο διακοπής

Μια προσθήκη στην προηγούμενη μέθοδο αναζήτησης έχει να κάνει με την προσθήκη ελέγχου έτσι ώστε ο υπολογισμός της απόστασης κατά intra-mb να διακοπεί πρόωρα σε περίπτωση που η απόσταση ξεφεύγει και γίνεται πολύ μεγαλύτερη από κάποιο κατώφλι ή πολύ μεγαλύτερη από τις ήδη υπάρχουσες μικρές.

Αυτό που συμβαίνει λοιπόν κατά τον έλεγχο για διακοπή είναι κάθε N διαδοχικές εικόνες να υπολογίζεται η μερική απόσταση κατά intra-mb, δηλαδή η απόσταση για τις εικόνες ως εκείνη τη στιγμή. Αν αυτή η απόσταση είναι πολύ μεγάλη τότε δεν έχει νόημα να συνεχιστεί ο υπολογισμός της απόστασης και έτσι ο αλγόριθμος θα προχωρήσει στην επόμενη θέση. Το N μπορεί να πάρει διαφορετικές τιμές, όπου με την τιμή 1 θα γίνεται ο έλεγχος σε κάθε δυνατή θέση, κάτι που έχει το πλεονέκτημα ότι ο υπολογισμός της απόστασης intra-mb θα διακοπεί αμέσως μόλις αυτή ξεπεράσει το επιλεχθέν κατώφλι, όμως έχει το μειονέκτημα ότι υπολογίζεται σε κάθε βήμα η μερική απόσταση κάτι που έχει μεγάλο υπολογιστικό κόστος. Αντίθετα, πολύ μεγάλες τιμές του N έχουν τον κίνδυνο το να αργήσει πολύ να διακοπεί ο υπολογισμός της απόστασης, άρα να μην υπάρχει ουσιαστικό πλεονέκτημα από τη χρήση της διακοπής.

3.8.3.3 Αναζήτηση μεγάλων τιμών

Για αυτή την τεχνική αναζήτησης, θα πρέπει να δημιουργηθεί ένας κατάλογος για το ποιές εικόνες έχουν συγκεκριμένες τιμές ποσοστού intra macroblock. Συγκεκριμένα, επειδή οι περισσότερες εικόνες έχουν μικρές τιμές intra macroblock κρατούνται στον κατάλογο μόνο οι εικόνες που έχουν από κάποιο επιλεχθέν ποσοστό και πάνω έτσι ώστε να μη μεγαλώνει πολύ το μέγεθος του καταλόγου αυτού. Ένα παράδειγμα θα ξεκαθαρίσει περαιτέρω τον τρόπο δημιουργίας του καταλόγου. Έστω το παράδειγμα με τιμές ποσοστών intra macroblock για ένα μικρό απόσπασμα που φαίνεται στον πίνακα 3.6. Αν το κατώφλι για τις τιμές που μπαίνουν στον κατάλογο είναι το 50 τότε ο κατάλογος

3.8. ΜΕΘΟΔΟΣ INTRA MACROBLOCK

για αυτόν τον πίνακα θα είναι

100: 1, 11

90: 6

60: 5

και 0 σε όλα τα άλλα στοιχεία από το 50 ως 99.

αριθμός εικόνας	1	2	3	4	5	6	7	8	9	10	11	12
ποσοστό intra mb	100	5	4	6	60	90	4	2	2	1	100	1

Πίνακας 3.6: Παράδειγμα τιμών

Ο κατάλογος αυτός λοιπόν θα πρέπει να δημιουργηθεί κατά τη δεικτοδότηση ενός βίντεο, αλλά θα μπορούσε εύκολα να δημιουργηθεί ακόμα και κατά την αναζήτηση, αφού χρειάζεται μόνο ένα «πέραςμα» κάθε βίντεο.

Εφόσον υπάρχει ο κατάλογος, μπορεί να γίνει η αναζήτηση με μεγάλες τιμές: Αρχικά βρίσκεται η μέγιστη τιμή ποσοστών intra macroblock του αποσπάσματος (φυσικά αγνοείται η πρώτη εικόνα). Αν υπάρχουν πολλές εικόνες με την ίδια μέγιστη τιμή τότε απλώς επιλέγεται η πρώτη εξ αυτών. Στη συνέχεια, με τη βοήθεια του καταλόγου βρίσκονται οι θέσεις του βίντεο της βάσης δεδομένων οι οποίες έχουν τιμή αρκετά κοντά στη μέγιστη τιμή του αποσπάσματος που βρέθηκε και βρίσκεται η intra mb απόσταση όταν ευθυγραμμιστούν οι δύο αυτές τιμές.

Στο σχήμα 3.25 φαίνεται ένα παράδειγμα: Η απόσταση κατά intra mb θα υπολογιστεί μόνο 5 φορές για αυτή την περίπτωση, κάτι που επιταχύνει πολύ την αναζήτηση.

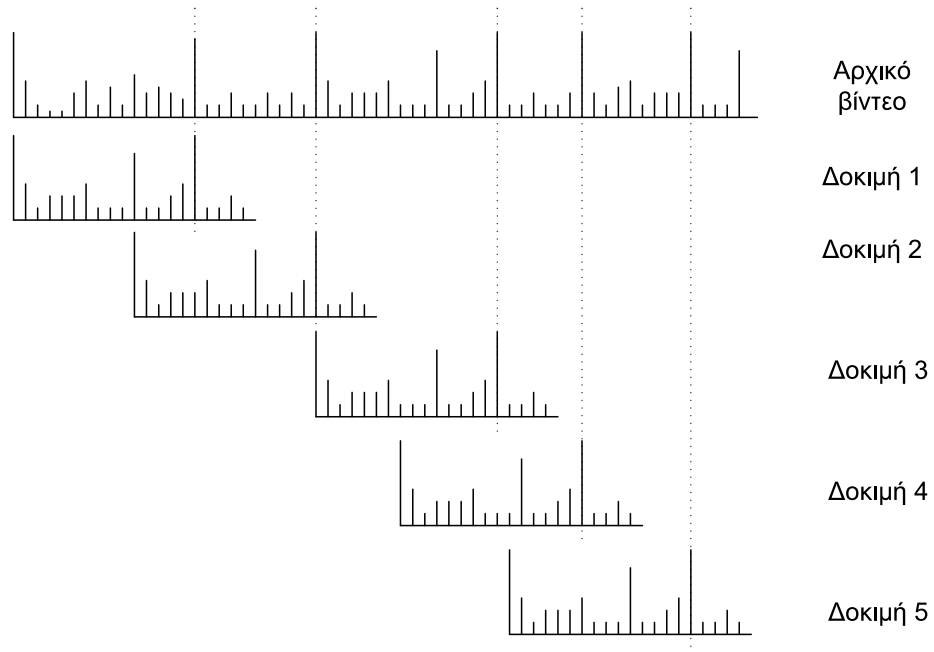
Για κάθε μια από τις υποψήφιες θέσεις υπάρχει η δυνατότητα φυσικά να χρησιμοποιηθεί ο έλεγχος διακοπής που περιγράφηκε στην προηγούμενη ενότητα, έτσι ώστε να γίνουν ακόμα λιγότερες πράξεις.

3.8.3.4 Αναζήτηση διαστημάτων

Ο τελευταίος τρόπος που προτείνεται για την αναζήτηση με τη βοήθεια των ποσοστών των intra macroblock κάθε εικόνας έχει να κάνει με τα διαστήματα και μοιάζει αρκετά με τη μέθοδο αναζήτησης βάσης δεδομένων που περιγράφηκε στην ενότητα 3.7.

Για την τεχνική αυτή δημιουργείται ακόμη ένας κατάλογος, ο οποίος περιέχει τα μήκη των διαστημάτων μεταξύ των εικόνων που έχουν μεγάλα ποσοστά intra macroblock. Βρίσκονται οι αποστάσεις όλων των εικόνων που έχουν ποσοστό intra macroblock πάνω από το κατώφλι και, αφού ταξινομηθούν, αποθηκεύονται σε έναν κατάλογο μαζί με τη θέση φυσικά της πρώτης εικόνας του διαστήματος.

Από το απόσπασμα εξάγονται οι αριθμοί δύο εικόνων οι οποίες βρίσκονται πάνω από το κατώφλι, αντίστοιχα με τον τρόπο με τον οποίο βρίσκονται οι αλλαγές πλάνου στην ενότητα 2.4. Πρέπει το απόσπασμα να έχει παραπάνω από



Σχήμα 3.25: Αναζήτηση με μεγάλες τιμές

μια τέτοια αλλαγή πλάνου έτσι ώστε να βρεθούν οι δύο, ενώ αν έχει παραπάνω από δύο τότε θα επιλεγούν οι δύο πρώτες και θα υπολογιστεί η μεταξύ τους απόσταση.

Στη συνέχεια, γίνεται χρήση του καταλόγου των αποστάσεων από όπου βρίσκονται τα ζεύγη αλλαγών πλάνου της βάσης δεδομένων που έχουν ίδια απόσταση μεταξύ τους με την απόσταση μεταξύ των αλλαγών πλάνου του αποσπάσματος. Τέλος, υπολογίζεται η απόσταση κατά intra mb μεταξύ του αποσπάσματος και του συγκεκριμένου σημείου του βίντεο της βάσης δεδομένων.

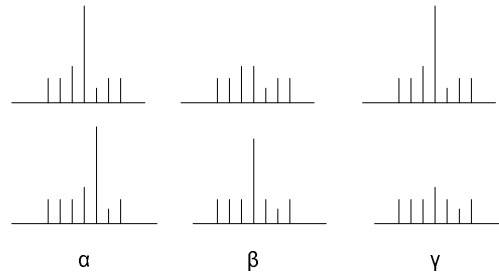
Σε αυτή την περίπτωση η απόσταση κατά intra mb υπολογίζεται πολύ λίγες φορές, ουσιαστικά μόνο σε εκείνα τα σημεία στα οποία τυχαίνει η απόσταση αλλαγών πλάνου να ταυτίζεται. Για να υπολογιστεί ακόμα λιγότερες φορές η απόσταση, μπορεί από το βίντεο να χρησιμοποιηθούν δύο η παραπάνω διάρκειες μεταξύ αλλαγών πλάνου και να γίνει η σύγκριση με τις συνεχόμενες διάρκειες στο βίντεο της βάσης δεδομένων.

3.8.3.5 Ανοχή σε jitter

Στην ενότητα 3.5.2 αποδείχθηκε ότι στις συνήθεις περιπτώσεις επανακωδικοποίησης βίντεο, όταν τα fps των βίντεο είναι σταθερά (η αλλάζουν αλλά επανέρχονται στην αρχική τους τιμή) μπορεί να δημιουργηθεί στο αλλαγμένο βίντεο jitter ενός ή δύο εικόνων, το οποίο δε συσσωρεύεται μεταξύ των εικόνων, δηλαδή δύο αλλαγές πλάνου για παράδειγμα δε θα διαφέρουν ποτέ περισσότερο από δύο εικόνες.

3.8. ΜΕΘΟΔΟΣ INTRA MACROBLOCK

Με τη μέχρι τώρα ανάλυση, δεν έχει παρουσιαστεί τρόπος για την ανοχή σε jitter. Στο σχήμα 3.26 α, φαίνονται τα ποσοστά intra macroblock δύο παραλλαγών ενός βίντεο τα οποία έχουν jitter μιας εικόνας μεταξύ τους. Με την απλή αναζήτηση θα βρεθεί μεγάλη διαφορά στις δύο εικόνες που έχουν την αλλαγή πλάνου και έτσι δε θα βρεθεί. Αντίθετα, με την αναζήτηση μεγάλων τιμών δε θα δημιουργηθεί πρόβλημα αφού το βίντεο θα συγχρονιστεί στο ίδιο σημείο. Σε αυτή την περίπτωση όμως, θα δημιουργηθεί πρόβλημα αν υπάρχει μια δεύτερη αλλαγή πλάνου αργότερα, αφού και πάλι τότε θα δημιουργηθεί μεγάλη διαφορά.



Σχήμα 3.26: Ανοχή σε jitter

Έτσι, χρειάζεται ένας γενικός τρόπος ώστε να μπορούν να λυθούν αυτά τα προβλήματα. Στον ορισμό 3.8.1 προηγουμένως ορίστηκε η intra-mb απόσταση δύο ακολουθιών. Ακολούθως θα οριστεί η intra-mb απόσταση με ανοχή σε jitter jt,

Ορισμός 3.8.3 Έστω δύο αποσπάσματα P και Q ίδιου μήκους k , όπου το P έχει ως υπογραφή την ακολουθία $\{p_1, p_2, \dots, p_k\}$ και το Q την $\{q_1, q_2, \dots, q_k\}$. Η intra-mb απόσταση αυτών με ανοχή σε jitter jt:

$$D_{intra,jt} = \frac{\sum_{i=2}^k \left(\min_{j=0}^{jt} \{(p_{i\pm j} - q_i)^2\} + \min_{j=0}^{jt} \{(p_i - q_{i\pm j})^2\} \right)}{2(k-1)}.$$

Η απόσταση λοιπόν δύο αποσπασμάτων όταν αυτά έχουν ανοχή σε jitter ίσο με jt βρίσκεται όταν για κάθε μια θέση σύγκρισης i ελέγχεται η διαφορά της θέσης i της Q από όλες τις θέσεις που βρίσκονται σε απόσταση jt από τη θέση i της P και κρατείται η μικρότερη διαφορά η οποία και αθροίζεται με τη μικρότερη διαφορά που έχει η θέση i της P από όλες τις θέσεις που βρίσκονται σε απόσταση jt από τη θέση i της Q . Η τιμή διαιρείται με το $2(k-1)$ αντίστοιχα με τον τύπο του ορισμού 3.8.1 που διαιρείται με το $k-1$ αφού εδώ υπάρχει άθροισμα δύο τιμών.

Αν λοιπόν τεθεί το $jt = 1$, ο τύπος θα δουλεύει σωστά για όλες τις περιπτώσεις του σχήματος 3.26. Στο α, όταν το P είναι στη θέση με την αλλαγή πλάνου, θα επιλεγεί η διαφορά με την επόμενη θέση του Q αφού εκεί είναι μικρότερη, ενώ για το Q θα επιλεγεί η διαφορά με την προηγούμενη θέση του P .

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

Ομοίως και για τη θέση του P αμέσως μετά την αλλαγή πλάνου, για το μεν P θα επιλεγεί η διαφορά μετά την αλλαγή πλάνου του Q και για το Q η αλλαγή πλάνου στο P . Στην περίπτωση τώρα το β , όπου υπάρχει αλλαγή πλάνου στο Q αλλά όχι στο P , το μεν P θα βρίσκει πάντοτε κάποια μικρή διαφορά με τις άλλες εικόνες, όμως στην περίπτωση της αλλαγής πλάνου του Q , πάντοτε θα μετράται η μεγάλη διαφορά που θα έχει αφού στο P δε θα υπάρχει αλλαγή πλάνου. Αυτός είναι και ο λόγος που πρέπει να εξετάζεται η ακτίνα και στο P και στο Q και να αθροίζονται, μιας και αν εξεταζόταν η ακτίνα μόνο στο Q , το P θα φαινόταν ότι ταιριάζει σε εκείνο το σημείο παρά την αλλαγή πλάνου. Ομοίως και στο γ , στο σημείο αλλαγής πλάνου του P δε θα ταιριάζει με κανένα του Q και έτσι θα αυξηθεί κατάλληλα η διαφορά ουτως ώστε να μην υπάρξει ταίριασμα στο σημείο εκείνο.

Με τον τρόπο αυτό λοιπόν, μπορεί να γίνει σύγκριση βίντεο ακόμα και αν έχει εισαχθεί αυθαίρετο jitter jt . Φυσικά, η πολυπλοκότητα αυξάνει, αφού αντί για μία διαφορά, τώρα θα χρειάζεται να υπολογιστούν $2 \times jt + 1$ διαφορές και να βρεθούν οι μικρότερες. Όμως, όπως έχει αναφερθεί, το jitter σε φυσιολογικές περιπτώσεις παίρνει τιμές 1 ή 2 εικόνες, και έτσι η πολυπλοκότητα της μεθόδου αυξάνει πολύ λίγο.

3.8.4 Γενικά για τη μέθοδο *intra macroblock*

Η μέθοδος *intra macroblock* είναι ελαφρώς πιο περίπλοκη από την πιθανοτική στην οποία ελέγχονται ακολουθίες από εικόνες κάθε μια εκ των οποίων απλά αντιπροσωπεύεται από 1 και 0 (ύπαρξη και μη ύπαρξη συμβάντος) και την μέθοδο βάσης δεδομένων στην οποία η διάρκεια κάθε πλάνου αντιπροσωπεύει πολλές εικόνες, αφού εδώ κάθε εικόνα αντιπροσωπεύεται από το ποσοστό *intra macroblock* που περιέχει και πρέπει να βρεθεί η *intra mb* απόσταση μεταξύ τους.

Εκτός από την απλή αναζήτηση όμως, μπορούν να χρησιμοποιηθούν τόσο η αναζήτηση μεγάλων τιμών όσο και η αναζήτηση αποστάσεων έτσι ώστε να αυξηθεί η ταχύτητα της μεθόδου αφού θα γίνει ένα γρήγορο κοσκίνισμα των υποψηφίων που δεν ταιριάζουν.

Η αναζήτηση διαστημάτων έχει τον κίνδυνο ότι μπορεί μια αλλαγή πλάνου να βρεθεί στο απόσπασμα αλλά όχι στο βίντεο της βάσης δεδομένων, και έτσι να χαθεί κάποιο αποτέλεσμα, αφού δεν υπάρχει εδώ πρόβλεψη για ανοχή σε χαμένες ή εσφαλμένες αναγνωρίσεις. Αντίθετα, η αναζήτηση μεγάλων τιμών ο κίνδυνος δεν υπάρχει σε τόσο μεγάλο βαθμό, μιας και οι «μεγάλες» τιμές του βίντεο της βάσης δεδομένων που θα ελεγχθούν μπορεί να είναι πάρα πολλές έτσι ώστε σίγουρα να περιληφθούν και οι αντίστοιχες «μεγάλες» τιμές του αποσπάσματος, με επιπλέον φυσικά υπολογιστικό κόστος.

3.9 Συμπεράσματα

Οι τρεις αυτές τεχνικές, έχουν κάποιες ομοιότητες και αρκετές διαφορές. Η πρώτη, βασίζεται σε μια θεωρητική ανάλυση του αριθμού των διαφορών στις αλλαγές πλάνου που αναγνωρίζονται σε ένα αρχικό βίντεο και σε ένα τροποποιημένο απόσπασμα αυτού. Η ανάλυση αυτή αποδεικνύει ότι αν το απόσπασμα συγκρίνεται με το τμήμα του βίντεο από το οποίο προέρχεται, ο αριθμός των διαφορών στις αλλαγές πλάνου θα είναι πολύ μικρότερος από οποιαδήποτε άλλη θέση.

Στη δεύτερη, δημιουργούνται ερωτήματα προς τη βάση δεδομένων για την εύρεση της ακολουθίας των αλλαγών πλάνου που υπάρχει στο απόσπασμα. Εδώ η αναζήτηση γίνεται με τη βοήθεια του συστήματος διαχείρισης βάσεων δεδομένων. Οι δύο αυτές τεχνικές χρησιμοποιούν ουσιαστικά μια ακολουθία από συμβάντα, τα οποία και στις δύο περιπτώσεις είναι οι εικόνες εκείνες που αναγνωρίζονται ως αλλαγές πλάνου και σαν κύριο ρόλο έχουν τη γρήγορη αναζήτηση υποψηφίων - παρότι θα μπορούσαν να χρησιμοποιηθούν και για ταίριασμα ακολουθίας όταν το απόσπασμα έχει αρκετές αλλαγές πλάνου. Παρότι και η πιθανοτική και η μέθοδος βάσης δεδομένων χρησιμοποιούν τελικά την ίδια ακριβώς πληροφορία (τις αλλαγές πλάνου και μόνο) διαφέρουν στον τρόπο με τον οποίο αναπαριστούν αυτή την πληροφορία: Στην πιθανοτική μέθοδο χρησιμοποιούνται ακολουθίες από 0 και 1 (συμβάν και μη συμβάν / αλλαγή πλάνου ή όχι) έτσι ώστε κάθε σύγκριση να θεωρηθεί ξεχωριστό πείραμα Bernoulli και να δημιουργηθούν οι διωνυμικές κατανομές των διαφορών. Στη μέθοδο της βάσης δεδομένων αντίθετα, αποθηκεύεται το μήκος των διαστημάτων μεταξύ των αλλαγών πλάνου, είτε σε χρόνο είτε σε εικόνες, έτσι ώστε να είναι εύκολη η δημιουργία ερωτημάτων προς τη βάση δεδομένων για την αναζήτηση συγκεκριμένων ακολουθιών διαστημάτων.

Η τρίτη μέθοδος χρησιμοποιεί την ακολουθία από το ποσοστό των intra macroblock κάθε μιας εικόνας έτσι ώστε να δημιουργήσει μια υπογραφή για όλες τις εικόνες του αποσπάσματος. Η υπογραφή αυτή συγκρίνεται με την υπογραφή του αρχικού βίντεο και ανιχνεύονται οι θέσεις με τη μικρότερη διαφορά. Η συγκεκριμένη τεχνική, ανάλογα με το πόσο ανεκτικό ή όχι το σύστημα θα είναι στις διαφορές και φυσικά με τη χρήση των βελτιώσεων που αναφέρθηκαν πριν⁶, μπορεί να χρησιμοποιηθεί και για αναζήτηση υποψηφίων αλλά και για ταίριασμα ακολουθίας.

Στον πίνακα 3.7 φαίνονται τα κυριότερα χαρακτηριστικά των τριών μεθόδων πρώτου επιπέδου της ιεραρχικής αναζήτησης που παρουσιάστηκαν στις ενότητες 3.6, 3.7 και 3.8. Η πιθανοτική και η intra mb εξετάζουν μία μία τις εικόνες, ενώ αντίθετα η μέθοδος βάσης δεδομένων εξετάζει μόνο τις διάρκειες κάθε πλάνου. Όλες οι μέθοδοι έχουν ανοχή στο jitter που παρουσιάζεται στις περισσότερες περιπτώσεις. Για τη μέθοδο intra mb δεν έχει νόημα η έννοια της εσφαλμένης ή χαμένης αναγνώρισης αλλαγής πλάνου, αφού εδώ οι εικόνες που ανιχνεύονται

⁶Μεγάλες τιμές / διαστήματα

ΚΕΦΑΛΑΙΟ 3. ΑΝΑΖΗΤΗΣΗ ΒΙΝΤΕΟ

	Πιθανοτική	ΒΔ	intra mb
Διακριτότητα	εικόνα	πλάνο	εικόνα
Χαρακτηριστικό	αλλαγή πλά- νου ή όχι	διάρκεια πλάνου	αριθμός intra mb
Ανοχή σε Jitter	Ναί	Ναί	Ναί
Ανοχή σε εσφαλ- μένες ανα- γνώρισεις	Ναί	Ναί	N/A
Ανοχή σε χαμέ- νες αναγνώρισεις	Ναί	Ναί (με κακή απόδοση)	N/A

Πίνακας 3.7: Σύγκριση μεθόδων πρώτου επιπέδου

ως αλλαγές πλάνου δεν παίζουν κάποιον ιδιαίτερο ρόλο (εκτός και αν χρησιμοποιηθεί η αναζήτηση αποστάσεων) ενώ οι άλλες δύο μέθοδοι υποστηρίζουν τόσο τις χαμένες όσο και τις εσφαλμένες αναγνώρισεις αλλαγών πλάνου.

Οι τεχνικές αναζήτησης που παρουσιάστηκαν έχουν ως κύριο χαρακτηριστικό τους τη μεγάλη ταχύτητά, τόσο κατά την αναζήτηση όσο και κατά τη δεικτοδότηση. Δόθηκε μεγάλη σημασία στη χρήση χαρακτηριστικών δεικτοδότησης τα οποία μπορούν να εξαχθούν από το πρωτόκολλο H.264 χωρίς να χρειάζεται πλήρης αποσυμπίεση του βίντεο. Όπως έχει αναφερθεί, το H.264 λόγω της δομής του και των εξελιγμένων τεχνικών συμπίεσης παρουσιάζει δυσκολίες στην εξαγωγή χαρακτηριστικών για αναζήτηση. Έτσι, χρησιμοποιήθηκαν οι αλλαγές πλάνου και ο τρόπος συμπίεσης κάθε macroblock, χαρακτηριστικά που μπορούν να εξαχθούν απευθείας από το συμπίεσμένο βίντεο.

Κεφάλαιο 4

Συστήματα αναζήτησης

ΓΙΑ την υποστήριξη των θεωρητικών και πειραματικών αποτελεσμάτων της διατριβής, σχεδιάστηκαν και υλοποιήθηκαν τρία ολοκληρωμένα συστήματα αναζήτησης βίντεο: Δύο συστήματα για αναζήτηση αποσπασμάτων σε ήδη υπάρχοντα αρχεία, βασισμένα το ένα σε τεχνολογίες ιστού και το άλλο για κονσόλα και ένα σύστημα για Windows για αναζήτηση αποσπασμάτων σε «ζωντανό» βίντεο (συνεχής αναζήτηση). Το σύστημα αναζήτησης αποσπασμάτων σε υπάρχοντα αρχεία που βασίζονταν σε τεχνολογίες ιστού είναι μια πλήρης εφαρμογή που θα μπορούσε να χρησιμοποιηθεί για ένα ολοκληρωμένο σύστημα αναζήτησης βίντεο. Το σύστημα της κονσόλας υλοποιεί και τις τρεις μεθόδους αναζήτησης που παρουσιάστηκαν στο προηγούμενο κεφάλαιο και χρησιμοποιήθηκε κυρίως για την εξαγωγή πειραματικών αποτελεσμάτων σχετικά με τις τεχνικές αναζήτησης, τα οποία και θα παρουσιαστούν στο επόμενο κεφάλαιο. Τέλος, το σύστημα της συνεχούς αναζήτησης χρησιμοποιεί τη μέθοδο των *intra macroblock* και μπορεί να χρησιμοποιηθεί σαν ένα πλήρες σύστημα καταγραφής στοιχείων για ένα συνεχές βίντεο, πχ τηλεοπτικό πρόγραμμα.

Ως πλατφόρμα εκτέλεσης των παραπάνω συστημάτων χρησιμοποιήθηκαν απλοί προσωπικοί υπολογιστές κάτι που σε συνδυασμό με τα αποτελέσματα του επόμενου κεφαλαίου για την ταχύτητα αναζήτησης, αποδεικνύει τη μεγάλη ταχύτητα των τεχνικών που προτείνει η εργασία. Επίσης φαίνεται ότι πλέον οι προσωπικοί υπολογιστές πληρούν με ευκολία τις απαιτήσεις για μια διαδικασία η οποία πριν κάποια χρόνια φάνταζε υπερβολικά περίπλοκη. Αυτό μπορεί να φανεί σε ακόμα μεγαλύτερο βαθμό στο σύστημα συνεχούς αναζήτησης, όπου όλες οι διαδικασίες (μετατροπή, δεικτοδότηση, αναζήτηση) πρέπει να γίνονται σε πραγματικό χρόνο.

Πριν την αναλυτική παρουσίαση των συστημάτων αυτών, είναι σημαντικό να αναφερθούν κάποια στοιχεία που αφορούν το πώς γίνεται η εξαγωγή των στοιχείων από τα διάφορα βίντεο αφού τέτοιες εφαρμογές πρέπει να μπορούν να συνεργαστούν με όσο γίνεται περισσότερα *formats* βίντεο. Έτσι, στην επόμενη ενότητα θα παρουσιαστούν δύο τεχνολογίες με τις οποίους μπορεί να γίνει εξαγωγή των στοιχείων από διάφορα βίντεο, και ακολούθως θα παρουσιαστούν

με τη σειρά τα συστήματα που αναφέρθηκαν.

4.1 Εξαγωγή στοιχείων από βίντεο

Μια απαίτηση ενός συστήματος αναζήτησης βίντεο είναι να μπορεί να εξάγει στοιχεία από όλα τα δημοφιλή format αναζήτησης βίντεο. Εφόσον υπάρχουν δεκάδες τέτοια format, είναι απαραίτητη η χρήση κάποιων εξωτερικών βιβλιοθηκών οι οποίες θα προσφέρουν γρήγορη πρόσβαση στα στοιχεία των βίντεο που είναι απαραίτητα. Από τις τρεις μεθόδους αναζήτησης που παρουσιάστηκαν στο κεφάλαιο 3, οι δύο βασίζονται στις αλλαγές πλάνου, άρα θα πρέπει να μπορεί να γίνει εύκολη ανίχνευση των αλλαγών πλάνου με τους τρόπους που περιγράφηκαν στο κεφάλαιο 2, ενώ η τρίτη βασίζεται στην εξαγωγή πληροφοριών για τον τύπο κάθε macroblock από ένα αρχείο H.264. Οι τεχνικές ανίχνευσης αλλαγών πλάνου βασίζονται η μία σε εξαγωγή πληροφοριών από H.264 και η άλλη σε δημιουργία ιστογράμματος φωτεινότητας από ένα ασυμπίεστο βίντεο.

Έτσι, αν το σύστημα αναζήτησης δεχθεί ως είσοδο ένα βίντεο με αυθαίρετο format τότε θα πρέπει είτε να μετατρέψει αυτό το βίντεο σε H.264 είτε απλώς να το αποσυμπίσει. Η αποσυμπίεση είναι έτσι ή αλλιώς προαπαιτούμενο για τη μετατροπή ενός βίντεο διαφορετικού είδους σε H.264 λόγω των εξελιγμένων εργαλείων συμπίεσης που χρησιμοποιούνται στο H.264. Έτσι, ενώ η απ' ευθείας μετατροπή από H.264 σε παλαιότερα formats έχει κέρδη στην ταχύτητα μετατροπής χωρίς υποβάθμιση στην ποιότητα σε σχέση με την πλήρη αποσυμπίεση και συμπίεση (για παράδειγμα μετατροπή από H.264 σε MPEG-2 στο [MFN⁺]), με το αντίθετο τα εξελιγμένα εργαλεία του H.264 δε μπορούν να χρησιμοποιηθούν και έτσι η συμπίεση είναι πολύ μικρότερη από ότι θα ήταν αν γινόταν πλήρης αποσυμπίεση ([AMB]).

Υπάρχουν δύο δημοφιλείς τεχνολογίες για το χειρισμό βίντεο: Το DirectShow της Microsoft [DSH, Pes03] και οι βιβλιοθήκες του FFMpeg [FFM] που είναι πρόγραμμα ανοιχτού κώδικα. Παρότι για τις υλοποιήσεις χρησιμοποιήθηκαν αποκλειστικά τα εργαλεία ανοιχτού κώδικα που βασίζονται στο FFMpeg έτσι ώστε οι υλοποιήσεις να μπορούν να χρησιμοποιηθούν τόσο σε Windows όσο και σε άλλα λειτουργικά συστήματα, τα εργαλεία του FFMpeg θα μπορούσαν εύκολα να αντικατασταθούν από αντίστοιχα βασισμένα στο DirectShow χωρίς να αλλάξει κάτι στον τρόπο λειτουργίας των μεθόδων αναζήτησης και ανίχνευσης αλλαγής πλάνου.

Από τα προηγούμενα φαίνεται ότι το πρωτόκολλο συμπίεσης H.264 έχει ιδιαίτερη θέση στην ανάλυση της εργασίας αυτής. Αυτό συμβαίνει επειδή, λόγω της φύσης του, είναι πολύ δύσκολη η εξαγωγή πληροφοριών από βίντεο συμπιεσμένο κατά H.264 χωρίς την πλήρη αποσυμπίεση του. Οι τεχνικές όμως που προτείνονται, δουλεύουν με το είδος του κάθε macroblock του βίντεο (αν είναι δηλαδή intra ή όχι), μια πληροφορία η οποία μπορεί να εξαχθεί εύκολα χωρίς την πλήρη αποσυμπίεση του H.264. Έτσι, με τους αλγόριθμους που προτείνονται εδώ μπορεί να γίνει εξαγωγή των χαρακτηριστικών δεικτοδότησης

4.1. ΕΞΑΓΩΓΗ ΣΤΟΙΧΕΙΩΝ ΑΠΟ ΒΙΝΤΕΟ

του H.264 χωρίς την πλήρη αποσυμπίεση του, αρκεί φυσικά το H.264 να είναι κατάλληλης μορφής, όπως αυτή έχει παρουσιαστεί στα προηγούμενα κεφάλαια.

Στη συνέχεια θα γίνει μια συνοπτική παρουσίαση των τεχνολογιών του DirectShow και του FFMpeg, θα αναφερθούν κάποια βασικά στοιχεία για τη χρήση καθώς και μια σύγκριση των πλεονεκτημάτων και μειονεκτημάτων τους.

4.1.1 FFMpeg

Το FFMpeg αποτελεί μια προσπάθεια για δημιουργία μιας πλήρους λύσης για την κωδικοποίηση και αποκωδικοποίηση ήχου και βίντεο σε πολλές πλατφόρμες. Το FFMpeg είναι λογισμικό ανοιχτού κώδικα και τα περισσότερα συστατικά του ακολουθούν την άδεια LGPL [LGP07]. Αυτό σημαίνει ότι μπορούν να χρησιμοποιηθούν τόσο σε εφαρμογές ανοιχτού κώδικα όσο και σε εμπορικές εφαρμογές¹. Οι βιβλιοθήκες libavcodec και libavformat είναι το σημαντικότερο στοιχεία του, με την πρώτη να περιλαμβάνει κωδικοποιητές και αποκωδικοποιητές για δεκάδες πρωτόκολλα ήχου και εικόνας, για παράδειγμα h.261, h.263, h.264, MJPEG, MPEG-1, MPEG-2, MPEG-4, QuickTime, Windows Media Video, AAC, AC3, MP2, MP3, PCM, Vorbis, Windows Media Audio και τη δεύτερη να υποστηρίζει δεκάδες file formats όπως asf, avi, flv, mp3, MPEG-1, MPEG2-PS, MPEG2-TS, MPEG-4, Ogg, WAV.

Παρότι κυρίως αναπτύσσεται για λειτουργικά που βασίζονται στο Unix η χρήση του στα windows είναι δυνατή τόσο με δημιουργία εφαρμογών που χρησιμοποιούν το ffmpeg αυτόνομα, όσο και με τη βοήθεια ενός DirectShow φίλτρου που ονομάζεται FFDSHOW και λειτουργεί σαν ένας κόμβος επικοινωνίας μεταξύ του DirectShow και του FFMpeg. Έτσι, τα Ffmpeg codecs μπορούν να χρησιμοποιηθούν με μηδαμινή εργασία από εφαρμογές που υποστηρίζουν το DirectShow.

Η υλοποίηση εφαρμογών που χρησιμοποιούν το FFMpeg για να αποκωδικοποιήσουν αρχεία βίντεο είναι αρκετά απλή: Αφού ανοιχτεί το βίντεο με τις μεθόδους της libavformat, αναγνωρίζεται το πρωτόκολλο συμπίεσης του και χρησιμοποιείται το κατάλληλο codec της libavcodec για να αποσυμπιέσει μια μια τις εικόνες. Από κάθε αποσυμπιεσμένη εικόνα, μπορούν να εξαχθούν μια σειρά από στοιχεία. Αν αυτά δεν επαρκούν μπορεί να γίνει και επέμβαση στη διαδικασία της αποσυμπίεσης: Κάθε codec είναι υλοποιημένο σε ένα ξεχωριστό αρχείο² οπότε το αρχείο αυτό μπορεί να αλλαχτεί κατάλληλα έτσι ώστε να εξαγονται στοιχεία όταν για παράδειγμα βρεθεί ένα intra macroblock.

4.1.2 DirectShow

Το DirectShow αποτελεί μια γενική λύση για πολυμέσα στα Windows. Αρχικά ήταν μέρος του υποσυστήματος DirectX για παιχνίδια, αλλά πλέον αποτελεί βασικό μέρος του Windows SDK. Βασίζεται στο Component Object

¹Αρκεί να γνωστοποιείται η χρήση της βιβλιοθήκης

²Για παράδειγμα, για το h.264 υπάρχει το αρχείο libavcodec/h264.c

ΚΕΦΑΛΑΙΟ 4. ΣΥΣΤΗΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ

Model - COM της Microsoft παρέχοντας δυνατότητες multimedia σε κάθε γλώσσα προγραμματισμού που υποστηρίζει το COM, δηλαδή μπορεί να χρησιμοποιηθούν όλες οι δυνατότητες του από δεκάδες γλώσσες προγραμματισμού. Το βασικό του στοιχείο είναι τα φίλτρα: Μια οποιαδήποτε πολύπλοκη λειτουργία multimedia (πχ αναπαραγωγή ενός βίντεο) διασπάται σε επιμέρους τμήματα. Κάθε ένα τμήμα, διεκπεραιώνεται από ένα «πρόγραμμα» που ονομάζεται φίλτρο. Ένα φίλτρο, έχει μια σειρά από εισόδους και εξόδους, έτσι ώστε τα φίλτρα να μπορούν να συνδεθούν με διάφορους τρόπους μεταξύ τους για να υλοποιήσουν διάφορες λειτουργίες. Για να γίνει μια οποιαδήποτε εργασία multimedia, πρέπει να δημιουργηθούν κάποια φίλτρα και να ενωθούν - αυτό λέγεται γράφημα φίλτρων (filter graph). Υπάρχουν τρία είδη φίλτρων:

- Φίλτρα πηγής (source filters) για την ανάγνωση μультιμεδία αρχείων
- Φίλτρα μετασχηματισμού (transform filter) για μετασχηματισμούς (πχ encoders)
- Φίλτρα εξόδου (renderer filters) για έξοδο (πχ στην οθόνη ή σε ένα αρχείο στο δίσκο)

Έτσι, για την υποστήριξη ενός format από το DirectShow αρκεί να υλοποιηθεί το κατάλληλο φίλτρο πηγής για το format αυτό, ενώ για υποστήριξη ενός πρωτοκόλλου συμπίεσης, αρκεί να υλοποιηθεί το φίλτρο μετασχηματισμού για αυτό το πρωτόκολλο.

Το DirectShow υποστηρίζεται από τις περισσότερες εταιρίες και έτσι υλοποιούνται φίλτρα για όλα τα γνωστά βίντεο φορματς, κάτι πολύ σημαντικό για τη χρήση του ως οικουμενικού αποκωδικοποιητή. Επίσης, με την υλοποίηση ενός κατάλληλου φίλτρου μετασχηματισμού, μπορεί να γίνεται η εξαγωγή των πληροφοριών ενός H.264.

4.1.3 Σύγκριση μεταξύ Ffmpeg και DirectShow

Παρότι είναι οι δύο δημοφιλέστερες τεχνολογίες διαχείρισης πολυμέσων, τα Ffmpeg και DirectShow έχουν περισσότερες διαφορές παρά ομοιότητες, κυρίως λόγω του ότι το ένα είναι πρόγραμμα ανοιχτού κώδικα και το άλλο όχι:

- Το Ffmpeg περιλαμβάνει codecs για σχεδόν όλα τα πρωτόκολλα συμπίεσης πολυμέσων, ενώ για το DirectShow θα πρέπει τα φίλτρα αποσυμπίεσης να βρεθούν και συνήθως να αγοραστούν.
- Το DirectShow υποστηρίζεται μόνο στα windows ενώ το Ffmpeg στις περισσότερες γνωστές πλατφόρμες.
- Το DirectShow λόγω των φίλτρων είναι αρκετά πιο ευέλικτο στη δημιουργία πολύπλοκων εφαρμογών πολυμέσων, όμως το Ffmpeg είναι ευκολότερο στη δημιουργία απλών εφαρμογών αφού δε χρειάζεται η διαχείριση των φίλτρων

4.2. ΣΥΣΤΗΜΑ ΙΣΤΟΥ

- Η επέμβαση στον κώδικα των codecs είναι δυνατή και σχετικά εύκολη στο Ffmpeg ενώ συνήθως είναι αδύνατη στο DirectShow, αφού λόγω φιλοσοφίας τα περισσότερα codecs είναι κλειστού κώδικα.
- Λόγω των φίλτρων πολλές φορές δημιουργούνται αμφιβολίες για το ποιός codec θα αποκωδικοποιήσει κάποιο βίντεο³, ενώ αντίθετα στο Ffmpeg τα codecs συνήθως περιέχονται στο binary της εφαρμογής και το ποιος θα κάνει τι είναι πολύ ξεκάθαρο.
- Το Ffmpeg έχει κάποια προβλήματα στη δημιουργία εμπορικών εφαρμογών λόγω του ότι κάποια (λίγα) από τα codecs του ακολουθούν την άδεια GPL [GPL07], μη επιτρέποντας ουσιαστικά τη δημιουργία εφαρμογών κλειστού κώδικα που θα τις χρησιμοποιούν.

Για τους παραπάνω λόγους είναι προφανές ότι στα πλαίσια της ερευνητικής εργασίας αυτής, το FFmpeg πλεονεκτεί καθολικά έναντι του DirectShow για αυτό και χρησιμοποιείται για την ανάπτυξη όλων των προγραμμάτων που χρησιμοποιήθηκαν για εξαγωγή των πειραματικών αποτελεσμάτων.

4.2 Σύστημα ιστού

Το ολοκληρωμένο σύστημα αναζήτησης που βασίζεται σε τεχνολογίες ιστού, χρησιμοποιήθηκε κυρίως ως demo περιβάλλον των τεχνικών που περιγράφηκαν στην εργασία αυτή. Χρησιμοποιήθηκε η τεχνική αναζήτησης με βάση τα intra macroblocks κάθε εικόνας που παρουσιάστηκε στην ενότητα 3.8.

4.2.1 Απαιτήσεις

Το σύστημα αναζήτησης ιστού πρέπει να έχει μια ομάδα από «γνωστά» και απο πριν δεικτοδοτημένα βίντεο τα οποία θα λειτουργούν ως βάση δεδομένων για την αναζήτηση. Έτσι, όταν πρέπει να αναζητηθεί κάποιο απόσπασμα, θα συγκρίνεται με τα βίντεο της βάσης δεδομένων. Αυτά τα γνωστά βίντεο είναι ήδη δεικτοδοτημένα και έτσι θα χρειάζεται η δεικτοδότηση μόνο του αποσπάσματος. Επίσης, πρέπει να υπάρχει η δυνατότητα να δεικτοδοτείται από πριν κάποιο απόσπασμα και να εισάγεται και αυτό σε μια ομάδα από «γνωστά» αποσπάσματα, κάτι που βοηθάει πολύ τα πειράματα και την παρουσίαση. Η διεπαφή χρήστη πρέπει να δίνει τη δυνατότητα για εισαγωγή ενός νέου βίντεο τόσο στη βάση δεδομένων των βίντεο όσο και των αποσπασμάτων, καθώς και αναζήτηση τόσο ενός γνωστού αποσπάσματος όσο και ενός αποσπάσματος που δεικτοδοτείται εκείνη τη στιγμή.

³Αντίστοιχα με τα θέματα που έχουν τα Windows με τις βιβλιοθήκες dll

4.2.2 Αρχιτεκτονική

Το σύστημα αναζήτησης ιστού μπορεί ουσιαστικά να διαχωριστεί σε τρία τμήματα: Στη βάση δεδομένων, όπου εισάγονται τα διάφορα βίντεο, στη διαδικασία εισαγωγής νέων στη βάση δεδομένων, και στη διαδικασία αναζήτησης. Η αρχιτεκτονική κάθε τμήματος θα παρουσιαστεί ξεχωριστά.

4.2.2.1 Βάση δεδομένων

Στην ενότητα 3.8.3, παρουσιάστηκαν οι τρόποι αναζήτησης. Στην απλή αναζήτηση, γίνεται η ολίσηση της υπογραφής του αποσπάσματος σε όλες τις δυνατές θέσεις και εξάγεται αυτή με τη μικρότερη διαφορά. Στους άλλους τρόπους, απλώς γίνεται αρχικά κάποια δεικτοδότηση η οποία αποκλείει αρκετές θέσεις. Σε κάθε περίπτωση, θα πρέπει να συγκριθούν πολλές συνεχόμενες θέσεις της βάσης δεδομένων (οι οποίες για παράδειγμα μπορεί να είναι 1000, όσες οι εικόνες ενός αποσπάσματος). Αυτό επαναφέρει τη συζήτηση της ενότητας 3.7.1 για τη δυσκολία ερώτησης της βάσης δεδομένων όταν υπάρχουν πολλές συνεχόμενες εγγραφές.

Έτσι, η βάση δεδομένων θα πρέπει να πάρει μια μορφή η οποία να μη θεωρείται από κάποιους βέλτιστη, ή ακόμα και αντί για βάση δεδομένων να μπορούσαν να χρησιμοποιηθούν απλά αρχεία, χωρίς σημαντική αλλαγή στην απόδοση του συστήματος.

Για να γίνουν καλύτερα κατανοητές οι αποφάσεις αυτές, θα εξεταστεί αρχικά το απλούστερο σχήμα της βάσης δεδομένων στην περίπτωση που παρουσιάζει πρόβλημα, και στη συνέχεια το σχήμα της στην περίπτωση που δεν παρουσιάζει πρόβλημα, αλλά δεν ακολουθεί κατά γράμμα τις αρχές σχεδίασης βάσεων δεδομένων.

Για τα βίντεο της βάσης δεδομένων θα πρέπει να υπάρχουν δύο πίνακες: Ένας πίνακας BINTEO, κάθε εγγραφή του οποίου θα έχει όλα τα στοιχεία ενός βίντεο, για παράδειγμα τίτλος, συντελεστές, διάρκεια, έτος παραγωγής, είδος, κοκ, και ένας πίνακας ΔΕΔΟΜΕΝΑ, με τα πεδία ΚΛΕΙΔΙ.BINTEO, ΑΡΙΘΜΟΣ.ΕΙΚΟΝΑΣ, ΠΟΣΟΣΤΟ. Μια πλειάδα (k, n, m) αυτού του πίνακα θα έχει το ποσοστό m των intra macroblock της εικόνας n του βίντεο με κλειδί k (ώστε να γίνει συσχετισμός με τον πίνακα BINTEO). Το κύριο κλειδί του πίνακα αυτού θα είναι το σύνολο (ΚΛΕΙΔΙ.BINTEO, ΑΡΙΘΜΟΣ.ΕΙΚΟΝΑΣ). Για τη σύγκριση τώρα ενός αποσπάσματος 1000 εικόνων με τα στοιχεία του πίνακα, θα πρέπει, κατά τα πρότυπα της ενότητας 3.7.1 να δημιουργηθεί μια ερώτηση *SQL* στην οποία θα ορίζονται 1000 μεταβλητές που θα διατρέχουν τον πίνακα ΔΕΔΟΜΕΝΑ, με επιπλέον χαρακτηριστικά να είναι συνεχόμενες (δηλαδή η μία μετά την άλλη) και να έχουν μια προς μια τα ποσοστά του αποσπάσματος. Μπορεί να γίνει κατανοητό ότι μετά την ανάλυση της ερώτησης αυτής από το *RDBMS* θα δημιουργήσει χιλιάδες απλά *SELECT* και *JOIN* και, η απόδοση της αναζήτησης θα είναι πολύ κακή.

Η λύση σε αυτό το πρόβλημα, είναι να χρησιμοποιηθεί μόνο ο πίνακας BIN-

4.2. ΣΥΣΤΗΜΑ ΙΣΤΟΥ

ΤΕΟ, στον οποίο θα προστεθεί ακόμη ένα πεδίο με όνομα ΔΕΔΟΜΕΝΑ, και το οποίο θα περιέχει ολόκληρη την υπογραφή του συγκεκριμένου βίντεο σε έναν πίνακα μεταβλητού μεγέθους. Προφανώς αυτή η αλλαγή κάνει εξαιρετικά επίπεδη τη δομή της βάσης δεδομένων, και έτσι, αντί για τη βάση θα μπορούσε να χρησιμοποιηθεί μια σειρά από απλά αρχεία, ένα για κάθε βίντεο. Για την αναζήτηση, θα αναχτάται ολόκληρος ο πίνακας των δεδομένων κάθε βίντεο, και θα γίνεται προγραμματιστικά η ολίσθηση μέσα σε αυτόν, με τους τρόπους που ήδη αναφέρθηκαν στην ενότητα 3.8.3.

Προφανώς, το παραπάνω σχήμα της βάσης είναι πολύ απλό. Στον πίνακα ΒΙΝΤΕΟ θα υπάρχουν και μια σειρά από επιπλέον πεδία, για παράδειγμα πεδία σχετικά με το έτος δημιουργίας του βίντεο, το είδος αυτού κοκ.

Ένα σημείο που θα πρέπει να ξεκαθαριστεί εδώ είναι το τι γίνεται με τους δείκτες (indexes) που θα πρέπει να δημιουργηθούν έτσι ώστε να γίνει η αναζήτηση με τους άλλους τρόπους, πέραν της απλής.

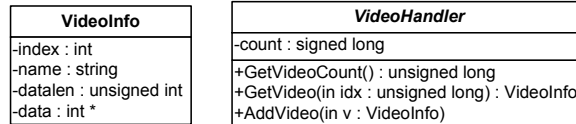
Για την αναζήτηση μεγάλων τιμών, δημιουργήθηκε ένας νέος πίνακας στη βάση, με όνομα ΚΑΤΑΛΟΓΟΣ_ΜΕΓΑΛΩΝ_ΤΙΜΩΝ, ο οποίος είχε τα πεδία ΚΛΕΙΔΙ.ΒΙΝΤΕΟ, ΠΟΣΟΣΤΟ, ΔΕΔΟΜΕΝΑ. Το κύριο κλειδί του πίνακα ήταν το σύνολο (ΚΛΕΙΔΙ.ΒΙΝΤΕΟ, ΠΟΣΟΣΤΟ). Δηλαδή, όταν το κλειδί είχε την τιμή (x, y) τότε το πεδίο ΔΕΔΟΜΕΝΑ περιείχε με τη σειρά όλες τις θέσεις του βίντεο x οι οποίες είχαν ποσοστό intra macroblock ίσο με y . Οι θέσεις αυτές περιέχονται σε ένα διάνυσμα. Έτσι, με ένα απλό SELECT μπορεί να επιλεγεί το τμήμα του καταλόγου που ενδιαφέρει ή ολόκληρος ο κατάλογος για κάποιο συγκεκριμένο βίντεο.

Για την αναζήτηση διαστημάτων, ομοίως δημιουργήθηκε ένας νέος πίνακας στη βάση ο οποίος είχε όνομα ΚΑΤΑΛΟΓΟΣ_ΔΙΑΣΤΗΜΑΤΩΝ. Τα πεδία του πίνακα αυτού ήταν ΚΛΕΙΔΙ.ΒΙΝΤΕΟ, ΜΗΚΟΣ, ΔΕΔΟΜΕΝΑ. Το κύριο κλειδί του πίνακα ήταν το σύνολο (ΚΛΕΙΔΙ.ΒΙΝΤΕΟ, ΜΗΚΟΣ). Δηλαδή, όταν το κλειδί είχε την τιμή (x, y) τότε το πεδίο ΔΕΔΟΜΕΝΑ θα περιέχει με τη σειρά όλες τις θέσεις του βίντεο x οι οποίες ήταν η αρχή ενός διαστήματος μήκους y , οπότε με ένα μόνο SELECT επιστρέφονται όλες οι θέσεις που χρειάζονται για τον έλεγχο.

Μια επιπλέον απαραίτητη προσθήκη στον πίνακα ΒΙΝΤΕΟ έχει να κάνει με το πεδίο ΕΙΔΟΣ, το οποίο καθορίζει το αν το συγκεκριμένο βίντεο είναι απόσπασμα ή κανονικό βίντεο ή και τα δύο. Δηλαδή καθορίζει αν αυτό το βίντεο θα αναζητάται σε άλλα βίντεο ή άλλα βίντεο θα αναζητώνται σε αυτό ή και τα δύο. Μια άλλη λύση θα ήταν να δημιουργηθούν δύο πίνακες, ένας με τα αποσπάσματα και ένας με τα μεγάλα βίντεο, όμως προτιμήθηκε η λύση που περιγράφεται μιας και είναι πιο τυπική στις αρχές των βάσεων δεδομένων.

Στο σχήμα 4.1 φαίνεται το διάγραμμα τάξεων για το επίπεδο πρόσβασης στη βάση δεδομένων. Υπάρχει η τάξη VideoInfo η οποία ενθυλακώνει ένα βίντεο, δηλαδή μια πλειάδα της βάσης δεδομένων και έχει ιδιότητες όπως το κλειδί του βίντεο (index), το όνομα (name), το μήκος της υπογραφής και την ίδια την υπογραφή (datalen και data). Εκτός από αυτές υπάρχουν και άλλες ιδιότητες που δε φαίνονται στο σχήμα, σε συνάρτηση με τα περιεχόμενα της

ΚΕΦΑΛΑΙΟ 4. ΣΥΣΤΗΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ



Σχήμα 4.1: Επίπεδο πρόσβασης στη βάση δεδομένων

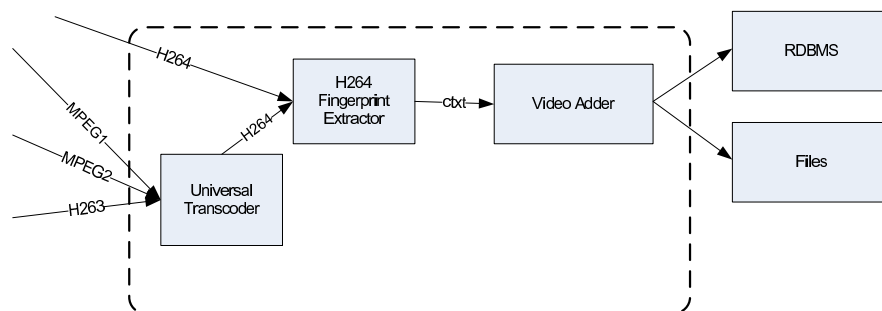
βάσης δεδομένων, για παράδειγμα παραπάνω πληροφορίες για το βίντεο όπως το έτος δημιουργίας ή το είδος του, και φυσικά οι κατάλογοι για τις μεγάλες τιμές και τα διαστήματα.

Η VideoHandler χρησιμοποιεί αντικείμενα της τάξης VideoInfo και έχει μια ιδιότητα που καθορίζει τον αριθμό των βίντεο που υπάρχουν ήδη στη βάση δεδομένων, αλλά και μεθόδους για την προσθήκη ενός νέου και την ανάκληση ενός ήδη υπάρχοντος VideoInfo.

Με την παραπάνω υλοποίηση μπορούν να χρησιμοποιηθούν πολλές, διαφορετικών ειδών βάσεις δεδομένων ταυτόχρονα, αφού κάθε μία θα είναι ένα αντικείμενο της VideoHandler. Έτσι, μπορεί κάποιος είτε να δημιουργήσει ένα αντικείμενο της VideoInfo από την αρχή (για παράδειγμα με δεικτοδότηση ενός βίντεο) και να το αποθηκεύσει στη βάση δεδομένων που χρησιμοποιείται, είτε μπορεί να εξετάσει όλα τα βίντεο που περιέχει η βάση δεδομένων με μορφή αντικειμένων της VideoInfo.

4.2.2.2 Δεικτοδότηση και εισαγωγή

Επιλέχτηκε για λόγους ευκολίας υλοποίησης οι εργασίες της δεικτοδότησης και της εισαγωγής στη βάση δεδομένων να γίνονται από διαφορετικά προγράμματα, όπως φαίνεται στο σχήμα 4.2.



Σχήμα 4.2: Δεικτοδότηση και εισαγωγή στη βάση

Έτσι, ανάλογα με το είδος του βίντεο, αυτό θα περάσει είτε από ένα οικουμενικό εργαλείο το οποίο θα μετατρέπεται σε H.264 και στη συνέχεια θα εξάγεται η υπογραφή του, είτε, αν είναι H.264 θα εξάγεται απευθείας η υπογραφή του. Έτσι, σε κάθε περίπτωση η υπογραφή θα εξάγεται από ένα αρχείο h.264 με τη βοήθεια του προγράμματος H264GetCtxt το οποίο χρησιμοποιή-

θηκε και στο κεφάλαιο 2 και αναλύεται παραπάνω στο παράρτημα. Η υπογραφή αυτή θα αποθηκεύεται στο σκληρό δίσκο σαν αρχείο, το οποίο απλώς θα περιέχει τον αριθμό της εικόνας και το ποσοστό σε *intra macroblock* αυτής. Αυτό το αρχείο κατόπιν θα χρησιμοποιείται από το πρόγραμμα προσθήκης βίντεο στη βάση δεδομένων, μαζί με τυχόν άλλες πληροφορίες που υπάρχουν έτσι ώστε να προστεθεί το βίντεο. Η προσθήκη θα μπορεί να γίνει είτε σε αρχεία είτε σε κανονικό RDBMS λόγω των δυνατοτήτων που δίδεται από την αρχιτεκτονική.

Φυσικά, αυτός ο διαχωρισμός σε στάδια έγινε στα πλαίσια της εργασίας, για την ευκολότερη υλοποίηση και την εξαγωγή πειραματικών αποτελεσμάτων. Υπό άλλες συνθήκες, το πρόγραμμα της προσθήκης στη βάση θα ήταν αυτό που περικλείεται από τη διακεκομμένη γραμμή του σχήματος: Ανάγνωση κάθε είδους αρχείου και απευθείας εισαγωγή αυτού στη βάση δεδομένων.

Για την εισαγωγή, απλώς αρχικοποιείται ένα αντικείμενο της *VideoHandler*, το οποίο να δείχνει είτε σε βάση δεδομένων είτε σε αρχεία στο δίσκο, διαβάζεται το αρχείο *etxt* με την υπογραφή του βίντεο, δημιουργείται ένα νέο αντικείμενο της *VideoInfo* με τα στοιχεία του βίντεο και εισάγεται με τη μέθοδο *AddVideo*.

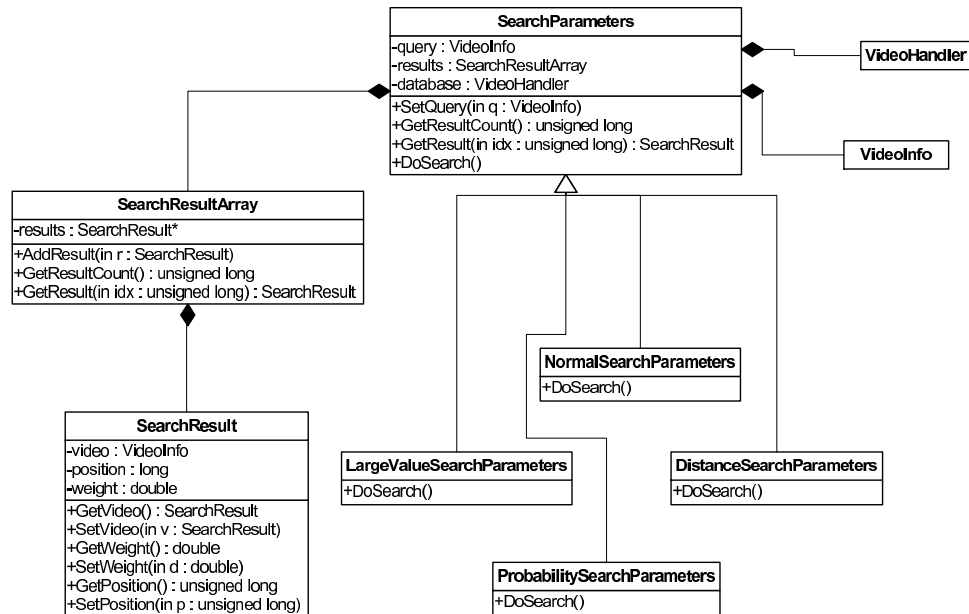
4.2.2.3 Αναζήτηση

Στο σχήμα 4.3 φαίνεται το διάγραμμα τάξεων για την αναζήτηση. Η σημαντικότερη τάξη του προγράμματος είναι η αφηρημένη τάξη *SearchParameters*. Αυτή, έχει ξεχωριστές υλοποιήσεις ανάλογα με τον τρόπο που γίνεται η αναζήτηση, δηλαδή τις *NormalSearchParameters* για την απλή αναζήτηση με ολίσθηση, *LargeValueSearchParameters* για την αναζήτηση μεγάλων τιμών, *DistanceSearchParameters* για την αναζήτηση με αποστάσεις και *ProbabilitySearchParameters* για την πιθανοτική αναζήτηση. Περιέχει ένα αντικείμενο της τάξης *VideoInfo* που ενθυλακώνει το απόσπασμα, ένα αντικείμενο της τάξης *VideoHandler* που ενθυλακώνει μια σύνδεση προς τη βάση δεδομένων όπου θα αναζητηθεί το απόσπασμα, καθώς και ένα αντικείμενο της τάξης *SearchResultArray* με τα αποτελέσματα. Όλες οι ιδιότητες της *SearchParameters* πλην της *SearchResultArray* θα πρέπει να αρχικοποιηθούν πριν την κλήση της αφηρημένης μεθόδου *DoSearch*. Αφού τελειώσει η *DoSearch*, το *SearchResultArray* θα περιέχει όλα τα αποτελέσματα της αναζήτησης.

Το *SearchResultArray* είναι ένας πίνακας που περιέχει αντικείμενα της *SearchResult*. Η διαφορά του από έναν απλό πίνακα είναι ότι η μέθοδος *AddResult* εισάγει το αποτέλεσμα στη σωστή σειρά, ανάλογα με το βάρος αυτού. Το *SearchResult* έχει σαν ιδιότητες, εκτός από ένα *VideoInfo* με τα στοιχεία του αποτελέσματος, το *position* με τη θέση όπου βρέθηκε το απόσπασμα καθώς και το *weight* με το «βάρος» αυτού. Βάρος είναι ουσιαστικά το πόσο βέβαιο είναι ότι πράγματι βρέθηκε το απόσπασμα σε εκείνο το σημείο και έχει να κάνει με το D_{intra} της ενότητας 3.8.3.

Αυτό που συμβαίνει λοιπόν είναι ότι όταν κληθεί το πρόγραμμα της αναζήτησης, χρειάζεται ως παραμέτρος ένα απόσπασμα καθώς και μια σύνδεση με τη βάση δεδομένων. Το απόσπασμα θα πρέπει να είναι ήδη δεικτοδοτημένο, όπως

ΚΕΦΑΛΑΙΟ 4. ΣΥΣΤΗΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ



Σχήμα 4.3: Διάγραμμα τάξεων για αναζήτηση

συμβαίνει και με το πρόγραμμα της εισαγωγής στη βάση, για ευκολία υλοποίησης. Στη συνέχεια, θα δημιουργηθεί ένα νέο αντικείμενο της τάξης SearchParameters, ανάλογα φυσικά με τον τρόπο αναζήτησης, δηλαδή με μεγάλες τιμές, με αποστάσεις κοκ. Το αντικείμενο αυτό θα αρχικοποιηθεί με το απόσπασμα και τη βάση δεδομένων, καθώς και τυχόν άλλες παραμέτρους που μπορεί να έχουν οι μέθοδοι αναζήτησης, και θα κληθεί η μέθοδος DoSearch. Αφού επιστρέψει αυτή, το αντικείμενο της SearchParameters περιέχει τα αποτελέσματα της αναζήτησης και μπορούν αυτά να ανακληθούν με τις GetResultCount και GetResult.

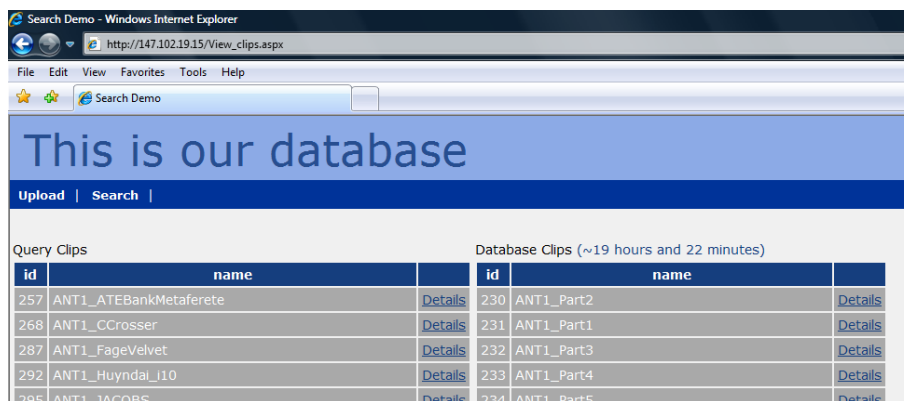
4.2.2.4 Διεπαφή χρήστη

Η διεπαφή χρήστη του συστήματος αναζήτησης ιστού ουσιαστικά αποτελείται από πέντε ιστοσελίδες: Την ιστοσελίδα παρουσίασης όλων των βίντεο, την ιστοσελίδα παρουσίασης καθενός βίντεο ξεχωριστά, την ιστοσελίδα ανεβασματος βίντεο, την ιστοσελίδα αναζήτησης και την ιστοσελίδα των αποτελεσμάτων.

4.2.2.4.1 Συνολική παρουσίαση Στην ιστοσελίδα αυτή φαίνονται όλα τα βίντεο της βάσης δεδομένων. Τα βίντεο αυτά χωρίζονται σε δύο επιμέρους κατηγορίες, όπως φαίνεται στο σχήμα 4.4: Στα αποσπάσματα και τα πλήρη βίντεο. Τα πρώτα είναι συνήθως μικρά βίντεο τα οποία μπορούν να αναζητηθούν μέσα στα δεύτερα, μεγαλύτερα βίντεο. Το είδος κάθε βίντεο μπορεί να καθοριστεί από το χρήστη κατά την εισαγωγή αυτού στη βάση. Ένα βίντεο μπορεί να ανήκει και στις δύο κατηγορίες, δηλαδή να μπορεί να αναζητηθεί μέσα σε άλλα,

4.2. ΣΥΣΤΗΜΑ ΙΣΤΟΥ

μεγαλύτερα του, αλλά και να μπορούν άλλα, μικρότερα του να αναζητηθούν σε αυτό. Για κάθε βίντεο υπάρχει μια σύνδεση με όνομα Details η οποία οδηγεί στην αναλυτική παρουσίαση αυτού.



Query Clips			Database Clips (~19 hours and 22 minutes)		
id	name	Details	id	name	Details
257	ANT1_ATEBankMetaferete	Details	230	ANT1_Part2	Details
268	ANT1_CCrosser	Details	231	ANT1_Part1	Details
287	ANT1_PageVelvet	Details	232	ANT1_Part3	Details
292	ANT1_Huyndai_I10	Details	233	ANT1_Part4	Details
295	ANT1_JACOBS	Details	234	ANT1_Part5	Details

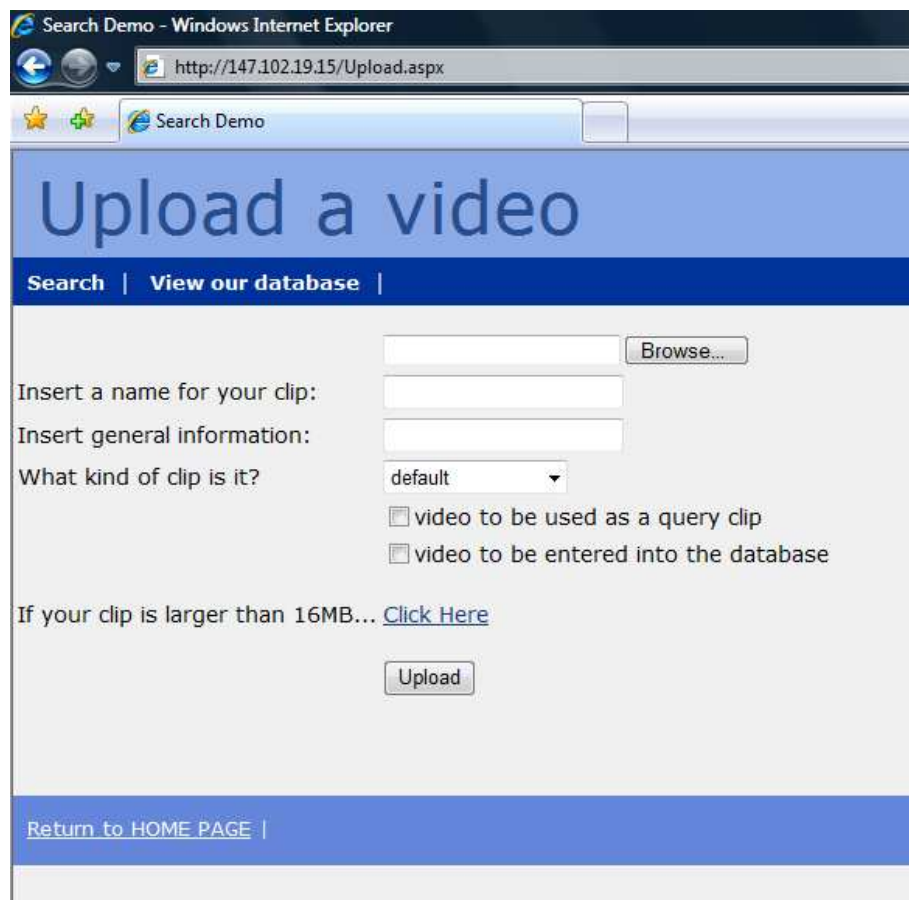
Σχήμα 4.4: Συνολική παρουσίαση των βίντεο

4.2.2.4.2 Αναλυτική παρουσίαση Εδώ φαίνονται αναλυτικότερες πληροφορίες για κάθε βίντεο. Η μορφή της σελίδας είναι ίδια και για τα αποσπάσματα και για τα μεγάλα βίντεο. Υπάρχει η δυνατότητα να κατέβει το βίντεο αυτό στον υπολογιστή κάθε χρήστη με τη βοήθεια της σύνδεσης Download αλλά και να αφαιρεθεί από τη βάση με το κουμπί Remove.

Τέλος, για το βίντεο αυτό φαίνονται και οι πρώτες εικόνες από κάθε πλάνο του. Η ανίχνευση αυτών γίνεται με τη μέθοδο των intra macroblock όπως αναλύθηκε στο κεφάλαιο 2, αφού ούτως ή άλλως, κάθε φορά που εισάγεται ένα βίντεο στη βάση δεδομένων γίνεται ανίχνευση των ποσοστών intra σε κάθε εικόνα και έτσι είναι πολύ εύκολη η εξαγωγή των αλλαγών πλάνου. Η παράθεση αυτών των εικόνων βοηθά πολύ στη γρήγορη οπτική ανασκόπηση του περιεχομένου του βίντεο.

4.2.2.4.3 Ανέβασμα Με τη σελίδα αυτή ο χρήστης μπορεί να ανεβάσει ένα νέο βίντεο στη βάση δεδομένων. Το βίντεο αυτό μπορεί να είναι είτε απόσπασμα, είτε πλήρες βίντεο, είτε και τα δύο, ανάλογα με την επιλογή του χρήστη.

Επίσης μπορούν να εισαχθούν κάποιες γενικές πληροφορίες για το βίντεο. Ένα σημείο που πρέπει να προσεχθεί έχει να κάνει με το μέγεθος του βίντεο σε bytes. Αν το βίντεο που πρόκειται να εισαχθεί είναι μεγάλο (κάτι που είναι αρκετά πιθανό να συμβεί στα βίντεο εντός των οποίων γίνεται η αναζήτηση) τότε θα δημιουργηθεί πρόβλημα με το χρόνο αναμονής του χρήστη μέχρι να ανέβει το βίντεο στο server και φυσικά μέχρι να γίνει η δεικτοδότηση του.



Σχήμα 4.5: Ανέβασμα βίντεο στη βάση δεδομένων

4.2.2.4.4 Αναζήτηση Στο σημείο αυτό γίνεται η πραγματική αναζήτηση στη βάση δεδομένων: Αποτελείται από δύο στήλες, την αριστερή όπου υπάρχουν τα αποσπάσματα βίντεο που έχουν ανεβεί προηγουμένως στη βάση δεδομένων και τη δεξιά που περιλαμβάνει τα πλήρη βίντεο της βάσης δεδομένων, τα βίντεο δηλαδή εντός των οποίων θα αναζητηθούν τα αποσπάσματα.

The screenshot shows a web interface for searching video clips. At the top, there is a blue header with the text "Search...". Below the header, there are two navigation links: "Upload" and "View our database". The main content area is divided into two sections. The left section is titled "Select query clips to be used for search" and contains a table with columns "Select", "id", and "name". The right section is titled "Select database videos to be searched" and contains a table with columns "Select", "id", and "name". Below these tables, there is a question "Do you want to search for another clip?" followed by a file path "C:\Documents and Settings" and a "Browse..." button. A "Search" button is also present. At the bottom, there is a link "Return to HOME PAGE".

Select	id	name
<input checked="" type="checkbox"/>	61	TellasCowboy
<input checked="" type="checkbox"/>	64	KeryKapnisto
<input type="checkbox"/>	65	CitroenC4
<input type="checkbox"/>	66	AllStarSxoleio
<input type="checkbox"/>	67	KinderBueno
<input checked="" type="checkbox"/>	68	KitKat
<input type="checkbox"/>	70	Kuprou
<input type="checkbox"/>	72	VodafoneSpiti

Select	id	name
<input checked="" type="checkbox"/>	47	MEGA.ctxt
<input checked="" type="checkbox"/>	48	STAR.ctxt
<input checked="" type="checkbox"/>	49	ANT1.ctxt

Σχήμα 4.6: Γραφικό περιβάλλον για αναζήτηση

Επίσης, υπάρχει και η δυνατότητα ο χρήστης να ανεβάσει ένα βίντεο που βρίσκεται στο σκληρό του δίσκο έτσι ώστε να αναζητηθεί αυτό στη βάση δεδομένων. Η συγκεκριμένη επιλογή θα πάρει περισσότερο χρόνο από την αναζήτηση των ήδη ανεβασμένων αποσπασμάτων, αφού θα χρειαστεί επιπλέον και η δεικτοδότηση του νέου βίντεο.

Μπορεί να γίνει επιλογή των βίντεο στη δεξιά στήλη, δηλαδή η αναζήτηση να γίνει μόνο σε συγκεκριμένα βίντεο της βάσης δεδομένων, το ίδιο και στην αριστερή, δηλαδή να αναζητηθούν το ένα μετά το άλλο περισσότερα από ένα βίντεο.

4.2.2.4.5 Αποτελέσματα Όταν πατηθεί το κουμπί search της αναζήτησης, θα εμφανιστεί η σελίδα των αποτελεσμάτων:

Για κάθε ένα από τα βίντεο που επιλέχθηκαν εμφανίζονται ξεχωριστά τα αποτελέσματα, δηλαδή το πού βρέθηκε αλλά και το πόσο χρόνο πήρε η αναζήτηση. Για κάθε ένα απόσπασμα γίνεται μια ξεχωριστή αναζήτηση στο τμήμα της βάσης δεδομένων που επιλέχτηκε.

ΚΕΦΑΛΑΙΟ 4. ΣΥΣΤΗΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ

The screenshot shows a web interface for search results. At the top, there is a blue header with the text "The results" and a small yellow box containing the text "JavaScript Switch Statement" and the URL "http://www.w3schools.com/js/js_switch.asp". Below the header is a navigation bar with links for "Upload", "Search", and "View our database". The main content area is divided into four sections, each representing a search query. Each section includes the query name, the search results, and the search time.

Query clip	Search Results	Search time
TellasCowboy	Found at: MEGA.ctxt at : 00:04:38 with loss: 4 Found at: STAR.ctxt at : 01:07:35 with loss: 7	843,75 milliseconds
KeryKapnisto	Not found	343,75 milliseconds
KitKat	Found at: MEGA.ctxt at : 00:30:53 with loss: 3	328,125 milliseconds
xrusos-odhgos.ts.avi	Not found	328,125 milliseconds

Indexing time = 6875 milliseconds.

[Go back](#) |

Σχήμα 4.7: Τα αποτελέσματα της αναζήτησης

4.2.3 Υλοποίηση

Η υλοποίηση του κυρίως συστήματος αναζήτησης έγινε στη γλώσσα προγραμματισμού C++. Η γραφική διεπαφή υλοποιήθηκε με τη βοήθεια του συστήματος ASP.NET και της C#, ενώ για την επικοινωνία μεταξύ της γραφικής διεπαφής που ήταν σε C# και του συστήματος αναζήτησης που ήταν σε C++ δημιουργήθηκε μια απλή βιβλιοθήκη την οποία καλούσε το C# πρόγραμμα.

Η υλοποίηση του κυρίως συστήματος αναζήτησης σε C++ είναι αρκετά απλή όπως μπορεί να φανεί και από την περιγραφή της αρχιτεκτονικής. Οι τάξεις που κάνουν την αναζήτηση `NormalSearchParameters`, `LargeValueSearchParameters`, `DistanceSearchParameters` και `ProbabilitySearchParameters` υλοποιούν αντίστοιχα τις μεθόδους που περιγράφηκαν στην ενότητα 3.8.3.

Αναφορικά με την υλοποίηση της γραφικής διεπαφής σε C#, όπως ήδη αναφέρθηκε και στην αρχιτεκτονική, χρησιμοποιήθηκαν διάφορες τεχνολογίες για την όσο το δυνατόν γρηγορότερη υλοποίηση.

Για το τμήμα της εξαγωγής των υπογραφών κάθε αρχείου βίντεο και τη δημιουργία του `ctxt` αρχείου, χρησιμοποιήθηκε το εργαλείο `Mplayer` [MPL] ως οικουμενικός μετατροπέας σε H.264 και το `FFmpeg` [FFM] για τη δημιουργία της υπογραφής. Με τη βοήθεια της βιβλιοθήκης `FFmpeg` γράφτηκε ένα πρόγραμμα που διαβάζει ένα H.264 αρχείο και για κάθε εικόνα αυτού γράφει σε ένα αρχείο `ctxt` στο δίσκο τον αριθμό της και το ποσοστό `intra macroblocks` αυτής, παρομοίως με την ενότητα 2.4.2.

Το `Mplayer` δημιουργούσε στο δίσκο ένα ενδιάμεσο H.264 αρχείο το οποίο και χρησιμοποιούσε το `FFmpeg` αφού τελείωνε το `Mplayer`.

4.3 Σύστημα κονσόλας

Για την ευκολότερη εξαγωγή πειραματικών αποτελεσμάτων, υλοποιήθηκε ένα σύστημα βασισμένο σε διεπαφή κειμένου το οποίο υλοποιούσε όλες τις μεθόδους αναζήτησης που παρουσιάστηκαν στο κεφάλαιο 3.

Το σύστημα αυτό ουσιαστικά αποτελείται από δύο προγράμματα, ένα για την προσθήκη βίντεο στη βάση δεδομένων προς αναζήτηση που ονομάστηκε `TextVideoAdd` και ένα για την αναζήτηση που ονομάστηκε `TextVideoSearch`.

Χαρακτηριστικό του συστήματος κονσόλας που αναπτύχθηκε είναι ότι χρησιμοποιούνται μόνο αρχεία `ctxt` για την αναζήτηση, δηλαδή τα βίντεο έχουν ήδη δεικτοδοτηθεί με τη βοήθεια των μεθόδων που παρουσιάστηκαν προηγουμένως⁴.

Η αρχιτεκτονική του συστήματος κονσόλας μπορεί να χωριστεί σε δύο τμήματα: Ένα τμήμα για την πιθανοτική και την αναζήτηση με `intra macroblocks` και ένα τμήμα για την αναζήτηση με βάση δεδομένων. Αυτός ο διαχωρισμός ήταν απαραίτητος αφού το τμήμα για την αναζήτηση με βάση δεδομένων θα έπρεπε να χρησιμοποιεί διαφορετική βάση δεδομένων.

⁴Δηλαδή με μετατροπή τους σε κατάλληλο H.264 αν δεν είναι ήδη και στη συνέχεια με εξαγωγή από αυτό του `ctxt` με το `H264GetCtxt`

ΚΕΦΑΛΑΙΟ 4. ΣΥΣΤΗΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ

Ετσι, η αρχιτεκτονική του τμήματος για αναζήτηση πιθανοτήτων και intra macroblock συστήματος είναι παρόμοια με αυτήν του συστήματος ιστού που παρουσιάστηκε στην προηγούμενη ενότητα, με τη διαφορά φυσικά ότι όλες οι λειτουργίες γίνονται από τη γραμμή εντολών και όχι από σελίδες ιστού. Ουσιαστικά, τα προγράμματα TextVideoAdd και TextVideoSearch χρησιμοποιούν τις βιβλιοθήκες αναζήτησης και τη βάση δεδομένων του συστήματος αναζήτησης ιστού, όπως αυτά παρουσιάστηκαν στο προηγούμενο κεφάλαιο.

Αντίστοιχα, στο τμήμα της αναζήτησης βάσης δεδομένων, χρησιμοποιήθηκε η αρχιτεκτονική που αναπτύχθηκε στην ενότητα 3.7. Η αναζήτηση γίνεται με βάση κυρίως ερωτήσεις στη βάση δεδομένων, όπως αναπτύχθηκε διεξοδικά στην ενότητα 3.7, ενώ δημιουργήθηκαν βιβλιοθήκες για τη δημιουργία των ερωτημάτων προς τη βάση σύμφωνα με όσα αναπτύχθηκαν στην ίδια ενότητα.

Τα αποτελέσματα των μετρήσεων που έγιναν θα παρουσιαστούν αναλυτικά στο επόμενο κεφάλαιο για όλες τις περιπτώσεις.

4.4 Συνεχής αναζήτηση

Το σύστημα συνεχούς αναζήτησης έχει τη δυνατότητα να αναζητά μια σειρά από αποσπάσματα που επιλέγει ο χρήστης σε ένα «ζωντανό» βίντεο, δηλαδή ένα βίντεο χωρίς συγκεκριμένη αρχή και τέλος. Χαρακτηριστική του εφαρμογή η αναζήτηση και επισήμανση διαφημίσεων σε τηλεοπτικό πρόγραμμα.

4.4.1 Απαιτήσεις

Γενικά το σύστημα θα πρέπει να αναζητά κάποια μικρά αποσπάσματα μέσα σε ένα ή περισσότερα βίντεο που παίζουν συνεχώς. Κάθε ένα από τα συνεχόμενα βίντεο αντιπροσωπεύει μια διαφορετική πηγή εκπομπής βίντεο (κανάλι) και θα είναι κάποιος τηλεοπτικός σταθμός επίγειας, ψηφιακής, δορυφορικής ή δικτυακής τηλεόρασης. Όταν βρεθεί ένα από τα αποσπάσματα σε κάποια από τις πηγές εκπομπής τότε θα εισάγεται σε μια βάση δεδομένων μια εγγραφή για την ώρα, το απόσπασμα και την πηγή στην οποία αυτό βρέθηκε.

Πρέπει να υπάρχει μια απλή διεπαφή χρήστη μέσω της οποίας ο χρήστης θα μπορεί να καθορίσει το ποιά θα είναι τα αποσπάσματα που θα αναζητούνται, καθώς και ποιά θα είναι τα κανάλια που θα παρακολουθούνται. Θα υπάρχουν όρια, ανάλογα με τις δυνατότητες του υπολογιστή που φιλοξενεί το σύστημα τόσο στον αριθμό των αποσπασμάτων που θα αναζητούνται ταυτόχρονα όσο και στον αριθμό των καναλιών που θα παρακολουθούνται ταυτόχρονα. Σε περίπτωση που κάποιο από αυτά τα όρια ξεπεραστεί, θα μπορεί να χρησιμοποιηθεί ένα αντίγραφο αυτού του συστήματος σε ένα δεύτερο υπολογιστή για την παρακολούθηση ακόμα περισσότερων καναλιών ή αποσπασμάτων.

4.4.2 Αρχιτεκτονική

Παρακάτω αναλύεται η αρχιτεκτονική του συστήματος συνεχούς αναζήτησης. Επειδή για την αναζήτηση χρησιμοποιείται τελικά η μέθοδος των intra macroblock, πολλά στοιχεία από αυτά που αναφέρονται είναι κοινά με αυτά της ενότητας 4.2.

4.4.2.1 Διαχωρισμός μεταξύ πηγών και αναζήτησης

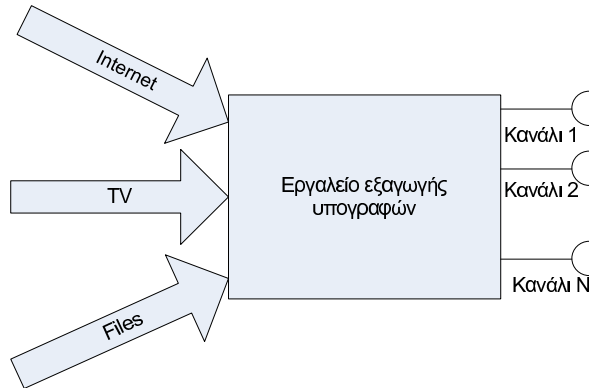
Ένα σημαντικό πρόβλημα του συστήματος συνεχούς αναζήτησης έχει να κάνει με τα διαφορετικά είδη πηγών βίντεο που υπάρχουν και θα έπρεπε ένα πλήρες σύστημα να υποστηρίζει. Οι πηγές αυτές είναι:

- Απλά αρχεία βίντεο αποθηκευμένα στο σκληρό δίσκο: Παρότι η περίπτωση αυτή δε χρειάζεται τη συνεχή αναζήτηση, θα πρέπει να υποστηρίζεται έτσι ώστε να μπορούν να γίνουν διάφορες δοκιμές αλλά και να αποδεικνύεται εύκολα ότι το σύστημα δουλεύει σωστά.
- Βίντεο από το διαδίκτυο: Αυτή είναι μια πολύ μεγάλη κατηγορία μιας και υπάρχουν πάρα πολλά διαφορετικά πρωτόκολλα με τη βοήθεια των οποίων γίνεται εκπομπή βίντεο στο διαδίκτυο, για παράδειγμα το πολύ γνωστό RTSP - Real Time Streaming Protocol της IETF [RTS] αλλά και το MMS - Microsoft Media Server Protocol της Microsoft [MMS].
- Τηλεόραση: Το σύστημα θα πρέπει να υποστηρίζει την παρακολούθηση πηγών αναλογικής, ψηφιακής, αλλά και δορυφορικής τηλεόρασης. Για να υπάρχει τέτοια υποστήριξη θα πρέπει να χρησιμοποιηθούν ειδικές κάρτες (TV-tuner) που να υποστηρίζουν τα διαφορετικά αυτά είδη τηλεόρασης. Συνήθως αυτές οι κάρτες συνοδεύονται από προγράμματα τα οποία μπορούν να εκπέμψουν το τηλεοπτικό πρόγραμμα σε μια δικτυακή διεύθυνση έτσι ώστε να μην είναι απαραίτητη η χρήση αυτών στον ίδιο υπολογιστή.

Λόγω όλων αυτών των τύπων πηγών που υπάρχουν, λήφθηκε η απόφαση να χρησιμοποιηθεί ένα γενικό εργαλείο το οποίο θα «συνδέεται» με κάθε είδους πηγή και θα εξάγει αυτόματα την υπογραφή, δηλαδή το ποσοστό των intra macroblock για κάθε μια εικόνα. Η ακολουθία αυτή των υπογραφών, που είναι τελικά μια ακολουθία από ακέραιους 0 ως 100 θα είναι διαθέσιμη μέσα από ένα TCP socket του υπολογιστή αυτού. Για κάθε κανάλι κάθε πηγής θα υπάρχει ένα διαφορετικό socket από οποίο συνεχώς θα αποστέλλονται ακέραιοι με τα ποσοστά των intra macroblock, έτσι όπως φαίνεται και στο σχήμα 4.8.

4.4.2.2 Η συνεχής αναζήτηση

Στην ενότητα αυτή θα εξηγηθεί ο τρόπος με τον οποίο γίνεται η συνεχής αναζήτηση: Χρησιμοποιήθηκε η απλή αναζήτηση με ολίση της μεθόδου των intra macroblock. Για την αναζήτηση θα πρέπει να «μαζευτεί» αρχικά μια σειρά



Σχήμα 4.8: Εξαγωγή υπογραφής από διαφορετικές πηγές

από ποσοστά intra macroblock εικόνων και να αποθηκευθεί σε έναν buffer και στη συνέχεια, κάθε ένα από τα αποσπάσματα τα οποία αναζητούνται να αναζητηθεί μέσα σε αυτόν τον buffer. Στη συνέχεια ο buffer αδειάζει και γεμίζει με την επόμενη σειρά από τιμές. Πρέπει εδώ να ληφθεί υπ όψιν και η περίπτωση όπου κάποιοι απόσπασμα τυχαίνει να βρίσκεται μεταξύ δύο γεμισμάτων buffer άρα δε θα βρεθεί ούτε την πρώτη ούτε τη δεύτερη φορά.

Έστω λοιπόν η υπογραφή για το απόσπασμα $Q = \{q_1, q_2, \dots, q_m\}$, αποτελούμενο από m εικόνες, και το συνεχές βίντεο $P = \{p_1, p_2, \dots\}$. Το συνεχές βίντεο χωρίζεται στα τμήματα $P_1 = \{p_1, p_2, \dots, p_\sigma\}$, $P_2 = \{p_{\tau+1}, p_2, \dots, p_{\tau+\sigma}\}$, ... κάθε ένα εκ των οποίων έχει μήκος σ εικόνες και προχωράει κατά τ εικόνες σε σχέση με το προηγούμενο. Πρέπει $\tau < \sigma$ έτσι ώστε να ληφθεί υπ όψιν η περίπτωση των συνόρων.

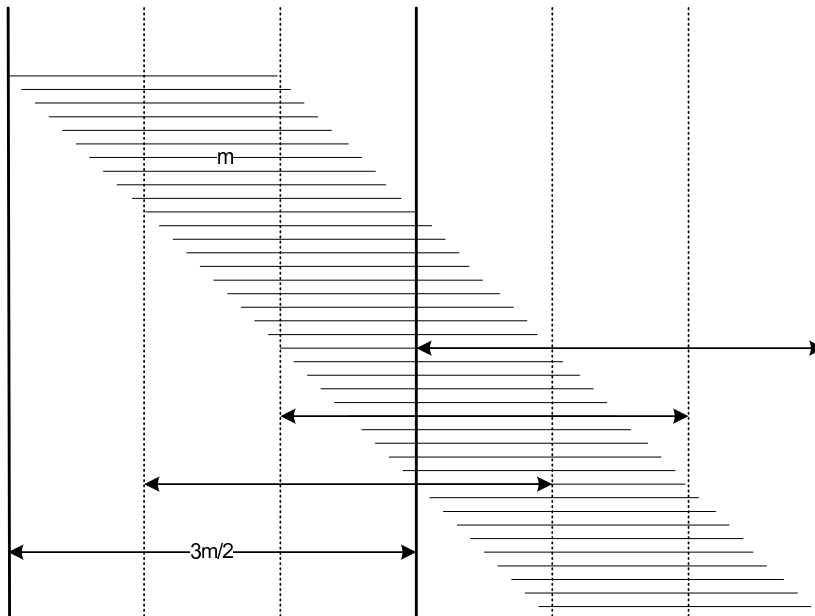
Η επιλογή που έγινε για τα σ, τ είναι $\sigma = 3/2m, \tau = m/2$. Με αυτή την επιλογή είναι σίγουρο ότι το απόσπασμα δε θα χαθεί, σε όποια θέση και αν βρίσκεται. Η απόδειξη είναι διαισθητική και φαίνεται εύκολα στο σχήμα 4.9: Κάθε δυνατή θέση του αποσπάσματος μήκους m θα βρεθεί μέσα σε κάποιο buffer μήκους $3m/2$ αν αυτός προχωρά κατά $m/2$ κάθε φορά όπως φαίνεται και στο σχήμα.

Σε περίπτωση που αναζητούνται περισσότερα αποσπάσματα κάθε φορά, τότε το m που χρησιμοποιήθηκε για τους ορισμούς των σ και τ θα πρέπει να είναι ίσο με το μήκος του μεγαλύτερου εξ αυτών. Για τα αποσπάσματα που έχουν μικρότερο μήκος από το m υπάρχει η πιθανότητα να βρεθεί κάποιο απόσπασμα δύο φορές σε δύο γειτονικές θέσεις, κάτι όμως που είναι εύκολο να προβλεφθεί.

4.4.2.3 Η αρχιτεκτονική της εφαρμογής

Η εφαρμογή γράφτηκε σε Visual C++ και μάλιστα χρησιμοποιώντας τη βιβλιοθήκη MFC για το γραφικό περιβάλλον. Εδώ θα αναλυθούν μόνο οι βασικές τάξεις που χρησιμοποιήθηκαν για τη συνεχή αναζήτηση και όχι τυχόν λεπτομέρειες υλοποίησης σχετικές με τη Visual C++, το MFC ή τα Windows. Επίσης

4.4. ΣΥΝΕΧΗΣ ΑΝΑΖΗΤΗΣΗ



Σχήμα 4.9: Εξήγηση λειτουργίας συνεχούς αναζήτησης

δε θα αναφερθούν οι λεπτομέρειες που αφορούν την αναζήτηση ενός μικρότερου αποσπάσματος σε ένα μεγαλύτερο, αφού χρησιμοποιούνται οι τεχνικές της μεθόδου των intra macroblock.

Στο σύστημα υπάρχουν ουσιαστικά τρία ξεχωριστά υποσυστήματα που τρέχουν ταυτόχρονα: Το GUI που λαμβάνει τα διάφορα συμβάντα του χρήστη όπως κλικ ποντικιού, μετακινήσεις παραθύρων, ακύρωση αναζήτησης κλπ, το socket το οποίο λαμβάνει συνεχώς τις πληροφορίες και, επειδή είναι ασύγχρονο θα χρησιμοποιεί το ίδιο thread με το GUI, δηλαδή τα μηνύματα από το socket θα παραδίδονται και θα εξυπηρετούνται από την ίδια ουρά, και τέλος το σύστημα που αναζητά συνεχώς τα αποσπάσματα στον buffer το οποίο και τρέχει σε δικό του thread.

Στο σχήμα 4.10 φαίνεται ένα τμήμα του διαγράμματος τάξεων για τη συνεχή αναζήτηση. Οι τάξεις που φαίνονται εδώ είναι οι εξής:

- CIndexData: Είναι η τάξη που αποθηκεύει την ακολουθία των δεδομένων στα οποία γίνεται η αναζήτηση. Δηλαδή είναι ουσιαστικά ο buffer που αναφέρθηκε προηγουμένως. Γενικά έχει γραφτεί έτσι ώστε να είναι όσο γίνεται πιο εύχρηστη, για αυτό και έχει πολλές μεθόδους. Οι ιδιότητες της έχουν να κάνουν με το μέγεθος των δεδομένων που έχει ήδη μέσα, με το μέγιστο μέγεθος των δεδομένων που χωράει (το σ), με τα ίδια τα δεδομένα που είναι ένας απλός πίνακας από ακέραιες τιμές από 0 ως 100 καθώς και με δύο σηματοδότες: Έναν που καθορίζει το πότε ο buffer είναι γεμάτος και έναν που καθορίζει το πότε δεν είναι γεμάτος. Ο μεν πρώτος σηματοδοτεί την έναρξη της αναζήτησης, ο δεύτερος σημα-

ΚΕΦΑΛΑΙΟ 4. ΣΥΣΤΗΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ

τοδοτεί τη μεταφορά δεδομένων από το socket στον buffer, μέχρι αυτός να γεμίσει ξανά. Σημειώνεται ότι ο buffer γεμίζει σιγά-σιγά, ανάλογα με το ρυθμό που έρχονται τα δεδομένα της υπογραφής από το socket, αλλά αδειάζει πάντοτε κατά $\tau (= \sigma/3)$. Οι μέθοδοι `AddData` και `GetIndexData` χρησιμοποιούνται για να προστεθούν δεδομένα ή να αναγνωστούν δεδομένα από τον buffer, η `GetIndexDataLen` επιστρέφει το μήκος του buffer ενώ η `ReleaseIndexData` τον αδειάζει.

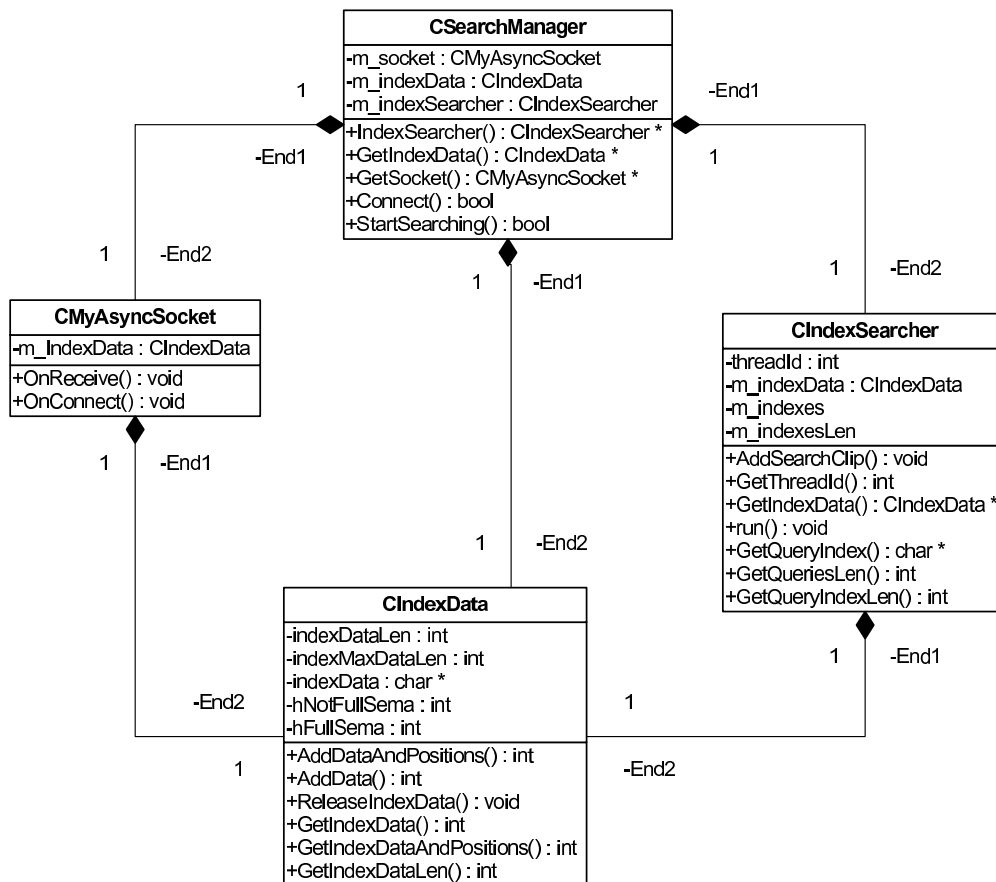
- `CSearchManager`: Είναι ο «διαχειριστής» των άλλων τάξεων. Ουσιαστικά, δημιουργεί ένα αντικείμενο της `CIndexData` που χρησιμοποιείται ως buffer και το δίνει σε αντικείμενα της `CMyAsyncSocket` και της `CIndexSearcher` που επίσης δημιουργεί έτσι ώστε να μοιράζονται αυτοί τον ίδιο buffer. Το αντικείμενο της `CMyAsyncSocket` θα συνδεθεί κατά τη δημιουργία του, ενώ το αντικείμενο της `CIndexSearcher` θα αρχίσει να εκτελείται στο δικό του thread.
- `CMyAsyncSocket`: Είναι ένα ασύγχρονο socket. Αυτό σημαίνει ότι η εκτέλεση του προγράμματος δε «μπλοκάρει» όταν το socket αναμένει για δεδομένα ή συνδέεται, αλλά συνεχίζεται κανονικά. Όταν γίνει η σύνδεση ή καταφθάσουν δεδομένα τότε θα κληθούν οι μέθοδοι `OnConnect` και `OnReceive` για το χειρισμό των συμβάντων. Το socket τοποθετεί τα δεδομένα στο αντικείμενο της `CIndexData` που λειτουργεί ως buffer, όταν αυτός φυσικά αδειάσει.
- `CIndexSearcher` Είναι η τάξη που κάνει την αναζήτηση. Εκτελείται σε ξεχωριστό thread και «μπλοκάρει» μέχρι να γεμίσει ο buffer. Όταν αυτό γίνει, τότε το thread ξυπνά και αναζητά κάθε ένα από τα απόσπασμα που βρίσκονται στον πίνακα `m_indexes` (και τα μήκη τους βρίσκονται στον πίνακα `m_indexesLen`) στον buffer `m_indexData`. Η μέθοδος `AddSearchClip` προσθέτει ένα απόσπασμα προς αναζήτηση, η `GetQueryIndex` επιστρέφει την υπογραφή του i αποσπασματος, η `GetQueryIndexLen` επιστρέφει το μήκος του i αποσπασματος, ενώ η `GetQueriesLen` επιστρέφει τον αριθμό των αποσπασμάτων που αναζητούνται.

Το παραπάνω διάγραμμα τάξεων αφορά μόνο το τμήμα της συνεχούς αναζήτησης. Όπως ήδη αναφέρθηκε η αναζήτηση από την τάξη `CIndexSearcher` γίνεται με παρόμοιο τρόπο με αυτόν της ενότητας 4.2. Δηλαδή, κάθε ένα από τα αποσπάσματα που έχουν επιλεγεί αναζητείται στον τρέχοντα κάθε φορά buffer αναζήτησης και αν βρεθεί το απόσπασμα σε εκείνο το σημείο τότε τα αποτελέσματα απλώς προστίθενται σε μια βάση δεδομένων η οποία περιγράφεται στην επόμενη ενότητα.

4.4.2.4 Η βάση δεδομένων

Το σχεσιακό μοντέλο της βάσης δεδομένων του συστήματος φαίνεται στο σχήμα 4.11. Υπάρχει ο πίνακας `ΑΠΟΣΠΑΣΜΑ` στον οποίο αποθηκεύονται

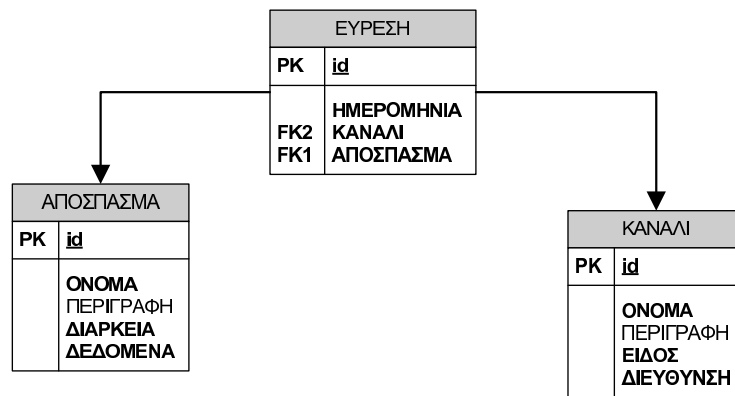
4.4. ΣΥΝΕΧΗΣ ΑΝΑΖΗΤΗΣΗ



Σχήμα 4.10: Διάγραμμα τάξεων για τη συνεχή αναζήτηση

ΚΕΦΑΛΑΙΟ 4. ΣΥΣΤΗΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ

πληροφορίες για τα αποσπάσματα προς αναζήτηση όπως το όνομα, μια μικρή περιγραφή, η διάρκεια αλλά και τα δεδομένα αυτών, δηλαδή η ακολουθία των ποσοστών intra macroblock κάθε εικόνας και τυχόν κατάλογοι αυτών όπως αναλύθηκαν στην ενότητα 3.8. Ο πίνακας ΚΑΝΑΛΙ έχει πληροφορίες για κάθε κανάλι προς παρακολούθηση, δηλαδή το όνομα, μια μικρή περιγραφή για αυτό, το είδος του (δηλαδή αν είναι δικτυακό ή αναλογικής τηλεόρασης κλπ) και τη διεύθυνση παρακολούθησης, η οποία θα είναι διαφορετικού τύπου ανάλογα με το είδος του καναλιού. Τέλος, στον πίνακα ΕΥΡΕΣΗ αποθηκεύονται οι εγγραφές για την ώρα που τυχόν βρέθηκε κάποιο απόσπασμα στο κανάλι.



Σχήμα 4.11: Σχηματικό μοντέλο της βάσης δεδομένων για τη συνεχή αναζήτηση

4.4.3 Υλοποίηση

Για το τμήμα της εξαγωγής των υπογραφών χρησιμοποιήθηκαν και εδώ από κοινού τα εργαλεία Mplayer [MPL] και FFmpeg [FFM]. Το Mplayer χρησιμοποιήθηκε απλά σαν ένας καθολικός μετατροπέας βίντεο σε H.264 που είναι απαραίτητο από τη μέθοδο αναζήτησης των intra macroblock ενώ, με τη βοήθεια της βιβλιοθήκης FFmpeg γράφτηκε ένα πρόγραμμα που διαβάζει ένα H.264 αρχείο και για κάθε εικόνα του αποστέλλει το ποσοστό των intra macroblocks του σε ένα καθορισμένο δικτυακό socket.

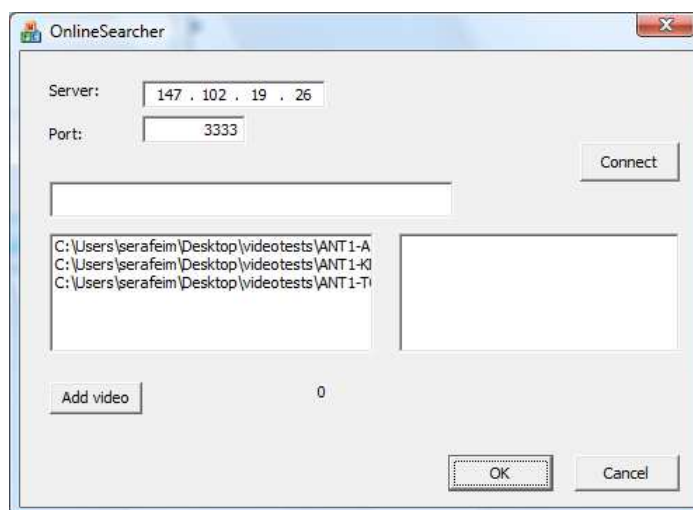
Το Mplayer επιλέχτηκε έτσι ώστε να μπορεί να χρησιμοποιηθεί κάθε είδους πηγή ως είσοδος στη συνεχή αναζήτηση, αφού υποστηρίζει τόσο ανάγνωση αρχείων από το δίσκο (και μάλιστα σχεδόν όλους τους γνωστούς τύπους αρχείων) όσο και ανάγνωση από μια σειρά από διαφορετικά δικτυακά πρωτόκολλα εκπομπής βίντεο στο διαδίκτυο. Τέλος, είναι δυνατή και η απευθείας ανάγνωση από διάφορες κάρτες τηλεόρασης (αναλογικής, ψηφιακής και δορυφορικής), αρκεί αυτές να υποστηρίζονται από το λειτουργικό σύστημα Linux. Έτσι, το Mplayer είναι ένα καθολικό εργαλείο που μπορεί να διαβάζει βίντεο από κάθε είδους πηγή και μάλιστα σχεδόν όλων των διαφορετικών τύπων που υπάρχουν και να το μετατρέπει σε H.264 κατάλληλης μορφής για την εξαγωγή των υπογραφών.

4.4. ΣΥΝΕΧΗΣ ΑΝΑΖΗΤΗΣΗ

Η εξαγωγή, γίνεται με ένα πρόγραμμα που χρησιμοποιεί τη βιβλιοθήκη FFmpeg, παρομοίως με την ενότητα 2.4.2 και την υλοποίηση του συστήματος αναζήτησης intra macroblock της ενότητας 4.2, απλώς σε αυτή την περίπτωση το ποσοστό intra macroblock κάθε εικόνας αποστέλλεται στους «ακροατές» ενός καθορισμένου socket.

Η σύνδεση μεταξύ του Mplayer και FFmpeg έγινε με τη βοήθεια unix named pipes. Δηλαδή, η έξοδος του Mplayer που ήταν ένα H.264 αρχείο γραφόταν σε ένα συγκεκριμένο named pipe, το οποίο ήταν και η είσοδος του προγράμματος του FFmpeg από το οποίο και δημιουργούσε την υπογραφή. Αυτό ήταν απαραίτητο αφού έπρεπε το σύστημα να μπορεί να λειτουργεί συνεχώς, κάτι που θα δημιουργούσε πρόβλημα αν χρησιμοποιούνταν απλά αρχεία στο δίσκο (αφού το μέγεθος τους θα γινόταν σύντομα πολύ μεγάλο).

Αναφορικά με το τμήμα της αναζήτησης, όπως ήδη αναφέρθηκε ήταν ένα πρόγραμμα σε Windows και MFC. Το γραφικό του περιβάλλον φαίνεται στο σχήμα 4.12. Ο χρήστης μπορεί να καθορίσει την IP διεύθυνση και TCP θύρα που θα συνδεθεί για να διαβάσει την υπογραφή του συνεχούς βίντεο. Υπάρχει μια λίστα αριστερά με τα αποσπάσματα τα οποία αναζητούνται και μπορούν να προστεθούν και άλλα με το κουμπί Add video. Στη δεξιά λίστα θα γράφονται οι χρόνοι στους οποίους βρέθηκε κάθε βίντεο παράλληλα με την εγγραφή στη βάση δεδομένων. Με το κουμπί Connect γίνεται η σύνδεση και αρχίζει η αναζήτηση σύμφωνα με όσα αναφέρθηκαν στην προηγούμενη ενότητα.



Σχήμα 4.12: Το γραφικό περιβάλλον του προγράμματος της συνεχούς αναζήτησης

Τέλος, πρέπει εδώ να σημειωθεί ότι το πλήρες σύστημα της συνεχούς αναζήτησης μπορούσε εύκολα να εκτελεστεί σε έναν απλό προσωπικό υπολογιστή και μάλιστα να παρακολουθούνται και δύο προγράμματα ταυτόχρονα.

ΚΕΦΑΛΑΙΟ 4. ΣΥΣΤΗΜΑΤΑ ΑΝΑΖΗΤΗΣΗΣ

Κεφάλαιο 5

Αποτελέσματα

Στο κεφάλαιο αυτό θα παρουσιαστούν μια σειρά από πειράματα που πραγματοποιήθηκαν με σκοπό τη μέτρηση της επίδοσης των τεχνικών που παρουσιάστηκαν στα προηγούμενα. Έγινε προσπάθεια να γίνουν αναλυτικές μετρήσεις για όλες τις προτάσεις της εργασίας έτσι ώστε να είναι εύκολη η επιβεβαίωση των συμπερασμάτων ακόμα και εποπτικά. Όλα τα πειράματα που ακολουθούν πραγματοποιήθηκαν σε ένα σύστημά με επεξεργαστή Intel Core 2 στα 2.40 GHz, με 2 GB RAM που έτρεχε Windows Vista.

Το κεφάλαιο χωρίζεται σε τρεις ενότητες: Στην ταχύτητα δεικτοδότησης, την ταχύτητα αναζήτησης και την επίδοση αναζήτησης.

5.1 Ταχύτητα δεικτοδότησης

Η ταχύτητα δεικτοδότησης έχει να κάνει με το πόσο γρήγορα μπορεί να εξαχθεί η υπογραφή ενός βίντεο, είτε αυτό το βίντεο λειτουργεί ως πλήρες βίντεο στο οποίο θα γίνει αναζήτηση είτε λειτουργεί ως απόσπασμα με το οποίο θα γίνει αναζήτηση. Στο κεφάλαιο 3 παρουσιάστηκαν ουσιαστικά τρεις τρόποι αναζήτησης: Δύο στους οποίους η υπογραφή είναι οι εικόνες που αναγνωρίζονται ως αλλαγές πλάνου και ένας στον οποίο η υπογραφή είναι το ποσοστό intra macroblock κάθε εικόνας σε ένα H.264 βίντεο καθορισμένης μορφής. Επίσης, στο κεφάλαιο 2, παρουσιάστηκαν δύο τεχνικές για την εξαγωγή των αλλαγών πλάνου: Μία που βασίζεται σε H.264 καθορισμένης μορφής παρόμοια με τη μέθοδο αναζήτησης, αφού η αποφασί για το αν μια εικόνα είναι ή όχι αλλαγή πλάνου εξαρτάται από το αν το ποσοστό intra macroblock της εικόνας ξεπερνά κάποιο καθορισμένο κατώφλι, και μια που βασίζεται σε ιστόγραμμα φωτεινότητας.

Έτσι τελικά η δεικτοδότηση μπορεί να χωριστεί σε δύο τμήματα: Στην εξαγωγή του ποσοστού intra macroblock από ένα H.264 και στην εξαγωγή του ιστογράμματος φωτεινότητας. Για την εξαγωγή του ποσοστού intra macroblock από ένα βίντεο αυθαίρετης μορφής, θα πρέπει αυτό το βίντεο να επανακωδικοποιηθεί σε H.264 κατάλληλης μορφής, ενώ για την εξαγωγή του

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

ιστογράμματος φωτεινότητας, θα πρέπει το βίντεο απλώς να αποσυμπίεστεί (ή να χρησιμοποιηθεί το DC ιστόγραμμα, όπως αναφέρθηκε στο κεφάλαιο 2).

Στην εργασία αυτή δε θα παρουσιαστούν μετρήσεις για τη δεικτοδότηση ενός αυθαίρετου αρχείου, αφού όπως αναφέρθηκε χρειάζεται αποκωδικοποίηση και επανακωδικοποίηση σε H264 και εργασίες που μετρούν την επίδοση αυτών των διαδικασιών υπάρχουν ήδη πολλές στη βιβλιογραφία [PSR93, CCPP02, ΗJKH03, OBL⁺04]. Θα παρουσιαστούν μετρήσεις όμως για την ταχύτητα εξαγωγής της υπογραφής (δηλαδή του ποσοστού intra macroblock) από ένα κατάλληλο H.264 και θα γίνει σύγκριση με το χρόνο που θα χρειαζόταν για την πλήρη αποσυμπίεση του βίντεο αυτού.

Ουσιαστικά χρησιμοποιήθηκαν δύο βίντεο, ένα με ανάλυση 704x384 το οποίο συμπίεστηκε με ποιότητα $Q = 26$, $Q = 30$ και $Q = 50$, το ίδιο βίντεο με τη μισή ανάλυση και $Q = 20$, δηλαδή 352x192 και τέλος ένα βίντεο πολύ μεγάλης ανάλυσης, ίσης με 1280x720 που συμπίεστηκε με $Q = 20$. Στα βίντεο αυτά έγινε αρχικά πλήρης αποσυμπίεση¹ και μετρήθηκε ο χρόνος που χρειαζόταν για την αποσυμπίεση. Ο αριθμός των εικόνων κάθε βίντεο διαιρέθηκε με αυτό το χρόνο και έτσι βρέθηκε το πόσες εικόνες αποσυμπίεζονται ανά δευτερόλεπτο. Στη συνέχεια, για τα ίδια βίντεο έγινε εξαγωγή της υπογραφής των intra macroblock χωρίς να γίνει πλήρης αποσυμπίεση και ομοίως εξάχθηκε ο αριθμός των εικόνων ανά δευτερόλεπτο. Τα αποτελέσματα φαίνονται στον πίνακα 5.1.

	704x384	704x384	704x384	352x192	1280x720
σε fps	Q26	Q30	Q50	Q26	Q20
Γρήγορο	1456	1679	2026	4905	218
Αργό	251	284	609	932	48

Πίνακας 5.1: Ταχύτητα εξαγωγής υπογραφής

Από αυτό τον πίνακα φαίνεται καθαρά η υπεροχή της μεθόδου εξαγωγής υπογραφής σε σχέση με την πλήρη αποσυμπίεση, αφού σχεδόν σε όλες τις περιπτώσεις η πλήρης αποσυμπίεση χρειάζεται τουλάχιστον πενταπλάσιο χρόνο. Στην περίπτωση του Q50 χρειάζεται περίπου 3.5 φορές περισσότερο χρόνο. Αυτό συμβαίνει επειδή η πλήρης αποσυμπίεση βελτιώνει την επίδοση της αρκετά όσο πέφτει η ποιότητα λόγω της λιγότερης πληροφορίας που υπάρχει στο συμπίεμένο βίντεο, και ειδικά στο Q50 η πληροφορία είναι ελάχιστη και η αποσυμπίεση πολύ γρήγορη. Αντίθετα, για τη μέθοδο εξαγωγής υπογραφής intra macroblock η βελτίωση όσο χειροτερεύει η ποιότητα, παρότι υπάρχει, είναι πολύ μικρή. Ακόμα και σε αυτή την περίπτωση όμως η εξαγωγή των intra macroblock είναι πολύ πιο γρήγορη από την πλήρη αποσυμπίεση.

Ίσως δημιουργηθεί η απορία του γιατί να υπάρχει βελτίωση στην ταχύτητα εξαγωγής όσο χειροτερεύει η ποιότητα του H.264, αφού άσχετα με την

¹Οπότε θα μπορούσε να δημιουργηθεί και το ιστόγραμμα φωτεινότητας, δηλαδή ο χρόνος που αναφέρεται είναι περίπου ο χρόνος για τη δημιουργία ιστογράμματος φωτεινότητας

5.2. ΤΑΧΥΤΗΤΑ ΑΝΑΖΗΤΗΣΗΣ

ποιότητα, για την εξαγωγή υπογραφής απλώς ελέγχεται ο τύπος κάθε macro-block της εικόνας. Η εξήγηση οφείλεται στο πολύ μικρότερο μεγέθους αρχείου καθώς στον τρόπο κωδικοποίησης του H.264. Έτσι, λόγω του μικρότερου αρχείου χρειάζονται λιγότερες αναγνώσεις από το δίσκο για την εξαγωγή της υπογραφής. Επίσης, το H.264 όσο χειροτερεύει η ποιότητα εισάγει όλο και περισσότερα skip macroblock, δηλαδή macroblock που έχουν μείνει ίδια μεταξύ των εικόνων. Τα skip αυτά macroblock όταν είναι συνεχόμενα μπορούν να ομαδοποιηθούν σε μεγάλες ομάδες με πολύ μικρή πληροφορία κωδικοποίησης, έτσι η υπογραφή για αυτά τα macroblock εξάγεται άμεσα.

Φαίνεται ότι ακόμα και στο βίντεο μεγάλης ανάλυσης η ταχύτητα εξαγωγής υπογραφής είναι πολύ μεγάλη (πάνω από 200 εικόνες / δευτερόλεπτο), ενώ για την πλήρη αποσυμπίεση η ταχύτητα είναι δύο φορές παραπάνω από τον πραγματικό χρόνο. Από τα παραπάνω, μπορεί να γίνει κατανοητό ότι για να εξαχθεί η υπογραφή σε ένα βίντεο διάρκειας δύο ωρών με κανονική ανάλυση, θα χρειαστούν περίπου 2 λεπτά, ενώ για την πλήρη αποσυμπίεση αυτού θα χρειαστούν πάνω από 10.

5.2 Ταχύτητα αναζήτησης

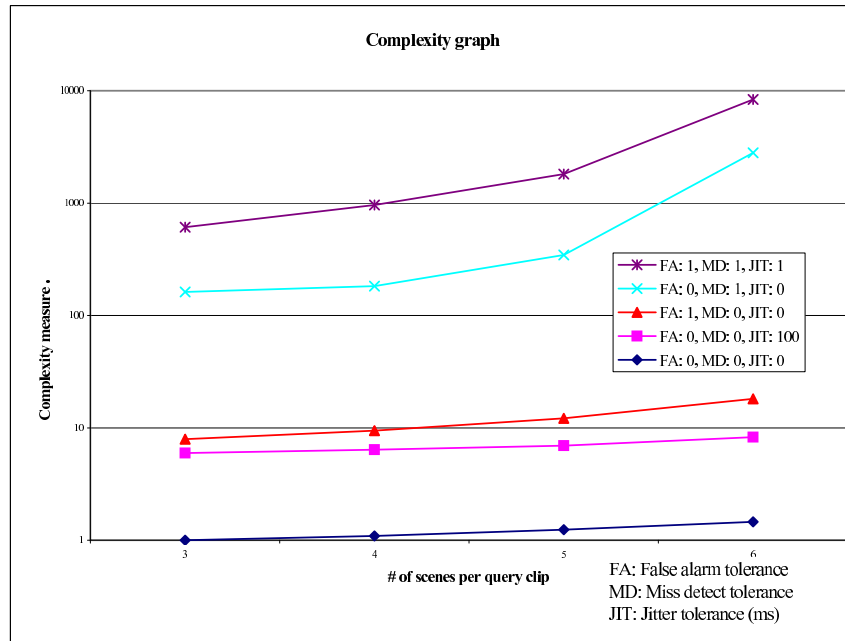
5.2.1 Μέθοδος βάσης δεδομένων

Επειδή για τη μέθοδο της βάσης δεδομένων πρέπει υποχρεωτικά να χρησιμοποιηθεί μια σχεσιακή βάση δεδομένων, αφού η μέθοδος βασίζεται στη δημιουργία κατάλληλων ερωτημάτων SQL προς τη βάση, τα αποτελέσματα θα παρουσιαστούν ξεχωριστά σε αυτή την ενότητα και δε θα γίνει σύγκριση με τα αποτελέσματα των μεθόδων intra macroblock και πιθανοτήτων.

Στην ενότητα 3.7 αναλύθηκε το πώς δουλεύει η μέθοδος: Δημιουργείται μια ερώτηση SQL με βάση τις διάρκειες του χρόνου μεταξύ των αλλαγών πλάνου. Ανάλογα πόση ανοχή χρειάζεται σε jitter, η διάρκειες αυτές έχουν μια ανοχή γύρω από τις τιμές του αποσπάσματος. Τέλος, ανάλογα με τον αριθμό των πιθανών εσφαλμένων και χαμένων αλλαγών πλάνου θα δημιουργηθούν μια σειρά από επιπλέον ερωτήματα προς τη βάση δεδομένων, στα οποία κάθε διάρκεια του αποσπάσματος θα θεωρείται ότι μπορεί να αποτελούνται από ένα σύνολο διαρκειών της βάσης και το αντίθετο.

Στην εικόνα 5.1 φαίνεται η πολυπλοκότητα κάθε περίπτωσης αναζήτησης στη βάση δεδομένων. Πιο συγκεκριμένα, αρχικά μετρήθηκε ο χρόνος που χρειαζόταν για να δωθεί απάντηση σε ένα ερώτημα που βασίστηκε σε απόσπασμα που περιείχε μόνο τρεις διάρκειες, χωρίς ανοχή σε εσφαλμένες ή χαμένες σκηνές. Στη συνέχεια, μετρήθηκαν οι χρόνοι για 4, 5 και 6 διάρκειες, για ανοχή σε jitter ίση με 100ms, για ανοχή σε μία εσφαλμένη σκηνή, για ανοχή σε μία χαμένη σκηνή και για ανοχή και στα δύο. Όλοι αυτοί οι χρόνοι τελικά διαιρέθηκαν με το χρόνο που χρειάζεται η απλούστερη περίπτωση και παρουσιάζονται στο λογαριθμικό γράφημα.

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ



Σχήμα 5.1: Πολυπλοκότητα ερωτημάτων προς βάση δεδομένων

Με τη μπλέ γραμμή, κάτω κάτω, φαίνεται η απλούστερη περίπτωση: Η πολυπλοκότητα του αποσπάσματος με 3 διάρκειες είναι ίση με 1, ενώ για 4, 5 και 6 διάρκειες αυξάνει ελαφρώς. Όταν προστεθεί το jitter, η πολυπλοκότητα αυξάνει αρκετά, δηλαδή ο χρόνος που θα χρειαστεί θα είναι περίπου 7-8 φορές παραπάνω. Αυτός ο χρόνος θα μπορούσε να μειωθεί αν η βάση είχε δεικτοδοτηθεί με ιδιαίτερο τρόπο που να βοηθά τις συγκρίσεις τιμών και όχι μόνο τις ισότητες. Με την κόκκινη γραμμή φαίνεται η περίπτωση που υπάρχει ανοχή σε 1 εσφαλμένη αλλαγή πλάνου (False Alarm tolerance = 1), και εδώ η πολυπλοκότητα είναι περίπου 10 φορές μεγαλύτερη από την απλή περίπτωση. Στην περίπτωση που υπάρχει ανοχή σε 1 χαμένη αλλαγή πλάνου (Missed Detect tolerance = 1) η πολυπλοκότητα αυξάνει πάρα πολύ αφού ξεκινά από 200 φορές παραπάνω στις 3 διάρκειες και φτάνει μέχρι και πάνω από 3000 στις 6 διάρκειες (γαλάζια γραμμή). Έχει ήδη αναφερθεί στην ενότητα 5.2.1 ότι λόγω του τρόπου που λειτουργεί η μέθοδος, θα πρέπει να αποφευχθεί όσο γίνεται η ανοχή σε χαμένες αλλαγές πλάνου και ο έλεγχος με παραπάνω από 4-5 διάρκειες μεταξύ των αλλαγών πλάνου. Τέλος, η μωβ γραμμή αναπαριστά την περίπτωση που υπάρχει ανοχή και σε εσφαλμένη και σε χαμένη αναγνώριση και φαίνεται καθαρά ότι η περίπτωση αυτή είναι προβληματική και ότι θα αργήσει πάρα πολύ να δώσει αποτελέσματα.

Στον πίνακα 5.2 φαίνονται ως παράδειγμα και οι χρόνοι που χρειάστηκαν για την πραγματοποίηση κάποιων ερωτημάτων για 3 και 5 διάρκειες και για όλες τις δυνατές περιπτώσεις ανοχής σε χαμένες και εσφαλμένες αναγνώρισεις αλλαγών πλάνου. Η διάρκεια των αποσπασμάτων δεν ενδιαφέρει, αφού τελικά χρησιμο-

5.2. ΤΑΧΥΤΗΤΑ ΑΝΑΖΗΤΗΣΗΣ

ποιούνται για τη δημιουργία των ερωτημάτων μόνον αυτές οι 3 ή 5 διάρκειες. Η βάση δεδομένων στην οποία έγινε η αναζήτηση περιέχει περίπου 7 ώρες βίντεο διαφόρων ειδών για παράδειγμα ντοκιμαντέρ, αθλητικά, κινηματογραφικές ταινίες κοκ.

Μπορεί να φανεί και εδώ η μικρή αύξηση από τις 3 στις 5 διάρκειες² και όταν υπάρχει false alarm tolerance αλλά και η μεγάλη αύξηση με την εισαγωγή του missed detection tolerance.

Διάρκειες	False Alarm tolerance	Missed Detect tolerance	Χρόνος αναζήτησης / ms
3	0	0	31
3	0	1	7718
3	1	0	78
3	1	1	21781
5	0	0	31
5	0	1	7688
5	1	0	158
5	1	1	36552

Πίνακας 5.2: Ταχύτητα αναζήτησης για τη μέθοδο της βάσης δεδομένων

Πρέπει να σημειωθεί ότι οι τιμές αυτές είναι τελείως ενδεικτικές αφού εξαρτώνται πολύ από τον υπολογιστή στον οποίο έτρεξαν τα ερωτήματα, το ποιά βάση δεδομένων χρησιμοποιήθηκε και φυσικά το πώς είχε βελτιστοποιηθεί (optimize) η απόδοση αυτής της βάσης δεδομένων για να τρέξει τα συγκεκριμένα ερωτήματα με την προσθήκη δεικτών κλπ, κάτι που δεν έγινε στα πλαίσια της συγκεκριμένης ερευνητικής εργασίας αφού αφορά άλλο ερευνητικό πεδίο.

5.2.2 Πιθανοτική και intra macroblock μέθοδοι

Για τις μεθόδους των πιθανοτήτων και intra macroblock οι μετρήσεις σχετικά με την ταχύτητα αναζήτησης έγιναν με τη βοήθεια του προγράμματος Text-VideoSearch στο οποίο ενσωματώθηκαν οι εξής μέθοδοι:

- prob1: Είναι η αναζήτηση με τη μέθοδο των πιθανοτήτων στην οποία το απόσπασμα ολισθαίνει σε όλες τις δυνατές θέσεις κάθε βίντεο της βάσης δεδομένων και μετρώνται οι διαφορές. Οι θέσεις που έχουν διαφορές κάτω από ένα κατώφλι επιλέγονται ως αποτελέσματα (ενότητα 3.6).
- prob2: Είναι η αναζήτηση με τη μέθοδο των πιθανοτήτων στην οποία όμως ο έλεγχος γίνεται μόνο σε συγκεκριμένες θέσεις των βίντεο της

²Όχι πάντα, αφού ανάλογα με το πώς είναι κατανομημένες οι διάρκειες σε ένα απόσπασμα μπορεί το ερώτημα που θα δημιουργηθεί να εκτελεστεί γρηγορότερα, λόγω του ότι κάποια επιμέρους select θα επιστρέψουν κενό σύνολο και δε θα χρειαστεί να γίνουν επιπλέον join

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

βάσης δεδομένων και στο απόσπασμα, δηλαδή εκεί όπου υπάρχουν συμβάντα και στα δύο (περίπου όπως περιγράφει ο αλγόριθμος της 3.6.4.1).

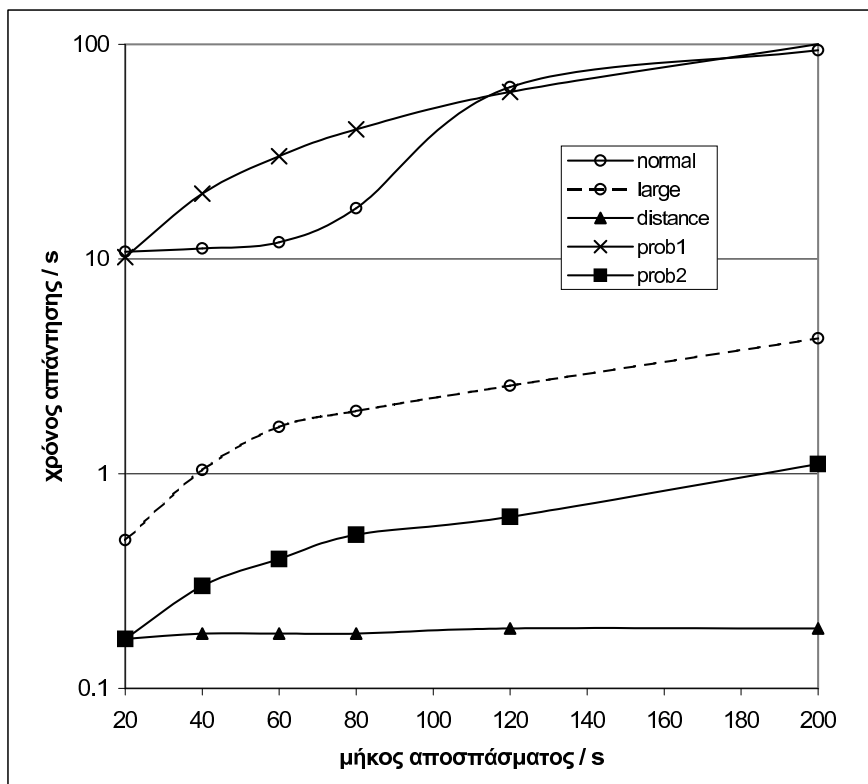
- **normal**: Είναι η απλή μέθοδος αναζήτησης στην οποία το απόσπασμα ολισθαίνει σε όλες τις θέσεις και μετράται σε κάθε σημείο η *intra-mb*: διαφορά. Υπάρχει ένα κατώφλι το οποίο διακόπτει την αναζήτηση αν η διαφορά γίνει πολύ μεγάλη, δηλαδή μπορεί να αποφασιστεί πριν τον πλήρη έλεγχο ότι η συγκεκριμένη θέση δεν είναι η σωστή (ενότητα 3.8.3.2).
- **large**: Είναι η αναζήτηση μεγάλων τιμών, στην οποία ο έλεγχος γίνεται μόνο γύρω από τις τιμές που υπερβαίνουν ένα κατώφλι, τόσο στο απόσπασμα όσο και στα βίντεο της βάσης δεδομένων (ενότητα 3.8.3.3).
- **distance**: Είναι η αναζήτηση αποστάσεων, στην οποία εξετάζονται οι αποστάσεις μεταξύ δύο εικόνων που έχουν τιμή πάνω από κάποιο κατώφλι (ενότητα 3.8.3.4).

Σε όλες τις μεθόδους που χρησιμοποιούν τελικά την απόσταστα *intra mb* οι μετρήσεις έγιναν με ανοχή σε *jitter* μιας εικόνας σύμφωνα με τη διαδικασία που περιγράφεται στην ενότητα 3.8.3.5. Δηλαδή ο υπολογισμός της απόστασης *intra-mb* ήταν περισσότερο περίπλοκος από την απλή περίπτωση που δε θα είχε ανοχή σε *jitter* αφού στην περίπτωση αυτή πρέπει να υπολογιστούν 6 διαφορές και τετράγωνα αντί για μόνο 1 στην απλή περίπτωση.

Για τα πειράματα, εξάχθηκαν αποσπάσματα μήκους 500, 1000, 1500, 2000, 3000 και 5000 εικόνων ή 20, 40, 60, 80, 120 και 200 δευτερολέπτων αν θεωρηθεί ότι τα βίντεο ήταν με δειγματοληψία 25 εικόνων ανα δευτερόλεπτο, με τα οποία πραγματοποιήθηκε αναζήτηση σε δύο διαφορετικές βάσεις δεδομένων, μία αποτελούμενη από περίπου 15,6 εκατομμύρια εικόνες ή 173 περίπου ώρες βίντεο ή 58 ξεχωριστά βίντεο και μία αποτελούμενη από 8,2 εκατομμύρια εικόνες η 91 περίπου ώρες βίντεο ή 31 ξεχωριστά βίντεο. Τα βίντεο της βάσης δεδομένων, εκτός από κάποια συγκεκριμένα βίντεο που θα χρησιμοποιηθούν στην επόμενη ενότητα για τις δοκιμές της επίδοσης στην ακρίβεια και ανάκληση, αποτελούνταν από συνεχή καταγραφή τηλεοπτικού προγράμματος από ψηφιακό δέκτη. Τα αποσπάσματα για κάθε μήκος εξάχθηκαν από τρία διαφορετικά βίντεο και οι χρόνοι είναι ο μέσος όρος έτσι ώστε να καλυφθούν τυχόν ειδικές περιπτώσεις σε αποσπάσματα

Τα αποτελέσματα για όλες τις μεθόδους φαίνονται στα σχήματα 5.2 για τη βάση δεδομένων των 173 ωρών βίντεο και 5.3 για τη βάση δεδομένων των 91 ωρών βίντεο. Από τις δύο αυτές εικόνες μπορούν να εξαχθούν μια σειρά από συμπεράσματα αναφορικά με την ταχύτητα αναζήτησης των μεθόδων που παρουσιάστηκαν. Εποπτικά φαίνεται ότι σε όλες τις μεθόδους, η αναζήτηση στη βάση δεδομένων με τις 173 ώρες βίντεο χρειάστηκε λίγο παραπάνω χρόνο από το διπλάσιο του χρόνου για τις 91 ώρες βίντεο. Αυτό είναι λίγο πολύ αναμενόμενο, αφού η βάση με τις 173 ώρες βίντεο, εκτός από το περίπου διπλάσιο μήκος σε εικόνες, έχει και περίπου τον διπλάσιο αριθμό από ξεχωριστά βίντεο,

5.2. ΤΑΧΥΤΗΤΑ ΑΝΑΖΗΤΗΣΗΣ



Σχήμα 5.2: Χρόνοι αναζήτησης για όλες τις μεθόδους για διάφορα μήκη αποσπάσματος για τη μεγαλύτερη βάση δεδομένων

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

έτσι, ασχέτως των δεικτών που μπορεί να χρησιμοποιούν οι διάφορες μέθοδοι για τη γρήγορη απόρριψη κάποιων βίντεο (όπως για παράδειγμα η αναζήτηση μεγάλων τιμών) θα πρέπει να γίνει ο έλεγχος στα διπλάσια βίντεο, άρα ο χρόνος αναλογικά θα διπλασιαστεί. Κάτι τέτοιο δεν θα ίσχυε σε περίπτωση που η βάση δεδομένων αποτελούνταν από ένα μόνο βίντεο διάρκειας 173 ή 91 ωρών αφού σε μια τέτοια περίπτωση για τις μεθόδους που χρησιμοποιούν δείκτες ο χρόνος αναζήτησης δε θα άλλαζε πολύ. Η περίπτωση του χρόνου αναζήτησης όταν η βάση αποτελούνταν μόνο από ένα ενιαίο βίντεο δεν ελέγχθηκε παραπάνω μιας και βίντεο με τόσο μεγάλες διάρκειες δε συναντώνται στην πράξη.

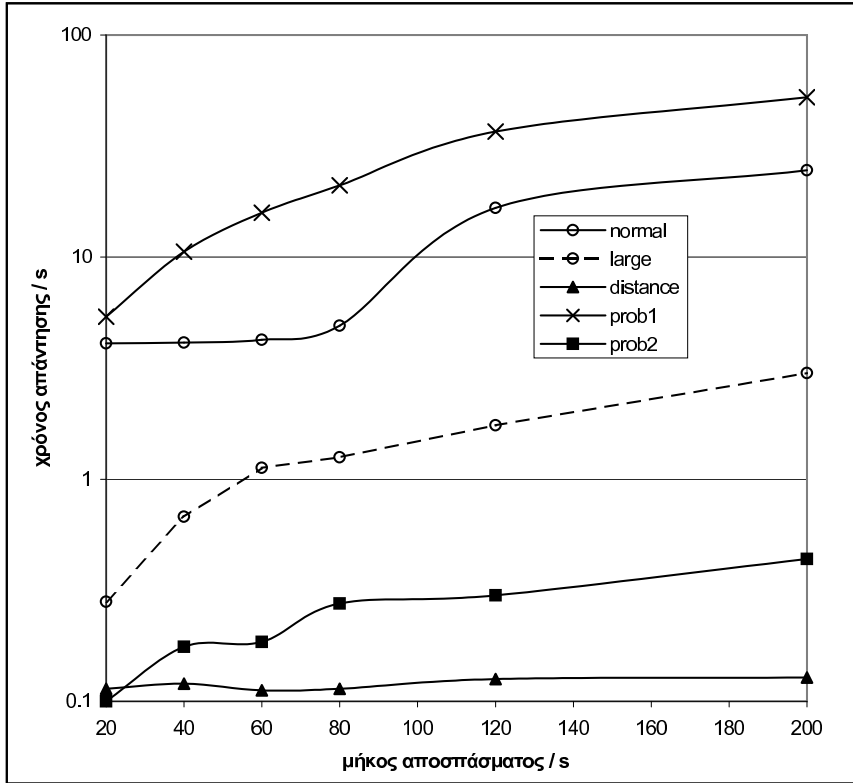
Η συμπεριφορά των μεθόδων όσο μεγαλώνει το μήκος του αποσπάσματος είναι παρόμοια και στην περίπτωση της μεγαλύτερης και της μικρότερης βάσης δεδομένων. Η καλύτερη επίδοση επιτυγχάνεται με διαφορά για τη μέθοδο αποστάσεων (γραμμή με τρίγωνα), όπου, ο χρόνος που χρειάζεται είναι λίγο πολύ σταθερός ασχέτως του μεγέθους του βίντεο, ενώ ο χρόνος αυτός είναι περίπου 110 ms στη μικρότερη και 190 ms στη μεγαλύτερη βάση δεδομένων. Ο χρόνος αυτός επιτυγχάνεται αφού, ελέγχονται μόνο τα σημεία των βίντεο της βάσης δεδομένων τα οποία έχουν αποστάσεις που συμφωνούν με αυτές των αποσπασμάτων. Η μέθοδος αυτή φυσικά έχει ένα μεγάλο μειονεκτήματα: Κάποια βίντεο μπορεί να χαθούν επειδή δε ταίριαζε μια απόσταση. Αυτό μπορεί να συμβεί επειδή μια εικόνα θα ξεπεράσει το κατώφλι στο ένα βίντεο και όχι στο άλλο. Το jitter δεν επηρεάζει, αφού αν στο απόσπασμα μια απόσταση είναι για παράδειγμα x , μπορεί να γίνει ο έλεγχος για αποστάσεις $x \pm jt$ όπου jt κάποιο επιλεγθέν μέγιστο jitter.

Οι μέθοδοι των μεγάλων τιμών (διακεκομένη γραμμή με κύκλους) και η γρήγορη πιθανοτική (γραμμή με τετράγωνα) ακολουθούν. Οι δύο αυτές μέθοδοι είναι λίγο πολύ παρόμοιες, αφού στην πρώτη επιλέγονται θέσεις στις οποίες το ποσοστό των intra macroblock ξεπερνά κάποιο κατώφλι τόσο στο απόσπασμα όσο και στο βίντεο της βάσης δεδομένων, γίνεται συγχρονισμός των δύο θέσεων στη συνέχεια βρίσκονται η intra mb απόσταση τους, ενώ στη δεύτερη επιλέγονται θέσεις με συμβάντα και μετρώνται τα συμβάντα γύρω από αυτές τις θέσεις (απλώς με *cor*). Το ότι στη γρήγορη πιθανοτική μέθοδο μετρώνται απλώς οι διαφορές στα συμβάντα και δε βρίσκεται η intra mb απόσταση είναι προφανώς και ο λόγος που έχει καλύτερη επίδοση από την μέθοδο μεγάλων τιμών.

Τη χειρότερη επίδοση όπως αναμένεται την έχουν η απλή αναζήτηση intra macroblock με έλεγχο διακοπής και η απλή πιθανοτική αναζήτηση. Στη μεν απλή πιθανοτική αναζήτηση ο χρόνος αυξάνει τελείως αναλογικά, αφού γίνεται απλή ολίσθηση του αποσπάσματος (τα γραφήματα είναι λογαριθμικά), ενώ στην αναζήτηση intra macroblock με έλεγχο διακοπής ακολουθείται μια πιο παράξενη καμπύλη η οποία οφείλεται στον έλεγχο διακοπής, αφού το κατώφλι που χρησιμοποιήθηκε έχει να κάνει με τις ήδη υπάρχουσες ελάχιστες αποστάσεις, έτσι, για μικρά αποσπάσματα μπορεί γρηγορότερα να διακοπεί ο έλεγχος για μια θέση αφού οι υπάρχουσες αποστάσεις δε θα είναι τόσο μεγάλες, αντίθετα για μεγαλύτερα αποσπάσματα οι υπάρχουσες αποστάσεις θα είναι αρκετά μεγάλες οπότε θα χρειάζονται περισσότεροι έλεγχοι μέχρι τελικά να διακοπεί η

5.3. ΑΠΟΔΟΣΗ ΑΝΑΖΗΤΗΣΗΣ

αναζήτηση.



Σχήμα 5.3: Χρόνοι αναζήτησης για όλες τις μεθόδους για διάφορα μήκη αποστάματος για τη μικρότερη βάση δεδομένων

Πρέπει εδώ να σημειωθεί ότι η αναζήτηση μπορεί ταυτόχρονα να γίνει από πολλούς υπολογιστές, κάθε ένας εκ των οποίων θα αναζητά κάποιο συγκεκριμένο τμήμα της βάσης δεδομένων. Λόγω του τρόπου που δουλεύει η αναζήτηση, ο διαχωρισμός των δεδομένων σε επιμέρους τμήματα τα οποία θα ανατεθούν σε ξεχωριστά συστήματα όσο και η συγκέντρωση των αποτελεσμάτων κάθε μιας αναζήτησης σε ένα συνολικό αποτέλεσμα είναι πολύ εύκολη και δεν απαιτεί σχεδόν καθόλου χρόνο. Έτσι, η ταχύτητα αναζήτησης είναι σχεδόν ανάλογη του αριθμού των υπολογιστών που χρησιμοποιούνται για την αναζήτηση.

5.3 Απόδοση αναζήτησης

Η απόδοση της αναζήτησης θα μετρηθεί με βάση τις μετρικές της ανάκλησης (recall) και ακρίβειας (precision) που παρουσιάστηκαν αναλυτικά στην ενότητα 3.3. Όπως έχει ήδη αναφερθεί, τα μεγέθη αυτά μετρώνται από ένα σύνολο από αποστάματα τα οποία προέρχονται και αναζητούνται σε μια βάση δεδομένων και αφορούν η μεν ανάκληση το πόσα από αυτά τα αποστάματα βρέθηκαν μέσα

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

στη βάση δεδομένων και η ακρίβεια το πόσα εσφαλμένα αποτελέσματα έδωσαν αυτά τα αποσπάσματα³.

5.3.1 Μέθοδος intra και πιθανοτήτων

Οι μέθοδοι αναζήτησης με βάση τις πιθανότητες και το ποσοστό των intra macroblock σε κάθε εικόνα υλοποιήθηκαν από κοινού στο πρόγραμμα TextVideoSearch το οποίο και χρησιμοποιήθηκε για τη διεξαγωγή αναλυτικών πειραμάτων για την απόδοση των μεθόδων αναζήτησης αυτών.

Συγκεκριμένα, χρησιμοποιήθηκε το ίδιο περιβάλλον με αυτό της ενότητας 5.2.2, δηλαδή μια βάση δεδομένων διάρκειας περίπου 180 ωρών, η οποία περιείχε κυρίως βίντεο τηλεοπτικού προγράμματος αλλά και κάποια συγκεκριμένα βίντεο από τα οποία είχαν εξαχθεί αποσπάσματα. Έγινε προσπάθεια να υπάρχουν βίντεο διαφορετικών κατηγοριών, δηλαδή:

- Μια κινηματογραφική ταινία
- Ένας ποδοσφαιρικός αγώνας
- Μια σειρά από μουσικά βίντεο κλιπ
- Ένα ντοκιμαντέρ

Τα βίντεο αυτά συμπίεστηκαν με το H.264 με αρχική ποιότητα $Q = 26$ και προστέθηκαν στη βάση αναζήτησης. Στη συνέχεια, τα αρχικά βίντεο (πριν συμπίεστούν) κόπηκαν σε δεκάδες τμήματα μήκους 500, 1000, και 1500 εικόνων (ή 20, 40 και 60 δευτερολέπτων αν υποθεθεί ότι τα βίντεο ήταν στα 25 fps). Τα τμήματα αυτά στη συνέχεια συμπίεστηκαν με $Q = 26$. Επίσης, τα τμήματα των 1000 εικόνων, επανασυμπίεστηκαν με διαφορετικά Q (30 και 40), αλλά και έχοντας τη μισή ανάλυση εικόνας για να εξεταστεί η συμπεριφορά της μεθόδου στις συνήθεις αλλαγές που μπορεί να υποστεί κάποιο βίντεο. Δεν έγινε προσθήκη λογότυπου στα βίντεο, αφού ήδη έχει φανεί στην ενότητα 3.8.2.3 ότι ακόμα και με την προσθήκη υποτίτλων (που δεν είναι ένα σταθερό λογότυπο) τα ποσοστά των intra macroblock σε κάθε εικόνα δεν αλλάζουν σχεδόν καθόλου, άρα δε θα εξεταστεί παραπάνω η συμπεριφορά⁴. Με αυτά τα αποσπάσματα εξετάστηκε η απόδοση των μεθόδων αναζήτησης τόσο για διαφορετικά μήκη αποσπασμάτων όσο και για τις συνηθέστερες αλλαγές που γίνονται στα βίντεο.

Πρέπει να σημειωθεί εδώ ότι για κάθε είδος πειράματος δημιουργήθηκαν περίπου 250 αποσπάσματα (δηλαδή κατά μέσο όρο περίπου 80 αποσπάσματα σε κάθε ένα από τα 4 βίντεο), ενώ για όλες τις περιπτώσεις (500, 1000, 1500 εικόνες, μισή ανάλυση, διαφορετικά Q) δημιουργήθηκαν περίπου 1500 διαφορετικά αποσπάσματα. Η βάση δεδομένων των 180 ωρών είναι επίσης ένα σημαντικό

³ Δηλαδή βρέθηκαν σε λάθος σημείο

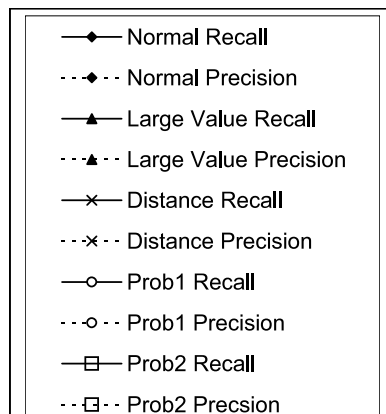
⁴ Παρότι η πιθανοτική μέθοδος δε βασίζεται στα ποσοστά των intra macroblock αλλά στις αλλαγές πλάνου, μπορεί να βρεθούν οι αλλαγές πλάνου μέσω των intra macroblock οπότε τελικά η συμπεριφορά και εδώ να μην αλλάξει.

5.3. ΑΠΟΔΟΣΗ ΑΝΑΖΗΤΗΣΗΣ

στοιχείο, αφού έτσι εξετάζεται με περισσότερη ακρίβεια το πόσα βίντεο επιστρέφονται από την αναζήτηση ως σωστά ενώ το απόσπασμα δεν προέρχεται από αυτά (false alarms).

Κάθε ένα από τα αποσπάσματα δεικτοδοτήθηκε απευθείας (αφού ήταν ήδη συμπιεσμένα σε H.264) και έγινε αναζήτηση με βάση την υπογραφή του. Σε κάθε αναζήτηση, ελεγχόταν αν βρέθηκε το σωστό βίντεο, καθώς και πόσα βίντεο επιστράφηκαν ως σωστά χωρίς να είναι και έτσι μετρήθηκαν η ανάκληση και η ακρίβεια.

Η απόφαση για το αν κάποιο αποτέλεσμα θα γίνει αποδεκτό ή όχι εξαρτάται από τον αριθμό των διαφορών στην πιθανοτική μέθοδο και από την intra macroblock απόσταση στην intra macroblock μέθοδο. Επειδή τα βέλτιστα μεγέθη των διαφορών ή της απόστασης αλλάζουν ανάλογα με τα χαρακτηριστικά των βίντεο οι μετρήσεις των recall και precision έγιναν με διάφορες τιμές των κριτηρίων απόφασης.



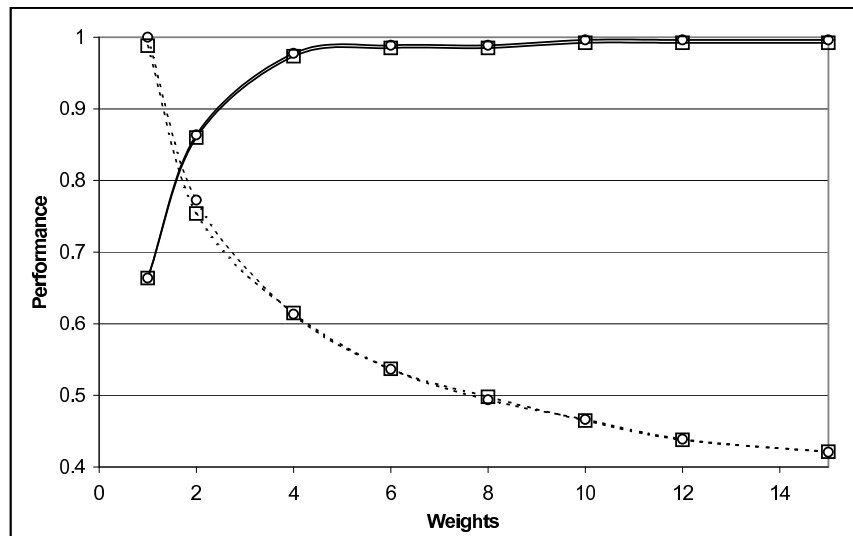
Σχήμα 5.4: Οι επεξηγήσεις των διαγραμμμάτων που ακολουθούν

Παρακάτω, θα παρουσιαστεί η συνολική επίδοση για όλα τα πειράματα. Η επίδοση για την ανάκληση και ακρίβεια των μεθόδων intra macroblock (δηλαδή της απλής μεθόδου, της μεθόδου των μεγάλων τιμών και της μεθόδου απόστασης) θα παρουσιαστούν σε ξεχωριστό διάγραμμα από την επίδοση για τις μεθόδους πιθανοτήτων (απλή και γρήγορη πιθανοτική, όπως παρουσιάστηκαν στην ενότητα 5.2.2).

Αρχικά, στο σχήμα 5.4 φαίνεται το υπόμνημα για τα σχήματα που θα ακολουθήσουν έτσι ώστε να μη χρειάζεται να επαναλαμβάνεται αυτό σε όλα τα σχήματα. Όπως φαίνεται, με συνεχή γραμμή παρουσιάζεται η ανάκληση (recall) και με διακεκομμένη η ακρίβεια (precision). Έτσι, είναι αναμενόμενο ότι όσο αυξάνει το κατώφλι που βρίσκεται στον οριζόντιο άξονα, δηλαδή τα κριτήρια αποδοχής ενός αποτελέσματος ως σωστού χαλαρώνουν, η ανάκληση θα αυξάνει αφού θα μειώνεται η πιθανότητα να χαθεί το σωστό αποτέλεσμα, αλλά η ακρίβεια θα μειώνεται αφού θα αυξάνεται η πιθανότητα να κριθεί ως σωστό κάποιο λανθασμένο αποτέλεσμα. Επίσης, με ρόμβο αναπαρίσταται η απλή ανα-

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

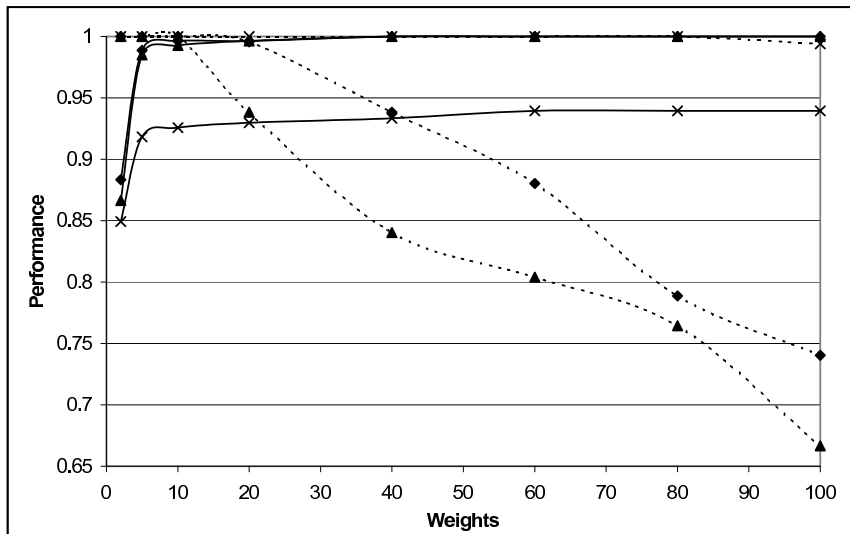
ζήτηση (normal), με τρίγωνο η αναζήτηση μεγάλων τιμών (large value), με \times η αναζήτηση αποστάσεων (distance), με κύκλο η απλή πιθανοτική αναζήτηση (prob1) και με τετράγωνο η γρήγορη πιθανοτική αναζήτηση (prob2). Μια πολύ σημαντική σημείωση για την ευκολότερη παρακολούθηση και σύγκριση των σχημάτων της επίδοσης είναι ότι κάθε ένα έχει διαφορετικό κάθετο άξονα, δηλαδή σε άλλα σχήματα ο άξονας κινείται για παράδειγμα από 0.7 ως 1 και σε άλλα από 0.4 ως 1, οπότε δε θα πρέπει να γίνει εποπτική σύγκριση των σχημάτων. Αυτό έγινε ώστε τα σχήματα να φαίνονται καλύτερα.



Σχήμα 5.5: Επίδοση πιθανοτήτων για απόσπασμα μήκους 1000 εικόνων

Στα σχήματα 5.5 και 5.6 παρουσιάζεται η επίδοση για τη μέθοδο των πιθανοτήτων και τη μέθοδο intra macroblock για αποσπάσματα μήκους 1000 εικόνων τα οποία επανασυμπίεστηκαν με τα ίδια χαρακτηριστικά με το αρχικό βίντεο. Όπως μπορεί εύκολα να φανεί από τα σχήματα, η επίδοση της ανάκλησης για τη μέθοδο των πιθανοτήτων είναι περίπου 1 όταν οι ανεκτές διαφορές μεταξύ του αποσπάσματος είναι από 4 και πάνω. Φυσικά, για τόσες διαφορές η ακρίβεια μειώνεται αρκετά αφού, όταν υπάρχει ανοχή σε 4 για παράδειγμα διαφορές, μπορεί να βρεθούν αποσπάσματα με 5 αλλαγές πλάνου - αυτά τα αποσπάσματα θα ταιριάζουν σχεδόν παντού οπότε προφανώς η ακρίβεια θα μειωθεί πολύ. Θα πρέπει λοιπόν όταν υπάρχει απόσπασμα με λίγες αλλαγές πλάνου, η ανοχή σε διαφορές να είναι μικρότερη από ότι όταν υπάρχει απόσπασμα με πολλές αλλαγές πλάνου έτσι ώστε η ανάκληση να παραμένει κοντά στο 1 αλλά και να μην πέφτει πολύ η ακρίβεια. Για ανοχή σε μια το πολύ διαφορά, η ακρίβεια πέφτει αρκετά αφού σε αποσπάσματα με αρκετές αλλαγές πλάνου είναι πολύ πιθανό μια από αυτές να διαφέρει (αντίστοιχα με τα αποτελέσματα του κεφαλαίου 2 για την αναγνώριση αλλαγών πλάνου). Στη μέθοδο των intra macroblock για μια απόσταση intra mb κάτω από 10 φαίνεται ότι και η ακρίβεια και η ανάκληση είναι πάνω από 90%. Η ανάκληση της μεθόδου των αποστάσεων είναι ελαφρώς

5.3. ΑΠΟΔΟΣΗ ΑΝΑΖΗΤΗΣΗΣ



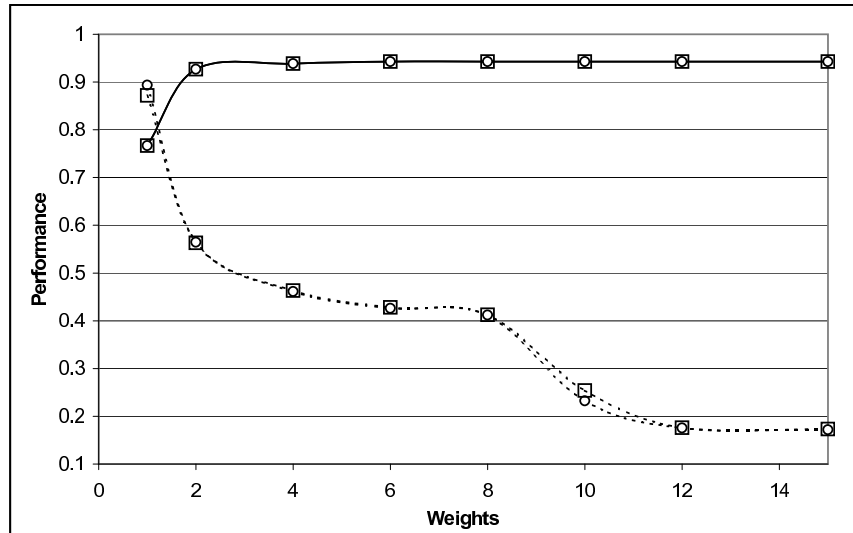
Σχήμα 5.6: Επίδοση *intra macroblock* για απόσπασμα μήκους 1000 εικόνων

μικρότερη αφού κάποιες αλλαγές πλάνου θα χαθούν, σε αντιστοιχία με τη μέθοδο των πιθανοτήτων. Επίσης, φαίνεται η ακρίβεια να είναι μεγαλύτερη για τη μέθοδο των αποστάσεων αφού πολλές εσφαλμένες απαντήσεις κόβονται λόγω των αποστάσεων των αλλαγών πλάνου, αντίθετα οι άλλες μέθοδοι είναι πιο ελαστικές και όσο αυξάνει το κατώφλι η ακρίβεια μειώνεται αρκετά αφού πλέον θα επιλέγονται όλο και περισσότερα λανθασμένα αποτελέσματα. Έτσι, όταν το απόσπασμα έχει παρόμοια χαρακτηριστικά με το βίντεο, θα πρέπει το κατώφλι να είναι αρκετά μικρό (κάτω από 10) ούτως ώστε να επιλέγονται τα σωστά βίντεο (μεγάλη ανάκληση) χωρίς να επιλέγονται πολλά εσφαλμένα βίντεο (μεγάλη ακρίβεια).

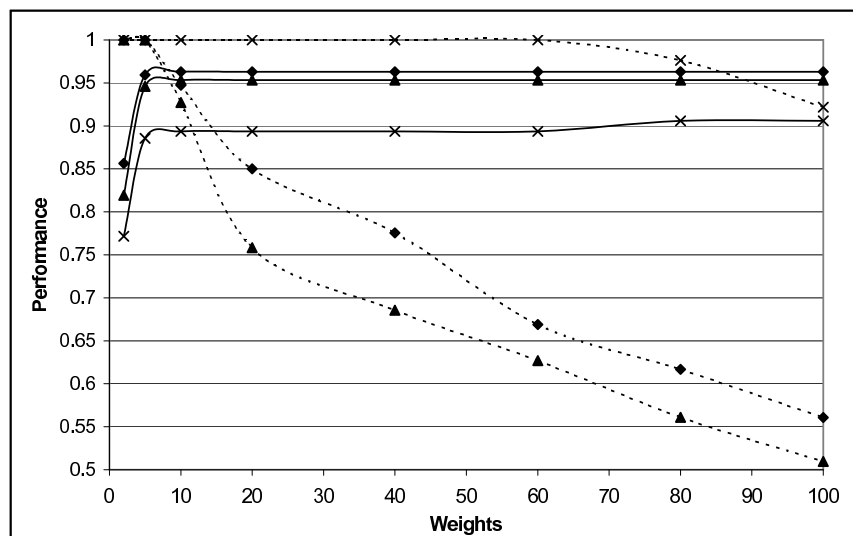
Στα σχήματα 5.7 και 5.8 φαίνεται η επίδοση αποσπάσματα ιδίων χαρακτηριστικών με το αρχικό με μήκος όμως 500 εικόνων. Για τη μέθοδο των πιθανοτήτων, η συμπεριφορά είναι λίγο χειρότερη, αφού παρότι η ανάκληση παραμένει κοντά στη μονάδα, η ακρίβεια έχει χειροτερέψει. Αυτό οφείλεται στο ότι λόγω του μικρότερου μέγεθους των βίντεο, είναι πιθανότερη η επιλογή κάποιων λανθασμένων βίντεο. Παρόμοια συμπεράσματα ισχύουν και για την *intra macroblock* μέθοδο: Η ανάκληση κινείται στα ίδια υψηλά επίπεδα, η ακρίβεια όμως είναι χειρότερη. Γενικά δηλαδή, η επίδοση για το συνδυασμό ανάκλησης και ακρίβειας για τα βίντεο με μικρότερο μήκος είναι ελαφρώς χειρότερη, κάτι που αναμένεται. Είναι ενδιαφέρον να εξηγηθεί το γιατί η επίδοση της ανάκλησης είναι ελαφρώς μικρότερη για τις μεθόδους των *intra macroblock* στα αποσπάσματα των 500 εικόνων από ότι στα αποσπάσματα των 1000 εικόνων. Αυτό οφείλεται στο ότι, λόγω του μεγαλύτερου μεγέθους του αποσπάσματος, κάποιες διαφορές που τυχόν υπάρχουν θα αμβλυνθούν περισσότερο από ότι στο μικρότερο μέγεθος και έτσι η ανάκληση θα είναι καλύτερη.

Για την περίπτωση των βίντεο ιδίων χαρακτηριστικών με μήκος όμως 1500

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

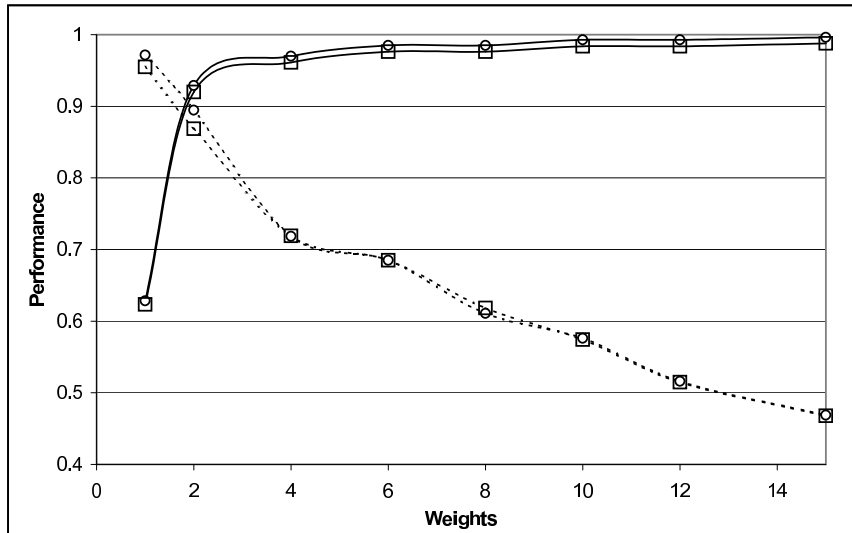


Σχήμα 5.7: Επίδοση πιθανοτήτων για απόσπασμα μήκους 500 εικόνων

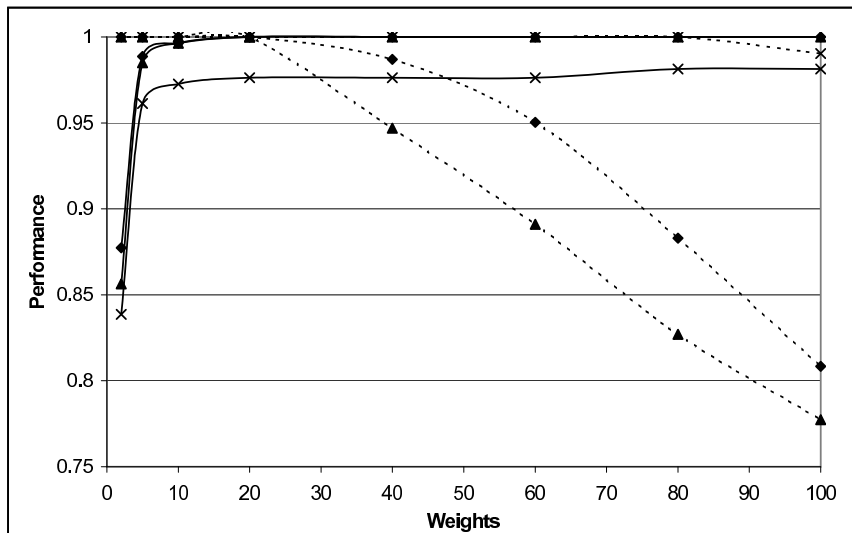


Σχήμα 5.8: Επίδοση intra macroblock για απόσπασμα μήκους 500 εικόνων

5.3. ΑΠΟΔΟΣΗ ΑΝΑΖΗΤΗΣΗΣ



Σχήμα 5.9: Επίδοση πιθανοτήτων για απόσπασμα μήκους 1500 εικόνων

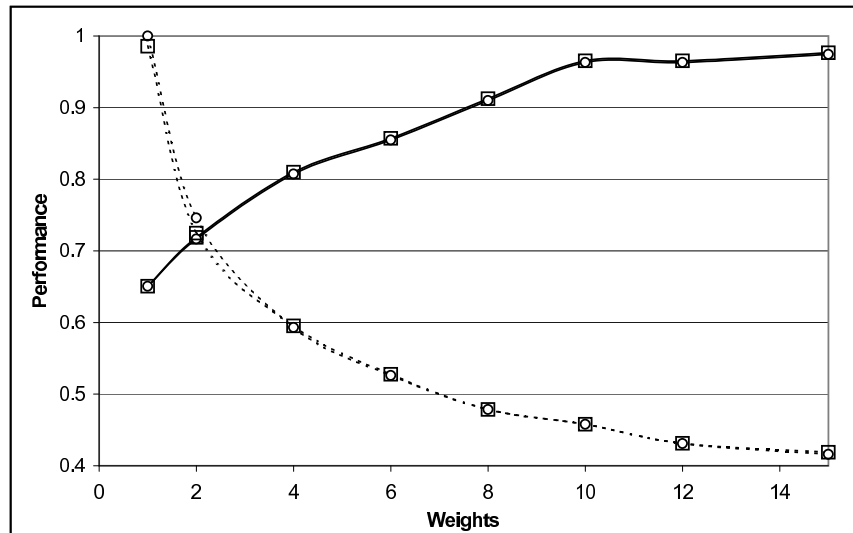


Σχήμα 5.10: Επίδοση *intra macroblock* για απόσπασμα μήκους 1500 εικόνων

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

εικόνων, που φαίνονται στα σχήματα 5.9 και 5.10 ισχύουν περίπου τα αντίθετα από ότι στην περίπτωση των 500 εικόνων. Για τη μέθοδο των πιθανοτήτων η ανάκληση παραμένει σταθερή ενώ η ακρίβεια αυξάνει αρκετά αφού μεγαλύτερα αποσπάσματα θα έχουν περισσότερες αλλαγές πλάνου. Ομοίως, για τη μέθοδο των πιθανοτήτων, η ακρίβεια αυξάνει αρκετά λόγω του μεγάλου μεγέθους, άρα στα εσφαλμένα σημεία θα αυξηθεί η απόσταση, ενώ ελαφρώς αυξάνει και η ανάκληση αφού στα σωστά σημεία, τυχόν διαφορές δε θα αυξηθούν τόσο την απόσταση όσο στα αποσπάσματα με μήκος 500 και 1000 εικόνες.

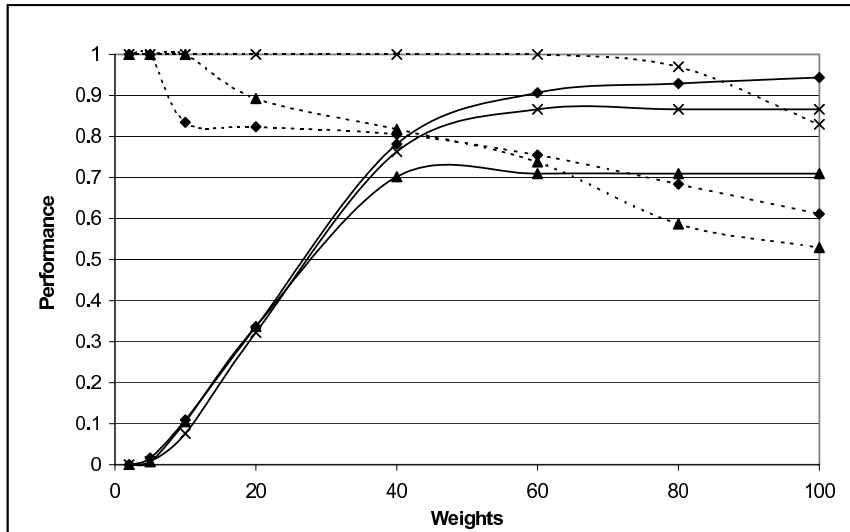
Φαίνεται λοιπόν ότι για αποσπάσματα μήκους 500, 1000 και 1500 εικόνων, δηλαδή 20, 40 και 60 δευτερολέπτων στα 25 fps τα αποτελέσματα της ανάκλησης είναι περίπου τα ίδια για όλες τις περιπτώσεις, τόσο για intra macroblock όσο και για πιθανότητες, ενώ η ακρίβεια αυξάνει όσο αυξάνει το μέγεθος του αποσπάσματος, ομοίως για όλες τις περιπτώσεις.



Σχήμα 5.11: Επίδοση πιθανοτήτων για απόσπασμα μήκους 1000 εικόνων με μισή ανάλυση από το αρχικό βίντεο

Όταν τα αποσπάσματα με τα οποία έγινε η αναζήτηση έχουν τη μισή οριζόντια και κάθετη ανάλυση (δηλαδή το 1/4 της συνολικής ανάλυσης) από ότι το αρχικό βίντεο, αναμένεται να αλλάξουν αρκετά τα αποτελέσματα σε σχέση με τα προηγούμενα. Οι αλλαγές αυτές φαίνονται εύκολα στα σχήματα 5.11 και 5.12. Για τη μέθοδο των πιθανοτήτων, θα πρέπει να μεγαλώσει η ανοχή σε διαφορές, αφού φαίνεται ότι χρειάζεται ανοχή σε 8 περίπου διαφορές έτσι ώστε η ανάκληση να φτάσει πάνω από το 90%, κάτι όμως που μειώνει αρκετά την ακρίβεια. Αν υπάρχει ανοχή 2 διαφορετικών αλλαγών πλάνου, ανάκληση και ακρίβεια είναι κοντά στο 75%. Αναφορικά με τη μέθοδο των intra macroblock, κατώφλια κάτω του 40 έχουν πολύ κακή επίδοση στην ανάκληση, αφού λόγω της πολύ μεγάλης διαφοράς του αρχικού βίντεο με το απόσπασμα, προστίθεται πάντοτε μια τέτοια διαφορά. Για ένα κατώφλι ίσο με 60, η επίδοση των μεθόδων

5.3. ΑΠΟΔΟΣΗ ΑΝΑΖΗΤΗΣΗΣ



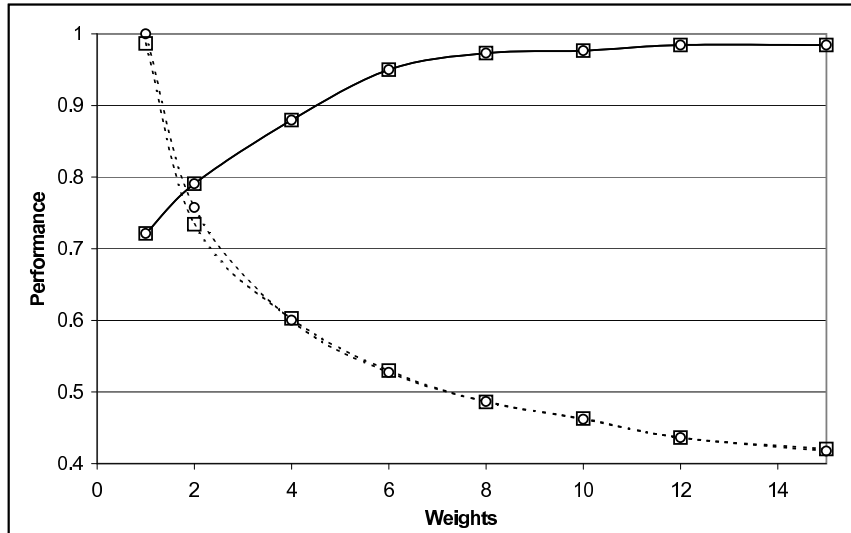
Σχήμα 5.12: Επίδοση *intra macroblock* για απόσπασμα μήκους 1000 εικόνων με μισή ανάλυση από το αρχικό βίντεο

στην ανάκληση είναι πάνω από 85% ενώ στην ακρίβεια περίπου 75%. Σημαντικό είναι ότι σε αυτή την περίπτωση η μέθοδος των αποστάσεων πετυχαίνει πολύ καλή ακρίβεια, ενώ δε διαφέρει σχεδόν καθόλου στην ανάκληση από τις άλλες δύο μεθόδους.

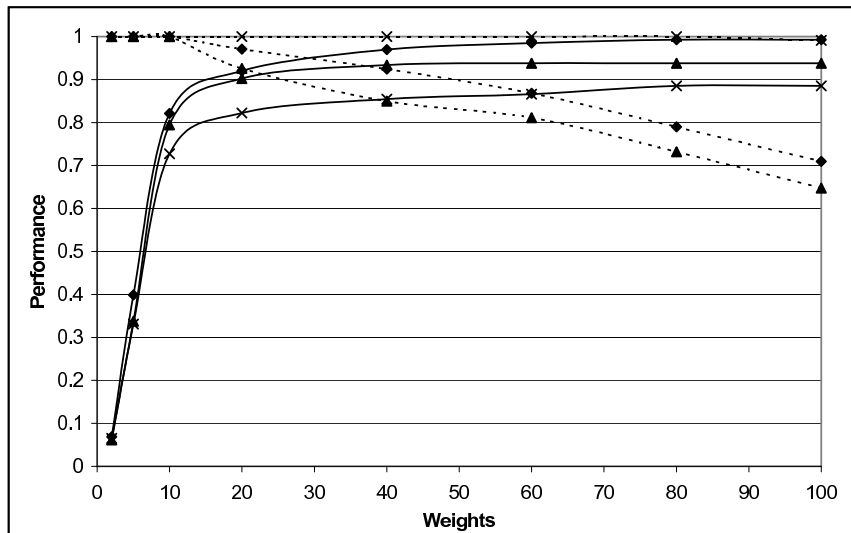
Στα σχήματα 5.13 και 5.14 φαίνεται η επίδοση για αποσπάσματα μήκους 1000 εικόνων που έχουν επανασυμπίεστεί με $Q = 30$ αντί για $Q = 26$ που ήταν το αρχικό βίντεο. Η επίδοση σε όλες τις περιπτώσεις είναι αρκετά καλή, απλώς, στη μέθοδο των πιθανοτήτων χρειάζεται η επιλογή ενός κατώφλιου ίσο με 6 διαφορές (αντί για 4 που ήθελε η περίπτωση του $Q = 26$), ενώ στην περίπτωση του *intra macroblock* θα πρέπει σαν κατώφλι να επιλεγεί μια *intra mb* απόσταση ίση με περίπου 20 (αντί για 10 στο $Q = 26$). Πρέπει εδώ να σημειωθεί ότι η διαφορά στην ποιότητα του $Q = 26$ από το $Q = 30$ είναι αρκετά μεγάλη, μιας και υπήρχε μια μείωση μεγέθους της τάξης του 40% στα αποσπάσματα που συμπίεστηκαν με $Q = 30$ σε σχέση με τα αποσπάσματα που συμπίεστηκαν με $Q = 26$. Φυσικά, η επίδοση της ακρίβειας στη μέθοδο των πιθανοτήτων πέφτει αρκετά λόγω του μεγαλύτερου κατώφλιου που πρέπει να επιλεγεί ώστε να επιτευχθεί καλή επίδοση στην ανάλυση. Αντίθετα, στη μέθοδο των *intra macroblock* παρότι επιλέγεται μεγαλύτερο κατώφλι, η επίδοση της ακρίβειας βρίσκεται πάνω από το 80%.

Τέλος, στα σχήματα 5.15 και 5.16 φαίνεται η επίδοση των μεθόδων όταν τα αποσπάσματα συμπίεστούν με $Q = 40$. Εδώ υπάρχουν πολύ μεγάλες διαφορές μεταξύ του αρχικού βίντεο και των αποσπασμάτων, αφού υπάρχει μια μείωση του μεγέθους κατά περίπου 80% σε ένα βίντεο που έχει συμπίεστεί με $Q = 40$ από ότι αν συμπίεστεί με $Q = 26$. Για τη μέθοδο των πιθανοτήτων πρέπει να επιλεγεί κάποιο κατώφλι ίσο με 8 έτσι ώστε να υπάρχει επίδοση στην ανάκληση

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

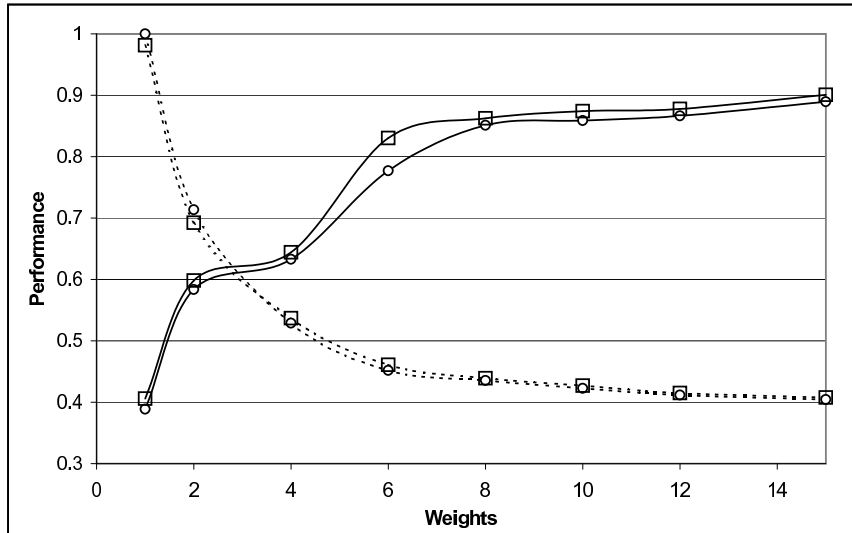


Σχήμα 5.13: Επίδοση πιθανοτήτων για απόσπασμα μήκους 1000 εικόνων με $Q = 30$

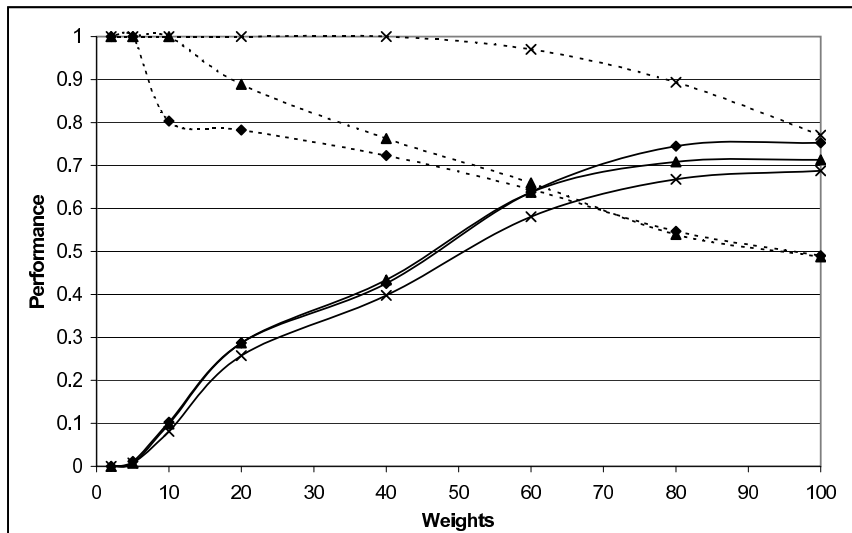


Σχήμα 5.14: Επίδοση *intra macroblock* για απόσπασμα μήκους 1000 εικόνων με $Q = 30$

5.3. ΑΠΟΔΟΣΗ ΑΝΑΖΗΤΗΣΗΣ



Σχήμα 5.15: Επίδοση πιθανοτήτων για απόσπασμα μήκους 1000 εικόνων με $Q = 40$



Σχήμα 5.16: Επίδοση *intra macroblock* για απόσπασμα μήκους 1000 εικόνων με $Q = 40$

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

ίση με περίπου 80% και στην ακρίβεια 45%. Είναι λογικό η επίδοση της ακρίβειας να πέφτει αρκετά με αυτό το μεγάλο κατώφλι αντίστοιχα με το $Q = 30$. Για τη μέθοδο των *intra macroblock* θα πρέπει το κατώφλι να είναι περίπου 70, έτσι ώστε να υπάρχει μια ανεκτή επίδοση στη ανάκληση, κάτι όμως που ρίχνει αρκετά την επίδοση στην ακρίβεια. Αυτό δεν ισχύει για τη μέθοδο των αποστάσεων η οποία αποτελεί μια καλή λύση, αφού μπορεί μεν να έχει ελαφρώς χειρότερη επίδοση από τις άλλες στην ανάκληση, επιτυγχάνει όμως μια πολύ καλή επίδοση στην ακρίβεια, ακόμα και για κατώφλι ίσο με 80.

Κάποια γενικά συμπεράσματα λοιπόν από τα προηγούμενα είναι ότι στη μέθοδο των πιθανοτήτων οι διαφορές μεταξύ της γρήγορης και απλής μεθόδου πιθανοτήτων είναι αμελητέες, έτσι όταν γίνεται αναζήτηση πιθανοτήτων είναι πολύ καλύτερο να χρησιμοποιείται η γρήγορη μέθοδος αναζήτησης αφού έχει πολύ μεγάλη διαφορά στην ταχύτητα αναζήτησης. Αντίθετα, για τη μέθοδο των *intra macroblock* καλύτερη επίδοση στην ανάκληση επιτυγχάνεται συνήθως με την απλή αναζήτηση, αφού εκεί εξετάζονται όλες οι δυνατές θέσεις κάποιες εκ των οποίων μπορεί να αγνοηθούν από τις άλλες μεθόδους, ενώ καλύτερη ακρίβεια υπάρχει στη μέθοδο των αποστάσεων η οποία είναι αρκετά αυστηρή στην αναζήτηση. Η ανάκληση στη μέθοδο των πιθανοτήτων είναι σχετικά καλύτερη από τη μέθοδο των *intra macroblock* για την περίπτωση που υπάρχουν πολύ μεγάλες αλλαγές στο απόσπασμα ($Q = 40$) αφού βασίζεται σε κάποια πιο γενικά κριτήρια δηλαδή την ύπαρξη ή όχι αλλαγής πλάνου. Αντίθετα, η ακρίβεια είναι καλύτερη στην περίπτωση της αναζήτησης με τα *intra macroblock*.

5.3.2 Μέθοδος βάσης δεδομένων

Και εδώ, όπως και στην ενότητα 5.2.1 τα αποτελέσματα για τη μέθοδο της βάσης δεδομένων θα παρουσιαστούν σε ξεχωριστή ενότητα αφού, επειδή χρησιμοποιείται αναγκαστικά διαφορετική βάση δεδομένων δε χρησιμοποιήθηκαν τα ίδια βίντεο για τις μετρήσεις. Χρησιμοποιήθηκε η ίδια βάση με τις 7 ώρες βίντεο όπως στην ενότητα 5.2.1, από όπου εξάχθηκαν αποσπάσματα τα οποία περιείχαν τουλάχιστον 4 αλλαγές πλάνου⁵. Τα αποσπάσματα αυτά επανακωδικοποιήθηκαν σε διαφορετική ανάλυση και ποιότητα, αντίστοιχα με την εργασία που έγινε στην ενότητα 5.3.1, απλώς εδώ τα αποτελέσματα θα παρουσιαστούν όλα μαζί και όχι ξεχωριστά όπως προηγουμένως.

Τα αποτελέσματα φαίνονται στον πίνακα 5.3. Έγιναν ουσιαστικά έξι διαφορετικά πειράματα, με διαφορετικές τιμές ανοχής σε jitter, εσφαλμένων και χαμένων αναγνωρίσεων. Όπως μπορεί να φανεί και από τον πίνακα, στο Π1, όπου η ανοχή σε jitter επιλέχθηκε ίση με 0 ms, τα αποτελέσματα είναι άριστα για την ακρίβεια αλλά πολύ άσχημα για την ανάκληση. Αυτό συμβαίνει επειδή στη βάση αποθηκεύονταν απόλυτοι χρόνοι για τη διάρκεια κάθε σκηνής και έτσι η πιθανότητα στο απόσπασμα μετά την επεξεργασία να διαφέρει έστω και λίγο

⁵Έτσι ώστε να υπάρχουν τουλάχιστον 3 διάρκειες μεταξύ αυτών για το ερώτημα στη βάση δεδομένων, αφού η διάρκεια από την αρχή του βίντεο ως την πρώτη αλλαγή πλάνου και από την τελευταία αλλαγή πλάνου ως το τέλος του βίντεο δε μπορούν να χρησιμοποιηθούν

5.3. ΑΠΟΔΟΣΗ ΑΝΑΖΗΤΗΣΗΣ

	Π1	Π2	Π3	Π4	Π5	Π6
FA	0	0	0	1	0	1
MD	0	0	0	0	1	1
jitter	0	50	100	100	100	100
Ανάκληση	0,22	0,47	0,65	0,76	0,74	0,89
Ακρίβεια	1	0,91	0,84	0,58	0,55	0,53

Πίνακας 5.3: Απόδοση για τη μέθοδο της βάσης δεδομένων

ένας χρόνος ήταν πολύ μεγάλη, έτσι η ανάκληση παίρνει πολύ μικρές τιμές. Αντίθετα, η ακρίβεια ίση με 1 σημαίνει ότι δεν βρέθηκε κανένα αποτέλεσμα που να μην είναι πραγματικό, αφού η πιθανότητα να βρεθεί η ίδια ακολουθία από διάρκειες σε ένα τυχαίο σημείο είναι ελάχιστη. Αντίθετα, στο Π2 που επιλέχθηκε ανοχή σε jitter ίση με 50 ms η ανάκληση βελτιώθηκε αισθητά αλλά μειώθηκε λίγο η ακρίβεια. Επίσης, ίδια συμπεριφορά εμφανίστηκε και για τα 100 ms με ακόμη περισσότερη βελτίωση της ανάκλησης και μείωση της ακρίβειας. Η τιμή των 100 ms χρησιμοποιήθηκε και για τα υπόλοιπα πειράματα μιας και είναι λογικό η διάρκειες των αλλαγών πλάνου να διαφέρουν μεταξύ τους κατά 100 ms που είναι περίπου 2 εικόνες στις 25 εικόνες ανά δευτερόλεπτο⁶.

Τα επόμενα δύο πειράματα λοιπόν έχουν να κάνουν με την εισαγωγή ανοχής για χαμένες και εσφαλμένες αναγνωρίσεις. Η αναμενόμενη συμπεριφορά είναι ότι η μεν ανάκληση θα βελτιωθεί, αφού πλέον θα βρίσκονται και τα βίντεο που περιέχουν εσφαλμένες ή χαμένες αλλαγές πλάνου, αλλά η ακρίβεια θα χειροτερέψει, αφού θα βρεθούν επίσης και κάποια παραπάνω λανθασμένα βίντεο λόγω των επιπλέον ερωτημάτων στη βάση. Μπορεί να φανεί ότι η βελτίωση στα Π4 και Π5 σε σχέση με το Π3 είναι παρόμοια, δηλαδή είτε υπάρχει ανοχή σε χαμένες είτε σε εσφαλμένες αναγνωρίσεις η βελτίωση στην απόδοση αναζήτησης θα είναι παρόμοια. Αυτό ενισχύει ακόμα περισσότερο την δικαιολόγηση της ενότητας 5.2.1 για αποφυγή εισαγωγής ανοχής σε χαμένες αλλαγές πλάνου, αφού η βελτίωση είναι παρόμοια είτε εισαχθεί ανοχή σε εσφαλμένες αλλαγές πλάνου, είτε σε χαμένες αλλαγές πλάνου. Μάλιστα, στα συγκεκριμένα πειραματικά δεδομένα, η εισαγωγή ανοχής σε μια εσφαλμένη αλλαγή πλάνου (FA=1) είχε καλύτερα αποτελέσματα τόσο στην ακρίβεια όσο και στην ανάκληση.

Τέλος, το πείραμα Π6 αφορά την εισαγωγή ταυτόχρονα ανοχής και σε εσφαλμένες και σε χαμένες αλλαγές πλάνου. Εδώ η ανάκληση αυξάνεται ακόμα περισσότερο και φθάνει περίπου το 90%, όμως η ακρίβεια πέφτει αρκετά αφού πλέον θα εμφανίζονται πολλά εσφαλμένα βίντεο στα αποτελέσματα και είναι είναι περίπου 50%, δηλαδή παρότι 9 στα 10 τυχαία βίντεο θα βρεθούν τελικά στο σημείο που πρέπει, ένα στα δύο αποτελέσματα κάθε αναζήτησης δε θα είναι το σωστό βίντεο. Παρόλα αυτά, η επίδοση 90% στην ανάκληση είναι πολύ καλή αν ληφθούν υπ' όψιν οι μεγάλες αλλαγές που έγιναν στα αποσπάσματα που χρησιμοποιήθηκαν για τις δοκιμές, ενώ η επίδοση 50% στην ακρίβεια ουσιαστικά

⁶ Αφού εκεί κάθε εικόνα έχει διάρκεια 1 s / 25 fps = 40 ms

ΚΕΦΑΛΑΙΟ 5. ΑΠΟΤΕΛΕΣΜΑΤΑ

δεν ενοχλεί τόσο, αφού όπως έχει αναφερθεί και αλλού, οι μέθοδοι που παρουσιάζονται κυρίως χρησιμοποιούνται σαν ένα πρώτο επίπεδο αναζήτησης και στη συνέχεια μπορεί είτε να χρησιμοποιηθεί στα αποτελέσματα μια απλή μέθοδος σύγκρισης ώστε να αποκλειστεί το λάθος αποτέλεσμα είτε ακόμα και να εμφανιστούν όλα τα αποτελέσματα στο χρήστη και αυτός να επιλέξει εποπτικά το σωστό.

Κεφάλαιο 6

Συμπεράσματα

ΟΙ περισσότερες υπάρχουσες εμπορικές εφαρμογές αναζήτησης βίντεο βασίζονται σε αναζήτηση με λέξεις κλειδιά όπου χρησιμοποιούνται οι γνωστές τεχνικές από τις μηχανές αναζήτησης κειμένου. Στον τομέα της ευθείας αναζήτησης, παρά τις κρίσιμες εφαρμογές που βασίζονται εκεί, οι εμπορικές εφαρμογές είναι ελάχιστες, μιας και τα προβλήματα της ταχύτητας δεικτοδότησης και αναζήτησης για μεγάλες κλίμακες γίνονται πολύ δύσκολα, ενώ πάντα υπάρχει η απαίτηση της ορθής λειτουργίας της αναζήτησης όταν το βίντεο υποστεί επεξεργασία.

Στα προηγούμενα παρουσιάστηκαν μια σειρά από τεχνικές οι οποίες επιλύουν το πρόβλημα της απ' ευθείας δεικτοδότησης των βίντεο εξάγωντας είτε τις αλλαγές πλάνου είτε το ποσοστό των intra macroblock των εικόνων ενός βίντεο. Το πλεονέκτημα αυτών είναι ότι η εξαγωγή τους μπορεί να γίνει ακόμα και από το συμπιεσμένο βίντεο χωρίς να χρειάζεται ιδιαίτερη επεξεργασία, κάτι που έχει μεγάλη σημασία σε σύγχρονα πρωτόκολλα βίντεο όπως το H.264 στα οποία δε λειτουργούν οι παραδοσιακές τεχνικές εξαγωγής χαρακτηριστικών αναζήτησης.

Με τη βοήθεια αυτών των χαρακτηριστικών, εξετάστηκαν μια σειρά από τεχνικές για την αναζήτηση αποσπασμάτων σε πλήρη βίντεο. Οι τεχνικές αυτές είναι εξαιρετικά γρήγορες και ευέλικτες σε τυχόν επεξεργασία του βίντεο και ανήκουν σε τρεις ξεχωριστές κατηγορίες:

- Τεχνικές που βασίζονται σε μια μέθοδο με βάση την πιθανότητα να διαφέρουν οι αλλαγές πλάνου δύο αποσπασμάτων,
- τεχνικές που υλοποιήθηκαν σε μια βάση δεδομένων και εξαρτώνται από το διάστημα μεταξύ των αλλαγών πλάνου του αποσπάσματος και
- τεχνικές που χρησιμοποιούν το ποσοστό των intra macroblock κάθε εικόνας του βίντεο.

Για κάθε μία από τις παραπάνω εξετάστηκαν διάφορες παραλλαγές οι οποίες βελτιώνουν είτε την ανοχή της σε χαμένες ή εσφαλμένες αλλαγές πλάνου είτε

ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ

την ταχύτητα της και έγιναν αναλυτικά πειράματα που εξετάζουν την επίδοση τους τόσο στο αν βρίσκουν τα αποσπάσματα που αναζητώνται (ανάκληση) όσο και αν επιστρέφουν ως σωστά κάποια αποτελέσματα που δε θα έπρεπε (ακρίβεια). Τα αποτελέσματα των πειραμάτων αποδεικνύουν ότι για όλες τις περιπτώσεις οι μέθοδοι δουλεύουν αρκετά καλά έχοντας πολύ καλές επιδόσεις τόσο στην ανάκληση όσο και στην ακρίβεια. Φυσικά, αυτό που ενδιαφέρει κυρίως είναι η ανάκληση μιας και οι μέθοδοι, λόγω των πολύ γενικών χαρακτηριστικών αναζήτησης και της ταχύτητας τους, έχουν σημαντική εφαρμογή σαν ένα πρώτο επίπεδο αναζήτησης, όπου θα γίνεται μια πολύ γρήγορη επιλογή κάποιων υποψηφίων οι οποίοι στη συνέχεια θα εξετάζονται από άλλες, πιο αναλυτικές μεθόδους.

Οι τεχνικές αυτές υλοποιήθηκαν σε πραγματικά συστήματα: Ένα σύστημα για την εξαγωγή αναλυτικών πειραματικών αποτελεσμάτων για όλες τις μεθόδους, ένα σύστημα που υλοποιούσε ένα web interface μέσω του οποίου μπορούσε να γίνει η αναζήτηση ενός αποσπάσματος και ένα σύστημα το οποίο μπορούσε να αναζητήσει κάποιο απόσπασμα σε συνεχές βίντεο, για παράδειγμα μια διαφήμιση σε τηλεοπτικό απόσπασμα. Η υλοποίηση και τα πειράματα αποδεικνύουν την εύκολη χρήση των τεχνικών αναζήτησης βίντεο σε πραγματικές εφαρμογές και την ταχύτητα τους. Επίσης, είναι εύκολη η επέκταση του συστήματος της αναζήτησης έτσι ώστε η αναζήτηση να γίνεται ταυτόχρονα από πολλούς υπολογιστές, κάτι που επιτρέπει τη γρήγορη αναζήτηση σε ακόμα μεγαλύτερο χώρο αναζήτησης.

Η έρευνα της ανά χείρας διατριβής μπορεί να επεκταθεί σε αρκετούς άξονες, για παράδειγμα: Σύμφωνα με την ενότητα 3.8.2, θα πρέπει να εξεταστούν παραπάνω οι μέθοδοι που βελτιώνουν τις διαφορές στην απόσταση intra mb όταν αλλάζει το Q . Αυτό είναι αρκετά προφανές στα πειράματα που έγιναν για αποσπάσματα με $Q = 40$ όταν το αρχικό βίντεο είχε $Q = 26$, όπου η επίδοση τόσο στην ανάκληση όσο και στην ακρίβεια ήταν αρκετά μειωμένη σε σχέση με τις άλλες περιπτώσεις. Ένας τρόπος για τη βελτίωση αυτών των διαφορών είναι η χρήση της πολλαπλής intra mb απόστασης όπου το βίντεο χωρίζεται σε τμήματα και για κάθε ένα βρίσκεται ξεχωριστά η intra mb απόσταση (για παράδειγμα $2 \times 2 = 4$ τμήματα). Επίσης προτάθηκε και η εύρεση ενός μέσου όρου διαφοράς για κάθε τμήμα των δύο βίντεο που συγκρίνονται και η αφαίρεση αυτών των διαφορών από τη συνολική. Άλλος τρόπος είναι η προσθήκη στην intra mb υπογραφή χαρακτηριστικών που έχουν να κάνουν με τη μεταβολή του ποσοστού των intra macroblock μεταξύ των εικόνων για παράδειγμα αύξηση ή μείωση. Προφανώς όλες αυτές οι βελτιώσεις αυξάνουν την πολυπλοκότητα αναζήτησης της μεθόδου ενώ βελτιώνουν την επίδοση - μένει να εξεταστεί το πόσο.

Είναι σημαντική η λειτουργία της μεθόδου εύρεσης αλλαγών πλάνου και εξαγωγής του ποσοστού των intra macroblock σε ένα H264 αυθαίρετης μορφής, το οποίο δηλαδή περιέχει και εικόνες B, και αυθαίρετα τοποθετημένες εικόνες I, αφού η περίπτωση αυτή εξετάζεται μόνο για τη μέθοδο των πιθανοτήτων όπου και προτείνεται μια λύση βασισμένη στο μοντέλο των πιθανοτήτων που

αναλύθηκε εκεί. Λύσεις για αυτές τις διαφορετικές μορφές του βίντεο έχουν προταθεί ήδη στη βιβλιογραφία και αναφέρονται στα κεφάλαια 2 και 3 (εξέταση αν οι εικόνες B προέρχονται από την προηγούμενη ή επόμενη εικόνα για την εύρεση των αλλαγών πλάνου, εξέταση των χαρακτηριστικών των macroblock I για να βρεθεί αν η εικόνα I είναι αλλαγή πλάνου ή όχι κοκ), πρέπει όμως να εξεταστεί η επίδοση των τεχνικών αναζήτησης που παρουσιάζονται εδώ όταν χρησιμοποιηθούν αυτές οι λύσεις για την εξαγωγή υπογραφής αυθαίρετων βίντεο.

Οι τεχνικές που παρουσιάστηκαν είχαν σαν κύριο σκοπό τους την πολύ γρήγορη δεικτοδότηση και αναζήτηση των αποσπασμάτων. Μια δεύτερη, πιο αναλυτική μέθοδος αναζήτησης μπορεί να χρησιμοποιηθεί στα αποτελέσματα αυτά, έτσι ώστε να αυξηθεί η ακρίβεια της αναζήτησης (χωρίς φυσικά να υποβαθμιστεί η ανάκληση). Τέτοιες μέθοδοι υπάρχουν ήδη στη βιβλιογραφία και μπορούν να προταθούν κι άλλες οι οποίες να συνεργάζονται αρμονικά με τις τεχνικές που παρουσιάστηκαν έτσι ώστε να αυξάνεται η επίδοση της αναζήτησης βίντεο χωρίς να μειώνεται η ταχύτητα αυτής. Η υλοποίηση ενός πλήρους συστήματος που θα χρησιμοποιεί συνδυασμό τεχνικών αναζήτησης (ακόμα και τεχνικών που παρουσιάστηκαν εδώ) είναι μια σημαντική μελλοντική επέκταση της εργασίας αυτής.

Το πιθανοτικό μοντέλο για τις διαφορές στον αριθμό των αλλαγών πλάνου μπορεί να επεκταθεί με διάφορους τρόπους, για παράδειγμα να περιλαμβάνονται περισσότερα από ένα είδος συμβάντων. Επίσης, μπορεί να χρησιμοποιηθεί για να βρεθούν κάποια θεωρητικά όρια για την ανίχνευση αναφορικά με το μέγεθος της βάσης δεδομένων, το μέγεθος του αποσπάσματος και το πόσο συχνά εμφανίζονται αλλαγές πλάνου.

Η μέθοδος της βάσης δεδομένων επιδέχεται μια σειρά από βελτιώσεις που έχουν να κάνουν με την εισαγωγή κατάλληλων δεικτών στη βάση δεδομένων έτσι ώστε να επιταχύνεται η αναζήτηση.

Τέλος, εκτός από την υλοποίηση των προγραμμάτων αναζήτησης από web και συνεχούς αναζήτησης, θα πρέπει να υλοποιηθούν οι τεχνικές αυτές σε ένα σύστημα αποθήκευσης βίντεο στο διαδίκτυο, όπως το Youtube. Η χρήση αυτών των τεχνικών, θα βοηθούσε πολύ στην εύρεση διπλών βίντεο έτσι ώστε να εξοικονομείται αποθηκευτικός χώρος στους servers αλλά και οι χρήστες να μη χρειάζεται να δουν τα διπλά βίντεο. Ομοίως, η υλοποίηση κάποιου πρωτοτύπου που θα χρησιμοποιούσε τις τεχνικές αυτές για την αναγνώριση της ανταλλαγής παράνομου υλικού θα βοηθούσε πάρα πολύ στον πόλεμο κατά της πειρατίας στο δίκτυο. Επίσης, μπορεί να εξεταστεί η συμπεριφορά της αναζήτησης όταν χρησιμοποιούνται περισσότεροι από ένας υπολογιστές ταυτόχρονα. Παρότι φαίνεται από την ανάλυση ότι η ταχύτητα θα αυξάνει αναλογικά με τον αριθμό των υπολογιστών λόγω του εύκολου διαμοιρασμού των δεδομένων σε πολλούς υπολογιστές, χρειάζονται πειραματικά αποτελέσματα για να αποδειχθεί αυτό.

ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Παράρτημα Α΄

Τι υλοποιήθηκε

Στα πλαίσια της διδακτορικής διατριβής υλοποιήθηκαν μια σειρά από προγράμματα, τόσο για έλεγχο των προτεινόμενων τεχνικών όσο και για εξαγωγή πειραματικών αποτελεσμάτων. Στα επόμενα, θα γίνει μια απλή αναφορά όλων αυτών των προγραμμάτων, για ευκολότερη αναφορά σε αυτά μέσα στο κείμενο της διατριβής.

Α΄.1 Αρχεία txt

Είναι αρχεία που περιέχουν κάποια αποτελέσματα τα οποία χρησιμοποιούνται σε πολλά σημεία: Τα αρχεία αυτά είναι απλά αρχεία της μορφής SSV - Semicolon Separated Values, δηλαδή περιέχουν τον αριθμό κάθε εικόνας και το ποσοστό των intra macroblock χωρισμένα με ερωτηματικά, ενώ κάθε εικόνα χωρίζεται από την άλλη με το newline. Η μορφή αυτή επιτρέπει την εύκολη επεξεργασία τόσο εποπτικά όσο και με το απ' ευθείας άνοιγμα του αρχείου σε πρόγραμμα όπως το Microsoft Excel για τη δημιουργία διαγραμμάτων.

Παράδειγμα ενός αρχείου txt:

```
1;1.000000
2;0.103030
3;0.084848
4;0.060606
5;0.036364
6;0.030303
7;0.000000
8;0.012121
9;0.036364
10;0.018182
```

A'.2 GetHistYUV

Ένα πρόγραμμα σε *C*, χρησιμοποιεί τη βιβλιοθήκη Ffmpeg για να δημιουργήσει το ιστόγραμμα κάθε ενός frame από ένα βίντεο MPEG κάθε είδους. Το βίντεο απλώς αποσυμπιέζεται, και, επειδή ένα αποσυμπιεσμένο MPEG είναι σε YUV, απλώς μετρώνται τα pixels με τη φωτεινότητα. Το ιστόγραμμα δημιουργείται με βάση την περιγραφή της 2.5 και οι τιμές για κάθε frame και κάθε κουβά γράφονται η μία κάτω από την άλλη.

A'.3 HistShotDet

Ένα πρόγραμμα σε Perl το οποίο δέχεται ένα αρχείο ιστογραμμάτων και, ελέγχοντας τη διαφορά μεταξύ των γειτονικών frame καθορίζει το ποιά frames αποτελούν αλλαγή πλάνου. Το αποτέλεσμα γράφεται σε μια λίστα από 0 και 1 για κάθε frame, όπου το 0 σημαίνει ότι το τρέχον frame δεν είναι αλλαγή πλάνου, και το 1 σημαίνει αλλαγή πλάνου.

A'.4 H264GetCtxt

Το πρόγραμμα αυτο είναι γραμμένο σε *C* και χρησιμοποιεί την Ffmpeg έτσι ώστε να εξάγει τα ποσοστά των intra macroblock ενός αρχείου H.264 συγκεκριμένης μορφής. Τα ποσοστά αυτά γράφονται το ένα κάτω από το άλλο, σε ένα αρχείο ctxt.

A'.5 CtxtShotDet

Είναι ένα πρόγραμμα σε Perl το οποίο δέχεται ένα αρχείο ctxt και, ελέγχοντας αν το ποσοστό των intra macroblock σε κάποιο frame είναι πάνω από κάθε κατώφλι, επιλέγει τις αλλαγές πλάνου. ο αποτέλεσμα γράφεται σε μια λίστα από 0 και 1 για κάθε frame, όπου το 0 σημαίνει ότι το τρέχον frame δεν είναι αλλαγή πλάνου, και το 1 σημαίνει αλλαγή πλάνου.

A'.6 ShotCutDetPerf

Ένα πρόγραμμα σε Perl το οποίο μετρά τα N_c , N_f και N_m μεταξύ δύο αρχείων με αλλαγές πλάνου με 0 και 1 και υπολογίζει τα Recall και Precision, θεωρώντας το ένα αρχείο ως «σωστό» και το άλλο ως «τροποποιημένο».

A'.7 TextVideoAdd

Ένα απλό πρόγραμμα το οποίο χρησιμοποιήθηκε για την εξαγωγή των πειραματικών αποτελεσμάτων της εργασίας. Προσθέτει ένα αρχείο ctxt στη βάση

δεδομένων αναζήτησης.

A'.8 TextVideoSearch

Ένα απλό πρόγραμμα το οποίο χρησιμοποιήθηκε για την εξαγωγή των πειραματικών αποτελεσμάτων της εργασίας. Δέχεται ως είσοδο ένα αρχείο `ctxt` και το αναζητά στη βάση δεδομένων που έχει δημιουργηθεί από το `TextVideoAdd`. Επίσης, μπορεί να πάρει παραμέτρους που καθορίζουν το ποιά μέθοδος θα χρησιμοποιηθεί κατά την αναζήτηση καθώς και το ποιά βάση δεδομένων θα χρησιμοποιηθεί.

A'.9 TextVideoSceneChangeGUI

Ένα πρόγραμμα με γραφικό περιβάλλον εργασίας που χρησιμοποιήθηκε για την εύκολη δημιουργία μιας σειράς πειραματικών αποτελεσμάτων. Το πρόγραμμα αυτό γράφτηκε με τη βοήθεια των Windows MFC και μια εικόνα του φαίνεται στο σχήμα 3.10.

Όπως φαίνεται από το σχήμα, το πρόγραμμα περιέχει δύο λίστες στις οποίες φορτώνονται μια σειρά από αρχεία `ctxt`. Στην αριστερή λίστα επιλέγεται ένα βίντεο που θα χρησιμοποιηθεί ως απόσπασμα και στη δεξιά ένα βίντεο στο οποίο θα αναζητηθεί το αριστερό βίντεο. Τα αριστερά βίντεο ονομάζονται `small` και τα δεξιά `big`. Υπάρχουν επιλογές για το ποιά θα είναι το κατώφλι πάνω από το οποίο θα θεωρηθεί ένα ποσοστό `intra macroblock` ως αλλαγή πλάνου (`small threshold`, `big threshold`), ποιές ακριβώς εικόνες θα επιλεγούν για την αναζήτηση και από τα δύο βίντεο (`small low`, `small high`, `big low`, `big high`), υπολογισμός του e για το επιλεγμένο `big` και κατώφλι, υπολογισμός των f και m για το `big`, `small` και όλες τις άλλες επιλογές, και τέλος επιλογές για το συγχρονισμό του επιλεχθέντος `small` στο `big` αλλά και σε όλα τα βίντεο της δεξιάς στήλης έτσι ώστε να δημιουργηθούν τα γραφήματα που παρουσιάζονται αναλυτικά στην ενότητα 3.6.6.

ΠΑΡΑΡΤΗΜΑ Α'. ΤΙ ΥΛΟΠΟΙΗΘΗΚΕ

Παράρτημα Β΄

Δημοσιεύσεις του συγγραφέα

- G. Doumenis, G. Stassinopoulos, F. Andritsopoulos and S. Papastefanos. Video Characterisation Identification and Search System. 2009 Patent
- George Stassinopoulos and Serafeim Papastefanos. Efficient selection of candidates in video content search. International Journal of Electronics and Communications, June 2009.
- G. Stassinopoulos, S. Papastefanos, and F. Andritsopoulos. Fast video search using scene durations. In Proceeding of the 1st ACM workshop on Analysis and retrieval of events/actions and workflows in video streams, pages 113 • 118, Vancouver, Canada, October 2008.
- F. Andritsopoulos, S. Papastefanos, V. Mpilili, and C. Theoharatos. An advanced direct searching technique applied on compressed video content repositories. In DIMEA 2008, Athens, Greece, September 2008.
- S. Papastefanos, F. Andritsopoulos, V. Mpilili, C. Theoharatos, and N. Achilleopoulos. Direct searching of multimedia content based on video characteristics extracted from compressed domain. In DIMEA 2008 Demo, Athens, Greece, September 2008.
- N. Achilleopoulos, C. Theoharatos, F. Andritsopoulos, and S. Papastefanos. An extensible scheme for direct searching in audiovisual archives: The divas system. In IEEE International Symposium on Consumer Electronics, Algarve, Portugal, April 2008.
- Gregory Doumenis, Serafeim Papastefanos, George Stassinopoulos, Fotis Andritsopoulos, Nikos Achilleopoulos, and Anton Mikhalev. Video index and search services based on content identification features.

ΠΑΡΑΡΤΗΜΑ Β'. ΔΗΜΟΣΙΕΥΣΕΙΣ ΤΟΥ ΣΥΓΓΡΑΦΕΑ

In IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, April 2008.

- F. Andritsopoulos, S. Papastefanos, G. Georgakarakos, and G. Doumenis. Reliable multicast h.264 video streaming for surveillance applications. In 18nth Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Athens, Greece, September 2007.
- Y. Royon, P. Parrend, S. Frenot, S. Papastefanos, H. Abdelnur, and D. Van de Poel. Multi-service, multi-protocol management for residential gateways. In BBEurope 2007, Antwerp, Belgium, December 2007.
- A. Nikolaidis, S. Papastefanos, G. Doumenis, G. Stassinopoulos, and M. Drakos. Local and remote management integration for exible service provisioning to the home. IEEE Communications Magazine, October 2007.
- M. Drakos, A. Nikolaidis, S. Papastefanos, and G. Stassinopoulos. Optimal admission control schemes in muse network. In BBEUROPE 2006, Geneva, Switzerland, December 2006.
- S. Papastefanos, A. Nikolaidis, M. Drakos, G. Doumenis, and G. Stassinopoulos. Performance evaluation of the CPE WAN management protocol. In BBEUROPE 2006, December 2006.
- A. Nikolaidis, S. Papastefanos, G. Stassinopoulos, M. Drakos, and G. Doumenis. Automating remote configuration mechanisms for home devices. IEEE Transactions on Consumer Electronics, May 2006.

References

- [ABC⁺03] A. Amir, M. Berg, S.-F. Chang, W. Hsu, G. Iyengar, C.-Y. Lin, M. Naphade, A. Natsev, C. Neti, H. Nock, J. Smith, B. Tseng, Y. Wu, and D. Zhang. IBM research TRECVID-2003 video retrieval system. In *Proc. TRECVID Workshop*, Gaithersburg, MD, 2003.
- [ACGW05] John Adcock, Matthew Cooper, Andreas Girgensohn, and Lynn Wilcox. Interactive video search using multilevel indexing. In *Conference on Image and Video Retrieval (CIVR)*, volume 3568, pages 205–214. Springer-Verlag New York, Inc., 2005.
- [ALT] Altavista video search. Website. <http://www.altavista.com/video/default>.
- [AMB] MPEG-2 to H.264 transcoding: Full re-encoding vs. stream manipulations. Website. <http://www.ambarella.com/docs/Transcoding.pdf>.
- [APMT08] F. Andritsopoulos, S. Papastefanos, V. Mpilili, and C. Theoharatos. An advanced direct searching technique applied on compressed video content repositories. In *DIMEA 2008*, Athens, Greece, September 2008.
- [ATAP08] N. Achilleopoulos, C. Theoharatos, F. Andritsopoulos, and S. Papastefanos. An extensible scheme for direct searching in audiovisual archives: The divas system. In *IEEE International Symposium on Consumer Electronics*, Algarve, Portugal, April 2008.
- [BDC⁺] Sarah De Bruyne, Davy Van Deursen, Jan De Cock, Wesley De Neve, Peter Lambert, and Rik Van de Walle. A compressed-domain approach for shot boundary detection on h.264/avc bit streams.

- [Bes04] Jesús Bescós. Real-time shot change detection over online MPEG-2 video. *IEEE Trans. Circuits Syst. Video Technol.*, 14(4), April 2004.
- [BGG99] Patrick Bouthemy, Marc Gelgon, and Fabrice Ganansia. A unified approach to shot change detection and camera motion characterization. *IEEE Trans. Circuits Syst. Video Technol.*, 9(7), October 1999.
- [BIT] Bittorrent. Website. <http://www.bittorrent.com/>.
- [BLI] Blinkx. Website. <http://www.blinkx.com/>.
- [BN98] D. Bhat and S. Nayar. Ordinal measures for image correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(4):415–423, April 1998.
- [BS99] A. P. Berman and L. G. Shapiro. A flexible image database system for content-based retrieval. *Computer vision and image understanding*, 75(1-2):175–195, 1999.
- [CAS] CastTV. Website. <http://www.casttv.com>.
- [CCPP02] F. Cavalli, R. Cucchiara, M. Piccardi, and A. Prati. Performance analysis of MPEG-4 decoder and encoder. In *4th EURASIP Video/Image Processing and Multimedia Communications*, pages 227–231, 2002.
- [CHMP04] Michael G. Christel, Chang Huang, Neema Moraveji, and Norman Papernick. Exploiting multiple modalities for interactive video retrieval. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP'04)*, volume 3, pages 1032–1035, Montreal, Quebec, Canada, May 2004.
- [CHO⁺98] Tae-Yun Chung, Min-Suk Hong, Young-Nam Oh, Dong-Ho Shin, and Sang-Hui Park. Digital watermarking for copyright protection of MPEG2 compressed video. *IEEE Transactions on Consumer Electronics*, 44(3):895–901, August 1998.
- [cSCZ00] Sen ching Samson Cheung and Avidesh Zakhor. Estimation of web video multiplicity. In *Proc. SPIE*, volume 3964, pages 34–36, 2000.
- [cSCZ03] Sen ching S. Cheung and Avidesh Zakhor. Efficient video similarity measurement with video signature. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):59–74, January 2003.

- [cSCZ05] Sen ching S. Cheung and Avidesh Zakhor. Fast similarity search and clustering of video sequences on the world-wide-web. *IEEE Transactions on Multimedia*, 7(3):534–537, June 2005.
- [CSI06] YARON CASPI, DENIS SIMAKOV, and MICHAL IRANI. Feature-based sequence-to-sequence matching. *International Journal of Computer Vision*, 68(1):53–64, 2006.
- [CSL99] Hyun Sung Chang, Sanghoon Sull, and Sang Uk Lee. Efficient video indexing scheme for content-based retrieval. *IEEE Trans. Circuits Syst. Video Technol.*, 9(8), December 1999.
- [CX06] Yuk Ying Chung and Fang Fei Xu. A secure digital watermarking scheme for MPEG-2 video copyright protection. In *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS'06)*, 2006.
- [DCH04] P. Duygulu, M. Chen, and A. Hauptmann. Comparison and combination of two novel commercial detection methods. In *IEEE International Conference on Multimedia and Expo, 2004. ICME'04*, 2004.
- [DNR05] Anastasios Dimou, Olivia Nemethova, and Markus Rupp. Scene change detection for h.264 using dynamic threshold techniques. In *Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Service*, Smolenice, Slovak Republic, June - July 2005.
- [DPS⁺08] Gregory Doumenis, Serafeim Papastefanos, George Stassinopoulos, Fotis Andritsopoulos, Nikos Achileopoulos, and Anton Mikhalev. Video index and search services based on content identification features. In *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, March 2008.
- [DSAP] G. Doumenis, G. Stassinopoulos, F. Andritsopoulos, and S. Papastefanos. Video characterization, identification and search system (patent).
- [DSH] Directshow at MSDN. Website. [http://msdn.microsoft.com/en-us/library/dd375454\(VS.85\).aspx](http://msdn.microsoft.com/en-us/library/dd375454(VS.85).aspx).
- [Esk03] Ahmet M. Eskicioglu. Protecting intellectual property in digital multimedia networks. *Computer*, 36(7):39–45, July 2003.
- [FAC] Facebook Web Site. Website. <http://www.facebook.com>.
- [Fal96] Christos Faloutsos. *Searching Multimedia Databases by Content*. Springer, September 1996.

- [FFM] FFmpeg. Website. <http://ffmpeg.mplayerhq.hu>.
- [FH06] Edward W. Felten and J. Alex Halderman. Digital rights management, spyware, and security. *IEEE Security and Privacy*, 4(1):18–23, Jan/Feb 2006.
- [FMR02] Gian Luca Foresti, Lucio Marcenaro, and Carlo Regazzoni. Automatic detection and indexing of video-event shots for surveillance applications. *IEEE Trans. Multimedia*, 4(4), dec 2002.
- [FMS99] Borko Furht, Edin Muharemagic, and Daniel Socek. *Multimedia Encryption and Watermarking*. Springer-Verlag New York, Inc., 1999.
- [GA05] J.M. Gauch and Shivadas A. Identification of new commercials using repeated video sequence detection. In *IEEE International Conference on Image Processing, 2005. ICIP 2005*, volume 3, September 2005.
- [GKS00] Ullas Gargi, Rangachar Kasturi, and Susan H. Strayer. Performance characterization of video-shot-change detection methods. *IEEE Trans. Circuits Syst. Video Technol.*, 10(1), February 2000.
- [GPL07] Gnu general public license. Website, June 2007. <http://www.gnu.org/licenses/gpl.html>.
- [HB01] Arun Hampapur and Ruud M. Bolle. Comparison of distance measures for video copy detection. In *IEEE International Conference on Multimedia and Expo, 2001. ICME 2001.*, pages 737–740, August 2001.
- [HF06] J. Alex Halderman and Edward W. Felten. Lessons from the Sony CD DRM episode. In *Proceedings of the 15th conference on USENIX Security Symposium*, volume 15. USENIX Association Berkeley, CA, USA, 2006.
- [HHB02] Arun Hampapur, Ki-Ho Hyun, and Ruud Bolle. Comparison of sequence matching techniques for video copy detection. In *Storage and Retrieval for Media Databases 2002*, volume 4676, pages 194–201, 201, San Jose, CA, USA, January 2002.
- [HJKH03] Michael Horowitz, Anthony Joch, Faouzi Kossentini, and Antti Hallapuro. H.264/avc baseline profile decoder complexity analysis. *IEEE Trans. Circuits Syst. Video Technol.*, 13(7), July 2003.

- [HJN02] Alexander G. Hauptmann, Rong Jin, and Tobun Dorbin Ng. Multi-modal information retrieval from broadcast video using ocr and speech recognition. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 160 – 161, Portland, Oregon, USA, 2002.
- [HKR93] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9), September 1993.
- [HL01] Chung-Lin Huang and Bing-Yao Liao. A robust scene-change detection method for video segmentation. *IEEE Trans. Circuits Syst. Video Technol.*, 11(12):1281–1288, DECEMBER 2001.
- [HSG99] Frank Hartung, Jonathan K. Su, and Bernd Girod. Spread spectrum watermarking: Malicious attacks and counterattacks. In *Proc. SPIE: Security and Watermarking of Multimedia Contents*, 1999.
- [Hu62] N.K. Hu. Visual pattern recognition by moment invariance. *IRE Transactions on Information Theory*, pages 179–187, 1962.
- [HWL03] Chu-Hong Hoi, Wei Wang, and Michael R. Lyu. A novel scheme for video similarity detection. *Lecture notes in computer science*, 2728/2003:541–546, 2003.
- [HZ06] T. C Hoad and J. Zobel. Detection of video sequences using compact signatures. *ACM Transactions on Information Systems*, 24(1):1–50, January 2006.
- [IPH] iPharro. Website. <http://www.ipharro.com/>.
- [JBC05] José M. Martínez José M. Menéndez Jesús Bescós, Guillermo Cisneros and Julián Cabrera. A unified model for techniques on video-shot transition detection. *IEEE Trans. Multimedia*, 7(2):2005, April 2005.
- [JVW99] Anil K. Jain, Aditya Vailaya, and Xiong Wei. Query by video clip. *Multimedia Syst.*, 7(5):369–384, 1999.
- [KL08] Tien-Ying Kuo and Yi-Chung Lo. Detection of h.264 shot change using intra predicted direction. In *Proceedings of the 2008 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 204–207, 2008.

- [KOS06] W. Kraaij, P. Over, and A. Smeaton. Trecvid 2006 - an overview. In *TRECVID 2006 Workshop*, November 2006.
- [KP02] Sang Hyun Kim and Rae-Hong Park. An efficient algorithm for video sequence matching using the modified hausdorff distance and the directed divergence. *IEEE Trans. Circuits Syst. Video Technol.*, 12(7), July 2002.
- [LEJ98] R. Lienhart, W. Effelsberg, and R. Jain. VisualGREP: A systematic method to compare and retrieve video sequences. In *Proceedings of the Conference on Storage and Retrieval for Image and Video Databases (SPIE)*, pages 271–283, 1998.
- [LELD05] E.I Lin, A.M. Eskicioglu, R.L. Lagendijk, and E.J. Delp. Advances in digital video content protection. *Proceedings of the IEEE*, 93:171–183, 2005.
- [LGP07] Gnu lesser general public license. Website, June 2007. <http://www.gnu.org/copyleft/lesser.html>.
- [LHZ07] Xiaosong Lou, Kai Hwang, and Runfang Zhou. Integrated copyright protection in peer-to-peer networks. In *27th International Conference on Distributed Computing Systems Workshops (ICDCSW'07)*, 2007.
- [Lie99] R. W. Lienhart. Comparison of automatic shot boundary detection algorithms. In *Proc. SPIE*, volume 3656, pages 290–301, December 1999.
- [LK07] T. Liu and R. Katpelly. A two-level queueing system for interactive browsing and searching of video content. *Multimed. Syst.*, 12(4-5), March 2007.
- [LKE97] R. Lienhart, C. Kuhmunch, and W. Effelsberg. On the detection and recognition of television commercials. In *Proceedings of the International Conference on Multimedia Computing and Systems*, pages 509–516, 1997.
- [LLB98] G.C. Langelaar, R. L. Lagendijk, and J. Biemond. Removing spatial spread spectrum watermarks by nonlinear filtering. In *Proceedings of the European Signal Processing Conference*, volume 4, pages 2281–2284, 1998.
- [LLZL04] Hongliang Li, Guizhong Liu, Zhongwei Zhang, and Yongli Li. Adaptive scene-detection algorithm for vbr video stream. *IEEE Trans. Multimedia*, 6(4):624–633, August 2004.

- [LTCJ⁺07] Julien Law-To, Li Chen, Alexis Joly, Ivan Laptev, Olivier Buisson, Valerie Gouet-Brunet, Nozha Boujemaa, and Fred Stentiford. Video copy detection: a comparative study. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 371 – 378, Amsterdam, The Netherlands, 2007.
- [LXY⁺06] Song Liu, Min Xu, Haoran Yi, Liang-Tien Chia, and Deepu Rajan. Multimodal semantic analysis and annotation for basketball video. *EURASIP Journal on Applied Signal Processing*, 2006:1–13, 2006.
- [LZ01] R. Lienhart and A. Zaccarin. A system for reliable dissolve detection in videos. In *International Conference on Image Processing, 2001*, volume 3, pages 406–409, Thessaloniki, Greece, 2001.
- [LZP99] Xiaoming Liu, Yueting Zhuang, and Yunhe Pan. A new approach to retrieve video by example video clip. In *Proceedings of the seventh ACM international conference on Multimedia*, pages 41–44, Orlando, Florida, USA, 1999.
- [MC05] M.E. Mizutani and S.S.F. Chang. Commercial detection in heterogeneous video streams using fused multi-modal and temporal features. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05)*, volume 2, pages 157–160, 2005.
- [MFN⁺] Sandro Moiron, Sergio Faria, Antonio Navarro, Vitor Silva, and Pedro Assuncao. Video transcoding from H.264/AVC to MPEG-2 with reduced computational complexity.
- [MHB03] Deirdre K. Mulligan, John Han, and Aaron J. Burstein. How drm-based content delivery systems disrupt expectations of ”-personal use”. In *Proceedings of the 3rd ACM workshop on Digital rights management*, pages 77–89. ACM New York, NY, USA, 2003.
- [Mia99] Jędrzej Zdzisław Miadowicz. *Story Tracking in Video News Broadcasts*. PhD thesis, Department of Electrical Engineering and Computer Science, University of Kansas, June 1999.
- [MMS] Microsoft media server protocol. Website. <http://msdn.microsoft.com/en-us/library/cc239490.aspx>.
- [Moh98] R. Mohan. Video sequence matching. In *Proc. of ICASSP98*, volume 6, pages 3697–3700, 1998.

- [MPL] MPlayer. Website. <http://www.mplayerhq.hu>.
- [NH01] Milind Ramesh Naphade and Thomas S. Huang. A probabilistic framework for semantic video indexing, filtering, and retrieval. *IEEE Trans. Multimedia*, 3(1), March 2001.
- [NKL01] Chung Wing Ng, Irwin King, and Michael R. Lyu. Video comparison using tree matching algorithms. In *Proceedings of the International Conference on Imaging Science, Systems and Technology*, volume 1, pages 184–190, Las Vegas, USA, 2001.
- [NYY00] M. R. Naphade, M. M. Yeung, and B.-L. Yeo. Novel scheme for fast and efficient video sequence matching using compact signatures. In *Proc. SPIE, Storage and Retrieval for Media Databases*, volume 3972, pages 564–572, December 2000.
- [OBL⁺04] Jörn Ostermann, Jan Bormans, Peter List, Detlev Marpe, Matthias Narroschke, Fernando Pereira, Thomas Stockhammer, and Thomas Wedi. Video coding with h.264/avc: Tools, performance, and complexity. *IEEE Circuits and Systems Magazine*, 4(1):7–28, 2004.
- [PAM⁺08] S. Papastefanos, F. Andritsopoulos, V. Mpilili, C. Theocharatos, and N. Achilleopoulos. Direct searching of multimedia content based on video characteristics extracted from compressed domain. In *DIMEA 2008 Demo*, Athens, Greece, September 2008.
- [PBG01] Tej Pratap Pandey, Asheesh Bhardwaj, and Sumana Gupta. Robust and efficient method for compressed domain video indexing. In *IEEE International Conference on Multimedia and Expo*, pages 1152–1155, 2001.
- [PC99] Soo-Chang Pei and Yu-Zuon Chou. Efficient MPEG compressed video analysis using macroblock type information. *IEEE Trans. Multimedia*, 1(4), December 1999.
- [Pei02] Marcus Peinado. Digital rights management in a multimedia environment. *SMPTE journal*, 111:159–163, 2002.
- [Pes03] Mark D. Pesce. *Programming Microsoft DirectShow for Digital Video and Television*. Microsoft Press, 2003.
- [PSR93] Ketan Patel, Brian C. Smith, and Lawrence A. Rowe. Performance of a software MPEG video decoder. In *Proceedings of the first ACM international conference on Multimedia*, pages 75 – 82, Anaheim, California, United States, 1993.

- [Ric03] Iain E. G. Richardson. *H.264 and MPEG-4 compression*. John Wiley & Sons, 2003.
- [RSB98] R. Weber, H.-J. Schek, and S. Blott. A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces. In *Proc. 24th Int. Conf. Very-Large Databases (VLDB' 98)*, pages 194–205, New York, August 1998.
- [RTS] Real time streaming protocol (RTSP). Website. <http://www.ietf.org/rfc/rfc2326.txt>.
- [Sam89] Hanan Samet. *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, 1989.
- [SB08] Klaus Schöffmann and Laszlo Böszörményi. Fast segmentation of h.264/avc bitstreams for on-demand video summarization. *Lecture Notes in Computer Science*, 4903/2008:265–276, 2008.
- [Shi99] Narayanan Shivakumar. *Detecting Digital Copyright Violations on the Internet*. PhD thesis, Stanford University, 1999.
- [SMOM02] D.A. Sadlier, S. Marlow, N. O'Connor, and N. Murphy. Automatic tv advertisement detection from mpeg bitstream. *Pattern Recognition*, 35:2719–2726, 2002.
- [SP98] M. Szummer and R. Picard. Indoor-outdoor image classification. In *IEEE Int. Workshop Content-based Access Image Video Databases*, Bombay, India, 1998.
- [SP09] G. Stassinopoulos and S. Papastefanos. Efficient selection of candidates in video content search. *International Journal of Electronics and Communications (AEU)*, June 2009.
- [SPA08] G. Stassinopoulos, S. Papastefanos, and F. Andritsopoulos. Fast video search using scene durations. In *Proceeding of the 1st ACM workshop on Analysis and retrieval of events/actions and workflows in video streams*, pages 113 – 118, Vancouver, Canada, October 2008.
- [SWKS07] C.G.M Snoek, M. Worring, D.C. Koelma, and A.W.M Smeulders. A learned lexicon-driven paradigm for interactive video retrieval. *IEEE Transactions on Multimedia*, 9(2):280–292, February 2007.
- [TCA⁺04] Cuneyt Taskiran, Jau-Yuen Chen, Alberto Albiol, Luis Torres, Charles A. Bouman, and Edward J. Delp. ViBE: A compressed video database structured for active browsing and search. *IEEE Transactions On Multimedia*, 6(1):103–118, February 2004.

- [TRA] TrackID. Website. <http://www.sonyericsson.com/product/trackid/>.
- [TRU] Truveo. Website. <http://www.truveo.com/>.
- [VFJZ01] A. Vailaya, M. Figueiredo, A. Jain, and H.-J. Zhang. Image classification for content-based indexing. *IEEE Trans. Image Process*, 10:117–130, January 2001.
- [WCGH99] Howard D. Wactlar, Michael G. Christel, Yihong Gong, and Alexander G. Hauptmann. Lessons learned from building a terabyte digital video library. *Computer*, 32(2):66–73, February 1999.
- [WDV⁺03] Hualu Wang, Ajay Divakaran, Anthony Vetro, Shih-Fu Chang, and Huifang Sun. Survey of compressed-domain features used in audio-visual indexing and analysis. *J. Vis. Commun. Image R.*, 14(2), June 2003.
- [WHN07] Xiao Wu, Alexander G. Hauptmann, and Chong-Wah Ngo. Practical elimination of near-duplicates from web video search. In *Proceedings of the 15th international conference on Multimedia*, pages 218 – 227, Augsburg, Bavaria, Germany, September 2007.
- [WSBL03] Thomas Wiegand, Gary J. Sullivan, Gisle Bjontegaard, and Ajay Luthra. Overview of the h.264/avc video coding standard. *IEEE Trans. Circuits Syst. Video Technol.*, 13(7), July 2003.
- [WSdR⁺07] M. Worring, C.G.M Snoek, O. de Rooij, G.P Nguyen, and A.W.M Smeulders. The Mediamill semantic video search engine. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007.*, volume 4, pages IV–1213 – IV–1216, 2007.
- [YL95] Boon-Lock Yeo and Bede Liu. Rapid scene analysis on compressed video. *IEEE Trans. Circuits Syst. Video Technol.*, 5(6), December 1995.
- [YLWZ04] Guizhong Yung Liu, Weiqiang Wang, Wen Gao, and Wei Zeng. A novel compressed domain shot segmentation algorithm on h.264/avc. In *International Conference on Image Processing (ICIP) 2004*, volume 4, pages 2235– 2238, October 2004.
- [YLZ02] J. Yang, Q. Li, and Y. Zhuang. In *Proceedings of the Eleventh International Conference on the World Wide Web*, pages 54–64, 2002.

- [YOU] Youtube Web Site. Website. <http://www.youtube.com>.
- [YOU06] Youtube serves up 100 million videos a day online. Website, July 2006. http://www.usatoday.com/tech/news/2006-07-16-youtube-views_x.htm.
- [YTR04] Junsong Yuan, Qi Tian, and Surendra Ranganath. Fast and robust search method for short video clips from large video collection. In *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3*, pages 866–869, 2004.
- [ZQL⁺01] Li Zhao, Wei Qi, S. Z. Li, S. Q. Yang, and H. J. Zhang. Content-based retrieval of video shot using the improved nearest feature line method. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001 (ICASSP '01)*, volume 3, pages 1625–1628, Salt Lake City, UT, USA, 2001.