

ΤΙΤΛΟΣ ΔΙΑΤΡΙΒΗΣ

ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΕΙΚΟΝΟΠΛΑΙΣΙΩΝ ΑΠΟ ΑΚΟΛΟΥΘΙΕΣ ΒΙΝΤΕΟ ΜΕ ΧΡΗΣΗ  
ΟΜΑΔΟΠΟΙΗΣΗΣ ΠΟΛΛΑΠΛΩΝ ΟΨΕΩΝ

Η

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύθεσης

του Τμήματος Πληροφορικής

Εξεταστική Επιτροπή

από τον

Ιωαννίδη Αντώνιο

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΤΟ ΛΟΓΙΣΜΙΚΟ

Ιούλιος 2012



## **ΕΥΧΑΡΙΣΤΙΕΣ**

---

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Λύκα Αριστείδη, τον διδάκτορα κ. Χασάνη Βασίλειο και τον υποψήφιο διδάκτορα κ. Τζώρτζη Γρηγόριο για την πολύτιμη βοήθειά τους κατά τη διάρκεια της εργασίας αυτής. Επίσης θα ήθελα να ευχαριστήσω τον πατέρα μου Ιορδάνη, την μητέρα μου Χρυσή και τον αδερφό μου Γρηγόριο για την υποστήριξή τους σε όλη τη διάρκεια των σπουδών μου.

## ΠΕΡΙΕΧΟΜΕΝΑ

---

	Σελ
ΕΥΧΑΡΙΣΤΙΕΣ	iii
ΠΕΡΙΕΧΟΜΕΝΑ	iv
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	vii
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	ix
ΠΕΡΙΛΗΨΗ	xii
EXTENDED ABSTRACT IN ENGLISH	xv
ΚΕΦΑΛΑΙΟ 1. Εισαγωγή	1
1.1. Περιγραφή του Προβλήματος	1
1.2. Αντικείμενο της Εργασίας	3
1.3. Δομή της Εργασίας	4
ΚΕΦΑΛΑΙΟ 2. Σημαντικά Στοιχεία Σχετικά με την Περίληψη Βίντεο	6
2.1. Τεχνικές Περίληψης Βίντεο	6
2.2. Πλήθος των Χαρακτηριστικών Εικονοπλαισίων	7
2.3. Κατηγορίες Μεθόδων για την Εξαγωγή Χαρακτηριστικών Εικονοπλαισίων	8
2.3.1. Μέθοδος Επαρκούς Αλλαγής Περιεχομένου (Sufficient Content Change)	9
2.3.2. Μέθοδος Μέγιστης Κάλυψης Εικονοπλαισίων (Maximum Frame Coverage)	11
2.3.3. Μέθοδος ελάχιστης συσχέτισης μεταξύ χαρακτηριστικών εικονοπλαισίων (Minimum Correlation Among Key-frames)	12
2.3.4. Μέθοδος με Χρήση του Σφάλματος Ανακατασκευής της Ακολουθίας (Shot Reconstruction Degree)	13
2.3.5. Μέθοδος των «Ενδιαφερόντων γεγονότων» (Interesting Events)	13
2.3.6. Μέθοδος Ομαδοποίησης (Clustering)	15
ΚΕΦΑΛΑΙΟ 3. Περιγραφείς Εικονοπλαισίων	18
3.1. Ιστογράμματα Χρώματος	18
3.2. Περιγραφείς CENTRIST	19

3.3. Περιγραφείς Wavelet	20
3.4. Περιγραφείς SIFT	22
3.5. Περιγραφείς CCH	24
3.6. Περιγραφείς SURF	25
ΚΕΦΑΛΑΙΟ 4. Ομαδοποίηση και Σύνολα Οπτικών Λέξεων (Bag of visual words)	26
4.1. Εισαγωγή	26
4.2. Αλγόριθμος K-means	27
4.3. Αλγόριθμος Fast Global K-means	28
4.4. Δημιουργία Ιστογράμματος Οπτικών Λέξεων	29
ΚΕΦΑΛΑΙΟ 5. Κυρτά Μικτά Μοντέλα με Βάρη (Weighed Multiview Convex Mixture Models)	32
5.1. Εισαγωγή	32
5.2. Κυρτά Μικτά Μοντέλα (Convex Mixture Models)	33
5.3. CMMs Πολλαπλών Όψεων (Multiple View CMMs)	35
5.4. Μοντέλα Προτύπων Πολλαπλών Όψεων με Βάρη (Exemplar-Based Weighted Multiple View Models)	37
5.4.1. Περιγραφή του Μοντέλου	37
5.4.2. Επιπλέον Στοιχεία	42
ΚΕΦΑΛΑΙΟ 6. Πειράματα	43
6.1. Σύνολο Δεδομένων	44
6.2. Καθορισμός του Ground Truth	45
6.3. Εξαγωγή Πληροφορίας από τις Ακολουθίες Βίντεο	52
6.3.1. Ιστογράμματα Χρώματος HSV	53
6.3.2. Περιγραφείς CENTRIST και Wavelet	54
6.3.3. Περιγραφείς CCH, SIFT και SURF	54
6.4. Συναρτήσεις Πυρήνα (Kernels)	56
6.5. Συνδυασμός Πολλών Όψεων με Βάρη (Weighted Multiview)	57
6.6. Κριτήρια Αξιολόγησης	59
ΚΕΦΑΛΑΙΟ 7. Πειραματικά Αποτελέσματα	61
7.1. Πρώτη Σειρά Πειραμάτων	61
7.1.1. Σύγκριση Single-view με Multi-view	62
7.1.2. Σύγκριση των Παραλλαγών Multi-view	68
7.1.3. Σύγκριση Επίδοσης Πυρήνων	74

7.1.4. Σύγκριση Βαρών κάθε Όψης	75
7.1.5. Σύγκριση Βαρών Πυρήνων της Μεθόδου Multi-kernel3	77
7.1.6. Σύγκριση Παραλλαγών Περιγραφέων CCH, SIFT και SURF	77
7.2. Δεύτερη Σειρά Πειραμάτων	80
7.2.1. Σύγκριση Single-view με Multi-view	80
7.2.2. Σύγκριση Επίδοσης Πυρήνων	82
7.2.3. Σύγκριση Βαρών κάθε Όψης	83
7.2.4. Σύγκριση Βαρών Πυρήνων της Μεθόδου Multi-kernel3	86
7.2.5. Σύγκριση Παραλλαγών Περιγραφέων CCH, SIFT και SURF	87
ΚΕΦΑΛΑΙΟ 8. Συμπεράσματα και Προτάσεις για Μελλοντική Εργασία	89
8.1. Συμπεράσματα	89
8.2. Προτάσεις για Μελλοντική Έρευνα	91
ΑΝΑΦΟΡΕΣ	93
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ	99

## **ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ**

---

Πίνακας	Σελ
Πίνακας 6-1 Χαρακτηριστικά του συνόλου των πλάνων που χρησιμοποιήθηκαν στα πειράματα	44
Πίνακας 7-1 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το μέσο όρο και τη διακύμανση	62
Πίνακας 7-2 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το συνολικό μέσο όρο	63
Πίνακας 7-3 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το μέσο όρο με βάση	63
Πίνακας 7-4 Αποτελέσματα της επίδοσης των πέντε διαφορετικών εκδοχών της μεθόδου πολλαπλών όψεων με βάση, με βάση το μέσο όρο και τη διακύμανση	68
Πίνακας 7-5 Αποτελέσματα της επίδοσης των πέντε διαφορετικών εκδοχών της μεθόδου πολλαπλών όψεων με βάση, με βάση το συνολικό μέσο όρο	69
Πίνακας 7-6 Αποτελέσματα της επίδοσης των πέντε διαφορετικών εκδοχών της μεθόδου πολλαπλών όψεων με βάση, με βάση το μέσο όρο με βάση	69
Πίνακας 7-7 Μέσοι όροι επίδοσης για κάθε πυρήνα	74
Πίνακας 7-8 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, χωρίς κατώφλι ελαχίστου βάρους	75
Πίνακας 7-9 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση κατωφλίου ελαχίστου βάρους 5%	75
Πίνακας 7-10 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση κατωφλίου ελαχίστου βάρους 10%	76
Πίνακας 7-11 Μέσοι όροι του ποσοστού συμμετοχής κάθε πυρήνα στη δημιουργία συνολικού πυρήνα (multi-kernel3) για κάθε όψη ξεχωριστά και συνολικά	77
Πίνακας 7-12 Μέσοι όροι και διακύμανση της επίδοσης κάθε παραλλαγής περιγραφέα CCH, SIFT και SURF	78
Πίνακας 7-13 Μέσος όρος επίδοσης, διακύμανση μέσου όρου επίδοσης και μέσος όρος διακύμανσης για κάθε παραλλαγή περιγραφέα	79
Πίνακας 7-14 Διαφορά (ως προς το μέσο όρο) κάθε παραλλαγής περιγραφέα CCH, SIFT και SURF σε σχέση με τον αντίστοιχο Multi περιγραφέα. Οι θετικές τιμές υποδεικνύουν υπεροχή του Multi περιγραφέα.	79

Πίνακας 7-15 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το μέσο όρο και τη διακύμανση	80
Πίνακας 7-16 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το συνολικό μέσο όρο	81
Πίνακας 7-17 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το μέσο όρο με βάση	82
Πίνακας 7-18 Μέσοι όροι επίδοσης για κάθε πυρήνα	83
Πίνακας 7-19 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση της όψης HSV-Multi	84
Πίνακας 7-20 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση της όψης HSV-Avg	84
Πίνακας 7-21 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση της όψης HSV-3D	85
Πίνακας 7-22 Μέσοι όροι του ποσοστού συμμετοχής κάθε πυρήνα στη δημιουργία συνολικού πυρήνα (multi-kernel3) για κάθε όψη ξεχωριστά	86
Πίνακας 7-23 Μέσοι όροι της επίδοσης κάθε παραλλαγής περιγραφέα ξεχωριστά για κάθε πυρήνα και συνολικά για όλους τους πυρήνες	87
Πίνακας 7-24 Διαφορά (ως προς το μέσο όρο) κάθε παραλλαγής περιγραφέα CCH, SIFT και SURF σε σχέση με τον αντίστοιχο Multi περιγραφέα. Οι θετικές τιμές υποδεικνύουν υπεροχή του Multi περιγραφέα.	88



## ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

---

Σχήμα	Σελ
Σχήμα 3.1 Μετασχηματισμός Census	19
Σχήμα 3.2 Διαδικασία εφαρμογής μετασχηματισμού Haar	21
Σχήμα 3.3 Στάδια επιλογής των σημαντικών σημείων. 1) Η αρχική εικόνα...	23
Σχήμα 3.4 Υπολογισμός ενός περιγραφέα σημαντικού σημείου.	24
Σχήμα 4.1 Εφαρμογή του αλγόριθμου K-means από την αρχική κατάσταση μέχρι την τελική.	27
Σχήμα 5.1 Αλγόριθμος EM για CMM πολλαπλών όψεων με βάρη	40
Σχήμα 7.1 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας γραμμικό (linear) πυρήνα	64
Σχήμα 7.2 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας γραμμικό (linear) πυρήνα	64
Σχήμα 7.3 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας γραμμικό (linear) πυρήνα	65
Σχήμα 7.4 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας chi-square πυρήνα	65
Σχήμα 7.5 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας chi-square πυρήνα	65
Σχήμα 7.6 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας chi-square πυρήνα	66
Σχήμα 7.7 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων	66
Σχήμα 7.8 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων	66
Σχήμα 7.9 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων	67
Σχήμα 7.10 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη	67
Σχήμα 7.11 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη	67

Σχήμα 7.12 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη	68
Σχήμα 7.13 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας γραμμικό (linear) πυρήνα	70
Σχήμα 7.14 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας γραμμικό (linear) πυρήνα	70
Σχήμα 7.15 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας γραμμικό (linear) πυρήνα	71
Σχήμα 7.16 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας chi-square πυρήνα	71
Σχήμα 7.17 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας chi-square πυρήνα	71
Σχήμα 7.18 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας chi-square πυρήνα	72
Σχήμα 7.19 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων (intersection)	72
Σχήμα 7.20 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων (intersection)	72
Σχήμα 7.21 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων (intersection)	73
Σχήμα 7.22 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη	73
Σχήμα 7.23 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη	73
Σχήμα 7.24 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη	74



## ΠΕΡΙΛΗΨΗ

---

Αντώνιος Ιωαννίδης του Ιορδάνη και της Χρυσής. MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Ιούλιος, 2012. Εξαγωγή χαρακτηριστικών εικονοπλαισίων από ακολουθίες βίντεο με χρήση ομαδοποίησης πολλαπλών όψεων. Επιβλέπων: Λύκας Αριστείδης.

Στις μέρες μας ένας τεράστιος αριθμός βίντεο υπάρχει στη διάθεση του χρήστη. Το γεγονός αυτό, οδήγησε τους ερευνητές στην ανάπτυξη τεχνικών για όσο γίνεται πιο αξιόπιστη περίληψη, αναζήτηση και ανάκτηση βίντεο. Το πρόβλημα της περίληψης ενός βίντεο, δηλαδή μίας σύντομης και περιεκτικής αναπαράστασης του βίντεο, είναι ένα 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 εξαγωγή χαρακτηριστικών εικονοπλαισίων (key-frames) που αναπαριστούν επαρκώς το συνολικό περιεχόμενο του πλάνου. Έτσι, ένα ολόκληρο βίντεο μπορεί να αναπαρασταθεί από λίγες μόνο εικόνες διατηρώντας ένα μεγάλο ποσοστό πληροφορίας. Το γεγονός αυτό βοηθάει στην κατανόηση του περιεχομένου του βίντεο, χωρίς να είναι απαραίτητη η παρακολούθηση ολόκληρου του βίντεο. Επιπλέον, η χρήση των χαρακτηριστικών εικονοπλαισίων κάνει ευκολότερη την οργάνωση και την ανάκτηση ακολουθιών βίντεο.

Στην εργασία αυτή, αρχικά αναφέρονται περιληπτικά οι προσεγγίσεις που έχουν προταθεί μέχρι σήμερα σχετικά με το πρόβλημα εξαγωγής χαρακτηριστικών εικονοπλαισίων από μία ακολουθία βίντεο. Στη συνέχεια παρουσιάζονται τα οπτικά χαρακτηριστικά όπως ιστόγραμμα χρώματος (HSV), περιγραφείς CENTRIST, SIFT, CCH, SURF και Wavelet, τα οποία χρησιμοποιήθηκαν στην εργασία αυτή. Για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων καθώς επίσης και για τον κατάλληλο

συνδυασμό των οπτικών χαρακτηριστικών χρησιμοποιήθηκε η μέθοδος ομαδοποίησης πολλαπλών όψεων με βάρη. Επιπλέον, εφαρμόστηκαν τρεις διαφορετικές συναρτήσεις ομοιότητας καθώς επίσης και συνδυασμοί αυτών για τα δεδομένα της κάθε όψης. Τέλος, παρουσιάζονται τα κριτήρια που χρησιμοποιήθηκαν για την αξιολόγηση της αποτελεσματικότητας της μεθόδου και αποτιμάται πειραματικά η επίδοση της μεθόδου χρησιμοποιώντας τα παραπάνω κριτήρια αξιολόγησης.



## **EXTENDED ABSTRACT IN ENGLISH**

---

Ioannidis Antonios I. MSc, Computer Science Department, University of Ioannina, Greece. July, 2012. Key-frame extraction from videos using multiple view clustering. Thesis Supervisor: Likas Aristidis.

Nowdays a huge amount of videos is available to everyone. This motivated the researchers in developing techniques for as reliable as possible video summarization, search and retrieval. The subject of video summarization, which is a brief and comprehensive representation of the video content, is one of the most important topics in analysis and processing of digital videos. The most common approach used for the representation of a shot, which is the aim of this thesis, is the extraction of a set of key-frames sufficiently representing the total content of the shot. As result, it becomes possible to represent the whole video content using only a few, cautiously picked, key-frames maintaining at the same time a great percentage of information. This fact is of great help in 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, initially various state-of-the art approaches for key-frame extraction are briefly described. Further on, the visual features used in this thesis are presented. These features are color histograms (HSV) and the descriptors CENTRIST, SIFT, CCH, SURF and Wavelet. For the key-frame extraction and for properly combining the information provided by the visual features (views) a multiple view clustering

method with weights has been applied. In addition, three different similarity functions and proper combinations of them were applied on each view. Finally, the evaluation criteria used for assessing the effectiveness of the proposed method are presented, along with empirical evaluation of the method's performance on several video sequences.



## ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

---

1.1 Περιγραφή του Προβλήματος

1.2 Αντικείμενο της Εργασίας

1.3 Δομή της Εργασίας

---

### 1.1. Περιγραφή του Προβλήματος

Τα τελευταία χρόνια έχει παρατηρηθεί τεράστια ανάπτυξη στον τομέα του ψηφιακού βίντεο. Η ανάπτυξη αυτή είναι αποτέλεσμα της ραγδαίας ανάπτυξης σε διάφορους τομείς της τεχνολογίας όπως για παράδειγμα την υπολογιστική ισχύ, τα ταχύτερα δίκτυα, τη μεγαλύτερη και οικονομικότερη χωρητικότητα των αποθηκευτικών μέσων. Τα παραπάνω σε συνδυασμό με την σμίκρυνση του υλικού οδήγησε στην κατασκευή φορητών συσκευών τεχνολογίας όπως notepads και smartphones, δίνοντας έτσι τη δυνατότητα σε πολλούς χρήστες να έχουν ανά πάσα στιγμή έναν ηλεκτρονικό υπολογιστή μαζί τους. Έτσι αναπτύχθηκαν καινούργιες εφαρμογές και τεχνολογίες προκειμένου να αξιοποιηθεί όσο γίνεται αποτελεσματικότερα και αποδοτικότερα η δυνατότητα αυτή. Ένα από τα χαρακτηριστικά των συσκευών αυτών είναι η εγγραφή και η αναπαραγωγή βίντεο. Το γεγονός αυτό σε συνδυασμό με τις ήδη υπάρχουσες ψηφιακές φωτογραφικές μηχανές και βιντεοκάμερες οδήγησε στη δημιουργία τεράστιων ποσοτήτων βίντεο και κατέστη απαραίτητη η ανάπτυξη εφαρμογών για όσο το δυνατόν καλύτερη διαχείριση των βίντεο αυτών όπως είναι η αρχειοθέτηση, η δημιουργία ευρετηρίου και η ανάκτησή τους.

Μία ακολουθία βίντεο συνήθως περιέχει ένα μεγάλο αριθμό από εικονοπλαίσια (frames), ο οποίος πολλές φορές είναι απαγορευτικός για αρκετές εφαρμογές. Για το

λόγο αυτό είναι πολύ σημαντική η ανάπτυξη ενός μηχανισμού ο οποίος θα επιτρέπει στον χρήστη να έχει άποψη για το περιεχόμενο ενός βίντεο χωρίς να είναι απαραίτητο να παρακολουθήσει ολόκληρο το βίντεο. Αυτό μπορεί να επιτευχθεί δημιουργώντας μία περίληψη του βίντεο.

Βασικός σκοπός της περίληψης ενός βίντεο είναι η διευκόλυνση της αναζήτησης σε μία βάση από βίντεο. Προκειμένου όμως η αναζήτηση αυτή να είναι όσο γίνεται πιο αποδοτική, πρέπει η περίληψη των διαθέσιμων βίντεο να είναι όσο γίνεται περισσότερο αντιπροσωπευτική του αρχικού βίντεο.

Επίσης, η βελτιστοποίηση της περίληψης ενός βίντεο, εκτός από την αποτελεσματικότερη αυτόματη αναζήτηση, δίνει σε έναν χρήστη καλύτερη άποψη για το βίντεο στο οποίο αναφέρεται η περίληψη. Κάτι τέτοιο είναι ιδιαίτερα χρήσιμο σε διάφορες εφαρμογές επεξεργασίας βίντεο, όπως το μοντάζ. Οι περιλήψεις μπορούν ωστόσο να χρησιμοποιηθούν και ανεξάρτητα βελτιώνοντας έτσι τον απαραίτητο χώρο αποθήκευσης, εύρους ζώνης αλλά και του χρόνου παρακολούθησης. Χαρακτηριστικό παράδειγμα είναι τα highlights ενός αθλητικού αγώνα.

Αναγνωρίζοντας έτσι την σημαντικότητα της περίληψης στον τομέα της επεξεργασίας βίντεο, η αυτόματη και αποτελεσματική περίληψη έχει αποτελέσει σημαντικό αντικείμενο έρευνας τα τελευταία χρόνια. Τα βίντεο είναι διαφόρων ειδών όπως ταινίες, ντοκιμαντέρ, αθλητικά, ειδήσεις και αρκετά άλλα. Αν και ορισμένες τεχνικές έχουν αρκετά ικανοποιητικά αποτελέσματα, η δημιουργία περιλήψεων είναι σε ένα μεγάλο βαθμό στο στάδιο της έρευνας. Βασικό στοιχείο περιορισμού είναι ότι κάθε βίντεο έχει διαφορετικά χαρακτηριστικά σε σχέση με άλλα βίντεο. Το γεγονός αυτό μας υποδεικνύει ότι η χρήση ενός οπτικού χαρακτηριστικού δεν αναμένεται να έχει ικανοποιητικά αποτελέσματα όταν εφαρμόζεται σε μία συλλογή διαφορετικών ακολουθιών βίντεο. Για παράδειγμα, δύο αντικείμενα που έχουν ίδιο χρώμα και διαφορετική υφή δεν μπορούν να διακριθούν χρησιμοποιώντας ιστογράμματα χρώματος, παρόλα αυτά χρησιμοποιώντας Wavelets η διάκριση είναι εφικτή λόγω της διαφορετικής υφής των δύο αντικειμένων. Παρόμοια προβλήματα μπορεί να προκύψουν λόγω αλλαγής της φωτεινότητας, της κλίμακας, της τοπολογίας ή κάποιας άλλης λεπτομέρειας. Επειδή μέχρι σήμερα, δεν υπάρχει κάποιος περιγραφέας που να

μπορεί να περιγράψει αποτελεσματικά όλα τα παραπάνω οπτικά χαρακτηριστικά, η ταυτόχρονη χρήση πολλών και διαφορετικών περιγραφών κρίνεται απαραίτητη. Επειδή όμως, η σημαντικότητα του κάθε περιγραφέα είναι ξεχωριστή για κάθε ακολουθία βίντεο και δεν είναι δυνατό να γνωρίζουμε εκ των προτέρων ποιος ή ποιοι τύποι περιγραφέα πρέπει να έχουν μεγαλύτερη βαρύτητα, εξετάζουμε τη χρήση μίας μεθόδου ομαδοποίησης πολλαπλών όψεων με βάρη. Ο αλγόριθμος αυτός προσδίδει αυτόματα βάρη σε κάθε περιγραφέα βασιζόμενος στα χαρακτηριστικά της ακολουθίας βίντεο που εξετάζεται. Με τον τρόπο αυτό φιλοδοξούμε να αντιμετωπίσουμε το παραπάνω σημαντικό πρόβλημα.

## 1.2. Αντικείμενο της Εργασίας

Προκειμένου να είμαστε σε θέση να εξάγουμε την περίληψη ενός βίντεο, πρέπει πρώτα να αναλύσουμε τη δομή του. Τα δεδομένα ενός βίντεο, οργανώνονται σε μία ιεραρχική δομή όπως περιγράφεται στα [12], [13]. Η δομή αυτή προκύπτει χωρίζοντας το βίντεο σε τρία τμήματα που περιέχουν σημαντικές πληροφορίες. Αυτά είναι τα πλάνα (shots), οι σκηνές (scenes) και τα κεφάλαια (chapters). Ένα πλάνο ορίζεται ως μία συνεχής ακολουθία εικονοπλασιών που έχουν καταγραφεί από μία μόνο κάμερα, μια σκηνή αποτελείται από διαδοχικά πλάνα τα οποία περιγράφουν μία ενέργεια ή ένα γεγονός και ένα κεφάλαιο είναι ένα σύνολο από σκηνές.

Στην εργασία αυτή ασχοληθήκαμε με πλάνα που προέρχονται από ένα μεγάλο εύρος ακολουθιών βίντεο και σκοπός μας είναι η περίληψή τους χρησιμοποιώντας χαρακτηριστικά εικονοπλαίσια (key-frames). Ακριβώς επειδή τα πλάνα προέρχονται από διαφορετικά πεδία, είναι σημαντική η δυνατότητα εντοπισμού του κατάλληλου χαρακτηριστικού, ή συνδυασμό χαρακτηριστικών, που οδηγεί στην όσο το δυνατόν καλύτερη περίληψη της ακολουθίας βίντεο. Για να επιτευχθεί κάτι τέτοιο, χρησιμοποιήθηκαν διάφορες μέθοδοι εξαγωγής πληροφορίας, καθώς επίσης και συνδυασμοί αυτών, που αναφέρονται σε διαφορετικά στοιχεία μίας εικόνας όπως το χρώμα, η τοπολογία, η υφή καθώς επίσης και λεπτομέρειες που μπορούν να αναδείξουν διαφορές ανάμεσα σε δύο εικόνες.

### 1.3. Δομή της Εργασίας

Αρχικά, στο 2<sup>ο</sup> κεφάλαιο περιγράφονται οι δύο βασικές κατηγορίες περίληψης βίντεο που είναι τα video skims και τα χαρακτηριστικά εικονοπλαίσια (key-frames). Στη συνέχεια αναφέρονται τρόποι εύρεσης του πλήθους των χαρακτηριστικών εικονοπλαισίων για ένα βίντεο. Τέλος, περιγράφονται συνοπτικά οι βασικές μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων και αναλύεται περισσότερο η μέθοδος της ομαδοποίησης, που είναι και αυτή που χρησιμοποιήθηκε στην εργασία.

Στο 3<sup>ο</sup> κεφάλαιο παρουσιάζονται οι περιγραφείς εικονοπλαισίων HSV, CENTRIST, Wavelet, SIFT, CCH και SURF οι οποίοι χρησιμοποιήθηκαν για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων.

Στο επόμενο κεφάλαιο περιγράφονται οι αλγόριθμοι ομαδοποίησης, K-means και Fast Global K-means, που χρησιμοποιήθηκαν στη εργασία. Στη συνέχεια παρουσιάζεται η μέθοδος δημιουργίας ιστογραμμάτων οπτικών λέξεων (visual words) που είναι μία ομαδοποιημένη και πιο συμπαγής αναπαράσταση πληροφορίας που σχετίζεται με τους περιγραφείς SIFT, CCH και SURF.

Στο 5<sup>ο</sup> κεφάλαιο περιγράφονται τα Κυρτά Μικτά Μοντέλα (Convex Mixture Models (CMM)), στη συνέχεια τα CMM πολλαπλών όψεων που δίνουν την δυνατότητα συνδυασμού όψεων και τέλος τα CMM πολλαπλών όψεων με βάρη που επιτρέπουν την ανάθεση βαρών σε κάθε όψη με σκοπό την καλύτερη δυνατή αξιοποίηση της διαθέσιμης πληροφορίας. Η μέθοδος αυτή χρησιμοποιείται στη συνέχεια για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων.

Στο 6<sup>ο</sup> κεφάλαιο περιγράφονται οι ακολουθίες βίντεο που χρησιμοποιήθηκαν στα πειράματα και το αντίστοιχο ground truth που σχετίζεται με τα key-frames. Επιπλέον, περιγράφεται η διαδικασία εξαγωγής της πληροφορίας από τις ακολουθίες καθώς επίσης και οι παράμετροι που χρησιμοποιήθηκαν για τους περιγραφείς SIFT, CCH και SURF. Κατόπιν, αναφέρονται συνοπτικά οι τρεις συναρτήσεις ομοιότητας που χρησιμοποιήθηκαν και οι τρόποι χρήσης αυτών σε συνδυασμό με τις διαθέσιμες όψεις. Τέλος, αναφέρονται οι παράμετροι εκτέλεσης της μεθόδου ομαδοποίησης με χρήση πολλαπλών όψεων με βάρη για τα πειράματα που εκτελέστηκαν.

Στο 7<sup>ο</sup> κεφάλαιο παρουσιάζονται οι πίνακες των αποτελεσμάτων από τις δύο σειρές πειραμάτων που εκτελέστηκαν. Το κεφάλαιο είναι χωρισμένο σε δύο μέρη, ένα για κάθε σειρά πειραμάτων, στα οποία αναλύονται τα πειραματικά αποτελέσματα που προέκυψαν.

Στο 8<sup>ο</sup> και τελευταίο κεφάλαιο αναφέρονται συνοπτικά τα συμπεράσματα που προέκυψαν από κάθε σειρά πειραμάτων αλλά και συνολικά και στη συνέχεια προτείνονται κάποιες ιδέες για μελλοντική έρευνα.

## ΚΕΦΑΛΑΙΟ 2. ΣΗΜΑΝΤΙΚΑ ΣΤΟΙΧΕΙΑ ΣΧΕΤΙΚΑ ΜΕ ΤΗΝ ΠΕΡΙΛΗΨΗ ΒΙΝΤΕΟ

---

### 2.1 Τεχνικές Περίληψης Βίντεο

### 2.2 Πλήθος των Χαρακτηριστικών Εικονοπλαισίων

### 2.3 Κατηγορίες Μεθόδων για την Εξαγωγή Χαρακτηριστικών Εικονοπλαισίων

---

Στο κεφάλαιο αυτό αναφέρονται κάποια σημαντικά στοιχεία που αφορούν την περίληψη ενός βίντεο. Αρχικά, περιγράφονται οι δύο βασικές τεχνικές που χρησιμοποιούνται για την περίληψη ενός βίντεο, που είναι τα video skims και τα χαρακτηριστικά εικονοπλαίσια (key-frames). Στη συνέχεια αναφέρεται το πλήθος των χαρακτηριστικών εικονοπλαισίων για ένα βίντεο και οι προσεγγίσεις που υπάρχουν για τον καθορισμό του. Τέλος, περιγράφονται συνοπτικά έξι βασικές μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων και αναλύεται περισσότερο η μέθοδος της ομαδοποίησης, που είναι και αυτή που χρησιμοποιήθηκε στην εργασία.

### 2.1. Τεχνικές Περίληψης Βίντεο

Υπάρχουν δύο βασικές τεχνικές περίληψης βίντεο:

- Video Skims
- Χαρακτηριστικά εικονοπλαίσια (key-frames)

Αναλυτικότερα,

**Video Skims:** Αυτός ο τύπος περίληψης προκύπτει από τη σύνδεση τμημάτων του αρχικού βίντεο. Τα τμήματα αυτά συνδέονται, είτε με ένα απότομο κόψιμο είτε με ένα σταδιακό εφέ, με τη σειρά που εμφανίζονται στο αρχικό βίντεο. Μία πολύ γνωστή εφαρμογή τους είναι τα trailer των ταινιών, αν και σε αντίθεση με τα trailer δεν έχουν σκοπό να παροτρύνουν τον χρήστη να παρακολουθήσει ολόκληρο το βίντεο, αλλά αντιθέτως να μην χρειαστεί να παρακολουθήσει ολόκληρο το βίντεο προκειμένου να κερδίσει χρόνο [4].

**Χαρακτηριστικά Εικονοπλαίσια:** Συνήθως λέγονται και key-frames ή στατικές περιλήψεις. Ουσιαστικά είναι μία ακολουθία από στατικές εικόνες, διαφορετικές μεταξύ τους, που περιέχουν όσο δυνατόν περισσότερη πληροφορία σχετικά με το βίντεο.

Ένα σημαντικό πλεονέκτημα που έχουν τα video skims σε σχέση με τα χαρακτηριστικά εικονοπλαίσια είναι ότι τα πρώτα εκτός από εικόνα περιέχουν ήχο καθώς επίσης και στοιχεία κίνησης. Κάτι τέτοιο διευκολύνει την εκφραστικότητα της περίληψης ενώ ταυτόχρονα την κάνει πιο ενδιαφέρουσα. Από την άλλη μεριά, τα χαρακτηριστικά εικονοπλαίσια μπορούν να δώσουν στον χρήστη μία καλύτερη και πιο γρήγορη κατανόηση σχετικά με το περιεχόμενο του βίντεο. Επίσης, η ανάλυση και ανάκτηση των χαρακτηριστικών εικονοπλαισίων απαιτεί πολύ λιγότερους υπολογιστικούς πόρους, κάτι πολύ σημαντικό για real-time εφαρμογές.

## 2.2. Πλήθος των Χαρακτηριστικών Εικονοπλαισίων

Ο προσδιορισμός του ιδανικού αριθμού των χαρακτηριστικών εικονοπλαισίων είναι ένα σημαντικό ζήτημα στην περίληψη ενός βίντεο. Για τον σκοπό αυτό έχουν προταθεί διάφορες εναλλακτικές προσεγγίσεις. Το μέγεθος του συνόλου των χαρακτηριστικών εικονοπλαισίων μπορεί να είναι:

- σταθερό εκ των προτέρων
- απροσδιόριστο

- καθορισμένο κατά τη διάρκεια της διαδικασίας

Αναλυτικότερα [14],

**Σταθερό πλήθος εκ των προτέρων (priori):** Το πλήθος των χαρακτηριστικών εικονοπλαισίων καθορίζεται πριν τη διαδικασία και αποτελεί περιορισμό του αλγορίθμου. Μπορεί να είναι ένας συγκεκριμένος αριθμός ή ένα ποσοστό του μήκους της ακολουθίας του βίντεο.

**Απροσδιόριστο πλήθος (posteriori):** Το πλήθος των χαρακτηριστικών εικονοπλαισίων δεν είναι γνωστό μέχρι την ολοκλήρωση της διαδικασίας. Εξαρτάται κυρίως από το ποσοστό της οπτικής αλλαγής στο πλάνο. Για παράδειγμα ένα βίντεο με πολύ δράση και κίνηση, όπως μία καταδίωξη, απαιτεί περισσότερα key-frames από ένα βίντεο που περιέχει στατικό περιεχόμενο, όπως ένα ηλιοβασίλεμα.

**Καθορισμένο πλήθος:** Ουσιαστικά η μεθοδολογία είναι ίδια με αυτή του απροσδιόριστου πλήθους, με τη διαφορά ότι πρέπει να καθοριστούν οι παράγοντες προσδιορισμού του κατάλληλου αριθμού χαρακτηριστικών εικονοπλαισίων πριν την ολοκλήρωση της διαδικασίας.

### 2.3. Κατηγορίες Μεθόδων για την Εξαγωγή Χαρακτηριστικών Εικονοπλαισίων

Στη συνέχεια αναφέρονται οι βασικές μέθοδοι εξαγωγής χαρακτηριστικών εικονοπλαισίων. Είναι σημαντικό να αναφερθεί ότι δεν υπάρχει κάποια μέθοδος που να θεωρείται καλύτερη σε όλες τις περιπτώσεις. Αντίθετα, η βέλτιστη μέθοδος εξαρτάται από το είδος του βίντεο στο οποίο πρόκειται να εφαρμοστεί. Υπάρχουν μέθοδοι με λίγες υπολογιστικές απαιτήσεις, κάτι που τις κάνει ταχύτερες και ιδανικές για real-time και online εφαρμογές, ενώ άλλες μέθοδοι απαιτούν πολλούς υπολογιστικούς πόρους και χρησιμοποιούνται σε πιο εξειδικευμένες εφαρμογές όπου η ποιότητα των αποτελεσμάτων είναι σημαντική και ο χρόνος επεξεργασίας δεν είναι πολύ περιορισμένος. Επιγραμματικά οι μέθοδοι αυτές είναι:

- Μέθοδος επαρκούς αλλαγής περιεχομένου (Sufficient Content Change)
- Μέθοδος μέγιστης κάλυψης εικονοπλαισίων (Maximum Frame Coverage)



- Μέθοδος ελάχιστης συσχέτισης μεταξύ χαρακτηριστικών εικονοπλαισίων (Minimum Correlation Among Key-frames)
- Μέθοδος με χρήση του σφάλματος ανακατασκευής της ακολουθίας (Shot Reconstruction Degree)
- Μέθοδος των «Ενδιαφερόντων γεγονότων» (Interesting Events)
- Μέθοδος ομαδοποίησης (Clustering)

### 2.3.1. Μέθοδος Επαρκούς Αλλαγής Περιεχομένου (Sufficient Content Change)

Η μέθοδος αυτή εφαρμόζεται επαναληπτικά και για την εφαρμογή της απαιτείται η γνώση της ακολουθίας των εικονοπλαισίων μέχρι το τρέχον σημείο. Κριτήριο για την επιλογή ενός χαρακτηριστικού εικονοπλαΐσιου είναι το πόσο διαφορετικό είναι ένα εικονοπλαΐσιο από το προηγούμενο χαρακτηριστικό εικονοπλαΐσιο που έχει επιλεγεί. Συνήθως το πρώτο εικονοπλαΐσιο είναι και το πρώτο χαρακτηριστικό εικονοπλαΐσιο. Πολύ σημαντική είναι η συνάρτηση ομοιότητας, για την οποία έχουν προταθεί διάφορες εκδοχές με δημοφιλέστερη τη διαφορά ιστογραμμάτων. Οι Zhang et al, [15], χρησιμοποιούν τον παραπάνω τρόπο για την επιλογή των χαρακτηριστικών εικονοπλαισίων, υπολογίζοντας την διαφορά των ιστογραμμάτων χρώματος ανάμεσα στο τρέχον εικονοπλαΐσιο και στο τελευταίο εξαγόμενο. Μία άλλη τεχνική που χρησιμοποιεί τα ιστογράμματα χρώματος είναι αυτή που περιγράφεται στο [29]. Σε αυτήν πρώτα υπολογίζεται μία τιμή ανομοιότητας ανάμεσα στο τρέχον εικονοπλαΐσιο  $k$  και στα  $N$  προηγούμενα, συγκρίνοντας το ιστόγραμμα χρώματος του τρέχοντος εικονοπλαΐσιου και του μέσου όρου των  $N$  προηγούμενων. Σε περίπτωση που η τιμή ανομοιότητας υπερβαίνει ένα προκαθορισμένο κατώφλι, το τρέχον εικονοπλαΐσιο θεωρείται ως key-frame. Ένα σημαντικό πρόβλημα της μεθόδου αυτής είναι ότι τις περισσότερες φορές το πρώτο εικονοπλαΐσιο κάθε πλάνου επιλέγεται ως key-frame, όπως και εικονοπλαΐσια που περιέχουν κάποια ασήμαντη κίνηση ή ενέργεια κάτι το οποίο δεν είναι επιθυμητό. Σε μία άλλη μέθοδο [30], συναθροίζονται οι διαφορές όλων των ιστογραμμάτων από διαδοχικά εικονοπλαΐσια σε κάθε πλάνο και σε ολόκληρη τη σκηνή. Στη συνέχεια, σύμφωνα με το ποσοστό διαφορών του κάθε

πλάνου σε σχέση με τις διαφορές σε ολόκληρη τη σκηνή επιλέγεται ένα υποσύνολο των αρχικών  $N$  εικονοπλαισίων και εφαρμόζεται ένας αλγόριθμος για την κατανομή των εικονοπλαισίων αυτών σε κάθε πλάνο. Μία άλλη τεχνική, από τους Zhao et al [31], βασίζεται σε ιστογράμματα χρώματος, κίνησης ή και σε άλλους περιγραφείς και εκμεταλλεύεται την οπτική δυναμική των εικονοπλαισίων, χωρίς να είναι απαραίτητο τα εικονοπλαίσια να είναι διαδοχικά. Ονόμασαν την τεχνική αυτή *Simplified Breakpoints*. Ένα εικονοπλαίσιο επιλέγεται ως key-frame εάν το ιστόγραμμα του διαφέρει από αυτό του προηγούμενου εικονοπλαισίου περισσότερο από ένα ορισμένο κατώφλι. Η διαδικασία τερματίζει όταν συμπληρωθεί ο αριθμός των απαιτούμενων key-frames. Στην εργασία [32] χρησιμοποιούν έναν αλγόριθμο προσαρμοσμένης δειγματοληψίας για την επιλογή των key-frames. Ο αλγόριθμος αυτός δειγματοληπτεί ομοιόμορφα από τον άξονα της καμπύλης αθροιστικών διαφορών των εικονοπλαισίων. Μία τεχνική που χρησιμοποιεί δύο κατώφλια για την επιλογή των key-frames παρουσιάζεται στο [33]. Το πρώτο κατώφλι,  $r$ , ελέγχει ποια από τα εικονοπλαίσια είναι ικανά για να εισαχθούν στο σύνολο και το δεύτερο κατώφλι,  $N$ , ορίζει το μέγιστο πλήθος των key-frames που μπορεί να έχει η ακολουθία. Η επιλογή των key-frames γίνεται μέσα από μία επαναληπτική διαδικασία τμηματικής ομαδοποίησης. Στο πρώτο επαναληπτικό βήμα της διαδικασίας η ακολουθία χωρίζεται σε διαδοχικές ομάδες ίδιου μεγέθους  $L$ . Για κάθε ομάδα ξεχωριστά υπολογίζεται η διαφορά του πρώτου με το τελευταίο εικονοπλαίσιο. Σε περίπτωση που η διαφορά των δύο εικονοπλαισίων είναι μεγαλύτερη από 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. Μετά την ολοκλήρωση της διαδικασίας για όλες τις ομάδες αν ο αριθμός των εικονοπλαισίων που έχουν εξαχθεί είναι ίσος ή μικρότερος από το  $N$  η διαδικασία σταματάει. Αν είναι μεγαλύτερος, τα επιλεγμένα εικονοπλαίσια χωρίζονται και πάλι σε διαδοχικές ομάδες και εφαρμόζεται η ίδια διαδικασία. Το μεγαλύτερο μειονέκτημα της τεχνικής αυτής είναι ο προσδιορισμός του κατάλληλου κατωφλίου  $r$  καθώς παίζει πολύ σημαντικό ρόλο στην ποιότητα του τελικού αποτελέσματος. Μία εναλλακτική τεχνική που δεν απαιτεί το  $r$  αλλά μόνο το  $N$  προτείνεται στο [34]. Η τεχνική αποτελείται από δύο στάδια. Στο πρώτο στάδιο γίνεται εκτίμηση του πλήθους των key-frames για κάθε πλάνο με βάση τη μεταβλητότητα του πλάνου αλλά και ολόκληρης της ακολουθίας.

Για τον υπολογισμό της μεταβλητότητας της ακολουθίας αθροίζονται οι διαφορές όλων των εικονοπλαισίων της ακολουθίας μεταξύ τους. Η διαδικασία σταματάει όταν το πλήθος των key-frames φτάνει το κατώφλι  $N$ . Σκοπός του δεύτερου σταδίου είναι η επίτευξη μίας καλύτερης κατανομής των key-frames προκειμένου να υπάρχει καλύτερη οπτική ροή ανάμεσα στα πλάνα της ακολουθίας. Το αρνητικό της μεθόδου αυτής είναι ότι προϋποθέτει εκ των προτέρων γνώση σχετικά με το περιεχόμενο του βίντεο, μία πληροφορία που δεν είναι πάντα διαθέσιμη. Τέλος, μία άλλη προσέγγιση, από τους Nara et al στο [35], προτείνει ένα νευρωνικό δίκτυο που χρησιμοποιεί την ένταση της κίνησης που υπολογίζεται σε ένα MPEG βίντεο. Η ένταση της κίνησης ταξινομείται σε πέντε κατηγορίες από ένα ευφρές σύστημα και ως key-frames επιλέγονται τα εικονοπλαίσια που παρουσιάζουν τη μεγαλύτερη ένταση κίνησης.

Η μέθοδος της επαρκούς αλλαγής περιεχομένου είναι σχετικά απλή και απαιτεί χαμηλούς υπολογιστικούς πόρους. Το γεγονός αυτό την καθιστά κατάλληλη για real-time και online εφαρμογές. Είναι σημαντικό να αναφέρουμε πως η απλότητά της οδηγεί πολλές φορές σε ασυμμετρίες, αφού για την εξαγωγή ενός χαρακτηριστικού εικονοπλαισίου λαμβάνονται υπόψη μόνο τα προηγούμενα εικονοπλαίσια της ακολουθίας και όχι το σύνολο των εικονοπλαισίων του βίντεο.

### 2.3.2. Μέθοδος Μέγιστης Κάλυψης Εικονοπλαισίων (Maximum Frame Coverage)

Στο [11] προτείνεται μία μέθοδος η οποία βασίζεται στο πλήθος των εικονοπλαισίων που μπορούν να αναπαρασταθούν από ένα τυχαίο εικονοπλαίσιο. Η μαθηματική συνάρτηση που περιγράφει τη μέθοδο αυτή είναι η εξής:

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

όπου,  $V$  το σύνολο των εικονοπλαισίων του πλάνου,  $r_i$  το  $i$ -οστό χαρακτηριστικό εικονοπλαίσιο του  $V$ ,  $C_i(\varepsilon)$  το σύνολο των εικονοπλαισίων του πλάνου τα οποία μπορούν να χρησιμοποιήσουν το  $i$ -οστό εικονοπλαίσιο της ακολουθίας ως αντιπροσωπευτικό με βάση κάποιο κατώφλι  $\varepsilon$  και  $k$  το πλήθος των χαρακτηριστικών εικονοπλαισίων.

Στην περίπτωση που υπάρχει περιορισμός σχετικά με το πλήθος των χαρακτηριστικών εικονοπλαισίων, τότε το πρόβλημα ανάγεται στην εύρεση της ελάχιστης τιμής  $\varepsilon$ , τέτοιας ώστε όλα τα εικονοπλαίσια της ακολουθίας να μπορούν να αντιπροσωπευτούν από τουλάχιστον ένα χαρακτηριστικό εικονοπλαίσιο. Η νέα σχέση είναι η παρακάτω:

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

Έχοντας τον ίδιο περιορισμό, μία άλλη εκδοχή του προβλήματος θα μπορούσε να είναι η εύρεση του συνόλου χαρακτηριστικών εικονοπλαισίων που μπορούν να αναπαραστήσουν όσο το δυνατόν περισσότερα εικονοπλαίσια:

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

Το πλεονέκτημα αυτής της μεθόδου, σε σχέση με την προηγούμενη, είναι ότι παράγει ένα περισσότερο συνοπτικό σύνολο από χαρακτηριστικά εικονοπλαίσια. Για να το πετύχει όμως αυτό απαιτείται ο υπολογισμός της διαφοράς κάθε ζεύγους εικονοπλαισίων, γεγονός που οδηγεί σε υψηλές υπολογιστικές απαιτήσεις, καθιστώντας έτσι τη μέθοδο ακατάλληλη για real-time και online εφαρμογές.

### 2.3.3. Μέθοδος ελάχιστης συσχέτισης μεταξύ χαρακτηριστικών εικονοπλαισίων (Minimum Correlation Among Key-frames)

Στη μέθοδο αυτή η εξαγωγή των χαρακτηριστικών εικονοπλαισίων βασίζεται στην ύπαρξη της ελάχιστης δυνατής συσχέτισης μεταξύ των εικονοπλαισίων. Έτσι, επιλέγονται τα εικονοπλαίσια που είναι όσο το δυνατόν ανόμοια σε σχέση με τα υπόλοιπα. Η μαθηματική σχέση που περιγράφει τη μέθοδο είναι η παρακάτω:

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

όπου  $Corr$  είναι η τιμή της συσχέτισης ανάμεσα στα εικονοπλαίσια (correlation measure)

### 2.3.4. Μέθοδος με Χρήση του Σφάλματος Ανακατασκευής της Ακολουθίας (*Shot Reconstruction Degree*)

Αυτή η μέθοδος παρουσιάζεται στο [16] και στηρίζεται σε ένα μέτρο που ονομάζεται SRD score (είναι γνωστό και ως shot reconstruction error). Αυτό μετρά την δυνατότητα ανακατασκευής του αρχικού βίντεο χρησιμοποιώντας μόνο τα χαρακτηριστικά εικονοπλαίσια. Έστω ότι έχουμε μία συνάρτηση παρεμβολής μεταξύ των εικονοπλαισίων  $I(t,R)$ , όπου  $R$  το σύνολο των χαρακτηριστικών εικονοπλαισίων,  $V$  το σύνολο όλων των εικονοπλαισίων και  $n=|V|$  το πλήθος των εικονοπλαισίων, η οποία υπολογίζει όλα ή μερικά χαρακτηριστικά της εικόνας τη χρονική στιγμή  $t$  στην ακολουθία του βίντεο. Το SRD score  $E(V,R)$  δίνεται από την σχέση:

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

όπου,  $D$  είναι η διαφορά ανάμεσα σε δύο εικονοπλαίσια. Έτσι δοθέντος του αριθμού των χαρακτηριστικών εικονοπλαισίων  $k$ , ως καλύτερο σύνολο  $r_1, r_2, \dots, r_k$  επιλέγεται εκείνο το οποίο έχει το ελάχιστο SRD score. Μαθηματικά περιγράφεται από τον παρακάτω τύπο:

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

### 2.3.5. Μέθοδος των «Ενδιαφερόντων γεγονότων» (*Interesting Events*)

Η παρακάτω μέθοδος έχει μία σημαντική διαφορά σε σχέση με τις προηγούμενες. Αντί να εστιάζει στο μήκος και στην ισορροπία της οπτικής κάλυψης του συνόλου των χαρακτηριστικών εικονοπλαισίων, προσπαθεί να εντοπίσει τα εικονοπλαίσια που περιέχουν σημαντικά συμβάντα. Για να επιτευχθεί κάτι τέτοιο χρειάζεται μία σχέση ανάμεσα στο πόσο σημαντικό είναι ένα εικονοπλαίσιο και στα πρότυπα κίνησης γύρω από τα εικονοπλαίσια, καθώς και στα χαρακτηριστικά του περιεχομένου του (π.χ. να περιέχει αυτοκίνητα).

Ένα μοντέλο που χρησιμοποιεί την δραστηριότητα κίνησης παρουσιάζεται στο [17]. Στο μοντέλο αυτό, ένα πλάνο χωρίζεται σε επιμέρους τμήματα από διαδοχικά

πρότυπα κίνησης ως προς τις επιταχύνσεις και τις επιβραδύνσεις. Για την εξαγωγή των key-frames εφαρμόζεται το τριγωνικό μοντέλο. Σε αυτό, η κάτω αριστερή κορυφή αντιπροσωπεύει το αρχικό σημείο κίνησης, η κάτω δεξιά το τελικό και η πάνω το σημείο με τη μέγιστη ταχύτητα. Έτσι, τα εικονοπλαίσια που βρίσκονται στις ανώτερες κορυφές του τριγώνου επιλέγονται ως key-frames αφού σε αυτά παρατηρείται η μεγαλύτερη αλλαγή από επιτάχυνση σε επιβράδυνση και επομένως αντιπροσωπεύουν τα εικονοπλαίσια με την μεγαλύτερη δράση. Στη μέθοδο που προτείνεται στο [36], για κάθε πλάνο εξάγεται μόνο ένα key-frame. Για την επιλογή του key-frame όλα τα εικονοπλαίσια κατατάσσονται σε προσκίνητο και παρασκήνιο και επιλέγεται το εικονοπλαίσιο που έχει το μεγαλύτερο λόγο προσκίνητο προς παρασκήνιο, καθώς θεωρείται ότι μεταφέρει την περισσότερη πληροφορία του συγκεκριμένου πλάνου. Στο [37] για την εξαγωγή των key-frames αναλύονται οι διαφορές ανάμεσα σε διαδοχικά εικονοπλαίσια. Η πολυπλοκότητα της ακολουθίας καθορίζεται από τις αλλαγές στο περιεχόμενό της, το οποίο εκφράζεται με διαφορετικούς περιγραφείς εικονοπλαισίων. Αρχικά, η ακολουθία χωρίζεται σε πλάνα τα οποία ταξινομούνται σε δύο κατηγορίες. Τα key-frames εξάγονται μόνο από την πρώτη κατηγορία αφού αυτή δεν περιέχει περιττή πληροφορία όπως για παράδειγμα ένα fade out εφέ. Στη συνέχεια για την καλύτερη αναπαράσταση του περιεχομένου ενός εικονοπλαισίου υπολογίζονται τρεις διαφορετικοί περιγραφείς: ιστόγραμμα χρώματος, ιστόγραμμα κατεύθυνσης ακμής και κατανομή συντελεστών wavelet. Έπειτα, ορίζεται ένα, διαφορετικό σε κάθε περίπτωση, μέτρο ανομοιότητας για κάθε περιγραφέα, ο συνδυασμός των οποίων δημιουργεί το τελικό μέτρο ανομοιότητας. Με τον τρόπο αυτό προκύπτει μία αθροιστική καμπύλη διαφορών η οποία περιγράφει τον τρόπο που αλλάζει το περιεχόμενο των εικονοπλαισίων κατά τη διάρκεια της ακολουθίας. Τέλος, για τον εντοπισμό των key-frames επιλέγονται οι περιοχές της καμπύλης που περιέχουν σημεία που εμφανίζουν τη μεγαλύτερη κλίση. Το ενδιάμεσο εικονοπλαίσιο κάθε τέτοιας περιοχής επιλέγεται ως key-frame.

Οι μέθοδοι αυτοί πετυχαίνουν πολύ καλά αποτελέσματα σε εφαρμογές ειδικού ενδιαφέροντος αλλά αποτυγχάνουν σε εφαρμογές γενικού περιεχομένου κυρίως γιατί οι μέθοδοι αυτοί βασίζονται σε κανόνες οι οποίοι εξάγονται εμπειρικά από συγκεκριμένα και μικρά σύνολα δεδομένων.

### 2.3.6. Μέθοδος Ομαδοποίησης (Clustering)

Η μέθοδος της ομαδοποίησης, που χρησιμοποιήθηκε σε αυτή την εργασία, βασίζεται στην δημιουργία ομάδων που χρησιμοποιούν τα εικονοπλαίσια μίας ακολουθίας βίντεο σαν σημεία στο χώρο των οπτικών χαρακτηριστικών (π.χ. περιγραφείς εικονοπλαισίων κτλ.). Μετά την δημιουργία των ομάδων γίνεται η επιλογή κάποιων «σημαντικών» σημείων από αυτά, τα οποία αντιπροσωπεύουν τα χαρακτηριστικά εικονοπλαίσια για το συγκεκριμένο τμήμα του βίντεο. Η διαδικασία αυτή μπορεί να εφαρμοστεί σε πλάνα ή ακόμα και σε ολόκληρα βίντεο. Συνήθως προϋποθέτει τα παρακάτω στάδια:

**Προεπεξεργασία των δεδομένων:** Πριν από την διαδικασία της δημιουργίας των ομάδων γίνεται μία προεπεξεργασία των δεδομένων. Σκοπός της προεπεξεργασίας αυτής είναι η ταχύτερη δημιουργία των ομάδων καθώς επίσης και η βελτίωση της αποτελεσματικότητας της μεθόδου. Ένας τρόπος προεπεξεργασίας είναι η μείωση των διαστάσεων (dimensionality reduction) του χώρου των χαρακτηριστικών στον οποίο βρίσκονται τα δεδομένα του βίντεο. Αυτό μπορεί να επιτευχθεί εύκολα χρησιμοποιώντας μία από τις γνωστές μεθόδους μείωσης διαστάσεων όπως η PCA, SVD και Factor Analysis.

**Ομαδοποίηση των δεδομένων:** Η ομαδοποίηση των δεδομένων γίνεται μετά την προεπεξεργασία των δεδομένων και μπορούν να χρησιμοποιηθούν διάφορες τεχνικές για τον διαμερισμό των εικονοπλαισίων σε ομάδες.

**Φιλτράρισμα των ομάδων:** Το βήμα αυτό γίνεται προκειμένου να παραληφθούν ορισμένες ομάδες, και τα αντίστοιχα χαρακτηριστικά εικονοπλαίσια, που περιέχουν αρκετό θόρυβο μετά την εφαρμογή των παραπάνω βημάτων.

**Εξαγωγή των αντιπροσωπευτικών εικονοπλαισίων από κάθε ομάδα:** Το τελευταίο βήμα είναι η επιλογή κάποιων αντιπροσωπευτικών σημείων από κάθε ομάδα, που θα είναι και τα χαρακτηριστικά εικονοπλαίσια του βίντεο. Μία συνηθισμένη τεχνική είναι η επιλογή των εικονοπλαισίων που βρίσκονται πιο κοντά στα κέντρα των ομάδων.

Στη βιβλιογραφία υπάρχει ένας αρκετά μεγάλος αριθμός από εργασίες στον τομέα της εξαγωγής χαρακτηριστικών εικονοπλαισίων που βασίζονται στη μέθοδο της ομαδοποίησης. Στο [38] τα εικονοπλαίσια κατανέμονται σε ομάδες. Αν ένα εικονοπλαίσιο είναι αρκετά όμοιο με το κέντρο κάποιας ομάδας τότε τοποθετείται σε αυτή. Αν η ομοιότητα αυτή δεν είναι μεγαλύτερη από ένα καθορισμένο κατώφλι, τότε σχηματίζεται μία νέα ομάδα που περιλαμβάνει το εικονοπλαίσιο αυτό. Το βασικό μειονέκτημα της μεθόδου αυτής είναι ο ακριβής προσδιορισμός των χαρακτηριστικών εικονοπλαισίων ο οποίος εξαρτάται από το κατώφλι που ρυθμίζει την πυκνότητα των ομάδων. Μία παραλλαγή αυτού του αλγορίθμου παρουσιάζεται στο [39] όπου το πλήθος των χαρακτηριστικών εικονοπλαισίων εξαρτάται από ένα κατώφλι που προσδιορίζει εάν δύο εικονοπλαίσια είναι όμοια. Στο [40] οι Hanjalic και Zhang βρίσκουν αυτόματα τον κατάλληλο αριθμό των ομάδων χρησιμοποιώντας μεθόδους ανάλυσης της ποιότητας των ομάδων (cluster validity techniques). Στη συνέχεια το medoid κάθε ομάδας, δηλαδή το εικονοπλαίσιο που έχει την μεγαλύτερη ομοιότητα με τα υπόλοιπα εικονοπλαίσια της ομάδας, επιλέγεται ως χαρακτηριστικό εικονοπλαίσιο. Μία διαφορετική τεχνική για την επιλογή των χαρακτηριστικών εικονοπλαισίων προτείνεται στο [41], όπου χρησιμοποιούνται περιορισμοί στη χρονική θέση των χαρακτηριστικών εικονοπλαισίων κατά τη διαδικασία της ομαδοποίησης. Αρχικά, καθορίζεται ένα σύνολο υποψήφιων εικονοπλαισίων, με βάση την ανομοιότητά τους με τα υπόλοιπα εικονοπλαίσια της ακολουθίας, και δίνεται ως είσοδος σε έναν ιεραρχικό αλγόριθμο ομαδοποίησης. Ένας συνηθισμένος τρόπος σύγκρισης εικονοπλαισίων είναι χρησιμοποιώντας τα ιστογράμματά τους. Αφού ολοκληρωθεί η διαδικασία της ομαδοποίησης, ορισμένες από τις ομάδες φιλτράρονται χρησιμοποιώντας ορισμένα κριτήρια και αφού οι ομάδες γίνουν όσες και το πλήθος των ζητούμενων χαρακτηριστικών εικονοπλαισίων ακολουθεί η επιλογή των κατάλληλων χαρακτηριστικών εικονοπλαισίων από κάθε ομάδα. Συνήθως οι ομάδες περιλαμβάνουν εικονοπλαίσια που είναι διασκορπισμένα σε όλη τη διάρκεια της ακολουθίας. Για το λόγο αυτό, προκειμένου να επιτευχθεί μία καλή περιγραφή του βίντεο, το αντιπροσωπευτικό εικονοπλαίσιο κάθε ομάδας προέρχεται από ένα χρονικό σημείο από το οποίο προέρχονται και τα περισσότερα εικονοπλαίσια της ομάδας του. Για το σκοπό αυτό 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. Μία γρήγορη κριτική για αυτή την τεχνική είναι ότι έχει το πλεονέκτημα να μεγιστοποιεί την χρονική απόσταση ανάμεσα στα χαρακτηριστικά εικονοπλαίσια, με αποτέλεσμα να μην είναι τόσο όμοια μεταξύ τους. Στο [42] παρουσιάζεται μία μέθοδος στην οποία χρησιμοποιούν ένα αλγόριθμο ομαδοποίησης που είναι βασισμένος σε Single Value Decomposition (SVD). Στο πρώτο στάδιο γίνεται δειγματοληψία των εικονοπλαισίων με χρονική σειρά και στην συνέχεια υπολογίζονται τα οπτικά χαρακτηριστικά των εικονοπλαισίων αυτών. Ο χώρος χαρακτηριστικών που εξάγεται από το SVD διαμερίζεται σε ομάδες και εξάγεται ένα χαρακτηριστικό εικονοπλαίσιο από κάθε ομάδα. Τέλος, στο [43] χρησιμοποιώντας μία split-merge προσέγγιση τα εικονοπλαίσια ομαδοποιούνται διαδοχικά με βάση την αξία της πληροφορίας (information value) που περιέχει το κάθε εικονοπλαίσιο. Τα εικονοπλαίσια που μεγιστοποιούν την αμοιβαία πληροφορία για κάθε ομάδα επιλέγονται ως χαρακτηριστικά εικονοπλαίσια.

Η ομαδοποίηση είναι η ευρύτερα χρησιμοποιούμενη μέθοδος για την εξαγωγή χαρακτηριστικών εικονοπλαισίων από ένα βίντεο. Παρόλα αυτά πρέπει να αναφερθεί ότι σε ορισμένες περιπτώσεις δεν είναι αποδοτική, όπως όταν υπάρχει μεγάλη μεταβλητότητα στις ομάδες που έχουν δημιουργηθεί, και τα εξαγόμενα χαρακτηριστικά εικονοπλαίσια δεν διατηρούν την χρονική σειρά του βίντεο. Αυτό έχει σαν αποτέλεσμα η επανεμφάνιση στοιχείων, κατά τη διάρκεια μίας ακολουθίας, να μην θεωρείται σημαντική.

## ΚΕΦΑΛΑΙΟ 3. ΠΕΡΙΓΡΑΦΕΙΣ ΕΙΚΟΝΟΠΛΑΙΣΙΩΝ

---

3.1 Ιστογράμματα Χρώματος

3.2 Περιγραφείς CENTRIST

3.3 Περιγραφείς Wavelet

3.4 Περιγραφείς SIFT

3.5 Περιγραφείς CCH

3.6 Περιγραφείς SURF

---

Σε αυτό το κεφάλαιο παρουσιάζονται αναλυτικά οι περιγραφείς εικονοπλαισίων που χρησιμοποιήθηκαν στην εργασία αυτή για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων.

### **3.1. Ιστογράμματα Χρώματος**

Τα ιστογράμματα χρώματος είναι η αναπαράσταση της κατανομής των χρωμάτων σε μία εικόνα. Το σύνολο των χρωμάτων χωρίζεται σε κάδους (bins) που περιέχουν ένα προκαθορισμένο εύρος χρωμάτων. Ένα ιστόγραμμα χρώματος αναπαριστά την κατανομή των pixel της εικόνας στους κάδους αυτούς. Ένα ιστόγραμμα χρώματος μπορεί να δημιουργηθεί για οποιοδήποτε χώρο χρωμάτων, αλλά συνήθως χρησιμοποιείται στον τρισδιάστατο χώρο όπως είναι το RGB ή HSV. Στην εργασία χρησιμοποιήθηκαν κανονικοποιημένα ιστογράμματα στον χώρο χρώματος HSV. Δηλαδή, για κάθε εικονοπλαίσιο υπολογίζεται ένα κανονικοποιημένο ιστόγραμμα με

8 κάδους για την απόχρωση H (Hue) και από 4 κάδους για κάθε ένα από τα κορεσμός S (Saturation) και αξία V (Value). Τα τρία αυτά ιστογράμματα ενώνονται και σχηματίζουν ένα διάνυσμα διάστασης 16 (8+4+4).

Το βασικότερο μειονέκτημα της μεθόδου αυτής είναι ότι περιγράφει μόνο την κατανομή χρώματος του αντικειμένου που μελετάται και αγνοεί σχήμα και επιφάνεια. Κάτι τέτοιο μπορεί να οδηγήσει σε ακριβώς ίδια ιστογράμματα χρώματος δύο αντικειμένων απλά και μόνο επειδή έχουν ίδια χρώματα. Ένα ακόμα πρόβλημα των ιστογραμμάτων χρώματος είναι η ευαισθησία τους στο θόρυβο, όπως αλλαγές στον φωτισμό.

### 3.2. Περιγραφείς CENTRIST

Είναι ένας πρόσφατος περιγραφέας εικόνας που έχει σκοπό την αναγνώριση τοπολογίας. Για την εφαρμογή του, αρχικά εφαρμόζεται ο μετασχηματισμός Census (CT). Ο μετασχηματισμός αυτός δεν έχει παραμέτρους και συγκρίνει την τιμή ενός pixel με την τιμή των 8 γειτονικών του pixel. Αν η τιμή του είναι μεγαλύτερη ή ίση από αυτή κάποιου γειτονικού τότε η τιμή 1 μπαίνει σε εκείνη τη θέση, αλλιώς μπαίνει η τιμή 0. Στη συνέχεια τα 8 bit που δημιουργήθηκαν ενώνονται με οποιαδήποτε σειρά (από αριστερά προς τα δεξιά ή από πάνω προς τα κάτω) και μετατρέπονται στο δεκαδικό σύστημα. Η τελική τιμή είναι η τιμή του κεντρικού pixel (Σχ. 3.1). Έτσι μπορεί να εξαχθεί το ιστόγραμμα CENTRIST (CENSus TRansform hISTogram).

$$\begin{array}{c|c|c} 32 & 64 & 96 \\ \hline 32 & 64 & 96 \\ \hline 32 & 32 & 96 \end{array} \Rightarrow \begin{array}{ccc} 1 & 1 & 0 \\ 1 & 0 & \\ 1 & 1 & 0 \end{array} \Rightarrow (11010110)_2 \Rightarrow CT = 214$$

Σχήμα 3.1 Μετασχηματισμός Census

Κάποια από τα πλεονεκτήματα του CENTRIST, όπως αναφέρεται στο [1], είναι το ότι δεν χρειάζεται παραμέτρους, υπολογίζεται πάρα πολύ γρήγορα και η υλοποίηση του είναι πολύ εύκολη. Μειονεκτήματά του είναι η ευαισθησία του στην περιστροφή

της εικόνας, στην κλιμάκωσή της εικόνας, στην ακριβή αναγνώριση σχημάτων ενώ αγνοεί τα χρώματα της εικόνας.

### 3.3. Περιγραφείς Wavelet

Τα wavelets, [25], είναι ένα μαθηματικό εργαλείο για την ιεραρχική αποσύνθεση αντικειμένων. Ανεξάρτητα από το τι είναι το αντικείμενο, όπως εικόνα, μία καμπύλη ή μία επιφάνεια, τα wavelets είναι μία τεχνική που περιγράφει τα επίπεδα λεπτομέρειας που υπάρχουν σε αυτό. Παρόλο που τα wavelets αρχικά εφαρμόστηκαν σε προβλήματα πιθανοτικής θεωρίας [23] και επεξεργασίας σήματος [24], στη συνέχεια η εφαρμογή τους επεκτάθηκε και σε άλλους τομείς όπως η επεξεργασία και η συμπίεση εικόνας.

Για καλύτερη κατανόηση, ακολουθεί η περιγραφή της μονοδιάστατης μεθόδου Haar. Έστω μία μονοδιάστατη “εικόνα” με ανάλυση τέσσερα pixel, με τιμές

$$[ 9 \ 7 \ 3 \ 5 ]$$

Η εικόνα αυτή μπορεί να αντικατασταθεί υπολογίζοντας έναν wavelet μετασχηματισμό. Για να γίνει αυτό, αρχικά εξάγεται η μέση τιμή των pixel σε ζευγάρια προκειμένου να σχηματιστεί μία νέα εικόνα με χαμηλότερη ανάλυση και τιμές

$$[ 8 \ 4 ]$$

Προκειμένου να υπάρχει η δυνατότητα ανάκτησης της αρχικής εικόνας, είναι απαραίτητη η αποθήκευση επιπρόσθετης πληροφορίας, που αφορά τους *συντελεστές λεπτομέρειας* (detail coefficients) που περιέχουν την χαμένη πληροφορία. Στην παραπάνω περίπτωση, οι συντελεστές είναι 1 για τα δύο πρώτα pixel, αφού  $8+1=9$  και  $8-1=7$ , και -1 για τα δύο τελευταία pixel, αφού  $4+(-1)=3$  και  $4-(-1)=5$ . Έτσι η αρχική εικόνα έχει αποσυντεθεί σε μία εικόνα χαμηλότερης ανάλυσης (δύο pixel) και σε ένα ζευγάρι συντελεστών. Επαναλαμβάνοντας την διαδικασία επιτυγχάνεται πλήρης αποσύνθεση:

Ανάλυση	Μέση Τιμή	Συντελεστές Λεπτομέρειας
4	[ 9 7 3 5 ]	
2	[ 8 4 ]	[ 1 -1 ]
1	[ 6 ]	[ 2 ]

Σχήμα 3.2 Διαδικασία εφαρμογής μετασχηματισμού Haar

Τελικά, ο wavelet μετασχηματισμός της αρχικής εικόνας με ανάλυση τεσσάρων pixel ορίζεται ως ένας συντελεστής, που περιγράφει την ολική μέση τιμή, ακολουθούμενος από τους συντελεστές λεπτομέρειας ταξινομημένους κατά αύξουσα ανάλυση. Έτσι, για την εικόνα του παραδείγματος, χρησιμοποιώντας τη μέθοδο Haar για μία διάσταση, ο wavelet μετασχηματισμός της αρχικής εικόνας είναι

$$[ 6 \ 2 \ 1 \ -1 ]$$

Αξίζει να σημειωθεί ότι κατά την παραπάνω διαδικασία δεν χάθηκε καθόλου πληροφορία. Επίσης, δοθέντος του μετασχηματισμού είναι δυνατή η κατασκευή της αρχικής εικόνας σε οποιαδήποτε ανάλυση ακολουθώντας την αντίστροφη διαδικασία. Η αποθήκευση της εικόνας με την παραπάνω αναπαράσταση έχει αρκετά πλεονεκτήματα. Ένα από αυτά είναι ότι ένα μεγάλο μέρος των συντελεστών λεπτομέρειας έχει πολύ μικρή τιμή με αποτέλεσμα να απαιτείται λιγότερος αποθηκευτικός χώρος. Επίσης, η απομάκρυνση των πολύ μικρών τιμών από την αναπαράσταση αυτή οδηγεί σε πολύ μικρά σφάλματα κατά τη διαδικασία της κατασκευής της αρχικής εικόνας επιτυγχάνοντας όμως ακόμα μεγαλύτερη συμπίεση.

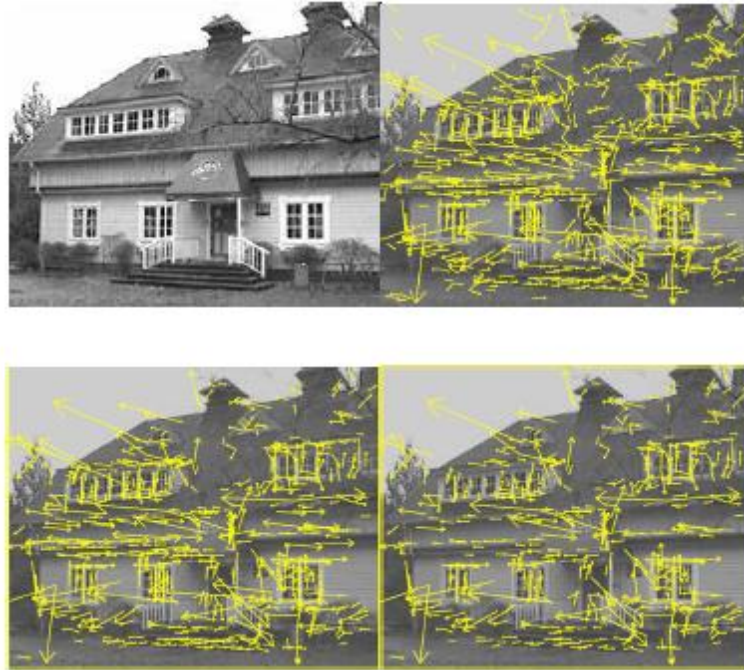
Εκτός όμως από τη συμπίεση της εικόνας, οι συντελεστές λεπτομέρειας περιέχουν πληροφορίες σχετικά με την κλίμακα και την κατεύθυνση της εικόνας, ενώ ταυτόχρονα αποτελούν μία βασική προσέγγιση της εικόνας. Ο συνδυασμός των παραπάνω οδήγησε στην χρήση των wavelets σε εφαρμογές σύγκρισης εικόνων με βάση την υφή (texture). Μία συνηθισμένη προσέγγιση χωρίζει την εικόνα σε

υποζώνες και χρησιμοποιεί τους συντελεστές της κάθε υποζώνης σαν περιγραφείς υφής.

### 3.4. Περιγραφείς SIFT

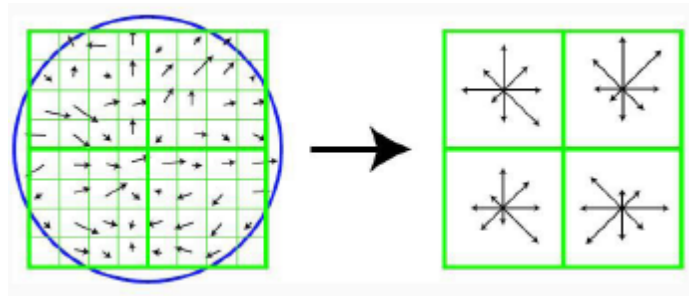
Για την εξαγωγή των περιγραφέων SIFT (Scale-Invariant Feature Transform) [5] απαιτούνται 4 στάδια (Σχ. 3.2):

1. Η εικόνα σαρώνεται σε όλο της το φάσμα με σκοπό να εντοπιστούν σημαντικά σημεία ή σημεία ενδιαφέροντος που παραμένουν αμετάβλητα αν αλλάξει η γωνία θέασης ή η κλίμακα της εικόνας. Αυτό επιτυγχάνεται αποδοτικά με τη χρήση Γκαουσιανών πυραμίδων.
2. Εξετάζονται όλα τα υποψήφια σημαντικά σημεία και απορρίπτονται αυτά που δεν είναι σταθερά. Ένα σημείο θεωρείται ότι δεν είναι σταθερό όταν έχει χαμηλή αντίθεση (γεγονός που το κάνει ευαίσθητο στο θόρυβο).
3. Σε κάθε σημαντικό σημείο ανατίθενται μία ή περισσότερες κατευθύνσεις με βάση την περιοχή γύρω από το σημείο αυτό. Πρέπει να σημειωθεί ότι όλη η περαιτέρω επεξεργασία των σημαντικών σημείων εφαρμόζεται σε δεδομένα επηρεασμένα από την κατεύθυνση, τη θέση και την κλίμακα που έχει το κάθε σημαντικό σημείο ξεχωριστά.
4. Δημιουργείται ένας τοπικός περιγραφέας εικόνας για κάθε σημαντικό σημείο. Ο περιγραφέας αυτός παρουσιάζει μεγάλη ανθεκτικότητα στη παραμόρφωση του σχήματος καθώς επίσης και στις αλλαγές του φωτισμού.



Σχήμα 3.3 Στάδια επιλογής των σημαντικών σημείων. 1) Η αρχική εικόνα 2) Οι αρχικές 832 θέσεις των σημαντικών σημείων στα ελάχιστα και στα μέγιστα των διαφορών της Γκαουσιανής συνάρτησης. Τα σημεία ενδιαφέροντος απεικονίζονται ως διανύσματα που δηλώνουν την κλίμακα, την κατεύθυνση και τη θέση 3) Τα 729 σημαντικά σημεία που απομένουν εφαρμόζοντας ένα κατώφλι ελάχιστης φωτεινότητας 4) Τα τελικά 536 σημαντικά σημεία που απομένουν αφού εφαρμοστεί και ένα επιπλέον κατώφλι στον λόγο της κύριας καμπυλότητας.

Ο τυπικός περιγραφές SIFT ενός σημαντικού σημείου δημιουργείται μέσω της κλιμάκωσης, της μετατόπισης της εικόνας στην περιοχή γύρω από το σημαντικό σημείο και της δημιουργίας κατάλληλων ιστογραμμάτων κατεύθυνσης (Σχ. 3.3). Πιο συγκεκριμένα δημιουργείται ένα 4x4 διάνυσμα από ιστογράμματα με 8 κάδους κατεύθυνσης το καθένα. Τέλος, το τελικό διάνυσμα 128 στοιχείων κανονικοποιείται για να γίνει ανθεκτικότερο στις αλλαγές φωτεινότητας. Για επιπλέον ανθεκτικότητα εφαρμόζεται φίλτρο που δεν επιτρέπει να υπάρχουν στοιχεία με τιμή πάνω από 0.2 και στη συνέχεια γίνεται και πάλι κανονικοποίηση. Πρέπει να αναφερθεί ότι οι περιγραφείς SIFT αποτελούν μία πολύ αποτελεσματική και δημοφιλή προσέγγιση για πολλά προβλήματα υπολογιστικής όρασης.



Σχήμα 3.4 Υπολογισμός ενός περιγραφέα σημαντικού σημείου.

### 3.5. Περιγραφείς CCH

Ένας άλλος τοπικός περιγραφέας εικόνων, γνωστός ως Contrast Context Histogram (CCH) παρουσιάζεται στο [6]. Περιγράφει την κατανομή της αντίθεσης (contrast) σε μία περιοχή γύρω από το σημαντικό σημείο δημιουργώντας έτσι έναν τοπικό περιγραφέα για εκείνη την περιοχή. Έχοντας μία εικόνα  $I$ , αρχικά εφαρμόζεται σε αυτή ένας Γκαουσιανός πυρήνας για να ομαλοποιηθεί και στη συνέχεια κατασκευάζεται μία πολυκλιμακωτή Λαπλασιανή πυραμίδα και εξάγονται τα σημαντικά σημεία ανιχνεύοντας γωνίες Harris [7]. Γύρω από κάθε σημαντικό σημείο  $p_c$  ορίζεται μία περιοχή  $R$  και η αντίθεση του σημείου στην περιοχή αυτή δίνεται από την παρακάτω σχέση:

$$C(p) = I(p) - I(p_c) \quad (3.1)$$

Η περιοχή  $R$  ορίζεται σε ένα κβαντισμένο και λογαριθμικά πολωμένο σύστημα συντεταγμένων  $(r, \theta)$ , όπου  $r_i = 0, \dots, r, r = \lfloor \log \sqrt{2n^2} \rfloor$  και  $\theta_j = \frac{2\pi}{l} m, m = 0, \dots, l - 1$ . Οι παράμετροι  $r$  και  $l$  ορίζουν την απόσταση και την κατεύθυνση του κβαντισμού αντίστοιχα. Για κάθε υπο-περιοχή  $R_{ij} = (r_i, \theta_i)$ , υπολογίζεται ένας θετικός και ένας αρνητικός κάδος για τις τιμές της αντίθεσης. Πιο συγκεκριμένα, για να ένα σημαντικό σημείο  $p_c$  και μία υπο-περιοχή  $R_{ij}$ , το θετικό και το αρνητικό ιστόγραμμα των κάδων ορίζεται από τις σχέσεις 3.2 και 3.3 αντίστοιχα.

$$H_{R_{ij}}^+(p_c) = \frac{\sum \{C(p) | p \in R_{ij} \text{ και } C(p) \geq 0\}}{\#R_{ij}^+}, \quad (3.2)$$

$$H_{R_{ij}}^-(p_c) = \frac{\sum \{C(p) | p \in R_{ij} \text{ και } C(p) < 0\}}{\#R_{ij}^-} \quad (3.3)$$



όπου, τα  $\#R_{ij}^+$  και  $\#R_{ij}^-$  ορίζουν τον αριθμό των θετικών και αρνητικών τιμών αντίθεσης στο  $R_{ij}$ . Συνδυάζοντας τα ιστογράμματα αντίθεσης όλων των υπο-περιοχών σε ένα διάνυσμα, ο περιγραφέας CCH του σημείου  $p_c$  της περιοχής  $R$  ορίζεται ως εξής:

$$CCH_{p_c} = (H_{R_{00}}^+, H_{R_{00}}^-, \dots, H_{R_{rl}}^+, H_{R_{rl}}^-) \quad (3.3)$$

Στην εργασία αυτή χρησιμοποιήθηκε  $r = 4$  και  $l = 8$ , όπως προτείνεται στο [6], έχοντας σαν αποτέλεσμα  $2 \times 4 \times 8 = 64$  διαστάσεις για κάθε τοπικό περιγραφέα CCH.

### 3.6. Περιγραφείς SURF

Οι περιγραφείς SURF (Speeded Up Robust Features) [2] βασίζονται στα ίδια χαρακτηριστικά που βασίζονται οι περιγραφείς SIFT μειώνοντας όμως ακόμα περισσότερο την πολυπλοκότητα επεξεργασίας. Το πρώτο βήμα είναι η ανάθεση μίας κατεύθυνσης για κάθε σημαντικό σημείο, λαμβάνοντας υπόψη την κυκλική περιοχή γύρω από το σημείο αυτό. Πιο συγκεκριμένα, υπολογίζονται τα Haar-wavelets, οριζόντια και κάθετα, γύρω από το σημαντικό σημείο. Αφού υπολογιστούν τα wavelets, σταθμίζονται με μία Gaussian έχοντας ως κέντρο το σημαντικό σημείο. Τα αποτελέσματα αντιπροσωπεύονται από διανύσματα στο χώρο με την οριζόντια τιμή στην τετμημένη και την κάθετη στην τεταγμένη. Η κυρίαρχη κατεύθυνση του σημαντικού σημείου υπολογίζεται αθροίζοντας τα κάθετα και οριζόντια διανύσματα χρησιμοποιώντας ένα κινούμενο παράθυρο (sliding window) καλύπτοντας γωνία  $\pi/3$ . Από το κάθε άθροισμα προκύπτει ένα νέο διάνυσμα. Το μεγαλύτερο από αυτά τα διανύσματα είναι και η κατεύθυνση του σημαντικού σημείου. Στη συνέχεια κατασκευάζεται μία τετράγωνη περιοχή βασισμένη στην κατεύθυνση του σημαντικού σημείου, από την οποία εξάγεται ο περιγραφείς. Το μέγεθος της περιοχής μπορεί να διαφέρει, στην εργασία χρησιμοποιήθηκε ο SURF-128 που χρησιμοποιεί τετράγωνα μεγέθους  $4 \times 4$ . Στα θετικά του SURF είναι η ταχύτητα υπολογισμού και ανθεκτικότητά του σε μικρή περιστροφή της εικόνας, αλλαγή της κλίμακας και στο θόρυβο.

## ΚΕΦΑΛΑΙΟ 4. ΟΜΑΔΟΠΟΙΗΣΗ ΚΑΙ ΣΥΝΟΛΑ ΟΠΤΙΚΩΝ ΛΕΞΕΩΝ (BAG OF VISUAL WORDS)

---

### 4.1 Εισαγωγή

### 4.2 Αλγόριθμος K-means

### 4.3 Αλγόριθμος fast global K-means

### 4.4 Δημιουργία Ιστογράμματος Οπτικών Λέξεων

---

Στο κεφάλαιο αυτό παρουσιάζονται συνοπτικά οι αλγόριθμοι ομαδοποίησης που χρησιμοποιήθηκαν για την επεξεργασία των περιγραφέων SIFT, CCH και SURF προκειμένου να κατασκευαστεί μία ομαδοποιημένη και πιο συμπαγής μορφή δεδομένων στη μορφή οπτικών λέξεων (visual words).

#### **4.1. Εισαγωγή**

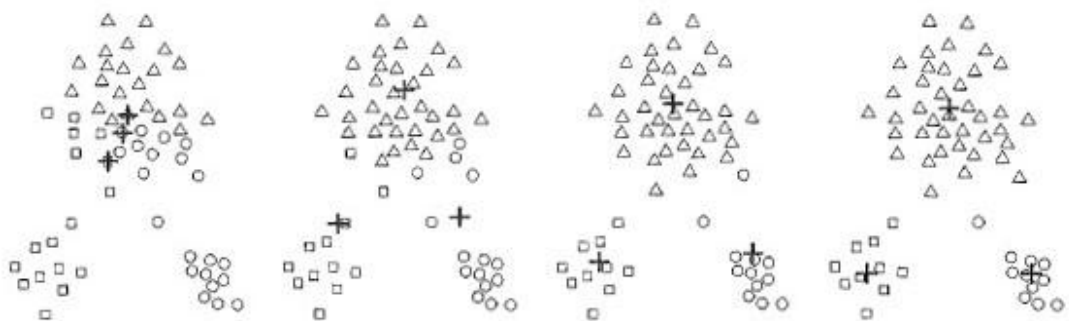
Ένας αλγόριθμος ομαδοποίησης προσπαθεί να βρει ομάδες αντικειμένων (ή δεδομένων) έχοντας κάποιο κριτήριο ομοιότητας. Επιπλέον, ένας αλγόριθμος ομαδοποίησης καθορίζει έναν αντιπρόσωπο για κάθε ομάδα (συνήθως το κέντρο της ομάδας). Προκειμένου να αποφασιστεί σε ποια ομάδα ανήκει κάθε αντικείμενο, οι περισσότεροι αλγόριθμοι υπολογίζουν την απόσταση ανάμεσα σε ένα αντικείμενο και στο κέντρο κάθε υποψήφιας ομάδας. Το τελικό αποτέλεσμα είναι τα κέντρα της κάθε ομάδας και η κατανομή των αντικειμένων σε αυτές. Οι αλγόριθμοι ομαδοποίησης που χρησιμοποιήθηκαν σε αυτή την εργασία είναι ο K-means και ο fast global K-means, οι οποίοι αναλύονται παρακάτω.

## 4.2. Αλγόριθμος K-means

Ο αλγόριθμος ομαδοποίησης K-means είναι αρκετά απλός. Αρχικά, επιλέγουμε  $K$  αρχικά κέντρα, το  $K$  ορίζεται από τον χρήστη και αντιπροσωπεύει το επιθυμητό πλήθος των ομάδων (clusters). Στη συνέχεια κάθε σημείο ανατίθεται στο κοντινότερο του κέντρο, και κάθε σύνολο σημείων που έχει ανατεθεί σε ένα κέντρο αποτελεί και μία ομάδα. Αφού ολοκληρωθεί αυτή η διαδικασία, τα κέντρα κάθε ομάδας επαναπροσδιορίζονται με βάση τα νέα σημεία της κάθε ομάδας. Η διαδικασία της ανάθεσης και του επαναπροσδιορισμού των κέντρων επαναλαμβάνεται μέχρις ότου να μην αλλάζει η ομάδα όλων των σημείων, ή ισοδύναμα, τα κέντρα των ομάδων να παραμένουν σταθερά. Πιο συνοπτικά, ο αλγόριθμος περιγράφεται από τα παρακάτω 4 βήματα:

1. Επέλεξε  $K$  αρχικά κέντρα
2. Ανάθεσε το κάθε σημείο στο κοντινότερο κέντρο δημιουργώντας  $K$  ομάδες
3. Επαναπροσδιόρισε τα κέντρα κάθε ομάδας
4. Επανάλαβε τα βήματα 2 και 3 μέχρις ότου να μην αλλάζουν τα κέντρα

Στο παρακάτω σχήμα (Σχ. 4.1) φαίνεται πώς συμπεριφέρεται ο αλγόριθμος K-means σε 4 επαναλήψεις.



Σχήμα 4.1 Εφαρμογή του αλγόριθμου K-means από την αρχική κατάσταση μέχρι την τελική.

Για την μέτρηση της ποιότητας της ομαδοποίησης χρησιμοποιείται το sum of squared errors (SSE), γνωστό και ως σφάλμα ομαδοποίησης (clustering error). Με άλλα λόγια, υπολογίζεται το σφάλμα κάθε σημείου, δηλαδή η απόστασή του από το κέντρο της ομάδας του, και στη συνέχεια υπολογίζεται το άθροισμα των τετραγώνων όλων των σφαλμάτων. Έτσι, αν υπάρχουν διαθέσιμες δύο διαφορετικές ομαδοποιήσεις που προέκυψαν από δύο διαφορετικές εκτελέσεις του K-means, προτιμάται αυτή με το χαμηλότερο SSE.

Στα θετικά του K-means είναι η απλότητα και η ευρεία χρήση του. Επίσης είναι αρκετά αποδοτικός, αν και πολλές φορές χρειάζονται πολλαπλές εκτελέσεις από διαφορετικές θέσεις των κέντρων. Αρκετές παραλλαγές του είναι ακόμα πιο αποδοτικές και λιγότερο εξαρτώμενες από την αρχικοποίηση των κέντρων. Στα αρνητικά είναι το ότι το αποτέλεσμα του εξαρτάται πολύ από την αρχικοποίηση των κέντρων και γι' αυτό συνήθως απαιτούνται πολλές εκτελέσεις του αλγόριθμου. Επίσης, δεν μπορεί να διαχειριστεί επιτυχώς μη-σφαιρικές ομάδες ή ομάδες με διαφορετικό μέγεθος και πυκνότητα. Επηρεάζεται επίσης σημαντικά από outliers, και η χρήση μεθόδων απαλοιφής outliers μπορεί να αυξήσει την απόδοσή σου.

### 4.3. Αλγόριθμος Fast Global K-means

Ο αλγόριθμος global K-means, [26], είναι ένας πολύ αποδοτικός αλγόριθμος ομαδοποίησης βασισμένος στην ελαχιστοποίηση του σφάλματος ομαδοποίησης, χωρίς αρχικοποίηση παραμέτρων και χρησιμοποιεί τον αλγόριθμο K-means σαν εσωτερική διαδικασία. Σε αντιπαράθεση με τον K-means, αντί να επιλέγει τυχαία αρχικά κέντρα, λειτουργεί αυξητικά προσπαθώντας να εισάγει βέλτιστα μία νέα ομάδα σε κάθε στάδιο. Για την επίλυση ενός προβλήματος ομαδοποίησης με  $K$  ομάδες ο αλγόριθμος global k-means ακολουθεί την παρακάτω διαδικασία:

1. Αρχικά υπάρχει μία ομάδα, με βέλτιστο κέντρο το κέντρο όλων των δεδομένων
2. Για να προστεθεί μία νέα ομάδα  $k$ :

2.1. Τα  $k-1$  κέντρα αρχικοποιούνται στις γνωστές βέλτιστες τιμές τους

2.2. Εκτελείται ο αλγόριθμος ομαδοποίησης K-means  $N$  φορές, όπου  $N$  είναι το πλήθος των δεδομένων, εξετάζοντας έτσι όλα τα πιθανά κέντρα της  $k$  ομάδας

2.3. Η καλύτερη λύση μετά από  $N$  εκτελέσεις του αλγορίθμου K-means είναι και η αποδεκτή λύση για το πρόβλημα ομαδοποίησης με  $k$  ομάδες

3. Εκτελούνται επαναληπτικά τα βήματα 2.1 έως 2.3 για  $k=2, \dots, K$

Προκειμένου να μειωθεί το υπολογιστικό κόστος του global K-means, χωρίς μεγάλη μείωση στην απόδοση της ομαδοποίησης, χρησιμοποιήθηκε ο fast global K-means [26]. Η μέθοδος αυτή υπολογίζει ένα άνω όριο  $E_n$  του τελικού σφάλματος ομαδοποίησης που προκύπτει από την εισαγωγή μίας νέας ομάδας σε ένα σημείο  $x_n$ . Η αρχική θέση του κέντρου της νέας ομάδας είναι το σημείο που ελαχιστοποιεί το  $E_n$  και ο αλγόριθμος K-means εκτελείται μόνο μία φορά για  $k$ . Επομένως, στην μέθοδο fast global K-means ο αλγόριθμος K-means εκτελείται μόνο  $K$  φορές, όπου  $K$  είναι το πλήθος των τελικών ομάδων. Ένα σημαντικό πλεονέκτημα της παραπάνω διαδικασίας είναι ότι παρέχει τις βέλτιστες λύσεις για όλα τα προβλήματα ομαδοποίησης με  $k \leq K$ .

#### 4.4. Δημιουργία Ιστογράμματος Οπτικών Λέξεων

Σε ορισμένες περιπτώσεις τα βασικά χαρακτηριστικά (low-level features) μίας εικόνας, όπως το χρώμα, δεν είναι ικανά να αποδώσουν επιτυχώς την πληροφορία ενός πλάνου. Για παράδειγμα, έστω ότι μία κάμερα παρακολουθεί την κίνηση ενός μαύρου αυτοκινήτου και κάποια στιγμή παύει να παρακολουθεί το αυτοκίνητο και στρέφεται σε μία μαύρη μοτοσυκλέτα. Εξετάζοντας την ακολουθία εικονοπλασίων που προκύπτει από το παραπάνω παράδειγμα και λαμβάνοντας υπόψη μόνο τη χρωματική πληροφορία, ο αλγόριθμος επικεντρώνεται στις αλλαγές του τοπίου γύρω από το αντικείμενο παρακολούθησης και πολύ λιγότερο στην αλλαγή του αυτοκινήτου με την μοτοσυκλέτα, αφού χρωματικά δεν έχουν σημαντική διαφορά. Το φαινόμενο αυτό οδηγεί σε απώλεια της σημασίας του πλάνου, καθώς δεν θεωρείται

σημαντική η αντικατάσταση ενός αντικειμένου παρακολούθησης με ένα άλλο. Προκειμένου να αντιμετωπιστεί το πρόβλημα αυτό μπορούν να χρησιμοποιηθούν τοπικοί περιγραφείς, όπως CCH, SIFT και SURF, οι οποίοι περιγράφουν επαρκώς τέτοια αντικείμενα ενδιαφέροντος χωρίς να επηρεάζονται από αλλαγές της γωνίας θέασης και της κλίμακάς τους. Οι περιγραφείς αυτοί μπορούν να ομαδοποιηθούν σε ένα μεγάλο αριθμό ομάδων. Κάθε ομάδα αντιμετωπίζεται ως μία οπτική λέξη και αντιπροσωπεύει ένα πρότυπο που μοιράζονται όλοι οι περιγραφείς της ομάδας. Αναθέτοντας κάθε περιγραφέα σε μία οπτική λέξη, υιοθετούμε την αναπαράσταση με χρήση συνόλου λέξεων (bag of words), η οποία στον χώρο της ανάλυσης εικόνας και βίντεο είναι γνωστή ως σύνολο οπτικών λέξεων (bag of visual words), [27]. Με τον τρόπο αυτό, είναι δυνατή η εξαγωγή και χρήση της σημασιολογικής πληροφορίας που περιέχει μία ακολουθία εικονοπλαισίων.

Η διαδικασία εξαγωγής οπτικών λέξεων εκτελείται για κάθε εικονοπλαίσιο στα οποία υπολογίζεται ένας διαφορετικός αριθμός από περιγραφείς. Πιο συγκεκριμένα για κάθε πλάνο που έχει  $n$  εικονοπλαίσια, εξάγεται για κάθε εικονοπλαίσιο ένα σύνολο  $D_n$  που περιέχει τους περιγραφείς του εικονοπλαισίου αυτού, ανάλογα με τον αλγόριθμο που χρησιμοποιείται (SIFT, CCH ή SURF). Στη συνέχεια, όλοι οι περιγραφείς συνενώνονται,  $D_s = D_1 U D_2 U \dots U D_n$ , προκειμένου να οριστεί το διάνυσμα περιγραφέων για ολόκληρο το πλάνο. Για την δημιουργία του ιστογράμματος οπτικών λέξεων απαιτούνται τρία βήματα.

**Προ-επεξεργασία περιγραφέων:** Προκειμένου να μειωθεί ο υπολογιστικός χρόνος της διαδικασίας εξαγωγής των οπτικών λέξεων, οι περιγραφείς του κάθε εικονοπλαισίου ομαδοποιούνται σε  $K$  ομάδες, στα πειράματά μας το  $K$  είναι το 10% του πλήθους των εξαχθέντων περιγραφέων των εικονοπλαισίων και χρησιμοποιείται ο αλγόριθμος fast global K-means για όσο το δυνατόν καλύτερη απόδοση και ταχύτητα. Έτσι, κάθε εικονοπλαίσιο αντιπροσωπεύεται από τα κέντρα των περιγραφέων  $D_n^K$ .

**Δημιουργία οπτικών λέξεων:** Στη συνέχεια, για την εξαγωγή των οπτικών λέξεων το σύνολο των κέντρων των περιγραφέων  $D = D_1^{K1} U D_2^{K2} U \dots U D_n^K$  ομαδοποιείται σε  $W$

ομάδες με τον αλγόριθμο K-means, όπου  $W$  είναι το πλήθος των οπτικών λέξεων (visual words).

**Δημιουργία ιστογράμματος οπτικών λέξεων:** Τέλος, για να δημιουργηθεί το ιστόγραμμα των οπτικών λέξεων κάθε ένας από τους αρχικούς περιγραφείς, κάθε εικονοπλαίσιο, ανατίθεται σε μία από τις  $W$  οπτικές λέξεις. Αυτό έχει ως αποτέλεσμα την δημιουργία ενός διανύσματος που περιέχει την συχνότητα της κάθε οπτικής λέξης μέσα σε ένα εικονοπλαίσιο, δηλαδή δημιουργείται το ιστόγραμμα οπτικών λέξεων του κάθε εικονοπλαισίου.

Για την ανάθεση των περιγραφέων στις οπτικές λέξεις έχουν προταθεί διάφορες μέθοδοι. Στις περισσότερες από αυτές χρησιμοποιείται η συχνότητα εμφάνισης του κάθε όρου (term frequency, tf) και η αντίστροφη συχνότητα (inverse document frequency, idf). Στο [8], υιοθετείται ο συνδυασμός tf-idf ενώ στις περισσότερες περιπτώσεις χρησιμοποιείται απευθείας tf [9, 10]. Στο [18], έχει χρησιμοποιηθεί η μέθοδος binary weighting η οποία επισημαίνει την ύπαρξη ή όχι μίας οπτικής λέξης με την τιμή 1 και 0 αντίστοιχα. Τέλος, στο [19], προτείνεται η μέθοδος soft-weighting, που είναι και η μέθοδος που χρησιμοποιήθηκε στην εργασία αυτή. Σε αυτή, για κάθε σημαντικό σημείο ενός εικονοπλαισίου, αντί να εντοπίζεται μόνο η κοντινότερη οπτική λέξη, επιλέγονται οι  $N$ -κοντινότερες οπτικές λέξεις.

Έστω ότι υπάρχει ένα οπτικό λεξικό με  $K$  οπτικές λέξεις, χρησιμοποιείται ένα διάνυσμα  $K$  διαστάσεων  $T=[t_1, \dots, t_k, \dots, t_K]$  όπου το κάθε στοιχείο  $t_k$  αντιπροσωπεύει το βάρος της  $k$  οπτικής λέξης σε ένα εικονοπλαίσιο τέτοιο ώστε

$$t_k = \sum_{i=1}^N \sum_{j \in M_i^k} \frac{1}{2^{i-1}} sim(j, k) \quad (4.1)$$

όπου  $M_i^k$  είναι το σύνολο των σημαντικών σημείων των οποίων ο  $i$ -οστός κοντινότερος γείτονας είναι η οπτική λέξη  $k$ . Η σχέση  $sim(j, k)$  αντιπροσωπεύει την κανονικοποιημένη ομοιότητα μεταξύ του σημαντικού σημείου  $j$  και της οπτικής λέξης  $k$ . Όπως αναφέρεται στο [19], μία τυπική τιμή για το  $N$  είναι το 4.

## ΚΕΦΑΛΑΙΟ 5. ΚΥΡΤΑ ΜΙΚΤΑ ΜΟΝΤΕΛΑ ΜΕ ΒΑΡΗ (WEIGHTED MULTIVIEW CONVEX MIXTURE MODELS)

---

### 5.1 Εισαγωγή

### 5.2 Κυρτά Μικτά Μοντέλα (Convex Mixture Models)

### 5.3 CMMs Πολλαπλών Όψεων (Multiple View CMMs)

### 5.4 Μοντέλα Προτύπων Πολλαπλών Όψεων με Βάρη (Exemplar-Based Weighted Multiple View Models)

---

#### 5.1. Εισαγωγή

Η ομαδοποίηση πολλαπλών όψεων (multi-view) εκμεταλλεύεται πολλαπλές αναπαραστάσεις ταυτόχρονα, προκειμένου να ομαδοποιήσει αποτελεσματικότερα τα δεδομένα σε σχέση με τη χρήση μόνο μίας αναπαράστασης (single-view). Έτσι, με την χρήση πολλών αναπαραστάσεων υπάρχει περισσότερη διαθέσιμη πληροφορία η οποία αναμένεται να οδηγήσει σε καλύτερη ομαδοποίηση των δεδομένων. Για την ομαδοποίηση πολλαπλών όψεων υπάρχουν δύο προσεγγίσεις:

- *Centralized*, όπου χρησιμοποιούνται ταυτόχρονα όλες οι διαθέσιμες αναπαραστάσεις για την ομαδοποίηση των δεδομένων, και
- *Distributed*, όπου ομαδοποιούνται τα δεδομένα χρησιμοποιώντας κάθε αναπαράσταση ξεχωριστά, με κάποιον single-view αλγόριθμο, και στη



συνέχεια τα αποτελέσματα συνδυάζονται προκειμένου να παραχθεί η τελική ομαδοποίηση.

Οι περισσότεροι αλγόριθμοι πολλαπλών όψεων χρησιμοποιούν εξίσου κάθε όψη για την ομαδοποίηση των δεδομένων, αδιαφορώντας για την ποιότητα της πληροφορίας που περιέχει κάθε όψη. Για παράδειγμα, μία όψη μπορεί να περιέχει θόρυβο, outliers ή η πληροφορία που περιέχει να μην σχετίζεται με το συγκεκριμένο κριτήριο ομαδοποίησης. Έτσι, η συμπερίληψη μίας τέτοιας όψης μπορεί να οδηγήσει σε μείωση της επίδοσης. Επομένως, η αναγνώριση και αφαίρεση τέτοιων όψεων είναι μία αρκετά σημαντική διαδικασία. Αυτό μπορεί να επιτευχθεί με την χρήση μίας centralized μεθόδου ομαδοποίησης η οποία αναθέτει διαφορετικά βάρη σε κάθε όψη και επιλέγει τα κατάλληλα βάρη αυτόματα. Τα βάρη αυτά, αντιπροσωπεύουν την ποιότητα της πληροφορίας που περιέχει η κάθε όψη και ανάλογα με αυτήν επηρεάζεται η συμμετοχή της κάθε όψης στην τελική ομαδοποίηση.

## 5.2. Κυρτά Μικτά Μοντέλα (Convex Mixture Models)

Τα CMMs [20] είναι απλοποιημένα μικτά μοντέλα που έχουν σαν σκοπό την ανάθεση των δεδομένων σε ομάδες και την εξαγωγή αντιπροσωπευτικών κέντρων από τα δεδομένα. Όταν εκπαιδεύονται τα μοντέλα αυτά, κάτι που γίνεται μεγιστοποιώντας την τιμή της πιθανοφάνειας (log-likelihood), όλα τα δεδομένα ανταγωνίζονται μεταξύ τους προκειμένου να επιλεγούν ως κέντρα των ομάδων. Στο τέλος, τα αντικείμενα που έχουν λάβει τις μεγαλύτερες *prior* πιθανότητες επιλέγονται ως πρότυπα (exemplars) και τα υπόλοιπα αντικείμενα ανατίθενται στο “κοντινότερο” σε αυτά πρότυπο. Αξίζει να σημειωθεί ότι οι τιμές *prior* των αντικειμένων είναι οι μοναδικές παράμετροι ενός Κυρτού Μικτού Μοντέλου.

Πιο συγκεκριμένα, δοθέντος ενός συνόλου δεδομένων  $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$ ,  $x_i \in \mathcal{R}^d$ , η κατανομή του CMM είναι  $Q(x) = \sum_{j=1}^N q_j f_j(x)$ ,  $x \in \mathcal{R}^d$  και όπου  $q_j \geq 0$  είναι η *prior* πιθανότητα του  $j$ -οστού αντικειμένου, ικανοποιώντας τον περιορισμό  $\sum_{j=1}^N q_j = 1$ , και  $f_j(x)$  είναι μία οικογένεια εκθετικής κατανομής. Πιο συγκεκριμένα,  $f_j(x) = C_\varphi(x) e^{-\beta d_\varphi(x, x_j)}$  με  $d_\varphi$  να είναι η απόσταση Bregman που αντιστοιχεί στην κατανομή

των αντικειμένων,  $C_\varphi(x)$  ανεξάρτητο από το  $x_j$ , και  $\beta$  μία σταθερά που καθορίζει τον τελικό αριθμό των exemplars [20].

Η ομαδοποίηση προκύπτει μεγιστοποιώντας την πιθανοφάνεια (log-likelihood) του συνόλου δεδομένων. Η πιθανοφάνεια ορίζεται στην παρακάτω σχέση:

$$\begin{aligned} L(\mathfrak{X}; \{q_j\}_{j=1}^N) &= \frac{1}{N} \sum_{i=1}^N \log \left[ \sum_{j=1}^N q_j f_j(x_i) \right] \\ &= \frac{1}{N} \sum_{i=1}^N \log \left[ \sum_{j=1}^N q_j e^{-\beta d_\varphi(x, x_j)} \right] + \text{σταθ} \quad (5.1) \end{aligned}$$

Αν ορίσουμε  $\hat{P}(x) = \frac{1}{N}$ ,  $x \in \mathfrak{X}$  να είναι η κατανομή των δεδομένων, μπορούμε να μετατρέψουμε το παραπάνω πρόβλημα μεγιστοποίησης πιθανοφάνειας σε μορφή Kullback-Leibler (KL) διαφοράς μεταξύ  $\hat{P}(x)$  και  $Q(x)$ , από τη στιγμή που η KL απόστασή τους είναι:

$$D(\hat{P}||Q) = - \sum_{i=1}^N \hat{P}(x_i) \log Q(x_i) - H(\hat{P}) = -L(\mathfrak{X}; \{q_j\}_{j=1}^N) + \text{σταθ}. \quad (5.2)$$

όπου  $H(\hat{P})$  είναι η εντροπία της κατανομής  $\hat{P}(x_i)$  η οποία δεν επηρεάζεται από τις παραμέτρους  $q_j$  του CMM. Επομένως, αντί να μεγιστοποιηθεί η σχέση 5.1, ελαχιστοποιείται η σχέση 5.2. Αυτό το κυρτό πρόβλημα ελαχιστοποίησης μπορεί να λυθεί χρησιμοποιώντας έναν επαναληπτικό αλγόριθμο. Οι *prior* πιθανότητες ενημερώνονται από την παρακάτω σχέση,

$$q_j^{(t+1)} = q_j^{(t)} \frac{\sum_{i=1}^N \hat{P}(x_i) f_j(x_i)}{\sum_{j'=1}^N q_{j'}^{(t)} f_{j'}(x_i)} \quad (5.3)$$

και ο αλγόριθμος εγγυάται ότι θα συγκλίνει στο ολικό ελάχιστο όσο  $q_j^{(0)} > 0, \forall j$  [20]. Όπως αναφέρθηκε και προηγουμένως, η *prior* πιθανότητα  $q_j$ , η οποία αναφέρεται στο σημείο  $j$ , περιγράφει πόσο πιθανό είναι το σημείο  $j$  να επιλεγεί ως πρότυπο (exemplar). Η δυνατότητα εντοπισμού του ολικού βέλτιστου, σε όλες τις

περιπτώσεις, κάνει το μοντέλο αυτό πολύ ελκυστικό καθώς αποφεύγεται το ζήτημα της αρχικοποίησης και το πρόβλημα των τοπικών ελαχίστων που έχουν τα τυπικά μικτά μοντέλα με αποτέλεσμα να απαιτούνται πολλαπλές εκτελέσεις του αλγορίθμου EM [45]. Επιπροσθέτως, στο [20] παρατίθενται πειραματικές ενδείξεις ότι ένα Gaussian CMM αποδίδει καλύτερα από ένα πλήρως παραμετροποιημένο Gaussian μικτό μοντέλο, παρόλο που το CMM έχει μικρότερη ευελιξία αφού οι μόνοι παράμετροι είναι *prior* τιμές  $q_j$ . Ένα ακόμα σημαντικό χαρακτηριστικό είναι ότι στους υπολογισμούς των  $q_j$  απαιτούνται μόνο οι αποστάσεις μεταξύ των σημείων  $d_\varphi(x_i, x_j)$ , εφόσον το  $C_\varphi(x_i)$  απαλείφεται, με αποτέλεσμα να μην απαιτούνται οι τιμές των δεδομένων σε περίπτωση που είναι διαθέσιμες μόνο οι αποστάσεις τους.

Ο διαχωρισμός των δεδομένων σε  $M$  ομάδες γίνεται χρησιμοποιώντας σαν κέντρα (exemplars) τα  $M$  αντικείμενα με τις μεγαλύτερες τιμές  $q_j$  και τα υπόλοιπα δεδομένα ανατίθενται στα exemplars με την μεγαλύτερη εκ των υστέρων (posterior) πιθανότητα.

Τέλος, η ομαδοποίηση χρησιμοποιώντας CMM απαιτεί την επιλογή μίας αποτελεσματικής τιμής για την σταθερά  $\beta$  ( $0 < \beta < \infty$ ). Στο [12], υιοθετείται η παρακάτω σχέση 5.4, η οποία οδηγεί σε καλά αποτελέσματα:

$$\beta_0 = N^2 \log N / \sum_{i,j=1}^N d_\varphi(x_i, x_j) \quad (5.4)$$

### 5.3. CMMs Πολλαπλών Όψεων (Multiple View CMMs)

Τα CMM πολλαπλών όψεων [21] υποθέτουν ότι τα στοιχεία της κάθε όψης έχουν δημιουργηθεί από κάποιο CMM και στόχος τους είναι να εντοπίσουν τα αντιπροσωπευτικά κέντρα λαμβάνοντας υπόψη τους όλες τις διαθέσιμες όψεις. Πιο συγκεκριμένα, έστω ότι δίνεται ένα σύνολο με  $N$  αντικείμενα και  $V$  περιγραφές (όψεις),  $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$ , όπου  $x_i$  είναι η αναπαράσταση του  $i$ -οστού αντικειμένου για όλες τις όψεις, δηλαδή  $x_i = \{x_i^1, x_i^2, \dots, x_i^V\}$ ,  $x_i^v \in \mathcal{R}^{d^v}$ . Το CMM της κάθε όψης  $V$  είναι:

$$Q^v(x^v) = \sum_{j=1}^N q_j f_j^v(x^v) = C_{\varphi_v}(x^v) \sum_{j=1}^N q_j e^{-\beta^v d_{\varphi_v}(x^v, x_j^v)}, x^v \in \mathcal{R}^{d^v} \quad (5.5)$$

Επειδή οι ιδιότητες κάθε μεμονωμένης όψης μπορεί να διαφέρουν σημαντικά, διαφορετικές όψεις μπορούν να έχουν διαφορετικές κατανομές  $f_j^v(x^v)$ . Από την άλλη, προκειμένου να συμμετέχουν όλες οι όψεις στην ομαδοποίηση χρησιμοποιούνται οι ίδιες *prior* πιθανότητες  $q_j$ .

Επιπλέον, μία κατανομή  $\hat{P}^v(x^v) = \frac{1}{N}, x^v \in \{x_1^v, x_2^v, \dots, x_N^v\}$ , σχετίζεται με κάθε όψη και ο αλγόριθμος πολλαπλών όψεων ελαχιστοποιεί το άθροισμα της KL (5.6) απόστασης ανάμεσα σε  $\hat{P}^v(x^v)$  και  $Q^v(x^v)$  για όλες τις όψεις.

$$\min_{q_1, \dots, q_N} \left\{ \sum_{v=1}^V D(\hat{P}^v || Q^v) \right\} = \min_{q_1, \dots, q_N} \left\{ - \sum_{v=1}^V \sum_{i=1}^N \hat{P}^v(x_i^v) \log Q^v(x_i^v) - \sum_{v=1}^V H(\hat{P}^v) \right\} \quad (5.6)$$

με τον περιορισμό  $\sum_{j=1}^N q_j = 1$ .

Για την ενημέρωση των *prior* πιθανοτήτων χρησιμοποιείται μία γενίκευση της εξίσωσης (5.3) όπου χρησιμοποιείται μία όψη και δίνεται από την παρακάτω σχέση:

$$q_j^{(t+1)} = \frac{q_j^{(t)}}{V} \sum_{v=1}^V \sum_{i=1}^N \frac{\hat{P}^v(x_i^v) f_j^v(x_i^v)}{\sum_{j'=1}^N q_{j'}^{(t)} f_{j'}^v(x_i^v)} \quad (5.7)$$

Όπως και στην περίπτωση της μίας όψης, απαιτούνται μόνο οι αποστάσεις μεταξύ των αντικειμένων για κάθε όψη και η τιμή  $C_{\varphi_v}(x_i^v)$  απαλείφεται.

Η *prior* πιθανότητα  $q_j$  που σχετίζεται με το  $j$ -οστό αντικείμενο αντιπροσωπεύει την πιθανότητα να επιλεγεί το αντικείμενο  $j$  ως exemplar λαμβάνοντας υπόψη όλες τις διαθέσιμες όψεις.

Ο διαχωρισμός των δεδομένων σε  $M$  ομάδες γίνεται χρησιμοποιώντας σαν κέντρα (exemplars) τα  $M$  αντικείμενα με τις μεγαλύτερες  $q_j$  τιμές και τα υπόλοιπα δεδομένα ανατίθενται στα exemplars με την μεγαλύτερη εκ των υστέρων (posterior) πιθανότητα. Για την επιλογή των τιμών  $\beta^v$  χρησιμοποιείται η σχέση 5.4 ως εξής:

$$\beta_0^v = N^2 \log N / \sum_{i,j=1}^N d_{\varphi_v}(x_i^v, x_j^v) \quad (5.8)$$

#### 5.4. Μοντέλα Προτύπων Πολλαπλών Όψεων με Βάρη (Exemplar-Based Weighted Multiple View Models)

Λόγω της καλής επίδοσης των CMMs στο [10], στην εργασία προτείνεται μία παραλλαγμένη centralized μέθοδος ομαδοποίησης πολλαπλών όψεων, στην οποία αποδίδονται βάρη σε κάθε όψη τα οποία καθορίζονται αυτόματα κατά τη διάρκεια της εκτέλεσης.

##### 5.4.1. Περιγραφή του Μοντέλου

Από την σχέση 5.6 παρατηρείται ότι όλες οι όψεις συμμετέχουν το ίδιο, ανεξάρτητα από το πόσο σχετίζονται με το πρόβλημα προς επίλυση. Σκοπός της μεθόδου είναι να εντοπίσει τα exemplars του συνόλου δεδομένων επιτρέποντας στις όψεις να συμμετέχουν με διαφορετικά βάρη η κάθε μία, ανάλογα με την ποιότητα της πληροφορίας, και να επιλέγει τα βάρη αυτά αυτόματα και κατά τη διάρκεια της εκτέλεσης. Αυτή η τεχνική γενικεύει την μέθοδο πολλαπλών όψεων και αποδεικνύεται χρήσιμη σε περιπτώσεις που υπάρχουν όψεις που περιέχουν πληροφορίες που δεν σχετίζονται με το πρόβλημα ομαδοποίησης ή περιέχουν θόρυβο.

Για να επιτευχθεί ο παραπάνω στόχος, προτείνεται μία μέθοδος ομαδοποίησης πολλαπλών όψεων με βάρη. Για την  $v$ -οστή όψη ορίζεται ένα CMM  $Q^v(x^v)$ , ίδιας μορφής με τη σχέση 5.5, και συσχετίζεται με αυτό ένα βάρος  $\pi^v$ . Πιο συγκεκριμένα, έστω ότι δίνεται ένα σύνολο με  $N$  αντικείμενα και  $V$  περιγραφές (όψεις),  $\mathfrak{X} = \{x_1, x_2, \dots, x_N\}$ , όπου  $x_i$  είναι η αναπαράσταση του  $i$ -οστού αντικειμένου για όλες τις όψεις, δηλαδή  $x_i = \{x_i^1, x_i^2, \dots, x_i^V\}$ ,  $x_i^v \in \mathcal{R}^{d^v}$ . Το μοντέλο που προτείνεται, το οποίο θα αναφέρεται ως CMM πολλαπλών όψεων με βάρη, περιγράφεται από τη σχέση

$$F(x = \{x^1, x^2, \dots, x^V\}) = \sum_{v=1}^V \pi^v Q^v(x^v) = \sum_{v=1}^V \pi^v \sum_{j=1}^N q_j f_j^v(x^v), x^v \in \mathcal{R}^{d^v} \quad (5.9)$$

και όπου

$$f_j^v(x^v) = C_{\varphi_v}(x^v) e^{-\beta^v d_{\varphi_v}(x^v, x_j^v)}, \pi^v \geq 0, \sum_{v=1}^V \pi^v = 1, q_j \geq 0, \sum_{j=1}^N q_j = 1$$

Παρατηρώντας τις παραπάνω σχέσεις είναι εύκολο να δει κανείς ότι πρόκειται για ένα μικτό μοντέλο το οποίο περιέχει  $V$  μικτές κατανομές CMM  $Q^v(x^v)$ , τα οποία αντιστοιχούν στις διαθέσιμες όψεις. Έτσι, θα μπορούσαμε να πούμε ότι τα βάρη  $\pi^v$  αντιπροσωπεύουν την *prior* πιθανότητα της κάθε όψης στο μικτό μοντέλο.

Η παραπάνω σχέση έχει ορισμένα πολύ σημαντικά χαρακτηριστικά. Τα CMMs κάθε όψης μπορούν να προέρχονται από διαφορετικές κατανομές  $f_j^v(x^v)$ , με διαφορετικές τιμές  $\beta^v$  και αποστάσεις  $d_{\varphi_v}$ . Για παράδειγμα, μία όψη μπορεί να χρησιμοποιεί Gaussian CMM και μία άλλη Bernouli CMM. Από την στιγμή που ένα CMM χρησιμοποιείται για κάθε όψη, όλα τα αντικείμενα εξετάζονται ως πιθανά πρότυπα (exemplars). Επιπλέον, οι τιμές  $q_j$  είναι ίδιες για όλες τις όψεις προκειμένου να εξαχθούν αντιπροσωπευτικά κέντρα βασισμένα σε όλες τις όψεις. Για το λόγο αυτό, αν η τιμή  $q_j$  ενός αντικειμένου είναι υψηλή, τότε είναι πολύ πιθανό να είναι ένα καλό exemplar για όλες τις διαθέσιμες όψεις. Τέλος, μία χαμηλή τιμή  $\pi^v$  υποδεικνύει ότι η όψη  $v$  εμπεριέχει λίγη πληροφορία σχετικά με την ομαδοποίηση των δεδομένων.

Από τη στιγμή που το  $F(x)$  μπορεί να θεωρηθεί μικτό μοντέλο, για να διαχωρίσουμε το σύνολο  $\mathfrak{X}$  πρέπει να μεγιστοποιήσουμε την πιθανοφάνεια (log-likelihood) που δίνεται από τον τύπο

$$L(\mathfrak{X}; \{\pi^v\}_{v=1}^V, \{q_j\}_{j=1}^N) = \sum_{i=1}^N \log \sum_{v=1}^V \pi^v Q^v(x_i^v) = \sum_{i=1}^N \log \left( \sum_{v=1}^V \pi^v \sum_{j=1}^N q_j f_j^v(x_i^v) \right) \quad (5.10)$$

Πρέπει να σημειωθεί ότι σε αντίθεση με το CMM πολλαπλών όψεων, αυτό το πρόβλημα βελτιστοποίησης δεν είναι κυρτό, γεγονός που οφείλεται στην εισαγωγή των βαρών  $\pi^v$ .

Τοπικά μέγιστα της πιθανοφάνειας μπορούν να εντοπιστούν χρησιμοποιώντας τον αλγόριθμο EM [45]. Ο αλγόριθμος αυτός επιλέγει τιμές για τις παραμέτρους και τις τροποποιεί επαναληπτικά, με σκοπό την αύξηση της πιθανοφάνειας μέχρις ότου επιτευχθεί ένα τοπικό μέγιστο. Στο μοντέλο που προτείνεται οι μοναδικές παράμετροι που αρχικοποιούνται είναι οι *prior* τιμές  $\pi^v$  και  $q_j$ , και αρχικοποιώντας αυτές ομοιόμορφα ( $\pi^{v(0)} = 1/V, q_j^v = 1/N$ ) αποφεύγονται οι πολλαπλές επανεκκινήσεις του EM. Φυσικά, σε περίπτωση που υπάρχει εκ των προτέρων πληροφορία σχετικά με την ποιότητα της κάθε όψης, οι τιμές  $\pi^v$  μπορούν να αρχικοποιηθούν κατάλληλα. Προκειμένου να επεξηγηθούν εν συντομία τα βήματα του αλγορίθμου EM, ορίζουμε ως  $\{\mathcal{X}, \mathcal{Z}\}$  το σύνολο δεδομένων, όπου  $\mathcal{Z} = \{z_1, z_2, \dots, z_N\}$  είναι οι κρυφές μεταβλητές που προσδιορίζουν από ποια όψη του μικτού μοντέλου προέρχεται το κάθε αντικείμενο,  $z_i \in \{1, 2, \dots, V\}$ .

**E-βήμα:** Οι κρυφές μεταβλητές περιγράφονται από τις εκ των υστέρων πιθανότητες  $P(z_i = v | x_i)$ , οι οποίες στην επανάληψη  $t$  υπολογίζονται για  $\forall i \in \{1, 2, \dots, N\}, \forall v \in \{1, 2, \dots, V\}$  από την σχέση

$$P^{(t)}(z_i = v | x_i) = \frac{\pi^{v(t)} Q^{v(t)}(x_i^v)}{\sum_{v=1}^V \pi^{v(t)} Q^{v(t)}(x_i^v)} \quad (5.11)$$

**M-βήμα:** Οι εκ των υστέρων πιθανότητες που προκύπτουν από το E-βήμα χρησιμεύουν στον υπολογισμό των νέων τιμών των παραμέτρων στο M-βήμα. Οι νέες τιμές προκύπτουν από τις σχέσεις 5.12 και 5.13

$$\pi^{v(t+1)} = \frac{1}{N} \sum_{i=1}^N P^{(t)}(z_i = v | x_i) \quad (5.12)$$

$$q_j^{(t'+1)} = \frac{q_j^{(t')}}{N} \sum_{i=1}^N \sum_{v=1}^V P^{(t)}(z_i = v | x_i) \frac{f_j^v(x_i^v)}{\sum_{j'=1}^N q_{j'}^{(t')} f_{j'}^v(x_i^v)} \quad (5.13)$$

Στο παρακάτω σχήμα φαίνεται ο αλγόριθμος της μεθόδου και στη συνέχεια ακολουθούν ορισμένες επισημάνσεις.

---

**Αλγόριθμος EM για CMM πολλαπλών όψεων με βάρη**


---

**Είσοδος:**  $\mathfrak{X} = \{x_1, x_2, \dots, x_N\}$ , όπου  $x_i = \{x_i^1, x_i^2, \dots, x_i^V\}$ ,  $x_i^v \in \mathcal{R}^{d^v}$

**Εξόδος:** Παράμετροι του μοντέλου:  $\{\pi^v\}_{v=1}^V, \{q_j\}_{j=1}^N$

---

1. // Αρχικοποίηση παραμέτρων
  2.  $\pi^{v(0)} = \frac{1}{V}, \forall v = 1, \dots, V$  και  $q_j^{(0)} = \frac{1}{N}, \forall j = 1, \dots, N, t = 0$
  3. Επανάλαβε
  4. // E-βήμα
  5. Για κάθε  $i=0$  έως  $N$
  6. Για κάθε  $v=1$  έως  $V$
  7. 
$$P^{(t)}(z_i = v | x_i) = \frac{\pi^{v(t)} Q^{v(t)}(x_i^v)}{\sum_{v=1}^V \pi^{v(t)} Q^{v(t)}(x_i^v)}$$
  8. Τέλος επανάληψης
  9. Τέλος επανάληψης
  10. // M-βήμα
  11. Για κάθε  $\pi^v, v=1$  έως  $V$  // ενημέρωση των βαρών  $\pi^v$
  12. 
$$\pi^{v(t+1)} = \frac{1}{N} \sum_{i=1}^N P^{(t)}(z_i = v | x_i)$$
  13. Τέλος επανάληψης
  14.  $t=t'$
  15. Επανάλαβε // ενημέρωση των *prior* τιμών  $q_j$
  16. Για κάθε  $q_j, j=1$  έως  $N$
  17. 
$$q_j^{(t'+1)} = \frac{q_j^{(t')}}{N} \sum_{i=1}^N \sum_{v=1}^V P^{(t)}(z_i = v | x_i) \frac{f_j^v(x_i^v)}{\sum_{j'=1}^N q_{j'}^{(t')} f_{j'}^v(x_i^v)}$$
  18. Τέλος επανάληψης
  19.  $t'=t'+1$
  20. Μέχρις ότου  $\sum_{j=1}^N |q_j^{(t')} - q_j^{(t'-1)}| < \epsilon'$
  21.  $t=t+1$
  22. Θέσε  $\{q_j^{(t)}\}_{j=1}^N = \{q_j^{(t')}\}_{j=1}^N$
  23. Μέχρις ότου  $\left| L\left(\mathfrak{X}; \{\pi^{v(t)}\}_{v=1}^V, \{q_j^{(t)}\}_{j=1}^N\right) - L\left(\mathfrak{X}; \{\pi^{v(t-1)}\}_{v=1}^V, \{q_j^{(t-1)}\}_{j=1}^N\right) \right| < \epsilon$
  24. Επέστρεψε  $\{\pi^v\}_{v=1}^V = \{\pi^{v(t)}\}_{v=1}^V, \{q_j\}_{j=1}^N = \{q_j^{(t)}\}_{j=1}^N$
- 

Σχήμα 5.1 Αλγόριθμος EM για CMM πολλαπλών όψεων με βάρη



Πρώτον, η νέα *prior* πιθανότητα  $q_j^{(t'+1)}$  εξαρτάται από την προηγούμενη  $q_j^{(t')}$  και για το λόγο αυτό απαιτείται μία εμφωλευμένη επανάληψη στο M-βήμα έτσι ώστε να γίνουν οι απαραίτητες ενημερώσεις των  $q_j$ . Η επανάληψη τερματίζει όταν η αλλαγή στην τιμή του  $q_j$  είναι μικρότερη από ένα όριο  $\varepsilon'$  (γραμμή 21). Δεύτερον, ο EM τερματίζει όταν η διαφορά της πιθανοφάνειας μεταξύ των βημάτων E και M γίνει μικρότερη από ένα όριο  $\varepsilon$  (γραμμή 24). Τρίτον, στους υπολογισμούς των εκ των υστέρων πιθανοτήτων πρέπει να ενσωματωθούν οι τιμές  $C_{\phi_v}(x^v)$  (απαλοίφονται για τον υπολογισμό των  $q_j$ ). Πρέπει να σημειωθεί ότι στη γενική περίπτωση, μόνο οι αποστάσεις μεταξύ των αντικειμένων δεν επαρκούν και απαιτούνται οι τιμές των αντικειμένων σε αντίθεση με τα CMM μίας και πολλαπλών όψεων. Όμως, όπως αναφέρθηκε και προηγουμένως, για συγκεκριμένες κατανομές  $f_j^v(x^v)$ , όπως η Gaussian κατανομή, οι τιμές των αντικειμένων δεν απαιτούνται. Τέταρτον, για την επιλογή της τιμής  $\beta_0^v$  υιοθετήθηκε η σχέση 5.8. Τέλος, από τον αλγόριθμο φαίνεται ότι τα βάρη  $\pi^v$  υπολογίζονται αυτόματα κατά τη διάρκεια του βήματος M.

Για τον διαχωρισμό του συνόλου  $\mathfrak{X}$  σε M διαφορετικές ομάδες  $C_1, C_2, \dots, C_M$ , μετά την ολοκλήρωση του EM, πρέπει να επιλεχθούν τα αντικείμενα που θα χρησιμοποιηθούν ως κέντρα (exemplars). Έτσι, επιλέγονται τα M αντικείμενα με τις μεγαλύτερες  $q_j$  τιμές, τα οποία σχηματίζουν το σύνολο  $\mathfrak{X}^E = \{x_1^E, x_2^E, \dots, x_M^E\} \subset \mathfrak{X}$ . Τα υπόλοιπα αντικείμενα  $x_i$  ανατίθενται στην ομάδα  $C_k$ , η οποία σχετίζεται με το  $k$ -οστό κέντρο, το οποίο έχει τη μεγαλύτερη εκ των υστέρων πιθανότητα  $P(C_k|x_i)$  (5.14).

$$P(C_k|x_i) = \frac{q_k^E \sum_{v=1}^V \pi^v f_k^{v^E}(x_i^v)}{\sum_{v=1}^V \pi^v \sum_{j=1}^M q_j f_j^v(x_i^v)} \quad (5.14)$$

Στην παραπάνω σχέση αναφέρονται η *prior* πιθανότητα και η κατανομή της  $v$ -οστής όψης που σχετίζεται με το κέντρο  $x_k^E$ , με τα  $q_k^E$  και  $f_k^{v^E}(x^v)$  αντίστοιχα. Η ανάθεση σε ομάδες δίνεται από τον τύπο

$$C_k = \{x_k^E\} \cup \{x_i | P(C_k|x_i) > P(C_l|x_i), \forall l \neq k, x_i \notin \mathfrak{X}^E\} \quad (5.15)$$

#### 5.4.2. Επιπλέον Στοιχεία

Στα δεδομένα υπάρχει η δυνατότητα να εφαρμοστεί ένας μετασχηματισμός  $\varphi^v$  έτσι ώστε η διαδικασία να εκτελεστεί σε ένα νέο χώρο διανυσμάτων (feature space). Η αναπαράσταση των δεδομένων σε ένα νέο χώρο μπορεί να οδηγήσει σε μία πιο ξεκάθαρη δομή των δεδομένων. Συνήθως χρησιμοποιείται μία συνάρτηση πυρήνα  $K^v(x_i^v, x_j^v)$  η οποία αντιστοιχεί σε εσωτερικά γινόμενα δημιουργώντας έναν πίνακα πυρήνα  $K^v \in \mathbb{R}^{N \times N}$ , όπου  $K_{ij}^v = \varphi^v(x_i^v)^T \varphi^v(x_j^v) = K^v(x_i^v, x_j^v)$ . Εφόσον στους υπολογισμούς χρησιμοποιούνται μόνο τα εσωτερικά γινόμενα δεν είναι απαραίτητο να είναι γνωστό το  $\varphi^v$ . Στην δική μας προσέγγιση, στην οποία χρησιμοποιούνται Gaussian CMM, έχουμε  $d_{\varphi^v}(\varphi^v(x_i^v), \varphi^v(x_j^v)) = \|\varphi^v(x_i^v) - \varphi^v(x_j^v)\|^2 = K_{ii}^v + K_{jj}^v - 2K_{ij}^v$ . Αυτό έχει σαν αποτέλεσμα οι αποστάσεις των αντικειμένων να υπολογίζονται από τις τιμές του πίνακα πυρήνα. Το γεγονός αυτό παρέχει στη μέθοδο τη δυνατότητα εφαρμογής πολλών πυρήνων (multi-kernel) χωρίς επίβλεψη, σε περίπτωση που υπάρχουν διαθέσιμες πολλές συναρτήσεις πυρήνα για κάθε όψη και είναι επιθυμητή η εύρεση του κατάλληλου συνδυασμού αυτών με σκοπό την όσο το δυνατόν καλύτερη ομαδοποίηση των δεδομένων. Ο αλγόριθμος αντιμετωπίζει την κάθε συνάρτηση πυρήνα σαν μία νέα όψη, υπολογίζει τις αποστάσεις των αντικειμένων και συνδυάζει τις συναρτήσεις πυρήνα χρησιμοποιώντας τα βάρη  $\pi^v$ .

Όσον αφορά την πολυπλοκότητα του EM στο μοντέλο μας, ο υπολογισμός των εκ των υστέρων πιθανοτήτων  $P(z_i = v | x_i)$  απαιτεί  $O(N^2V)$  υπολογισμούς, ενώ οι ενημερώσεις των βαρών  $\pi^v$  και τον *prior* πιθανοτήτων κοστίζουν  $O(NV)$  και  $O(N^2V)$  υπολογισμούς αντίστοιχα. Υποθέτοντας  $t$  EM επαναλήψεις μέχρι να επιτευχθεί σύγκλιση και  $t'$  επαναλήψεις σε κάθε εμφωλευμένη επανάληψη του βήματος M κατά τη διαδικασία υπολογισμού των *prior* πιθανοτήτων, η συνολική πολυπλοκότητα είναι  $O(N^2Vt + NVt + N^2Vt't') = O(N^2Vt't')$ . Τέλος, αν δεν είναι διαθέσιμες οι αποστάσεις μεταξύ των δεδομένων  $d_{\varphi^v}(x_i^v, x_j^v)$  της κάθε όψης, συνήθως απαιτείται επιπλέον  $O(N^2Vd)$  υπολογισμούς, όπου  $d = \max\{d^1, d^2, \dots, d^V\}$ .

## ΚΕΦΑΛΑΙΟ 6. ΠΕΙΡΑΜΑΤΑ

---

6.1 Σύνολο Δεδομένων

6.2 Καθορισμός του Ground Truth

6.3 Εξαγωγή Πληροφορίας από τις Ακολουθίες Βίντεο

6.4 Συναρτήσεις Πυρήνα (Kernels)

6.5 Συνδυασμός Πολλών Όψεων με Βάρη (Weighted Multiview)

6.6 Κριτήριο Αξιολόγησης

---

Στο κεφάλαιο αυτό περιγράφονται οι ακολουθίες βίντεο που χρησιμοποιήθηκαν στα πειράματα και τα εικονοπλάισια που θεωρούνται ως ground truth τους. Επιπλέον, περιγράφεται η διαδικασία εξαγωγής της πληροφορίας από τις ακολουθίες αυτές καθώς επίσης και οι παράμετροι που χρησιμοποιήθηκαν για τους περιγραφείς SIFT, CCH και SURF. Επίσης, αναφέρονται συνοπτικά οι τρεις συναρτήσεις πυρήνα (kernels) που χρησιμοποιήθηκαν και οι τρόποι χρήσης αυτών σε συνδυασμό με τις διαθέσιμες περιγραφές (όψεις). Τέλος, αναφέρονται οι παράμετροι εκτέλεσης της μεθόδου ομαδοποίησης πολλαπλών όψεων με βάρη για τα πειράματα που εκτελέστηκαν. Είναι σημαντικό να αναφερθεί ότι για την εξαγωγή των αποτελεσμάτων πραγματοποιήθηκαν δύο σειρές πειραμάτων χρησιμοποιώντας διαφορετικό ground truth επιλεγμένο από διαφορετικό χρήστη και η οπτική αξιολόγηση των αποτελεσμάτων πραγματοποιήθηκε από δύο άτομα.

### 6.1. Σύνολο Δεδομένων

Κατά την διάρκεια της πειραματικής διαδικασίας χρησιμοποιήθηκαν 23 διαφορετικά πλάνα, ανάλυσης 320x240, καλύπτοντας έτσι ένα μεγάλο φάσμα περιπτώσεων. Ορισμένες από τις κατηγορίες που περιγράφουν τα πλάνα είναι κίνηση αυτοκινήτων, κατεδαφίσεις κατασκευών, αυτοκινητιστικά ατυχήματα, εναλλαγή φωτεινών σηματοδοτών, κίνηση σε περιορισμένο χώρο κ.α. Στον πίνακα 6.1 παρουσιάζεται το πλήθος των εικονοπλαισίων κάθε βίντεο και το πλήθος των χαρακτηριστικών εικονοπλαισίων που περιγράφει επαρκώς το περιεχόμενο της ακολουθίας.

Πίνακας 6-1 Χαρακτηριστικά του συνόλου των πλάνων που χρησιμοποιήθηκαν στα πειράματα

Video ID	Πλήθος Εικονοπλαισίων	Πλήθος key-frames	
		Ground Truth 1	Ground Truth 2
1	389	6	4
2	279	5	6
3	399	5	6
4	479	4	4
5	519	6	7
6	572	6	6
7	512	6	4
8	524	5	4
9	195	4	5
10	256	6	4
11	270	4	5
12	180	5	4
13	127	3	5
14	120	6	3
15	112	3	3
16	103	3	4
17	175	6	4
18	463	3	6
19	118	3	5
20	176	4	3
21	276	9	4
22	712	6	6
23	806	5	10
<b>Συνολικά</b>	<b>7762</b>	<b>113</b>	<b>112</b>

## 6.2. Καθορισμός του Ground Truth

Η επιλογή του συνόλου ground truth έγινε με τέτοιο τρόπο ώστε να επιτευχθεί μία περίληψη που αντιπροσωπεύει επαρκώς και αποτελεσματικά το περιεχόμενο των ακολουθιών βίντεο που χρησιμοποιήθηκαν στην εργασία αυτή. Επιπλέον, η επιλογή αυτή έγινε προσπαθώντας ο αριθμός των χαρακτηριστικών εικονοπλαισίων να είναι όσο γίνεται μικρότερος, χωρίς όμως να χαθεί σημαντική πληροφορία. Είναι σημαντικό να αναφερθεί ότι στο ground truth σε ορισμένες περιπτώσεις η επανεμφάνιση ενός στοιχείου, όπως ανθρώπου ή αυτοκινήτου, κατά τη διάρκεια του βίντεο κρίθηκε αρκετά σημαντική προκειμένου να συμπεριληφθεί στην περίληψη της ακολουθίας. Το γεγονός αυτό αναπόφευκτα οδηγεί σε μείωση της επίδοσης του εφαρμοζόμενου αλγορίθμου, καθώς είναι γνωστό ότι οι αλγόριθμοι ομαδοποίησης δεν λαμβάνουν υπόψη τους την σειρά εμφάνισης των εικονοπλαισίων (βλ. 2.3.6), αλλά ταυτόχρονα αυξάνει την αξιοπιστία των αποτελεσμάτων μας αφού το σύνολο δεδομένων δεν είναι ιδανικό για τον αλγόριθμο που χρησιμοποιήθηκε. Στη συνέχεια παρουσιάζονται τα εικονοπλαίσια που αποτελούν το ground truth της πρώτης σειράς πειραμάτων για την κάθε ακολουθία:



Εικόνα 6-1 Τα εικονοπλαίσια του ground truth της 1<sup>ης</sup> ακολουθίας



Εικόνα 6-2 Τα εικονοπλάισια του ground truth της 2<sup>ης</sup> ακολουθίας



Εικόνα 6-3 Τα εικονοπλάισια του ground truth της 3<sup>ης</sup> ακολουθίας



Εικόνα 6-4 Τα εικονοπλάισια του ground truth της 4<sup>ης</sup> ακολουθίας



Εικόνα 6-5 Τα εικονοπλαίσια του ground truth της 5<sup>ης</sup> ακολουθίας



Εικόνα 6-6 Τα εικονοπλαίσια του ground truth της 6<sup>ης</sup> ακολουθίας



Εικόνα 6-7 Τα εικονοπλαίσια του ground truth της 7<sup>ης</sup> ακολουθίας



Εικόνα 6-8 Τα εικονοπλαίσια του ground truth της 8<sup>ης</sup> ακολουθίας



Εικόνα 6-9 Τα εικονοπλαίσια του ground truth της 9<sup>ης</sup> ακολουθίας



Εικόνα 6-10 Τα εικονοπλαίσια του ground truth της 10<sup>ης</sup> ακολουθίας



Εικόνα 6-11 Τα εικονοπλαίσια του ground truth της 11<sup>ης</sup> ακολουθίας





Εικόνα 6-12 Τα εικονοπλαίσια του ground truth της 12<sup>ης</sup> ακολουθίας



Εικόνα 6-13 Τα εικονοπλαίσια του ground truth της 13<sup>ης</sup> ακολουθίας



Εικόνα 6-14 Τα εικονοπλαίσια του ground truth της 14<sup>ης</sup> ακολουθίας



Εικόνα 6-15 Τα εικονοπλαίσια του ground truth της 15<sup>ης</sup> ακολουθίας



Εικόνα 6-16 Τα εικονοπλαίσια του ground truth της 16<sup>ης</sup> ακολουθίας



Εικόνα 6-17 Τα εικονοπλαίσια του ground truth της 17<sup>ης</sup> ακολουθίας



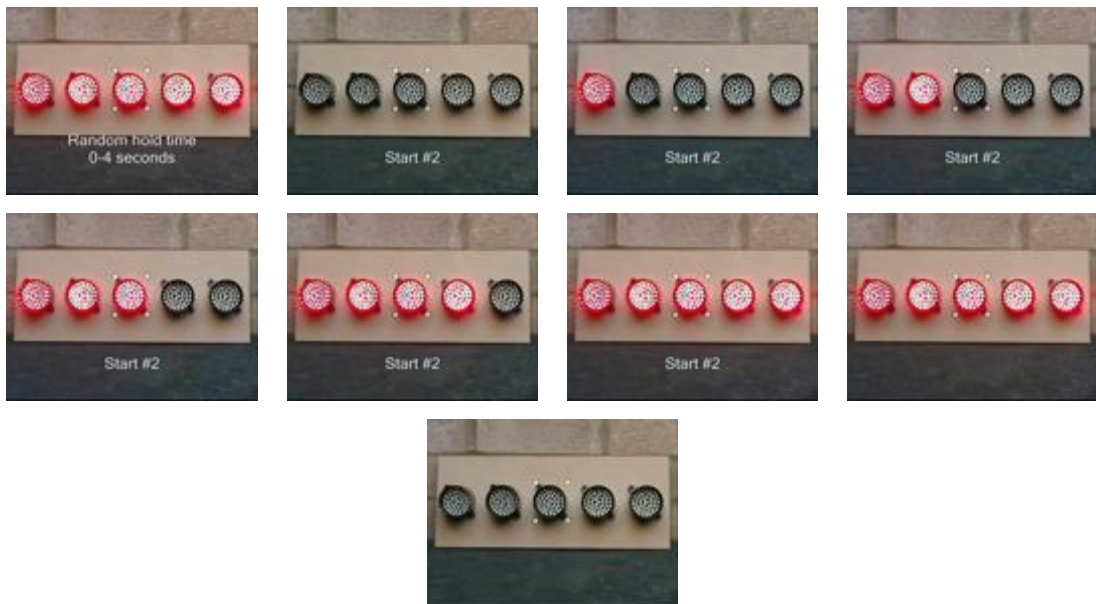
Εικόνα 6-18 Τα εικονοπλαίσια του ground truth της 18<sup>ης</sup> ακολουθίας



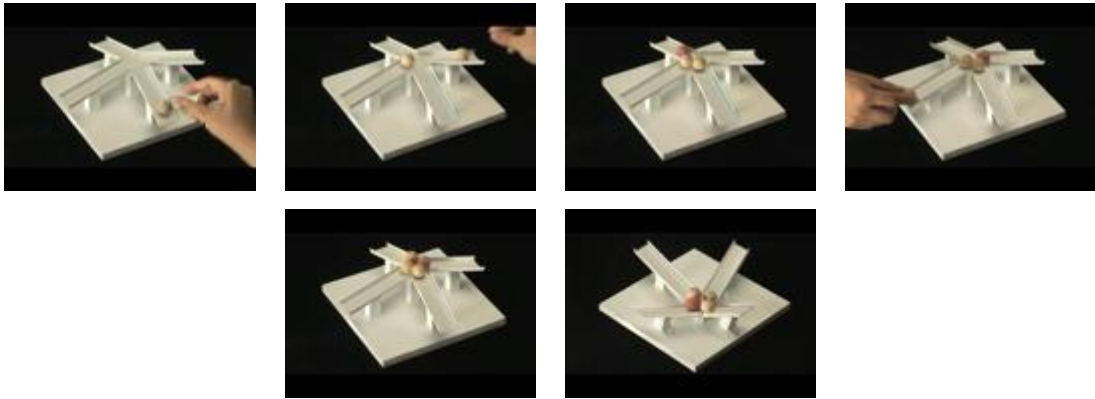
Εικόνα 6-19 Τα εικονοπλαίσια του ground truth της 19<sup>ης</sup> ακολουθίας



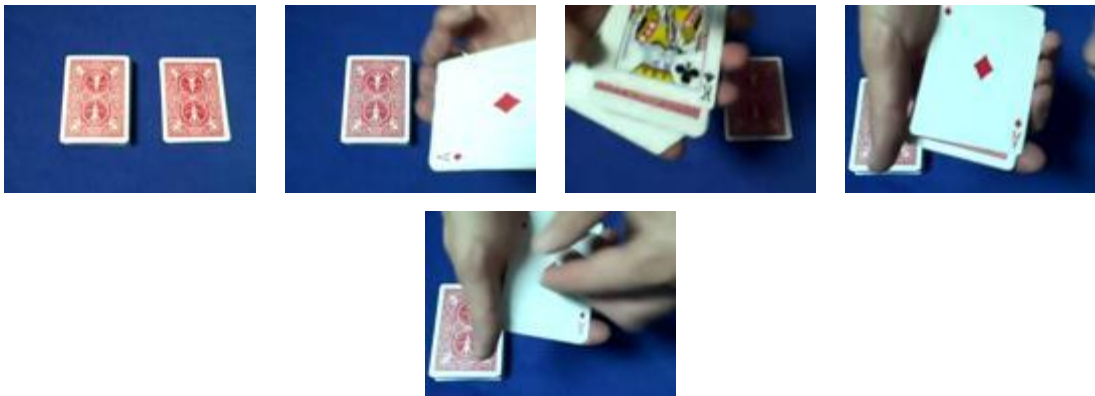
Εικόνα 6-20 Τα εικονοπλαίσια του ground truth της 20<sup>ης</sup> ακολουθίας



Εικόνα 6-21 Τα εικονοπλαίσια του ground truth της 21<sup>ης</sup> ακολουθίας



Εικόνα 6-22 Τα εικονοπλαίσια του ground truth της 22<sup>ης</sup> ακολουθίας



Εικόνα 6-23 Τα εικονοπλαίσια του ground truth της 23<sup>ης</sup> ακολουθίας

Για τη δεύτερη σειρά πειραμάτων το ground truth επιλέχθηκε από διαφορετικό χρήστη δημιουργώντας έτσι μία κάπως διαφορετική περίληψη για κάθε ακολουθία βίντεο. Επίσης, είναι σημαντικό να παρατηρήσει κανείς ότι στα δύο ground truth που χρησιμοποιήθηκαν (πίνακας 6.1) υπάρχουν διαφορές όσον αφορά το πλήθος χαρακτηριστικών εικονοπλαισίων που αντιστοιχεί σε κάθε ακολουθία.

### 6.3. Εξαγωγή Πληροφορίας από τις Ακολουθίες Βίντεο

Όπως αναφέρθηκε παραπάνω, οι περιγραφείς (όψεις) που χρησιμοποιήθηκαν στην εργασία είναι:

- Ιστογράμματα χρώματος HSV

- Περιγραφείς CENTRIST
- Περιγραφείς Wavelet
- Περιγραφείς CCH
- Περιγραφείς SIFT
- Περιγραφείς SURF

### 6.3.1. Ιστογράμματα Χρώματος HSV

Για τα ιστογράμματα χρώματος HSV χρησιμοποιήθηκαν δύο διαφορετικές μορφές. Η πρώτη, γνωστή ως HSV-1D, περιγράφεται από ένα διάνυσμα μεγέθους  $1 \times 16$  και αποτελείται από 8 κάδους για την απόχρωση H (Hue) και από 4 κάδους για κάθε ένα από τα κορεσμός S (Saturation) και αξία V (Value). Η δεύτερη, γνωστή ως HSV-3D, περιγράφεται από ένα διάνυσμα μεγέθους  $1 \times 128$  και προκύπτει από τον πολλαπλασιασμό των παραπάνω κάδων. Η διαφορά στις δύο μεθόδους είναι ότι στην HSV-1D ένα pixel έχει μία ξεχωριστή τιμή για κάθε ένα από τα H, S και V, ενώ στην HSV-3D η τιμή του κάθε pixel προκύπτει από συνδυασμό των τιμών H, S και V που έχει.

1<sup>η</sup> σειρά πειραμάτων: Χρησιμοποιήθηκε η μορφή HSV-3D

2<sup>η</sup> σειρά πειραμάτων: Αρχικά, χρησιμοποιήθηκε η μορφή HSV-3D μόνη της. Στη συνέχεια έγινε συνδυασμός των HSV-3D και HSV-1D έχοντας ίδια βάρη (HSV-AVG) και τέλος έγινε συνδυασμός τους χρησιμοποιώντας τα βάρη που προέκυψαν από την παρακάτω διαδικασία:

1. Για κάθε μία ακολουθία ξεχωριστά
2. Εξαγωγή των όψεων HSV-1D και HSV-3D
3. Εκτέλεση του αλγορίθμου ομαδοποίησης με είσοδο τις δύο όψεις και εξαγωγή των βαρών  $W_{1D}$  και  $W_{3D}$

4. Δημιουργία μίας τελικής όψης, HSV–Multi, συνδυάζοντας τις όψεις HSV-1D και HSV-3D χρησιμοποιώντας τα αντίστοιχα βάρη  $W_{1D}$  και  $W_{3D}$

### 6.3.2. Περιγραφείς *CENTRIST* και *Wavelet*

*CENTRIST*: Χρησιμοποιήθηκε ο μετασχηματισμός Census (βλ. 3.2) έχοντας ως τελικό αποτέλεσμα ένα ιστόγραμμα με 254 στοιχεία για κάθε εικονοπλαίσιο της ακολουθίας.

*Wavelet*: Για τους περιγραφείς wavelet κάθε εικονοπλαίσιο χωρίζεται σε 9 τμήματα και σε κάθε ένα από αυτά εφαρμόστηκε ο μετασχηματισμός Haar (βλ. 3.3) χρησιμοποιώντας υποζώνες μεγέθους  $3 \times 3$ , με αποτέλεσμα την δημιουργία ενός διανύσματος  $81 (3 \times 3 \times 9)$  διαστάσεων. Στη συνέχεια το διάνυσμα κανονικοποιείται προκειμένου όλες οι διαθέσιμες όψεις να έχουν το ίδιο εύρος τιμών.

### 6.3.3. Περιγραφείς *CCH*, *SIFT* και *SURF*

*CCH*: Στους περιγραφείς αυτούς, αφού εξαχθούν τα σημαντικά σημεία κάθε εικονοπλαισίου δημιουργείται ένας πίνακας μεγέθους  $P \times 2$ , όπου  $P$  είναι το πλήθος των σημαντικών σημείων, και ένας άλλος μεγέθους  $P \times 64$ . Ο πρώτος πίνακας αναφέρεται στη θέση  $(x,y)$  των σημαντικών σημείων, και ο άλλος περιέχει πληροφορία σχετικά με την κατανομή της αντίθεσης γύρω από τα σημαντικά σημεία.

*SIFT*: Από τους περιγραφείς αυτούς παράγονται δύο πίνακες. Ο ένας μεγέθους  $P \times 2$ , ο οποίος περιέχει πληροφορία σχετικά με τη θέση του κάθε σημαντικού σημείου, και ο άλλος μεγέθους  $P \times 128$ , που περιέχει επιπλέον πληροφορία σχετικά με την κλιμάκωση και την μετατόπιση της εικόνας γύρω από κάθε σημαντικό σημείο.

*SURF*: Και σε αυτούς τους περιγραφείς υπάρχουν δύο διαθέσιμοι πίνακες. Ένας σχετικά με τη θέση των σημαντικών σημείων και ένας που περιέχει επιπλέον πληροφορία σχετικά με την κλιμάκωση και την μετατόπιση της εικόνας γύρω από κάθε σημαντικό σημείο. Το μέγεθος των πινάκων είναι  $P \times 2$  και  $P \times 64$  αντίστοιχα.

Επειδή συνήθως το πλήθος των σημαντικών σημείων σε ένα εικονοπλαίσιο είναι σχετικά μεγάλο, για επιτάχυνση της διαδικασίας χρησιμοποιήθηκε ο αλγόριθμος Fast Global K-means (βλ. 4.3) προκειμένου όλα τα σημαντικά σημεία να αντικατασταθούν από αντιπροσωπευτικά κέντρα πλήθους  $0.1 \times P$  (10% του αρχικού  $P$ ).

Στη συνέχεια εφαρμόστηκε η διαδικασία εξαγωγής ιστογραμμάτων οπτικών λέξεων (βλ. 4.4) με πλήθος λέξεων 20, 50 και 100. Κατά τη διαδικασία αυτή χρησιμοποιήθηκαν δύο διαφορετικοί τρόποι χρήσης των περιγραφών.

*1<sup>η</sup> σειρά πειραμάτων:* Όλα τα σημαντικά σημεία ανατέθηκαν σε 50 λέξεις. Για τους περιγραφείς CCH και SURF η ομαδοποίηση έγινε με βάση την πληροφορία σχετικά με τα σημαντικά σημεία και στους πίνακες συμβολίζονται ως  $CCH_{desc}$  και  $SURF_{desc}$  ενώ οι περιγραφείς SIFT ομαδοποιήθηκαν με βάση τη θέση των σημαντικών σημείων και συμβολίζονται ως  $SIFT_{locs}$ .

*2<sup>η</sup> σειρά πειραμάτων:* Χρησιμοποιήθηκαν οι όψεις CCH-Multi, SIFT-Multi και SURF-Multi οι οποίες προκύπτουν από τον συνδυασμό των 20, 50 και 100 οπτικών λέξεων. Οι όψεις που προέκυψαν από την ομαδοποίηση σε 20, 50 και 100 λέξεις συνδυάστηκαν με τον παρακάτω τρόπο:

1. Για κάθε μία ομάδα περιγραφών CCH, SIFT και SURF
2. Για κάθε μία ακολουθία ξεχωριστά
3. Δημιουργία τριών νέων όψεων  $V_{20}$ ,  $V_{50}$  και  $V_{100}$  χρησιμοποιώντας 20, 50 και 100 λέξεις αντίστοιχα
4. Εκτέλεση του αλγορίθμου με είσοδο τις τρεις νέες όψεις,  $V_{20}$ ,  $V_{50}$  και  $V_{100}$ , και εξαγωγή βαρών  $W_{20}$ ,  $W_{50}$  και  $W_{100}$  για κάθε μία από αυτές αντίστοιχα.
5. Δημιουργία μίας τελικής όψης, CCH-Multi, SIFT-Multi και SURF-Multi αντίστοιχα, συνδυάζοντας τις τρεις όψεις  $V_{20}$ ,  $V_{50}$  και  $V_{100}$  χρησιμοποιώντας τα αντίστοιχα βάρη  $W_{20}$ ,  $W_{50}$  και  $W_{100}$ .

Επιπλέον, στο τέλος κάθε σειράς πειραμάτων έγινε περαιτέρω έρευνα σχετικά με την επίδοση κάθε περιγραφέα CCH, SIFT και SURF για 20, 50 και 100 λέξεις ξεχωριστά προκειμένου να συγκρίνουμε την απόδοσή τους σε σχέση με την απόδοση του

συνδυασμού τους. Είναι σημαντικό να αναφερθεί ότι στις σειρές πειραμάτων δεν χρησιμοποιήθηκαν τα συμπεράσματα που προκύπτουν από τα περαιτέρω πειράματα σχετικά με την επίδοση κάθε περιγραφέα προκειμένου τα αποτελέσματα να είναι πιο αντικειμενικά.

#### 6.4. Συναρτήσεις Πυρήνα (Kernels)

Στην εργασία αυτή χρησιμοποιήθηκαν πέντε διαφορετικοί πυρήνες ως συναρτήσεις ομοιότητας για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων. Οι πυρήνες αυτοί είναι οι γραμμικός (linear), chi-square, τομής ιστογραμμάτων (intersection) συνδυασμός του γραμμικού και του chi-square καθώς επίσης και συνδυασμός και των τριών. Έστω  $x$  και  $y$  τα διανύσματα περιγραφής δύο εικόνων (π.χ. ιστογράμματα χρώματος). Οι συναρτήσεις πυρήνα ορίζεται ως εξής:

*Γραμμικός (Linear)*: Είναι η πιο απλή μέθοδος πυρήνα και περιγράφεται από την σχέση:

$$k(x, y) = x^T y + c \quad (6.1)$$

*Chi-square*: Ο πυρήνας Chi-square, [28], είναι ένας πυρήνας απλός και αποτελεσματικός σε περιπτώσεις σύγκρισης ιστογραμμάτων. Επίσης, έχει αποδειχθεί ότι είναι ανθεκτικός σε περιπτώσεις διαχωρισμού εικόνων που στηρίζονται στο χρώμα και την υφή. Ο πυρήνας Chi-square περιγράφεται από την σχέση:

$$k(x, y) = 1 - \sum_{i=1}^n \frac{(x_i - y_i)^2}{\frac{1}{2}(x_i + y_i)} \quad (6.2)$$

*Τομής Ιστογραμμάτων (Intersection)*: Ο πυρήνας αυτός είναι γνωστός και σαν πυρήνας ελαχίστου και έχει αποδειχθεί χρήσιμος σε περιπτώσεις ομαδοποίησης εικόνων:

$$k(x, y) = \sum_{i=1}^n \min(x_i, y_i) \quad (6.3)$$



*Συνδυασμός όλων των πυρήνων (Multi-kernel3):* Για την επίτευξη του κατάλληλου συνδυασμού των τριών παραπάνω πυρήνων ακολουθήθηκε η παρακάτω διαδικασία:

1. Για όλες τις όψεις ξεχωριστά.
2. Για όλες τις ακολουθίες ξεχωριστά.
3. Δημιουργία τριών νέων όψεων  $v_L$ ,  $v_C$  και  $v_I$  οι οποίες προκύπτουν από την εφαρμογή των τριών πυρήνων, γραμμικού, chi-square και τομής ιστογραμμάτων (intersection) αντίστοιχα, στην αρχική όψη.
4. Εκτέλεση του αλγορίθμου έχοντας ως είσοδο τις τρεις νέες όψεις ( $v_L$ ,  $v_C$  και  $v_I$ ) και εξαγωγή βαρών για κάθε μία από αυτές  $w_L$ ,  $w_C$  και  $w_I$  αντίστοιχα.
5. Δημιουργία μίας τελικής όψης συνδυάζοντας τους τρεις πυρήνες χρησιμοποιώντας τα αντίστοιχα βάρη τους ( $w_L$ ,  $w_C$  και  $w_I$ ).

*Συνδυασμός δύο πυρήνων (Multi-kernel2):* Για τον συνδυασμό του γραμμικού πυρήνα και του chi-square ακολουθήθηκε η διαδικασία που ακολουθήθηκε και για τον συνδυασμό όλων των πυρήνων με μόνη διαφορά ότι συμμετείχαν μόνο ο γραμμικός και ο chi-square πυρήνας.

Είναι σημαντικό να αναφερθεί ότι οι παραπάνω πυρήνες δεν έχουν αρχικές παραμέτρους.

*1<sup>η</sup> σειρά πειραμάτων:* Χρησιμοποιήθηκαν όλοι οι πυρήνες εκτός από τον συνδυασμό δύο πυρήνων (Multi-kernel2)

*2<sup>η</sup> σειρά πειραμάτων:* Χρησιμοποιήθηκαν όλοι οι διαθέσιμοι πυρήνες

## **6.5. Συνδυασμός Πολλών Όψεων με Βάρη (Weighted Multiview)**

Η χρήση του αλγορίθμου εξαγωγής χαρακτηριστικών εικονοπλαισίων πολλαπλών όψεων με βάρη έγινε με πέντε διαφορετικούς τρόπους. Σε κάθε περίπτωση τα αρχικά βάρη των όψεων είναι ίδια.

*Απευθείας χρήση:* Για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων ο αλγόριθμος εκτελείται έχοντας την δυνατότητα να αναθέσει βάρη σε όλες τις διαθέσιμες πληροφορίες (όψεις), όπως αυτός κρίνει καλύτερα.

*Με κατώφλι συμμετοχής (Threshold):* Σε αυτή την περίπτωση αρχικά έγινε απευθείας χρήση του αλγορίθμου και στη συνέχεια έγινε επανεκτέλεσή του, στα αρχικά δεδομένα, χρησιμοποιώντας μόνο τις όψεις που είχαν συμμετοχή (βάρος) πάνω από κάποιο, καθορισμένο από το χρήστη, κατώφλι. Ορίστηκαν δύο διαφορετικά κατώφλια με τιμή συμμετοχής τουλάχιστον 10% και τουλάχιστον 5%.

*Εξαγωγή βαρών και δημιουργία μίας τελικής όψης (Multi2One):* Σε αυτή την περίπτωση αρχικά έγινε απευθείας χρήση του αλγορίθμου και στη συνέχεια χρησιμοποιήθηκαν τα βάρη για την δημιουργία μίας τελικής όψης. Στη συνέχεια ο αλγόριθμος εκτελέστηκε ξανά στα αρχικά δεδομένα χρησιμοποιώντας μόνο την νέα όψη. Για παράδειγμα, αν υποθέσουμε ότι έχουμε τις  $N$  όψεις  $V_1...V_N$  και  $W_1...W_N$  είναι τα αντίστοιχα βάρη τους μετά την απευθείας εκτέλεση του αλγορίθμου, η τελική όψη προκύπτει από τη σχέση

$$V = V_1 * W_1 + \dots + V_N * W_N \quad (6.4)$$

*Εξαγωγή βαρών και χρήση μόνο της όψης με το μεγαλύτερο βάρος (Max Weight):* Για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων αρχικά έγινε απευθείας χρήση του αλγορίθμου και στη συνέχεια έγινε εξαγωγή των χαρακτηριστικών εικονοπλαισίων χρησιμοποιώντας μόνο μία όψη (single-view), αυτή με το μεγαλύτερο βάρος.

*1<sup>η</sup> σειρά πειραμάτων:* Ο αλγόριθμος χρησιμοποιήθηκε με όλους τους παραπάνω τρόπους εκτέλεσης. Οι όψεις που συνδυάστηκαν είναι HSV3D, CENTRIST, Wavelet, CCH<sub>desc</sub>, SIFT<sub>locs</sub> και SURF<sub>desc</sub>.

*2<sup>η</sup> σειρά πειραμάτων:* Χρησιμοποιήθηκε μόνο η απευθείας χρήση του αλγορίθμου. Δημιουργήθηκαν τρεις ομάδες συνδυασμού. Οι όψεις CENTRIST, Wavelet, CCH-Multi, SIFT-Multi και SURF-Multi συμμετείχαν και στις τρεις ομάδες, ενώ κάθε ομάδα συμπεριλάμβανε και μία εκ των HSV3D, HSV-AVG και HSV-Multi.

## 6.6. Κριτήρια Αξιολόγησης

Η αξιολόγηση των αποτελεσμάτων βασίστηκε στην οπτική προσέγγιση συγκρίνοντας τα χαρακτηριστικά εικονοπλαίσια που προέκυψαν από την εκτέλεση του αλγορίθμου με αυτά του ground truth. Η αξιολόγηση έγινε από δύο άτομα και για τις δύο σειρές πειραμάτων και χρησιμοποιήθηκε η τομή των δύο αξιολογήσεων. Μετά την αξιολόγηση των αποτελεσμάτων, και την εξαγωγή ποσοστού επιτυχίας για κάθε ακολουθία ξεχωριστά, χρησιμοποιήθηκαν τρεις διαφορετικοί μέθοδοι για την εξαγωγή του μέσου όρου της απόδοσης του αλγορίθμου. Οι τρεις μέθοδοι περιγράφονται στη συνέχεια:

*Μέσος Όρος και Διακύμανση (M.O. και std):* Έστω ότι υπάρχουν  $N$  ακολουθίες  $V=\{V_1, \dots, V_N\}$  με ποσοστό εύρεσης ορθών χαρακτηριστικών εικονοπλαισίων  $P=\{P_1, \dots, P_N\}$  αντίστοιχα. Ο μέσος όρος κάθε κατηγορίας υπολογίζεται ως,

$$M.O. = \frac{1}{N} \sum_{i=1}^N P_i \quad (6.5)$$

και η διακύμανση υπολογίζεται από τη σχέση 6.6,

$$std = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - M.O.)^2} \quad (6.6)$$

*Συνολικός Μέσος Όρος ( $M_{all}$ ):* Έστω ότι υπάρχουν  $N$  ακολουθίες  $V=\{V_1, \dots, V_N\}$  και με  $F_i$  συμβολίζουμε το πλήθος των ορθών χαρακτηριστικών εικονοπλαισίων που προέκυψαν από τον αλγόριθμο, για κάθε ακολουθία  $V_i$ . Ο συνολικός μέσος όρος προκύπτει από την σχέση 6.7,

$$M_{all} = \frac{1}{GT_{all}} \sum_{i=1}^N F_i \quad (6.7)$$

όπου  $GT_{all}$ , το συνολικό πλήθος των χαρακτηριστικών εικονοπλαισίων του ground truth.

*Μέσος Όρος με Βάρη (MW)*: Έστω ότι υπάρχουν  $N$  ακολουθίες  $V=\{V_1, \dots, V_N\}$  με ποσοστό επιτυχίας  $P=\{P_1, \dots, P_N\}$  αντίστοιχα. Το σύνολο  $GT=\{GT_1, \dots, GT_N\}$  αντιπροσωπεύει το πλήθος των χαρακτηριστικών εικονοπλασιών της κάθε ακολουθίας. Το βάρος της κάθε ακολουθίας προκύπτει από την σχέση:

$$W_i = \frac{GT_i}{GT_{all}}, \quad i = 1, \dots, N \quad (6.8)$$

όπου  $GT_{all}$ , το συνολικό πλήθος των χαρακτηριστικών εικονοπλασιών του ground truth. Ο μέσος όρος με βάρη προκύπτει από τη σχέση:

$$MW = \sum_{i=1}^N W_i P_i \quad (6.9)$$

## ΚΕΦΑΛΑΙΟ 7. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

---

### 7.1 Πρώτη Σειρά Πειραμάτων

### 7.2 Δεύτερη Σειρά Πειραμάτων

---

Στο κεφάλαιο αυτό παρουσιάζονται οι πίνακες των αποτελεσμάτων από τις δύο σειρές πειραμάτων που εκτελέστηκαν κατά τη διάρκεια της έρευνας. Το κεφάλαιο είναι χωρισμένο σε δύο μέρη, ένα για κάθε σειρά πειραμάτων, στα οποία αναλύονται τα πειραματικά αποτελέσματα. Στα αποτελέσματα συμπεριλαμβάνονται και εκτελέσεις με χρήση μίας μόνο όψης (single-view) προκειμένου να υπάρχει ένα μέτρο σύγκρισης σχετικά με την αποδοτικότητα του αλγορίθμου μας.

#### 7.1. Πρώτη Σειρά Πειραμάτων

Παρακάτω ακολουθούν τα πειραματικά αποτελέσματα της πρώτης σειράς πειραμάτων. Για την αξιολόγησή τους χρησιμοποιήθηκε το ground truth που περιγράφεται στο κεφάλαιο 6.2. Σε κάθε πίνακα υπογραμμισμένη είναι η καλύτερη επίδοση που είχε η κάθε μέθοδος συνολικά και με έντονα γράμματα είναι η καλύτερη επίδοση συγκρίνοντας όλες τις μεθόδους, για κάθε συνάρτηση πυρήνα.

### 7.1.1. Σύγκριση Single-view με Multi-view

Πίνακας 7-1 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το μέσο όρο και τη διακύμανση

Μέθοδος	Linear		Chi-Square		Intersection		Multi-Kernel3	
	M.O.(%)	std	M.O.(%)	std	M.O.(%)	std	M.O.(%)	std
<b>CCH<sub>desc</sub></b>	61,38	22,79	62,83	16,56	49,18	19,58	<u>67,87</u>	<u>17,50</u>
<b>CENTRIST</b>	63,57	20,03	60,10	18,53	49,54	20,15	<u>70,77</u>	<u>20,13</u>
<b>HSV3D</b>	63,74	15,83	<u>67,70</u>	<u>17,69</u>	57,75	22,22	62,85	15,49
<b>SIFT<sub>locs</sub></b>	63,77	16,08	<u>64,69</u>	<u>17,23</u>	55,95	20,84	60,60	20,48
<b>SURF<sub>desc</sub></b>	54,32	18,15	52,03	16,66	52,93	17,80	<u>68,16</u>	<u>16,86</u>
<b>Wavelet</b>	62,87	16,04	<u>64,83</u>	<u>18,56</u>	56,67	21,06	56,38	15,21
<b>Multiview</b>	68,38	17,23	71,21	18,37	63,12	19,81	<b><u>77,32</u></b>	<b><u>20,15</u></b>

Από τον παραπάνω πίνακα παρατηρούμε ότι η μέθοδος πολλαπλών όψεων με βάρη (weighted multi-view) που προτείνουμε έχει καλύτερη επίδοση από οποιαδήποτε άλλη μέθοδο βασισμένη σε μία όψη (single-view) και για τους τέσσερις πυρήνες που χρησιμοποιήθηκαν. Η καλύτερη επίδοση επιτεύχθηκε χρησιμοποιώντας τη μέθοδο πολλαπλών όψεων με βάρη σε συνδυασμό με τη χρήση πολλαπλών πυρήνων (multi-kernel). Επίσης, μία σημαντική παρατήρηση είναι ότι δεν υπάρχει κάποια single-view μέθοδος που να υπερισχύει ξεκάθαρα όλων των υπολοίπων, γεγονός που αποδεικνύει ότι οι ακολουθίες που χρησιμοποιήθηκαν περιλαμβάνουν όλα τα είδη πληροφορίας ομοιόμορφα κατανεμημένα.

Ένα σημαντικό πρόβλημα στα αποτελέσματα των πειραμάτων είναι η αυξημένη διακύμανση. Το φαινόμενο αυτό παρατηρείται εξαιτίας του διαφορετικού πλήθους χαρακτηριστικών εικονοπλαισίων της κάθε ακολουθίας σε συνδυασμό με το σχετικά μικρό πλήθος ακολουθιών. Για την υποστήριξη των αποτελεσμάτων η μέση τιμή της

επίδοσης υπολογίστηκε και με δύο ακόμα τρόπους (όπως αναφέρθηκε στο κεφάλαιο 6.6), τα αποτελέσματα των οποίων φαίνονται στους πίνακες 7.2 και 7.3.

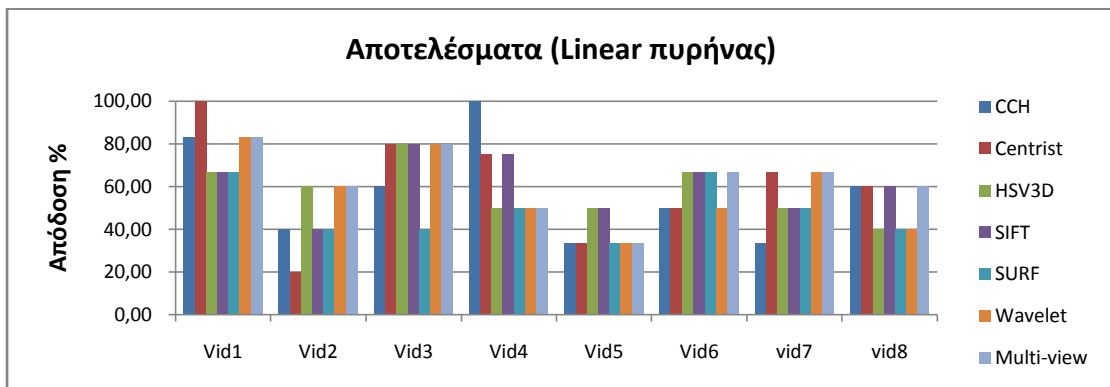
Πίνακας 7-2 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το συνολικό μέσο όρο

Μέθοδος	Linear	Chi-Square	Intersection	Multi-Kernel3
	$M_{all}(\%)$	$M_{all}(\%)$	$M_{all}(\%)$	$M_{all}(\%)$
<b>CCH<sub>desc</sub></b>	59,29	61,06	48,67	<u>74,34</u>
<b>CENTRIST</b>	63,72	60,18	50,44	<u>76,11</u>
<b>HSV3D</b>	62,83	67,26	57,52	<u>67,26</u>
<b>SIFT<sub>locs</sub></b>	64,60	<u>65,49</u>	55,75	62,83
<b>SURF<sub>desc</sub></b>	54,87	53,98	54,87	<u>75,22</u>
<b>Wavelet</b>	61,95	<u>63,72</u>	57,52	62,83
<b>Multiview</b>	67,26	69,03	63,72	<b><u>79,65</u></b>

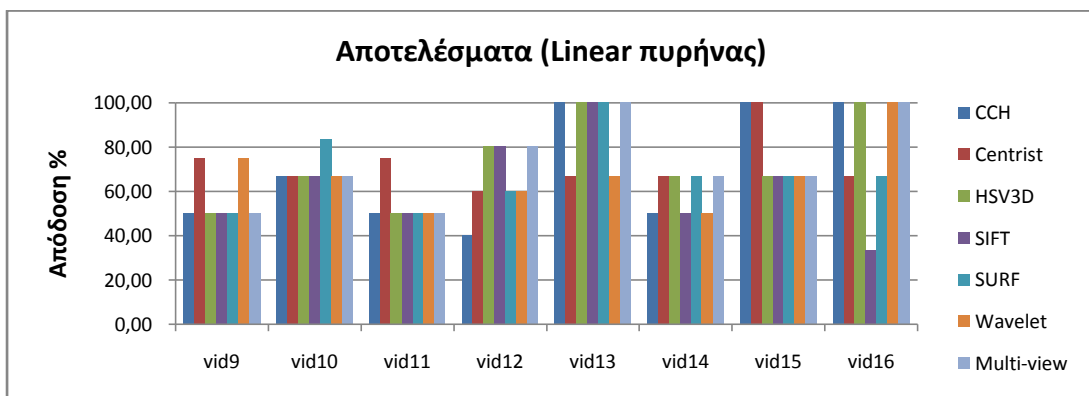
Πίνακας 7-3 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το μέσο όρο με βάση το βάρος

Μέθοδος	Linear	Chi-Square	Intersection	Multi-Kernel3
	$MW(\%)$	$MW(\%)$	$MW(\%)$	$MW(\%)$
<b>CCH<sub>desc</sub></b>	59,27	61,04	48,65	<u>67,26</u>
<b>CENTRIST</b>	63,69	60,15	50,42	<u>70,97</u>
<b>HSV3D</b>	62,80	<u>67,23</u>	57,49	61,95
<b>SIFT<sub>locs</sub></b>	64,57	<u>65,46</u>	55,73	59,29
<b>SURF<sub>desc</sub></b>	54,84	53,95	54,85	<u>68,14</u>
<b>Wavelet</b>	61,92	<u>63,69</u>	57,49	57,52
<b>Multiview</b>	67,22	69,00	63,69	<b><u>75,22</u></b>

Όπως φαίνεται και από τους πίνακες 7.2 και 7.3 η μέθοδος πολλαπλών όψεων με βάρη υπερσχύει όλων των υπόλοιπων single-view μεθόδων και για τους τέσσερις πυρήνες που χρησιμοποιήθηκαν. Και πάλι η καλύτερη επίδοση επιτεύχθηκε χρησιμοποιώντας τη μέθοδο πολλαπλών όψεων με βάρη σε συνδυασμό με τη χρήση πολλαπλών πυρήνων (multi-kernel). Όπως παρατηρήθηκε και προηγουμένως, δεν υπάρχει κάποια single-view μέθοδος που να υπερσχύει σε όλες τις περιπτώσεις όλων των υπολοίπων. Αναλυτικά τα αποτελέσματα για κάθε βίντεο παρουσιάζονται στα παρακάτω διαγράμματα:

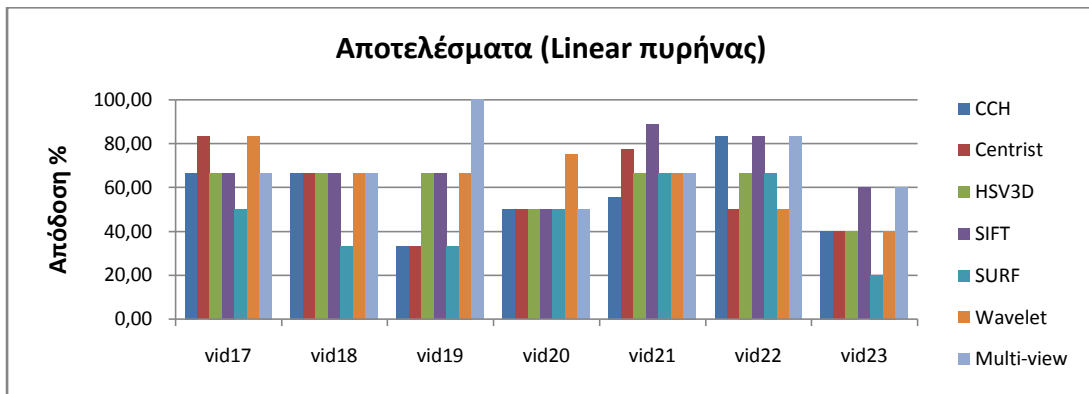


Σχήμα 7.1 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας γραμμικό (linear) πυρήνα

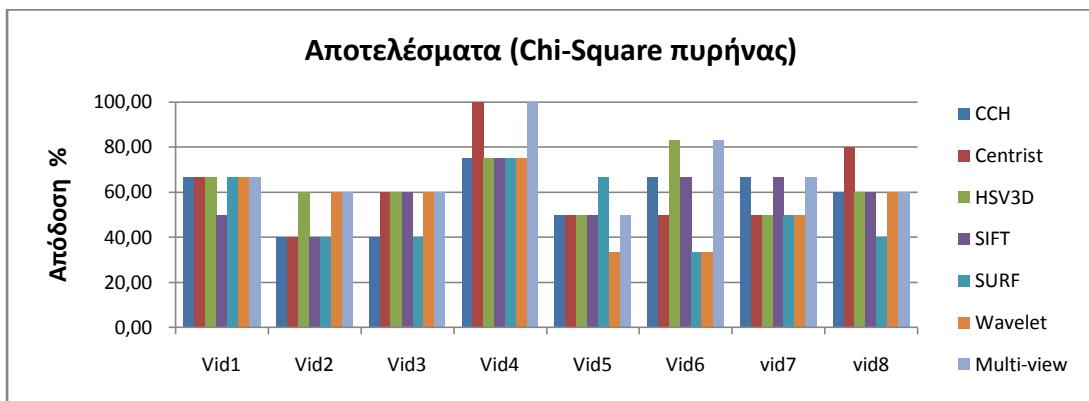


Σχήμα 7.2 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας γραμμικό (linear) πυρήνα

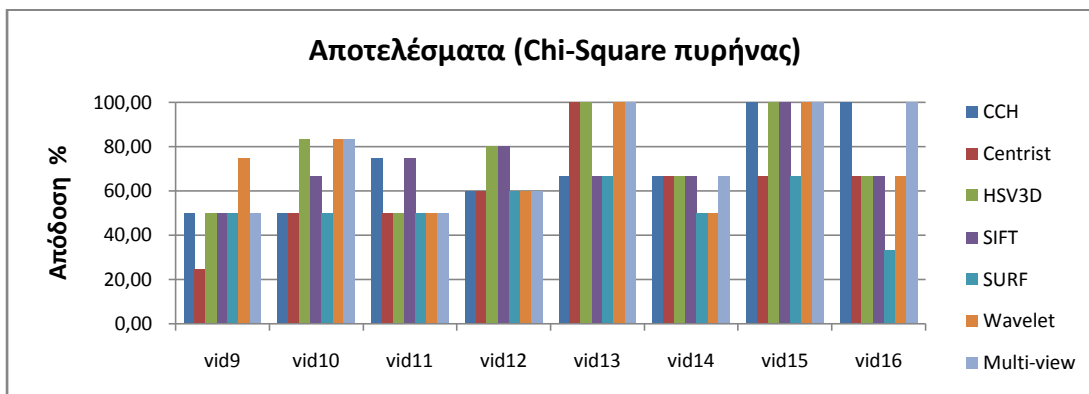




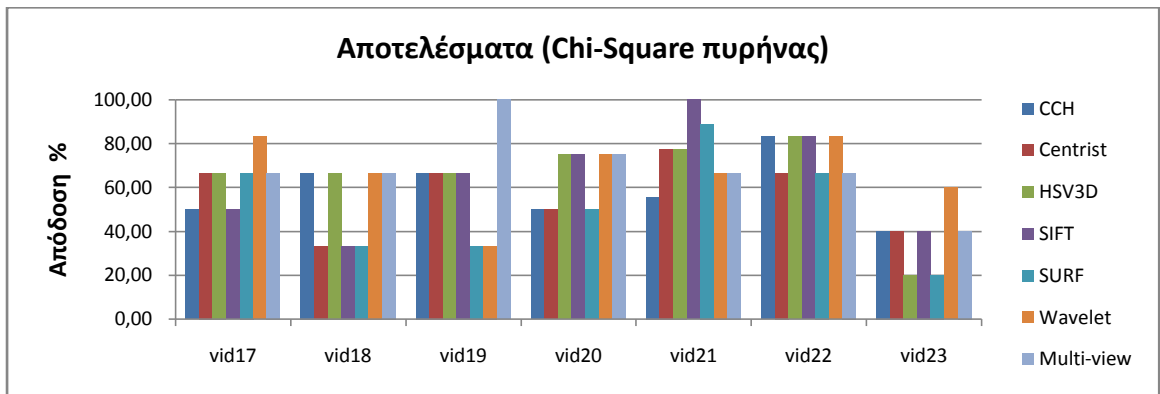
Σχήμα 7.3 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας γραμμικό (linear) πυρήνα



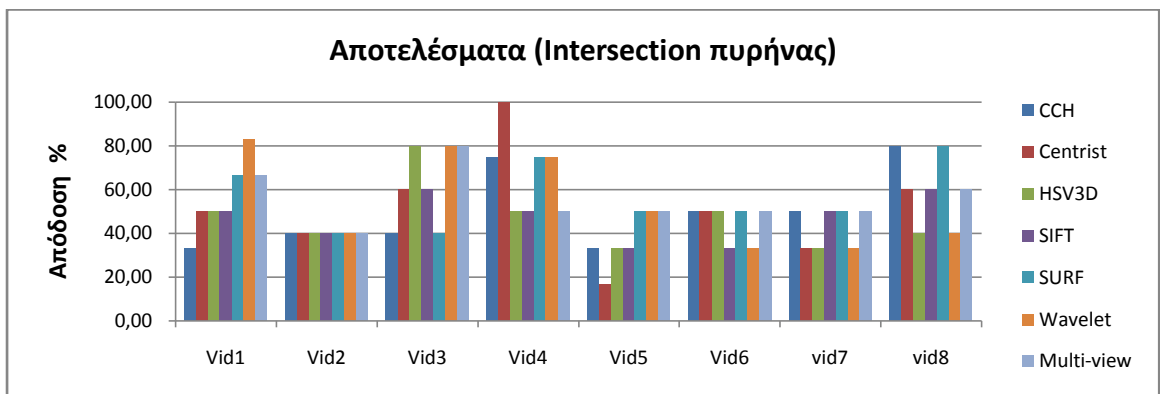
Σχήμα 7.4 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας chi-square πυρήνα



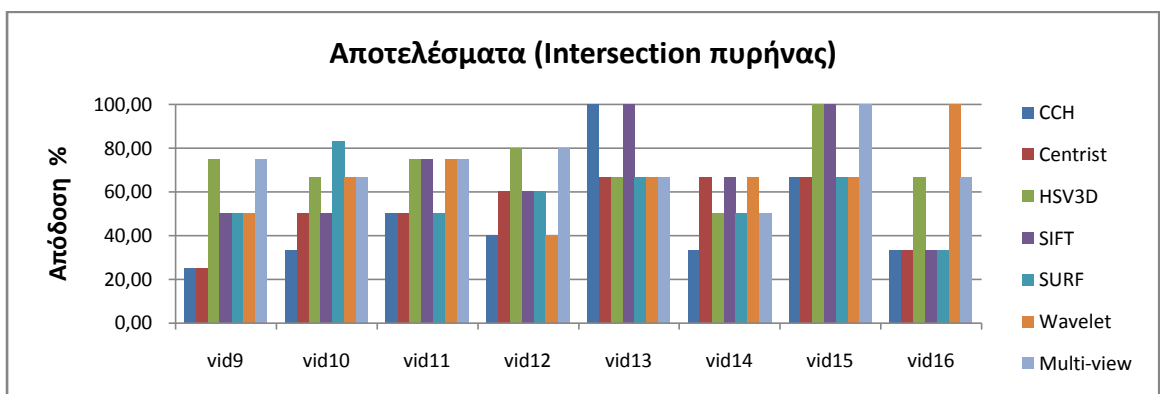
Σχήμα 7.5 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας chi-square πυρήνα



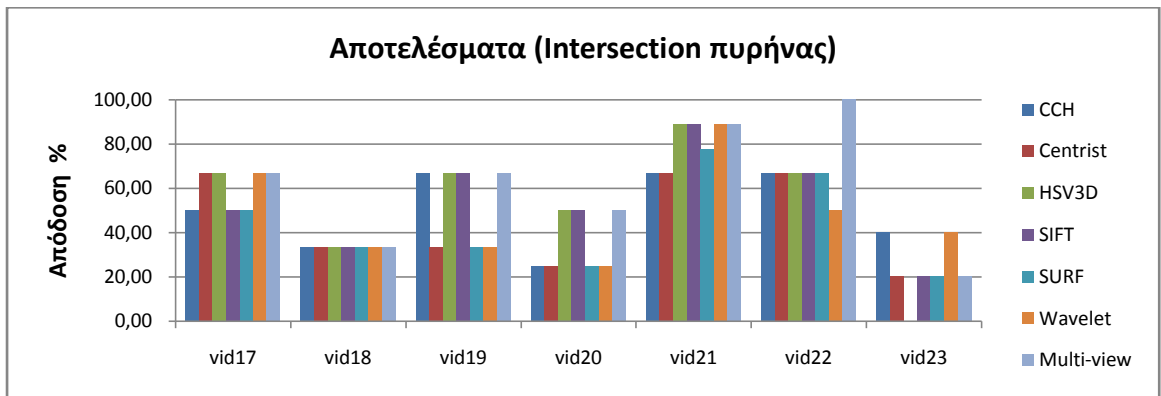
Σχήμα 7.6 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας chi-square πυρήνα



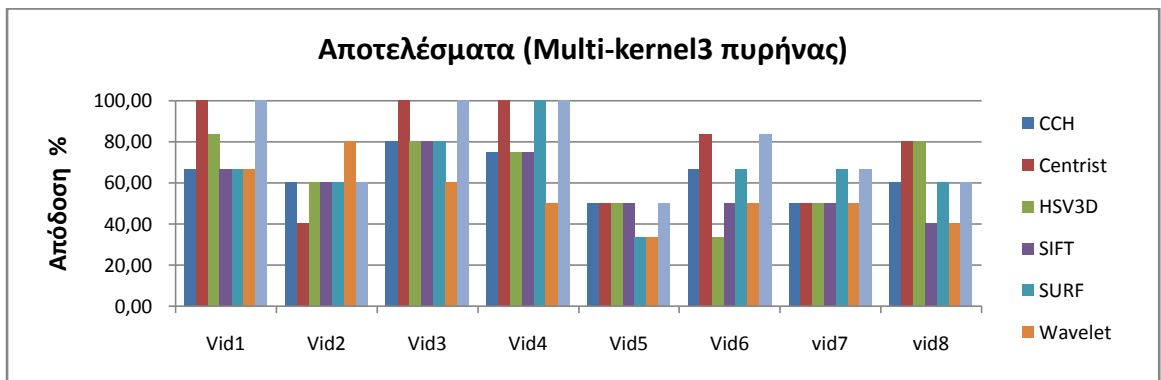
Σχήμα 7.7 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων



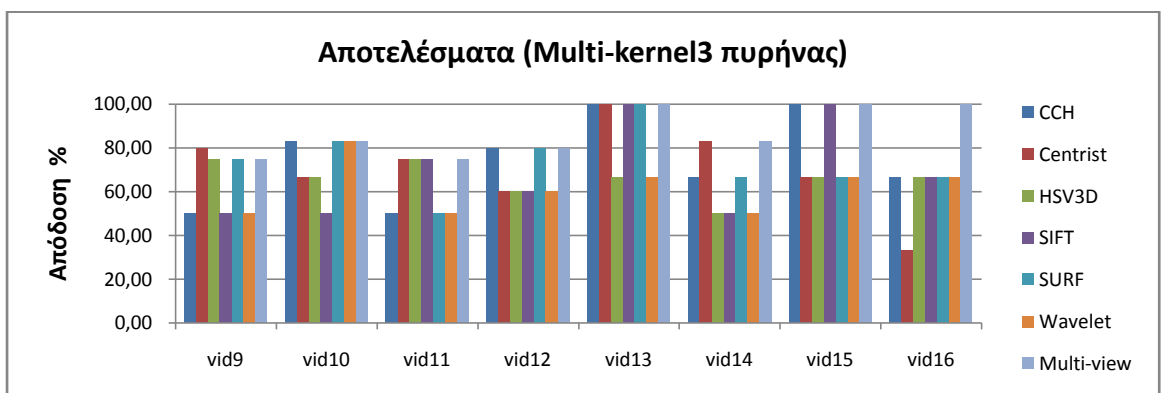
Σχήμα 7.8 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων



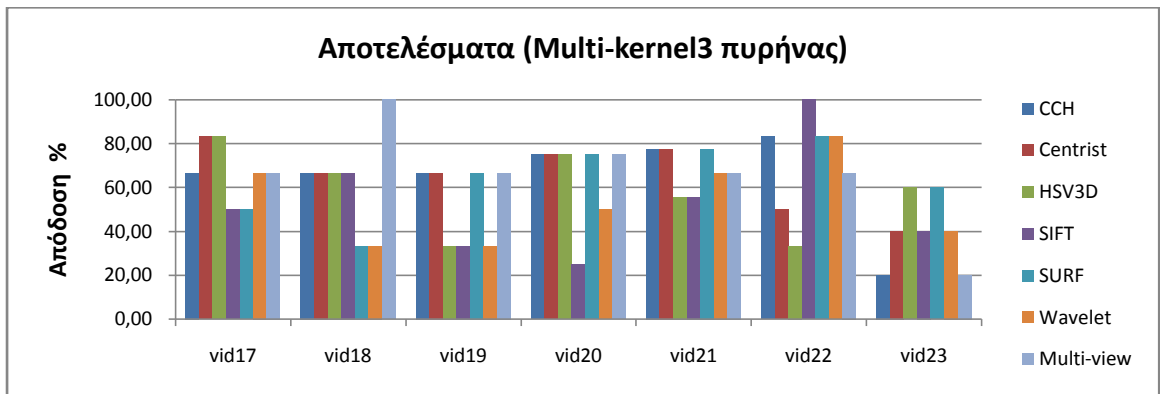
Σχήμα 7.9 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων



Σχήμα 7.10 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη



Σχήμα 7.11 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη



Σχήμα 7.12 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη

### 7.1.2. Σύγκριση των Παραλλαγών Multi-view

Παρακάτω ακολουθούν οι πίνακες 7.4, 7.5 και 7.6, για τις τρεις εκδοχές μέσου όρου, που παρουσιάζουν την επίδοση των πέντε διαφορετικών παραλλαγών της μεθόδου πολλαπλών όψεων με βάρη.

Πίνακας 7-4 Αποτελέσματα της επίδοσης των πέντε διαφορετικών εκδοχών της μεθόδου πολλαπλών όψεων με βάρη, με βάση το μέσο όρο και τη διακύμανση

Μέθοδος	Linear		Chi-Square		Intersection		Multi-Kernel3	
	M.O.(%)	std	M.O.(%)	std	M.O.(%)	std	M.O.(%)	std
<b>Multiview</b>	68,38	17,23	71,21	18,37	63,12	19,81	<u>77,32</u>	<u>20,15</u>
<b>Threshold 10%</b>	71,28	16,47	70,99	16,52	60,22	20,31	<u>72,25</u>	<u>20,67</u>
<b>Threshold 5%</b>	<b>72,49</b>	<b>15,85</b>	<b>72,95</b>	<b>18,16</b>	<b>64,20</b>	<b>19,09</b>	<b><u>78,33</u></b>	<b><u>19,62</u></b>
<b>Multi2One</b>	65,67	15,09	<u>68,31</u>	<u>18,88</u>	59,12	23,24	67,63	22,79
<b>Max Weigth</b>	63,09	17,24	<u>70,99</u>	<u>16,52</u>	58,33	23,43	66,59	19,91

Πίνακας 7-5 Αποτελέσματα της επίδοσης των πέντε διαφορετικών εκδοχών της μεθόδου πολλαπλών όψεων με βάρη, με βάση το συνολικό μέσο όρο

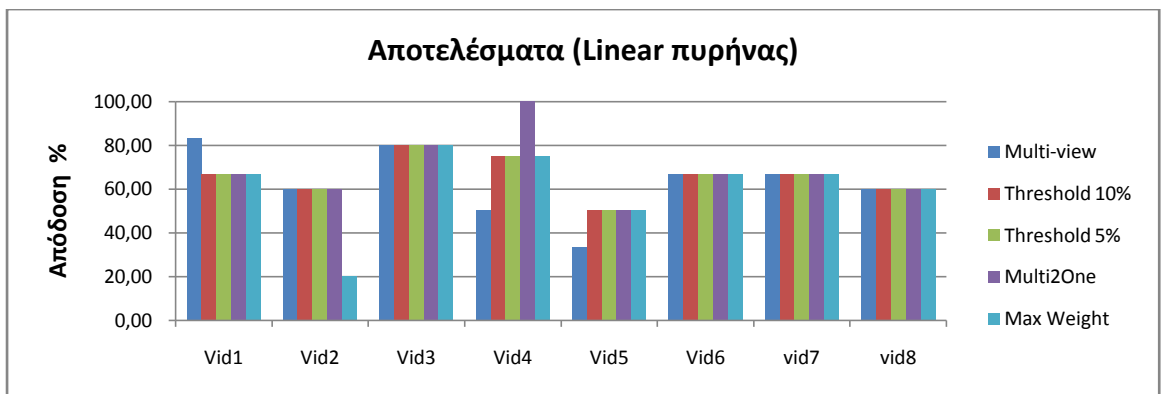
Μέθοδος	Linear	Chi-Square	Intersection	Multi-Kernel3
	M <sub>all</sub> (%)	M <sub>all</sub> (%)	M <sub>all</sub> (%)	M <sub>all</sub> (%)
<b>Multiview</b>	67,26	69,03	63,72	<b><u>79,65</u></b>
<b>Threshold 10%</b>	69,91	69,91	61,06	<u>73,45</u>
<b>Threshold 5%</b>	71,68	71,68	64,60	<b><u>79,65</u></b>
<b>Multi2One</b>	65,49	69,03	60,18	<u>73,45</u>
<b>Max Weigth</b>	61,95	<u>69,91</u>	57,52	<u>69,91</u>

Πίνακας 7-6 Αποτελέσματα της επίδοσης των πέντε διαφορετικών εκδοχών της μεθόδου πολλαπλών όψεων με βάρη, με βάση το μέσο όρο με βάρη

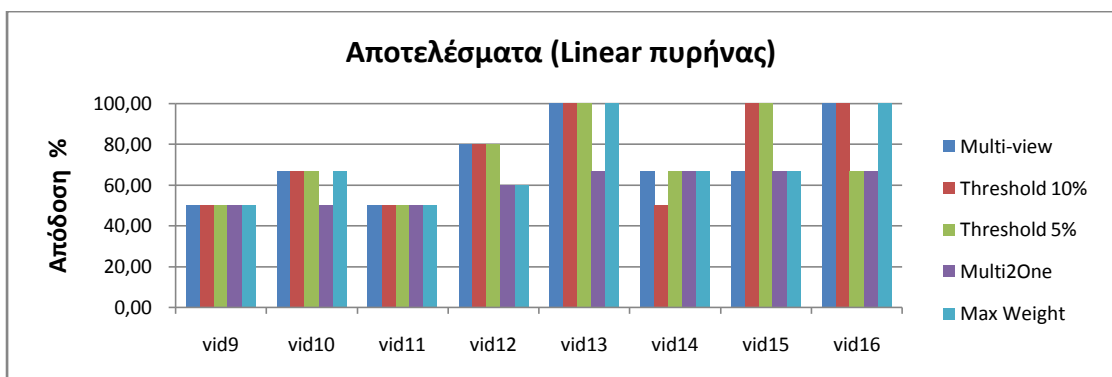
Μέθοδος	Linear	Chi-Square	Intersection	Multi-Kernel3
	MW(%)	MW(%)	MW(%)	MW(%)
<b>Multiview</b>	67,22	69,00	63,69	<u>75,22</u>
<b>Threshold 10%</b>	69,89	69,89	61,04	<u>70,80</u>
<b>Threshold 5%</b>	71,65	71,66	64,58	<b><u>76,99</u></b>
<b>Multi2One</b>	65,46	<u>69,01</u>	60,17	68,14
<b>Max Weigth</b>	61,92	<u>69,89</u>	57,50	66,37

Από τους τρεις παραπάνω πίνακες προκύπτει ότι μεγαλύτερη αποτελεσματικότητα για όλους τους πυρήνες έχει η χρήση κατωφλίου συμμετοχής τουλάχιστον 5%. Το γεγονός αυτό μας δείχνει ότι ακόμα και η μικρή συμμετοχή, μέχρι και 5%, μίας όψης έχει θετική επιρροή στο τελικό αποτέλεσμα. Αυτό γίνεται πιο ξεκάθαρο από την διαφορά των μεθόδων με όριο συμμετοχής 10% και 5%. Στη μέθοδο με όριο συμμετοχής 10% δεν περιλαμβάνονται όψεις με συμμετοχή κάτω από 10%, γεγονός που, όπως φαίνεται, επηρεάζει αρνητικά το τελικό αποτέλεσμα.

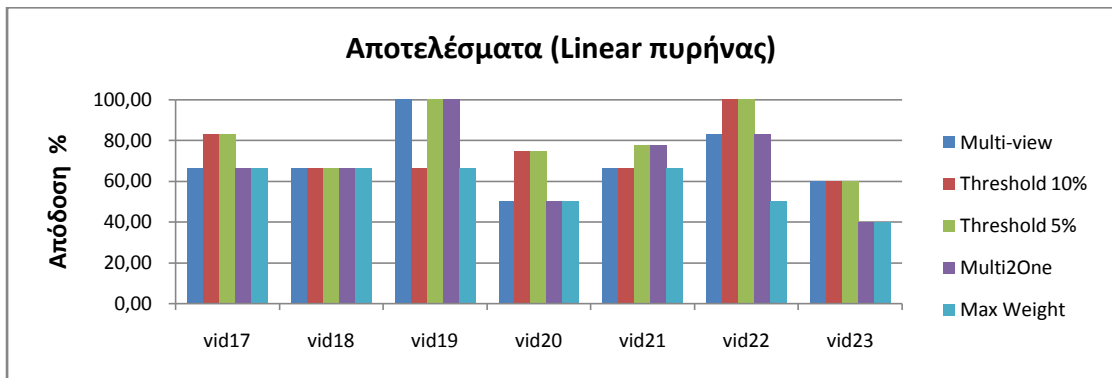
Συνολικά, προκύπτει ότι η χρήση της μεθόδου πολλαπλών όψεων με βάρη με ταυτόχρονη χρήση κατωφλίου συμμετοχής τουλάχιστον 5% έχει την καλύτερη επίδοση για όλους τους πυρήνες, ενώ με την τεχνική πολλαπλών πυρήνων (multi-kernel) έχει την καλύτερη επίδοση όλων. Η δεύτερη μέθοδος σε αποτελεσματικότητα είναι η χρήση πολλαπλών όψεων με βάρη καθώς υπερಿಸχύει της μεθόδου πολλαπλών όψεων με βάρη με ταυτόχρονη χρήση κατωφλίου συμμετοχής τουλάχιστον 10% στους τρεις από τους τέσσερις πυρήνες. Για τις υπόλοιπες μεθόδους δεν μπορούμε να βγάλουμε ασφαλή συμπεράσματα καθώς η απόδοσή τους κινείται σε παρόμοια επίπεδα. Τα αποτελέσματα των πειραμάτων φαίνονται αναλυτικά στα παρακάτω διαγράμματα:



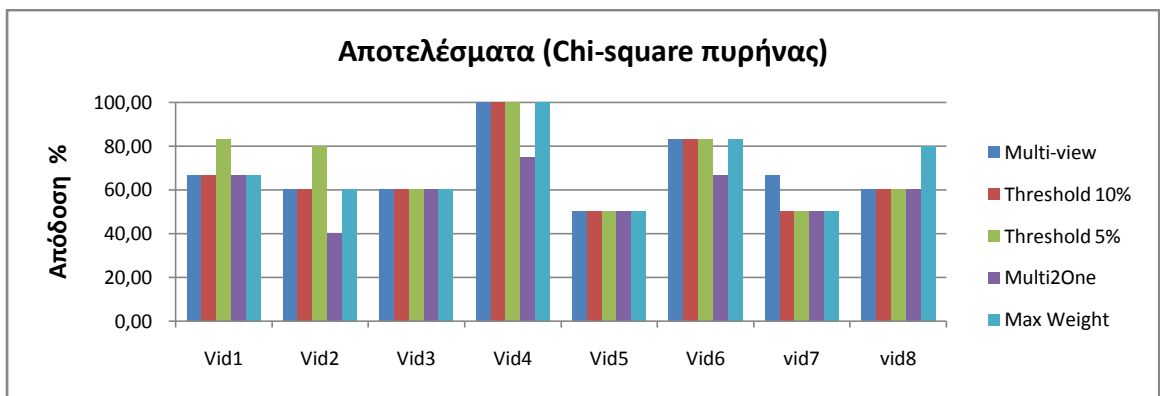
Σχήμα 7.13 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας γραμμικό (linear) πυρήνα



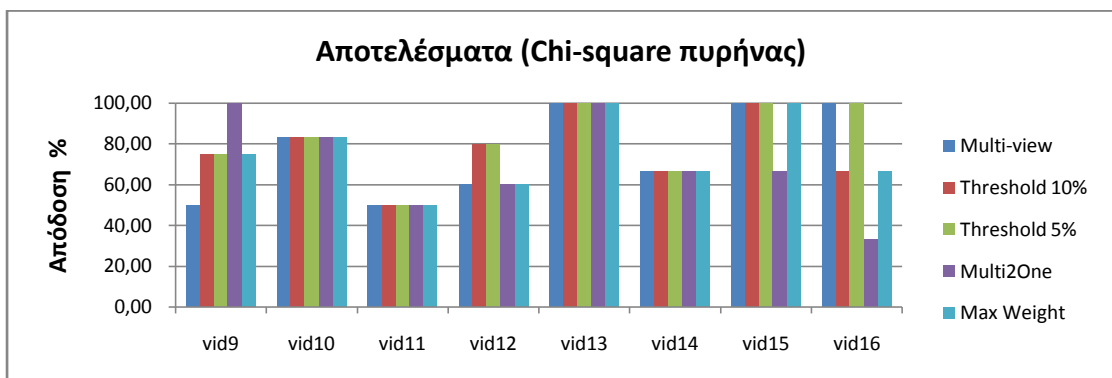
Σχήμα 7.14 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας γραμμικό (linear) πυρήνα



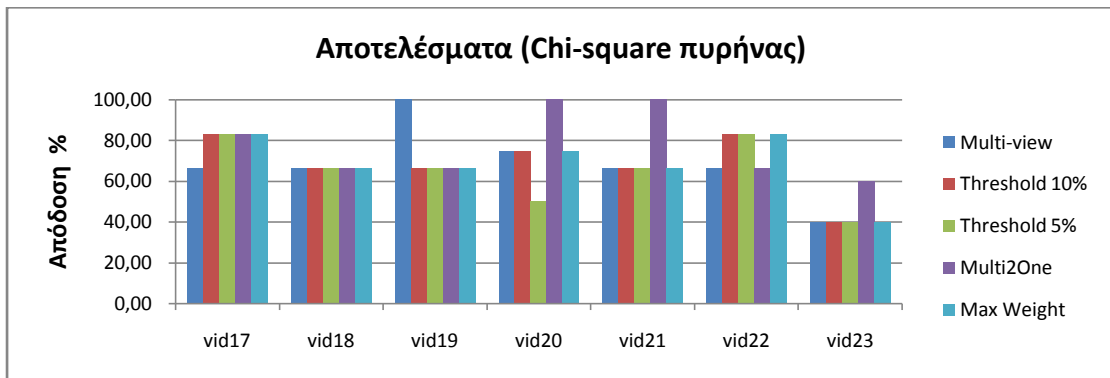
Σχήμα 7.15 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας γραμμικό (linear) πυρήνα



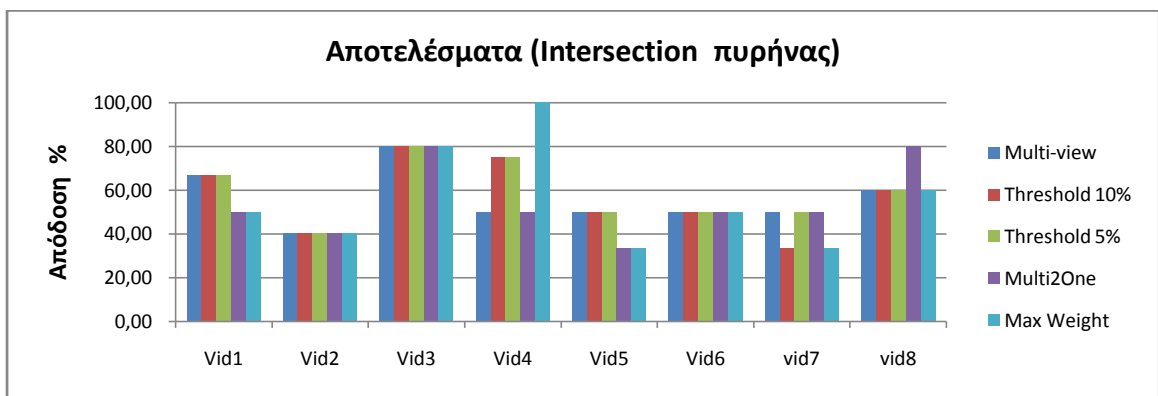
Σχήμα 7.16 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας chi-square πυρήνα



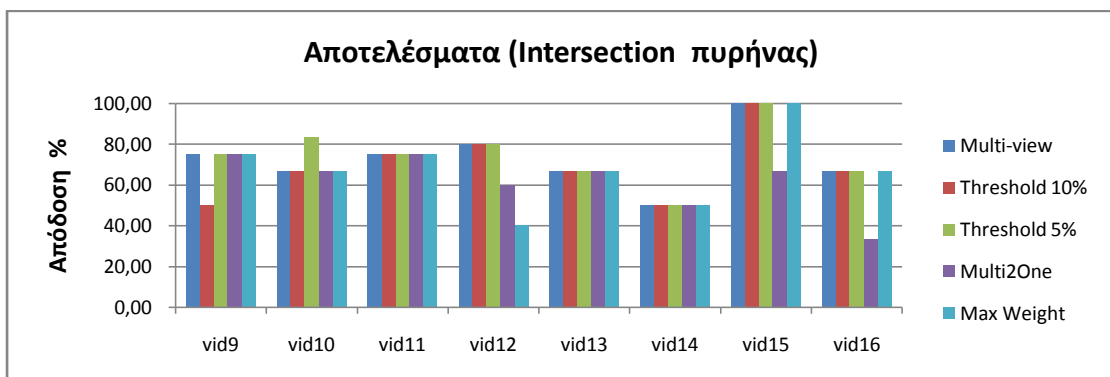
Σχήμα 7.17 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας chi-square πυρήνα



Σχήμα 7.18 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας chi-square πυρήνα

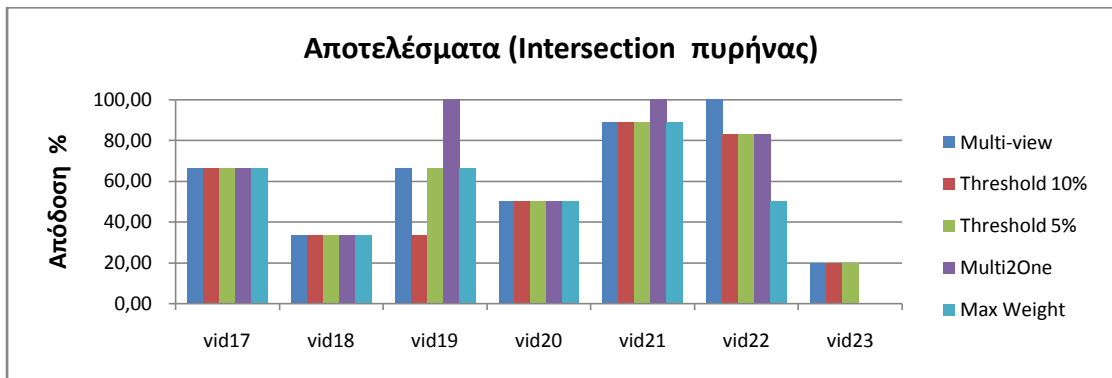


Σχήμα 7.19 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων (intersection)

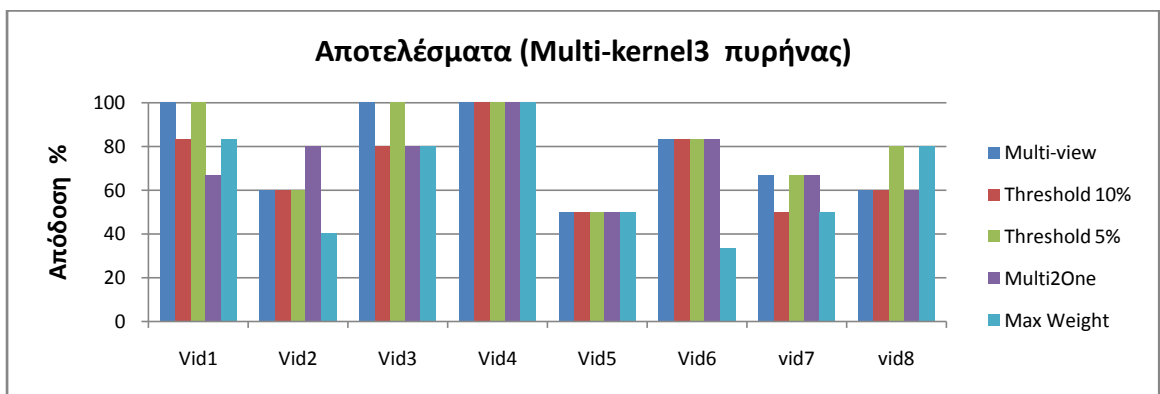


Σχήμα 7.20 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων (intersection)

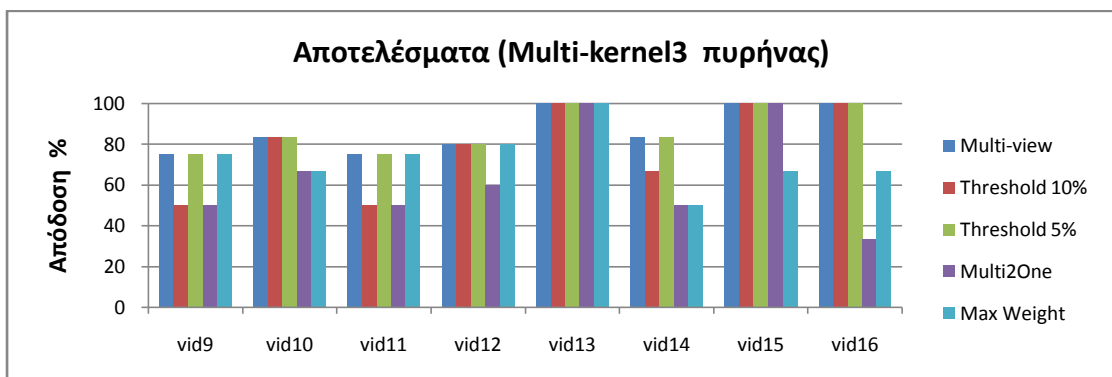




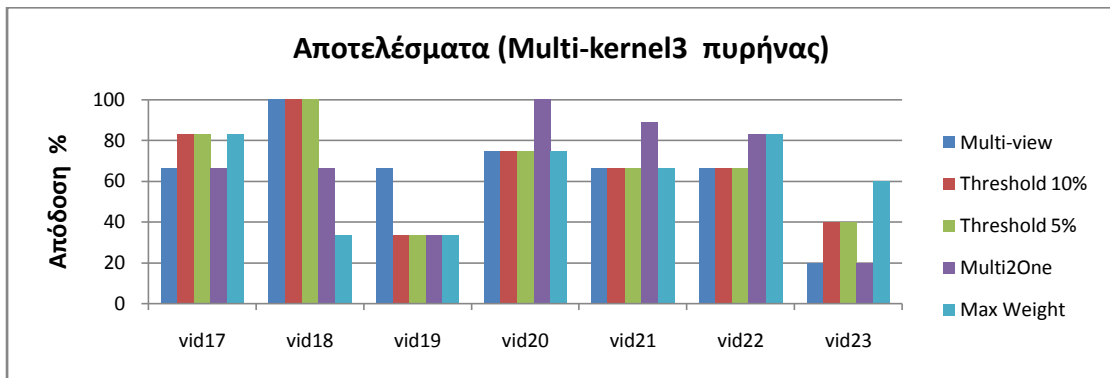
Σχήμα 7.21 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας πυρήνα τομής ιστογραμμάτων (intersection)



Σχήμα 7.22 Αποτελέσματα πειραμάτων για τις ακολουθίες 1 έως 8 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη



Σχήμα 7.23 Αποτελέσματα πειραμάτων για τις ακολουθίες 9 έως 16 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη



Σχήμα 7.24 Αποτελέσματα πειραμάτων για τις ακολουθίες 17 έως 23 χρησιμοποιώντας πυρήνα που προκύπτει από τον συνδυασμό πυρήνων με βάρη

### 7.1.3. Σύγκριση Επίδοσης Πυρήνων

Όσον αφορά τους πυρήνες που χρησιμοποιήθηκαν στον πίνακα 7.7 παρουσιάζονται οι μέσοι όροι για κάθε έναν από αυτούς.

Πίνακας 7-7 Μέσοι όροι επίδοσης για κάθε πυρήνα

Μέθοδος	Linear		Chi-Square		Intersection		Multi-kernel3	
	M.O.	std	M.O.	std	M.O.	std	M.O.	std
M.O. - std	65,87	17,25	67,50	17,34	58,42	20,86	69,50	18,63
<b>M<sub>all</sub> (%)</b>	65,19		66,96		58,78		<b>73,53</b>	
<b>MW (%)</b>	65,16		66,94		58,75		<b>68,91</b>	

Από τον πίνακα 7.7 προκύπτει ότι η αποτελεσματικότερη τεχνική είναι να χρησιμοποιηθεί συνδυασμός των τριών πυρήνων (multi-kernel3) για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων. Το αποτέλεσμα αυτό είναι και το αναμενόμενο καθώς ο αλγόριθμος έχοντας στη διάθεση του και τους τρεις πυρήνες επιλέγει κάθε φορά να αποδώσει το κατάλληλο βάρος σε κάθε έναν από αυτούς ανάλογα με την περίπτωση. Παρατηρώντας ατομικά τους τρεις πυρήνες, στα ίδια επίπεδα φαίνονται να είναι ο chi-square με τον γραμμικό ενώ ο τομής ιστογραμμάτων έχει λίγο χαμηλότερη επίδοση από τους δύο παραπάνω.

#### 7.1.4. Σύγκριση Βαρών κάθε Όψης

Στη συνέχεια ακολουθούν οι πίνακες που περιέχουν τον μέσο όρο των βαρών που αποδόθηκαν στις έξι όψεις που χρησιμοποιήθηκαν. Παρουσιάζονται και οι τρεις τρόποι χρήσης της μεθόδου πολλαπλών όψεων με βάρη.

Πίνακας 7-8 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, χωρίς κατώφλι ελαχίστου βάρους

Multiview	Linear		Chi-Square		Intersection		Multi-Kernel3	
	Mean	std	Mean	std	Mean	std	Mean	std
No Threshold								
HSV3D	<b>44,42</b>	30,96	<b>43,12</b>	33,74	<b>48,11</b>	31,41	<b>44,34</b>	35,00
CENTRIST	<b>25,36</b>	28,20	<b>16,58</b>	31,29	<b>14,49</b>	27,57	<b>26,35</b>	33,60
Wavelet	<b>16,35</b>	18,61	<b>29,59</b>	29,16	<b>30,16</b>	30,35	<b>17,45</b>	20,65
CCH <sub>desc</sub>	<b>3,81</b>	7,73	<b>3,21</b>	7,53	<b>1,77</b>	2,68	<b>4,71</b>	13,70
SIFT <sub>locs</sub>	<b>6,12</b>	9,37	<b>5,81</b>	11,43	<b>3,46</b>	6,05	<b>4,64</b>	7,71
SURF <sub>desc</sub>	<b>3,93</b>	6,22	<b>1,69</b>	3,58	<b>2,00</b>	3,28	<b>2,51</b>	4,58

Πίνακας 7-9 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση κατωφλίου ελαχίστου βάρους 5%

Multiview	Linear		Chi-Square		Intersection		Multi-Kernel3	
	Mean	std	Mean	std	Mean	std	Mean	std
5% Threshold								
HSV3D	<b>44,60</b>	32,07	<b>40,27</b>	31,91	<b>50,14</b>	31,69	<b>45,22</b>	35,44
CENTRIST	<b>26,72</b>	31,34	<b>18,10</b>	31,32	<b>15,10</b>	28,27	<b>27,21</b>	34,82
Wavelet	<b>16,71</b>	20,36	<b>32,27</b>	29,49	<b>29,70</b>	29,01	<b>17,40</b>	20,71
CCH <sub>desc</sub>	<b>2,57</b>	6,84	<b>2,70</b>	7,41	<b>0,77</b>	1,97	<b>3,31</b>	9,29
SIFT <sub>locs</sub>	<b>6,09</b>	10,82	<b>5,60</b>	10,68	<b>2,68</b>	6,61	<b>4,81</b>	8,82
SURF <sub>desc</sub>	<b>3,30</b>	6,35	<b>1,07</b>	3,58	<b>1,61</b>	3,81	<b>2,06</b>	4,81

Πίνακας 7-10 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση κατωφλίου ελάχιστου βάρους 10%

<b>Multiview</b>	<b>Linear</b>		<b>Chi-Square</b>		<b>Intersection</b>		<b>Multi-Kernel3</b>	
	<b>Mean</b>	<b>std</b>	<b>Mean</b>	<b>std</b>	<b>Mean</b>	<b>std</b>	<b>Mean</b>	<b>std</b>
<b>10% Threshold</b>								
<b>HSV3D</b>	<b>45,84</b>	32,57	<b>44,64</b>	36,37	<b>52,91</b>	34,09	<b>46,15</b>	37,33
<b>CENTRIST</b>	<b>26,51</b>	31,94	<b>15,27</b>	31,88	<b>14,33</b>	30,67	<b>25,88</b>	35,56
<b>Wavelet</b>	<b>17,52</b>	21,31	<b>31,60</b>	30,13	<b>30,21</b>	30,34	<b>17,60</b>	20,80
<b>CCH<sub>desc</sub></b>	<b>2,60</b>	7,49	<b>2,64</b>	10,22	<b>0,00</b>	0,00	<b>4,44</b>	16,83
<b>SIFT<sub>locs</sub></b>	<b>5,73</b>	11,48	<b>4,58</b>	10,82	<b>1,93</b>	6,64	<b>4,40</b>	8,81
<b>SURF<sub>desc</sub></b>	<b>1,79</b>	5,83	<b>1,28</b>	4,42	<b>0,62</b>	2,97	<b>1,53</b>	4,29

Από τους παραπάνω πίνακες παρατηρούμε ότι σε όλες τις περιπτώσεις η όψη HSV3D έχει τη μεγαλύτερη συμμετοχή, κατά μέσο όρο, στην εξαγωγή των χαρακτηριστικών εικονοπλαισίων. Επίσης, οι περιγραφείς CCH, SIFT και SURF δεν έχουν μεγάλο βάρος. Αυτό οφείλεται στο γεγονός ότι οι όψεις HSV, CENTRIST και Wavelet έχουν block μορφή, η οποία βοηθάει στην ομαδοποίηση, ενώ οι όψεις CCH, SIFT και SURF δεν έχουν τόσο σαφή block μορφή. Επιπλέον, μπορούμε να παρατηρήσουμε ότι το βάρος των CCH, SIFT και SURF παίρνει τις μικρότερες τιμές στην περίπτωση χρήσης κατωφλίου ελάχιστης συμμετοχής 10%. Το γεγονός αυτό επιβεβαιώνει το συμπέρασμά μας ότι ακόμα και η μικρή συμμετοχή, της τάξης του 5%, μίας όψης είναι προτιμότερη από την αφαίρεσή της αφού η χρήση κατωφλίου ελάχιστης συμμετοχής 5% απέφερε τα καλύτερα αποτελέσματα. Η σημαντικότερη όμως πληροφορία είναι ότι σε κάθε περίπτωση όλες οι όψεις συμμετέχουν στην επιλογή των χαρακτηριστικών εικονοπλαισίων και ανεξάρτητα από το βάρος τους παίζουν σημαντικό ρόλο στην εξαγωγή του τελικού αποτελέσματος, κάτι που επιβεβαιώνεται από την συνολικά καλύτερη επίδοση της μεθόδου πολλαπλών όψεων με βάρη. Επίσης, η μεγάλη διακύμανση που παρατηρείται στα βάρη κάθε όψης δείχνει ότι δεν υπάρχει κάποια όψη που να υπερισχύει απόλυτα των υπολοίπων αλλά αντιθέτως ανάλογα με την ακολουθία που εξετάζεται τα βάρη κατανέμονται κατάλληλα.

### 7.1.5. Σύγκριση Βαρών Πυρήνων της Μεθόδου Multi-kernel3

Επιπλέον, παρουσιάζεται η κατά μέσο όρο κατανομή των βαρών σε κάθε πυρήνα κατά τη διαδικασία δημιουργίας ενός συνολικού πυρήνα που προκύπτει από τον συνδυασμό των πυρήνων γραμμικός (linear), chi-square και τομής ιστογραμμάτων (intersection). Οι τιμές των μέσων όρων συμμετοχής για κάθε όψη ξεχωριστά φαίνονται στον πίνακα 7.11.

Πίνακας 7-11 Μέσοι όροι του ποσοστού συμμετοχής κάθε πυρήνα στη δημιουργία συνολικού πυρήνα (multi-kernel3) για κάθε όψη ξεχωριστά και συνολικά

Όψη	Linear		Chi-square		Intersection	
	Mean	std	Mean	std	Mean	std
<b>CCH<sub>desc</sub></b>	<b>74,10</b>	16,58	<b>24,56</b>	15,32	<b>1,34</b>	2,21
<b>CENTRIST</b>	<b>84,44</b>	15,30	<b>14,70</b>	14,38	<b>0,86</b>	1,83
<b>HSV3D</b>	<b>55,18</b>	12,74	<b>42,53</b>	12,39	<b>2,29</b>	3,08
<b>SIFT<sub>locs</sub></b>	<b>65,37</b>	13,54	<b>32,16</b>	11,61	<b>2,47</b>	2,95
<b>SURF<sub>desc</sub></b>	<b>59,43</b>	13,02	<b>37,44</b>	10,63	<b>3,13</b>	3,03
<b>Wavelet</b>	<b>76,56</b>	13,02	<b>20,94</b>	10,63	<b>2,50</b>	3,03
<b>Συνολικά</b>	<b>69,18</b>	14,03	<b>28,72</b>	12,49	<b>2,10</b>	2,69

Από τον παραπάνω πίνακα παρατηρούμε ότι τη μεγαλύτερη συμμετοχή στην εξαγωγή του multi-kernel πυρήνα έχει ο γραμμικός (linear) πυρήνας ενώ ο πυρήνας τομής ιστογραμμάτων (intersection) συμμετέχει ελάχιστα. Το γεγονός αυτό δικαιολογείται εν μέρει αν κοιτάξει κανείς τον πίνακα 7.7 στον οποίο φαίνεται ότι ο πυρήνας τομής ιστογραμμάτων (intersection) κινείται σε χαμηλότερα επίπεδα επίδοσης από τον γραμμικό και τον chi-square πυρήνα.

### 7.1.6. Σύγκριση Παραλλαγών Περιγραφέων CCH, SIFT και SURF

Τέλος, παρουσιάζονται τα αποτελέσματα κάθε περιγραφέα CCH, SIFT και SURF για 20, 50 και 100 οπτικές λέξεις καθώς επίσης και για τον συνδυασμό αυτών (Multi).

Πίνακας 7-12 Μέσοι όροι και διακύμανση της επίδοσης κάθε παραλλαγής περιγραφέα CCH, SIFT και SURF

Όψη	Linear		ChiSquare		Intersection	
	M.O.	Std	M.O.	Std	M.O.	Std
CCH20	<u>61,04</u>	<u>17,91</u>	48,96	15,33	60,00	19,59
CCH50	61,38	22,79	<b><u>62,83</u></b>	<b><u>16,56</u></b>	49,18	19,58
CCH100	58,53	20,59	50,31	18,07	<u>60,70</u>	<u>20,64</u>
CCH-Multi	61,06	15,67	<u>61,09</u>	<u>13,65</u>	57,32	17,63
SIFT20	59,49	14,69	60,46	18,70	<u>64,61</u>	<u>24,14</u>
SIFT50	63,77	16,08	<u>64,69</u>	<u>17,23</u>	55,95	20,84
SIFT100	64,98	19,43	56,11	22,24	<b><u>68,16</u></b>	<b><u>17,17</u></b>
SIFT-Multi	<u>67,13</u>	<u>19,69</u>	63,86	15,00	61,16	19,13
SURF20	<b><u>63,74</u></b>	<b><u>13,72</u></b>	55,77	13,79	58,16	16,58
SURF50	<u>54,32</u>	<u>18,15</u>	52,03	16,66	52,93	17,80
SURF100	60,82	16,69	47,73	17,24	<u>60,63</u>	<u>14,53</u>
SURF-Multi	<u>61,06</u>	<u>18,23</u>	57,95	16,61	59,86	20,26

Συγκρίνοντας τις όψεις μεταξύ τους δεν μπορούμε να εξάγουμε ασφαλή συμπεράσματα σχετικά με την απόδοσή τους. Αν παρατηρήσουμε όμως συνολικά την επίδοση κάθε όψης και για τους τρεις πυρήνες, εξετάζοντας το συνολικό μέσο όρο, τη διακύμανση του συνολικού μέσου όρου αλλά και το μέσο όρο της διακύμανσης κάθε πυρήνα (βλ. πίνακα 7.13) συμπεραίνουμε ότι συνολικά οι περιγραφείς Multi είναι αποτελεσματικότεροι από τους περιγραφείς που χρησιμοποιούν 20, 50 και 100 οπτικές λέξεις. Επιπλέον, μία πολύ σημαντική παρατήρηση είναι ότι η διακύμανση των μέσων όρων των περιγραφέων Multi είναι πολύ χαμηλή, γεγονός που αποδεικνύει ότι έχουν σταθερή επίδοση ανεξάρτητα από τον πυρήνα που χρησιμοποιήθηκε. Τέλος, παρατηρείται ότι και ο μέσος όρος της διακύμανσης είναι χαμηλότερος στους περιγραφείς Multi, εξαιρουμένου του περιγραφέα SURF, γεγονός που αποδεικνύει ότι έχουν πιο σταθερή επίδοση, σε σχέση με τους single περιγραφείς, ανεξάρτητα από την ακολουθία βίντεο που εξετάζεται. Βέβαια, πρέπει να αναφερθεί ότι και πάλι η διακύμανση κυμαίνεται σε υψηλά επίπεδα.

Πίνακας 7-13 Μέσος όρος επίδοσης, διακύμανση μέσου όρου επίδοσης και μέσος όρος διακύμανσης για κάθε παραλλαγή περιγραφέα

Όψη	Συνολικά		
	M.O.	Std M.O.	M.O. Std
CCH20	56,67	6,69	17,61
CCH50	57,80	7,50	19,65
CCH100	56,51	5,48	19,77
<b>CCH-Multi</b>	<b>59,82</b>	<b>2,17</b>	<b>15,65</b>
SIFT20	61,52	2,72	19,18
SIFT50	61,47	4,80	18,05
SIFT100	63,08	6,25	19,61
<b>SIFT-Multi</b>	<b>64,05</b>	<b>2,99</b>	<b>17,94</b>
SURF20	59,23	4,09	14,70
SURF50	53,09	1,16	17,53
SURF100	56,39	7,50	16,15
<b>SURF-Multi</b>	<b>59,62</b>	<b>1,57</b>	<b>18,37</b>

Πίνακας 7-14 Διαφορά (ως προς το μέσο όρο) κάθε παραλλαγής περιγραφέα CCH, SIFT και SURF σε σχέση με τον αντίστοιχο Multi περιγραφέα. Οι θετικές τιμές υποδεικνύουν υπεροχή του Multi περιγραφέα.

Όψη	Linear	ChiSquare	Intersection	M.O. Διαφοράς
CCH20	0,02	12,13	-2,68	<b>3,16</b>
CCH50	-0,32	-1,74	8,14	<b>2,03</b>
CCH100	2,54	10,77	-3,38	<b>3,31</b>
SIFT20	7,63	3,41	-3,45	<b>2,53</b>
SIFT50	3,36	-0,82	5,21	<b>2,58</b>
SIFT100	2,15	7,75	-7,00	<b>0,97</b>
SURF20	-2,68	2,17	1,69	<b>0,39</b>
SURF50	6,74	5,92	6,93	<b>6,53</b>
SURF100	0,24	10,22	-0,77	<b>3,23</b>

## 7.2. Δεύτερη Σειρά Πειραμάτων

Παρακάτω ακολουθούν τα πειραματικά αποτελέσματα της δεύτερης σειράς πειραμάτων. Για την αξιολόγησή των αποτελεσμάτων του αλγορίθμου χρησιμοποιήθηκε ground truth επιλεγμένο από διαφορετικό χρήστη προκειμένου τα αποτελέσματα να είναι περισσότερο αντικειμενικά. Σε κάθε πίνακα υπογραμμισμένη είναι η καλύτερη επίδοση που είχε η κάθε μέθοδος συνολικά και με έντονα γράμματα είναι η καλύτερη επίδοση συγκρίνοντας όλες τις μεθόδους, για κάθε πυρήνα.

### 7.2.1. Σύγκριση Single-view με Multi-view

Στον πίνακα 7.15 παρουσιάζονται τα αποτελέσματα της επίδοσης του αλγορίθμου για κάθε όψη ξεχωριστά αλλά και για το συνδυασμό αυτών.

Πίνακας 7-15 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το μέσο όρο και τη διακύμανση

Μέθοδος	Linear		Chi-square		Intersection		Multi-Kernel3		Multi-Kernel2											
	M.O.	Std	M.O.	Std	M.O.	Std	M.O.	Std	M.O.	Std										
HSV - Multi	65,39%	17,79%	68,26%	18,91%	62,30%	16,13%	<u>68,35%</u>	<u>13,88%</u>	67,04%	15,31%										
HSV - Avg	66,09%	13,94%	<u>67,74%</u>	<u>12,16%</u>	65,70%	13,25%	66,13%	12,20%	67,22%	13,94%										
HSV3D	<u>70,70%</u>	<u>14,87%</u>	68,09%	11,61%	61,22%	14,58%	62,35%	17,60%	63,39%	16,80%										
CENTRIST	<u>68,96%</u>	<u>17,86%</u>	67,70%	15,76%	54,96%	19,37%	61,70%	15,98%	60,61%	15,82%										
Wavelet	65,30%	13,92%	64,74%	15,82%	53,91%	19,49%	67,13%	16,54%	<u>69,09%</u>	<u>18,03%</u>										
CCH - Multi	63,13%	19,53%	62,17%	18,71%	53,48%	16,29%	<u>69,57%</u>	<u>9,31%</u>	67,87%	11,51%										
SIFT - Multi	63,09%	15,64%	65,00%	13,54%	58,57%	14,49%	<u>71,57%</u>	<u>15,00%</u>	71,13%	10,03%										
SURF - Multi	64,87%	13,00%	66,83%	12,32%	65,00%	21,04%	71,74%	10,49%	<u>72,22%</u>	<u>9,11%</u>										
HSV-Multi	69,04%	13,86%	72,26%	14,74%	63,52%	14,80%	<u>75,52%</u>	<u>13,77%</u>	73,43%	9,47%										
HSV-Avg											72,13%	14,62%	76,30%	12,03%	63,04%	16,08%	75,26%	9,74%	<b>75,35%</b>	<b>9,46%</b>
HSV3D											71,35%	14,48%	<u>75,70%</u>	<u>15,05%</u>	62,87%	16,20%	73,35%	12,64%	73,35%	10,41%

Από τον παραπάνω πίνακα παρατηρούμε ότι συνολικά την καλύτερη επίδοση έχει η μέθοδος πολλαπλών όψεων με βάρη χρησιμοποιώντας την όψη HSV-Avg και πυρήνα multi-kernel2. Χωρίς μεγάλη διαφορά στην επίδοση ακολουθούν οι άλλες δύο μέθοδοι συνδυασμού πολλαπλών όψεων που χρησιμοποιούν HSV3D και HSV-Multi.



Επίσης, παρατηρώντας τις καλύτερες επιδόσεις της κάθε μεθόδου, οι υπογραμμισμένες τιμές του πίνακα, φαίνεται ότι όλες οι μέθοδοι single-view κινούνται σε ποσοστό επιτυχίας περίπου 70% ενώ οι μέθοδοι multi-view έχουν μεγαλύτερο ποσοστό επιτυχίας κατά περίπου 5%. Όμως, το σημαντικότερο πλεονέκτημα των multi-view μεθόδων είναι η σταθερή απόδοσή τους πάνω από 70%, αν εξαιρέσουμε την περίπτωση του intersection πυρήνα όπου όλες οι μέθοδοι παρουσιάζουν πτώση επίδοσης, ανεξάρτητα από τον πυρήνα που χρησιμοποιήθηκε.

Για την υποστήριξη των αποτελεσμάτων λόγω της αυξημένης διακύμανσης η μέση τιμή της επίδοσης της κάθε μεθόδου υπολογίστηκε με δύο ακόμα τρόπους (βλ. κεφάλαιο 6.6). Τα αποτελέσματά τους παρουσιάζονται στους πίνακες 7.16 και 7.17 .

Πίνακας 7-16 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το συνολικό μέσο όρο

Μέθοδος	Linear	Chi-square	Intersection	Multi-Kernel3	Multi-Kernel2	
	M <sub>all</sub>	M <sub>all</sub>	M <sub>all</sub>	M <sub>all</sub>	M <sub>all</sub>	
HSV - Multi	65,18%	66,07%	62,50%	<u>68,75%</u>	66,96%	
HSV - Avg	66,07%	66,96%	66,07%	66,96%	<u>67,86%</u>	
HSV3D	<u>69,64%</u>	66,96%	61,61%	59,82%	60,71%	
CENTRIST	<u>67,86%</u>	<u>67,86%</u>	54,46%	60,71%	59,82%	
Wavelet	61,61%	64,29%	52,68%	66,96%	<u>68,75%</u>	
CCH - Multi	58,93%	60,71%	52,68%	<u>69,64%</u>	67,86%	
SIFT - Multi	61,61%	66,07%	58,93%	<u>71,43%</u>	70,54%	
SURF - Multi	63,39%	66,07%	64,29%	71,43%	<u>72,32%</u>	
HSV-Multi	66,96%	71,43%	64,29%	<u>75,00%</u>	72,32%	
HSV-Avg						Multiview
HSV3D						
HSV3D	70,54%	<u>75,89%</u>	64,29%	73,21%	73,21%	

Πίνακας 7-17 Αποτελέσματα της επίδοσης κάθε μεθόδου με βάση το μέσο όρο με βάρη

Μέθοδος	Linear	Chi-square	Intersection	Multi-Kernel3	Multi-Kernel2	
	MW	MW	MW	MW	MW	
HSV - Multi	64,27%	66,54%	62,35%	<u>68,43%</u>	67,00%	
HSV - Avg	65,74%	67,26%	65,94%	66,85%	<u>67,74%</u>	
HSV3D	<u>69,34%</u>	67,25%	60,92%	60,40%	61,33%	
CENTRIST	67,95%	<u>68,18%</u>	54,04%	60,18%	59,29%	
Wavelet	62,24%	64,34%	52,47%	66,69%	<u>68,47%</u>	
CCH - Multi	59,25%	60,46%	52,60%	<u>70,40%</u>	68,12%	
SIFT - Multi	61,67%	65,13%	58,72%	<u>71,65%</u>	70,85%	
SURF - Multi	63,74%	66,88%	64,72%	<u>72,19%</u>	72,01%	
HSV-Multi	Multiview	66,96%	71,63%	64,33%	<u>75,19%</u>	72,28%
HSV-Avg		71,95%	<u>75,54%</u>	63,74%	73,69%	74,67%
HSV3D		70,55%	<u>75,50%</u>	64,29%	72,93%	72,93%

Οι πίνακες 7.16 και 7.17 επιβεβαιώνουν το συμπέρασμα ότι η μέθοδος που προτείνουμε έχει σταθερά πολύ καλά αποτελέσματα ανεξάρτητα από τον πυρήνα που θα χρησιμοποιηθεί, με μοναδική εξαίρεση τον πυρήνα τομής ιστογραμμάτων για τον οποίο παρατηρείται πτώση στην επίδοση όλων των μεθόδων. Οι αποδόσεις της μεθόδου πολλαπλών όψεων με βάρη είναι πολύ κοντά μεταξύ τους και δε μας δίνουν τη δυνατότητα να αποφανθούμε με βεβαιότητα σχετικά με το ποιος συνδυασμός είναι αποτελεσματικότερος.

### 7.2.2. Σύγκριση Επίδοσης Πυρήνων

Στη συνέχεια αναλύεται η επίδοση του κάθε πυρήνα ξεχωριστά. Οι μέσοι όροι επίδοσης του κάθε πυρήνα παρουσιάζονται στον πίνακα 7.18 .

Πίνακας 7-18 Μέσοι όροι επίδοσης για κάθε πυρήνα

Μέθοδος	Linear		Chi-square		Intersection		Multi-Kernel3		Multi-Kernel2	
	M.O. - std	67,28%	15,41%	68,62%	14,60%	60,42%	16,52%	<b>69,33%</b>	<b>13,38%</b>	69,15%
M <sub>all</sub>	65,75%		68,02%		60,55%		<b>68,83%</b>		68,67%	
MW	65,79%		68,06%		60,38%		<b>68,96%</b>		68,61%	

Όπως και στην πρώτη σειρά πειραμάτων, η καλύτερη επίδοση, κατά μέσο όρο, παρατηρείτε χρησιμοποιώντας συνδυασμό πυρήνων (Multi-Kernel3 και Multi-Kernel2) όντας λίγο καλύτεροι από τον chi-square. Ακολουθεί ο γραμμικός πυρήνας, και τελευταίος σε επίδοση πυρήνας είναι αυτός της τομής ιστογραμμάτων.

### 7.2.3. Σύγκριση Βαρών κάθε Όψης

Εκτός από την επίδοση της κάθε όψης, είναι σημαντικό να παρατηρήσουμε το ποσοστό με το οποίο συμμετείχε κάθε όψη στη διαδικασία εξαγωγής των χαρακτηριστικών εικονοπλαισίων όταν συνδυάστηκαν όλες μαζί. Παρακάτω παρουσιάζονται τρεις πίνακες που περιέχουν το βάρος κάθε όψης σε κάθε περίπτωση. Η διαφορά στις τρεις περιπτώσεις εντοπίζεται στην επιλογή του ιστογράμματος χρώματος που χρησιμοποιείται, HSV-Multi, HSV-Avg και HSV-3D.

Πίνακας 7-19 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση της όψης HSV-Multi

		HSV-Multi	CENTRIST	Wavelet	CCH-Multi	SIFT-Multi	SURF-Multi
Linear	M.O.	<b>45,74%</b>	<b>14,88%</b>	<b>16,51%</b>	<b>9,99%</b>	<b>7,80%</b>	<b>5,09%</b>
	std	33,01%	19,29%	24,47%	13,23%	9,41%	6,33%
Chi-square	M.O.	<b>58,67%</b>	<b>6,25%</b>	<b>18,22%</b>	<b>6,05%</b>	<b>6,77%</b>	<b>4,04%</b>
	std	32,43%	14,48%	23,27%	14,00%	8,82%	5,61%
Intersection	M.O.	<b>62,48%</b>	<b>4,58%</b>	<b>21,26%</b>	<b>4,15%</b>	<b>4,79%</b>	<b>2,73%</b>
	std	29,36%	8,50%	25,23%	6,74%	7,43%	4,38%
Multi-Kernel $\beta$	M.O.	<b>45,15%</b>	<b>17,58%</b>	<b>16,29%</b>	<b>9,65%</b>	<b>6,73%</b>	<b>4,59%</b>
	std	34,25%	26,66%	20,55%	16,62%	8,98%	6,30%

Πίνακας 7-20 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση της όψης HSV-Avg

		HSV-Avg	CENTRIST	Wavelet	CCH-Multi	SIFT-Multi	SURF-Multi
Linear	M.O.	<b>38,69%</b>	<b>19,19%</b>	<b>17,44%</b>	<b>11,92%</b>	<b>8,14%</b>	<b>4,61%</b>
	std	33,67%	24,61%	24,75%	19,07%	9,07%	5,84%
Chi-square	M.O.	<b>42,61%</b>	<b>8,27%</b>	<b>24,10%</b>	<b>7,80%</b>	<b>10,23%</b>	<b>6,99%</b>
	std	31,86%	15,72%	27,33%	14,79%	11,59%	9,62%
Intersection	M.O.	<b>53,45%</b>	<b>6,17%</b>	<b>26,50%</b>	<b>5,07%</b>	<b>5,86%</b>	<b>2,96%</b>
	std	32,45%	10,48%	31,06%	9,18%	8,48%	4,61%
Multi-Kernel $\beta$	M.O.	<b>36,72%</b>	<b>20,94%</b>	<b>18,90%</b>	<b>11,13%</b>	<b>6,61%</b>	<b>5,71%</b>
	std	34,97%	28,40%	21,77%	20,05%	7,65%	6,66%

Πίνακας 7-21 Μέσοι όροι και διακύμανση των βαρών κάθε όψης, για κάθε πυρήνα, με χρήση της όψης HSV-3D

		HSV3D	CENTRIST	Wavelet	CCH-Multi	SIFT-Multi	SURF-Multi
Linear	M.O.	<b>38,23%</b>	<b>19,84%</b>	<b>17,10%</b>	<b>12,07%</b>	<b>8,46%</b>	<b>4,29%</b>
	std	32,60%	25,20%	23,56%	19,31%	9,70%	5,84%
Chi-square	M.O.	<b>37,27%</b>	<b>8,62%</b>	<b>26,47%</b>	<b>8,50%</b>	<b>11,54%</b>	<b>7,60%</b>
	std	29,48%	15,91%	27,84%	14,73%	12,10%	9,54%
Intersection	M.O.	<b>49,48%</b>	<b>6,92%</b>	<b>26,48%</b>	<b>6,30%</b>	<b>6,86%</b>	<b>3,96%</b>
	std	32,76%	10,79%	29,83%	11,17%	9,22%	5,54%
Multi-Kernel3	M.O.	<b>35,81%</b>	<b>20,43%</b>	<b>19,39%</b>	<b>12,05%</b>	<b>7,00%</b>	<b>5,32%</b>
	std	34,48%	28,06%	22,71%	22,32%	8,64%	6,44%

Στους τρεις παραπάνω πίνακες είναι εύκολο να παρατηρήσει κανείς ότι σε όλες τις περιπτώσεις το μεγαλύτερο βάρος δίνεται στην όψη HSV ενώ ακολουθούν οι όψεις Wavelet και CENTRIST. Στις τελευταίες θέσεις, όπως και στην πρώτη σειρά πειραμάτων, βρίσκονται οι περιγραφείς CCH, SIFT και SURF. Η σημαντικότερη όμως πληροφορία που εξάγεται από τους παραπάνω πίνακες είναι ότι όλες οι όψεις συμμετέχουν στην επιλογή των χαρακτηριστικών εικονοπλαισίων και ανεξάρτητα από το βάρος τους παίζουν σημαντικό ρόλο στην εξαγωγή του τελικού αποτελέσματος, κάτι που επιβεβαιώνεται από την συνολικά καλύτερη επίδοση της μεθόδου πολλαπλών όψεων με βάρη. Επίσης, η μεγάλη διακύμανση στο βάρος κάθε όψης δείχνει ότι δεν υπάρχει κάποια όψη που να υπερισχύει απόλυτα των υπολοίπων αλλά αντιθέτως ανάλογα με την ακολουθία βίντεο που εξετάζεται τα βάρη κατανέμονται κατάλληλα.

#### 7.2.4. Σύγκριση Βαρών Πυρήνων της Μεθόδου Multi-kernel3

Στη συνέχεια, παρουσιάζεται η κατά μέσο όρο κατανομή των βαρών σε κάθε πυρήνα κατά τη διαδικασία δημιουργίας ενός συνολικού πυρήνα που προκύπτει από τον συνδυασμό των πυρήνων γραμμικός (linear), chi-square και τομής ιστογραμμάτων (intersection). Οι τιμές των μέσων όρων συμμετοχής για κάθε όψη ξεχωριστά φαίνονται στον πίνακα 7.22.

Πίνακας 7-22 Μέσοι όροι του ποσοστού συμμετοχής κάθε πυρήνα στη δημιουργία συνολικού πυρήνα (multi-kernel3) για κάθε όψη ξεχωριστά

Όψη	Linear		Chi-square		Intersection	
	Mean	Std	Mean	Std	Mean	Std
<b>HSV-Multi</b>	<b>59,58%</b>	15,60%	<b>38,71%</b>	14,81%	<b>1,70%</b>	2,03%
<b>HSV-Avg</b>	<b>85,16%</b>	14,38%	<b>13,67%</b>	12,81%	<b>1,17%</b>	2,19%
<b>HSV3D</b>	<b>57,30%</b>	14,55%	<b>40,41%</b>	13,60%	<b>2,29%</b>	2,97%
<b>CENTRIST</b>	<b>69,16%</b>	15,05%	<b>28,30%</b>	13,75%	<b>2,54%</b>	2,32%
<b>Wavelet</b>	<b>62,85%</b>	11,26%	<b>35,20%</b>	10,66%	<b>1,95%</b>	2,12%
<b>CCH-Multi</b>	<b>71,03%</b>	15,49%	<b>27,06%</b>	14,14%	<b>1,91%</b>	2,47%
<b>SIFT-Multi</b>	<b>67,01%</b>	15,65%	<b>30,59%</b>	14,10%	<b>2,40%</b>	3,14%
<b>SURF-Multi</b>	<b>71,36%</b>	15,27%	<b>27,02%</b>	14,11%	<b>1,62%</b>	2,81%

Από τον παραπάνω πίνακα βλέπουμε ότι μεγαλύτερη συμμετοχή στη δημιουργία του συνολικού πυρήνα έχει ο γραμμικός πυρήνας ενώ ο πυρήνας τομής ιστογραμμάτων συμμετέχει ελάχιστα. Το γεγονός αυτό δικαιολογείται εν μέρει από τη χαμηλότερη απόδοσή του σε σχέση με τους άλλους δύο πυρήνες (βλ. πίνακα 7.16).

### 7.2.5. Σύγκριση Παραλλαγών Περιγραφέων CCH, SIFT και SURF

Τέλος, στον παρακάτω πίνακα παρουσιάζονται τα αποτελέσματα κάθε περιγραφέα CCH, SIFT και SURF για 20, 50 και 100 οπτικές λέξεις καθώς επίσης και για τον συνδυασμό αυτών (Multi).

Πίνακας 7-23 Μέσοι όροι της επίδοσης κάθε παραλλαγής περιγραφέα ξεχωριστά για κάθε πυρήνα και συνολικά για όλους τους πυρήνες

Όψη	Linear	ChiSquare	Intersection	Συνολικά	
	M.O.	M.O.	M.O.	M.O.	Std του M.O
<b>CCH20</b>	55,85%	<u>61,41%</u>	55,20%	57,48%	3%
<b>CCH50</b>	<u>60,73%</u>	59,69%	56,43%	58,95%	2%
<b>CCH100</b>	<u>63,35%</u>	57,88%	57,62%	59,62%	3%
<b>CCH-Multi</b>	<u>63,14%</u>	62,13%	53,43%	<b>59,57%</b>	5%
<b>SIFT20</b>	60,67%	<u>62,52%</u>	59,40%	60,86%	2%
<b>SIFT50</b>	56,76%	<u>62,01%</u>	54,40%	57,72%	4%
<b>SIFT100</b>	59,95%	<u>62,56%</u>	52,01%	58,17%	5%
<b>SIFT-Multi</b>	63,06%	<u>64,98%</u>	58,53%	<b>62,19%</b>	3%
<b>SURF20</b>	60,20%	<u>62,05%</u>	56,53%	59,59%	3%
<b>SURF50</b>	<u>64,33%</u>	63,72%	58,35%	62,13%	3%
<b>SURF100</b>	61,86%	<u>65,20%</u>	56,68%	61,25%	4%
<b>SURF-Multi</b>	64,80%	<b>66,76%</b>	64,94%	<b>65,50%</b>	1%

Από τον παραπάνω πίνακα αν παρατηρήσουμε το μέσο όρο της επίδοσης της κάθε όψης για όλους τους πυρήνες συμπεράνουμε ότι οι συνδυασμοί των οπτικών λέξεων επιτυγχάνουν ίδιο ή καλύτερο αποτέλεσμα, κάτι που φαίνεται καλύτερα στον επόμενο πίνακα. Επίσης, η όψη SURF-Multi παρουσιάζει ιδιαίτερα σταθερή επίδοση ανεξάρτητα από τον πυρήνα που εφαρμόζεται έχοντας ταυτόχρονα και την καλύτερη επίδοση από τις υπόλοιπες όψεις. Τέλος, αν εξετάσουμε ξεχωριστά την επίδοση των 20, 50 και 100 οπτικών λέξεων φαίνεται ότι οι περιγραφείς CCH, SIFT και SURF αποδίδουν καλύτερα με 100, 25 και 50 οπτικές λέξεις αντίστοιχα.

Πίνακας 7-24 Διαφορά (ως προς το μέσο όρο) κάθε παραλλαγής περιγραφέα CCH, SIFT και SURF σε σχέση με τον αντίστοιχο Multi περιγραφέα. Οι θετικές τιμές υποδεικνύουν υπεροχή του Multi περιγραφέα.

Όψη	Linear	ChiSquare	Intersection	Μ.Ο. Διαφοράς
<b>CCH20</b>	7,29%	0,72%	-1,77%	<b>2,08%</b>
<b>CCH50</b>	2,40%	2,44%	-3,00%	<b>0,61%</b>
<b>CCH100</b>	-0,22%	4,25%	-4,19%	<b>-0,05%</b>
<b>SIFT20</b>	2,39%	2,46%	-0,87%	<b>1,33%</b>
<b>SIFT50</b>	6,30%	2,97%	4,13%	<b>4,47%</b>
<b>SIFT100</b>	3,12%	2,42%	6,52%	<b>4,02%</b>
<b>SURF20</b>	4,61%	4,71%	8,41%	<b>5,91%</b>
<b>SURF50</b>	0,48%	3,04%	6,58%	<b>3,37%</b>
<b>SURF100</b>	2,94%	1,56%	8,26%	<b>4,25%</b>



## ΚΕΦΑΛΑΙΟ 8. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

---

### 8.1 Συμπεράσματα

### 8.2 Προτάσεις για Μελλοντική Έρευνα

---

#### **8.1. Συμπεράσματα**

Στην εργασία αυτή μελετήθηκε το πρόβλημα της εξαγωγής χαρακτηριστικών εικονοπλαισίων από ακολουθίες βίντεο. Για την εξαγωγή των χαρακτηριστικών εικονοπλαισίων εφαρμόστηκε η μέθοδος ομαδοποίησης πολλαπλών όψεων με βάρη σε 23 ακολουθίες βίντεο. Οι όψεις ή οπτικά χαρακτηριστικά που χρησιμοποιήθηκαν στην εργασία αυτή είναι τα ιστογράμματα χρώματος HSV και οι περιγραφείς CENTRIST, Wavelet, CCH, SIFT και SURF με ποικίλες παραλλαγές οι οποίες αναλύονται στα κεφάλαια τρία και έξι της εργασίας αυτής. Επιπλέον, εκτός από την προεπεξεργασία των δεδομένων για καλύτερη επίδοση και ταχύτητα εφαρμόστηκαν σε αυτά τρεις διαφορετικοί πυρήνες καθώς επίσης και συνδυασμοί αυτών. Εκτελέστηκαν δύο σειρές πειραμάτων, τα αποτελέσματα των οποίων αξιολογήθηκαν οπτικά από δύο άτομα και η μέση επίδοση της κάθε μεθόδου υπολογίστηκε με τρεις διαφορετικούς τρόπους για μεγαλύτερη αξιοπιστία.

Από την πρώτη σειρά πειραμάτων το κύριο συμπέρασμα είναι ότι η μέθοδος ομαδοποίησης πολλαπλών όψεων με βάρη (weighted multi-view) αποδίδει, για όλους τους πυρήνες που χρησιμοποιήθηκαν, καλύτερα από οποιαδήποτε μεμονωμένη όψη (single-view). Στη συνέχεια εξετάστηκε περαιτέρω η μέθοδος πολλαπλών όψεων

χρησιμοποιώντας τέσσερις διαφορετικές παραλλαγές της. Αποδοτικότερη, για όλους τους πυρήνες, αποδείχθηκε η μέθοδος πολλαπλών όψεων με χρήση κατωφλίου συμμετοχής 5% επιτυγχάνοντας λίγο καλύτερα αποτελέσματα από την βασική μέθοδο πολλαπλών όψεων με βάρη. Από αυτό συμπεράνουμε ότι ακόμα και μία πολύ μικρή συμμετοχή μίας όψης μπορεί να επηρεάσει το τελικό αποτέλεσμα και το όριο του 5% φαίνεται να έχει θετική επιρροή. Όσον αφορά τους πυρήνες που χρησιμοποιήθηκαν, καλύτερη επίδοση κατά μέσο όρο είχε ο συνδυασμός τριών πυρήνων, γραμμικού, chi-square και τομής ιστογραμμάτων, ενώ εξετάζοντάς τους ατομικά οι πυρήνες γραμμικός και chi-square έχουν παρόμοια επίδοση αφήνοντας σε χαμηλότερα επίπεδα τον πυρήνα τομής ιστογραμμάτων. Τέλος, από τα βάρη που απέδωσε η μέθοδος πολλαπλών όψεων σε κάθε όψη παρατηρήθηκε ότι όλες οι όψεις συμμετείχαν στην εξαγωγή των χαρακτηριστικών εικονοπλαισίων με μεγαλύτερη συμμετοχή της όψης HSV3D. Όσο για τον συνδυασμό των πυρήνων, μεγαλύτερη συμμετοχή είχαν ο γραμμικός και ο chi-square πυρήνας ενώ ο πυρήνας τομής ιστογραμμάτων συμμετείχε ελάχιστα.

Από την δεύτερη σειρά πειραμάτων επιβεβαιώνεται το βασικό συμπέρασμα ότι η μέθοδος ομαδοποίησης με χρήση πολλαπλών όψεων με βάρη αποδίδει καλύτερα από όλες τις μεθόδους που χρησιμοποιούν μόνο μία όψη. Χρησιμοποιήθηκε ένα βασικό σύνολο όψεων το οποίο περιείχε τις όψεις CENTRIST, Wavelet, CCH-Multi, SIFT-Multi και SURF-Multi. Σε αυτό προστέθηκαν οι όψεις HSV3D, HSV-Multi και HSV-Avg δημιουργώντας έτσι τρία σύνολα όψεων. Και τα τρία σύνολα έχουν παρόμοια επίδοση χρησιμοποιώντας τους πυρήνες chi-square, multi-kernel3 και multi-kernel2 αντίστοιχα. Εξετάζοντας συνολικά τον μέσο όρο επίδοσης του κάθε πυρήνα, καλύτερος φαίνεται να είναι ο multi-kernel3 έχοντας πολύ λίγο καλύτερα αποτελέσματα από τον multi-kernel2. Τρίτος σε μέσο όρο επίδοσης πυρήνας είναι ο chi-square ακολουθούμενος από τον γραμμικό αφήνοντας τελευταίο τον πυρήνα τομής ιστογραμμάτων. Όσον αφορά τα βάρη που αποδόθηκαν σε κάθε όψη επιβεβαιώνεται το γεγονός ότι όλες οι όψεις συμμετέχουν στην εξαγωγή των χαρακτηριστικών εικονοπλαισίων. Σχετικά με το ποσοστό συμμετοχής κάθε όψης τα ιστογράμματα χρώματος HSV έχουν τη μεγαλύτερη συμμετοχή με τους περιγραφείς Wavelet και CENTRIST να ακολουθούν. Τέλος, κατά τη διαδικασία δημιουργίας του

πυρήνα multi-kernel3 μεγαλύτερη συμμετοχή έχουν ο γραμμικός και ο chi-square πυρήνας, ενώ ο πυρήνας τομής ιστογραμμάτων συμμετέχει ελάχιστα, όπως και στην πρώτη σειρά πειραμάτων.

Από τα περαιτέρω πειράματα που έγιναν και για τις δύο σειρές πειραμάτων σχετικά με τις παραλλαγές των περιγραφέων CCH, SIFT και SURF για 20, 50 και 100 οπτικές λέξεις καθώς επίσης και για τον συνδυασμό αυτών, συμπεραίνεται ότι η μέθοδος Multi, αν και δεν έχει σε όλες τις περιπτώσεις την καλύτερη αποτελεσματικότητα, παρουσιάζει σταθερή επίδοση ανεξάρτητα από τον πυρήνα που χρησιμοποιείται. Επιπλέον, φαίνεται να έχει πιο σταθερή επίδοση από τις υπόλοιπες μεθόδους καθώς παρουσιάζει χαμηλότερη διακύμανση από αυτές.

Συνολικά, από τις δύο σειρές πειραμάτων προκύπτουν δύο βασικά συμπεράσματα. Το πρώτο είναι πως η μέθοδος ομαδοποίησης πολλαπλών όψεων με βάρη έχει συνολικά καλύτερη επίδοση από οποιαδήποτε μέθοδο που χρησιμοποιεί μόνο μία όψη. Το δεύτερο είναι ότι όλες οι όψεις συμμετέχουν στο τελικό αποτέλεσμα και το ποσοστό συμμετοχής τους μεταβάλλεται ανάλογα με τον τύπο της ακολουθίας βίντεο που εξετάζεται. Επίσης, είναι πολύ σημαντικό το γεγονός ότι παρόλο που τα δύο ground truth είναι αρκετά διαφορετικά όσον αφορά το πλήθος των χαρακτηριστικών εικονοπλαισίων για κάθε ακολουθία, η μέθοδος που προτείνουμε δεν επηρεάζεται από αυτό το γεγονός και διατηρείται σε αρκετά υψηλά επίπεδα επίδοσης.

## **8.2. Προτάσεις για Μελλοντική Έρευνα**

Ένα ζήτημα που παρουσιάζει ενδιαφέρον ζήτημα για περαιτέρω έρευνα είναι αρχικά η αξιολόγηση της παραπάνω μεθόδου σε ένα αρκετά μεγαλύτερο πλήθος ακολουθιών βίντεο, προκειμένου τα αποτελέσματα και τα συμπεράσματα να είναι πιο αξιόπιστα. Επιπλέον, μία άμεση επέκταση της μεθόδου σχετίζεται με τη χρήση περισσότερων όψεων προκειμένου να καλυφθεί ακόμα μεγαλύτερο εύρος πληροφορίας από κάθε εικονοπλαίσιο. Τέλος, θα μπορούσαν να χρησιμοποιηθούν περισσότεροι πυρήνες και συνδυασμοί αυτών προκειμένου να επιτευχθεί όσο το δυνατόν καλύτερο αποτέλεσμα.



## ΑΝΑΦΟΡΕΣ

---

- [1] Jianxin Wu and James M. Rehg. CENTRIST: A Visual Descriptor for Scene Categorization. 2009
- [2] Herbert Bay, Tinne Tuytelaars and Luc Van Gool. SURF: Speeded Up Robust Features. 2006
- [4] Michael G. Christel, Michael A. Smith, C. Roy Taylor, David B. Winkler. Evolving Video Skims into Useful Multimedia Abstractions. April 1998
- [5] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91-110, 2004.
- [6] C.-R. Huang, C.-S. Chen, and P.-C. Chung. Contrast context histogram---An efficient discriminating local descriptor for object recognition and image matching. 2007.
- [7] C. Harris and M. Stephens. A combined corner and edge detector. In *The Fourth Alvey Vision Conference*, pages 147-151, 1988.
- [8] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *ICCV*, 2003.
- [9] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE CVPR*, 2006.
- [10] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: An in-depth study. In *INRIA Technical Report RR-5737*, 2005.

- [11] 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, 1999.
- [12] I. Koprinska and S. Carrato. Temporal video segmentation: A survey. *Signal Processing: Image Communication*, 16(5):477-500, January 2001.
- [13] 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*, Vol.9, No.4. pp. 1280-1289, June 1999.
- [14] Truong, B. T. and Venkatesh, S.2007. Video abstraction: A systematic review and classification. *ACM Trans. Multimedia Computing, Communications and Applications*, Vol. 3, No. 1, Article 3, Publication date: February 2007.
- [15] 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.
- [16] 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*.
- [17] Tianming Liu, Hong-Jiang Zhang, and Feihu Qi. A Novel Video Key-Frame-Extraction Algorithm Based on Perceived Motion Energy Model, *IEEE Transactions on Circuits and Systems for Video Technology*. 2003;13(10):1006-1013.
- [18] E. Nowak et al. Sampling strategies for bag-of-features image classification. In *Proc. ECCV, 2006*.
- [19] Yu-Gang Jiang, Chong-Wah Ngo and Jun Yang. Towards Optimal Bag-of-Features for Object Categorization and Semantic Video Retrieval. *CIVR'07, Amsterdam, The Netherlands, 2007*
- [20] D. Lashkari and P. Golland, "Convex clustering with exemplar-based models," in *Advances in Neural Information Processing Systems 20*. Cambridge, MA: MIT Press, pp. 825–832, 2008

- [21] G. Tzortzis and A. Likas, "Convex mixture models for multi-view clustering," in Proc. 19th Int. Conf. Artif. Neural Netw. Part II, Limassol, Cyprus, pp. 205–214, 2009
- [22] G. Tzortzis and A. Likas, "Multiple View Clustering Using a Weighted Combination of Exemplar-Based Mixture Models". IEEE Transactions on neural networks, vol. 21, No. 12, December 2010
- [23] Ingrid Daubechies. Orthonormal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics, 41(7):909–996, October 1988.
- [24] Stephane Mallat. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7):674–693, July 1989.
- [25] Eric J. Stollnitz, Tony D. DeRose, David H. Salesin. Wavelets for Computer Graphics: A Primer Part 1. University of Washington, May 1995.
- [26] A. Likas, N. Vlassis, and J. J. Verbeek. The global K-means clustering algorithm. Pattern Recognition, 36(2):451-461, 2003.
- [27] V. Chasanis, A. Kalogeratos and A. Likas. Movie Segmentation into Scenes and Chapters Using Locally Weighted Bag of Visual Words. CIVR 2009
- [28] J. Puzicha, T. Hofmann, and J. Buhmann, "Non-Parametric Similarity Measures for Unsupervised Texture Segmentation and Image Retrieval," in Proc. CVPR, 1997.
- [29] B. Günsel and A. M. Tekalp, "Content-based video abstraction," in Proc. IEEE Int. Conf. Image Processing, Chicago, IL, pp. 128–132, 1998.
- [30] 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, pp. 64–67, 1996.
- [31] 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, 217-220, 2000.

- [32] 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, 475-479, 2000.
- [33] 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.
- [34] 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.
- [35] 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, 5307:439-447, 2004.
- [36] 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, 5307:61-69, 2004.
- [37] 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.
- [38] 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.
- [39] Z. Rasheed and M. Shah. Detection and representation of scenes in videos. IEEE Transactions on Multimedia, 7(6):1097-1105, December 2005.
- [40] A. Hanjalic and H. J. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis," IEEE Transactions on circuits and systems for video technology, Vol. 9, No.8, December 1999.
- [41] Girgensohn A., Boreczky J. Time-Constrained Keyframe Selection Technique. Multimedia Tools and Application, 11:347-358, 2000.



- [42] Gong Y. and Liu X. Generating optimal video summaries. Proc. IEEE Int. Conference on Multimedia and Expo, 3:1559-1562, 2000.
- [43] 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.
- [44] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, “Clustering with Bregman divergences,” J. Mach. Learn. Res., vol. 6, pp. 1705–1749, Dec. 2005.
- [45] C. M. Bishop, Pattern Recognition and Machine Learning. New York: Springer-Verlag, 2006.



## ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

---

Ο Αντώνιος Ιωαννίδης με καταγωγή από την Καστανιώτισσα Ευβοίας γεννήθηκε το Δεκέμβριο του 1987. Το 2004 αποφοίτησε από το 1<sup>ο</sup> Ενιαίο Λύκειο Ξάνθης με βαθμό 18.8 και εισήχθη στο Τμήμα Πληροφορικής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης από οποίο αποφοίτησε με βαθμό 7.15 το 2009. Αφού ολοκλήρωσε την στρατιωτική του θητεία, έγινε δεκτός στο Πρόγραμμα Μεταπτυχιακών Σπουδών του Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων τον Οκτώβριο του 2010 και ολοκλήρωσε τις υποχρεώσεις του τον Ιούλιο του 2012 αποκτώντας Μεταπτυχιακό Δίπλωμα με ειδίκευση στο «Λογισμικό».

