



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Δημιουργία Περιλήψεων από Ακολουθίες  
Κινούμενων Εικόνων – Ανίχνευση Πλάνων και  
Επιλογή Χαρακτηριστικών Καρέ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεώργιος Π. Κούμουλος

**Επιβλέπων :** Στέφανος Δ. Κόλλιας  
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2005





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Δημιουργία Περιλήψεων από Ακολουθίες  
Κινούμενων Εικόνων – Ανίχνευση Πλάνων και  
Επιλογή Χαρακτηριστικών Καρέ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Γεώργιος Π. Κούμουλος

**Επιβλέπων :** Στέφανος Δ. Κόλλιας  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την ....<sup>η</sup> Οκτωβρίου 2005.

.....  
Στέφανος Κόλλιας  
Καθηγητής Ε.Μ.Π.

.....  
Ανδρέας-Γεώργιος Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π.

.....  
Παναγιώτης Τσανάκας  
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2005

.....

**ΓΕΩΡΓΙΟΣ Π. ΚΟΥΜΟΥΛΟΣ**

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

**Copyright © Γεώργιος Κούμουλος, 2005**

**Με επιφύλαξη παντός δικαιώματος. All rights reserved.**

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## Περίληψη

Τα τελευταία χρόνια ο διαρκώς αυξανόμενος όγκος διαθέσιμου περιεχομένου βίντεο έχει οδηγήσει στην ανάγκη για την αυτόματη ανάλυση, σύνοψη και εξαγωγή πληροφορίας από το βίντεο. Κάθε ακολουθία κινούμενων εικόνων αποτελείται από ένα σύνολο πλάνων, που περιέχουν καρέ με χρονική συνέχεια, ενώ τα γειτονικά πλάνα συνδέονται στα όριά τους με κάποιο είδος μετάβασης. Η ανίχνευση των ορίων αυτών και η χρονική κατάτμηση της ακολουθίας αποτελούν συνήθως το πρώτο βήμα για κάθε είδους ανάλυση του βίντεο, ενώ το επόμενο και πιο σημαντικό βήμα αφορά στη σύνοψη ή δημιουργία περίληψης από το βίντεο. Η τελευταία επιτυγχάνεται συνήθως με την επιλογή ενός αριθμού χαρακτηριστικών καρέ από κάθε πλάνο, ούτως ώστε να αναπαρασταθεί το περιεχόμενό του με σύντομο και ταυτόχρονα περιεκτικό τρόπο.

Σε αυτή τη διπλωματική εργασία κατασκευάζεται ένα σύστημα περίληψης βίντεο, το οποίο υποστηρίζει ανίχνευση αλλαγής πλάνου, εξαγωγή χαρακτηριστικών ιδιοτήτων και επιλογή χαρακτηριστικών καρέ. Το σύστημα συνδυάζει μεθόδους απευθείας στο κωδικοποιημένο πεδίο MPEG και εντοπίζει αυτόματα τις αλλαγές πλάνου του βίντεο. Στη συνέχεια εξάγονται οι χαρακτηριστικές ιδιότητες από κάθε καρέ της ακολουθίας και η τιμή τους δίνει ένα σημείο στον χώρο των χαρακτηριστικών, οπότε ολόκληρη η ακολουθία του βίντεο αντιπροσωπεύεται από μία τροχιά σε αυτόν τον πολυδιάστατο χώρο. Με μαθηματικές μεθόδους ορισμού της καμπυλότητας καθορίζονται τα χαρακτηριστικά σημεία της καμπύλης, στα οποία παρατηρείται τοπικά ακραία συμπεριφορά. Το σύστημα εξάγει αυτόματα τα χαρακτηριστικά καρέ μέσω των σημείων τοπικών μεγίστων και τοπικών ελαχίστων της καμπυλότητας, ανεξάρτητα από το περιεχόμενο του βίντεο.

Στα πλαίσια της εργασίας συγκρίνονται δύο διαφορετικοί τρόποι υπολογισμού της καμπυλότητας και δύο άλλες ήδη υλοποιημένες μέθοδοι, που εξάγουν χαρακτηριστικά σημεία καμπύλης. Τα πειράματα αφορούν στην αξιολόγηση της αποτελεσματικότητας των μεθόδων για την εξαγωγή χαρακτηριστικών καρέ του βίντεο. Η επίδοση των μεθόδων επιδιώκεται να βελτιωθεί με τη βοήθεια του υπολογιστικού μοντέλου της οπτικής προσοχής, με το οποίο εντοπίζεται μία σημαντική περιοχή της εικόνας για την εξαγωγή των χαρακτηριστικών ιδιοτήτων από αυτή. Τέλος, μέσα από τα αποτελέσματα των πειραμάτων αναλύονται οι δυνατότητες, οι ελλείψεις και τα περιθώρια βελτίωσης και επέκτασης του συστήματος περίληψης βίντεο.

### Λέξεις – Κλειδιά:

περίληψη βίντεο, κωδικοποίηση MPEG, ανίχνευση πλάνων, χαρακτηριστικό καρέ, χώρος χαρακτηριστικών, διάνυμα χαρακτηριστικών, πολυδιάστατη καμπύλη, καμπυλότητα, μοντέλο οπτικής προσοχής



## **Abstract**

Nowadays, the enlarging available video data has led to the emerging need for automatic analysis, synopsis and extraction of information from videos. Every video sequence consists of a number of shots, each of them containing temporally associative frames, while contiguous shots are connected to each other with some type of transition at their boundaries. The first step for any kind of video analysis seems to be the detection of these boundaries followed by a temporal segmentation of video, while the next and more important step is related with the synopsis or the summarization of video. This is achieved by selecting certain number of characteristic frames (keyframes) from each shot, so that the content of the shot is represented in a short and also meaningful way.

In this diploma thesis, a video summarization system is being constructed, in order to accomplish shot change detection, feature extraction and keyframe selection. This system combines methods working directly on MPEG compressed domain and automatically locates shot changes of video. Features from each frame of the sequence are then extracted and their values produce a point in the feature space. Therefore the entire video sequence is represented by a trajectory in this multidimensional space. According to mathematical methods defining the curvature of a curve, the characteristic points of the curve are determined, in areas where locally extreme behaviour is observed. The system automatically extracts the keyframes via points of local maxima and local minima of curvature, irrespective of the video content.

In this diploma work, two different ways of calculation of curvature and two other already proposed methods that export characteristic points of a curve are compared. Experiments concern in the effectiveness of these methods on the part of video keyframes selection. We attempt to improve performance of these methods by applying the computational model of Visual Attention, in order to extract features from the salient region of the image that this model locates. Finally, through the results of these experiments we aim to indicate capabilities, disadvantages and cases of improving and expanding this video summarization system.

### **Key – Words:**

video summarization, MPEG coding, shot detection, keyframe, feature space, feature vector, multidimensional curve, curvature, visual attention model





## Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2004-2005 στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων (IVML) του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή κ. Στέφανο Κόλλια για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου την εργασία αυτή και για τη δυνατότητα που μου έδωσε να ασχοληθώ με το συγκεκριμένο ενδιαφέρον θέμα. Θα ήθελα επίσης να ευχαριστήσω τον Ερευνητή ΕΠΙΣΕΥ – ΕΜΠ Δρ Ιωάννη Αβρίθη και τον Υπογύφιο Διδάκτορα Κωνσταντίνο Ραπαντζίκο, διότι συνέβαλαν στη συγγραφή αυτής της εργασίας με την καθοδήγηση, τις πολύτιμες συμβουλές και υποδείξεις τους. Να ευχαριστήσω τέλος την οικογένεια και τους φίλους μου, που με στήριξαν και με βοήθησαν, ο καθένας με τον τρόπο του, κατά τη διάρκεια των φοιτητικών μου σπουδών.



# Περιεχόμενα

<b>ΚΕΦΑΛΑΙΟ 1</b>	<b>Εισαγωγή .....</b>	<b>17</b>
1.1	Αντικείμενο της Μελέτης .....	17
1.2	Σκοπός της Εργασίας .....	18
1.3	Υπάρχουσες Μέθοδοι Εξαγωγής Χαρακτηριστικών Καρέ .....	20
1.4	Οργάνωση της Εργασίας.....	24
<b>ΚΕΦΑΛΑΙΟ 2</b>	<b>Χρονική Κατάτμηση του Βίντεο σε Πλάνα .....</b>	<b>27</b>
2.1	Εισαγωγικά .....	27
2.2	Σύντομη Περιγραφή της Κωδικοποίησης MPEG.....	29
2.2.1	Το πρότυπο MPEG .....	29
2.2.2	Τα τρία είδη των frames στο MPEG.....	30
2.2.2.1	I-frames (Intra frames).....	30
2.2.2.2	P-frames (Predicted frames) .....	30
2.2.2.3	B-frames (Bidirectional Interpolated frames).....	31
2.2.3	Περιγραφή των MacroBlocks .....	32
2.3	Υπάρχουσες Τεχνικές για την Ανίχνευση Πλάνων .....	35
2.3.1	Μέθοδοι Εφαρμοσμένες σε Ασυμπίεστο Βίντεο .....	35
2.3.1.1	Διαφορές των Pixels .....	35
2.3.1.2	Διαφορές Ιστογράμματος.....	37
2.3.1.3	Στατιστικές Μετρικές .....	38
2.3.1.4	Μετρικές Βασισμένες σε Ακμές .....	39
2.3.2	Μέθοδοι Εφαρμοσμένες σε Συμπίεσμένο Βίντεο.....	40
2.3.2.1	Ανίχνευση σε I-frame .....	40
2.3.2.2	Ανίχνευση σε P-frame .....	41
2.3.2.3	Ανίχνευση σε B-frame .....	42
2.4	Συνδυασμός Μεθόδων και Κατασκευή Κανόνων .....	43

2.4.1	Κανόνες για τα P-frames .....	44
2.4.2	Κανόνες για τα I-frames .....	46
2.4.3	Κανόνες για τα B-frames .....	49
2.4.4	Σύστημα Κανόνων και Διάγραμμα Απόφασης.....	51

### **Κ Ε Φ Α Λ Α Ι Ο 3      Εξαγωγή και Ομαλοποίηση του Διανύσματος Χαρακτηριστικών ..... 57**

3.1	Εισαγωγικά .....	57
3.2	Εξαγωγή του Διανύσματος Χαρακτηριστικών .....	58
3.3	Ομαλοποίηση του Διανύσματος Χαρακτηριστικών .....	62
3.3.1	Φίλτρα Ομαλοποίησης μιας Καμπύλης .....	62
3.3.1.1	Ορθογώνια Συνάρτηση (Μέση Τιμή) .....	63
3.3.1.2	Τριγωνική Συνάρτηση .....	63
3.3.1.3	Φίλτρο Ενδιάμεσης Τιμής.....	64
3.3.1.4	Γκαουσιανή Συνάρτηση.....	64
3.3.2	Σύγκριση Ομαλοποίησης Μέσης Τιμής και Γκαουσιανής .....	66
3.3.3	Παράδειγμα Ομαλοποίησης Διανύσματος Χαρακτηριστικών .....	69

### **Κ Ε Φ Α Λ Α Ι Ο 4      Μέθοδοι Καμπυλότητας για την Εξαγωγή Χαρακτηριστικών Σημείων Καμπύλης ..... 73**

4.1	Εισαγωγικά .....	73
4.1.1	Εξαγωγή Χαρακτηριστικών Σημείων - Καρέ.....	73
4.1.2	Έννοιες στον N-διάστατο Ευκλείδειο Χώρο .....	74
4.2	Η Καμπυλότητα μιας Καμπύλης.....	75
4.3	Περιγραφή Καμπύλης με Φυσική Παράμετρο .....	79
4.4	Η Φυσική Παράμετρος s και η Τυχαία Παράμετρος t.....	82
4.5	Γενικευμένες Καμπυλότητες στον χώρο $R^n$ .....	85
4.6	Σύγκριση των Μεθόδων Καμπυλότητας.....	87
4.6.1	Διδιάστατο Μοντέλο.....	88
4.6.2	Τρισδιάστατο Μοντέλο.....	93
4.6.3	Διάνυσμα Χαρακτηριστικών 3D.....	99

<b>Κ Ε Φ Α Λ Α Ι Ο 5</b>	<b>Άλλες Μέθοδοι για την Εξαγωγή Χαρακτηριστικών Σημείων Καμπύλης.....</b>	<b>109</b>
5.1	Η Μέθοδος Απλοποίησης Καμπύλης .....	111
5.1.1	Εισαγωγικά .....	111
5.1.2	Επαναληπτικός Αλγόριθμος Διαμέρισης N-D Καμπύλης.....	112
5.1.3	Περιγραφή του Αλγορίθμου .....	115
5.1.4	Γενίκευση του Αλγορίθμου του Ramer .....	118
5.2	Η Μέθοδος Προσέγγισης Περιγράμματος.....	119
5.2.1	Εισαγωγικά .....	119
5.2.2	Περιγραφή των Αλγορίθμων .....	120
5.2.2.1	Αλγόριθμος Global-searching.....	120
5.2.2.2	Αλγόριθμος Pair-growing.....	122
5.2.3	Προσαρμογή για την Περίπτωση N-D Καμπύλης.....	123
 <b>Κ Ε Φ Α Λ Α Ι Ο 6</b>	 <b>Υπολογιστικό Μοντέλο Οπτικής Προσοχής.....</b>	 <b>127</b>
6.1	Εισαγωγικά .....	128
6.2	Περιγραφή του Μοντέλου Οπτικής Προσοχής.....	129
6.3	Εξαγωγή των Πρωταρχικών Οπτικών Χαρακτηριστικών .....	130
6.4	Κατασκευή του Χάρτη Ενδιαφέροντος .....	134
6.5	Εφαρμογή της Οπτικής Προσοχής.....	137
6.5.1	Αποτελέσματα στις Εικόνες και Παρατηρήσεις.....	140
6.5.2	Εφαρμογή στην Εξαγωγή Χαρακτηριστικών από το Καρέ.....	142
 <b>Κ Ε Φ Α Λ Α Ι Ο 7</b>	 <b>Πειράματα, Αποτελέσματα και Παρατηρήσεις.....</b>	 <b>145</b>
7.1	Η Υλοποίηση της Εφαρμογής.....	146
7.2	Πειράματα στην Ανίχνευση Πλάνων.....	148
7.2.1	Χαρακτηριστικά του Δείγματος Βίντεο.....	148
7.2.2	Ενδεικτικά Διαγράμματα .....	150
7.2.3	Αποτελέσματα και Παρατηρήσεις.....	154
7.3	Πειράματα στην Εξαγωγή των Χαρακτηριστικών Καρέ.....	165
7.3.1	Χαρακτηριστικά του Δείγματος Βίντεο.....	165

7.3.2	Διαγράμματα και Παρατηρήσεις .....	166
7.3.2.1	Αποτελέσματα για την ακολουθία FORMULA1 .....	166
7.3.2.2	Αποτελέσματα για την ακολουθία RALLY2.....	176
7.3.2.3	Αποτελέσματα για την ακολουθία STEADY .....	186
7.3.3	Συγκεντρωτικά Αποτελέσματα .....	191
7.4	Πειράματα στην Εφαρμογή της Οπτικής Προσοχής .....	202
 <b>Κ Ε Φ Α Λ Α Ι Ο 8 Συμπεράσματα και Επίλογος .....</b>		<b>209</b>
8.1	Συμπεράσματα .....	209
8.2	Μελλοντικές Επεκτάσεις .....	213
 <b>Π α ρ ά ρ τ η μ α .....</b>		<b>217</b>
Π.1	Απόδειξη του τύπου της δεύτερης παραγώγου ως προς $s$ .....	217
Π.2	Απόδειξη του τύπου 3 για την καμπυλότητα .....	218
Π.3	Γενικευμένες Καμπυλότητες ή Συναρτήσεις Καμπυλότητας .....	220
 <b>Β ι β λ ι ο γ ρ α φ ί α .....</b>		<b>225</b>







# ΚΕΦΑΛΑΙΟ 1

## *Εισαγωγή*

### 1.1 Αντικείμενο της Μελέτης

Στη σύγχρονη κοινωνία της πληροφορίας και της γνώσης, το διαδίκτυο αποτελεί ένα από τα κυριότερα μέσα επικοινωνίας και ίσως την πιο περιεκτική πηγή δεδομένων, τα οποία προέρχονται από διάφορα σημεία του πλανήτη, ανανεώνονται συνεχώς, αφορούν ποικίλες θεματικές κατηγορίες και προορίζονται για πολυδιάστατη και διαφορετική χρήση. Στον όγκο αυτό των δεδομένων το βίντεο κατέχει σημαντική θέση, ως μέσο καταγραφής οποιασδήποτε στιγμιαίας λεπτομέρειας του οπτικού περιβάλλοντος, την οποία αντιλαμβάνεται το ανθρώπινο μάτι. Με λίγα λόγια περιέχει οπτικά δεδομένα, τα οποία ο άνθρωπος δεν θα μπορούσε να αποκτήσει παρά μόνο αν βρισκόταν στο σημείο όπου η κάμερα κατέγραφε τα γεγονότα που συνέβαιναν. Έτσι, η πληροφορία που περιέχεται στο ψηφιακό βίντεο στο διαδίκτυο παρέχει στον χρήστη τη δυνατότητα της μόρφωσης, της επικοινωνίας, της ψυχαγωγίας, αλλά και της ενημέρωσης σε θέματα που εξελίσσονται, και για τον λόγο αυτό είναι απαραίτητη σε κάθε πτυχή της καθημερινής ζωής του σύγχρονου ανθρώπου.

Είναι πολύ σημαντικό το γεγονός ότι ο χρήστης μπορεί στο σημείο που βρίσκεται να συλλέξει οπτική πληροφορία από διαφορετικές πηγές, πληροφορία που προέρχεται από γεγονότα, σχετικά ή όχι, τα οποία μπορεί να συμβαίνουν ή να συνέβησαν ακόμη και ταυτόχρονα, και να τα χρησιμοποιήσει για μελέτη, γνώσεις, συγκρίσεις και συμπεράσματα για τα οποία σε άλλη περίπτωση απαιτείται κόπος και οπωσδήποτε χρόνος για να επιτευχθούν. Το πλεονέκτημα αυτής της ταχύτατης απόκτησης οπτικής εμπειρίας δυστυχώς περιορίζεται, όχι μόνο από το περιορισμένο εύρος ζώνης που

παρέχει το διαδίκτυο (και το οποίο ισχύει και για τις υπόλοιπες μορφές δεδομένων), αλλά κυρίως λόγω του μεγάλου όγκου των οπτικών δεδομένων. Το βίντεο απαιτεί μεγάλη χωρητικότητα κατά την αποθήκευσή του, ακόμα και σε ψηφιακή συμπίεσμένη μορφή (ειδικά αν απαιτείται ποιοτική εικόνα), οπότε δημιουργούνται ογκώδεις βάσεις δεδομένων και ακόμη μεγαλύτερες απαιτήσεις στο ρυθμό μεταφοράς των δεδομένων. Η πολλαπλή χρησιμότητα ενός αρχείου οπτικού περιεχομένου (π.χ. βίντεο με απόσπασμα ειδήσεων ενδιαφέρει δημοσιογράφους, οικονομολόγους, πολιτικούς, κοινωνιολόγους, ιστορικούς αναλυτές,) οδηγεί σε αυξημένη ζήτηση και ταυτόχρονη πρόσβαση από πολλά σημεία του δικτύου, οπότε και η καθυστέρηση στη μεταφορά είναι αναπόφευκτη, υποβιβάζοντας το πλεονέκτημα της ταχύτητας που χαρακτηρίζει τον σύγχρονο τρόπο πληροφόρησης.

Από την άλλη πλευρά, ένα απόσπασμα βίντεο αποτελείται από μία ακολουθία καρτέ (εικόνων), τα οποία απέχουν συγκεκριμένο χρονικό διάστημα (συνήθως αντιστοιχούν 25 καρτέ στο δευτερόλεπτο), ενώ πολλά από αυτά ενδέχεται να έχουν παρόμοιο περιεχόμενο, και άρα να μεταφέρουν πλεονάζουσα πληροφορία. Η αναπαράσταση του περιεχομένου του βίντεο με μερικά χαρακτηριστικά καρτέ (και όχι με ολόκληρη την ακολουθία των εικόνων) έχει αποδειχθεί ως η πιο αποδοτική λύση στο πρόβλημα του μεγάλου όγκου των δεδομένων του βίντεο. Πριν από περίπου δέκα χρόνια, ενώ είχε ήδη ξεκινήσει η μελέτη σε αρχικό ερευνητικό στάδιο, δόθηκε πλέον μεγαλύτερη βαρύτητα στο συγκεκριμένο ζήτημα και υπήρξαν νέες προοπτικές για την σύντομη και περιεκτική αναπαράσταση των ακολουθιών κινούμενων εικόνων. Συνεχώς μέχρι και σήμερα εξελίσσεται το θέμα που αφορά στη δημιουργία περιλήψεων από βίντεο (video summarization), στην τοποθέτηση δεικτών με βάση το περιεχόμενο (content-based indexing), στην περιήγηση του περιεχομένου με αποδοτικό τρόπο (video browsing) και γενικά στην εύκολη ανάκτηση αρχείων βίντεο από το διαδίκτυο (video retrieval).

## 1.2 Σκοπός της Εργασίας

Στο σύγγραμμα αυτό μελετάται αναλυτικά το ζήτημα της δημιουργίας περιλήψεων από ακολουθίες κινούμενων εικόνων και ειδικότερα εξετάζονται αποδοτικές μέθοδοι που έχουν ως στόχο την αναπαράσταση του περιεχομένου της ακολουθίας με τα

κατάλληλα επιλεγμένα καρέ, δίνοντας μία όσο το δυνατόν πιο συνοπτική και περιεκτική παρουσίαση του βίντεο. Ο σκοπός της διπλωματικής εργασίας είναι η αναζήτηση και σύγκριση τέτοιων μεθόδων, η αναφορά σε εναλλακτικές τεχνικές που ενδείκνυνται για την ίδια εφαρμογή, η κατασκευή ενός συστήματος αυτόματης εξαγωγής χαρακτηριστικών καρέ που να δίνει γρήγορα και αξιόπιστα αποτελέσματα, αυτόνομα και ανεξάρτητα από το περιεχόμενο της εισόδου του (να μην χρειάζεται επίβλεψη ή εκπαίδευση σε συγκεκριμένο δείγμα), και τέλος η μελέτη των δυνατοτήτων του συστήματος αυτού και οι προτάσεις για βελτίωση και επέκταση.

Κάθε ακολουθία κινούμενων εικόνων (βίντεο) αποτελείται από ένα πλήθος πλάνων, που περιέχουν εικόνες με χρονική συνέχεια και συνδέονται στα όριά τους με κάποιο είδος μετάβασης. Η ανίχνευση των ορίων αυτών και η κατάτμηση της ακολουθίας είναι το πρώτο βήμα για τη σύνοψη του βίντεο, ενώ το επόμενο και πιο σημαντικό βήμα αφορά στην επιλογή μερικών χαρακτηριστικών καρέ από κάθε πλάνο ούτως ώστε να αναπαρασταθεί το περιεχόμενό του με τον πιο σύντομο και ταυτόχρονα περιεκτικό τρόπο.

Στην εργασία αυτή κατασκευάζεται ένα σύστημα που συνδυάζει μεθόδους απευθείας στο κωδικοποιημένο MPEG πεδίο και εντοπίζει αυτόματα τις αλλαγές πλάνου του βίντεο. Στη συνέχεια οι χαρακτηριστικές ιδιότητες εξάγονται από κάθε καρέ της ακολουθίας και η τιμή τους δίνει ένα σημείο στον χώρο των χαρακτηριστικών, οπότε ολόκληρη η ακολουθία βίντεο αντιπροσωπεύεται από μία τροχιά στον πολυδιάστατο χώρο. Με μαθηματικές μεθόδους ορισμού της καμπυλότητας καθορίζονται τα χαρακτηριστικά σημεία πάνω στην καμπύλη, όπου παρατηρείται τοπικά ακραία συμπεριφορά. Οι τιμές τοπικών μεγίστων και τοπικών ελαχίστων της καμπυλότητας αντιπροσωπεύουν απότομη και μικρή απόκλιση της τροχιάς αντίστοιχα, ενώ τα καρέ που αντιστοιχούν σε αυτά τα σημεία λαμβάνονται ως χαρακτηριστικά καρέ.

Δύο διαφορετικοί τρόποι υπολογισμού της καμπυλότητας και δύο άλλες ήδη υλοποιημένες μέθοδοι, που χειρίζονται σημεία καμπύλης, συγκρίνονται μέσω πειραμάτων για την αποτελεσματικότητά τους στην εξαγωγή χαρακτηριστικών καρέ του βίντεο. Η επίδοση των μεθόδων επιχειρείται να βελτιωθεί με τη βοήθεια του υπολογιστικού μοντέλου της οπτικής προσοχής, με το οποίο εντοπίζεται μία σημαντική περιοχή της εικόνας για την εξαγωγή των χαρακτηριστικών από αυτή. Τέλος, μέσα από τα αποτελέσματα των πειραμάτων αναλύονται οι δυνατότητες, οι ελλείψεις και τα περιθώρια επέκτασης του συστήματος περίληψης βίντεο.

### 1.3 Υπάρχουσες Μέθοδοι Εξαγωγής Χαρακτηριστικών Καρέ

Για τη δημιουργία μιας σύνοψης του βίντεο, που να είναι περιεκτική σε νόημα αλλά και όσο το δυνατόν συντομότερη, η πιο σωστή αντιμετώπιση είναι ο προσδιορισμός μερικών καρέ (χαρακτηριστικά καρέ), τα οποία θα υποκαθιστούν το νοηματικό περιεχόμενο ολόκληρης της ακολουθίας. Αυτό συνεπάγεται ότι θα πρέπει να χρησιμοποιούνται περισσότερα καρέ σε σημεία όπου υπάρχει μεγαλύτερη πληροφορία (απότομες αλλαγές, κίνηση αντικειμένων, γενικά νοηματικές αλλαγές) και λιγότερα στα υπόλοιπα σημεία, ώστε ούτε να χάνεται αλλά ούτε να επαναλαμβάνεται πληροφορία. Τα καρέ αυτά, που θα αναπαριστούν το περιεχόμενο του βίντεο και θα χρησιμοποιούνται στην περιήγηση (browsing) και την διευκόλυνση της ανάκτησης (retrieval) του βίντεο, θα ονομάζονται χαρακτηριστικά καρέ (characteristic frames) ή αλλιώς “καρέ-κλειδιά” (“keyframes”). Τα υπόλοιπα ενδιάμεσα καρέ της ακολουθίας που βρίσκονται μεταξύ των χαρακτηριστικών καρέ θεωρείται ότι περιέχουν πλεονάζουσα πληροφορία χωρίς κάποιο αξιοσημείωτο συμβάν και γι’ αυτό αντιπροσωπεύονται πλήρως από τα χαρακτηριστικά καρέ στα άκρα τους.

Μία σημαντική πρώτη εργασία για την εξαγωγή αντιπροσωπευτικών καρέ από το βίντεο έγινε το 1993 από τους Zhang et al. ([Zha95a]), οι οποίοι προσπάθησαν να ορίσουν διαφορετικό αριθμό επιλεγόμενων καρέ για κάθε πλάνο ανάλογα με το περιεχόμενό του, και όχι όπως οι περισσότερες μέχρι τότε εργασίες που όριζαν ίδιο αριθμό καρέ ή απλώς ανάλογο της διάρκειας κάθε πλάνου. Σύμφωνα με αυτούς, το πρώτο καρέ κάθε πλάνου πάντοτε επιλέγεται ως το πρώτο χαρακτηριστικό καρέ (keyframe) και στη συνέχεια υπολογίζεται ένα μέτρο διαφοράς (σε συμπιεσμένο ή ασυμπιεστο βίντεο, το μέτρο θα μπορούσε για παράδειγμα να είναι το χρωματικό ιστόγραμμα - color histogram) μεταξύ του keyframe και κάθε καρέ της ακολουθίας. Όταν η διαφορά είναι πολύ μεγάλη, τότε το τρέχον καρέ επιλέγεται και αυτό ως keyframe και αποτελεί τη βάση για τη σύγκριση με τα επόμενα πλαίσια στην ακολουθία, για την εξαγωγή νέου keyframe, ενώ χρησιμοποιούνται δύο τιμές κατωφλίων για τον καθορισμό της πυκνότητας των επιλεγόμενων καρέ. Παρόλο που η μέθοδος επιλέγει δυναμικά τα καρέ-κλειδιά, το κύριο μειονέκτημα είναι η χρήση των thresholds τα οποία θα πρέπει σε κάθε βίντεο να επιλέγονται κατάλληλα και δεν υπάρχει ενιαία τιμή που να λειτουργεί καλά σε όλα τα είδη περιεχομένου. Ακόμα, το πρώτο καρέ της ακολουθίας δεν είναι και τόσο κατάλληλο σαν keyframe, αφού μπορεί να μην περιέχει σημαντική

πληροφορία για το περιεχόμενο του πλάνου, δεδομένου ότι ανήκει στο όριο του πλάνου και μπορεί ακόμα να ανήκει και σε μετάβαση πλάνου (shot transition).

Οι Zhuang et al. ([Zhu98]) προτείνουν τη χρήση ενός είδους ομαδοποίησης (clustering), τεχνική που δανείζονται από τους χώρους της Αναγνώρισης Προτύπων (Pattern Recognition) και της Ανάλυσης Λόγου (Speech Analysis). Το σχήμα της ομαδοποίησης χωρίς επίβλεψη που περιγράφουν εφαρμόζεται σε όλα τα καρέ ενός πλάνου, τα οποία τελικά ομαδοποιούνται σε συγκεκριμένο αριθμό από ομάδες (clusters). Το μέτρο για την ομαδοποίηση μπορεί να είναι οποιαδήποτε μετρική για τη σύγκριση της ομοιότητας μεταξύ εικόνων όπως το χρωματικό ιστόγραμμα, ενώ μία παράμετρος καθορίζει την πυκνότητα κάθε ομάδας, αρά και τον συνολικό αριθμό των ομάδων που θα δημιουργηθούν για ένα πλάνο. Εκείνες οι ομάδες που περιέχουν μεγάλο αριθμό πλαισίων θεωρούνται ως πλέον σημαντικές σε πληροφορία, το περιεχόμενό τους θεωρείται απαραίτητο να αναπαρασταθεί και ονομάζονται ομάδες-κλειδιά (“key clusters”), οπότε ως keyframe επιλέγεται ένα καρέ που είναι πιο κοντά στο κέντρο βάρους (centroid) κάθε ομάδας-κλειδιού.

Οι προηγούμενες προσεγγίσεις που βασίζονται σε ανάλυση του χρωματικού περιεχομένου του βίντεο έχουν υιοθετηθεί συχνά, λόγω των ιδιοτήτων του ιστογράμματος να μένει ανεπηρέαστο από τον προσανατολισμό της εικόνας, τη μικρή μετακίνηση αντικειμένων και την επίδραση του θορύβου. Όμως σε βίντεο όπου συνεχώς υπάρχει σχετικά μεγάλη κίνηση κάποιων αντικειμένων ή της κάμερας, η τεχνική δεν δουλεύει διότι δεν μπορεί να διακρίνει περιοχές που ξεχωρίζουν στην μεταβολή του χρωματικού ιστογράμματος. Γι’ αυτόν τον λόγο είναι σημαντική η εφαρμογή τεχνικών που λαμβάνουν υπ’ όψη τους την κίνηση μέσα στο βίντεο, διότι εκμεταλλεύονται καλύτερα τις χρονικές δυναμικές μεταβολές του περιεχομένου. Για παράδειγμα οι Divakaran et al. ([Div01]) προτείνουν μία μέθοδο περίληψης για το βίντεο που βασίζεται σε περιγραφείς κίνησης που εξάγονται απευθείας από το συμπιεσμένο πεδίο του κωδικοποιημένου MPEG βίντεο. Υπολογίζοντας την τυπική απόκλιση του μέτρου όλων των διανυσμάτων κίνησης (motion vectors) σε κάθε P-frame εξάγουν χαρακτηριστικά δραστηριότητας κίνησης (motion activity) και μετά σε κάθε I-frame βρίσκουν έναν περιγραφέα δραστηριότητας κίνησης (motion activity descriptor) από το μέσο όρο όλων των δραστηριοτήτων κίνησης σε όλα τα P-frames που ανήκουν στο προηγούμενο GOP. Ο μέσος όρος όλων των τιμών των περιγραφέων δραστηριότητας κίνησης στα I-frames του πλάνου δίνει τη δραστηριότητα κίνησης ολόκληρου του πλάνου, η ένταση της οποίας καθορίζει το πλήθος των keyframes, δηλαδή το βαθμό της περίληψης του πλάνου. Το

μέτρο της δραστηριότητας κίνησης σε κάθε καρέ (από τα I- ή P-frames) χαρακτηρίζει την καταλληλότητά του ως χαρακτηριστικό καρέ.

Είναι λογικό ότι μέθοδοι που απαιτούν για κάθε βίντεο τον καθορισμό μια τιμής για κάποια χρησιμοποιούμενα κατώφλια δεν έχουν πολύ μεγάλη πρακτική αξία, αν υποθέσουμε ότι θέλουμε ένα αυτόματο χωρίς επίβλεψη ευφυές σύστημα που θα λειτουργεί ικανοποιητικά για κάθε είσοδο που του δίνουμε με αποτέλεσμα ένα πλήθος από keyframes ανάλογα με το περιεχόμενο. Μια τεχνική που προτάθηκε από τους Doulamis et al. ([Dou00]) βασίζεται στη μαθηματική μεθοδολογία που εφαρμόζεται πάνω σε χαμηλού επιπέδου (low-level) χαρακτηριστικά, όπως για παράδειγμα το χρωματικό ιστόγραμμα της ακολουθίας των εικόνων ή μία κατάτμηση (segmentation) στο πεδίο χρώματος και κίνησης. Από τις τιμές των χαρακτηριστικών προκύπτει το διάνυσμα των χαρακτηριστικών που ορίζεται σε έναν πολυδιάστατο χώρο, τον χώρο των χαρακτηριστικών, οπότε από όλα τα καρέ προκύπτει μία καμπύλη στον χώρο αυτόν. Ως keyframes λαμβάνονται τα κατάλληλα σημεία της καμπύλης που χαρακτηρίζουν τη μορφή της τροχιάς αυτής στο χώρο και τα οποία αντιστοιχούν σε μικρή ή μεγάλη καμπυλότητα, η οποία υπολογίζεται ως μέτρο της δεύτερης παραγώγου του διανύσματος ως προς τον χρόνο.

Αντίστοιχα με την προηγούμενη εργασία που χειρίζεται το βίντεο στον χώρο των χαρακτηριστικών, οι DeMenthlon et al. ([DeM98]) προτείνουν την ανάλυση της τροχιάς με εργαλεία που έχουν αναπτυχθεί για επίπεδες καμπύλες και τη μέθοδο που ονομάζεται απλοποίηση καμπύλης. Μέσω ενός αλγορίθμου διαμέρισης της N-διάστατης καμπύλης, η τροχιά του βίντεο απλοποιείται σε τμήματα μικρότερης διάστασης, των οποίων τα άκρα λαμβάνονται ως σημεία που θα συνοψίσουν το περιεχόμενο. Μία παράμετρος σύγκλισης  $\epsilon$  καθορίζει τα διαφορετικά επίπεδα λεπτομέρειας, δηλαδή το πόσο αναλυτική θα είναι η σύνοψη του βίντεο. Η μείωση των διαστάσεων της καμπύλης επιτρέπει την καλύτερη αναπαράστασή της ως προβολή στις δύο ή τρεις διαστάσεις. Μία ανάλογη μέθοδος προτείνεται από τους Wang et al. ([Wan98a], [Wan98b]) με σκοπό την προσέγγιση του περιγράμματος ενός διδιάστατου σχήματος. Τοποθετούν διαδοχικά σε ένα σύνολο τα σημεία τα πιο απομακρυσμένα από το τμήμα που ενώνει τα ήδη επιλεγμένα σημεία, μέχρι να επιτευχθεί ένας προκαθορισμένος αριθμός σημείων. Με τον τρόπο αυτό η μέθοδος προσομοιάζει μια απλοποίηση της καμπύλης στη 1 διάσταση και εξάγει χαρακτηριστικά σημεία, μεταξύ των οποίων η καμπύλη μπορεί να θεωρηθεί ευθύγραμμη. Η προσαρμογή για πολυδιάστατη καμπύλη είναι απλή και περιορίζεται στον ορισμό της απόστασης σημείου από ευθύγραμμο τμήμα που ενώνει δύο άλλα σημεία.

Επίσης σημαντική είναι η εργασία των Zhang et al. ([Zha95b]), όπου μετά τη διαδικασία της κοπής του βίντεο σε πλάνα και της εξαγωγής των χαρακτηριστικών καρέ για κάθε πλάνο, επιχειρείται μία σύγκριση μεταξύ του περιεχομένου των πλάνων και μία ταξινόμησή τους σε κλάσεις με εξέχουσα σημασία. Μπορούν να οριστούν και να αναζητηθούν σημαντικά γεγονότα (στόρυ) μέσω της ανάλυσης χαρακτηριστικών χαμηλού ή υψηλού επιπέδου (low-level ή high-level) από τα keyframes (όπως χρώμα-color, υφή-texture, σχήμα-shape, ακμές-edges) ή μέσω χρονικών χαρακτηριστικών του πλάνου. Οι Yeung et al. ([Yeu97]) παρουσιάζουν μία τεχνική για την διαφοροποίηση των χαρακτηριστικών καρέ μιας ομάδας πλάνων (στόρυ) με βάση την κυριαρχία σε σημαντικότητα του καθενός. Έτσι κατασκευάζουν μια συλλογή εικόνων του βίντεο (ένα video poster) το οποίο περιέχει σύντομα, περιεκτικά και ευπαρουσίαστα το περιεχόμενο του στόρυ και η συλλογή για όλα τα στόρυ του βίντεο δίνει τη συνολική περίληψη με χρονική και νοηματική σειρά.

Τέλος, οι Hanjalic et al. ([Han99]) υλοποιούν την τεχνική της ομαδοποίησης σε όλα τα πλαίσια του βίντεο και μετά εφαρμόζουν μια ανάλυση των αξιόπιστων ομάδων (cluster-validity analysis) για τον διαχωρισμό κάποιων βέλτιστων ομάδων (optimal clusters). Τα χαρακτηριστικά καρέ ορίζονται ως τα κέντρα βάρους των βέλτιστων ομάδων, οι οποίες αποτελούν τα στόρυ του βίντεο. Μεταβάλλοντας τις παραμέτρους της μεθόδου, εξαγουν τα χαρακτηριστικά καρέ σε διαφορετικούς βαθμούς λεπτομέρειας και κατασκευάζουν μια ιεραρχική δομή με επίπεδα λεπτομέρειας. Η πολυεπίπεδη σύνοψη που προκύπτει είναι κατάλληλη για τη γρήγορη περιήγηση του περιεχομένου (browsing) και κυρίως για τη διεξαγωγή ερωτήσεων (queries) πάνω στο βίντεο, με το υψηλότερο επίπεδο να περιέχει τα στόρυ, το επόμενο τα πλάνα που ανήκουν στα στόρυ και αμέσως μετά την ιεραρχική δομή των χαρακτηριστικών καρέ για κάθε πλάνο, με το τελευταίο επίπεδο να περιέχει τα πιο λεπτομερή χαρακτηριστικά καρέ.

Στην εργασία αυτή υιοθετείται η αναπαράσταση του βίντεο στον πολυδιάστατο χώρο των χαρακτηριστικών, διότι με αυτόν τον τρόπο εκφράζεται το περιεχόμενο του βίντεο με βάση τη σημασία του και τις νοηματικές μεταβολές του. Οι μέθοδοι που διαμερίζουν την καμπύλη με επαναληπτικό τρόπο απαιτούν τον καθορισμό κάποιου κριτηρίου σύγκλισης ή του αριθμού των εξαγόμενων σημείων ξεχωριστά για κάθε βίντεο και γι' αυτό εμφανίζουν δυσκολίες σε μια αυτόματη εφαρμογή. Χρησιμοποιείται η έννοια της καμπυλότητας για την εξαγωγή των ακραίων σημείων, που επιτρέπει στο σύστημα να λειτουργεί αυτόνομα ανεξάρτητα από το βίντεο που δέχεται στην είσοδό του. Η καμπυλότητα ως δεύτερη παράγωγος ως προς τον χρόνο ([Dou00]) εκφράζει την

επιτάχυνση (ή επιβράδυνση) του περιεχομένου στην ακολουθία, το πόσο γρήγορα μεταβάλλεται από καρέ σε καρέ. Επιχειρείται εδώ να οριστεί η καμπυλότητα με έναν πιο ορθό από μαθηματικής άποψης τρόπο, ο οποίος αφορά στην θεώρηση της μεταβολής με βάση την μετατόπιση  $s$  πάνω στην καμπύλη. Αυτό έχει ως αποτέλεσμα την εξέταση της μεταβολής του νοηματικού περιεχομένου ανεξάρτητα από την ταχύτητα με την οποία συμβαίνει. Τελικός στόχος είναι το μέτρο της καμπυλότητας να εκφράζει τον τρόπο της μεταβολής στον χώρο των χαρακτηριστικών, άρα οι πιο απότομες μεταβολές να δίνουν μεγαλύτερη καμπυλότητα απ' ότι οι πιο ομαλές (ακόμα και αν η απότομη μεταβολή υπήρχε σε διάρκεια πολλών καρέ με κοντινά σημεία, δηλαδή με μικρή τιμή στη δεύτερη παράγωγο ως προς τον χρόνο). Εξάλλου, στη θεωρία της διαφορικής γεωμετρίας, η καμπυλότητα (ή οι καμπυλότητες για πολυδιάστατη καμπύλη) ορίστηκαν ώστε να περιγράφουν πλήρως τη γεωμετρική μορφή μιας καμπύλης.

## 1.4 Οργάνωση της Εργασίας

Η μελέτη της ακολουθίας κινούμενων εικόνων ξεκινάει από την κατάταμσή της σε πλάνα, αφού έτσι αρχικά διαχωρίζονται τμήματα του βίντεο με κοινό περιεχόμενο. Το θέμα αυτό μελετάται στο Κεφάλαιο 2, όπου πρώτα αναφέρονται τεχνικές που εργάζονται σε ασυμπίεστο αλλά και κωδικοποιημένο MPEG βίντεο, με τα πλεονεκτήματα και τα μειονεκτήματά τους και κατόπιν κατασκευάζεται σύστημα κανόνων για τον εντοπισμό αλλαγής πλάνου. Για τη διαδικασία αυτή χρησιμοποιείται η κωδικοποιημένη πληροφορία του MPEG και με συνδυασμό μεθόδων εφαρμόζεται ένα πολύπλοκο σύστημα κανόνων για την ανίχνευση των πλάνων (shot detection).

Στο Κεφάλαιο 3 εξετάζεται το θέμα της εξαγωγής των χαρακτηριστικών ιδιοτήτων (features) από κάθε καρέ, διαδικασία που έπεται της κοπής του βίντεο στα πλάνα. Οι τιμές των χαρακτηριστικών προσδιορίζουν ένα σημείο στον χώρο των χαρακτηριστικών, έναν χώρο πολλών διαστάσεων όπου τα διαδοχικά σημεία (καρέ) σχηματίζουν μία καμπύλη, στην οποία τα σημεία απότομης καμπύλωσης αντιστοιχούν σε δραματικές αλλαγές στο περιεχόμενο της ακολουθίας. Αφού πρώτα ομαλοποιηθεί η καμπύλη με ένα γκαουσιανό φίλτρο, τα χαρακτηριστικά καρέ που θα συνοψίσουν το



βίντεο εξάγονται από την καμπύλη είτε μέσω του ορισμού της καμπυλότητας είτε μέσω δύο μεθόδων επιλογής χαρακτηριστικών σημείων της καμπύλης.

Στο Κεφάλαιο 4 παρουσιάζεται η μαθηματική μέθοδος της καμπυλότητας (curvature) για την ανάλυση και την περιγραφή της καμπύλης στον πολυδιάστατο χώρο, η οποία έχει τις βάσεις στη διαφορική γεωμετρία των καμπυλών. Για την καμπυλότητα δίνονται τέσσερις διαφορετικοί τύποι, αλλά χρησιμοποιούνται μόνο οι δύο στην τελική εφαρμογή του συστήματος, ο πρώτος και ο τέταρτος. Ο πρώτος περιλαμβάνει χρονικές παραγώγους και χρονική εξέλιξη του βίντεο (των καρέ), ενώ ο τέταρτος βασίζεται σε θεωρία για πολυδιάστατες καμπύλες. Στο τέλος του Κεφαλαίου παρατίθενται παραδείγματα με εφαρμογή και συγκρίσεις των τεσσάρων τύπων για την καμπυλότητα.

Στο Κεφάλαιο 5 περιγράφονται αναλυτικά δύο άλλες μέθοδοι χειρισμού της πολυδιάστατης καμπύλης και επιλογής χαρακτηριστικών σημείων της, όπως έχουν παρουσιαστεί στη βιβλιογραφία. Η πρώτη αφορά την απλοποίηση της καμπύλης (curve simplification), που είναι η υποβάθμιση των διαστάσεων της καμπύλης σε μικρότερη διάσταση, δηλαδή σε έναν χώρο μικρότερης διάστασης όπου “χωράει” μια σύνοψη της καμπύλης, την οποία και χρησιμοποιούμε για τα χαρακτηριστικά σημεία. Η δεύτερη αναφέρεται στην προσέγγιση του περιγράμματος ενός σχήματος (shape contour approximation), η οποία έχει παρουσιαστεί για διδιάστατη καμπύλη και γι’ αυτό εδώ προσαρμόζεται για τις  $N$  διαστάσεις.

Το Κεφάλαιο 6 πραγματεύεται την υπολογιστική μέθοδο της οπτικής προσοχής (visual attention), η οποία χρησιμοποιείται για να βελτιώσει τα αποτελέσματα των μεθόδων εξαγωγής χαρακτηριστικών καρέ. Το μοντέλο της οπτικής προσοχής εντοπίζει σε κάθε καρέ περιοχές με ιδιαίτερα σημαντικό περιεχόμενο, από τις οποίες εξάγονται τα χαρακτηριστικά και κατασκευάζεται η καμπύλη στον χώρο. Στο τέλος του Κεφαλαίου υπάρχουν τα αποτελέσματα της εφαρμογής της οπτικής προσοχής σε ακίνητες εικόνες και οι αντίστοιχες παρατηρήσεις.

Στο Κεφάλαιο 7 παρουσιάζονται τα πειράματα που διενεργούνται σε ένα δείγμα από βίντεο για τον έλεγχο των δυνατοτήτων του συστήματος, σημειώνονται οι επιδόσεις για την ανίχνευση πλάνων και συγκρίνονται οι δύο μέθοδοι της καμπυλότητας με τις δύο εναλλακτικές μεθόδους του Κεφαλαίου 5 για τη σύνοψη της καμπύλης. Καταγράφονται οι παρατηρήσεις πάνω στις μετρήσεις, ενώ εξετάζεται και ο βαθμός βελτίωσης των αποτελεσμάτων από την εφαρμογή του μοντέλου της οπτικής προσοχής.

Τέλος, στο Κεφάλαιο 8 παρατίθενται παρατηρήσεις και συμπεράσματα που προκύπτουν από τα συνολικά αποτελέσματα του συστήματος περίληψης βίντεο, με βάση

τις επιδόσεις, τις δυνατότητες αλλά και τις ελλείψεις του. Επίσης, γίνεται μία σύντομη αναφορά στις προοπτικές βελτίωσης και επέκτασης του συστήματος, καθώς αναφέρονται μέθοδοι που θεωρούνται άξιες προσοχής για να υλοποιηθούν μελλοντικά.

## ΚΕΦΑΛΑΙΟ 2

# *Χρονική Κατάτμηση του Βίντεο σε*

## *Πλάνα*

### 2.1 Εισαγωγικά

Το πρώτο και βασικότερο βήμα, που πρέπει πάντα να προηγείται της διαδικασίας της περιήγησης και της δημιουργίας περίληψης είναι η χρονική κατάτμηση του βίντεο σε ξεχωριστά μέρη, καθένα από τα οποία αποτελεί ένα πλάνο. Το πλάνο ορίζεται ως η δομική μονάδα του βίντεο που αποτελείται από διαδοχικές κινούμενες εικόνες που παρουσιάζουν χρονική συνέχεια ([Bro00]). Το πλάνο είναι η καταγεγραμμένη ακολουθία από συνεχόμενα καρέ που προκύπτουν από μία συγκεκριμένη κάμερα με εγγραφή χωρίς διακοπή ([Bor96], [Cal02]). Στην αγγλική βιβλιογραφία χρησιμοποιείται περισσότερο ο όρος “shot” και μερικές φορές ο όρος “scene”, οπότε εδώ λαμβάνεται ως αντικειμενικώς αποδεκτή έκφραση το “shot”. Πολλά συνεχόμενα πλάνα που συσχετίζονται μεταξύ τους, εστιάζοντας σε κάποιο συγκεκριμένο αντικείμενο ή γενικότερα σε κάποιο κοινό θέμα ενδιαφέροντος αποτελούν μία σκηνή (story), δηλαδή μία μικρή αυτοτελή ιστορία ([Bor96], [Cal02]). Οι σκηνές αυτές μπορούν να ταξινομηθούν σε διάφορα επίπεδα σημαντικότητας και περιεχομένου και έτσι να υλοποιηθεί ένα πολυεπίπεδο μοντέλο για μια ιεραρχική δομή του βίντεο ([Han99]). Οι αγγλικοί όροι που χρησιμοποιούνται είναι οι “scene” και “story”. Για την αποφυγή σύγχυσης, από εδώ και στο εξής για το πλάνο (αδιάκοπη εγγραφή της κάμερας) θα χρησιμοποιείται ο όρος πλάνο (shot) και για τη

συλλογή σημασιολογικά συσχετιζόμενων πλάνων θα χρησιμοποιείται ο όρος σκηνή (scene).

Μέσα σε ένα βίντεο οι διάφορες γειτονικά πλάνα, που μπορεί να αφορούν την ίδια ή και διαφορετική σκηνή, συνδέονται μεταξύ τους με κάποιους συγκεκριμένους τρόπους μετάβασης, που ονομάζονται μεταβάσεις πλάνου (shot transitions) ([Bor96], [Cal02]). Οι κυριότεροι από αυτούς είναι οι παρακάτω:

1. **Cut :** Πρόκειται για την πιο απότομη μετάβαση (τομή), η οποία πραγματοποιείται σε ένα μόνο καρέ, αφού το τελευταίο καρέ του προηγούμενου πλάνου ακολουθείται από το πρώτο καρέ του επόμενου πλάνου και η μετάβαση γίνεται αμέσως.
2. **Fade :** Είναι η αργή μεταβολή στη φωτεινότητα (αύξηση ή μείωση) σε κάθε εικόνα, με αποτέλεσμα τη σταδιακή εμφάνιση (fade in) ή εξαφάνιση (fade out) του πλάνου, από ή προς μία μαύρη εικόνα αντίστοιχα.
3. **Dissolve :** Η διάλυση, που μπορεί να θεωρηθεί ότι είναι ένα ταυτόχρονο fade-out και fade-in, όπως και το fade, πραγματοποιείται σε ένα σύνολο από συνεχόμενα καρέ. Συμβαίνει όταν η εικόνα του προηγούμενου πλάνου γίνεται σταδιακά πιο σκοτεινή, ενώ η εικόνα του επόμενου πλάνου σταδιακά αυξάνει σε φωτεινότητα.
4. **Wipe :** Συμβαίνει όταν η εικόνα του πρώτου πλάνου κινείται με συγκεκριμένο τρόπο και σχήμα, βγαίνοντας από το καρέ και δίνοντας σταδιακά τη θέση της στην εικόνα του επόμενου πλάνου.

Επομένως η χρονική κατάτμηση του βίντεο σε πλάνα αφορά στην ανίχνευση των μεταβάσεων μεταξύ των πλάνων, την ταξινόμηση τους σε μια από τις προηγούμενες κατηγορίες και τον εντοπισμό των ορίων των πλάνων (shot boundaries).

Από εδώ και στο εξής θα χρησιμοποιούνται πολλές φορές οι αγγλικοί όροι για καλύτερη ακρίβεια και πιστότητα στην ορολογία της ανάλυσης βίντεο, αλλά και για αποφυγή σύγχυσης κατά την διατύπωση, ειδικότερα εξαιτίας του ότι κάποιοι όροι δεν χρησιμοποιούνται συχνά στην ελληνική γλώσσα.

Ο πιο διαδεδομένος τρόπος αποθήκευσης και κωδικοποίησης κινούμενων εικόνων είναι η κωδικοποίηση MPEG, πρότυπο το οποίο βασίζεται στην κίνηση των blocks από κάθε πλαίσιο στο επόμενο ([ISO93]). Αν εκμεταλλευτούμε την πληροφορία που υπάρχει αποθηκευμένη στο MPEG, όπως είναι τα διανύσματα κίνησης, ο τύπος και

οι ιδιότητες των Macroblocks, ή το σφάλμα κωδικοποίησης, η ανίχνευση των πλάνων μπορεί να γίνει απευθείας στο κωδικοποιημένο πεδίο του MPEG προτύπου, εξοικονομώντας χρόνο αλλά επιτυγχάνοντας και καλύτερα αποτελέσματα. Δεδομένου ότι πρόκειται για εφαρμογές πάνω σε μεγάλες βάσεις δεδομένων από βίντεο μεγάλης διάρκειας και δεδομένου ότι η ανάλυση των πλάνων αποτελεί μόνο το πρώτο βήμα στην ανάλυση και την περιγραφή του περιεχομένου του βίντεο, η ταχύτητα των υπολογισμών αποτελεί βασικό παράγοντα της ποιότητας της μεθόδου.

Στο Κεφάλαιο αυτό αρχικά υπάρχει μία σύντομη περιγραφή του τρόπου κωδικοποίησης του MPEG προτύπου, με τα είδη των frames και την περιγραφή των Macroblocks. Στη συνέχεια παρατίθενται τεχνικές εντοπισμού αλλαγών πλάνου που έχουν ήδη προταθεί και εφαρμόζεται, είτε σε ασυμπίεστο είτε σε συμπιεσμένο (MPEG κωδικοποιημένο) βίντεο, ενώ παράλληλα αναφέρονται τα πλεονεκτήματα και τα μειονεκτήματά τους. Στην τελευταία παράγραφο κατασκευάζεται ένα σύστημα κανόνων που έχει στόχο την ανίχνευση των πλάνων (που χωρίζονται με cut) στο πεδίο MPEG με την μεγαλύτερη δυνατή ακρίβεια, αλλά και ταχύτητα. Χρησιμοποιείται συνδυασμός μεθόδων και στα τρία είδη των frames του MPEG, επιλέγοντας τις τεχνικές που έχουν καλές επιδόσεις σύμφωνα με τη βιβλιογραφία, αλλά και επιπλέον τεχνικές που δίνουν καλά αποτελέσματα σε διάφορες δοκιμές. Τελικά κατασκευάζεται το τελικό διάγραμμα απόφασης που περιλαμβάνει τον ταξινομητή, ο οποίος και ταξινομεί τα καρέ ανάλογα με αν αποτελούν το πρώτο ή κάποιο ενδιάμεσο καρέ ενός πλάνου.

## **2.2 Σύντομη Περιγραφή της Κωδικοποίησης MPEG**

### **2.2.1 Το πρότυπο MPEG**

Το πρότυπο MPEG αποτελεί τον πλέον δημοφιλή και αποτελεσματικό τρόπο για την κωδικοποίηση και τη συμπίεση κινούμενων εικόνων, με σκοπό την αποθήκευση, τη μεταφορά και την αναπαραγωγή της. Από το 1988 που πρωτοεμφανίστηκε μέχρι και σήμερα έχουν διατυπωθεί διάφορες εκδόσεις του προτύπου (MPEG 1 – 2 – 4 – 7), καθεμιά από τις οποίες ανταποκρίνεται σε διαφορετικές απαιτήσεις και προδιαγραφές κωδικοποίησης ([Κολλ01]).

Το πρώτο μοντέλο είναι το MPEG 1, πάνω στο οποίο βασίστηκαν όλα τα επόμενα. Καλύπτει εφαρμογές κωδικοποίησης βίντεο και ήχου, με τυπικό ρυθμό μετάδοσης 1.5 Mbits/sec (1.25 Mbits/sec για το video και 250 Kbits/sec για δύο κανάλια audio), ενώ είναι κατάλληλο για αποθήκευση ήχου και βίντεο σε CD-ROM και γενικά για εφαρμογές σε υπολογιστικά συστήματα.

Το βίντεο αποτελείται από ακολουθία κινούμενων εικόνων, όπου οι διαδοχικές εικόνες που μεταδίδονται με σταθερό ρυθμό (συνήθως 25 εικόνες ανά sec) παρουσιάζουν χωρική και χρονική συσχέτιση μεταξύ τους, γεγονός που εκμεταλλεύεται το MPEG για να συμπίεσει την πληροφορία. Τα πλαίσια (frames) του βίντεο ομαδοποιούνται ανά ομάδες GOP (Group of Pictures), τα οποία κωδικοποιούνται και αποκωδικοποιούνται ανεξάρτητα. Υπάρχουν τρία είδη κωδικοποιημένων πλαισίων, όπως περιγράφεται παρακάτω.

## *2.2.2 Τα τρία είδη των frames στο MPEG*

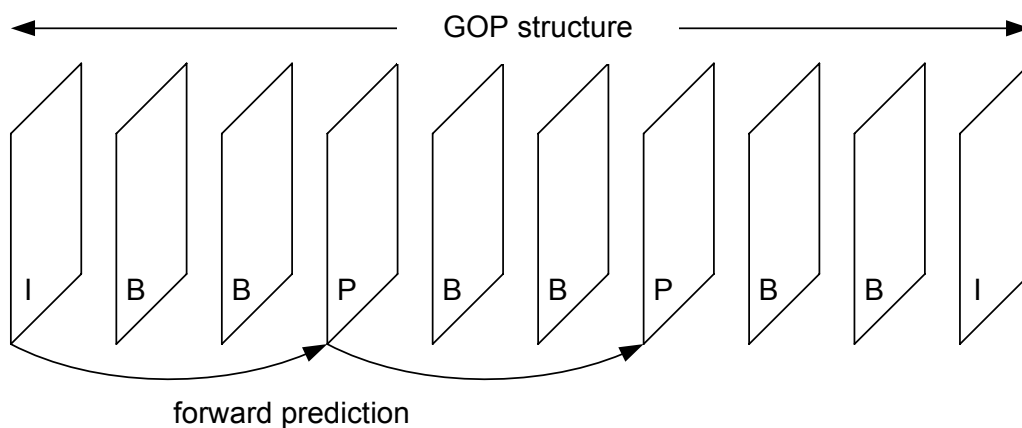
### **2.2.2.1 I-frames (Intra frames)**

Τα I-frames είναι εικόνες που δεν χρησιμοποιούν εκτίμηση κίνησης, αλλά μεταδίδουν συνολικά την πληροφορία της εικόνας σε μορφή κωδικοποιημένη σύμφωνα με το πρότυπο JPEG. Η διαδικασία που ακολουθείται είναι η εξής: η εικόνα χωρίζεται σε υποεικόνες (blocks) των 8x8 ή 16x16 pixels και εφαρμόζεται ο διακριτός μετασχηματισμός συνημιτόνου (Discrete Cosine Transformation - DCT) σε καθεμιά από αυτές. Στην μετασχηματισμένη πλέον εικόνα στο DCT πεδίο εφαρμόζεται η διαδικασία του κβαντισμού και της συμπίεσης-κωδικοποίησης (Huffman), οπότε τελικά κατασκευάζεται η εικόνα σε συμπιεσμένη μορφή, αλλά με όλη την πληροφορία και γι' αυτό χρησιμοποιείται και ως πλαίσιο αναφοράς για τα υπόλοιπα είδη πλαισίων.

### **2.2.2.2 P-frames (Predicted frames)**

Στα P-frames χρησιμοποιείται η διαδικασία της εκτίμησης και αντιστάθμισης κίνησης (motion estimation and compensation), δηλαδή μεταδίδεται η πληροφορία της κίνησης των blocks από το προηγούμενο πλαίσιο αναφοράς (το προηγούμενο I ή P

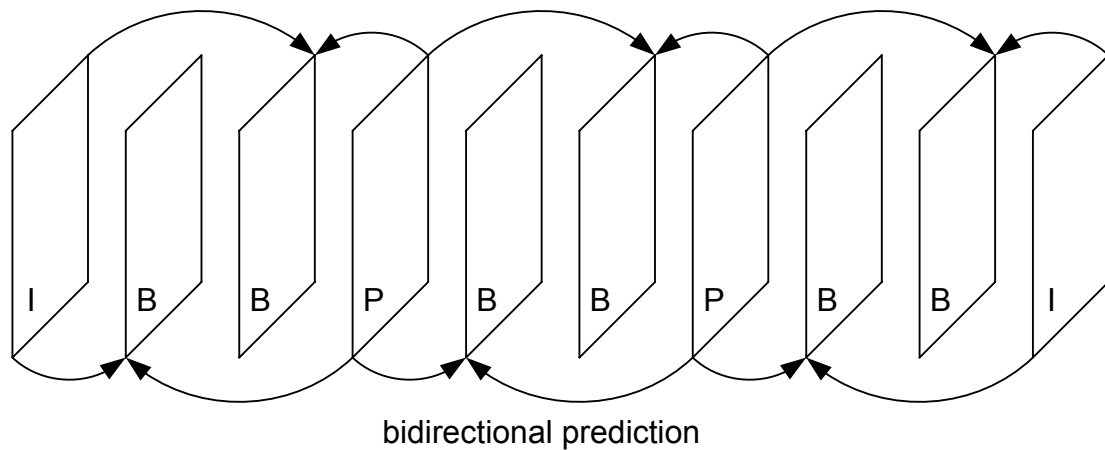
frame) προς το τρέχον πλαίσιο P, καθώς και το σφάλμα που προκύπτει από αυτή τη διαδικασία. Η πληροφορία της κίνησης αναπαρίσταται μέσω των διανυσμάτων κίνησης, ενώ το σφάλμα μεταδίδεται με τη μορφή κωδικοποιημένης με τον DCT μετασχηματισμό εικόνας. Υπάρχει σημαντική ελευθερία στον τρόπο εξαγωγής της πληροφορίας κίνησης, γεγονός που επιτρέπει συνεχώς τη βελτίωση της ποιότητας της κωδικοποίησης και της εκμετάλλευσης των ιδιοτήτων του προτύπου. Στο σχήμα 2.1 φαίνεται ο τρόπος της αναφοράς για την προς τα εμπρός πρόβλεψη (forward prediction) των P-frames.



Σχ. 2.1 : Προς τα εμπρός πρόβλεψη στα P-frames του MPEG

### 2.2.2.3 B-frames (Bidirectional Interpolated frames)

Στα B-frames πραγματοποιείται και πάλι η διαδικασία εκτίμησης και αντιστάθμισης κίνησης, μόνο που εδώ χρησιμοποιείται το προηγούμενο πλαίσιο αναφοράς (το προηγούμενο I ή P frame) ή το επόμενο πλαίσιο αναφοράς (το επόμενο I ή P frame) ή ακόμη και τα δύο ταυτόχρονα, ενώ το frame συντίθεται μέσω διαδικασίας χωρο-χρονικά ισοσταθμισμένης γραμμικής παρεμβολής. Το σφάλμα που προκύπτει από αυτή τη διαδικασία μεταδίδεται και εδώ, κωδικοποιημένο στο DCT πεδίο. Η πληροφορία της κίνησης αναπαρίσταται μέσω διανυσμάτων κίνησης, που εδώ είναι δύο ειδών: διανύσματα προς τα εμπρός κίνησης (forward motion vectors) με αναφορά στο προηγούμενο πλαίσιο αναφοράς (forward prediction), και διανύσματα προς τα πίσω κίνησης (backward motion vectors) με αναφορά στο επόμενο πλαίσιο αναφοράς (backward prediction). Η διαδικασία της αναφοράς για την κίνηση προς τις δύο κατευθύνσεις (bidirectional prediction) για τα B-frames φαίνεται στο σχήμα 2.2.

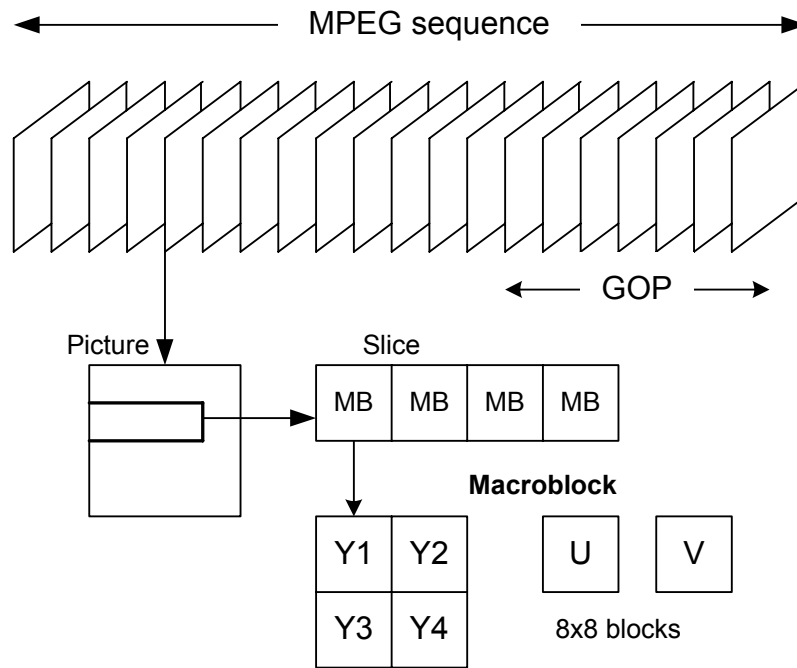


Σχ. 2.2 : Πρόβλεψη διπλής κατεύθυνσης στα B-frames του MPEG

### 2.2.3 Περιγραφή των MacroBlocks

Για την κωδικοποίηση της εικόνας στα Intra frames (I), αλλά και για την εκτίμηση της κίνησης στα Inter frames (P, B), η εικόνα χωρίζεται σε blocks 16x16 pixels το καθένα από τα οποία τελικά αποτελεί ένα Macroblock, που έχει σημαντικό ρόλο στην κωδικοποίηση του MPEG. Ένα Macroblock αποτελείται συνολικά από έξι blocks μεγέθους 8x8 pixels, τα οποία είναι τέσσερα blocks φωτεινότητας ( $Y_1, Y_2, Y_3, Y_4$ ) και δύο blocks για κάθε χρωματικό κανάλι (U, V). Και αυτό διότι ο λόγος δειγματοληψίας (sampling ratio) για τα τρία κανάλια Y:U:V είναι 4:1:1, που είναι και ο συνηθέστερος χρησιμοποιούμενος λόγος ([Sik97]). Παρακάτω (σχ. 2.3) δίνεται σχηματικά η θέση και η δομή ενός Macroblock (MB) μέσα στην ακολουθία MPEG.





Σχ. 2.3 : Περιγραφή ενός Macroblock (MB) στο MPEG

Σύμφωνα με το πρότυπο MPEG υπάρχουν τρεις διαφορετικοί τύποι κωδικοποίησης ενός Macroblock ([Sik97]):

- **Inter MB:** Σε αυτά τα Macroblocks χρησιμοποιείται πρόβλεψη εκτίμησης κίνησης (είτε προς τα εμπρός είτε προς τα πίσω), με βάση το προηγούμενο και το επόμενο πλαίσιο αναφοράς. Μεταδίδεται ο τύπος, η θέση του MB, το διάνυσμα ή τα διανύσματα κίνησης (αν υπάρχει κίνηση) και οι DCT συνιστώσες που περιέχουν το σφάλμα.
- **Skipped MB:** Είναι τα Macroblocks που ο αλγόριθμος εκτιμά ότι δεν μεταβάλλονται (προκύπτει δηλαδή πολύ μικρή διαφορά), άρα είναι αυτά που μένουν όπως είναι στο προηγούμενο πλαίσιο. Δεν μεταδίδεται καμία πληροφορία, ούτε διάνυσμα κίνησης, ούτε κωδικοποιημένο σφάλμα. Αυτά τα Macroblocks έχουν ίδιες τιμές με τα αντίστοιχα στο αμέσως προηγούμενο πλαίσιο αναφοράς.
- **Intracoded MB:** Σε αυτή την περίπτωση δεν χρησιμοποιείται εκτίμηση κίνησης, γιατί ο αλγόριθμος βρίσκει πολύ μεγάλη διαφορά από οποιοδήποτε Macroblock που ανήκει στην περιοχή αναζήτησης στο πλαίσιο αναφοράς, οπότε μεταδίδεται ολόκληρη η πληροφορία. Δεν μεταδίδεται κανένα διάνυσμα κίνησης, αλλά ο τύπος, η θέση του MB

και οι DCT συνιστώσες, που περιέχουν ακριβώς τις τιμές του MB κωδικοποιημένες με τη μέθοδο του JPEG.

Ο τύπος κάθε MB αναφέρεται στην επικεφαλίδα του MB (header). Στα I-frames ο τύπος των Macroblocks είναι μόνο intracoded, αφού αυτά τα πλαίσια δεν χρησιμοποιούν καθόλου πρόβλεψη, αλλά κωδικοποιούνται εξ' αρχής. Στα P-frames υπάρχουν και τα τρία είδη Macroblocks, αλλά τα inter MB περιορίζονται μόνο σε προς τα εμπρός πρόβλεψη (forward prediction). Στα B-frames υπάρχουν επίσης και οι τρεις τύποι Macroblocks, αλλά τα inter MB περιλαμβάνουν MB που χρησιμοποιούν κίνηση προς τα εμπρός, προς τα πίσω ή προς τις δύο κατευθύνσεις (bidirectional interpolated prediction) με δύο πλαίσια αναφοράς.

Τέλος, θα πρέπει να σημειωθεί ότι η σειρά των frames (με βάση τον τύπο τους) είναι διαφορετική στην περίπτωση της αναπαραγωγής και διαφορετική στην περίπτωση της μετάδοσης (ή αποθήκευσης). Αυτό συμβαίνει διότι στον δέκτη για να γίνει σωστά η αναπαραγωγή θα πρέπει να γίνει η αποκωδικοποίηση, η οποία για τα I-frames είναι η αποκωδικοποίηση του JPEG και για τα P-frames χρησιμοποιούνται οι αποκωδικοποιημένες εικόνες από τα I-frames αναφοράς (για τα I και P δεν αλλάζει η σειρά). Όμως για την αποκωδικοποίηση των B-frames πρέπει να έχει προηγηθεί η αποκωδικοποίηση των πλαισίων αναφοράς για τις δύο κατευθύνσεις (forward και backward prediction), τα οποία είναι το προηγούμενο και το επόμενο I- ή P-frame. Γι' αυτόν τον λόγο όλα τα B-frames αναγκαστικά μεταδίδονται αμέσως μετά το επόμενο I- ή P-frame που τα ακολουθεί στη σειρά της αναπαραγωγής.

Στο σχήμα 2.4 δίνεται ένα παράδειγμα όπου φαίνεται η διαφορά στην ακολουθία της σειράς αναπαραγωγής (display) και της σειράς μετάδοσης (transmit):

<b><u>Sequence in Display Order:</u></b>																		
I	B	B	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B	I
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

<b><u>Sequence in Transmit Order:</u></b>																		
I	P	B	B	P	B	B	I	B	B	P	B	B	P	B	B	I	B	B
1	4	2	3	7	5	6	10	8	9	13	11	12	16	14	15	19	17	18

**Σχ. 2.4 :** Δύο τρόποι διάταξης των frames στην ακολουθία MPEG

## 2.3 Υπάρχουσες Τεχνικές για την Ανίχνευση Πλάνων

Επειδή το αντικείμενο αυτής της εργασίας δεν στοχεύει μόνο στην κατάτμηση του βίντεο σε πλάνα, αλλά οφείλει να μελετήσει και το θέμα της εξαγωγής χαρακτηριστικών καρέ από κάθε πλάνο, ο εντοπισμός των μεταβάσεων του πλάνου περιορίζεται μόνο σε τομές (cuts). Τα cuts είναι ο βασικότερος και ο πιθανότερος τρόπος μετάβασης από το ένα πλάνο στο επόμενο, και αυτό ισχύει για τα περισσότερα είδη του περιεχομένου του βίντεο. Η πλειοψηφία των μεταβάσεων ανήκουν στην κατηγορία cuts, μικρό ποσοστό ανήκει στην κατηγορία dissolves, ενώ ακόμη λιγότερες μεταβάσεις χαρακτηρίζονται ως wipes ή fade-ins, fade-outs. Τα dissolves εντοπίζονται σε όλα τα είδη, τα fades σχεδόν αποκλειστικά σε κινηματογραφικές ταινίες ή σε ντοκιμαντέρ, ενώ τα wipes σε ζωντανές εκπομπές, σε αναμεταδόσεις αθλητικών γεγονότων και σε ειδήσεις. Πάντως, περισσότερο από το 80 % των μεταβάσεων ανήκει στην κατηγορία cuts και γι' αυτό θεωρείται σημαντική όσο και αποτελεσματική η μεμονωμένη μελέτη τους. Παρακάτω παρατίθενται διάφορες μέθοδοι ανίχνευσης πλάνων που έχουν ήδη υλοποιηθεί και δίνουν ικανοποιητικά αποτελέσματα, τουλάχιστον όσον αφορά την κατηγορία ανίχνευσης απότομης αλλαγής (cut detection) ([Bor96], [Lie99]).

### 2.3.1 Μέθοδοι Εφαρμοσμένες σε Ασυμπιεστο Βίντεο

#### 2.3.1.1 Διαφορές των Pixels

Ο πιο απλός τρόπος για την ανίχνευση ενός cut μεταξύ δύο διαδοχικών καρέ είναι ο ορισμός ενός μέτρου της διαφοράς τους (pair-wise difference). Αυτό μπορεί να υλοποιηθεί μετρώντας το πλήθος των pixels των οποίων η απόλυτη διαφορά μεταξύ των διαδοχικών εικόνων  $i-1$  και  $i$  υπερβαίνει ένα κατώφλι  $t$  ([Zha93]). Αθροίζοντας το πλήθος των pixels που αλλάζουν και διαιρώντας δια του πλήθους των συνολικών pixels της εικόνας, προκύπτει ένα ποσοστό  $D_1$  που εκφράζει το μέγεθος της μεταβολής από τη μία εικόνα στην άλλη και καθορίζει την ύπαρξη ενός cut, αν αυτή υπερβαίνει ένα δεύτερο κατώφλι  $T$ .

Ο παραπάνω κανόνας εκφράζεται με τις σχέσεις (για εικόνα  $M \times N$  pixels):

$$\text{Για κάθε pixel } P(x, y) : \quad DP_i(x, y) = \begin{cases} 1, & \text{if } |P_i(x, y) - P_{i-1}(x, y)| > t \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Μέτρο της μεταβολής:} \quad D_1(i) = \frac{\sum_{x=1}^M \sum_{y=1}^N DP_i(x, y)}{M * N} < T \quad (\text{εξ. 2.1})$$

Η μέθοδος αυτή λειτουργεί ικανοποιητικά, επιλέγοντας βέβαια κατάλληλο ζεύγος  $t$  και  $T$  (thresholds), το οποίο μετά από πειράματα έχει εκτιμηθεί ότι μπορεί να είναι 10 και 0.75 αντίστοιχα, ώστε να καλύπτονται πολλές περιπτώσεις βίντεο ([Whi03]). Το κύριο μειονέκτημα της μεθόδου έγκειται στο ότι είναι πολύ ευαίσθητη σε κίνηση, είτε της κάμερας είτε ενός μεγάλου αντικειμένου. Σε αυτή την περίπτωση πολλά pixels μεταβάλλονται με αποτέλεσμα να αυξάνεται το μέτρο της μεταβολής  $D_i$  και συνεπώς να εντοπίζεται εσφαλμένα η αλλαγή πλάνου. Προσπάθεια για βελτίωση αυτής της αδυναμίας έγινε από τους Zhang et al. [Zha93], με χρήση φίλτρου ομαλοποίησης (smoothing), αλλά χωρίς να εξαλειφθεί εντελώς.

Άλλες εναλλακτικές μέθοδοι για μέτρηση της μεταβολής των pixels είναι το απόλυτο άθροισμα των διαφορών των pixels ( $D_2$ ), το άθροισμα των απόλυτων διαφορών των pixels ( $D_3$ ) και το άθροισμα των τετραγώνων των διαφορών των pixels ( $D_4$ ), που δίνονται από τους τύπους ([For00], [Bes02]):

$$D_2 = \left| \sum_{x=1}^M \sum_{y=1}^N \{P_i(x, y) - P_{i-1}(x, y)\} \right| / (M * N) \quad (\text{εξ. 2.2})$$

$$D_3 = \sum_{x=1}^M \sum_{y=1}^N |P_i(x, y) - P_{i-1}(x, y)| / (M * N) \quad (\text{εξ. 2.3})$$

$$D_4 = \sum_{x=1}^M \sum_{y=1}^N (P_i(x, y) - P_{i-1}(x, y))^2 / (M * N) \quad (\text{εξ. 2.4})$$

Σημειώνεται ότι τα αθροίσματα διαιρούνται με το πλήθος των pixels, έτσι ώστε να προκύψει μία μέση μεταβολή ανά pixel, και στη συνέχεια χρησιμοποιείται και πάλι ένα κατάλληλο threshold για τον εντοπισμό ενός cut.

Τέλος, μια εικόνα θα μπορούσε να θεωρηθεί ως μονοδιάστατο διάνυσμα μήκους  $M*N$ , τοποθετώντας τα pixels σε μία διάσταση. Γνωρίζοντας ότι η προβολή του ενός διανύσματος πάνω στο άλλο, ή αλλιώς το εσωτερικό γινόμενο, αποτελεί μέτρο της

ομοιότητάς τους (παίρνει τη μέγιστη τιμή όταν τα διανύσματα συμπίπτουν), λαμβάνουμε ως μέτρο μεταβολής το:

$$D_5 = 1 - \frac{\vec{V}_{i-1} \bullet \vec{V}_i}{\|\vec{V}_{i-1}\| \cdot \|\vec{V}_i\|} \quad (\text{εξ. 2.5})$$

Όλες οι παραπάνω μετρικές ανήκουν στην ίδια κατηγορία των ντετερμινιστικών μεθόδων που βασίζονται στη σύγκριση pixel προς pixel (point-wise methods) και γι' αυτό επιτυγχάνουν στον εντοπισμό των πραγματικών cuts χωρίς να χάνουν απότομες αλλαγές πλάνου (θα έπρεπε να είναι σχεδόν ίδια τα καρέ των δύο πλάνων για να αποτύχει η μέθοδος). Έχουν όμως το πρόβλημα του θορύβου και της κίνησης κάμερας και αντικειμένων, που αυξάνουν το μέτρο μεταβολής και εμφανίζουν ανύπαρκτες αλλαγές πλάνου, αλλά και το μειονέκτημα του απαιτούμενου χρόνου υπολογισμού.

### 2.3.1.2 Διαφορές Ιστογράμματος

Το ιστόγραμμα (grey-level ή color histogram) είναι ο πιο γνωστός και σταθερός τρόπος για τον εντοπισμό αλλαγής πλάνου ([Bor96], [Bro00]), κυρίως διότι μένει ανεπηρέαστο από οποιαδήποτε μεγάλη ή μικρή κίνηση μέσα στην εικόνα και αντιμετωπίζει τον θόρυβο σε ικανοποιητικό βαθμό. Για να υπολογιστεί η διαφορά μεταξύ των δύο εικόνων, αφού κατασκευαστεί το ιστόγραμμα, υπολογίζεται το άθροισμα των απόλυτων διαφορών των στοιχείων (bins) του ιστογράμματος κάθε εικόνας. Για color histogram με χρωματικές συνιστώσες r, g, b, με B bins για κάθε χρώμα και H(r, g, b) τον αριθμό των pixels που ανήκουν στο χρώμα r, g, b, το μέτρο της διαφοράς των δύο εικόνων δίνεται από τη σχέση ([Lie99]):

$$D_6 = \sum_{r=0}^{B-1} \sum_{g=0}^{B-1} \sum_{b=0}^{B-1} |H_i(r, g, b) - H_{i-1}(r, g, b)| \Big/ (M * N) \quad (\text{εξ. 2.6})$$

Σημειώνεται ότι ο μικρός αριθμός των bins επιτυγχάνει τη μείωση της επίδρασης του θορύβου που υπεισέρχεται σε ένα ή σε περισσότερα διαδοχικά καρέ, των απότομων αυξομειώσεων της φωτεινότητας (π.χ. φλας φωτογραφικής μηχανής), της κίνησης αντικειμένων ή της μεταβολής τους σε σχήμα ή χρώμα, αλλά κυρίως της κίνησης της

κάμερας. Η απόφαση για εντοπισμό μετάβασης πλάνου γίνεται και εδώ με κάποιο κατώφλι, ή ακόμα και δύο κατώφλια σε περίπτωση σταδιακής μετάβασης (dissolve, fade), που όμως θα πρέπει να δοκιμάζονται σε κάθε διαφορετικό βίντεο για να βρεθεί μια κατάλληλη τιμή.

### 2.3.1.3 Στατιστικές Μετρικές

Οι στατιστικές μέθοδοι βασίζονται στην ιδέα του διαχωρισμού της εικόνας σε περιοχές και στη σύγκριση στατιστικών μετρήσεων πάνω στα pixels των αντίστοιχων περιοχών των δύο εικόνων ([Bor96]), ενώ μπορούν να υπολογιστούν και σε ολόκληρη την εικόνα. Οι στατιστικές μετρήσεις είναι η μέση τιμή και η τυπική απόκλιση ή η έκφραση κάποιας πιθανότητας και εφαρμόζονται πάνω στις τιμές ενός χαρακτηριστικού (φωτεινότητα  $Y$ , χρωματικά κανάλια  $C_b$ ,  $C_r$ ) ή στις τιμές του ιστογράμματος ([For00]). Η μέση τιμή  $\mu$ , η διασπορά  $V$  και η τυπική απόκλιση  $\sigma$  για  $N$  τιμές (pixels) ενός χαρακτηριστικού  $C$  σε μια εικόνα δίνονται από τις σχέσεις:

$$\mu(C) = \frac{\sum_{i=1}^N C_i}{N} \quad (\text{εξ. 2.7})$$

$$V(C) = \mu\{(C - \mu(C))^2\} = \mu(C^2) - \mu^2(C) \quad (\text{εξ. 2.8})$$

$$\sigma(C) = \sqrt{V(C)} = \sqrt{\frac{\sum_{i=1}^N (C_i - \mu(C))^2}{N}} \quad (\text{εξ. 2.9})$$

Έχουν προταθεί διάφοροι τύποι για την έκφραση του λόγου πιθανότητας (Likelihood Ratio) της ομοιότητας μεταξύ των δύο εικόνων. Αν  $a$ ,  $b$  θεωρηθούν δύο διαδοχικά καρέ,  $C$  ένα χαρακτηριστικό των pixels,  $\mu_a$ ,  $\mu_b$ ,  $\sigma_a$ ,  $\sigma_b$  η μέση τιμή και η τυπική απόκλιση για τις δύο εικόνες για το  $C$ , και  $\sigma_{ab}^2$  η συνδιακύμανση, μερικές εκφράσεις για την πιθανότητα ομοιότητας των δύο εικόνων δίνονται παρακάτω ([For00], [Bes02]):

$$\text{Student t-test:} \quad t = \sqrt{N} \cdot \frac{|\mu_a - \mu_b|}{\sqrt{\sigma_a^2 + \sigma_b^2}} \quad (\text{εξ. 2.10})$$

$$\text{Likelihood Ratio}_1: \quad LR_1 = \frac{\left[ \frac{\sigma_a^2 + \sigma_b^2}{2} + \left( \frac{\mu_a - \mu_b}{2} \right)^2 \right]^2}{\sigma_a^2 \cdot \sigma_b^2} \quad (\text{εξ. 2.11})$$

$$\text{Likelihood Ratio}_2: LR_2 = \frac{\sigma_{ab}^4}{\sigma_a^2 \cdot \sigma_b^2} \quad (\text{εξ. 2.12})$$

Οπότε οι αντίστοιχες μετρικές που δείχνουν ένα ποσοστό της διαφοράς μεταξύ των δύο καρτέ είναι:

$$D_7 = t \quad (\text{εξ. 2.13})$$

$$D_8 = 1 - \min\left(LR_1, \frac{1}{LR_1}\right) \quad (\text{εξ. 2.14})$$

$$D_9 = 1 - \min\left(LR_2, \frac{1}{LR_2}\right) \quad (\text{εξ. 2.15})$$

Οι στατιστικές μέθοδοι γενικά δεν επηρεάζονται από θόρυβο, ενώ είναι ευαίσθητες στην κίνηση μόνο σε περίπτωση που η εικόνα χωριστεί σε περιοχές μικρότερες από τα κινούμενα αντικείμενα. Με εφαρμογή σε όλη την εικόνα ή σε τέσσερις μεγάλες περιοχές, η κίνηση των αντικειμένων εξαλείφεται, αλλά και πάλι η κίνηση της κάμερας επηρεάζει σημαντικά τα αποτελέσματα. Είναι επίσης σχετικά πιο αργές από τις προηγούμενες μεθόδους, λόγω της πολυπλοκότητας υπολογισμού των στατιστικών τύπων.

#### 2.3.1.4 Μετρικές Βασισμένες σε Ακμές

Η μέθοδος αυτή βασίζεται στην ανίχνευση της θέσης των ακμών στα δύο καρτέ και στον υπολογισμό των pixels των ακμών που εισέρχονται και εξέρχονται από την εικόνα. Αν  $\rho_{in}$  είναι το ποσοστό των ακμών που εισέρχονται και  $\rho_{out}$  είναι το ποσοστό των ακμών που εξέρχονται, οι Zabih et al. ([Zab99]) ορίζουν ως βασικό μέτρο ανομοιότητας των δύο εικόνων την τιμή:

$$D_{10} = \max(\rho_{in}, \rho_{out}) \quad (\text{εξ. 2.16})$$

Ο αλγόριθμος που χρησιμοποιείται είναι αρκετά πολύπλοκος, δεδομένου ότι για τις δύο εικόνες υπολογίζει με φιλτράρισμα δύο άλλες εικόνες ακμών, ανιχνεύει τις ακμές, στη συνέχεια υπολογίζει τα pixels που εξέρχονται και τέλος υπολογίζει το μέτρο

ανομοιότητας. Οι Ford et al. ([For00]) για να μειώσουν την πολυπλοκότητα προτείνουν τον υπολογισμό των ακμών σε κάθε εικόνα,  $edges(a)$  και  $edges(b)$ , και ορίζουν το μέτρο:

$$D_{11} = \frac{edges(a) - edges(b)}{edges(a) + edges(b)} \quad (\text{εξ. 2.17})$$

Η μέθοδος αυτή έχει αποδειχθεί πολύ πιο ακριβής στην ανίχνευση των cuts συγκριτικά με τις υπόλοιπες μεθόδους και είναι πολύ λιγότερο ευαίσθητη σε θόρυβο και σε κίνηση. Παρόλ' αυτά έχει υψηλή πολυπλοκότητα και μεγάλες υπολογιστικές απαιτήσεις, γεγονός που δυσχεραίνει τη χρήση της σε βίντεο μεγάλης διάρκειας.

### 2.3.2 Μέθοδοι Εφαρμοσμένες σε Συμπιεσμένο Βίντεο

Η ανίχνευση αλλαγής πλάνου και ειδικότερα η ανίχνευση ενός cut, όπως φαίνεται από τις μεθόδους που αναφέρθηκαν παραπάνω, είναι μία διαδικασία με γνωστή κυρίως απλή υλοποίηση και αναμενόμενα αποτελέσματα, τα οποία συνήθως είναι και αρκετά ικανοποιητικά. Όμως η υψηλή πολυπλοκότητα (αφού πρόκειται για εξαντλητικό αλγόριθμο που επαναλαμβάνεται για όλα τα pixels της εικόνας) συνεπάγεται και αυξημένο χρονικό κόστος. Για τον λόγο αυτό και επειδή ο πιο διαδεδομένος τρόπος αποθήκευσης αρχείων βίντεο είναι η κωδικοποίηση MPEG, είναι προτιμότερο η ανίχνευση των πλάνων να γίνεται απευθείας στην κωδικοποιημένη MPEG ακολουθία, χωρίς να προηγείται η αποκωδικοποίηση που απαιτεί και αυτή κάποιο χρόνο. Παρακάτω παρουσιάζονται κάποιες προσεγγίσεις για την ανίχνευση των πλάνων στο MPEG κωδικοποιημένο βίντεο.

#### 2.3.2.1 Ανίχνευση σε I-frame

Στην ακολουθία MPEG, το I-frame είναι μια εικόνα με μεγάλη πληροφορία, κωδικοποιημένη με παρόμοιο τρόπο με το πρότυπο JPEG ([Κολλ01]). Τα I-frames είναι οι εικόνες της ακολουθίας με την καλύτερη ποιότητα, οι οποίες χρησιμοποιούνται και ως εικόνες αναφοράς για την αποκωδικοποίηση των υπόλοιπων frames. Για την ανίχνευση αλλαγής πλάνου σε ακολουθία κωδικοποιημένη στο MPEG πρότυπο ο πιο απλός τρόπος



είναι η εφαρμογή μιας από τις μεθόδους του ασυμπίεστου βίντεο σε δύο γειτονικά I-frames (δηλαδή τα πιο κοντινά).

Εκτός από τη μέθοδο των ακμών, οι υπόλοιπες μέθοδοι μπορούν να εφαρμοστούν απευθείας στο κωδικοποιημένο συμπιεσμένο πεδίο, και αντί των τιμών των pixels της εικόνας να χρησιμοποιηθούν οι DC τιμές των DCT blocks της εικόνας στο συμπιεσμένο πεδίο ([Bes02]). Στο πρότυπο JPEG, η εικόνα έχει χωριστεί σε υποεικόνες (blocks) και σε κάθε block έχει εφαρμοστεί ο διακριτός μετασχηματισμός συνημιτόνου (DCT). Έτσι για κάθε block προκύπτει μια DC συνιστώσα, που δείχνει τον μέσο όρο της έντασης των pixels του block, και άλλες AC συνιστώσες που δείχνουν τη διακύμανση των τιμών της έντασης.

Εφαρμόζοντας λοιπόν τη μέθοδο μόνο με τις DC τιμές από τα blocks της εικόνας, είναι δυνατός ο εντοπισμός πιθανής αλλαγής πλάνου, η οποία μπορεί να συμβαίνει σε οποιοδήποτε σημείο μεταξύ των δύο I-frames. Κάθε ένα από τα frames μέχρι και το δεύτερο I είναι πιθανό να είναι η αρχή του νέου πλάνου.

Η χρήση των DC τιμών (μέσος όρος έντασης του DCT μπλοκ) συμβάλλει στην εξάλειψη του θορύβου και κυρίως στη μείωση της ευαισθησίας των μεθόδων σημείου προς σημείο (point-wise) σε κίνηση αντικειμένων και κάμερας, προσεγγίζοντας τον τρόπο της βελτίωσης που επιχείρησαν οι Zhang et al. [Zha93] με φιλτράρισμα (smoothing) σε 3x3 blocks. Επίσης, οι Bescós et al. ([Bes02]) εφάρμοσαν πολλές από τις γνωστές μετρικές μεθόδους (διαφορές των pixels, στατιστικές τεχνικές) σε ασυμπίεστο βίντεο και σε συμπιεσμένο κωδικοποιημένο MPEG βίντεο. Παρατήρησαν ότι στο συμπιεσμένο βίντεο με χρήση DC τιμών επιτυγχάνεται μεγαλύτερος βαθμός διαχωρισιμότητας των πλάνων και ότι τελικά είναι προτιμότερος ο συνδυασμός μεθόδων.

### 2.3.2.2 Ανίχνευση σε P-frame

Τα P-frames στην ακολουθία MPEG είναι συμπιεσμένες εικόνες με πρόβλεψη κίνησης που χρησιμοποιούν ως frame αναφοράς το προηγούμενο I-frame και διανύσματα κίνησης που δείχνουν την κατεύθυνση της κίνησης των Macroblocks (MB) της εικόνας από το I στο P frame. Αν συμβεί μια αλλαγή πλάνου μεταξύ του I-frame και του τρέχοντος P-frame, τότε πολύ λίγα (σχεδόν κανένα) MB του P θα έχουν πρόβλεψη κίνησης από το I ([Boc00], [Men95]). Αυτό συμβαίνει επειδή το σφάλμα πρόβλεψης προκύπτει μεγάλο και γι' αυτό ο κωδικοποιητής κωδικοποιεί πολλά Macroblocks που

ξεπερνούν κάποιο όριο σφάλματος ως intracoded MB (που σημαίνει ότι μεταδίδει όλη την πληροφορία για αυτά μέσα στο DCT block του P-frame). Στην αλλαγή πλάνου, όπου το frame θα είναι πολύ διαφορετικό από το προηγούμενο frame αναφοράς, το πλήθος των intracoded MB θα αυξηθεί απότομα και η αλλαγή μπορεί να εντοπιστεί παρακολουθώντας τον λόγο των intracoded MB προς τα συνολικά Macroblocks της εικόνας ([Boc00]):

$$ratio = \frac{Intracoded MB}{Total MB} \quad (\text{εξ. 2.18})$$

όπου Total MB είναι τα συνολικά Macroblocks στα οποία έχει χωριστεί η εικόνα για την κωδικοποίηση MPEG, που αντιστοιχούν σε μέγεθος 16x16 pixels της εικόνας.

Οι Meng et al. ([Men95]) προτείνουν τον λόγο των intracoded MB προς τα MB που δεν είναι intracoded, έτσι ώστε να προκύπτουν πιο απότομες και καθαρές αιχμές (peaks) σε σημεία αλλαγής πλάνου. Και στους δύο λόγους απαιτείται ένα κατώφλι για τον καθορισμό των peaks, το οποίο θα πρέπει να προσαρμόζεται ανάλογα με το βίντεο.

### 2.3.2.3 Ανίχνευση σε B-frame

Τα B-frames της κωδικοποίησης MPEG χρησιμοποιούν πρόβλεψη κίνησης με δύο frames αναφοράς, το προηγούμενο και το επόμενο, που μπορεί να είναι ένα I- ή ένα P-frame. Υπάρχουν διανύσματα κίνησης που δείχνουν κίνηση προς τα εμπρός ή και προς τα πίσω (γίνεται παρεμβολή). Όταν συμβεί μια αλλαγή πλάνου κοντά σε ένα B-frame, τότε ανάλογα με τη θέση της αλλαγής, τα διανύσματα στο B-frame θα δείχνουν κατά πλειοψηφία προς τα εμπρός ή κατά πλειοψηφία προς τα πίσω. Δηλαδή τα διανύσματα προς το frame αναφοράς που ανήκει στο ίδιο πλάνο με το τρέχον B-frame θα είναι περισσότερα από τα διανύσματα προς την αντίθετη κατεύθυνση.

Οι Boccignone et al. ([Boc00]) αναφέρουν ως μέτρο για τον έλεγχο της αλλαγής πλάνου τον λόγο των διανυσμάτων της προς τα πίσω κίνησης προς τα συνολικά διανύσματα κίνησης του frame, ενώ οι Meng et al. ([Men95]) προτείνουν τον λόγο του πλήθους των προς τα πίσω διανυσμάτων προς το πλήθος των προς τα εμπρός διανυσμάτων. Πάντως, δεν αναφέρουν συγκεκριμένες συνθήκες για κάθε περίπτωση της θέσης της αλλαγής πλάνου, όπως για παράδειγμα οι Calic et al. [Cal01], που

διατυπώνουν λεπτομερώς κάθε περίπτωση, δηλαδή αν το πρώτο frame του νέου πλάνου είναι το πρώτο ή το δεύτερο B-frame, ή αν είναι το επόμενο frame αναφοράς (I ή P). Παρακολουθώντας απλά το πλήθος των προς τα εμπρός διανυσμάτων (forward predicted) σε σχέση με το πλήθος των προς τα πίσω διανυσμάτων (backward predicted), είναι δυνατή η ανίχνευση των αλλαγών πλάνου σε κάθε τύπου frame της ακολουθίας.

## 2.4 Συνδυασμός Μεθόδων και Κατασκευή Κανόνων

Όπως ήδη αναφέρθηκε στην προηγούμενη παράγραφο, οι αλγόριθμοι που επεξεργάζονται τις εικόνες σε επίπεδο pixel (ασυμπίεστο - uncompressed video) οδηγούν σε σχετικά καλά ποσοστά επιτυχίας στις αλλαγές πλάνου. Παρόλ' αυτά παρουσιάζουν αυξημένη υπολογιστική πολυπλοκότητα, καθώς εφαρμόζουν τεχνικές σε όλα τα pixels της εικόνας και εκτελούν φιλτραρίσματα (συνελίξεις) και πράξεις που απαιτούν επαναλήψεις εισάγοντας σημαντική καθυστέρηση χρόνου. Επίσης, πολλές από αυτές τις μεθόδους απαιτούν τον προσδιορισμό της τιμής για ένα κατώφλι (ή και περισσότερα μερικές φορές), το οποίο θα πρέπει να προσαρμόζεται ανάλογα με το βίντεο και άρα δεν έχει μεγάλη πρακτική αξία, αν κάθε φορά χρειάζεται νέα εκτίμηση της τιμής του.

Στις περισσότερες προηγούμενες εργασίες πάνω στην έρευνα της ανίχνευσης πλάνων η προσοχή εστιάζεται συνήθως σε μία μέθοδο και επιχειρείται η σύγκριση αυτής με τις υπόλοιπες μεθόδους. Πιο σπάνια χρησιμοποιείται ένας συνδυασμός από περισσότερες τεχνικές ([Nar98], [DeS02]), όπου τα βήματα των μεθόδων συσχετίζονται και με τη βοήθεια ενός δέντρου απόφασης κατασκευάζεται ένα σύστημα ταξινόμησης με πολύ καλές επιδόσεις. Όπως παρατηρούν και οι Bescós et al. ([Bes02]), στην έρευνα που έκαναν συγκρίνοντας διάφορες μεθόδους, ο συνδυασμός κάποιων τεχνικών είναι προτιμότερος από την εφαρμογή ενός μόνο αλγορίθμου. Αφού πρώτα έδειξαν ότι στις περισσότερες μεθόδους που δουλεύουν σε επίπεδο pixel (pixel-wise), αν εφαρμοστούν στις DC συνιστώσες του DCT πεδίου, αποδίδουν καλύτερα αποτελέσματα απ' ότι στο επίπεδο pixel, προσπάθησαν να συνδυάσουν τεχνικές (στο DCT πλέον) μεταξύ τους ανά δύο και να μετρήσουν τη σύγκλιση της συνδυασμένης μεθόδου, εξάγοντας ένα μέτρο για τη σύγκριση και την επιλογή του πιο ικανοποιητικού συνδυασμού.

Με βάση από τις προηγούμενες παρατηρήσεις, στο παρόν σύστημα είναι λογικό να επιλεγεί η ανίχνευση μιας αλλαγής πλάνου (shot change) να γίνει απευθείας στην συμπίεσμένη μορφή του MPEG. Χρησιμοποιούνται διαφορετικές μέθοδοι που εκμεταλλεύονται την κωδικοποιημένη πληροφορία στα frames του MPEG, αλλά θα πρέπει να συνδυαστούν με κάποια από τις γνωστές μεθόδους ( $D_1 - D_9$ ) η οποία θα εφαρμοστεί στις DC συνιστώσες του I-frame. Η επεξεργασία στο πεδίο MPEG επιτυγχάνει γρηγορότερους υπολογισμούς αλλά και βελτιωμένα αποτελέσματα, αφού μπορεί να μειωθεί η επίδραση θορύβου ή μεγάλης κίνησης μέσω της χρήσης των DC τιμών (μέσος όρος κάθε μπλοκ). Η προσοχή εστιάζεται στις απότομες αλλαγές πλάνου (cuts) και στο τέλος οι παραπάνω μέθοδοι συνδυάζονται σε ένα δέντρο αποφάσεων όπου με συγκεκριμένα βήματα θα καταλήγει σε έναν ταξινομητή (classifier) ο οποίος θα ταξινομεί τα καρέ σε δύο κατηγορίες, δηλαδή α) σε καρέ που αποτελούν το πρώτο καρέ του νέου πλάνου (εντοπισμός αλλαγής πλάνου), και β) σε καρέ που ανήκουν στο ίδιο πλάνο με το προηγούμενό τους καρέ.

#### 2.4.1 Κανόνες για τα P-frames

Όταν βρισκόμαστε σε ένα P-frame της ακολουθίας, υπάρχουν πολλοί τρόποι για να διαπιστώσουμε αν από το προηγούμενο frame αναφοράς (I ή P) μέχρι το τρέχον frame είναι πιθανό να συμβαίνει μία αλλαγή πλάνου (δηλαδή στα B-frames που παρεμβάλλονται ή στο τρέχον). Ένας τρόπος είναι να υπολογίσουμε το μέτρο του σφάλματος, αθροίζοντας τις τετραγωνικές ή τις απόλυτες τιμές των DC συνιστωσών, οπότε στα σημεία που αυτό το μέτρο παρουσιάζει απότομες ακμές είναι πιθανή η αλλαγή πλάνου. Δεν λαμβάνεται όμως υπόψη η ύπαρξη των intracoded MB, που υποτίθεται ότι δίνουν μεγάλο σφάλμα και γι' αυτό κωδικοποιούνται εξ' αρχής, άρα δεν γνωρίζουμε πόσο είναι αυτό το σφάλμα ώστε να προστεθεί στο παραπάνω μέτρο. Ειδικά σε περιπτώσεις γρήγορης κίνησης, το μέτρο αυτό δεν εγγυάται σωστά αποτελέσματα, αφού το σφάλμα αυξάνει παρόλο που υπάρχει κίνηση MB από το προηγούμενο πλαίσιο αναφοράς ([Men95]).

Ένας άλλος τρόπος θα ήταν ο έλεγχος των απότομων μεταβολών στον αριθμό των προς τα εμπρός διανυσμάτων πρόβλεψης. Αν συμβεί αλλαγή πλάνου στο διάστημα που ελέγχουμε, το πλήθος των διανυσμάτων από το προηγούμενο πλαίσιο αναφοράς θα μειωθεί. Όμως σε περιπτώσεις ακινησίας αντικειμένων ή κάμερας τα διανύσματα δεν θα

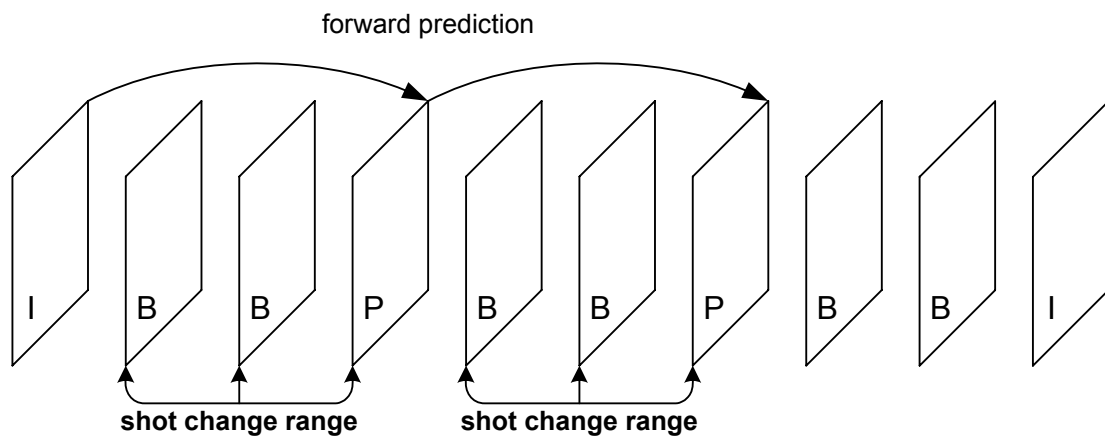
υπάρχουν, αφού τα περισσότερα Macroblocks θα είναι skipped MB, οπότε λανθασμένα θα θεωρήσουμε ότι υπάρχει αλλαγή πλάνου.

Από τους τρεις τύπους των Macroblocks, ο πιο αξιόπιστος για έλεγχο (monitoring) είναι τα intracoded MB, που η ύπαρξή τους δείχνει ότι το σφάλμα κατά τη διαδικασία της εκτίμησης κίνησης είναι μεγάλο, και άρα δεν υπάρχει MB στην γύρω περιοχή στο πλαίσιο αναφοράς που να μοιάζει με το τρέχον. Μετρώντας το πλήθος των intracoded MB και διαιρώντας με το πλήθος των συνολικών Macroblocks της εικόνας, προκύπτει το ποσοστό των intracoded MB:

$$P_{metric} = \frac{\# intracoded MB}{\# total MB} \quad (\text{εξ. 2.19})$$

το οποίο χρησιμοποιείται ως μέτρο έκφρασης της πιθανότητας αλλαγής πλάνου (με κατώφλι  $P_{metric} > P_{thres}$ ) στα frames που αναφέρθηκαν ([Boc00]).

Οι Meng et al. ([Men95]) πρότειναν τη διαίρεση των intracoded MB με εκείνα τα MB που χρησιμοποιούν εκτίμηση κίνησης. Ως MB με εκτίμηση κίνησης θεωρούνται τα inter MB αλλά και τα skipped MB, αφού δεν μπορούν να απορριφθούν τα τελευταία. Άρα η ύπαρξη μεγαλύτερης ή μικρότερης κίνησης δεν εισάγει κάποια επίδραση στον υπολογισμό και γι' αυτό το διάγραμμα δεν βελτιώνει σημαντικά, δηλαδή δεν δημιουργεί πιο απότομες ακμές (peaks). Μια αποτελεσματική προσέγγιση θα ήταν η εκτίμηση της κλίμακας της κίνησης (ακίνησια-αργή-γρήγορη) και ο συνυπολογισμός του πλήθους των intracoded MB και του μεγέθους (μέτρου) ταχύτητας για την κίνηση.



Σχ. 2.5 : Εύρος ανίχνευσης αλλαγής πλάνου σε ένα P-frame (Display Order)

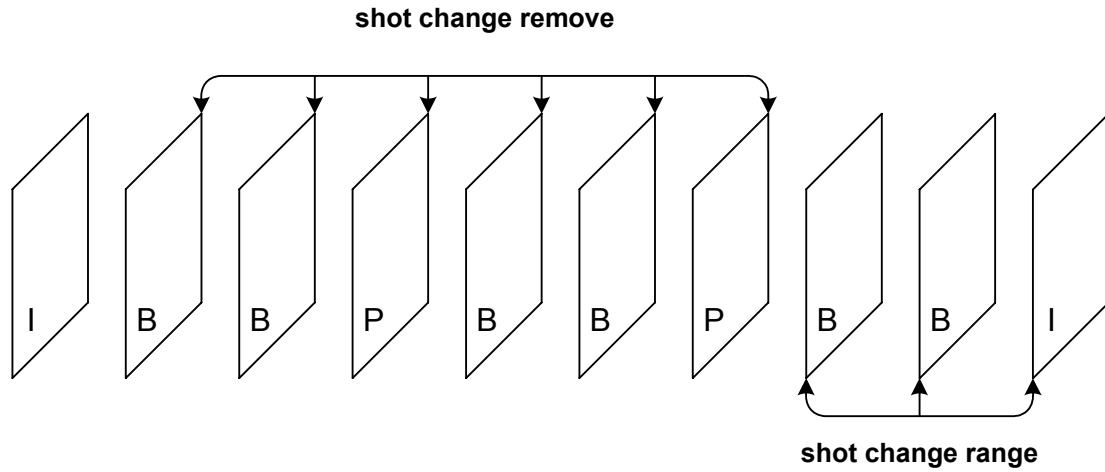
Είναι απαραίτητο να επισημανθεί ότι στα P-frames μπορεί να καθοριστεί μια πιθανή αλλαγή πλάνου που συμβαίνει μόνο στα προηγούμενα B-frames ή στο τρέχον P-frame (shot change range), όπως φαίνεται και στο σχήμα 2.5 σε Display Order. Στην περίπτωση του I-frame και των προηγούμενων B-frames απαιτείται ο ορισμός άλλου τρόπου μέτρησης, όπως παρουσιάζεται και παρακάτω. Αφού θεωρηθεί πιθανή αλλαγή πλάνου στο P-frame, ακολουθεί έλεγχος στα B-frames που ακολουθούν σε Transmit Order (τα B-frames που ακολουθούν στη σειρά μετάδοσης είναι αυτά που προηγούνται στη σειρά αναπαραγωγής), από τον οποίο προκύπτει η τελική απόφαση και η ακριβής θέση του πρώτου frame του νέου πλάνου.

### 2.4.2 Κανόνες για τα I-frames

Όπως αναφέρθηκε στην προηγούμενη παράγραφο, η περίπτωση που δεν λαμβάνεται υπόψη στα P-frames πρέπει να αντιμετωπιστεί στα I-frames. Με αυτήν τη λογική οδηγούμαστε στην ιδέα να υλοποιήσουμε μία από τις μεθόδους  $D_1 - D_9$  στις DCT συνιστώσες του I-frame, που είναι και η πραγματική πληροφορία της εικόνας, δηλαδή θα προκύψουν ακόμη και βελτιωμένα αποτελέσματα από τη μέθοδο που θα χρησιμοποιηθεί ([Bes02]). Σημαντική είναι η επίδραση της κίνησης, αφού εξετάζουμε δύο πλαίσια που απέχουν μεταξύ τους όσο το πλήθος των πλαισίων του GOP και μπορεί στο ενδιάμεσο να παρεμβάλλεται γρήγορη κίνηση που να μεταβάλλει τη δομή της εικόνας, αλλά αυτό το πρόβλημα επιλύεται καλύτερα και από περιπτώσεις pixel-wise μεθόδων, αφού χρησιμοποιούμε τις DC τιμές, που είναι ο μέσος όρος ενός 16x16 block (U, V) ή ενός 8x8 block (Y). Μόνο σε περίπτωση που υπάρχει γρήγορη κίνηση στο μεγαλύτερο μέρος της εικόνας θεωρείται εσφαλμένα πιθανή αλλαγή πλάνου.

Με τον τρόπο αυτό υπολογίζεται η μεταβολή από το ένα I-frame στο επόμενο I-frame και η μεγάλη μεταβολή ή απότομη αύξηση της τιμής δείχνει μια μεγάλη κίνηση ή μια αλλαγή πλάνου μεταξύ των δύο I-frames. Στη συνέχεια, η μία επιλογή που έχουμε είναι να θεωρήσουμε πιθανή αλλαγή πλάνου στο διάστημα των δύο B-frames που προηγούνται και στο τρέχον I-frame (shot change range) και να ακολουθήσει ο έλεγχος στα B-frames που ακολουθούν σε Transmit Order για την τελική απόφαση. Η άλλη επιλογή που έχουμε είναι να θεωρήσουμε πιθανή αλλαγή πλάνου στο υπόλοιπο μέρος του GOP που προηγείται (shot change remove). Στην τελευταία όμως περίπτωση γνωρίζουμε ότι έχουν ήδη προσπελαστεί τα P-frames, έχει ελεγχθεί ο κανόνας για αυτά,

αλλά επίσης και για τα B-frames που τα ακολουθούν και στα οποία έχει ληφθεί η τελική απόφαση. Επομένως εδώ δεν χρειάζεται να ληφθεί απόφαση, παρά μόνο στην περίπτωση που έχει αποφασιστεί αλλαγή πλάνου στο GOP που προηγείται (μέσω του ελέγχου στα P- και B-frames) και υπολογισθεί ότι η μεταβολή από I σε I είναι σχετικά μικρή, τότε αναιρείται η προκαθορισμένη αλλαγή πλάνου (θεωρείται ότι ο έλεγχος στα P-frames είχε σφάλμα). Σχηματικά οι δύο επιλογές, απόφασης και αναίρεσης απόφασης φαίνονται στο σχέδιο του σχήματος 2.6.



**Σχ. 2.6 :** Εύρος ανίχνευσης αλλαγής πλάνου σε ένα I-frame (Display Order)

Οι μέθοδοι που χρησιμοποιούμε για την εκτίμηση της πιθανής αλλαγής πλάνου όταν βρισκόμαστε σε I-frame, δανεισμένες από pixel-wise τεχνικές, είναι οι τρεις παρακάτω, οι οποίες χρησιμοποιούνται με συνδυασμούς ανά δύο ή ανά τρεις μεταξύ τους, διαζευκτικούς ή συζευκτικούς.

Καταρχήν υπολογίζεται η μετρική  $D_I$  που επιτελεί τον κύριο έλεγχο στα I-frames, όπου το μέτρο υπολογίζεται για τις DC τιμές του U ( $D_{IU}$ ) και του V ( $D_{IV}$ ) πεδίου και διαιρώντας δια του δύο προκύπτει ένας μέσος όρος:

$$\text{Για DC τιμή } P(x, y) : DP_i(x, y) = \begin{cases} 1, & \text{if } |P_i(x, y) - P_{i-1}(x, y)| > 10 \\ 0, & \text{otherwise} \end{cases}$$

$$I\ metric_1 = \frac{D_{IU} + D_{IV}}{2} = \frac{\sum_{x=1}^M \sum_{y=1}^N DP_{U_i}(x, y) + \sum_{x=1}^M \sum_{y=1}^N DP_{V_i}(x, y)}{2 * M * N} \quad (\text{εξ. 2.20})$$

Χρησιμοποιώντας δύο κατώφλια (το πρώτο δεδομένο  $t = 10$  και το άλλο όπως ορίζεται παρακάτω), αν ισχύει (**Imetric**<sub>1</sub> < **Ithres**<sub>11</sub>) και έχει αποφασιστεί αλλαγή πλάνου στο προηγούμενο GOP, τότε αναιρείται αυτή η αλλαγή, ενώ αν ισχύει (**Imetric**<sub>1</sub> > **Ithres**<sub>12</sub>), ο αλγόριθμος προχωράει στις επόμενες μετρικές (εδώ φαίνεται ότι αυτό το μέτρο είναι το πιο βασικό, διότι από την αρχή “χοντρικά” αποφασίζει αν θα συνεχίσει ή όχι στην εκτίμηση του shot change range).

Μετά από την D<sub>1</sub>, σε περίπτωση που ικανοποιείται το **Ithres**<sub>12</sub>, υπολογίζουμε δύο άλλες μετρικές, την D<sub>4</sub> και την D<sub>8</sub>, τις οποίες θυμίζουμε εδώ:

$$D_4 = \sum_{x=1}^M \sum_{y=1}^N (P_i(x, y) - P_{i-1}(x, y))^2 / (M * N)$$

$$LR_1 = \frac{\left[ \frac{\sigma_a^2 + \sigma_b^2}{2} + \left( \frac{\mu_a - \mu_b}{2} \right)^2 \right]^2}{\sigma_a^2 \cdot \sigma_b^2}$$

$$D_8 = 1 - \min \left( LR_1, \frac{1}{LR_1} \right)$$

Η μετρική D<sub>4</sub> υπολογίζεται για τα U (D<sub>4U</sub>) και V (D<sub>4V</sub>), καθώς και για το I (D<sub>4I</sub>) ομαδοποιώντας τις τέσσερις DC τιμές που ανήκουν στο ίδιο MB. Η μετρική D<sub>8</sub> υπολογίζεται για τα U (D<sub>8U</sub>) και V (D<sub>8V</sub>). Οπότε λαμβάνουμε ως μέτρα τις τιμές:

$$I metric_2 = \frac{0.0625 * D_{4I} + D_{4U} + D_{4V}}{3} \quad (\text{εξ. 2.21})$$

$$I metric_3 = D_{8U} \quad (\text{εξ. 2.22})$$

$$I metric_4 = D_{8V} \quad (\text{εξ. 2.23})$$

Εδώ χρησιμοποιείται ένα πολύπλοκο σύστημα συζευκτικών και διαζευκτικών προτάσεων, συμπληρώνοντας τον αρχικό χώρο απόφασης που έχει μόνο το μέτρο **Imetric**<sub>1</sub>, όπου οι κανόνες είναι οι εξής:

- Ανίχνευση απότομης μεταβολής: αν ισχύει (**Imetric**<sub>2</sub> > 6 \* **Ithres**<sub>2</sub>), τότε θεωρούμε ότι συμβαίνει απότομη αύξηση στο μέτρο μεταβολής, συγκρίνοντάς το με κατώφλι που προκύπτει δυναμικά σαν μέση τιμή της προηγούμενης τιμής του

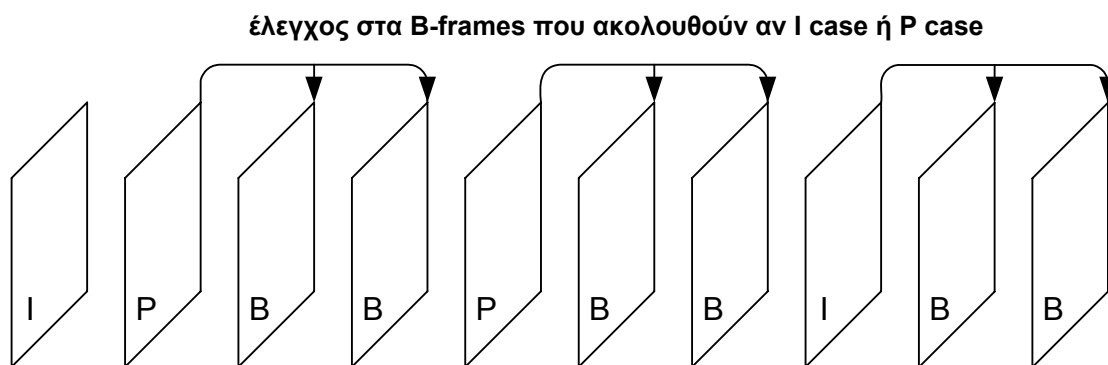


$Imetric_2$  και του μέσου όρου μέχρι αυτό το σημείο της ακολουθίας όλων των τιμών  $Imetric_2$  (το τελευταίο δείχνει τη σχετική κίνηση του βίντεο), ή:

- Ανίχνευση στατιστικής μεταβολής: αν ισχύει ( **$Imetric_3 > Ithres_3$** ) και ( **$Imetric_4 > Ithres_3$** ), τότε θεωρούμε ότι έχουμε πιθανή αλλαγή πλάνου, αφού προκύπτει μεγάλη τιμή στην πιθανότητα ανομοιοτητας για τα χρωματικά πεδία U και V, που δείχνει την διαφορά σε μέση τιμή και διασπορά για τα δύο frames (δηλαδή το πώς διαφέρουν οι κατανομές στα χρώματα στα δύο frames), ή:
- Ανίχνευση διαφοράς από προηγούμενο frame: πρόκειται για την περίπτωση που η μεταβολή που μετράει το  $Imetric_2$  δεν είναι πολύ μεγάλη (απότομη μεταβολή), αλλά είναι μεγάλη σε σχέση με την προηγούμενη τιμή της (δυναμικό κατώφλι  $Ithres_4$ ). Αν ισχύει ( **$Imetric_2 > Ithres_4$** ) και ( **$Imetric_3 > Ithres_5$**  ή  **$Imetric_4 > Ithres_5$** ), τότε θεωρούμε ότι έχουμε πιθανή αλλαγή πλάνου.

### 2.4.3 Κανόνες για τα B-frames

Όταν σε μία από τις δύο προηγούμενες περιπτώσεις χαρακτηρίσουμε πιθανή την ύπαρξη μιας αλλαγής πλάνου (I case και P case), είναι λογική συνέπεια των όσων έχουν αναφερθεί ότι θα πρέπει να ορισθούν κάποιοι μετρικοί κανόνες επιπλέον για τα B-frames που ακολουθούν στην σειρά μετάδοσης (Transmit Order), όπως φαίνονται και στο σχήμα 2.7 τα B-frames που θα πρέπει να εξεταστούν. Με τον τρόπο αυτό θα συμπληρωθεί ο χώρος απόφασης με άλλο ένα κριτήριο το οποίο θα ενισχύει την απόφαση για την αλλαγή πλάνου σε θετική περίπτωση, θα την αποτρέπει σε αρνητική, αλλά και θα εντοπίζει την ακριβή θέση του πρώτου πλαισίου του νέου πλάνου.



**Σχ. 2.7 :** Περίπτωση ελέγχου των B-frames (Transmit Order)

Επειδή στα B-frames έχουμε το πλεονέκτημα που παρέχει η πληροφορία της κίνησης προς τα εμπρός και προς τα πίσω (forward and backward prediction), χρησιμοποιούμε το ποσοστό των διανυσμάτων που δείχνουν προς τα εμπρός κίνηση (forward), από το προηγούμενο πλαίσιο αναφοράς, και το ποσοστό των διανυσμάτων που δείχνουν προς τα πίσω κίνηση (backward), από το επόμενο πλαίσιο αναφοράς. Ανάλογα με τις τιμές των ποσοστών αυτών χαρακτηρίζουμε την κύρια κατεύθυνση της αναφοράς σε αυτό το B-frame (dominant reference), που θα είναι είτε στο προηγούμενο πλαίσιο αναφοράς είτε στο επόμενο πλαίσιο αναφοράς είτε θα μένει απροσδιόριστη. Ως συνολικά θεωρούμε όλα τα διανύσματα που θα μπορούσαν να υπάρχουν, δηλαδή το ίδιο πλήθος με τα Macroblocks της εικόνας. Ορίζοντας τις μετρικές:

$$B metric_1 = \frac{forward\ vectors}{total\ vectors} \quad (\epsilon\xi. 2.24)$$

$$B metric_2 = \frac{backward\ vectors}{total\ vectors} \quad (\epsilon\xi. 2.25)$$

$$B metric_3 = \frac{forward\ vectors}{backward\ vectors} \quad (\epsilon\xi. 2.26)$$

$$B metric_4 = \frac{backward\ vectors}{forward\ vectors} \quad (\epsilon\xi. 2.27)$$

Για τον προσδιορισμό της κύριας κατεύθυνσης αναφοράς (dominant reference), χρησιμοποιούνται οι παρακάτω κανόνες (και επίσης 5 κατώφλια):

- Αν ισχύει:

$$(Bmetric_1 < Bthres_{ff}) \text{ και } (Bmetric_2 > Bthres_{sf})$$

ή αν ισχύει:

$$(Bmetric_4 > Bthres_r) \text{ και } (Bmetric_1 < Bthres_{fe}) \text{ και } (Bmetric_2 > Bthres_{sc})$$

τότε έχουμε αναφορά στο επόμενο πλαίσιο αναφοράς (**Next R**). Αλλιώς:

- Αν ισχύει:

$$(Bmetric_2 < Bthres_{ff}) \text{ και } (Bmetric_1 > Bthres_{sf})$$

ή αν ισχύει:

$$(Bmetric_3 > Bthres_r) \text{ και } (Bmetric_2 < Bthres_{fe}) \text{ και } (Bmetric_1 > Bthres_{sc})$$

τότε έχουμε αναφορά στο προηγούμενο πλαίσιο αναφοράς (**Previous R**). Αλλιώς:

- Αν ισχύει οποιαδήποτε άλλη περίπτωση:

τότε έχουμε απροσδιόριστη κύρια αναφορά.

Τέλος, όταν αναλυθεί και το τελευταίο B-frame (δηλαδή όταν φτάσουμε στο επόμενο R, δηλαδή I- ή P-frame), γνωρίζοντας για όλα τα συσχετιζόμενα B-frames τον τρόπο που αναφέρονται στα προηγούμενα και στα επόμενα (κύρια αναφορά), μπορούμε να βγάλουμε τα εξής συμπεράσματα:

1. Αν όλα τα B-frames χαρακτηρίζονται από Next R (αναφέρονται στο επόμενο πλαίσιο R), τότε το πρώτο B-frame είναι και το πρώτο του νέου πλάνου.
2. Αν όλα τα B-frames μέχρι το k-τάξης (το πλαίσιο με αύξων αριθμό k) χαρακτηρίζονται ως Prev R, ενώ τα επόμενα χαρακτηρίζονται από Next R, τότε το πρώτο B-frame που έχει Next R θα είναι και το πρώτο του νέου πλάνου (k+1).
3. Αν όλα τα B-frames χαρακτηρίζονται από Previous R (αναφέρονται στο προηγούμενο πλαίσιο R), τότε το επόμενο R-frame είναι το πρώτο του νέου πλάνου (δηλαδή το I- ή P-frame που εξετάστηκε μόλις πριν τα B-frames, αφού αναλύουμε την ακολουθία σε Transmit Order).
4. Σε κάθε άλλη περίπτωση δεν υπάρχει αλλαγή πλάνου.

#### 2.4.4 Σύστημα Κανόνων και Διάγραμμα Απόφασης

Συνολικά οι κανόνες που χρησιμοποιούνται είναι συγκεντρωμένοι εδώ, μαζί με τις τιμές των κατωφλίων που χρησιμοποιούνται, αλλά και τις μεθόδους για τα δυναμικά υπολογιζόμενα κατώφλια:

##### **ΚΑΝΟΝΕΣ :**

**Κανόνας 1<sup>ος</sup>:**  $(P_{metric} > P_{thres})$

**Κανόνας 2<sup>ος</sup>:**  $(I_{metric_1} < I_{thres_{11}})$

**Κανόνας 3<sup>ος</sup>:**  $(I_{metric_1} > I_{thres_{12}})$

**Κανόνας 4<sup>ος</sup>:**  $(I_{metric_2} > 6 * I_{thres_2})$

**Κανόνας 5<sup>ος</sup>:**  $((I_{metric_3} > I_{thres_3}) \text{ AND } (I_{metric_4} > I_{thres_3}))$

**Κανόνας 6<sup>ος</sup>:**  $((I_{metric_2} > I_{thres_4}) \text{ AND } (I_{metric_3} > I_{thres_5}) \text{ OR } (I_{metric_4} > I_{thres_5}))$

**Κανόνας 7<sup>ος</sup>:**  $((Bmetric_1 < Bthres_{ff}) \text{ AND } (Bmetric_2 > Bthres_{sf})) \text{ OR } ((Bmetric_4 > Bthres_r) \text{ AND } (Bmetric_1 < Bthres_{fc}) \text{ AND } (Bmetric_2 > Bthres_{sc}))$

**Κανόνας 8<sup>ος</sup>:**  $((Bmetric_2 < Bthres_{ff}) \text{ AND } (Bmetric_1 > Bthres_{sf})) \text{ OR } ((Bmetric_3 > Bthres_r) \text{ AND } (Bmetric_2 < Bthres_{fc}) \text{ AND } (Bmetric_1 > Bthres_{sc}))$

**Κανόνας 9<sup>ος</sup>:** Όλα τα B-frames έχουν κύρια αναφορά Next R.

**Κανόνας 10<sup>ος</sup>:** Όλα τα B-frames έχουν κύρια αναφορά Previous R.

**Κανόνας 11<sup>ος</sup>:** Από τα i B-frames, τα j πρώτα έχουν κύρια αναφορά Previous R, ενώ τα επόμενα (i-j) έχουν κύρια αναφορά Next R.

#### **ΤΙΜΕΣ ΣΤΑ ΚΑΤΩΦΛΙΑ:**

Pthres	=	0.6
Ithres <sub>0</sub>	=	1000
Ithres <sub>11</sub>	=	0.7
Ithres <sub>12</sub>	=	0.6
Ithres <sub>3</sub>	=	0.1
Ithres <sub>5</sub>	=	0.01
Bthres <sub>ff</sub>	=	0.05
Bthres <sub>sf</sub>	=	0.2
Bthres <sub>fc</sub>	=	0.1
Bthres <sub>sc</sub>	=	0.15
Bthres <sub>r</sub>	=	20

Τα κατώφλια για τα B-frames έχουν τη σημασία της επιλογής του κατάλληλου ποσοστού των διανυσμάτων ως εξής (δείκτες: f=few, s=some, c=coarse, f=fine):

- Bthres<sub>ff</sub> : κατώφλι για σχετικά μεγάλο ποσοστό, “λεπτομερώς”
- Bthres<sub>sf</sub> : κατώφλι για σχετικά μικρό ποσοστό, “λεπτομερώς”
- Bthres<sub>fc</sub> : κατώφλι για σχετικά μεγάλο ποσοστό, “χοντρικά”
- Bthres<sub>sc</sub> : κατώφλι για σχετικά μικρό ποσοστό, “χοντρικά”
- Bthres<sub>r</sub> : κατώφλι για τον λόγο αντίθετα κατευθυνόμενων διανυσμάτων

Τα “χοντρικά” κατώφλια είναι πιο “χαλαρά” και χρησιμοποιούνται σε συνδυασμό με το κατώφλι  $B_{thres_i}$  για να ενισχύσουν την απόφαση, ενώ τα “λεπτομερή” κατώφλια εξετάζονται αυτοτελώς, και γι’ αυτό είναι λίγο πιο “αυστηρά”.

Το δυναμικό κατώφλι  $I_{thres_2}$  προκύπτει όπως αναφέρθηκε σαν μέση τιμή της προηγούμενης τιμής του  $I_{metric_2}$  και του μέσου όρου όλων των τιμών  $I_{metric_2}$  των I-frames μέχρι το τρέχον σημείο της ακολουθίας. Ο τελευταίος χρησιμοποιείται επειδή δείχνει γενικά το μέγεθος της μεταβολής (κίνηση, αλλαγές πλάνου) του βίντεο.

Το κατώφλι  $I_{thres_4}$  προκύπτει και αυτό δυναμικά, από τις τιμές των προηγούμενων I-frames, μόνο αν ανήκουν στο ίδιο πλάνο, ως εξής:  
(έστω  $I_{thres_0}$  ένα αρχικό κατώφλι ίσο με 1000 και  $i$  το τρέχον I-frame)

1. Αν τα  $I_i$  και  $I_{i-1}$  δεν ανήκουν στο ίδιο πλάνο:

$$I_{thres_4} = 2 * I_{thres_0}$$

2. Αν τα  $I_i$  και  $I_{i-1}$  ανήκουν στο ίδιο πλάνο:

$$I_{thres_4} = 0.8 * I_{thres_4(i-1)} + 0.2 * I_{thres_0}$$

3. Αν τα  $I_i$  και  $I_{i-1}$  και  $I_{i-2}$  ανήκουν στο ίδιο πλάνο:

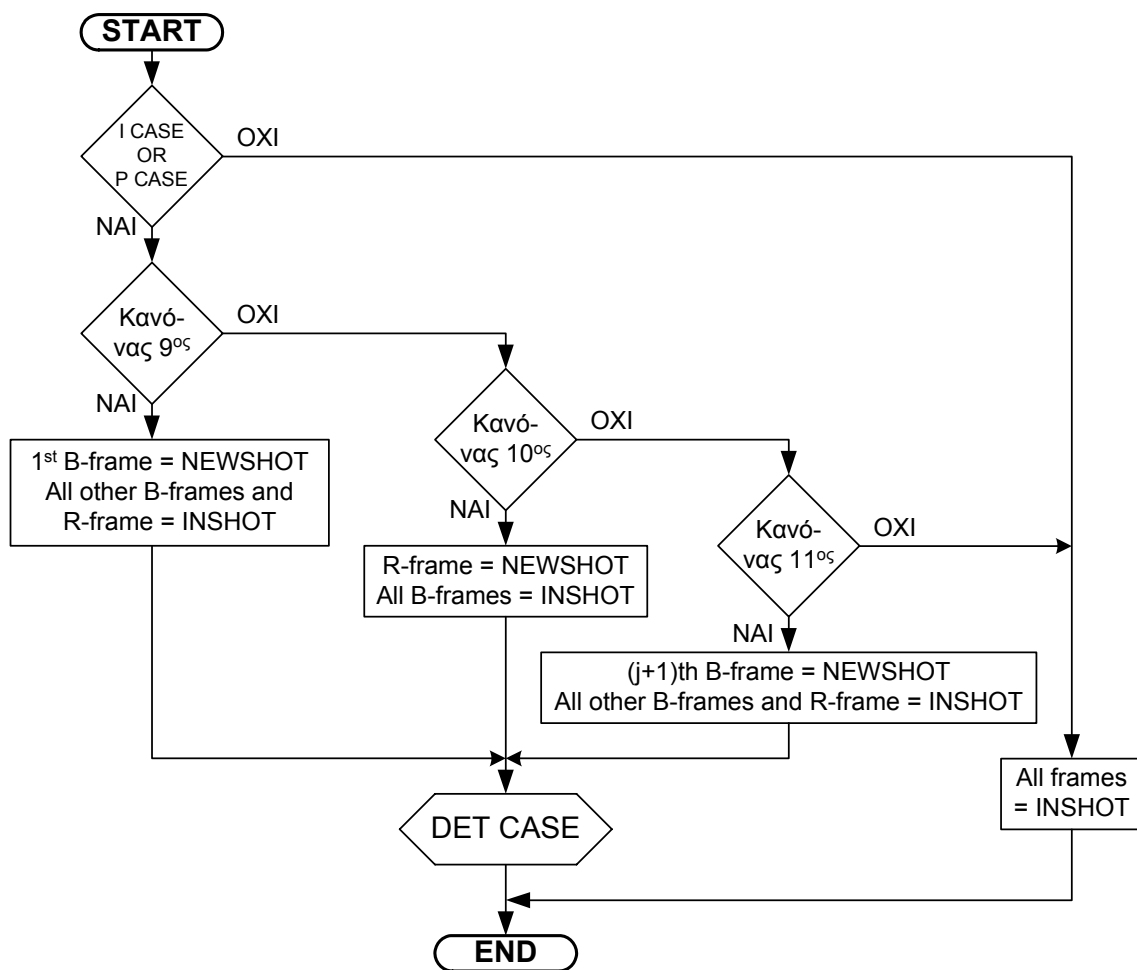
$$I_{thres_4} = 0.7 * I_{thres_4(i-1)} + 0.3 * I_{thres_4(i-2)}$$

4. Αν τα  $I_i$  και  $I_{i-1}$  και  $I_{i-2}$  και  $I_{i-3}$  ανήκουν στο ίδιο πλάνο:

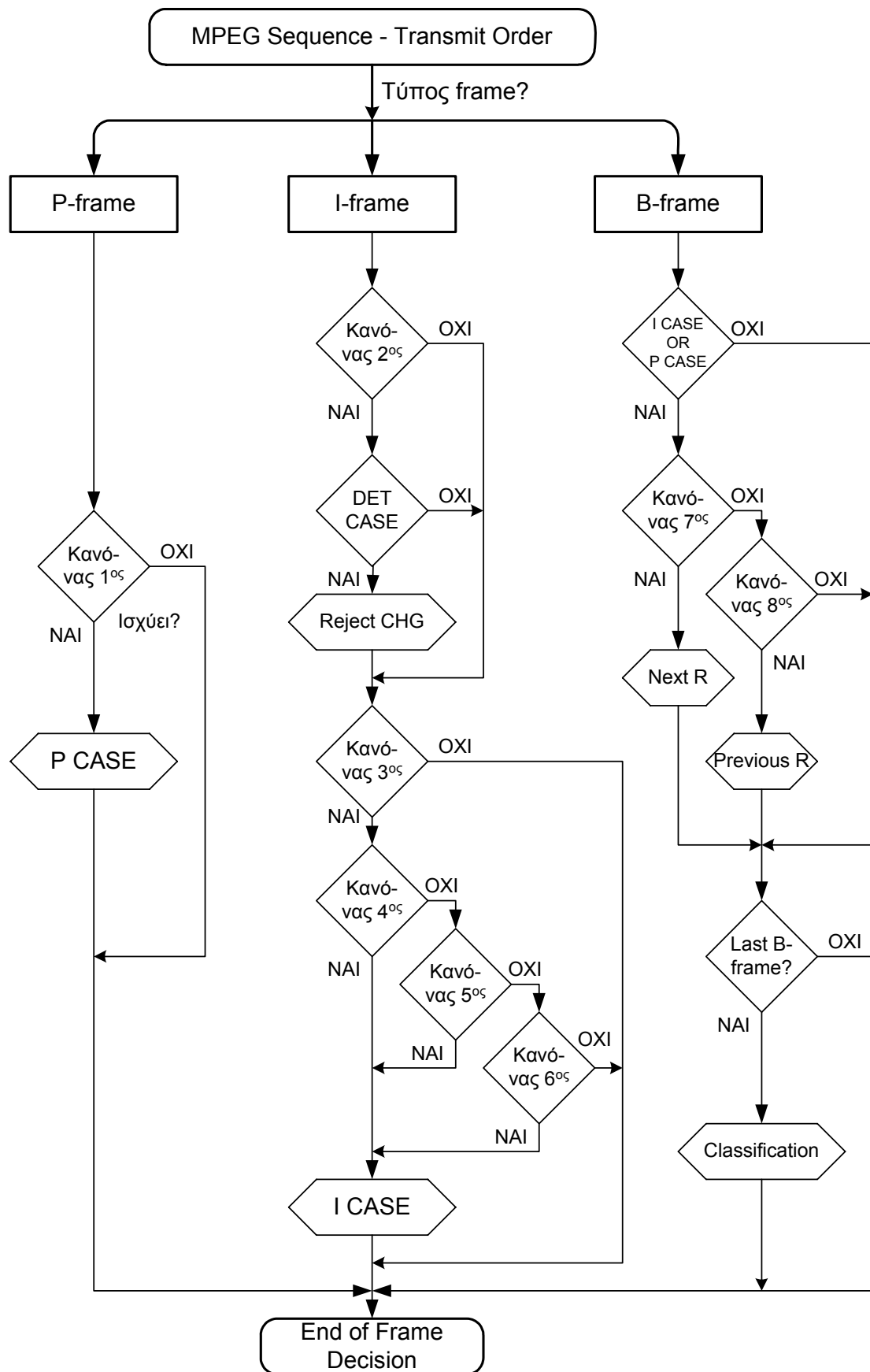
$$I_{thres_4} = 0.6 * I_{thres_4(i-1)} + 0.3 * I_{thres_4(i-2)} + 0.1 * I_{thres_4(i-3)}$$

Εδώ παρουσιάζεται το λογικό διάγραμμα ροής για τη λήψη της τελικής απόφασης – ταξινόμησης (classification), που επιτελεί σε κάθε frame τον έλεγχο των συνθηκών των κανόνων της μεθόδου που περιγράφηκε για την ανίχνευση της αλλαγής πλάνου με απότομη τομή (cut), όπου φαίνεται ο συνδυασμός των κανόνων στο χώρο απόφασης. Η διαδικασία classification ακολουθείται μόνο μετά από R-frame στο οποίο ικανοποιείται ένας από τους κανόνες 3 έως 6. Μόλις εξετασθεί η ισχύς των κανόνων 7 και 8 σε όλα τα B-frames που ακολουθούν το R-frame, τότε καλείται ο ταξινομητής για να ταξινομήσει τα καρέ που αντιστοιχούν στην διαδικασία (το προηγούμενο R-frame και τα B-frames που μόλις εξετάστηκαν). Οι δύο κλάσεις στις οποίες τα ταξινομεί είναι η κλάση INSHOT (όταν το frame ανήκει στο εσωτερικό ενός πλάνου) και η κλάση NEWSHOT (όταν το frame είναι το πρώτο σε ένα νέο πλάνο). Ακολουθούν επεξηγήσεις των βασικών ενεργειών, το διάγραμμα του ταξινομητή (σχ. 2.8) και τέλος το συνολικό διάγραμμα απόφασης (σχ. 2.9).

- P CASE** : Η υπόθεση πιθανής αλλαγής πλάνου από κανόνα σε ένα P-frame
- I CASE** : Η υπόθεση πιθανής αλλαγής πλάνου από κανόνα σε ένα I-frame
- DET CASE** : Η απόφαση αλλαγής πλάνου σε ένα frame από κανόνες στα B-frames
- Reject CHG**: Η απόρριψη (αναίρεση) αλλαγής πλάνου λόγω του 2<sup>ου</sup> κανόνα
- Next R** : Κύρια αναφορά ενός B-frame στο επόμενο frame αναφοράς
- Previous R** : Κύρια αναφορά ενός B-frame στο προηγούμενο frame αναφοράς
- NEWSHOT**: Το frame είναι το πρώτο του νέου πλάνου (αλλαγή πλάνου)
- INSHOT** : Το frame δεν είναι πρώτο, αλλά βρίσκεται μέσα σε ένα πλάνο



**Σχ. 2.8 :** Διάγραμμα της Ταξινόμησης των καρέ (Classification)



Σχ. 2.9 : Συνολικό Διάγραμμα Απόφασης για Αλλαγή Πλάνου





## ΚΕΦΑΛΑΙΟ 3

# *Εξαγωγή και Ομαλοποίηση του Διανύσματος Χαρακτηριστικών*

### 3.1 Εισαγωγικά

Μετά την διαδικασία του εντοπισμού κάθε αλλαγής πλάνου που συμβαίνει στο βίντεο, η ακολουθία των κινούμενων εικόνων διαχωρίζεται σε χρονικά ανεξάρτητα δομικά στοιχεία (πλάνα), από τα οποία το καθένα έχει χρονική και χωρική αυτοτέλεια (αφού πρόκειται για μία συνεχόμενη αδιάκοπη καταγραφή της κάμερας), και γι' αυτό τα καρέ που το αποτελούν ακολουθούν την ίδια νοηματική και σημασιολογική εξέλιξη. Υπάρχει περίπτωση μερικά πλάνα μεταξύ τους να παρουσιάζουν ομοιότητες σε σημασία, περιεχόμενο ή κίνηση, αλλά αυτό προς το παρόν δεν θα εξεταστεί, θεωρώντας κάθε μία ανεξάρτητη από τις υπόλοιπες. Υποθέτοντας ότι η μέθοδος της κοπής σε πλάνα λειτουργεί σωστά (δηλαδή ότι τα πλάνα που προκύπτουν είναι και τα πραγματικά), έχουμε ένα πρώτο βήμα στην προσπάθεια αναπαράστασης του περιεχομένου του βίντεο.

Το επόμενο βήμα είναι η εξαγωγή των αντιπροσωπευτικών καρέ κάθε πλάνου, που θα συνοψίζουν το περιεχόμενό του. Με σκοπό τη σύγκριση και την επιλογή τέτοιων καρέ, εξάγονται οι χαρακτηριστικές ιδιότητες από κάθε καρέ (χαρακτηριστικά), οι οποίες μπορεί να είναι το χρωματικό ιστόγραμμα, η κατάτμηση σε χρωματικές περιοχές, κάποιο μέτρο της κίνησης κτλ. Εδώ χρησιμοποιείται το χρωματικό ιστόγραμμα, το οποίο εξάγεται από τα πεδία φωτεινότητας (luminance) και χρώματος (chrominance) της

εικόνας, οπότε κατασκευάζεται το διάνυσμα των χαρακτηριστικών με στοιχεία τις τιμές του ιστογράμματος (των χαρακτηριστικών ιδιοτήτων).

Το διάνυσμα των χαρακτηριστικών είναι ένα διάνυσμα με πολλές συνιστώσες ορισμένο στον Ευκλείδειο  $N$ -διάστατο χώρο (όπου  $N$  είναι το πλήθος των συνιστωσών του), που ονομάζουμε χώρο των χαρακτηριστικών. Κάθε καρέ του βίντεο αναπαρίσταται με ένα σημείο στον χώρο των χαρακτηριστικών και ολόκληρη η ακολουθία βίντεο αποτελεί μία καμπύλη στον χώρο αυτόν (αν ενώσουμε όλα αυτά τα σημεία). Στη συνέχεια υιοθετείται η αναπαράσταση του βίντεο στον πολυδιάστατο χώρο των χαρακτηριστικών, διότι με αυτόν τον τρόπο εκφράζεται το περιεχόμενο του βίντεο με βάση τη σημασία του και τις νοηματικές μεταβολές του.

Τα σημεία του βίντεο δεν αποτελούν μία συνεχή γραμμή, αφού έχουμε διακριτές τιμές του διανύσματος χαρακτηριστικών, και πάντοτε υπεισέρχεται θόρυβος στις τιμές του ιστογράμματος, οπότε η καμπύλη που προκύπτει στον χώρο παρουσιάζει διακυμάνσεις. Επειδή η εφαρμογή μαθηματικών πράξεων θα πολλαπλασιάσει τον θόρυβο και θα αλλοιώσει το αποτέλεσμα, ομαλοποιούμε την καμπύλη προκειμένου να έχει λιγότερες διακυμάνσεις και περισσότερο συνεχή συμπεριφορά.

### 3.2 Εξαγωγή του Διανύσματος Χαρακτηριστικών

Η μέθοδος που υιοθετείται εδώ είναι η κατασκευή ενός διανύσματος χαρακτηριστικών (**feature vector**) με διάσταση ίση με το πλήθος των χαρακτηριστικών που εξάγονται από το βίντεο και συνιστώσες τα χαρακτηριστικά αυτά ([Dou97], [Dou98]). Ο σκοπός είναι η μετατροπή του τρόπου αναπαράστασης της εικόνας από το πεδίο των pixels της εικόνας στο πεδίο των χαρακτηριστικών, το οποίο είναι καταλληλότερο για την περιγραφή του περιεχομένου του βίντεο, αφού χρησιμοποιεί αναπαράσταση που προσεγγίζει την ανθρώπινη οπτική αντίληψη και γι' αυτόν τον λόγο επιτρέπει την εξαγωγή περισσότερο σημαντικής πληροφορίας. Τα χαρακτηριστικά εκείνα που δίνουν ένα διάνυσμα χαρακτηριστικών περισσότερο αντιπροσωπευτικό και περιεκτικό είναι το χρώμα, η κίνηση, η υφή, όπως και το πλήθος, το σχήμα και η θέση αντικειμένων μέσα στην εικόνα. Στην παραπάνω σχετική εργασία υλοποιούν τη μέθοδό τους απευθείας στο κωδικοποιημένο MPEG χρησιμοποιώντας DC και AC συντελεστές

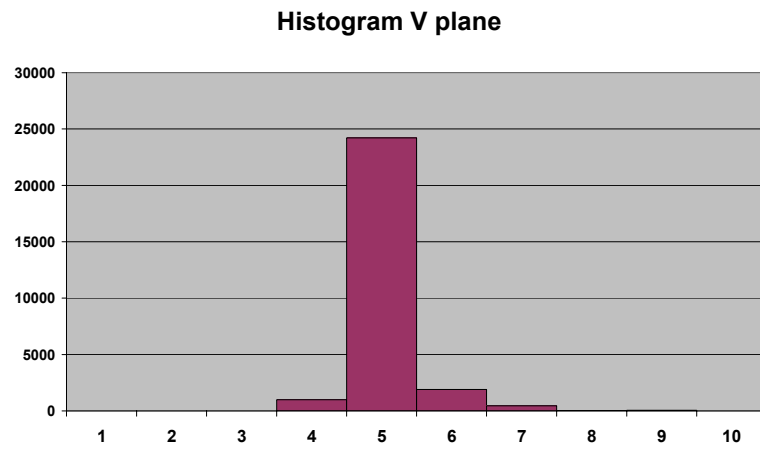
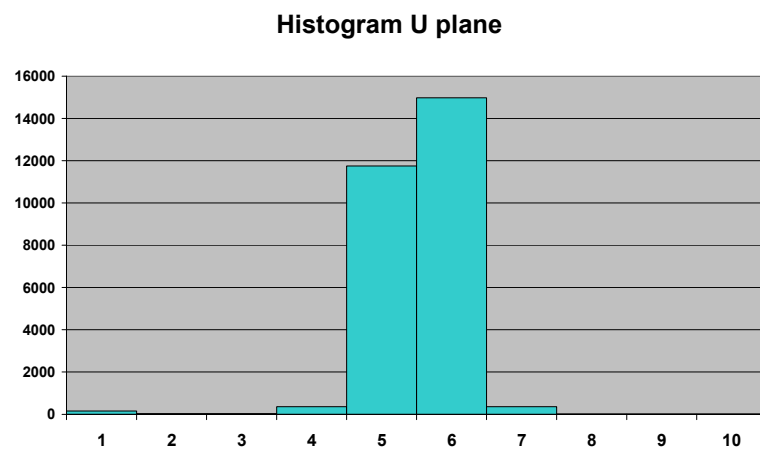
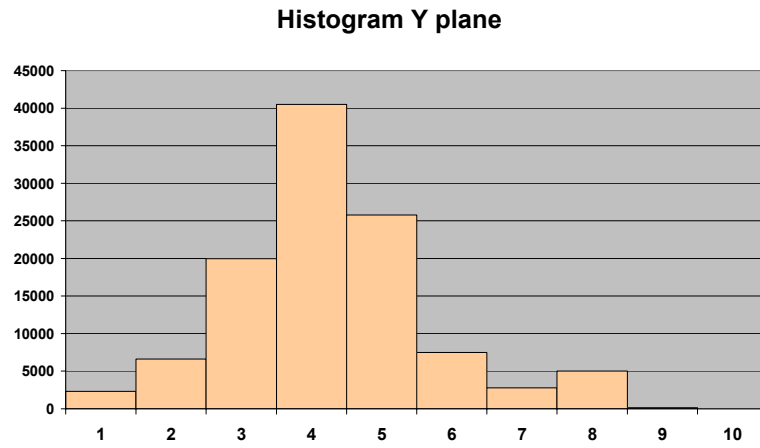
από το DCT πεδίο μετασχηματισμού, εκμεταλλευόμενοι πλεονεκτήματα του MPEG και εξοικονομώντας υπολογιστικό χρόνο. Επίσης χρησιμοποιούν κατάτμηση της εικόνας με βάση το χρώμα και την κίνηση (color and motion segmentation) για να εξάγουν χαρακτηριστικά.

Σε αυτήν εδώ την εργασία, αμέσως μετά την ανίχνευση των πλάνων, το βίντεο, που είναι κωδικοποιημένο σε μορφή MPEG 1 και αποτελείται από το πεδίο της φωτεινότητας (Y) και δύο χρωματικές συνιστώσες (U, V), αποκωδικοποιείται σε ακολουθία συνεχόμενων ασυμπίεστων frames με τρία ξεχωριστά πεδία, Y, U και V (και μάλιστα το Y να έχει τις διπλάσιες διαστάσεις από τα U και V). Η εξαγωγή των χαρακτηριστικών λαμβάνει χώρα στα ασυμπίεστα frames, διότι έτσι έχουμε μεγαλύτερη λεπτομέρεια (έχουμε μεγαλύτερη ανάλυση από το αν παίρναμε μόνο τους DC συντελεστές του DCT) και επίσης μπορούμε να εφαρμόσουμε τεχνικές που έχουν ήδη δοκιμαστεί σε ασυμπίεστη ακολουθία εικόνων, όπως είναι το ιστόγραμμα, η ομαλοποίηση και η εστίαση της οπτικής προσοχής (visual attention).

Ως βασικά χαρακτηριστικά χρησιμοποιούμε τη συνιστώσα της φωτεινότητας και τις δύο συνιστώσες χρώματος, κατασκευάζοντας το χρωματικό ιστόγραμμα (ουσιαστικά τρία ιστογράμματα ) με δέκα στοιχεία (bins) στο καθένα. Το χρωματικό ιστόγραμμα του πρώτου καρέ από το βίντεο FORMULA1 (σχ. 3.1) δίνεται ως παράδειγμα στα τρία διαγράμματα του σχήματος 3.2 (Y, U, V), ενώ οι τιμές των συνιστωσών των χαρακτηριστικών αυτών (δέκα για το καθένα) φαίνονται στον πίνακα 3.1.



**Σχ. 3.1 :** Εικόνα παραδείγματος για κατασκευή του χρωματικού ιστογράμματος και εξαγωγή του διανύσματος χαρακτηριστικών



**Σχ. 3.2 :** Το χρωματικό ιστόγραμμα της εικόνας με 10 bins για κάθε χαρακτηριστικό (για τα πεδία φωτεινότητας Y, και χρώματος U και V αντίστοιχα)

**Πίνακας 3.1 :** Τιμές των συνιστωσών του διανύσματος χαρακτηριστικών

1	2	3	4	5	6	7	8	9	10
2298	6605	19972	40510	25803	7505	2763	5011	125	0

11	12	13	14	15	16	17	18	19	20
160	25	22	359	11745	14978	359	0	0	0

21	22	23	24	25	26	27	28	29	30
0	0	0	996	24226	1903	445	26	52	0

Το ιστόγραμμα προκύπτει μετρώντας τα pixels που ανήκουν σε κάθε στοιχείο του, χωρίζοντας το διάστημα από 0 – 255 σε δέκα ομάδες των 25 τιμών. Αν αθροίσουμε τις τιμές όλων των bins για ένα συγκεκριμένο χαρακτηριστικό (Y ή U ή V) θα προκύψει ο συνολικός αριθμός των pixels της εικόνας. Εδώ επισημαίνεται ότι το ιστόγραμμα με μικρότερο αριθμό στοιχείων δείχνει την κατανομή φωτεινότητας και χρωμάτων σε πιο “χοντρές” κλάσεις, εκφράζοντας μία συνοπτική αναπαράσταση της εικόνας.

Βάζοντας στη σειρά τις τιμές των τριών ιστογραμμάτων Y, U και V, προκύπτει το διάνυσμα των χαρακτηριστικών (**feature vector**) που ορίζεται στον ευκλείδειο χώρο των τριάντα διαστάσεων (**30-D Euclidean Space**) ο οποίος αποτελεί και τον πολυδιάστατο χώρο των χαρακτηριστικών, μέσα στον οποίο κάθε καρέ αναπαρίσταται με ένα σημείο, το οποίο και το χαρακτηρίζει από τη σκοπιά του περιεχομένου του. Αν ενώσουμε όλα αυτά τα σημεία που προκύπτουν από όλα τα καρέ της ακολουθίας με συνεχή γραμμή, προκύπτει μια καμπύλη C ορισμένη στον πολυδιάστατο χώρο (30-D), η οποία αντιπροσωπεύει την νοηματική (με βάση τα συγκεκριμένα χαρακτηριστικά) εξέλιξη του βίντεο. Πάνω σε αυτή την καμπύλη, οι απότομες ή μεγάλες μεταβολές στο περιεχόμενο εκφράζονται με σημεία που ανήκουν σε καμπυλωτό τμήμα της καμπύλης, ενώ οι αργές ή ανύπαρκτες μεταβολές εκφράζονται με σχεδόν ευθύγραμμα τμήματα. Οπότε τα χαρακτηριστικά καρέ που επιθυμούμε να εξάγουμε θα αντιπροσωπεύουν το περιεχόμενο του βίντεο, καλύπτοντας τις απότομες αλλαγές αλλά και τα σταθερά πλάνα, και

εξετάζοντας την χρονική μεταβολή του διανύσματος χαρακτηριστικών, θα αντιστοιχούν σε σημεία ακραίας καμπυλότητας, δηλαδή μέγιστης και ελάχιστης αντίστοιχα.

### 3.3 Ομαλοποίηση του Διανύσματος Χαρακτηριστικών

#### 3.3.1 Φίλτρα Ομαλοποίησης μιας Καμπύλης

Η καμπύλη που προκύπτει από τη χρονική εξέλιξη (ακολουθία των frames) του διανύσματος χαρακτηριστικών, λόγω του ότι οι τιμές της δεν αποτελούν συνεχή συνάρτηση με το χρόνο, αλλά είναι διακριτά σημεία που αντιστοιχούν σε τιμές χαρακτηριστικών από κάθε καρέ, παρουσιάζει διακυμάνσεις οι οποίες οφείλονται σε θόρυβο. Αν επιχειρήσουμε να εφαρμόσουμε μαθηματικές μεθόδους σε αυτήν την καμπύλη, λόγω των πράξεων που θα γίνουν πολλές φορές (παραγωγίσεις, διαφορίσεις, αθροίσματα και πολλαπλασιασμοί) η επίδραση του θορύβου θα πολλαπλασιαστεί και θα παραμορφώσει τα αποτελέσματα, δίνοντας πολλά πιθανά ακραία σημεία, χωρίς να ξεχωρίζουν τα κύρια ακραία σημεία όπου παρατηρείται μεγάλη καμπυλότητα.

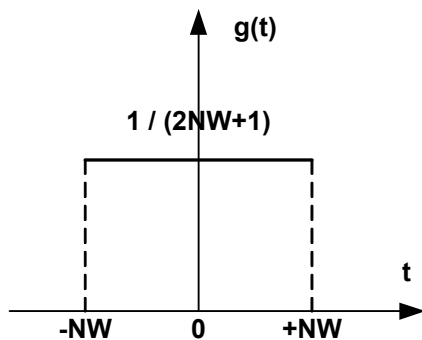
Γι' αυτόν τον λόγο κρίνεται απαραίτητη η εφαρμογή ενός είδους ομαλοποίησης (smoothing) της καμπύλης, μειώνοντας τις τοπικές μικρές αιχμές, αλλά διατηρώντας τις μεγάλες μεταβολές (ακραία σημεία) και τα καμπύλα τμήματα ανεπηρέαστα στη μορφή τους. Αν η  $f(t)$  αποτελεί μία συνάρτηση ως προς το χρόνο (υποθέτουμε ότι η  $f$  ορίζεται σε διακριτές τιμές του χρόνου, αφού η διακριτή περίπτωση μας ενδιαφέρει), η ομαλοποίηση (smoothing) είναι η εφαρμογή μιας συνέλιξης ενός πυρήνα (kernel) με την συνάρτηση  $f(t)$  με διακριτό παράθυρο  $NW$ , δηλαδή σε κάθε σημείο για την ομαλοποίηση χρησιμοποιούνται  $NW$  προηγούμενα σημεία και  $NW$  επόμενα σημεία, δηλαδή συνολικά ο πυρήνας της συνέλιξης έχει μέγεθος  $(2*NW + 1)$  σημείων ([Mok02]).

Από εδώ και στο εξής το  $NW$  θα αναφέρεται ως παράθυρο συνέλιξης ή ομαλοποίησης, ενώ το  $(2*NW + 1)$  ως εύρος συνέλιξης. Προσοχή πρέπει να δοθεί στο γεγονός ότι για τον υπολογισμό του ομαλοποιημένου  $f(t)$  χρησιμοποιούνται στο παράθυρο οι αρχικές τιμές της συνάρτησης και όχι οι ομαλοποιημένες που έχουν μέχρι το σημείο αυτό υπολογιστεί. Ανάλογα με την συνάρτηση που θα χρησιμοποιηθεί ως

φίλτρο (δηλαδή ο πυρήνας της συνέλιξης), έχουμε διάφορες μεθόδους για την εφαρμογή της ομαλοποίησης, τέσσερις από τις οποίες περιγράφονται στη συνέχεια.

### 3.3.1.1 Ορθογώνια Συνάρτηση (Μέση Τιμή)

Η πιο απλή περίπτωση ομαλοποίησης είναι η εφαρμογή της μέσης τιμής (mean value), δηλαδή υπολογίζεται η μέση τιμή από τα σημεία της  $f(t)$  στο εύρος της συνέλιξης και στη συνέχεια αντικαθίσταται η κεντρική τιμή με την μέση τιμή. Αυτό συνεπάγεται τη χρήση μιας ορθογώνιας συνάρτησης (σταθερή με μέτρο ίσο με το αντίστροφο του εύρους) ως πυρήνα της συνέλιξης, όπως φαίνεται στο σχήμα 3.3, αλλά και στον τύπο της ομαλοποίησης:



$f(t)$  = συνάρτηση προς ομαλοποίηση

$$g(t) = \frac{1}{2 \cdot NW + 1}, \quad \forall t$$

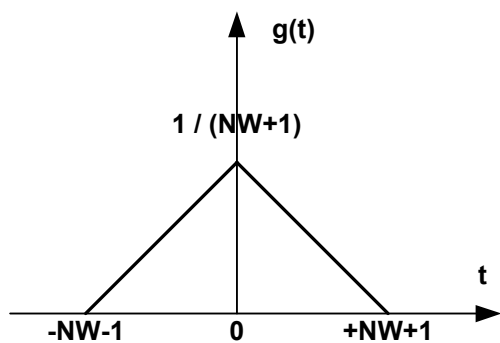
Ομαλοποιημένη συνάρτηση:

$$\begin{aligned} \text{smooth}(t) &= \\ &= \sum_{c=t-NW}^{t+NW} f(c) \cdot g(c-t) = \sum_{c=t-NW}^{t+NW} \frac{f(c)}{2 \cdot NW + 1} \end{aligned}$$

Σχ. 3.3 : Συνάρτηση Ομαλοποίησης Μέσης Τιμής

### 3.3.1.2 Τριγωνική Συνάρτηση

Μία άλλη απλή περίπτωση ομαλοποίησης είναι η εφαρμογή ενός τριγωνικού φίλτρου (triangular function), δηλαδή υπολογίζεται το γινόμενο κάθε σημείου της  $f(t)$  στο εύρος της συνέλιξης επί την τιμή του πυρήνα και στη συνέχεια αντικαθίσταται η κεντρική τιμή με το άθροισμα των γινομένων αυτών. Το ύψος του τριγώνου λαμβάνεται τέτοιο ώστε για σταθερή συνάρτηση να προκύπτει για το σημείο ίδια τιμή και μετά την ομαλοποίηση. Η τριγωνική συνάρτηση προσεγγίζει ουσιαστικά την γκαουσιανή και φαίνεται στο σχήμα 3.4 και στον διπλανό τύπο.



$f(t)$  = συνάρτηση προς ομαλοποίηση

$$g(t) = \begin{cases} \frac{t + NW + 1}{(NW + 1)^2}, & -NW - 1 \leq t \leq 0 \\ \frac{-t + NW + 1}{(NW + 1)^2}, & 0 \leq t \leq NW + 1 \end{cases}$$

Ομαλοποιημένη συνάρτηση:

$$\text{smooth}(t) = \sum_{c=t-NW}^{t+NW} f(c) \cdot g(c-t)$$

Σχ. 3.4 : Τριγωνική Συνάρτηση Ομαλοποίησης

### 3.3.1.3 Φίλτρο Ενδιάμεσης Τιμής

Στην περίπτωση αυτή η ομαλοποίηση της συνάρτησης  $f(t)$  γίνεται με τη χρήση ενός φίλτρου ενδιάμεσης τιμής (median value), το οποίο λειτουργεί ως εξής: για την ομαλοποίηση στο σημείο  $t$  οι τιμές της συνάρτησης  $f$

$$f(c), \quad c \in [t - NW, t + NW],$$

που είναι  $2 \cdot NW + 1$  το πλήθος, ταξινομούνται από τη μικρότερη στη μεγαλύτερη και μετά επιλέγεται η ενδιάμεση τιμή (κεντρική), δηλαδή η τιμή που βρίσκεται στη θέση  $NW + 1$ . Αυτή θεωρείται και ως ομαλοποιημένη τιμή της συνάρτησης  $f$  στο σημείο  $t$ . Πρέπει να σημειωθεί ότι το φίλτρο ενδιάμεσης τιμής δεν είναι κατάλληλο για τον σκοπό που το χρειαζόμαστε, αφού σε πολλές περιπτώσεις ακμών αντικαθιστώντας με την ενδιάμεση τιμή οι ακμές της καμπύλης χάνονται και προφανώς παραμορφώνεται το δείγμα από το οποίο θα καθοριστεί το περιεχόμενο του βίντεο.

### 3.3.1.4 Γκαουσιανή Συνάρτηση

Η πιο διαδεδομένη συνάρτηση που χρησιμοποιείται για τον σκοπό της ομαλοποίησης είναι η γκαουσιανή ([Mok02]), δηλαδή η συνάρτηση του Gauss που χρησιμοποιείται για την ανάλυση των σφαλμάτων σε πειράματα (κανονική ή γκαουσιανή κατανομή). Η χρήση της γκαουσιανής συνάρτησης (Gaussian function) έχει αποδειχθεί ότι συγκριτικά με τις υπόλοιπες έχει τη μεγαλύτερη ικανότητα να μειώνει την επίδραση

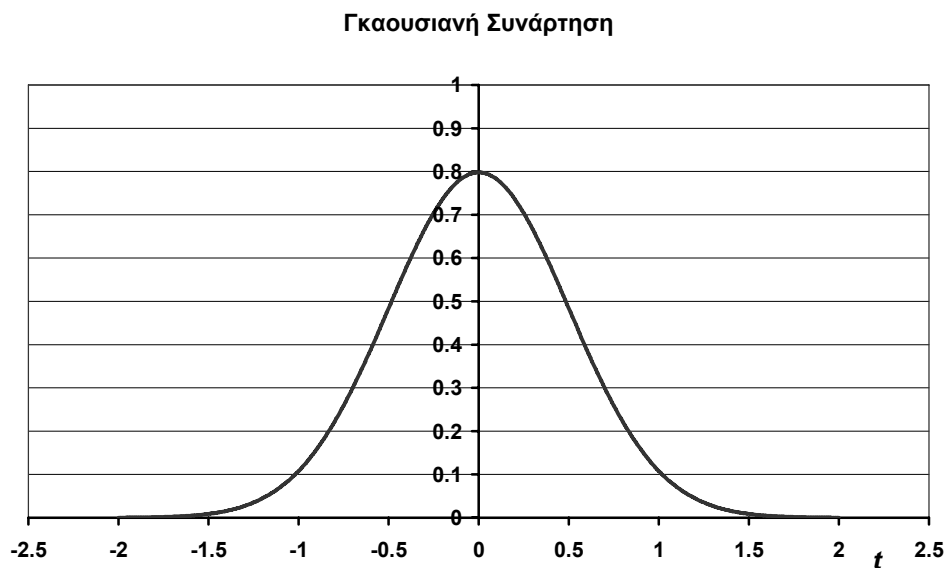


του θορύβου. Στο σχήμα 3.5 δίνεται η μορφή της (για μέση τιμή μηδέν  $\mu = 0$  και τυπική απόκλιση  $\sigma = 0.5$ ) και ο τύπος της συνάρτησης – πυρήνας της συνέλιξης:

$f(t)$  = συνάρτηση προς ομαλοποίηση

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{(t-\mu)^2}{2 \cdot \sigma^2}\right\}$$

Ομαλοποιημένη συνάρτηση:  $\text{smooth}(t) = \sum_{c=t-NW}^{t+NW} f(c) \cdot g(c-t)$



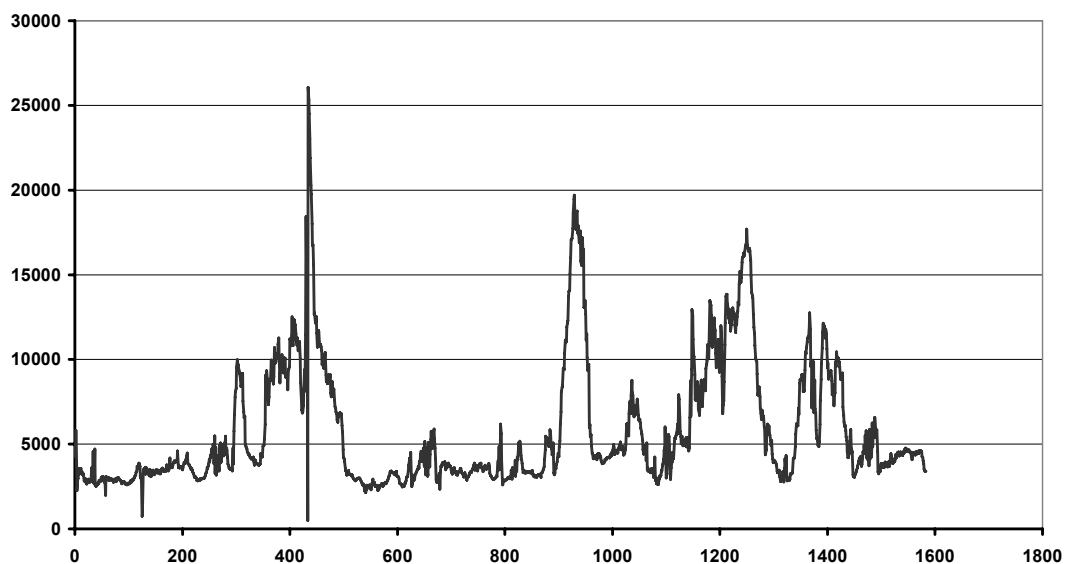
**Σχ. 3.5 :** Γκαουσιανή Συνάρτηση Ομαλοποίησης ( $\mu = 0$ ,  $\sigma = 0.5$ )

Οι Mokhtarian et al. ([Mok02]) προτείνουν την χρήση ενός μοντέλου πολλαπλής κλίμακας (multiscale approach) κατά την εφαρμογή της γκαουσιανής συνάρτησης, χρησιμοποιώντας διαφορετικές τιμές για το παράθυρο της συνέλιξης  $NW$ , υπολογίζοντας περισσότερες ή λιγότερες γειτονικές τιμές για την ομαλοποίηση του  $f(t)$ . Έτσι επιτυγχάνουν διάφορα επίπεδα ομαλοποίησης (multi scale) για την καμπύλη, με περισσότερες ή λιγότερες αιχμές ώστε να τροφοδοτήσουν τη μέθοδό τους με το αποτέλεσμα. Η κατάλληλη κλίμακα (scale) που θα χρησιμοποιηθεί εξαρτάται από τις απαιτήσεις, αλλά μπορεί να οριστεί μία κλίμακα (scale) που να επιτυγχάνει ικανοποιητική ομαλοποίηση για τις περισσότερες καμπύλες, ώστε να εξαλείφει το μεγαλύτερο μέρος του θορύβου.

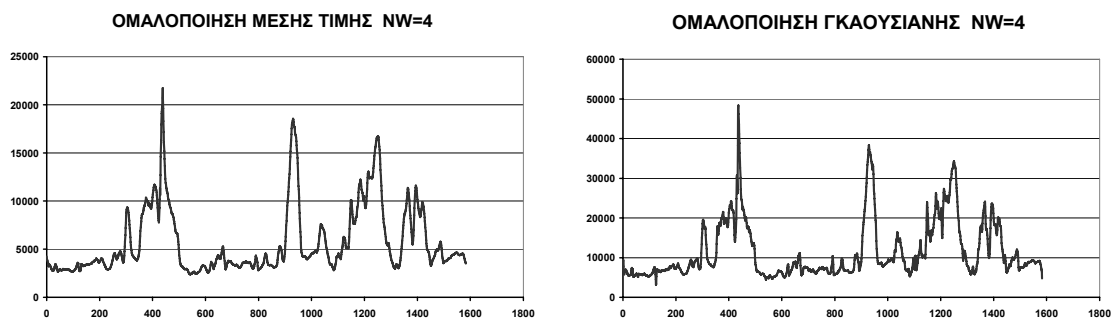
### 3.3.2 Σύγκριση Ομαλοποίησης Μέσης Τιμής και Γκαουσιανής

Για να ομαλοποιηθεί η πολυδιάστατη καμπύλη στις τριάντα διαστάσεις, η ομαλοποίηση εφαρμόζεται σε κάθε συνιστώσα του διανύσματος χαρακτηριστικών ξεχωριστά. Στα σχήματα 3.6 – 3.10 παρουσιάζονται τα αποτελέσματα της ομαλοποίησης, εφαρμόζοντας τη συνάρτηση μέσης τιμής και την γκαουσιανή συνάρτηση στην καμπύλη του διανύσματος χαρακτηριστικών που προκύπτει από το βίντεο “FORMULA1”. Παρατίθενται τα διαγράμματα με την καμπύλη της όγδοης συνιστώσας του διανύσματος χαρακτηριστικών (8<sup>η</sup> συνιστώσα του Y), η αρχική καμπύλη χωρίς ομαλοποίηση και κατόπιν οι καμπύλες μετά την ομαλοποίηση με καθεμιά από τις δύο μεθόδους, με τιμές για το παράθυρο συνέλιξης: 4, 14, 24 και 34.

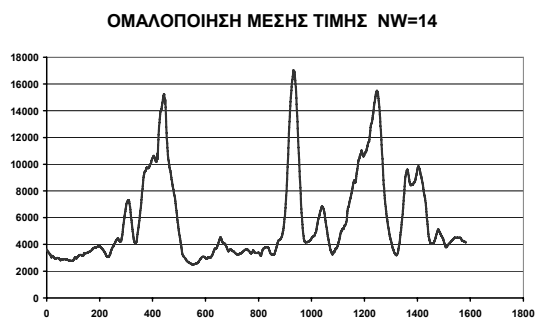
Αρχική Καμπύλη χωρίς Ομαλοποίηση



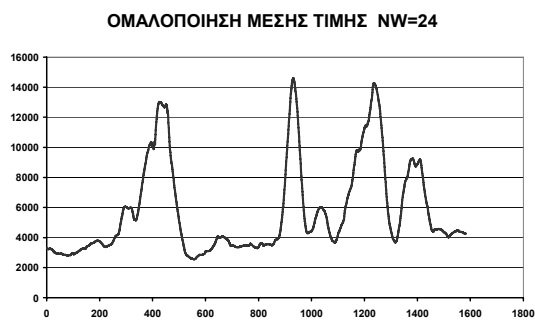
Σχ. 3.6 : Καμπύλη του 8<sup>ου</sup> στοιχείου του feature vector χωρίς ομαλοποίηση



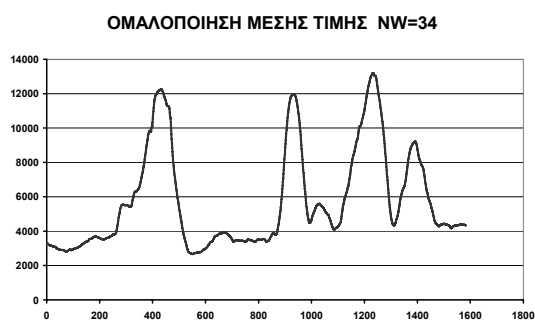
Σχ. 3.7 : Ομαλοποίηση με παράθυρο NW = 4 (μέση τιμή και γκαουσιανή)



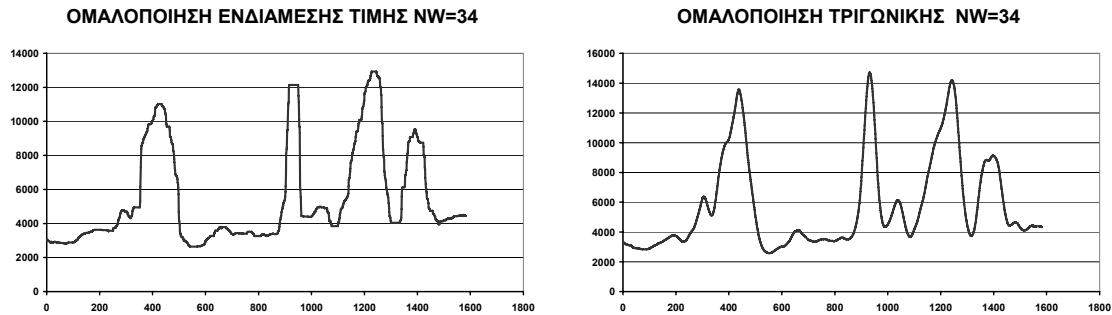
**Σχ. 3.8 :** Ομαλοποίηση με παράθυρο NW = 14 (μέση τιμή και γκαουσιανή)



**Σχ. 3.9 :** Ομαλοποίηση με παράθυρο NW = 24 (μέση τιμή και γκαουσιανή)



**Σχ. 3.10 :** Ομαλοποίηση με παράθυρο NW = 34 (μέση τιμή και γκαουσιανή)



**Σχ. 3.11 :** Ομαλοποίηση με παράθυρο  $NW = 34$  (ενδιάμεση τιμή και τριγωνική)

Στο σχήμα 3.11 παρουσιάζονται η ομαλοποίηση με τις μεθόδους της ενδιάμεσης τιμής και της τριγωνικής συνάρτησης για παράθυρο  $NW = 34$ , όπου φαίνεται ότι η ενδιάμεση τιμή εξομαλύνει μερικά σημεία της καμπύλης σε εντελώς ευθύγραμμα τμήματα, γι' αυτό και δεν χρησιμοποιείται, ενώ με την τριγωνική συνάρτηση η ομαλοποιημένη καμπύλη προσεγγίζει τη μορφή της αντίστοιχης γκαουσιανής, αλλά με εξομαλυνμένες τοπικές διακυμάνσεις.

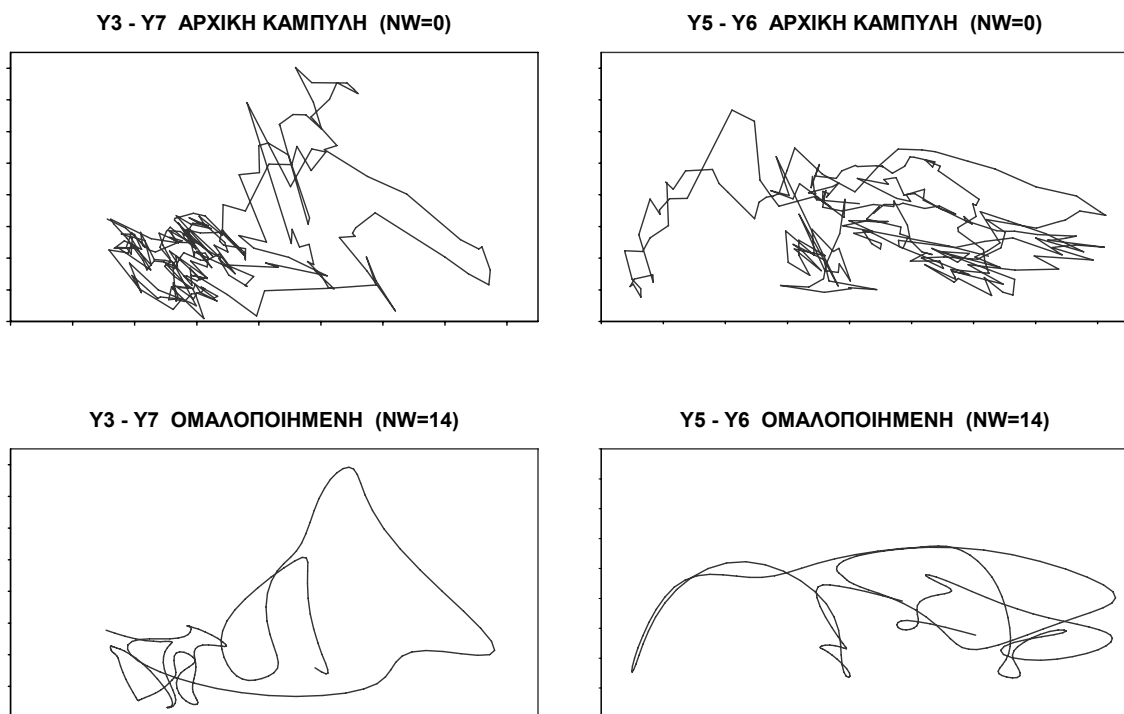
Συγκρίνοντας τα αποτελέσματα, είναι φανερό ότι η ομαλοποίηση έχει σαν αποτέλεσμα τη μεγαλύτερη εξομάλυνση για την μέση τιμή σε σχέση με την γκαουσιανή συνάρτηση για αντίστοιχες τιμές του  $NW$ . Αυτό οφείλεται στον θόρυβο, ο οποίος με τη μέση τιμή αθροίζεται και μοιράζεται (μέση τιμή) μεταξύ των τιμών του παραθύρου και άρα στα περισσότερα σημεία έχουμε μία “χοντρική” τιμή της συνάρτησης. Αυτό έχει ως αποτέλεσμα τελικά μετά την ομαλοποίηση να παραμένουν μόνο οι πολύ μεγάλες αιχμές (τέσσερις μεγάλες αιχμές φαίνονται στο σχ. 3.10, όπου χρησιμοποιείται μεγάλο παράθυρο  $NW$ ), ενώ σημεία με τοπικά μέγιστα (μικρές αιχμές), όπως για παράδειγμα στο frame 300 στο σχήμα 3.6, με τη μέση τιμή εξαφανίζονται σχεδόν τελείως.

Έτσι για την μέση τιμή όταν θέλουμε να επιτύχουμε “καλή” ομαλοποίηση, όπως φαίνεται από  $NW = 24$  και μετά, χάνουμε κάποιες μικρές αιχμές (peaks) και έτσι πολλά τμήματα της καμπύλης γίνονται επίπεδα παραμορφώνοντας το αρχικό σήμα, γεγονός που δεν συμβαίνει στην περίπτωση της γκαουσιανής. Επίσης, κρίνουμε καταλληλότερο παράθυρο συνέλιξης με την γκαουσιανή συνάρτηση το  $NW = 34$  ή μεγαλύτερο, που όπως φαίνεται από τα διαγράμματα δίνει μία ικανοποιητική καμπύλη που διατηρεί τις μεγάλες αιχμές, αλλά και τα τοπικά μέγιστα (λεπτομέρειες, δηλαδή σημεία μεγάλης και μικρής καμπύλωσης), συγκρίνοντάς το με την μέση τιμή με παράθυρο  $NW = 24$ .

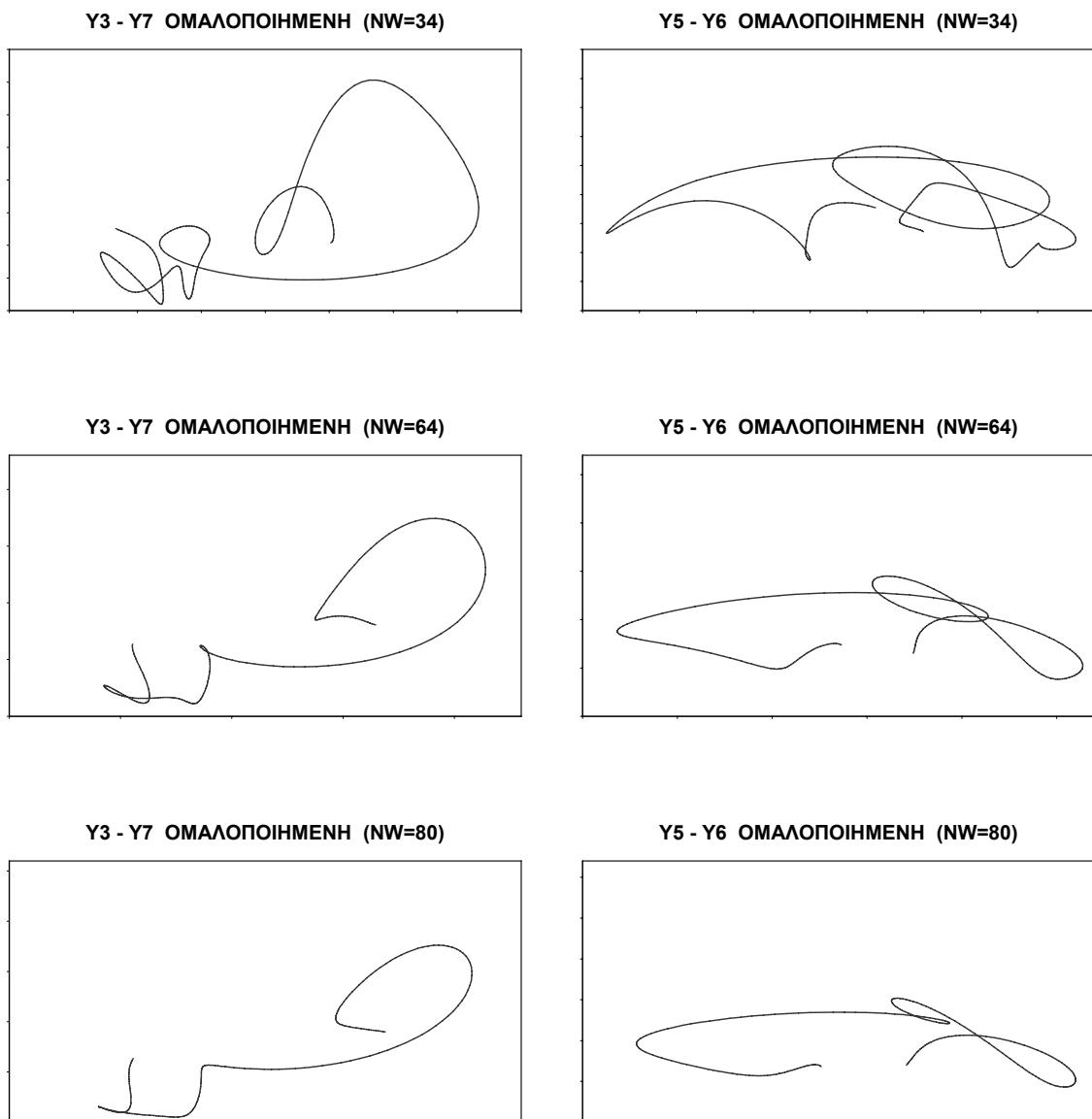
### 3.3.3 Παράδειγμα Ομαλοποίησης Διανύσματος Χαρακτηριστικών

Όπως αναλύθηκε, η καμπύλη που προέρχεται από τα χαρακτηριστικά των καρέ του βίντεο προκύπτει με συνεχείς διακυμάνσεις και η ομαλοποίησή της έγκειται στην εφαρμογή του φίλτρου ομαλοποίησης σε κάθε συνιστώσα του διανύσματος χαρακτηριστικών (σαν να ήταν μία ξεχωριστή συνάρτηση του χρόνου). Έτσι ο θόρυβος εξαλείφεται και η καμπύλη γίνεται πιο ομαλή (smoothed) με σημεία καμπύλωσης που διακρίνονται και χαρακτηριστικά τέτοια ώστε οι μέθοδοι χειρισμού της N-διάστατης καμπύλης να εξάγουν κατάλληλα αποτελέσματα.

Η ομαλοποίηση με γκαουσιανό φίλτρο (Gaussian smoothing) εφαρμόζεται ως παράδειγμα στην καμπύλη που προκύπτει από το βίντεο “RALLY1”. Για καλύτερη παρατήρηση, η καμπύλη προβάλλεται στα επίπεδα Y3 – Y7 και Y5 – Y6 (δύο ζεύγη από τις συνιστώσες της), και συγκεκριμένα για τα frames από 471 έως 720. Παραθέτουμε εδώ τις καμπύλες χωρίς και με ομαλοποίηση σε 5 κλίμακες (διαφορετικό παράθυρο εφαρμογής ομαλοποίησης για  $NW = 0, 14, 34, 64, 80$ ), ώστε να γίνει κατανοητή η λειτουργία και η χρησιμότητα της διαδικασίας (σχ. 3.12).



Σχ. 3.12α : Ομαλοποίηση Γκαουσιανής σε προβολές καμπύλης στις κλίμακες  $NW = 0, 14$



**Σχ. 3.12β :** Ομαλοποίηση Γκαουσιανής σε προβολές καμπύλης στις κλίμακες NW = 34, 64, 80

Όπως βλέπουμε οι πολύ μεγάλες αλλαγές στην καμπύλωση που υπάρχουν εξαρχής διατηρούνται σε όλες τις κλίμακες, ενώ οι μικρότερες καμπυλώσεις σιγά σιγά εξαλείφονται και τελικά εξαφανίζονται σε πολύ μεγάλο παράθυρο NW. Το πόσο μεγάλο θα είναι το παράθυρο εφαρμογής της συνέλιξης εξαρτάται από το πόσο γενική επιθυμούμε να είναι η περίληψη του πλάνου από τα χαρακτηριστικά καρέ που θα εξάγουμε. Φυσικά για μεγαλύτερη κλίμακα προκύπτουν ομαλότερες καμπύλες του διανύσματος χαρακτηριστικών και γενικά ομαλότερες γραφικές παραστάσεις με ξεκάθαρα ακραία σημεία, όπου όμως έχουν εξαλειφθεί οι μικρές μεταβολές. Επίσης, επειδή στα όρια (αρχή και τέλος) της ακολουθίας δεν μπορούμε να εφαρμόσουμε το

γκαουσιανό φίλτρο λόγω έλλειψης τιμών (αλλιώς θα παραμορφώσουμε την καμπύλη), το παράθυρο NW καθορίζει NW frames στην αρχή και NW frames στο τέλος του βίντεο τα οποία αναγκαστικά τα αγνοούμε.

Η λεπτομέρεια λοιπόν που θέλουμε να έχει η προσέγγιση που θα κάνουμε με τα χαρακτηριστικά καρέ και η ανοχή στα frames αρχής και τέλους που απορρίπτουμε θα καθορίσουν το παράθυρο που θα εφαρμόσουμε. Επιλέγουμε παράθυρο  $NW = 64$  για πλάνα με αριθμό frames 1280 ή μεγαλύτερο (δηλαδή με διάρκεια ίση ή μεγαλύτερη από 51.2 sec για ρυθμό πλαισίων 25 frames/sec), ενώ για τις υπόλοιπες ακολουθίες με αριθμό frames μικρότερο του 1280 επιλέγουμε  $NW = (\text{αριθμός frames}) / 20$ . Αυτό έχει αποτέλεσμα να “χάνουμε” περίπου 10% της διάρκειας του βίντεο που βρίσκεται στην αρχή και στο τέλος (μόνο για πλάνα με αριθμό frames μικρότερο του 1280), ποσοστό που μπορεί να θεωρηθεί μικρό ώστε να μην απορρίπτεται σημαντική πληροφορία, ενώ όπως φαίνεται από το σχ. 3.12, η ομαλοποίηση σε αυτήν την κλίμακα δίνει περίπου 6 σημεία μεγάλης καμπύλωσης που μαζί με τα ενδιάμεσα 5 σημεία μικρής καμπύλωσης (ευθείες) δίνουν συνολικά 11 σημεία, τα οποία συνοψίζουν το περιεχόμενο του βίντεο για τα frames από 470 έως 720, δηλαδή για 250 frames (10 sec) έχουμε 1 keyframe ανά 23 frames (ανά 1 sec περίπου).

Με τον παραπάνω τρόπο λαμβάνουμε τιμή για το παράθυρο ομαλοποίησης ανάλογη της διάρκειας του πλάνου (για πλάνα με αριθμό frames μικρότερο του 1280), οπότε έτσι επιτρέπουμε στην καμπύλη να διατηρήσει τις απότομες μεταβολές της ακόμα και σε μικρής διάρκειας πλάνα. Αυτό αρχικά φαίνεται παράδοξο, γιατί θα μπορούσε κάποιος να θεωρήσει ότι με αυτόν τον τρόπο μεγάλα και μικρά πλάνα (σε διάρκεια) θα ομαλοποιούνται ανάλογα με αποτέλεσμα ίδιο αριθμό χαρακτηριστικών καρέ και για τις δύο κατηγορίες. Όμως είναι σωστό από δύο απόψεις: πρώτον ο αριθμός των χαρακτηριστικών καρέ δεν πρέπει να εξαρτάται άμεσα από την διάρκεια του πλάνου, αλλά από τον βαθμό μεταβολής του περιεχομένου του και κατόπιν από την διάρκειά της, και δεύτερον διατηρούμε μεταβολές περισσότερο σε μικρότερης διάρκειας πλάνα, ώστε σε σύντομα πλάνα με απότομες μεταβολές να έχουμε περισσότερα χαρακτηριστικά καρέ απ’ ότι σε εκτενέστερα πλάνα με αργά μεταβαλλόμενο περιεχόμενο, αφού τότε το σύντομο πλάνο θα έχει ίσως και μεγαλύτερη σημασία από το εκτενέστερο.





## ΚΕΦΑΛΑΙΟ 4

# *Μέθοδοι Καμπυλότητας για την Εξαγωγή Χαρακτηριστικών Σημείων Καμπύλης*

### 4.1 Εισαγωγικά

#### 4.1.1 Εξαγωγή Χαρακτηριστικών Σημείων - Καρέ

Στην εργασία αυτή, όπως αναφέρθηκε, επιλέγεται η αναπαράσταση του βίντεο στον πολυδιάστατο χώρο των χαρακτηριστικών, όπου κάθε καρέ αντιστοιχεί σε ένα σημείο και ολόκληρη η ακολουθία αντιστοιχεί σε μία καμπύλη που εκφράζει τον τρόπο εξέλιξης του περιεχομένου.

Η καμπύλη του διανύσματος χαρακτηριστικών θα αποτελείται από μεγάλα σχεδόν ευθύγραμμα τμήματα, που θα αντιστοιχούν σε αργή κίνηση της κάμερας (σταθερά πλάνα) και γενικά σε σταθερή μεταβολή του περιεχομένου του βίντεο, και από καμπύλα τμήματα όπου η καμπύλη αλλάζει κατεύθυνση απότομα ή και ομαλά, τα οποία θα αντιστοιχούν σε καρέ του βίντεο όπου συμβαίνουν δραματικές αλλαγές στο περιεχόμενο ή απότομη κίνηση της κάμερας. Τα ευθύγραμμα τμήματα θα πρέπει να αντικατασταθούν από λίγα καρέ (δυνατότητα μεγάλης αφαίρεσης), ενώ τα καμπύλα τμήματα θα πρέπει να αντικατασταθούν από περισσότερα καρέ ώστε το περιεχόμενο

αυτών των καρέ να ανταποκρίνεται στο περιεχόμενο της καμπύλης και να το περιγράφει πλήρως.

Χρησιμοποιείται η έννοια της καμπυλότητας για την εξαγωγή των ακραίων σημείων της καμπύλης, που επιτρέπει στο σύστημα να λειτουργεί αυτόνομα ανεξάρτητα από το βίντεο που δέχεται στην είσοδό του. Από τη μαθηματική θεωρία της διαφορικής γεωμετρίας για τις καμπύλες στον  $N$ -διάστατο Ευκλείδειο χώρο, μέσω της παραμετρικής αναπαράστασης της καμπύλης, προκύπτουν διαφορετικοί τρόποι ορισμού της καμπυλότητας.

Αρχικά εξετάζεται η καμπυλότητα ως δεύτερη παράγωγος ως προς τον χρόνο ([Dou00]), η οποία εκφράζει την επιτάχυνση (ή την επιβράδυνση) του περιεχομένου μέσα στην ακολουθία, δηλαδή το πόσο γρήγορα μεταβάλλονται τα χαρακτηριστικά από καρέ σε καρέ. Αν η καμπυλότητα δίνει μεγάλο μέτρο (απόλυτη τιμή), αυτό σημαίνει ότι οι αλλαγές στο βίντεο γίνονται με επιταχυνόμενο ρυθμό, ενώ σε αντίθετη περίπτωση οι μεταβολές γίνονται με σχεδόν σταθερό ρυθμό, δηλαδή με σταθερή ταχύτητα.

Επιδιώκεται στη συνέχεια ο κατάλληλος ορισμός της καμπυλότητας με βάση τη μαθηματική έννοια, που συγκεκριμένα αναφέρεται στην παραγωγή σε σχέση με την μετατόπιση  $s$  πάνω στην καμπύλη. Αυτό έχει ως αποτέλεσμα την εξέταση της μεταβολής του νοηματικού περιεχομένου ανεξάρτητα από την ταχύτητα με την οποία συμβαίνει. Στην τελική φάση αναζητείται η μέθοδος ώστε το μέτρο της καμπυλότητας να εκφράζει τον τρόπο της μεταβολής της καμπύλης στον χώρο των χαρακτηριστικών, μέσω της θεώρησης των πολλών διαστάσεων. Τα τοπικά ακρότατα της καμπυλότητας θα επιλέγονται ως χαρακτηριστικά σημεία της καμπύλης (άρα και ως χαρακτηριστικά καρέ του βίντεο), όπου τα μέγιστα θα αντιστοιχούν στην μέγιστη επιταχυνόμενη μεταβολή, ενώ τα ελάχιστα σε σταθερή μεταβολή.

#### *4.1.2 Έννοιες στον $N$ -διάστατο Ευκλείδειο Χώρο*

Η καμπύλη του διανύσματος χαρακτηριστικών, αφού ομαλοποιηθεί με τον τρόπο που περιγράφηκε παραπάνω, πρέπει να αναλυθεί με μαθηματικές μεθόδους ούτως ώστε να εξαχθούν τα σημεία όπου παρατηρούνται απότομες ή και ακραίες αλλαγές στην κατεύθυνσή της. Ο χώρος των χαρακτηριστικών, που έχει δέκα διαστάσεις για κάθε χαρακτηριστικό (φωτεινότητα και δύο χρωματικές συνιστώσες), είναι τριάντα

διαστάσεων και θεωρείται 30-διάστατος Ευκλείδειος χώρος, όπου ισχύουν οι γνωστές ιδιότητες κάθε Ευκλείδειου χώρου, και κατά πρώτον το μέτρο ή η ευκλείδεια νόρμα.

Αν  $x, y$  είναι δύο σημεία του  $N$ -διάστατου Ευκλείδειου χώρου, τότε ορίζεται το εσωτερικό γινόμενο μεταξύ των δύο διανυσμάτων ως εξής:

$$x \cdot y = \sum_{i=1}^N x_i \cdot y_i = x_1 \cdot y_1 + x_2 \cdot y_2 + \dots + x_N \cdot y_N \quad (\text{εξ. 4.1})$$

Μέσω του εσωτερικού γινομένου ορίζεται το μέτρο του διανύσματος  $x$  (ευκλείδεια νόρμα) ως εξής:

$$\|x\| = \sqrt{x \cdot x} = \sqrt{\sum_{i=1}^N (x_i)^2} \quad (\text{εξ. 4.2})$$

Τέλος, η απόσταση μεταξύ των σημείων  $x$  και  $y$  ορίζεται μέσω της προηγούμενης Ευκλείδειας νόρμας:

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (\text{εξ. 4.3})$$

Ακόμη, μπορούμε να ορίσουμε και την (εσωτερική) γωνία  $\theta$  μεταξύ των διανυσμάτων  $x$  και  $y$  μέσω του εσωτερικού γινομένου και της νόρμας:

$$\theta = \cos^{-1} \left( \frac{x \cdot y}{\|x\| \cdot \|y\|} \right) \quad (\text{εξ. 4.4})$$

## 4.2 Η Καμπυλότητα μιας Καμπύλης

Μία καμπύλη γραμμή θεωρούμε ότι καμπυλώνει (curving) όταν τείνει να “ξεφύγει” από την ευθεία γραμμή. Το πόσο πολύ τείνει να αποκλίνει από την ευθεία γραμμή, δηλαδή με τι ρυθμό θα απομακρύνεται, εκφράζεται με την καμπυλότητα (curvature) που αποτελεί τον βαθμό της καμπύλωσης. Η τροχιά των διανυσμάτων των χαρακτηριστικών όλων των καρέ ενός πλάνου δίνει την χρονική εξέλιξη του πλάνου και γι’ αυτό επιδιώκουμε να επιλέξουμε τα κατάλληλα σημεία πάνω στην καμπύλη αυτή, που

θα χαρακτηρίζουν την καμπύλη και θα προσφέρουν αρκετή πληροφορία ώστε το σχήμα και η μορφή της καμπύλης να μπορούν να αναπαραχθούν (κατά προσέγγιση) γνωρίζοντας μόνο αυτά τα σημεία και χρησιμοποιώντας παρεμβολή ([Αντ99]). Ως τέτοια σημεία λαμβάνονται τα ακραία σημεία (extreme points) μέγιστης και ελάχιστης (τοπικής) καμπυλότητας.

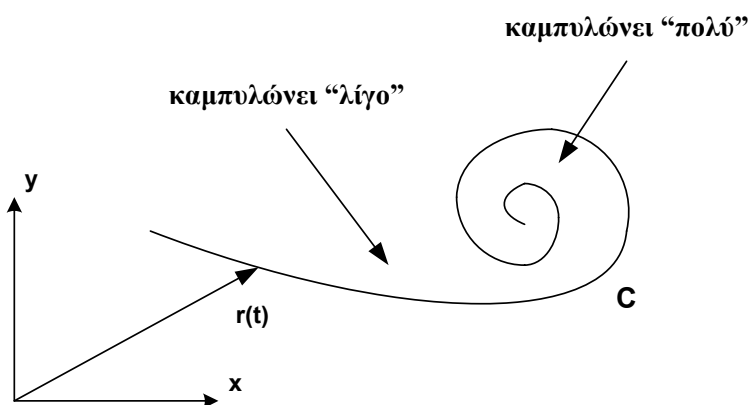
Έστω ότι μία συνεχής καμπύλη  $C$  στον διδιάστατο χώρο αναπαρίσταται με την εξίσωση ( $t$  μία τυχαία παράμετρος):

$$\vec{r}(t) = (x(t), y(t)), \quad t \in I, \quad I \subseteq \mathbb{R} \quad (\text{εξ. 4.5})$$

Αν η τυχαία παράμετρος  $t$  εκφράζει τον χρόνο, η πρώτη παράγωγος του διανύσματος  $\mathbf{r}(t)$  εκφράζει τον ρυθμό μεταβολής του διανύσματος, δηλαδή την ταχύτητα με την οποία αλλάζει θέση. Η δεύτερη παράγωγος του διανύσματος  $\mathbf{r}(t)$  εκφράζει τον ρυθμό με τον οποίο αλλάζει η πρώτη παράγωγος, δηλαδή εκφράζει την επιτάχυνση ή την επιβράδυνση του διανύσματος. Ως μέσο μέτρησης της καμπύλωσης θεωρείται το μέτρο  $|D(t)|$  της δεύτερης παραγώγου του διανύσματος σε συνάρτηση με τον χρόνο (παράμετρος  $t$ ), οπότε παρακολουθώντας τα τοπικά ακρότατα (μέγιστα και ελάχιστα) του μέτρου αυτού εξάγουμε τα χαρακτηριστικά καρέ που αντιστοιχούν σε αυτές τις χρονικές στιγμές ([Αντ99]). Η παράγωγος του διανύσματος ως προς το χρόνο είναι το διάνυσμα με συνιστώσες τις παραγώγους των συνιστωσών του αρχικού διανύσματος:

$$\vec{r}'(t) = (x'(t), y'(t)), \quad t \in I \quad (\text{εξ. 4.6})$$

$$\vec{r}''(t) = (x''(t), y''(t)), \quad t \in I \quad (\text{εξ. 4.7})$$



Σχ. 4.1 : Σημεία μικρής και σημεία μεγάλης καμπύλωσης

Για το διάνυσμα χαρακτηριστικών, που ορίζεται στις  $N$  διαστάσεις, η καμπυλότητα ως μέτρο της δεύτερης παραγώγου ως προς τον χρόνο δίνεται από την παρακάτω σχέση:

Αν το διάνυσμα χαρακτηριστικών είναι  $\vec{f}(t) = [x_1(t), x_2(t), \dots, x_N(t)]$ ,  $t \in I$ , τότε η καμπύλη έχει μέτρο καμπύλωσης:

#### ΚΑΜΠΥΛΟΤΗΤΑ - ΤΥΠΟΣ 1:

(εξ. 4.8)

$$curvature1 = \|\vec{f}''(t)\| = \left\| \frac{d^2 \vec{f}}{dt^2} \right\|$$

Συγκεκριμένα στην ακολουθία βίντεο ως στοιχειώδη μεταβολή του χρόνου λαμβάνουμε το χρονικό διάστημα μεταξύ δύο διαδοχικών καρέ, αφού μεταξύ των δύο αυτών εικόνων δεν έχουμε περισσότερη πληροφορία για να γνωρίζουμε τι μπορεί να συμβαίνει. Γι' αυτό στην εφαρμογή των τύπων για το διάνυσμα χαρακτηριστικών με τις παραγώγους ως προς το χρόνο λαμβάνουμε ως  $\mathbf{dt} = \mathbf{1}$  και υπολογίζουμε τις διακριτές παραγώγους στο διακριτό πεδίο του χρόνου. Η πρώτη παράγωγος προκύπτει ως διαφορά των τιμών διαδοχικών στιγμιοτύπων του διανύσματος χαρακτηριστικών, ενώ η δεύτερη παράγωγος προκύπτει ως διαφορά των διαδοχικών τιμών της πρώτης παραγώγου:

$$\text{feature vector : } \vec{f}(t) = [x_1(t), x_2(t), \dots, x_N(t)], \quad t \in [t_1, t_2, \dots, t_m]$$

για  $m$  συνεχόμενα καρέ

$$1^{\text{η}} \text{ παράγωγος : } \vec{d}_1(t) = \frac{d\vec{f}}{dt} = \vec{f}(t) - \vec{f}(t-1), \quad t \in [t_2, t_3, \dots, t_m]$$

$$2^{\text{η}} \text{ παράγωγος : } \vec{d}_2(t) = \frac{d^2 \vec{f}}{dt^2} = \vec{d}_1(t) - \vec{d}_1(t-1), \quad t \in [t_3, t_4, \dots, t_m]$$

[ Από εδώ και στο εξής ο συμβολισμός  $\mathbf{r}$  θα χρησιμοποιείται για καμπύλες των δύο και τριών διαστάσεων, ενώ ο συμβολισμός  $\mathbf{f}$  θα χρησιμοποιείται για το διάνυσμα χαρακτηριστικών που είναι διάνυσμα ορισμένο στον  $N$ -διάστατο Ευκλείδειο χώρο με  $N$  συνιστώσες. ]

Η ιδέα της παραγωγίσης ως προς τον χρόνο  $t$  βασίζεται στη φυσική έννοια της ταχύτητας και της επιτάχυνσης. Αν για παράδειγμα θεωρήσουμε μία σφαίρα που εκτελεί

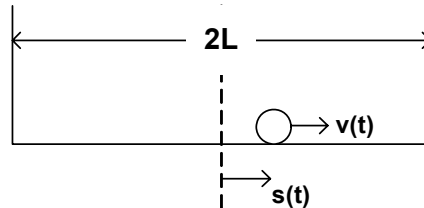
περιοδική κίνηση σε μία διάσταση, όπως δείχνει το σχήμα 4.2, μπορούμε να θεωρήσουμε ως χαρακτηριστικό της κίνησής της την μετατόπιση  $s(t)$  από το κέντρο του διαστήματος μήκους  $2L$ , η οποία λόγω περιοδικής κίνησης δίνεται από τον τύπο:

$$s(t) = L \cdot \sin(t)$$

με πρώτη και δεύτερη παράγωγο:

$$\dot{s}(t) = L \cdot \cos(t)$$

$$\ddot{s}(t) = -L \cdot \sin(t)$$

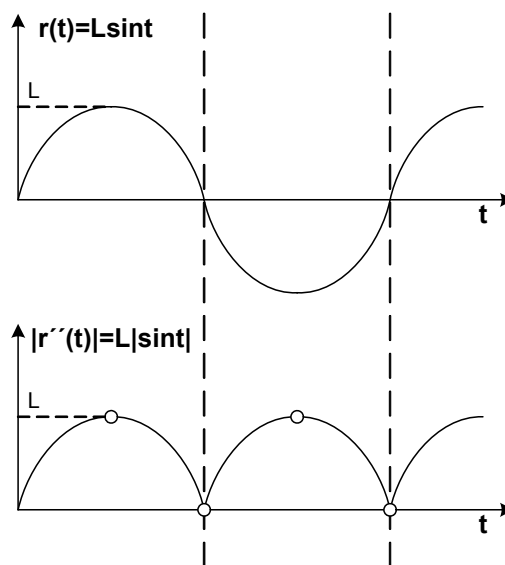


**Σχ. 4.2 :** Περιοδική κίνηση σφαίρας σε μία διάσταση

Στη συγκεκριμένη περίπτωση το διάνυσμα των χαρακτηριστικών έχει μία μόνο συνιστώσα ( $r(t) = s(t)$ ), την μετατόπιση  $s$ , οπότε η καμπυλότητα ως μέτρο της δεύτερης παραγώγου ως προς  $t$  είναι:

$$curvature1 = \|\ddot{r}(t)\| = \|\ddot{s}(t)\| = |-L \cdot \sin t| = L \cdot |\sin t|$$

Στις γραφικές παραστάσεις του σχήματος 4.3 παρουσιάζεται το διάνυσμα χαρακτηριστικών και το μέτρο της δεύτερης παραγώγου ( $curvature1$ ), όπου παρατηρούμε τα σημεία στα οποία παρουσιάζει τοπικό μέγιστο και τοπικό ελάχιστο.



**Σχ. 4.3 :** Η μετατόπιση  $s(t)$  και το μέτρο της δεύτερης παραγώγου ως προς  $t$

Τα σημεία τοπικού μεγίστου είναι τα σημεία στα οποία το διάνυσμα των χαρακτηριστικών μεταβάλλεται με την μεγαλύτερη επιτάχυνση (ή επιβράδυνση), δηλαδή τα σημεία στα οποία η σφαίρα εφάπτεται στα τοιχώματα, όπου η μετατόπισή της αλλάζει κατεύθυνση (η πιο απότομη μεταβολή της ταχύτητας, από θετικές σε αρνητικές τιμές).

Τα σημεία τοπικού ελαχίστου είναι τα σημεία στα οποία το διάνυσμα των χαρακτηριστικών έχει ταχύτητα τέτοια που μεταβάλλεται με τον μικρότερο δυνατό ρυθμό (ή ταχύτητα σταθερή), δηλαδή τα σημεία στα οποία η σφαίρα περνάει από το κέντρο του διαστήματος και στα οποία η επιτάχυνσή της μηδενίζεται (στιγμιαία αποκτά σταθερή ταχύτητα).

Εξάγουμε τα σημεία που αντιστοιχούν στα τοπικά ακρότατα της καμπυλότητας και τα οποία μας δείχνουν στιγμές ακραίας επιτάχυνσης (πολύ μεγάλης και πολύ μικρής) των συνιστωσών του διανύσματος των χαρακτηριστικών ως προς τον χρόνο.

### 4.3 Περιγραφή Καμπύλης με Φυσική Παράμετρο

Σχετικά με την παραμετροποίηση της καμπύλης, η παράμετρος  $t$  που χρησιμοποιείται παραπάνω είναι ο χρόνος, συναρτήσει του οποίου μεταβάλλονται οι συνιστώσες του διανύσματος. Το  $t$  αποτελεί μία τυχαία παράμετρο με την οποία εκφράζεται η παραμετρική εξίσωση ορισμού της καμπύλης. Θα μπορούσε να χρησιμοποιηθεί οποιαδήποτε άλλη παράμετρος αντί της  $t$  για την παραμετροποίηση της καμπύλης, συναρτήσει της οποίας θα μεταβάλλονταν οι συνιστώσες του διανύσματος χαρακτηριστικών. Σύμφωνα με τη θεωρία της διαφορικής γεωμετρίας ([Car76], [Χαιν98], [Spi99]) η πιο κατάλληλη αναπαράσταση μιας καμπύλης  $C$  στον Ευκλείδειο χώρο είναι η φυσική αναπαράσταση της καμπύλης, δηλαδή ο ορισμός της εξίσωσης της καμπύλης με παράμετρο το μήκος του τόξου  $s$ . Το  $s$  είναι το μήκος του τόξου του τμήματος της καμπύλης που διανύεται για το διάστημα που η τυχαία παράμετρος μεταβάλλεται από  $0$  σε  $t$ . Το μήκος τόξου  $s$  είναι μία συνάρτηση του  $t$ ,  $s = s(t)$  και ονομάζεται **φυσική παράμετρος της καμπύλης**, ενώ η τιμή της δίνεται από τον παρακάτω τύπο (για διδιάστατη καμπύλη):

$$s = s(t) = \int_0^s ds = \int_0^t \left| \frac{d\vec{r}(t)}{dt} \right| dt = \int_0^t \sqrt{\left( \frac{dx}{dt} \right)^2 + \left( \frac{dy}{dt} \right)^2} dt \quad (\text{εξ. 4.9})$$

όπου το στοιχειώδες  $ds$  δίνεται από το μέτρο της πρώτης παραγώγου του διανύσματος  $\mathbf{r}(t)$  ως προς το χρόνο:

$$ds(t) = \left| \frac{d\vec{r}(t)}{dt} \right| dt \quad (\text{εξ. 4.10})$$

Όταν ως παράμετρος ληφθεί ο χρόνος  $t$ , τότε οι παράγωγοι που υπολογίζουμε ως προς τον χρόνο δείχνουν τον ρυθμό μεταβολής (ταχύτητα) και την επιτάχυνση της μεταβολής του διανύσματος, αλλά μόνο το πώς μεταβάλλονται με την πάροδο του χρόνου. Αυτό σημαίνει ότι υπάρχει το ενδεχόμενο να βρισκόμαστε σε σημεία μικρής καμπυλότητας (σχ. 4.1, καμπυλώνει “λίγο”) και το διάνυσμα  $\mathbf{r}(t)$  να διατρέχει με μεγάλη ταχύτητα και επιτάχυνση την τροχιά της καμπύλης (αυτό προκύπτει από μεγάλη τιμή των παραγώγων των συνιστωσών  $x(t)$  και  $y(t)$ ), αλλά και αντίστροφα, στα σημεία μεγάλης καμπυλότητας (σχ. 4.1, καμπυλώνει “πολύ”), μπορεί το διάνυσμα να προχωράει αργά ή να επιβραδύνεται πάνω στην καμπύλη. Αυτό συνεπάγεται ότι η καμπυλότητα ως μέτρο της δεύτερης παραγώγου ως προς το χρόνο θα δώσει μεγάλες τιμές στην πρώτη περίπτωση (σημεία μικρής καμπύλωσης) και μικρές τιμές στη δεύτερη περίπτωση (σημεία μεγάλης καμπύλωσης), αντίθετα αποτελέσματα από αυτά που περιμένουμε.

Για τον παραπάνω λόγο θεωρούμε τη φυσική παράμετρο της καμπύλης  $s$  (μήκος τόξου) και τις παραγώγους των συνιστωσών και του διανύσματος  $\mathbf{f}(t)$  ως προς το  $s$ . Αυτό συνεπάγεται τον υπολογισμό των ρυθμών μεταβολής του διανύσματος καθώς μεταβάλλεται το μήκος τόξου, δηλαδή ανεξάρτητα από τον χρόνο και το πόσο γρήγορα ή αργά κινούμαστε πάνω στην καμπύλη, βλέπουμε πόσο καμπυλώνει η τροχιά για μια συγκεκριμένη μετακίνησή μας ( $ds$ ) πάνω στην καμπύλη  $C$ . Ο τύπος της καμπυλότητας (**ορισμός της καμπυλότητας**) που χρησιμοποιείται είναι το μέτρο της δεύτερης παραγώγου του διανύσματος ως προς  $s$ :

#### ΚΑΜΠΥΛΟΤΗΤΑ - ΤΥΠΟΣ 2:

(εξ. 4.11)

$$curvature2 = \left\| \vec{f}''(s) \right\| = \left\| \frac{d^2 \vec{f}}{ds^2} \right\|$$



Το στοιχειώδες  $ds$  στην περίπτωση των πολλών διαστάσεων δίνεται από το μέτρο της πρώτης παραγώγου του διανύσματος  $\mathbf{f}(\mathbf{t})$  ως προς τον χρόνο:

$$ds(t) = \left| \frac{d\vec{f}(t)}{dt} \right| \quad (\text{εξ. 4.12})$$

ενώ το μήκος τόξου είναι το ολοκλήρωμα όλων των  $ds$  πάνω στην καμπύλη μέχρι το σημείο  $t$ :

$$\begin{aligned} s &= s(t) = \int_0^t ds = \int_0^t ds(t)dt = \\ &= \int_0^t \left| \frac{d\vec{f}(t)}{dt} \right| dt = \int_0^t \sqrt{\left( \frac{dx_1}{dt} \right)^2 + \left( \frac{dx_2}{dt} \right)^2 + \dots + \left( \frac{dx_N}{dt} \right)^2} dt \quad (\text{εξ. 4.13}) \end{aligned}$$

Στην ακολουθία βίντεο, όπου ως παράμετρο έχουμε τον χρόνο που είναι η εξέλιξη των διαδοχικών καρέ, γνωρίζοντας τις τιμές του διανύσματος χαρακτηριστικών για αυτές τις χρονικές στιγμές, πρέπει να ορίσουμε τις παραγώγους ως προς το μήκος τόξου  $s$ . Αυτό μπορεί να γίνει εύκολα, αν θεωρήσουμε το  $s$  σαν τη μετατόπιση του διανύσματος  $\mathbf{f}$  στον Ευκλείδειο χώρο, μεταξύ των δύο διαδοχικών χρονικών στιγμών. Έτσι έχουμε το  $s$  ως συνάρτηση του  $t$ , και μάλιστα είναι μία αύξουσα (όχι γνησίως) συνάρτηση, με “στοιχειώδεις” αυξήσεις  $ds$  (στοιχειώδεις επειδή το  $dt$  θεωρείται στοιχειώδες) οι οποίες προκύπτουν υπολογίζοντας την απόσταση δύο διαδοχικών στιγμιοτύπων του διανύσματος χαρακτηριστικών (δηλαδή το μέτρο της διακριτής παραγώγου). Το  $s$  προκύπτει ως άθροισμα όλων των  $ds$  από το 0 μέχρι το σημείο  $t$ . Διαιρώντας με το στοιχειώδες  $ds$  προκύπτουν οι 1<sup>η</sup> και 2<sup>η</sup> παράγωγοι του  $\mathbf{f}$ :

$$\text{feature vector : } \vec{f}(t) = [x_1(t), x_2(t), \dots, x_N(t)], \quad t \in [t_1, t_2, \dots, t_m]$$

για  $m$  συνεχόμενα καρέ

$$\begin{aligned} \text{Στοιχειώδες } ds : \quad ds(t) &= \left\| \vec{f}(t) - \vec{f}(t-1) \right\| = \\ &= \sqrt{(x_1(t) - x_1(t-1))^2 + \dots + (x_N(t) - x_N(t-1))^2} \end{aligned}$$

$$\text{1<sup>η</sup> παράγωγος : } \vec{d}_1 = \frac{d\vec{f}}{ds} = \frac{\vec{f}(t) - \vec{f}(t-1)}{ds}, \quad t \in [t_2, t_3, \dots, t_m]$$

$$\text{2<sup>η</sup> παράγωγος : } \vec{d}_2 = \frac{d^2 \vec{f}}{ds^2} = \frac{\vec{d}_1(t) - \vec{d}_1(t-1)}{ds}, \quad t \in [t_3, t_4, \dots, t_m]$$

Σε περίπτωση που το  $f(t)$  μένει σταθερό οπότε  $ds = 0$ , οι τιμές των παραγώγων προκύπτουν ίσες με μηδέν, αφού δεν υπάρχει μεταβολή. Σύμφωνα με τη θεωρία ([Car76]), παραμετροποιημένη διαφορίσιμη καμπύλη που έχει μη μηδενική πρώτη παράγωγο για όλες τις τιμές του  $t$  ονομάζεται κανονική και το μήκος τόξου ως συνάρτηση του  $t$  είναι γνησίως αύξουσα. Στο βίντεο υπάρχει σπάνια περίπτωση δύο διαδοχικά σημεία να συμπίπτουν, και γι' αυτό η συνάρτηση  $s(t)$  είναι απλώς αύξουσα.

[Ένας άλλος τρόπος για τον υπολογισμό των παραγώγων ως προς  $s$  είναι η χρήση του κανόνα της αλυσίδας και του τύπου:

$$f'' = \frac{\ddot{f} \cdot \dot{s} - \dot{f} \cdot \ddot{s}}{\dot{s}^3},$$

$$\text{όπου } \dot{s} = \frac{ds}{dt}, \ddot{s} = \frac{d^2s}{dt^2}, \dot{f} = \frac{df}{dt}, \ddot{f} = \frac{d^2f}{dt^2}, f'' = \frac{d^2f}{ds^2}$$

ο οποίος αποδεικνύεται στο Παράρτημα (παράγραφος Π.1). Μέσω του τύπου αυτού μπορεί να υπολογιστεί απευθείας η δεύτερη παράγωγος ως προς το  $s$  συναρτήσει των παραγώγων ως προς  $t$ , αλλά με τον μόνο περιορισμό η συνάρτηση  $s(t)$  να είναι συνεχής, γνησίως μονότονη και ένα προς ένα (αντιστρέψιμη). Επειδή στο βίντεο υπάρχει περίπτωση γειτονικά καρέ να έχουν το ίδιο διάνυσμα (σταθερή εικόνα), οπότε η  $s(t)$  να μην είναι γνησίως αύξουσα αλλά απλώς αύξουσα συνάρτηση, δεν χρησιμοποιούμε τον τύπο διότι δεν τηρούνται αυστηρά οι περιορισμοί. ]

## 4.4 Η Φυσική Παράμετρος $s$ και η Τυχαία Παράμετρος $t$

Για την περιγραφή της μορφής μιας καμπύλης στον τρισδιάστατο χώρο, εκτός από την καμπυλότητα που εκφράζει την τάση της καμπύλης να αποκλίνει από την ευθεία γραμμή (δηλαδή όταν από τη μία διάσταση τείνει να “πάει” στις δύο), υπάρχει και η στρέψη που εκφράζει την τάση της καμπύλης να αποκλίνει από το επίπεδο (δηλαδή όταν από τις δύο διαστάσεις τείνει να “πάει” στις τρεις). Μια καμπύλη δύο διαστάσεων έχει μηδενική στρέψη, ενώ μια καμπύλη μιας διάστασης (ευθεία γραμμή) έχει μηδενική στρέψη και μηδενική καμπυλότητα.

Όπως παρουσιάζεται στο βιβλίο του Ιωάννη Θ. Χαΐνη ([Χαι98]), καθώς και στο βιβλίο του Michael Spivak ([Spi99]), στο σημείο όπου γίνεται διαφορική ανάλυση των καμπυλών έως και 3 διαστάσεων, δίνονται διαφορετικοί τύποι για παράμετρο  $t$  και για παράμετρο  $s$ . Έτσι δεν γίνεται παρερμηνευση των ορισμών καμπυλότητας και στρέψης, αφού ανάλογα με την παράμετρο της εξίσωσης που χρησιμοποιούμε ισχύουν και οι αντίστοιχοι τύποι, οι οποίοι φυσικά δίνουν το ίδιο αποτέλεσμα αφού προκύπτουν από τους ορισμούς καμπυλότητας και στρέψης από το **τρίεδρο του Frenet** ή τις λεγόμενες σχέσεις “Serret-Frenet formulas” (βλέπε βιβλιογραφία). Οι τύποι αυτοί για την **καμπυλότητα  $\kappa$**  και τη **στρέψη  $\sigma$**  είναι οι παρακάτω (η τελεία εκφράζει παραγώγιση ως προς  $t$ , ο τόνος εκφράζει παραγώγιση ως προς  $s$ ):

**Φυσική παράμετρος  $s$ :**

Εξίσωση της καμπύλης :  $\vec{r} = \vec{r}(s)$

Καμπυλότητα :  $\kappa(s) = \|\vec{r}''\|$  (εξ. 4.14)

Στρέψη:  $\sigma(s) = \frac{\vec{r}' \cdot (\vec{r}'' \times \vec{r}''')}{\|\vec{r}''\|^2}$  (εξ. 4.15)

**Τυχαία παράμετρος  $t$ :**

Εξίσωση της καμπύλης :  $\vec{r} = \vec{r}(t)$

Καμπυλότητα :  $\kappa(t) = \frac{\|\dot{\vec{r}} \times \ddot{\vec{r}}\|}{\|\dot{\vec{r}}\|^3}$  (εξ. 4.16)

Στρέψη:  $\sigma(t) = \frac{\dot{\vec{r}} \cdot (\ddot{\vec{r}} \times \dddot{\vec{r}})}{\|\dot{\vec{r}} \times \ddot{\vec{r}}\|^2}$  (εξ. 4.17)

Με τους παραπάνω τύπους για την τυχαία παράμετρο, που ισχύουν χωρίς περιορισμούς θα μπορούσαμε να υπολογίσουμε την καμπυλότητα από τις τιμές του διανύσματος χαρακτηριστικών, αφού γνωρίζουμε ότι οι τιμές για τις συνιστώσες του διανύσματος δίνονται συναρτήσει του χρόνου  $t$ . Οι τύποι έχουν αποδειχθεί μόνο για τις τρεις διαστάσεις. Κάνοντας μία παραδοχή, ότι ισχύουν και για περισσότερες από τις τρεις διαστάσεις, απομένει να οριστεί μια γωνία μεταξύ δύο διανυσμάτων του  $N$ -διάστατου Ευκλείδειου χώρου. Από τον ορισμό του μέτρου του εξωτερικού γινομένου και τον παραπάνω τύπο για παράμετρο  $t$  (βλέπε απόδειξη στο Παράρτημα, παράγραφος

Π.2), προκύπτει η σχέση που δίνει την καμπυλότητα συναρτήσει των χρονικών παραγώγων:

### ΚΑΜΠΥΛΟΤΗΤΑ - ΤΥΠΟΣ 3:

(εξ. 4.18)

$$curvature3 = |\vec{f}''(s)| = \frac{\sqrt{\dot{s}^2 \cdot \|\ddot{\vec{f}}\|^2 - (\dot{\vec{f}} \cdot \ddot{\vec{f}})^2}}{\dot{s}^3}$$

Οι παράγωγοι ως προς τον χρόνο (δείχνονται με τελεία) είναι οι διαφορές των τιμών των διαδοχικών στιγμιοτύπων:

$$\text{feature vector : } \vec{f}(t) = [x_1(t), x_2(t), \dots, x_N(t)], \quad t \in [t_1, t_2, \dots, t_m]$$

για m συνεχόμενα καρέ

$$1^{\text{η}} \text{ παράγωγος του feature vector : } \frac{d\vec{f}}{dt} = \vec{f}(t) - \vec{f}(t-1), \quad t \in [t_2, t_3, \dots, t_m]$$

$$2^{\text{η}} \text{ παράγωγος του feature vector : } \frac{d^2 \vec{f}}{dt^2} = \vec{d}_1(t) - \vec{d}_1(t-1), \quad t \in [t_3, t_4, \dots, t_m]$$

$$1^{\text{η}} \text{ παράγωγος του } s : \frac{ds}{dt} = ds = \left| \frac{d\vec{f}}{dt} \right| = \|\vec{f}(t) - \vec{f}(t-1)\|, \quad t \in [t_2, t_3, \dots, t_m]$$

[ Δεν περιοριστήκαμε στον τύπο 2 για την καμπυλότητα, παρόλο που θεωρητικά είναι σωστός και καλύπτει και την περίπτωση του N-διάστατου Ευκλείδειου χώρου καθώς αποτελεί τον ορισμό της καμπυλότητας, για δύο λόγους. Ο ένας είναι ότι ο πολυδιάστατος χώρος απαιτεί διαφορετική αντιμετώπιση δεδομένου ότι οι περισσότερες διαστάσεις λογικά θα περιπλέξουν περισσότερο τον τύπο της καμπυλότητας και ιδιαίτερα επειδή πλέον δεν θα υπάρχει μία μόνο στρέψη (αφού έχουμε πολλά επίπεδα και όχι μόνο ένα). Ο άλλος λόγος είναι ο εξής: η χρήση της παραμέτρου s επιβάλλει την εκμετάλλευση της φυσικής αναπαράστασης της καμπύλης για την εξαγωγή συμπερασμάτων σχετικά με τα ακραία σημεία της. Για την σύγκριση μεταξύ των διαδοχικών σημείων απαιτείται η χρήση ίσων ds, αλλά στο βίντεο ενώ τα dt είναι ίσα (και ίσα με 1), η μετατόπιση ds έχει πολλές διακυμάνσεις. Έτσι δεν έχουμε “συνεχή” καμπύλη ως προς το s, όπως έχουμε ως προς το t (δεχόμαστε ότι το dt είναι πολύ μικρό). Αναγκαστικά λόγω έλλειψης πληροφορίας μεταξύ των καρέ, πρέπει να εξετάσουμε τα ds ως ευθύγραμμα τμήματα – αποστάσεις μεταξύ διαδοχικών σημείων και να αγνοήσουμε

την ενδιάμεση πιθανή καμπύλωση. Γι' αυτό προτιμούμε τους τύπους που εκφράζουν την καμπυλότητα με βάση τις παραγώγους ως προς  $t$ , ώστε να είναι δυνατή η γραφική αναπαράσταση και η σύγκριση των τιμών. ]

## 4.5 Γενικευμένες Καμπυλότητες στον χώρο $R^n$

Για την μελέτη των καμπυλών στον  **$n$ -διάστατο Ευκλείδειο χώρο  $R^n$**  απαιτείται όπως είπαμε διαφορετική αντιμετώπιση των περιπτώσεων της διαφορικής γεωμετρίας, και επειδή ο τύπος 3 έχει προκύψει καταχρηστικά γενικεύοντας τη θεωρία που ισχύει για τις τρεις διαστάσεις, αναζητούμε έναν πιο γενικό ορισμό που να καλύπτει την περίπτωση του πολυδιάστατου χώρου, όπου θα πρέπει να ορίζονται μία καμπυλότητα και  $n-2$  στρέψεις.

Σύμφωνα με τον Herman Gluck ([Glu66]) αναπτύσσεται ένας απλός αλγόριθμος που βασίζεται στην διαδικασία της ορθοκανονικοποίησης Gram-Schmidt και υπολογίζει τις **υψηλότερες καμπυλότητες (higher curvatures)** μίας καμπύλης στον  $n$ -διάστατο Ευκλείδειο χώρο, οι οποίες είναι η καμπυλότητα και οι στρέψεις που αναφέρθηκαν παραπάνω. Η διαδικασία είναι σχετικά απλή, αφού ουσιαστικά κατασκευάζει ένα πλαίσιο Frenet (Frenet frame) ορίζοντας τα διανύσματα Frenet (Frenet vectors), το οποίο αντιστοιχεί στο τρίεδρο του Frenet που ορίζεται για τις τρεις διαστάσεις (βλέπε βιβλιογραφία), και στη συνέχεια ορίζει τις καμπυλότητες ως εσωτερικά γινόμενα των διανυσμάτων Frenet (της παραγώγου του τρέχοντος διανύσματος επί το επόμενο διάνυσμα). Παρόμοια οδός ακολουθείται και από τον Michael Spivak ([Spi99]) για την ταξινόμηση καμπυλών στον  $R^n$  σε κατάλληλη Ευκλείδεια κίνηση, όπου και πάλι χρησιμοποιείται η ορθοκανονικοποίηση Gram-Schmidt για τον υπολογισμό  **$n-1$  συναρτήσεων καμπυλότητας (curvature functions)**.

Τέλος, σε διάφορους δικτυακούς τόπους ([WWW]) γίνεται επίσης εκτενής ανάλυση της Διαφορικής Γεωμετρίας των Καμπυλών (Differential Geometry of Curves), όπου αντίστοιχα με τις παραπάνω αναφορές, ορίζονται οι  **$n-1$  γενικευμένες καμπυλότητες (generalized curvatures)** μια καμπύλης στον  $n$ -διάστατο χώρο. Δεν χρησιμοποιείται η φυσική αναπαράσταση της καμπύλης, αλλά απευθείας η τυχαία παράμετρος  $t$ , και για τον λόγο αυτό υιοθετούμε αυτή την μεθοδολογία για να εξάγουμε

το αποτέλεσμα που επιθυμούμε. Η διαδικασία είναι ίδια με τις προηγούμενες, όπου κατασκευάζεται το **πλαίσιο Frenet (Frenet frame)**, που αποτελείται από ορθοκανονικοποιημένα διανύσματα (με την **μέθοδο Gram-Schmidt**):

$$e_1(t), e_2(t), \dots, e_n(t)$$

τα οποία ονομάζονται **διανύσματα Frenet (Frenet vectors)**.

Οι γενικευμένες καμπυλότητες (**generalized curvatures**) δίνονται από τον τύπο:

$$k_i(t) = \frac{\dot{\vec{e}}_i(t) \cdot \vec{e}_{i+1}(t)}{\|\dot{\vec{f}}(t)\|} \quad (\text{εξ. 4.19})$$

για  $i = 1, 2, \dots, n-1$ , όπου  $\mathbf{f}(t)$  είναι το διάνυσμα που περιγράφει την εξίσωση της καμπύλης με παράμετρο τον χρόνο  $t$  (στην περίπτωσή μας είναι το διάνυσμα των χαρακτηριστικών).

Εκτενής ανάλυση της μεθόδου και των υπολογισμών βρίσκεται στο Παράρτημα, παράγραφος Π.3. Εδώ παραθέτουμε μόνο τον τύπο για την **πρώτη γενικευμένη καμπυλότητα (first generalized curvature)**, η οποία είναι και αυτή που μας ενδιαφέρει:

$$k_1(t) = \frac{\dot{\vec{e}}_1(t) \cdot \vec{e}_2(t)}{\|\dot{\vec{f}}(t)\|} \quad (\text{εξ. 4.20})$$

Οι καμπυλότητες περιγράφουν πλήρως τη γεωμετρική μορφή της καμπύλης στον  $N$ -διάστατο χώρο, όπου η πρώτη καμπυλότητα δείχνει το μέτρο της τάσης της καμπύλης να αποκλίνει από την ευθεία γραμμή, ενώ οι υπόλοιπες  $n-2$  καμπυλότητες εκφράζουν το μέτρο της τάσης της καμπύλης να αποκλίνει από ένα υπερεπίπεδο, δηλαδή από έναν χώρο με διαστάσεις λιγότερες από  $n$  (δηλαδή ένας χώρος  $m$  διαστάσεων που είναι υποσύνολο του  $n$ -D χώρου).

Αντικαθιστώντας τις τιμές από τους τύπους που ορίζονται κατά την διαδικασία (βλέπε Παράρτημα), έχουμε τις σχέσεις:

$$\text{feature vector :} \quad \vec{f}(t) = [x_1(t), x_2(t), \dots, x_N(t)], \quad t \in [t_1, t_2, \dots, t_m]$$

για  $m$  συνεχόμενα καρέ

$$\text{1st Gram-Schmidt vector:} \quad \vec{u}_1(t) = \dot{\vec{f}}(t)$$

$$\text{1st Frenet vector :} \quad \vec{e}_1(t) = \frac{\vec{u}_1(t)}{\|\vec{u}_1(t)\|} = \frac{\dot{\vec{f}}(t)}{\|\dot{\vec{f}}(t)\|}$$

**2<sup>nd</sup> Gram-Schmidt vector:**  $\vec{u}_2(t) = \ddot{\vec{f}}(t) - \left( \ddot{\vec{f}}(t) \cdot \vec{e}_1 \right) \cdot \vec{e}_1$

**2<sup>nd</sup> Frenet vector :**  $\vec{e}_2(t) = \frac{\vec{u}_2(t)}{\|\vec{u}_2(t)\|}$

**1<sup>st</sup> Generalized Curvature :**  $k_1(t) = \frac{\dot{\vec{e}}_1(t) \cdot \vec{e}_2(t)}{\|\dot{\vec{f}}(t)\|} = \frac{\|\vec{u}_2\|}{\|\dot{\vec{f}}(t)\| \cdot \|\vec{u}_1\|}$

Αντικαθιστώντας το μέτρο της πρώτης παραγώγου του διανύσματος χαρακτηριστικών με την μετατόπιση  $ds$  πάνω στην καμπύλη και τα  $u_1$  και  $u_2$  από τους παραπάνω τύπους, η καμπυλότητα δίνεται από τον τύπο:

**ΚΑΜΠΥΛΟΤΗΤΑ - ΤΥΠΟΣ 4:**

(εξ. 4.21)

$$curvature4 = \frac{\left\| \ddot{\vec{f}}(t) - \left[ \frac{\ddot{\vec{f}}(t) \cdot \dot{\vec{f}}(t)}{ds(t)} \right] \cdot \left[ \frac{\dot{\vec{f}}(t)}{ds(t)} \right] \right\|}{[ds(t)]^2}$$

[ Όλα τα μεγέθη που χρησιμοποιούνται στον τύπο έχουν οριστεί παραπάνω. ]

## 4.6 Σύγκριση των Μεθόδων Καμπυλότητας

Στο σημείο αυτό συγκρίνονται οι παραπάνω μέθοδοι υπολογισμού της καμπυλότητας, με την εφαρμογή τους σε μία διδιάστατη και σε μία τρισδιάστατη καμπύλη (θεωρητικά μοντέλα), καθώς και με την εφαρμογή τους στο διάνυσμα χαρακτηριστικών 30 διαστάσεων που προκύπτει από ένα σύντομο απόσπασμα βίντεο.

Δίνονται εδώ συγκεντρωμένοι οι 4 τύποι που χρησιμοποιούνται για την καμπυλότητα:

**Πίνακας 4.1 : ΚΑΜΠΥΛΟΤΗΤΑ στον  $\mathbf{R}^n$**

**ΤΥΠΟΣ 1:**  $curvature1 = \|\vec{f}''(t)\| = \left\| \frac{d^2 \vec{f}}{dt^2} \right\|$

**ΤΥΠΟΣ 2:**  $curvature2 = \|\vec{f}''(s)\| = \left\| \frac{d^2 \vec{f}}{ds^2} \right\|$

**ΤΥΠΟΣ 3:**  $curvature3 = |\vec{f}''(s)| = \frac{\sqrt{\dot{s}^2 \cdot \|\ddot{\vec{f}}\|^2 - \left(\dot{\vec{f}} \cdot \ddot{\vec{f}}\right)^2}}{\dot{s}^3}$

**ΤΥΠΟΣ 4:**  $curvature4 = \frac{\left\| \ddot{\vec{f}}(t) - \left[ \frac{\ddot{\vec{f}}(t) \cdot \dot{\vec{f}}(t)}{ds(t)} \right] \cdot \left[ \frac{\dot{\vec{f}}(t)}{ds(t)} \right] \right\|}{[ds(t)]^2}$

#### 4.6.1 Διδιάστατο Μοντέλο

Η καμπύλη που χρησιμοποιείται για τη σύγκριση των μεθόδων στο επίπεδο (δύο διαστάσεις) έχει τύπο (όπου  $t \in [0, 360]$ ):

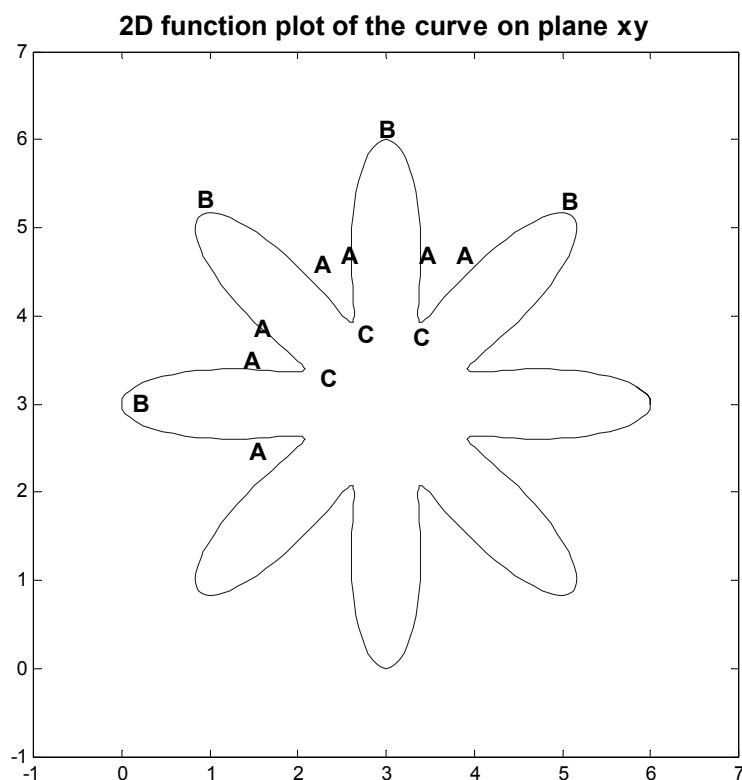
$\vec{r}(t) = [x(t), y(t)]$ , με συνιστώσες:

$$x(t) = 3 + \left[ 2 + \cos\left(\frac{16\pi}{360} \cdot t\right) \right] \cdot \cos\left(\frac{2\pi}{360} \cdot t\right) \quad (\text{εξ. 4.22})$$

$$y(t) = 3 + \left[ 2 + \cos\left(\frac{16\pi}{360} \cdot t\right) \right] \cdot \sin\left(\frac{2\pi}{360} \cdot t\right) \quad (\text{εξ. 4.23})$$



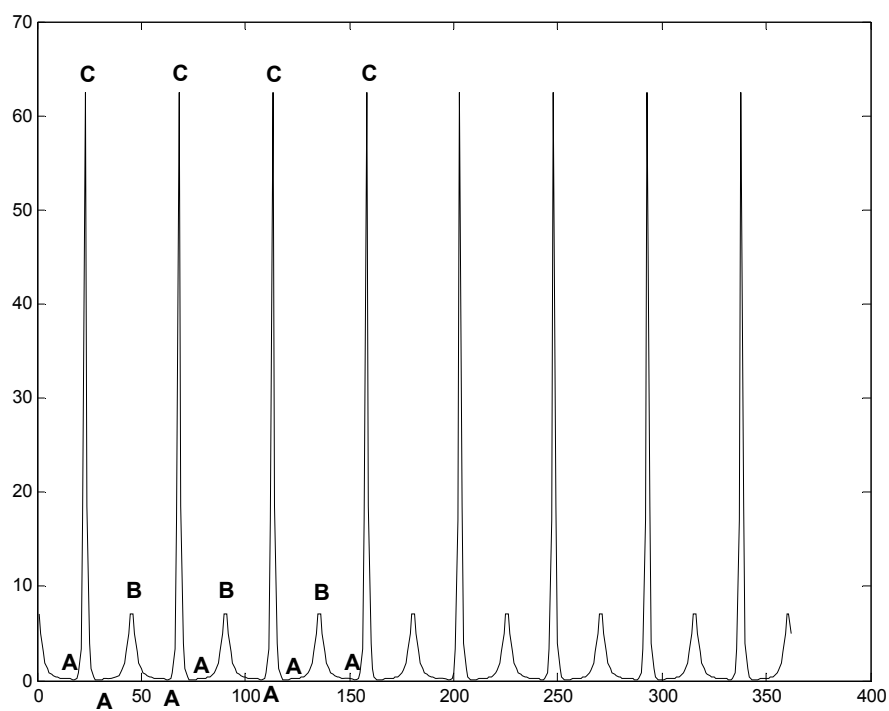
Η μορφή της καμπύλης είναι μία “μαργαρίτα” στο επίπεδο, όπως φαίνεται και στο σχήμα 4.4, η οποία έχει πολύ σημαντικά χαρακτηριστικά ως προς τον τρόπο καμπύλωσής της. Παρουσιάζει σημεία πολύ μικρής καμπύλωσης (τα πλάγια τμήματα στα “πέταλα” της μαργαρίτας, σημεία A), αλλά και σημεία μεγάλης καμπύλωσης (οι βάσεις και τα άκρα των “πετάλων”), τα οποία καμπυλώνουν πιο ομαλά στα άκρα των πετάλων (σημεία B) και πιο απότομα στα σημεία όπου ενώνονται τα πέταλα (σημεία C).



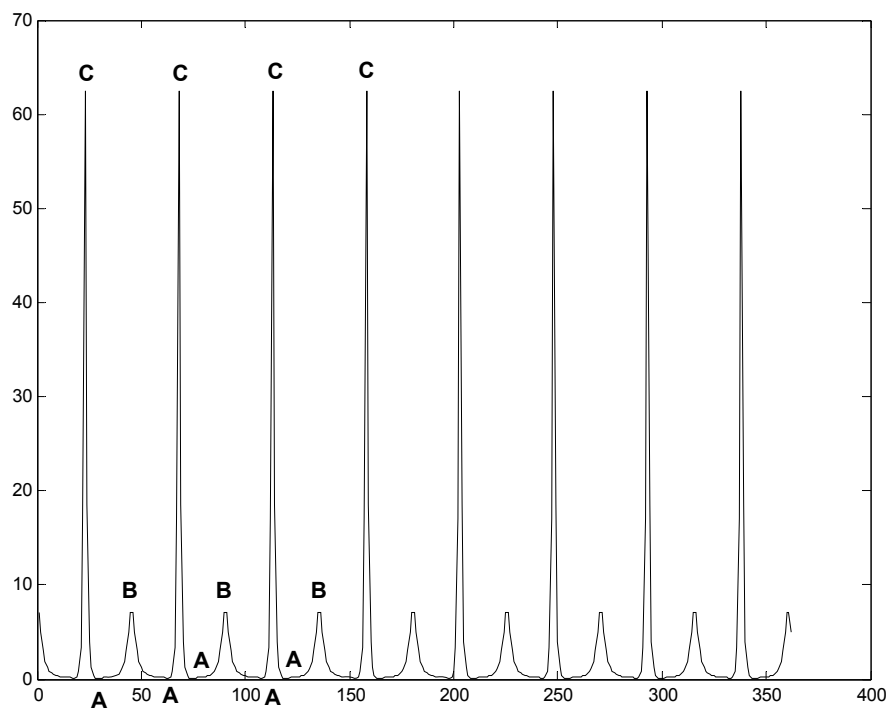
**Σχ. 4.4 :** Το διδιάστατο μοντέλο της επίπεδης καμπύλης

Στα διαγράμματα 4.5 – 4.8 απεικονίζεται η καμπυλότητα για κάθε μία από τις τέσσερις προαναφερθείσες μεθόδους, σε συνάρτηση με την παράμετρο  $t$ :





Σχ. 4.7 : 2D Καμπυλότητα - Τύπος 3 (Καταχρηστικός τύπος)



Σχ. 4.8 : 2D Καμπυλότητα - Τύπος 4 (Γενικευμένη καμπυλότητα)

Αν η καμπύλη του σχήματος 4.4 ήταν η τροχιά του διανύσματος χαρακτηριστικών  $\mathbf{f}(\mathbf{t})$  για τα καρέ ενός βίντεο, θα έπρεπε τα σημεία C να θεωρούνται τα πιο σημαντικά ως προς το περιεχόμενο, ενώ τα σημεία B θα έπρεπε να σχετίζονται με μικρότερης κλίμακας μεταβολές στο περιεχόμενο. Τα σημεία A θα αντιστοιχούσαν σε καρέ με σχεδόν σταθερό περιεχόμενο. Όπως προκύπτει από τον τύπο της καμπύλης, τα σημεία B παρουσιάζονται στις γωνίες  $t$  με μοίρες  $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$  (8 σημεία). Τα σημεία C παρουσιάζονται στις γωνίες  $t$  με μοίρες  $22.5^\circ, 67.5^\circ, 112.5^\circ, 157.5^\circ, 202.5^\circ, 247.5^\circ, 292.5^\circ, 337.5^\circ$  (8 σημεία). Τα σημεία A μικρής μεταβολής είναι τα σημεία μεταξύ των B και C.

Από το πρώτο από αυτά τα διαγράμματα (σχ. 4.5), όπου παρουσιάζεται η καμπυλότητα ως μέτρο της δεύτερης παραγώγου του διανύσματος  $\mathbf{r}(\mathbf{t})$  ως προς τον χρόνο  $t$ , φαίνονται οι αιχμές (peaks) που παρουσιάζονται στα ακραία σημεία της καμπύλης, δηλαδή αντιστοιχούν στα σημεία όπου αλλάζει απότομα η κατεύθυνση (σημεία B και C). Όπως παρατηρούμε από τη γραφική παράσταση τα σημεία B και C θεωρούνται εξίσου απότομα σε μεταβολή της κατεύθυνσης, και μάλιστα τα σημεία B παρουσιάζουν λίγο μεγαλύτερη τιμή καμπυλότητας. Αυτό σημαίνει ότι στα σημεία B το περιεχόμενο του βίντεο θεωρείται ότι αλλάζει με τον ίδιο βαθμό που αλλάζει και στα σημεία C, γεγονός που αντιβαίνει την αρχική διαπίστωση ότι τα σημεία με βάση τη σημαντικότητά τους θα πρέπει να ταξινομηθούν ως εξής: C, B, A.

Από τα υπόλοιπα διαγράμματα, τα οποία είναι παρόμοια, παρατηρούμε άνισες αιχμές, όπου οι μεγαλύτερες αντιστοιχούν στα σημεία C, ενώ οι μικρότερες αντιστοιχούν στα σημεία B. Όπως φαίνεται σε σύγκριση με το διάγραμμα του σχ. 4.5, οι υπόλοιπες μέθοδοι εντοπίζουν ιδανικά την καμπύλωση όσο και τη διαφορά στο βαθμό καμπύλωσης μεταξύ B και C, αλλά και στα μεταξύ τους σημεία (σημεία A) δίνουν μικρή τιμή της καμπυλότητας (σχεδόν μηδέν), δείχνοντας τη μη καμπύλωση της καμπύλης. Η αδυναμία του τύπου 1 να εκφράσει τη σχεδόν μηδενική καμπύλωση στα σημεία A και την διαφορά στον βαθμό καμπύλωσης μεταξύ μεγάλων καμπυλώσεων (B και C) οφείλεται στο ότι η τιμή του  $\text{curvature}_1$  εξαρτάται από την ταχύτητα που διαγράφεται η καμπύλη, γεγονός που εξηγήθηκε στην παράγραφο 4.2.

Θεωρώντας τον τύπο 4 ως τον πιο γενικό, αφού προκύπτει από θεωρία που ισχύει για πολλές διαστάσεις, συγκρίνουμε τα αποτελέσματα των τύπων 3 και 4 και παρατηρούμε ότι συμπίπτουν οι γραφικές παραστάσεις. Επίσης, ο τύπος 2 διαφέρει μόνο στον λόγο των μεγάλων αιχμών (peaks) προς τις μικρότερες, αλλά όμως έχει ίδια μορφή με τον γενικό τύπο 4. Στο διδιάστατο μοντέλο μπορούν να θεωρηθούν αποδεκτές οι 3

προσεγγίσεις για την καμπυλότητα, εκτός από τον τύπο 1 που δείχνει ακατάλληλος για την περιγραφή της καμπύλης.

#### 4.6.2 Τρισδιάστατο Μοντέλο

Η καμπύλη που χρησιμοποιείται για την σύγκριση των μεθόδων στον τρισδιάστατο χώρο προέκυψε από την συνάρτηση `peaks` του MATLAB που παριστάνει μία τρισδιάστατη επιφάνεια, πάνω στην οποία πήραμε μία καμπύλη γραμμή (που έχει προβολή στο επίπεδο  $xy$  μία έλλειψη). Το διάνυσμα που περιγράφει την καμπύλη αυτή έχει συνιστώσες:

$$x(t) = 8 \cdot \cos\left(\frac{2\pi}{360} \cdot t\right) \quad (\text{εξ. 4.24})$$

$$y(t) = 1.625 \cdot \sin\left(\frac{2\pi}{360} \cdot t\right) \quad (\text{εξ. 4.25})$$

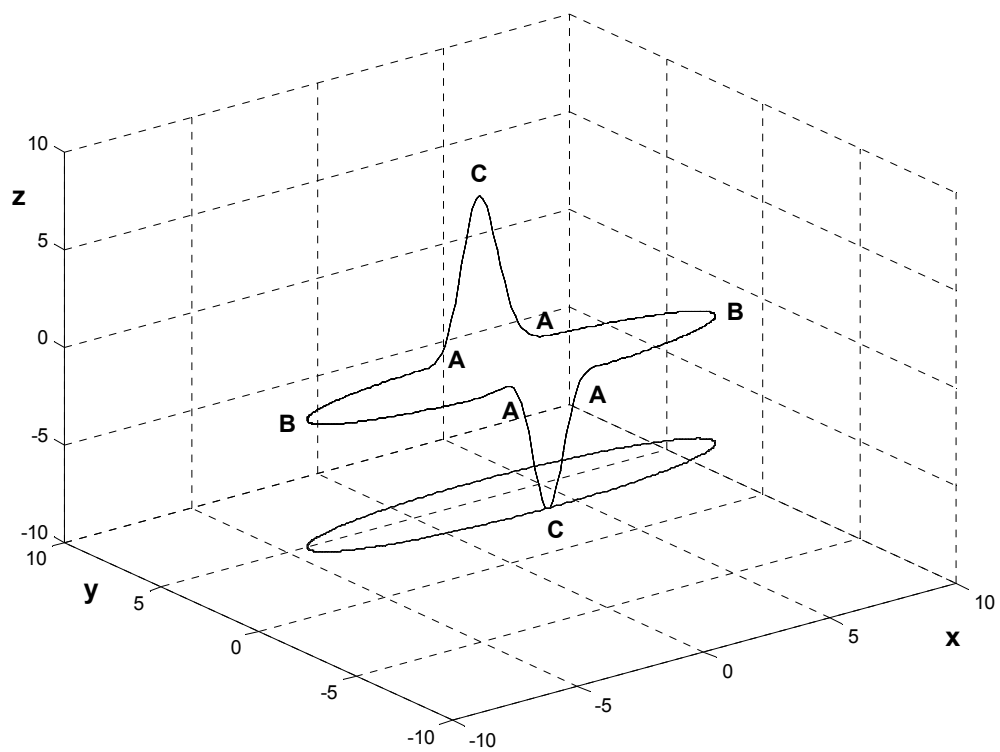
$$z(t) = \text{peaks}(x(t), y(t)) \quad (\text{εξ. 4.26})$$

με  $t \in [0, 360]$ .

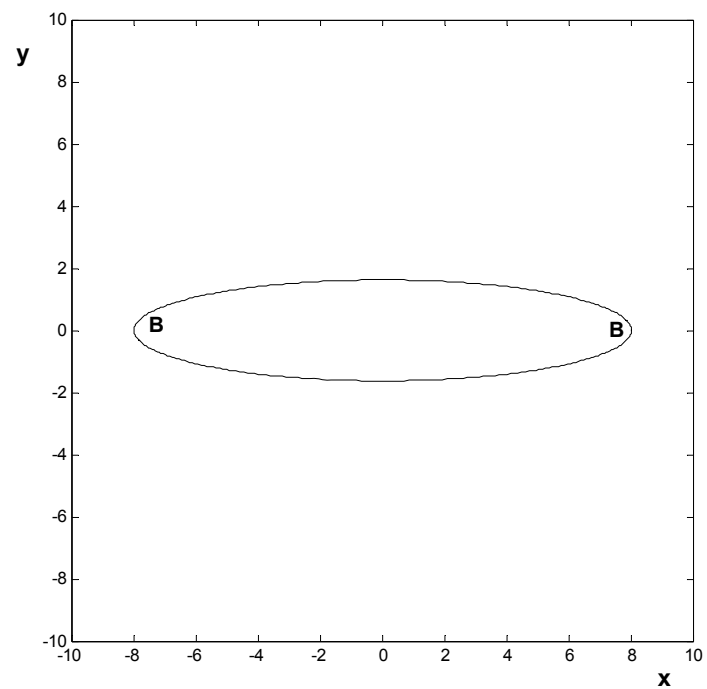
[ Η συνάρτηση `peaks()` του MATLAB έχει τύπο:

$$\begin{aligned} Z(t) &= \text{peaks}(x(t), y(t)) \\ &= 3 \cdot (1-x)^2 \cdot \exp(-(x.^2) - (y+1).^2) \\ &\quad - 10 \cdot (x/5 - x.^3 - y.^5) \cdot \exp(-x.^2 - y.^2) \\ &\quad - 1/3 \cdot \exp(-(x+1).^2 - y.^2) \end{aligned} \quad ]$$

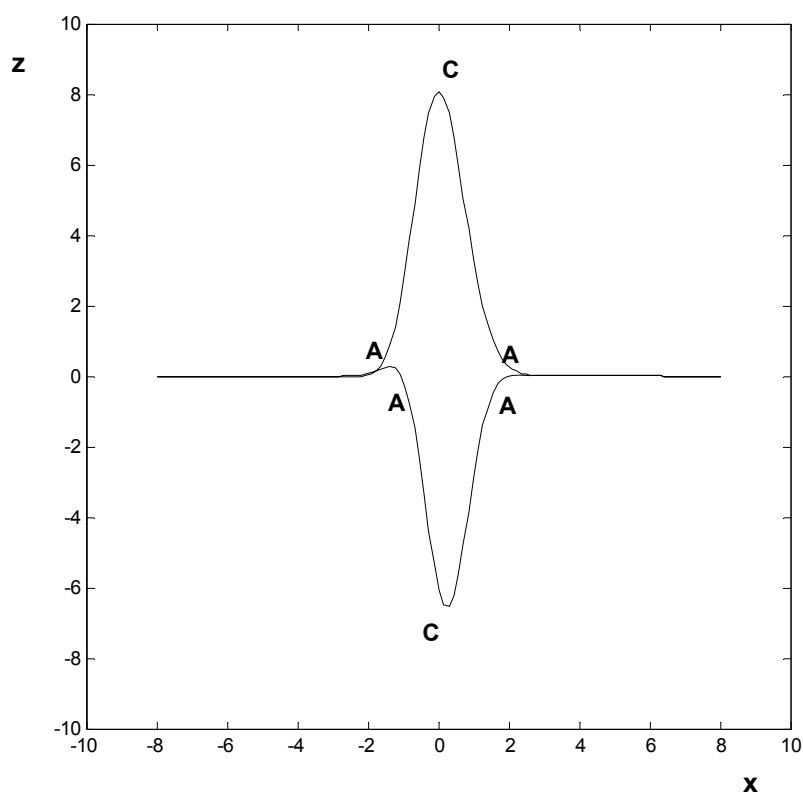
Η τρισδιάστατη απεικόνιση της παραπάνω καμπύλης φαίνεται στο σχήμα 4.9. Τα σημαντικά χαρακτηριστικά ως προς τον τρόπο καμπύλωσής της είναι ότι παρουσιάζει τρία είδη καμπύλωσης. Από το σχήμα 4.9, αλλά κυρίως από τα σχήματα 4.10 και 4.11 που παρουσιάζουν τις προβολές της καμπύλης στα επίπεδα  $x-y$  και  $x-z$ , φαίνεται ότι παρουσιάζεται μικρή καμπύλωση στα σημεία A (σχ. 4.11), λίγο μεγαλύτερη στα σημεία B (σχ. 4.10) και την μεγαλύτερη καμπύλωση στα σημεία C (σχ. 4.11). Επομένως σύμφωνα με τον βαθμό της καμπύλωσης και αντίστοιχα της σημαντικότητας, εάν επρόκειτο για καμπύλη από ακολουθία βίντεο, τα σημεία θα ταξινομηθούν ως εξής: C, B, A (από την πιο απότομη στην πιο ομαλή καμπύλωση). Τα σημεία που βρίσκονται ενδιάμεσα στα παραπάνω δεν παρουσιάζουν καμπύλωση και αναμένουμε η καμπυλότητα σε αυτά να είναι πολύ μικρή.



**Σχ. 4.9 :** Το τρισδιάστατο μοντέλο της καμπύλης σε απεικόνιση στο χώρο (όπου φαίνεται και η προβολή του στο επίπεδο  $x-y$ )



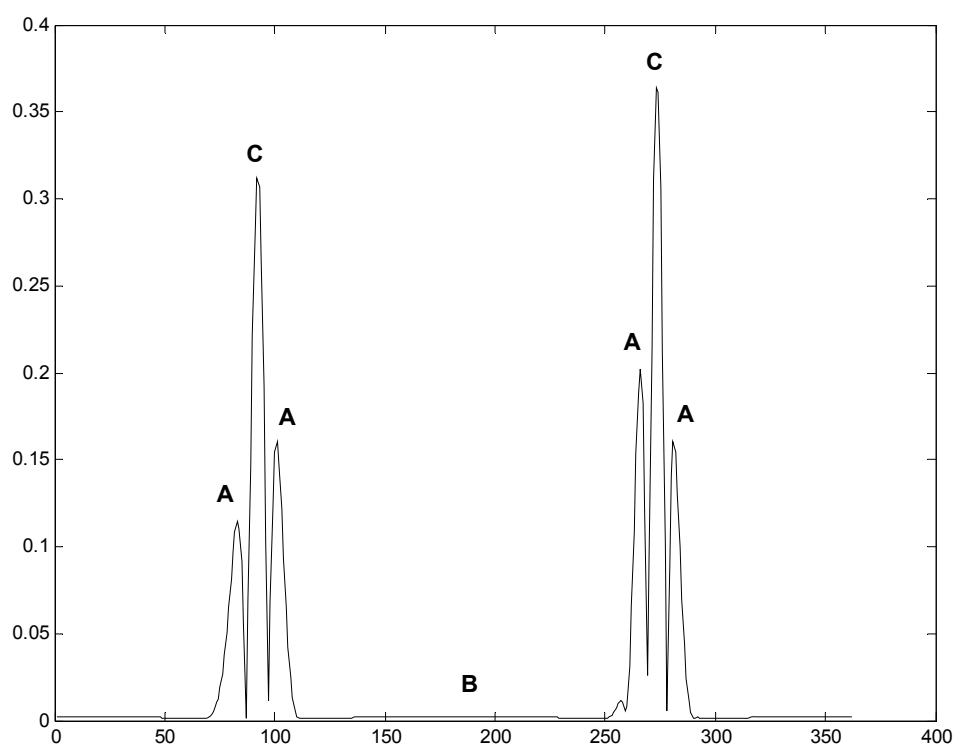
**Σχ. 4.10 :** Η προβολή της 3-D καμπύλης στο επίπεδο  $x-y$



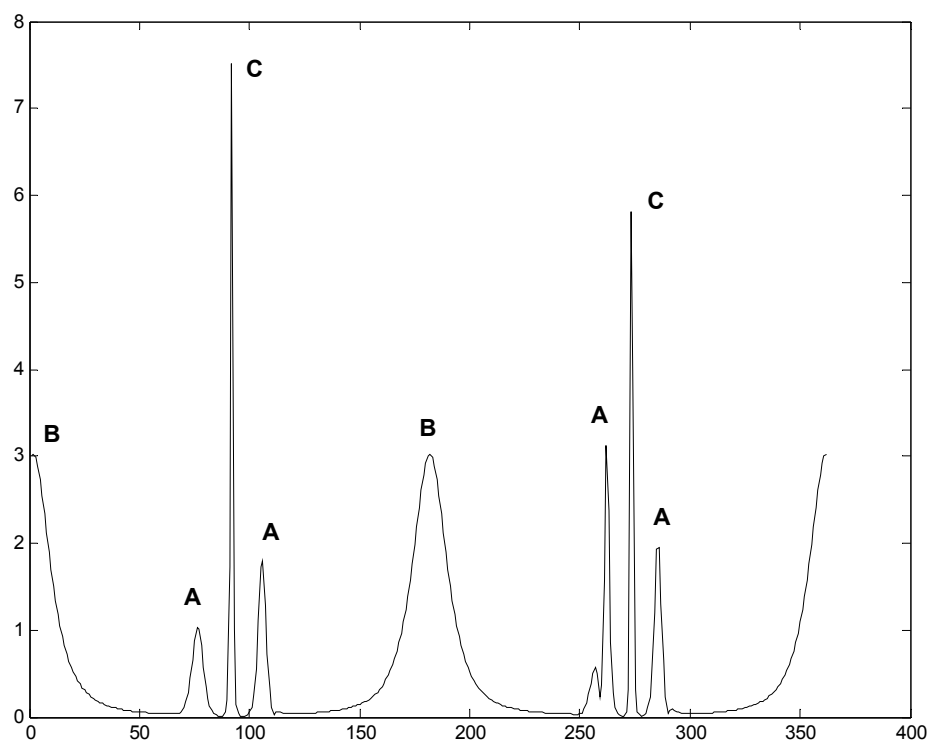
**Σχ. 4.11 :** Η προβολή της 3-D καμπύλης στο επίπεδο x-z

Η σύγκριση μεταξύ των προβολών της καμπύλης στα επίπεδα μπορεί να γίνει με συνέπεια, αφού τα συστήματα των συντεταγμένων που χρησιμοποιούνται είναι ορθοκανονικά, οπότε η ίδια καμπύλωση θα φαίνεται ανάλογη στα δύο διαγράμματα. Όπως προείπαμε στην συγκεκριμένη καμπύλη υπάρχουν σημεία διαφορετικής καμπύλωσης, τα οποία στο τρισδιάστατο διάγραμμα δεν είναι εμφανή, αλλά εντοπίζονται συγκρίνοντας τις προβολές.

Στα σχήματα 4.12 – 4.15 παρατίθενται τα διαγράμματα που απεικονίζουν την καμπυλότητα (άξονας των τεταγμένων) για κάθε μία από τις 4 προαναφερθείσες μεθόδους, σε συνάρτηση με την παράμετρο  $t$  (άξονας των τετμημένων).

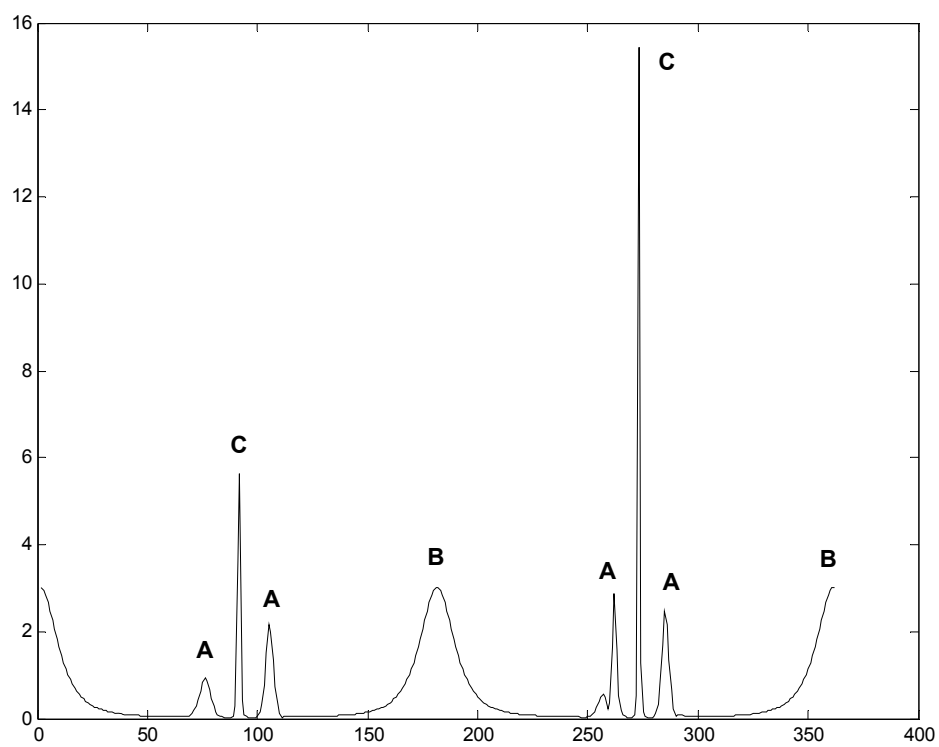


Σχ. 4.12 : 3D Καμπυλότητα - Τύπος 1 (2<sup>η</sup> παράγωγος ως προς t)

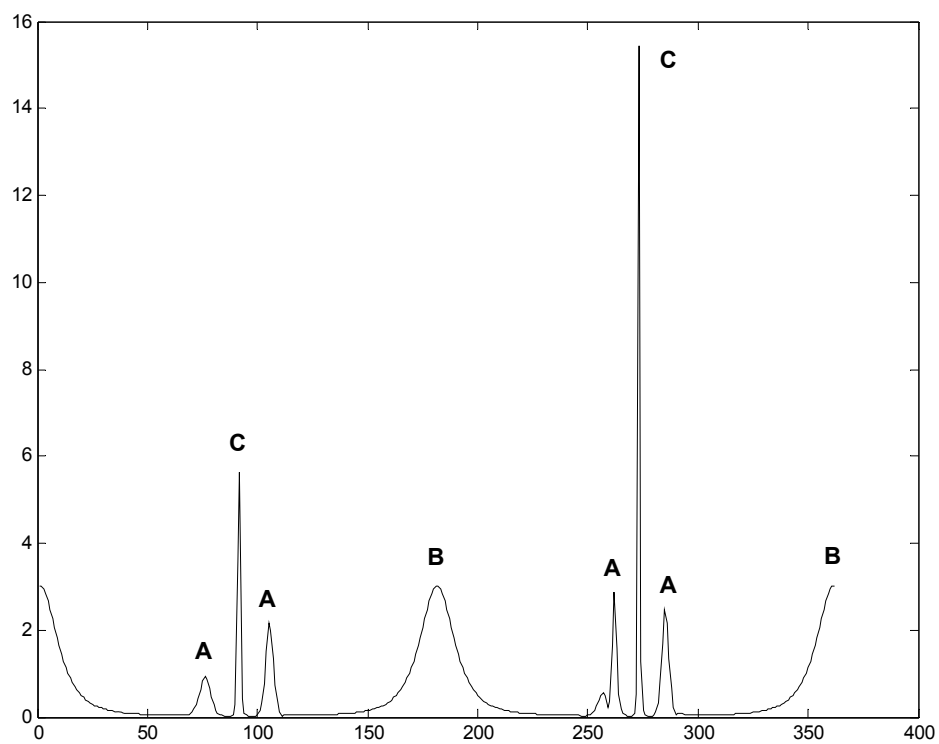


Σχ. 4.13 : 3D Καμπυλότητα - Τύπος 2 (2<sup>η</sup> παράγωγος ως προς s)





**Σχ. 4.14 :** 3D Καμπυλότητα - Τύπος 3 (Καταχρηστικός τύπος)



**Σχ. 4.15 :** 3D Καμπυλότητα - Τύπος 4 (Γενικευμένη καμπυλότητα)

Τα σημεία A παρουσιάζονται περίπου στις γωνίες  $t$  με μοίρες  $75^\circ$ ,  $105^\circ$ ,  $260^\circ$  και  $280^\circ$  (4 σημεία). Τα σημεία B παρουσιάζονται στις γωνίες  $t$  με μοίρες  $0^\circ$  και  $180^\circ$  (2 σημεία). Τα σημεία C παρουσιάζονται στις γωνίες  $t$  με μοίρες  $90^\circ$  και  $270^\circ$  (2 σημεία). Τα σημεία μεταξύ των A, B και C είναι σημεία πολύ μικρής καμπύλωσης.

Παρατηρώντας τα σχήματα 4.12, 4.13, 4.14 και 4.15, το πρώτο συμπέρασμα είναι ότι τα διαγράμματα για τους τύπους 3 και 4 συμπίπτουν όπως και στο διδιάστατο μοντέλο, και άρα για τις 2 και τις 3 διαστάσεις ο τύπος 3 ισχύει (εφόσον προκύπτει από θεωρία που έχει αποδειχθεί και στη βιβλιογραφία για 3 διαστάσεις).

Μία σημαντική επισήμανση είναι οι ατέλειες που και εδώ επιδεικνύει ο τύπος 1, αφού τα δύο σημεία B (στις  $0^\circ$  και  $180^\circ$ ) τα παρουσιάζει με σχεδόν μηδενική καμπυλότητα. Στα σημεία B το διάνυσμα κινείται αργά και η παράγωγος ως προς το χρόνο παραμένει σε μικρές τιμές, οπότε μειώνεται η τιμή της αιχμής (που όμως υπάρχει). Στα σημεία A και C φαίνονται σωστά οι αιχμές της καμπυλότητας, επειδή σε αυτά το διάνυσμα κινείται γρήγορα (με λίγα λόγια η ταχύτητα του διανύσματος πάνω στην καμπύλη καθορίζει το μέγεθος της καμπυλότητας). Επίσης, το διάγραμμα παρουσιάζει στα A και C μεγάλο διάστημα με υψηλές τιμές (όπως φαίνεται τα μικρά και τα μεγάλα peaks εφάπτονται), το οποίο οφείλεται στο ότι το διάνυσμα περνάει με μεγάλη ταχύτητα από τα A, C και A και “βλέπει” μεγάλη καμπυλότητα. Κανονικά θα έπρεπε να ελαττώνεται πολύ η τιμή μεταξύ των δύο αιχμών, αφού στα σημεία αυτά δεν υπάρχει καμπύλωση (η τροχιά είναι σχεδόν ευθύγραμμη).

Ο τύπος 2 όπως παρατηρούμε έχει πολύ καλύτερα αποτελέσματα από τον τύπο 1, αφού παρουσιάζει αιχμές στα σημεία B και ελαττώνει την τιμή της καμπυλότητας μεταξύ των σημείων A και C. Αυτό που παρουσιάζει εσφαλμένα είναι το peak στο σημείο C για  $t = 90^\circ$ , στο οποίο από το διάγραμμα φαίνεται ότι η καμπύλωση είναι μεγαλύτερη απ’ ότι στο C με  $t = 270^\circ$ . Στην πραγματικότητα, παρατηρώντας το σχήμα 4.11, το χαμηλότερο σημείο C που αντιστοιχεί στις  $270^\circ$  παρουσιάζει ελαφρώς μεγαλύτερη καμπύλωση από το C στις  $90^\circ$ .

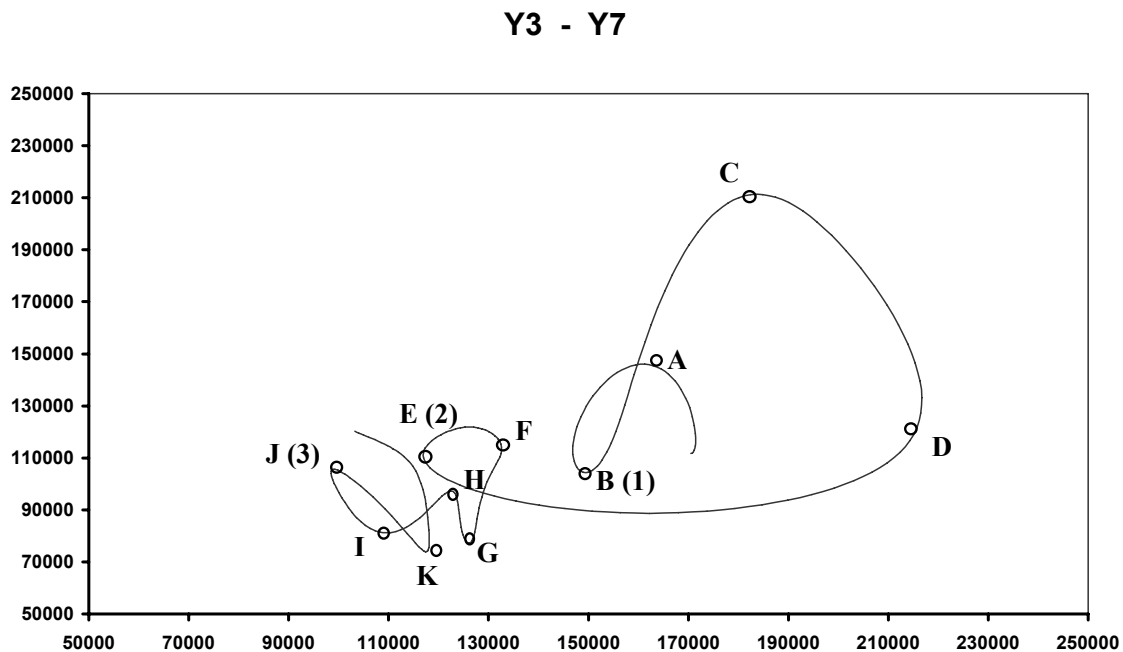
Σε αντίθεση με το διδιάστατο μοντέλο, εδώ τα διαγράμματα των τύπων 2 και 4 διαφέρουν στην τιμή του peak για τα σημεία C. Φαίνεται ότι το διάγραμμα του τύπου 4 δείχνει σωστά την καμπυλότητα για τα C (μεγαλύτερη για  $t = 90^\circ$  απ’ ότι για  $t = 270^\circ$ ). Αυτό μπορεί να οφείλεται στον τρόπο υπολογισμού των διακριτών παραγώγων, όπου χρησιμοποιείται μεταβαλλόμενο  $ds$ , και όχι σε σφάλμα του τύπου curvature<sup>2</sup>.

[ Η υλοποίηση των δύο θεωρητικών μοντέλων έγινε σε περιβάλλον MATLAB. ]

### 4.6.3 Διάνυσμα Χαρακτηριστικών 30-D

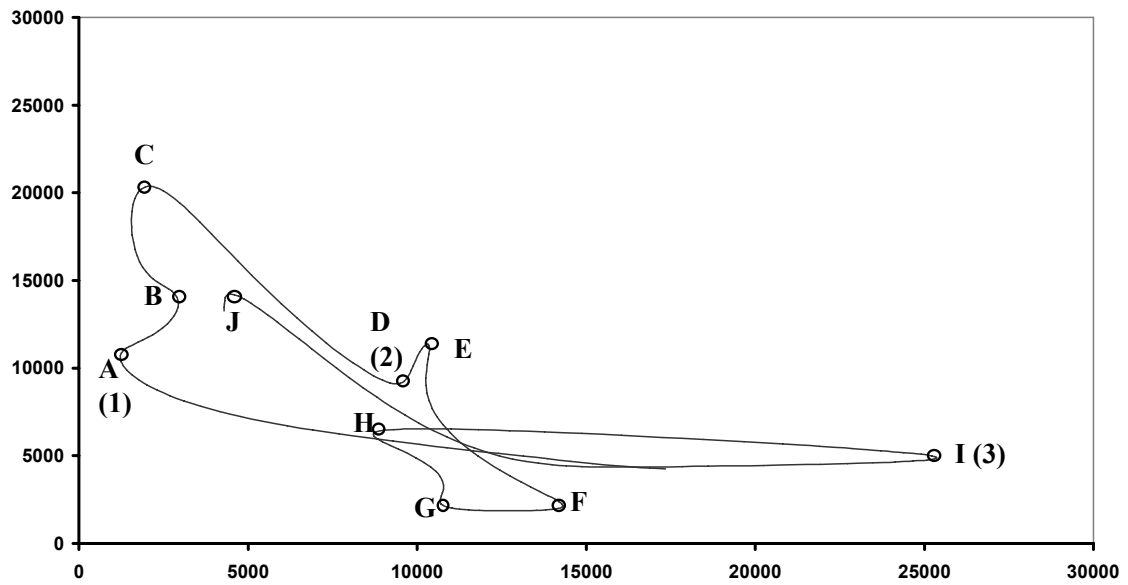
Εδώ χρησιμοποιούμε την καμπύλη του διανύσματος χαρακτηριστικών στον 30-διάστατο Ευκλείδειο χώρο που προκύπτει από τα χαρακτηριστικά ενός σύντομου αποσπάσματος βίντεο που προέρχεται από το βίντεο “RALLY1” και συγκεκριμένα από τα frames 471 έως 720. Το απόσπασμα αυτό προέρχεται από αγώνες αυτοκινήτου, με την κάμερα να καταγράφει πλάνο με ότι φαίνεται προς την κατεύθυνση του αυτοκινήτου μαζί με συγκεκριμένο κομμάτι του εσωτερικού του αυτοκινήτου (που όμως δεν αλλάζει σε περιεχόμενο).

Στα σχήματα 4.16 – 4.18 φαίνονται τρία διαγράμματα που απεικονίζουν την προβολή του διανύσματος χαρακτηριστικών σε τρία επίπεδα δύο διαστάσεων, δηλαδή σε ζεύγη από συνιστώσες του διανύσματος χαρακτηριστικών. Το διάνυσμα αυτό έχει συνολικά 30 συνιστώσες, 10 για κάθε χαρακτηριστικό, αφού πρόκειται για χρωματικό ιστόγραμμα της εικόνας (Y, U, V). Κάθε στοιχείο (bin) του ιστογράμματος συμβολίζεται με  $Y_x$ ,  $U_x$ ,  $V_x$ , αντίστοιχα για κάθε χαρακτηριστικό, όπου το x δείχνει τον αριθμό του bin (τον αύξοντα αριθμό του στοιχείου). Για παράδειγμα αν θέλουμε να συμβολίσουμε το πέμπτο bin του πεδίου Y (φωτεινότητα), γράφουμε Y5.



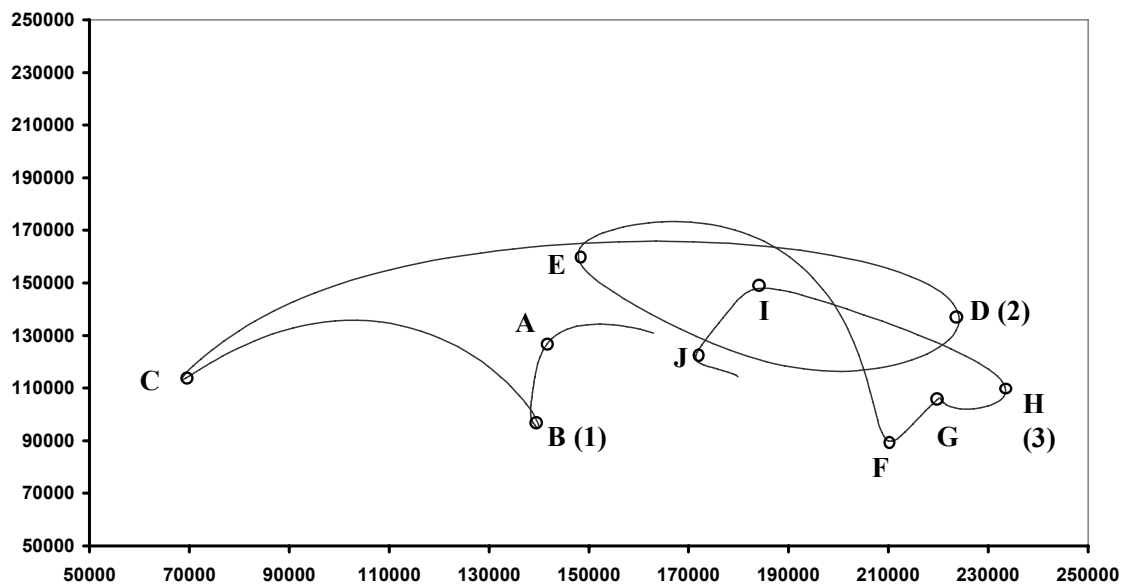
Σχ. 4.16 : Η προβολή της 30-D καμπύλης στο επίπεδο Y3-Y7

### U4 - U6



Σχ. 4.17 : Η προβολή της 30-D καμπύλης στο επίπεδο U4-U6

### Y5 - Y6



Σχ. 4.18 : Η προβολή της 30-D καμπύλης στο επίπεδο Y5-Y6

Οι παραπάνω γραφικές παραστάσεις προέρχονται από τις τιμές δύο συνιστωσών του διανύσματος χαρακτηριστικών, όπως δίνονται παρακάτω, αγνοώντας τις υπόλοιπες συνιστώσες:

Σχ. 4.16 : άξονας των τετμημένων: 3<sup>ο</sup> στοιχείο του Y

άξονας των τεταγμένων: 7<sup>ο</sup> στοιχείο του Y

Σχ. 4.17 : άξονας των τετμημένων: 4<sup>ο</sup> στοιχείο του U

άξονας των τεταγμένων: 6<sup>ο</sup> στοιχείο του U

Σχ. 4.18 : άξονας των τετμημένων: 5<sup>ο</sup> στοιχείο του Y

άξονας των τεταγμένων: 6<sup>ο</sup> στοιχείο του Y

Για τη σωστή αναπαράσταση του διανύσματος χαρακτηριστικών των 30 διαστάσεων σε επίπεδο δύο διαστάσεων, θα έπρεπε να απεικονιστεί σε διδιάστατο διάγραμμα κάθε δυνατό ζεύγος μεταξύ των 30 συνιστωσών, ώστε να παρατηρηθεί η μεταβολή κάθε στοιχείου ως προς τη μεταβολή κάθε άλλου, στο ίδιο σύστημα συντεταγμένων. Η σύγκριση μεταξύ  $30 * 30 = 900$  διαγραμμάτων είναι αδύνατη, αλλά και υπερβολική, αφού η εξέταση πολλών από αυτά θα ήταν περιττή καθώς θα έδιναν σχεδόν ευθύγραμμες καμπύλες, με πολύ λίγα σημεία καμπύλωσης. Τα τμήματα των καμπύλων που έχουν ενδιαφέρον περιορίζονται σε μερικά στοιχεία από κάθε χαρακτηριστικό, στην περιοχή στην οποία μεταβάλλονται τα χαρακτηριστικά της εικόνας (για παράδειγμα από τα 10 στοιχεία του U θα μεταβάλλονται μόνο τα 2, από τα οποία το ένα περιέχει τα χρώματα ενός αντικειμένου που εισέρχεται μέσα στο πλάνο και το άλλο τα χρώματα της περιοχής που καλύπτει το αντικείμενο).

Για τους παραπάνω προφανείς λόγους οικονομίας χώρου και χρόνου, μετά από σύγκριση των τιμών και των μεταβολών των περισσότερων από τις συνιστώσες του διανύσματος χαρακτηριστικών για το παράδειγμά μας, επιλέχθηκαν τα τρία παραπάνω ενδεικτικά διαγράμματα, τα οποία θεωρήθηκε ότι αντιπροσωπεύουν καλύτερα τα σημεία καμπύλωσης της καμπύλης  $f(t)$  στον 30-διάστατο Ευκλείδειο χώρο. Τα υπόλοιπα ζεύγη συνιστωσών (bins) που παραλείφθηκαν είτε είχαν παρόμοιες καμπυλώσεις είτε παρουσίαζαν πολύ λίγα σημεία ενδιαφέροντος είτε παρέμεναν σχεδόν σταθερά.

Γνωρίζουμε ότι στην καμπυλότητα λαμβάνονται υπόψη εξίσου όλες οι συνιστώσες, οπότε οι μεταβολές που γίνονται σε μία συνιστώσα θα πρέπει να είναι της ίδιας κλίμακας με όλες τις υπόλοιπες μεταβολές για να αυξήσουν την καμπυλότητα, αλλιώς η μικρής κλίμακας μεταβολή ενός στοιχείου σε σχέση με τα άλλα δεν θα έχει καμία επίδραση. Επομένως αυτό που θα πρέπει να εξεταστεί είναι η μεταβολή

κατεύθυνσης της τροχιάς που συμβαίνει ταυτόχρονα σε πολλά ζεύγη απ' τις συνιστώσες του διανύσματος (όπως στα παραπάνω διαγράμματα) και η σύγκριση του σχετικού βαθμού καμπύλωσης σε σχέση με τα υπόλοιπα σημεία. Για να επιτευχθεί αυτό θα πρέπει τα διαγράμματα να είναι σχεδιασμένα σε ορθοκανονικούς άξονες, ώστε να συγκρίνονται με ανάλογο τρόπο τα σημεία καμπύλωσης σε κάθε γραφική παράσταση και να εξάγεται η σχετική καμπύλωση για τον 30-D χώρο. Παρατηρούμε τα σχήματα 4.16 – 4.18 και γνωρίζοντας για κάθε σημείο το αντίστοιχο καρέ, αναζητούμε σημεία τοπικά μεγάλης καμπύλωσης για κάθε διάγραμμα, που προκύπτουν για τα καρέ (frames) του πίνακα 4.2.

**Πίνακας 4.2 : Σημεία σημαντικής καμπύλωσης στο επίπεδο**

<b>Σημείο Καμπύλωσης</b>	<b>Διάγραμμα Y3-Y7</b>	<b>Διάγραμμα U4-U6</b>	<b>Διάγραμμα Y5-Y6</b>
<b>A</b>	486	<b>501</b>	480
<b>B</b>	<b>502</b>	519	<b>502</b>
<b>C</b>	528	545	532
<b>D</b>	545	<b>567</b>	<b>567</b>
<b>E</b>	<b>570</b>	580	596
<b>F</b>	594	609	636
<b>G</b>	617	626	651
<b>H</b>	636	644	<b>668</b>
<b>I</b>	652	<b>666</b>	688
<b>J</b>	<b>671</b>	713	708
<b>K</b>	692	–	–

Παρατηρώντας τον πίνακα 4.2 και με την παραδοχή ότι σημεία που απέχουν μέχρι και 4-5 καρέ μπορεί να αντιστοιχούν στην ίδια περιοχή καμπύλωσης, βλέπουμε ότι και τα τρία διαγράμματα παρουσιάζουν σημεία καμπύλωσης για κοινές τιμές του t (σε πολύ κοντινά καρέ όπου οι τρεις καμπύλες παρουσιάζουν αυξημένη καμπύλωση). Ορίζουμε λοιπόν περιοχές frames όπου παρατηρείται το παραπάνω φαινόμενο και θεωρούμε ότι στον 30-D χώρο η καμπύλωση πιθανόν να είναι μεγάλη (πίνακας 4.3).

**Πίνακας 4.3 :** Περιοχές μεγάλης καμπύλωσης στον 30-D Χώρο

Περιοχή καμπύλωσης N <sup>ο</sup>	Περιοχή των frames
1	501 – 502
2	567 – 570
3	666 – 671

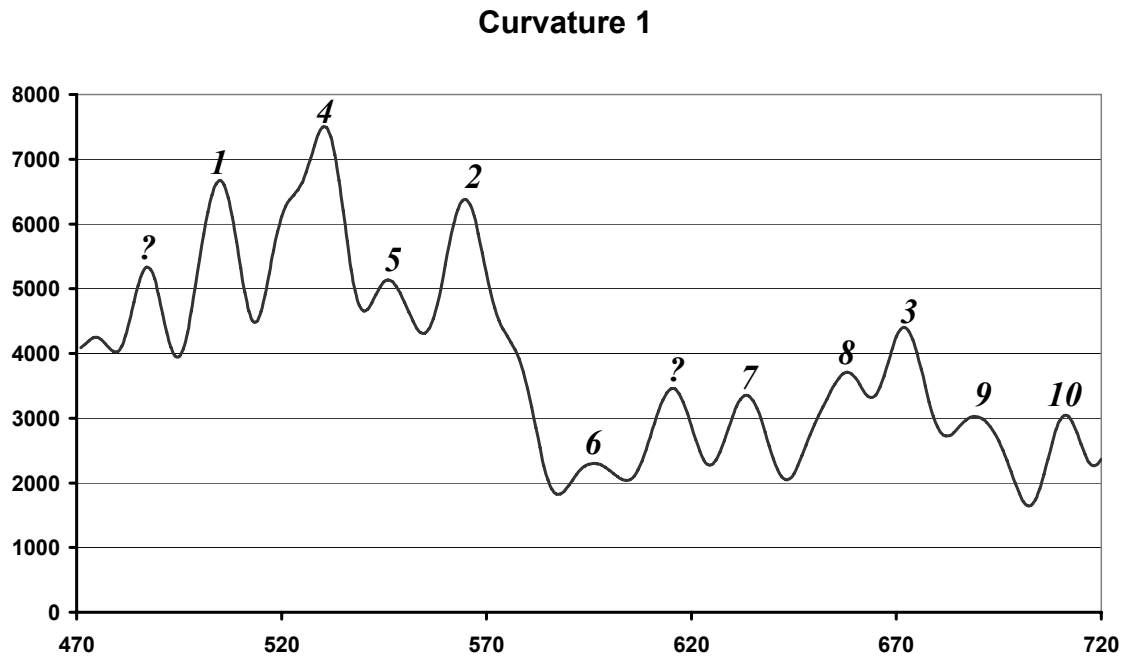
Παρατηρώντας επίσης τον πίνακα 4.2 βρίσκουμε κοινά σημεία καμπύλωσης εξετάζοντας ανά δύο τα τρία διαγράμματα. Ορίζουμε τότε περιοχές frames όπου έχουμε τέτοια κοινά σημεία σε δύο διαγράμματα και θεωρούμε ότι και στον 30-D χώρο θα υπάρχει καμπύλωση, που δεν θα είναι όμως το ίδιο μεγάλη όπως προηγουμένως, αλλά μέτριου βαθμού καμπύλωση (βλέπε πίνακα 4.4).

**Πίνακας 4.4 :** Περιοχές μέτριας καμπύλωσης στον 30-D Χώρο

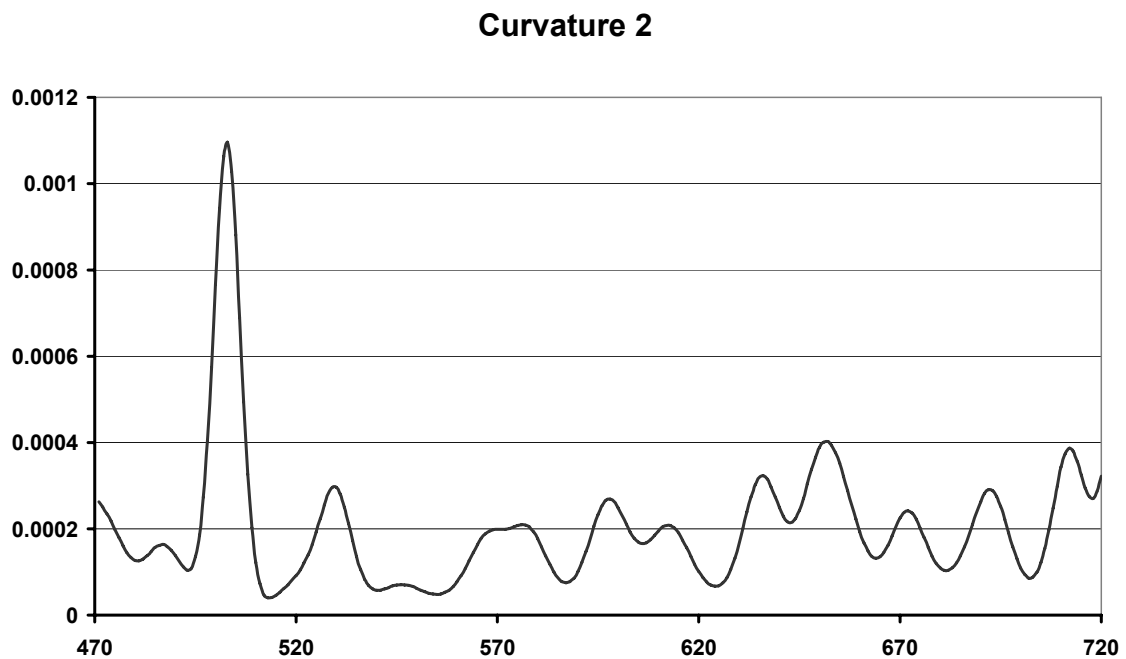
Περιοχή καμπύλωσης N <sup>ο</sup>	Περιοχή των frames
4	528 – 532
5	545
6	594 – 596
7	636
8	651 – 652
9	688 – 692
10	708 – 713

Από τους τελευταίους δύο πίνακες (4.3 και 4.4) προκύπτουν τα πιθανά σημεία καμπύλωσης της καμπύλης του διανύσματος χαρακτηριστικών στον πολυδιάστατο χώρο, όπως προκύπτουν από τις προβολές της σε τρία επίπεδα όπου φαίνονται κάποια χαρακτηριστικά σημεία. Τα ονομάζουμε πιθανά, διότι δεν λαμβάνουμε υπόψη τις υπόλοιπες συνιστώσες. Παρόλ' αυτά, περιμένουμε ότι οι μέθοδοι της καμπυλότητας θα δίνουν αποτέλεσμα τουλάχιστον για τα σημεία στα οποία φαίνεται ότι η προβολή της καμπύλης ταυτόχρονα και στα τρία επίπεδα του συστήματος συντεταγμένων αλλάζει απότομα κατεύθυνση (περιοχές 1-3, από τον πίνακα 4.2).

Στα σχήματα 4.19 – 4.21 δίνεται το διάγραμμα της καμπυλότητας σε συνάρτηση με τον χρόνο (frames), όπως υπολογίζεται από τους τύπους 1, 2 και 4 αντίστοιχα (η υλοποίηση των μεθόδων έγινε σε γλώσσα προγραμματισμού C).



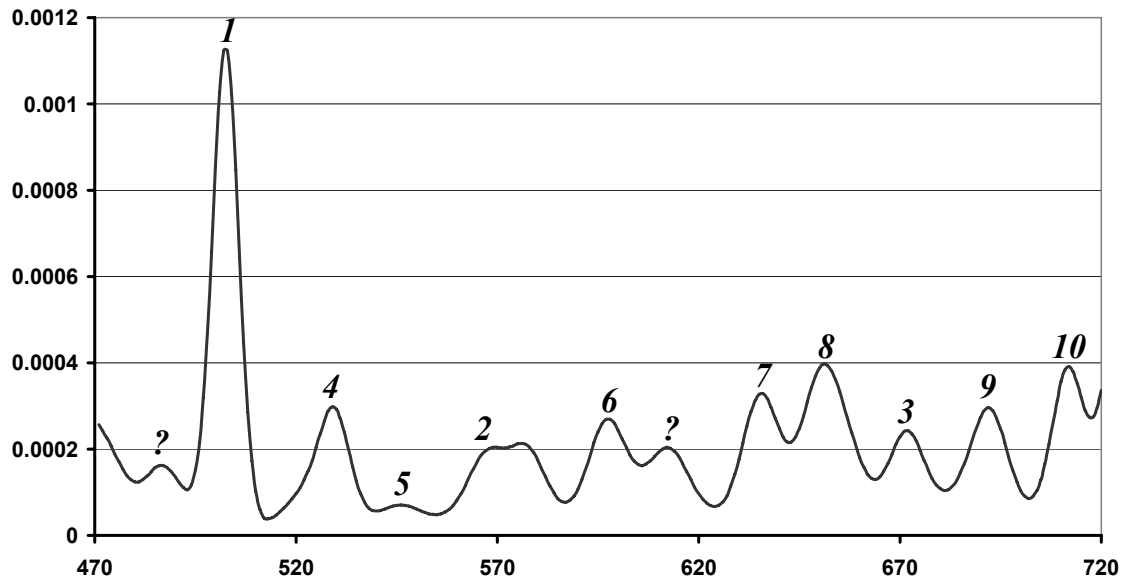
Σχ. 4.19 : 30-D Καμπυλότητα - Τύπος 1 (2<sup>η</sup> παράγωγος ως προς t)



Σχ. 4.20 : 30-D Καμπυλότητα - Τύπος 2 (2<sup>η</sup> παράγωγος ως προς s)



### Curvature 4



Σχ. 4.21 : 30-D Καμπυλότητα - Τύπος 4 (Γενικευμένη καμπυλότητα)

Η πρώτη παρατήρηση που κάνουμε είναι ότι τα διαγράμματα για τους τύπους 2 και 4 είναι και σε αυτή την περίπτωση όμοια, γεγονός που αποδεικνύει την ισχύ της γενικότητας του τύπου 4. Από εδώ και στο εξής θα χρησιμοποιούμε τον τύπο 4 ως ορισμό της καμπυλότητας στον N-διάστατο Ευκλείδειο χώρο και για έναν ακόμη λόγο: μέσω της θεωρίας των γενικευμένων καμπυλοτήτων ορίζονται N-1 καμπυλότητες, όπου οι N-2 τελευταίες μπορούν να χρησιμοποιηθούν επιπλέον της πρώτης (η πρώτη δίνεται από τον τύπο 4) για τον έλεγχο της αλλαγής καμπύλωσης μιας N-διάστατης καμπύλης.

Συγκρίνοντας τις γραφικές παραστάσεις των σχημάτων 4.19 και 4.21 (για το διάστημα των frames 471 - 720), μπορούμε καταρχήν να παρατηρήσουμε ότι ο τύπος 4 δίνει ένα μέγιστο που ξεχωρίζει από τα υπόλοιπα, αλλά και τοπικά μέγιστα σε κλίμακα περίπου στο ένα τέταρτο του ολικού μεγίστου, ενώ ο τύπος 1 δεν παρουσιάζει ολικό μέγιστο που να διακρίνεται από τα υπόλοιπα, αλλά έχει τοπικά μέγιστα με τιμές στην περιοχή του ολικού μεγίστου (για τα frames 471-580), και τοπικά μέγιστα με τιμές περίπου στο μισό του ολικού μεγίστου (για τα frames 581-720).

Στα διαγράμματα αυτά έχουν σημειωθεί τοπικές αιχμές (peaks) που αντιστοιχούν σε περιοχές καμπύλωσης στον 30-D χώρο και όπως φαίνεται όλες οι περιοχές που ορίζονται στους πίνακες 4.2 και 4.3 (περιοχές 1 έως 10) παριστάνονται με κάποιο μέγιστο στις καμπύλες και των δύο διαγραμμάτων. Υπάρχουν επίσης κάποιες αιχμές

(σημειώνονται με ερωτηματικό ?), οι οποίες δεν αντιστοιχούν σε καμπύλωση που έχουμε ορίσει και εμφανίζονται σε κοινά καρέ και στα δύο διαγράμματα (υποθέτουμε ότι οφείλονται στη μεταβολή συνιστωσών που δεν εξετάσαμε).

Παρατηρούμε στο πρώτο διάγραμμα ότι το σημείο 4 που αντιστοιχεί στο σημείο C της καμπύλης του σχ. 4.16 και στο σημείο C της καμπύλης του σχ. 4.18, έχει καμπυλότητα μεγαλύτερη από τα σημεία 1 και 2 (σύμφωνα με τον τύπο 1). Από τα δεδομένα που έχουμε γνωρίζουμε ότι στα σημεία 1 και 2 καμπυλώνουν και οι τρεις προβολές, ενώ στο σημείο 4 καμπυλώνουν μόνο οι δύο. Περιμένουμε λοιπόν μικρότερη καμπυλότητα στο σημείο 4 απ' ότι στα σημεία 1-2, το οποίο φαίνεται να διορθώνεται κάπως στο δεύτερο διάγραμμα όπου τα σημεία 4 και 2 είναι κοντά, ενώ το σημείο 1 έχει μεγαλύτερη τιμή.

Τα σημεία ? δεν αντιστοιχούν σε περιοχή καμπύλωσης που ήδη έχουμε ορίσει. Το πρώτο σημείο αντιστοιχεί περίπου στο frame 490, ενώ το δεύτερο στο frame 610. Για το σημείο 490 (βλέπε και πίνακα 4.1), παρατηρούμε ότι προσεγγίζει τα σημεία A του σχ. 4.16 και του σχ. 4.18, όπου οι δύο προβολές καμπυλώνουν ελαφρώς, άρα είναι λογική η αύξηση της καμπυλότητας. Για το σημείο 610 (βλέπε και πίνακα 4.1), παρατηρούμε ότι βρίσκεται κοντά στο σημείο G του σχ. 4.16 και στο σημείο F του σχ. 4.17, όπου οι δύο προβολές καμπυλώνουν απότομα, οπότε θα αυξάνει και η καμπυλότητα. Για το δεύτερο σημείο ? αναμένουμε μεγαλύτερη αιχμή στην καμπυλότητα (όπως φαίνεται στον τύπο 4 και όχι στον τύπο 1).

Από τα παραπάνω μοντέλα (δύο θεωρητικά 2-D, 3-D και ένα πραγματικό 30-D) και εφόσον δεν μπορούμε να ελέγξουμε την καμπύλωση σε όλες τις προβολές για το 30-D, βασιζόμαστε στα εξής:

1. Ο τύπος 1 για το 2-D μοντέλο δίνει εσφαλμένα αποτελέσματα διότι δεν ξεχωρίζει τον βαθμό καμπύλωσης σε διαφορετικά σημεία. Το ίδιο ισχύει και για το 3-D μοντέλο, όπου αγνοούνται σημεία με σημαντική καμπύλωση (υποβαθμίζεται πολύ η καμπυλότητά τους).
2. Ο τύπος 4 δίνει εμφανώς βελτιωμένα αποτελέσματα για τα δύο μοντέλα 2-D και 3-D, όπως αποδείχθηκε παραπάνω. Επίσης η καμπυλότητα που βασίζεται στον τύπο 4 προκύπτει ίδια με τον τύπο 3, ενώ ο τύπος 2 είναι παρόμοιας μορφής, για τα 2-D, 3-D και 30-D μοντέλα.
3. Στο 30-D μοντέλο παρατηρώντας τις προβολές της καμπύλης σε τρία επίπεδα και το διάγραμμα της καμπυλότητας (τύποι 1 και 4), βλέπουμε ότι τα σημαντικά

σημεία καμπύλωσης αντιστοιχούν σε αιχμές στα δύο διαγράμματα καμπυλότητας, οπότε καμία μέθοδος δεν αγνοεί κάποιο από αυτά τα σημεία.

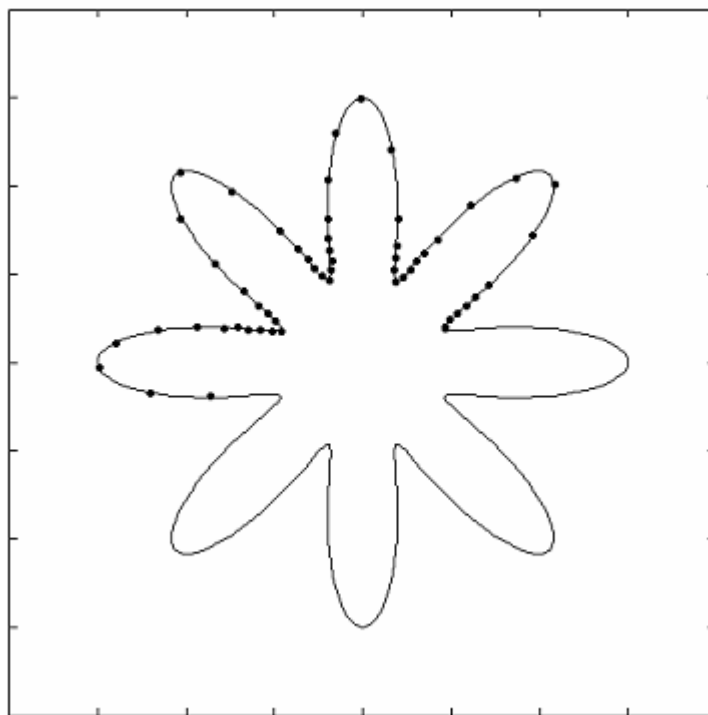
4. Γενικά οι διαφορές μεταξύ του τύπου 1 και του τύπου 4 δεν είναι στα σημεία (καρέ) όπου παρουσιάζονται τα μέγιστα, αλλά στην τιμή που έχουν αυτά τα μέγιστα. Όπως έχουμε ήδη αναφέρει, ο υπολογισμός του  $\text{curvature}$  εξαρτάται από την ταχύτητα που διαγράφεται η καμπύλη (χρονική μεταβολή – παράγωγος ως προς  $t$ ), η οποία και καθορίζει το μέτρο της καμπυλότητας. Γι' αυτό και αναζητούμε τον ορισμό της καμπυλότητας που περιγράφει την καμπύλη ανεξάρτητα από την ταχύτητα του διανύσματος (παράγωγος ως προς  $s$ ).

Από τις παραπάνω επισημάνσεις καταλήγουμε στο συμπέρασμα ότι η θεωρία της γενικευμένης καμπυλότητας και ο τύπος 4 (θα συμβολίζεται C4) ισχύουν στον διδιάστατο, τρισδιάστατο και πολυδιάστατο χώρο, γι' αυτό και υιοθετούνται ως ορισμός της καμπυλότητας για την εκτίμηση των σημείων “ακραίας” συμπεριφοράς (τοπικά μεγάλη μεταβολή της καμπύλωσης). Ο τύπος 1 (θα συμβολίζεται C1) βλέπουμε ότι δεν επιτυγχάνει στην περίπτωση των θεωρητικών 2-D και 3-D, αλλά στις προβολές της 3D-D καμπύλης υποδεικνύει τα περισσότερα σημεία όπου συμβαίνουν ακραίες μεταβολές.

Η μέθοδος C1 δίνει την επιτάχυνση της χρονικής μεταβολής του περιεχομένου της εικόνας (μεταβολή από καρέ σε καρέ), οπότε σε γρήγορες αλλαγές του ιστογράμματος θα έχουμε τοπικά μέγιστα, αλλά όταν οι αλλαγές είναι αργές το μέγιστο θα έχει πολύ μικρή τιμή. Μία όμοια αλλαγή (ίδια μορφή της ακολουθίας των σημείων στον χώρο) θα δίνει μεγαλύτερη καμπυλότητα, αν συμβαίνει ταχύτατα (σε λίγα καρέ).

Αυτό διορθώνεται με την μέθοδο C4, όπου θεωρείται η καμπύλωση με βάση την μετατόπιση  $ds$  πάνω στην καμπύλη. Η καμπύλωση εκφράζεται ανεξάρτητα από την ταχύτητα, αλλά με βάση τη μετακίνηση του διανύσματος χαρακτηριστικών στον χώρο των χαρακτηριστικών (μεταβολή του περιεχομένου), οπότε ολοκληρωτική απότομη αλλαγή στο περιεχόμενο θα δώσει μεγάλη τιμή στην καμπυλότητα, ακόμα και αν συμβαίνει αργά (σε διάρκεια πολλών καρέ). Δηλαδή όμοιες αλλαγές, με ίδια μορφή στον χώρο των χαρακτηριστικών, θα δίνουν την ίδια καμπυλότητα. Για την εξήγηση των παραπάνω βλέπε σχήμα 4.22. Το μειονέκτημα είναι ότι οι “στοιχειώδεις” μετατοπίσεις  $ds$  δεν είναι ίσες (όπως είναι τα  $dt = 1$ ) ούτε πραγματικά στοιχειώδεις, με αποτέλεσμα να έχουμε τις προσεγγιστικές τιμές για την καμπυλότητα C4 (αυτό θα δίνει παραμορφωμένες τιμές στη καμπυλότητα, που όμως μπορούμε να δεχτούμε ότι

προσεγγιστικά είναι σωστές). Θα είμαστε επιφυλακτικοί με τα αποτελέσματα, τόσο για τον τύπο C1 όσο και για τον τύπο C4.



**Σχ. 4.22:** Διδιάστατη καμπύλη με διαφορετικές πυκνότητες σημείων-καρέ στις περιοχές χαρακτηριστικών σημείων (καμπύλες)

## ΚΕΦΑΛΑΙΟ 5

# *Άλλες Μέθοδοι για την Εξαγωγή*

## *Χαρακτηριστικών Σημείων*

### *Καμπύλης*

Στο κεφάλαιο αυτό παρουσιάζονται δύο εναλλακτικοί τρόποι για τον χειρισμό μιας καμπύλης στον πολυδιάστατο Ευκλείδειο χώρο και την εξαγωγή των σημείων που παρουσιάζουν ακραία συμπεριφορά, με αποτέλεσμα τη δημιουργία μίας περίληψης της μορφής της καμπύλης. Οι μέθοδοι αυτοί βασίζονται σε δύο έρευνες, από τις οποίες η πρώτη αναφέρεται στην απλοποίηση της καμπύλης (Curve Simplification) μέσω της προβολής σε άξονες αναφοράς (στον πολυδιάστατο χώρο των χαρακτηριστικών), οι οποίοι περνούν από τα ενδιαφέροντα σημεία που τελικά εξάγονται με επαναληπτικό τρόπο και με κάποιο κριτήριο σύγκλισης. Η δεύτερη μέθοδος, που βασίζεται στην ανάλυση των τοπικών χαρακτηριστικών, δηλαδή στη σύγκριση των αποστάσεων και την επιλογή σημείων που προσεγγίζουν όσο το δυνατόν καλύτερα το περίγραμμα της καμπύλης (Shape Contour Approximation), προέρχεται από τον κλάδο της σύγκρισης ομοιότητας μεταξύ σχημάτων (Shape Similarity).

Στην εργασία αυτή επιθυμούμε να συγκρίνουμε τις μεθόδους της καμπυλότητας με άλλες υπάρχουσες μεθόδους εξαγωγής χαρακτηριστικών καρέ. Ο βασικός λόγος που επιλέχθηκαν οι δύο μέθοδοι του Κεφαλαίου αυτού είναι το ότι αναλύουν πολυδιάστατες καμπύλες και εξάγουν τα χαρακτηριστικά τους σημεία. Τα σημεία αυτά μπορούν να

συγκριθούν με τα σημεία που δίνουν οι μέθοδοι της καμπυλότητας και μάλιστα αντιστοιχούν στα τοπικά μέγιστα της καμπυλότητας. Έτσι γίνεται εύκολη η αντιπαράθεση των αποτελεσμάτων και η άμεση σύγκριση των σημείων επάνω στην καμπύλη προσφέρει έδαφος για την εξαγωγή συμπερασμάτων σχετικά με τις επιδόσεις των μεθόδων μας.

Όσον αφορά στη μέθοδο απλοποίησης καμπύλης, αυτή έχει βασιστεί σε μαθηματική θεωρία διαμέρισης επίπεδης καμπύλης. Οι συγγραφείς χρησιμοποιούν αλγόριθμο για τη διαμέριση της καμπύλης και τον προσαρμόζουν για την απλοποίηση  $N$ -διάστατης καμπύλης σε  $n$  διαστάσεις ( $n < N$ ), συγκεκριμένα για εφαρμογή σε βίντεο. Στην παρούσα εργασία υιοθετείται η ίδια μεθοδολογία, αλλά καθορίζεται συγκεκριμένο  $n$  ανάλογα με το  $N$ , ενώ το κριτήριο σύγκλισης  $\varepsilon$  (παράμετρος λεπτομέρειας) ορίζεται ως ένα ποσοστό της μέγιστης προβολής σε άξονα.

Η μέθοδος της προσέγγισης περιγράμματος σχήματος έχει υλοποιηθεί με σκοπό τη σύγκριση της μορφής μεταξύ αντικειμένων (δύο διαστάσεις). Γι' αυτό εδώ, αφού περιγραφεί αναλυτικά καθένας από τους αλγορίθμους που προτείνουν οι συγγραφείς, δίνεται η γενίκευση της μεθόδου σε περισσότερες διαστάσεις μέσω του ορισμού απόστασης σημείου από ευθεία στον  $N$ -διάστατο Ευκλείδειο χώρο.

Οι δύο παραπάνω μέθοδοι δεν εξάγουν αυτόνομα τα σημεία της καμπύλης, αλλά απαιτείται εκ των προτέρων ο καθορισμός του κριτηρίου σύγκλισης και του αριθμού των χαρακτηριστικών σημείων αντίστοιχα. Δεν υπάρχουν προκαθορισμένες τιμές που να επιτρέπουν στις μεθόδους να λειτουργούν πάντοτε ικανοποιητικά. Για να γίνει η επιθυμητή σύγκριση των μεθόδων, επιλέγεται το ποσοστό για την πρώτη μέθοδο ώστε να εξάγεται αριθμός σημείων περίπου στον μέσο όρο των σημείων των μεθόδων καμπυλότητας. Ο αριθμός των σημείων της μεθόδου προσέγγισης περιγράμματος επιλέγεται περίπου ίσος με τον μέσο όρο των υπολοίπων.

## 5.1 Η Μέθοδος Απλοποίησης Καμπύλης

### 5.1.1 Εισαγωγικά

Το βίντεο όπως έχουμε αναφέρει μπορεί να παρασταθεί από την τροχιά (καμπύλη) του διανύσματος χαρακτηριστικών στον πολυδιάστατο χώρο των χαρακτηριστικών. Η ανάλυση της καμπύλης αυτής, όπως προτείνεται από τους DeMenthlon et al. ([DeM98]), μπορεί να γίνει με εργαλεία που έχουν αναπτυχθεί για επίπεδες καμπύλες, όπως είναι ο κλασικός δυαδικός αλγόριθμος διαμέρισης καμπύλης, προσαρμοσμένος για τις πολλές διαστάσεις. Ελέγχοντας τη διάσταση και το κριτήριο σύγκλισης για κάθε μέρος της καμπύλης που προκύπτει από τη διαμέριση, αυτό μπορεί να απλοποιηθεί και να αναπαρασταθεί σε ένα χώρο μικρότερης διάστασης. Έτσι η καμπύλη των χαρακτηριστικών αναπαρίσταται με μία δομή δένδρου, με τα μέρη της διαμέρισης να βρίσκονται στα επίπεδα που αντιστοιχούν σε διαφορετικές διαστάσεις. Οι σύνδεσμοι μεταξύ αυτών των μερών λαμβάνονται ως ενδιαφέροντα σημεία, δηλαδή ως keyframes, ενώ το κριτήριο σύγκλισης καθορίζει τον βαθμό της περίληψης της ακολουθίας των εικόνων (δηλαδή την λεπτομέρεια).

Το διάνυσμα των χαρακτηριστικών σε συνάρτηση με τον χρόνο αντιστοιχεί σε σημεία στον πολυδιάστατο χώρο, τα οποία σχηματίζουν μία τροχιά από ευθύγραμμα τμήματα και καμπύλες, μικρές ή μεγαλύτερες. Απλοποιώντας την καμπύλη σε μικρότερη διάσταση, ουσιαστικά αντικαθιστούμε ευθύγραμμα τμήματα ή μικρές καμπύλες από ένα τμήμα που ανήκει σε μικρή διάσταση, ενώ για κομμάτια με σημεία μεγάλης καμπύλωσης δεν αρκεί ένα τμήμα, αλλά πρέπει το κομμάτι αυτό να διαμεριστεί σε περισσότερα τμήματα μικρής διάστασης. Η μέθοδος ([DeM98]) προτείνει την διαμέριση της καμπύλης και τον εντοπισμό των τμημάτων μικρότερης διάστασης που την αποτελούν, οπότε ως σημαντικά σημεία λαμβάνονται τα άκρα αυτών των τμημάτων, χωρίς να απαιτείται ο ακριβής εντοπισμός των συνδέσμων του δένδρου που αναφέρθηκε προηγουμένως.

Επίσης, εντοπίζοντας τα τμήματα αυτά σε διάφορα επίπεδα, δημιουργούμε ένα δυαδικό δένδρο, στο οποίο κάθε διακλάδωση δημιουργεί δύο “παιδιά” που αποτελούν μία πιο λεπτομερή περιγραφή του “πατέρα” τους. Άλλο ένα πλεονέκτημα της μεθόδου είναι το γεγονός ότι απλοποιώντας την καμπύλη σε χαμηλότερη διάσταση, διευκολύνεται η τρισδιάστατη απεικόνιση, διότι η απεικόνιση για παράδειγμα 15D καμπύλης σε 3D είναι πιο αξιόπιστη από την απεικόνιση 1800D σε 3D.

### 5.1.2 Επαναληπτικός Αλγόριθμος Διαμέρισης N-D Καμπύλης

Οι τεχνικές για την προσέγγιση ή την απλοποίηση μιας καμπύλης γραμμής, που έχουν ευρέως αναπτυχθεί για δύο διαστάσεις (επίπεδο), έχουν σκοπό να συνοψίσουν το γενικό σχήμα με μερικά σημαντικά σημεία. Η μεταβολή της επίπεδης καμπύλης μεταξύ αυτών των σημείων είναι μικρή και θεωρείται ότι περιέχει θόρυβο, γι' αυτό και δεν χρειάζεται να ασχοληθούμε με το περιεχόμενο των καρέ που αντιστοιχούν σε αυτό το τμήμα, οπότε προσεγγίζουμε την καμπύλη με τα ευθύγραμμα τμήματα (αποτελούνται από μία διάσταση) που ενώνουν τα σημεία ενδιαφέροντος.

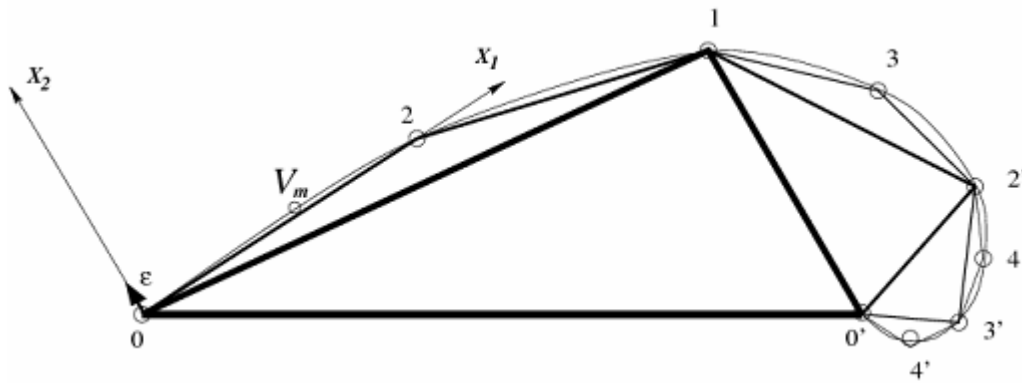
Πολλές από τις ιδέες που έχουν αναπτυχθεί για τις επίπεδες καμπύλες μπορούν να επεκταθούν και στις περισσότερες διαστάσεις για την εφαρμογή στο διάνυσμα χαρακτηριστικών, όπως η μεθοδολογία που αναφέρθηκε στην προηγούμενη παράγραφο. Η πιο γνωστή μέθοδος για τις επίπεδες καμπύλες είναι ο επαναληπτικός δυαδικός αλγόριθμος διαμέρισης καμπύλης, που προτάθηκε από τον U. Ramer ([Ram72]) το 1972 για την επεξεργασία εικόνας και αναφέρθηκε επίσης από τους D.H. Douglas και T.K. Peucker το 1973 για χαρτογραφικές εφαρμογές ([DPe73]). Ο αλγόριθμος αποτελεί το κατάλληλο μαθηματικό εργαλείο, αλλά και την καλύτερη μέθοδο από την πλευρά της οπτικής αξιοπιστίας του αποτελέσματος, για την αναπαράσταση της καμπύλης, σύμφωνα με τους J. Hersberger και J. Snoeyink ([Her92]).

Σύμφωνα με τους τελευταίους συγγραφείς, ο αλγόριθμος συνήθως περιγράφεται ως εξής: Για να προσεγγίσουμε μία αλυσίδα σημείων από το  $V_i$  έως το  $V_j$ , ξεκινάμε από το ευθύγραμμο τμήμα μεταξύ των  $V_i$  και  $V_j$ . Αν το μακρύτερο σημείο  $V_m$  (της αλυσίδας) απέχει από το ευθύγραμμο τμήμα  $V_iV_j$  απόσταση  $d_m$  μικρότερη από μία παράμετρο λεπτομέρειας  $\epsilon$ , τότε θεωρούμε ότι το ευθύγραμμο τμήμα προσεγγίζει την αλυσίδα των σημείων μεταξύ των  $V_i$  και  $V_j$  (κριτήριο σύγκλισης). Διαφορετικά, η αλυσίδα χωρίζεται στο σημείο  $V_m$  σε δύο τμήματα  $V_iV_m$  και  $V_mV_j$ , στα οποία εφαρμόζεται επαναληπτικά ο ίδιος αλγόριθμος (βλέπε σχ. 5.1).

Για επίπεδη καμπύλη (όπως για παράδειγμα αυτή του σχήματος 5.1), ο αλγόριθμος εκφράζεται επίσης και με την κατασκευή τοπικού συστήματος συντεταγμένων (πλαίσιο αναφοράς). Για να προσεγγίσουμε το διδιάστατο τμήμα  $V_iV_j$  με μονοδιάστατα τμήματα, κατασκευάζουμε ένα πλαίσιο αναφοράς με πρώτο άξονα τον  $V_iV_j$  και δεύτερο άξονα τον κάθετο σε αυτόν. Βρίσκουμε το σημείο  $V_m$  που έχει το



μεγαλύτερο μέτρο στην δεύτερη συντεταγμένη στο τοπικό αυτό σύστημα. Αν η τιμή αυτή είναι μικρότερη από το  $\varepsilon$ , τότε η δεύτερη διάσταση σε αυτό το σύστημα αναφοράς μπορεί να αγνοηθεί και η διδιάστατη γραμμή να προσεγγιστεί με την μονοδιάστατη γραμμή  $V_i V_j$  (προβολή σε έναν άξονα). Αυτό δείχνει την τοπική ελάττωση της διάστασης από 2D σε 1D.

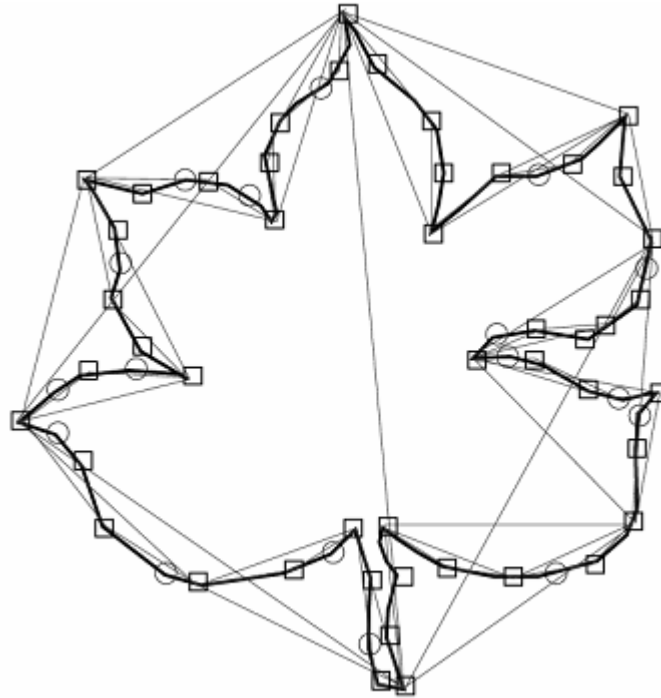


**Σχ. 5.1 :** Παράδειγμα εφαρμογής του επαναληπτικού δυαδικού αλγορίθμου διαμέρισης σε μια επίπεδη καμπύλη C ([DeM98])

Στο σχήμα 5.1 φαίνεται η διαδικασία που περιγράφηκε παραπάνω για το τμήμα  $(0, 2)$ , στο οποίο κατασκευάζεται το τοπικό ευκλείδειο σύστημα αναφοράς  $(Ox_1, Ox_2)$  και εντοπίζεται το σημείο  $V_m$  που έχει τη μεγαλύτερη απόλυτη τιμή για την συντεταγμένη  $x_2$  (η οποία είναι μικρότερη από  $\varepsilon$ , οπότε η καμπύλη δεν διαμερίζεται πλέον). Παρατηρούμε ότι αρχικά το τμήμα  $(0,0')$ , δηλαδή (αρχή, τέλος) της καμπύλης, διαμερίζεται στα τμήματα  $(0,1)$  και  $(1,0')$ , όπου το 1 είναι το σημείο με την μεγαλύτερη τεταγμένη (1<sup>η</sup> διαμέριση). Κατόπιν βρίσκονται τα σημεία που έχουν τη μεγαλύτερη τεταγμένη στα δύο τοπικά πλαίσια αναφοράς, που είναι τα 2 και 2' (2<sup>η</sup> διαμέριση). Από τα νέα τμήματα που δημιουργούνται  $(0,2)$ ,  $(2,1)$ ,  $(1,2')$  και  $(2',0')$ , τα δύο πρώτα δεν διαμερίζονται περαιτέρω, ενώ τα δύο άλλα χωρίζονται στα σημεία 3 και 3' (3<sup>η</sup> διαμέριση). Στην επόμενη διαμέριση χωρίζονται μόνο τα τμήματα  $(2',3')$  και  $(3',0')$  στα σημεία 4 και 4' (4<sup>η</sup> διαμέριση), ενώ άλλη διαμέριση δεν απαιτείται. Έτσι η καμπύλη C του σχ. 5.1 προσεγγίζεται από τα σημεία 0, 2, 1, 3, 2', 4, 3', 4', 0', τα οποία συνοψίζουν το περίγραμμά της, αφού είναι περισσότερα στο κομμάτι όπου καμπυλώνει περισσότερο και λιγότερα στο τμήμα 0 έως 1, όπου είναι λιγότερο καμπυλόγραμμα.

Στο σχήμα 5.2 φαίνεται μία πιο πολύπλοκη επίπεδη καμπύλη, στην οποία έχει εφαρμοστεί ο επαναληπτικός δυαδικός αλγόριθμος διαμέρισης μέχρι ένα συγκεκριμένο

επίπεδο λεπτομέρειας, δηλαδή μέχρι κάποιο βάθος του δυαδικού δένδρου. Στο σχήμα φαίνονται τα τμήματα που δημιουργούνται σε κάθε επίπεδο διαμέρισης, ενώ με τετράγωνο σημειώνονται όλα τα σημεία που έχουν προκύψει μετά από κάθε διαμέριση σε κάθε διακλάδωση του δένδρου. Με κύκλο σημειώνονται τα σημεία που θα προστεθούν στο επόμενο επίπεδο λεπτομέρειας (επόμενη διακλάδωση), περιγράφοντας με μεγαλύτερη αναλυτικότητα το σχήμα της καμπύλης.



**Σχ. 5.2 :** Απλοποίηση σύνθετης καμπύλης μέσω του δυαδικού αλγορίθμου διαμέρισης καμπύλης σε δύο επίπεδα λεπτομέρειας – επίπεδα δυαδικού δένδρου ([DeM98])

Ο Ramer ανέφερε την δυνατότητα του αλγορίθμου να εφαρμοστεί και σε καμπύλες που ορίζονται σε πολυδιάστατο χώρο ([Ram72]), ελαττώνοντας την διάστασή τους από  $N$ -D σε 1D. Μία προτιμότερη προσαρμογή του αλγορίθμου για τον  $N$ -διάστατο χώρο προτείνεται από τους DeMenthlon et al. ([DeM98]), και αφορά στην απλοποίηση της καμπύλης σε τμήματα μικρότερης διάστασης  $n$  ( $n < N$ ) και όχι σε μία διάσταση. Τα μέτρα των συντεταγμένων για τις διαστάσεις που αγνοούνται είναι μικρότερα από τα μέτρα των συντεταγμένων για τις διαστάσεις που παραμένουν και μικρότερα από την παράμετρο  $\epsilon$ . Αυτή η γενίκευση του αλγορίθμου έχει ενδιαφέρον διότι επιτρέπει την απλοποίηση της καμπύλης του διανύσματος χαρακτηριστικών ( $N$ -D) σε χώρο μικρότερων διαστάσεων ( $n$ -D), όπου φαίνονται οι μεταβολές στο περιεχόμενο του βίντεο

(κίνηση κάμερας, ζουμ, εισερχόμενα αντικείμενα), από μια απλοποίηση σε χώρο μίας διάστασης (μία ευθεία γραμμή), όπου το μέτρο ενός μόνο μεγέθους (1D) δεν μπορεί να περιγράψει πλήρως τις μεταβολές στο περιεχόμενο του βίντεο.

### 5.1.3 Περιγραφή του Αλγορίθμου

Στον επαναληπτικό δυαδικό αλγόριθμο διαμέρισης πολυδιάστατης καμπύλης (γενίκευση του αλγορίθμου όπως προαναφέρθηκε [DeM98]), για ένα τμήμα καμπύλης  $V_i V_j$  κατασκευάζεται ένα τοπικό σύστημα συντεταγμένων με  $n$  ορθογώνιους άξονες. Έστω ότι το διάνυσμα χαρακτηριστικών είναι ορισμένο σε  $N$ -διάστατο Ευκλείδειο χώρο, όπου διαγράφει μία  $N$ -D καμπύλη. Θα απλοποιήσουμε την καμπύλη αυτή σε έναν  $n$ -D χώρο μικρότερης διάστασης, όπου το  $n$  δίνεται από τη σχέση:

$$n = \left\lfloor \frac{N-1}{10} + 1 \right\rfloor = \left\lfloor \frac{N+9}{10} \right\rfloor \quad (\text{στην περίπτωση } N = 30 \text{ προκύπτει } n = 3)$$

Ως σημείο εκκίνησης επιλέγεται το  $V_i$ , ενώ ο πρώτος άξονας  $A_1$  επιλέγεται να είναι ο άξονας που ενώνει τα  $V_i$  και  $V_j$  με κατεύθυνση από το  $V_i$  προς το  $V_j$  (βλέπε σχ. 5.3). Στη συνέχεια, μόλις κατασκευαστεί διαδοχικά κάθε άξονας  $A_k$  ( $k = 1, \dots, n$ ), υπολογίζεται το μοναδιαίο διάνυσμα  $a_k$  που αντιστοιχεί σε αυτόν, που για τον  $A_1$  είναι:

$$\hat{a}_1 = \frac{\overline{V_i V_j}}{\left| \overline{V_i V_j} \right|} \quad (\text{εξ. 5.1})$$

Στη συνέχεια υπολογίζεται η συντεταγμένη  $x_{sk}$  κάθε σημείου  $V_s$  ( $s = i, \dots, j$ ) της καμπύλης με χρήση του εσωτερικού γινομένου:

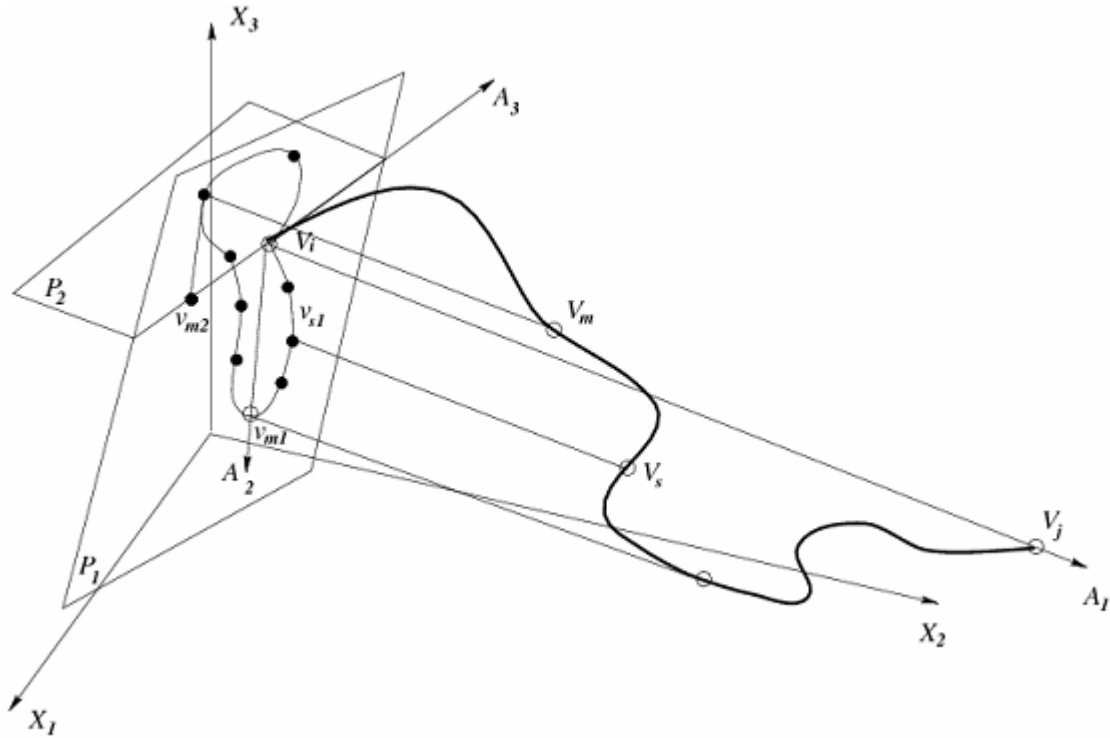
$$x_{sk} = \overline{V_i V_s} \cdot \hat{a}_k \quad (\text{εξ. 5.2})$$

όπου το  $a_k$  είναι το μοναδιαίο διάνυσμα που μόλις κατασκευάστηκε.

Έτσι έχω όλες τις προβολές  $x_{sr}$  των σημείων ( $s = i, \dots, j$ ) σε όλους τους άξονες που έχουν μέχρι τώρα κατασκευαστεί ( $r = 1, \dots, k$ ).

Για να βρεθεί ένας νέος άξονας ή για να αποφασιστεί αν θα πρέπει να διαμεριστεί το τμήμα της καμπύλης  $V_i V_j$  (όταν έχουν κατασκευαστεί όλοι οι άξονες), υπολογίζεται η τετραγωνική απόσταση (ευκλείδεια νόρμα)  $V_i P_{sk}^2$  από το σημείο  $V_i$  προς την προβολή

( $P_{sk}$ ) κάθε σημείου  $V_s$  της καμπύλης στο υπερεπίπεδο που είναι κάθετο στον  $k$  τάξης άξονα (ο τελευταίος που κατασκευάστηκε). Το υπερεπίπεδο δεν είναι ένα επίπεδο με δύο διαστάσεις (συντεταγμένες), αλλά έχει συνολικά  $N-k$  διαστάσεις (δηλαδή από τις  $N$  αφαιρούμε τις διαστάσεις που ορίζουν οι  $k$  άξονες).



**Σχ. 5.3 :** Διαμέριση καμπύλης μέσω του δυαδικού αλγορίθμου διαμέρισης καμπύλης (εφαρμογή απλοποίησης σε τρισδιάστατο χώρο  $n=3$ ) ([DeM98])

Οι αποστάσεις των προβολών αυτών προκύπτουν από το Πυθαγόρειο Θεώρημα και καθεμιά ισούται με την απόσταση του  $V_i$  από το  $V_s$  μείον το άθροισμα των τετραγώνων όλων των συντεταγμένων του σημείου  $V_s$  για όλους τους άξονες 1 έως  $k$ :

$$V_i P_{sk}^2 = V_i V_s^2 - \sum_{r=1}^k x_{sr}^2 \quad \forall s, \quad s \in \{i, \dots, j\} \quad (\text{εξ. 5.3})$$

Επιλέγουμε το  $s = m$  το οποίο δίνει την μεγαλύτερη απόσταση  $V_i P_{mk}^2$ , έστω ότι αυτή είναι η  $d_m$ , καθώς και το αντίστοιχο σημείο της καμπύλης  $V_m$ , με προβολή στο  $k$  υπερεπίπεδο το σημείο  $P_m$ .

Αν το  $k$  είναι μικρότερο του  $n$  (δηλαδή αν δεν έχουν κατασκευαστεί όλοι οι άξονες του  $n$ -D χώρου), τότε κατασκευάζεται νέος άξονας ( $k+1$  τάξης) που ορίζεται από

τα  $V_i$  και  $P_m$  με κατεύθυνση προς το  $P_m$  (βλέπε σχ. 5.3). Η διαδικασία της προβολής επαναλαμβάνεται για τον νέο άξονα, ο οποίος έχει μοναδιαίο διάνυσμα:

$$\hat{a}_{k+1} = \frac{\overline{V_i P_{mk}}}{\left| \overline{V_i P_{mk}} \right|} \quad (\text{εξ. 5.4})$$

όπου το διάνυσμα  $V_i P_{mk}$  δίνεται από τον τύπο (ανάλυση σε συνιστώσες):

$$\overline{V_i P_{mk}} = \overline{V_i V_m} - \sum_{r=1}^k x_{mr} \cdot \hat{a}_r \quad (\text{εξ. 5.5})$$

Αν το  $k$  έχει φτάσει στον αριθμό  $n$  των διαστάσεων του νέου χώρου (οπότε δεν κατασκευάζεται πλέον νέος άξονας), γίνεται σύγκριση της τιμής της απόστασης  $d_m$  με την τιμή της παραμέτρου λεπτομέρειας  $\varepsilon$ . Αν η  $d_m$  βρεθεί μικρότερη από το  $\varepsilon$ , τότε θεωρούμε ότι η διάσταση  $n+1$  για αυτό το κομμάτι της καμπύλης είναι πολύ μικρή και μπορεί να αγνοηθεί. Όλοι οι υπόλοιποι άξονες  $A_{n+1}, \dots, A_N$  περιέχονται στο υπερεπίπεδο  $n$  το κάθετο στον  $A_n$ , επομένως η προβολή  $P_{sn}$  κάθε σημείου της καμπύλης στο υπερεπίπεδο  $n$  θα περιέχει όλες τις συντεταγμένες του  $x_{s(n+1)}, \dots, x_{sN}$ .

Η συντεταγμένη ενός σημείου είναι η προβολή του στον αντίστοιχο άξονα, γι' αυτό και οι παραπάνω συντεταγμένες είναι οι προβολές του  $V_i P_{sn}$  στους άξονες  $A_{n+1}, \dots, A_N$ . Ξέρουμε ότι η προβολή διανύσματος έχει μέτρο μικρότερο από ή ίσο με το μέτρο του διανύσματος. Το μέτρο του διανύσματος  $V_i P_{sn}$  έχει άνω όριο το:

$$\left| \overline{V_i P_{sn}} \right| \leq \left| \overline{V_i P_{mn}} \right| < \varepsilon \quad (\text{εξ. 5.6})$$

Άρα όλες οι συντεταγμένες  $x_{s(n+1)}, \dots, x_{sN}$  (ως προβολές) θα έχουν μέτρο μικρότερο του διανύσματος  $V_i P_{sn}$  δηλαδή μικρότερο του  $\varepsilon$ , και άρα μπορούν να απορριφθούν (αφού ικανοποιούν το κριτήριο σύγκλισης: μέτρο  $< \varepsilon$ ).

Σε αυτήν την περίπτωση ( $d_m < \varepsilon$ ) όπου ισχύει το κριτήριο σύγκλισης, το τμήμα  $V_i V_j$  “ταιριάζει” απόλυτα στον  $n$ -D χώρο (σύμφωνα πάντα με τον βαθμό λεπτομέρειας) και θεωρείται ότι περιγράφεται πλήρως από τα σημεία στα άκρα του, τα οποία είναι σημεία του  $n$ -D χώρου. Αυτό σημαίνει ότι το τμήμα  $V_i V_j$  όπως αναλύθηκε σε  $n$  συνιστώσες είναι επαρκές για να περιγράψει την καμπύλωση της καμπύλης των  $N$ -D διαστάσεων και γι' αυτό δεν αναλύεται σε περαιτέρω συνιστώσες. Τα σημεία του,  $V_s$ , μπορούν να αναπαρασταθούν με  $n$  συνιστώσες και συνολικά αποτελούν ένα ευθύγραμμο τμήμα στον  $n$ -D χώρο.

Στην περίπτωση που δεν ικανοποιείται το κριτήριο σύγκλισης ( $d_m > \varepsilon$ ), δηλαδή όταν ικανοποιείται η συνθήκη διαμέρισης, το τμήμα της καμπύλης πρέπει να χωριστεί σε δύο τμήματα, διότι δεν ταιριάζει στον n-D χώρο, αφού χρειάζεται περισσότερες από n συνιστώσες για να αναπαρασταθεί. Το τμήμα της καμπύλης διαμερίζεται στο σημείο  $V_m$  σε δύο άλλα καμπύλα τμήματα  $V_i V_m$  και  $V_m V_j$ , στα οποία εφαρμόζεται σε καθένα ξεχωριστά ο αλγόριθμος για την απλοποίηση καθενός στις n διαστάσεις (επαναληπτικά). Στην περίπτωση της παρούσας υλοποίησης, ως βαθμό λεπτομέρειας  $\varepsilon$  χρησιμοποιούμε ένα ποσοστό της μέγιστης απόστασης  $d_m$  που βρίσκεται αρχικά στην πρώτη επανάληψη (περίπου το 40 % της απόστασης αυτής).

#### 5.1.4 Γενίκευση του Αλγορίθμου του Ramer

Ο αλγόριθμος για την απλοποίηση καμπύλης του N-D χώρου σε τμήματα καμπύλης που έχουν διάσταση n (που “ταιριάζουν” στον n-D χώρο) περιληπτικά περιγράφεται ως εξής:

- Εκφράζουμε τις συντεταγμένες των σημείων  $V_s$  (για κάθε  $s = i, \dots, j$ ) του τμήματος  $V_i V_j$  της καμπύλης στο τοπικό σύστημα αναφοράς, όπου ως πρώτος άξονας  $A_1$  λαμβάνεται το  $V_i V_j$ .  
Βρίσκοντας την μεγαλύτερη απόσταση του  $V_i$  από τις προβολές ( $P_{s1}$ ) όλων των σημείων  $V_s$  στο υπερεπίπεδο το κάθετο στον άξονα  $A_1$  και ενώνοντας το  $V_i$  με το  $P_{s1}$  που δίνει την απόσταση αυτή, προκύπτει ο δεύτερος άξονας  $A_2$ .  
Ο k-άξονας προκύπτει από την προβολή ( $P_{s(k-1)}$ ) όλων των σημείων  $V_s$  στο υπερεπίπεδο το κάθετο στον άξονα  $A_{k-1}$  και ενώνοντας το  $V_i$  με το  $P_{s(k-1)}$  που δίνει την μέγιστη απόσταση (για κάθε  $k = 1, \dots, n$ ).
- Αφού κατασκευαστούν και οι n άξονες  $A_1, A_2, \dots, A_n$ , προβάλλουμε τα σημεία  $V_s$  στο υπερεπίπεδο το κάθετο στον άξονα  $A_n$ , βρίσκουμε την μέγιστη από τις αποστάσεις του σημείου  $V_i$  από τις προβολές  $P_{sn}$ , έστω  $P_{mn}$  η προβολή που δίνει τη μέγιστη απόσταση και έστω  $d_m$  αυτή η απόσταση.
- Αν  $d_m < \varepsilon$ , τότε το τμήμα  $V_i V_j$  της καμπύλης ταιριάζει στον n-D χώρο και δεν διαμερίζεται περαιτέρω.

- Αν  $d_m \geq \varepsilon$ , τότε διαμερίζουμε το τμήμα  $V_i V_j$  σε δύο άλλα τμήματα  $V_i V_m$  και  $V_m V_j$  και εφαρμόζουμε σε καθένα επαναληπτικά τον ίδιο αλγόριθμο για την προσέγγισή τους με τμήματα του  $n$ -D χώρου.

## 5.2 Η Μέθοδος Προσέγγισης Περιγράμματος

### 5.2.1 Εισαγωγικά

Η δεύτερη εναλλακτική μέθοδος που χρησιμοποιείται για την περιγραφή μιας καμπύλης βασίζεται στην θεωρία της ανάκτησης όμοιων σχημάτων (similar shape retrieval) μέσω της προσέγγισης του περιγράμματος ενός αντικειμένου. Οι J. Wang et al. ([Wan98a]) προτείνουν τη δομή ενός τέτοιου συστήματος ανάκτησης το οποίο χειρίζεται σχήματα σε διάφορες μορφές, κλίμακες και κατευθύνσεις, μέσω τοπικών αλλά και ολικών χαρακτηριστικών του σχήματος. Με μία μέθοδο δημιουργίας ακολουθίας κόμβων, εξάγονται ακραία σημεία (σημαντικά σε τοπικό επίπεδο) του περιγράμματος τα οποία ανεξάρτητα από το μέγεθος και τη θέση του αντικειμένου περιγράφουν με συγκεκριμένο τρόπο το σχήμα του. Όσον αφορά τα ολικά χαρακτηριστικά, αυτά συγκρίνονται με τα χαρακτηριστικά γνωστών σχημάτων μίας βάσης δεδομένων και αφού απορριφθούν τα ανόμοια σχήματα, στη συνέχεια η μέθοδος προχωράει σε τοπικά χαρακτηριστικά, όπου συγκρίνεται η Ευκλείδεια απόσταση μεταξύ των όμοιων σχημάτων με σκοπό την εξαγωγή της μεγαλύτερης δυνατής ομοιότητας και την τελική απόφαση για το σχήμα του αντικειμένου και το είδος του.

Τα ολικά χαρακτηριστικά αναφέρονται στο συνολικό σχήμα του αντικειμένου, όπως είναι η περίμετρος, η επιφάνεια που καλύπτει, η κυρτότητά του, όπως και ο προσανατολισμός του ως προς τους άξονες. Αυτά περιγράφουν τη συνολική μορφή του αντικειμένου, καθώς δεν περιλαμβάνουν τις λεπτομέρειες του περιγράμματός του. Σε αντίθεση με αυτά, τα τοπικά χαρακτηριστικά περιλαμβάνουν πληροφορίες για τις τοπικές λεπτομέρειες του περιγράμματος ενός σχήματος, όπως είναι οι ακμές του, οι γωνίες του, τα σημεία καμπύλωσης, και καθορίζουν με μεγαλύτερη ακρίβεια το σχέδιο του αντικειμένου, όμως πρόκειται για μέθοδο πιο ευαίσθητη σε παραμόρφωση της εικόνας και σε θόρυβο. Στην συγκεκριμένη εργασία θα αναλυθούν μόνο τα τοπικά

χαρακτηριστικά, καθώς αυτά είναι που ενδιαφέρουν στην μελέτη και στην προσέγγιση της καμπύλης στον N-διάστατο χώρο.

Το περίγραμμα ενός διδιάστατου σχεδίου αποτελείται από μία (καμπύλη) γραμμή που συντίθεται από πάρα πολλά σημεία. Τα σημεία αυτά μπορεί να προέρχονται από ψηφιακή επεξεργασία σήματος ή εικόνας, ή μπορεί να προκύπτουν από κάποιον αλγόριθμο, πάντως δεν είναι πρακτικό να κρατήσουμε όλα τα σημεία για λόγους πολυπλοκότητας χώρου και χρόνου και αυξημένης υπολογιστικής πολυπλοκότητας ([Wan98a]). Για τον σκοπό αυτό οι J. Wang et al. προτείνουν δύο αλγορίθμους για την επιλογή μερικών σημαντικών σημείων του περιγράμματος, τα οποία περιγράφουν με τον καλύτερο τρόπο τα χαρακτηριστικά και τις λεπτομέρειες του σχήματος, βασιζόμενοι στον τρόπο προσέγγισης με τη χρήση πολυεπίπεδου πολυγώνου. Η μέθοδος είναι παρόμοια με την απλοποίηση καμπύλης όπως περιγράφηκε στην προηγούμενη παράγραφο (μέθοδος Curve Simplification).

Ο πρώτος αλγόριθμος περιγράφεται στην πρώτη εργασία των J. Wang et al. ([Wan98a]) και ονομάζεται αλγόριθμος Global-searching (Global-searching algorithm), ενώ στη δεύτερη εργασία τους ([Wan98b]) προτείνουν έναν παρόμοιο αλγόριθμο με βελτιωμένη χρονική πολυπλοκότητα, που ονομάζεται αλγόριθμος Pair-growing (Pair-growing algorithm). Οι αλγόριθμοι αυτοί διατυπώθηκαν για 2-D καμπύλες, αλλά η γενίκευση σε περισσότερες διαστάσεις είναι απλή διαδικασία, αφού περιορίζεται στον ορισμό απόστασης σημείου από ευθεία. Παρακάτω δίνεται η περιγραφή αυτών των αλγορίθμων και η προσαρμογή τους για την περίπτωση που μας ενδιαφέρει.

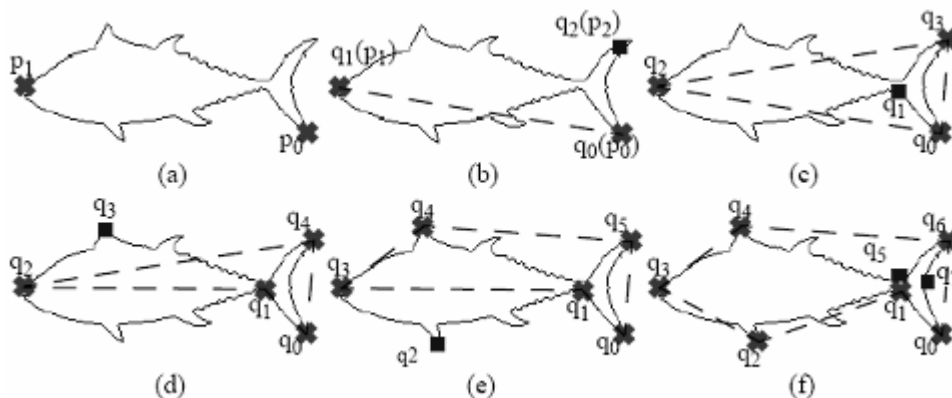
## 5.2.2 Περιγραφή των Αλγορίθμων

### 5.2.2.1 Αλγόριθμος Global-searching

Αν συμβολίσουμε με  $P$  το σύνολο από τα αρχικά διατεταγμένα σημεία που δίνουν το περίγραμμα του σχήματος, με  $Q$  το σύνολο από τα σημαντικά ενδιαφέροντα σημεία που πρέπει να εξάγουμε για την προσέγγιση του περιγράμματος, και με  $\max$  τον αριθμό των σημαντικών αυτών σημείων, τότε ο αλγόριθμος Global-searching έχει ως εξής:



1. Βρίσκουμε δύο σημεία του περιγράμματος (δηλαδή που ανήκουν στο σύνολο  $P$ )  $P_0$  και  $P_1$  τα οποία απέχουν την μεγαλύτερη απόσταση μεταξύ τους και από τα οποία εκκινεί ο αλγόριθμος.
2. Βρίσκουμε σημείο  $P_2$  του συνόλου  $P$  που απέχει την μέγιστη απόσταση από το ευθύγραμμο τμήμα  $P_0P_1$  και αφού ταξινομήσουμε τα σημεία  $P_0$ ,  $P_1$  και  $P_2$ , με τη σειρά που βρίσκονται στο  $P$ , τα τοποθετούμε στο σύνολο ενδιαφέροντος  $Q$  (το οποίο αρχικά είναι κενό).
3. Έστω  $M$  το πλήθος των σημείων του συνόλου  $Q$ . Για κάθε δύο γειτονικά σημεία του  $Q$ , έστω  $q_i$  και  $q_{i+1}$  ( $i = 0, 1, \dots, M-1$ ), αναζητούμε το σημείο  $P_{\max_i}$  που βρίσκεται μεταξύ τους στο σύνολο  $P$  και απέχει την μέγιστη απόσταση από το ευθύγραμμο τμήμα  $q_iq_{i+1}$ , έστω  $s_i$  η απόσταση αυτή. Από όλα τα σημεία  $P_{\max_i}$  που βρήκαμε για κάθε τέτοιο ζεύγος γειτονικών σημείων ( $i = 0, 1, \dots, M-1$ ), επιλέγεται εκείνο που έχει την μεγαλύτερη απόσταση  $s_i$  (τα συγκρίνουμε μεταξύ τους) και μόνο αυτό τελικά τοποθετείται στο σύνολο  $Q$ , το οποίο κατόπιν αναδιατάσσεται.
4. Το βήμα (3) επαναλαμβάνεται για το νέο διατεταγμένο σύνολο  $Q$  που μόλις κατασκευάστηκε, έως ότου το πλήθος των σημείων του  $Q$  φτάσει τον μέγιστο αριθμό  $\max$  των σημαντικών σημείων που αρχικά υποθέσαμε για την προσέγγιση του περιγράμματος (κριτήριο σύγκλισης).



**Σχ. 5.4 :** Εφαρμογή του αλγορίθμου Global-searching σε έξι επίπεδα ([Wan98b])

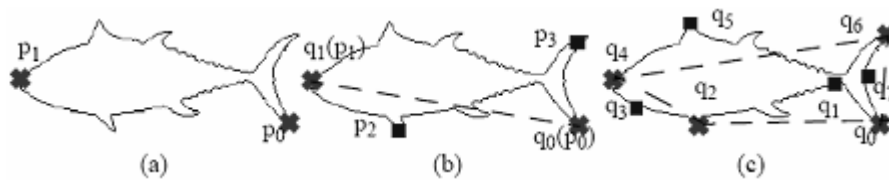
Στο σχήμα 5.4 παρουσιάζεται μία εφαρμογή του Global-searching algorithm σε επίπεδη καμπύλη, όπου φαίνονται έξι επίπεδα επανάληψης της μεθόδου. Σε κάθε επίπεδο το νέο σημείο που εντοπίζεται και προστίθεται στο σύνολο  $Q$  σημειώνεται με τετράγωνο, ενώ τα υπόλοιπα σημεία που υπάρχουν ήδη στο  $Q$  σημειώνονται με  $x$ . Αρχικά λαμβάνονται τα δύο πιο απομακρυσμένα σημεία  $P_0$  και  $P_1$  (a) και στη συνέχεια μαζί με το σημείο  $P_2$  κατασκευάζεται το  $Q$ , πρώτη επανάληψη (b). Στην δεύτερη επανάληψη εύρεσης σημείου με τη μεγαλύτερη απόσταση από τα ευθύγραμμα τμήματα προστίθεται στο  $Q$  το σημείο  $q_1$  (c) και το  $Q$  αναδιατάσσεται  $Q = \{q_0, q_1, q_2, q_3\}$ . Στα επόμενα βήματα τοποθετείται κάθε φορά ένα παραπάνω σημείο με αποτέλεσμα την δημιουργία του ανανεωμένου διατεταγμένου συνόλου  $Q$  στα επίπεδα λεπτομέρειας 3 (d), 4 (e), 5 και 6 (f), όπου τελικά με την εύρεση 7 σημαντικών σημείων το περίγραμμα του αντικειμένου συνοψίζεται στα σημεία του συνόλου  $Q = \{q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7\}$ .

### 5.2.2.2 Αλγόριθμος Pair-growing

Συμβολίζουμε και πάλι με  $P$  το σύνολο των αρχικών σημείων του περιγράμματος, με  $Q$  το σύνολο από τα σημαντικά σημεία που θα προσεγγίζουν το περίγραμμα, και με  $\max$  τον αριθμό των σημαντικών αυτών σημείων, οπότε ο αλγόριθμος Pair-growing περιγράφεται ως εξής:

1. Βρίσκουμε δύο σημεία του περιγράμματος από το σύνολο των διατεταγμένων σημείων  $P$ , τα οποία τα συμβολίζουμε  $P_0$  και  $P_1$  και τα οποία απέχουν την μεγαλύτερη απόσταση μεταξύ τους (από αυτά εκκινεί ο αλγόριθμος).
2. Τα δύο σημεία  $P_0$  και  $P_1$  τοποθετούνται στο σύνολο ενδιαφέροντος  $Q$  (το οποίο αρχικά είναι κενό).
3. Βρίσκουμε δύο σημεία  $P_2$  και  $P_3$ , τα οποία βρίσκονται μεταξύ των  $P_0, P_1$  και  $P_1, P_0$  (αντίστοιχα) και είναι εκείνα τα σημεία που απέχουν τη μεγαλύτερη απόσταση από τα ευθύγραμμα τμήματα  $P_0P_1$  και  $P_1P_0$  αντίστοιχα. Τα  $P_2$  και  $P_3$  τοποθετούνται στο σύνολο  $Q$ , το οποίο κατόπιν αναδιατάσσεται.

4. Έστω  $M$  το πλήθος των σημείων του συνόλου  $Q$ . Για κάθε δύο γειτονικά σημεία του  $Q$ , έστω  $q_i$  και  $q_{i+1}$  ( $i = 0, 1, \dots, M-1$ ,  $M$  με  $q_M = q_0$ ), αναζητούμε το σημείο  $P_{\max_i}$  που βρίσκεται μεταξύ τους στο σύνολο  $P$  και απέχει την μέγιστη απόσταση από το ευθύγραμμο τμήμα  $q_i q_{i+1}$ , έστω  $s_i$  η απόσταση αυτή. Κάθε τέτοιο σημείο  $P_{\max_i}$  (για κάθε  $i = 0, 1, \dots, M-1$ ) τοποθετείται στο σύνολο  $Q$ , το οποίο κατόπιν και αναδιατάσσεται.
5. Το βήμα (4) επαναλαμβάνεται για κάθε νέο σύνολο  $Q$  που κατασκευάζεται έως ότου ο αριθμός των σημείων του  $Q$  γίνει ίσος με τον μέγιστο αριθμό  $\max$  των σημαντικών σημείων (κριτήριο σύγκλισης).



Σχ. 5.5 : Εφαρμογή του αλγορίθμου Pair-growing σε δύο επίπεδα ([Wan98b])

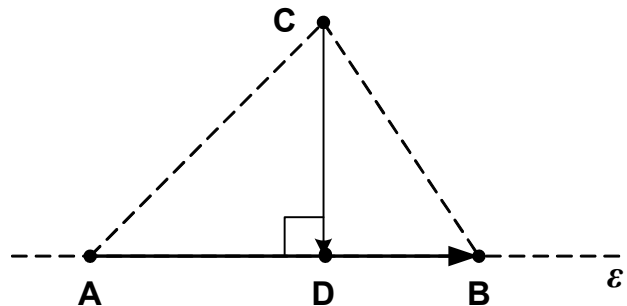
Στο σχήμα 5.5 παρουσιάζεται μία εφαρμογή του Pair-growing algorithm σε επίπεδη καμπύλη, όπου φαίνονται τα δύο επίπεδα επανάληψης της μεθόδου. Σε κάθε επίπεδο τα νέα σημεία που προστίθενται στο σύνολο  $Q$  σημειώνονται με τετράγωνο, ενώ τα υπόλοιπα σημεία που υπάρχουν ήδη στο  $Q$  σημειώνονται με  $x$ . Αρχικά λαμβάνονται τα δύο πιο απομακρυσμένα σημεία  $P_0$  και  $P_1$  (a) και στη συνέχεια μαζί με τα σημεία  $P_2$  και  $P_3$  κατασκευάζεται το  $Q$  (b), το οποίο και διατάσσεται με την ορθή σειρά των σημείων (πρώτη επανάληψη). Στο επόμενο βήμα (c), για κάθε ζεύγος σημείων εντοπίζονται τα ενδιάμεσα σημεία με τη μεγαλύτερη απόσταση από τα τμήματα και τοποθετούνται και αυτά στο  $Q$ . Τελικά, μετά και την δεύτερη επανάληψη, προκύπτει το αναδιατεταγμένο σύνολο  $Q = \{q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7\}$ .

### 5.2.3 Προσαρμογή για την Περίπτωση $N$ -D Καμπύλης

Για να εφαρμόσουμε τις δύο παραπάνω μεθόδους στην περίπτωση του διανύσματος χαρακτηριστικών, το οποίο παράγει μία καμπύλη στον πολυδιάστατο Ευκλείδειο χώρο, ως αρχικά σημεία  $P_0$  και  $P_1$ , και στις δύο μεθόδους, λαμβάνονται το

πρώτο και το τελευταίο στιγμιότυπο του διανύσματος χαρακτηριστικών, δηλαδή τα δύο σημεία του N-D χώρου που αντιστοιχούν στο πρώτο και στο τελευταίο καρέ του πλάνου. Αυτά τα δύο σημεία τοποθετούνται στο σύνολο ενδιαφέροντος  $Q$  και οι αλγόριθμοι προχωρούν κανονικά από αυτό το σημείο, με την διαπίστωση ότι στην περίπτωση του αλγορίθμου Pair-growing το τρίτο βήμα θα δώσει ένα μόνο νέο σημείο (παρόμοια με το βήμα 2 του Global-searching). Αυτό συμβαίνει διότι η καμπύλη που εξετάζουμε δεν είναι κλειστή, όπως το επίπεδο κλειστό περίγραμμα που αναλύεται από τους J. Wang et al., αλλά πρόκειται καμπύλη με αρχή και τέλος.

Επίσης, για να προσαρμόσουμε τους δύο αλγορίθμους στην περίπτωση καμπύλης στον N-D χώρο, αρκεί να ορίσουμε στον N-διάστατο Ευκλείδειο χώρο την απόσταση ενός σημείου  $C$  από την ευθεία  $\varepsilon$  που διέρχεται από δύο άλλα σημεία, έστω τα σημεία  $A$  και  $B$ . Αν υποθέσουμε ότι  $D$  είναι το σημείο που αποτελεί την προβολή του σημείου  $C$  πάνω στην ευθεία που διέρχεται από τα σημεία  $A$  και  $B$ , όπως φαίνεται και από το σχήμα 5.6, ανάγουμε το πρόβλημα στην εύρεση του μέτρου του διανύσματος  $CD$ . Η γραμμή που ενώνει τα σημεία  $A$  και  $B$  είναι ευθεία μιας διάστασης, αλλά το επίπεδο που ορίζεται από τα σημεία  $A$ ,  $B$  και  $C$  είναι ένα υπερεπίπεδο  $N-1$  διαστάσεων.



**Σχ. 5.6 :** Εύρεση της απόστασης σημείου από ευθεία στον N-D χώρο

Το διάνυσμα  $CD$  είναι ορθογώνιο με το διάνυσμα  $AB$  με την γνωστή έννοια της ορθογωνιότητας που γενικεύεται για τον N-διάστατο χώρο με τη βοήθεια του ορισμού του εσωτερικού γινομένου και της Ευκλείδειας νόρμας. Για τη ζητούμενη απόσταση  $\text{dist}$  του  $C$  από την ευθεία  $\varepsilon$ , που είναι το μέτρο του διανύσματος  $CD$ , πρέπει να εντοπίσουμε το σημείο  $D$ . Όμως, για να αποφύγουμε τον υπολογισμό των συντεταγμένων του  $D$ , βρίσκουμε το μέτρο του διανύσματος  $AD$  το οποίο προκύπτει από το εσωτερικό γινόμενο των διανυσμάτων  $AC$  και  $AB$ :

$$\overline{AC} \cdot \overline{AB} = \overline{AD} \cdot \overline{AB} = \|\overline{AD}\| \cdot \|\overline{AB}\| \Rightarrow$$

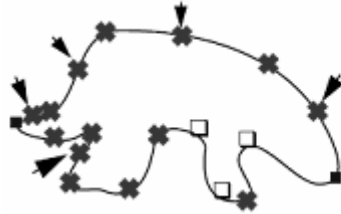
$$\|\overline{AD}\| = \frac{\overline{AC} \cdot \overline{AB}}{\|\overline{AB}\|} \quad (\text{εξ. 5.7})$$

όπου το εσωτερικό γινόμενο “ $\cdot$ ” και η Ευκλείδεια νόρμα “ $\|\cdot\|$ ” διανυσμάτων στον N-D Ευκλείδειο χώρο έχουν οριστεί στο τέλος του Κεφαλαίου 3 (εξ. 3.1, 3.2, 3.3).

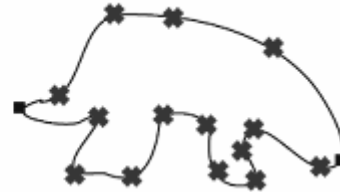
Από το Πυθαγόρειο Θεώρημα προκύπτει η ζητούμενη απόσταση  $dist$  (το μέτρο του διανύσματος  $CD$ ), που χρησιμοποιείται στους παραπάνω αλγόριθμους:

$$\|\overline{AC}\|^2 = \|\overline{AD}\|^2 + \|\overline{CD}\|^2 \Rightarrow$$

$$\Rightarrow dist = \|\overline{CD}\| = \sqrt{\|\overline{AC}\|^2 - \|\overline{AD}\|^2} \quad (\text{εξ. 5.8})$$



(a) Σημεία αλγορίθμου Pair-growing



(b) Σημεία αλγορίθμου Global-searching

**Σχ. 5.7 :** Σύγκριση της προσέγγισης Pair-growing και Global-searching ([Wan98b])

Όπως φαίνεται από το παραπάνω σχήμα (σχ. 5.7), και σύμφωνα με τους J. Wang et al. ([Wan98b]) ο αλγόριθμος Pair-growing, που αναπτύχθηκε με σκοπό την γρηγορότερη σύγκλιση και την μικρότερη πολυπλοκότητα, έχει ως αποτέλεσμα την υπόδειξη πολλών περιττών σημείων όπως αυτά που σημειώνονται με βέλος στο σχήμα 5.7 (a), τα οποία δεν χρειάζονται για την περιγραφή του σχήματος, αφού δεν βρίσκονται σε τμήματα με απότομες μεταβολές. Επίσης ο αλγόριθμος αυτός αγνοεί πολλά χαρακτηριστικά σημεία που καθορίζουν το σχήμα του αντικειμένου, όπως τα σημεία που σημειώνονται με λευκό τετράγωνο στο σχήμα 5.7 (a). Αντίθετα, ο αλγόριθμος Global-searching εξάγει λιγότερα σημεία στο τμήμα του περιγράμματος που καμπυλώνει λιγότερο και περισσότερα σημεία (όπως και απαιτείται) στα τμήματα με τις απότομες καμπυλώσεις (βλέπε σχήμα 5.7 (b)).

Όπως παρατηρούμε και από τα σχήματα 5.4 και 5.5, για την εξαγωγή του ίδιου αριθμού σημαντικών σημείων (8) ο αλγόριθμος Global-searching απαιτεί την πραγματοποίηση έξι επαναλήψεων, ενώ ο αλγόριθμος Pair-growing απαιτεί μόνο δύο επαναλήψεις για να εξάγει το ίδιο πλήθος σημείων. Όμως το σημείο  $q_3$  (βλέπε σχ. 5.5 (c)) είναι περιττό, αφού δεν δίνει σημείο καμπύλωσης, ενώ λείπει σημείο που θα έδειχνε δραματική καμπύλωση του περιγράμματος, όπως το σημείο  $q_5$  (βλέπε σχ. 5.4 (f)).

Δεδομένου λοιπόν ότι επιθυμούμε την όσο το δυνατόν πιο πλήρη αναπαράσταση της N-D καμπύλης, με όλα τα σημαντικά σημεία μεγάλης καμπύλωσης (απότομης αλλαγής περιεχομένου), χρησιμοποιούμε τον Global-searching με κόστος την πολυπλοκότητα χρόνου.

## ΚΕΦΑΛΑΙΟ 6

# *Υπολογιστικό Μοντέλο Οπτικής*

## *Προσοχής*

Στο κεφάλαιο αυτό αναλύεται η μεθοδολογία που οδηγεί στη δημιουργία ενός υπολογιστικού μοντέλου οπτικής προσοχής (**visual attention system**), που βασίζεται στο ανθρώπινο οπτικό σύστημα ([Itt98]). Το σύστημα αυτό, λαμβάνοντας στην είσοδό του μια ψηφιακή εικόνα, έχει σκοπό να διακρίνει περιοχές της εικόνας που θεωρούνται σημαντικές και ξεχωρίζουν από τις υπόλοιπες. Επίσης, έχει τη δυνατότητα να καθορίζει ένα βαθμό ενδιαφέροντος για κάθε περιοχή της εικόνας, ούτως ώστε η οπτική προσοχή να μπορεί να κατευθυνθεί σε ιδιαίτερα σημαντικές περιοχές και να οδηγηθεί σταδιακά από αυτές στις λιγότερο σημαντικές περιοχές.

Η υπολογιστική αυτή διαδικασία έχει στόχο μέσω της επεξεργασίας που πραγματοποιεί να μειώσει τη δεδομένη πολυπλοκότητα του προβλήματος κατανόησης του περιεχομένου μιας εικόνας (image understanding). Επιλέγονται έτσι περιοχές με τα πιο προφανή και εντυπωσιακά σημεία, οι οποίες και αναλύονται περαιτέρω, αντί να αναλυθεί ολόκληρη η πληροφορία της εικόνας. Στην εργασία αυτή ο σκοπός της εφαρμογής της οπτικής προσοχής (**Visual Attention**) είναι ο εντοπισμός των σημαντικών περιοχών σε κάθε καρέ της ακολουθίας του βίντεο και η εξαγωγή των χαρακτηριστικών (ιστόγραμμα  $Y$ ,  $U$ ,  $V$ ) μόνο από αυτές.

## 6.1 Εισαγωγικά

Η μεθοδολογία που αναπτύσσεται εδώ για την κατασκευή του συστήματος οπτικής προσοχής με βάση τη σημαντικότητα των περιοχών της εικόνας έχει μελετηθεί και αναλυθεί από τους L. Itti, C. Koch και E. Niebur το 1998 ([Itt98]). Βασίζονται καταρχήν στο βιολογικό ισοδύναμο του οπτικού συστήματος που επιθυμούν να κατασκευάσουν. Σύμφωνα με έρευνες, τα θηλαστικά έχουν έναν αξιόλογο τρόπο να μεταφράζουν ένα σύνθετη εικόνα σε άμεσο χρόνο (real-time), παρά την περιορισμένη ταχύτητα των νευρώνων που διαθέτουν για αυτήν την επεξεργασία.

Έχει διαπιστωθεί ότι μέσω ειδικών διαδικασιών επιλέγουν ένα υποσύνολο των πληροφοριών που προσφέρονται από τα αισθητήρια όργανα, προκειμένου να προχωρήσουν στην περαιτέρω επεξεργασία τους, με στόχο τη μείωση της πολυπλοκότητας ανάλυσης της εικόνας. Η επιλογή αυτή γίνεται ταυτόχρονα με δύο τρόπους, ο ένας είναι γρήγορος, από κάτω προς τα πάνω (bottom-up), με βάση τη σημασία, και ο άλλος ο πιο αργός, από πάνω προς τα κάτω (top-down), με βάση συγκεκριμένο στόχο. Οδηγεί στον εντοπισμό μιας συγκεκριμένης περιοχής του οπτικού πεδίου, διαδικασία που ονομάζεται “εστίαση της προσοχής” (**focus of attention**).

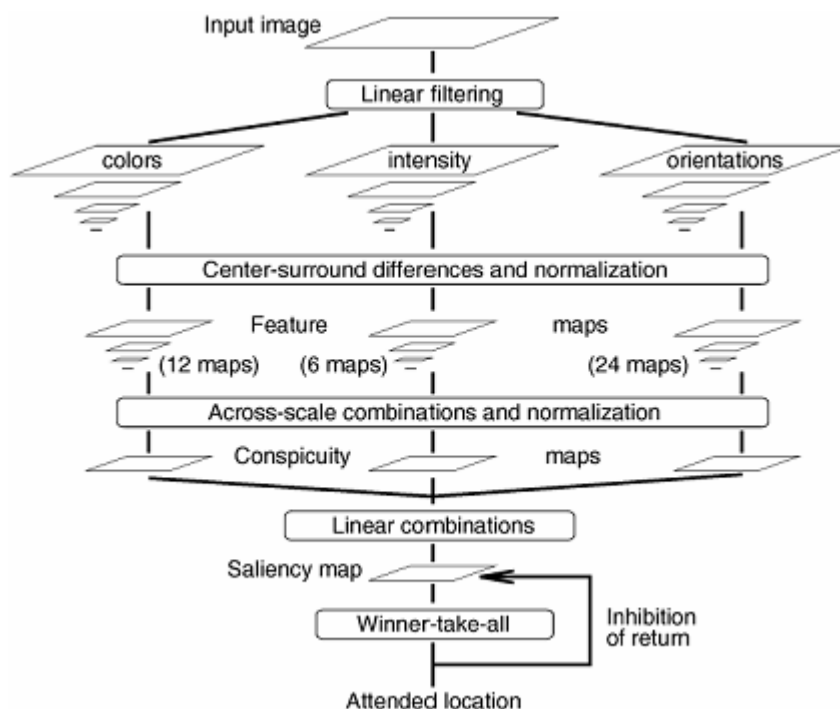
Πολλά μοντέλα που αναφέρονται στην οπτική προσοχή περιγράφουν την επεξεργασία της πληροφορίας του οπτικού πεδίου με δυναμικό τρόπο. Η πληροφορία αυτή προέρχεται από μια μικρή περιοχή της εικόνας, η οποία επιλέγεται μέσω υπολογισμών των νευρώνων του εγκεφάλου. Η διαδικασία της επιλογής βασίζεται όπως έχει αποδειχθεί σε διεργασίες με top-down και bottom-up ελέγχους, δηλαδή με κριτήριο τον στόχο και το περιεχόμενο της εικόνας αντίστοιχα.

Στην εργασία αυτή περιγράφεται το σύστημα που προτάθηκε από τους Itti και Koch ([Itt98]). Είναι ένα μοντέλο με βιολογικό ανάλογο, βασισμένο σε προτάσεις που έχουν υιοθετηθεί από πολλά οπτικά μοντέλα. Αποτελεί ένα από κάτω προς τα επάνω (bottom-up) σύστημα, που υλοποιείται μέσω της θεωρίας που επιδιώκει να εξηγήσει τη διαδικασία της bottom-up ανθρώπινης όρασης. Η οπτική διαδικασία από επάνω προς τα κάτω (top-down) είναι πολύ δύσκολο να αναλυθεί και να υλοποιηθεί (βρίσκεται πιο κοντά στην ανθρώπινη αντίδραση), οπότε αγνοείται και το σύστημα λειτουργεί αυτόνομα. Η περιγραφή του συστήματος και τα βήματα της διαδικασίας που προτείνουν οι παραπάνω συγγραφείς δίνονται αναλυτικά στις επόμενες παραγράφους.



## 6.2 Περιγραφή του Μοντέλου Οπτικής Προσοχής

Όπως αναφέρθηκε, το μοντέλο Οπτικής Προσοχής (Visual Attention) βασίζεται στα δεδομένα της εικόνας (bottom-up) και γι' αυτό δεν περιορίζεται σε περιπτώσεις γνώσης εκ των προτέρων (π.χ. του στόχου). Αρχικά αναλύει την είσοδο (οπτικό ερέθισμα) σε ένα σύνολο χαρτών προσοχής (**conspicuity maps**), που περιλαμβάνουν τοπικά χαρακτηριστικά. Οι διαφορετικές χωρικές κατανομές των χαρακτηριστικών ιδιοτήτων ανταγωνίζονται μέσα σε κάθε χάρτη και τελικά επικρατούν περιοχές που τοπικά ξεχωρίζουν από το περιβάλλον τους (γειτονιά). Τέλος, όλοι οι χάρτες λαμβάνονται υπόψη για την κατασκευή ενός βασικού χάρτη ενδιαφέροντος (**saliency map**), όπου κάθε χάρτης προσοχής καθορίζει διαφορετικά τα σημεία που “προεξέχουν” τοπικά μέσα σε ολόκληρη την εικόνα. Όλα τα παραπάνω φαίνονται αναλυτικά στο σχήμα 6.1, που δείχνει τη δομή του μοντέλου.



Σχ. 6.1 : Γενική Δομή του μοντέλου Οπτικής Προσοχής ([Itt98])

Στο βιολογικό ισοδύναμο, πιστεύεται ότι ένα παρόμοιο μοντέλο βρίσκεται στην πίσω πλευρά του εγκεφάλου του ανθρώπου. Ο χάρτης ενδιαφέροντος (ή σημαντικότητας) προκαλεί μετατοπίσεις της προσοχής και το μοντέλο δεν απαιτεί καθοδήγηση (γνώση, top-down) για να εστιάσει την προσοχή. Σύμφωνα με το παραπάνω σύστημα επιλέγεται

γρήγορα μικρός αριθμός περιοχών ενδιαφέροντος από την συνολική εικόνα για περαιτέρω επεξεργασία (μεγαλύτερης πολυπλοκότητας).

Αν ως είσοδος στο σύστημα δοθεί μία εικόνα με ανάλυση 640x480 pixels, κατασκευάζονται εννέα χωρικές κλίμακες, που προκύπτουν χρησιμοποιώντας δυαδικές Γκαουσιανές πυραμίδες (dyadic Gaussian pyramids) και εφαρμόζοντας προοδευτικά φιλτράρισμα με κατωδιαβατό φίλτρο (lowpass filtering) και υποδειγματοληψία της εικόνας. Προκύπτει τελικά μία μείωση της ανάλυσης της εικόνας σε οριζόντια και κατακόρυφη διάσταση, με κλίμακες από 1:1 (scale 0) έως 1:256 (scale 8). Σε περίπτωση διαφορετικής ανάλυσης της εικόνας, ο αριθμός των επιπέδων καθορίζεται ανάλογα (συνήθως επιδιώκουμε στο τελευταίο επίπεδο να έχουμε ανάλυση τουλάχιστον 4x4 pixels).

Οι τυπικοί οπτικοί νευρώνες είναι πιο ευαίσθητοι σε σημεία που διαφέρουν σε σχέση με τα σημεία γύρω τους, δηλαδή σε τοπικές ασυνέχειες, και γι' αυτόν τον λόγο κάθε χαρακτηριστικό υπολογίζεται από γραμμικές πράξεις μεταξύ κέντρου και γειτονιάς (center-surround). Αυτό υλοποιείται ως διαφορά μεταξύ κέντρου στο λεπτομερές επίπεδο και γειτονιάς στο λιγότερο λεπτομερές επίπεδο. Το κέντρο  $c$  ανήκει στο σύνολο  $\{2, 3, 4\}$  και η γειτονιά  $s$  είναι  $s = c + \delta$ , όπου το  $\delta$  ανήκει στο σύνολο  $\{3, 4\}$ . Χρησιμοποιώντας διάφορες τιμές για τα  $s$  και  $\delta$  προκύπτει η εξαγωγή χαρακτηριστικών σε πολλά επίπεδα (multiscale feature extraction).

### 6.3 Εξαγωγή των Πρωταρχικών Οπτικών Χαρακτηριστικών

Συμβολίζοντας με  $r$ ,  $g$ ,  $b$ , τα κανάλια του κόκκινου, πράσινου και μπλε χρώματος της εικόνας, κατασκευάζεται μια εικόνα έντασης  $I$  και η αντίστοιχη Γκαουσιανή πυραμίδα  $I(\sigma)$ , για τις κλίμακες  $\sigma \in [0..8]$ . Η ένταση (φωτεινότητα)  $I$  είναι:

$$I = \frac{r + g + b}{3} \quad (\text{εξ. 6.1})$$

Στη συνέχεια τα κανάλια  $r$ ,  $g$ ,  $b$  κανονικοποιούνται με το  $I$  με σκοπό τον διαχωρισμό της απόχρωσης από την ένταση (διαιρούνται με την τιμή της έντασης σε κάθε σημείο). Επειδή διαφορές στο χρωματικό πεδίο δεν διακρίνονται σε μικρές τιμές της φωτεινότητας, η κανονικοποίηση λαμβάνει χώρα σε σημεία όπου η ένταση είναι

μεγαλύτερη από το  $1 / 10$  της μέγιστης τιμής της, ενώ στα υπόλοιπα σημεία μικρότερης έντασης τα  $r$ ,  $g$ ,  $b$  μηδενίζονται. Δημιουργούνται τέσσερα διαφορετικά χρωματικά κανάλια που είναι απομακρυσμένα μεταξύ τους, για το κόκκινο, το πράσινο, το μπλε και το κίτρινο χρώμα, που είναι αντίστοιχα:

$$R = r - \frac{g+b}{2} \quad (\text{εξ. 6.2})$$

$$G = g - \frac{r+b}{2} \quad (\text{εξ. 6.3})$$

$$B = b - \frac{r+g}{2} \quad (\text{εξ. 6.4})$$

$$Y = \frac{r+g}{2} - \frac{|r-g|}{2} - b \quad (\text{εξ. 6.5})$$

Στις παραπάνω σχέσεις, σε περίπτωση αρνητικών τιμών το χρωματικό κανάλι μηδενίζεται. Από τα κανάλια αυτά κατασκευάζονται οι Γκαουσιανές πυραμίδες  $R(\sigma)$ ,  $G(\sigma)$ ,  $B(\sigma)$  και  $Y(\sigma)$  αντίστοιχα για τις κλίμακες  $\sigma \in [0..8]$ .

Στη συνέχεια, από τις Γκαουσιανές πυραμίδες δημιουργούνται οι χάρτες των χαρακτηριστικών, με τις χαρακτηριστικές ιδιότητες των τοπικών διαφορών στην ένταση ή στο χρώμα. Οι τιμές των χαρτών αυτών προκύπτουν μέσω της διαφοράς κέντρου-γειτονιάς, με το κέντρο σε λεπτομερές επίπεδο και τη γειτονιά σε λιγότερο λεπτομερές επίπεδο (πιο “χοντρικό”). Ορίζεται η διαφορά επί κλίμακας (**across-scale difference**), η οποία είναι η παρεμβολή (interpolation) στο πιο λεπτομερές επίπεδο (finer scale) και η αφαίρεση σημείου από σημείο (point-by-point), συμβολίζεται δε με το “ $\Theta$ ”.

Όσον αφορά στην ένταση  $I$ , γνωρίζουμε ότι ο ανθρώπινος οργανισμός την ανιχνεύει μέσω νευρώνων που είναι ευαίσθητοι σε σκοτεινό κέντρο μέσα σε φωτεινή γειτονιά ή σε φωτεινό κέντρο που περιβάλλεται από σκοτεινή γειτονιά. Αυτό εκφράζεται ταυτόχρονα με την απόλυτη τιμή της διαφοράς across-scale difference, κατασκευάζοντας έξι χάρτες  $I(c, s)$ , όπου  $c \in \{2, 3, 4\}$  και  $s = c + \delta$ , με  $\delta \in \{3, 4\}$ :

$$I(c, s) = |I(c) \Theta I(s)| \quad (\text{εξ. 6.6})$$

Ένα δεύτερο σύνολο χαρτών κατασκευάζεται ομοίως για τα 4 χρωματικά κανάλια που υπολογίστηκαν, με τη διαφορά ότι εδώ ισχύει η “διπλή χρωματική αντίθεση”, όπου ο νευρώνας για το κέντρο διεγείρεται από ένα χρώμα και δεν διεγείρεται από κάποιο άλλο, ενώ το αντίθετο συμβαίνει για τη γειτονιά (για τα ίδια χρώματα). Τέτοια χρωματική και χωρική αντίθεση αποδεικνύεται ότι ισχύει για τα χρώματα κόκκινο / πράσινο και μπλε /

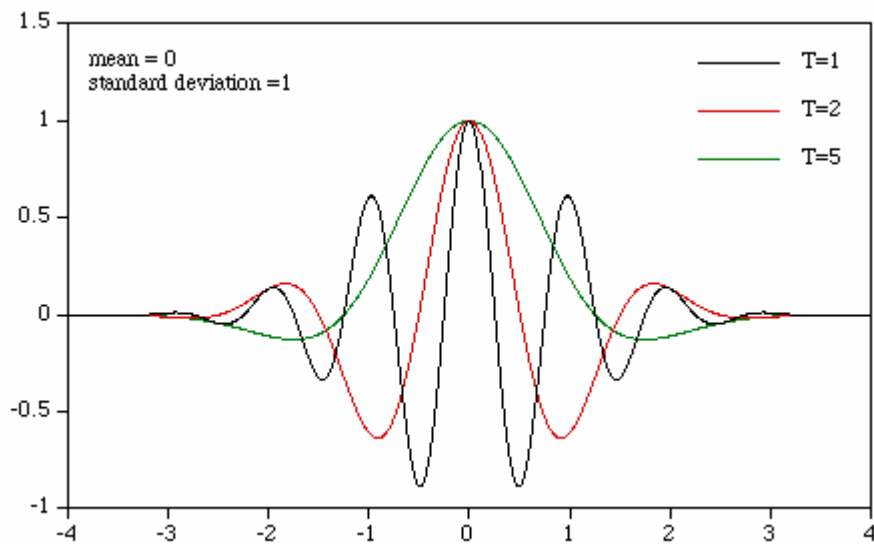
κίτρινο. Αυτό περιγράφεται ταυτόχρονα από τους έξι χάρτες  $RG(c, s)$  και τους έξι χάρτες  $BY(c, s)$ , που προκύπτουν αντίστοιχα με τις πράξεις:

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \quad (\text{εξ. 6.7})$$

$$BY(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \quad (\text{εξ. 6.8})$$

Τέλος, πληροφορίες σχετικά με τον προσανατολισμό του περιεχομένου της εικόνας εξάγονται από την εικόνα της έντασης  $I$ , εφαρμόζοντας προσανατολισμένα φίλτρα Gabor (Gabor filters) και κατασκευάζοντας προσανατολισμένες Gabor πυραμίδες  $O(\sigma, \theta)$ , όπου η κλίμακα είναι  $\sigma \in [0..8]$  και η επιθυμητή γωνία προσανατολισμού  $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ .

Ένα φίλτρο Gabor, που είναι το γινόμενο μιας Γκαουσιανής διδιάστατης συνάρτησης και ενός συνημιτόνου ορισμένου σε δικτυωτό πλέγμα, γνωρίζουμε ότι προσεγγίζει την ευαισθησία με την οποία οι νευρώνες αντιλαμβάνονται τον προσανατολισμό στο ερέθισμα που δέχονται. Ένα παράδειγμα φίλτρου Gabor με μία παράμετρο (μονοδιάστατο φίλτρο) δίνεται στο σχήμα 6.2, όπου φαίνεται η μορφή του για τρεις διαφορετικές τιμές της περιόδου  $T$  του συνημιτόνου.



**Σχ. 6.2 :** Ένα φίλτρο Gabor μιας παραμέτρου, για διαφορετικές τιμές του  $T$

Το διδιάστατο φίλτρο Gabor που μας ενδιαφέρει (για να εφαρμοστεί στην εικόνα των δύο διαστάσεων) έχει δύο παραμέτρους  $x$  και  $y$  και αναλυτικό τύπο:

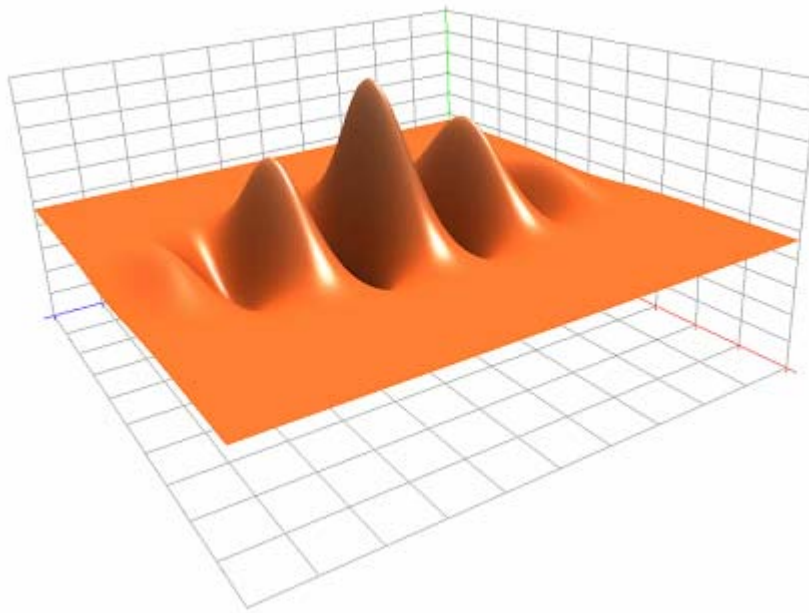
$$gabor(x, y) = \exp\left\{-\frac{x^2 + y^2}{2 \cdot \sigma^2}\right\} \cdot \cos\left(\frac{2 \cdot \pi \cdot x'}{T}\right) \quad (\text{εξ. 6.9})$$

όπου το  $x'$  δίνεται συναρτήσει των  $x, y, \theta$ :

$$x' = x \cdot \cos(\theta) + y \cdot \sin(\theta) \quad (\text{εξ. 6.10})$$

όπου τα  $x$  και  $y$  αποτελούν το πλέγμα, συνήθως διαστάσεων  $5 \times 5$  ή  $7 \times 7$ .

Αν απεικονίσουμε τη συνάρτηση Gabor δύο παραμέτρων σε τρισδιάστατο σύστημα, όπου οι δύο άξονες αποτελούν τις παραμέτρους  $x$  και  $y$  και ο τρίτος κατακόρυφος άξονας έχει την τιμή της συνάρτησης, τότε η επιφάνεια που δημιουργείται έχει τη μορφή του σχήματος 6.3, απ' όπου γίνεται κατανοητή και η λειτουργία του φίλτρου στον διαχωρισμό περιοχών με συγκεκριμένο προσανατολισμό.



**Σχ. 6.3 :** Τρισδιάστατη απεικόνιση φίλτρου Gabor με δύο παραμέτρους  $x, y$

Με αυτόν τον τρόπο κατασκευάζονται οι χάρτες για τα χαρακτηριστικά του προσανατολισμού, εκφράζοντας την τοπική αντίθεση στον προσανατολισμό μεταξύ κέντρου και γειτονιάς, έξι σε πλήθος για κάθε γωνία προσανατολισμού, που δίνονται από τη σχέση:

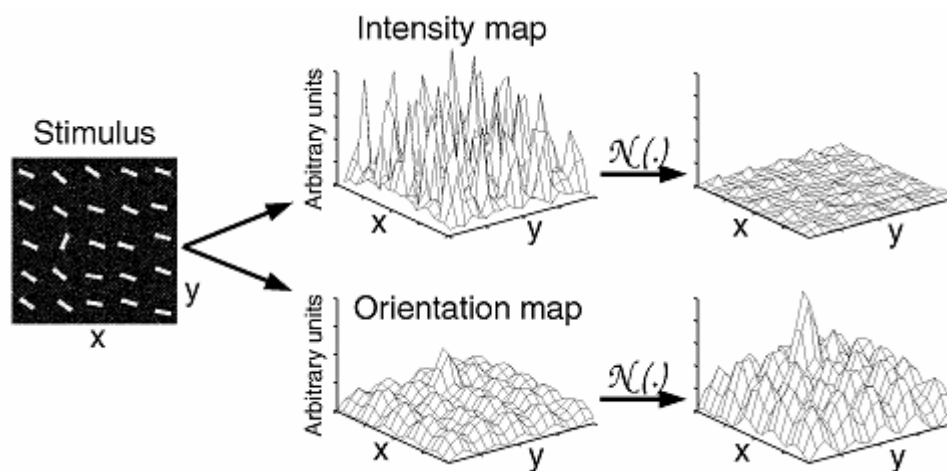
$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)| \quad (\text{εξ. 6.11})$$

Συνολικά έχουμε 42 χάρτες χαρακτηριστικών (**feature maps**), από τους οποίους 6 αντιστοιχούν στην ένταση  $I$  (intensity maps), 12 στο χρώμα  $C$  (color maps) και 24 στον προσανατολισμό  $O$  (orientation maps).

## 6.4 Κατασκευή του Χάρτη Ενδιαφέροντος

Στο σημείο αυτό κατασκευάζεται ο χάρτης ενδιαφέροντος (saliency map), ο οποίος έχει σκοπό να αναπαραστήσει τον βαθμό σημαντικότητας (ή ενδιαφέροντος) σε κάθε σημείο του οπτικού πεδίου, ώστε μέσω της σύγκρισης να επιτρέπει την ακολουθιακή επιλογή σημείων ενδιαφέροντος με φθίνουσα σημασία. Οι χάρτες των χαρακτηριστικών (feature maps) συνδυάζονται (bottom-up) για να δώσουν την είσοδο για τον χάρτη ενδιαφέροντος, που μοντελοποιείται ως δυναμικό νευρωνικό δίκτυο.

Οι χάρτες των χαρακτηριστικών περιέχουν πληροφορία που δεν είναι άμεσα συγκρίσιμη, αφού προέρχονται από διαφορετικό τρόπο υπολογισμού και διαφορετικό εύρος για κάθε χαρακτηριστικό. Γι' αυτό είναι δύσκολος ο συνδυασμός τους σε έναν χάρτη. Επίσης, το γεγονός ότι και οι 42 χάρτες συνδυάζονται μαζί για το τελικό αποτέλεσμα, σημαίνει ότι υπάρχει περίπτωση σημαντικά αντικείμενα που βρίσκονται σε μερικούς μόνο χάρτες να καλύπτονται από θόρυβο ή ασήμαντες περιοχές που βρίσκονται σε πολλούς χάρτες.



Σχ. 6.4 : Ο Τελεστής Κανονικοποίησης  $N(.)$  και η λειτουργία του ([Itt98])

Για τους παραπάνω λόγους και εξαιτίας της έλλειψης γνώσης (επίβλεψη top-down), οι Itti et al. ([Itt98]) προτείνουν την εφαρμογή ενός **τελεστή κανονικοποίησης**  $N(.)$ , ο οποίος προάγει χάρτες (ολόκληρους) στους οποίους εμφανίζεται μικρός αριθμός απότομων αιχμών (peaks), ενώ καταστέλλει χάρτες (μειώνει τις τιμές των στοιχείων τους) στους οποίους εμφανίζεται μεγάλος αριθμός συγκρίσιμων αιχμών. Στο σχήμα 6.4

φαίνεται σχηματικά πώς λειτουργεί και ποια είναι η χρησιμότητα αυτής της ιδέας. Ο τελεστής λαμβάνει ως είσοδο έναν χάρτη χαρακτηριστικών και περιλαμβάνει τα εξής:

1. Κανονικοποίηση των τιμών του χάρτη σε καθορισμένο εύρος  $[0..M]$ , με σκοπό την εξασφάλιση κοινού εύρους στις τιμές των χαρτών.
2. Εντοπισμός του ολικού μεγίστου  $M$  του χάρτη και υπολογισμός της μέσης τιμής  $\bar{m}$  όλων των υπολοίπων τοπικών μεγίστων.
3. Πολλαπλασιασμός κάθε τιμής του χάρτη επί την τιμή  $(M - \bar{m})^2$ .

Μέσω της διαδικασίας αυτής εξετάζονται μόνο τοπικά μέγιστα, αφού ο παραπάνω τελεστής αγνοεί περιοχές που είναι ομοιογενείς και προάγει περιοχές που περιέχουν σημεία ενδιαφέρουσας δραστηριότητας (activation spots). Με τη διαφορά μεταξύ ολικού μεγίστου και μέσης τιμής τοπικών μεγίστων υπολογίζεται το πόσο διαφέρει η πιο “δραστήρια” περιοχή από τις υπόλοιπες. Όταν η διαφορά αυτή είναι μεγάλη, σημαίνει ότι η περιοχή αυτή ξεχωρίζει περισσότερο από τη γειτονιά της και γι’ αυτό ο χάρτης πολλαπλασιάζεται με μεγάλη τιμή, ώστε να παίζει καθοριστικό ρόλο στο αποτέλεσμα. Το αντίθετο συμβαίνει όταν η διαφορά αυτή είναι μικρή.

Στη συνέχεια οι χάρτες των χαρακτηριστικών συνδυάζονται με εφαρμογή του τελεστή κανονικοποίησης και κατασκευάζονται 3 ενδιάμεσοι **χάρτες προσοχής (conspicuity maps)**, ο χάρτης  $\bar{I}$  για την ένταση, ο χάρτης  $\bar{C}$  για το χρώμα και ο χάρτης  $\bar{O}$  για τον προσανατολισμό, οι οποίοι ορίζονται στην κλίμακα του Χάρτη Ενδιαφέροντος ( $\sigma = 4$ ). Προκύπτουν από την άθροιση επί κλίμακας (**across-scale addition**) που συμβολίζεται με το “ $\oplus$ ” και αποτελείται από τη χωρική μείωση με παρεμβολή (interpolation) κάθε χάρτη στο επίπεδο  $\sigma = 4$  και την άθροιση σημείου με σημείο (point-by-point). Οι χάρτες προσοχής είναι:

$$\bar{I} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(I(c, s)) \quad (\text{εξ. 6.12})$$

$$\bar{C} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))] \quad (\text{εξ. 6.13})$$

Σχετικά με τους χάρτες του προσανατολισμού, αρχικά δημιουργούνται τέσσερις ενδιάμεσοι χάρτες ξεχωριστά για κάθε γωνία  $\theta$ , από τον συνδυασμό των 6 χαρτών

χαρακτηριστικών για τη συγκεκριμένη γωνία, και στη συνέχεια συνδυάζονται σε έναν χάρτη προσοχής ως εξής:

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N \left( \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(O(c, s, \theta)) \right) \quad (\text{εξ. 6.14})$$

Ο λόγος που κατασκευάστηκαν τρεις διαφορετικοί χάρτες προσοχής και που η κανονικοποίηση έγινε ξεχωριστά στον καθένα είναι ότι τα όμοια χαρακτηριστικά ανταγωνίζονται μεταξύ τους ισότιμα για τη σημαντικότητα (**saliency**), ενώ οι διαφορετικές μονάδες (ένταση, χρώμα, προσανατολισμός) ανταγωνίζονται ανεξάρτητα για την επίδραση στον τελικό χάρτη ενδιαφέροντος. Οι τρεις χάρτες προσοχής κανονικοποιούνται και αθροίζονται δίνοντας την τελική είσοδο  $S$  του χάρτη ενδιαφέροντος:

$$S = \frac{1}{3} \cdot [N(\bar{I}) + N(\bar{C}) + N(\bar{O})] \quad (\text{εξ. 6.15})$$

Ο **χάρτης ενδιαφέροντος (saliency map – SM)** που έχει αρχική είσοδο το  $S$ , μεταβάλλει δυναμικά τις τιμές του. Οποιαδήποτε στιγμή, το μέγιστο του χάρτη SM καθορίζει την εμφανέστερη θέση της εικόνας στην οποία θα πρέπει να κατευθυνθεί η εστίαση της προσοχής (**focus of attention - FOA**). Ο χάρτης ενδιαφέροντος τροφοδοτεί ένα νευρωνικό δίκτυο που περιλαμβάνει νευρώνες Winner-Take-All (WTA), μέσω του οποίου η εστίαση της προσοχής μετατοπίζεται σε όλο και λιγότερο σημαντικές περιοχές (με μικρότερο οπτικό ενδιαφέρον).

Όπως προκύπτει από την παραπάνω διαδικασία, ο χάρτης ενδιαφέροντος (SM) κατασκευάζεται σε κλίμακα  $\text{scale} = 4$ , που σημαίνει ότι έχει διαστάσεις  $40 \times 30$ , αν έχουμε ως αρχική είσοδο εικόνα  $640 \times 480$ , δηλαδή ο χάρτης έχει ανάλυση ίση με το  $1 / 2^4$  των αρχικών διαστάσεων. Αυτό βασίζεται στο βιολογικό μοντέλο της ανθρώπινης οπτικής προσοχής και τις διεργασίες που γίνονται στον ανθρώπινο εγκέφαλο. Πολλά μοντέλα που έχουν προταθεί υποστηρίζουν ότι οι νευρώνες που ευθύνονται για την οπτική αντίληψη πραγματοποιούν ενός είδους υποβάθμιση της ανάλυσης (δηλαδή της πολυπλοκότητας) της εικόνας που δέχονται ως ερέθισμα (bottom-up), με σκοπό την απομόνωση σημαντικών περιοχών οι οποίες ξεχωρίζουν “χοντρικά” (coarsely) από τη γειτονιά τους.



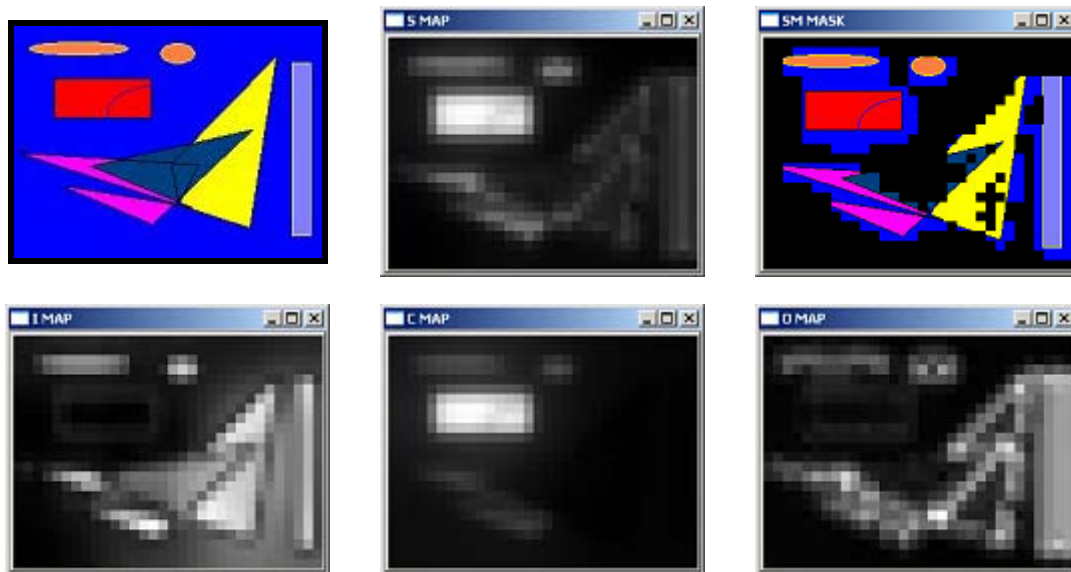
Στη συνέχεια προχωρούν σε επεξεργασία στην πιο λεπτομερή ανάλυση των περιοχών αυτών, π.χ. με διαδικασίες που βασίζονται στη γνώση του στόχου (πιο αργή ανάλυση).

Σύμφωνα με τη συγκεκριμένη εργασία τους, οι Itti et al. ([Itt98]) επιχειρούν τη σύγκριση των αποτελεσμάτων του μοντέλου που προτείνουν με μια άλλη μέθοδο από το μοντέλο του Περιεχομένου Χωρικής Συχνότητας (Spatial Frequency Content - SFC), που βασίζεται στις τοπικές κατανομές της χωρικής συχνότητας που προκύπτουν από το ανθρώπινο μάτι σε περίπτωση εικόνας σε κλίμακα του γκρι (grayscale). Το SFC εμφανίζει μεγαλύτερες τιμές σε περιοχές ενδιαφέροντος απ' ότι στις υπόλοιπες περιοχές (μέση τιμή). Παρατήρησαν τα ίδια αποτελέσματα στον εντοπισμό της προεξέχουσας περιοχής (salient region), όμως στην περίπτωση που δοκίμασαν εικόνες με θόρυβο (είτε λευκό θόρυβο είτε θόρυβο με χρώμα), είχαν σαν αποτέλεσμα το μοντέλο SFC να μην μπορεί να εντοπίσει την σημαντική περιοχή (υπέδειξε πολλές ασήμαντες περιοχές ως σημαντικές), ενώ το μοντέλο Οπτικής Προσοχής να εντοπίζει κανονικά την πιο σημαντική περιοχή (λιγότερο ευαίσθητο στην επίδραση του θορύβου).

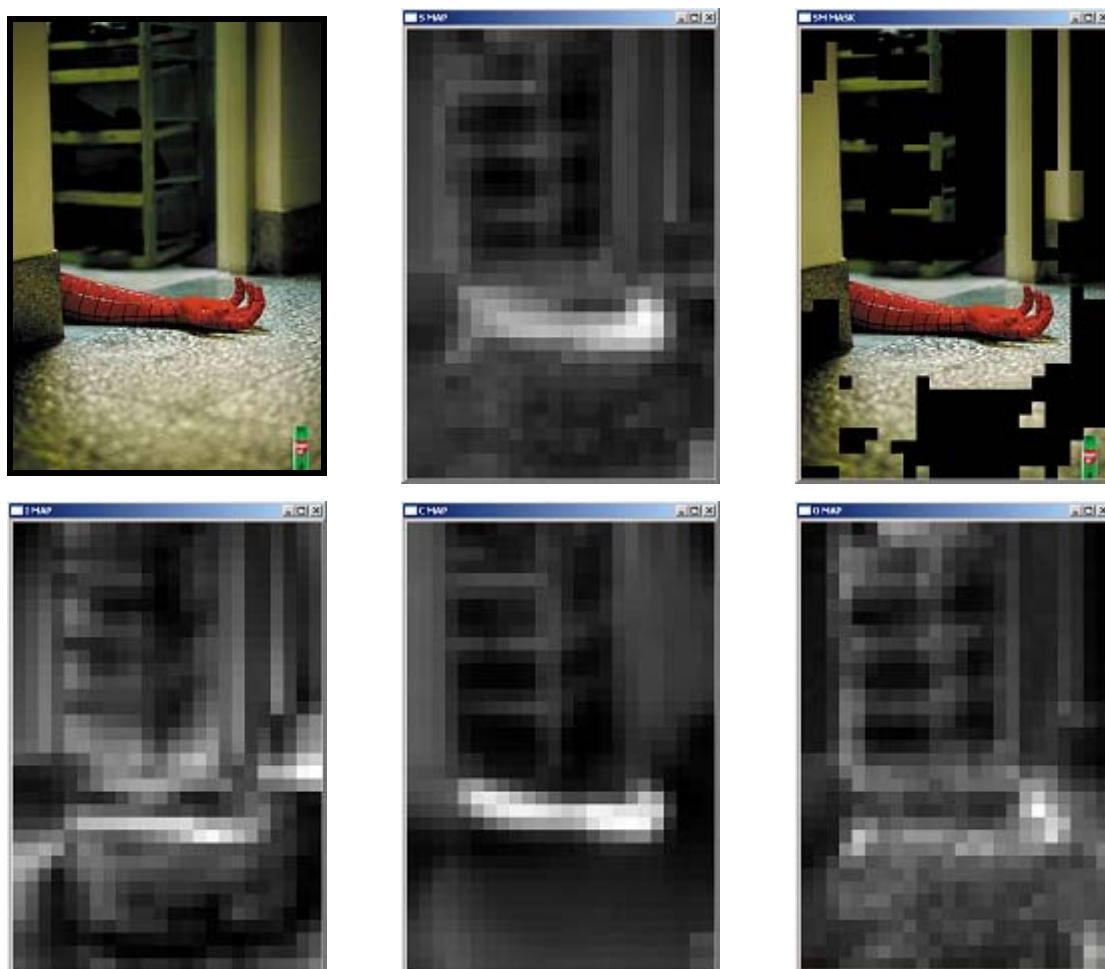
## 6.5 Εφαρμογή της Οπτικής Προσοχής

Εφαρμόζοντας τους υπολογισμούς του μοντέλου Οπτικής Προσοχής που περιγράφηκε αναλυτικά παραπάνω σε τέσσερις εικόνες και παρατηρώντας τα αποτελέσματα για τους ενδιάμεσους χάρτες προσοχής (conspicuity maps I, C, O) και το τελικό αποτέλεσμα (S) μπορούμε να κατανοήσουμε τη λογική του μοντέλου, αλλά και τα βήματα της διαδικασίας, όπως είναι ο τελεστής κανονικοποίησης ο οποίος συμβάλλει στην υποβάθμιση περιοχών που έχουν πολλές μεγάλες τιμές (και όχι τιμή που ξεχωρίζει σε λίγα σημεία), ώστε να μην επηρεάσουν το τελικό αποτέλεσμα με θόρυβο.

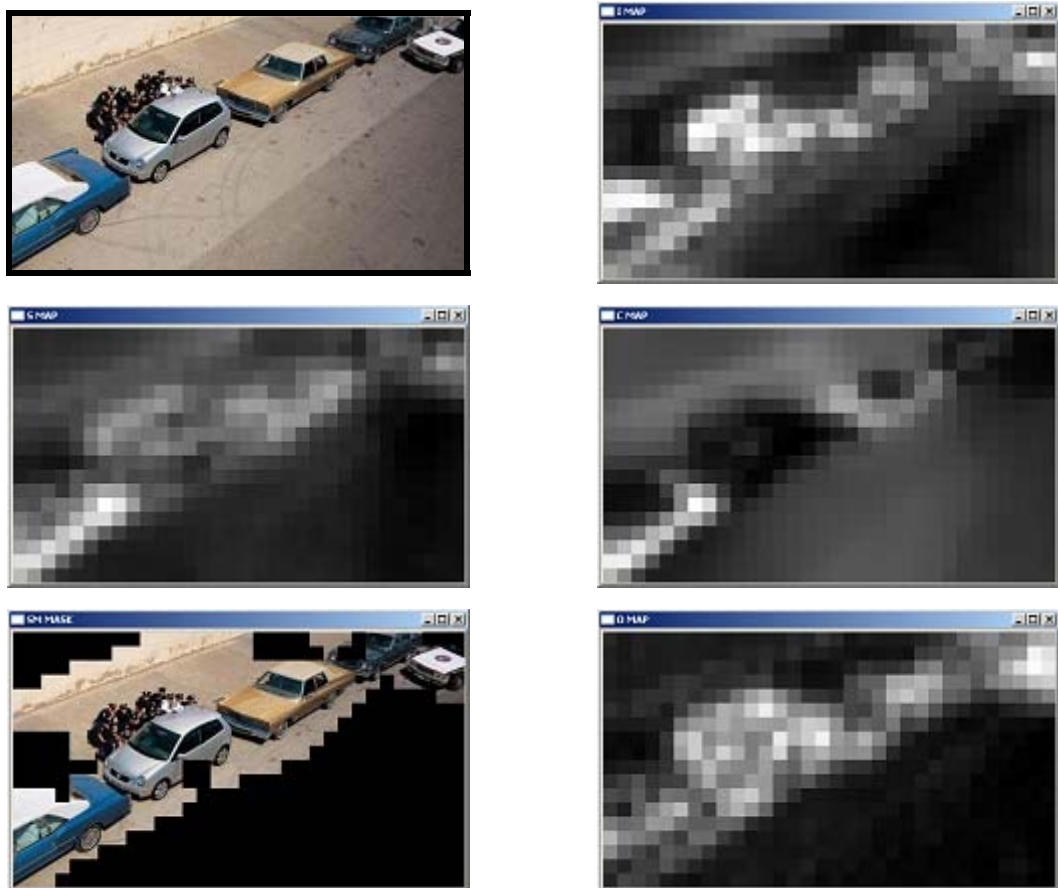
Ανάλογα με τις διαστάσεις κάθε εικόνας, καθορίζεται η κλίμακα (scale) για τους χάρτες (δηλαδή η ανάλυση). Για την απεικόνιση στις διαστάσεις της αρχικής εικόνας χρησιμοποιείται κάποιο είδος παρεμβολής (π.χ. γραμμική). Εδώ χρησιμοποιείται παρεμβολή του πιο κοντινού γείτονα (nearest neighbour - NN), ώστε στις τελικές εικόνες που απεικονίζουν τον χάρτη να διατηρείται εποπτικά η ανάλυση στην κλίμακα του χάρτη ενδιαφέροντος, όπου σύμφωνα με το βιολογικό ισοδύναμο η σημαντική περιοχή επιλέγεται σε μικρή λεπτομέρεια (βλέπε σχήματα 6.5 – 6.8).



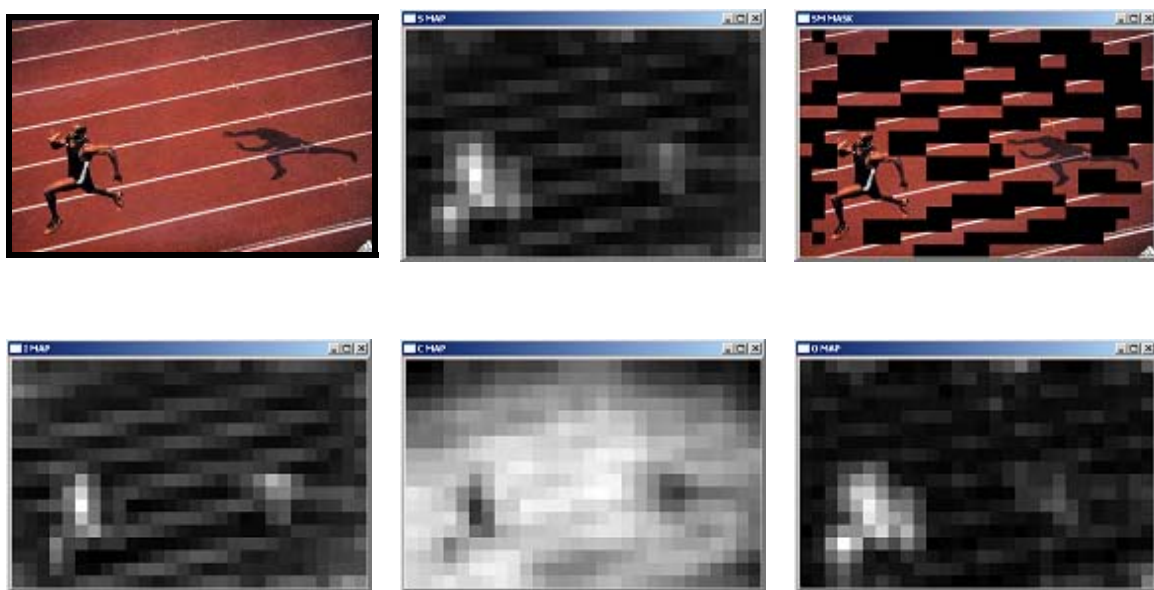
Σχ. 6.5 : Visual Attention - **Εικόνα 1** με ανάλυση 256x192 pixels



Σχ. 6.6 : Visual Attention - **Εικόνα 2** με ανάλυση 384x560 pixels



Σχ. 6.7 : Visual Attention - **Εικόνα 3** με ανάλυση 512x288 pixels



Σχ. 6.8 : Visual Attention - **Εικόνα 4** με ανάλυση 448x288 pixels

### 6.5.1 Αποτελέσματα στις Εικόνες και Παρατηρήσεις

Στη διαδικασία της Οπτικής Προσοχής προσθέτουμε και ένα ακόμη βήμα. Αφού κατασκευαστεί ο χάρτης ενδιαφέροντος SM, οι τιμές του μας δίνουν σε κάθε σημείο (περιοχή) ένα μέτρο για τον βαθμό έλξης της προσοχής της όρασης. Με αυτόν τον τρόπο χωρίζουμε την εικόνα σε δύο είδη περιοχών, στις σημαντικές περιοχές (που έχουν ενδιαφέρον) και στις περιοχές που δεν διακρίνονται άμεσα. Αυτό μπορεί να γίνει με πολλούς τρόπους, όπως με ένα νευρωνικό δίκτυο (ή και να χωριστεί η εικόνα σε περισσότερες περιοχές). Εδώ επιλέγουμε τον καθορισμό μιας τιμής κατώφλιου, η οποία καθορίζει το όριο της σημαντικής περιοχής (salient region), αφού σε σημεία που δίνουν τιμές στον χάρτη μικρότερες από το κατώφλι, μηδενίζουμε (“μαυρίζουμε”,  $r = g = b = 0$ ) το αντίστοιχο pixel, αλλιώς το αφήνουμε όπως έχει (εφαρμογή μάσκας που επιτρέπει μόνο σημαντικές περιοχές να εμφανίζονται στην έξοδό της). Ως κατώφλι λαμβάνουμε την τιμή:

$$mask = mean(Smap) + c \cdot std(Smap) \quad (\text{εξ. 6.16})$$

όπου  $mean$  είναι η μέση τιμή και  $std$  η τυπική απόκλιση των τιμών του χάρτη ενδιαφέροντος που συμβολίζεται  $Smap$ .

Η παράμετρος  $c$  συνήθως παίρνει τιμές μεταξύ  $-1$  και  $+1$ . Στη συγκεκριμένη υλοποίηση, μετά από πειράματα σε εικόνες και σε καρέ από διάφορα βίντεο, επιλέχθηκε η τιμή που έδωσε τα πιο ικανοποιητικά αποτελέσματα (ώστε το κατώφλι να επιτρέπει πραγματικά σημαντικές περιοχές να εμφανίζονται στην έξοδο). Η τιμή που χρησιμοποιήθηκε σε όλα τα πειράματα για την παράμετρο  $c$  είναι:

$$c = -0.2$$

Μελετώντας αναλυτικά τις εικόνες των παραπάνω σχημάτων, ξεχωριστά για κάθε περίπτωση καταλήγουμε στις παρακάτω παρατηρήσεις (όπου αρχική εικόνα ονομάζουμε την εικόνα-είσοδο στο μοντέλο και τελική εικόνα εκείνη που προκύπτει από την εφαρμογή της παραπάνω μάσκας):

**Εικόνα 1:** Στο σχήμα 6.5 έχουμε στην πρώτη σειρά την αρχική εικόνα 1 (είσοδος στο σύστημα), τον χάρτη ενδιαφέροντος ή σημαντικότητας και την εφαρμογή του χάρτη πάνω στην εικόνα με την μάσκα που αναφέρθηκε παραπάνω. Στην δεύτερη σειρά βρίσκονται οι χάρτες προσοχής, για την ένταση (I), το χρώμα (C) και τον προσανατολισμό (O) αντίστοιχα. Με την πρώτη ματιά βλέπουμε ότι στην τελική εικόνα με την μάσκα έχουν παραμείνει μόνο σημαντικές περιοχές, που είναι τα σχήματα της εικόνας χωρίς τον μπλε φόντο. Πιο αναλυτικά, το κόκκινο ορθογώνιο όπως και το μπλε φόντο έχουν την ίδια χαμηλή ένταση γι' αυτό δίνουν χαμηλή τιμή στο χαρακτηριστικό I, στο εσωτερικό αλλά και στο όριο μεταξύ τους (βλέπε χάρτη I). Όμως το ορθογώνιο διαφέρει σε χρώμα από τη γειτονιά του γι' αυτό δίνει μεγάλη τιμή στον χάρτη C, όπου μετά τον τελεστή κανονικοποίησης είναι και η μοναδική περιοχή που είναι σημαντική σε χρώμα. Ο χάρτης προσανατολισμού O προκύπτει από τον I, γι' αυτό δίνει τις ακμές των σχημάτων εκτός από το κόκκινο ορθογώνιο που έχει μικρή τιμή I. Έτσι στον τελικό χάρτη λαμβάνουν μέρος όλα τα χαρακτηριστικά, αλλά επειδή το κόκκινο ορθογώνιο στον C είναι το μόνο με υψηλές τιμές, η νέα κανονικοποίηση (βλέπε τύπο 8) θα προάγει τον χάρτη C περισσότερο από τους υπόλοιπους και έτσι τελικά στον χάρτη ενδιαφέροντος (Smap) η περιοχή αυτή θα είναι η πιο ξεχωριστή, ενώ οι υπόλοιπες δεν χάνονται αλλά υπάρχουν στον χάρτη Smap με μικρότερες τιμές (όπως φαίνεται και από το σχήμα).

**Εικόνα 2:** Στο σχήμα 6.6, όπως και στο προηγούμενο, έχουμε την αρχική εικόνα 2, τον χάρτη ενδιαφέροντος και την εικόνα με εφαρμογή της μάσκας στην πρώτη σειρά, ενώ στην δεύτερη σειρά βρίσκονται οι χάρτες προσοχής (I, C, O αντίστοιχα). Στην εικόνα με τη μάσκα βλέπουμε ότι έχει παραμείνει μεγάλο μέρος της αρχικής εικόνας, διότι το κόκκινο κομμάτι έχει μεγάλες τιμές στο C (αντίθεση με τη γειτονιά του), κάποιες περιοχές έχουν αντίθεση στην ένταση (μεγάλες τιμές στο I), ενώ άλλες δίνουν μεγάλες τιμές στο O (λόγω αντίθεσης στον προσανατολισμό τους). Τελικά όμως στην τελική εικόνα έχουν απορριφθεί περιοχές με ομοιόμορφο περιεχόμενο όπως φαίνεται και από την τελική εικόνα στο σχήμα.

**Εικόνα 3:** Στο σχήμα 6.7 η διάταξη είναι κατά στήλη, στην πρώτη στήλη η αρχική εικόνα 3, ο χάρτης ενδιαφέροντος και η τελική εικόνα, ενώ στη δεύτερη στήλη οι χάρτες προσοχής (I, C, O αντίστοιχα). Εδώ φαίνεται ότι περιοχές που διαφέρουν από τη γειτονιά τους σε φωτεινότητα (μεγάλες τιμές στο I) είναι τα δύο λευκά και το ασημί αυτοκίνητο,

όπως και το πλήθος αστυνομικών. Σε προσανατολισμό μεγάλες τιμές δίνουν όλα τα αντικείμενα, αυτοκίνητα και πλήθος. Σε χρώμα πιο πολύ ξεχωρίζουν το μπεζ και το θαλασσί αυτοκίνητο. Στον χάρτη saliency map όλες οι σημαντικές περιοχές έχουν κάποια επίδραση και στη τελική εικόνα το αποτέλεσμα είναι το ιδανικό, διαχωρίζοντας ως σημαντικά τα σωστά τμήματα που ελκύουν την προσοχή και προκαλούν εντύπωση.

**Εικόνα 4:** Στο σχήμα 6.8, όπου η διάταξη είναι κατά σειρές όπως και στις δύο πρώτες εικόνες, φαίνεται η επίδραση του τελεστή κανονικοποίησης, αντίθετη από αυτή στην εικόνα 1. Ο χάρτης C έχει στα περισσότερα σημεία του μεγάλες τιμές (λευκό χρώμα) και ο τελεστής κανονικοποίησης καταστέλλει συνολικά ολόκληρο τον χάρτη, με αποτέλεσμα να μην προσθέτει θόρυβο στον χάρτη saliency, οπότε και θα κάλυπτε τις προεξέχουσες περιοχές που προέρχονται από τους άλλους δύο χάρτες. Ο χάρτης O, που βασίζεται στην ένταση I, διακρίνει την περιοχή του αθλητή και όχι τις γραμμές του δαπέδου, οι οποίες δεν έχουν γύρω τους σημεία με διαφορετικό προσανατολισμό και γι' αυτό το φίλτρο Gabor δεν δίνει μεγάλες τιμές. Τελικά ο χάρτης C δεν επηρεάζει καθόλου το αποτέλεσμα (λόγω του τελεστή που εφαρμόζεται στην εξ. 6.13), και η τελική εικόνα έχει περιοχές σημαντικές σε ένταση (I) και σημαντικές σε προσανατολισμό (O), όπως είναι οι γραμμές, η σκιά, ο αθλητής (μεγάλο I), και ο αθλητής (μεγάλο O) αντίστοιχα.

### 6.5.2 Εφαρμογή στην Εξαγωγή Χαρακτηριστικών από το Καρέ

Στην εργασία αυτή, όπως είπαμε και στην αρχή του Κεφαλαίου, ο σκοπός της προηγούμενης ανάλυσης για την εφαρμογή της οπτικής προσοχής (Visual Attention) είναι ο εντοπισμός των σημαντικών περιοχών σε κάθε καρέ της ακολουθίας του βίντεο και η εξαγωγή των χαρακτηριστικών (χρωματικό ιστόγραμμα για τα πεδία Y, U και V) μόνο από αυτές τις περιοχές, οι οποίες περιέχουν καθοριστική πληροφορία για την σημασία του συγκεκριμένου καρέ.

Υλοποιώντας την μέθοδο αυτή πριν την εξαγωγή των χαρακτηριστικών και την κατασκευή της καμπύλης στο χώρο, σε κάθε καρέ εφαρμόζουμε τη μάσκα που προκύπτει από τον χάρτη ενδιαφέροντος και έτσι κρατάμε μέσα στο καρέ μόνο τις σημαντικές περιοχές (όλες οι υπόλοιπες παίρνουν συγκεκριμένο Y, U, V, δηλαδή ίδια pixels π.χ. με χρώμα μαύρο). Έτσι οι διακυμάνσεις στο ιστόγραμμα προέρχονται από μεταβολές που

γίνονται σε αξιοπρόσεκτο περιεχόμενο της εικόνας (π.χ. αντικείμενα) και όχι σε μεγάλες ομοιόμορφες περιοχές, όπου είναι λιγότερο πιθανό να εστιάσει το ανθρώπινο μάτι.

Κατά την εξέλιξη του βίντεο, η οπτική προσοχή θα παρακολουθεί μία συγκεκριμένη περιοχή και τις αλλαγές που γίνονται στο ιστόγραμμά της. Όσο μετατοπίζεται η περιοχή αυτή, θα μετατοπίζεται και η οπτική προσοχή. Το ιστόγραμμα θα μεταβληθεί σε περίπτωση που κάποιο χαρακτηριστικό αλλάξει μέσα στη σημαντική περιοχή, όπως η φωτεινότητα ή το χρώμα. Μπορεί ακόμα να συμβεί μεταβολή στα όρια της περιοχής, οπότε να μεταβληθεί το εύρος της ή να προστεθούν νέα σημεία, αν κάποιο εντυπωσιακό γεγονός ξεκινήσει να συμβαίνει σε κάποιο σημείο της εικόνας (π.χ. είσοδος ενός αντικειμένου).

Στις παραπάνω περιπτώσεις θα συμβεί μεταβολή του ιστογράμματος και θα δημιουργηθούν σημεία με ακραία συμπεριφορά στη νέα καμπύλη που κατασκευάζεται. Η καμπύλη αυτή έχει σημεία που παρουσιάζουν μικρότερες διακυμάνσεις απ' ότι η αρχική καμπύλη, και αυτό διότι τώρα στα bins του ιστογράμματος δεν υπολογίζονται οι περιοχές που απορρίπτονται. Άρα η νέα καμπύλη προκύπτει πιο ομαλή, εκφράζοντας την εξέλιξη του ιστογράμματος μόνο από τη σημαντική περιοχή, και παρουσιάζει σημεία απότομης καμπύλωσης σπανιότερα από την αρχική. Η καμπυλότητα θα δίνει τοπικά μέγιστα στα καρέ όπου αλλάζει με επιταχυνόμενο ρυθμό το περιεχόμενο των σημαντικών περιοχών και δεν θα επηρεάζεται από μεταβολές που γίνονται σε λιγότερο εντυπωσιακές περιοχές.

Τελικά η νέα καμπύλη θα είναι πιο περιεκτική όσον αφορά τις νοηματικές αλλαγές (αυτός είναι και ο σκοπός του μοντέλου οπτικής προσοχής), απ' ότι αν παίρναμε χαρακτηριστικά από ολόκληρη την εικόνα. Ευελπιστούμε με την μεθοδολογία αυτή να έχουμε βελτιωμένα αποτελέσματα στο σύστημά μας ως προς την επιλογή των πιο κατάλληλων χαρακτηριστικών καρέ. Αναλυτικά αποτελέσματα από την εφαρμογή δίνονται στο Κεφάλαιο που περιλαμβάνει τα πειράματα.





## ΚΕΦΑΛΑΙΟ 7

# *Πειράματα, Αποτελέσματα και Παρατηρήσεις*

Στα πλαίσια της παρούσας εργασίας υλοποιήθηκε ένα υπολογιστικό σύστημα που περιλαμβάνει το σύστημα κανόνων για την ανίχνευση αλλαγής πλάνου, τους ορισμούς για την καμπυλότητα, τις δύο άλλες μεθόδους για την εξαγωγή χαρακτηριστικών σημείων καμπύλης και το μοντέλο οπτικής προσοχής που εφαρμόζεται σε διδιάστατη εικόνα. Ο σκοπός της υλοποίησης αυτής είναι η διεξαγωγή πειραμάτων πάνω σε βίντεο από συγκεκριμένο δείγμα και η εξαγωγή των αριθμητικών μετρήσεων όπως προκύπτουν από τις διάφορες μεθόδους. Επιδιώκεται η αξιολόγηση των αριθμητικών αποτελεσμάτων, αλλά και η εποπτική αξιολόγηση των καρτέ που προκύπτουν στην έξοδο του συστήματος, είτε ως καρτέ αλλαγής πλάνου είτε ως χαρακτηριστικά καρτέ του βίντεο. Γίνεται έτσι δυνατή η σύγκριση μεταξύ των μεθόδων της καμπυλότητας και των δύο άλλων εναλλακτικών μεθόδων με σκοπό την εξαγωγή συμπερασμάτων σχετικά με την καταλληλότητα και την αποτελεσματικότητά τους.

Η υλοποίηση περιλαμβάνει τη διαδικασία της ανίχνευσης των πλάνων απευθείας στην κωδικοποιημένη μορφή του MPEG. Από τα frames, που βρίσκονται σε σειρά μετάδοσης (Transmit Order) και διαβάζονται από το σύστημα χωρίς να αλλάξει η σειρά ή να αποκωδικοποιηθούν, προκύπτουν μεγέθη (όπως οι DC τιμές ή το πλήθος των διανυσμάτων κίνησης) που τροφοδοτούν τους κανόνες και τελικά η απόφαση για αλλαγή πλάνου προκύπτει real-time ταυτόχρονα με την ανάγνωση του MPEG. Αφού χωριστεί το βίντεο σε πλάνα, ακολουθεί η αποκωδικοποίηση κάθε καρτέ και η κατασκευή του τρέχοντος σημείου του διανύσματος χαρακτηριστικών από τα στοιχεία του χρωματικού

ιστογράμματος (Y, U, V) του καρέ. Αφού κατασκευαστεί το διάνυσμα χαρακτηριστικών για ένα πλάνο (δηλαδή η καμπύλη στον χώρο), εφαρμόζονται οι μέθοδοι της καμπυλότητας, η μέθοδος της απλοποίησης καμπύλης και η μέθοδος προσέγγισης περιγράμματος. Υπάρχει η επιλογή της οπτικής προσοχής, που εφαρμόζεται μετά την αποκωδικοποίηση του καρέ και πριν την εξαγωγή των χαρακτηριστικών. Το διάνυσμα των χαρακτηριστικών μπορεί να ομαλοποιηθεί με κάποιο από τα φίλτρα που αναφέρθηκαν στην αντίστοιχη παράγραφο, ενώ υπάρχει η δυνατότητα να ομαλοποιηθεί η πρώτη ή η δεύτερη παράγωγός του, ή ακόμα και η τιμή της καμπυλότητας που εξάγεται τελικά. Το σύστημα, με είσοδο ένα βίντεο σε μορφή MPEG1, δίνει στην έξοδό του τα καρέ που δείχνουν την αλλαγή πλάνου, το διάνυσμα χαρακτηριστικών για κάθε πλάνο (δηλαδή τις συντεταγμένες των σημείων στον πολυδιάστατο χώρο), τις τιμές της καμπυλότητας (αν εφαρμοστεί κάποια μέθοδος καμπυλότητας) και τέλος τα χαρακτηριστικά καρέ που προκύπτουν από τη μέθοδο που χρησιμοποιείται.

## 7.1 Η Υλοποίηση της Εφαρμογής

Ολόκληρη η εφαρμογή αυτής της εργασίας έχει υλοποιηθεί σε γλώσσα προγραμματισμού C, ενώ η πλατφόρμα που χρησιμοποιήθηκε είναι το περιβάλλον προγραμματισμού “Microsoft Visual Studio, 6.0 release”. Για την επεξεργασία και την ανάλυση αρχείων εικόνας και αρχείων βίντεο χρησιμοποιήθηκαν δομές, τύποι, μεταβλητές και συναρτήσεις από δύο διαφορετικές έτοιμες (υλοποιημένες σε κώδικα C) βιβλιοθήκες.

Η πρώτη είναι η βιβλιοθήκη Dali VM (Dali Virtual Machine του Πανεπιστημίου Cornell), που περιέχει ρουτίνες που χειρίζονται εικόνες σε μορφή YUV και RGB, καθώς και βίντεο σε μορφή MPEG1. Οι συναρτήσεις που ορίζει μπορούν να κατασκευάσουν και να χρησιμοποιήσουν δεδομένα από αρχεία MPEG video, MPEG audio, MPEG system, JPEG, GIF, PPM, και WAV αρχεία ήχου, ή ακόμα και βίντεο σε μορφή AVI. Οι δομές δεδομένων που χρησιμοποιεί αποδεικνύονται χρήσιμα εργαλεία για τις συγκεκριμένες επεξεργασίες, ενώ ο χρόνος ανάγνωσης και αποκωδικοποίησης του MPEG είναι αρκετά γρήγορος σε σχέση με άλλες υλοποιήσεις ([wwwDaliVM]). Η βιβλιοθήκη αυτή χρησιμοποιήθηκε για την ανάγνωση και αποκωδικοποίηση του MPEG,

για την μέθοδο της ανίχνευσης των πλάνων και γενικά για τον χειρισμό των δομών των πεδίων Y, U, V της εικόνας.

Η δεύτερη βιβλιοθήκη χρησιμοποιήθηκε κυρίως για την εφαρμογή του υπολογιστικού μοντέλου της οπτικής προσοχής, καθώς προσφέρει υλοποιημένες μεθόδους επεξεργασίας εικόνας. Προέρχεται από την ερευνητική ομάδα της εταιρείας Intel, ονομάζεται OpenCV (Intel Open Source Computer Vision Library) και αποτελείται και αυτή από συναρτήσεις σε γλώσσα C και μερικές κλάσεις στην αντικειμενοστραφή γλώσσα C++, οι οποίες εκτός του ότι ορίζουν δομές και ρουτίνες για τον χειρισμό των δεδομένων της εικόνας, υλοποιούν επίσης και πολλές μεθόδους και αλγορίθμους που σχετίζονται με την Επεξεργασία Εικόνας και την Όραση Υπολογιστών ([wwwOpenCV]). Το πλεονέκτημα της OpenCV είναι η ανεξαρτησία από άλλες βιβλιοθήκες και κυρίως οι υλοποιήσεις που παρέχει, όπως είναι οι Γκαουσιανές πυραμίδες (Gaussian pyramids), η παρεμβολή (linear, nearest neighbour) και οι μορφολογικές διεργασίες, που φάνηκαν χρήσιμες για το σύστημα Visual Attention.

Τα αρχεία που είναι κωδικοποιημένα με το πρότυπο MPEG 1 (για τις προδιαγραφές και την προτυποποίηση βλέπε βιβλιογραφία, [ISO93]), αποτελούνται από μια ακολουθία δεδομένων, το ονομαζόμενο System Stream (συνολική πληροφορία σε σειρά), το οποίο περιλαμβάνει ένα “ρεύμα εικόνας” (Video Stream) και ένα “ρεύμα ήχου” (Audio Stream). Αυτά τα δύο διαφορετικά είδη δεδομένων βρίσκονται πεπλεγμένα μέσα στην συνολική ακολουθία, με τη μορφή πακέτων που επαναλαμβάνονται ανά τακτά χρονικά διαστήματα. Από τις δύο μορφές πληροφορίας απομονώνουμε το βίντεο, το οποίο και θα επεξεργαστούμε στη συνέχεια, αναλύοντας ένα προς ένα τα καρέ που προέρχονται από τα πακέτα βίντεο και λαμβάνοντας τελικά την ακολουθία των MPEG frames (I, P και B frames) μαζί με κάθε πληροφορία που μεταφέρουν. Το σύστημα που υλοποιήθηκε εδώ υποστηρίζει μορφή video stream MPEG 1 σταθερού ρυθμού μετάδοσης δεδομένων (constant bitrate), ανεξάρτητα από την τιμή του (δεν περιλαμβάνει χειρισμό variable bitrate). Κανονικά το MPEG 1 σύμφωνα με το πρότυπο υποστηρίζει ακολουθία βίντεο ρυθμού μέχρι 1.25 Mbits/sec και ανάλυσης μέχρι και 352x240, αλλά αυτό δεν σημαίνει ότι δεν μπορεί να κωδικοποιηθεί βίντεο στην μορφή MPEG 1 με μεγαλύτερη ανάλυση ή μεγαλύτερο ρυθμό δεδομένων, απλά τα χαρακτηριστικά του δεν ικανοποιούν τους περιορισμούς (δεν είναι constrained, δηλαδή δεν είναι μέσα στα τυπικά όρια του προτύπου).

## 7.2 Πειράματα στην Ανίχνευση Πλάνων

### 7.2.1 Χαρακτηριστικά του Δείγματος Βίντεο

Για την εφαρμογή του αυτόματου συστήματος ανίχνευσης πλάνων σε ακολουθία κινούμενων εικόνων όπως περιγράφηκε στο Κεφάλαιο 2 χρησιμοποιήθηκαν οι τιμές για τα κατώφλια (thresholds) που αναφέρονται στην παράγραφο 2.4.4, έτσι ώστε το σύστημα να μην χρειάζεται καθοδήγηση από εξωτερικό παράγοντα και να λειτουργεί όσο το δυνατόν πιο αποτελεσματικά. Υπάρχει βέβαια και η δυνατότητα επίβλεψης με σκοπό την προσαρμογή του συστήματος σε κάθε περίπτωση, ανάλογα με το είδος του περιεχομένου του βίντεο που δέχεται στην είσοδό του. Για παράδειγμα οι τιμές των thresholds θα μπορούσαν να επιλέγονται από τον χρήστη πριν την μέθοδο (με βάση την προσωπική του γνώση), ή ακόμα να καθορίζεται από τον χρήστη το περιεχόμενο του βίντεο (π.χ. αθλητικός αγώνας, ειδήσεις, ταινία δράσης, κτλ.), οπότε ένα σύστημα απόφασης ανάλογα με τα δεδομένα (γνώση του επιβλέποντα, περιεχόμενο του βίντεο, είδος δραστηριότητας, βαθμός κίνησης) να επιλέγει κατάλληλες τιμές και να βελτιώνει το επιθυμητό αποτέλεσμα.

Πρέπει να σημειωθεί ότι το σύστημα των κανόνων που έχει διατυπωθεί στο Κεφάλαιο 2 για την επίλυση του προβλήματος του εντοπισμού αλλαγής πλάνου, απαιτεί στην είσοδό του MPEG 1 βίντεο με σταθερή ακολουθία GOP (Group-of-Pictures), δηλαδή σταθερό αριθμό και δομή επαναλαμβανόμενων I, P και B frames, δηλαδή δεν είναι σχεδιασμένο για να ανταποκρίνεται σε οποιαδήποτε μεταβαλλόμενη ακολουθία. Επίσης, απαραίτητη προϋπόθεση είναι η παρουσία δύο τουλάχιστον B-frames μεταξύ των R-frames (I ή P), αφού αυτά είναι που θα καθορίσουν την τελική απόφαση (ενώ δουλεύει και με ένα B-frame, αλλά δεν εξάγει αξιόπιστα αποτελέσματα).

Το δείγμα που χρησιμοποιείται για τα πειράματα της ανίχνευσης αλλαγής πλάνου (αφορά στις απότομες αλλαγές, αιχμές-cuts) αποτελείται από δέκα (10) αρχεία βίντεο κωδικοποιημένα με το πρότυπο MPEG 1. Τα αρχεία αυτά έχουν συνολική διάρκεια 54 λεπτά και 56 δευτερόλεπτα, καταλαμβάνουν 617 MB στον δίσκο και το καθένα έχει μέγεθος από 34 MB έως και 96 MB περίπου. Οι απότομες αλλαγές πλάνου (cuts) που περιέχονται σε αυτά είναι συνολικά 516 στο πλήθος, ενώ σε κάθε βίντεο συμβαίνουν από 24 μέχρι και 135 απότομες αλλαγές πλάνου. Τα χαρακτηριστικά του MPEG 1 είναι διαφορετικά για κάθε αρχείο, με ανάλυση 320x240 ή 384x288 pixels, ρυθμό πλαισίων 25

ή 29.97 frames/sec, ρυθμό δεδομένων από 1000 μέχρι 2048 Kbits/sec και διάρκεια που κυμαίνεται από 4 έως 10 λεπτά περίπου.

Παρακάτω δίνεται η λίστα με τα αρχεία MPEG και η ονομασία τους. Αμέσως μετά στον πίνακα που ακολουθεί βρίσκονται τα χαρακτηριστικά του καθενός αναλυτικά:

#### **Αρχεία MPEG1 για Ανίχνευση Πλάνων (cuts)**

1 <sup>ο</sup> αρχείο	:	MRBEAN
2 <sup>ο</sup> αρχείο	:	ANT1NEWS
3 <sup>ο</sup> αρχείο	:	ALTERNEWS
4 <sup>ο</sup> αρχείο	:	NETNEWS1
5 <sup>ο</sup> αρχείο	:	NETNEWS2
6 <sup>ο</sup> αρχείο	:	SPORTS
7 <sup>ο</sup> αρχείο	:	THEATER
8 <sup>ο</sup> αρχείο	:	ET1
9 <sup>ο</sup> αρχείο	:	ABCNEWS
10 <sup>ο</sup> αρχείο	:	FOOTBALL

<b>Βίντεο</b>	<b>Μέγεθος</b>	<b>Ανάλυση</b>	<b>Kbits/sec</b>	<b>Fps</b>	<b>Διάρκεια</b>	<b>Frames</b>	<b>Πλάνα</b>
<b>1</b>	57.7 MB	384x288	1024 Kbps	25	0:06:20	9508	<b>55</b>
<b>2</b>	62.2 MB	384x288	2048 Kbps	25	0:03:45	5638	<b>25</b>
<b>3</b>	82.3 MB	384x288	2048 Kbps	25	0:04:58	7468	<b>55</b>
<b>4</b>	34.7 MB	384x288	1024 Kbps	25	0:03:48	5713	<b>43</b>
<b>5</b>	50.1 MB	320x240	1024 Kbps	25	0:05:30	8263	<b>50</b>
<b>6</b>	95.9 MB	384x288	2000 Kbps	25	0:05:56	8909	<b>33</b>
<b>7</b>	82.7 MB	320x240	1000 Kbps	29.97	0:09:18	13965	<b>136</b>
<b>8</b>	44.5 MB	384x288	1000 Kbps	25	0:05:00	7515	<b>62</b>
<b>9</b>	56.3 MB	384x288	1500 Kbps	25	0:04:29	8089	<b>35</b>
<b>10</b>	50.7 MB	384x288	1000 Kbps	25	0:05:42	8556	<b>32</b>

**Πίνακας 7.1 :** Χαρακτηριστικά για κάθε βίντεο του δείγματος ανίχνευσης πλάνων

Η δομή των frames (GOP structure) στα παραπάνω MPEG βίντεο, εκτός από το 7 και το 9 είναι:

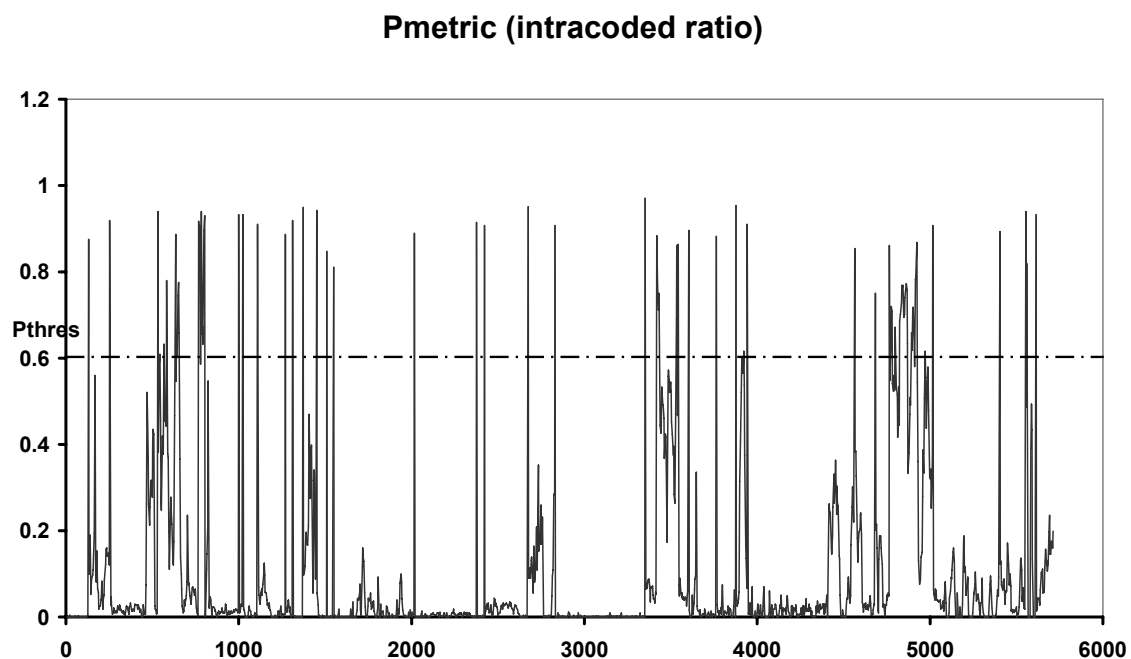
I B B P B B P B B P B B

Στα βίντεο 7 και 9 το GOP είναι:

I B B P B B P B B

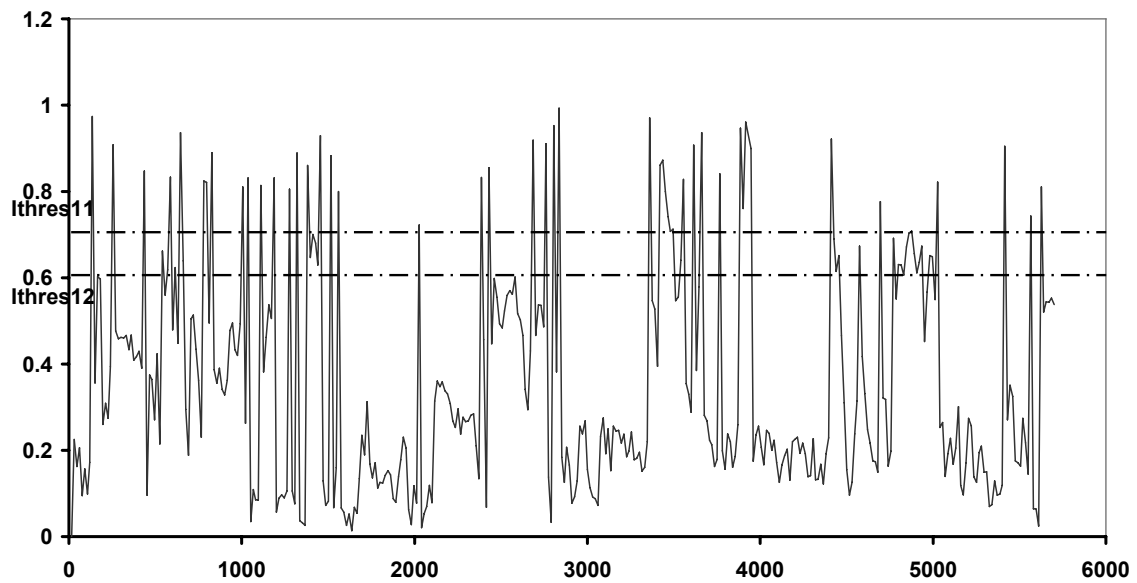
### 7.2.2 Ενδεικτικά Διαγράμματα

Εδώ γίνεται αναλυτική παρουσίαση για το 4<sup>ο</sup> βίντεο (NETNEWS1), για το οποίο κατασκευάζονται τα διαγράμματα, που αντιστοιχούν στους κανόνες που παρουσιάζονται στην παράγραφο 2.4 και αφορούν στα μέτρα σύγκρισης για τα I-frames:  $Imetric_1$  (“χοντρική” διαφορά των frames),  $Imetric_2$  (τετραγωνική διαφορά),  $Imetric_3$  (στατιστική διαφορά για το U-πεδίο),  $Imetric_4$  (στατιστική διαφορά για το V-πεδίο), όπως και στο μέτρο  $Pmetric$  (ποσοστό των intracoded Macroblocks) για τα P-frames. Όλα τα προηγούμενα έχουν οριστεί αναλυτικά στο Κεφάλαιο 2. Στα επόμενα πέντε σχήματα φαίνονται τα διαγράμματα αυτά (NETNEWS1), όπου αναφέρεται το μέτρο που παριστάνουν γραφικά και το κατώφλι ή τα κατώφλια που χρησιμοποιούνται για αυτό.



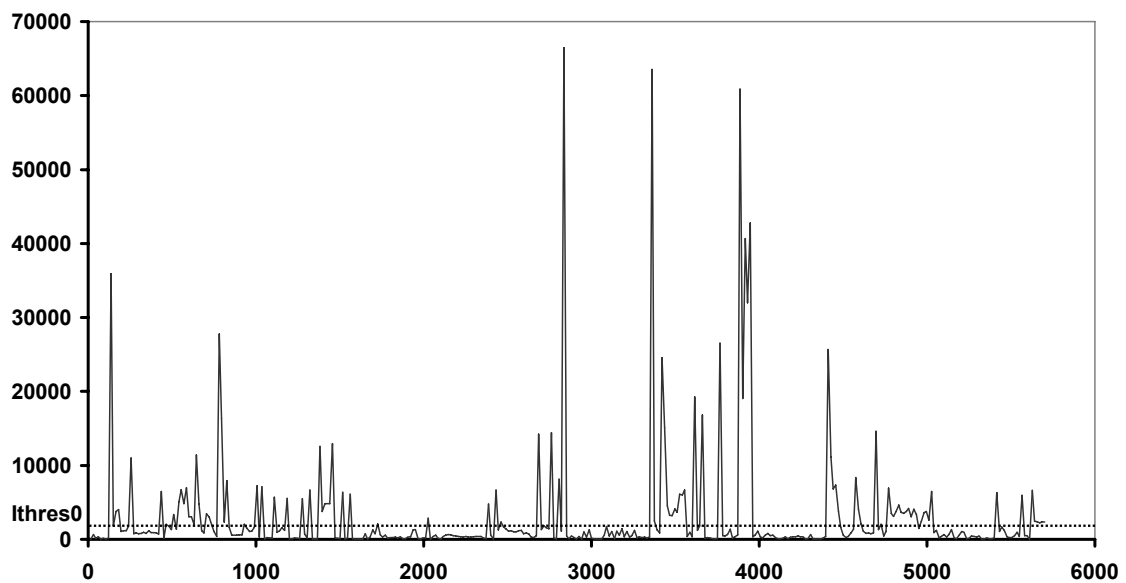
Σχ. 7.1 : Διάγραμμα  $Pmetric$  = Ποσοστό των intracoded MB

### Imetric 1 (coarse difference)



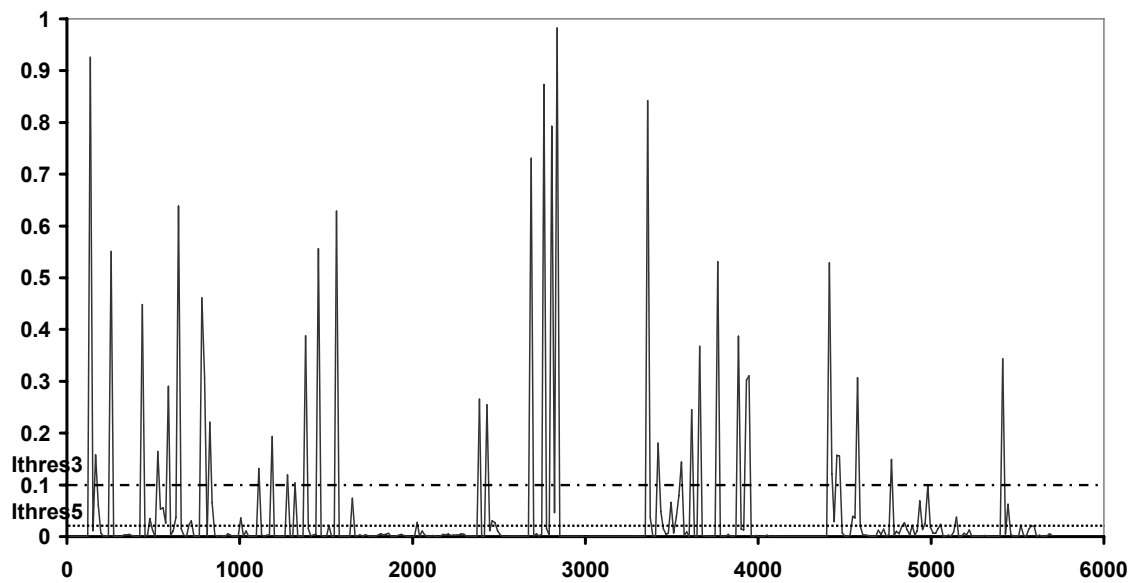
Σχ. 7.2 : Διάγραμμα Imetric1 = “Χοντρική” Διαφορά

### Imetric 2 (squared difference)



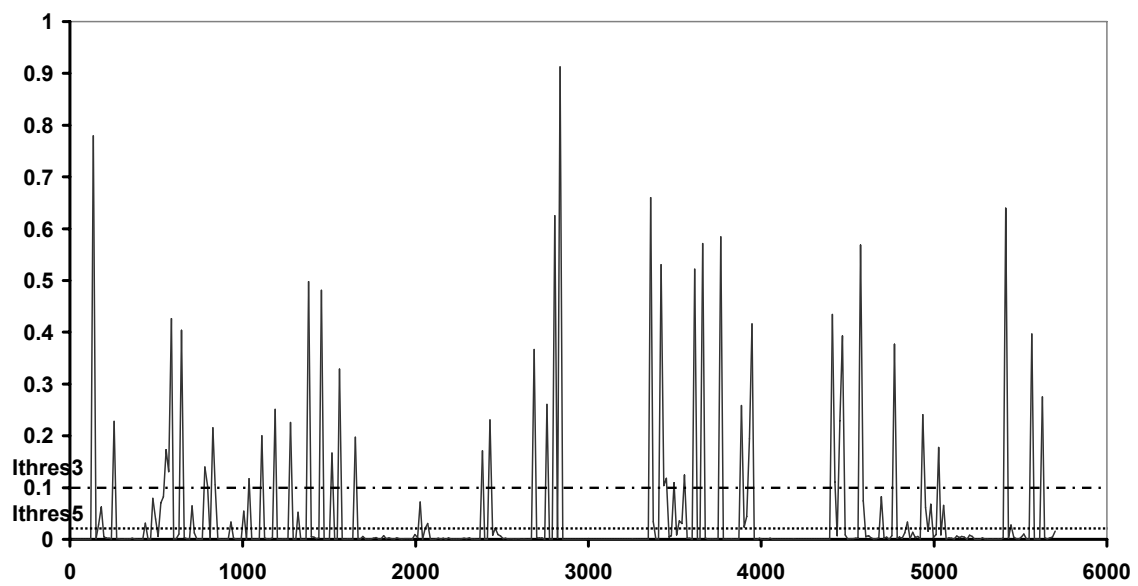
Σχ. 7.3 : Διάγραμμα Imetric2 = Τετραγωνική Διαφορά

### Imetric 3 (statistical difference U)



Σχ. 7.4 : Διάγραμμα Imetric3 = Στατιστική Διαφορά U-πεδίου

### Imetric 4 (statistical difference V)

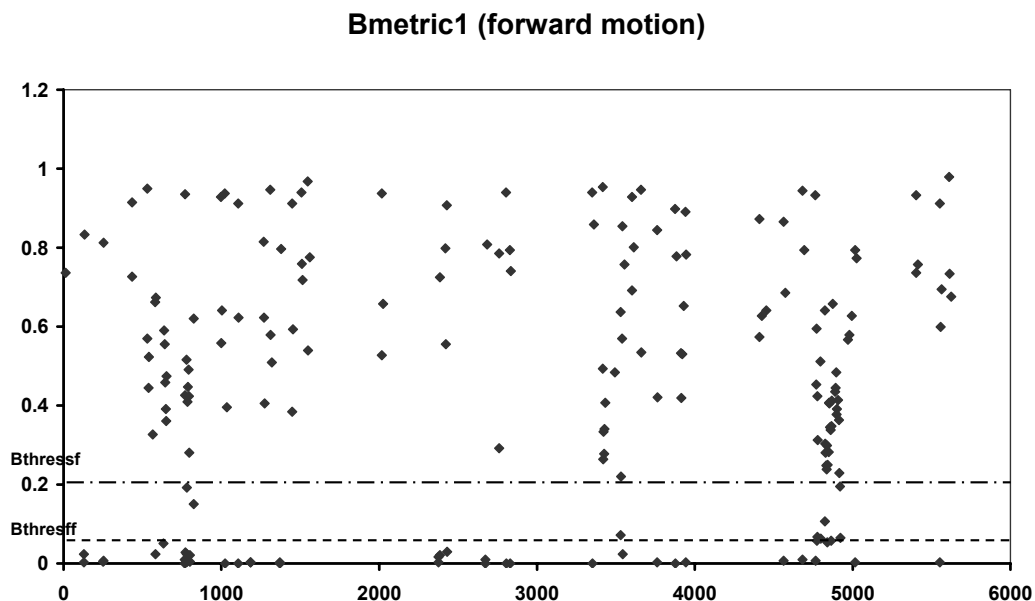


Σχ. 7.5 : Διάγραμμα Imetric4 = Στατιστική Διαφορά V-πεδίου

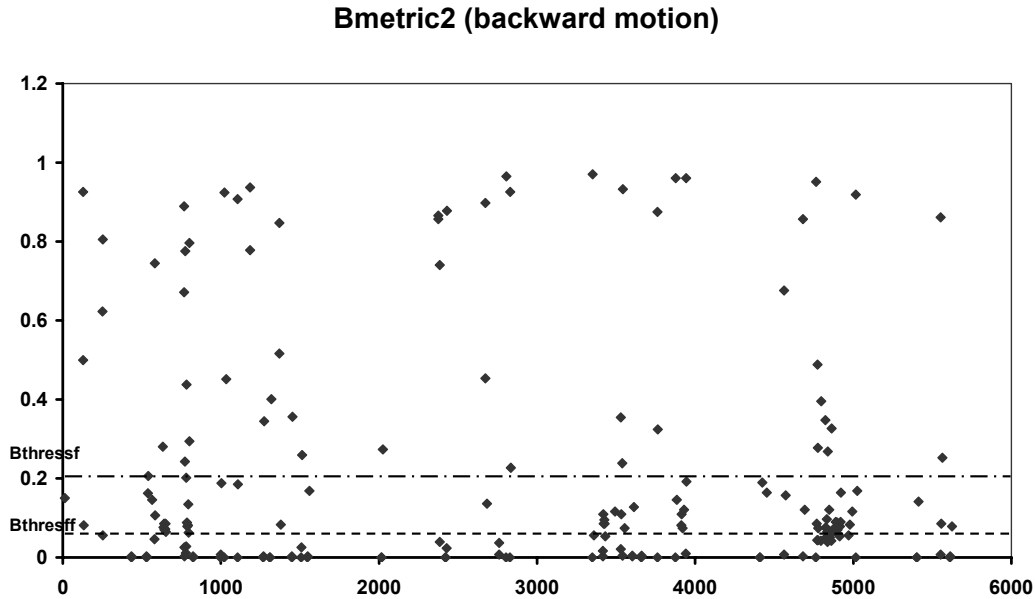


Όπως φαίνεται στο διάγραμμα του σχ. 7.3 χρησιμοποιείται το γενικό κατώφλι  $I_{thres0}$  που είναι ορισμένο (fixed) για να “κόβει” τον περισσότερο θόρυβο. Στη συνέχεια για το μέτρο διαφοράς  $I_{metric2}$  χρησιμοποιούνται δυναμικά κατώφλια  $I_{thres2}$  και  $I_{thres4}$  (ορίζονται στο σχετικό Κεφάλαιο) τα οποία προσαρμόζουν το πρόβλημα στην εκάστοτε συνθήκη που ισχύει τοπικά στο βίντεο (π.χ. κίνηση μέσα στο πλάνο κτλ.). Στα σχήματα 7.4 και 7.5 φαίνονται τα δύο κατώφλια που χρησιμοποιούνται, το πρώτο ( $I_{thres5}$ ) έχει σκοπό να “κόψει” τον θόρυβο (και γι’ αυτό χρησιμοποιείται βοηθητικά μαζί με άλλο μέτρο), ενώ το δεύτερο ( $I_{thres6}$ ) ξεχωρίζει τις σχετικά μεγάλες τιμές και γι’ αυτό χρησιμοποιείται αυτοτελώς (μαζί για τα δύο U και V κανάλια) για τον εντοπισμό πιθανής αλλαγής πλάνου.

Επομένως, οι δύο βασικές μέθοδοι που πρωταρχικά εντοπίζουν μία πιθανή αλλαγή πλάνου είναι ο έλεγχος των παραπάνω μέτρων στα P-frames και στα I-frames αντίστοιχα. Στην συνέχεια ελέγχονται οι μετρικές για τα B-frames που ακολουθούν σε Transmit Order. Ελέγχεται το πλήθος των διανυσμάτων κίνησης που δείχνουν προς τα εμπρός και προς τα πίσω (forward και backward motion vectors, fmv και bmv). Τα ποσοστά τους καθορίζουν την κύρια αναφορά (κίνηση) σε κάθε B-frame και τέλος από την συνολική αναφορά που των συνεχόμενων B-frames προκύπτει η τελική απόφαση. Στα δύο σχήματα που ακολουθούν δίνονται τα μέτρα  $B_{metric1}$  (ποσοστό των προς τα εμπρός) και  $B_{metric2}$  (ποσοστό των προς τα πίσω), όπως ορίζονται στην παράγραφο 2.4, και μόνο για τις περιπτώσεις που ελέγχονται (όταν είναι πιθανή η αλλαγή πλάνου).



Σχ. 7.6 : Διάγραμμα  $B_{metric1}$  = Ποσοστό των fmv  
(τα σημεία δεν ενώνονται διότι δεν αποτελούν μετρήσεις σε κάθε B-frame)



**Σχ. 7.7 :** Διάγραμμα Bmetric2 = Ποσοστό των bmv  
(τα σημεία δεν ενώνονται διότι δεν αποτελούν μετρήσεις σε κάθε B-frame)

Όπως φαίνεται από τα διαγράμματα, οι τιμές των δύο κατωφλίων έχουν επιλεγεί ώστε να διαχωρίζουν τις σχετικά μεγάλες και τις σχετικά μικρές τιμές των ποσοστών (πλήθος των διανυσμάτων), αλλά και να αποκλείουν μία ζώνη η οποία περιέχει απροσδιόριστες τιμές (ενδιάμεσες μεταξύ μεγάλων και μικρών) που οφείλονται σε θόρυβο (ο οποίος προέρχεται από χαμηλή ποιότητα του βίντεο ή από την μέθοδο εκτίμησης κίνησης – motion estimation που χρησιμοποιήθηκε κατά την κωδικοποίηση του αρχείου).

### 7.2.3 Αποτελέσματα και Παρατηρήσεις

Για την εκτίμηση της ποιότητας των αποτελεσμάτων και της σύγκρισης της μεθόδου του παρόντος συστήματος με άλλες υλοποιημένες τεχνικές, χρησιμοποιούνται τα ποσοστά RECALL (ανάκληση πληροφορίας) και PRECISION (ακρίβεια) ως δείκτες επίδοσης του συστήματος. Έστω ότι με AS (actual shots) συμβολίζουμε το πλήθος των πραγματικών αλλαγών πλάνου που συμβαίνουν στην ακολουθία βίντεο, με CD (correctly detected) το πλήθος των αλλαγών πλάνου που ανιχνεύει η μέθοδός μας και που είναι πραγματικές, FP (false positive) τις αλλαγές που ανιχνεύονται από την μέθοδο αλλά στην

ουσία δεν είναι αλλαγές πλάνου, και με DS (detected shots) τον αριθμό των αλλαγών που ανιχνεύει συνολικά η μέθοδος (άρα είναι  $DS = CD + FP$ ).

Οι δείκτες επίδοσης RECALL και PRECISION ορίζονται ως εξής:

$$RECALL = \frac{CD}{AS}$$

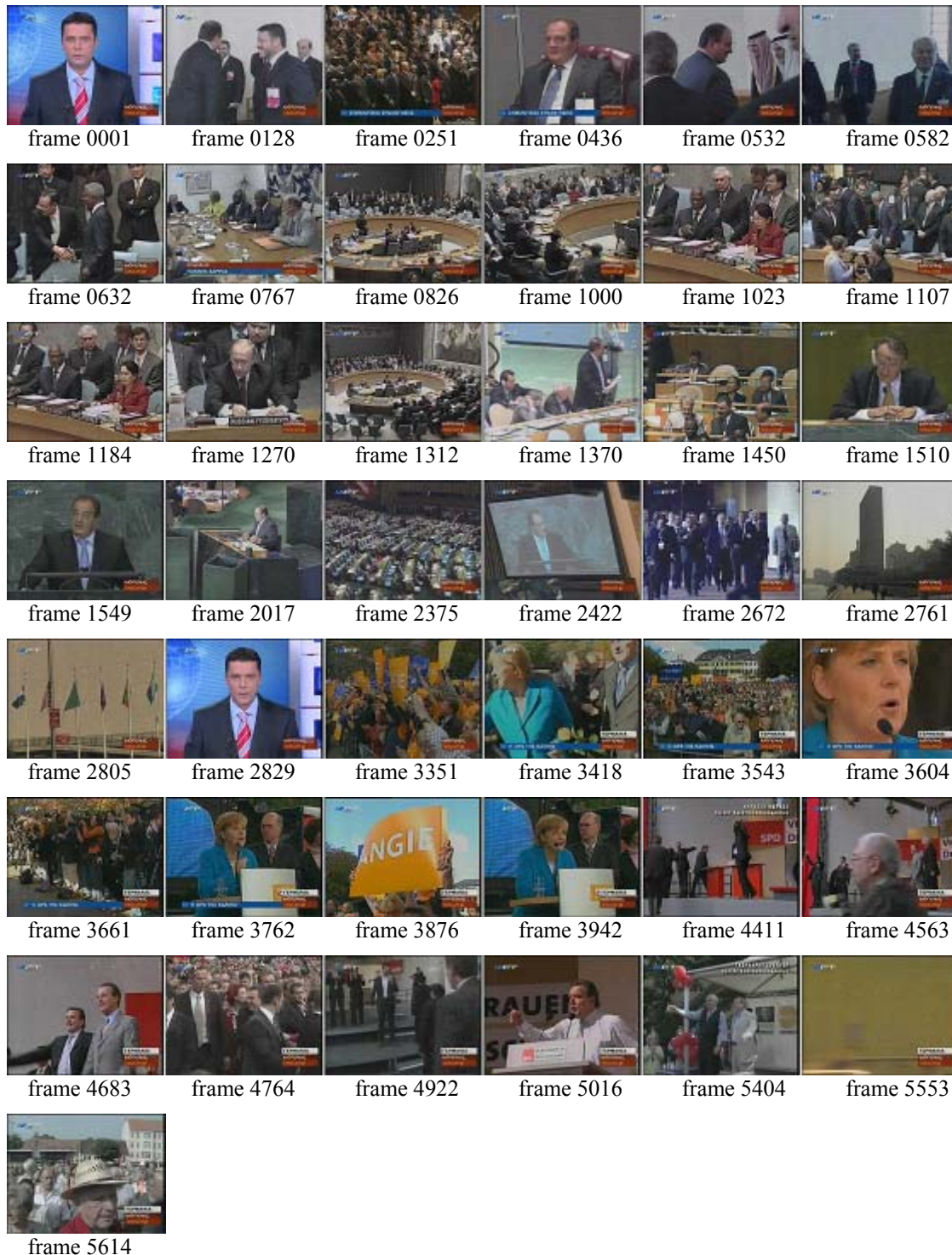
$$PRECISION = \frac{CD}{DS} = \frac{CD}{CD + FP}$$

Στον πίνακα 7.2 δίνονται οι τιμές των συνολικών FRAMES, AS, CD, FP και οι τιμές των δεικτών recall και precision που προκύπτουν από το σύστημά μας με είσοδο τα 10 βίντεο (το δείγμα του πειράματος).

A/A	Όνομα	FRAMES	AS	CD	FP	RECALL	PRECISION
1	MRBEAN	9508	54	53	7	0.9815	0.8833
2	ANT1NEWS	5638	24	22	6	0.9167	0.7857
3	ALTERNEWS	7468	54	51	31	0.9444	0.6220
4	NETNEWS1	5713	42	37	4	0.8810	0.9024
5	NETNEWS2	8263	49	48	2	0.9796	0.9600
6	SPORTS	8909	32	32	1	1.0000	0.9697
7	THEATER	13965	135	135	2	1.0000	0.9854
8	ET1	7515	61	58	0	0.9508	1.0000
9	ABCNEWS	8089	34	33	3	0.9706	0.9167
10	FOOTBALL	8556	31	30	1	0.9677	0.9677
<b>ΣΥΝΟΛΟ :</b>		<b>83624</b>	<b>516</b>	<b>499</b>	<b>57</b>	<b>0.9670</b>	<b>0.8975</b>

**Πίνακας 7.2 :** Αποτελέσματα της Ανίχνευσης Αλλαγής Πλάνου (recall και precision)

Στο επόμενο σχήμα παραθέτουμε στη σειρά το πρώτο καρέ κάθε πλάνου, για τα 43 πλάνα της ακολουθίας NETNEWS1, για την οποία κατασκευάστηκαν τα διαγράμματα στην παράγραφο 7.2.2. Τα πλάνα αυτά είναι τα πραγματικά πλάνα που υπάρχουν στο βίντεο (actual shots - AS).



**Σχ. 7.8 :** Το πρώτο καρέ κάθε πλάνου της ακολουθίας NETNEWS1 (AS – actual shots), όπως επιθυμούμε να εντοπίσει το σύστημα ανίχνευσης πλάνων

Για την παραπάνω ακολουθία NETNEWS1 το σύστημα των κανόνων που κατασκευάσαμε “χάνει” συνολικά 5 αλλαγές πλάνου και υποδεικνύει εσφαλμένα 4 αλλαγές πλάνου. Οι αλλαγές πλάνου όπου αποτυγχάνει το σύστημα συμβαίνουν στα frames 532, 632, 4563, 4764, 4922, και φαίνονται στο σχήμα 7.9, ενώ οι εσφαλμένα εντοπισμένες αλλαγές είναι στα frames 773, 782, 2385, 2430, και παρουσιάζονται στο σχήμα 7.10, μαζί με τα γειτονικά τους καρέ, ώστε να κατανοηθεί η αιτία της εσφαλμένης υπόδειξης.



**Σχ. 7.9 :** Οι αλλαγές πλάνου που αποτυγχάνει να εντοπίσει η μέθοδος (missed)



**Σχ. 7.10 :** Οι αλλαγές πλάνου που εσφαλμένα εντοπίζει η μέθοδος (FP – false positive)  
(με έντονα γράμματα σημειώνεται το πρώτο καρέ του νέου πλάνου)

Για τα καρέ του σχ. 7.9 παρατηρούμε από τα διαγράμματα των σχ. 7.1 – 7.5, δηλαδή από τα αριθμητικά αποτελέσματα τρεξίματος της μεθόδου, ότι οι κανόνες για τα Pmetric, Imetric<sub>1</sub> ικανοποιούνται, ενώ το μέτρο Imetric<sub>2</sub> που χρησιμοποιεί δυναμικά thresholds βλέπουμε ότι στα συγκεκριμένα σημεία έχει τιμή τουλάχιστον τετραπλάσια από τα προηγούμενα I-frames (εκτός από το frame 4936). Τα μέτρα Imetric<sub>3</sub> και Imetric<sub>4</sub> ικανοποιούνται ταυτόχρονα, εκτός από το 532 και 4922. Από τις προηγούμενες διαπιστώσεις καταλήγουμε στο ότι σε όλα τα παραπάνω καρέ η μέθοδος μέσω των I- ή P-frames που ακολουθούν εντοπίζει μία πιθανή αλλαγή πλάνου (με κάποιον από τους κανόνες 1 έως 6) και τελικά για την απόρριψη της αλλαγής πλάνου ευθύνονται οι κανόνες που ελέγχονται στα B-frames (που ακολουθούν σε Transmit Order).

Αυτό συμβαίνει διότι το πλήθος των διανυσμάτων κίνησης (motion vectors) δεν βρίσκεται μέσα στα αναμενόμενα όρια κατωφλίου όταν συμβαίνει αλλαγή πλάνου, που μπορεί να οφείλεται στο ότι το MPEG χρησιμοποιεί σταθερό αριθμό bits ανά sec, για να κωδικοποιήσει την πληροφορία, οπότε αν υπάρχουν περιθώρια χώρου πολλά Macroblocks μεταδίδονται ως intracoded (λόγω αυξημένου σφάλματος αντιστάθμισης κίνησης του MPEG, motion estimation-compensation) και δεν χρησιμοποιούν vectors. Δύο άλλοι λόγοι είναι η ομοιότητα των δύο γειτονικών πλάνων κοντά στο όριό τους, οπότε υπάρχουν forward και backward motion vectors, και η ύπαρξη μεγάλης κίνησης στην αρχή του νέου πλάνου. Όμως δεν ισχύει κάποιος από τους δύο τελευταίους λόγους, αφού ομοιότητα πλάνων δεν υπάρχει (μπορεί μερικές να έχουν ίδιο περιεχόμενο, αλλά δεν έχουν ίδιο στόχο κοντά στο όριό τους), αλλά ούτε και μεγάλη κίνηση στα πρώτα frames του νέου πλάνου (εκτός από το 4563, που και εκεί υπάρχει κίνηση μόνο στο 1 / 4 της εικόνας). Επομένως δεν μπορούμε με κάποιο τρόπο να βελτιώσουμε το αποτέλεσμα, παρά μόνο αν καταργήσουμε τους κανόνες στα B-frames, πράγμα που θα οδηγούσε σε αυξημένο FP (δηλαδή περισσότερες εσφαλμένες αποφάσεις για αλλαγή πλάνου σε σημεία εσωτερικά των πλάνων).

Παρατηρώντας το σχήμα 7.10 βλέπουμε ότι οι δύο πρώτες εσφαλμένες αλλαγές πλάνου (frames 773, 782) συμβαίνουν εξαιτίας του φλας φωτογραφικής μηχανής που εμφανίζεται μόνο σε ένα καρέ, και μάλιστα είναι frame αναφοράς (I ή P). Το φλας που εμφανίζεται σε ένα καρέ αλλάζει όχι μόνο την φωτεινότητα του πλάνου (κανάλι Y), αλλά και τα κανάλια χρώματος (U και V), αφού είναι μία πολύ απότομη μεταβολή (αποτελεί την πιο απότομη μεταβολή που μπορεί να συμβεί στο βίντεο, περισσότερο ακραία και από μετακίνηση αντικειμένου που καλύπτει ολόκληρη την εικόνα). Αυτή η στιγμιαία μεταβολή τυχαίνει εδώ να συμβαίνει πάνω σε ένα frame αναφοράς R1 (I ή P). Αν το



επόμενο frame αναφοράς R2 είναι P-frame, επειδή χρησιμοποιεί διανύσματα κίνησης από το R1, θα φαίνεται πολύ διαφορετικό με πολύ λίγα διανύσματα να δείχνουν από το A προς το B, αυξημένο σφάλμα για τα blocks που περιέχουν κίνηση και αυξημένο αριθμό intracoded MB. Αν το R1 είναι I-frame και το R2 είναι επίσης I-frame, τότε πάλι έχουμε μεγάλη διαφορά που προκύπτει από την μέθοδο μεταξύ των δύο. Αν όμως το R1 είναι P-frame και το R2 είναι I-frame, τότε επειδή οι κανόνες της μεθόδου μετρούν τη διαφορά με το προηγούμενο I, θα προκύψει μικρή διαφορά αν τα δύο I ανήκουν στο ίδιο πλάνο. Στις δύο πρώτες περιπτώσεις λοιπόν (R1=IήP και R2=P, R1=I και R2=I), οι μεγάλες διαφορές στις μετρικές (παράγραφος 2.4) θεωρούν πιθανή αλλαγή πλάνου, ενώ οι κανόνες στα B-frames δείχνουν κύρια αναφορά στο επόμενο frame αναφοράς (αφού σε εκείνο μοιάζουν πιο πολύ όπως βλέπουμε και από το σχ. 7.10) και αποφασίζουν αλλαγή πλάνου.

Παρατηρώντας τις δύο τελευταίες εσφαλμένες αλλαγές πλάνου (frames 2385, 2430) στο σχ. 7.10 βλέπουμε ότι τα καρέ περιέχουν σχετικά μεγάλα αντικείμενα, είναι αρκετά όμοια μεταξύ τους και ότι η αλλαγή γίνεται πριν από κάποιο I-frame. Και στις δύο αυτές περιπτώσεις παρατηρούμε ότι τα προηγούμενα από αυτά τα I-frames ανήκουν σε διαφορετικό πλάνο, εν τω μεταξύ αλλάζει το πλάνο (η αλλαγή έχει ήδη εντοπιστεί σε κάποιο P-frame), τα μέτρα για τους κανόνες στα παρόντα I-frames προκύπτουν μεγάλα και τα B-frames δείχνουν το πρώτο προς το προηγούμενο P-frame και το δεύτερο προς αυτό το I-frame. Γι' αυτό προκύπτει σφάλμα αλλαγής πλάνου το οποίο δεν μπορούμε να αντιμετωπίσουμε, αλλά συμβαίνει σπάνιες φορές.

Πρέπει στο σημείο αυτό να τονίσουμε ότι η αποτυχία του συστήματος να διαχωρίσει το στιγμιαίο φλας από πραγματική αλλαγή πλάνου έλαβε χώρα στις δύο παραπάνω περιπτώσεις, αλλά συνολικά στην ακολουθία NETNEWS1 υπάρχουν πολλά σημεία με φλας, αφού το βίντεο αποτελεί απόσπασμα ειδήσεων ελληνικού τηλεοπτικού καναλιού όπου το πρώτο κομμάτι αφορά συμβούλιο του ΟΗΕ και η παρουσία φωτογραφικών φλας είναι εκτεταμένη, ειδικά σε κοντινά πλάνα όπου και επηρεάζουν περισσότερο το περιεχόμενο. Η μέθοδος κατορθώνει να αποφύγει το σφάλμα σε δέκα περίπου άλλα σημεία φλας, διότι αυτά συμβαίνουν πάνω σε B-frames. Σε αυτήν την περίπτωση δεν υπάρχει αύξηση στα μέτρα που υπολογίζονται σε I και P frames, αλλά και αν ακόμα υπάρξει ένδειξη για πιθανή αλλαγή, απορρίπτεται στα B-frames όπου η κύρια αναφορά του καρέ με το φλας μένει απροσδιόριστη. Επειδή συνήθως τα B-frames αποτελούν το 66% του συνόλου των frames σε μία ακολουθία MPEG, (συνήθως ανά τρία frames MPEG, το πρώτο είναι I ή P και τα άλλα δύο είναι B-frames), το σύστημα

επιτυγχάνει κατά το πλείστον στην αναγνώριση των μεταβολών φλας, ενώ η αποτυχία του περιορίζεται σε πιθανότητα 33% (ή και μικρότερη, αφού στις 2 από τις 3 περιπτώσεις κάνει λάθος, όπως αναλύσαμε παραπάνω). Στο σχήμα 7.11 ενδεικτικά παρουσιάζονται 3 περιπτώσεις στις οποίες το σύστημα επιτυγχάνει στο ίδιο βίντεο (NETNEWS1), όπου σημειώνονται με έντονα γράμματα τα B-frames με το φλας.



**Σχ. 7.11 :** Σημεία με φωτογραφικό φλας (NETNEWS1) όπου η μέθοδος επιτυγχάνει (με έντονα γράμματα σημειώνεται το καρέ όπου υπάρχει το φωτογραφικό φλας)

Η ικανότητα αυτή του συστήματος φαίνεται και στο 9<sup>ο</sup> βίντεο ABCNEWS, που αποτελεί απόσπασμα από ειδήσεις ξένου τηλεοπτικού καναλιού, όπου σε συνέντευξη τύπου εμφανίζονται πάλι μεμονωμένα καρέ με φωτογραφικό φλας όπου το σύστημα επιτυγχάνει πλήρως στην διαδικασία και δεν γίνεται σύγχυση της παρουσίας φλας και της αλλαγής πλάνου. Ενδεικτικά παρουσιάζονται μερικά συνεχόμενα καρέ όπου εμφανίζεται αυτή η περίπτωση στο σχήμα 7.12.





frame 7075 (I)



frame 7076 (B)



**frame 7077 (B)**



frame 7078 (P)



frame 7180 (P)



**frame 7181 (B)**



frame 7182 (B)



frame 7183 (I)



frame 7210 (I)



frame 7211 (P)



**frame 7212 (B)**



frame 7213 (P)

**Σχ. 7.12 :** Σημεία με φωτογραφικό φλας (ABCNEWS) όπου η μέθοδος επιτυγχάνει (με έντονα γράμματα σημειώνεται το καρέ όπου υπάρχει το φωτογραφικό φλας)

### **Συμπεράσματα:**

Από τα αριθμητικά αποτελέσματα που παρατίθενται παραπάνω, στον πίνακα 7.2, συμπεραίνουμε άμεσα τα εξής:

1. Με την σκέψη ότι αρχικά ως στόχο είχαμε την κατασκευή ενός συστήματος που να εξάγει αποτελέσματα σχετικά με την ανίχνευση απότομης αλλαγής πλάνου (cut) σε σχετικά σύντομο χρονικό διάστημα με αρκετή επιτυχία, χωρίς να χάνει πολλές αλλαγές, εργαστήκαμε απευθείας στο MPEG και σχεδιάσαμε σύστημα που σε γλώσσα προγραμματισμού C χρειάζεται περίπου 20 φορές μικρότερο χρόνο (κατά μέσο όρο) από την διάρκεια της ακολουθίας βίντεο που δέχεται ως είσοδο (που φυσικά εξαρτάται από την ανάλυση του βίντεο και το bitrate). Ο χρόνος υπολογισμού σε δευτερόλεπτα για καθένα από τα 10 βίντεο δίνεται στο σχήμα 7.13 (το πρόγραμμα έτρεξε σε επεξεργαστή Pentium III 733MHz 384 RAM):

Βίντεο	1	2	3	4	5	6	7	8	9	10
Διάρκεια (sec)	380	225	298	228	330	356	558	300	269	342
Χρόνος Υπολογισμού (sec)	14.0	17.0	24.0	9.0	12.0	25.0	22.0	13.0	16.0	13.0
Λόγος	27	13	12	25	28	14	25	23	17	26

Σχ. 7.13 : Χρόνος Υπολογισμού για την μέθοδο Ανίχνευσης Αλλαγής Πλάνου

2. Τα αποτελέσματα από τον πίνακα 7.2 είναι περισσότερο από ικανοποιητικά, με ακριβή εντοπισμό των πλάνων (ανίχνευση του πρώτου καρέ νέου πλάνου), αν σκεφτούμε ότι στο δείγμα των MPEG υπάρχουν τα εξής περιεχόμενα:

- α) απόσπασμα από τηλεοπτική σειρά (MRBEAN) με πολλές αλλαγές cuts, όπου όλα τα πλάνα είναι από έναν συγκεκριμένο χώρο (δωμάτιο) με χαμηλό φωτισμό και μοιάζουν στο περιεχόμενό τους,
- β) αποσπάσματα από ειδήσεις (ALTERNEWS, NETNEWS1, NETNEWS2, ABCNEWS) με ρεπορτάζ, φωτογραφικά φλας, γρήγορα zoom-in ή zoom-out, κίνησης της κάμερας, ποικιλία πλάνων,
- γ) αποσπάσματα από αθλητικό ρεπορτάζ (ANT1NEWS, SPORTS), με καλαθοσφαίριση, ποδόσφαιρο και σχετικές συνεντεύξεις, όπου υπάρχει πολλή κίνηση και εναλλαγή περιεχομένου με δράση, αλλά και πολλές λεπτομέρειες λόγω των χρωμάτων,
- δ) απόσπασμα από θεατρική σκηνή (THEATER), όπου υπάρχουν πλάνα τριών προσώπων, κοντινά και μακρινά πλάνα των πρωταγωνιστών και γενικό πλάνο ολόκληρης της θεατρική σκηνής (δυσκολία διαχωρισμού λόγω ίδιου background - θεατρική σκηνή),
- ε) εκπομπή πολιτισμού όπου υπάρχει συνέντευξη με σταθερό πλάνο και αποσπάσματα από πλάνα ταινιών, θεατρικών και μουσικής (πρόβλημα λόγω σκοτεινών πλάνων), και

στ) απόσπασμα ποδοσφαιρικού αγώνα 6 λεπτών, όπου είναι πολύ δύσκολη η ανίχνευση των αλλαγών και συχνή η εσφαλμένη απόφαση νέου πλάνου, αφού το φόντο είναι το ίδιο πράσινο γήπεδο για τα περισσότερα πλάνα, οι λεπτομέρειες του πλάνου (γήπεδο, παίκτες, μπάλα, θεατές, πολλά διαφορετικά χρώματα) “μπερδεύουν” με διανύσματα κίνησης, και η σχετική κίνηση αντικειμένων και κυρίως της κάμερας “γεμίζει” με intracoded blocks.

3. Η τιμή για τους δύο δείκτες επίδοσης (πίνακας 7.2) προκύπτει στις περισσότερες περιπτώσεις πολύ καλή. Ο δείκτης ανάκλησης RECALL βλέπουμε ότι έχει συνήθως τιμή μεγαλύτερη του 0.95 (στα 7 από τα δέκα βίντεο), που σημαίνει ότι έχει 95 % επιτυχία στην ανίχνευση αλλαγών πλάνου (στις 100 πραγματικές αλλαγές πλάνου, καταφέρνει να εντοπίσει τις 95). Ο δείκτης PRECISION κυμαίνεται σε λίγο χαμηλότερες τιμές, παρόλ’ αυτά βρίσκεται συνήθως σε τιμές πάνω από 0.9 (για τα 7 από τα 10 βίντεο), δηλαδή η πιθανότητα του συστήματος να έχει εντοπίσει ακριβώς τα πραγματικά πλάνα είναι μεγαλύτερη του 90 % (στις 100 αλλαγές πλάνου που εντοπίζει, οι 90 είναι πραγματικές). Παρόλο που έχουμε τιμή του PRECISION ακόμα και 0.622, συνολικά το σύστημα παρουσιάζει πολύ καλή ανταπόκριση στην είσοδο που του δίνουμε, αφού συνολικά για ολόκληρη την ακολουθία από τα βίντεο του δείγματος με το παραπάνω ποικίλο περιεχόμενο, που αποτελείται συνολικά από 83624 frames, με διάρκεια περίπου 55 min, και με πραγματικές αλλαγές πλάνου 516 (cuts), έχουμε **RECALL = 0.967** και **PRECISION = 0.898**.
4. Συγκρίνοντας με παρόμοιες μεθόδους που εργάζονται στο κωδικοποιημένο MPEG πεδίο, βλέπουμε ότι βρίσκμαστε σε ικανοποιητικά ποσοστά για τους δείκτες επίδοσης. Η σύγκριση με μεθόδους που δουλεύουν σε ασυμπίεστο βίντεο δεν είναι σωστή, διότι τέτοιες μέθοδοι κάνουν υπολογισμούς σε όλα τα pixels της εικόνας, με αναμενόμενο καλύτερο αποτέλεσμα αλλά με πολύ μεγαλύτερο υπολογιστικό κόστος. Σύμφωνα με τους G. Boccignone et al. ([Boc00]), που χρησιμοποίησαν μία βάση από 18 διαφορετικά βίντεο (6 βίντεο με ταινίες, 6 βίντεο με ειδήσεις και 6 βίντεο με διαφημίσεις) για δοκιμή στην απόκριση του συστήματός τους που δούλευε σε MPEG για ανίχνευση cuts, είχαν σαν αποτέλεσμα συνολικά 17% missed cuts και 9% false cuts αντίστοιχα. Οι αντίστοιχες δικές μας τιμές για το σύνολο του δείγματος είναι 3.29% για τις χαμένες αλλαγές cuts και 10.25% για τις εσφαλμένες false positive cuts. Οι De Santo et al. ([DeS02]) αναφέρουν τις μεθόδους των Yeο et al., Pei et al.

και Lee et al., οι οποίοι στις εργασίες τους παραθέτουν ως αποτέλεσμα για precision και recall τις τιμές: 0.93 - 0.93, 0.99 - 1, και 0.97 - 0.99 αντίστοιχα, όμως οι τρεις προηγούμενοι συγγραφείς χρησιμοποιούν ως δείγμα βίντεο από 99000 frames (41 cuts), 36000 frames (269 cuts) και 80887 frames (611 cuts). Οι De Santo et al. επιθυμούν να χρησιμοποιήσουν μεγαλύτερο και αντιπροσωπευτικότερο δείγμα, γι' αυτό και τροφοδοτούν την μέθοδο με δείγμα εκπαίδευσης (training set) από 64343 frames (543 cuts) και δείγμα δοκιμής (test set) από 69971 frames (528 cuts). Με το παραπάνω δείγμα, προτείνουν έναν συνδυασμό από τις μεθόδους των Yeo, Pei, Lee. Ξεχωριστά η κάθε μέθοδος, για το σύνολο του test set, δίνει precision και recall: 0.88 και 0.82, 0.91 και 0.92, 0.02 και 0.64, αντίστοιχα. Ο συνδυασμός των μεθόδων και οι διαφορετικοί τρόποι που επιχείρησαν δεν κατάφερε να τους δώσει ποσοστά μεγαλύτερα από 0.93 για κάθε δείκτη επίδοσης.

5. Τέλος, επισημαίνουμε ότι το σύστημα επιδεικνύει σχετικά μικρή ευαισθησία σε απότομες στιγμιαίες μεταβολές, όπως είναι το φωτογραφικό φλας ή άλλες απότομες μεταβολές φωτεινότητας ή χρωμάτων που συμβαίνουν αποκλειστικά σε ένα-δύο καρέ της ακολουθίας, και επιτυγχάνει με την βοήθεια των κανόνων στα B-frames. Στη σπάνια περίπτωση όπου αλλαγή πλάνου συμβαίνει ταυτόχρονα με φλας, συνήθως “πιάνεται” στο επόμενο I-frame.
6. Στο 3<sup>ο</sup> βίντεο (ακολουθία ALTERNEWS) παρουσιάζεται αυξημένος αριθμός false positive, δηλαδή cuts που εντοπίστηκαν από την μέθοδο ενώ δεν είναι πραγματικά. Αυτό οφείλεται στο ότι το περιεχόμενο του βίντεο αποτελείται από καρέ που έχουν πολύ μεγάλη λεπτομέρεια στην εικόνα και μικρό bitrate, με αποτέλεσμα κατά την κωδικοποίηση MPEG να δημιουργούνται: αυξημένο σφάλμα (error) σε όλα τα frames, αυξημένος αριθμός intracoded MB στα P-frames, παραμορφωμένες DC τιμές στο DCT πεδίο των I-frames (μεγάλη διαφορά από προηγούμενο I) και αλλοιωμένο το πλήθος των motion vectors στα B-frames. Όλα αυτά “παραπλανούν” την μέθοδο, που βασίζεται στην κωδικοποιημένη πληροφορία του MPEG μόνο (οποσδήποτε χρειάζονται οι κανόνες στα B-frames και μάλιστα για την καθοριστική τελική απόφαση), και γι' αυτό σε αυτές τις περιπτώσεις δίνει αυξημένο αριθμό “εικονικών” false cuts, χωρίς όμως να αποτυγχάνει στον εντοπισμό των πραγματικών αλλαγών πλάνου.

## 7.3 Πειράματα στην Εξαγωγή των Χαρακτηριστικών Καρέ

### 7.3.1 Χαρακτηριστικά του Δείγματος Βίντεο

Το δείγμα που χρησιμοποιείται για τα πειράματα με στόχο την κατασκευή της καμπύλης του διανύσματος χαρακτηριστικών και την εξαγωγή των χαρακτηριστικών καρέ (keyframes) αποτελείται από έξι (6) αρχεία βίντεο κωδικοποιημένα με το πρότυπο MPEG 1. Δεν χρησιμοποιήθηκε το δείγμα της ανίχνευσης πλάνων, διότι στα βίντεο εκείνα δεν υπήρχαν οι απαραίτητες μεταβολές ώστε να παρατηρηθεί η συμπεριφορά της καμπύλης. Τα αρχεία αυτά έχουν συνολική διάρκεια 5 λεπτά και 35 δευτερόλεπτα, καταλαμβάνουν 75 MB στον σκληρό δίσκο και το καθένα έχει μέγεθος από 3 MB έως και 20 MB περίπου. Αποτελούνται κυρίως από 1 πλάνο, διότι ο στόχος είναι η παρατήρηση των καμπυλώσεων και όχι τα στατιστικά αποτελέσματα.

Παρακάτω, στον πίνακα 7.3, δίνεται η λίστα με τα αρχεία MPEG και η ονομασία τους. Περιγράφονται επίσης τα χαρακτηριστικά για κάθε βίντεο αναλυτικά.

A/A	Όνομα	Μέγεθος (MB)	Ανάλυση	Kbps	Fps	Διάρκεια	Frames	Πλάνα
1	FORMULA1	17.6	384x288	2048	25	0:01:03	1588	1
2	FORMULA2	18.8	384x288	1024	25	0:02:07	3178	9
3	RALLY1	15.8	384x288	2048	25	0:00:56	1423	1
4	RALLY2	7.97	384x288	2048	25	0:00:28	703	1
5	RALLY3	11.5	384x288	2048	25	0:00:41	1033	1
6	STEADY	3.07	320x240	1000	25	0:00:20	510	1

**Πίνακας 7.3 :** Χαρακτηριστικά για κάθε βίντεο του δείγματος για την εξαγωγή χαρακτηριστικών καρέ

Τα τέσσερα πρώτα βίντεο αφορούν πλάνα που έχουν καταγραφεί με κάμερα μέσα από ένα αγωνιστικό όχημα (τα δύο πρώτα από formula one και τα τρία επόμενα από

αυτοκίνητο rally). Η κάμερα καταγράφει τον δρόμο που βρίσκεται εμπρός από το όχημα, καθώς και ότι άλλο συναντάει στην διαδρομή του (άλλα αυτοκίνητα ή αντικείμενα εκτός του δρόμου, όπως θεατές, κτίρια, ταμπέλες). Το τελευταίο βίντεο περιέχει ένα πλάνο με σταθερή κάμερα, με στόχο την είσοδο μιας κατοικίας. Σταθερή εικόνα, κάποια στιγμή ένα πρόσωπο βγαίνει από την κατοικία και φεύγει από το πλάνο μπροστά από την κάμερα, μετά πάλι σταθερή εικόνα, δύο πρόσωπα που έρχονται, χαιρετιούνται και φεύγουν, και τέλος πάλι σταθερή εικόνα. Δηλαδή συμβαίνουν συνολικά τέσσερις σημαντικές μεταβολές του περιεχομένου (2 φορές είσοδος και έξοδος αντικειμένων).

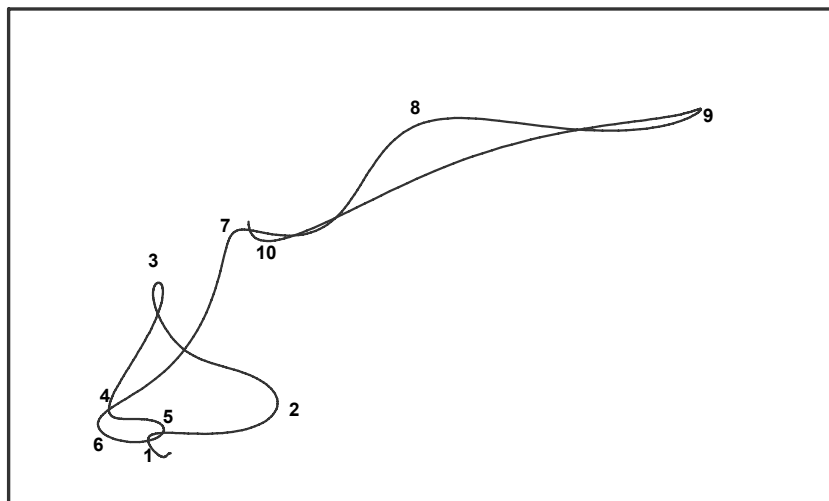
Τα κωδικοποιημένα στη μορφή MPEG αρχεία βίντεο, αφού πρώτα περάσουν από το βήμα της ανίχνευσης πλάνων, ακολουθεί η αποκωδικοποίησή τους (το ιστόγραμμα εξάγεται στο επίπεδο pixels), η κατασκευή του χρωματικού ιστογράμματος, του διανύσματος χαρακτηριστικών και της καμπύλης στον N-D Ευκλείδειο χώρο, και τέλος η εφαρμογή μίας από τις τεχνικές που αναλύθηκαν στα αντίστοιχα Κεφάλαια, που είναι οι δύο τύποι της καμπυλότητας 1 (C1) και 4 (C4), βλέπε Κεφάλαιο 4, καθώς και οι δύο μέθοδοι απλοποίησης καμπύλης (Curve Simplification – CS) και προσέγγισης περιγράμματος (Shape Contour Approximation – SCA), βλέπε Κεφάλαιο 5.

### *7.3.2 Διαγράμματα και Παρατηρήσεις*

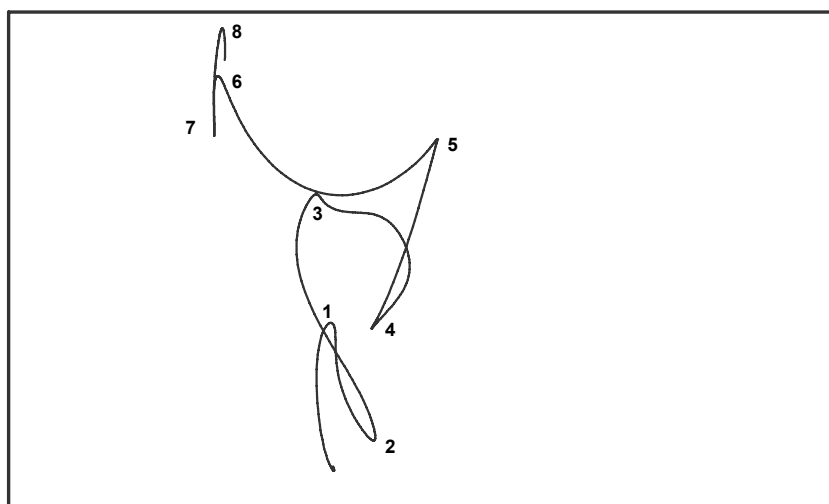
Κατασκευάζοντας την καμπύλη του διανύσματος χαρακτηριστικών και προβάλλοντάς την σε 3 επίπεδα, επιλέγοντας δύο συνιστώσες από το ίδιο κανάλι του ιστογράμματος (Y ή U ή V), προκύπτουν τα διδιάστατα διαγράμματα για διάφορα βίντεο από το δείγμα, όπως παρουσιάζονται στις επόμενες παραγράφους. Στη συνέχεια εφαρμόζουμε μία από τις μεθόδους C1, C4, CS, SCA, και από τα αποτελέσματά τους οδηγούμαστε σε σχετικές συγκρίσεις και χρήσιμα συμπεράσματα. Στη συνέχεια γίνονται 3 πειράματα, στα βίντεο FORMULA1, RALLY2 και STEADY.

#### **7.3.2.1 Αποτελέσματα για την ακολουθία FORMULA1**

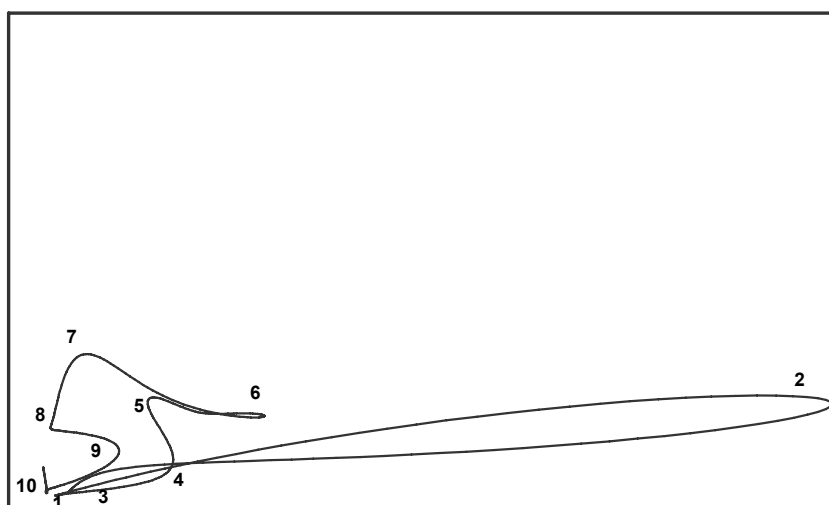
Στα σχήματα 7.14, 7.15, 7.16 φαίνονται τα διαγράμματα για την ακολουθία “FORMULA1”, με τα επίπεδα Y3 – Y7, U4 – U6 και V4 – V7 αντίστοιχα, για τα frames 601 έως 1100 (συνολικά 500 frames = 20 sec).



Σχ. 7.14 : Η προβολή της καμπύλης του FORMULA1 στο Y3 – Y7 και τα αντίστοιχα επιθυμητά χαρακτηριστικά σημεία

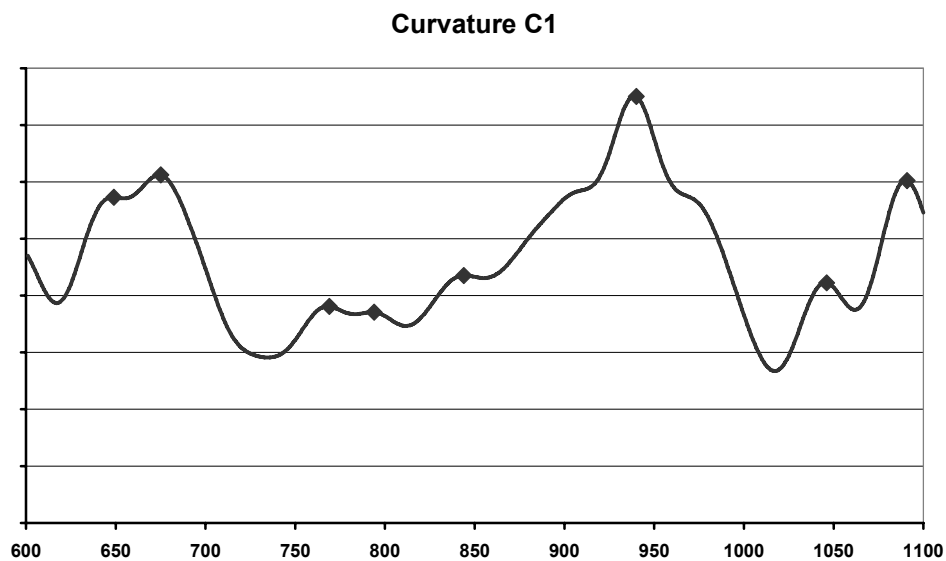


Σχ. 7.15 : Η προβολή της καμπύλης του FORMULA1 στο U4 – U6 και τα αντίστοιχα επιθυμητά χαρακτηριστικά σημεία

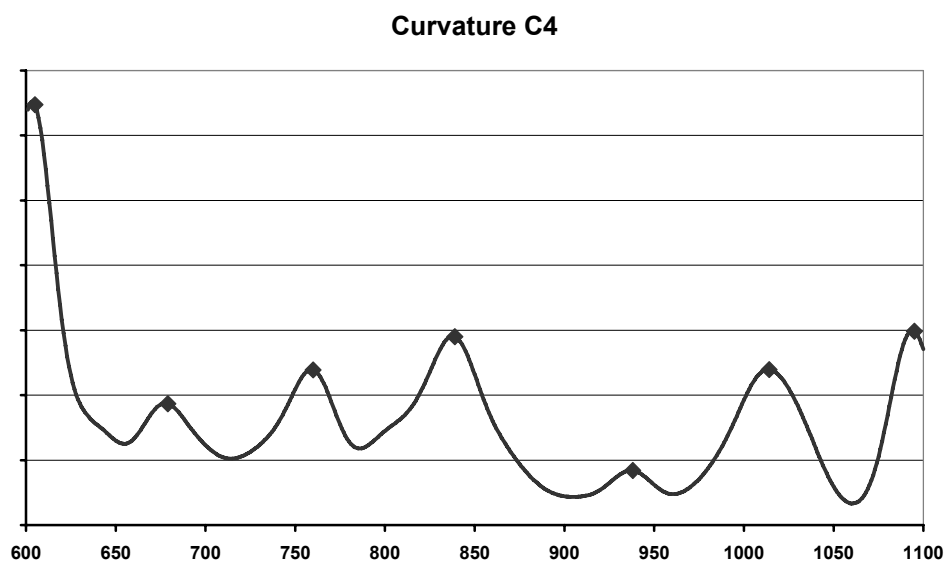


Σχ. 7.16 : Η προβολή της καμπύλης του FORMULA1 στο V4 – V7 και τα αντίστοιχα επιθυμητά χαρακτηριστικά σημεία

Παρακάτω, σχ. 7.17 και 7.18, παραθέτουμε τις τιμές που προκύπτουν από την μέθοδο καμπυλότητας και τους τύπους 1 και 4, δηλαδή C1 και C4 αντίστοιχα, για τα frames 601 έως 1100, όπου παρατηρούμε τα μέγιστα και τα ελάχιστα της κάθε μεθόδου. Το μέτρο της καμπυλότητας έχει ομαλοποιηθεί με γκαουσιανό φίλτρο με παράθυρο ίσο με  $NW / 2$  (βλέπε παράγραφο 3.3.1.2), δηλαδή το μισό από το παράθυρο ομαλοποίησης της καμπύλης, ώστε να προκύπτει ομαλή τιμή με ξεκάθαρα μέγιστα και ελάχιστα, αλλά χωρίς να παραμορφώνεται το αποτέλεσμα.



Σχ. 7.17 : Οι τιμές καμπυλότητας 1 για το βίντεο FORMULA1 (C1)



Σχ. 7.18 : Οι τιμές καμπυλότητας 4 για το βίντεο FORMULA1 (C4)



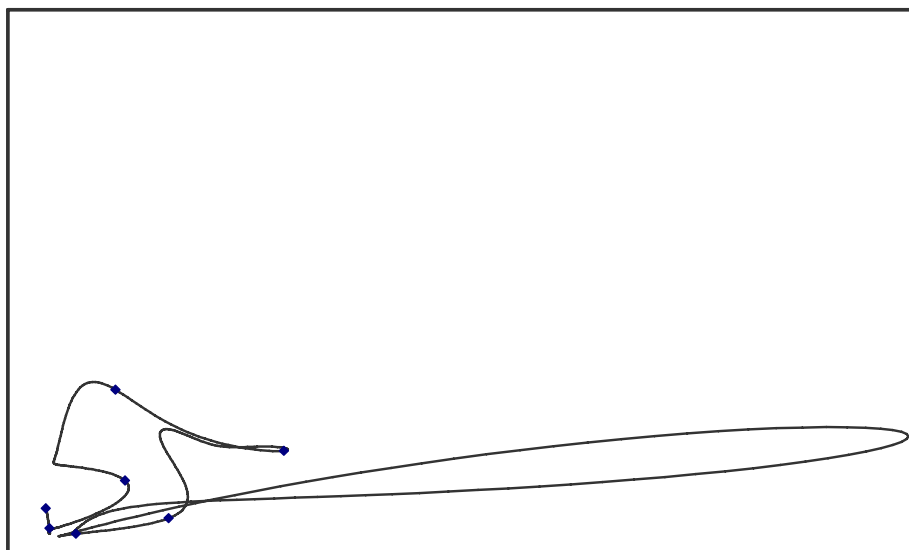
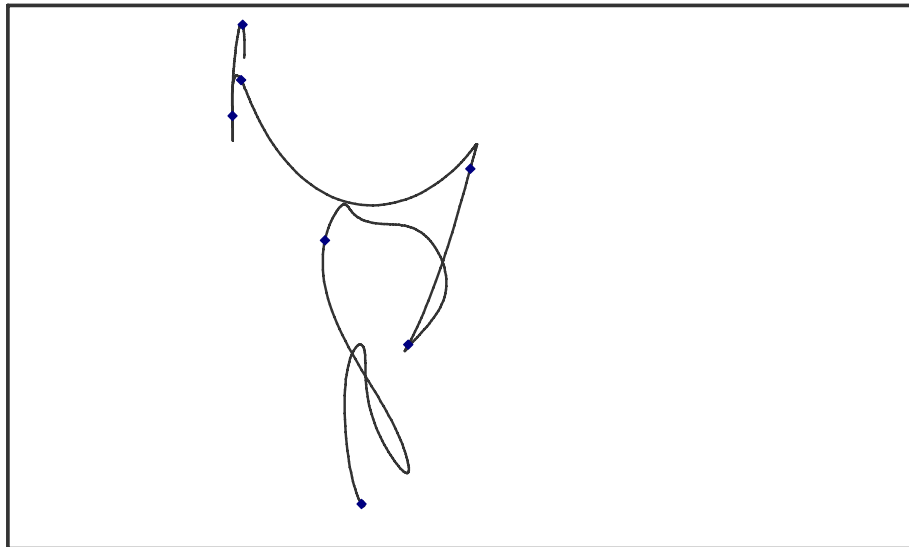
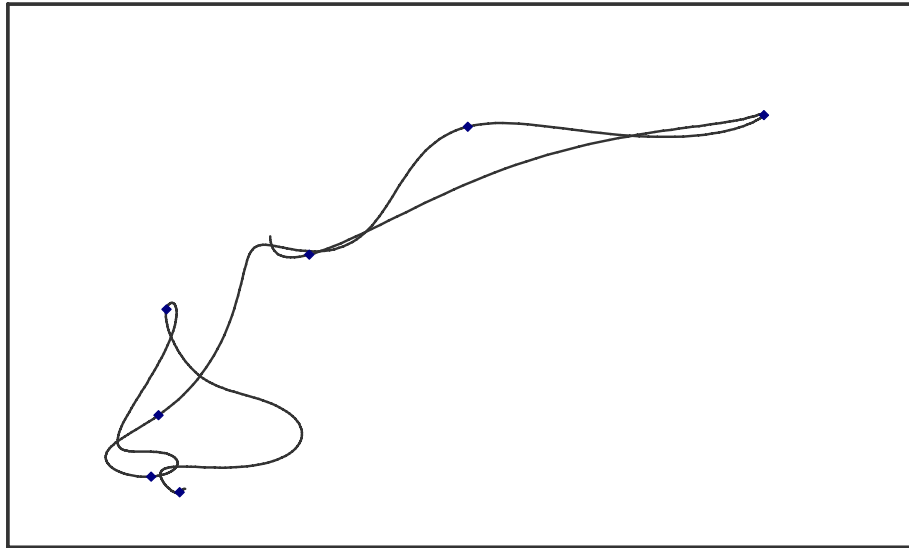
Αρχικά σημειώνουμε τα τοπικά μέγιστα στα διαγράμματα. Επίσης, καταγράφουμε τα χαρακτηριστικά σημεία που προκύπτουν από τις μεθόδους CS (με  $\varepsilon = 0.2$ ) και SCA (με δεδομένο αριθμό σημείων ίσο με 20). Ο αριθμός των σημείων που δίνουν οι CS και SCA επιλέχτηκε να είναι περίπου ίσος με το πλήθος των σημείων μεγίστων που δίνουν οι καμπυλότητες. Στη συνέχεια συγκεντρώνονται μαζί τα χαρακτηριστικά σημεία των CS, SCA (μεταξύ των frames 600 και 1100) και τα καρέ με τοπικό μέγιστο στις καμπυλότητες, όλα φαίνονται στον πίνακα 7.4.

CS keyframes	SCA keyframes	C1 Local Maximum	C4 Local Maximum
603	602	649	604
747	758	676	678
838	861	767	760
880	935	793	838
966	1017	845	937
1006	1093	940	1015
1085	–	1046	1095
–	–	1091	–

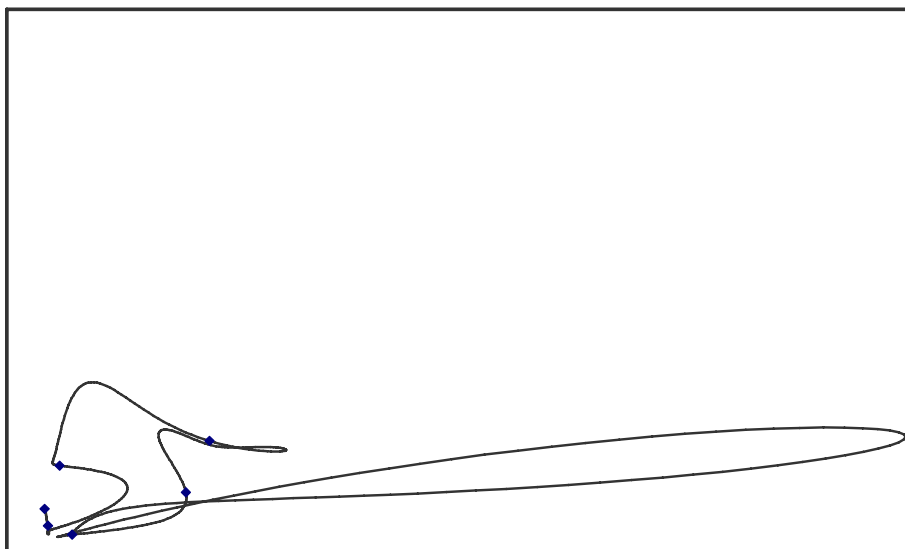
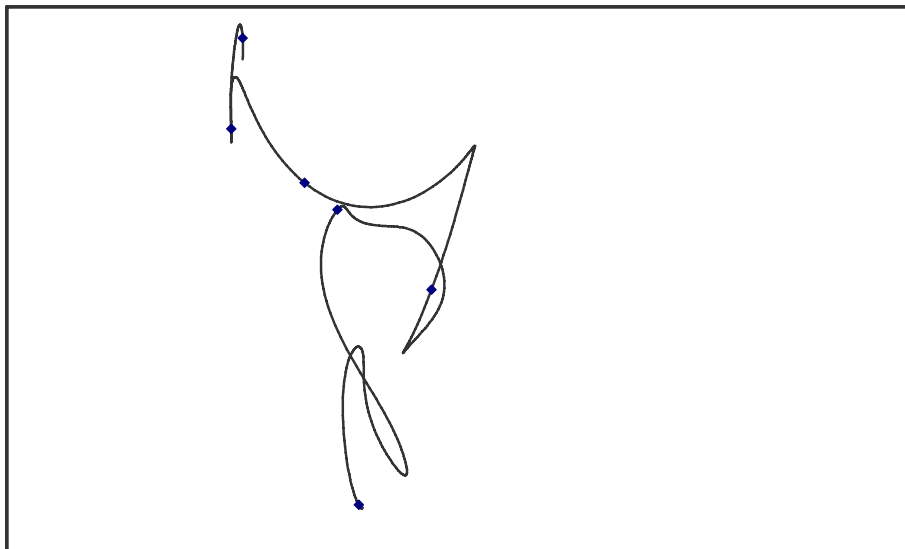
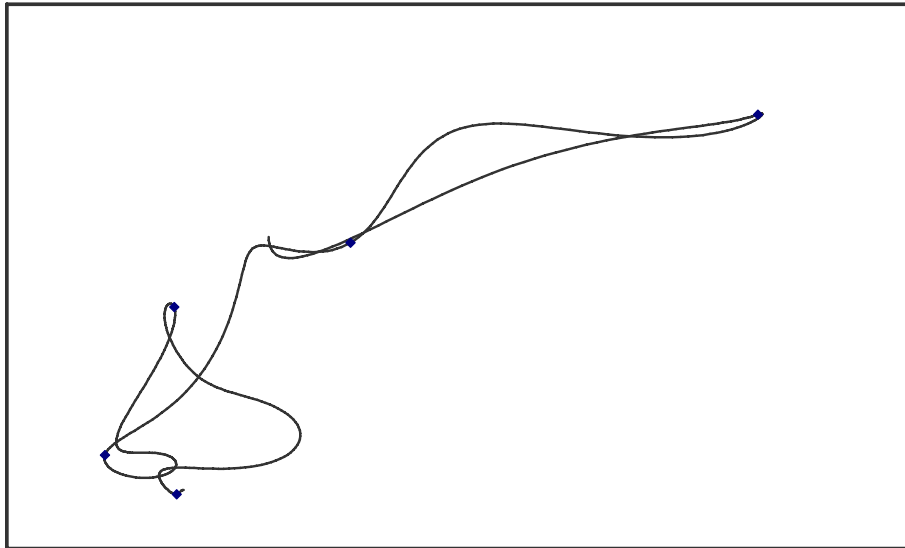
**Πίνακας 7.4 :** Τα χαρακτηριστικά καρέ που προκύπτουν από όλες τις μεθόδους

Από τα σημεία του πίνακα 7.4 και παρατηρώντας τις καμπύλες των σχημάτων 7.14, 7.15 και 7.16 τα σημεία στα οποία αντιστοιχούν, επιχειρούμε να συγκρίνουμε τις μεθόδους C1, C4, CS και SCA (επιθυμούμε να διαπιστώσουμε ποια μέθοδος προσεγγίζει καλύτερα την καμπύλη, με τα χαρακτηριστικά σημεία που βρίσκει, δηλαδή τα σημεία ακραίας καμπύλωσης).

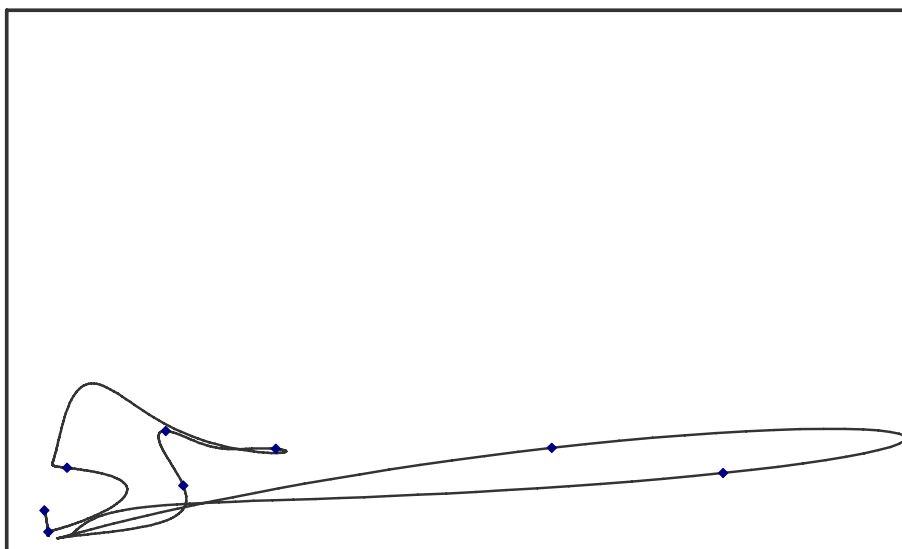
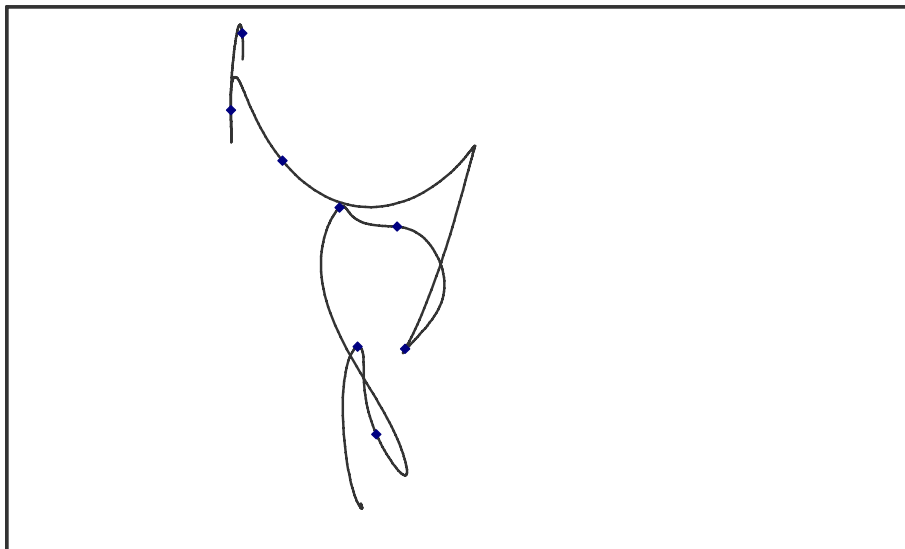
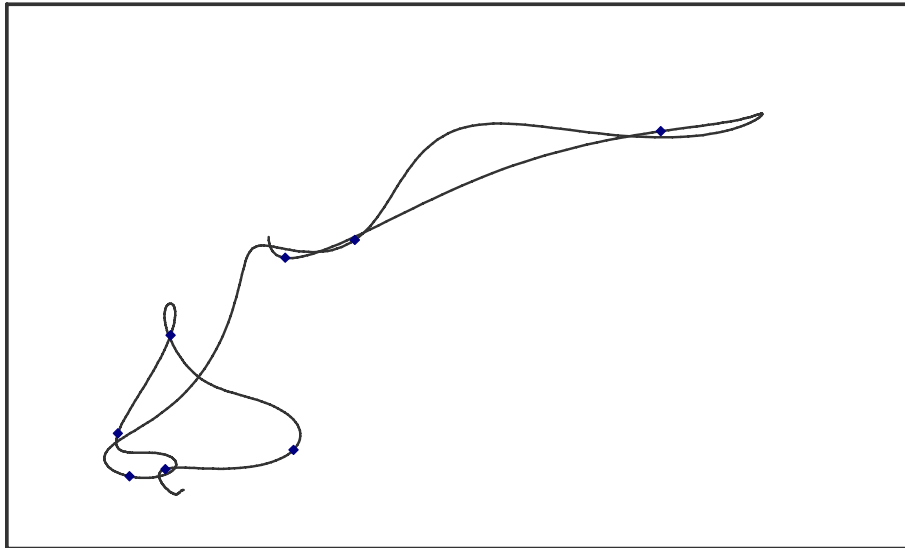
Παρακάτω παραθέτουμε για κάθε μία από τις 4 μεθόδους τα 3 διαγράμματα με τις προβολές της καμπύλης και σημειωμένα τα χαρακτηριστικά σημεία που η κάθε μέθοδος υπολογίζει, στα σχήματα 7.19, 7.20, 7.21 και 7.22.



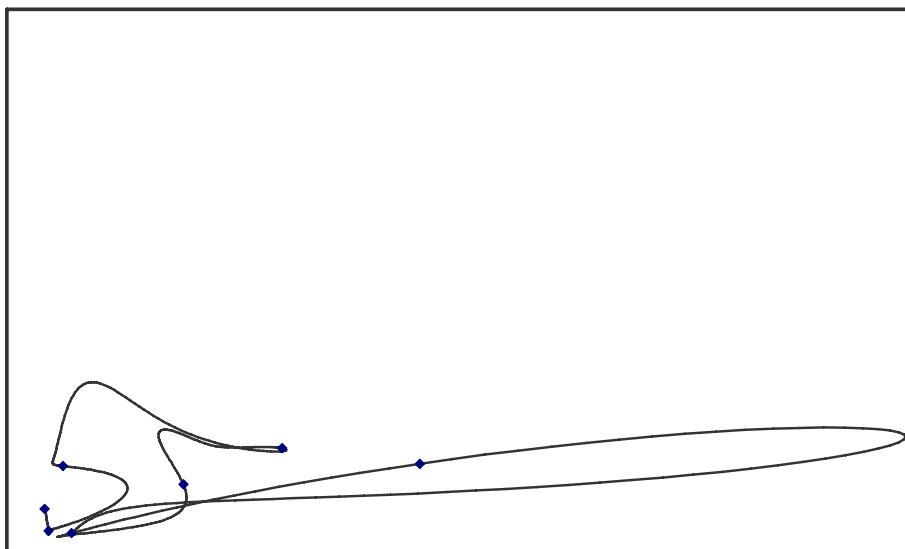
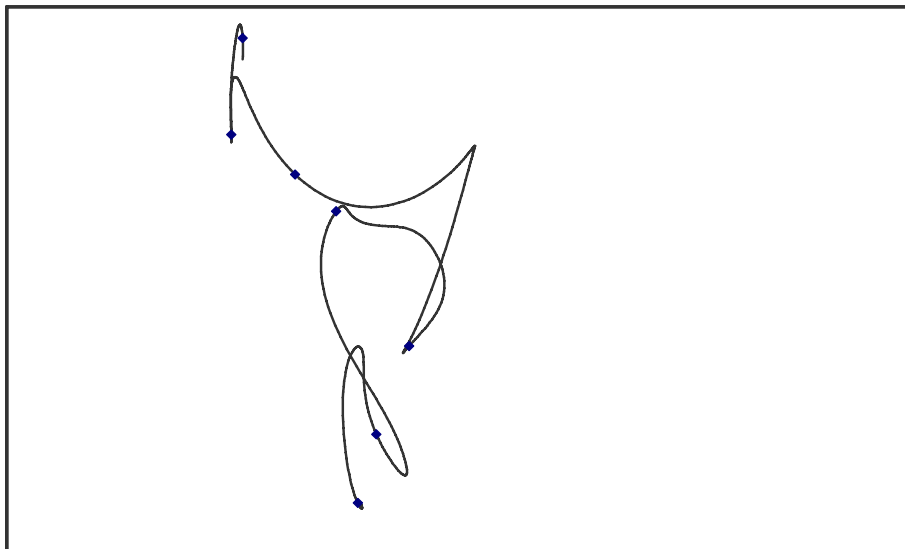
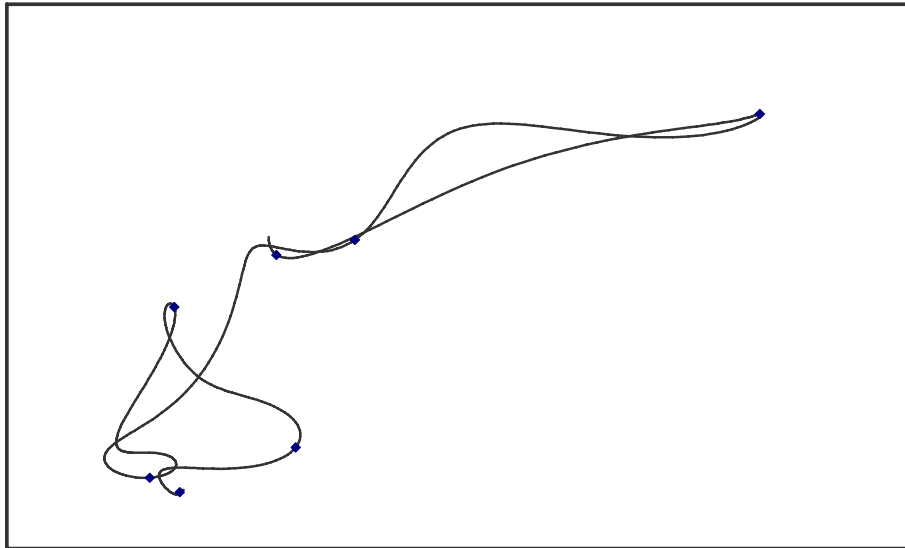
**Σχ. 7.19 :** Τα χαρακτηριστικά σημεία της μεθόδου CS στις  
διδιάστατες προβολές (FORMULA1)



**Σχ. 7.20 :** Τα χαρακτηριστικά σημεία της μεθόδου SCA στις  
 διδιάστατες προβολές (FORMULA1)



Σχ. 7.21 : Τα χαρακτηριστικά σημεία της μεθόδου C1 στις  
 διδιάστατες προβολές (FORMULA1)



Σχ. 7.22 : Τα χαρακτηριστικά σημεία της μεθόδου C4 στις  
διδιάστατες προβολές (FORMULA1)

## Παρατηρήσεις

Συγκρίνοντας ξεχωριστά κάθε επίπεδο προβολής της καμπύλης του βίντεο FORMULA1 και τα σημεία που ορίζει κάθε μέθοδος για την συγκεκριμένη επίπεδη καμπύλη, καταλήγουμε στα εξής (τα διαγράμματα είναι σχεδιασμένα σε ορθοκανονικό σύστημα αξόνων για αντικειμενική σύγκριση):

1. Στο διάγραμμα Y3 – Y7 υπάρχουν “χοντρικά”, όπως εύκολα διαπιστώνεται με την παρατήρηση 10 βασικά σημεία καμπύλωσης (βλέπε σχ. 7.14). Θα παρακολουθήσουμε πόσα από αυτά τα σημεία προσεγγίζονται από τις 4 μεθόδους που χρησιμοποιήσαμε, σύμφωνα με τα διαγράμματα των σχ. 7.19 – 7.22 (η αρίθμηση των σημείων φαίνεται στο σχ. 7.14). Θεωρούμε κοντινά σημεία όμοια με τα σημεία καμπύλωσης για απλούστευση (εκτός από σημεία που βρίσκονται στο ενδιάμεσο μεταξύ σημαντικών σημείων), αφού ούτως ή άλλως γειτονικά καρέ δεν διαφέρουν πολύ σε περιεχόμενο. Η μέθοδος CS εντοπίζει συνολικά 7 χαρακτηριστικά σημεία (1, 3, 5, 6, 8, 9, 10), η μέθοδος SCA 5 χαρακτηριστικά σημεία (3, 6, 9, 10), η καμπυλότητα C1 υποδεικνύει 8 σημεία μέγιστης καμπύλωσης (1, 2, 3, 4, 6, 7, 9, 10) και η καμπυλότητα C4 7 σημεία (1, 2, 3, 5, 7, 9, 10). Όπως βλέπουμε η μέθοδο C1 δίνει τα πιο πολλά σημαντικά σημεία, και ακολουθούν οι C4 και CS, από τις οποίες η C4 δίνει το σημείο 5 αντί για τα 4 και 6 της C1.

2. Στο διάγραμμα U4 – U6 υπάρχουν 8 βασικά σημεία καμπύλωσης (βλέπε σχ. 7.15). Όπως και προηγουμένως, παρακολουθούμε πόσα από αυτά τα σημεία προσεγγίζονται από τις 4 μεθόδους από τα διαγράμματα των σχ. 7.19 – 7.22. Η μέθοδος CS εντοπίζει συνολικά 6 χαρακτηριστικά σημεία (3, 4, 5, 6, 7, 8), η μέθοδος SCA 4 χαρακτηριστικά σημεία (3, 4, 7, 8), η καμπυλότητα C1 υποδεικνύει 6 σημεία μέγιστης καμπύλωσης (1, 2, 3, 4, 7, 8) και η καμπυλότητα C4 5 σημεία (2, 3, 4, 7, 8). Το σημείο που βρίσκει η μέθοδος CS και όχι οι C1 και C4 είναι το 5, καθοριστικό σημείο για αυτήν την επίπεδη καμπύλη, που όμως η καμπυλότητα της N-D καμπύλης δεν παρουσιάζει μέγιστο, διότι η μεταβολή καλύπτεται από τις υπόλοιπες συνιστώσες, αφού όλες μαζί υπολογίζονται για την μέτρηση της καμπυλότητας. Εδώ τα περισσότερα σημεία τα δίνουν οι C1 και CS.

3. Στο διάγραμμα V4 – V7 υπάρχουν 10 βασικά σημεία καμπύλωσης (βλέπε σχ. 7.16). Με παρόμοια διαδικασία όπως και προηγουμένως, παρατηρούμε ότι η μέθοδος CS εντοπίζει συνολικά 6 χαρακτηριστικά σημεία (1, 4, 6, 7, 9, 10), η μέθοδος SCA 5 χαρακτηριστικά σημεία (1, 4, 6, 8, 10), η καμπυλότητα C1 υποδεικνύει 6 σημεία μέγιστης καμπύλωσης (2, 4, 5, 6, 8, 10) και η καμπυλότητα C4 6 σημεία (1, 2, 4, 6, 8, 10). Όπως παρατηρούμε όλες οι μέθοδοι έχουν τον ίδιο αριθμό σημαντικών σημείων εκτός από την SCA. Η διαφορά τους είναι ότι η CS δείχνει τα σημεία 7 και 9, ενώ η C1 και C4 δείχνουν τα 2 και 5 αντίστοιχα, όπου το 2 φαίνεται ότι είναι το πιο σημαντικό σημείο στο διάγραμμα αυτό (εκεί υπάρχει η απότομη καμπύλωση), αλλά και το 5 σε σχέση με το 7.

4. Εξετάζοντας μεμονωμένα, η SCA έχει την μικρότερη ικανότητα μεταξύ των 4 μεθόδων (πάντοτε δίνει τα λιγότερα σημεία), ενώ οι υπόλοιπες μέθοδοι δίνουν καλά αποτελέσματα με την C1 να δίνει πάντοτε τα περισσότερα σημεία. Η CS ακολουθεί με 1 λιγότερο σημείο σε 1 επίπεδο και αμέσως μετά η C4 με 1 λιγότερο σημείο σε 2 επίπεδα.

5. Αν τώρα εξετάσουμε συνολικά τα 3 διαγράμματα των προβολών και τα σημεία καμπύλωσης που παρουσιάζουν σε κοινά καρέ, συγκεντρώνουμε τις τιμές στον πίνακα 7.5(α,β), όπου βλέπουμε στον πρώτο πίνακα τα επικρατέστερα πραγματικά σημεία καμπύλωσης (όπως προκύπτουν από τις προβολές) και στον δεύτερο πίνακα τις στήλες που δείχνουν τον αριθμό των keyframes των μεθόδων που αντιστοιχούν σε σωστά σημεία καμπύλωσης (N=ΝΑΙ, O=ΟΧΙ, όσα σημεία απέχουν από τα πραγματικά έως και 7 frames τα δεχόμαστε ως σωστά), καθώς και το ποσοστό επιτυχίας (αριθμός σωστών keyframes / αριθμός συνολικών ενδιαφερόντων σημείων) και ποσοστό σφάλματος (αριθμός εσφαλμένων keyframes / αριθμός συνολικών υποδεικνυόμενων keyframes).

Πραγματικά Σημεία Μεγάλης Καμπύλωσης της N-D καμπύλης					
603	636	682	753	800	838
862	914	934	965	1013	1092

**Πίνακας 7.5α :** Πραγματικά Σημεία Μεγάλης Καμπύλωσης της N-D καμπύλης

A/A	CS keyframes		SCA keyframes		C1 Local Maximum		C4 Local Maximum	
1.	603	N	602	N	649	O	604	N
2.	747	N	758	N	676	N	678	N
3.	838	N	861	N	767	O	760	N
4.	880	O	935	N	793	N	838	N
5.	966	N	1017	N	845	N	937	N
6.	1006	N	1093	N	940	N	1015	N
7.	1085	N	–	–	1046	O	1095	N
8.	–	–	–	–	1091	N	–	–
ποσοστό επιτυχίας	6/12 = 0.5		6/12 = 0.5		5/12 = 0.42		7/12 = 0.58	
ποσοστό σφάλματος	1/7 = 0.14		0/7 = 0.0		3/8 = 0.38		0/7 = 0.0	

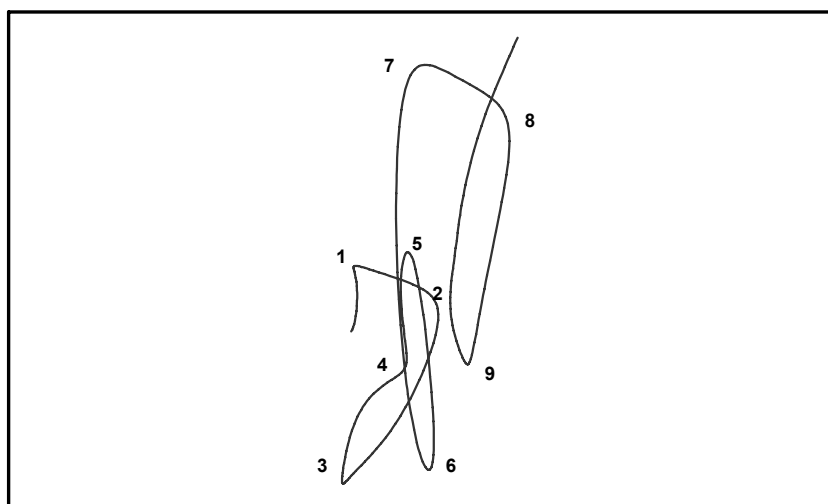
**Πίνακας 7.5β :** Τα ποσοστά επιτυχίας και σφάλματος για όλες τις μεθόδους

Πρέπει να τονίσουμε ότι τα παραπάνω ποσοστά είναι εικονικά, αφού προκύπτουν από παρατηρήσεις στις τρεις μόνο από τις 30 συνιστώσες του διανύσματος χαρακτηριστικών. Παρόλ’ αυτά, χρησιμοποιούμε τα αποτελέσματα ενδεικτικά για να διαπιστώσουμε κοντινά ποσοστά επιτυχίας από όλες τις μεθόδους, αλλά το μεγαλύτερο ποσοστό από την μέθοδο C4 της γενικευμένης καμπυλότητας. Η μέθοδος C1 δίνει τα λιγότερα σωστά χαρακτηριστικά σημεία, ενώ σε κάθε επίπεδο ξεχωριστά ήταν η πιο αποτελεσματική.

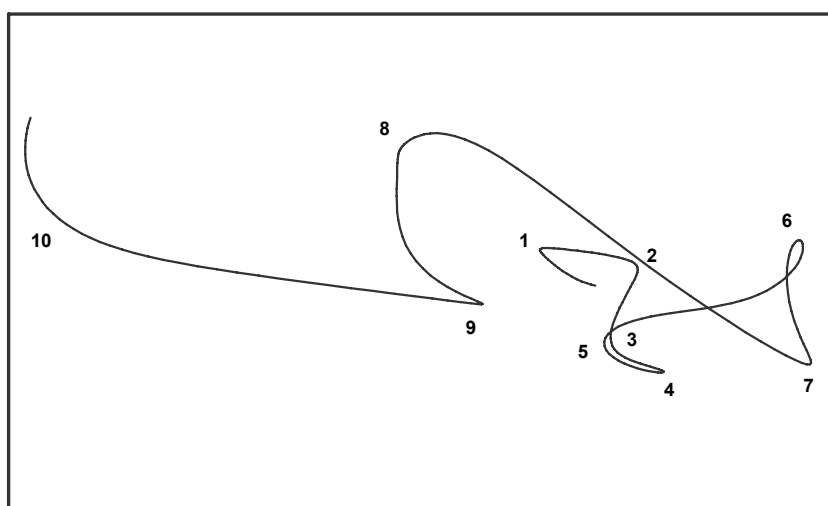
### 7.3.2.2 Αποτελέσματα για την ακολουθία RALLY2

Όπως και προηγουμένως, εργαζόμαστε ανάλογα και για το βίντεο “RALLY2” με 703 συνολικά frames. Στα σχήματα 7.23, 7.24, 7.25 φαίνονται τα διαγράμματα στα επίπεδα Y1 – Y6, Y4 – Y6 και Y5 – Y6 αντίστοιχα (frames 61 – 310).

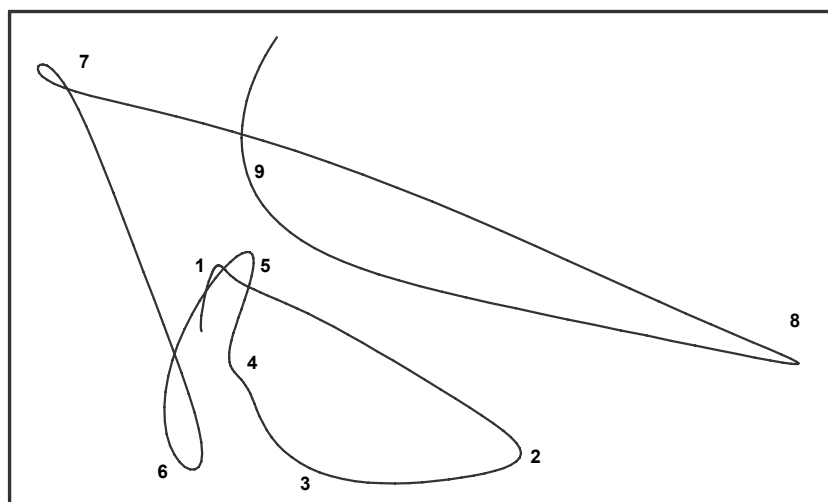




**Σχ. 7.23 :** Η προβολή της καμπύλης του RALLY2 στο  $Y1 - Y6$  και τα αντίστοιχα επιθυμητά χαρακτηριστικά σημεία

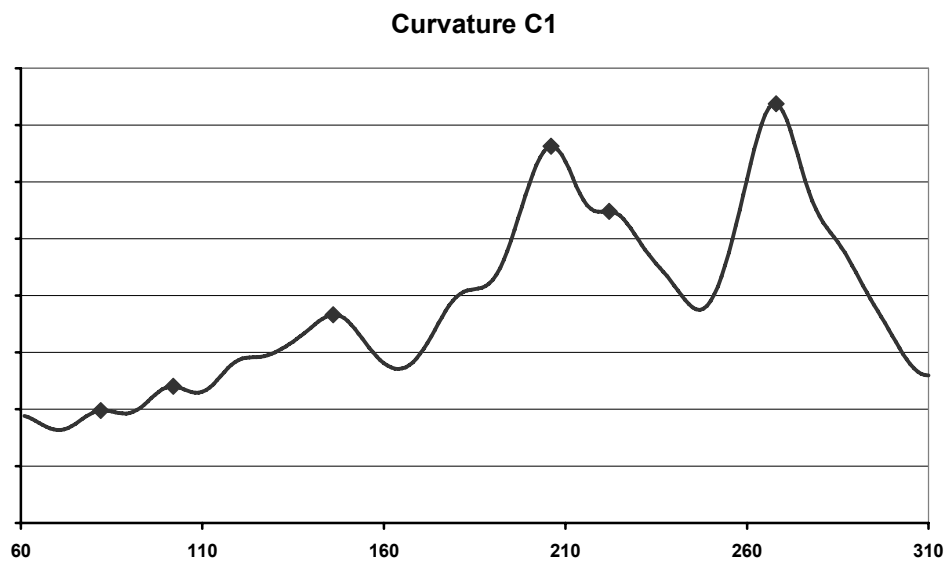


**Σχ. 7.24 :** Η προβολή της καμπύλης του RALLY2 στο  $Y4 - Y6$  και τα αντίστοιχα επιθυμητά χαρακτηριστικά σημεία

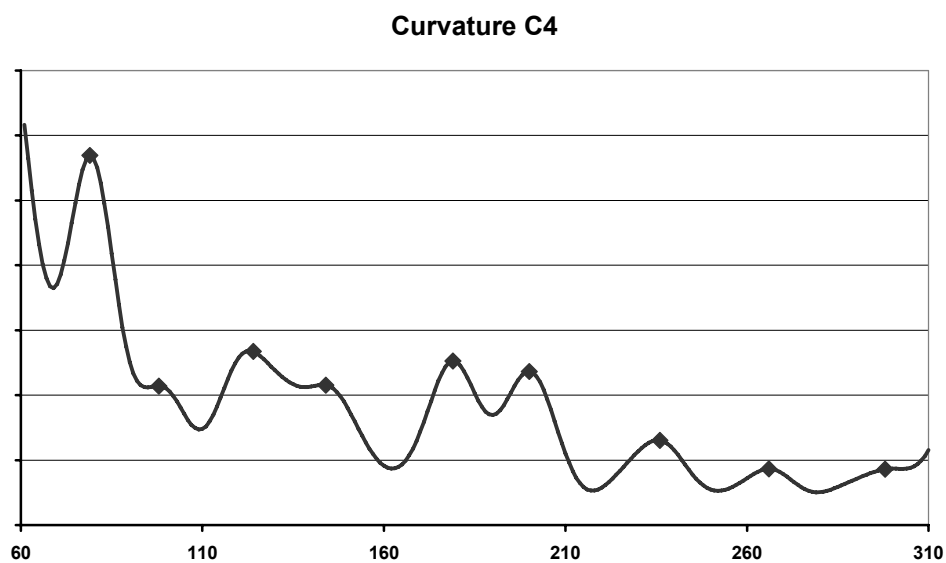


**Σχ. 7.25 :** Η προβολή της καμπύλης του RALLY2 στο  $Y5 - Y6$  και τα αντίστοιχα επιθυμητά χαρακτηριστικά σημεία

Χρησιμοποιήθηκαν μόνο συνιστώσες φωτεινότητας, διότι τα χρωματικά κανάλια παρουσιάζουν χαμηλή μεταβολή λόγω του περιεχομένου του βίντεο. Παρακάτω, σχ. 7.26 και 7.27, παραθέτουμε τις τιμές που προκύπτουν από την μέθοδο καμπυλότητας και τους τύπους 1 και 4, δηλαδή C1 και C4 αντίστοιχα, για τα frames 61 έως 310, όπου παρατηρούμε τα μέγιστα και τα ελάχιστα της κάθε μεθόδου. Το σημαντικό σημείο που πρέπει να αναφερθεί είναι ότι η C1 έχει δύο πολύ μεγάλα μέγιστα σε δύο σημεία όπου η C4 έχει ένα μέτριο και ένα πολύ μικρό μέγιστο.



Σχ. 7.26 : Οι τιμές καμπυλότητας 1 για το βίντεο RALLY2 (C1)



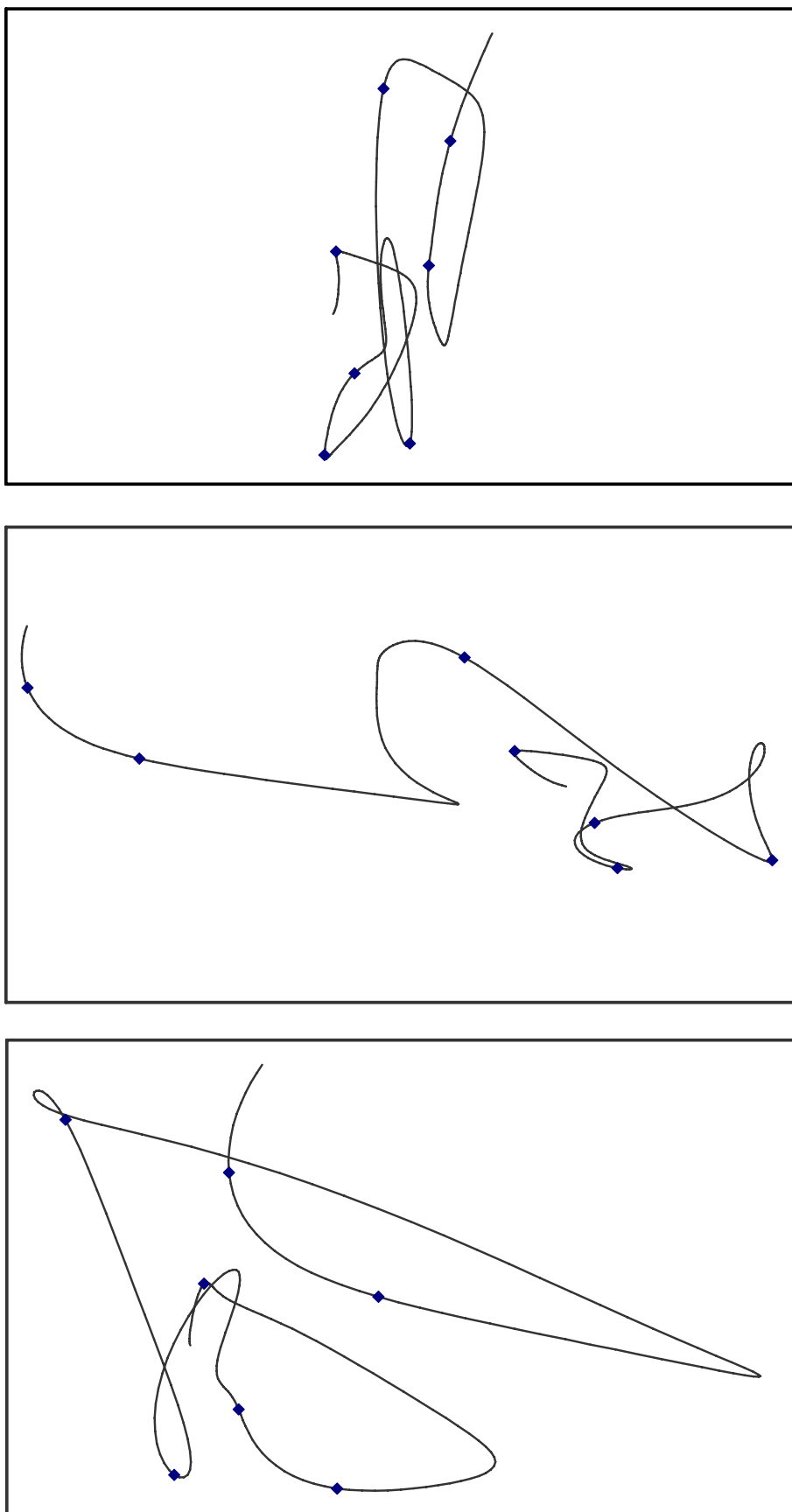
Σχ. 7.27 : Οι τιμές καμπυλότητας 4 για το βίντεο RALLY2 (C4)

Σημειώνουμε τα τοπικά μέγιστα στα διαγράμματα. Επίσης, καταγράφουμε τα χαρακτηριστικά σημεία που προκύπτουν από τις μεθόδους CS (με  $\varepsilon = 0.2$ ) και SCA (με δεδομένο αριθμό σημείων ίσο με 20). Ο αριθμός των σημείων που δίνουν οι CS και SCA επιλέχτηκε να είναι περίπου ίσος με το πλήθος των σημείων μεγίστων που δίνουν οι καμπυλότητες (περίπου ίσος με 7, μέσος όρος C1 και C4). Στη συνέχεια συγκεντρώνονται μαζί τα χαρακτηριστικά σημεία των CS, SCA (μεταξύ των frames 600 και 1100) και τα καρέ με τοπικό μέγιστο στις καμπυλότητες, στον πίνακα 7.6.

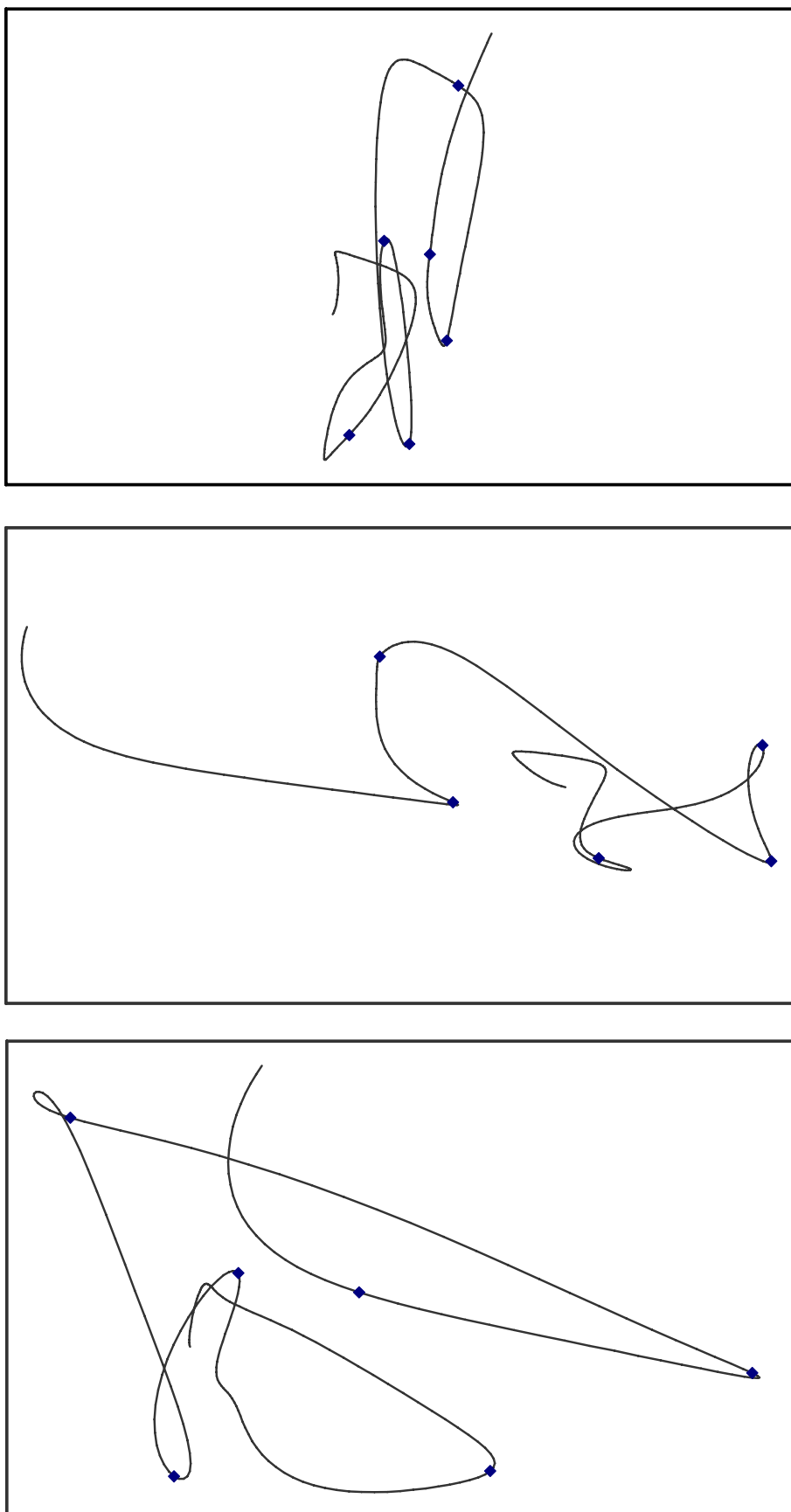
CS keyframes	SCA keyframes	C1 Local Maximum	C4 Local Maximum
79	122	83	79
136	178	102	98
151	200	146	123
200	236	206	143
222	263	222	179
285	286	268	200
300	–	–	235
–	–	–	266
–	–	–	300

**Πίνακας 7.6 :** Τα χαρακτηριστικά καρέ που προκύπτουν από όλες τις μεθόδους

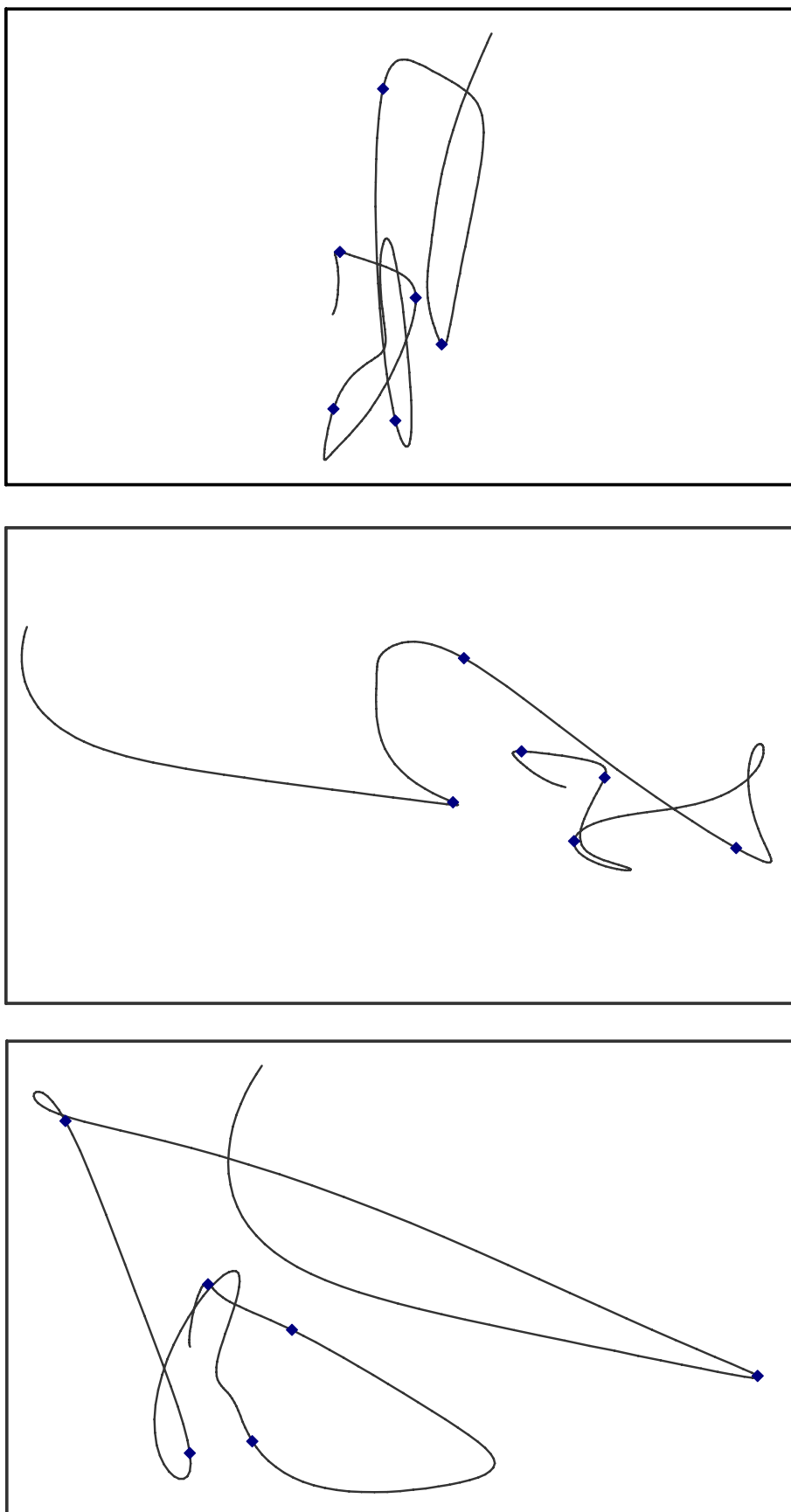
Από τα σημεία του πίνακα 7.6 επιχειρούμε να συγκρίνουμε τις μεθόδους C1, C4, CS και SCA. Παρακάτω παραθέτουμε για τις 4 μεθόδους τα 3 διαγράμματα με τις προβολές της καμπύλης και σημειωμένα τα χαρακτηριστικά σημεία που η κάθε μέθοδος υπολογίζει, στα σχήματα 7.28, 7.29, 7.30 και 7.31.



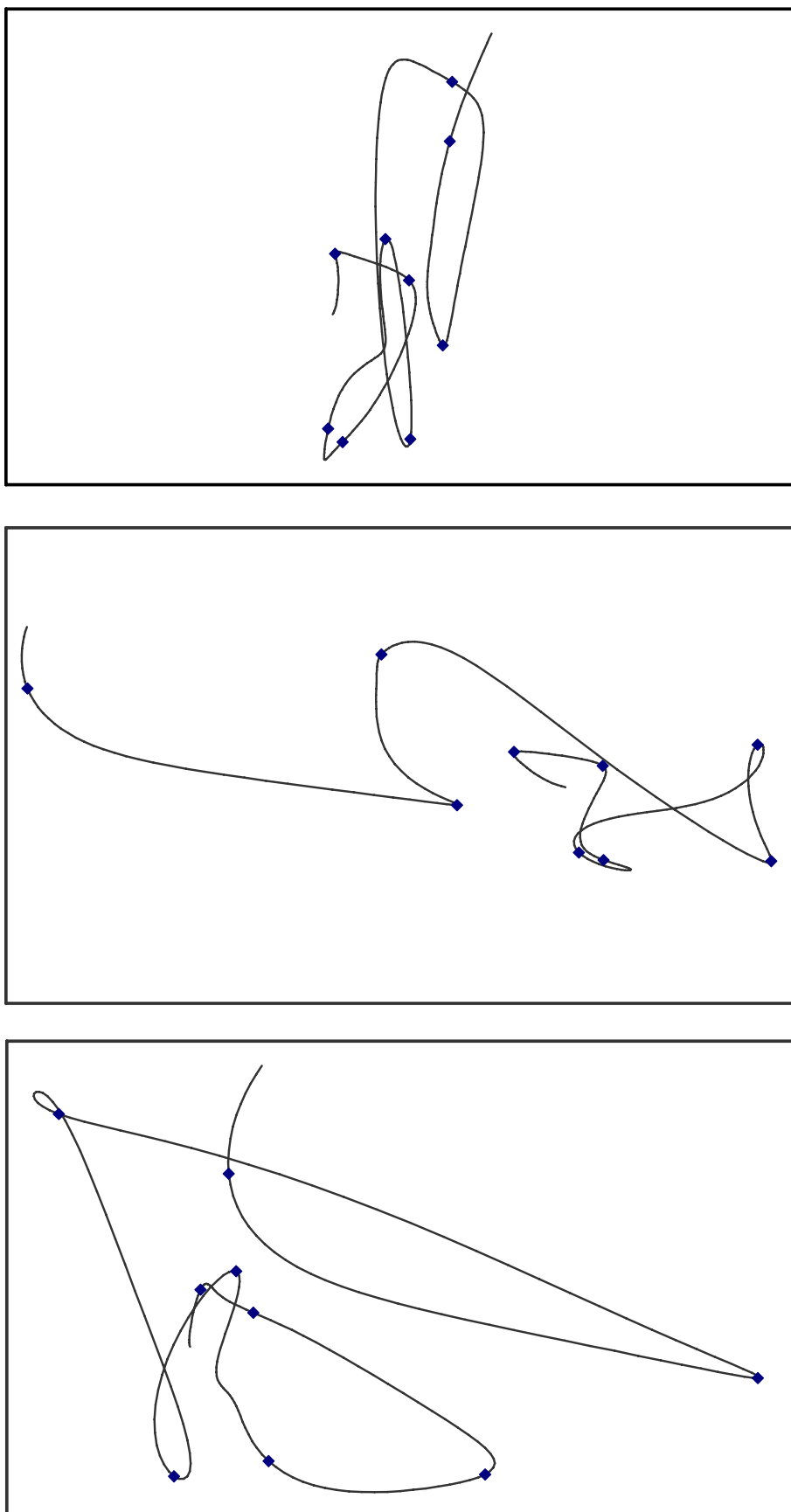
**Σχ. 7.28 :** Τα χαρακτηριστικά σημεία της μεθόδου CS στις διδιάστατες προβολές (RALLY2)



Σχ. 7.29 : Τα χαρακτηριστικά σημεία της μεθόδου SCA στις διδιάστατες προβολές (RALLY2)



**Σχ. 7.30 :** Τα χαρακτηριστικά σημεία της μεθόδου C1 στις διδιάστατες προβολές (RALLY2)



Σχ. 7.31 : Τα χαρακτηριστικά σημεία της μεθόδου C4 στις διδιάστατες προβολές (RALLY2)

## **Παρατηρήσεις**

Συγκρίνοντας ξεχωριστά κάθε επίπεδο προβολής της καμπύλης του βίντεο RALLY2 και τα σημεία που ορίζει κάθε μέθοδος για την συγκεκριμένη επίπεδη καμπύλη, καταλήγουμε στα εξής (τα διαγράμματα είναι σχεδιασμένα σε ορθοκανονικό σύστημα αξόνων για αντικειμενική σύγκριση):

1. Στο διάγραμμα Y1 – Y6 υπάρχουν “χοντρικά”, όπως εύκολα διαπιστώνεται με την παρατήρηση 9 βασικά σημεία καμπύλωσης (βλέπε σχ. 7.23). Θα παρακολουθήσουμε πόσα από αυτά τα σημεία προσεγγίζονται καλύτερα από τις 4 μεθόδους που χρησιμοποιήσαμε, σύμφωνα με τα διαγράμματα των σχ. 7.28 – 7.31 (η αρίθμηση των σημείων φαίνεται στο σχ. 7.23). Η μέθοδος CS εντοπίζει συνολικά 6 χαρακτηριστικά σημεία (1, 3, 4, 6, 7, 9), η μέθοδος SCA 5 σημεία (3, 5, 6, 8, 9), η καμπυλότητα C1 υποδεικνύει 6 σημεία μέγιστης καμπύλωσης (1, 2, 3, 6, 7, 9) και η καμπυλότητα C4 7 σημεία (1, 2, 3, 5, 6, 8, 9).

2. Στο διάγραμμα Y4 – Y6 υπάρχουν 10 βασικά σημεία καμπύλωσης. Παρόμοια, με παρατήρηση, προκύπτει ότι η μέθοδος CS εντοπίζει συνολικά 6 χαρακτηριστικά σημεία (1, 4, 5, 7, 8, 10), η μέθοδος SCA 5 σημεία (4, 6, 7, 8, 9), η καμπυλότητα C1 υποδεικνύει 6 σημεία μέγιστης καμπύλωσης (1, 2, 5, 7, 8, 9) και η καμπυλότητα C4 9 σημεία (1, 2, 4, 5, 6, 7, 8, 9, 10).

3. Στο διάγραμμα Y5 – Y6 υπάρχουν 9 βασικά σημεία καμπύλωσης. Παρόμοια, με παρατήρηση, προκύπτει ότι η μέθοδος CS εντοπίζει συνολικά 6 χαρακτηριστικά σημεία (1, 3, 4, 6, 7, 9), η μέθοδος SCA 5 σημεία (2, 5, 6, 7, 8), η καμπυλότητα C1 υποδεικνύει 5 σημεία μέγιστης καμπύλωσης (1, 3, 6, 7, 8) και η καμπυλότητα C4 8 σημεία (1, 2, 3, 5, 6, 7, 8, 9).

4. Όπως φαίνεται η καλύτερη μέθοδος που δίνει τα πιο πολλά ορθά χαρακτηριστικά σημεία στα τρία επίπεδα είναι η μέθοδος C4, αμέσως μετά είναι οι μέθοδοι C1 και CS, ενώ τελευταία έρχεται η μέθοδος SCA (όπως και στο προηγούμενο πείραμα).

5. Αν εξετάσουμε συνολικά τα 3 διαγράμματα των προβολών και τα σημεία καμπύλωσης που αυτά παρουσιάζουν σε κοινά καρέ, συγκεντρώνουμε τις τιμές στον



πίνακα 7.7, όπου βλέπουμε με τα επικρατέστερα πραγματικά σημεία καμπύλωσης (όπως προκύπτουν από τις προβολές) και τον αριθμό των keyframes των μεθόδων που αντιστοιχούν σε σωστά σημεία καμπύλωσης (N=ΝΑΙ, O=ΟΧΙ, όσα σημεία απέχουν από τα πραγματικά έως και 5 frames τα δέχομαι ως σωστά), καθώς και τα ποσοστά επιτυχίας και ποσοστό σφάλματος.

Πραγματικά Σημεία Μεγάλης Καμπύλωσης της N-D καμπύλης				
82	100	135	147	166
180	202	235	266	295

**Πίνακας 7.7α :** Πραγματικά Σημεία Μεγάλης Καμπύλωσης της N-D καμπύλης

A/A	CS keyframes		SCA keyframes		C1 Local Maximum		C4 Local Maximum	
1.	79	N	122	O	83	N	79	N
2.	136	N	178	N	102	N	98	O
3.	151	N	200	N	146	N	123	O
4.	200	N	236	N	206	N	143	N
5.	222	O	263	N	222	O	179	N
6.	285	O	286	O	268	N	200	N
7.	300	N	–	–	–	–	235	N
8.	–	–	–	–	–	–	266	N
9.	–	–	–	–	–	–	300	N
ποσοστό επιτυχίας	5/10 = 0.5		4/10 = 0.4		5/10 = 0.5		7/10 = 0.7	
ποσοστό σφάλματος	2/7 = 0.29		2/6 = 0.33		1/6 = 0.17		2/9 = 0.22	

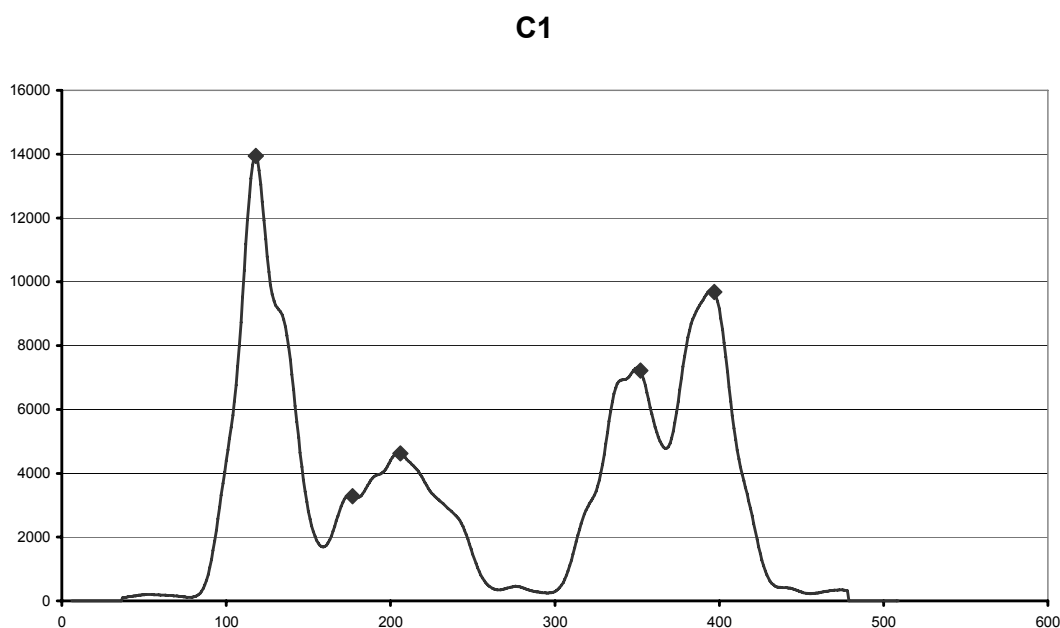
**Πίνακας 7.7β :** Τα ποσοστά επιτυχίας και σφάλματος για όλες τις μεθόδους

Όπως παρατηρούμε από τον προηγούμενο πίνακα, συνολικά για την πολυδιάστατη καμπύλη η μέθοδος C4 έχει το μεγαλύτερο ποσοστό επιτυχίας για την ανίχνευση σημαντικών χαρακτηριστικών σημείων (frames). Αμέσως μετά σε επιτυχία έρχονται οι μέθοδοι CS και C1, από τις οποίες η C1 έχει και το μικρότερο ποσοστό σφάλματος μεταξύ όλων των μεθόδων. Η μέθοδος SCA έχει το μικρότερο ποσοστό επιτυχίας και το μεγαλύτερο ποσοστό σφάλματος, παρουσιάζει δηλαδή την χειρότερη περίπτωση στην εξαγωγή των χαρακτηριστικών σημείων.

### 7.3.2.3 Αποτελέσματα για την ακολουθία STEADY

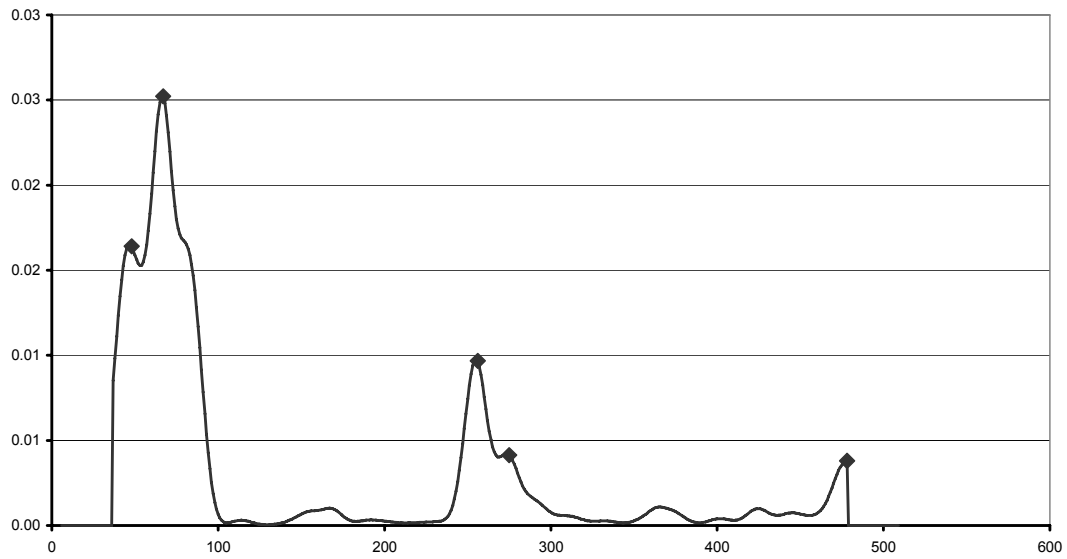
Στο πείραμα αυτό επιδιώκεται να αναλυθεί η συμπεριφορά των δύο μεθόδων C1 και C4 στην περίπτωση βίντεο που έχει διαστήματα σταθερού περιεχομένου, περίπτωση σπάνια, αλλά που πρέπει να εξεταστεί περισσότερο επειδή σε τέτοιο βίντεο οι μεταβολές θα είναι πολύ ακραίες και τα διαγράμματα της καμπυλότητας θα χαρακτηρίζονται από διογκωμένη μορφή ώστε να παρατηρηθούν πιο εύκολα τα στοιχεία που χαρακτηρίζουν (θετικά ή αρνητικά) τις μεθόδους.

Εξετάζουμε το βίντεο “STEADY” με 510 συνολικά frames, υπολογίζοντας τις καμπυλότητες C1 και C4 (με  $NW = 32$  για την καμπύλη και  $NW = 16$  για το τελικό διάγραμμα), οι οποίες φαίνονται στα σχήματα 7.32 και 7.33.



**Σχ. 7.32 :** Οι τιμές καμπυλότητας 1 για το βίντεο STEADY (C1) και τα τοπικά μέγιστα

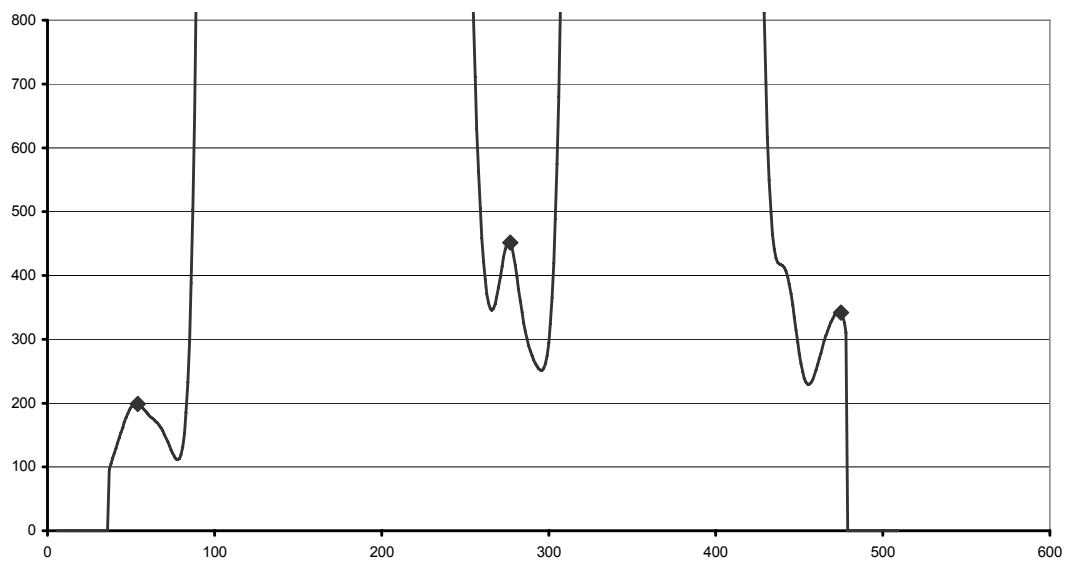
#### C4



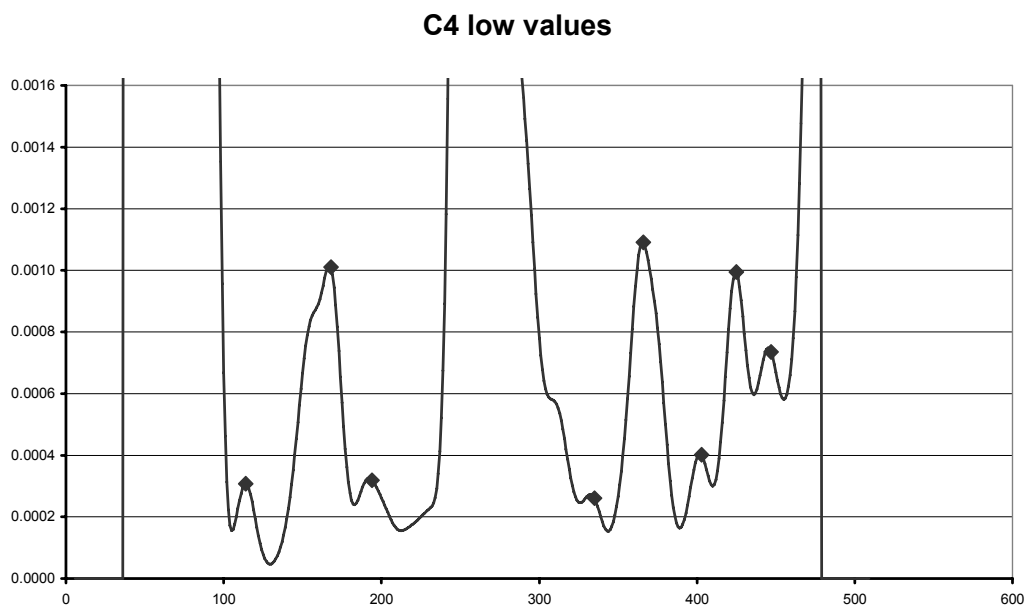
**Σχ. 7.33 :** Οι τιμές καμπυλότητας 4 για το βίντεο STEADY (C4) και τα τοπικά μέγιστα

Με την πρώτη ματιά στα διαγράμματα, φαίνεται ότι το ένα παρουσιάζει μέγιστες τιμές στο σημείο που το άλλο έχει ελάχιστες τιμές, αλλά με την παρατήρηση των πολύ μικρών τιμών (low) βλέπουμε ότι και σε αυτές τις τιμές κάθε διάγραμμα παρουσιάζει τοπικά μέγιστα (σχ. 7.34 και 7.35).

#### C1 low values

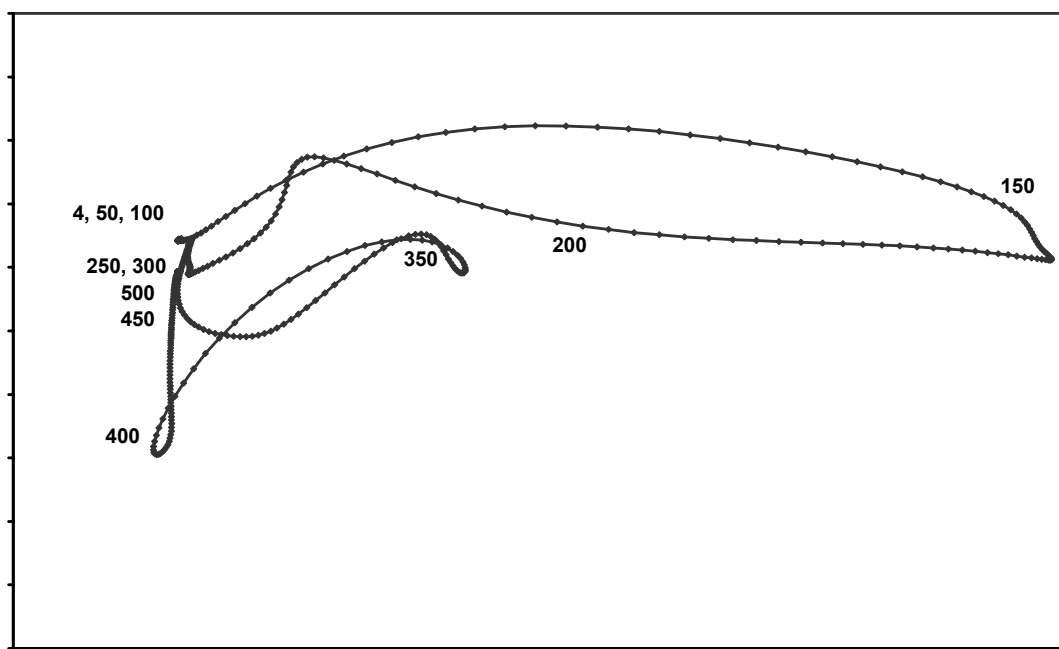


**Σχ. 7.34 :** Οι χαμηλές τιμές C1 για το βίντεο STEADY (C1 low) και τα τοπικά μέγιστα



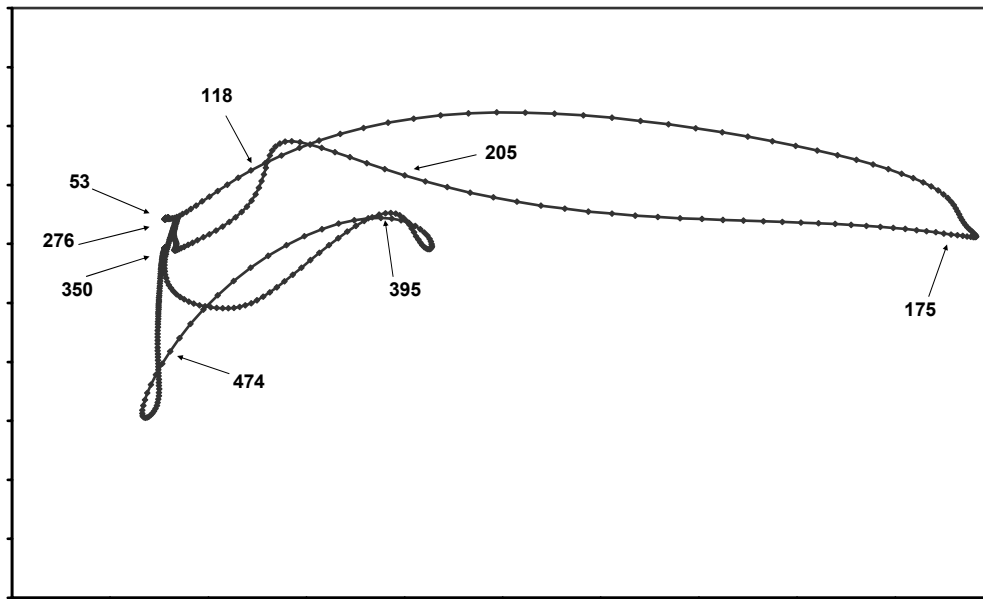
**Σχ. 7.35 :** Οι χαμηλές τιμές C4 για το βίντεο STEADY (C4 low) και τα τοπικά μέγιστα

Μια τόσο μεγάλη διαφορά στις τιμές μεταξύ των τοπικών μεγίστων του ίδιου μέτρου καμπυλότητας (όπως ισχύει για το C1 και το C4) εξηγείται από την ακραία συμπεριφορά της καμπύλης του βίντεο, αφού έχουμε διαδοχικές εναλλαγές μεταξύ σχεδόν απόλυτης ακινησίας (άρα συμπίκνωση πολύ κοντινών σημείων της καμπύλης) και μεγάλης μεταβολής (δηλαδή απομακρυσμένα σημεία της καμπύλης). Αυτό φαίνεται στο σχ. 7.36, όπου σημειώνονται ανά 50 τα frames που αντιστοιχούν στα σημεία.

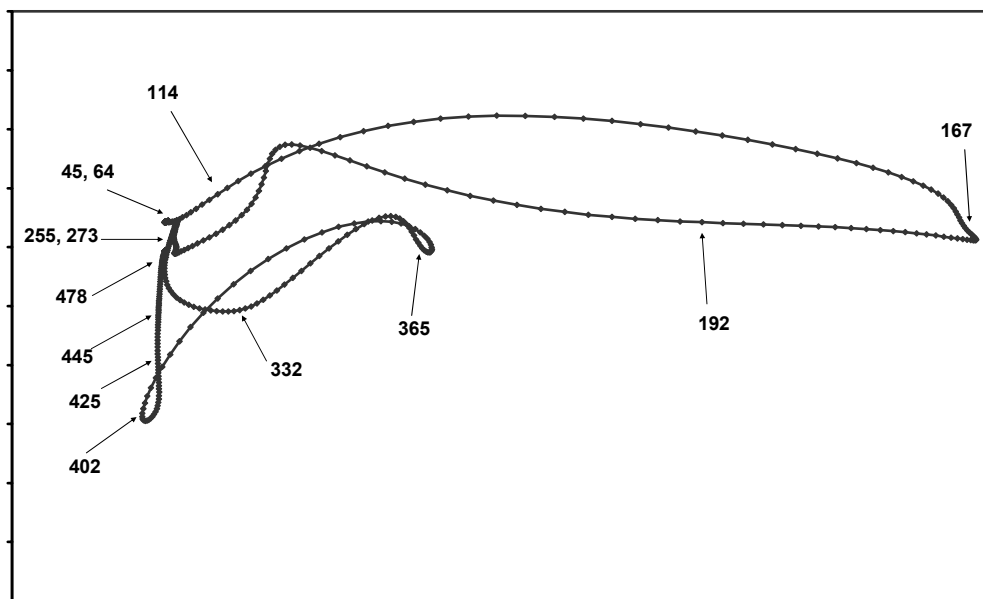


**Σχ. 7.36 :** Η προβολή της καμπύλης του STEADY στο επίπεδο Y3 – Y8, όπου σημειώνεται ο αριθμός των αντίστοιχων frames του βίντεο (ανά 50 τιμές)

Όπως βλέπουμε από την προβολή της καμπύλης, το βίντεο χωρίζεται σε δύο νοηματικά μεταβαλλόμενα τμήματα και σε τρία τμήματα ακινησίας. Τα τμήματα χωρίς μεταβολή είναι τα σημεία αριστερά στην καμπύλη που αντιστοιχούν στην αρχή, στη μέση και στο τέλος του βίντεο (frames 4-50-100, 250-300 και 450-500). Τα τμήματα στα οποία συμβαίνουν δραματικές αλλαγές είναι τα δύο μακροσκελή στελέχη της καμπύλης (frames 150-200 και 350-400). Όλα τα σημεία τοπικών μεγίστων για τις C1 και C4 φαίνονται στα επόμενα σχήματα.



Σχ. 7.37 : Τα χαρακτηριστικά σημεία της μεθόδου C1 στην προβολή Y3 – Y8



Σχ. 7.38 : Τα χαρακτηριστικά σημεία της μεθόδου C4 στην προβολή Y3 – Y8

## Παρατηρήσεις

1. Η καμπυλότητα  $C1$ , που αποτελεί την επιτάχυνση του διανύσματος χαρακτηριστικών ως προς τον χρόνο, παρουσιάζει μόνο 1 μέγιστο για κάθε τμήμα ακινησίας (σχ. 7.34) και 2-3 μέγιστα για τα τμήματα μεταβολών (σχ. 7.32). Αυτό οφείλεται στην ταχύτητα της μεταβολής, όπου στα σταθερά τμήματα η ταχύτητα είναι σχεδόν μηδέν, ενώ στα άλλα δύο στελέχη η μεταβολή γίνεται με συνεχώς απότομες αυξομειώσεις (γρήγορη κίνηση) και προκαλεί αύξηση της καμπυλότητας  $C1$  (επιτάχυνση).
2. Αντίθετα, η  $C4$  δεν επηρεάζεται από την ταχύτητα μεταβολής (γρήγορη ή αργή κίνηση μέσα στο βίντεο), αλλά από τον τρόπο καμπύλωσης στον χώρο. Στα τμήματα ακινησίας τα μέγιστα είναι 1-2 (σχ. 7.33), ενώ στα άλλα στελέχη των μεταβολών είναι 3-5 (σχ. 7.35).
3. Από τα διαγράμματα των σχημάτων 7.37, 7.38 παρατηρούμε ότι και οι δύο μέθοδοι έχουν τοπικά μέγιστα και για τα 2 είδη των τμημάτων (σταθερό και μεταβλητό περιεχόμενο). Όμως η  $C4$  δίνει συνολικά περισσότερα σημεία στα σταθερά τμήματα, αναλογικά με την  $C1$ , στα οποία η καμπυλότητα είναι πολύ μεγαλύτερη απ' ό,τι στα σημεία μεταβολής. Αυτό συμβαίνει διότι το “στοιχειώδες”  $ds$  είναι διαφορετικό κάθε φορά με αποτέλεσμα να παραμορφώνεται η τιμή που υπολογίζουμε για την καμπυλότητα και να εμφανίζεται μεγαλύτερη εκεί όπου ελαττώνεται πολύ το  $ds$  (συμπύκνωση των σημείων).
4. Η καμπυλότητα  $C1$  φαίνεται ότι δίνει καλύτερα αποτελέσματα, διότι οι τιμές της καμπυλότητας είναι μεγαλύτερες για τα σημεία όπου έχω τις αλλαγές στο βίντεο. Αυτό όμως είναι σχετικό με την ταχύτητα με την οποία γίνονται οι μεταβολές και γι' αυτό δεν εκφράζει το μέγεθος της σημασίας του περιεχομένου (σε συνολική μεγάλη μετακίνηση του διανύσματος χαρακτηριστικών θα αντιστοιχούμε σχετικά μεγάλη σημασία). Μεταβολή μεγάλης σημασίας που γίνεται αργά θα δώσει τιμή στην  $C1$  μικρότερη απ' ό,τι σε μεταβολή μικρότερης σημασίας, που όμως συμβαίνει ταχύτατα.
5. Και οι δύο μέθοδοι  $C1$  και  $C4$  φαίνεται να υπολειτουργούν η κάθε μία για διαφορετικό λόγο, αλλά επειδή στο συγκεκριμένο βίντεο η  $C1$  δίνει διαφορετικά

διαγράμματα με λιγότερα σημεία σε σταθερό πλάνο αντίθετα με την C4, θα δώσουμε και για τις δύο τα αποτελέσματα σε εικόνες (εξαγόμενα χαρακτηριστικά καρέ) στο Κεφάλαιο των Πειραμάτων.

### 7.3.3 Συγκεντρωτικά Αποτελέσματα

Τα αποτελέσματα των μεθόδων CS, SCA, C1 και C4 για καθένα από τα 6 βίντεο του δείγματος παρουσιάζονται στους πίνακες που ακολουθούν, όπου η μέθοδος CS εφαρμόστηκε με κριτήριο σύγκλισης  $\varepsilon = \alpha \cdot \text{μέγιστη απόσταση}$ , όπου το  $\alpha$  καθορίζεται ανάλογα με το βίντεο ώστε να δίνει περίπου τον μέσο αριθμό σημείων των C1 και C4, όπου στα 6 αρχεία του δείγματος θέσαμε τις τιμές (με τη σειρά) 0.2, 0.8, 0.1, 0.25, 0.3 και 0.2. Η μέθοδος SCA εφαρμόστηκε με αριθμό σημείων περίπου κοντά στον μέσο όρο του πλήθους που προκύπτει από τις C1, C4 και CS. Οι μέθοδοι C1 και C4 εξάγουν αυτόματα τα χαρακτηριστικά σημεία (λαμβάνοντας τα τοπικά μέγιστα της καμπυλότητας). Η ομαλοποίηση της καμπύλης γίνεται με τις γνωστές συνθήκες που αναφέρονται στην παράγραφο 3.3.3 (μόνο για το βίντεο STEADY ειδικά, με παράθυρο ομαλοποίησης  $NW = (\text{αριθμός frames}) / 16$ ).

**Πίνακας 7.8 :** Τα keyframes που προκύπτουν από τις μεθόδους CS, SCA, C1 και C4 για την ακολουθία FORMULA1 (1 πλάνο)

CS	SCA	C1	C4
4	4	76	80
256	162	148	161
346	264	266	211
420	338	313	246
435	433	434	320
535	504	533	391
564	544	568	437
603	602	649	603
747	758	674	678
838	861	768	759
880	935	793	838
966	1017	844	937
1006	1093	939	1014
1085	1246	1045	1094
1118	1313	1090	1152
1147	1348	1149	1196
1314	1399	1197	1243
1442	1436	1253	1316
1480	1484	1321	1345
1587	1587	1347	1391
		1391	1484
		1431	1523
		1485	

**Πίνακας 7.9 :** Τα keyframes που προκύπτουν από τις μεθόδους CS, SCA, C1 και C4 για την ακολουθία FORMULA2 (9 πλάνα)

Πλάνα	CS	SCA	C1	C4
1 <sup>ο</sup>	4 168 924	4 275 435 924	57 97 133 166	71 101
2 <sup>ο</sup>	925 964 990 1032	925 1018 1043 1067	934 940 945	939
3 <sup>ο</sup>	1067 1068 1192 1230	1068 1174 1214 1230	1086	1081 1091 1100 1108 1113
4 <sup>ο</sup>	1231 1462 1676	1231 1374 1390 1676	1259 1276	1261
5 <sup>ο</sup>	1677 1733 1812 1849 1923	1677 1718 1764 1923	1691 1708 1720	1697 1709 1718 1733 1741 1756
6 <sup>ο</sup>	1924 2013 2081 2170 2352 2437	1924 2236 2295 2437	1953 1977 1992 2012 2034 2043	1956 1974 1991 2015
7 <sup>ο</sup>	2438 2483 2549	2438 2459 2482 2549	2452 2458 2462 2471	2447
8 <sup>ο</sup>	2550 2595 2776	2550 2626 2677 2776	2563 2576	2572 2583
9 <sup>ο</sup>	2777 2868 2902 3048 3112 3177	2777 2930 3043 3177	2802 2820 2854 2902 2922 2956 2967 2985 2997 3014 3041 3076 3116 3137 3156	2802 2820 2834 2862 2898 2929 2946 2972 2981 3003 3028 3044 3090 3111 3125 3136 3148



**Πίνακας 7.10 :** Τα keyframes που προκύπτουν από τις μεθόδους CS, SCA, C1 και C4 για την ακολουθία RALLY1 (1 πλάνο)

CS	SCA	C1	C4
4	4	81	83
90	115	186	142
149	224	253	199
226	301	356	253
299	352	386	297
323	388	422	348
405	426	537	391
425	467	681	427
463	501	763	469
479	531	826	501
542	572	847	529
553	608	951	578
587	668	1001	597
667	707	1029	650
728	764	1117	713
815	816	1142	765
883	928	1185	815
970	995	1243	878
1036	1038	1274	923
1107	1105	1317	994
1143	1145	1341	1039
1160	1181		1101
1260	1240		1148
1291	1274		1182
1422	1422		1234
			1288
			1310
			1358

**Πίνακας 7.11 :** Τα keyframes που προκύπτουν από τις μεθόδους CS, SCA, C1 και C4 για την ακολουθία RALLY2 (1 πλάνο)

CS	SCA	C1	C4
4	4	42	51
79	178	83	79
136	200	102	98
151	236	146	123
200	263	206	143
222	286	222	179
285	321	268	200
300	392	317	235
369	407	395	266
379	435	429	300
406	463	486	324
462	480	563	373
490	493	635	390
523	518	664	431
583	610		463
607	634		518
650	649		558
702	702		582
			609
			633
			668

**Πίνακας 7.12 :** Τα keyframes που προκύπτουν από τις μεθόδους CS, SCA, C1 και C4 για την ακολουθία RALLY3 (1 πλάνο)

CS	SCA	C1	C4
4	4	60	103
337	123	102	133
356	196	178	204
377	344	272	239
415	418	353	274
490	468	466	338
562	562	588	382
650	596	651	417
674	694	695	468
756	739	775	526
778	774	842	596
831	807	913	653
915	883		698
1032	915		738
	1032		805
			881
			916
			981

**Πίνακας 7.13 :** Τα keyframes που προκύπτουν από τις μεθόδους CS, SCA, C1 και C4 για την ακολουθία STEADY (1 πλάνο)

CS	SCA	C1	C4
4	4	53	47
118	112	118	66
152	119	175	114
194	166	205	167
213	211	276	192
336	292	350	255
374	336	395	273
389	374	474	332
399	401		365
509	509		402
			425
			445
			478

Παραθέτουμε εδώ τα χαρακτηριστικά καρέ που προκύπτουν από το σύστημα αυτόματης εξαγωγής χαρακτηριστικών σημείων με τις μεθόδους καμπυλότητας C1 και C4, με είσοδο την ακολουθία RALLY2 και την ακολουθία RALLY3. Τα keyframes για τα βίντεο αυτά φαίνονται στα σχήματα 7.39 έως 7.42, διατεταγμένα με τη σειρά που εμφανίζονται στην ακολουθία αναπαραγωγής.



keyframe 42



keyframe 83



keyframe 102



keyframe 146



keyframe 206



keyframe 222



keyframe 268



keyframe 317



keyframe 395



keyframe 429



keyframe 486



keyframe 563

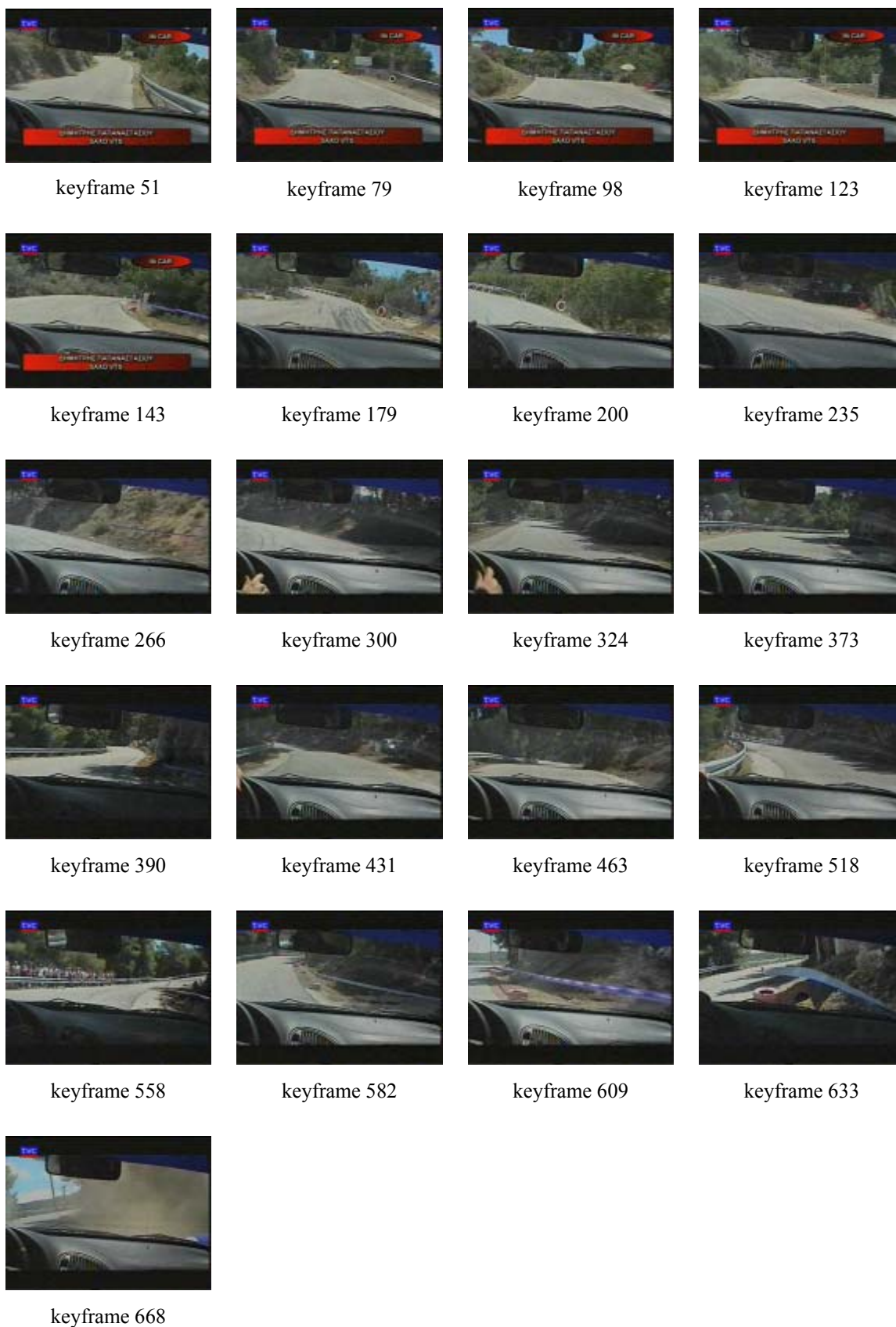


keyframe 635



keyframe 664

**Σχ. 7.39 :** Τα χαρακτηριστικά καρέ για την ακολουθία RALLY2 με την μέθοδο C1



**Σχ. 7.40 :** Τα χαρακτηριστικά καρέ για την ακολουθία RALLY2 με την μέθοδο C4



keyframe 60



keyframe 102



keyframe 178



keyframe 272



keyframe 353



keyframe 466



keyframe 588



keyframe 651



keyframe 695



keyframe 775



keyframe 842



keyframe 913

**Σχ. 7.41 :** Τα χαρακτηριστικά καρέ για την ακολουθία RALLY3 με την μέθοδο C1





keyframe 103



keyframe 133



keyframe 204



keyframe 239



keyframe 274



keyframe 338



keyframe 382



keyframe 417



keyframe 468



keyframe 526



keyframe 596



keyframe 653



keyframe 698



keyframe 738



keyframe 805



keyframe 881



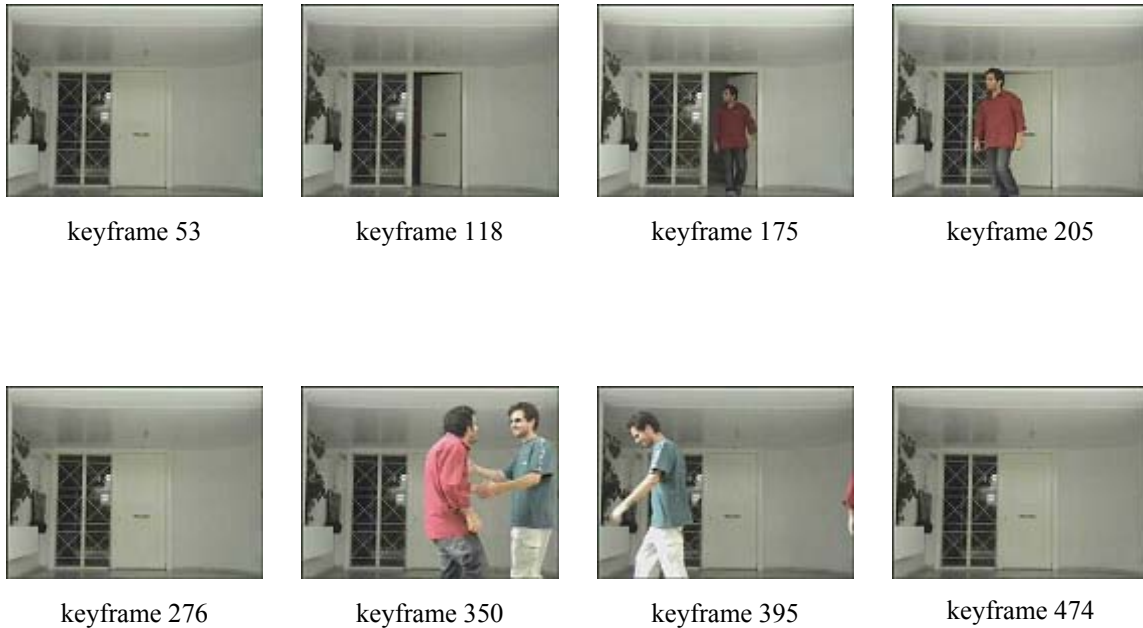
keyframe 916



keyframe 981

Σχ. 7.42 : Τα χαρακτηριστικά καρέ για την ακολουθία RALLY3 με την μέθοδο C4

Για το βίντεο STEADY, δίνονται τα χαρακτηριστικά καρέ με τις μεθόδους C1 και C4 στα σχήματα 7.43 και 7.44 αντίστοιχα, με σκοπό την καλύτερη κατανόηση των αποτελεσμάτων στην ακραία αυτή περίπτωση ακινησίας και κίνησης. Στις περιοχές με την σταθερή εικόνα είναι γνωστό ότι η C4 εσφαλμένα θεωρεί πολλά σημαντικά σημεία, όπως εξηγήθηκε παραπάνω.



**Σχ. 7.43 :** Τα χαρακτηριστικά καρέ για την ακολουθία STEADY με την μέθοδο C1



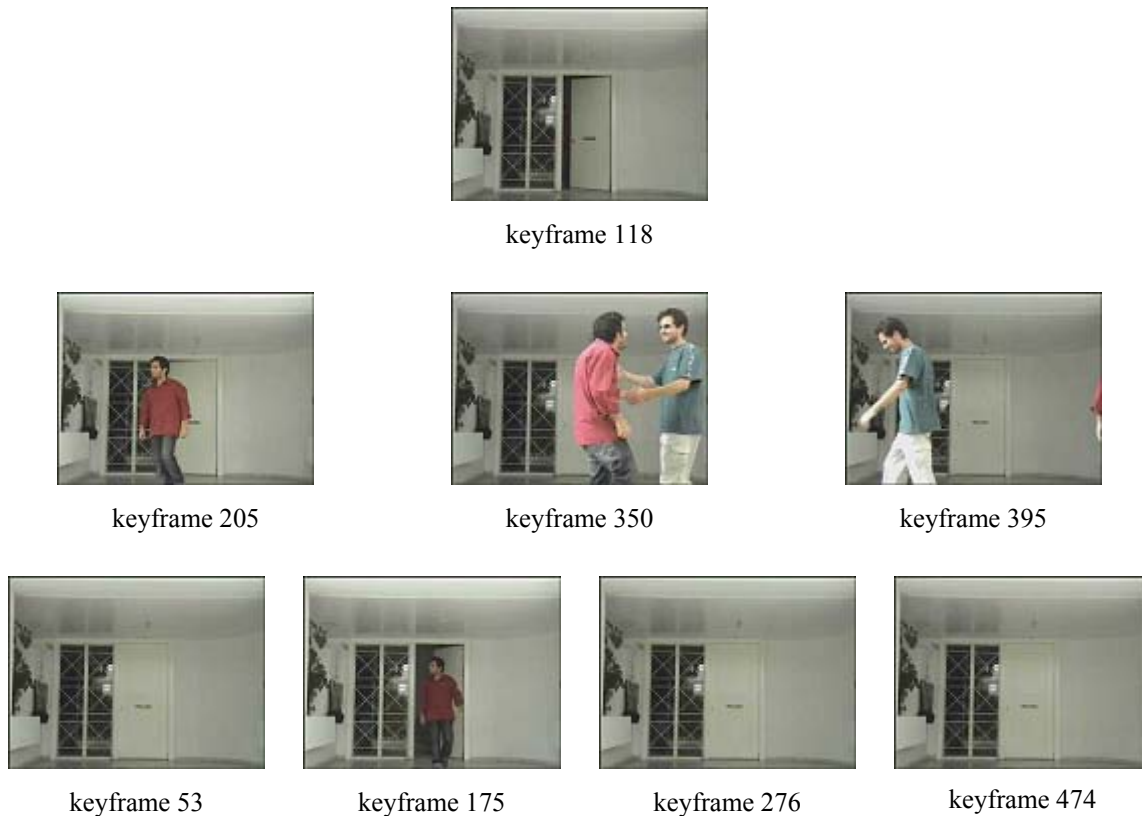
**Σχ. 7.44 :** Τα χαρακτηριστικά καρέ για την ακολουθία STEADY με την μέθοδο C4

Μπορούμε επίσης να ταξινομήσουμε τα χαρακτηριστικά καρέ σε ένα ιεραρχικό μοντέλο με επίπεδα λεπτομέρειας για την περιγραφή του βίντεο. Γνωρίζοντας τις τιμές της καμπυλότητας, θεωρούμε ως πιο σημαντικά εκείνα τα χαρακτηριστικά καρέ που παρουσιάζουν μεγαλύτερες τιμές στην καμπυλότητα και τα τοποθετούμε ανάλογα σε επίπεδα φθίνουσας σημασίας, τα οποία μπορούν να καθοριστούν με πολλούς τρόπους (π.χ. αυξανόμενος αριθμός καρέ σε κάθε χαμηλότερο επίπεδο ή καρέ που έχουν καμπυλότητα πάνω από ένα ποσοστό του ολικού μεγίστου).

Στο σχήμα 7.45 φαίνονται τα χαρακτηριστικά καρέ (keyframes) που αντιστοιχούν σε τοπικά μέγιστα της μεθόδου C1 για το βίντεο STEADY, σε ιεραρχικό μοντέλο. Στο πρώτο επίπεδο παρουσιάζεται το keyframe με την μεγαλύτερη τιμή καμπυλότητας, στο



επόμενο επίπεδο δίνονται τα keyframes με τις επόμενες 3 μικρότερες τιμές και στο τρίτο επίπεδο όλα τα υπόλοιπα. Στο τελευταίο επίπεδο θα μπορούσαν να προστεθούν τα καρέ που δίνουν τοπικό ελάχιστο μεταξύ των τοπικών μεγίστων, για μεγαλύτερη λεπτομέρεια, αντιπροσωπεύοντας έτσι σημεία σταθερής μεταβολής του περιεχομένου.



**Σχ. 7.45 :** Τα χαρακτηριστικά καρέ για την ακολουθία STEADY με την μέθοδο C1 σε ιεραρχία

Από το σχήμα 7.42, με την μέθοδο C1, έχουμε ως βασικότερο keyframe το 118, που όπως φαίνεται δεν παρουσιάζει την πιο σημαντική αλλαγή (δεν πρόκειται για δραματική αλλαγή στο feature vector όπως συμβαίνει στο δεύτερο επίπεδο), αλλά η αλλαγή αυτή συμβαίνει με μεγαλύτερο επιταχυνόμενο ρυθμό από τις αλλαγές στα frames 205, 350 και 395, με αποτέλεσμα να δίνει και το ολικό μέγιστο της καμπυλότητας. Για την C4 δεν κατασκευάσαμε ιεραρχική δομή, αφού τα περισσότερα καρέ αφορούν το σταθερό πλάνο, όμως βλέπουμε ότι η C4 δίνει περισσότερα keyframes (3) για την διαδικασία συνάντησης των δύο προσώπων από την C1 (2).

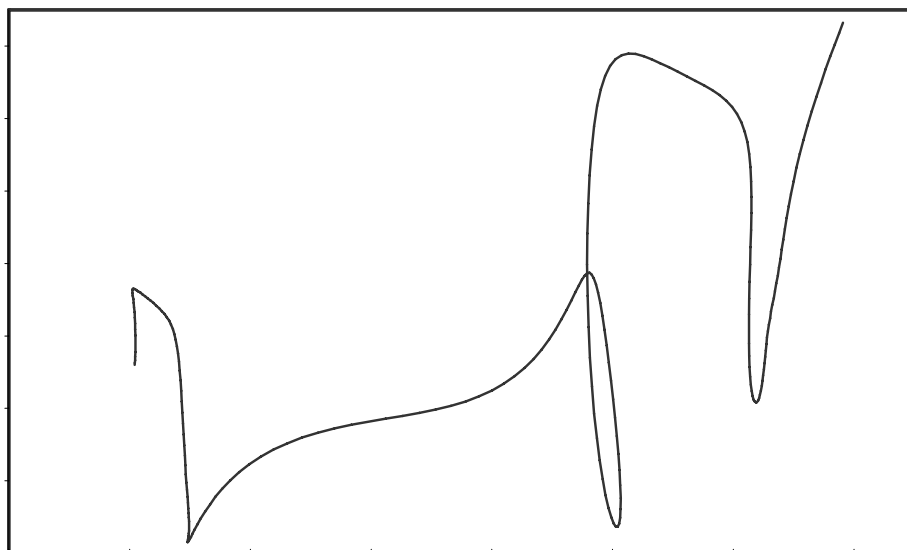
Για λόγους οικονομίας χώρου δεν μπορούμε να παρουσιάσουμε το σύνολο των εξαγόμενων keyframes για όλο το δείγμα και όλες τις μεθόδους που αναλύθηκαν.

## 7.4 Πειράματα στην Εφαρμογή της Οπτικής Προσοχής

Όπως παρατηρούμε από το σχήμα 7.40, υπάρχουν χαρακτηριστικά καρέ που είναι όμοια με τα γειτονικά τους, κυρίως με την μέθοδο C4, όπου keyframes με παραπλήσιο περιεχόμενο είναι τα 98-123-143, 300-324-373, και 431-463. Αυτό σημαίνει ότι στη σύνοψη που κάνουμε με τις παραπάνω μεθόδους περιέχονται, εκτός από σημαντικά καρέ, και καρέ με παρόμοια σημασία που αποτελούν περιττή πληροφορία.

Θα επιχειρήσουμε να χρησιμοποιήσουμε την τεχνική της Οπτικής Προσοχής (Visual Attention), ούτως ώστε να εξάγουμε σε κάθε καρέ μία σημαντική περιοχή (που προκαλεί την εντύπωση) από την οποία να λαμβάνουμε τα χαρακτηριστικά και να κατασκευάζουμε νέα καμπύλη στον 30-D χώρο.

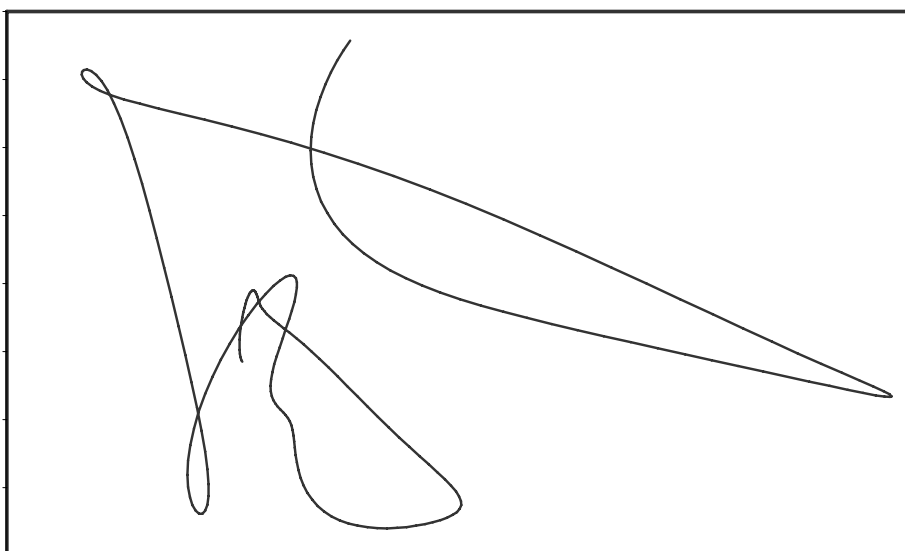
Εφαρμόζουμε την παραπάνω μέθοδο εστίασης της προσοχής με είσοδο το βίντεο RALLY2, για το οποίο υπολογίσαμε τα διαγράμματα Y1-Y6, Y4-Y6, Y5-Y6 και παρουσιάσαμε τα keyframes για κάθε μέθοδο. Εδώ θα πειραματιστούμε μόνο με την μέθοδο C4, η οποία βλέπουμε ότι δίνει μεγάλη πλεονάζουσα πληροφορία. Κατασκευάζοντας την νέα καμπύλη, την προβάλλουμε στα ίδια επίπεδα (για τα frames 61 έως 310) και παραθέτουμε εδώ τα διαγράμματα με τις προβολές (σχ. 7.46, 7.47, 7.48) ώστε να γίνει άμεση σύγκριση με την καμπύλη χωρίς την οπτική προσοχή.



Σχ. 7.46 : Εφαρμογή Οπτικής Προσοχής (RALLY2): προβολή της καμπύλης στο Y1 – Y6



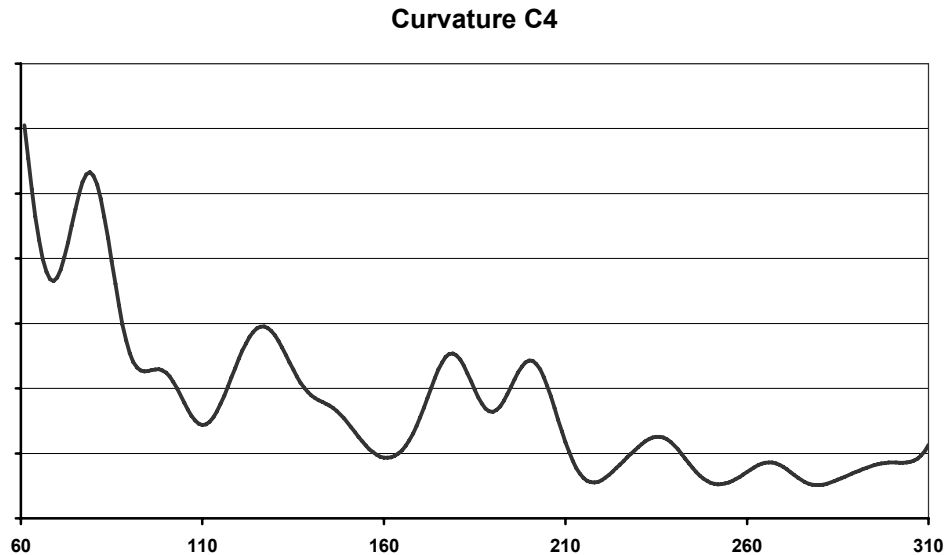
**Σχ. 7.47 :** Εφαρμογή Οπτικής Προσοχής (RALLY2): προβολή της καμπύλης στο Y4 – Y6



**Σχ. 7.48 :** Εφαρμογή Οπτικής Προσοχής (RALLY2): προβολή της καμπύλης στο Y5 – Y6

Παρατηρώντας τα διαγράμματα αυτά και συγκρίνοντάς τα με τα αντίστοιχα διαγράμματα της παραγράφου 7.3.2.2 (σχ. 7.23 – 7.25), βλέπουμε ότι οι καμπύλες διαφοροποιούνται, αλλά διατηρούν την κύρια μορφή τους και αυτό ισχύει για τις 3 προβολές. Σε μερικά σημεία έχουμε πιο ομαλές μεταβολές των τιμών της καμπύλης, αλλά το κύριο χαρακτηριστικό των νέων προβολών είναι ότι τώρα η τροχιά συνήθως παρουσιάζει απότομες αλλαγές της πορείας της. Η ομαλοποιημένη καμπύλη φαίνεται στις προβολές Y1 – Y6 και Y4 – Y6, ενώ η προβολή Y5 – Y6 είναι σχεδόν όμοια με την αρχική. Από το σχήμα 7.43 μπορούμε να παρατηρήσουμε ότι τώρα η προβολή

αποτελείται από πολύ καμπύλα σημεία (με μεγάλη καμπυλότητα) τα οποία συνδέονται με τμήματα σχεδόν ευθύγραμμα. Η καμπυλότητα περιμένουμε να δίνει τώρα πιο ξεκάθαρες αιχμές, με πολλές χαμηλές τιμές μεταξύ των τοπικών μεγίστων. Στο σχήμα 7.49 παρουσιάζεται το μέτρο της καμπυλότητας που υπολογίζεται με τη μέθοδο C4.



**Σχ. 7.49 :** Εφαρμογή Οπτικής Προσοχής (RALLY2): μέθοδος καμπυλότητας 4 (C4)

Όπως φαίνεται με την άμεση σύγκριση του διαγράμματος της καμπυλότητας C4 από τα σχήματα 7.49 και 7.27 (με και χωρίς εφαρμογή της οπτικής προσοχής), τα τοπικά μέγιστα που έχει τώρα η τιμή της C4 μεταξύ των frames 60 και 310 είναι λιγότερα από πριν (8 αντί για 9). Βλέπουμε ότι το τοπικό μέγιστο στο σημείο 143 (frame) δεν υπάρχει τώρα στην τιμή της καμπυλότητας. Αυτό σημαίνει ότι στο σημείο αυτό η καμπύλη έχει ομαλοποιηθεί τόσο ώστε δεν παρουσιάζει καμπύλωση, και αυτό οφείλεται στην εξαγωγή των χαρακτηριστικών μόνο από τη σημαντική περιοχή, δηλαδή από ένα υποσύνολο της εικόνας.

Αυτή είναι και η αναμενόμενη συμπεριφορά του συστήματος με την εφαρμογή της Οπτικής Προσοχής, δηλαδή ο σκοπός ήταν να μεταβάλλουμε την καμπύλη, εξάγοντας τα χαρακτηριστικά από περισσότερο σημαντικές περιοχές της εικόνας, ούτως ώστε το διάνυσμα χαρακτηριστικών να περιγράφει όσο το δυνατόν καλύτερα την εξέλιξη του περιεχομένου του βίντεο, και όχι απλά την μεταβολή του ιστογράμματος ολόκληρης της εικόνας. Με αυτό τον τρόπο επιδιώκουμε μία πιο περιεκτική περιγραφή του βίντεο, η οποία θα μας δώσει ακόμα καλύτερα αποτελέσματα ως προς την σύνοψη που θα κάνουμε

λαμβάνοντας ακόμα λιγότερα χαρακτηριστικά καρέ. Τα χαρακτηριστικά καρέ που θα απορριφθούν είναι εκείνα που βρίσκονται σε ακραία σημεία της αρχικής καμπύλης και οφείλονται σε μεταβολή του ιστογράμματος μιας μη σημαντικής περιοχής του καρέ (γι' αυτό και θεωρούνται περιττά).

Παραθέτουμε εδώ (σχ. 7.50) τα 20 keyframes που προέκυψαν με την εφαρμογή της οπτικής προσοχής και στη συνέχεια με τον χειρισμό της καμπύλης με την μέθοδο καμπυλότητας C4.

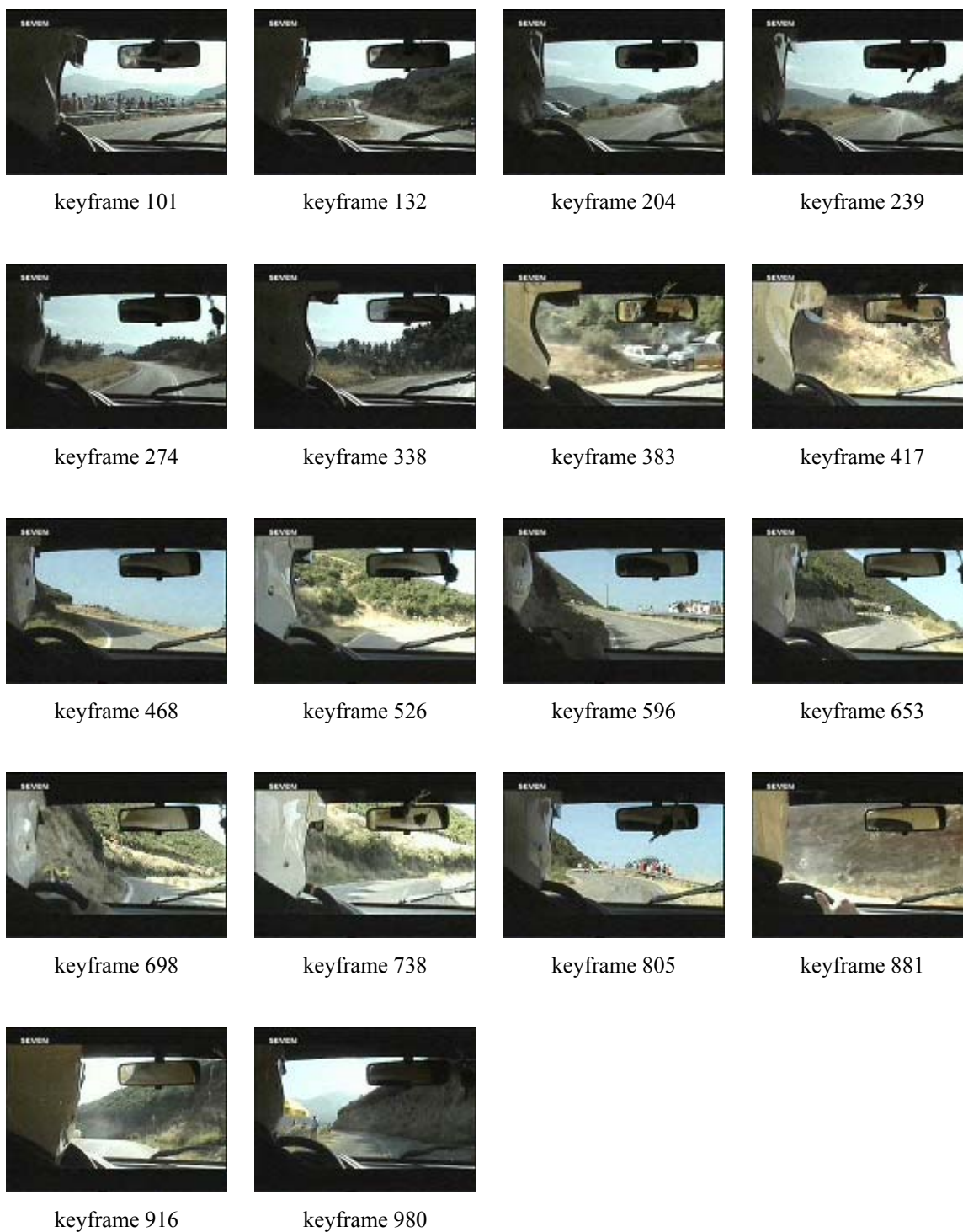


**Σχ. 7.50 :** Τα χαρακτηριστικά καρέ για την ακολουθία RALLY2 με εφαρμογή της Οπτικής Προσοχής και της μεθόδου καμπυλότητας C4

Από το διάγραμμα της C4 παρατηρήσαμε ότι δεν υπάρχει τοπικό μέγιστο στο σημείο 143. Από την ομάδα των όμοιων καρέ 98-123-143 απορρίπτεται το καρέ 143, οπότε η περιοχή αυτή αντιπροσωπεύεται από τα καρέ 98 και 127 (βλέπε σχήμα 7.50). Έτσι με την εφαρμογή της οπτικής προσοχής δεν χάνεται το σημαντικό σημείο αυτό του βίντεο (η περιοχή των frames 98 και 127) από την περίληψη που γίνεται, αλλά απορρίπτεται και περιττή πληροφορία (το καρέ 143).

Για τις δύο άλλες ομάδες που περιέχουν περιττή πληροφορία (frames 300-324-373 και 431-463), βλέπουμε ότι τώρα δεν απορρίπτεται κάποιο από αυτά τα καρέ. Η μέθοδος της οπτικής προσοχής σε αυτήν την περίπτωση δεν ομαλοποίησε αρκετά την καμπύλη ώστε να μην παρουσιάζει αιχμή σε κάποιο από τα καρέ. Περιμέναμε μεγαλύτερη βελτίωση στην μέθοδο με λιγότερα πλεονάζοντα χαρακτηριστικά σημεία και τελικά η βελτίωση περιορίστηκε σε ένα μόνο σημείο από τα εξεταζόμενα frames (60-310).

Αντίστοιχα, εφαρμόζουμε την μέθοδο Οπτικής Προσοχής για την ακολουθία RALLY3, για την οποία προκύπτουν τα 18 χαρακτηριστικά καρέ του σχήματος 7.51. Συγκρίνοντας τα χαρακτηριστικά καρέ όπως προκύπτουν από τη μέθοδο C4 με και χωρίς την εφαρμογή της οπτικής προσοχής (δηλαδή τα σχήματα 7.51 και 7.42 αντίστοιχα), παρατηρούμε ότι τελικά στο βίντεο αυτό η μέθοδος δεν βελτιώνει το αποτέλεσμα. Τα χαρακτηριστικά καρέ που εξάγονται στις δύο περιπτώσεις είναι ακριβώς τα ίδια (διαφέρουν το πολύ κατά 1 frame μεταξύ τους). Περιμέναμε βελτίωση του αποτελέσματος, διότι και εδώ υπάρχει πλεονάζουσα πληροφορία, όπως είναι τα frames 101-132 και 204-239.



**Σχ. 7.51 :** Τα χαρακτηριστικά καρέ για την ακολουθία RALLY3 με εφαρμογή της Οπτικής Προσοχής και της μεθόδου καμπυλότητας C4





## ΚΕΦΑΛΑΙΟ 8

# *Συμπεράσματα και Επίλογος*

### 8.1 Συμπεράσματα

Σε αυτή τη διπλωματική εργασία, όπως αναφέρεται και στο εισαγωγικό κεφάλαιο, ο στόχος είναι η μελέτη και η υλοποίηση ενός συστήματος αυτόματης δημιουργίας περίληψης από μία ακολουθία βίντεο, ούτως ώστε να εξαχθεί ένας αριθμός από χαρακτηριστικά καρέ (keyframes), τα οποία θα συμβάλλουν στην γρηγορότερη αναζήτηση και ανάκτηση αρχείων βίντεο στο διαδίκτυο (retrieval) και στην αποδοτικότερη περιήγηση μιας ακολουθίας βίντεο (fast browsing), χωρίς να απαιτείται η παρακολούθηση ολόκληρης της διάρκειας (που απαιτεί χρόνο), αλλά και χωρίς να παραλείπεται σημαντική πληροφορία.

Στο πρώτο στάδιο της διαδικασίας βρίσκεται η κατάτμηση του βίντεο σε πλάνα, στην οποία έχει αφιερωθεί πλήθος ερευνητικών μεθόδων από διάφορες πηγές, μερικές από τις οποίες μελετήθηκαν και συγκρίθηκαν στο Κεφάλαιο 2 και τελικά αναπτύχθηκε ένα σύστημα με συνδυασμό τεχνικών, το οποίο δέχεται ως είσοδο κωδικοποιημένο MPEG βίντεο και δίνει στην έξοδό του τα καρέ στα οποία συμβαίνει μία απότομη αλλαγή πλάνου (cut). Στο σύστημα κανόνων που κατασκευάστηκε, με μετρικές που χρησιμοποιήθηκαν από τη βιβλιογραφία και κάποιες νέες μεθόδους που προστέθηκαν (όπως η έννοια της κύριας αναφοράς στα B-frames), χρησιμοποιήθηκε μία τέτοια διάταξη των κανόνων που να επιτρέπει την γρήγορη επεξεργασία του βίντεο.

Το σύστημα ανίχνευσης αλλαγής πλάνου χαρακτηρίζεται από ταχείς υπολογισμούς (επεξεργάζεται απ' ευθείας στο MPEG, δεν ελέγχει όλα τα συνεχόμενα frames, χρησιμοποιεί DC τιμές και όχι όλα τα pixels της εικόνας), αλλά και από πολύ

καλές επιδόσεις, όπως αποδεικνύεται από τα πειράματα που έγιναν. Χρησιμοποιώντας ποικίλου περιεχομένου ακολουθίες το σύστημα παρουσιάζει κατά μέσο όρο 96.7 % επιτυχία στον εντοπισμό των πραγματικών αλλαγών (δείκτης recall), ενώ το ποσοστό του σφάλματος διατηρείται σε χαμηλά επίπεδα, περίπου κατά μέσο όρο 10.25 %, που σημαίνει ότι εντοπίζοντας 100 αλλαγές, μόνο οι 10 από αυτές είναι εσφαλμένες. Επίσης, χαρακτηριστικό του συστήματος είναι η ικανότητά του να μην επηρεάζεται από απότομες αλλαγές που συμβαίνουν σε ένα ή δύο καρέ (όπως είναι το φωτογραφικό φλας ή παράσιτα), γεγονός που οφείλεται στον ορισμό της κύριας αναφοράς στα B-frames και στον τελικό κανόνα απόφασης που χρησιμοποιεί τα B-frames. Σε αυτά τα σημεία όπου άλλες μέθοδοι αποτυγχάνουν θεωρώντας αλλαγή πλάνου, η μέθοδος αυτή επιτυγχάνει τις περισσότερες φορές. Το μειονέκτημά της είναι ότι αυξάνει το ποσοστό σφάλματος (μειώνεται ο δείκτης precision) σε βίντεο κωδικοποιημένα με χαμηλό bitrate ή γενικά με χαμηλή ποιότητα (δηλαδή σε μεγάλη συμπίεση), όπου τα διανύσματα κίνησης των B-frames αποπροσανατολίζουν τη μέθοδο. Το ίδιο πρόβλημα συμβαίνει και σε περιπτώσεις μεγάλης κίνησης στο περιεχόμενο των καρέ, το οποίο μπορεί να επιλυθεί με πληροφορία κίνησης από τα motion vectors.

Το επόμενο βήμα για την περίληψη είναι η επιλογή κάποιων χαρακτηριστικών καρέ που να αναπαριστούν το περιεχόμενο κάθε πλάνου. Προς αυτή την κατεύθυνση εργάζονται τεχνικές, όπως περιγράφηκε στην παράγραφο 3.1 του Κεφαλαίου 3, οι οποίες εστιάζουν σε διαφορετικούς τρόπους ανάλυσης των χαρακτηριστικών του βίντεο. Εδώ υιοθετήθηκε η μαθηματική αναπαράσταση των χαρακτηριστικών του βίντεο μέσω του διανύσματος χαρακτηριστικών, με μία καμπύλη στον πολυδιάστατο χώρο των χαρακτηριστικών (feature space), και στη συνέχεια εφαρμόστηκαν μαθηματικές μέθοδοι για την εξαγωγή των χαρακτηριστικών σημείων. Στο Κεφάλαιο 4 αναπτύχθηκε η μέθοδος όπου χρησιμοποιείται ο ορισμός της καμπυλότητας (curvature) για να οριστεί ένα μέτρο της καμπύλωσης (4 τύποι από τους οποίους χρησιμοποιούνται οι δύο, C1 και C4), ενώ το Κεφάλαιο 5 αναφέρθηκε στις δύο μεθόδους που χρησιμοποιούνται για τις συγκρίσεις, απλοποίηση καμπύλης CS και προσέγγιση περιγράμματος σχήματος SCA.

Στις εφαρμογές που έγιναν, τόσο με μοντελοποιημένες 2D και 3D καμπύλες όσο και σε πειράματα σε πραγματικό βίντεο, είναι άμεσα εμφανής η διαφορά των C1 και C4 στις τιμές των μεγίστων που παρουσιάζουν. Τα τοπικά μέγιστα των τιμών τους, σύμφωνα με την θεωρία, δίνουν τα σημεία που ζητάμε. Πρέπει να σημειώσουμε εδώ ότι η C1 βασίζεται στην παράγωγο ως προς τον χρόνο και γι' αυτό εκφράζει την εξέλιξη στο βίντεο με τον χρόνο (δηλαδή την επιτάχυνση της μεταβολής που συμβαίνει στο

περιεχόμενο), ενώ η C4 βασίζεται στην παράγωγο ως προς το διάστημα που διανύει το διάλυμα χαρακτηριστικών πάνω στην καμπύλη και γι' αυτό εκφράζει την μεταβολή με βάση το περιεχόμενο (δηλαδή όχι πόσο γρήγορα αλλάζει το περιεχόμενο από το ένα καρέ στο άλλο όπως η C1, αλλά πώς μεταβάλλεται ως προς τη θέση του στον χώρο των χαρακτηριστικών).

Από τα πειράματα του Κεφαλαίου 7 προκύπτει ότι όλες οι μέθοδοι εκτός από την SCA δίνουν καλά αποτελέσματα, με τις C1 και C4 να δίνουν τα περισσότερα σωστά χαρακτηριστικά σημεία στις προβολές της καμπύλης για τα δύο πειράματα αντίστοιχα. Σε γενικό επίπεδο, θεωρώντας τα κοινά χαρακτηριστικά σημεία των προβολών, οι C1, C4 και CS εξάγουν περίπου με τα ίδια ποσοστά χαρακτηριστικά σημεία κοντά στα πραγματικά σημεία απότομης αλλαγής της καμπύλωσης. Η μέθοδος C4 εμφανίζει το μεγαλύτερο ποσοστό επιτυχίας σύμφωνα και με τα δύο πειράματα που διεξήχθησαν (βλέπε πίνακες 7.5β και 7.7β). Επομένως οι καλύτερες μέθοδοι προκύπτουν να είναι οι C1 και C4, αφού αυτές δίνουν τα καλύτερα ποσοστά επιτυχίας στον εντοπισμό χαρακτηριστικών σημείων είτε αυτοτελώς για κάθε προβολή (μέση τιμή και για τις 3) είτε από κοινού για τις 3 προβολές (πίνακες 7.5 και 7.7).

Η μέθοδος SCA δεν έχει ικανοποιητική απόδοση (χαμηλό ποσοστό επιτυχίας και υψηλό σφάλματος), γεγονός που ήταν αναμενόμενο, αφού μπορεί να θεωρηθεί ότι προσεγγίζει την μέθοδο CS, ως απλοποίηση της καμπύλης στη μία διάσταση. Η προσέγγιση όμως της 30D καμπύλης με πολλά ευθύγραμμα τμήματα (1D) προφανώς δεν διατηρεί τα χαρακτηριστικά της και γι' αυτό υπάρχει απώλεια πληροφορίας. Τα ποσοστά επιτυχίας είναι μικρά (περίπου στο 50 %) καταρχάς επειδή προκύπτουν για επιθυμητά χαρακτηριστικά σημεία στις προβολές μερικών συνιστωσών της καμπύλης, αλλά και για τον λόγο του ότι οι τιμές της καμπυλότητας ομαλοποιήθηκαν ώστε να παραμείνουν μόνο οι απότομες αιχμές της (δηλαδή οι μεγάλες μεταβολές). Οι μέθοδοι CS και SCA ρυθμίστηκαν ώστε να δίνουν ίδιο αριθμό σημείων με τις C1 και C4, οπότε για όλες τις μεθόδους έχουμε περιορισμένο αριθμό σημείων που εξάγουν (σημεία με πιο ακραία συμπεριφορά).

Η σύγκριση των πλεονεκτημάτων και μειονεκτημάτων των C1, C4, CS και SCA είναι απαραίτητη, ώστε ναδειχθεί και πρακτικά η αξία της μεθόδου της καμπυλότητας. Η μέθοδος SCA δεν αρκεί για τον ορισμό ακραίων σημείων, αφού όπως είπαμε η καμπύλη 30D δεν “χωράει” σε μία διάσταση. Εκτός από αυτό, η μέθοδος απαιτεί τον καθορισμό των εξαγόμενων σημείων εκ των προτέρων, άρα παρουσιάζει σοβαρό μειονέκτημα αφού για να ορίσουμε αριθμό σημείων ανά πλάνο και ανά βίντεο είναι

πρακτικά αδύνατο. Δεν έχει πρακτική αξία όταν πρέπει κάθε φορά να εφαρμόζεται κάποια άλλη μέθοδος για να γνωρίζουμε έναν τέτοιο κατάλληλο αριθμό σημείων (εκτός αν ορίζαμε αριθμό ανάλογο της διάρκειας του πλάνου, αλλά αυτό μας περιορίζει την ικανότητα της περίληψης δυναμικά, με βάση το περιεχόμενο).

Τέλος, η μέθοδος CS έχει θεωρητική μαθηματική βάση στη διαμέριση καμπύλης με βάση τις προβολές σε άξονες ([DeM98]). Απαιτεί όμως και αυτή εκ των προτέρων τον ορισμό του κριτηρίου σύγκλισης, ενός αριθμού  $\epsilon$  που θα καθορίζει την περαιτέρω ή όχι διαμέριση της καμπύλης. Με βάση τα πειράματα και τις δοκιμές που έγιναν και αφότου χρησιμοποιήσαμε ως  $\epsilon$  ένα ποσοστό της μέγιστης απόστασης η οποία υπολογίζεται στην πρώτη επανάληψη του αλγορίθμου, καταλήξαμε στο συμπέρασμα ότι ούτε αυτό το ποσοστό μπορεί να είναι δεδομένο και ίδιο για όλα τα βίντεο. Αυτό διότι ανάλογα με το περιεχόμενο, έχουμε διαφορετική κατανομή των σημείων του διανύσματος χαρακτηριστικών και οι αποστάσεις μεταξύ τους δεν είναι γνωστές εκ των προτέρων. Οπότε η χρήση της μεθόδου CS, παρόλο που εξάγει δυναμικό αριθμό χαρακτηριστικών σημείων, περιορίζει την αυτόματη ικανότητα του συστήματος να εξάγει χαρακτηριστικά καρέ από οποιαδήποτε ακολουθία βίντεο δέχεται ως είσοδο. Αυτό επιτυγχάνεται πλήρως με τις μεθόδους C1 και C4.

Η καμπυλότητα είναι ένα μέτρο της καμπύλωσης της καμπύλης, δηλαδή τα μέγιστα δεν δείχνουν μόνο το πού καμπυλώνει αλλά και το πόσο καμπυλώνει, οπότε το διάγραμμα της καμπυλότητας θα μπορούσε να χρησιμοποιηθεί για μια κατηγοριοποίηση (classification) των ήδη εξαχθέντων χαρακτηριστικών σημείων και να γίνει μία κλιμακωτή περίληψη με τα χαρακτηριστικά καρέ σε επίπεδα λεπτομέρειας. Υλοποιείται έτσι ένα ιεραρχικό μοντέλο για την παρουσίαση των χαρακτηριστικών καρέ, με επίπεδα που θα περιέχουν keyframes ανάλογα με την τιμή της καμπυλότητας που παρουσιάζουν (δηλαδή το μέτρο της σημασίας τους). Στο ανώτερο επίπεδο θα υπάρχει το keyframe με το ολικό μέγιστο καμπυλότητας (που θα αντιπροσωπεύει ολόκληρο το πλάνο), ενώ στα επόμενα επίπεδα θα περιέχονται τα λιγότερο σημαντικά keyframes (δομή που περιγράφει με όλο και πιο αναλυτικό τρόπο το βίντεο, ως μία ιεραρχική περιεκτική σύνοψη, ιεραρχία των keyframes).

Στο Κεφάλαιο 6 για την βελτίωση των αποτελεσμάτων χρησιμοποιούμε το υπολογιστικό μοντέλο της οπτικής προσοχής (Visual Attention – VA), όπως διατυπώθηκε από τους Itti et al. ([Itt98]), εφαρμόζοντας σε κάθε καρέ και λαμβάνοντας μία σημαντική περιοχή. Εξάγουμε τα χαρακτηριστικά από αυτήν την περιοχή και επιθυμούμε με το διάνυσμα χαρακτηριστικών να αναπαραστήσουμε περισσότερο

ενδιαφέροντα σημεία και άρα περισσότερο αξιοπρόσεκτες μεταβολές στο περιεχόμενο του βίντεο. Με τις δοκιμές προκύπτει η νέα καμπύλη ομαλοποιημένη σε σχέση με την αρχική, αλλά με περισσότερο απότομες καμπυλώσεις και συνεπώς με πιο απότομες αιχμές στην καμπυλότητα. Εξάγονται τα χαρακτηριστικά σημεία που οφείλονται σε μεταβολές στο περιεχόμενο της σημαντικής περιοχής, ενώ χαρακτηριστικά σημεία λόγω μεταβολών της υπόλοιπης περιοχής (μη σημαντικής) θα πρέπει να έχουν ομαλοποιηθεί.

Το μοντέλο της Οπτικής Προσοχής (Visual Attention) εφαρμόστηκε στην περίπτωση δύο βίντεο από το δείγμα. Στο πρώτο από αυτά γνωρίζουμε ότι αρχικά η μέθοδος C4 εξάγει μερικά περιττά keyframes, αφού υπάρχουν 3 ομάδες γειτονικών keyframes τα οποία έχουν παρόμοιο περιεχόμενο. Μετά την εφαρμογή της οπτικής προσοχής, στα νέα αποτελέσματα απουσιάζει το ένα από τα περιττά keyframes της μιας ομάδας, αλλά χωρίς να αντιμετωπίζεται η πλεονάζουσα πληροφορία των άλλων ομάδων περιττών καρέ. Η καμπύλη παρουσιάζεται ελαφρώς ομαλοποιημένη όπως περιμέναμε, αλλά αυτό δεν συμβαίνει σε τέτοιο βαθμό που να έχουμε δραματικές επιδράσεις στα αποτελέσματα. Στο δεύτερο βίντεο παρατηρήσαμε ότι τα αποτελέσματα της μεθόδου δεν βελτιώνονται με την εφαρμογή του μοντέλου, αφού προκύπτουν ακριβώς τα ίδια χαρακτηριστικά καρέ.

## 8.2 Μελλοντικές Επεκτάσεις

Στην παράγραφο αυτή γίνεται αναφορά σε ενδεχόμενους τρόπους μετατροπής και επέκτασης του συστήματος περίληψης βίντεο με σκοπό τη βελτίωση των αποτελεσμάτων και τη διεύρυνση των δυνατοτήτων του.

Όσον αφορά στο κομμάτι την κατάτμησης σε πλάνα, μία βελτίωση του συστήματος αποτελεί η χρήση πληροφορίας κίνησης. Μία μέθοδος μέσω μετρικών από τα διανύσματα κίνησης (motion vectors) του MPEG προσφέρει γνώση σχετικά με την κίνηση στο πλάνο και έτσι μπορούν να διαχωριστούν σημεία μεγάλης κίνησης από σημεία με αλλαγή πλάνου. Επίσης, για την επέκταση σε ανίχνευση dissolves ή fades μπορεί να χρησιμοποιηθεί το ιστόγραμμα από τις DC τιμές του DCT domain στα P-frames, ενώ για την ανίχνευση wipes είναι προτιμότερη η εφαρμογή μεθόδων σε αποκωδικοποιημένο βίντεο για μεγαλύτερη ακρίβεια.

Στην εξαγωγή των χαρακτηριστικών από τα καρέ του βίντεο, εκτός από το χρωματικό ιστόγραμμα που αποτελεί δοκιμασμένη και αποτελεσματική μέθοδο, μπορεί να προστεθεί πληροφορία κίνησης, όπως έχει χρησιμοποιηθεί συχνά από πολλούς ερευνητές. Αφού το σύστημα δέχεται ως είσοδο αρχεία κωδικοποιημένα στο MPEG, μπορούμε από τα διανύσματα κίνησης να πάρουμε την πληροφορία που θέλουμε (και αυτό μπορεί να γίνει κατά την διαδικασία της ανίχνευσης πλάνων για εξοικονόμηση χρόνου). Σύμφωνα με τους Divakaran et al. ([Div01]) από τα motion vectors των P-frames μπορούμε να εξάγουμε ένα μέτρο κίνησης. Για όλο το GOP αθροίζονται όλα τα μέτρα κίνησης και προκύπτει ο περιγραφέας δραστηριότητας κίνησης για το I-frame που ακολουθεί, ενώ η δραστηριότητα κίνησης για όλο το πλάνο προκύπτει από το άθροισμα από όλα τα I-frames. Λαμβάνοντας ως ένα ακόμη χαρακτηριστικό την δραστηριότητα κίνησης (διευρύνουμε το διάνυσμα χαρακτηριστικών), θα εντοπίζουμε μετακινήσεις εσωτερικά στην εικόνα, όπου το ιστόγραμμα συνήθως μένει σταθερό.

Γενικότερη και αποδεκτή λύση για την περιγραφή της καμπύλης στις πολλές διαστάσεις παρουσιάζεται η γενικευμένη καμπυλότητα, δηλαδή η θεώρηση που βασίζεται στην μετατόπιση  $ds$  πάνω στην καμπύλη και δεν εξαρτάται από την ταχύτητα διαγραφής της. Πρέπει όμως να αντιμετωπιστεί το μειονέκτημα που εμφανίζεται με τα διάφορα  $ds$  να είναι άνισα και μη στοιχειώδη (ενώ για τα  $dt$  δεν υπάρχει τέτοιο πρόβλημα διότι είναι ίσα με το χρονικό διάστημα μεταξύ δύο συνεχόμενων frames). Αυτό μπορεί να εξαλειφθεί υπολογίζοντας το μικρότερο  $ds_{min}$  που εμφανίζεται μεταξύ δύο διαδοχικών frames και να χωριστούν όλα τα υπόλοιπα  $ds$  σε τμήματα μήκους  $ds_{min}$ . Έτσι, με κάποιο είδος παρεμβολής (με κάποια πολυωνυμική προσέγγιση, όχι όμως γραμμική παρεμβολή) θα υπολογιστεί η καμπυλότητα ανά ίσα τμήματα της καμπύλης που μπορούν να θεωρηθούν στοιχειώδη. Τα σημεία υπολογισμού δεν θα αντιστοιχούν όλα σε καρέ του βίντεο και το keyframe θα εξαχθεί ως το σημείο με αντίστοιχο καρέ στο βίντεο που βρίσκεται πιο κοντά στο τοπικό μέγιστο.

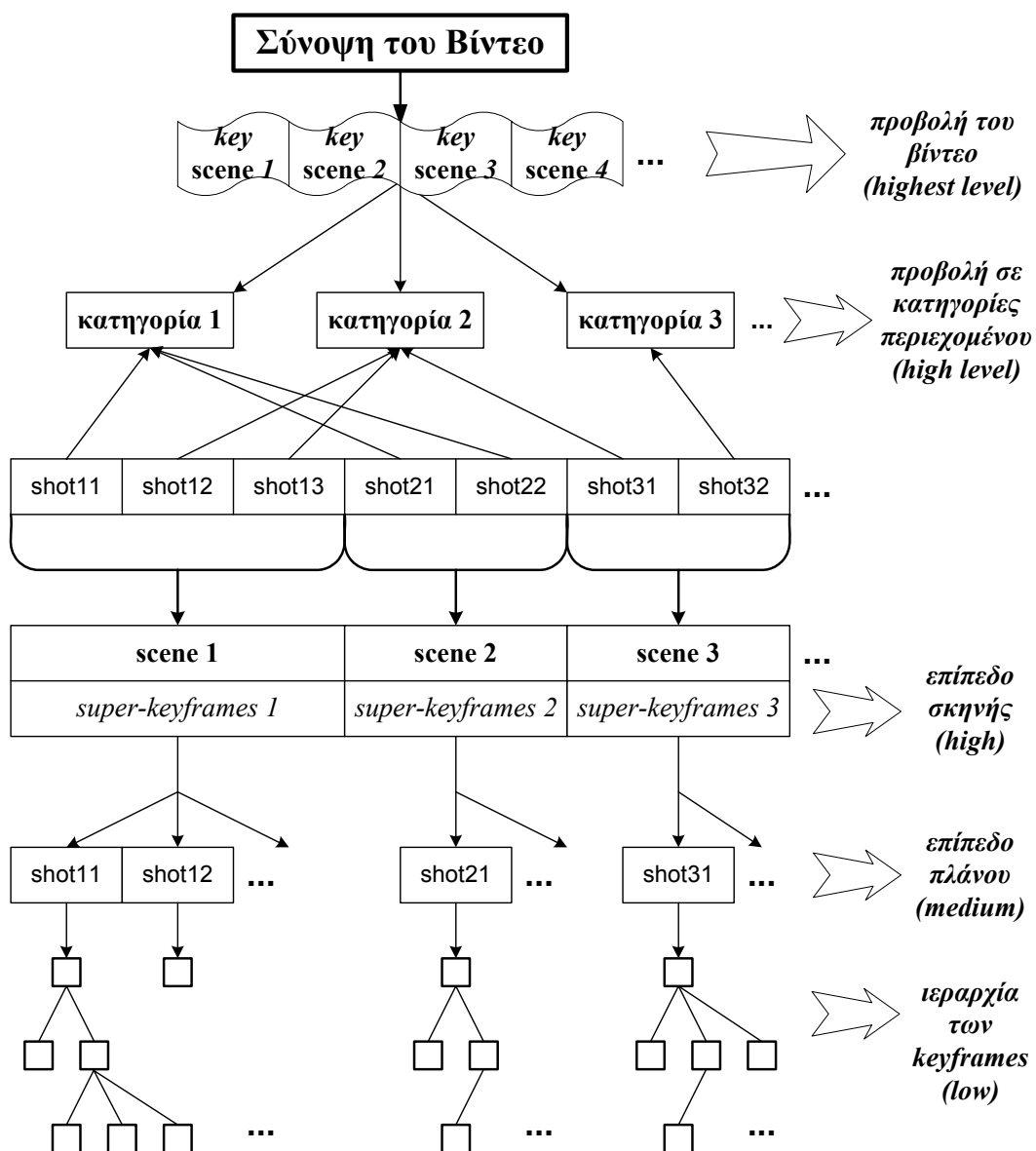
Κατά την εφαρμογή της μεθόδου της καμπυλότητας για την εξαγωγή των χαρακτηριστικών σημείων της καμπύλης, με τον τύπο 4 (C4), βασιζόμαστε στη θεωρία των γενικευμένων καμπυλοτήτων για πολυδιάστατες καμπύλες ( $n$  – διάστατος χώρος). Η πρώτη από τις  $n - 1$  καμπυλότητες αποτελεί την τάση απόκλισης της καμπύλης από την ευθεία γραμμή (γενικευμένος ορισμός για τον  $nD$  χώρο), ενώ οι υπόλοιπες αποτελούν τις στρέψεις ως προς τα υπερεπίπεδα. Οι στρέψεις εκφράζουν την τάση απομάκρυνσης από τα υπερεπίπεδα (δηλαδή από τους χώρους των 2, 3, ...,  $n - 1$  διαστάσεων). Μία πιο γενική θεώρηση της καμπύλης μας οδηγεί στην διαπίστωση ότι οι  $n - 1$  γενικευμένες

καμπυλότητες περιγράφουν πλήρως την γεωμετρική μορφή της καμπύλης. Γι' αυτό η χρήση ταυτόχρονα και των  $n - 1$  αυτών τιμών για κάθε σημείο και ο συνδυασμός τους με κάποιο σύστημα κανόνων θα δώσει έναν πιο γενικό και ακριβή τρόπο εκτίμησης της κίνησης στον χώρο, δηλαδή σε έναν πιο βελτιωμένο τρόπο εξαγωγής χαρακτηριστικών σημείων από την καμπύλη (για περισσότερα βλέπε βιβλιογραφία [Spi99], [Glu66], [WWW]).

Για τη βελτίωση των αποτελεσμάτων εξαγωγής χαρακτηριστικών καρέ εφαρμόσαμε την μέθοδο Visual Attention σε κάθε καρέ της ακολουθίας, από την οποία προκύπτει μία πιο σύντομη περίληψη, που όμως περιέχει ακόμα περιττά όμοια καρέ. Εφαρμόζοντας στις δύο διαστάσεις (εικόνα) εξάγουμε σημαντικές περιοχές, αλλά δεν λαμβάνουμε υπ' όψη την κίνηση που υπάρχει στο βίντεο και την εστίαση του ενδιαφέροντος σε σημεία όπου υπάρχει κάποιο εντυπωσιακό γεγονός. Αυτό μπορεί να εφαρμοστεί με βάση την μέθοδο που προτείνουν οι Rapantzikos et al. ([Rap05]) για την εφαρμογή ενός τρισδιάστατου μοντέλου Οπτικής Προσοχής, το οποίο στις δύο διαστάσεις έχει την εικόνα (κάθε καρέ) και στην τρίτη διάσταση έχει τον χρόνο, δηλαδή τη χρονική εξέλιξη των καρέ. Το χωροχρονικό αυτό μοντέλο οπτικής προσοχής επιτυγχάνει τον εντοπισμό σημαντικών περιοχών με δύο έννοιες: με την χωρική έννοια, όπως το ανθρώπινο μάτι εστιάζει σύμφωνα με τα γνωστά σε περιοχές που ξεχωρίζουν από τη "γειτονιά" τους, και με την τρισδιάστατη έννοια καθώς σε μια ακολουθία βίντεο το μάτι εστιάζει σε περιοχές όπου υπάρχει μεταβολή κατά την εξέλιξη των καρέ, άρα ξεχωρίζουν τρισδιάστατοι όγκοι (salient volumes) σε αναλογία με τις περιοχές (salient regions). Αν εξάγουμε τα χαρακτηριστικά από τους σημαντικούς όγκους (salient volumes) θα κατασκευάσουμε μία καμπύλη ακόμα πιο περιεκτική, αφού θα εκφράζει την μεταβολή μόνο των pixels που περιέχουν κάποιο σημαντικό γεγονός. Επειδή το περιεχόμενο αυτών των όγκων είναι και το σημαντικό νόημα του πλάνου, η περίληψη με τα χαρακτηριστικά καρέ που θα εξάγουμε θα είναι πιο περιεκτική και με λιγότερη πλεονάζουσα πληροφορία.

Τέλος, με βάση την ιδέα της ταξινόμησης των χαρακτηριστικών καρέ σύμφωνα με την σημασία τους, μπορούμε να κατασκευάσουμε ένα συνολικό ιεραρχικό πρότυπο πάνω στο οποίο θα υλοποιείται η τεχνική της περίληψης ([Yeu97], [Han99]). Το κατώτερο επίπεδο θα αφορά τα πλάνα και θα περιέχει τα keyframes για κάθε πλάνο σε μια εσωτερική ιεραρχία, όπως κατασκευάστηκε στην παράγραφο 7.3.3, ανάλογα με την λεπτομέρεια που επιθυμούμε (σημαντικότητα των keyframes). Το αμέσως ανώτερο θα είναι το πρώτο high level επίπεδο που θα περιέχει τις σκηνές με ομαδοποιημένα τα πλάνα

που αφορούν την ίδια συνεχόμενη δραστηριότητα (και αποτελούν αυτοτελείς γεγονός), όπου θα έχουν επιλεγεί κάποια keyframes (super-keyframes) από κάθε πλάνο ώστε να αντιπροσωπεύουν τη σκηνή. Στο επόμενο επίπεδο θα γίνεται μία ομαδοποίηση (clustering) των πλάνων για την κατηγοριοποίησή τους σε μεγάλες κατηγορίες scenes (όπως για παράδειγμα δράση, συνομιλίες, τοπία, πρόσωπα, εντός – εκτός πόλης κτλ.). Τέλος, στο ανώτατο επίπεδο θα περιέχονται οι πιο σημαντικές σκηνές (key scenes), οι οποίες και θα συνθέτουν τη σύνοψη του βίντεο στο υψηλότερο δυνατό επίπεδο (highest level). Παρακάτω δίνεται σχηματικά το ιεραρχικό μοντέλο της πολυεπίπεδης σύνοψης του βίντεο, με βάση και την ιδέα των Hanjalic et al. ([Han99]).



Σχ. 8.1: Ιεραρχικό μοντέλο μιας πολυεπίπεδης σύνοψης του βίντεο



# Παράρτημα

## Π.1 Απόδειξη του τύπου της δεύτερης παραγώγου ως προς s

Στο Κεφάλαιο 4 στον υπολογισμό των παραγώγων ως προς s αναφέρεται η χρήση του τύπου:

$$f'' = \frac{\ddot{f} \cdot \dot{s} - \dot{f} \cdot \ddot{s}}{\dot{s}^3}, \quad (\text{εξ. Π.1})$$

$$\text{όπου } \dot{s} = \frac{ds}{dt}, \ddot{s} = \frac{d^2s}{dt^2}, \dot{f} = \frac{df}{dt}, \ddot{f} = \frac{d^2f}{dt^2}, f'' = \frac{d^2f}{ds^2}$$

ο οποίος αποδεικνύεται με τη βοήθεια του κανόνα της αλυσίδας για τις μερικές παραγώγους.

Υποθέτουμε ότι η  $f(t)$  είναι μία βαθμωτή ή μία διανυσματική συνάρτηση και ότι η συνάρτηση  $s(t)$  είναι μία βαθμωτή και μάλιστα γνησίως μονότονη συνάρτηση, ώστε να είναι αντιστρέψιμη ως προς το  $t$  (δηλαδή ορίζεται η  $t(s)$ ). Έχουμε τους παρακάτω δύο εναλλακτικούς τρόπους απόδειξης:

### 1<sup>ος</sup> τρόπος:

$$f'' = \frac{d^2f}{ds^2} = \frac{d}{ds} \left( \frac{df}{ds} \right) = \frac{d}{ds} \left( \frac{df}{dt} \cdot \frac{dt}{ds} \right) = \frac{d}{ds} \left( \frac{df/dt}{ds/dt} \right) \Rightarrow$$

$$f'' = \frac{d}{dt} \left( \frac{df/dt}{ds/dt} \right) \cdot \frac{dt}{ds} = \frac{\frac{d^2f}{dt^2} \cdot \frac{ds}{dt} - \frac{df}{dt} \cdot \frac{d^2s}{dt^2}}{\left( ds/dt \right)^3} \Rightarrow$$

$$f'' = \frac{\ddot{f} \cdot \dot{s} - \dot{f} \cdot \ddot{s}}{\dot{s}^3}$$

**2<sup>ος</sup> τρόπος:**

$$\dot{f} = \frac{df}{dt} = \frac{df}{ds} \cdot \frac{ds}{dt} = f' \cdot \dot{s}$$

$$\ddot{f} = \frac{d^2 f}{dt^2} = \frac{d}{dt} \left( \frac{df}{dt} \right) = \frac{d}{dt} (\dot{f}) = \frac{d}{dt} (f' \cdot \dot{s}) \Rightarrow$$

$$\ddot{f} = \frac{df'}{dt} \cdot \dot{s} + f' \cdot \frac{d\dot{s}}{dt} = \frac{df'}{ds} \cdot \frac{ds}{dt} \cdot \dot{s} + f' \cdot \frac{d\dot{s}}{dt} \Rightarrow$$

$$\ddot{f} = \frac{d^2 f}{ds^2} \cdot \left( \frac{ds}{dt} \right)^2 + \frac{df}{ds} \cdot \frac{d^2 s}{dt^2} = f'' \cdot \dot{s}^2 + \frac{f'}{\dot{s}} \cdot \ddot{s} \Rightarrow$$

$$f'' = \frac{\ddot{f} \cdot \dot{s} - \dot{f} \cdot \ddot{s}}{\dot{s}^3}$$

## Π.2 Απόδειξη του τύπου 3 για την καμπυλότητα

Αν  $\mathbf{r} = \mathbf{r}(t)$  είναι η παράσταση μιας καμπύλης στον τρισδιάστατο Ευκλείδειο χώρο, από τη βιβλιογραφία είναι γνωστός ο τύπος:

$$k = \frac{|\vec{r} \times \ddot{\vec{r}}|}{|\ddot{\vec{r}}|^3} \quad (\text{εξ. Π.2})$$

που ορίζει την καμπυλότητα με βάση το διάνυσμα  $\mathbf{r}$  που είναι συνάρτηση της τυχαίας παραμέτρου  $t$ .

Για να εφαρμόσουμε τον ορισμό αυτόν στην περίπτωση των  $N$  διαστάσεων, προχωρούμε σε μια προσαρμογή του παραπάνω τύπου με βάση τους ορισμούς εσωτερικού και εξωτερικού γινομένου και Ευκλείδειας νόρμας στον  $N$ -διάστατο χώρο. Έστω ότι το  $\mathbf{r}$  είναι διάνυσμα με  $N$  συνιστώσες, το οποίο έχει πρώτη και δεύτερη παράγωγο ως προς  $t$ .

Για δύο διανύσματα  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $N$  διαστάσεων, δίνεται το εσωτερικό τους γινόμενο και το μέτρο του εξωτερικού τους γινομένου αντίστοιχα ως εξής:

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \cdot \|\vec{b}\| \cdot \cos \theta \quad (\text{εξ. Π.3})$$

$$\|\vec{a} \times \vec{b}\| = \|\vec{a}\| \cdot \|\vec{b}\| \cdot \sin \theta \quad (\text{εξ. Π.4})$$

όπου  $\theta$  η γωνία που ορίζεται μεταξύ των δύο διανυσμάτων.

Ένας άλλος ορισμός του εσωτερικού γινομένου δίνεται από τον τύπο:

$$\vec{a} \cdot \vec{b} = a_1 \cdot b_1 + a_2 \cdot b_2 + \dots + a_N \cdot b_N = \sum_{i=1}^N a_i \cdot b_i \quad (\text{εξ. Π.5})$$

Πρέπει να τονίσουμε ότι ο τύπος της εξίσωσης Π.3 προκύπτει από τη σχέση που ισχύει μεταξύ του εσωτερικού γινομένου και των μέτρων του  $a$ , του  $b$  και του εξωτερικού γινομένου (βλέπε βιβλιογραφία [Χαιν98]):

$$\|\vec{a} \times \vec{b}\|^2 = \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 - (\vec{a} \cdot \vec{b})^2 \quad (\text{εξ. Π.6})$$

Στον  $N$ -διάστατο χώρο δεν γνωρίζουμε την γωνία  $\theta$  μεταξύ των διανυσμάτων, οπότε προσπαθούμε να την απαλείψουμε ώστε να πάρουμε τον τύπο για το μέτρο του εξωτερικού τους γινομένου. Έτσι, από τους τύπους των εξισώσεων Π.4 και Π.6 έχουμε:

$$\|\vec{a} \times \vec{b}\| = \sqrt{\|\vec{a}\|^2 \cdot \|\vec{b}\|^2 - (\vec{a} \cdot \vec{b})^2} = \sqrt{\|\vec{a}\|^2 \cdot \|\vec{b}\|^2 - \left(\sum_i a_i \cdot b_i\right)^2} \quad (\text{εξ. Π.7})$$

Αρα αν τα διανύσματα  $a$  και  $b$  είναι οι 1<sup>η</sup> και 2<sup>η</sup> παράγωγοι του  $r$  ως προς  $t$ , τότε ισχύει:

$$\|\vec{\dot{r}} \times \vec{\ddot{r}}\| = \sqrt{\|\vec{\dot{r}}\|^2 \cdot \|\vec{\ddot{r}}\|^2 - (\vec{\dot{r}} \cdot \vec{\ddot{r}})^2} = \sqrt{\|\vec{\dot{r}}\|^2 \cdot \|\vec{\ddot{r}}\|^2 - \left(\sum_i \dot{r}_i \cdot \ddot{r}_i\right)^2} \quad (\text{εξ. Π.8})$$

Επίσης, για την μετατόπιση  $s$  πάνω στην καμπύλη στον  $N$ -διάστατο χώρο ισχύει:

$$s = s(t) = \int_{t_0}^t \left\| \frac{d\vec{r}}{dt} \right\| dt \Rightarrow \frac{ds}{dt} = \left\| \frac{d\vec{r}}{dt} \right\| \quad (\text{εξ. Π.9})$$

Επειδή ως  $dt$  δεχόμαστε την χρονική απόσταση μεταξύ δύο καρέ του βίντεο, θέτουμε  $dt = 1$  και έτσι προκύπτει ότι η στοιχειώδης μετατόπιση  $ds$  είναι:

$$ds = \dot{s} = \frac{ds}{dt} = \left\| \frac{d\vec{r}}{dt} \right\| = \|\vec{\dot{r}}\| \quad (\text{εξ. Π.10})$$

Χρησιμοποιώντας τις δύο σχέσεις από τις εξισώσεις Π.8 και Π.10, όπως και τον ορισμό της καμπυλότητας μέσω του μέτρου του εξωτερικού γινομένου (εξ. Π.2), προκύπτει ο τύπος 3 της καμπυλότητας (εξ. 4.18) που χρησιμοποιήθηκε στο Κεφάλαιο 4 για την καμπυλότητα που δίνεται με βάση την πρώτη και δεύτερη παράγωγο του διανύσματος και την πρώτη παράγωγο της μετατόπισης  $s$  (παράγωγοι ως προς  $t$ ):

$$curvature3 = |\vec{f}''(s)| = \frac{\sqrt{\dot{s}^2 \cdot \|\ddot{\vec{f}}\|^2 - \left(\dot{\vec{f}} \cdot \ddot{\vec{f}}\right)^2}}{\dot{s}^3} \quad (\text{εξ. Π.11})$$

όπου με  $\vec{f}$  συμβολίζεται το διάνυσμα που περιγράφει την καμπύλη (δηλαδή το  $\vec{r}$ ).

Ένας άλλος τρόπος για την εξαγωγή ενός αντίστοιχου τύπου με βάση τις παραγώγους ως προς  $t$  μπορεί να προκύψει από τον νόμο των συνημιτόνων. Για δύο διανύσματα  $\vec{u}$  και  $\vec{v}$  ισχύει:

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2 \cdot \|\vec{u}\| \cdot \|\vec{v}\| \cdot \cos \theta \quad (\text{εξ. Π.12})$$

Από την παραπάνω σχέση λύνουμε ως προς το συνημίτονο και από την τριγωνομετρική ταυτότητα παίρνουμε το ημίτονο:

$$|\sin \theta| = \sqrt{1 - \cos^2 \theta} \quad (\text{εξ. Π.13})$$

οπότε από την εξίσωση Π.4 εκφράζουμε το μέτρο του εξωτερικού γινομένου συναρτήσει των μεγεθών που επιθυμούμε.

## Π.3 Γενικευμένες Καμπυλότητες ή Συναρτήσεις

### Καμπυλότητας

Έστω  $C$  μία καμπύλη ορισμένη στον  $N$ -διάστατο Ευκλείδειο χώρο, η οποία έχει παραμετρική εξίσωση:

$$C : \vec{r} = \vec{r}(t), t \in I, I \rightarrow \mathbb{R}^n$$

Έστω επίσης ότι για αυτήν την καμπύλη, η διανυσματική συνάρτηση  $r$  είναι παραγωγίσιμη τάξης  $n$ , δηλαδή υπάρχουν οι πρώτες  $n$  παράγωγοι και είναι συνεχείς συναρτήσεις.

Για τον υπολογισμό των  $n-1$  συναρτήσεων καμπυλότητας κατασκευάζουμε το πλαίσιο του Frenet (Frenet Frame, αντίστοιχα με το τρίεδρο Frenet-Serret formulas για τις τρεις διαστάσεις), το οποίο αποτελείται από τα  $n$  ορθοκανονικά διανύσματα  $e_1(t)$ ,  $e_2(t)$ , ...,  $e_n(t)$ , τα διανύσματα Frenet, τα οποία προκύπτουν με εφαρμογή της ορθοκανονικοποίησης Gram-Schmidt. Η διαδικασία αυτή περιλαμβάνει την κατασκευή των  $n$  ορθοκανονικών διανυσμάτων  $u$ , η οποία και γίνεται ταυτόχρονα με την κατασκευή του πλαισίου του Frenet (Frenet Frame) με διαδοχικό τρόπο ως εξής (τα μεγέθη  $u_i$ ,  $r^{(n)}$ ,  $e_i$  αφορούν διανυσματικά μεγέθη, ενώ οι παράγωγοι προκύπτουν ως προς  $t$ ):

$$u_1 = r'(t) \Rightarrow e_1(t) = \frac{u_1}{\|u_1\|} = \frac{r'(t)}{\|r'(t)\|} \quad (\text{εξ. Π.14})$$

$$u_2 = r'' - \frac{(r'' \cdot e_1) \cdot e_1}{\|e_1\|} \Rightarrow e_2(t) = \frac{u_2}{\|u_2\|} \quad (\text{εξ. Π.15})$$

.....

.....

$$u_n = r^{(n)} - \sum_{i=1}^{n-1} (r^{(n)} \cdot e_i) \cdot e_i \Rightarrow e_n(t) = \frac{u_n}{\|u_n\|} \quad (\text{εξ. Π.16})$$

Οι γενικευμένες καμπυλότητες  $k_1, k_2, \dots, k_{n-1}$ , προκύπτουν από την παραγωγή των διανυσμάτων του πλαισίου Frenet και την προβολή της παραγώγου στο αμέσως επόμενο διάνυσμα Frenet, μέσω του εσωτερικού γινομένου τους. Για την περίπτωση των συναρτήσεων ως προς  $t$ , διαιρούμε με το μέτρο της πρώτης παραγώγου του  $r$ . Έτσι οι  **$n-1$**  γενικευμένες καμπυλότητες ή συναρτήσεις καμπυλότητας είναι:

**Συναρτήσεις Καμπυλότητας:**

$$k_i(t) = \frac{e'_i(t) \cdot e_{i+1}(t)}{\|r'(t)\|} \quad (\text{εξ. Π.17})$$

(για  $i = 1, 2, \dots, n-1$ )

Η έννοια των συναρτήσεων καμπυλότητας εξηγείται εδώ αναλυτικά, ξεκινώντας από τις γνωστές μικρότερες διαστάσεις (1, 2, 3), όπου είναι ευκολότερο να γίνει αντιληπτή η ιδιότητα της καμπύλωσης στον χώρο.

Στην μία διάσταση, δηλαδή στην ευθεία γραμμή δεν έχουμε καμπύλωση, γι' αυτό και η καμπυλότητά της είναι μηδέν. Για την περιγραφή της γραμμής, όσον αφορά την γεωμετρική της σημασία δεν χρειάζεται να γνωρίζουμε κάποια παράμετρο, αφού δεν ενδιαφέρει η παραμετρική της εξίσωση και η μετατόπιση  $s$  που γίνεται πάνω στη γραμμή, αλλά η γεωμετρική μορφή της, η οποία είναι πάντοτε ίδια.

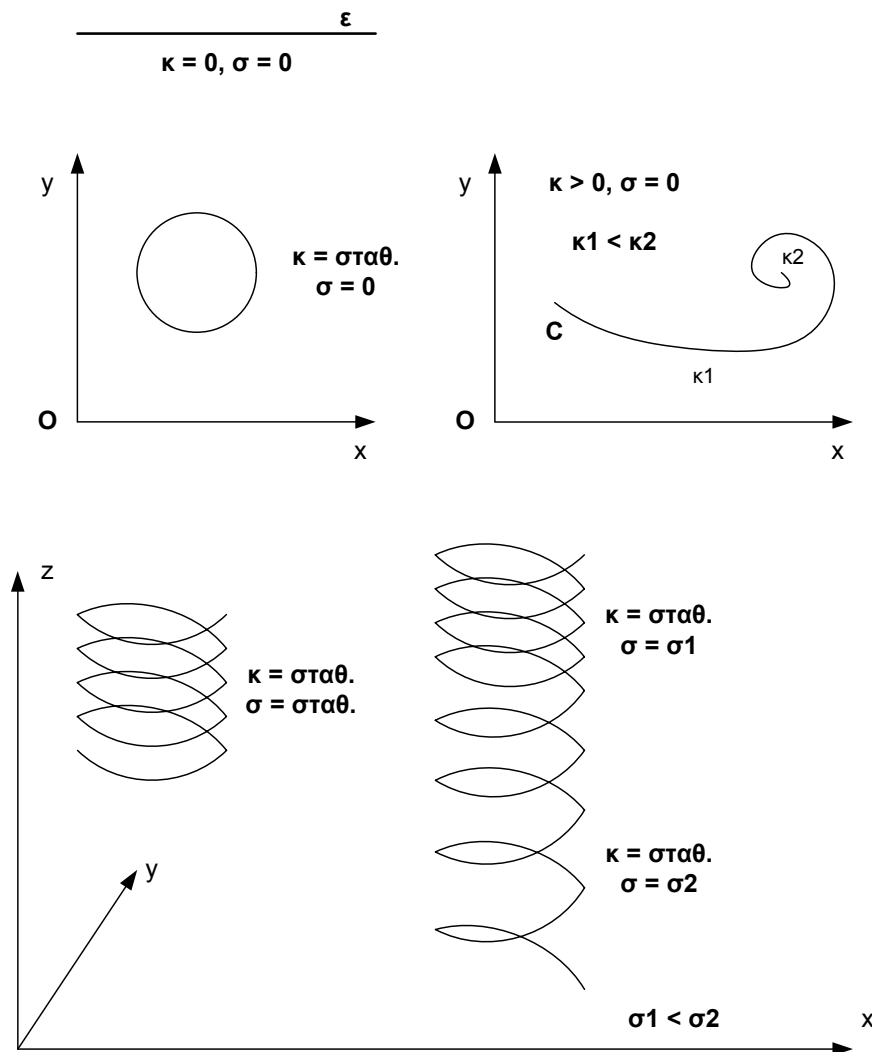
Στις δύο διαστάσεις, η καμπύλη μπορεί να καμπυλώνει, οπότε να παίρνει διαφορετική γεωμετρική μορφή από την απλή μορφή μιας ευθείας γραμμής. Η καμπυλότητα  $\kappa$  χρησιμοποιείται για να εκφράσει την τάση της καμπύλης να “φεύγει” από την ευθεία γραμμή.

Στις τρεις διαστάσεις εκτός από την καμπυλότητα  $\kappa$  που ορίστηκε για το διδιάστατο επίπεδο και που έχει και εδώ την ίδια σημασία, χρησιμοποιούμε και την στρέψη  $\sigma$  για να εκφράσουμε την τάση της καμπύλης να “ξεφεύγει” από το επίπεδο των δύο διαστάσεων. Με την καμπυλότητα και τη στρέψη μπορούμε να περιγράψουμε πλήρως μια καμπύλη και να γνωρίζουμε την γεωμετρική μορφή της ανεξάρτητα από τη θέση της στο χώρο, η οποία δεν επηρεάζει και δεν εξαρτάται από τη γεωμετρική συμπεριφορά της. Η στρέψη είναι η δεύτερη καμπυλότητα ( $\kappa_2$ ) των συναρτήσεων καμπυλότητας. Επομένως σε κάθε διάσταση αναζητούμε παραμέτρους που περιγράφουν την τάση της καμπύλης να “αποκλίνει” από τους χώρους των προηγούμενων διαστάσεων.

Ανάλογα με τα παραπάνω, στις  $n$  διαστάσεις ορίζονται συνολικά  $n-1$  συναρτήσεις καμπυλότητας  $\kappa_i$  οι οποίες εκφράζουν την τάση της καμπύλης να “αποκλίνει” από τους χώρους με τις προηγούμενες  $n-1$  διαστάσεις, δηλαδή από την ευθεία γραμμή ( $\kappa_1$ ), από το διδιάστατο επίπεδο ( $\kappa_2$ ), από τον τρισδιάστατο χώρο ( $\kappa_3$ ), και γενικά η καμπυλότητα  $\kappa_i$  εκφράζει την τάση της καμπύλης να “αποκλίνει” από τον χώρο διάστασης  $i$ . Οι συναρτήσεις καμπυλότητας περιγράφουν πλήρως την γεωμετρική συμπεριφορά της καμπύλης μέσα στον  $n$ -διάστατο Ευκλείδειο χώρο.

Στο παρακάτω σχήμα (σχ. Π.1) φαίνονται γραφικά οι παρατηρήσεις για την σημασία της καμπυλότητας, ως μέσο περιγραφής της γεωμετρικής συμπεριφοράς της καμπύλης. Συγκεκριμένα παρατηρούμε ότι σε διδιάστατο κύκλο έχουμε σταθερή καμπυλότητα (ίση με το αντίστροφο της ακτίνας του κύκλου), ενώ σε τρισδιάστατη έλικα έχουμε σταθερή καμπυλότητα, αλλά μεταβαλλόμενη στρέψη, αν η κατακόρυφη ταχύτητα

$v$  δεν είναι σταθερή. Στην περίπτωση σταθερής ταχύτητας  $v$  έχουμε σταθερό αριθμό σπειρών ανά μονάδα ύψους και άρα σταθερή αλλαγή της μορφής, ενώ για μεταβαλλόμενη  $v$  έχουμε μεταβαλλόμενο αριθμό σπειρών, και η στρέψη είναι το μέτρο που μας δείχνει το σημείο όπου αραιώνουν περισσότερο οι σπείρες και άρα το σημείο όπου έχουμε μεγαλύτερη καμπύλωση (αλλαγή της γεωμετρικής μορφής με επιταχυνόμενο ρυθμό).



**Σχ. Π.1 :** Η σημασία των συναρτήσεων καμπυλότητας για την περιγραφή της γεωμετρικής μορφής μια καμπύλης





# *Βιβλιογραφία*

- [Avr99] Y. Avrithis, A. Doulamis, N. Doulamis and S. Kollias, “A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases”, *Computer Vision and Image Understanding*, vol. 75 (1/2), pp. 3-24, July 1999.
- [Bes02] J. Bescós, A. Movilla, J.M. Menéndez and G. Cisneros, “Real-time temporal segmentation of MPEG video”, in *Proc. ICIP’2002*, Rochester, NY, Sep. 2002.
- [Boc00] G. Boccignone, M. De Santo and G. Percannella, “An algorithm for video cut detection in MPEG sequences”, *Proc. of the IS&T/SPIE International Conference on Storage and Retrieval of Media Databases 2000*, pp. 523-530, San Jose, CA, Jan. 2000.
- [Bor96] J.S. Boreczky, L.A. Rowe, “Comparison of Video Shot Boundary Detection Techniques”, *Storage and Retrieval for Image and Video Databases IV*, *Proceedings SPIE Vol. 2670*, pp. 170-179, 1996.
- [Bro00] P. Browne, A. Smeaton, N. Murphy, N. O'Connor, S. Marlow and C. Berrut, “Evaluating and Combining Digital Video Shot Boundary Detection Algorithms”, in *Proceedings of IMVIP 2000*, Belfast, Northern Ireland, September 2000.
- [Cal01] J. Calic and E. Izquierdo, “Towards Real-Time Shot Detection in the MPEG Compressed Domain”, *Proceedings of WIAMIS'2001 - Workshop on Image Analysis for Multimedia Interactive Services*, Tampere, Finland, May 2001.
- [Cal02] J. Calic, S. Sav, E. Izquierdo, S. Marlow, N. Murphy and N.E. O'Connor, “Temporal Video Segmentation for Real-Time Key Frame Extraction”, *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002)*, Orlando, USA, 17-23 May, 2002.
- [Car76] Manfredo P. do Carmo, “*Differential Geometry of Curves and Surfaces*”, Prentice-Hall Inc., USA, Upper Saddle River, New Jersey, 1976.
- [DeM98] D. DeMenthon, V. Kobla and D. Doermann, “Video Summarization by Curve Simplification”, *Proc. 6th ACM Int. Multimedia Conf.*, Bristol, U.K., 1998.
- [DeS02] M. De Santo, G. Percannella, C. Sansone, R. Santoro, M. Vento, “Automatic Cut Detection in MPEG Movies: A Multi-expert Approach”, *Lecture Notes in Computer Science*, Volume 2396, Page 699, Jan 2002.

- [Div01] A. Divakaran, R. Radhakrishnan and K.A. Peker, "Video summarization using descriptors of motion activity: A motion activity based approach to key-frame extraction from video shots", *Journal of Electronic Imaging*, vol. 10, iss. 4, pp. 909–916, October 2001.
- [Dou00] A. Doulamis, N. Doulamis, G. Akrivas and S. Kollias, "Non-sequential video content representation using temporal variation of feature vectors", *IEEE Transactions on Consumer Electronics*, vol. 46, no. 3, August 2000.
- [Dou97] A. Doulamis, Y. Avrithis, N. Doulamis and S. Kollias, "Indexing and retrieval of the most characteristic frames/scenes in video databases", in *Proc. of Workshop on Image Analysis for Multimedia Interactive Systems (WIAMIS)*, pp. 105–110, Louvain-la-Neuve, Belgium, 1997.
- [Dou98] N. Doulamis, A. Doulamis, Y. Avrithis and S. Kollias, "Video content representation using optimal extraction of frames and scenes", in *Proc. of IEEE Int. Conference on Image Processing (ICIP'98)*, Chicago IL, USA, Oct. 1998.
- [DPe73] D.H. Douglas and T.K. Peucker, "Algorithms for the Reduction of the Number of Points required to represent a Line or its Caricature", *The Canadian Cartographer*, vol. 10 (iss. 2), pp. 112-122, 1973.
- [For00] R.M. Ford, C. Robson, D. Temple, M. Gerlach, "Metrics for shot boundary detection in digital video sequences", *Multimedia Systems*, Volume 8, Issue 1, pages 37 - 46, Jan 2000.
- [Glu66] Herman Gluck, "Higher Curvatures of Curves in Euclidean Space", *The American Mathematical Monthly*, Jstor, Vol. 73, No. 7, pp. 699-704, Aug. - Sep. 1966.
- [Glu67] Herman Gluck, "Higher Curvatures of Curves in Euclidean Space, II", *The American Mathematical Monthly*, Jstor, Vol. 74, No. 9, pp. 1049-1056, Nov. 1967.
- [Han99] A. Hanjalic and H. 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, pp. 1280–1289, Dec. 1999.
- [Her92] J. Hershberger and J. Snoeyink, "Speeding up the Douglas-Peucker Line-Simplification Algorithm", <http://www.cs.ubc.ca/cgi-bin/tr/1992/TR-92-07>.
- [ISO93] ISO/IEC 11172-2, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s – Video", Geneva, 1993.
- [Itt98] L. Itti, C. Koch and E. Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis", *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 20, no. 11, pp. 1254-1259, 1998.
- [LeG91] Didier Le Gall, "MPEG: a video compression standard for multimedia applications", *Source Communications of the ACM archive*, Volume 34, Issue 4, pp. 46-58, April 1991.

- [Lie99] R. Lienhart, "Comparison of automatic shot boundary detection algorithms", in Proc. of SPIE Storage and Retrieval for Image and Video Databases VII, San Jose, CA, U.S.A., volume 3656, pages 290–301, January 1999.
- [Men95] J. Meng, Y. Juan and S.F. Chang, "Scene Change Detection in a MPEG Compressed Video Sequence", in IS&T/SPIE Symp. Proc., vol. 2419, San Jose, 1995.
- [Mok02] F. Mokhtarian, S. Abbasi, "Shape similarity retrieval under affine transforms", Pattern Recognition, Volume 35, Number 1, pp. 31-41, January 2002.
- [Nap98] M. Naphade, R. Mehrotra, A. Ferman, J. Warnick, T. Huang and A. Tekalp, "A high performance algorithm for shot boundary detection using multiple cues", in Proc. IEEE Int. Conf. Image Processing, vol. 2, pp. 884-887, (Chicago, IL), 1998.
- [Ram72] U. Ramer, "An Iterative Procedure for the Polynomial Approximation of Plane Curves", Computer Graphics and Image Processing, vol. 1, pp. 244–256, 1972.
- [Rap05] K. Rapantzikos, Y. Avrithis and S. Kollias, "On the use of spatiotemporal visual attention for video classification", Proc. of Proc. of Int. Workshop on Very Low Bitrate Video Coding (VLBV '01), Sardinia, Italy, September 15-16, 2005.
- [Sik97] Thomas Sikora, "MPEG Digital Video Coding Standards", Signal Processing Magazine, IEEE, Vol. 14, No. 5, pp. 82-10, 1997.
- [Sme99] A.F. Smeaton, J. Gilvarry, G. Gormley, B. Tobin S. Marlow and N. Murphy, "An Evaluation of Alternative Techniques for Automatic Detection of Shot Boundaries in Digital Video", in Irish Machine Vision and Image Processing Conference, 1999.
- [Spi99] Michael Spivak, "A Comprehensive Introduction to Differential Geometry", Volume Two, Publish Or Perish Inc., USA, Texas, Houston, 1999.
- [Wan98a] J. Wang and R. Acharya, "A Vertex-based Shape Coding Approach for Similar Shape Retrieval", Proceedings of the 1998 ACM symposium on Applied Computing, pp. 520-524, Atlanta, Georgia, USA, February 1998.
- [Wan98b] J. Wang, W. Yang and R. Acharya, "Efficient Access to and Retrieval from a Shape Image Database", Proceedings of the IEEE Workshop on Content - Based Access of Image and Video Libraries (CBAIVL), pp. 63, 1998 .
- [Whi03] A. Whitehead, "Fast Feature-based Video Segmentation and Annotations", International Symposium on Signal Processing and Applications, Paris, France, July 2003.
- [Yeu97] M.M. Yeung and B. Yeo, "Video Visualization for Compact Presentation and Fast Browsing of Pictorial Content", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, No. 5, pp. 771-785, Oct. 1997.
- [Zab99] R. Zabih, J. Miller and K. Mai, "A Feature Based Algorithm for detecting and Classifying Production Effects", ACM Journal of Multimedia Systems, Vol 7, pp. 119-128, 1999.

- [Zha93] H. Zhang, A. Kankanhalli, S. Smoliar, “Automatic partitioning of full-motion video”, ACM/Springer Multimedia Systems, vol. 1, No. 1., pp. 10-28, 1993.
- [Zha95a] H. Zhang, C.Y. Low and S.W. Smoliar, “Video parsing and browsing using compressed data”, in Multimedia Tools and Applications, vol. 1, pp. 89–111, Norwell, 1995.
- [Zha95b] H. Zhang, C.Y. Low, S.W. Smoliar and J.H. Wu, “Video parsing, retrieval and browsing: An integrated and content-based solution”, in Proc. ACM Multimedia’95, pp. 15–24, San Francisco, CA, Nov. 5–9, 1995.
- [Zhu98] Y. Zhuang, Y. Rui, T. S. Huang and S. Mehrotra, “Adaptive key frame extraction using unsupervised clustering”, in Proc. IEEE Int. Conference on Image Processing (ICIP’98), vol. 1, pp. 866–870, Chicago IL, USA, Oct. 1998.
- [Κοκολ99] Γ. Κοκολάκης, Ι. Σπηλιώτης, “Εισαγωγή στη Θεωρία Πιθανοτήτων και Στατιστική”, Εκδόσεις Συμεών, Έκδ. 3η, Αθήνα, 1999.
- [Κολλ01] Στέφανος Κόλλιας, “Επεξεργασία, Ανάλυση και Τεχνολογία Εικόνων – Βίντεο”, ΕΜΠ, Αθήνα, Οκτώβριος 2001.
- [Χαιν98] Ι.Θ. Χαΐνη, “Μαθήματα Μαθηματικής Αναλύσεως”, Τόμος Β, Ελλάδα, Αθήνα 1998.
- [WWW] Το διαδίκτυο (internet – world wide web)
- [wwwDaliVM] Δικτυακός τόπος (Πανεπιστήμιο Cornell):  
<http://www.cs.cornell.edu/zeno/projects/dali/Default.html> και  
<http://www.cs.cornell.edu/dali/>
- [wwwOpenCV] Δικτυακός τόπος (Ερευνητική Ομάδα Intel):  
<http://www.intel.com/research/mrl/research/opencv/> και  
<http://www.sourceforge.net/projects/opencvlibrary>



