

ΤΙΤΛΟΣ ΔΙΑΤΡΙΒΗΣ
ΜΕΘΟΔΟΙ ΕΞΑΓΩΓΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΙΚΟΝΟΠΛΑΙΣΙΩΝ
ΑΠΟ ΑΚΟΛΟΥΘΙΕΣ ΒΙΝΤΕΟ

Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

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

από τον

Ελευθέριο Χατζιωαννίδη

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ

Φεβρουάριος 2010

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Αριστείδη Λύκα καθώς και τον κ. Βασίλειο Χασάνη για την πολύτιμη βοήθεια που μου πρόσφεραν κατά την εκπόνηση της εργασίας.

ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ
ΕΥΧΑΡΙΣΤΙΕΣ	ii
ΠΕΡΙΕΧΟΜΕΝΑ	iii
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	vi
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	vii
ΠΕΡΙΛΗΨΗ	ix
EXTENDED ABSTRACT IN ENGLISH	xi
ΚΕΦΑΛΑΙΟ 1. Εισαγωγή	1
1.1. Περιγραφή του προβλήματος	1
1.2. Αντικείμενο της εργασίας	3
1.3. Δομή της εργασίας	4
ΚΕΦΑΛΑΙΟ 2. Βασικές έννοιες και σχετική βιβλιογραφία	6
2.1. Κατηγορίες περίληψης βίντεο	7
2.2. Το μέγεθος του συνόλου των key-frames	7
2.3. Μονάδα αναπαράστασης	8
2.4. Κατηγοριοποίηση των μεθόδων για την εξαγωγή των key-frames	9
2.4.1. Μέθοδος της Επαρκούς Αλλαγής Περιεχομένου	9
2.4.2. Μέθοδος Μέγιστης Κάλυψης Εικονοπλαισίων	12
2.4.3. Μέθοδος Ομαδοποίησης	13
2.4.4. Μέθοδος Ελάχιστης συσχέτισης μεταξύ των key-frames	16
2.4.5. Μέθοδος με χρήση του Σφάλματος ανακατασκευής της ακολουθίας	16
2.4.6. Μέθοδος των ‘Ενδιαφερόντων γεγονότων’	17
ΚΕΦΑΛΑΙΟ 3. Περιγραφείς Εικονοπλαισίων	19
3.1. Ιστογράμματα χρώματος	19
3.2. Περιγραφείς SIFT	20
3.3. Περιγραφείς CCH	22

3.4. Περιγραφείς κίνησης	24
3.5. Σύνολα από οπτικές λέξεις	26
ΚΕΦΑΛΑΙΟ 4. Ομαδοποίηση Εικονοπλαισίων	28
4.1. Ο τυπικός φασματικός αλγόριθμος ομαδοποίησης	28
4.2. Ο αλγόριθμος k-means	29
4.3. Ο αλγόριθμος Fast Global k-means	31
ΚΕΦΑΛΑΙΟ 5. Μεθοδολογίες εξαγωγής χαρακτηριστικών εικονοπλαισίων	33
5.1. Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε ιστογράμματα χρώματος, κίνησης, CCH και SIFT περιγραφών	33
5.2. Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε συνδυασμό λύσεων ομαδοποίησης	35
5.3. Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε κατάτμηση της ακολουθίας βίντεο	37
ΚΕΦΑΛΑΙΟ 6. Μέτρα Αξιολόγησης	40
6.1. Μέτρα αξιολόγησης που προτείνονται στη βιβλιογραφία	40
6.1.1. Average Shot Fidelity	41
6.1.2. Shot Reconstruction Degree	41
6.2. Προτεινόμενα μέτρα αξιολόγησης	43
6.2.1. Χρονική απόσταση από το Ground Truth	43
6.2.2. Οπτική Αξιολόγηση	44
6.2.3. Ομοιότητα Περιεχομένου	44
6.2.4. Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες	45
ΚΕΦΑΛΑΙΟ 7. Πειραματικά αποτελέσματα	47
7.1. Το σύνολο δεδομένων	47
7.2. Το σύνολο Ground Truth	48
7.3. Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε ένα χαρακτηριστικό	53
7.3.1. Χρονική απόσταση από το Ground Truth	53
7.3.2. Οπτική Αξιολόγηση	62
7.3.3. Ορθή κατανομή εικονοπλαισίων στις ομάδες	63
7.4. Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε συνδυασμό χαρακτηριστικών	65
7.5. Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε κατάτμηση βίντεο	70

7.6. Σύγκριση μεθόδων	72
ΚΕΦΑΛΑΙΟ 8. Συμπεράσματα και μελλοντική εργασία	74
ΑΝΑΦΟΡΕΣ	76
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ	80

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας	Σελ
Πίνακας 7.1 Το σύνολο δεδομένων.	48
Πίνακας 7.2 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό, κριτήριο «Χρονική απόσταση – Distance = 10».	54
Πίνακας 7.3 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό, κριτήριο «Χρονική απόσταση – μεταβλητή».	55
Πίνακας 7.4 Επιθυμητές αποστάσεις GT με K-F.	56
Πίνακας 7.5 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό, κριτήριο «Οπτική Αξιολόγηση».	63
Πίνακας 7.6 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό, κριτήριο «Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες».	64
Πίνακας 7.7 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με συνδυασμό χαρακτηριστικών, κριτήριο «Χρονική απόσταση - μεταβλητή».	66
Πίνακας 7.8 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με συνδυασμό χαρακτηριστικών, κριτήριο «Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες».	67
Πίνακας 7.9 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων συνδυασμού λύσεων ομαδοποίησης, κριτήριο «Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες».	70
Πίνακας 7.10 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με κατάτμηση βίντεο, κριτήριο «Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες».	71
Πίνακας 7.11 Συγκριτικός πίνακας όλων των μεθόδων	72
Πίνακας 7.12 Σύγκριση αλγορίθμων που χρησιμοποιούν περιγραφείς SIFT	73

ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα	Σελ
Σχήμα 3.1 Στάδια επιλογής των σημείων ενδιαφέροντος.	21
Σχήμα 3.2 Υπολογισμός ενός descriptor σημείου ενδιαφέροντος.	21
Σχήμα 3.3 Ιστογράμμο αντίθεσης για μια γωνία ρc σε ένα log polar σύστημα συντεταγμένων.	23
Σχήμα 3.4 “Block matching” σε ένα macro block πλευράς 16 pixels και σταθερά αναζήτησης $p=7$.	24
Σχήμα 3.5 Διαδικασία αλγορίθμου “Three Step Search”.	26
Σχήμα 5.1 Συνοπτική περιγραφή της μεθόδου βασισμένη σε ιστογράμματα χρώματος, κίνησης, CCH και SIFT περιγραφέων.	34
Σχήμα 5.2 Δημιουργία πίνακα ομοιότητας της μεθόδου εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε συνδυασμό λύσεων ομαδοποίησης.	35
Σχήμα 5.3 Συνοπτική περιγραφή της μεθόδου συνδυασμού λύσεων ομαδοποίησης.	36
Σχήμα 5.4 Χρονική εξομάλυνση του ιστογράμματος οπτικών λέξεων ενός εικονοπλαισίου, χρησιμοποιώντας ένα Γκαουσιανό πυρήνα εξομάλυνσης.	37
Σχήμα 5.5 Επιλογή χαρακτηριστικών εικονοπλαισίων από την ομαλοποιημένη καμπύλη διαφορών.	38
Σχήμα 7.1 Τα εικονοπλαίσια του Ground Truth της ακολουθίας SPORTS1.	49
Σχήμα 7.2 Τα εικονοπλαίσια του Ground Truth της ακολουθίας SPORTS2.	49
Σχήμα 7.3 Τα εικονοπλαίσια του Ground Truth της ακολουθίας SPORTS3.	49
Σχήμα 7.4 Τα εικονοπλαίσια του Ground Truth της ακολουθίας SPORTS4.	50
Σχήμα 7.5 Τα εικονοπλαίσια του Ground Truth της ακολουθίας FOOTBALL.	50
Σχήμα 7.6 Τα εικονοπλαίσια του Ground Truth της ακολουθίας BASKETBALL1.	50

Σχήμα 7.7 Τα εικονοπλάισια του Ground Truth της ακολουθίας BASKETBALL2.	50
Σχήμα 7.8 Τα εικονοπλάισια του Ground Truth της ακολουθίας TRAFFIC.	51
Σχήμα 7.9 Τα εικονοπλάισια του Ground Truth της ακολουθίας POLITIC.	51
Σχήμα 7.10 Τα εικονοπλάισια του Ground Truth της ακολουθίας CLASSIC1.	51
Σχήμα 7.11 Τα εικονοπλάισια του Ground Truth της ακολουθίας CLASSIC2.	52
Σχήμα 7.12 Τα εικονοπλάισια του Ground Truth της ακολουθίας COMEDY.	52
Σχήμα 7.13 Τα εικονοπλάισια του Ground Truth της ακολουθίας RAIN.	52
Σχήμα 7.14 Τα εικονοπλάισια του Ground Truth της ακολουθίας ZOOM.	53
Σχήμα 7.15 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία SPORTS1.	57
Σχήμα 7.16 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία SPORTS2.	58
Σχήμα 7.17 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία SPORTS4.	59
Σχήμα 7.18 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία RAIN.	60
Σχήμα 7.19 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία ZOOM.	61
Σχήμα 7.20 Αξιολόγηση των μεθόδων εξαγωγής χαρακτηριστικών εικονοπλαισίων με συνδιασμό χαρακτηριστικών προσθέτοντας βάρη, με το κριτήριο S_D .	69

ΠΕΡΙΛΗΨΗ

Ελευθέριος Χατζηιωαννίδης του Γεωργίου και της Ελένης.

MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Φεβρουάριος, 2010.

Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων από ακολουθίες βίντεο.

Επιβλέπων : Αριστείδης Λύκας.

Η μεγάλη αύξηση της διαθεσιμότητας των ψηφιακών βίντεο τα τελευταία χρόνια έχει δημιουργήσει την ανάγκη ανάπτυξης τεχνικών που επιτυγχάνουν αξιόπιστη περίληψη, αναζήτηση και ανάκτηση ακολουθιών βίντεο. Το πρόβλημα της περίληψης ενός βίντεο, δηλαδή μιας σύντομης και περιεκτικής αναπαράστασης της ακολουθίας των εικονοπλαισίων είναι από τα σημαντικότερα ζητήματα ανάλυσης και επεξεργασίας ψηφιακού βίντεο. Η πιο συνηθισμένη προσέγγιση για την αναπαράσταση ενός πλάνου, που αποτελεί και το αντικείμενο μελέτης της παρούσας εργασίας, είναι η αναπαράσταση του περιεχομένου του πλάνου από μια ακολουθία χαρακτηριστικών εικονοπλαισίων (key-frames). Έτσι ένα ολόκληρο βίντεο μπορεί να αναπαρασταθεί από λίγες μόνο εικόνες που βοηθούν ώστε να γίνει κατανοητό το περιεχόμενο χωρίς να είναι απαραίτητη η παρακολούθηση ολόκληρου του βίντεο.

Στην εργασία αυτή καταρχήν συνοψίζονται οι προσεγγίσεις που έχουν προταθεί μέχρι τώρα στο πρόβλημα της εξαγωγής χαρακτηριστικών εικονοπλαισίων από μια ακολουθία βίντεο. Στη συνέχεια μελετάται μια μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε ομαδοποίηση με βάση οπτικά χαρακτηριστικά όπως ιστογράμματα χρώματος, περιγραφείς SIFT και CCH και διανύσματα κίνησης, και προτείνονται δύο επιπλέον προσεγγίσεις: μια μέθοδος που βασίζεται σε συνδυασμό λύσεων ομαδοποίησης και μια μέθοδος που βασίζεται σε κατάτμηση της ακολουθίας βίντεο. Επίσης παρουσιάζονται τα μειονεκτήματα των κριτηρίων αξιολόγησης των

μεθόδων εξαγωγής εικονοπλαισίων που προτείνονται στη βιβλιογραφία και προτείνονται καινούργια κριτήρια αξιολόγησης. Τέλος αποτιμάται πειραματικά η αποτελεσματικότητα των διαφόρων μεθόδων εξαγωγής χαρακτηριστικών εικονοπλαισίων χρησιμοποιώντας τα προτεινόμενα κριτήρια αξιολόγησης.

EXTENDED ABSTRACT IN ENGLISH

Chadjioannidis Eleftherios, G.

MsC, Computer Science Department, University of Ioannina, Greece, February, 2010.

Thesis Title : Key-frame extraction methods for video summarization

Thesis Supervisor : Aristidis Likas.

The rapid increase in the availability of digital videos has stimulated the research interest for the development of new techniques concerning the reliable indexing, summarization and content-based search and retrieval of video sequences. The aim of the video summarization task is to provide a brief and comprehensive summary of video content. The most usual approach is to first segment the video into shots and then create a summary of the shot content using a set of characteristic key-frames. In this way, the whole video can be represented using a small number of representative frames. This is of great help for obtaining a rapid assessment of the video content without needing to watch the whole video. In addition the video summary is considerably useful in performing indexing and retrieval of similar videos.

In this thesis, first various state-of-the art approaches for key-frame extraction are briefly described and then three methods of extracting characteristic key-frames are examined and compared. The first method is based on frame clustering using visual features for frame similarity such as colour histograms, SIFT and CCH descriptions and motion vectors . For frame clustering the spectral clustering algorithm has been employed. The second approach is a new one and is based on a combination of clustering solutions obtained using each visual descriptor separately. The third approach is also a novel one and is based on segmentation of the video sequence.

An interesting issue arising in key-frame extraction concerns the evaluation of the quality of the obtained key-frames. The weaknesses of the currently used evaluation metrics are demonstrated and some novel evaluation criteria are proposed. Finally the methods under consideration are comparatively evaluated in several experiments using video sequences of different characteristics.

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1 Περιγραφή του προβλήματος

1.2 Αντικείμενο της εργασίας

1.3 Δομή της εργασίας

1.1. Περιγραφή του προβλήματος

Η επανάσταση στο ψηφιακό video που έχει ξεκινήσει εδώ και κάποια χρόνια οδηγείται από τη ραγδαία ανάπτυξη σε διάφορους τομείς της επιστήμης της πληροφορικής όπως λόγου χάριν από τη υπολογιστική ισχύ, τη μεγαλύτερη και λιγότερο ακριβή χωρητικότητα των αποθηκευτικών μέσων αλλά και τα ταχύτερα δίκτυα. Αυτή η επανάσταση έχει επιφέρει πολλές καινούργιες εφαρμογές και ως επακόλουθο αυτού, την έρευνα σε πολλές καινούργιες τεχνολογίες που έχουν σαν στόχο την αύξηση της αποτελεσματικότητας αλλά και της αποδοτικότητας τους, όπως για παράδειγμα της αρχειοθέτησης, της δημιουργίας ευρετηρίου και της ανάκτησης βίντεο.

Μια ακολουθία βίντεο συνήθως περιέχει ένα μεγάλο αριθμό από εικονοπλαίσια (frames), ο οποίος πολλές φορές είναι απαγορευτικός για πολλές εφαρμογές και γι' αυτό αποτελεί επιτακτική ανάγκη η ύπαρξη ενός μηχανισμού ο οποίος θα επιτρέπει τον χρήστη να αποκτά γρήγορα άποψη πάνω στο περιεχόμενο ενός βίντεο χωρίς να χρειάζεται να το παρακολουθήσει ολόκληρο. Αυτός ο μηχανισμός ονομάζεται περίληψη του βίντεο ή αλλιώς video abstraction.

Οι τεχνικές περίληψης βίντεο έχουν σχεδιαστεί κυρίως για να διευκολύνουν την αναζήτηση σε μια βάση από βίντεο. Ωστόσο ένα ολοκληρωμένο σύστημα αυτόματης ανάκτησης βίντεο παρουσιάζει ακόμη σημαντικά προβλήματα. Υπάρχουν σημασιολογικές έννοιες οι οποίες είναι πολύ δύσκολο ή και πολλές φορές ακατόρθωτο να εξαχθούν αυτόματα, ενώ σε άλλες περιπτώσεις δεν μπορούν να διαμορφωθούν επακριβείς και σαφείς ερωτήσεις για την αναζήτηση ενός βίντεο. Ο κυρίαρχος περιορισμός που υπάρχει στην περίληψη βίντεο, είναι ότι μπορεί να είναι αποτελεσματικό μόνο εάν ο αριθμός των ακολουθιών βίντεο είναι σχετικά μικρός. Η αναζήτηση σε ένα μεγάλο αριθμό από περιλήψεις βίντεο μπορεί να είναι αρκετά κουραστική και χρονοβόρα για τον χρήστη. Συνδυάζοντας την δημιουργία περιλήψεων βίντεο και την αυτόματη ανάθεση εννοιών στις περιλήψεις, η αναζήτηση της επιθυμητής πληροφορία μπορεί να γίνει πιο αποδοτική. Η μηχανή αναζήτησης βίντεο λειτουργεί ως φίλτρο σε όλα τα βίντεο της βάσης σύμφωνα με την ερώτηση του χρήστη και επιστρέφονται έτσι οι περιλήψεις των βίντεο από τις οποίες ο χρήστης μπορεί να προσδιορίσει τα επιθυμητά βίντεο.

Στην προσπάθεια να διευκολύνουν την αναζήτηση σε μια μεγάλη συλλογή από βίντεο, οι περιλήψεις μπορούν ωστόσο να βοηθήσουν τον χρήστη και στην περιήγηση ενός βίντεο. Αυτό είναι ιδιαίτερος χρήσιμο σε διάφορες εφαρμογές επεξεργασίας βίντεο. Οι περιλήψεις μπορούν επίσης να χρησιμοποιηθούν και από μόνες τους σαν ένα υλικό που μπορεί να μοιραστεί, να κωδικοποιηθεί και να χρησιμοποιηθεί από τον ίδιο τον χρήστη. Διατηρώντας μόνο το απαραίτητο περιεχόμενο μιας ακολουθίας βελτιώνεις την αποθήκευση, το εύρος ζώνης αλλά και το χρόνο παρακολούθησης. Μια τυπική εφαρμογή περίληψης είναι τα αθλητικά τηλεοπτικά highlights. Η περίληψη τέλος, έχει πολλά πλεονεκτήματα και από την υπολογιστική σκοπιά. Μια σύντομη και περιεκτική περίληψη μειώνει σημαντικά το υπολογιστικό κόστος της ανάλυσης και της ανάκτησης του περιεχομένου ενός βίντεο.

Αναγνωρίζοντας λοιπόν τη σπουδαιότητα της περίληψης μέσα στο διευρυμένο πεδίο της διαχείρισης οπτικού περιεχομένου, υπάρχει τα τελευταία χρόνια ένας σημαντικός αριθμός εργασιών οι οποίες επιδιώκουν αυτόματες και αντικειμενικές μεθόδους για τη δημιουργία περιλήψεων. Οι μέχρι τώρα τεχνικές έχουν σαν αντικείμενο έρευνας διάφορα πεδία όπως ταινίες, ντοκιμαντέρ, αθλητικά, νέα, home videos και βλέπουν το

πρόβλημα από διάφορες οπτικές γωνίες κάνοντας κάποιες υποθέσεις σχετικά με το τι αποτελεί τελικά μια καλή περίληψη. Η αλήθεια είναι ότι πολλές από αυτές είχαν αρκετά ικανοποιητικά αποτελέσματα ειδικά στη δημιουργία αθλητικών highlights. Όμως η δημιουργία περιλήψεων σε ένα βίντεο παραμένει κατά μεγάλο ποσοστό, στο στάδιο της έρευνας. Οι πιο πολλές πρακτικές εφαρμογές παραμένουν αρκετά περιορισμένες και ως προς την πολυπλοκότητα τους αλλά και ως προς την έκταση της ανάπτυξης τους. Χαρακτηριστικό παράδειγμα αυτού είναι το ότι εξυπηρετητές αναζήτησης βίντεο όπως λόγου χάριν, το Yahoo και Alta Vista χρησιμοποιούν ένα απλό χαρακτηριστικό εικονοπλαίσιο (key-frame) για να απεικονίσουν το βίντεο ενώ το Google παρέχει μια λίστα από key-frames που περιγράφουν πιο αναλυτικά το περιεχόμενο του εκάστοτε βίντεο.

1.2. Αντικείμενο της εργασίας

Η αρχειοθέτηση, η περίληψη, η αναζήτηση και η αναπαράσταση μιας ακολουθίας βίντεο αποτελούν σημαντικά ζητήματα στη διαχείριση βίντεο. Η ανάλυση της δομής του βίντεο είναι ένα από τα πρώτα βήματα της οργάνωσης του περιεχομένου ενός βίντεο. Τα δεδομένα του βίντεο οργανώνονται σε μια ιεραρχική δομή όπως φαίνεται στα [16], [11]. Αυτή η δομή προκύπτει χωρίζοντας το βίντεο σε τμήματα που περιέχουν σημαντικές πληροφορίες όπως είναι τα πλάνα (shots), οι σκηνές (scenes) και τα κεφάλαια (chapters).

Ένα πλάνο ορίζεται ως μια συνεχής ακολουθία εικονοπλαισίων που έχουν καταγραφεί από μια μόνο κάμερα, μια σκηνή αποτελείται από διαδοχικά πλάνα τα οποία περιγράφουν μια ενέργεια ή ένα γεγονός και ένα κεφάλαιο είναι ένα σύνολο από σκηνές.

Στην εργασία αυτή ασχολούμαστε με τμήματα ενός βίντεο που το περιεχόμενο τους αντιστοιχεί σε ένα πλάνο. Η πιο διαδεδομένη προσέγγιση αναπαράστασης ενός πλάνου είναι τα χαρακτηριστικά εικονοπλαίσια (key-frames). Μελετάμε τρεις μεθόδους εξαγωγής χαρακτηριστικών εικονοπλαισίων και προτείνουμε ορισμένα μέτρα αξιολόγησης, σύμφωνα με τα οποία εξετάζουμε τη συμπεριφορά αυτών των

μεθόδων έχοντας χρησιμοποιήσει διαφορετικά οπτικά χαρακτηριστικά όπως περιγράφεις SIFT, CCH, ιστογράμματα χρώματος και διανύσματα κίνησης.

1.3. Δομή της εργασίας

Αρχικά στο κεφάλαιο 2 της εργασίας περιγράφονται οι βασικές κατηγορίες περίληψης βίντεο που είναι τα χαρακτηριστικά εικονοπλαίσια (key-frames) και τα video skims. Περιγράφονται οι έξι βασικές κατηγορίες των μεθόδων εξαγωγής χαρακτηριστικών εικονοπλαισίων και σε κάθε μια από αυτές περιγράφονται οι σημαντικότερες εργασίες που έχουν προταθεί στη βιβλιογραφία.

Το κεφάλαιο 3 αφορά τους οπτικούς περιγραφείς εικονοπλαισίων. Περιγράφονται αναλυτικά τα ιστογράμματα χρώματος, οι περιγραφείς SIFT και CCH όπως και οι περιγραφείς κίνησης. Τέλος παρουσιάζεται η διαδικασία παραγωγής του συνόλου οπτικών λέξεων (visual words) για ένα πλάνο μιας ακολουθίας βίντεο.

Στο κεφάλαιο 4 περιγράφονται οι αλγόριθμοι ομαδοποίησης εικονοπλαισίων που χρησιμοποιήθηκαν στην εργασία. Γίνεται αναλυτική περιγραφή του φασματικού αλγόριθμου ομαδοποίησης, του γνωστού αλγόριθμου k-means όπως και μια παραλλαγή αυτού που είναι ο fast global k-means.

Το κεφάλαιο 5 αφορά τις μεθοδολογίες εξαγωγής χαρακτηριστικών εικονοπλαισίων που χρησιμοποιήσαμε στην παρούσα εργασία. Οι τρεις μεθοδολογίες με τη σειρά που περιγράφονται είναι οι εξής : α) μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε ιστογράμματα χρώματος και περιγραφείς SIFT, CCH και κίνησης, β) μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε συνδυασμό λύσεων ομαδοποίησης και γ) μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε κατάτμηση της ακολουθίας βίντεο.

Στο κεφάλαιο 6 περιγράφονται τα μέτρα αξιολόγησης της ποιότητας των εξαγόμενων key-frames που προτείνονται στη βιβλιογραφία, όπως είναι το Fidelity measure και το Shot Reconstruction Degree καθώς και τα μέτρα αξιολόγησης που προτείνουμε εμείς.

Τέλος το κεφάλαιο 7 αφορά την συγκριτική πειραματική αποτίμηση των προτεινόμενων μεθόδων. Περιγράφεται αναλυτικά το σύνολο πλάνων που χρησιμοποιήθηκε, παρουσιάζονται τα συγκριτικά πειραματικά αποτελέσματα και γίνεται αξιολόγηση των προτεινόμενων μεθόδων.

ΚΕΦΑΛΑΙΟ 2. ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΣΧΕΤΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

-
- 2.1 Κατηγορίες περίληψης βίντεο
 - 2.2 Το μέγεθος του συνόλου των key-frames
 - 2.3 Μονάδα αναπαράστασης
 - 2.4 Κατηγοριοποίηση των μεθόδων για την εξαγωγή key-frames
 - 2.4.1 Μέθοδος της Επαρκούς αλλαγής περιεχομένου (Sufficient Content Change)
 - 2.4.2 Μέθοδος Μεγίστης Κάλυψης Εικονοπλαισίων (Maximum Frame Coverage)
 - 2.4.3 Μέθοδος Ομαδοποίησης (clustering)
 - 2.4.4 Μέθοδος Ελάχιστης συσχέτισης μεταξύ Key-frames (Minimum Correlation Among Key-frames)
 - 2.4.5 Μέθοδος με χρήση του Σφάλματος ανακατασκευής της ακολουθίας (Shot Reconstruction Degree)
 - 2.4.6 Μέθοδος των ‘Ενδιαφερόντων γεγονότων’ (Interesting Events)
-

Στο κεφάλαιο αυτό αρχικά παρουσιάζονται κάποιες βασικές έννοιες σχετικά με το πρόβλημα της περίληψης μιας ακολουθίας βίντεο. Στη συνέχεια περιγράφεται αναλυτικά η μια από τις δυο κατηγορίες μεθόδων περίληψης βίντεο, (που είναι και το αντικείμενο της εργασίας), που βασίζεται στην εξαγωγή των χαρακτηριστικών εικονοπλαισίων. Διακρίνονται έξι βασικές κατηγορίες μεθόδων εξαγωγής χαρακτηριστικών εικονοπλαισίων και περιγράφονται οι κυριότερες μέθοδοι που έχουν προταθεί μέχρι σήμερα.

2.1. Κατηγορίες περίληψης βίντεο

Υπάρχουν δυο βασικές κατηγορίες περίληψης βίντεο : Key-frames και Video Skims.

Key-frames: Συνήθως λέγονται και χαρακτηριστικά εικονοπλαίσια ή στατικές περιλήψεις. Ένα σύνολο από key-frames είναι ουσιαστικά μια συλλογή από τις πιο βασικές εικόνες που μπορούν να εξαχθούν από μια ακολουθία βίντεο.

Video skims: Συνήθως λέγονται και περιληπτικές ακολουθίες. Αυτός ο τύπος της περίληψης αποτελείται από μια συλλογή από τμήματα του βίντεο που έχουν εξαχθεί από την αρχική ακολουθία. Τα τμήματα αυτά συνδέονται εκατέρωθεν είτε με ένα απότομο κόψιμο είτε από ένα σταδιακό εφέ. Ουσιαστικά μπορεί να χαρακτηριστεί σαν ένα βίντεο κλιπ αλλά με μικρότερη διάρκεια. Μια πολύ δημοφιλής εφαρμογή αυτού του είδους της περίληψης αποτελούν τα trailers των ταινιών.

Ένα πλεονέκτημα που έχουν τα video skims απέναντι σε ένα σύνολο από key-frames είναι η ικανότητα που έχουν τα πρώτα να περιέχουν ήχο αλλά και στοιχεία κίνησης που διευκολύνουν την εκφραστικότητα της περίληψης. Επίσης σαφώς είναι πιο ενδιαφέρον να παρακολουθείς ένα skim από το να βλέπεις τις εικόνες των key-frames. Από την άλλη μεριά όμως η χρησιμοποίηση των key-frames σαφώς αυξάνει τις πιθανότητες πιο ορθής χρήσης της περίληψης για σκοπούς όπως αναζήτηση ή και περιήγηση σε ένα βίντεο, καθώς επιτρέπει το χρήστη να κατανοήσει το περιεχόμενο του βίντεο γρηγορότερα. Τέλος τα key-frames προσφέρουν σημαντική βοήθεια στη μείωση της υπολογιστικής πολυπλοκότητας σε διάφορες εφαρμογές όπως στην ανάλυση και ανάκτηση ενός βίντεο.

2.2. Το μέγεθος του συνόλου των key-frames

Υπάρχουν διάφορες εναλλακτικές προσεγγίσεις για τον υπολογισμό του αριθμού των key-frames. Το μέγεθος του συνόλου των key-frames μπορεί να είναι σταθερό εκ των προτέρων, είναι γνωστό ως *priori*, απροσδιόριστο σύνολο γνωστό και ως *posteriori*

και σύνολο καθορισμένο κατά την διάρκεια της διαδικασίας. Οι περισσότερες από τις τεχνικές χρησιμοποιούν μια από τις τρεις παραπάνω προσεγγίσεις [27].

Σταθερός αριθμός εκ των προτέρων : Ο αριθμός των key-frames καθορίζεται πριν τη διαδικασία και χρησιμοποιείται ως περιορισμός στον αλγόριθμο. Μπορεί να θεωρηθεί ως ένας συγκεκριμένος αριθμός ή ως ποσοστό του μήκους της ακολουθίας.

Απροσδιόριστος αριθμός : Σ' αυτή την περίπτωση, δεν είναι γνωστός ο αριθμός των εξαγόμενων key-frames μέχρι να ολοκληρωθεί η διαδικασία. Αυτός κυρίως καθορίζεται από το βαθμό της οπτικής αλλαγής στο πλάνο. Μια ακολουθία με πλούσια δράση και κίνηση χρειάζεται περισσότερα key-frames για να περιγραφεί από ότι μια άλλη η οποία εμφανίζει στατικό περιεχόμενο.

Καθορισμένο σύνολο : Στην ουσία και αυτή η προσέγγιση είναι μεθοδολογία απροσδιόριστου συνόλου με τη διαφορά όμως ότι πρέπει να καθοριστούν οι μεθοδολογίες προσδιορισμού του κατάλληλου αριθμού από key-frames πριν ολοκληρωθεί η διαδικασία εξαγωγής.

2.3. Μονάδα αναπαράστασης

Ένα αρκετά σημαντικό ζήτημα είναι να προσδιορίσουμε ποια είναι η βασική μονάδα που αντιπροσωπεύει ένα key-frame. Στην βιβλιογραφία τα key-frames μπορούν να εξαχθούν για να αναπαραστήσουν ένα πλάνο σαν ένα ενδιαμέσο στάδιο ή να αναπαραστήσουν ευθέως ένα ολόκληρο clip. Οι πιο απλές τεχνικές που βασίζονται σε πλάνα είναι η επιλογή του πρώτου, του μεσαίου και του τελευταίου εικονοπλαισίου στο πλάνο ως key-frame. Ωστόσο όμως οι περισσότερες τεχνικές επηρεάζονται από την οπτική δυναμική του πλάνου και εξάγονται key-frames για κάθε πλάνο ξεχωριστά ενώ υπάρχουν άλλες τεχνικές που λειτουργούν χωρίς να έχουν κάποια ιδιαίτερη γνώση για τα όρια του πλάνου, οπότε εξάγονται key-frames για όλο το βίντεο.

Στις shot-based τεχνικές η επεξεργασία των πλάνων γίνεται ανεξάρτητα, συνεπώς κάθε πλάνο θεωρείται σαν ένα ξεχωριστό clip. Επιπλέον και εφόσον δεν υποστούν

κάποια επεξεργασία αργότερα οι εν λόγω τεχνικές είναι πολύ πιθανόν να εξάγουν key-frames αρκετά όμοια με κάποια άλλα σε διαφορετικά πλάνα λόγω της ομοιότητας των πλάνων. Τα συστήματα αναζήτησης βίντεο που χρησιμοποιούν ένα ή περισσότερα εικονοπλάισια για την αναπαράσταση ενός πλάνου δεν κλιμακώνονται για μεγάλα βίντεο δεδομένου ότι η αναπαράσταση μπορεί να αποτελείται από μεγάλο αριθμό εικόνων κάτι το οποίο μπορεί να αποβεί χρονοβόρο αλλά και μη αποδοτικό. Απ' την άλλη μεριά όμως οι τεχνικές που βασίζονται σε πλάνα είναι περισσότερο αποδοτικές όταν εφαρμόζονται σε ένα τμήμα του βίντεο που μπορούμε να το επεξεργαστούμε σε λογικό χρόνο.

2.4. Κατηγοριοποίηση των μεθόδων για την εξαγωγή key-frames

Στην συνέχεια περιγράφονται οι κυριότερες μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων που έχουν υλοποιηθεί μέχρι σήμερα. Η βέλτιστη μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων εξαρτάται από το είδος της εφαρμογής που πρόκειται να υλοποιηθεί. Υπάρχουν μέθοδοι οι οποίες είναι γρήγορες και χαμηλών υπολογιστικών απαιτήσεων, και οι οποίες είναι κατάλληλες για real-time και online εφαρμογές, αλλά υπάρχουν και αυτές οι οποίες είναι πιο πολύπλοκες υπολογιστικά, οι οποίες χρησιμοποιούνται συνήθως σε εξειδικευμένες εφαρμογές όπου η ποιότητα των αποτελεσμάτων παίζει πρωτεύοντα ρόλο ακόμη αν και ο χρόνος επεξεργασίας είναι αρκετά μεγάλος.

2.4.1. Μέθοδος επαρκούς αλλαγής περιεχομένου (Sufficient Content Change)

Αυτή η μέθοδος εφαρμόζεται επαναληπτικά και για την εφαρμογή της απαιτείται η γνώση της ακολουθίας των εικονοπλαισίων μέχρι το τρέχον χρονικό σημείο. Ένα εικονοπλαίσιο επιλέγεται ως key-frame εάν το οπτικό περιεχόμενο του είναι αρκετά διαφορετικό από το προηγούμενο που έχει ήδη επιλεγεί. Η διαδικασία συνήθως ξεκινά επιλέγοντας ως πρώτο key-frame το πρώτο εικονοπλαίσιο του πλάνου. Όσο για τη συνάρτηση ομοιότητας του περιεχομένου των εικονοπλαισίων έχουν προταθεί

στην βιβλιογραφία μια ποικιλία από μετρικές με πιο δημοφιλή τη διαφορά ιστογραμμάτων. Οι Zhang *et al* [29] επιλέγουν key-frames με τον παραπάνω τρόπο για κάθε πλάνο, υπολογίζοντας τη διαφορά των ιστογραμμάτων χρώματος ανάμεσα στο τρέχον εικονοπλαίσιο και στο τελευταίο εξαγόμενο. Στο [7] πρώτα υπολογίζουν μια τιμή ανομοιότητας ανάμεσα στο τρέχον εικονοπλαίσιο k και στα N προηγούμενα, συγκρίνοντας το ιστόγραμμα χρώματος του τρέχοντος εικονοπλαισίου και του μέσου όρου των N προηγούμενων. Εάν η τιμή ανομοιότητας υπερβαίνει ένα προϋπολογισμένο κατώφλι τότε το τρέχον εικονοπλαίσιο θεωρείται ως key-frame. Ένα βασικό πρόβλημα αυτής της μεθόδου είναι ότι συνήθως επιλέγεται ως key-frame το πρώτο εικονοπλαίσιο κάθε πλάνου όπως επίσης και εικονοπλαίσια που έχουν σαν περιεχόμενο κάποια ασήμαντη ενέργεια κάτι το οποίο δεν είναι επιθυμητό. Στο [8] συναθροίζουν τις διαφορές των ιστογραμμάτων από διαδοχικά εικονοπλαίσια σε κάθε πλάνο και σε ολόκληρη τη σκηνή. Για κάθε πλάνο επιλέγουν ένα υποσύνολο των δοθέντων N εικονοπλαισίων σύμφωνα με το ποσοστό των διαφορών του κάθε πλάνου σε σχέση με τις διαφορές σε ολόκληρη τη σκηνή. Στην συνέχεια χρησιμοποιείται ένας αλγόριθμος για να κατανέμει τα key-frames σε κάθε πλάνο. Μια άλλη μέθοδος βασισμένη πάλι σε ιστογράμματα χρώματος, κίνησης ή και άλλους περιγραφείς η οποία εκμεταλλεύεται την οπτική δυναμική των εικονοπλαισίων σε μια ακολουθία υπολογίζοντας διαφορές σε ζεύγη εικονοπλαισίων όχι απαραίτητα διαδοχικά, είναι αυτή των Zhao *et al*. [28]. Ανέπτυξαν μια μέθοδο με την ονομασία *Simplified Breakpoints* όπου τα key-frames επιλέγονται αναλύοντας τις τιμές που προκύπτουν. Ένα εικονοπλαίσιο επιλέγεται ως key-frame εάν το ιστόγραμμα χρώματος διαφέρει από αυτό του προηγούμενου εικονοπλαισίου περισσότερο από ένα δοθέν κατώφλι. Όταν το σύνολο των επιλεγθέντων εικονοπλαισίων πλησιάζει τον απαιτούμενο αριθμό από key-frames η διαδικασία τερματίζει. Στο [25] παρουσιάζεται μια μέθοδος στην οποία προσδιορίζονται δυο κατώφλια. Το r το οποίο ελέγχει ποια από τα εικονοπλαίσια πρέπει να εισαχθούν στο σύνολο και το N που ήταν ο μεγαλύτερος ανεκτός αριθμός από key-frames της ακολουθίας. Η εξαγωγή των key-frames βασίζεται σε μια επαναληπτική διαδικασία τμηματικής ομαδοποίησης. Στο πρώτο επαναληπτικό βήμα η ακολουθία διαιρείται σε διαδοχικές ομάδες ιδίου μεγέθους L . Σε κάθε ομάδα υπολογίζεται η διαφορά του πρώτου με το τελευταίο εικονοπλαίσιο. Εάν η διαφορά υπερβαίνει το κατώφλι r όλα τα εικονοπλαίσια αυτής της ομάδας επιλέγονται ως key-frames. Διαφορετικά επιλέγεται μόνο το πρώτο και το

τελευταίο εικονοπλαίσιο αυτής της ομάδας. Εάν ο συνολικός αριθμός των εικονοπλαισίων που έχουν εξαχθεί είναι ίσος ή μικρότερος από το N η διαδικασία εξαγωγής σταματάει. Σε αντίθετη περίπτωση μια νέα ακολουθία δημιουργείται από τα εξαγόμενα εικονοπλαίσια και εφαρμόζεται πάλι η ίδια διαδικασία. Το μεγαλύτερο μειονέκτημα αυτής της διαδικασίας είναι η δυσκολία που έχει ο καθορισμός του κατωφλίου r καθώς παίζει πολύ σημαντικό ρόλο στην ποιότητα που θα έχει το τελικό σύνολο των εξαγόμενων key-frames. Στο [9] προτείνεται μια εναλλακτική μέθοδος. Δεν απαιτείται κάποιο κατώφλι, παρά μόνο ο μέγιστος επιτρεπτός αριθμός των key-frames για ένα δοθέν video. Η μέθοδος αυτή αποτελείται από δυο στάδια. Πρώτα γίνεται η εκτίμηση του αριθμού των key-frames για κάθε πλάνο με βάση τη μεταβλητότητα του περιεχομένου του πλάνου αλλά και ολόκληρης της ακολουθίας. Η μεταβλητότητα του περιεχομένου μιας ακολουθίας καθορίζεται από το άθροισμα όλων των διαφορών frame-to-frame για ολόκληρη την ακολουθία. Η διαδικασία ολοκληρώνεται όταν το άθροισμα των key-frames της ακολουθίας πλησιάζει τον αριθμό N , δηλαδή τον μέγιστο επιτρεπόμενο αριθμό των key-frames. Η καλύτερη κατανομή των key-frames σε κάθε πλάνο επιτυγχάνεται στο δεύτερο βήμα όπου χρησιμοποιείται ένας αλγόριθμος που βελτιστοποιεί την κατανομή της οπτικής ροής μεταξύ των πλάνων. Ωστόσο και αυτή η μέθοδος μπορεί να είναι προβληματική σε ορισμένες περιπτώσεις εξαιτίας της έλλειψης γνώσης του περιεχομένου του βίντεο. Στο [13] πρότειναν μια προσέγγιση στην οποία τα key-frames επιλέγονται από έναν αλγόριθμο προσαρμοσμένης δειγματοληψίας, ο οποίος δειγματοληπτει ομοιόμορφα από τον y άξονα της καμπύλης των αθροιστικών διαφορών των εικονοπλαισίων. Τέλος μια άλλη προσέγγιση προτάθηκε το 2004 από τους Nara et al. [23]. Πρότειναν ένα νευρωνικό δίκτυο χρησιμοποιώντας την ένταση της κίνησης που υπολογίζεται σε ένα MPEG video. Ένα ευφρές σύστημα ταξινομεί την ένταση της κίνησης σε πέντε κατηγορίες και εκείνα τα εικονοπλαίσια που δείχνουν μεγάλη ένταση κίνησης επιλέγονται ως key-frames.

Η μέθοδος της επαρκούς αλλαγής περιεχομένου είναι απλή στην υλοποίησή της, χωρίς μεγάλες απαιτήσεις σε υπολογιστική ισχύ. Αυτά τα χαρακτηριστικά την καθιστούν κατάλληλη για real-time και online εφαρμογές. Η ακολουθιακή επεξεργασία των εικονοπλαισίων βέβαια μπορεί να οδηγήσει σε σύνολα key-frames στα οποία να υπάρχουν ασυμμετρίες αφού για την εξαγωγή ενός εικονοπλαισίου

λαμβάνουμε υπόψη μόνο τα προηγούμενα από αυτό και όχι το σύνολο των εικονοπλαισίων της ακολουθίας βίντεο.

2.4.2. Μέθοδος Μέγιστης Κάλυψης Εικονοπλαισίων (Maximum Frame Coverage)

Η μέθοδος αυτή όπως έχει προταθεί στο [3] βασίζεται στο πλήθος των εικονοπλαισίων που μπορούν να αναπαρασταθούν από ένα τυχαίο εικονοπλαίσιο. Έστω V σύνολο των εικονοπλαισίων του πλάνου, r_i το i -στο key-frame του V , $C_i(\varepsilon)$ το σύνολο των εικονοπλαισίων του πλάνου τα οποία μπορούν να χρησιμοποιήσουν το i -στο εικονοπλαίσιο της ακολουθίας ως αντιπροσωπευτικό με βάση κάποιο κατώφλι ομοιότητας ε και k ο αριθμός των key-frames. Το επιθυμητό σύνολο των key-frames χωρίς περιορισμούς δίνεται από τον παρακάτω τύπο

$$\{r_1, r_2, \dots, r_k\} = \arg \min_{r_i} \{k \mid C_{r_1}(\varepsilon) \cup C_{r_2}(\varepsilon) \cup \dots \cup C_{r_k}(\varepsilon) = V\} \quad (2.1)$$

Όταν ο αριθμός των key-frames δίνεται ως περιορισμός, τότε το ζήτημα αναγάγεται είτε σε πρόβλημα εύρεσης της ελαχίστης τιμής ε , τέτοιας ώστε όλα τα εικονοπλαίσια να μπορούν να καλυφτούν από τουλάχιστον ένα εικονοπλαίσιο

$$\{r_1, r_2, \dots, r_k\} = \arg \min_{r_i} \{\varepsilon \mid C_{r_1}(\varepsilon) \cup C_{r_2}(\varepsilon) \cup \dots \cup C_{r_k}(\varepsilon) = V\} \quad (2.2)$$

είτε ως εύρεση του συνόλου των εικονοπλαισίων που μπορούν να αναπαραστήσουν όσο το δυνατόν περισσότερα.

$$\{r_1, r_2, \dots, r_k\} = \arg \min_{r_i} \{|C_{r_1}(\varepsilon) \cup C_{r_2}(\varepsilon) \cup \dots \cup C_{r_k}(\varepsilon)|\} \quad (2.3)$$

Το πλεονέκτημα αυτής της μεθόδου είναι ότι το πλήθος των εικονοπλαισίων που αντιπροσωπεύονται από ένα εικονοπλαίσιο δεν χρειάζεται όπως στην προηγούμενη μέθοδο να είναι ένα συνεχόμενο τμήμα του βίντεο, με αποτέλεσμα να δημιουργείται ένα περισσότερο συνοπτικό σύνολο key-frames. Από την άλλη πλευρά η απαίτηση

υπολογισμού της διαφοράς για κάθε ζεύγος εικονοπλαισίων οδηγεί σε υψηλές υπολογιστικές απαιτήσεις, με αποτέλεσμα να μην είναι κατάλληλη για real-time και online εφαρμογές.

2.4.3. Μέθοδος Ομαδοποίησης (Clustering)

Οι μέθοδοι που βασίζονται στην δημιουργία ομάδων χρησιμοποιούν τα εικονοπλάισια μιας ακολουθίας βίντεο σαν σημεία στο χώρο των οπτικών χαρακτηριστικών (π.χ. χρώμα, υφή κτλ.). Μετά την σχηματοποίηση των ομάδων γίνεται επιλογή κάποιων χαρακτηριστικών σημείων από αυτά, τα οποία αντιπροσωπεύουν τα χαρακτηριστικά εικονοπλάισια για το συγκεκριμένο τμήμα της ακολουθίας βίντεο. Η διαδικασία της δημιουργίας ομάδων μπορεί να εφαρμοστεί τόσο σε πλάνα όσα και σε ολόκληρα clip και συνήθως προϋποθέτει τα ακόλουθα τέσσερα στάδια.

- *Προεπεξεργασία των δεδομένων.* Πριν την διαδικασία του σχηματισμού ομάδων προηγείται μια προεπεξεργασία των δεδομένων με σκοπό να μειωθεί ο υπολογιστικός φόρτος της διαδικασίας δημιουργίας ομάδων αλλά και να βελτιωθούν τα αποτελέσματα της μεθόδου. Ένα βήμα που μπορεί να προηγηθεί της δημιουργίας ομάδων είναι η μείωση των διαστάσεων (dimensionality reduction) του χώρου των χαρακτηριστικών στον οποίο βρίσκονται τα δεδομένα της ακολουθίας βίντεο. Η μείωση των διαστάσεων μπορεί να γίνει με μια από τις μεθόδους μείωσης των διαστάσεων όπως για παράδειγμα PCA, SVD, Factor Analysis. Εκτός από την μείωση των διαστάσεων, ως βήμα προεπεξεργασίας των δεδομένων μπορεί να θεωρηθεί και η εφαρμογή μιας από τις άλλες υπάρχουσες μεθόδους (π.χ. της μεθόδου της επαρκούς διαφοράς περιεχομένου) ώστε να πάρουμε ένα αρχικό σύνολο χαρακτηριστικών εικονοπλαισίων που θα σχηματίσουν στην συνέχεια ένα σύνολο από το οποίο θα επιλέξουμε τελικά ένα βελτιωμένο σύνολο χαρακτηριστικών εικονοπλαισίων.
- *Ομαδοποίηση των δεδομένων.* Είναι το επόμενο στάδιο καθώς μετά την προεπεξεργασία των δεδομένων μπορούν να εφαρμοστούν διάφορες τεχνικές για τον διαμερισμό των εικονοπλαισίων σε ομάδες.

- *Φιλτράρισμα των ομάδων.* Εξαιτίας του γεγονότος ότι πολλές φορές τα δεδομένα που προκύπτουν από την προηγούμενη διαδικασία μπορούν να περιέχουν αρκετό θόρυβο ορισμένες ομάδες και τα αντίστοιχα key-frames παραλείπονται.
- *Εξαγωγή των αντιπροσωπευτικών εικονοπλαισίων από κάθε ομάδα.* Μετά τον σχηματισμό των ομάδων επιλέγονται κάποια χαρακτηριστικά σημεία από την καθεμία. Η συνηθισμένη τεχνική για την δημιουργία του συνόλου των key-frames είναι η επιλογή εκείνων των εικονοπλαισίων που βρίσκονται πιο κοντά στα κέντρα των ομάδων.

Στη βιβλιογραφία υπάρχει αρκετά μεγάλος αριθμός από εργασίες που βασίζονται στη μέθοδο της ομαδοποίησης. Στο [31] ένα εικονοπλαίσιο αντιστοιχίζεται σε μια υπάρχουσα ομάδα εάν είναι αρκετά όμοιο με το κέντρο αυτής της ομάδας. Εάν αυτή η ομοιότητα είναι μικρότερη από ένα προκαθορισμένο κατώφλι, σχηματίζεται μια νέα ομάδα που περιέχει το εικονοπλαίσιο. Ένα βασικό μειονέκτημα αυτού του αλγορίθμου είναι ο ακριβής προσδιορισμός του αριθμού των key-frames ο οποίος εξαρτάται από ένα κατώφλι που ρυθμίζει την πυκνότητα των ομάδων. Στο [24] παρουσιάζεται μια παραλλαγή αυτού του αλγορίθμου όπου ο τελικός αριθμός των key-frames εξαρτάται από ένα κατώφλι το οποίο προσδιορίζει εάν δυο εικονοπλαίσια είναι όμοια. Στο [10] οι Hanjalic and Zhang βρίσκουν αυτόματα τον αριθμό των ομάδων εφαρμόζοντας cluster- validity ανάλυση. Στη συνέχεια για κάθε ομάδα επιλέγεται ως key-frame το κέντρο της ομάδας το οποίο είναι το εικονοπλαίσιο εκείνο του οποίου η μέση ομοιότητα με τα υπόλοιπα εικονοπλαίσια του συνόλου είναι μέγιστη. Στο [5] προτείνεται μια διαφορετική τεχνική για τον προσδιορισμό key-frames χρησιμοποιώντας περιορισμούς στη θέση των key-frames κατά τη διαδικασία της ομαδοποίησης. Το πρώτο στάδιο είναι απλό. Για να καθοριστεί ένα σύνολο υποψηφίων εικονοπλαισίων που θα χρησιμοποιηθεί ως είσοδος σε έναν ιεραρχικό αλγόριθμο ομαδοποίησης επιλέγονται ως υποψήφια για key-frames εκείνα τα εικονοπλαίσια που είναι αρκετά ανόμοια από άλλα. Στην συνέχεια ακολουθεί η ιεραρχική ομαδοποίηση για την παραγωγή των ζητούμενων key-frames. Για τη σύγκριση των εικονοπλαισίων επιλέγεται μεταξύ άλλων η τεχνική των ιστογραμμάτων. Στη συνέχεια ορισμένες από τις ομάδες φιλτράρονται χρησιμοποιώντας κάποια κριτήρια και αφού επιλέγονται τελικά τόσες ομάδες όσα τα

ζητούμενα key-frames, ακολουθεί το τελικό στάδιο που είναι η επιλογή των key-frames από κάθε ομάδα. Συνήθως οι ομάδες περιέχουν εικονοπλαίσια που είναι διασκορπισμένα σε όλη τη διάρκεια του video. Για το λόγο αυτό θεωρείται ότι για να έχουν μια καλή περιγραφή του video, είναι επιθυμητό να επιλέγεται έναν αντιπροσωπευτικό εικονοπλαίσιο από ένα χρονικό σημείο στο οποίο εκείνο η ομάδα υπερισχύει. Γι' αυτό το σκοπό καθορίζεται η πιο μεγάλη ακολουθία από εικονοπλαίσια για κάθε επιλεγμένη ομάδα, η οποία δεν διακόπτεται από εικονοπλαίσια που ανήκουν σε άλλη ομάδα. Τέλος επιλέγεται το εικονοπλαίσιο που είναι πιο κοντά στο κέντρο αυτής της ακολουθίας. Μια γρήγορη κριτική για αυτή την προσέγγιση είναι ότι έχει το πλεονέκτημα να μεγιστοποιεί την απόσταση στο χρόνο ανάμεσα στα αντιπροσωπευτικά εικονοπλαίσια, με αποτέλεσμα τα key-frames που επιλέγονται να μην είναι τόσο όμοια μεταξύ τους. Στο [6] χρησιμοποιούν έναν αλγόριθμο ομαδοποίησης ο οποίος είναι βασισμένος σε Single Value Decomposition (SVD). Τα εικονοπλαίσια δειγματοληπτούνται χρονικά και από αυτά κατόπιν υπολογίζονται τα οπτικά χαρακτηριστικά. Ο χώρος των χαρακτηριστικών που εξάγεται από το SVD στην συνέχεια διαμερίζεται σε ομάδες και εξάγεται ένα key-frame από κάθε μια από τις ομάδες. Στο [2] χρησιμοποιώντας μια split-merge προσέγγιση ομαδοποιούνται διαδοχικά εικονοπλαίσια με βάση την αξία πληροφορίας (information value) κάθε εικονοπλαίσιου. Ως key-frames επιλέγονται εκείνα τα εικονοπλαίσια τα οποία μεγιστοποιούν την αμοιβαία πληροφορία για κάθε ομάδα.

Είναι φανερό πως η ομαδοποίηση είναι η πιο δημοφιλής μέθοδος για την εξαγωγή key-frames από ένα βίντεο. Ωστόσο όμως πρέπει να επισημανθεί πως ορισμένες φορές η κατάλληλη εξαγωγή key-frames από ορισμένες ομάδες δεν είναι αποδοτική κυρίως όταν αυτές οι εμφανίζουν μεγάλη μεταβλητότητα. Επιπλέον ένα ακόμη ζήτημα στο οποίο υστερούν οι εν λόγω τεχνικές είναι όταν θέλουμε τα εξαγόμενα key-frames να διατηρούν την χρονική εξέλιξη της ακολουθίας.

2.4.4. Μέθοδος Ελάχιστης συσχέτισης μεταξύ Key-frames (Minimum Correlation among Key-frames)

Αυτή η κατηγορία μεθόδων βασίζεται στο γεγονός ότι σε ένα σύνολο από key-frames θα πρέπει να υπάρχει ελάχιστη συσχέτιση μεταξύ των στοιχείων του. Με βάση αυτό το κριτήριο οι μέθοδοι επιλέγουν εκείνα τα εικονοπλάγια που είναι ανόμοια σε σχέση με τα υπόλοιπα. Η βέλτιστη εξαγωγή key-frames με βάση το κριτήριο της ελάχιστης συσχέτισης διατυπώνεται ως εξής :

$$\{r_1, r_2, \dots, r_k\} = \arg \min_{r_i} \{Corr(f_{r_1}, f_{r_2}, \dots, f_{r_k})\} \quad (2.4)$$

όπου το $Corr$ είναι η τιμή της συσχέτισης (correlation measure).

2.4.5. Μέθοδος με χρήση του Σφάλματος ανακατασκευής της ακολουθίας (Shot Reconstruction Degree)

Αυτή η προσέγγιση παρουσιάζεται στο [18] και βασίζεται σε ένα μέτρο που ονομάζεται SRD score (είναι γνωστό και ως shot reconstruction error). Αυτό μετρά την ικανότητα του συνόλου των key-frames για την ανακατασκευή της αρχικής ακολουθίας του βίντεο. Έστω ότι έχουμε μια συνάρτηση παρεμβολής μεταξύ των εικονοπλαισίων $I(t, R)$, όπου R το σύνολο των key-frames, V το σύνολο των εικονοπλαισίων και n το πλήθος των εικονοπλαισίων, η οποία υπολογίζει όλα ή μερικά χαρακτηριστικά της εικόνας την χρονική στιγμή t στην ακολουθία του βίντεο στο σύνολο των key-frames. Το SRD score $E(V, R)$ του συνόλου δίνεται από τον τύπο :

$$E(V, R) = \sum_{i=1}^n D(f_i, I(i, R)) \quad (2.5)$$

όπου D είναι η διαφορά ανάμεσα σε δυο εικονοπλαίσια. Έτσι δοθέντος του αριθμού των key-frames k , ως καλύτερο σύνολο key-frames $\{r_1, r_2, \dots, r_k\}$ επιλέγεται εκείνο το σύνολο το οποίο έχει το ελάχιστο SRD score και προσδιορίζεται από τον παρακάτω τύπο :

$$\{r_1, r_2, \dots, r_k\} = \arg \min_{r_i} \{E(V, R), 1 \leq r_i \leq n\} r_k \leq n \quad (2.6)$$

2.4.6. Μέθοδος των 'Ενδιαφερόντων γεγονότων' (Interesting Events).

Σε αντίθεση με τις προηγούμενες προσεγγίσεις, οι οποίες εστίαζαν κυρίως στο να μεγιστοποιήσουν το μήκος και την ισορροπία της οπτικής κάλυψης του συνόλου των key-frames, οι μέθοδοι οι οποίες βασίζονται σε ενδιαφέροντα γεγονότα προσπαθούν να αναγνωρίσουν τα εικονοπλαίσια τα οποία είναι σημασιολογικά αρκετά σημαντικά. Οι περισσότερες μέθοδοι σε αυτήν την κατηγορία θεωρούν μια σχέση ανάμεσα στο πόσο ενδιαφέρον είναι ένα εικονοπλαίσιο και στα πρότυπα κίνησης γύρω από τα εικονοπλαίσια, καθώς και στα χαρακτηριστικά του περιεχομένου του (π.χ. να περιέχει ανθρώπινα πρόσωπα).

Οι Liu, Zhang, και Qi [26] πρώτα κατασκευάζουν ένα αντιπροσωπευτικό μοντέλο της δραστηριότητας της κίνησης. Με βάση αυτό το μοντέλο, ένα πλάνο χωρίζεται σε επιμέρους τμήματα από διαδοχικά πρότυπα κίνησης ως προς τις επιταχύνσεις και τις επιβραδύνσεις. Τα key-frames τα οποία εξάγονται από αυτά τα τμήματα βασίζονται σε ένα τριγωνικό μοντέλο. Η κάτω αριστερή κορυφή αντιπροσωπεύει το αρχικό σημείο της κίνησης, η δεξιά το τελικό και η πάνω κορυφή του τριγώνου αντιπροσωπεύει το σημείο με τη μέγιστη ταχύτητα. Έτσι τελικά επιλέγονται ως key-frame εκείνα τα εικονοπλαίσια που βρίσκονται στις ανώτατες κορυφές του τριγώνου οι οποίες αντιπροσωπεύουν σημεία στα οποία η κίνηση αλλάζει από επιτάχυνση σε επιβράδυνση. Ο λόγος που επιλέγονται αυτά τα σημεία είναι γιατί συνήθως αντιπροσωπεύουν τα πιο εξέχοντα σημεία της δράσης. Σε μια άλλη μέθοδο [30], μόνο ένα key-frames εξάγεται από κάθε πλάνο. Τα εικονοπλαίσια κατατάσσονται σε προσκήνιο και παρασκήνιο και στη συνέχεια εκείνο με το μεγαλύτερο λόγο

προσκήνιο προς παρασκήνιο επιλέγεται ως key-frame καθώς θεωρείται ότι μεταφέρει την περισσότερη πληροφορία του συγκεκριμένου πλάνου. Στο [4] πρότειναν μια άλλη προσέγγιση για την εξαγωγή key-frames, καθώς αναλύοντας τις διαφορές ανάμεσα σε διαδοχικά εικονοπλάισια, ο αλγόριθμος καθορίζει την πολυπλοκότητα της ακολουθίας σε σχέση με τις αλλαγές στο περιεχόμενο, οι οποίες εκφράζονται με διαφορετικούς περιγραφείς εικονοπλαισίων. Αρχικά χωρίζουν την ακολουθία σε πλάνα τα οποία ταξινομούνται σε δυο κατηγορίες από τις οποίες μόνο μια επιλέγεται για να δώσει τα key-frames καθώς δεν περιέχει περιττή πληροφορία όπως για παράδειγμα ένα fade out effect. Στην συνέχεια και για να αντιμετωπίσουν καλύτερα το ζήτημα της αναπαράστασης ενός εικονοπλαισίου υπολογίζουν τρεις διαφορετικούς περιγραφείς : ιστόγραμμα χρώματος, ιστόγραμμα κατεύθυνσης ακμής και κατανομή συντελεστών wavelet. Κατόπιν και αφού ορίσουν ένα διαφορετικό μέτρο ανομοιότητας για κάθε περιγραφέα, υπολογίζουν ως συνδυασμό αυτών των τριών το τελικό μέτρο ανομοιότητας. Έτσι προκύπτει μια αθροιστική καμπύλη διαφορών, η οποία περιγράφει πως αλλάζει το περιεχόμενο των εικονοπλαισίων στην διάρκεια του χρόνου. Τέλος για να εντοπίσουν τα key-frames επιλέγουν τις περιοχές εκείνες στην καμπύλη που προσδιορίζονται από σημεία που εμφανίζουν την μεγαλύτερη κλίση. Το ενδιάμεσο εικονοπλάισιο κάθε περιοχής επιλέγεται ως το key-frame.

Αν και οι μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων οι οποίες βασίζονται στην εύρεση 'ενδιαφερόντων' γεγονότων σε ένα σύνολο εικονοπλαισίων είναι επιτυχημένες σε συγκεκριμένες εφαρμογές ειδικού ενδιαφέροντος, αποτυγχάνουν σε εφαρμογές γενικού περιεχομένου κυρίως επειδή οι μέθοδοι αυτοί βασίζονται σε κανόνες οι οποίοι εξάγονται εμπειρικά από συγκεκριμένα μικρά σύνολα δεδομένων.

ΚΕΦΑΛΑΙΟ 3. ΠΕΡΙΓΡΑΦΕΙΣ ΕΙΚΟΝΟΠΛΑΙΣΙΩΝ

- 3.1 Ιστογράμματα Χρώματος
 - 3.2 Περιγραφείς SIFT (SIFT Descriptors)
 - 3.3 Περιγραφείς CCH (CCH Descriptors)
 - 3.4 Περιγραφείς κίνησης (Motion Descriptors)
 - 3.5 Σύνολα από οπτικές λέξεις (Bag of Visual Words)
-

Στο κεφάλαιο αυτό παρουσιάζονται αναλυτικά οι περιγραφείς των εικονοπλαισίων που χρησιμοποιούμε στους αλγορίθμους της εργασίας. Κατόπιν περιγράφεται η διαδικασία κατά την οποία συνδυάζονται τα σύνολα των περιγραφέων ενός εικονοπλαίσιου και αποκτάται τελικά μια πιο συμπαγής αναπαράσταση.

3.1. Ιστογράμματα Χρώματος

Τα ιστογράμματα χρώματος είναι το πιο διαδεδομένα χαρακτηριστικά στο πρόβλημα της περίληψης ενός βίντεο. Είναι αρκετά εύρωστα στην κίνηση των αντικειμένων και της κάμερας και παρέχουν έναν καλό συνδυασμό ανάμεσα στην ακρίβεια της ανίχνευσης και στην ταχύτητα της διαδικασίας. Εμείς έχουμε επιλέξει να χρησιμοποιήσουμε τα κανονικοποιημένα ιστογράμματα στον HSV χώρο χρώματος. Ως εκ τούτου υπολογίζεται για κάθε εικονοπλαίσιο ένα κανονικοποιημένο ιστόγραμμα με 8 κάδους για την απόχρωση H_H και από 4 κάδους για κάθε ένα από

τα H_s , H_v . Τα τρία αυτά ιστογράμματα συνενώνονται σε ένα διάνυσμα διάστασης $4+4+8$ στην περίπτωση της μιας διάστασης (1-D) και σε ένα διάνυσμα διάστασης 8×4 στην δισδιάστατη περίπτωση (2-D), το οποίο τελικά παριστάνει το τελικό ιστόγραμμα για κάθε εικονοπλαίσιο :

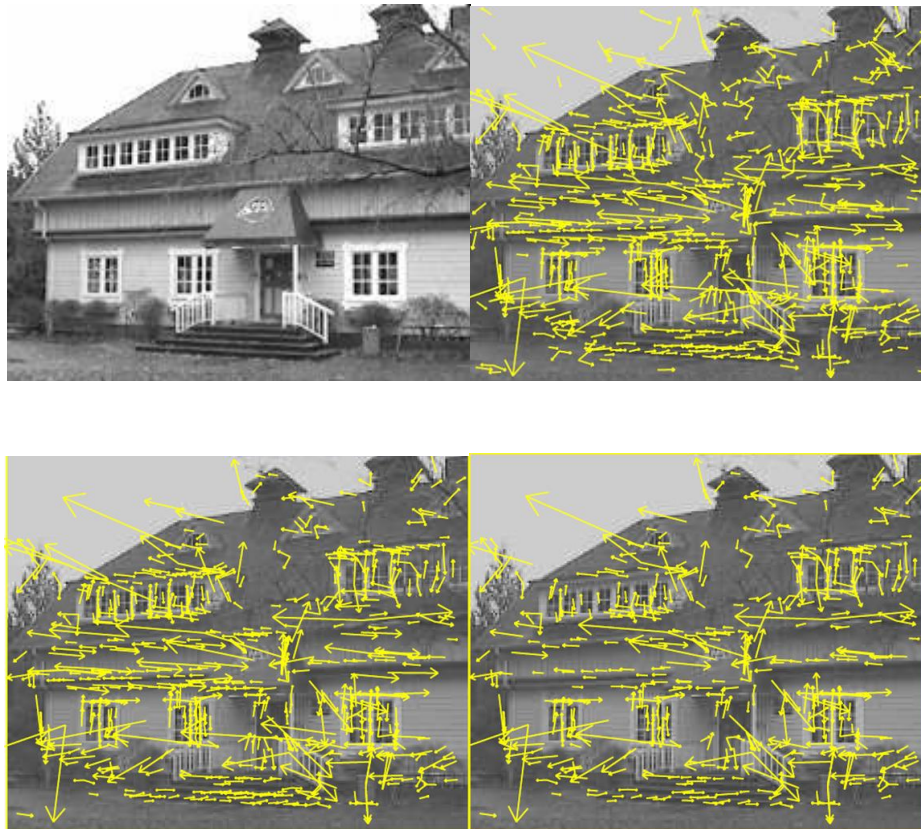
$$H = [H_H H_S H_V] \text{ (1-D) } \text{ ή } H = [H_{H_1} H_{S_1} H_{H_2} H_{S_2} H_{H_3} H_{S_3} \dots H_{H_8} H_{S_4}] \text{ (2-D)} \quad (3.1)$$

3.2 Περιγραφείς SIFT (SIFT Descriptors)

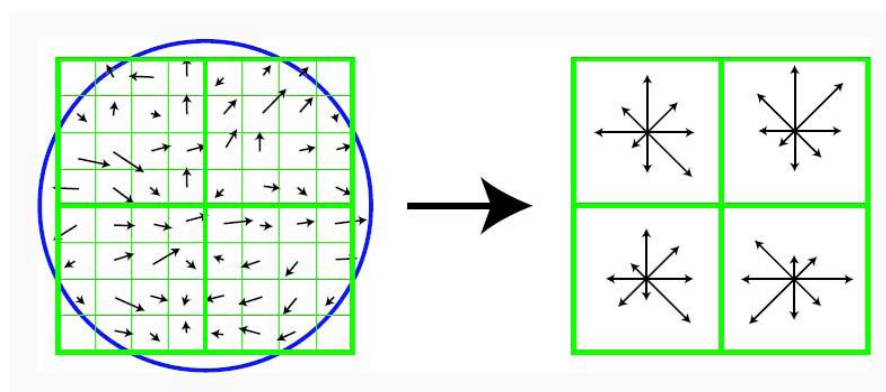
Η μέθοδος εξαγωγής των περιγραφών SIFT, η οποία περιγράφεται στο [20] αποτελείται από τέσσερα κυρίαρχα στάδια. Στο πρώτο στάδιο, η εικόνα σαρώνεται σε όλο της το φάσμα για να εντοπιστούν χαρακτηριστικά σημεία ή σημεία ενδιαφέροντος. Κατασκευάζεται μια γκαουσιανή πυραμίδα και εντοπίζονται οι τοπικές κορυφές σε μια σειρά από διαφορές από Γκαουσιανές εικόνες. Στο δεύτερο στάδιο, εξαλείφονται τα ασταθή σημεία ενδιαφέροντος. Στο τρίτο στάδιο, για κάθε σημείο ενδιαφέροντος, αναγνωρίζονται οι επικρατέστερες μετατοπίσεις που συμβαίνουν σε κάθε τοπική περιοχή της εικόνας.

Τέλος, στο τέταρτο στάδιο δημιουργείται ένας τοπικός περιγραφέας εικόνας για κάθε σημείο ενδιαφέροντος βασισμένος στο τοπικό διάνυσμα κλίσης. Στο σχήμα (3.1) παρουσιάζουμε τα στάδια επιλογής των σημείων ενδιαφέροντος.

Ο τυπικός περιγραφέας SIFT ενός σημείου ενδιαφέροντος δημιουργείται μέσω της κλιμάκωσης και της μετατόπισης της εικόνας στην περιοχή γύρω από το σημείο ενδιαφέροντος και της δημιουργίας κατάλληλων ιστογραμμάτων κλίσης (σχ. 3.2). Πιο συγκεκριμένα δημιουργείται ένα 4×4 διάνυσμα από ιστογράμματα καθένα από τα οποία αποτελείται από 8 κάδους μετατόπισης. Αυτό το διάνυσμα των 128 στοιχείων ($4 \times 4 \times 8$) κατόπιν κανονικοποιείται και εφαρμόζεται ένα κατώφλι για να απομακρυνθούν τα στοιχεία με πολύ μικρές τιμές. Οι SIFT περιγραφείς αποτελούν μια πολύ δημοφιλή προσέγγιση για αρκετά προβλήματα υπολογιστικής όρασης.



Σχήμα 3.1. Σταδία επιλογής των σημείων ενδιαφέροντος. Α) Η αρχική εικόνα Β) Οι αρχικές 832 θέσεις των σημείων ενδιαφέροντος στα ελάχιστα και μέγιστα των διαφορών της γκαουσιανής συνάρτησης. Τα σημεία ενδιαφέροντος απεικονίζονται ως διανύσματα που δηλώνουν την κλίμακα, τον προσανατολισμό και τη θέση. Γ) Τα 729 σημεία ενδιαφέροντος που απομένουν εφαρμόζοντας ένα κατώφλι ελάχιστης φωτεινότητας. Δ) Τα τελικά 536 σημεία ενδιαφέροντος που απομένουν αφού εφαρμοστεί και ένα επιπλέον κατώφλι στον λόγο της κύριας καμπυλότητας.



Σχήμα 3.2 Υπολογισμός ενός descriptor σημείου ενδιαφέροντος (Η εικόνα έχει ληφθεί από το [20]).

3.3 Περιγραφείς CCH (CCH Descriptors)

Ένας αντίστοιχος τοπικός περιγραφέας εικόνων, καλούμενος ως contrast context histogram (CCH) προτείνεται στο [14] και εφαρμόζεται επιτυχώς στο πρόβλημα ανίχνευσης πλάνων [15]. Απεικονίζει την κατανομή της φωτεινότητας σε μια τοπική περιοχή γύρω από ένα σημείο ενδιαφέροντος και χρησιμεύει ως ένας τοπικός περιγραφέας για αυτήν την περιοχή. Δοθείσης μιας εικόνας I , πρώτα εφαρμόζεται σε αυτή ένας Γκαουσιανός πυρήνας για να ομαλοποιηθεί και κατόπιν κατασκευάζεται μια πολυκλιμακωτή Λαπλασιανή πυραμίδα και εξάγονται τα εξέχοντα σημεία ανιχνεύοντας Harris corners [12]. Για κάθε εξέχον σημείο p_c ορίζεται μια περιοχή R διάστασης $n \times n$. Η αντίθεση ενός σημείου p σε αυτή την περιοχή δίνεται από την παρακάτω σχέση :

$$C(p) = I(p) - I(p_c) \quad (3.2)$$

όπου $I(p)$ η φωτεινότητα του σημείου p .

Η περιοχή R προσδιορίζεται σε ένα κβαντισμένο log-polar σύστημα συντεταγμένων (r, θ) όπου $r_i = 0, \dots, r$, $r = \left\lfloor \log(\sqrt{2n^2}) \right\rfloor$ και $\theta_j = \frac{2\pi}{l} m, m = 0, \dots, l-1$. Οι παράμετροι r, l καθορίζουν την απόσταση και τη μετατόπιση αντίστοιχα. Για κάθε υποπεριοχή $R_{ij} = (r_i \theta_j)$ υπολογίζονται ένας θετικός και ένας αρνητικός κάδος ιστογράμματος με βάση τις τιμές της αντίθεσης. Ειδικότερα, δοθέντος ενός εξέχοντος σημείου p_c και μιας υποπεριοχής R_{ij} οι θετικοί και αρνητικοί κάδοι ιστογράμματος υπολογίζονται από τις παρακάτω σχέσεις αντίστοιχα

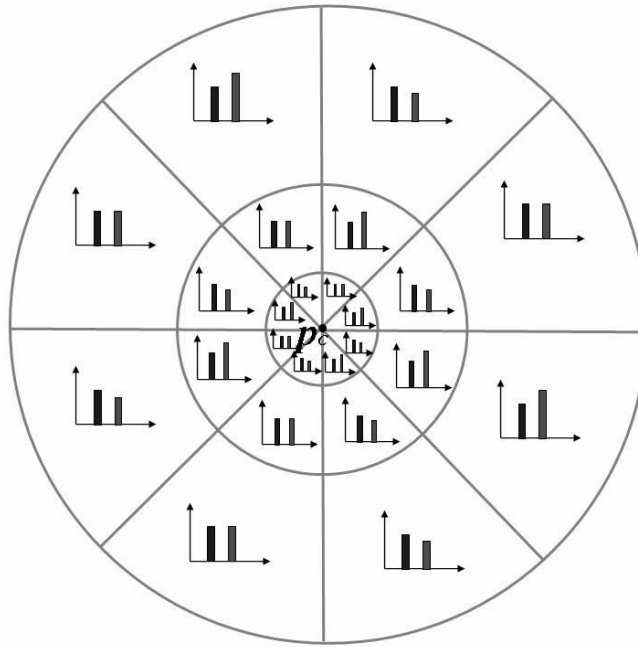
$$H_{R_{ij}}^+(p_c) = \frac{\sum \{C(p) \mid p \in R_{ij} \text{ και } C(p) \geq 0\}}{\# R_{ij}^+} \quad (3.3)$$

$$H_{R_{ij}}^-(p_c) = \frac{\sum \{C(p) \mid p \in R_{ij} \text{ και } C(p) < 0\}}{\# R_{ij}^-} \quad (3.4)$$

όπου οι $\#R_{ij}^+$, $\#R_{ij}^-$ καθορίζουν τις θετικές και αρνητικές τιμές αντίθεσης στο R_{ij} . Ο CCH περιγραφέας p_c προκύπτει από τη συνένωση όλων των ιστογραμμάτων αντίθεσης σε όλες τις υποπεριοχές σε ένα διάνυσμα και καθορίζεται ως εξής :

$$CCHp_c = (H_{R_{00}}^+, H_{R_{00}}^-, \dots, H_{R_{rl}}^+, H_{R_{rl}}^-) \quad (3.5)$$

Στην δική μας προσέγγιση χρησιμοποιήσαμε $r=3$ και $l=8$ όπως προτείνεται στο [14] δημιουργώντας έτσι $2 \times 4 \times 8 = 64$ διαστάσεις για κάθε CCH περιγραφέα. Στο σχήμα (3.3) παρουσιάζουμε το ιστόγραμμα αντίθεσης για μια γωνία p_c σε ένα log-polar σύστημα συντεταγμένων.

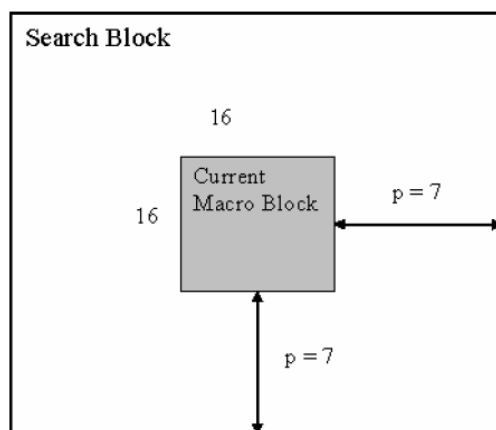


Σχήμα 3.3 Ιστόγραμμα αντίθεσης για μια γωνία p_c σε ένα log-polar σύστημα συντεταγμένων.

3.4. Περιγραφείς Κίνησης (Motion Descriptors)

Μια άλλη κατηγορία χαρακτηριστικών είναι τα διανύσματα κίνησης (motion vectors) που χρησιμοποιούνται σε block matching αλγόριθμους για την εκτίμηση κίνησης στην συμπίεση βίντεο.

Η βασική ιδέα του block matching όπως περιγράφεται στο [1], είναι ο διαχωρισμός του τρέχοντος εικονοπλαισίου σε ένα πίνακα από ‘macro blocks’, τα οποία στη συνέχεια συγκρίνονται με τα αντίστοιχα ή γειτονικά block του προηγούμενου εικονοπλαισίου. Αυτό έχει σαν αποτέλεσμα τη δημιουργία ενός διανύσματος που ορίζει την κίνηση σε ένα macro block από μια θέση του τρέχοντος εικονοπλαισίου σε μια άλλη στο προηγούμενο εικονοπλαίσιο. Η κίνηση στο τρέχον εικονοπλαίσιο εκφράζεται από την κίνηση που υπολογίζεται από όλα τα macro blocks που το απαρτίζουν. Η περιοχή αναζήτησης του macro block match περιορίζεται σε p pixels γύρω από το υπό εξέταση macro block στο προηγούμενο εικονοπλαίσιο. Το p καλείται σταθερά αναζήτησης. Όσο πιο μεγάλη είναι η κίνηση τόσο μεγαλύτερο πρέπει να είναι το p . Από την άλλη μεριά όμως όσο μεγαλύτερο είναι το p τόσο πιο ακριβή υπολογιστικά γίνεται η διαδικασία εκτίμησης της κίνησης. Συνήθως το macro block ορίζεται ως ένα τετράγωνο στο οποίο η κάθε μια πλευρά του αποτελείται από 16 pixels και η σταθερά αναζήτησης $p=7$ pixels. Αυτές είναι οι τιμές που χρησιμοποιήσαμε και εμείς στα πειράματά μας. Η ιδέα αναπαρίσταται στο σχήμα (3.4).



Σχήμα 3.4 “Block matching” σε ένα macro block πλευράς 16 pixels και σταθερά αναζήτησης $p=7$.

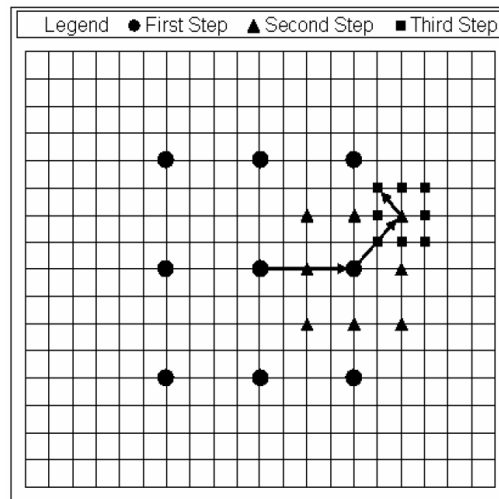
Η ομοιότητα ενός macro block με ένα άλλο προκύπτει από μια συνάρτηση κόστους. Εκείνο το macro block που εμφανίζει το ελάχιστο κόστος θεωρείται ότι ταιριάζει περισσότερο με το τρέχον block. Στα πειράματά μας χρησιμοποιήσαμε την πιο δημοφιλή συνάρτηση κόστους που καλείται ως MAD (Mean Absolute Difference) και δίνεται ως :

$$MAD = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |C_{ij} - R_{ij}| \quad (3.6)$$

Όπου N είναι η πλευρά του macro block, C_{ij} και R_{ij} είναι τα pixels του τρέχοντος και του προηγούμενου macro block αντίστοιχα.

Στη βιβλιογραφία υπάρχουν αρκετοί αλγόριθμοι για την εκτίμηση της κίνησης άλλοτε πιο ακριβοί υπολογιστικά αλλά με μεγαλύτερη ακρίβεια και άλλοτε πιο απλοί αλλά με όχι καλά αποτελέσματα. Στα πειράματά μας χρησιμοποιήσαμε τον αλγόριθμο TSS (Three Step Search) [1], ο οποίος αποτελεί μια από τις πιο απλές προσεγγίσεις και χρονολογείται στα μέσα της δεκαετίας του '80. Η βασική ιδέα παρουσιάζεται στο σχήμα (3.5). Ο αλγόριθμος ξεκινάει την αναζήτηση από το κέντρο και στην περίπτωση όπου η σταθερά αναζήτησης έχει την τιμή 7 αναθέτει στο μήκος βήματος S την τιμή 4. Κατόπιν αναζητεί σε οκτώ θέσεις γύρω από το (0,0) με απόσταση $\pm S$ pixels. Από αυτές τις εννέα θέσεις επιλέγει εκείνη με το μικρότερο κόστος θέτοντας την ως την αρχική θέση για τη νέα αναζήτηση. Κατόπιν θέτει στο μήκος βήματος S την τιμή $S=S/2$ και επαναλαμβάνει την ίδια αναζήτηση για δυο επαναλήψεις μέχρι το S να πάρει την τιμή 1. Στο σημείο στο οποίο βρεθεί η τοποθεσία με την ελάχιστη τιμή στη συνάρτηση κόστους, το macro block της τοποθεσίας αποτελεί το βέλτιστο ταίριασμα.

Αφού υπολογιστούν τα διανύσματα κίνησης για κάθε block (αριθμός block=300), κατόπιν υπολογίζονται οι γωνίες μετατόπισης των block και τοποθετούνται σε ένα ιστόγραμμα 21 κάδων (20 κάδους από 18° ο καθένας και ένας κάδος για τις γωνίες 0°), όπου στη συνέχεια αφού κανονικοποιηθεί δημιουργείται τελικά το ιστόγραμμα κίνησης.



Σχήμα 3.5 Διαδικασία αλγορίθμου “Three Step Search”. Το διάνυσμα κίνησης είναι το (5,-3).

3.5. Σύνολα από οπτικές λέξεις (Bag of Visual Words)

Για κάθε εικονοπλαίσιο υπολογίζεται ένας διαφορετικός αριθμός από περιγραφείς οι οποίοι περιγράφουν τα κύρια αντικείμενα ή σημεία ενδιαφέροντος σε αυτό. Πιο συγκεκριμένα δοθέντος ενός πλάνου s και του αντίστοιχου συνόλου από τα n εικονοπλαίσια $F = \{f_1, \dots, f_n\}$, για κάθε ένα $\forall f_i, i = 1, \dots, n$ εξάγεται ένα σύνολο από περιγραφείς D_{f_i} , SIFT ή CCH χρησιμοποιώντας τους αλγόριθμους που παρουσιάζονται στο [20], [14] αντίστοιχα.

Για να επιταχυνθεί η διαδικασία, το σύνολο D_{f_i} των περιγραφέων κάθε εικονοπλαισίου f_i , συνοψίζεται για να αποκτηθεί μια πιο συμπαγής αναπαράσταση. Πιο συγκεκριμένα το σύνολο D_{f_i} ομαδοποιείται σε M ομάδες (στα πειράματά μας επιλέξαμε $M=30$) χρησιμοποιώντας τον αλγόριθμο *fast global k-means* [17]. Έτσι τελικά κάθε εικονοπλαίσιο αντιπροσωπεύεται από τα κέντρα των M ομάδων $D_{f_i}^M$. Για να εξαχθούν τώρα οι οπτικές λέξεις από τους περιγραφείς, το σύνολο των αντιπροσωπευτικών κέντρων των περιγραφέων για όλα τα n εικονοπλαίσια του

πλάνου $D_s = D_{f_1}^M \cup D_{f_2}^M \cup \dots \cup D_{f_n}^M$ ομαδοποιείται σε k ομάδες $\{C_1, C_2, \dots, C_k\}$, χρησιμοποιώντας τον k-means αλγόριθμο όπου το k δηλώνει το μέγεθος του λεξικού των οπτικών λέξεων (στα πειράματά μας χρησιμοποιήσαμε $k=20, k=50, k=100$). Για να κατασκευαστεί το ιστόγραμμα των οπτικών λέξεων (bag of visual words) για κάθε εικονοπλαίσιο f_i , κάθε ένας από τους περιγραφείς D_{f_i} , αντιστοιχίζεται σε μια από τις k οπτικές λέξεις (ομάδες) έχοντας έτσι σαν αποτέλεσμα να δημιουργηθεί ένα διάνυσμα που περιέχει τη συχνότητα της κάθε οπτικής λέξης μέσα στο εικονοπλαίσιο. Έτσι δοθέντος ενός εικονοπλαισίου f_i και του συνόλου των περιγραφέων του $D = \{d_{t_1}, \dots, d_{t_D}\}$, το ιστόγραμμα οπτικών λέξεων VH_i δίνεται από την παρακάτω σχέση :

$$VH_i(l) = \frac{\#\{d_{t_j} \in C_l, j = 1, \dots, D\}}{|D|}, l = 1, \dots, k \quad (3.7)$$

ΚΕΦΑΛΑΙΟ 4. ΟΜΑΔΟΠΟΙΗΣΗ ΕΙΚΟΝΟΠΛΑΙΣΙΩΝ

4.1 Ο τυπικός φασματικός αλγόριθμος ομαδοποίησης (Spectral Clustering Algorithm)

4.2 Ο αλγόριθμος k-means

4.3 Ο Αλγόριθμος Fast Global k-means

Στο κεφάλαιο αυτό παρουσιάζονται οι αλγόριθμοι που χρησιμοποιήθηκαν από τη βιβλιογραφία για την εξαγωγή των key-frames.

4.1. Ο τυπικός φασματικός αλγόριθμος ομαδοποίησης (Spectral Clustering Algorithm)

Ο φασματικός αλγόριθμος ομαδοποίησης χρησιμοποιείται για να πραγματοποιηθεί η εξαγωγή των key-frames σε ένα πλάνο του βίντεο ομαδοποιώντας τα εικονοπλαίσια σε ομάδες. Κατόπιν ως key-frame χαρακτηρίζεται το medoid κάθε ομάδας δηλαδή εκείνο το εικονοπλαίσιο της ομάδας το οποίο έχει τη μεγαλύτερη μέση ομοιότητα με όλα τα εικονοπλαίσια της ομάδας. Τα κύρια βήματα του αλγορίθμου [22] παρουσιάζονται παρακάτω. Υποθέτοντας ότι έχουμε ένα σύνολο αντικειμένων $S = \{s_1, s_2, \dots, s_N\}$ και θέλουμε να το χωρίσουμε σε k ομάδες.

- Υπολογίζουμε τον πίνακα ομοιότητας $A \in R^{N \times N}$ από τα ζευγάρια των αντικειμένων του συνόλου S .

- Καθορίζουμε τον πίνακα D ως τον διαγώνιο πίνακα του οποίου τα (i,i) στοιχεία του είναι το άθροισμα της i -οστής γραμμής του πίνακα A και κατασκευάζουμε τον Λαπλασιανό πίνακα $L = I - D^{-1/2} A D^{-1/2}$.
- Υπολογίζουμε τα k κύρια ιδιοδιανύσματα $x_1, x_2, \dots, x_k$ του πίνακα L για να δημιουργήσουμε έναν $N \times K$ πίνακα $X = [x_1, x_2, \dots, x_k]$.
- Κανονικοποιούμε κάθε γραμμή του πίνακα X ώστε να έχει μέτρο ένα και σχηματίζουμε τον πίνακα Y με τον εξής τρόπο :

$$y_{ij} = x_{ij} / (\sum_l x_{il}^2)^{1/2} \quad (4.1)$$

- Ομαδοποιούμε τις γραμμές του Y σε k ομάδες χρησιμοποιώντας τον k-means αλγόριθμο.
- Τελικά αναθέτουμε το αντικείμενο s_i στην ομάδα j εάν και μόνο αν η γραμμή i του πίνακα Y έχει ανατεθεί στην ομάδα j .

Σε ότι αφορά το δικό μας πρόβλημα εξαγωγής key-frame υποθέτουμε ότι έχουμε ένα σύνολο $H = \{H_1, \dots, H_N\}$ όπου το H_n είναι ένα διάνυσμα χαρακτηριστικών (ένα από τα κανονικοποιημένα ιστογράμματα που περιγράφονται στο κεφάλαιο 3), του νιοστού εικονοπλαισίου. Η συνάρτηση απόστασης που χρησιμοποιούμε είναι η Ευκλείδεια απόσταση ανάμεσα στα ιστογράμματα των εικονοπλαισίων. Έτσι κάθε στοιχείο του πίνακα ομοιότητας A υπολογίζεται ως εξής :

$$a(i, j) = 1 - \frac{1}{\sqrt{2}} \sqrt{\sum_{h \in bins} (H_i(h) - H_j(h))^2} \quad (4.2)$$

4.2. Ο αλγόριθμος k-means

Ο αλγόριθμος k-means (k-μέσων) είναι ένας αλγόριθμος [21] που ομαδοποιεί αντικείμενα βάσει των χαρακτηριστικών των k ομάδων. Υποθέτει ότι τα χαρακτηριστικά του αντικειμένου ορίζουν ένα διάνυσμα και ο σκοπός του είναι να

ελαχιστοποιήσει τη συνολική διακύμανση των ομάδων (συνάρτηση σφάλματος ομαδοποίησης).

Τα βασικά βήματα του αλγόριθμου είναι τα εξής:

1. Επιλογή του αριθμού των ομάδων.
2. Τυχαία δημιουργία k ομάδων και ορισμός των κεντροειδών των ομάδων.
3. Μεταβίβαση του κάθε σημείου στο κεντροειδές της κοντινότερης ομάδας.
4. Υπολογισμός των νέων κεντροειδών των ομάδων.
5. Επανάληψη των 2 - 4 μέχρι να συγκλίνει ο αλγόριθμος σε κάποιο κριτήριο.

Ο αλγόριθμος ξεκινά διαχωρίζοντας τα αρχικά σημεία σε k αρχικά σύνολα είτε τυχαία είτε χρησιμοποιώντας ευρετικές μεθόδους. Στη συνέχεια υπολογίζει το κεντροειδές του κάθε συνόλου, υλοποιεί νέο διαχωρισμό ώστε το κάθε σημείο να σχετίζεται με το κοντινότερο κεντροειδές. Έπειτα τα κεντροειδή ξαναυπολογίζονται για τις νέες ομάδες και ο αλγόριθμος επαναλαμβάνει τα δυο βήματα ωσότου τα σημεία δεν μπορούν να αλλάξουν ομάδες (ή εναλλακτικά τα κεντροειδή παραμένουν αμετάβλητα).

Όσον αφορά την απόδοση ο αλγόριθμος δεν εγγυάται ότι θα αγγίξει το βέλτιστο. Η ποιότητα της τελική λύσης εξαρτάται πολύ από το αρχικό σύνολο ομάδων και μπορεί να είναι πολύ χαμηλότερη από το ολικό βέλτιστο. Επίσης, ένα άλλο μειονέκτημα του αλγόριθμου είναι ότι ο αριθμός των ομάδων πρέπει να οριστεί εξ αρχής.

4.3. Ο Αλγόριθμος Fast Global k-means

Στην μέθοδο που υλοποιήσαμε και πιο συγκεκριμένα στο πέμπτο στάδιο του φασματικού αλγόριθμου ομαδοποίησης αντί να χρησιμοποιήσουμε την απλή προσέγγιση του k-means αλγόριθμου χρησιμοποιήσαμε μια γρήγορη έκδοση του πολύ αποδοτικού global k-means αλγορίθμου [17]. Ο Global k-means είναι ένας αυξητικός, ντετερμινιστικός αλγόριθμος ο οποίος υπερνικά το σημαντικό πρόβλημα της αρχικοποίησης της απλής προσέγγισης του k-means. Το συγκεκριμένο πρόβλημα της αρχικοποίησης τείνει να γίνει αρκετά σοβαρό στην περίπτωση της ομαδοποίησης των εικονοπλαισίων επηρεάζοντας σημαντικά την γρήγορη προσέγγιση του και την ποιότητα των key-frames. Χρησιμοποιώντας όμως τον Global k-means τα παραγόμενα key-frames αποτελούν μια αρκετά λογική αναπαράσταση του περιεχομένου του πλάνου ενός βίντεο. Στη συνέχεια θα περιγράψουμε σύντομα τον αλγόριθμο.

Υποθέτοντας ότι έχουμε ένα σύνολο δεδομένων $X = \{x_1, \dots, x_N\}$, $x_i \in R^d$ και θέλουμε να το διαμερίσουμε σε K ξένες μεταξύ τους ομάδες $C_1, C_2, \dots, C_K$. Αυτός ο αλγόριθμος είναι από τη φύση του αυξητικός. Βασίζεται στην ιδέα ότι η επιθυμητή κατάτμηση του συνόλου δεδομένων σε K ομάδες, μπορεί να αποκτηθεί δια μέσου της τοπικής αναζήτησης (χρησιμοποιώντας τον απλό k-means αλγόριθμο) ξεκινώντας από μια αρχική κατάσταση με α) τα $K-1$ κέντρα να τοποθετούνται στις βέλτιστες θέσεις για το $(K-1)$ πρόβλημα ομαδοποίησης και β) το εναπομένον K -οστό κέντρο να τοποθετείται στην κατάλληλη θέση που αντιστοιχεί σε κάποιο σημείο από το σύνολο δεδομένων. Βασισμένο έτσι σε αυτή την ιδέα το K -πρόβλημα ομαδοποίησης λύνεται όπως περιγράφεται παρακάτω. Αρχίζοντας με $k=1$ βρίσκει την βέλτιστη λύση όπου είναι το κέντρο του συνόλου X . Για να λύσει το πρόβλημα με δυο ομάδες ο αλγόριθμος k-means εκτελείται N φορές (όπου το N είναι το μέγεθος του συνόλου δεδομένων) από τις ακόλουθες αρχικές θέσεις των κέντρων των ομάδων : το πρώτο κέντρο της ομάδας τοποθετείται πάντα στην βέλτιστη θέση με $k=1$, ενώ το δεύτερο κέντρο κατά την n εκτέλεση αρχικά τοποθετείται στη θέση του στοιχείου x_n . Η βέλτιστη λύση που προκύπτει μετά από N επαναλήψεις του k-means θεωρείται ως η λύση για $k=2$. Γενικά εάν θέλουμε να λύσουμε το πρόβλημα για k ομάδες

πραγματοποιούμε N εκτελέσεις του k -means όπου η κάθε εκτέλεση n θα αρχίζει από τα $k-1$ κέντρα που θα έχουν τοποθετηθεί αρχικά στις θέσεις που αντιστοιχούν στην λύση που πρόέκυψε λύνοντας το $k-1$ πρόβλημα ομαδοποίησης, ενώ το k -οστό κέντρο θα τοποθετείται αρχικά στη θέση του στοιχείου x_n . Ένα μεγάλο πλεονέκτημα αυτού του αλγορίθμου είναι ότι παρέχει όλες τις λύσεις για όλα τα k -προβλήματα ομαδοποίησης με $k \leq K$.

Το υπολογιστικό κόστος του global k -means μπορεί να μειωθεί χωρίς σημαντική φθορά στην ποιότητα της λύσης χρησιμοποιώντας τον fast global k -means αλγόριθμο [17]. Αυτή η μέθοδος υπολογίζει ένα άνω όριο E_n από το τελικό σφάλμα ομαδοποίησης που αποκτάται αρχικοποιώντας ένα νέο κέντρο στη θέση x_n . Η αρχική θέση για το καινούργιο κέντρο της ομάδας επιλέγεται ως εκείνο το σημείο x_i το οποίο ελαχιστοποιεί το E_n και έτσι ο k -means εκτελείται μόνο μια φορά για κάθε k . Η εφαρμογή του αλγορίθμου fast global k -means απαιτεί μια απλή εκτέλεση του k -means για κάθε τιμή (m) του αριθμού των ομάδων: $m = 1, \dots, K$.

ΚΕΦΑΛΑΙΟ 5. ΜΕΘΟΔΟΛΟΓΙΕΣ ΕΞΑΓΩΓΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΙΚΟΝΟΠΛΑΙΣΙΩΝ

5.1 Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε ιστογράμματα χρώματος, κίνησης, CCH και SIFT περιγραφών.

5.2 Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε συνδυασμό λύσεων ομαδοποίησης.

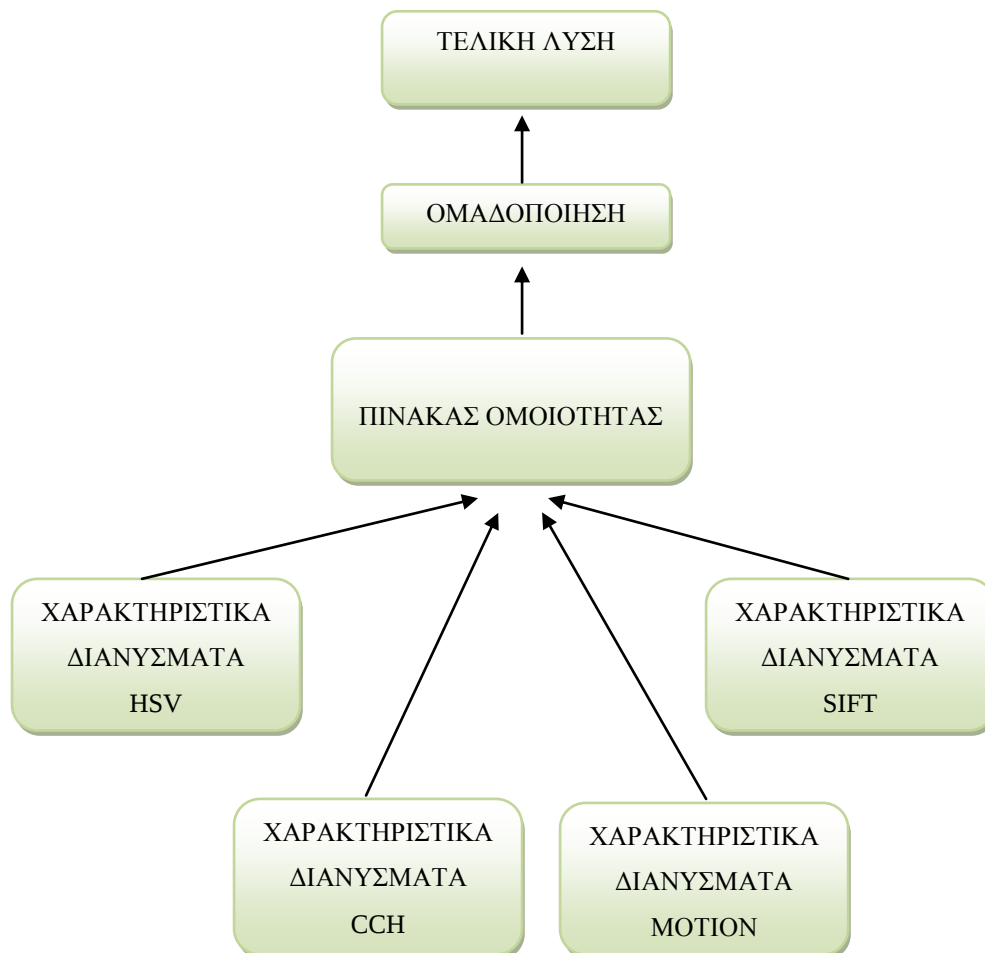
5.3 Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε κατάτμηση της ακολουθίας βίντεο.

Στο κεφάλαιο αυτό παρουσιάζονται οι τρεις μεθοδολογίες εξαγωγής χαρακτηριστικών εικονοπλαισίων με τις οποίες ασχοληθήκαμε.

5.1. Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε ιστογράμματα χρώματος, κίνησης, CCH και SIFT περιγραφών.

Στη μέθοδο αυτή αρχικά υπολογίζονται τα χαρακτηριστικά διανύσματα όπως περιγράφηκαν στο κεφάλαιο 3. Στα πειράματα χρησιμοποιήσαμε όλες τις περιπτώσεις των χαρακτηριστικών, χρώματος, κίνησης, CCH και SIFT περιγραφείς αλλά και συνδυασμούς αυτών. Στη συνέχεια υπολογίζονται οι ομοιότητες των χαρακτηριστικών διανυσμάτων των εικονοπλαισίων και δημιουργείται ο πίνακας ομοιότητας (σχέση 4.2). Στις περιπτώσεις όπου υπάρχει συνδυασμός χαρακτηριστικών αμέσως μετά τη δημιουργία των πινάκων ομοιότητας για κάθε χαρακτηριστικό ακολουθεί ένα επιπλέον στάδιο κατά το οποίο ανάλογα με το βάρος που δίνεται σε κάθε χαρακτηριστικό δημιουργείται ένα καινούργιος πίνακας ομοιότητας ο οποίος αποτελεί συνδυασμό των προηγούμενων. Ο αριθμός των

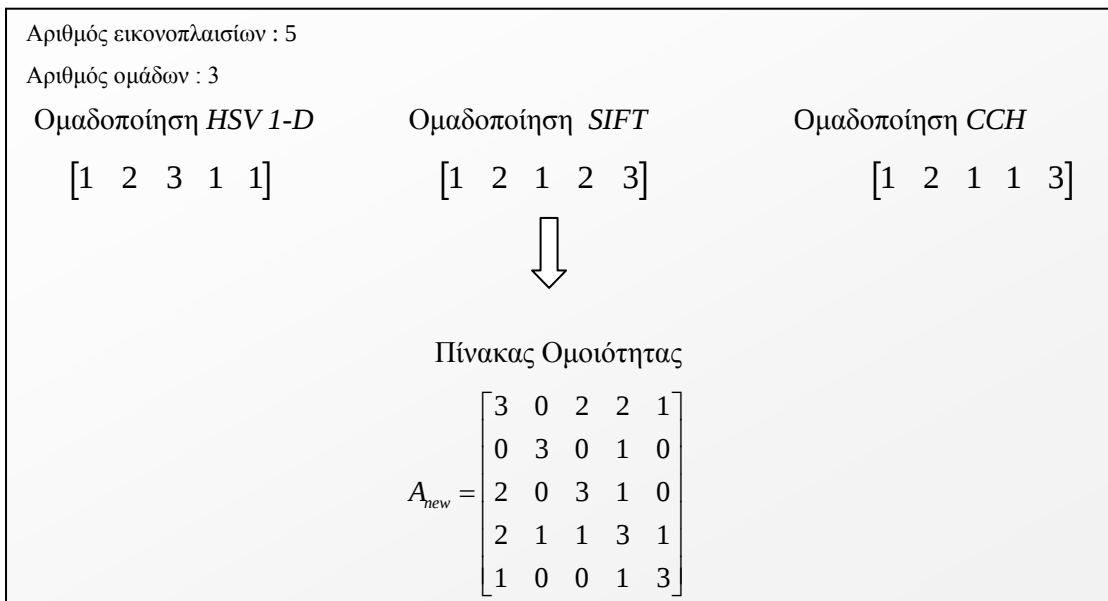
εξαγόμενων key-frames για κάθε ακολουθία εικονοπλαισίων επιλέγεται να είναι ίδιος με τον αριθμό των key-frames του συνόλου ground truth. Στην συνέχεια ακολουθεί η διαδικασία της ομαδοποίησης χρησιμοποιώντας τον φασματικό αλγόριθμο ομαδοποίησης. Τελικά, για κάθε ομάδα επιλέγεται ως key-frame, εκείνο το εικονοπλαίσιο που έχει τη μέση μέγιστη ομοιότητα με τα υπόλοιπα εικονοπλαίσια της ομάδας.



Σχήμα 5.1 Συνοπτική περιγραφή της μεθόδου που βασίζεται σε ιστογράμματα χρώματος, κίνησης και περιγραφών SIFT και CCH.

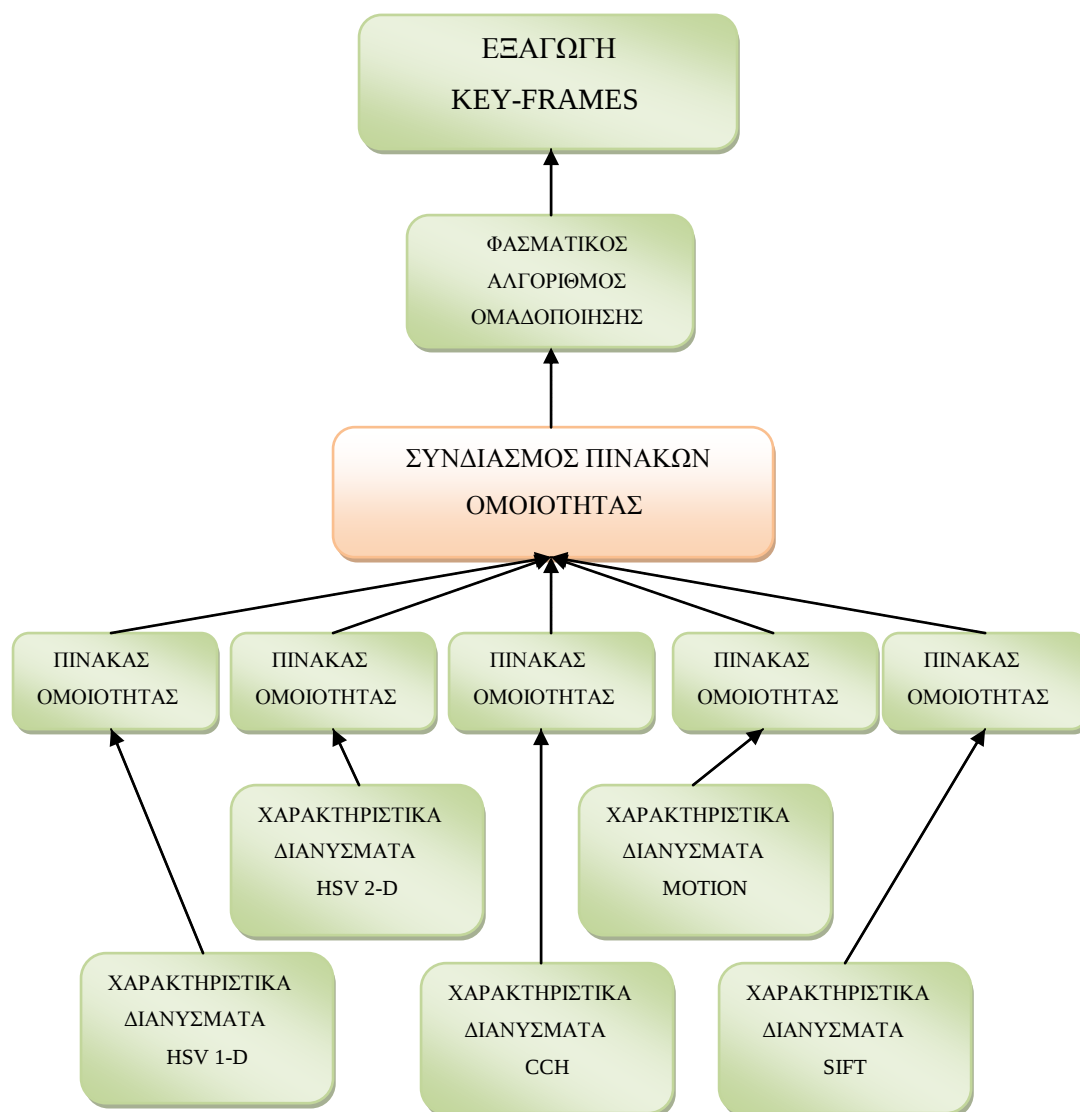
5.2. Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε συνδυασμό λύσεων ομαδοποίησης.

Η μέθοδος η οποία βασίζεται σε συνδυασμό λύσεων ομαδοποίησης είναι η εξής. Αρχικά υπολογίζονται τα χαρακτηριστικά διανύσματα (κεφάλαιο 3). Στα πειράματά μας χρησιμοποιήσαμε χαρακτηριστικά χρώματος, SIFT και CCH περιγραφείς. Στη συνέχεια υπολογίζονται οι ομοιότητες των χαρακτηριστικών διανυσμάτων των εικονοπλαισίων και δημιουργείται ο πίνακας ομοιότητας για κάθε ένα χαρακτηριστικό. Ο αριθμός των εξαγόμενων key-frames για κάθε ακολουθία εικονοπλαισίων επιλέγεται να είναι ίδιος με τον αριθμό των key-frames του συνόλου ground truth. Ακολουθεί η διαδικασία της ομαδοποίησης χρησιμοποιώντας τον φασματικό αλγόριθμο ομαδοποίησης. Η διαφορά αυτής της μεθόδου με την προηγούμενη (5.1) είναι ουσιαστικά στο επόμενο στάδιο. Αφού έχουν ομαδοποιηθεί τα εικονοπλαίσια για κάθε ένα χαρακτηριστικό ξεχωριστά, δημιουργείται ένας καινούργιος πίνακας ομοιότητας A_{new} διάστασης $n \times n$ όπου n το μέγεθος του συνόλου των εικονοπλαισίων. Το (i, j) στοιχείο του πίνακα A_{new} δηλώνει σε πόσα από τα χαρακτηριστικά (χρώμα, SIFT και CCH) το i -οστό εικονοπλαίσιο ανήκει στην ίδια ομάδα με το j -οστό εικονοπλαίσιο.



Σχήμα 5.2 Δημιουργία πίνακα ομοιότητας της μεθόδου εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε συνδυασμό λύσεων ομαδοποίησης.

Στη συνέχεια ο πίνακας κανονικοποιείται και ακολουθεί και πάλι η διαδικασία της ομαδοποίησης χρησιμοποιώντας τον φασματικό αλγόριθμο. Τέλος για κάθε ομάδα επιλέγεται ως key-frame, εκείνο το εικονοπλαίσιο που έχει τη μέση μέγιστη ομοιότητα με τα υπόλοιπα εικονοπλαίσια της ομάδας.



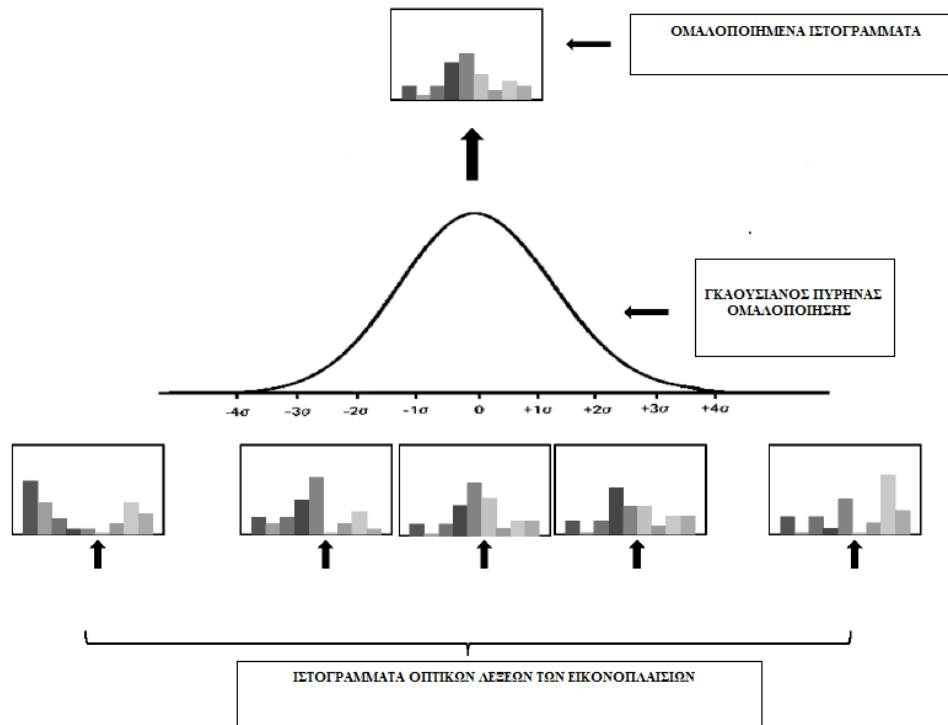
Σχήμα 5.3 Συνοπτική περιγραφή της μεθόδου που βασίζεται σε συνδιασμό λύσεων ομαδοποίησης.

5.3. Μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων βασισμένη σε κατάτμηση της ακολουθίας βίντεο.

Στη μέθοδο αυτή χρησιμοποιείται ένας τοπικός πυρήνας εξομάλυνσης, ο οποίος ομαλοποιεί το ιστόγραμμα οπτικών λέξεων ενός εικονοπλαισίου, λαμβάνοντας υπόψη τα ιστογράμματα των γειτονικών εικονοπλαισίων. Το ομαλοποιημένο ιστόγραμμα SH_t από ένα ιστόγραμμα οπτικών λέξεων VH_t ενός εικονοπλαισίου f_t (όπου το t εκφράζει τη χρονική σειρά του εικονοπλαισίου), δίνεται από την παρακάτω σχέση :

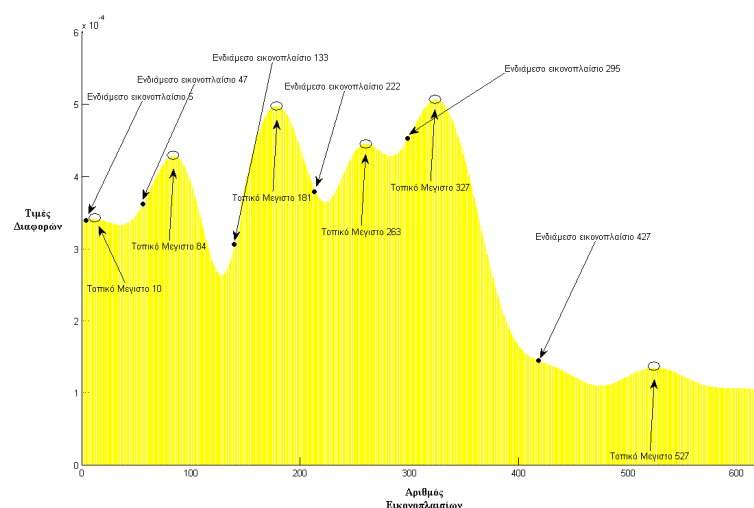
$$SH_t = \sum_{n=-\infty}^{\infty} VH_t \cdot K_{\sigma}(t-n) \quad (5.1)$$

όπου το K_{σ} είναι ο κανονικοποιημένος Γκαουσιανός πυρήνας με μέσο μηδέν και τυπική απόκλιση σ .



Σχήμα 5.4 Χρονική εξομάλυνση του ιστογράμματος οπτικών λέξεων ενός εικονοπλαισίου, χρησιμοποιώντας ένα Γκαουσιανό πυρήνα εξομάλυνσης.

Στο σχήμα (5.4) παρουσιάζεται μια οπτική προσέγγιση της διαδικασίας ομαλοποίησης. Πρώτα από όλα υπολογίζονται τα ιστογράμματα οπτικών λέξεων (VH_i) για κάθε εικονοπλαίσιο (κατώτερο επίπεδο). Κατόπιν τα ιστογράμματα οπτικών λέξεων κάθε εικονοπλασίου ομαλοποιούνται χρονικά με τα γειτονικά ιστογράμματα οπτικών λέξεων χρησιμοποιώντας ένα Γκαουσιανό πυρήνα με αποτέλεσμα τη δημιουργία του ομαλοποιημένου ιστογράμματος (SH_i) των εικονοπλασίων. Ο αριθμός των γειτονικών ιστογραμμάτων που συνεισφέρουν στην ομαλοποίηση εξαρτάται αποκλειστικά από τη σταθερά σ . Όσο πιο μικρή είναι η τιμή της σταθεράς σ τόσο περισσότερη λεπτομέρεια μπορεί να συμπεριλάβει η απεικόνιση. Είναι προφανές έτσι ότι και ο αριθμός των key-frames που εξάγονται από τη συγκεκριμένη διαδικασία εξαρτάται από τη σταθερά σ . Στη συνέχεια υπολογίζονται οι διαφορές των ιστογραμμάτων και κατόπιν επιλέγονται από τη καμπύλη των διαφορών τα τοπικά μέγιστα. Τα σημεία αυτά, οι κορυφές δηλαδή της καμπύλης είναι εκείνα τα σημεία στα οποία παρατηρείται η μεγαλύτερη αλλαγή στο περιεχόμενο της ακολουθίας. Επομένως θέτοντας ως όρια τα τοπικά μέγιστα της καμπύλης ουσιαστικά πετυχαίνουμε το διαχωρισμό της ακολουθίας σε τμήματα τα οποία αποτελούνται από εικονοπλαίσια με όμοιο περιεχόμενο και τελικά πετυχαίνουμε έτσι ένα είδος ομαδοποίησης. Στη συνέχεια μπορεί να γίνει η επιλογή των αντιπροσώπων των ομάδων είτε επιλέγοντας τα medoids κάθε ομάδας (τμήμα της ακολουθίας) είτε τα centroids είτε ακόμη και τα ενδιάμεσα στοιχεία του κάθε τμήματος της ακολουθίας.



Σχήμα 5.5 Επιλογή χαρακτηριστικών εικονοπλασίων από την ομαλοποιημένη καμπύλη διαφορών.

Στα πειράματά μας σε αυτή την εργασία επιλέξαμε ως χαρακτηριστικά εικονοπλαίσια των ακολουθιών βίντεο που εξετάσαμε τα ενδιάμεσα στοιχεία κάθε τμήματος της ακολουθίας (σχήμα 5.5). Επίσης αυξομειώσαμε την τιμή της σταθεράς σ σε κάθε ακολουθία βίντεο που εξετάσαμε με σκοπό ο ακριβής αριθμός των key-frames που πρέπει να έχει η κάθε μια ακολουθία να είναι ίδιος με τον αριθμό του συνόλου των εικονοπλαισίων του ground truth για κάθε ακολουθία.

ΚΕΦΑΛΑΙΟ 6. ΜΕΤΡΑ ΑΞΙΟΛΟΓΗΣΗΣ

6.1 Μέτρα Αξιολόγησης που προτείνονται στη βιβλιογραφία

6.1.1 Average Shot Fidelity

6.1.2 Shot Reconstruction Degree

6.2 Προτεινόμενα μέτρα αξιολόγησης

6.2.1 Χρονική απόσταση από το Ground Truth

6.2.2 Οπτική Αξιολόγηση

6.2.3 Ομοιότητα Περιεχομένου

6.2.4 Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες

Ένα αρκετά σημαντικό θέμα στο πρόβλημα εξαγωγής key-frames έχει να κάνει με την αποτίμηση των εξαγόμενων εικονοπλαισίων και κατά πόσο αυτά μπορούν να είναι αντιπροσωπευτικά για την αναπαράσταση του περιεχομένου ενός πλάνου.

6.1. Μέτρα Αξιολόγησης που προτείνονται στη βιβλιογραφία

Υπάρχουν αρκετά κριτήρια ομοιότητας που μπορούν να χρησιμοποιηθούν για να εκτιμήσουν την αποδοτικότητα των αλγορίθμων. Δυο από αυτά είναι το Fidelity measure όπως παρουσιάστηκε στο [3] και το Shot Reconstruction Degree όπως προτάθηκε στο [19].

6.1.1. Average Shot Fidelity

Αυτό το κριτήριο ομοιότητας συγκρίνει κάθε key-frame μέσα στο πλάνο με όλα τα εικονοπλαίσια μέσα σε αυτό. Δοθέντος μιας ακολουθίας από εικονοπλαίσια $F = \{F_1, F_2, \dots, F_N\}$ και το σύνολο των key-frames $KF = \{KF_1, KF_2, \dots, KF_M\}$, η απόσταση μεταξύ του συνόλου KF και του εικονοπλαισίου F_n ορίζεται ως εξής :

$$d(F_n, KF) = \min_j \text{Diff}(F_n, KF_j), j = 1, \dots, N_{kf} \quad (6.1)$$

όπου N_{kf} είναι ο αριθμός των key-frames και $\text{Diff}(F_i, KF_j)$ είναι η τομή των ιστογραμμάτων ανάμεσα σε δυο εικονοπλαίσια F_i και F_j και ορίζεται ως εξής :

$$\text{Diff}(F_i, F_j) = \sum_{h \in \text{bins}} \min(H_i(h), H_j(h)) \quad (6.2)$$

όπου το H_i και H_j είναι τα διανύσματα των χαρακτηριστικών F_i και F_j εικονοπλαισίων αντίστοιχα. Ωστόσο όπως αναφέρεται στο [19] το κριτήριο αυτό δεν μπορεί να συλλάβει αρκετά καλά την δυναμική ενός πλάνου καθώς δεν εστιάζει σε λεπτομέρειες. Για αυτό το λόγο υπολογίζουμε το Average Shot Fidelity (ASF) το οποίο είναι ένα κριτήριο που υπολογίζεται χρησιμοποιώντας τις ελάχιστες αποστάσεις ανάμεσα στο σύνολο των key-frames και στα εικονοπλαίσια του πλάνου και ορίζεται ως εξής :

$$\text{ASF}(F, KF) = 1 - \frac{1}{N} \sum_{n=1}^N d(F_n, KF) \quad (6.3)$$

6.1.2. Shot Reconstruction Degree

Ολόκληρη η ακολουθία ενός πλάνου μπορεί να αναπαρασταθεί από τα key-frames χρησιμοποιώντας έναν προσθετικό αλγόριθμο. Όσο καλύτερα η ανακατασκευασμένη ακολουθία προσεγγίζει την αρχική ακολουθία του πλάνου, τόσο καλύτερα θα

περιγράφει και το σύνολο των key-frames το περιεχόμενο του πλάνου. Πιο συγκεκριμένα δοθείσας μιας ακολουθίας εικονοπλαισίων F , ένα σύνολο από key-frames KF και έναν αλγόριθμο $IA()$, μπορούμε να ανακατασκευάσουμε κάθε εικονοπλαίσιο από ένα ζευγάρι key-frames KF χρησιμοποιώντας την παρακάτω σχέση:

$$\tilde{F}_n = IA(KF_{n_j}, KF_{n_{j+1}}), n_j \leq n \leq n_{j+1} \quad (6.4)$$

όπου IA είναι ένας αλγόριθμος παρεμβολής. Στην παρούσα εργασία χρησιμοποιήσαμε ένα γραμμικό αλγόριθμο παρεμβολής όπως στο [19].

Το Shot Reconstruction Degree ορίζεται ως εξής :

$$SRD(F, KF) = \sum_{n=0}^{N-1} Sim(F_n, \tilde{F}_n) \quad (6.5)$$

όπου $Sim()$ η ομοιότητα μεταξύ του ανακατασκευασμένου εικονοπλαισίου $\tilde{F}_n$ και του αρχικού F_n και δίνεται από την παρακάτω ισότητα :

$$Sim(F_n, \tilde{F}_n) = \log(1 / Diff(F_n, \tilde{F}_n)) \quad (6.6)$$

στην οποία το $Diff(F_i, F_j)$ είναι η τομή των ιστογραμμάτων των εικονοπλαισίων F_i, F_j όπως ορίστηκε στη σχέση (6.2)

Ένα στοιχείο το οποίο καθιστά πολλές φορές αδύνατη την χρησιμοποίηση των παραπάνω κριτηρίων είναι ότι τα δυο αυτά κριτήρια εξαρτώνται αποκλειστικά από τα χαρακτηριστικά διανύσματα των εικονοπλαισίων και έτσι είναι αδύνατη η σύγκριση μεθόδων που χρησιμοποιούν διαφορετικά χαρακτηριστικά. Έτσι για γίνει εφικτή η σύγκριση τέτοιων μεθόδων χρησιμοποιήσαμε προτείνουμε και άλλα κριτήρια ομοιότητας τα οποία παρουσιάζονται παρακάτω.

6.2. Προτεινόμενα μέτρα αξιολόγησης

Στην παρούσα εργασία προτείνουμε επιπλέον τέσσερα κριτήρια ομοιότητας τα οποία βασίζονται i) στην χρονική απόσταση των key-frames με τα εικονοπλαίσια του ground truth, ii) στην ομοιότητα του περιεχομένου των εικονοπλαισίων των δυο παραπάνω συνόλων, iii) σε συνδυασμό αυτών των δυο και iv) στην ποιότητα της λύσης της ομαδοποίησης.

6.2.1. Χρονική απόσταση από το Ground Truth

Ένα από τα κριτήρια που χρησιμοποιήσαμε για τη σύγκριση των διάφορων μεθόδων είναι η χρονική απόσταση κάθε εικονοπλασίου του συνόλου των key-frames από τα εικονοπλαίσια του συνόλου ground-truth.

Δοθείσας μιας ακολουθίας από εικονοπλαίσια $F = \{F_1, F_2, \dots, F_N\}$, του συνόλου των key-frames $KF = \{KF_1, KF_2, \dots, KF_{N_{kf}}\}$ και του συνόλου ground truth $GT = \{GT_1, GT_2, \dots, GT_{N_{kf}}\}$, ορίζεται μια συνάρτηση κόστους S_{KF_i} η οποία δίνεται :

$$S_{KF_i}(i) = \begin{cases} 1 & \min d_1(KF_i, GT) < dist \\ 0 & \min d_1(KF_i, GT) > dist \end{cases} \quad i = 1 \dots N_{kf} \quad (6.7)$$

όπου $dist$ είναι η ελάχιστη αποδεκτή χρονική απόσταση μεταξύ του key-frame και του Ground Truth και $d_1(KF_i, GT)$ η χρονική απόσταση του i key-frame από τα εικονοπλαίσια του Ground Truth. Στην παρούσα εργασία για τα πειράματα στην μια περίπτωση η παράμετρος $dist$ ορίστηκε σταθερή $dist=10$ και στην άλλη μεταβαλλόμενη ανάλογα με το μέγεθος της ακολουθίας των εικονοπλαισίων. Τελικά το κριτήριο που αξιολογεί την παραγόμενη ακολουθία από key-frames ορίζεται ως εξής :

$$S_D = \frac{\sum_{i=1}^{N_{kf}} S_{KF_i}(i)}{N_{kf}} \times 100 \quad (6.8)$$

6.2.2. Οπτική Αξιολόγηση

Στην παρούσα εργασία προτείνεται ως κριτήριο επίσης η αξιολόγηση της ομοιότητας του συνόλου των key-frames των διάφορων μεθόδων, με το σύνολο ground truth με βάση την οπτική εξέταση. Έτσι δοθέντος μιας ακολουθίας από εικονοπλαίσια $F = \{F_1, F_2, \dots, F_N\}$, του συνόλου των key-frames $KF = \{KF_1, KF_2, \dots, KF_{N_{kf}}\}$ και του συνόλου ground truth $GT = \{GT_1, GT_2, \dots, GT_{N_{kf}}\}$, ορίζεται μια συνάρτηση κόστους $S_{KF_2}(i)$ $i = 1 \dots N_{kf}$ η οποία παίρνει την τιμή ένα στην περίπτωση όπου το i key-frame ταιριάζει οπτικά με κάποιο από εικονοπλαίσια του συνόλου Ground Truth. Τελικά το κριτήριο αξιολόγησης ορίζεται ως εξής :

$$S_V = \frac{\sum_{i=1}^{N_{kf}} S_{KF_2}(i)}{N_{kf}} \times 100 \quad (6.9)$$

6.2.3. Ομοιότητα Περιεχομένου

Δοθείσης μιας ακολουθίας από εικονοπλαίσια $F = \{F_1, F_2, \dots, F_N\}$, του συνόλου των key-frames $KF = \{KF_1, KF_2, \dots, KF_{N_{kf}}\}$ και του συνόλου ground truth $GT = \{GT_1, GT_2, \dots, GT_{N_{kf}}\}$, ορίζεται μια συνάρτηση κόστους $S_{KF_3}(i)$ $i = 1 \dots N_{kf}$ η οποία παίρνει την τιμή ένα στην περίπτωση την οποία η ομοιότητα του περιεχομένου του i key-frame υπερβαίνει σε σχέση με τα εικονοπλαίσια του συνόλου Ground Truth ένα καθορισμένο κατώφλι t_k και παράλληλα ικανοποιεί τη σχέση $d_1(KF_i, GT) < dist$ όπως ορίστηκε στην παράγραφο (6.2.1). Πιο συγκεκριμένα η συνάρτηση κόστους $S_{KF_3}(i)$ ορίζεται ως :

$$S_{KF_3}(i) = \begin{cases} 1 & d_2(KF_i, GT) > t_k \quad \text{και} \quad d_1(KF_i, GT) < dist \\ 0 & \text{αλλιώς} \end{cases} \quad i = 1 \dots N_{kf} \quad (6.10)$$

όπου $d_2(KF_i, GT)$ η ευκλείδεια απόσταση των ιστογραμμάτων των εικονοπλαισίων, t_k η ελάχιστη αποδεκτή ομοιότητα περιεχομένου μεταξύ του key-frame και του Ground Truth, $d_1(KF_i, GT)$ η χρονική απόσταση του i key-frame από τα εικονοπλάσια του Ground Truth και $dist$ η ελάχιστη αποδεκτή χρονική απόσταση μεταξύ του key-frame και του Ground Truth. Τελικά το κριτήριο αξιολόγησης ορίζεται ως εξής :

$$S_C = \frac{\sum_{i=1}^{N_{Kf}} S_{KF_3}(i)}{N_{Kf}} \times 100 \quad (6.11)$$

6.2.4. Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες

Δοθείσης μιας ακολουθίας από εικονοπλάσια $F = \{F_1, F_2, \dots, F_N\}$, του συνόλου ground truth $GT = \{GT_1, GT_2, \dots, GT_{N_{Kf}}\}$ και μιας ομαδοποίησης των εικονοπλαισίων σε k ομάδες $\{C_1, C_2, \dots, C_k\}$ προτείνεται ένα κριτήριο το οποίο εξετάζει κατά πόσο τα εικονοπλάσια GT_i του Ground Truth έχουν κατανεμηθεί σε διαφορετικές ομάδες. Έτσι ορίζεται μια συνάρτηση κόστους η οποία είναι η εξής :

$$S_{KF_4}(i) = \begin{cases} 1 & GT_i \in C_j \text{ και } GT_l \notin C_j, \forall l=1\dots k, l \neq i \\ 0 & \text{αλλιώς} \end{cases} \quad i=1\dots k \quad (6.12)$$

Τελικά το κριτήριο αξιολόγησης ορίζεται ως εξής :

$$S_R = \frac{\sum_{i=1}^k S_{KF_4}(i)}{N_{Kf}} \times 100 \quad (6.13)$$

Το κριτήριο της ορθής κατανομής των εικονοπλαισίων που προτείνεται στην παρούσα εργασία αποδεικνύεται ιδιαίτερα χρήσιμο για την αξιολόγηση των πειραμάτων που παρουσιάζονται στο κεφάλαιο 7 καθώς μπορούμε να συγκρίνουμε μεθόδους που χρησιμοποιούν διαφορετικά χαρακτηριστικά κάτι το οποίο είναι αδύνατο με τα κριτήρια που προτείνονται στη βιβλιογραφία (SDR και Fidelity), και χωρίς να χρειάζεται να καθορίζονται παράγοντες από τον χρήστη όπως γίνεται στα υπόλοιπα κριτήρια που προτείναμε (χρονική απόσταση από το Ground Truth, οπτική αξιολόγηση και ομοιότητα περιεχομένου).

ΚΕΦΑΛΑΙΟ 7. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

- 7.1 Το σύνολο δεδομένων
 - 7.2 Το σύνολο Ground Truth
 - 7.3 Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε ένα χαρακτηριστικό
 - 7.3.1 Χρονική απόσταση από το Ground Truth
 - 7.3.2 Οπτική Αξιολόγηση
 - 7.3.3 Ορθή κατανομή εικονοπλαισίων στις ομάδες
 - 7.4 Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε συνδυασμό χαρακτηριστικών
 - 7.5 Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε κατάτμηση βίντεο
 - 7.6 Σύγκριση μεθόδων
-

7.1. Το σύνολο δεδομένων

Στην παρούσα εργασία το σύνολο δεδομένων που χρησιμοποιήσαμε για να εκτιμήσουμε την αποδοτικότητα των αλγορίθμων που παρουσιάζουμε αποτελείται από δεκατέσσερις ακολουθίες εικονοπλαισίων που αντιστοιχούν σε ένα πλάνο του βίντεο από όπου τις έχουμε εξαγάγει. Από αυτές οι εφτά είναι σχετικές με αθλητικά γεγονότα και πιο συγκεκριμένα πλάνα ομαδικών αθλημάτων και ολυμπιακών αγωνισμάτων, η μια περιέχει ένα zoom in πλάνο και οι υπόλοιπες αναφέρονται σε πλάνα εσωτερικού χώρου αλλά και εξωτερικά με πολύ κίνηση.

Πίνακας 7.1 Το σύνολο δεδομένων

Ακολουθία	Αριθμός εικονοπλαισίων	Αριθμός Key- frames (Ground Truth)
SPORTS1	206	7
SPORTS2	434	7
SPORTS3	320	7
SPORTS4	177	6
FOOTBALL	696	5
BASKETBALL1	143	4
BASKETBALL2	288	5
TRAFFIC	1316	9
POLITIC	364	8
CLASSIC1	719	5
CLASSIC2	781	8
COMEDY	633	7
RAIN	368	5
ZOOM	659	5

Ο αριθμός των key-frames που αναφέρεται στον πίνακα (7.1), αντιστοιχεί στο ground truth που χρησιμοποιήσαμε για κάθε ακολουθία.

7.2. Το σύνολο Ground Truth

Η επιλογή του συνόλου ground truth έγινε με τρόπο ώστε να προσεγγίζει όσο το δυνατό καλύτερα την περίληψη που θα ήθελε να πετύχει κάποιος βλέποντας αρκετές φορές και έχοντας κατανοήσει το περιεχόμενο των ακολουθιών που μελετήσαμε. Η επιλογή των εικονοπλαισίων προσπαθήσαμε όσο είναι δυνατόν να περιέχει εικονοπλαίσια με βέλτιστη ποιότητα, να μην επαναλαμβάνονται όμοια ως προς το περιεχόμενο εικονοπλαίσια και γενικά να είναι όσο το δυνατόν μικρότερος ο αριθμός των εικονοπλαισίων με την προϋπόθεση να μην χάνεται σημαντική πληροφορία. Παρακάτω παρουσιάζονται για κάθε ένα πλάνο που μελετήσαμε τα εικονοπλαίσια του ground truth:

SPORTS1



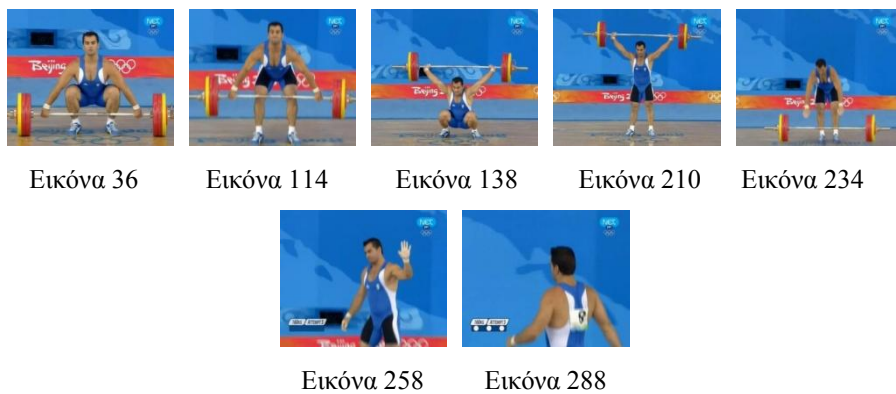
Σχήμα 7.1 Τα εικονοπλαίσια του Ground Truth της ακολουθίας SPORTS1

SPORTS2



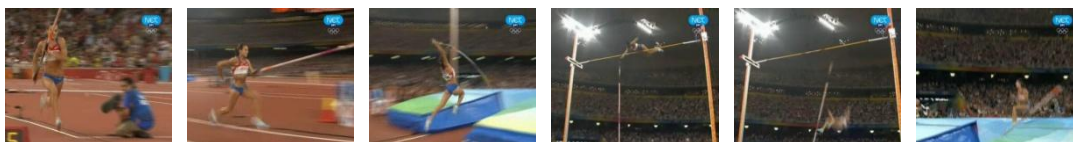
Σχήμα 7.2 Τα εικονοπλαίσια του Ground Truth της ακολουθίας SPORTS2

SPORTS3



Σχήμα 7.3 Τα εικονοπλαίσια του Ground Truth της ακολουθίας SPORTS3

SPORTS4



Εικόνα 35

Εικόνα 83

Εικόνα 100

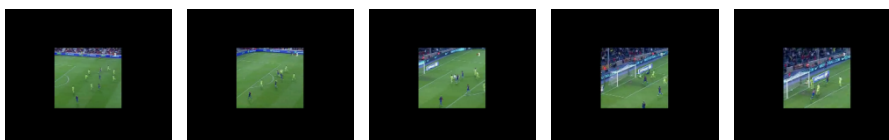
Εικόνα 135

Εικόνα 154

Εικόνα 177

Σχήμα 7.4 Τα εικονοπλαίσια του Ground Truth της ακολουθίας SPORTS4

FOOTBALL



Εικόνα 24

Εικόνα 67

Εικόνα 111

Εικόνα 149

Εικόνα 207

Σχήμα 7.5 Τα εικονοπλαίσια του Ground Truth της ακολουθίας FOOTBALL

BASKETBALL1



Εικόνα 16

Εικόνα 53

Εικόνα 88

Εικόνα 136

Σχήμα 7.6 Τα εικονοπλαίσια του Ground Truth της ακολουθίας BASKETBALL1

BASKETBALL2



Εικόνα 136

Εικόνα 136

Εικόνα 136

Εικόνα 136

Εικόνα 136

Σχήμα 7.7 Τα εικονοπλαίσια του Ground Truth της ακολουθίας BASKETBALL2

TRAFFIC



Εικόνα 63

Εικόνα 222

Εικόνα 268

Εικόνα 356

Εικόνα 542



Εικόνα 727

Εικόνα 1016

Εικόνα 1088

Εικόνα 1233

Σχήμα 7.8 Τα εικονοπλαίσια του Ground Truth της ακολουθίας TRAFFIC

POLITIC



Εικόνα 1233

Εικόνα 1233

Εικόνα 1233

Εικόνα 1233

Εικόνα 1233



Εικόνα 1233

Εικόνα 1233

Εικόνα 1233

Σχήμα 7.9 Τα εικονοπλαίσια του Ground Truth της ακολουθίας POLITIC

CLASSIC1



Εικόνα 33

Εικόνα 153

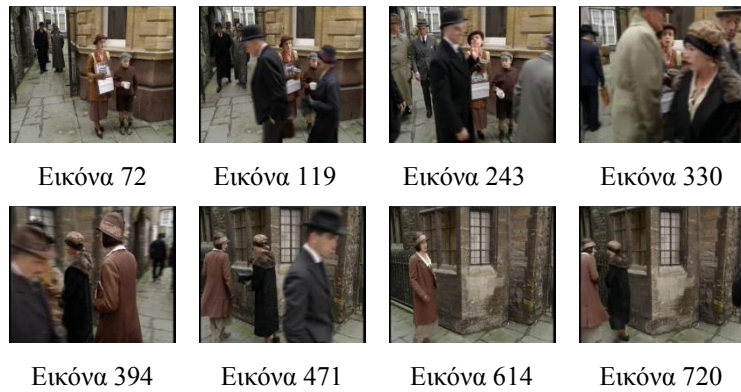
Εικόνα 253

Εικόνα 367

Εικόνα 625

Σχήμα 7.10 Τα εικονοπλαίσια του Ground Truth της ακολουθίας CLASSIC1

CLASSIC2



Σχήμα 7.11 Τα εικονοπλαίσια του Ground Truth της ακολουθίας CLASSIC2

COMEDY

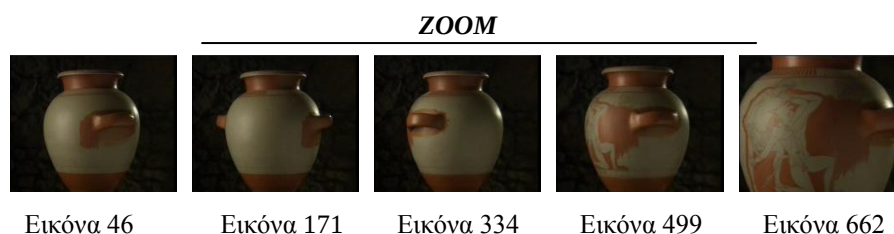


Σχήμα 7.12 Τα εικονοπλαίσια του Ground Truth της ακολουθίας COMEDY

RAIN



Σχήμα 7.13 Τα εικονοπλαίσια του Ground Truth της ακολουθίας RAIN



Σχήμα 7.14 Τα εικονοπλάισια του Ground Truth της ακολουθίας ZOOM

7.3. Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε ένα χαρακτηριστικό

Στην ενότητα αυτή παρουσιάζονται τα αποτελέσματα των πειραμάτων που έγιναν χρησιμοποιώντας τις μεθόδους που περιγράφονται στο κεφάλαιο 5, στην ενότητα 5.1 και βασίζονται σε ένα χαρακτηριστικό.

7.3.1. Χρονική απόσταση από το Ground Truth

Ένα από τα πρώτα κριτήρια σύμφωνα με το οποίο συγκρίνουμε την αποτελεσματικότητα των μεθόδων που παρουσιάσαμε είναι η χρονική απόσταση των εξαγόμενων key-frames από το σύνολο ground truth. Αρχικά θέσαμε σταθερή την απόσταση και ίση με δέκα. Τα αποτελέσματα που πήραμε παρουσιάζονται στον πίνακα (7.2):

Πίνακας 7.2 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό

Κριτήριο S_D : Χρονική απόσταση από το Ground Truth (Distance =10)

Βίντεο	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ								MOTION
	HSV 1-D	HSV 2-D	CCH 20 VSWD	CCH 50 VSWD	CCH 100 VSWD	SIFT 20 VSWD	SIFT 50 VSWD	SIFT 100 VSWD	
SPORTS1	57,1%	42,9%	57,1%	71,4%	57,1%	57,1%	57,1%	57,1%	57,1%
SPORTS2	28,6%	14,3%	0,0%	42,9%	28,6%	28,6%	57,1%	42,9%	14,3%
SPORTS3	42,9%	57,1%	42,9%	85,7%	85,7%	57,1%	57,1%	57,1%	42,9%
SPORTS4	50,0%	50,0%	66,7%	50,0%	50,0%	50,0%	50,0%	66,7%	66,7%
FOOTBALL	80,0%	80,0%	60,0%	40,0%	20,0%	40,0%	60,0%	60,0%	40,0%
BASKETBALL1	100,0%	100,0%	50,0%	50,0%	75,0%	50,0%	50,0%	75,0%	75,0%
BASKETBALL2	60,0%	80,0%	20,0%	40,0%	60,0%	60,0%	60,0%	40,0%	60,0%
TRAFFIC	22,2%	11,1%	33,3%	22,2%	11,1%	22,2%	11,1%	22,2%	11,1%
POLITIC	37,5%	37,5%	37,5%	62,5%	25,0%	37,5%	25,0%	37,5%	50,0%
CLASSIC1	0,0%	20,0%	0,0%	0,0%	0,0%	20,0%	20,0%	0,0%	20,0%
CLASSIC2	37,5%	37,5%	12,5%	25,0%	25,0%	0,0%	12,5%	12,5%	37,5%
COMEDY	0,0%	0,0%	14,3%	28,6%	42,9%	42,9%	42,9%	57,1%	14,3%
RAIN	60,0%	40,0%	20,0%	20,0%	40,0%	20,0%	20,0%	20,0%	20,0%
ZOOM	20,0%	20,0%	20,0%	0,0%	20,0%	0,0%	0,0%	20,0%	20,0%
Σύνολο	39,8%	38,6%	30,7%	39,8%	37,5%	34,1%	36,4%	39,8%	36,4%

Είναι προφανές πως τα αποτελέσματα του παραπάνω πίνακα δεν οδηγούν εύκολα σε ασφαλή συμπεράσματα. Ένα σημαντικό μειονέκτημα της σταθερής απόστασης είναι ότι οδηγεί σε λανθασμένα συμπεράσματα στην περίπτωση που γίνεται σύγκριση ακολουθιών με δυσανάλογο πλήθος (για παράδειγμα BASKETBALL1=143 εικονοπλάισια και TRAFFIC=1316 εικονοπλάισια). Επίσης κάτι το οποίο μπορεί να συμβάλει στο να μην μπορεί να γίνει η σύγκριση με την παραπάνω μέθοδο σωστά είναι το περιεχόμενο της ακολουθίας που εξετάζουμε. Η σταθερή χρονική απόσταση των εικονοπλαισίων θα δώσει τελείως διαφορετικά αποτελέσματα σε μια ακολουθία βίντεο με στατικό περιεχόμενο (παράδειγμα ZOOM), καθώς οι διαφορές στο περιεχόμενο των διαδοχικών εικονοπλαισίων είναι ελάχιστες και τελείως διαφορετικά αποτελέσματα σε μια ακολουθία που στο περιεχόμενο της παρατηρείται έντονη κίνηση (παράδειγμα TRAFFIC). Έτσι οδηγηθήκαμε στο συμπέρασμα ότι θα έπρεπε η απόσταση αυτή να μην είναι σταθερή αλλά μεταβλητή ανάλογα με το πλήθος των εικονοπλαισίων και το περιεχόμενο της ακολουθίας (αν είναι στατικό ή όχι) για κάθε ακολουθία βίντεο. Παρακάτω στον πίνακα (7.4) παρουσιάζονται τα αποτελέσματα των μεθόδων εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό

χρησιμοποιώντας το κριτήριο S_D με ξεχωριστή απόσταση για κάθε ακολουθία βίντεο:

Πίνακας 7.3: Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό
Κριτήριο S_D : Χρονική απόσταση Ground Truth (μεταβλητή απόσταση)

Βίντεο	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ								MOTION
	HSV 1-D	HSV 2-D	CCH 20 VSWD	CCH 50 VSWD	CCH 100 VSWD	SIFT 20 VSWD	SIFT 50 VSWD	SIFT 100 VSWD	
SPORTS1	57,1%	57,1%	71,4%	71,4%	71,4%	71,4%	71,4%	71,4%	71,4%
SPORTS2	28,6%	42,9%	14,3%	42,9%	71,4%	71,4%	42,9%	71,4%	28,6%
SPORTS3	57,1%	71,4%	71,4%	85,7%	85,7%	85,7%	85,7%	57,1%	57,1%
SPORTS4	66,7%	66,7%	83,3%	83,3%	83,3%	83,3%	66,7%	83,3%	83,3%
FOOTBALL	100,0%	100,0%	100,0%	80,0%	80,0%	60,0%	100,0%	100,0%	80,0%
BASKETBALL1	100,0%	100,0%	75,0%	100,0%	75,0%	100,0%	100,0%	100,0%	75,0%
BASKETBALL2	80,0%	100,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%
TRAFFIC	77,8%	66,7%	77,8%	88,9%	77,8%	100,0%	88,9%	77,8%	77,8%
POLITIC	87,5%	75,0%	62,5%	75,0%	75,0%	75,0%	75,0%	75,0%	87,5%
CLASSIC1	80,0%	80,0%	80,0%	80,0%	80,0%	60,0%	60,0%	80,0%	60,0%
CLASSIC2	87,5%	75,0%	87,5%	87,5%	87,5%	87,5%	75,0%	87,5%	75,0%
COMEDY	71,4%	71,4%	85,7%	100,0%	85,7%	71,4%	85,7%	85,7%	71,4%
RAIN	100,0%	100,0%	100,0%	100,0%	100,0%	80,0%	80,0%	100,0%	80,0%
ZOOM	80,0%	80,0%	60,0%	80,0%	80,0%	80,0%	60,0%	60,0%	20,0%
Σύνολο	75,0%	75,0%	73,9%	81,8%	80,7%	79,5%	76,1%	79,5%	68,2%

Οι τιμές που χρησιμοποιήσαμε ως κατώφλι για την χρονική απόσταση των εικονοπλαισίων από το ground truth φαίνονται στον πίνακα (7.3):

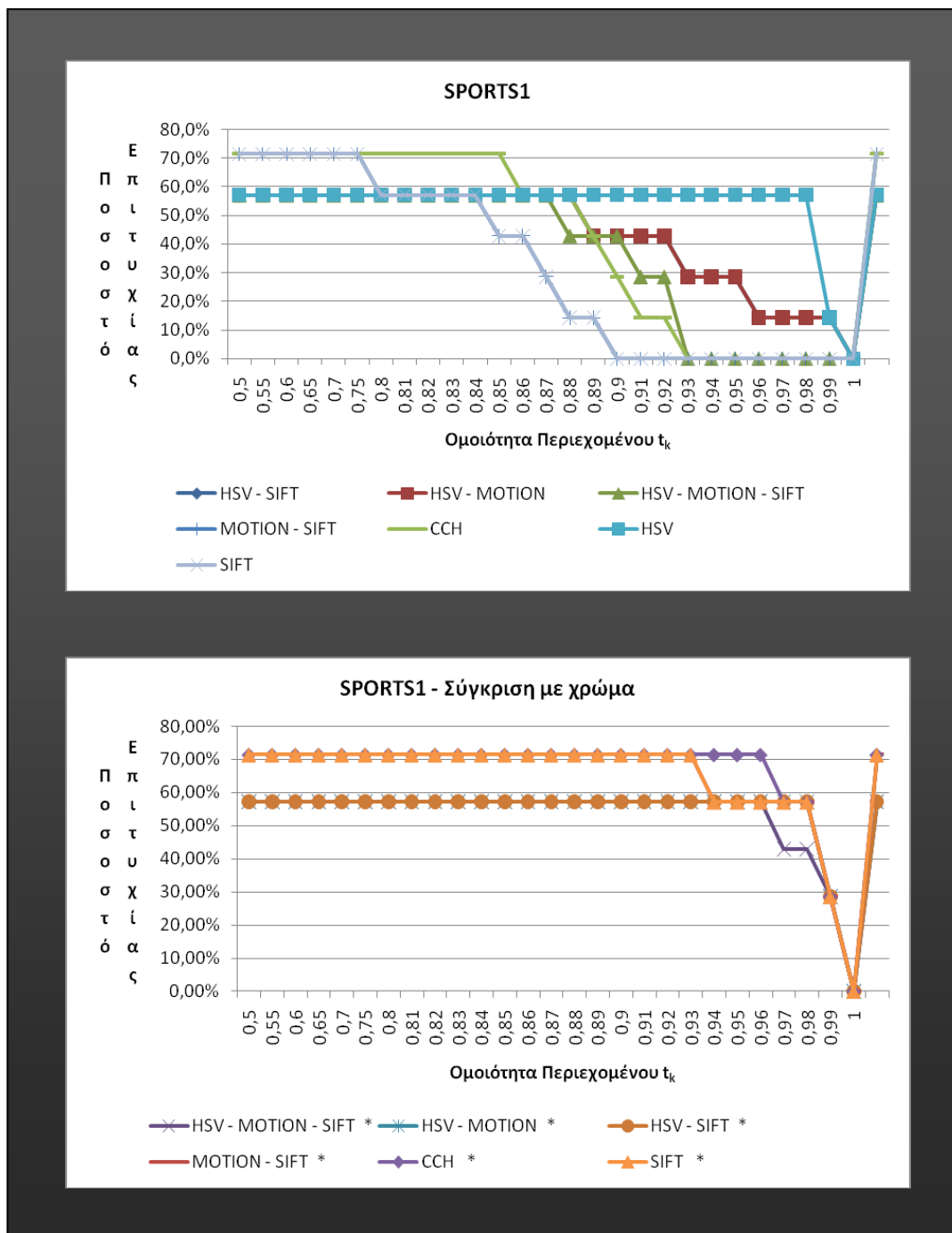
Πίνακας 7.4 Επιθυμητές Αποστάσεις GT με K - F

Βίντεο	Αριθμός εικονοπλαισίων	Αριθμός Key- frames	Όριο Απόστασης
SPORTS1	206	7	20
SPORTS2	434	7	20
SPORTS3	320	7	20
SPORTS4	177	6	20
FOOTBALL	696	5	20
BASKETBALL1	143	4	20
BASKETBALL2	288	5	20
TRAFFIC	1316	9	100
POLITIC	364	8	50
CLASSIC1	719	5	100
CLASSIC2	781	8	100
COMEDY	633	7	100
RAIN	368	5	50
ZOOM	659	5	50

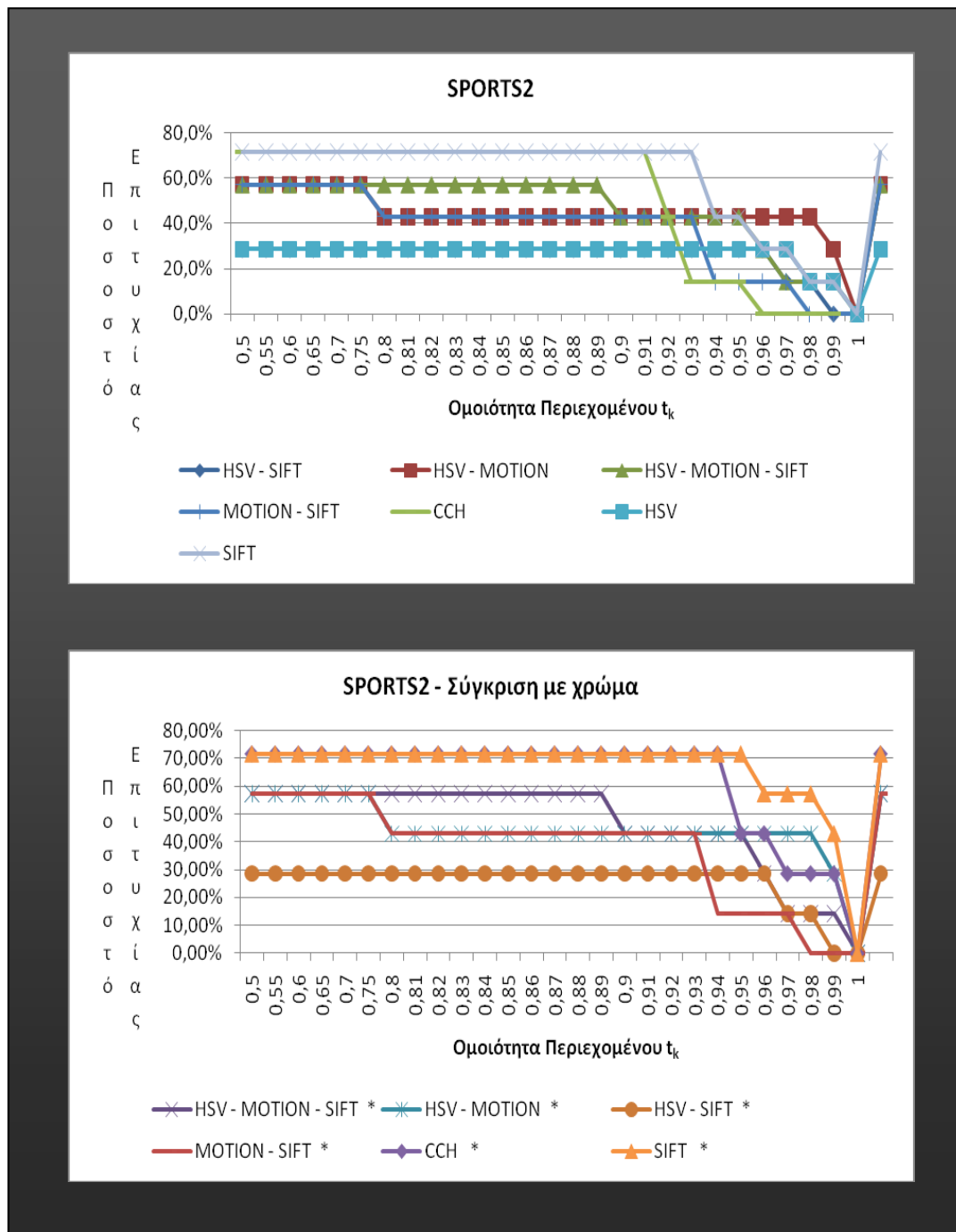
Η βελτίωση στα αποτελέσματα όλων των μεθόδων είναι εμφανής σε σχέση με το προηγούμενο κριτήριο ομοιότητας (πίνακας 7.2). Καλύτερη αναπαράσταση της ακολουθίας από τα key-frames φαίνεται να δίνουν οι μέθοδοι που βασίζονται στους CCH και SIFT περιγραφείς. Υστερούν αρκετά ως προς τις υπόλοιπες, οι μέθοδοι που βασίζονται σε διανύσματα κίνησης.

Στη συνέχεια για να γίνει πιο λεπτομερής η διαδικασία της αξιολόγησης, συγκρίναμε συνδυαστικά μαζί με την χρονική απόσταση από το ground truth και την ομοιότητα του περιεχομένου των key-frames με σύνολο ground truth. Ένα σημαντικό ζήτημα σε αυτήν την περίπτωση είναι ως προς ποιο χαρακτηριστικό θα γίνει η σύγκριση της ομοιότητας του περιεχομένου των εικονοπλαισίων : ως προς το χαρακτηριστικό ή το συνδυασμό χαρακτηριστικών που χρησιμοποιεί ο αλγόριθμος κάθε φορά για την εξαγωγή των key-frames ή μόνο ως προς το χρώμα.

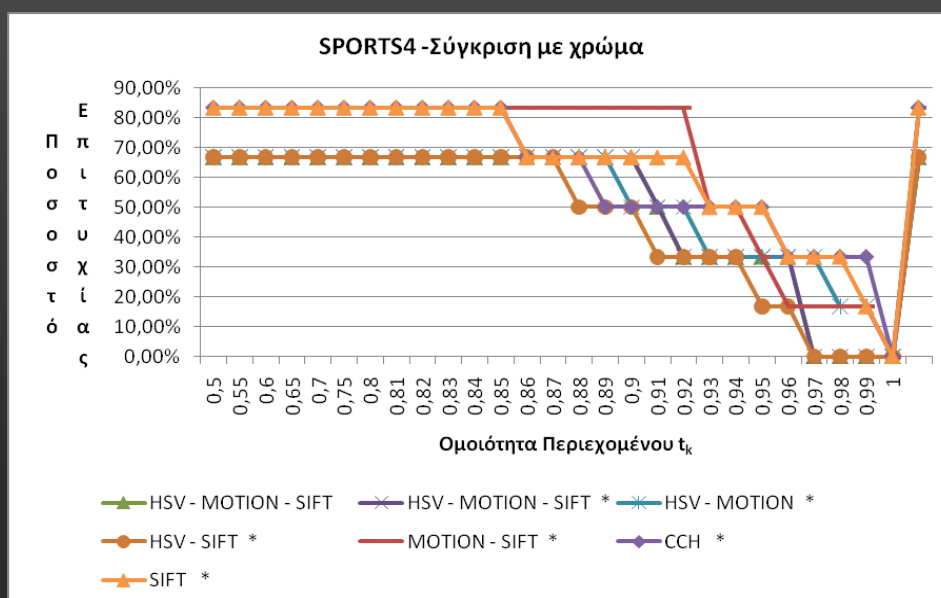
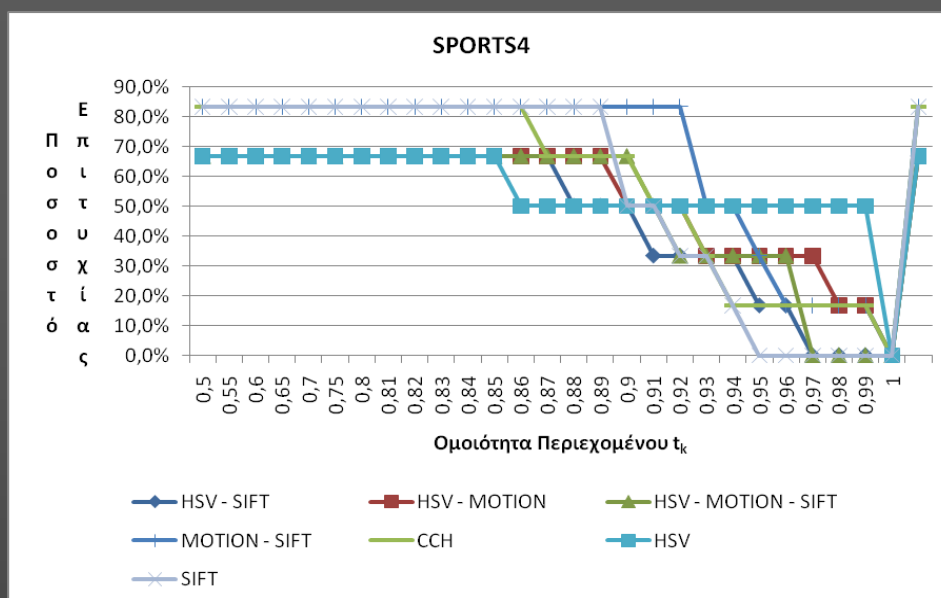
Ενδεικτικά ορισμένα από τα αποτελέσματα που πήραμε παρουσιάζονται στα παρακάτω γραφήματα:



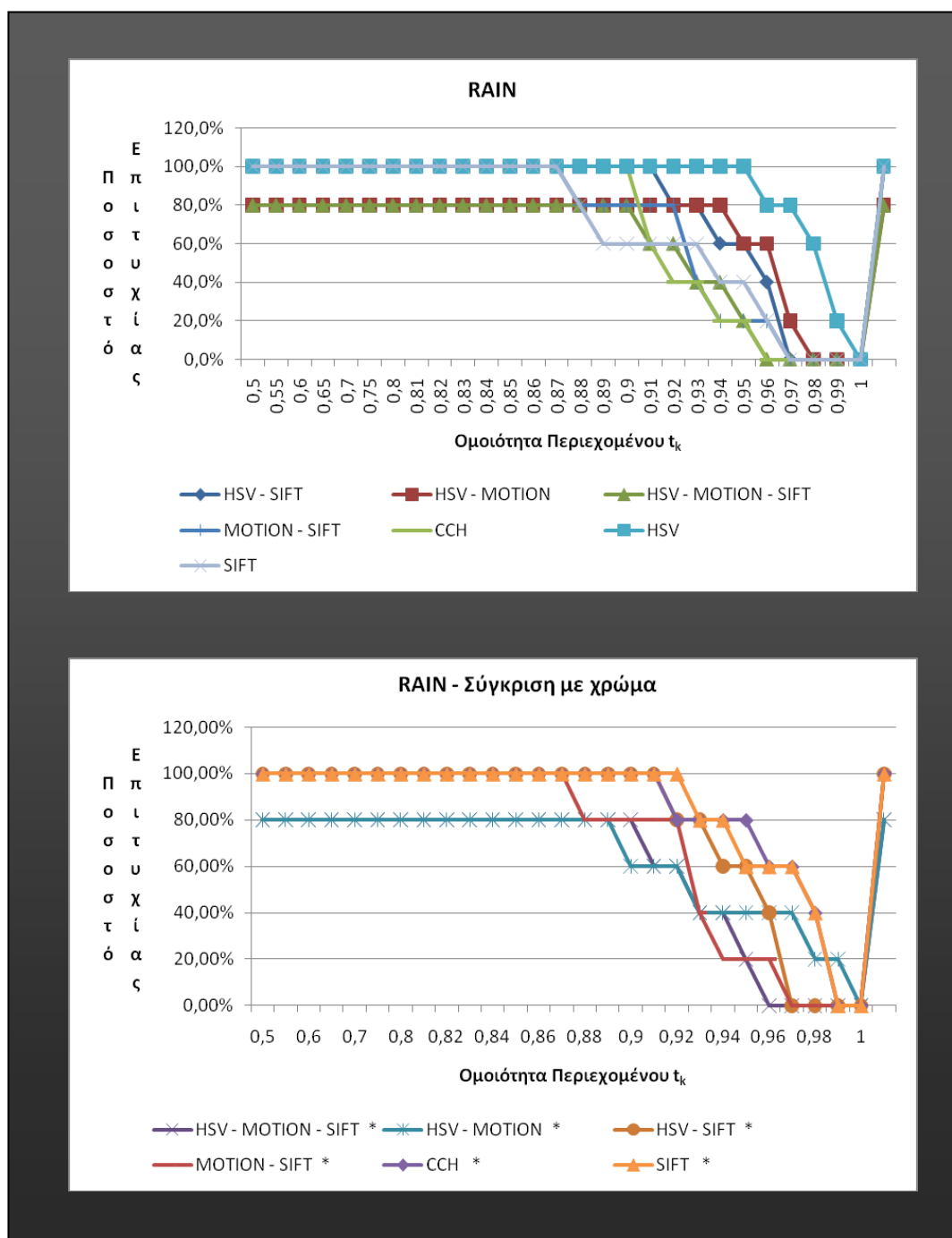
Σχημα 7.15 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία SPORTS1



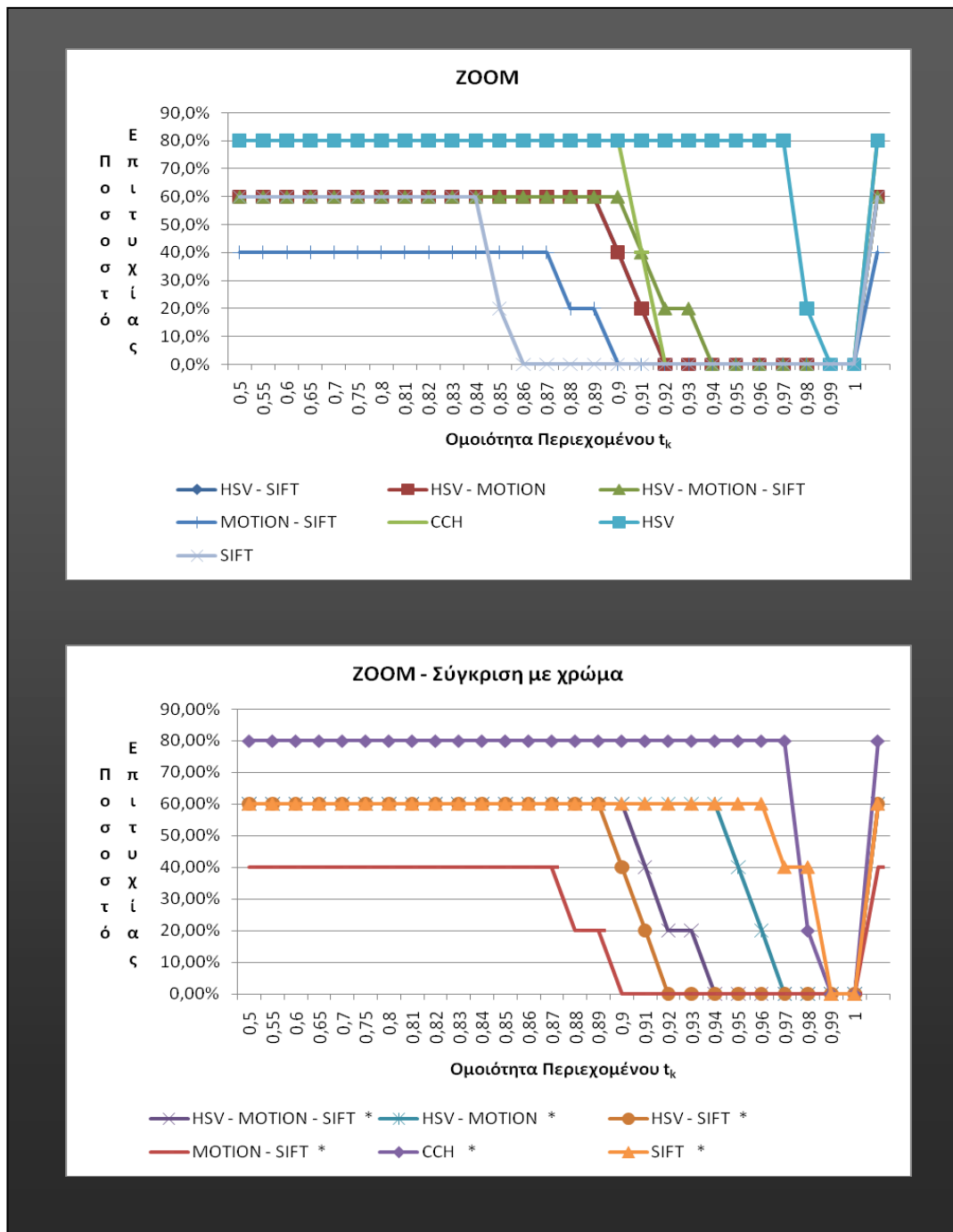
Σχημα 7.16 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία SPORTS2.



Σχημα 7.17 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία SPORTS4.



Σημα 7.18 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία RAIN.



Σχημα 7.19 Τα αποτελέσματα του κριτηρίου S_C ως προς τα χαρακτηριστικά του αλγορίθμου και ως προς το χρώμα, στην ακολουθία ZOOM.

Ο οριζόντιος άξονας παριστάνει τις κατώτατες τιμές της συνάρτησης ομοιότητας πάνω από τις οποίες ένα εικονοπλαίσιο του ground truth και ένα key-frame θεωρείται όμοιο. Ως συνάρτηση ομοιότητας χρησιμοποιήθηκε η Ευκλείδεια απόσταση. Η τελευταία τιμή του οριζόντιου άξονα αναφέρεται μόνο στην χρονική απόσταση και όχι στην ομοιότητα περιεχομένου. Όπου δεν υπάρχει () στις μεθόδους ο έλεγχος της ομοιότητας έγινε ως προς τα χαρακτηριστικά που χρησιμοποίησε ο αλγόριθμος στην συγκεκριμένη μέθοδο. Στις μεθόδους με (*) ο έλεγχος της ομοιότητας έγινε ως προς το χρώμα.*

Σύμφωνα με τα παραπάνω σχήματα δεν υπάρχει σαφής ένδειξη για το κατώφλι που μπορούμε να χρησιμοποιήσουμε για να αυτοματοποιήσουμε τη διαδικασία και επιπλέον απ'ότι φαίνεται η χρονική απόσταση αρκεί για να έχουμε καλά αποτελέσματα.

7.3.2. Οπτική Αξιολόγηση

Ένα άλλο κριτήριο που εφαρμόσαμε στις μεθόδους με ένα χαρακτηριστικό είναι η οπτική αξιολόγηση. Τα αποτελέσματα και εδώ όπως φαίνονται στον πίνακα (7.5) δεν δείχνουν κάποια από τις μεθόδους να υπερισχύει. Πιο κοντά στη βέλτιστη λύση είναι οι μέθοδοι που χρησιμοποιούν ιστογράμματα χρώματος και περιγραφείς SIFT (με σύνολο οπτικών λέξεων 100) ενώ χειρότερη από όλες φαίνεται να είναι η μέθοδος που χρησιμοποιεί διανύσματα κίνησης:

Πίνακας 7.5 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό

Κριτήριο S_V : Οπτική Αξιολόγηση									
Βίντεο	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ								MOTION
	HSV 1-D	HSV 2-D	CCH 20 VSWD	CCH 50 VSWD	CCH 100 VSWD	SIFT 20 VSWD	SIFT 50 VSWD	SIFT 100 VSWD	
SPORTS1	28,6%	42,9%	42,9%	42,9%	28,6%	57,1%	71,4%	71,4%	57,1%
SPORTS2	28,6%	14,3%	28,6%	28,6%	57,1%	57,1%	57,1%	42,9%	28,6%
SPORTS3	57,1%	71,4%	71,4%	85,7%	85,7%	85,7%	85,7%	85,7%	57,1%
SPORTS4	50,0%	50,0%	50,0%	50,0%	50,0%	50,0%	33,3%	50,0%	66,7%
FOOTBALL	100,0%	100,0%	100,0%	80,0%	60,0%	40,0%	80,0%	100,0%	80,0%
BASKETBALL1	75,0%	75,0%	75,0%	75,0%	75,0%	100,0%	75,0%	75,0%	75,0%
BASKETBALL2	80,0%	80,0%	80,0%	60,0%	100,0%	80,0%	60,0%	60,0%	80,0%
TRAFFIC	55,6%	44,4%	55,6%	22,2%	33,3%	55,6%	44,4%	44,4%	11,1%
POLITIC	62,5%	75,0%	62,5%	75,0%	75,0%	50,0%	50,0%	50,0%	62,5%
CLASSIC1	60,0%	60,0%	40,0%	40,0%	40,0%	40,0%	60,0%	60,0%	40,0%
CLASSIC2	50,0%	25,0%	37,5%	37,5%	37,5%	25,0%	37,5%	25,0%	37,5%
COMEDY	71,4%	85,7%	57,1%	57,1%	71,4%	85,7%	57,1%	85,7%	42,9%
RAIN	100,0%	80,0%	60,0%	60,0%	80,0%	40,0%	60,0%	80,0%	60,0%
ZOOM	80,0%	80,0%	60,0%	80,0%	80,0%	80,0%	80,0%	80,0%	80,0%
Σύνολο	61,4%	60,2%	56,8%	54,5%	60,2%	59,1%	59,1%	62,5%	52,3%

7.3.3. Ορθή κατανομή εικονοπλαισίων στις ομάδες

Τέλος ένα τελευταίο μέτρο που χρησιμοποιήσαμε για να συγκρίνουμε τις διάφορες μεθόδους, είναι η ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες που έχουν δημιουργηθεί κατά την διαδικασία ομαδοποίησης. Τα αποτελέσματα που πήραμε παρουσιάζονται στον πίνακα (7.6):

Πίνακας 7.6 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με ένα χαρακτηριστικό

Κριτήριο S_R : Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες											
Βίντεο	Key-frames	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ									
		HSV 1-D	%	HSV 2-D	%	CCH	%	SIFT	%	MOTION	%
SPORTS1	7	5	71,4%	5	72,0%	5	72,0%	5	72,0%	5	72,0%
SPORTS2	7	5	71,4%	6	86,0%	6	86,0%	5	86,0%	4	58,0%
SPORTS3	7	5	71,4%	6	86,0%	5	72,0%	7	100%	5	72,0%
SPORTS4	6	4	66,7%	4	67,0%	4	67,0%	4	67,0%	4	67,0%
FOOTBALL	5	5	100%	5	100%	4	80,0%	5	100%	3	60,0%
BASKETBALL1	4	4	100%	4	100%	3	75,0%	3	75,0%	4	100%
BASKETBALL2	5	4	80,0%	4	80,0%	4	80,0%	5	100%	3	60,0%
TRAFFIC	9	8	88,9%	7	78,0%	7	78,0%	7	78,0%	5	56,0%
POLITIC	8	7	87,5%	7	87,0%	6	75,0%	6	75,0%	6	75,0%
CLASSIC1	5	5	100%	4	80,0%	4	80,0%	4	80,0%	4	80,0%
CLASSIC2	8	6	75,0%	7	87,5%	5	62,5%	7	87,5%	6	75,0%
COMEDY	7	5	71,4%	6	86,0%	6	86,0%	5	72,0%	5	74,2%
RAIN	5	5	100%	5	100%	5	100%	5	100%	4	80,0%
ZOOM	5	4	80,0%	4	80,0%	4	80,0%	5	100%	4	80,0%
Σύνολο	88	72	82%	74	84%	68	78%	73	83%	62	71%

Για την ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες, στις μεθόδους με SIFT και CCH περιγραφείς, χρησιμοποιήσαμε 100 οπτικές λέξεις.

Συγκρίνοντας τα αποτελέσματα του πίνακα (7.6) με τα αποτελέσματα των παραπάνω πινάκων (7.2, 7.4, 7.5) παρατηρούμε σημαντική βελτίωση των ποσοστών επιτυχίας όλων των μεθόδων. Αυτό ίσως οφείλεται στο ότι τα κριτήρια της χρονικής απόστασης αλλά και της οπτικής αξιολόγησης τελικά ίσως να μην αποτελούν και την πλέον αξιόπιστη λύση σύγκρισης της αποτελεσματικότητας αυτών των μεθόδων καθώς εισέρχεται στην σύγκριση το στοιχείο του ανθρώπινου παράγοντα. Πιο αποτελεσματική αποδεικνύεται η μέθοδος εξαγωγής χαρακτηριστικών εικονοπλαισίων η οποία χρησιμοποιεί ιστογράμματα χρώματος διάστασης 2-D, ενώ

υψηλά ποσοστά επιτυχίας παρουσιάζει και η μέθοδος που χρησιμοποιεί ιστογράμματα χρώματος διάστασης $1-D$ όπως και περιγραφών SIFT. Όπως και στα υπόλοιπα μέτρα αξιολόγησης, υστερούν, οι μέθοδοι που χρησιμοποιούν διανύσματα κίνησης. Ένα άλλο συμπέρασμα που προκύπτει είναι πως οι μέθοδοι που χρησιμοποιούν τους περιγραφείς SIFT έχουν καλύτερα αποτελέσματα σε ακολουθίες βίντεο όπου υπάρχει zoom της κάμερας. Αυτό συμβαίνει γιατί οι SIFT περιγραφείς είναι αμετάβλητοι σε περιστροφές ή αλλαγές στο μέγεθος των αντικειμένων, οπότε στην περίπτωση όπου υπάρχει zoom ενώ η κατανομή χρώματος των εικονοπλαισίων μπορεί να αλλάξει, η κατανομή των οπτικών λέξεων παραμένει η ίδια έχοντας σαν συνέπεια να δίνει καλύτερα αποτελέσματα.

7.4 . Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε συνδυασμό χαρακτηριστικών

Μια εναλλακτική των μεθόδων που χρησιμοποιούν ένα χαρακτηριστικό, είναι οι μέθοδοι που βασίζονται σε συνδυασμό των χαρακτηριστικών που περιγράφονται στο κεφάλαιο 5. Στα πειράματα μας χρησιμοποιήσαμε τρεις μεθόδους με όλους τους δυνατούς συνδυασμούς των χαρακτηριστικών χρώματος, περιγραφών SIFT και διανυσμάτων κίνησης και μια με τα τρία χαρακτηριστικά μαζί. Τα ιστογράμματα χρώματος είναι διάστασης $1-D$ και οι περιγραφείς SIFT χρησιμοποιούν 100 οπτικές λέξεις. Τα πειραματικά αποτελέσματα αυτών των μεθόδων φαίνονται στον πίνακα (7.7):

Πίνακας 7.7 Μέθοδοι εξαγωγής χαρακτηριστικών
εικονοπλαισίων με συνδυασμό χαρακτηριστικών

Κριτήριο S_D : Χρονική απόσταση Ground Truth (μεταβλητή απόσταση)				
Βίντεο	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ			
	MOTION - SIFT	SIFT - HSV	MOTION - HSV	HSV - MOTION - SIFT
SPORTS1	71,4%	57,1%	57,1%	57,1%
SPORTS2	57,1%	28,6%	57,1%	57,1%
SPORTS3	42,9%	71,4%	57,1%	57,1%
SPORTS4	83,3%	66,7%	66,7%	66,7%
FOOTBALL	100,0%	100,0%	80,0%	100,0%
BASKETBALL1	75,0%	75,0%	100,0%	100,0%
BASKETBALL2	60,0%	80,0%	60,0%	40,0%
TRAFFIC	66,7%	88,9%	77,8%	77,8%
POLITIC	75,0%	75,0%	62,5%	87,5%
CLASSIC1	60,0%	60,0%	60,0%	60,0%
CLASSIC2	87,5%	75,0%	87,5%	75,0%
COMEDY	57,1%	85,7%	71,4%	85,7%
RAIN	100,0%	100,0%	80,0%	80,0%
ZOOM	40,0%	60,0%	60,0%	60,0%
Σύνολο	69,3%	72,7%	69,3%	71,6%

Η απόσταση είναι μεταβλητή σε κάθε ακολουθία σύμφωνα με τον πίνακα (7.3). Καλύτερη από όλες τις μεθόδους είναι ο συνδυασμός ιστογραμμάτων χρώματος και περιγραφών SIFT. Καμία μέθοδος ως προς το σύνολο όλων των ακολουθιών δεν έχει υψηλό ποσοστό επιτυχίας.

Στη συνέχεια εξετάσαμε τις παραπάνω μεθόδους με το κριτήριο της ορθής κατανομής των εικονοπλαισίων του Ground Truth στις ομάδες. Τα αποτελέσματα φαίνονται στον πίνακα (7.8):

Πίνακας 7.8 Μέθοδοι εξαγωγής χαρακτηριστικών
εικονοπλαισίων με συνδυασμό χαρακτηριστικών

Κριτήριο S_R : Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες				
Βίντεο	ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ			
	MOTION - SIFT	SIFT - HSV	MOTION - HSV	HSV - MOTION - SIFT
SPORTS1	71,4%	71,4%	71,4%	71,4%
SPORTS2	57,1%	71,4%	57,1%	57,1%
SPORTS3	71,4%	85,7%	57,1%	71,4%
SPORTS4	66,7%	83,3%	83,3%	83,3%
FOOTBALL	100,0%	100,0%	80,0%	100,0%
BASKETBALL1	75,0%	75,0%	100,0%	75,0%
BASKETBALL2	80,0%	100,0%	60,0%	80,0%
TRAFFIC	77,8%	77,8%	77,8%	77,8%
POLITIC	75,0%	87,5%	75,0%	75,0%
CLASSIC1	100,0%	100,0%	100,0%	100,0%
CLASSIC2	87,5%	87,5%	75,0%	75,0%
COMEDY	71,4%	71,4%	57,1%	71,4%
RAIN	80,0%	100,0%	80,0%	80,0%
ZOOM	80,0%	100,0%	80,0%	100,0%
Σύνολο	77,3%	85,2%	73,9%	78,4%

Σύμφωνα με τα αποτελέσματα του πίνακα (7.8) φαίνεται πως η μέθοδος που βασίζεται σε συνδυασμό HSV και SIFT περιγραφών δίνει καλύτερα αποτελέσματα συγκριτικά με τις άλλες μεθόδους. Χειρότερη από τέσσερις μεθόδους φαίνεται να είναι αυτή που χρησιμοποιεί συνδυασμό διανυσμάτων κίνησης και ιστογράμματα χρώματος.

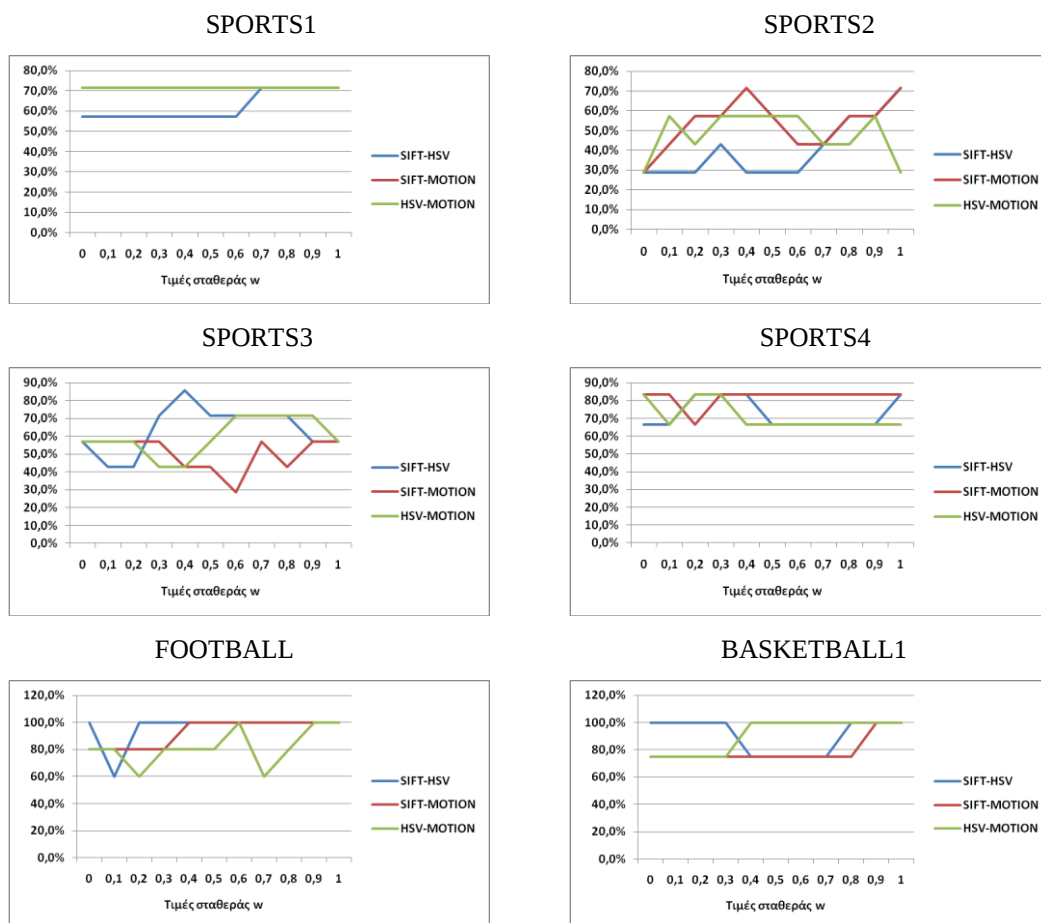
Στις παραπάνω μεθόδους σε κάθε περίπτωση η συνεισφορά κάθε χαρακτηριστικού στη δημιουργία του τελικού πίνακα ομοιότητας όπως περιγράφεται στο κεφάλαιο 4 είναι ίδια. Έτσι θέλοντας να μελετήσουμε τη συμπεριφορά κάθε ενός χαρακτηριστικού ξεχωριστά, προσθέσαμε βάρη σε κάθε ένα από τα χαρακτηριστικά και μελετήσαμε ξεχωριστά κάθε μια από τις ακολουθίες ως προς το κριτήριο της χρονικής απόστασης των key-frames από το ground truth.

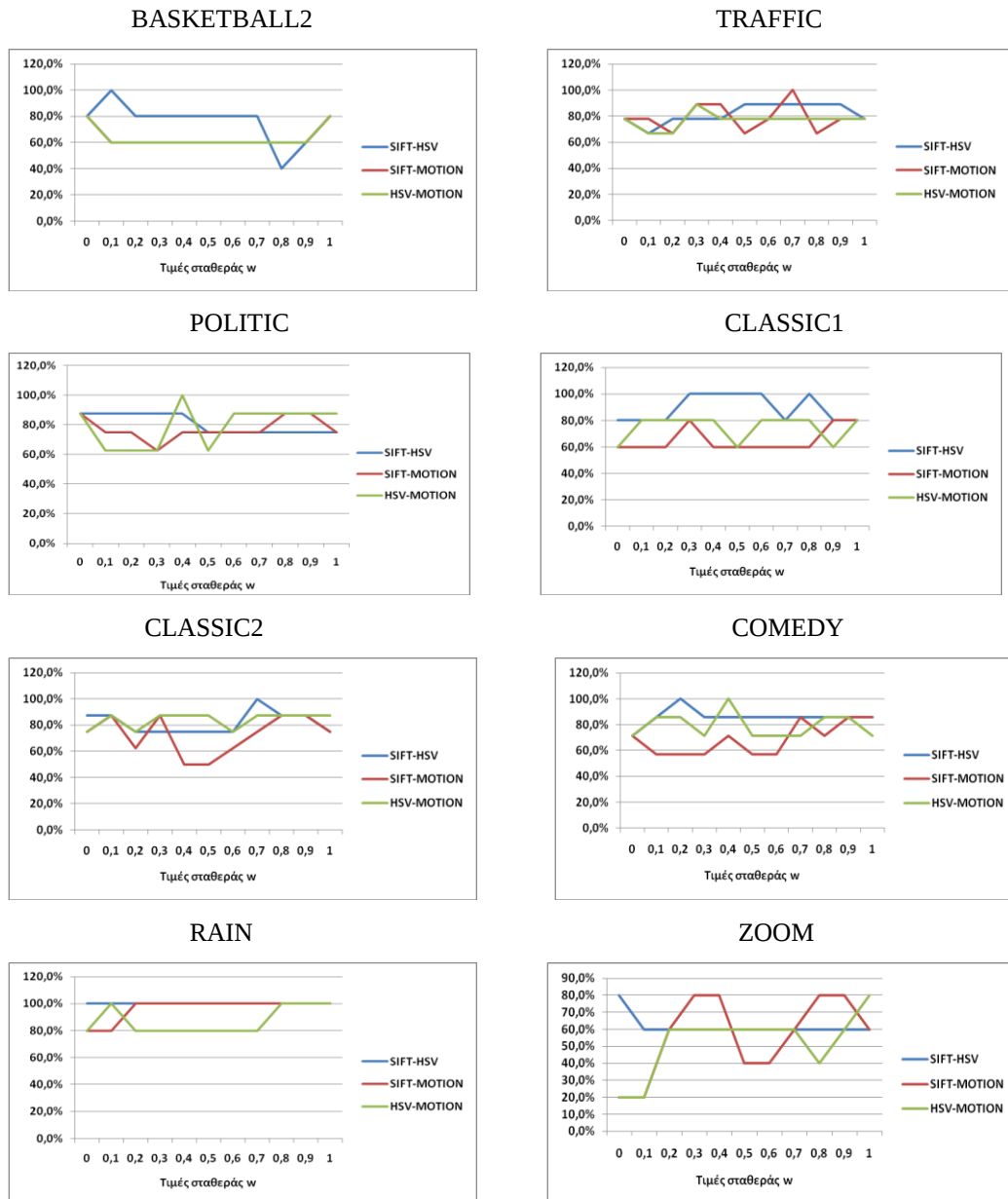
Οπότε ο τελικός πίνακας ομοιότητας A προκύπτει με βάση τη σχέση :

$$A = wa_1 + (1 - w)a_2$$

όπου a_1, a_2 οι πίνακες ομοιότητας κάθε χαρακτηριστικού και $0 \leq w \leq 1$.

Τα αποτελέσματα των πειραμάτων φαίνονται στα παρακάτω γραφήματα:





Σχημα 7.20 Αξιολόγηση των μεθόδων εξαγωγής χαρακτηριστικών εικονοπλαισίων με συνδιασμό χαρακτηριστικών προσθέτοντας βάρη, με το κριτήριο S_D .

Σύμφωνα με τα παραπάνω σχήματα δεν υπάρχει σαφής ένδειξη για τα βάρη που μπορούμε να χρησιμοποιήσουμε σε κάθε χαρακτηριστικό για να αυτοματοποιήσουμε τη διαδικασία.

7.5. Πειραματικά αποτελέσματα μεθόδων που βασίζονται σε συνδυασμό λύσεων ομαδοποίησης και σε κατάτμηση βίντεο

Το κριτήριο σύμφωνα με το οποίο μελετήσαμε την αποδοτικότητα των μεθόδων που βασίζονται σε συνδυασμό λύσεων ομαδοποίησης και σε κατάτμηση της ακολουθίας βίντεο είναι η ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες. Τα αποτελέσματα φαίνονται στον πίνακα (7.9):

Πίνακας 7.9 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων συνδυασμού λύσεων ομαδοποίησης		
Κριτήριο S_R : Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες		
ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ		
Βίντεο	HSV - SIFT - CCH	HSV - SIFT
SPORTS1	71,4%	71,4%
SPORTS2	71,4%	85,7%
SPORTS3	71,4%	85,7%
SPORTS4	66,7%	66,7%
FOOTBALL	80,0%	100,0%
BASKETBALL1	100,0%	75,0%
BASKETBALL2	100,0%	80,0%
TRAFFIC	66,7%	77,8%
POLITIC	75,0%	75,0%
CLASSIC1	80,0%	80,0%
CLASSIC2	75,0%	62,5%
COMEDY	85,7%	71,4%
RAIN	100,0%	100,0%
ZOOM	80,0%	100,0%
Σύνολο	78,4%	79,5%

Στον πίνακα (7.9) συγκρίνοντας τους δυο αλγόριθμους συνδυασμού λύσεων ομαδοποίησης παρατηρούμε ότι έχουν παρόμοια υψηλά ποσοστά επιτυχίας. Καλύτερος είναι ο αλγόριθμος που χρησιμοποιεί τις ομαδοποιήσεις που προήρθαν από ιστογράμματα χρώματος, CCH και SIFT περιγραφών με διαφορά 5%.

Πίνακας 7.10 Μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων με κατάτμηση βίντεο

Κριτήριο S_R : Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες

ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ	
Βίντεο	SIFT
SPORTS1	71,4%
SPORTS2	85,7%
SPORTS3	85,7%
SPORTS4	83,3%
FOOTBALL	100,0%
BASKETBALL1	75,0%
BASKETBALL2	100,0%
TRAFFIC	66,7%
POLITIC	75,0%
CLASSIC1	80,0%
CLASSIC2	75,0%
COMEDY	85,7%
RAIN	80,0%
ZOOM	100,0%
Σύνολο	81,8%

Ο πίνακας (7.10) δείχνει πως η μέθοδος κατάτμησης δίνει καλύτερα αποτελέσματα σε ακολουθίες βίντεο στις οποίες υπάρχει zoom της κάμερας όπως και σε κάποιες από τις ακολουθίες με αθλητικό περιεχόμενο. Αντίθετα δεν δίνει καλή περίληψη σε ακολουθίες που στο περιεχόμενο τους εμφανίζεται έντονη κίνηση και όχι σταθερή λήψη (π.χ. TRAFFIC).

7.6. Σύγκριση μεθόδων

Το κριτήριο σύμφωνα με το οποίο φαίνεται να είναι εφικτή η σύγκριση όλων των μεθόδων που παρουσιάστηκαν στο κεφάλαιο 5, είναι η ορθή κατανομή των εικονοπλαισίων του ground truth στις ομάδες. Τα αποτελέσματα φαίνονται συνοπτικά στον πίνακα (7.11):

Πίνακας 7.11. Συγκριτικός πίνακας όλων των μεθόδων	
Κριτήριο S_R : Ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες	
ΜΕΘΟΔΟΣ	ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ
HSV	81.8%
HSV-2D	84.1%
CCH	77.3%
SIFT	83%
MOTION	70.5%
MOTION - SIFT	77.3%
HSV - MOTION	73.9%
HSV - SIFT	85.2%
HSV - MOTION - SIFT	78,40%
ΣΥΝΔΙΑΣΜΟΣ ΛΥΣΕΩΝ HSV-SIFT	79,50%
ΣΥΝΔΙΑΣΜΟΣ ΛΥΣΕΩΝ HSV-SIFT-CCH	78,40%
ΚΑΤΑΤΜΗΣΗ	81,80%

Καλύτερη αναπαράσταση της ακολουθίας από όλες τις μεθόδους φαίνεται να δίνει ο συνδυασμός ιστογραμμάτων χρώματος και περιγραφών SIFT κάτι που αποδεικνύει τη χρησιμότητα των δυο αυτών χαρακτηριστικών. Επίσης αρκετά υψηλά ποσοστά επιτυχίας έχουν οι μέθοδοι με ιστογράμματα χρώματος διάστασης 2-D. Τελευταία έρχεται η μέθοδος που χρησιμοποιεί διανύσματα κίνησης κάτι που αποδεικνύει πως τα διανύσματα κίνησης δεν αποτελούν και την πλέον αξιόπιστη λύση περίληψης ακολουθίας βίντεο με τον τρόπο που χρησιμοποιήθηκαν στις μεθόδους αυτής της εργασίας.

Τέλος χρησιμοποιώντας τα κριτήρια Shot Reconstruction Degree και Fidelity που περιγράφονται στην παράγραφο 6.1 συγκρίναμε τις δύο μεθόδους που χρησιμοποιούν περιγραφείς SIFT (ομαδοποίησης και κατάτμησης). Τα αποτελέσματα είναι τα εξής:

Πίνακας 7.12 Σύγκριση αλγορίθμων που χρησιμοποιούν περιγραφείς SIFT

Κριτήρια : Shot Reconstruction Degree και Fidelity measure				
Βίντεο	Αλγόριθμος ομαδοποίησης		Αλγόριθμος κατάτμησης	
	Μέτρα αξιολόγησης		Μέτρα αξιολόγησης	
	SDR	FIDELITY	SDR	FIDELITY
SPORTS1	1.591.143	0.5563	1.555.889	0.5547
SPORTS2	7.428.266	0.8071	7.379.379	0.8078
SPORTS3	4.954.165	0.7866	4.981.383	0.7851
SPORTS4	2.373.515	0.7459	2.360.129	0.7397
FOOTBALL	2.153.135	0.6188	2.157.141	0.6116
BASKETBALL1	2.313.715	0.8028	2.308.882	0.7974
BASKETBALL2	1.918.677	0.7369	1.897.115	0.7445
TRAFFIC	1,96E+07	0.7712	1,97E+07	0.7740
POLITIC	6.617.166	0.8324	6.394.541	0.8301
CLASSIC1	1,25E+07	0.8171	1,26E+07	0.8204
CLASSIC2	1,42E+07	0.8373	1,42E+07	0.8341
COMEDY	1,06E+07	0.8375	1,02E+07	0.8222
RAIN	5.654.322	0.7634	5.634.561	0.7596
ZOOM	6.698.987	0.6066	6.588.969	0.6053

Στον πίνακα (7.12) φαίνεται πως οι ακολουθίες βίντεο που εξετάσαμε σε αυτήν την εργασία δίνουν καλύτερη περίληψη εικονοπλαισίων στη μέθοδο που περιγράφεται στην παράγραφο 5.1 χρησιμοποιώντας περιγραφείς SIFT, σε σύγκριση με τη μέθοδο της παραγράφου 5.3 (κατάτμηση βίντεο), χρησιμοποιώντας πάλι περιγραφείς SIFT. Αυτό ισχύει και για τα δυο κριτήρια ομοιότητας SDR και Fidelity.

ΚΕΦΑΛΑΙΟ 8. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

Στην εργασία αυτή μελετήθηκε το πρόβλημα της εξαγωγής χαρακτηριστικών εικονοπλαισίων από ακολουθίες βίντεο. Αφού παρουσιάστηκαν οι βασικές κατηγορίες μεθόδων στη συνέχεια δόθηκε ιδιαίτερη έμφαση στις μεθόδους που βασίζονται σε ιστογράμματα χρώματος, περιγραφείς SIFT και CCH και διανύσματα κίνησης, στις μεθόδους που βασίζονται σε συνδυασμούς λύσεων ομαδοποίησης όπως επίσης και στις μεθόδους που βασίζονται σε κατάτμηση της ακολουθίας βίντεο. Προτάθηκαν τέσσερα κριτήρια αξιολόγησης σύμφωνα με τα οποία εκτιμήθηκε η αποτελεσματικότητα της κάθε μεθόδου. Αξιολογώντας τα αποτελέσματα του κριτηρίου της χρονικής απόστασης των χαρακτηριστικών εικονοπλαισίων από τα εικονοπλαίσια του Ground Truth (S_D) έχοντας σταθερή απόσταση παρατηρούμε ότι σχεδόν όλες οι μέθοδοι δίνουν ποσοστά επιτυχίας κάτω από 50%. Μεταβάλλοντας την απόσταση στη συνέχεια τα ποσοστά επιτυχίας αυξάνονται σημαντικά όμως υπάρχει το ερώτημα κατά πόσο ορθή είναι η επιλογή της τιμής του ορίου της απόστασης για κάθε ακολουθία. Πιο αξιόπιστο από τα κριτήρια που προτάθηκαν φαίνεται να είναι η ορθή κατανομή των εικονοπλαισίων του Ground Truth στις ομάδες (S_R) καθώς σύμφωνα με αυτό οι περισσότεροι αλγόριθμοι σχεδόν σε όλες τις μεθόδους φαίνεται να δίνουν αρκετά καλά αποτελέσματα. Ένας από τους λόγους που μπορεί να συμβαίνει αυτό είναι ότι σε αυτό το κριτήριο δεν απαιτείται κάποια ρύθμιση από τον χρήστη κάτι που ισχύει για τα υπόλοιπα κριτήρια.

Η χρήση του κατωφλίου (t_k) στην ομοιότητα περιεχομένου δυο εικονοπλαισίων μέσω των πειραμάτων που πραγματοποιήσαμε, δεν δίνει σαφή ένδειξη ώστε να μπορούσαμε να αυτοματοποιήσουμε τη διαδικασία και επιπλέον απ'ότι φαίνεται η χρονική απόσταση ως κριτήριο αρκεί από μόνη της ώστε να έχουμε καλά αποτελέσματα. Ακόμη από τα πειράματα που κάναμε φαίνεται πως δεν υπάρχει

σαφής ένδειξη για τα βάρη που πρέπει να έχουν τα χαρακτηριστικά ως προς τη συμμετοχή τους στη δημιουργία του τελικού πίνακα ομοιότητας στις μεθόδους που χρησιμοποιούν συνδυασμό χαρακτηριστικών.

Η χρήση των περιγραφών SIFT φαίνεται πως βοηθά σε ακολουθίες βίντεο όπου υπάρχει zoom της κάμερας ενώ αρνητικά επηρεάζονται οι αλγόριθμοι στην παρουσία διανυσμάτων κίνησης καθώς οι περισσότερες από τις μεθόδους που χρησιμοποιούν διανύσματα κίνησης δεν δίνουν και τόσο καλά αποτελέσματα

Σύμφωνα το κριτήριο S_R καλύτερη από όλες είναι η μέθοδος που χρησιμοποιεί συνδυασμό ιστογραμμάτων χρώματος και περιγραφών SIFT, ενώ στις μεθόδους συνδυασμού λύσεων ομαδοποίησης καλύτερη είναι η μέθοδος που χρησιμοποιεί ιστογράμματα χρώματος, περιγραφείς SIFT και CCH.

Ένα ζήτημα που παρουσιάζει ενδιαφέρον για μελλοντική εργασία είναι η αυξομείωση του αριθμού των οπτικών λέξεων στο σύνολο οπτικών λέξεων που χρησιμοποιούν οι περιγραφείς SIFT και CCH και κατά πόσο αυτό είναι ικανή να επηρεάσει την τελική λύση. Επίσης χρήσιμη θα ήταν η μελέτη ενός αρκετά μεγαλύτερου συνόλου δεδομένων καθώς θα ήταν ίσως πιο εύκολη εξαγωγή συμπερασμάτων για τις μεθόδους που προταθήκαν στην εργασία. Τέλος ένα αρκετά ενδιαφέρον ζήτημα είναι η δημιουργία και χρήση ενός συνόλου οπτικών λέξεων χρησιμοποιώντας άλλους περιγραφείς όπως επίσης και η αυτόματη ανάθεση βαρύτητας στους περιγραφείς με βάση τις ιδιότητες της ακολουθίας των εικονοπλαισίων.

ΑΝΑΦΟΡΕΣ

- [1] Aroh Barjatya,, IEEE. Block Matching Algorithms For Motion Estimation, 2004.
- [2] Z. Cernekova, I. Pitas, and C. Nikou. Information theory-based shot cut/fade detection and video summarization. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(1):82-91, January 2006.
- [3] H. S. Chang, S. Sull, and S. U. Lee. Efficient video indexing scheme for content-based retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(8):1269-1279, December 1999.
- [4] C. Gianluigi and S. Raimondo. An innovative algorithm for key frame extraction in video summarization. *Journal of Real-Time Image Processing*, 1(1):69-88, 2006.
- [5] Girgensohn A., Boreczky J. Time-Constrained Keyframe Selection Technique. *Multimedia Tools and Application*, 2000;11:347-358.
- [6] Gong Y. and Liu X. Generating optimal video summaries. *Proc. IEEE Int. Conference on Multimedia and Expo*, 2000;3:1559-1562.
- [7] B. Gunsel and A. M. Tekalp, “Content-based video abstraction,” in *Proc. IEEE Int. Conf. Image Processing*, Chicago, IL, 1998, pp. 128–132.
- [8] A. Hanjalic, R. L. Lagendijk, and J. Biemond, “A new key-frame allocation method for representing stored video-streams,” in *Proc. 1st Int. Workshop Image Databases Multi Media Search*, Amsterdam, The Netherlands, 1996, pp. 64–67.

- [9] Hanjalic A., Lagendijk R. L., Biemond J. A new Method for Key Frame Based Video Content Representation. In: Image Databases and Multimedia Search, World Scientific Singapore, 1998.
- [10] A. Hanjalic and H. J. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis," *IEEE Trans.*
- [11] A. Hanjalic, R. L. Lagendijk, and J. Biemond. Automated high-level movie segmentation for advanced video-retrieval systems. *IEEE Transactions on Circuits and Systems for Video Technology*, 1999.
- [12] C. Harris and M. Stephens. A combined corner and edge detector. In *The Fourth Alvey Vision Conference*, pages 147-151, 1988.
- [13] Hoon S. H., Yoon K., and Kweon I. A new Technique for Shot Detection and Key Frames. Selection in Histogram Space. Proc. 12th Workshop on Image Processing and Image Understanding, 2000;475-479.
- [14] C.-R. Huang, C.-S. Chen, and P.-C. Chung. Contrast context histogram - a discriminating local descriptor for image matching. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, volume 4, pages 53-56, Los Alamitos, CA, USA, 2006.
- [15] C.-R. Huang, H.-P. Lee, and C.-S. Chen. Shot change detection via local keypoint matching. *IEEE Transactions on Multimedia*, 10(6):1097-1108, October 2008.
- [16] I. Koprinska and S. Carrato. Temporal video segmentation: A survey. *Signal Processing: Image Communication*, 16(5):477-500, January 2001.
- [17] A. Likas, N. Vlassis, and J. J. Verbeek. The global k-means clustering algorithm. *Pattern Recognition*, 36(2):451-461, 2003.

- [18] Liu, T and Kender, J. R. An efficient error-minimizing algorithm for variable-rate temporal video sampling. In Proceedings of the European Conference on Computer Vision (ECCV) 2002.
- [19] T. Liu, X. Zhang, J. Feng, and K.-T. Lo. Shot reconstruction degree: a novel criterion for key frame selection. *Pattern Recognition Letters*, 25(12):1451-1457, 2004.
- [20] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91-110, 2004.
- [21] J. B. MacQueen. Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1:281-297.
- [22] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*, pages 849-856. MIT Press, 2001.
- [23] Narasimha R., Savakis A., Rao R. M. and De Queiroz R. A Neural Network Approach to key Frame extraction. *Proc. of SPIE-IS&T Electronic Imaging Storage and retrieval Methods and Applications for Multimedia*, 2004;5307:439-447.
- [24] Z. Rasheed and M. Shah. *Detection and representation of scenes in videos*. *IEEE Transactions on Multimedia*, 7(6):1097-1105, December 2005.
- [25] X. Sun, M. S. Kankanhalli, Y. Zhu, and J. Wu, “Content-based representative frame extraction for digital video,” draft version; see also *Proc. IEEE Multimedia Computing and Systems*, Austin, TX, 1998.
- [26] Tianming Liu, Hong-Jiang Zhang, and Feihu Qi. A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model, October 2003.

- [27] Truong, B.T and Venkatesh, S.2007. Video abstraction : A systematic review and classification. *ACM Trans. Multimedia Comput. Commun. Appl.* 3,1, Article 3, February 2007.
- [28] Li Zhao, Wei Qi, Stan Z. Li, S.Q.Yang, H.J. Zhang. Key-frame Extraction and Shot Retrieval Using Nearest Feature Line (NFL). *Proc. ACM Int. Workshops on Multimedia Information Retrieval*, 2000;217-220.
- [29] H. J. Zhang, J.Wu, D. Zhong, and S.W. Smoliar, “An integrated system for content-based video retrieval and browsing,” *Pattern Recognit.*, vol. 30, no. 4, pp. 643–658, 1997.
- [30] Zhonghua Sun, Fu Ping. *Combination of Color and Object Outline Based Method in Video Segmentation*. *Proc. SPIE Storage and Retrieval Methods and Applications for Multimedia*, 2004;5307:61-69.
- [31] Y. Zhuang, Y. Rui, T. Huang, and S. Mehrotra. Adaptive key frame extraction using unsupervised clustering. In *ICIP '98:Proc of International Conference on Image Processing*, volume 1, pages 866-870, Chicago, IL, USA, 1998.

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Ο Ελευθέριος Χατζιωαννίδης γεννήθηκε στην Κοζάνη το 1984. Το 2001 εισήχθη στο Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων από το οποίο αποφοίτησε με βαθμό «Λίαν Καλώς» το 2007. Παρακολούθησε το Πρόγραμμα Μεταπτυχιακών Σπουδών του Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων από τον Οκτώβριο του 2007 και απεφοίτησε τον Μάρτιο του 2010 αποκτώντας δίπλωμα με ειδίκευση στις «Τεχνολογίες-Εφαρμογές».

