

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«Τεχνικές κατάτμησης και ανάλυσης μεταβολών σε
χρονοσειρές εικόνων (video)»**

Ιωάννης Σταθόπουλος

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

Ευριπίδης Πετράκης (Αν.Καθηγητής) Επιβλέπων

Μιχάλης Ζερβάκης (Καθηγητής)

Σταύρος Χριστοδουλάκης (Καθηγητής)

**ΧΑΝΙΑ
ΟΚΤΩΒΡΙΟΣ, 2004**

Περίληψη

Ο κύριος στόχος αυτής της διπλωματικής εργασίας είναι η ανίχνευση των βασικών μεταβολών που συμβαίνουν σε ένα βίντεο με σκοπό την τμηματοποίηση και κατάτμηση του σε στιγμιότυπα. Οι βασικές μεταβολές στο βίντεο χωρίζονται σε δυο κατηγορίες: απότομες μεταβολές (*abrupt transition*) και σταδιακές μεταβολές (*gradual transition*). Στην πρώτη κατηγορία εντάσσονται τα Cut και Flash ενώ στην δεύτερη τα Dissolve, Zoom, Pan και Tilt. Οι μεταβολές οι οποίες προκαλούν μετάβαση από ένα σύνολο όμοιων εικονοπλαισίων σε ένα διαφορετικό αποτελούν σημασιολογικά τα όρια με τα οποία τμηματοποιούνται οι χρονοσειρές εικόνων σε στιγμιότυπα (*shots*). Στην συγκεκριμένη μελέτη προτείνεται ένα αυτοματοποιημένο σύστημα που επιτυγχάνει ταυτόχρονα με την κατάτμηση του βίντεο σε στιγμιότυπα την ανίχνευση και τον χαρακτηρισμό μεταβολών μέσα σε κάθε στιγμιότυπο ανάλογα με τον τύπο της κίνησης που παρατηρείται (Zoom, Pan, Tilt).

Αρχική επιδίωξη της μελέτης αυτής είναι η ανάλυση και υλοποίηση “μεθόδων κατωφλίωσης” (*thresholding methods*) γνωστών από την βιβλιογραφία “Μέθοδος διπλής σύγκρισης” ([7] Twin Comparison) και “Μέθοδος διπλής σύγκρισης σε παράθυρο” ([20] Twin Comparison Window) αλλά και κάποιων που προτείνονται σε παλιότερες διπλωματικές εργασίες “Μέθοδος παραθύρου ολίσθησης” ([22] Sliding Window) και “Προσαρμοστική μέθοδο” ([22] Adaptive method). Με αυτόν τον τρόπο μελετάται το πρόβλημα από το πρωταρχικό στάδιο της τμηματοποίησης σήματος και εξετάζονται οι εγγενείς αδυναμίες της κάθε μεθόδου ξεχωριστά.

Στη συνέχεια από τις χρονοσειρές εικόνων εξάγεται πληροφορία για τις μεταβολές της έντασης διαδοχικών εικονοπλαισίων όπως και χαρακτηριστικά από τις μεταβολές των διανυσμάτων κίνησης. Τα δεδομένα αυτά αναπαριστώνται σαν μια συνεχή ροή τιμών (σήμα). Σε αυτά εφαρμόζονται οι μέθοδοι κατωφλίωσης και σε συνδυασμό με τα κριτήρια που επιβάλλει ο κάθε μέθοδος ανίχνευσης μεταβολών εξάγονται τα συμπεράσματα για τον χαρακτηρισμό των μεταβολών και εν συνεχεία την τμηματοποίηση του περιεχομένου του βίντεο. Η χρήση των μεθόδων έχει σκοπό την ανίχνευση των δυο βασικών κατηγοριών μεταβολών (απότομες και σταδιακές) αλλά και την προσπάθεια χαρακτηρισμού του είδους αυτών (Cut-Flash, Dissolve-Zoom-Pan-Tilt). Ο συνδυασμός των παραπάνω μεθόδων καταφέρνει να εξαλείψει

μεταβολές όπως (flash,zoom,pan,tilt) οι οποίες δεν αποτελούν όριο στιγμιότυπου (*Shot boundary*) και συνεπώς δεν είναι ορθό να συνυπολογίζονται στα τελικά αποτελέσματα. Στο τελευταίο κομμάτι της εργασίας μετρήθηκε η απόδοση των μεθόδων κατωφλιώσης οι οποίες εφαρμόστηκαν για τον υπολογισμού κατωφλίου στις μεθόδους ανίχνευσης μεταβολών βίντεο χρησιμοποιώντας σαν μέτρο τις ποσότητες **precision** και **recall**. Οι παραπάνω διαδικασίες εφαρμόστηκαν σε μια συλλογή από διαφορετικές κατηγορίες βίντεο (αθλητικά , ντοκιμαντέρ, ειδήσεις, animation) των οποίων η αρχική ανάλυση και κατάτμηση έγινε από εξωτερικό παρατηρητή, παρέχοντας έτσι την πραγματική πληροφορία ανάλυσης και κατάτμησης (*ground truth*) για τα δεδομένα. Η μέτρηση της απόδοσης των μεθόδων που εξετάστηκαν έγινε συγκρίνοντας τα αποτελέσματα των μεθόδων με τις μετρήσεις του εξωτερικού παρατηρητή.

Κεφαλαίο 1	6
Εισαγωγή - Σχετικές Εργασίες	6
1.1 Εισαγωγή	6
1.2 Σχετικές εργασίες	9
1.2.1) Συμπιεσμένο ή ασυμπιεστο βίντεο	9
α) Μέθοδοι βασισμένες σε ιστογράμματα έντασης	9
β) Μέθοδοι βασισμένες σε συγκρίσεις μεταξύ εικονοστοιχείων (pixels) ..	12
γ) Μέθοδοι βασισμένες σε συγκρίσεις μεταξύ περιοχών (blocks) εικονοστοιχείων	14
δ) Μέθοδοι βασισμένες σε μεταβολές ακμών	17
ε) Μέθοδοι βασισμένες σε πληροφορία κίνησης	18
Κεφαλαίο 2	20
Συμπύεση Βίντεο	20
2.1 Motion Pictures Expert Group (MPEG)	20
2.2 Δομή ροής δεδομένων του βίντεο σε MPEG 1	25
2.3 Βασική πηγή πληροφορίας (Source Input Format)	27
2.4 Είδη εικονοπλαισίων	31
2.5 Είδη macroblocks που εμφανίζονται σε κάθε τύπο εικονοπλαισίου	33
2.6 Ενδο-εικονοπλαισιακή κωδικοποίηση (Intraframe coding) με χρήση του Διακριτού Συνημιτονικού Μετασχηματισμού (DCT : Discrete Cosine Transform) ..	35
2.7 Δια-εικονοπλαισιακή κωδικοποίηση (Interframe coding) με την χρήση υπολογισμού της κίνησης (Motion Estimation)	37
2.8 Αποκωδικοποίηση των εικονοπλαισίων σε μια σειρά DC ελαττωμένες εικόνες	41
2.8.1 Επακριβής εξαγωγή των DC συντελεστών	43
2.8.2 Προσεγγιστική εξαγωγή των DC συντελεστών	46
2.9 Αντιστάθμιση και κανονικοποίηση της κίνησης “Motion compensation and motion vector normalization”	51
2.9.1 Υπολογισμός κανονικοποιημένων διανυσμάτων μεταξύ I και P εικονοπλαισίων	53
2.9.2 Υπολογισμός κανονικοποιημένων διανυσμάτων μεταξύ B εικονοπλαισίων	54
Κεφάλαιο 3	63
Μελέτη Μεθόδων Κατωφλίωσης	63
3.1 Εισαγωγή στην μελέτη μεθόδων κατωφλίωσης	63
3.2 Μέθοδος Διπλής Σύγκρισης - Twin Comparison method	66
3.3 Μέθοδος Διπλής Σύγκρισης σε παράθυρο - Twin Comparison Window method	71
3.4 Μέθοδος Παραθύρου Ολίσθησης - Sliding Window method	76
3.5 Προσαρμοστική Μέθοδος - Adaptive method	85
Κεφάλαιο 4	92
Κατάτμηση του βίντεο και χαρακτηρισμός της κίνησης	92
4.1 Η βασική ιδέα κατάτμησης και ανάλυση των προβλημάτων εφαρμογής της ..	92
4.2 Μέθοδος ανίχνευσης φλας (flash model)	94
4.3 Πληροφορία κίνησης για ανίχνευση zoom, pan και tilt	98
4.3.1 Ιστόγραμμα γωνιών διανυσμάτων κίνησης	99
4.3.2 Μέθοδοι ανίχνευσης zoom, pan, tilt	103
4.4 Ανίχνευση απότομων και σταδιακών μεταβολών	106
4.4.1 Μέθοδος διαφοράς Ιστογραμμάτων τιμών έντασης	106
4.4.2 Μέθοδος διαφορών των Intracoded macroblocks	110

4.5 Αλγόριθμος ανίχνευσης ορίων στιγμιότυπου.....	112
Κεφάλαιο5	114
Πειραματικά αποτελέσματα- Συμπεράσματα	114
5.1 Πειραματικά αποτελέσματα	114
5.2 Αποτελέσματα από την Μέθοδο 1, ανίχνευσης φλας.....	116
5.3 Αποτελέσματα από την Μέθοδο 2, ανίχνευσης zoom, pan και tilt.....	117
5.4 Αποτελέσματα από την Μέθοδο 3, ανίχνευσης απότομων και σταδιακών μεταβολών με την χρήση διαφορών ιστογραμμάτων έντασης.	119
5.5 Αποτελέσματα από την Μέθοδο 4, ανίχνευσης απότομων και σταδιακών μεταβολών με την χρήση intracoded macroblocks	121
5.6 Αποτελέσματα από τον Αλγόριθμο 1, υπολογισμό ορίων στιγμιότυπου	123
5.7 Συμπεράσματα –Μελλοντική εργασία.....	125
ΑΝΑΦΟΡΕΣ.....	127

Κεφαλαίο 1

Εισαγωγή - Σχετικές Εργασίες

1.1 Εισαγωγή

Η ραγδαία ανάπτυξη της τηλεπικοινωνιακής τεχνολογίας είχε σαν αποτέλεσμα την εμφάνιση υπέρογκων ποσοτήτων πολυμεσικών δεδομένων σε ψηφιακά πληροφοριακά συστήματα ευρείας χρήσης όπως “VOD” (*video on demand*), ψηφιακή βιβλιοθήκη (*digital library*), μηχανές αναζήτησης πολυμεσικής πληροφορίας κλπ. Καθώς η ποσότητα και η πολυπλοκότητα της πληροφορίας από βίντεο μεγαλώνουν, απαιτούνται ευφυέστερες τεχνικές διαχείρισης της. Για παράδειγμα η διαχείριση και η αποθήκευση του βίντεο (*video storage and management*), η δεικτοδότηση (*video indexing*), η σύνοψη (*video abstraction*), η κατάτμηση του βίντεο (*video segmentation*) με τη χρήση ανίχνευσης αλλαγών σκηνής είναι οι βασικές διεργασίες οι οποίες απαιτείται να υλοποιηθούν για να οργανωθεί η πληροφορία των δεδομένων βίντεο. Οι διαδικασίες αυτές οργανώνουν την τεράστια σε όγκο πληροφορία που περιέχει το βίντεο και με πρακτικό τρόπο μπορεί να αναζητηθεί και να μεταδοθεί (Πχ : μέσω του διαδικτύου που είναι και το βασικότερο μέσο μετάδοσης πληροφοριών) .

Μια ακολουθία βίντεο είναι μια χρονικά εξελισσόμενη σκηνή, όπου το περιεχόμενο της αλλάζει σύμφωνα με την κίνηση κάποιου αντικείμενου, την κίνηση της κάμερας (*camera motion*), τα κοψίματα (*cuts*) και τα ειδικά εφε (*special effect*). Η διαδικασία της κατάτμησης του βίντεο (*video segmentation*) η οποία είναι το πρώτο βήμα στην ανάλυση και ανάκτηση του βίντεο (*content-based analysis and retrieval*) αναφέρεται στην κατάτμηση του εισερχόμενου βίντεο σε χρονικά κομμάτια με κοινά χαρακτηριστικά. Η προσπάθεια διαχωρισμού του βίντεο από κάποιο εξωτερικό παρατηρητή είναι μάλλον μια μη αποδοτική και ανεπαρκής διαδικασία αφού υπάρχουν τεράστια ποσά πληροφορίας που πρέπει να αναλυθούν. Συνεπώς απαιτείται αυτόματη κατάτμηση που επιτυγχάνεται με την ανίχνευση των μεταβολών μεταξύ στιγμιότυπων βίντεο (**shots** = μια σειρά από όμοια συνεχόμενα εικονοπλαίσια

που προέρχονται από μια απλή λειτουργία της κάμερας και αποτελούν το αρχικό σύνολο ομαδοποίησης του βίντεο). Οι μεταβολές αυτές μπορεί να είναι απότομες ή ομαλές. Απότομες μεταβολές χαρακτηρίζονται τα Cuts και Flash ενώ ομαλές μεταβολές τα Dissolves, Zoom, Pan και Tilt . Από τις απότομες μεταβολές το Cut χρησιμοποιείται για να διαχωρίσει τα στιγμιότυπα, να ορίσει δηλαδή το τέλος του τρέχοντος και την έναρξη ενός νέου στιγμιότυπου, ενώ από τις ομαλές μεταβολές το Dissolve χρησιμοποιείται για να γεφυρώσει γεγονότα που συμβαίνουν σε ξεχωριστό χρόνο και μέρος. Επομένως αποτελεί και αυτό σημείο αλλαγής στιγμιότυπου.

Κατά τη διαδικασία της αυτόματης ανίχνευσης των μεταβολών σε ένα βίντεο προκύπτουν διαφόρων ειδών προβλήματα. Έτσι ενώ είναι σχετικά εύκολο να ανιχνευτούν οι απότομες αλλαγές (*abrupt transition*), σημαντικό πρόβλημα προκύπτει στην προσπάθεια ανίχνευσης σταδιακών αλλαγών (*gradual transition*) και της κίνησης της κάμερας (*camera motion*), αφού πολλές φορές υπάρχει επικάλυψη των αλλαγών αυτών, ενώ άλλες φορές παρόμοια σήματα προκύπτουν από διαφορετικές καταστάσεις και είναι δύσκολο να διαχωριστούν.

Αρχικά θα παρουσιαστούν τεχνικές που έχουν υλοποιηθεί σε προηγούμενες μελέτες κατάτμησης βίντεο σε στιγμιότυπα και οι οποίες βασίζονται στην ανίχνευση μεταβολών του βίντεο. Στην συνέχεια στο Κεφάλαιο 2 θα αναφερθούν τα πρότυπα MPEG και θα αναλυθούν οι βασικές λειτουργίες συμπίεσης που εκτελούνται σε MPEG 1 και MPEG 2 (DC approximation, Motion Compensation, Motion Estimation).

Στο Κεφάλαιο 3 ξεκινάει το πρακτικό μέρος της εργασίας στο οποίο παρουσιάζονται οι δυο γνωστές από την βιβλιογραφία μέθοδοι κατωφλίωσης **“Μέθοδος Διπλής Σύγκρισης”** ([7] Twin Comparison) και **“Μέθοδος Διπλής Σύγκρισης σε Παράθυρο”** ([20] Twin Comparison Window) όπως επίσης κάποιες προτεινόμενες από παλιότερες διπλωματικές εργασίες μέθοδοι **“Μέθοδος Παραθύρου Ολίσθησης”** ([22] Sliding Window), **“Προσαρμοστική Μέθοδος”** ([22] Adaptive method). Πέραν από την υλοποίηση αυτών των μεθόδων παρουσιάζονται γραφικά οι αποκρίσεις τους σε διαφορετικά σήματα τονίζοντας τις αδυναμίες και τις ιδιαιτερότητες τους στην διαδικασία υπολογισμού κατωφλίου.

Στο επόμενο κεφάλαιο, Κεφάλαιο 4, καθορίζονται ποιες ακριβώς μεταβολές ορίζουν αλλαγή στιγμιότυπου και ποιες μεταβολές (false positives: flash, zoom, pan και tilt) μειώνουν την απόδοση των αποτελεσμάτων της κατάτμησης. Στη συνέχεια περιγράφεται και υλοποιείται η μέθοδος ανίχνευσης flash όπως αναλύεται στην

εργασία του [20] Zhang. Η επόμενη μέθοδος είναι η μέθοδος ανίχνευσης zoom,pan και tilt με την χρήση της διασποράς των γωνιών των διανυσμάτων κίνησης σε κάθε εικονοπλαίσιο. Μετέπειτα αναλύονται οι μέθοδοι ανιχνεύσεις απότομων και σταδιακών μεταβολών με την χρήση διαφοράς ιστογραμμάτων έντασης μεταξύ διαδοχικών εικονοπλαισίων όπως επίσης και με την διαφορά του πλήθους των Intracoded macroblocks μεταξύ διαδοχικών εικονοπλαισίων. Τελειώνοντας παρουσιάζεται ένας αλγόριθμος ο οποίος συνδυάζει όλες τις μεθόδους για να εξάγει τα όρια αλλαγής στιγμιότυπου που καθορίζονται μονάχα από τις δυο μεταβολές ,τα Cuts και τα Dissolves.

Στο Κεφάλαιο 5, παραθέτονται οι πειραματικά αποτελέσματα. Οι ποσότητες οι οποίες αντιπροσωπεύουν την απόδοση της κάθε μεθόδου είναι οι precision και recall. Με βάση τις πειραματικές αυτές μετρήσεις εξάχθηκαν χρήσιμα συμπεράσματα για την λειτουργία των μεθόδων τα οποία μπορούν να αποτελέσουν τους βασικούς άξονες μελλοντικής εργασίας στην ανάλυση και τμηματοποίηση βίντεο.

1.2 Σχετικές εργασίες

1.2.1) Συμπιεσμένο ή ασυμπιεστο βίντεο

α) Μέθοδοι βασισμένες σε ιστογράμματα έντασης.

Αρκετές μέθοδοι έχουν προταθεί όσον αφορά την χρήση ιστογράμματος (**histogram-based methods**) για ανίχνευση μεταβολών σε βίντεο. Μια πρώτη προσέγγιση γίνεται στο άρθρο του **Tonomura [1]** ο οποίος προτείνει μια μέθοδο που βασίζεται στον υπολογισμό της απόστασης μεταξύ ιστογραμμάτων (gray-level histogram). Η ανίχνευση του ορίου ενός στιγμιότυπου προκύπτει από την ανισότητα

$$\left(\sum_{v=0}^V |H(I_t, v) - H(I_{t-1}, v)| \right) > T \quad (1.3.1)$$

όπου $H(I_t, v)$ είναι η τιμή v του ιστογράμματος του I_t εικονοπλαισίου και V ο αριθμός των τιμών αναπαράστασης του χρώματος στο ιστόγραμμα. Αν το άθροισμα των διαφορών των τιμών του ιστογράμματος ξεπερνά το κατώφλι T δηλώνει ότι υπάρχει σημαντική αλλαγή στο περιεχόμενο του βίντεο και επομένως μπορεί να θεωρηθεί σαν όριο ενός νέου στιγμιότυπου

Κάτι παρόμοιο παρουσιάζεται και στο άρθρο του **Nagasaka [2]** οποίος προτείνει την ίδια μέθοδο με την παραπάνω αλλά χρησιμοποιώντας μόνο 64 τιμές για την αναπαράσταση των χρωμάτων μια για κάθε 4 τιμές από το αντίστοιχη κλίμακα 0-255.

$$\left(\sum_{v=0}^{63} |H_{64}(I_t, v) - H_{64}(I_{t-1}, v)| \right) > T \quad (1.3.2)$$

Η διαφορά ιστογραμμάτων όπως περιγράφεται στους τύπους (1.3.1 και 1.3.2) έχει εφαρμογή πέραν από την ένταση στο RGB και gray-level χώρο χρωμάτων και στα (HSV, YIQ, L*a*b*, L*u*v*, και Munsell) όπου σαν ποσότητες δεν χρησιμοποιούνται οι τιμές της έντασης αλλά Χροιά-Hue και Κορεσμός-Saturation στο HSV, I και Q στο YIQ, a* και b* στο L*a*b*, u* και v* στο L*u*v* και Χροιά-hue και Χρώμα-chroma στο Munsell χρωματικό χώρο. Ο **Gargi** [3] παρουσιάζει στο άρθρο του την διάφορα ιστογραμμάτων στους παραπάνω χρωματικούς χώρους χρησιμοποιώντας δυο χαρακτηριστικά στοιχεία χρώματος όπως φαίνεται και στην παρακάτω ανισότητα .

$$\left(\sum_{k=1}^2 \sum_{v=0}^V |H(I_t, C_k, v) - H(I_{t-1}, C_k, v)| \right) > T \quad (1.3.3)$$

Στην εργασία του ο **Pye** [4] υπολογίζει τρεις διαφορές ιστογραμμάτων, θεωρώντας ξεχωριστά τα τρία χρώματα του RGB χρωματικού χώρου. Η υψηλότερη τιμή από τις τρεις συγκρίνεται με ένα κατώφλι για την ανίχνευση μεταβολής του βίντεο

$$\left(\max_{k \in \{R, G, B\}} \sum_{v=0}^V |H(I_t, C_k, v) - H(I_{t-1}, C_k, v)| \right) > T \quad (1.3.4)$$

Ο **Toole** [5] χρησιμοποιεί το κριτήριο της ομοιότητας συνημίτονου μεταξύ δυο ιστογραμμάτων. Αρχικά τρία ιστογράμματα 64 τιμών αντιπροσωπεύουν τα Y, U και V στοιχεία από κάθε frame. Στη συνέχεια τα τρία ιστογράμματα συγκεντρώνονται σε ένα ιστόγραμμα 192 τιμών για κάθε εικονοπλαίσιο. Τέλος αυτά τα ιστογράμματα συγκρίνονται χρησιμοποιώντας το κριτήριο της ομοιότητας συνημίτονου σύμφωνα με τον παρακάτω τύπο

$$\left(1 - \frac{\sum_{v=0}^V (H_{YUV}(I_t, v) H_{YUV}(I_{t-1}, v))}{\sum_{v=0}^V H_{YUV}(I_t, v)^2 \sum_{v=0}^V H_{YUV}(I_{t-1}, v)^2} \right) > T \quad (1.3.5)$$

ενώ σχεδόν παρόμοια μελέτη παρουσιάζεται και στο άρθρο του **Cabedo** [6].

Μια από τις πιο γνωστές αναφορές στο χώρο της ανάλυσης και κατάτμησης βίντεο είναι η μελέτη του **Zhang** [7] ο οποίος προτείνει την μέθοδο της Διπλής Σύγκρισης (Twin Comparison) η οποία περιγράφεται αναλυτικά στο Κεφαλαίο 3.2, αφού αποτέλεσε και μια από τις μεθόδους που υλοποιήθηκαν και μελετήθηκαν σε αυτή τη διπλωματική εργασία.

Επίσης ενδιαφέρουσα αναφορά είναι του **Zhao** [8] στην οποία προτείνεται η χρήση κάποιων βαρών στην σύγκριση ιστογραμμάτων για να τονιστούν κάποια χαρακτηριστικά που είναι περισσότερο σημαντικά. Αρχικά υπολογίζει την κανονική διαφορά ιστογραμμάτων όπως περιγράφηκε και στις παραπάνω μελέτες. Στην συνέχεια χρησιμοποιείται κάποια συνάρτηση υπολογισμού βέλτιστων βαρών και τέλος υπολογίζεται η νέα διαφορά των ιστογραμμάτων με βάρη από τον τύπο

$$\left(\sum_{k=1}^3 \sum_{v=0}^V w(k, v) |H(I_t, C_k, v) - H(I_{t-1}, C_k, v)| \right) > T \quad (1.3.6)$$

όπου $w(k, v)$ αντιπροσωπεύει τα βέλτιστα βάρη .

Ένας διαφορετικός τρόπος σύγκρισης ιστογραμμάτων είναι ο υπολογισμός της τομής ιστογραμμάτων η οποία υπολογίζεται με διάφορους τελεστές όπως για παράδειγμα η min function

$$\left(1 - \frac{1}{XY} \sum_{v=0}^V \min (H(I_t, v), H(I_{t-1}, v)) \right) > T \quad (1.3.7)$$

Στο άρθρο του **Gargi** [31] χρησιμοποιείται ο υπολογισμός της τομής των ιστογραμμάτων ο οποίος ορίζεται λίγο διαφορετικά σύμφωνα με την παρακάτω σχέση

$$\left(1 - \frac{1}{XY} \sum_{v=0}^V \frac{\min (H(I_t, v), H(I_{t-1}, v))}{\max (H(I_t, v), H(I_{t-1}, v))} \right) > T \quad (1.3.8)$$

Πέραν όμως από την επικάλυψη ιστογραμμάτων στο άρθρο του **Dalianas** [9] παρουσιάζεται η σύγκριση ιστογραμμάτων βασισμένη στο χ^2 τεστ σύμφωνα με τον παρακάτω τύπο

$$\left(\sum_{v=0}^V \frac{(H(I_t, v) - H(I_{t-1}, v))^2}{\max(H(I_t, v), H(I_{t-1}, v))} \right) > T \quad (1.3.9)$$

β) Μέθοδοι βασισμένες σε συγκρίσεις μεταξύ εικονοστοιχείων (pixels) .

Στην συνέχεια θα γίνει αναφορά σε μεθόδους βασισμένες σε πράξεις και συγκρίσεις μεταξύ εικονοστοιχείων (**pixel -based methods**). Μια από τις πρώτες αναφορές στη βιβλιογραφία σε τέτοιου είδους μεθόδους είναι από το **Nagasaka [2]** όπου η ανίχνευση αλλαγής στιγμιότυπου υπολογίζεται μέσω του αθροίσματος των διαφορών έντασης δυο εικονοστοιχείων που βρίσκονται στο ίδιο σημείο δυο διαδοχικών εικονοπλαισίων.

$$\left(\left| \sum_{i=1}^X \sum_{j=1}^Y P(I_t, i, j) - \sum_{i=1}^X \sum_{j=1}^Y P(I_{t-1}, i, j) \right| \right) > T \quad (1.3.10)$$

Με παρόμοιο τρόπο ο **Zhang [7]** περιγράφει την ίδια μέθοδο χρησιμοποιώντας Boolean λογική .

$$D(I_t, I_{t-1}, i, j) = \begin{cases} 1 & \text{if } P(I_t, i, j) \neq P(I_{t-1}, i, j) \\ 0 & \text{otherwise} \end{cases} \quad (1.3.11)$$

και τέλος το ποσό των διαφορετικών εικονοστοιχείων υπολογίζεται και συγκρίνεται με ένα κατώφλι από όπου εξάγεται και το αν υπάρχει αλλαγή στιγμιότυπου

$$\left(\sum_{i=1}^X \sum_{j=1}^Y D(I_t, I_{t-1}, i, j) \right) > T \quad (1.3.12)$$

Οι μέθοδοι αυτοί που βασίζονται σε πράξεις και συγκρίσεις εικονοστοιχείων επηρεάζονται πολύ από τον θόρυβο λόγω ψηφιοποίησης και λόγω της κίνησης της

κάμερας, για αυτό συνήθως προτείνεται να γίνει μια εξομάλυνση (smoothing) των εικόνων με κάποιο φίλτρο 3×3 πριν να υπολογιστούν οι ποσότητες D .

Η κατηγορία των pixel-based μεθόδων ολοκληρώνεται με την ανάλυση της απόκλισης των εντάσεων των εικονοστοιχείων μέσα στο χρόνο. Στο άρθρο του **Taniguchi** [10] περιγράφεται μια μέθοδος στην οποία κάθε εικονοστοιχείο αποκτά μια ετικέτα η οποία και το χαρακτηρίζει, τέτοιες ετικέτες μπορεί να παίρνουν τη μορφή “constant”, “step(I_t)”, “linear(I_{t1} , I_{t2})” και “no label”. Αυτές οι ετικέτες συμβολίζουν αντίστοιχα εικονοστοιχεία με σταθερή τιμή, εικονοστοιχεία των οποίων οι τιμές αλλάζουν μέσα στο I_t εικονοπλαίσιο, εικονοστοιχεία που αλλάζουν προοδευτικά τις τιμές τους μεταξύ των εικονοπλαισίων I_{t1} και I_{t2} και τέλος εικονοστοιχεία με τυχαίες τιμές έντασης.

Έτσι δυο Boolean συνθήκες $\Theta_1(I_{t1}, I_{t2}, i, j)$ και $\Theta_2(I_{t1}, I_{t2}, i, j)$ εισάγονται με σκοπό να ορίσουν την σταθερότητα του συνόλου των τιμών των εικονοστοιχείων $P(I_t, i, j)$ με $t_1 < t < t_2$

και ορίζονται ως εξής :

$$\begin{aligned} \Theta_1(I_{t1}, I_{t2}, i, j) &= \begin{cases} \text{true} & \text{if } \left(\max_{t_1 \leq t \leq t_2} P(I_t, i, j) - \min_{t_1 \leq t \leq t_2} P(I_t, i, j) \right) < T \\ \text{false} & \text{otherwise} \end{cases} \\ \Theta_2(I_{t1}, I_{t2}, i, j) &= \begin{cases} \text{true} & \text{if } \left(\begin{aligned} &\max_{t_1 \leq t \leq t_2} (P(I_t, i, j) + (t - t_1)\theta_{t1, t2}) \\ &- \min_{t_1 \leq t \leq t_2} (P(I_t, i, j) + (t - t_1)\theta_{t1, t2}) \end{aligned} \right) < T \\ \text{false} & \text{otherwise} \end{cases} \end{aligned} \quad (1.3.13)$$

με $\theta_{t1, t2}$

$$\theta_{t1, t2} = \frac{|P(I_{t1}, i, j) - P(I_{t2}, i, j)|}{t_2 - t_1} \quad (1.3.14)$$

όπου $t_2 - t_1$ είναι σταθερή τιμή και όχι μικρότερη από το χρονικό διάστημα μεταξύ δυο εικονοπλαισίων. Αυτές τις συνθήκες ομοιότητας Θ_1 και Θ_2 αφού υπολογιστούν χρησιμοποιούνται για να καθορίσουν τις ετικέτες label $L(I_{t0}, I_{tf}, i, j)$ που παίρνουν τιμές σύμφωνα με το παρακάτω τύπο

$$L(I_{t_0}, I_{t_f}, i, j) = \begin{cases} \text{constant} & \text{if } \Theta_1(I_{t_0}, I_{t_f}, i, j) \\ \text{step}(I_t) & \text{if } \left(\begin{array}{l} \Theta_1(I_{t_0}, I_{t-1}, i, j) \\ \wedge \neg \Theta_1(I_{t-1}, I_t, i, j) \end{array} \right) \\ \text{linear}(I_{t_1}, I_{t_2}) & \text{if } \left(\begin{array}{l} \Theta_1(I_{t_0}, I_{t_1}, i, j) \\ \wedge \neg \Theta_1(I_{t_1}, I_{t_2}, i, j) \\ \wedge \Theta_2(I_{t_1}, I_{t_2}, i, j) \end{array} \right) \\ \text{no label} & \text{otherwise} \end{cases} \quad (1.3.15)$$

αφού κάθε εικονοστοιχείο χαρακτηριστεί από μια από τις παραπάνω ετικέτες τότε ανιχνεύεται απότομη μεταβολή (**cut**) στο I_t εικονοπλάνο όταν ικανοποιείται η παρακάτω ανισότητα

$$\left(\frac{\sum_{i=1}^X \sum_{j=1}^Y L_{\text{step}(I_t)}(I_{t_0}, I_{t_f}, i, j)}{XY - \sum_{i=1}^X \sum_{j=1}^Y L_{\text{no label}}(I_{t_0}, I_{t_f}, i, j)} \right) > T \quad (1.3.16)$$

Ενώ αντίστοιχα σταδιακή μεταβολή (**dissolve**) μεταξύ I_{t1} και I_{t2} εικονοπλαισίων ανιχνεύεται αν ισχύει η παρακάτω ανισότητα

$$\left(\frac{\sum_{i=1}^X \sum_{j=1}^Y L_{\text{linear}(I_{t_1}, I_{t_2})}(I_{t_0}, I_{t_f}, i, j)}{XY - \sum_{i=1}^X \sum_{j=1}^Y L_{\text{no label}}(I_{t_0}, I_{t_f}, i, j)} \right) > T \quad (1.3.17)$$

γ) Μέθοδοι βασισμένες σε συγκρίσεις μεταξύ περιοχών (blocks) εικονοστοιχείων.

Φεύγοντας από αυτή την κατηγορία μεθόδων πολλές αναφορές έχουν γίνει και σε επίπεδο περιοχών της εικόνας (**block based methods**). Είναι ένας ενδιαμέσος τρόπος επεξεργασίας μεταξύ των **pixel based** και **histogram based** μεθόδων και έχει σκοπό να αυξήσει την ποιότητα των μεθόδων ανίχνευσης στιγμιότυπων αλλά συγχρόνως να μειώσει και τον υπολογιστικό χρόνο. Το πλεονέκτημα της χρήσης είναι ότι μπορούν οι ήδη υπάρχοντες αλγόριθμοι να εφαρμοστούν στα τμήματα (blocks) όπως ακριβώς στα εικονοστοιχεία και τα ιστογράμματα και επίσης ότι τα blocks σαν

πληροφορία δεν είναι τόσο ευαίσθητη στον θόρυβο και στην κίνηση της κάμερας και των αντικειμένων.

Στην μελέτη που παρουσιάζει ο **Kasturi [11]** προτείνει τον υπολογισμό της ομοιότητας των blocks δειγμάτων της εικόνας. Όπως ακριβώς στις pixel based μεθόδους, δυο διαδοχικά εικονοπλαίσια συγκρίνονται. Η ομοιότητα βασίζεται σε χαρακτηριστικά των blocks όπως η μέση τιμή (mean) και η διασπορά (variance). Έτσι η τιμή της πιθανότητας όπως ορίζεται για το block b είναι

$$L(I_t, I_{t-1}, b) = \frac{\left(\frac{\sigma_{t,b}^2 + \sigma_{t-1,b}^2}{2} + \left(\frac{\mu_{t,b} - \mu_{t-1,b}}{2} \right)^2 \right)^2}{\sigma_{t,b}^2 \sigma_{t-1,b}^2} \quad (1.3.18)$$

όπου $\mu_{t,b}$ και $\sigma_{t,b}$ είναι η μέση τιμή και η απόκλιση του block b της εικόνας I_t στη συνέχεια η τιμή αυτή συγκρίνεται με ένα κατώφλι

$$L_D(I_t, I_{t-1}, b) = \begin{cases} 1 & \text{if } L(I_t, I_{t-1}, b) > T_D \\ 0 & \text{in other cases} \end{cases} \quad (1.3.19)$$

όπου T_D ορίζεται σαν τιμή ανοχής (tolerance value). Η ανίχνευση τώρα της αλλαγής στιγμιότυπου γίνεται με τον έλεγχο της παρακάτω ανισότητας

$$\sum_{b=1}^B c_b L_D(I_t, I_{t-1}, b) > T \quad (1.3.20)$$

όπου c_b χρησιμοποιείται για να δώσει μικρότερη ή μεγαλύτερη βαρύτητα στο block αν και τις περισσότερες φορές παίρνει την σταθερή τιμή 1 για όλα τα blocks

Ο **Lee [12]** στην εργασία του υπολογίζει διαφορές των blocks αλλά στο HSV χώρο χρωμάτων. Πρώτα οι RGB εικόνες μετατρέπονται σε HSV και υπολογίζονται οι τιμές της Χροιάς - Hue και Κορεσμού - Saturation για κάθε block. Έτσι δυο blocks συγκρίνονται χρησιμοποιώντας τις παρακάτω συναρτήσεις διαφορών

$$\begin{aligned}
D_H(I_{t_1}, I_{t_2}, b) &= |\mu(I_{t_1}, b, C_H) - \mu(I_{t_2}, b, C_H)| \\
D_S(I_{t_1}, I_{t_2}, b) &= |\mu(I_{t_1}, b, C_S) - \mu(I_{t_2}, b, C_S)|
\end{aligned}
\tag{1.3.21}$$

όπου $\mu(I_t, b, C_k)$ είναι η μέση τιμή του block b στο εικονοπλαίσιο I_t . Στη συνέχεια οι διαφορές αυτές συγκρίνονται με κάποια κατώφλι τα οποία καθορίζουν και κατά πόσο κάθε block αλλάζει υπολογίζοντας και τον αριθμό των blocks που άλλαξαν.

$$D(I_{t_1}, I_{t_2}, b) = \begin{cases} 1 & \text{if } (D_H(I_{t_1}, I_{t_2}, b) > T_H) \vee (D_S(I_{t_1}, I_{t_2}, b) > T_S) \\ 0 & \text{otherwise} \end{cases}
\tag{1.3.22}$$

Τέλος η αναλογία του αριθμού των blocks τα οποία έχουν αλλάξει προς τα συνολικά blocks συγκρίνεται με κάποιο άλλο κατώφλι το οποίο καθορίζει και τον αν υπάρχει αλλαγή στιγμιότυπου .

$$\frac{1}{B} \sum_{b=1}^B D(I_t, I_{t-\Delta}, b) > T
\tag{1.3.23}$$

όπου το Δ δηλώνει ανά ποσά εικονοπλάγια υπολογίζεται η διαφορά των blocks.

Μια εντελώς διαφορετική προσέγγιση στο πρόβλημα της ανίχνευσης αλλαγών στιγμιότυπων περιγράφεται στο άρθρο του ο **Xiong [13]** ο οποίος προτείνει την επεξεργασία όχι του συνολικού αριθμού των δειγμάτων (εικονοστοιχεία ή blocks) στο χώρο και στο χρόνο. Έτσι εξετάζεται η ανισότητα των διαφορών σε ένα υποσύνολο δειγμάτων όπως και φαίνεται από την παρακάτω σχέση

$$\left(\sum_{b \in B'} D_\mu(I_t, I_{t+\Delta}, b) \right) > T
\tag{1.3.25}$$

όπου το B' αντιπροσωπεύει ένα υποσύνολο δειγμάτων ενώ το $D_\mu(I_{t_1}, I_{t_2}, b)$ ορίζεται ως εξής

$$D_\mu(I_{t_1}, I_{t_2}, b) = \begin{cases} 1 & \text{if } |\mu_{t_1, b} - \mu_{t_2, b}| > T_\mu \\ 0 & \text{in other cases} \end{cases}
\tag{1.3.26}$$

με $\mu_{t_1, b}$ η μέση τιμή του block b την χρονική στιγμή t_1 .

δ) Μέθοδοι βασισμένες σε μεταβολές ακμών

Άλλη μια προσπάθεια επίλυσης του προβλήματος δίνεται από τον **Zabih [14]** ο οποίος χρησιμοποιεί τις αλλαγές ακμών για να ορίσει και τα όρια των στιγμιότυπων. Οι ακμές εξάγονται βάση του αλγορίθμου Canny και κανονικοποιούνται υπολογίζοντας τις εισερχόμενες ακμές και τις εξερχόμενες από κάθε εικονοπλαίσιο σύμφωνα με τις σχέσεις.

$$\begin{aligned}
 P(C_{out}, I_t) &= 1 - \frac{\sum_{i=1}^X \sum_{j=1}^Y E(I_{t-1}, i + \alpha_{t-1,t}, j + \beta_{t-1,t}) E_d(I_t, i, j)}{\sum_{i=1}^X \sum_{j=1}^Y E(I_{t-1}, i, j)} \\
 P(C_{in}, I_t) &= 1 - \frac{\sum_{i=1}^X \sum_{j=1}^Y E_d(I_{t-1}, i + \alpha_{t-1,t}, j + \beta_{t-1,t}) E(I_t, i, j)}{\sum_{i=1}^X \sum_{j=1}^Y E(I_{t-1}, i + \alpha_{t-1,t}, j + \beta_{t-1,t})}
 \end{aligned} \tag{1.3.27}$$

όπου E και E_d αντιστοιχεί στο περίγραμμα της εικόνας και στην διογκωμένη μορφή (dilation version) ενώ το $(\alpha_{t-1,t}, \beta_{t-1,t})$ αντιπροσωπεύει το ολικό διάνυσμα κίνησης μεταξύ των δυο εικονοπλαισίων I_t και I_{t-1} . Το μέτρο της ανομοιότητας $ECF(I_t)$ καλείται το κλάσμα αλλαγής ακμής και υπολογίζεται ως εξής

$$ECF(I_t) = \max(P(C_{out}, I_t), P(C_{in}, I_t)) \tag{1.3.28}$$

Τέλος αυτή η τιμή συγκρίνεται με ένα κατώφλι για να καθορίσει και τις αλλαγές στα στιγμιότυπα .

$$ECF(I_t) > T \tag{1.3.29}$$

Κάτι παρόμοιο παρουσιάζεται και στο άρθρο του **Smeaton [15]** με την μόνη διαφορά ότι εξετάζεται ο λόγος της αλλαγής ακμών σε περισσότερα από ένα εικονοπλαίσια αντί της ανάλυσης της ανομοιογένειας μόνο σε ένα εικονοπλαίσιο

$$(ECF(I_t) - ECF(I_{t-1})) > T \quad (1.3.30)$$

Ο **Lienhart [16]** εκμεταλλεύεται την πληροφορία που παίρνει από τις ακμές για να ορίσει κριτήρια ανίχνευσης σταδιακών μεταβολών όπως είναι τα dissolves, fade-in και fade-out. Ομοίως και εδώ εξάγονται οι ακμές με τον αλγόριθμο Canny και ορίζονται δυο κατώφλι τα οποία διαχωρίζουν τις ακμές σε έντονες και ασθενείς ακμές σύμφωνα με τις παρακάτω σχέσεις.

$$E_w(I_t, i, j) = \begin{cases} E(I_t, i, j) & \text{if } T_w \leq E(I_t, i, j) \leq T_s \\ 0 & \text{in other cases} \end{cases}$$

$$E_s(I_t, i, j) = \begin{cases} E(I_t, i, j) & \text{if } T_s \leq E(I_t, i, j) \\ 0 & \text{in other cases} \end{cases} \quad (1.3.31)$$

όπου T_w και T_s αντιστοιχούν στο χαμηλό και υψηλό κατώφλι για τον διαχωρισμό των ασθενών και έντονων ακμών. Ενώ E_w και E_s αντιπροσωπεύουν τις εικόνες τους. Στη συνέχεια υπολογίζεται η συνάρτηση EC (*edge based contrast*)

$$EC(I_t) = 1 + \frac{\sum_{i=1}^X \sum_{j=1}^Y E_s(I_t, i, j) - \sum_{i=1}^X \sum_{j=1}^Y E_w(I_t, i, j) - 1}{\sum_{i=1}^X \sum_{j=1}^Y E_s(I_t, i, j) + \sum_{i=1}^X \sum_{j=1}^Y E_w(I_t, i, j) + 1} \quad (1.3.32)$$

Έτσι dissolve ανιχνεύεται όταν η τρέχουσα τιμή του EC είναι τοπικά ελάχιστη.

ε) Μέθοδοι βασισμένες σε πληροφορία κίνησης

Πολλά είναι τα άρθρα τα οποία επικεντρώνουν το ενδιαφέρον τους σε μεθόδους βασισμένες στην κίνηση (**motion based method**) και εκμεταλλεύονται με διάφορους τρόπους την πληροφορία που εξάγεται από κίνηση της κάμερας (camera motion), τα διανύσματα κίνησης (motion vectors), και την οπτική ροή (optical flow)

Ο **Akutsu [17]** από την άλλη χρησιμοποιεί την ομαλότητα της κίνησης (motion smoothness) για να ανιχνεύει μεταβολές στιγμιότυπων. Η μέθοδος αυτή υπολογίζει τα διανύσματα κίνησης και στη συνέχεια υπολογίζεται το μέτρο και η γωνία που σχηματίζει με το επίπεδο το διάνυσμα. Στη συνέχεια γίνεται κάποια

συσχέτιση των διανυσμάτων του ενός με το επόμενο εικονοπλαίσια για να υπολογιστεί και η ομοιότητα μεταξύ των εικονοπλαισίων (interframe similarity) ενώ το motion smoothness ορίζεται σαν ο λόγος μεταξύ του αριθμού των blocks που έχουν αλλάξει σημαντικά και της απόστασης τους από την προηγούμενη θέση τους.

Συνοψίζοντας όλες τις παραπάνω μεθόδους και τα αποτελέσματα τους καταλήγουμε στο ότι η προσπάθεια μιας συνδυαστικής μελέτης χρησιμοποιώντας πληροφορία από μεταβολές της έντασης αλλά και πληροφορία από τα διανύσματα κίνησης οδηγεί σε μια αποδοτικότερη αντιμετώπιση του προβλήματος που είναι και ο βασικός στόχος της συγκεκριμένης διπλωματικής εργασίας.

Κεφαλαίο 2

Συμπίεση Βίντεο

2.1 Motion Pictures Expert Group (MPEG)

Το MPEG αποτελεί ένα σύνολο τυποποιήσεων που έχουν αναπτυχθεί από τον Διεθνή Οργανισμό Τυποποιήσεων, για τη μετάδοση συμπιεσμένου βίντεο με μεταβλητό bitrate. Η γενική ιδέα πίσω από την κωδικοποίηση του βίντεο χρησιμοποιώντας MPEG είναι η καταγραφή των μεταβολών / διαφορών μεταξύ των εικονοπλαισίων του βίντεο, διορθώνοντας σε τακτά διαστήματα την μετάδοση με πλήρεις, αναλυτικές εικόνες

Ο στόχος είναι η ανάπτυξη μιας μεθόδου για αποθήκευση και μετάδοση μέσω δικτύου ψηφιακού βίντεο και ήχου. Το MPEG standard σήμερα, αποτελείται από εξι φάσεις, MPEG-1, MPEG-2, MPEG-3, MPEG-4, MPEG-7 και MPEG-21 οι οποίες ορίζονται από τις ανάλογες συστάσεις της ISO/IEC. Θα παρατεθούν γενικά τα χαρακτηριστικά και οι τεχνικές που εφαρμόζονται στο πρότυπο MPEG 1-2 . [23,25]



Σχήμα 2.1.1 : Το σύνολο των τυποποιήσεων MPEG

Ανοικτά Πρότυπα Συμπίεσης Video (Open Video Compression Standards)

1. **MPEG-1:** Το πρότυπο αυτό βασίζεται στη παρατήρηση ότι το σύνολο των διαδοχικών εικόνων μιας ακολουθίας video θα διαφέρει λίγο με αποτέλεσμα να έχουμε αρκετή πλεονάζουσα πληροφορία. Η συμπίεση που επιτυγχάνει είναι της τάξης του 26:1. Το πρότυπο MPEG-1 είχε σαν αρχικό στόχο την αποθήκευση συγχρονισμένου ήχου και έγχρωμης κινούμενης εικόνας σε οπτικό δίσκο (CD-ROM). Αυτός ήταν ο λόγος όπου για το ηχητικό κανάλι δεσμεύτηκαν 200 ως 320 Kbps (για ποιότητα ήχου εφάμιλλης του CD) ενώ για την κινούμενη εικόνα είναι 1.15 έως 1.2 Mbps. Η ποιότητα της εικόνας είναι επιπέδου ενός οικιακού video (VCR). [23]
2. **MPEG-2:** Το MPEG-2 απευθύνεται σε εφαρμογές με μεγάλες απαιτήσεις στην ποιότητα του video. Μάλιστα αρχικά προοριζόταν για συμπίεση video με ποιότητα στούντιο και απαιτούμενο bandwidth από 4 έως 6Mbps. Καθώς όμως ενσωμάτωσε και το MPEG-3, το οποίο εγκαταλείφθηκε στην συνέχεια, υποστηρίζει πλέον και διάφορα είδη τηλεόρασης υψηλής ευκρίνειας (HDTV). Το πρότυπο απαιτεί μεταβλητούς ρυθμούς μετάδοσης, καθώς ο βαθμός συμπίεσης μεταβάλλεται αντιστρόφως ανάλογα με την πολυπλοκότητα των πλαισίων που κωδικοποιούνται. Οι διαφορές του από το MPEG-1 είναι η υψηλότερη ποιότητα, υποστήριξη δειγματοληψίας χρωματικού χώρου YCbCr σε 4:2:2 και 4:4:4, Interlaced βίντεο ως είσοδο, field/frame DCT και εκτίμηση κίνησης, βελτιωμένες μέθοδοι κβάντισης και κωδικοποίησης (frame-based), Scalable bitstreams, δυνατότητα κωδικοποίησης πολυκαναλικού ήχου, επέκταση του MPEG-1 Audio. [23]

Τα τέσσερα βαθμωτά επίπεδα κωδικοποίησης του video stream MPEG-2 (χαμηλά, μεσαία και υψηλά στρώματα)

Πίνακας 2.2.1

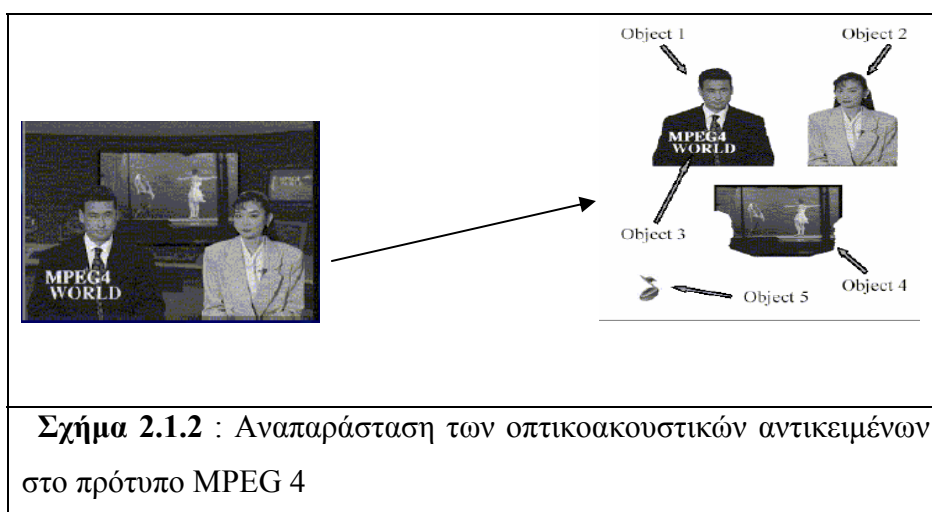
Επίπεδο	Μέγιστος αριθμός εικονοστοιχείων	Μέγιστος αριθμός γραμμών	Μέγιστος αριθμός εικονοπλαισίων ανά sec	Ρυθμός Μετάδοσης
Low	352	288	30	80 Mbps
Main	720	576	30	60 Mbps
High- 1440	1440	1152	60	14 Mbps
High	1920	1152	60	4 Mbps

3. **MPEG-3:** Το MPEG-3 χρησιμοποιούνταν για κωδικοποίηση και συμπίεση του

HDTV. Με την ραγδαία όμως ανάπτυξη του MPEG-2 εγκαταλείφθηκε εφόσον

αυτή η ανάγκη καλύφθηκε με το MPEG-2. [23]

3. **MPEG-4:** Επιτρέπει κωδικοποίηση σε επίπεδο οπτικοακουστικών αντικειμένων (AudioVisual objects, AV Objects ή Media Objects), ενοποιεί συνθετικά και φυσικά τα οπτικοακουστικά αντικείμενα, επιτρέπει ορισμό διάδρασης σε επίπεδο αντικειμένων. [23]



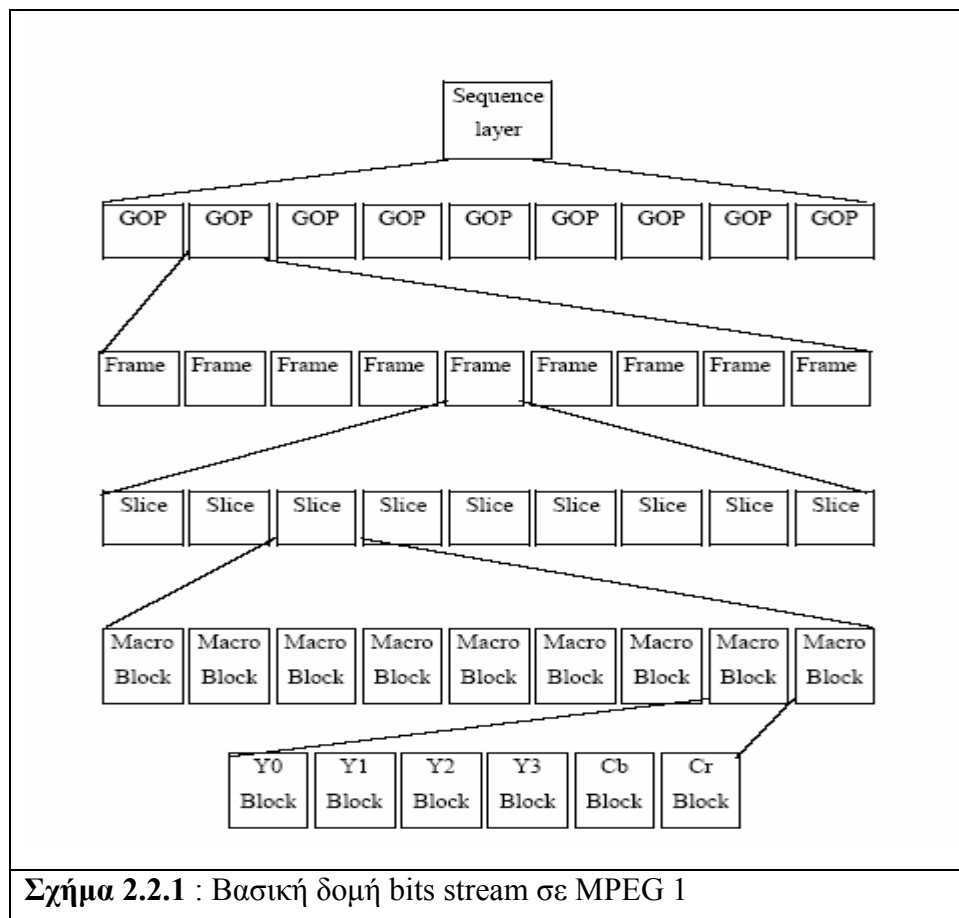
Τα βασικά χαρακτηριστικά του πρότυπου είναι η κωδικοποίηση με χαμηλό ρυθμό μετάδοσης (<64 Kbps), διαδραστική και σε επίπεδο αντικειμένου κωδικοποίηση, ανεκτικότητα σε σφάλματα, κωδικοποίηση γραφικών και βίντεο στον τρισδιάστατο χώρο, διαχείριση δικαιωμάτων πνευματικής ιδιοκτησίας. Με την ολοκλήρωση του, το πρότυπο MPEG-4 θέτει σε λειτουργία ένα ευρύ φάσμα καινούριων εφαρμογών, περιλαμβάνοντας multimedia εφαρμογές σε κινητά δίκτυα (mobile networks), video-τηλεφωνία με απλή (παλιά) υπηρεσία τηλεφώνου ή με ασύρματα δίκτυα. [23]

5. **MPEG-7:** Πρότυπο για την περιγραφή πολυμεσικών δεδομένων σε όσο πιο ευρεία γκάμα εφαρμογών γίνεται. Το MPEG-7 ορίζει μόνο τον τρόπο με τον οποίο περιγράφονται τα πολυμεσικά δεδομένα και όχι τις μεθόδους ανάλυσης που προηγούνται για την εξαγωγή των χαρακτηριστικών ή τις μηχανές αναζήτησης που έπονται για την χρήση αυτών. Οπτικοακουστικά δεδομένα που μπορούν να έχουν MPEG-7 περιγραφές είναι: ακίνητες εικόνες, γραφικά 3D μοντέλα, ηχος, ομιλία, βίντεο, πληροφορία σύνθεσης (σενάριο) των πιο πάνω σε μια πολυμεσική παρουσίαση. Τα κύρια στοιχεία του προτύπου είναι: Εργαλεία περιγραφής: Περιγραφείς (Descriptors, D), που ορίζουν την σύνταξη και την σημασιολογία κάθε χαρακτηριστικού (στοιχείου μεταδεδομένου). Σχήματα περιγραφών (Description Schemes, DS), που καθορίζουν την δομή και την σημασιολογία των σχέσεων μεταξύ των τμημάτων τους, που μπορεί να είναι περιγραφείς και σχήματα περιγραφών. [23]
6. **MPEG-21:** Πρότυπο για τον ορισμό ανοιχτού συστήματος-πλαισίου (framework) για την μετάδοση και κατανάλωση πολυμεσικού περιεχομένου με σκοπό την χρήση του από όλους τους κύριους παίκτες (παροχείς περιεχομένου, υπηρεσιών, διανομείς, χρήστες) σε ολόκληρη την αλυσίδα παραγωγής, μετάδοσης και κατανάλωσης. Κύριος στόχος είναι ο ορισμός της απαραίτητης τεχνολογίας για την υποστήριξη ανταλλαγής, πρόσβασης, κατανάλωσης, και γενικότερα διαχείρισης Ψηφιακών Στοιχείων

(Digital Item) με τρόπο αποτελεσματικό, διάφανο και διαλειτουργικό πάνω από ετερογενή δίκτυα και με χρήση διαφορετικών τερματικών συσκευών.
[23]

2.2 Δομή ροής δεδομένων του βίντεο σε MPEG 1

Η δομή με την οποία μεταφέρονται τα bits σε μια ροή βίντεο είναι χωρισμένη σε διαδοχικά επίπεδα όπως φαίνεται και από το παρακάτω σχήμα [25]



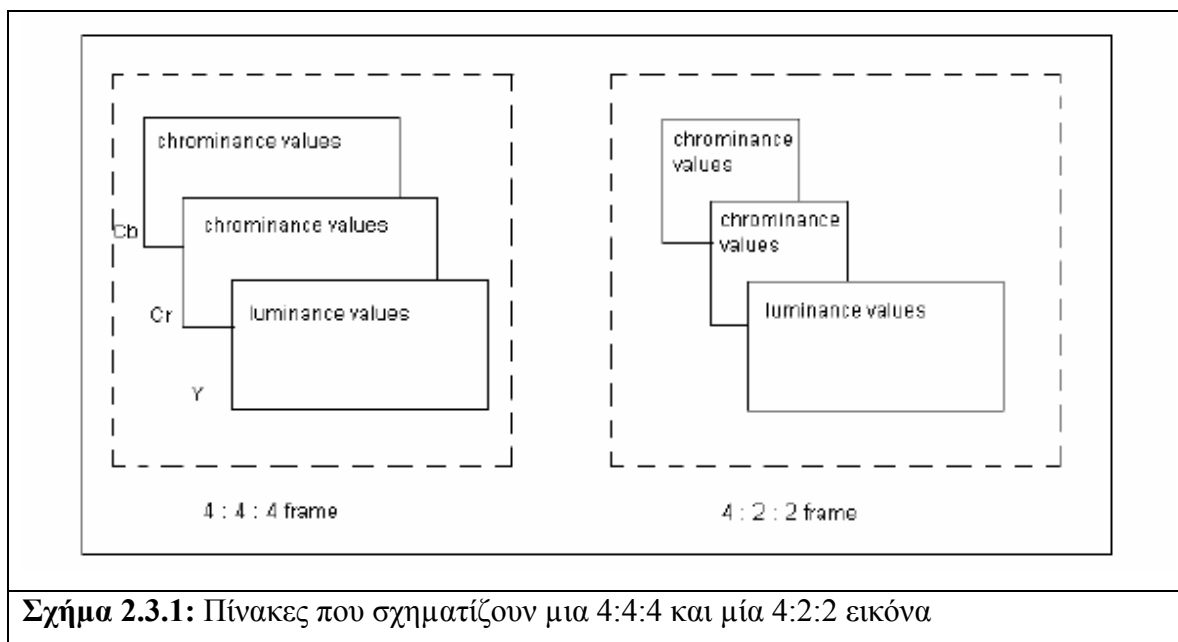
1. Το επίπεδο ακολουθίας “**sequence layer**” : Περιέχει γενική πληροφορία για το βίντεο όπως η κατακόρυφη και η οριζόντια διάσταση του εικονοπλαισίου, τον λόγο ύψους με μήκος του εικονοπλαισίου, η συχνότητα μετάδοσης των εικονοπλαισίων, το μέγεθος του buffer και τους πίνακες κβαντοποίησης

2. Το επίπεδο ομάδας εικόνων “**Group of pictures (GOP) layer**”: Είναι ένα σύνολο εικόνων που ομαδοποιούνται για να υποστηρίξουν μεγαλύτερη ευελιξία και αποδοτικότητα στο κωδικοποίητη και αποκωδικοποιητή.
3. Το επίπεδο εικονοπλασίου “**frame layer (picture layer)**” : Είναι η πρωταρχική κωδικοποιήσιμη ενότητα, η οποία περιέχει πληροφορία σχετικά με την σειρά με την οποία εμφανίζονται τα εικονοπλαίσια, την κατηγορία των εικονοπλασίων (Intra, Predicted or Bi-directionally predicted) την ακρίβεια και το εύρος μέσα στο οποίο κάθε διάνυσμα κίνησης εμφανίζεται σε ένα εικονοπλαίσιο.
4. Το επίπεδο του κομματιού “**Slice layer**”: Είναι υπεύθυνο για την διαχείριση των λαθών. Αν για παράδειγμα ο αποκωδικοποιητής συναντήσει ένα κατεστραμμένο κομμάτι για να αποφευχθεί η αποδιοργάνωση της ροής του βίντεο ξεπερνάει αυτό το σημείο πηγαίνοντας στο διαδοχικά επόμενο κομμάτι του εικονοπλασίου χωρίς να παρατηρηθεί αισθητά ένα πρόβλημα στη ροή του βίντεο.
5. Το επίπεδο Μακρό-τμήματος “**Macroblock layer**”: Είναι η βασική ενότητα κωδικοποίησης και περιέχει πληροφορία για τα διανύσματα κίνησης που σχετίζονται με το κάθε macroblock.
6. Το επίπεδο τμήματος “**Block layer**” : Είναι η μικρότερη ενότητα κωδικοποίησης και περιέχει πληροφορία για τους DCT συντελεστές των εικονοστοιχείων.

2.3 Βασική πηγή πληροφορίας (Source Input Format)

- **Εικονοπλαίσιο “Frame”**

Ένα εικονοπλαίσιο περιέχει όλη την απαραίτητη πληροφορία για το χρώμα και την φωτεινότητα, έτσι ώστε να εμφανιστεί η εικόνα. Αυτή η πληροφορία οργανώνεται σε τρεις πίνακες που περιέχουν τις τιμές για την φωτεινότητα “luminance” και χρώμα “chrominance”. Το μέγεθος αυτών των πινάκων διαφέρει ανάλογα με τον ρυθμό δειγματοληψίας χρώματος που εφαρμόζεται. Οι αντίστοιχοι πίνακες για ρυθμό 4:4:4 και 4:2:2 φαίνονται παρακάτω.



Σχήμα 2.3.1: Πίνακες που σχηματίζουν μια 4:4:4 και μία 4:2:2 εικόνα

Πιο αναλυτικά: όπως είναι γνωστό για την κωδικοποίηση της εικόνας δε χρησιμοποιούνται τρία ανεξάρτητα σήματα για τις συνιστώσες R,G,B παρόλο που κάθε χρώμα μπορεί να περιγραφεί από αυτές και άλλο ένα για τη φωτεινότητα, αλλά στην πράξη χρησιμοποιείται μόνο αυτό της φωτεινότητας (Y) και δύο άλλα (Cr) και (Cb) που περιέχουν τις πληροφορίες για τα R,G,B αν συνδυαστούν με τη φωτεινότητα σύμφωνα με τους παρακάτω τύπους :

$$Y = 0,30R + 0,59G + 0,14B$$

$$Cr = 0,74(R-Y) - 0,27(B-Y) = 0,60R + 0,28G + 0,32B$$

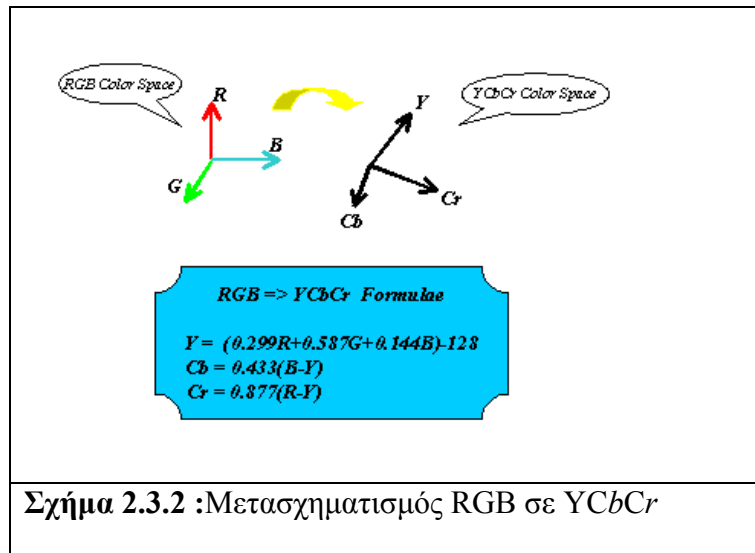
$$Cb = 0,48(R-Y) + 0,41(B-Y) = 0,21R + 0,52G + 0,31B \quad (\text{για το NTSC})$$

και

$$Y = 0,30R + 0,59G + 0,11B$$

$$Cr = 0,493(B-Y) = -0,15R - 0,29G + 0,44B$$

$$Cb = 0,877(R-Y) = 0,62R - 0,52G - 0,10B \quad (\text{για το PAL})$$



Τα σήματα Cr και Cb δειγματοληπτούνται στα 6,75 Hz και το Y στα 13,5 Hz (δηλαδή στη διπλή συχνότητα). Η συχνότητα δειγματοληψίας του Y είναι διπλάσια των άλλων δυο γιατί το σύστημα όρασης του ανθρώπου (HVS-Human Visual System) είναι περισσότερο ευαίσθητο στις μεταβολές της Y φωτεινότητας. Έτσι γίνεται πλήρης κωδικοποίηση στο Y στοιχείο ενώ στα άλλα δυο στοιχεία Cb Cr επειδή υπάρχει μικρότερη ευαισθησία, υποδειγματίζονται κατά μέθοδο της κωδικοποίησης. Κάνοντας αυτή την διαδικασία μειώνονται τα δεδομένα χωρίς να επηρεάζεται η οπτική ποιότητα.[27]

- **Κομματα “Slices”**

Ένα MPEG εικονοπλαίσιο όπως και αναφέρθηκε νωρίτερα χωρίζεται σε κομμάτια (slices). Ένα slice αποτελείται από μια σειρά από macroblocks ακολουθώντας μια σάρωση του εικονοπλαισίου από αριστερά προς τα δεξιά και από πάνω προς τα κάτω.

Στην κωδικοποιημένη ροή των bits κάθε slice ξεκινάει με μια επικεφαλίδα η οποία δεν έχει κωδικοποιηθεί όπως γίνεται σε όλα τα δεδομένα που υπάρχουν στην ροή

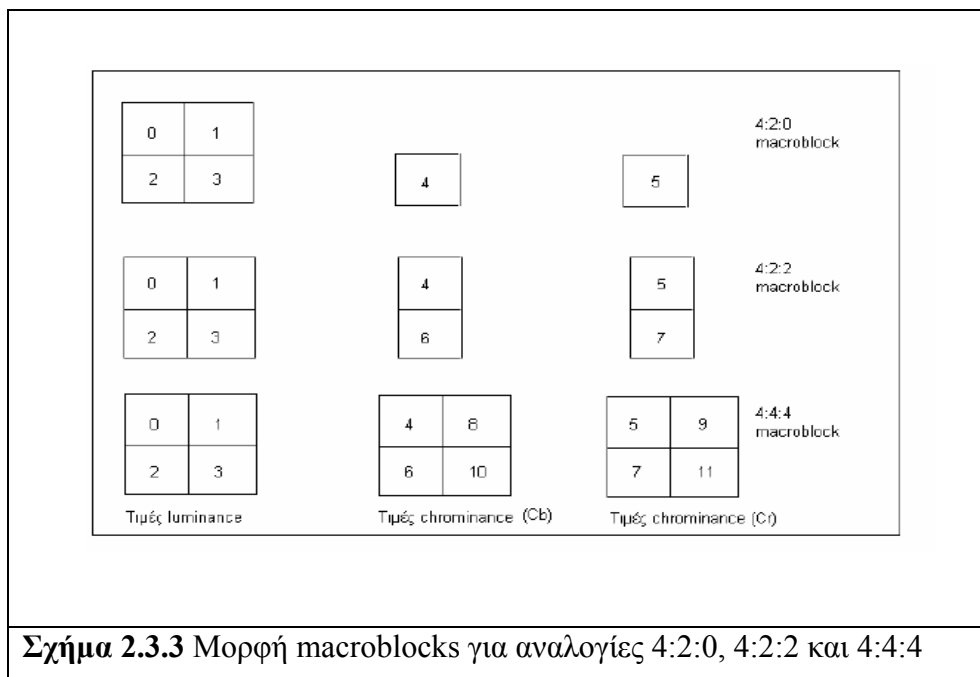
μετάδοσης bits και συνεπώς είναι το χαμηλότερο επίπεδο ενοτήτων στη κωδικοποιημένη ροή που μπορεί να έχει κάποιος πρόσβαση χωρίς να χρειαστεί αποκωδικοποίηση. Η παρουσία τμημάτων δεν συμβάλλει στη συμπίεση της εικόνας, παρά μόνο στην ύπαρξη πληροφορίας συγχρονισμού μέσα σε μία εικόνα σε ότι αφορά την απεικόνισή της. Σε περίπτωση ενός λάθους στη μετάδοση, και την απώλεια πληροφορίας σχετικά με την εικόνα, η διεργασία δεν απορρίπτει ολόκληρο το εικονοπλαιο, παρά χρησιμοποιεί την πληροφορία στο τμήμα και συνεχίζει με την αποκωδικοποίηση του επόμενου τμήματος. Συμπεραίνουμε ότι έχοντας πολλά τμήματα σε μία εικόνα, έχουμε μεγάλη ανεκτικότητα σε λάθη. Παράλληλα όμως, αυτό συνεπάγεται ένα μεγάλο overhead για κάθε εικόνα, οπότε πρέπει να βρούμε την χρυσή τομή ανάμεσα σε ποιότητα, παρουσία λαθών και απαιτούμενο εύρος ζώνης..[27]

- **Τμήματα - Μακροτμήματα “ Blocks - Macroblocks”**

Τα blocks και τα macroblocks συνθέτουν μία ιεραρχία στη δομή του MPEG : κάθε εικονοπλαίσιο διαιρείται σε έναν αριθμό blocks, τα οποία ομαδοποιούνται σε macroblocks και αυτά με την σειρά τους σε slices. Κάθε block αποτελείται από οχτώ γραμμές και κάθε γραμμή περιέχει οχτώ δείγματα τιμών luminance και chrominance από ένα εικονοπλαίσιο. Έτσι συνολικά, έχουμε 64 τιμές chrominance ή luminance σε ένα block, κάθε block συνεπώς αντιπροσωπεύει 64 εικονοστοιχεία. Κάθε macroblock περιέχει σταθερά τέσσερα blocks με τιμές luminance και έναν αριθμό από blocks με τιμές chrominance. Ο αριθμός των blocks chrominance σε ένα macroblock εξαρτάται από το ρυθμό δειγματοληψίας χρώματος. Έτσι υπάρχουν :

- macroblocks ρυθμού 4:2:0 : Περιέχουν δύο blocks με πληροφορία chrominance.
- macroblocks ρυθμού 4:2:2 : Περιέχουν τέσσερα blocks με πληροφορία chrominance.
- macroblocks ρυθμού 4:4:4 : Περιέχουν οχτώ blocks με πληροφορία chrominance.

Η παρακάτω εικόνα αποσαφηνίζει τη δομή ενός macroblock για τους τρεις ρυθμούς δειγματοληψίας χρώματος.[27]

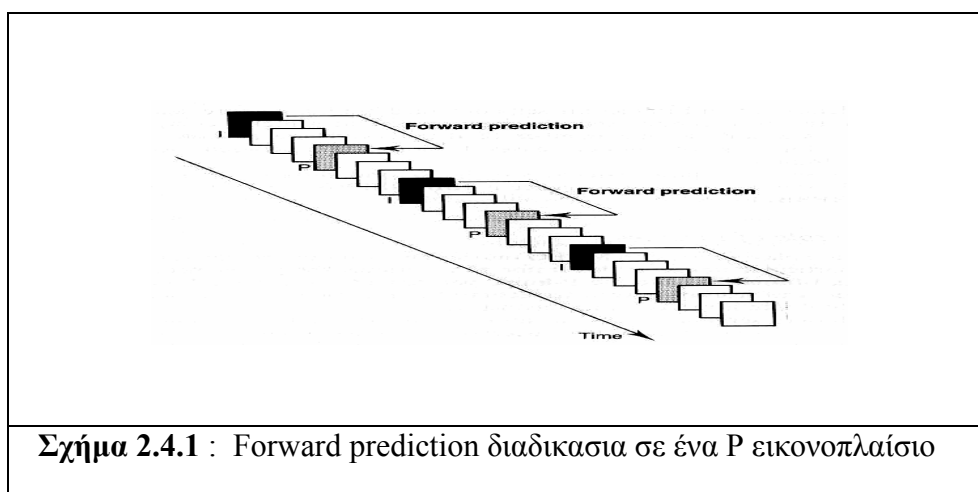


Πάνω στους πίνακες αυτούς εφαρμόζεται ο μετασχηματισμός DCT και οι υπόλοιπες διαδικασίες συμπίεσης καθώς και η διαδικασία της πρόβλεψης της κίνησης.

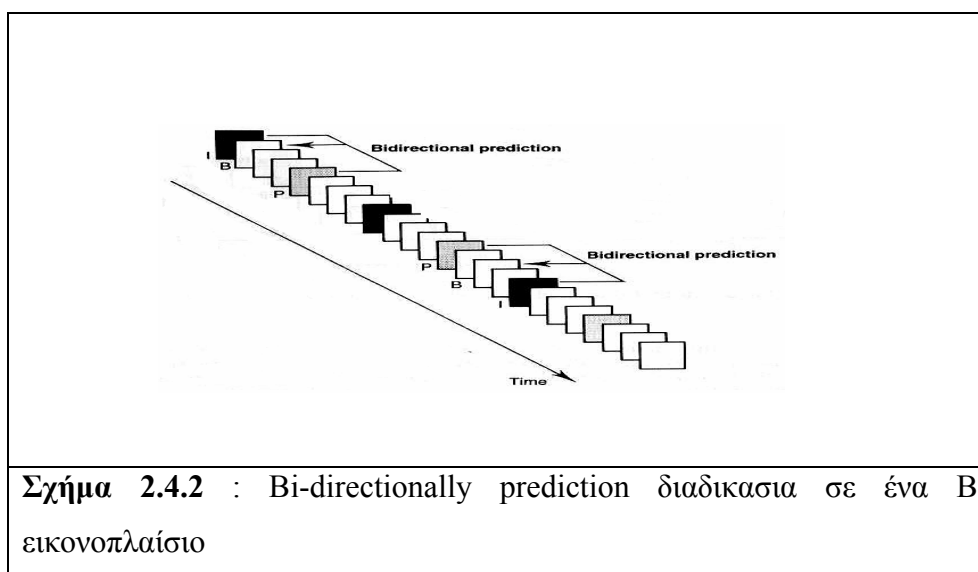
2.4 Είδη εικονοπλαίσιων

I-εικονοπλαίσιο “Intra coded frame”: Το είδος αυτό των εικονοπλαίσιων κάνει χρήση του intra frame-coding. Τα εικονοπλαίσια **τύπου I** είναι τα μόνα που είναι κωδικοποιημένα στο σύνολό τους με μεθόδους συμπίεσης εικόνας (DCT μετασχηματισμού) και η αποκωδικοποίηση μπορεί να γίνει χωρίς αναφορά σε κάποιο άλλο εικονοπλαίσιο. Είναι κατά συνέπεια τα μεγαλύτερα σε μέγεθος και αποτελούν σημεία αναφοράς κατά την τυχαία προσπέλαση ενός σήματος. Επειδή η παρουσία τους είναι απαραίτητη σαν σημείο χρονικής αναφοράς και για να αποφευχθεί η διάδοση των σφαλμάτων που δημιουργούν τα P εικονοπλαίσια επιβάλλεται να μεταδίδονται ανά τακτά χρονικά εικονοπλαίσια. [29]

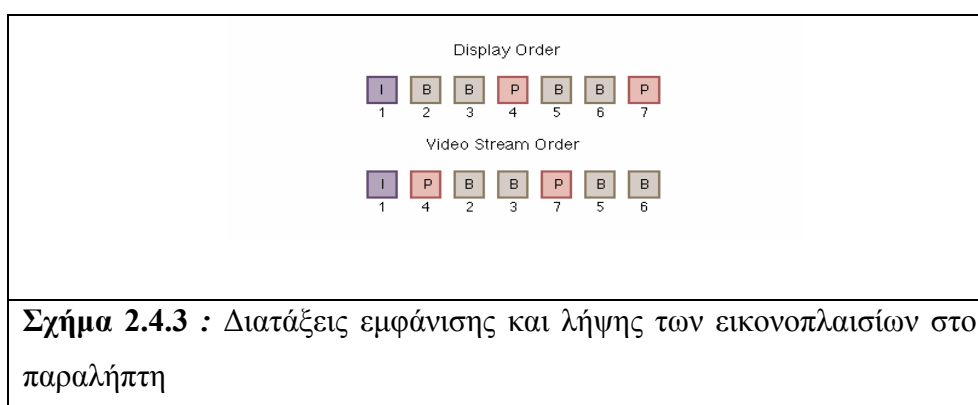
P-εικονοπλαίσιο “Predicted coded frame” : Το είδος αυτό των εικονοπλαίσιων κάνει χρήση του “forward predicted frame coding”. Τα εικονοπλαίσια **τύπου P** κωδικοποιούνται χρησιμοποιώντας κάποιο εικονοπλαίσιο αναφοράς και πιο συγκεκριμένα ένα προηγούμενο I ή P εικονοπλαίσιο. Με τη σειρά τους μπορούν να αποτελέσουν και αυτά σημείο αναφοράς για επόμενα εικονοπλαίσια και αυτός είναι και ο λόγος που συμβάλλουν στην εισαγωγή και διάδοση σφαλμάτων, αφού η διαδικασία της πρόβλεψης κίνησης δεν μπορεί να είναι 100% ακριβής. Δεν έχουν το μέγεθος των I εικονοπλαίσιων παρουσιάζουν μεγαλύτερο ποσοστό συμπίεσης. [29]



B-εικονοπλαίσιο “Bi-directionally predicted coded frames”: Τα εικονοπλαίσια τύπου **B** είναι εικονοπλαίσια που κωδικοποιούνται βασιζόμενα σε δυο εικονοπλαίσια αναφοράς ένα προηγούμενο I ή P χρησιμοποιώντας “forward predicted frame coding” και ένα επόμενο I ή P χρησιμοποιώντας “backward predicted frame coding”. Ένα B εικονοπλαίσιο δεν μπορεί ποτέ να αποτελέσει εικονοπλαίσιο αναφοράς για την κωδικοποίηση κάποιου αλλού και επειδή έχει υποστεί το μεγαλύτερο ποσοστό συμπίεσης από τα άλλα δυο είναι και το μικρότερο σε μέγεθος. [29]



Στο ακόλουθο σχήμα γίνεται αντιληπτός ο τρόπος μετάδοσης των εικονοπλαισίων από τον αποστολέα, έτσι ώστε να είναι δυνατή η αποκωδικοποίηση στον παραλήπτη



2.5 Είδη macroblocks που εμφανίζονται σε κάθε τύπο εικονοπλαίσιο

Σε ένα **I εικονοπλαίσιο** υπάρχει μόνο ένα είδος macroblock

1. **Intra macroblocks**: Δεν υπάρχει κάποιο διάνυσμα κίνησης που να ορίζει κάποια μετακίνηση block μέσα στην περιοχή του και κωδικοποιείται χρησιμοποιώντας μόνο την πληροφορία του.

Σε ένα **P εικονοπλαίσιο** υπάρχουν τρεις κατηγορίες macroblocks

1. **Intra macroblocks**: Δεν υπάρχει κάποιο διάνυσμα κίνησης που να ορίζει κάποια μετακίνηση block μέσα στην περιοχή του και κωδικοποιείται χρησιμοποιώντας μόνο την πληροφορία του.

2. **Forward Predicted macroblocks**: Έχουν διανύσματα κίνησης που έχουν εξαχθεί χρησιμοποιώντας forward predicted μέθοδο δηλαδή διάνυσμα σε σχέση με ένα προηγούμενο σημείο αναφοράς.

3. **Skipped macroblocks**: Είναι τα macroblocks που εμφανίζονται ακριβώς τα ίδια και στο προηγούμενο εικονοπλαίσιο.

Σε ένα **B εικονοπλαίσιο** υπάρχουν πέντε κατηγορίες macroblocks

1. **Forward Predicted macroblocks**: Έχουν διανύσματα κίνησης που έχουν εξαχθεί χρησιμοποιώντας forward predicted μέθοδο δηλαδή διανύσματα σε σχέση με ένα προηγούμενο σημείο αναφοράς.

2. **Backward Predicted macroblocks**: Έχουν διανύσματα κίνησης που έχουν εξαχθεί χρησιμοποιώντας backward predicted μέθοδο δηλαδή διανύσματα σε σχέση με ένα επόμενο σημείο αναφοράς.

3. **Interpolated macroblocks (Forward and Backward predicted):** Έχουν διανύσματα κίνησης που έχουν εξαχθεί χρησιμοποιώντας backward predicted και forward predicted μέθοδο δηλαδή διανύσματα σε σχέση με ένα επόμενο και προηγούμενο σημείο αναφοράς
4. **Intra macroblocks:** Δεν υπάρχει κάποιο διάνυσμα κίνησης που να ορίζει κάποια μετακίνηση block μέσα στην περιοχή του και κωδικοποιείται χρησιμοποιώντας μόνο την πληροφορία του
5. **Skipped macroblocks:** Είναι τα macroblocks που εμφανίζονται ακριβώς τα ίδια και στο προηγούμενο εικονοπλαίσιο.

Με την περιγραφή αυτή γίνεται αντιληπτό ότι όταν κατηγοριοποιούμε τα εικονοπλαίσια σε P και B δεν σημαίνει πως κατά την διαδικασία κωδικοποίησης τους είναι απαραίτητο όλα τα macroblocks που ανήκουν στην κάθε κατηγορία να κωδικοποιηθούν με τον ίδιο τρόπο που τους επιβάλλει το είδος του εικονοπλαισίου στο οποίο ανήκουν, όπως αναφέρεται στο κεφάλαιο 2.4 .

2.6 Ενδο-εικονοπλαισιακή κωδικοποίηση (Intraframe coding) με χρήση του Διακριτού Συνημιτονικού Μετασχηματισμού (DCT : Discrete Cosine Transform)

Ο Διακριτός Συνημιτονικός Μετασχηματισμός (Discrete Cosine Transform) είναι μία μέθοδος που βρίσκει μεγάλη εφαρμογή στην ψηφιακή συμπίεση γενικά αλλά και στο MPEG ειδικότερα. Με το μετασχηματισμό DCT μπορεί να μεταφερθεί η πληροφορία που περικλείεται στην εικόνα από το πεδίο του χώρου στο πεδίο της συχνότητας (αφηρημένο πεδίο), όπου η περιγραφή της μπορεί να γίνει με σημαντικά μικρότερο πλήθος bits, για διάφορους λόγους.

Ο μετασχηματισμός DCT ορίζεται ως εξής : Αρχικά οι τιμές των εικονοστοιχείων $f(x,y)$ μετασχηματίζονται από το πεδίο τιμών $[0,255]$ σε $[-127,128]$ στη συνέχεια το κάθε εικονοπλαίσιο χωρίζεται σε μπλοκς των 8×8 εικονοστοιχείων και εκεί εφαρμόζεται ο παρακάτω τύπος

$$F(u,v) = 1/4 C(u)C(v) \sum_{x=0}^7 \sum_{y=0}^7 f(x,y) \cos((2x+1)u\pi/16) \cos((2y+1)v\pi/16)$$
$$u, v, x, y = 0, 1, 2, \dots, 7$$
$$C(j) = 1/\sqrt{2}, j = 0$$
$$C(j) = 1, j > 0$$
(2.6.1)

υπολογίζεται η τιμή $DCT(i,j)$ που είναι η τιμή του συντελεστή του μετασχηματισμού στο πεδίο της συχνότητας. Έτσι αντιστοιχίζονται οι τιμές των εικονοστοιχείων στις τιμές συντελεστών.

Κατά την αναπαραγωγή γίνεται η αντίστροφη διαδικασία με τη βοήθεια του μετασχηματισμού IDCT (Inverse Discrete Cosine Transform - Αντίστροφος Διακριτός Μετασχηματισμός Συνημίτονων), που περιγράφεται από τον τύπο:

$$f(x,y) = 1/4 \sum_{u=0}^7 \sum_{v=0}^7 C(u)C(v) F(u,v) \cos((2x+1)u\pi/16) \cos((2y+1)v\pi/16)$$
(2.6.2)

Το αποτέλεσμα είναι ότι μπορεί να εξαχθεί σχεδόν ανέπαφη η αρχική πληροφορία (εκτός από κάποια αναπόφευκτα σφάλματα στρογγυλοποίησης)

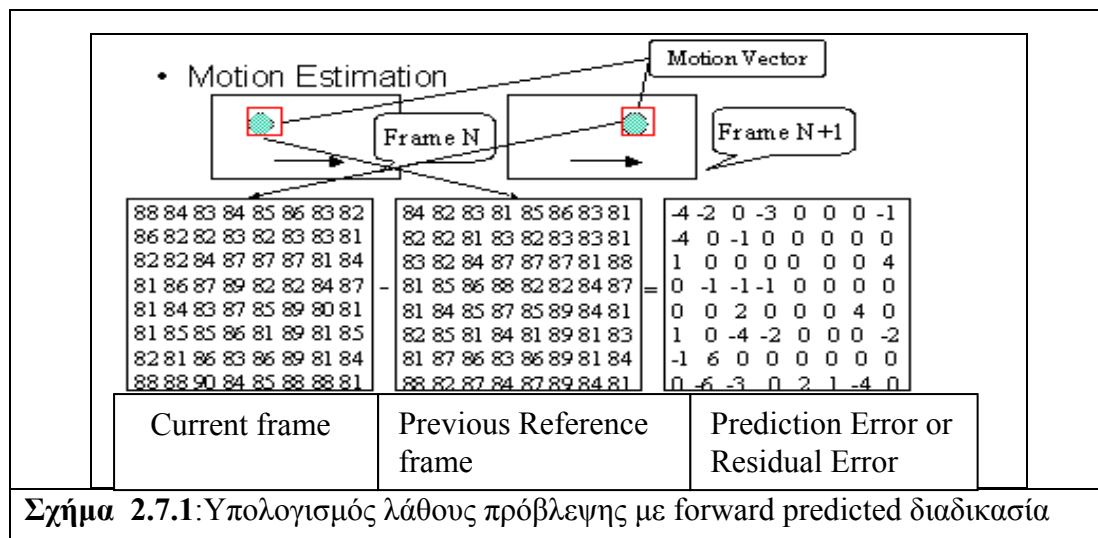
88 84 83 84 85 86 83 82		67 51 -6 2 -2 0 5 -5
86 82 82 83 82 83 83 81		-4 1 2 1 5 1 -3 0
82 82 84 87 87 87 81 84	DCT	2 3 4 6 -2 2 1 5
81 86 87 89 82 82 84 87	→	-3 -1 0 2 0 -2 2 -4
81 84 83 87 85 89 80 81		4 3 1 -1 -2 1 -3 1
81 85 85 86 81 89 81 85		1 -2 0 -3 2 -1 1 1
82 81 86 83 86 89 81 84		3 0 -1 0 -1 -1 0 -2
88 88 90 84 85 88 88 81		-1 -1 -5 5 2 -2 2 0
Σχήμα 2.6.1: Εφαρμογή του μετασχηματισμού DCT σε ένα 8x8 block.		

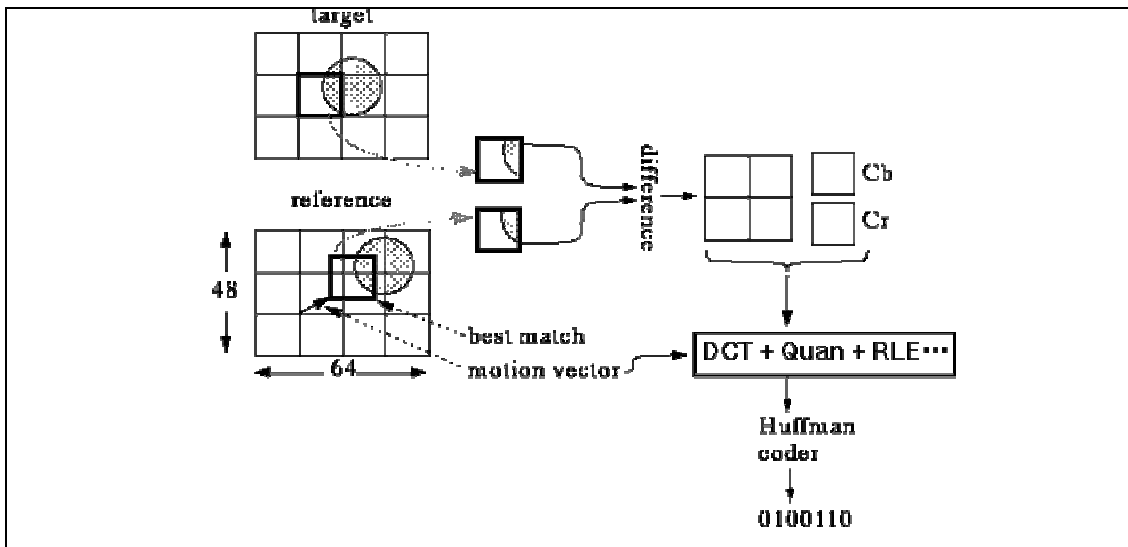
Συνεπώς ακολουθώντας αυτήν την διαδικασία κωδικοποιούνται τα blocks 8x8 των Intra macroblocks των I,P και B εικονοπλαισίων. [26]

2.7 Δια-εικονοπλαισιακή κωδικοποίηση (Interframe coding) με την χρήση υπολογισμού της κίνησης (Motion Estimation)

Ένας τρόπος με τον οποίο περιγράφονται πιο αποτελεσματικά εικόνες με κίνηση είναι η τμηματική πρόβλεψη κίνησης. Με τη βοήθεια αυτής της μεθόδου υλοποιείται η δια-εικονοπλαισιακή κωδικοποίηση με την οποία περιγράφεται η αλληλουχία εικόνων ως σειρά ομοιοτήτων και διαφορών. Έτσι μπορεί να περιγραφεί η διαδοχικά επόμενη εικόνα με το να χωριστεί σε ίσα τμήματα και να φτιαχτεί ένας πίνακας που να περιέχει τμήματα που έχουν μείνει ίδια και διανύσματα που να δείχνουν τη νέα θέση των τμημάτων που άλλαξαν θέση. Έτσι αν έχει ήδη αποσταλεί η πρώτη εικόνα μπορεί να σταλεί και η δεύτερη σαν ένα πίνακα X διανυσμάτων και ορισμένων σταθερών τμημάτων, που προφανώς έχει πολύ μικρότερο μέγεθος.

Η διάφορα της θέσης των δυο blocks ονομάζεται διάνυσμα κίνησης **Motion Vector**, ενώ η διαφορά των τιμών φωτεινότητας των δυο blocks ονομάζεται λάθος πρόβλεψης **Prediction Error**. Ο κωδικοποιητής υπολογίζει τα δυο αυτά μεγέθη τα οποία εν συνεχεία χρησιμοποιεί ο αποκωδικοποιητής μαζί με το τρέχων εικονοπλαίσιο για ανακατασκευάσει το επόμενο εικονοπλαίσιο. Η αντίστροφη αυτή διαδικασία του υπολογισμού κίνησης είναι η αντιστάθμιση κίνησης **Motion Compensation**, η οποία θα περιγράψει στην συνέχεια.

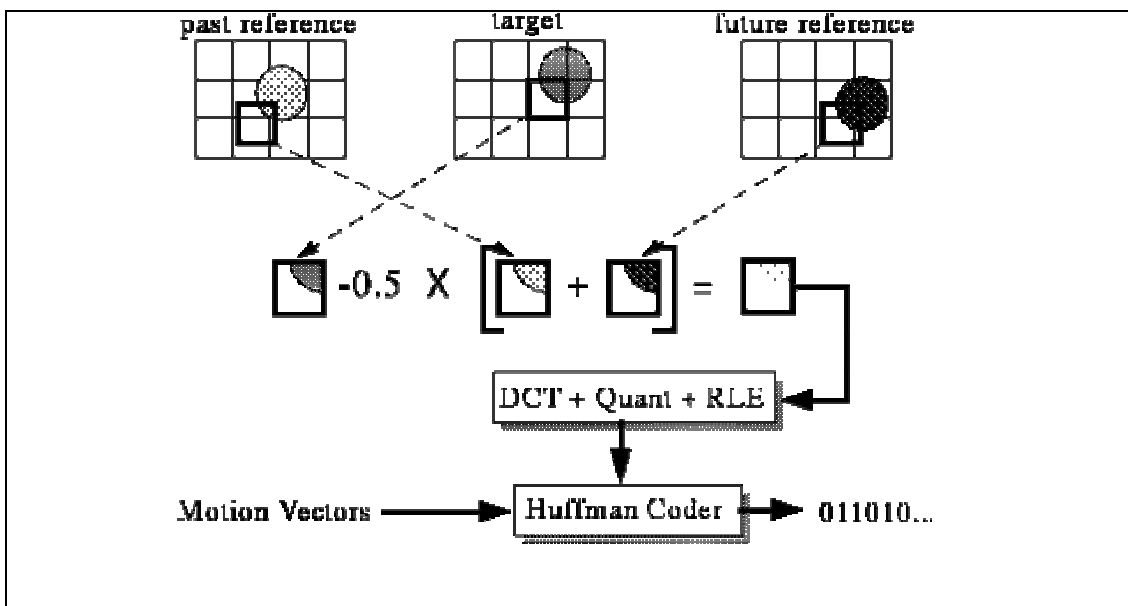




Σχήμα 2.7.2 :Η διαδικασία κωδικοποίησης χρησιμοποιώντας δια-εικονοπλαισιακή κωδικοποίηση (inter frame coding). Ο υπολογισμός της διαφοράς τιμών και του διανύσματος κίνησης ενός block σε σχέση ένα μόνο εικονοπλαίσιο αναφοράς . Ένα τέτοιο block μπορεί να είναι κάποιο που ανήκει σε μια από τις παρακάτω κατηγορίες macroblocks

Forward predicted macroblock

Backward predicted macroblock



Σχήμα 2.7.3 :Η διαδικασία κωδικοποίησης χρησιμοποιώντας δια-εικονοπλαισιακή κωδικοποίηση (inter frame coding). Ο υπολογισμός της διαφοράς τιμών και του διανύσματος κίνησης ενός block σε σχέση δυο εικονοπλαίσια αναφοράς . Ένα τέτοιο block ανήκει μόνο σε ένα Bi-directionally predicted macroblock

Η σύνταξη του MPEG καθορίζει πως θα αναπαρίσταται η πληροφορία και όχι πως θα υπολογίζονται τα διανύσματα κίνησης για το λόγο αυτό εμφανίζονται διάφορες υλοποιήσεις της μεθόδου εύρεσης των διανυσμάτων κίνησης οι οποίες στηρίζονται όλες στην ελαχιστοποίηση μίας συνάρτησης που υπολογίζει την ταύτιση του τρέχοντος με το macroblock αναφοράς.

Αν και μπορεί να χρησιμοποιηθεί κάθε συνάρτηση σφάλματος που υπάρχει, η πιο συχνά χρησιμοποιούμενη συνάρτηση είναι η **Μέση Απόλυτη Διαφορά -Mean Absolute Difference(MAD)** η οποία δίνεται από τον παρακάτω τύπο :

$$MAD(dx, dy) = \frac{1}{mn} \sum_{i=-n/2}^{n/2} \sum_{j=-m/2}^{m/2} |F(i, j) - G(i + dx, j + dy)| \quad (2.7.1)$$

$F(i,j)$: το μπλοκ στο τρέχων εικονοπλαίσιο

$G(i,j)$: το ίδιο μπλοκ στο προηγούμενο ή επόμενο εικονοπλαίσιο

(dx,dy) : το διάνυσμα για την περιοχή αναζήτησης

Το macroblock που παράγει το μικρότερο σφάλμα αντιστοιχεί στην τιμή του διανύσματος που αναζητείται. [26]

Μια ακόμα συνάρτηση υπολογισμού είναι η **Μέση Διάφορα Τετράγωνων (Mean Squared Difference-MSD)** με τύπο :

$$MSD(dx, dy) = \frac{1}{mn} \sum_{i=-n/2}^{n/2} \sum_{j=-m/2}^{m/2} |F(i, j) - G(i + dx, j + dy)|^2 \quad (2.7.2)$$

Μια επίσης συνήθης συνάρτηση είναι η **Cross-Correlation Difference (CCD)** που προκύπτει από τον παρακάτω τύπο :

$$CCF(dx, dy) = \frac{\sum_i \sum_j F(i, j)G(i + dx, j + dy)}{\left(\sum_i \sum_j F^2(i, j) \right)^{1/2} \left(\sum_i \sum_j G^2(i + dx, j + dy) \right)^{1/2}} \quad (2.7.3)$$

και τέλος μια συνάρτηση ομαδοποίησης με βάση τις διαφορές των εικονοστοιχείων η **Pixel Difference Classification (PDC)** η οποία προκύπτει από τον τύπο :

$$PDC(dx, dy) = \sum_i \sum_j T(dx, dy, i, j)$$

$$T(dx, dy, i, j) = \begin{cases} 1 & \text{if } |F(i, j) - G(i + dx, j + dy)| \leq t \\ 0 & \text{otherwise} \end{cases} \quad (2.7.4)$$

t: προκαθορισμένο threshold.

Κάθε εικονοστοιχείο κατηγοριοποιείται σαν matching pixel (T=1) ή mismatching pixel (T=0) και η μέγιστη τιμή **PDC** δίνει το σωστό macroblock

Η πιο απλή από θέμα υλοποίησης αλλά και η πιο πολύπλοκη από πλευράς υπολογιστικής πολυπλοκότητας είναι η πλήρης αναζήτηση (**Exhaustive search**) η οποία καλύπτει κάθε pixel στην περιοχή αναζήτησης. Ονομαστικά υπάρχουν και άλλες τεχνικές οι οποίες η κάθε μια μπορεί να υπολογίσει το καλύτερο ταίριασμα των macroblocks:

1. **Three - step search**
2. **2-D logarithmic searches**
3. **Conjugate direction search**
4. **Parallel hierarchical 1-D search (not included)**
5. **Pixel difference classification (not included)**

2.8 Αποκωδικοποίηση των εικονοπλαισίων σε μια σειρά DC ελαττωμένες εικόνες .

Σε συμπιεσμένα αρχεία βίντεο όπως είναι και το MPEG η εξαγωγή της πληροφορίας συνεπάγεται και την πλήρη αποκωδικοποίηση του βίντεο σε απλά εικονοπλαίσια σαν μια σειρά από διαδοχικές εικόνες. Επειδή όμως αυτή η διαδικασία είναι αρκετά χρονοβόρα στην εφαρμογή των μεθόδων κατάτμησης και απαιτεί μεγάλο αποθηκευτικό χώρο αποφεύγεται να χρησιμοποιηθεί.

Η αναζήτηση γρηγορότερων αλγορίθμων σε συμπιεσμένο βίντεο οδήγησε σε μια μορφή ημισυμπιεσμένου βίντεο, δηλαδή βίντεο που έχει υποστεί εν μέρει αποκωδικοποίηση. Τέτοια μορφή είναι η ακολουθία μια σειράς από διαδοχικές DC εικόνες. Τί είναι όμως DC εικόνες; Οι DC εικόνες είναι οι εικόνες που προκύπτουν αν στα αρχικά εικονοπλαίσια εφαρμοστεί συμπίεση βασισμένη στο Διακριτό Συνημιτονικό Μετασχηματισμό (DCT). Το αποτέλεσμα είναι μια μικρότερης ανάλυσης εικόνα η οποία όμως διατηρεί αρκετά σημαντική πληροφορία όσον αφορά την ένταση και συνεπώς μπορεί να χρησιμοποιηθεί για την μελέτη μεθόδων τμηματοποίησης. Όπως αναφέρθηκε στο κεφαλαίο 2.6 λοιπόν ένα μπλοκ από 8x8 εικονοστοιχεία μπορεί να κωδικοποιηθεί στο χώρο της συχνότητας και να δώσει ένα νέο πίνακα στον οποίο όμως όλη η πληροφορία είναι συγκεντρωμένη στον DC συντελεστή, δηλαδή στο στοιχείο (0,0). Λαμβάνοντας υπόψη αυτό το γεγονός χρησιμοποιείται μονάχα αυτή η τιμή για να περιγραφεί κάθε μπλοκ από 8x8 τιμές εικονοστοιχείων . Επομένως στα I εικονοπλαίσια η διαδικασία είναι αρκετά απλή αφού το μόνο που χρειάζεται είναι να χρησιμοποιηθούν μονάχα οι DC συντελεστές από τα ήδη κωδικοποιημένα με DCT Intra macroblocks

Σε μια εικόνα f μεγέθους $N \times M$, ορίζουμε κάθε εικονοστοιχείο στο l BxB block σαν $x_l(i,j)$ και ο DCT μετασχηματισμός του σαν $c_l(i,j)$. Καθορίζουμε την ελαττωμένη εικόνα $f(n)$ ότι είναι του τύπου R_3 , αν ισχύουν οι εξής συνθήκες

1. Έχει μέγεθος $\frac{N}{8} \times \frac{M}{8}$

2. Οι DCT συντελεστές $c_l^{(3)}$ του l block $\frac{B}{8} \times \frac{B}{8}$ υπολογίζονται με βάση τους τύπους :

$$c_l^{(3)}(i, j) = \frac{1}{8} c_l(i, j) H_3(i, j), \quad i, j = 0, 1, \dots, \frac{B}{8} - 1 \quad (2.8.1)$$

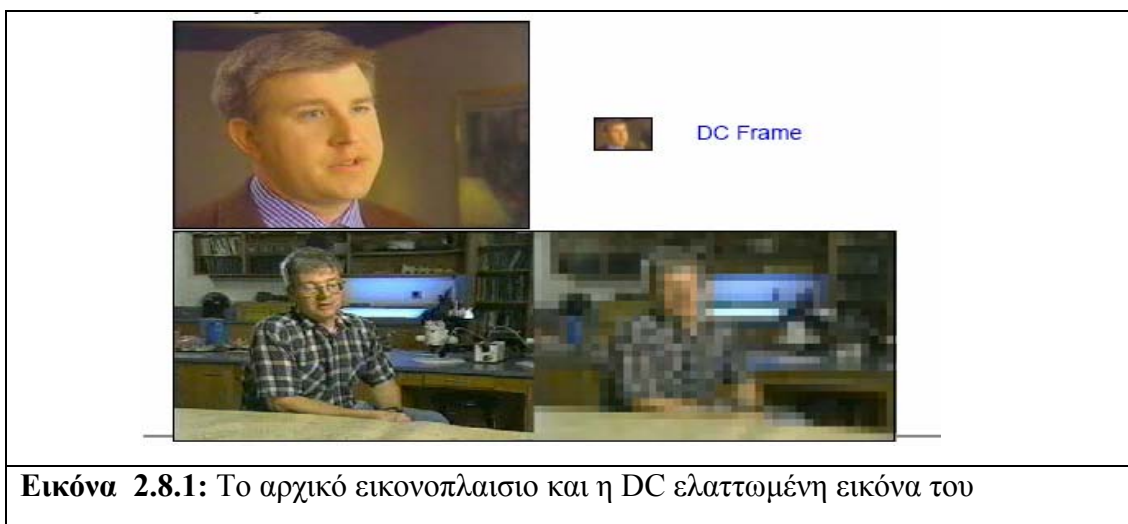
$$\text{Όπου} \quad H_3(i, j) = \begin{cases} 1 & , i = j = 0 \\ 0 & , \text{otherwise} \end{cases} \quad (2.8.2)$$

$$c(i, j) = \frac{1}{4} k(i) k(j) \left[\sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \frac{(2x+1)i\pi}{16} \cos \frac{(2y+1)j\pi}{16} \right] \quad (2.8.3)$$

Η συνθήκη (2) εννοεί ότι οι ελαττωμένες εικόνες υπολογίζονται λαμβάνοντας υποψιν κάποιους από τους συντελεστές (DC) και όχι όλους τους συντελεστές (DC+AC) του DCT της f . Άρα για την ανακατασκευή μιας εικόνας τύπου R_3 αυτό που χρειάζεται είναι ο υπολογισμός των DC συντελεστών από κάθε εικονοπλαίσιο.

Για τα P και B εικονοπλαίσια ακολουθείται μια πιο περίπλοκη διαδικασία αφού εκεί δεν έχουμε μόνο Intra macroblocks αλλά και Forward Predicted ή Backward Predicted ή Bi-directionally Predicted macroblocks. Η τεχνική υπολογισμού εκεί θα περιγράφει στην επόμενη ενότητα.

Ένα εικονοπλαίσιο που εξάγεται από ένα MPEG αρχείο με αρχική ανάλυση 320x288 μπορεί να μειωθεί σε μια εικόνα διαιρεμένη με 8 κάθε διάσταση της και να προκύψει η DC εικόνα 40x36, όπως φαίνεται και στο παρακάτω Σχήμα 4.1.1. Η διαδικασία αυτή έχει προταθεί στο άρθρο του Boon-Lock Yeo [21], “Efficient processing of compressed images and video”

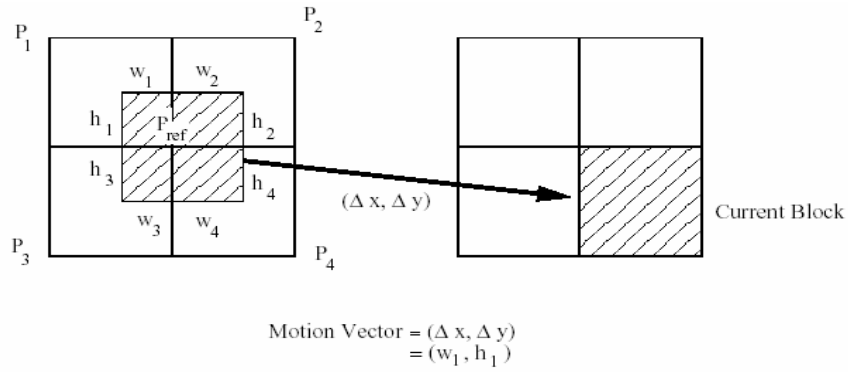


Εικόνα 2.8.1: Το αρχικό εικονοπλαίσιο και η DC ελαττωμένη εικόνα του

2.8.1 Επακριβής εξαγωγή των DC συντελεστών

Η διαδικασία εκτίμησης των συντελεστών DCT διαφέρει ανάλογα με το είδος του εικονοπλαισίου το οποίο επεξεργαζόμαστε. Η δημιουργία των DC εικόνων από ένα I εικονοπλαίσιο είναι μια απλή διαδικασία στην οποία το μόνο που χρειάζεται είναι να υπολογιστούν οι DC συντελεστές για κάθε block 8x8. Αυτό έχει γίνει κατά την αρχική διαδικασία κωδικοποίησης του βίντεο και συνεπώς με την αντίστροφη διαδικασία αποκωδικοποίησης μπορούμε να φτάσουμε στο σημείο να πάρουμε του συντελεστές αυτούς άμεσα. Παρόλο που οι DC εικόνες χρειάζονται μόνο τους DC συντελεστές από κάθε MB, στην διαδικασία ανακατασκευής τους υπολογίζονται όλοι οι DCT συντελεστές ($DCT=DC+AC$) των I και P εικονοπλαισίων διότι αποτελούν εικονοπλαίσια αναφοράς και χρειάζονται για να υπολογιστούν οι DC συντελεστές των B εικονοπλαισίων. Στην τελική όμως διαδικασία κατασκευής DC εικόνων έχουμε συνεισφορά μόνο των DC συντελεστών των MBs.

Η εξαγωγή των DCT συντελεστών σε ένα P εικονοπλαίσιο από ένα I είναι μια πιο περίπλοκη διαδικασία όπως περιγράφεται στο άρθρο του Kobla “Compressed domain video indexing techniques using DCT and motion vector information in MPEG video” [24], όπου οι DCT συντελεστές του block 8x8 ανακτώνται από τους DCT συντελεστές του αντιστοίχου block στο εικονοπλαίσιο αναφοράς. Υποθέτουμε ότι στην εικόνα 2.8.1.1 P_{ref} είναι το block αναφοράς του τρέχοντος block, και τα P_1 , P_2 , P_3 , P_4 είναι τα τέσσερα γειτονικά blocks από όπου εξάγεται το P_{ref} και το διάνυσμα κίνησης (Δx , Δy)



Σχήμα 2.8.1.1: Πως υπολογίζονται οι συντελεστές DCT για ένα block (8 x8) με βάση το block αναφοράς του.

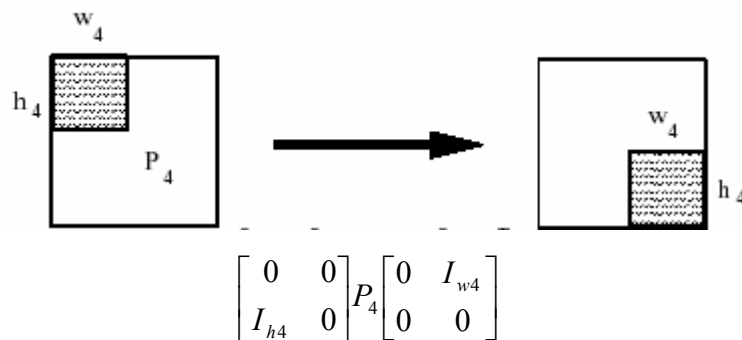
Αν το κάθε block το θεωρήσουμε σαν ένα πίνακα 8x8, τότε μπορούμε να το περιγράψουμε στο χώρο μέσα στο MB από την παρακάτω σχέση.

$$P_{ref} = \sum_{i=1}^4 S_{i1} P_i S_{i2} \quad (2.8.1.1)$$

Όπου S_{ij} είναι οι πίνακες της μορφής

$$L_n = \begin{bmatrix} 0 & 0 \\ I_n & 0 \end{bmatrix} \quad \eta \quad R_n = \begin{bmatrix} 0 & I_n \\ 0 & 0 \end{bmatrix}$$

Ένα παράδειγμα για το πώς ενεργεί ο S_{ij} στο P_4 φαίνεται στο σχήμα 4.3



Σχήμα 2.8.1.1: Ο πολλαπλασιασμός του P_4 με τους δυο πίνακες μετέφερε την πληροφορία από πάνω αριστερά, που ήταν αρχικά, σε κάτω δεξιά

Υπάρχουν τέσσερις πιθανές υποπεριοχές στις οποίες μπορεί να μετακινηθεί ένα block: πάνω αριστερά , πάνω δεξιά , κάτω αριστερά , κάτω δεξιά.

Πιθανές υποπεριοχές μετακίνησης των blocks μέσα στο macroblock			
Πίνακας 2.8.1.1			
Υποπεριοχές	Θέσεις	S_{i1}	S_{i2}
P_1	Κάτω δεξιά	$\begin{bmatrix} 0 & 0 \\ I_{h1} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & I_{w1} \\ 0 & 0 \end{bmatrix}$
P_2	Κάτω αριστερά	$\begin{bmatrix} 0 & I_{h2} \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & I_{w2} \\ 0 & 0 \end{bmatrix}$
P_3	Πάνω δεξιά	$\begin{bmatrix} 0 & 0 \\ I_{h3} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ I_{w3} & 0 \end{bmatrix}$
P_4	Πάνω αριστερά	$\begin{bmatrix} 0 & 0 \\ I_{h4} & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & I_{w4} \\ 0 & 0 \end{bmatrix}$

Στη συνέχεια ορίζουμε τον DCT μετασχηματισμό ενός block σε ένα P εικονοπλαίσιο σαν $DCT(P)$

$$DCT(P) = TPT^t \quad (2.8.1.2)$$

Όπου T είναι ο DCT πίνακας όπως ορίστηκε στην 4.2 και αντίστοιχα ο ανάστροφος του. Λόγω τώρα της γραμμικότητας του DCT ισχύει η παρακάτω σχέση.

$$DCT(AB) = DCT(A)DCT(B) \quad (2.8.1.3)$$

Έτσι εφαρμόζοντας DCT στην σχέση (4.8) μπορεί να γράφει

$$DCT(P_{ref}) = \sum_{i=1}^4 DCT(S_{i1})DCT(P_i)DCT(S_{i2}) \quad (2.8.1.4)$$

Πρέπει να σημειωθεί πάλι ότι ενώ ενδιαφερόμαστε ουσιαστικά μόνο για τους DC συντελεστές, πρέπει να ανακτήσουμε όλους τους DCT συντελεστές από τα I ή P εικονοπλαίσια γιατί χρησιμοποιούνται σαν εικονοπλαίσια αναφοράς στην πρόβλεψη των P και B εικονοπλαίσια. Ενώ για τα B εικονοπλαίσια αρκεί να υπολογιστούν οι DC συντελεστές σύμφωνα με τον παρακάτω τύπο ($DC = (DCT)_{00}$).

$$\begin{aligned} (DCT(P_{ref}))_{00} &= \sum_{i=1}^4 \left(\sum_{m=0}^7 \sum_{l=0}^7 (DCT(S_{i1}))_{0m} (DCT(P_i))_{ml} (DCT(S_{i2}))_{l0} \right) \\ &= \sum_{i=1}^4 \left(\sum_{m=0}^7 \sum_{l=0}^7 w_{ml}^i (DCT(P_i))_{ml} \right) \end{aligned} \quad (2.8.1.5)$$

Όπου για το (i,j) στοιχείο του P ορίζουμε το $(P)_{ij}$ και γράφουμε

$$w_{ml}^i = (DCT(S_{i1}))_{0m} \times (DCT(S_{i2}))_{l0}.$$

Η σχέση (2.8.1.5) ορίζει ότι η DC τιμή του block P_{ref} είναι ένας γραμμικός συνδυασμός των DCT συντελεστών των P_1, P_2, P_3, P_4 με τα βάρη w_{ml}^i . Έτσι για να μπορέσουμε να ανακτήσουμε την DC τιμή από ένα block από ένα B εικονοπλαίσιο χρειαζόμαστε το πολύ 256 πολλαπλασιασμούς. Ενώ για ένα P εικονοπλαίσιο που χρειάζονται και οι AC συντελεστές το σύνολο των πολλαπλασιασμών θα είναι 256 x 64.

2.8.2 Προσεγγιστική εξαγωγή των DC συντελεστών.

Είναι προφανές από την σχέση (4.12) ότι έχει πολύ μεγάλο υπολογιστικό κόστος η ανακατασκευή των DC συντελεστών για ένα B εικονοπλαίσιο και ακόμα περισσότερο η ανακατασκευή όλων των DCT συντελεστών από ένα P

εικονοπλαίσιο. Για να μειωθεί αυτό το κόστος προτείνεται μια προσεγγιστική μέθοδος γνωστή σαν “πρώτου βαθμού προσέγγιση” (*first order approximation*) [22][24]. Υποθέτουμε χωρίς έλλειψη της γενικότητας ότι $0 \leq \Delta x \leq 8$ και $0 \leq \Delta y \leq 8$. Έτσι, στην εικόνα 2.8.1.1, έχουμε

$$h_1 = h_3 = \Delta x, w_1 = w_3 = \Delta y, h_2 = h_4 = 8 - h_1, w_2 = w_4 = 8 - w_1.$$

Ορίζουμε με $DC(P)$ τις DC τιμές του P , ενώ οι τιμές μετά την εφαρμογή της “first order approximation” ορίζονται $DC(P)^l$.

Για να υπολογίζουμε τον DC συντελεστή με την “first order approximation”, αθροίζουμε τις τιμές των DC τιμών των γειτονικών P_1, P_2, P_3, P_4 πολλαπλασιάζοντας την κάθε τιμή με ένα αντίστοιχο βάρος. Το βάρος αυτό ορίζεται σαν ο λόγος του εμβαδού του κομματιού που υπερκαλύπτει το P_{ref} την κάθε υποπεριοχή προς το συνολικό εμβαδόν της υποπεριοχής (64) και ορίζεται σαν

$$weight(i) = \frac{h_i w_i}{64}$$

$$DC(P_{ref})^1 = \sum_{i=1}^4 \frac{h_i w_i}{64} DC(P_i) \quad (2.8.2.1)$$

Από την προσέγγιση αυτής της μεθόδου προκύπτουν δυο αξιοσημείωτες παρατηρήσεις.

Παρατήρηση 1: Ο συντελεστής w_{00}^i στην σχέση (2.8.1.5) είναι ίσος με την ποσότητα $weight(i) = \frac{h_i w_i}{64}$.

Αυτό προκύπτει απλά από το γεγονός ότι w_{00}^i είναι

$$\left(\frac{1}{8} \sum_{n,m} (S_{i1})_{n,m} \right) \left(\frac{1}{8} \sum_{n,m} (S_{i2})_{n,m} \right) = \frac{h_i w_i}{64}$$

Ξαναγράφοντας την (4.12)
$$DC(P_{ref}) = \sum_{i=1}^4 \frac{h_i w_i}{64} DC(P_i) + c$$

Όπου

$$c = \sum_{i=1}^4 \left(\sum_{(m,l) \neq (0,0)} w_{ml}^i (DCT(P_i))_{ml} \right)$$

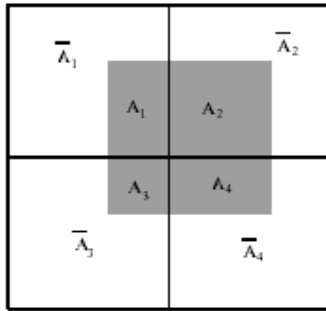
(2.8.2.2)

Μπορούμε να δούμε την ανακατασκευή του $DC(P_{ref})$ σαν το άθροισμα που δίνει η μέθοδος “ first order approximation ” και τον όρο λάθους “error term” c το οποίο δεν εξαρτάται από τους DC συντελεστές των P_1, \dots, P_4 . Μπορούμε να το δηλώσουμε $\hat{c} = \frac{c}{8}$ κανονικοποιημένο λάθος “normalized error term”. Αυτή η κανονικοποίηση λαμβάνει υπόψιν της το γεγονός ότι η τιμή DC είναι 8 φορές η μέση τιμή των εντάσεων του block.

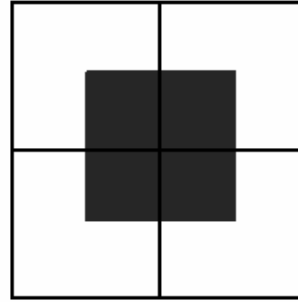
Παρατήρηση 2: Η απόλυτη μέγιστη τιμή του “κανονικοποιημένου λάθους” είναι $\frac{3}{4} \times \max_{x,y} f(x,y)$, όπου $\max_{x,y} f(x,y)$ δηλώνεται σαν η μέγιστη τιμή έντασης ενός εικονοπλαισίου. Για 8-bit εικόνας αυτή η τιμή είναι 255.

Απόδειξη : Παρόλο που στη σχέση (2.8.2.2) δουλεύουμε στο μετασχηματισμένο επίπεδο DCT, θα μπορούσαμε κάνοντας τις παρακάτω μετατροπές να επανέλθουμε στο επίπεδο του χώρου. Μπορούμε να γράψουμε το “normalized error” σαν

$$\begin{aligned} \hat{c} &= \frac{1}{64} \sum_{i=1}^4 \left(\left(1 - \frac{h_i w_i}{64} \right) \sum_{A_i} f(x,y) - \frac{h_i w_i}{64} \sum_{A_i} f(x,y) \right) \\ &\leq \frac{1}{64} \sum_{i=1}^4 \left(\left(1 - \frac{h_i w_i}{64} \right) \sum_{A_i} f(x,y) \right) \\ &\leq \max_{x,y} f(x,y) \times \sum_{i=1}^4 \left(1 - \frac{h_i w_i}{64} \right) \frac{h_i w_i}{64} \\ &\leq \frac{3}{4} \max_{x,y} f(x,y) \end{aligned}$$



Σχήμα 2.8.2.1: Προσέγγιση διευθέτηση του Block



Σχήμα 2.8.2.2: Η χειρότερη περίπτωση για προσέγγιση διευθέτησης ενός block

Όπου $f(n, m) = \begin{cases} 0 & , (n, m) \in \bar{A}_i \\ \max_{x,y} f(x, y) & , (n, m) \in A_i \end{cases}$ και $h_i = h_j$ και $w_i = w_j$, για όλα $i, j = 1, 2, 3, 4$.

Έτσι η χειρότερη προσέγγιση συμβαίνει όταν ακριβώς το $1/4$ του P_{ref} επικαλύπτει κάθε P_i με τιμές $\max_{x,y} f(x, y)$ στο P_{ref} και 0 οπουδήποτε αλλού. Μια τέτοια κατάσταση φαίνεται στο σχημα 4.5. Στην πράξη πολύ σπάνια συμβαίνει μια τέτοια μεταβολή. Τα πειραματικά αποτελέσματα έδειξαν ότι η ανακατασκευή των DC συντελεστών με την μέθοδο “first order approximation” είναι αρκετά καλή.

Το λάθος εξαρτάται από τις διαφοροποιήσεις των τιμών στο block και του διανύσματος $(\Delta x, \Delta y)$. Μάλιστα σε μια πολύ ειδική περίπτωση ,αν κάθε από τα 4 blocks είναι σταθερά ,τότε η εκτίμηση των τιμών DC είναι η ακριβής εξαγωγή όπως παρουσιάστηκε στην προηγούμενη ενότητα .Επίσης όταν $\Delta x = \Delta y = 0$ δεν θα υπάρχει λάθος εκτίμησης “error term=0”. Ενώ από την άλλη μεγάλη τιμή λάθους θα έχουμε όταν $\Delta x = \Delta y = 4$.

Παρακάτω φαίνονται οι DC εικόνες όπως εξάχθηκαν με βάση την προσέγγιση “first order approximation” από τα αρχικά τους εικονοπλαίσια.



Σχημα 2.8.2.3: Ανακατασκευή των DC εικόνων με την μέθοδο “first order approximation”

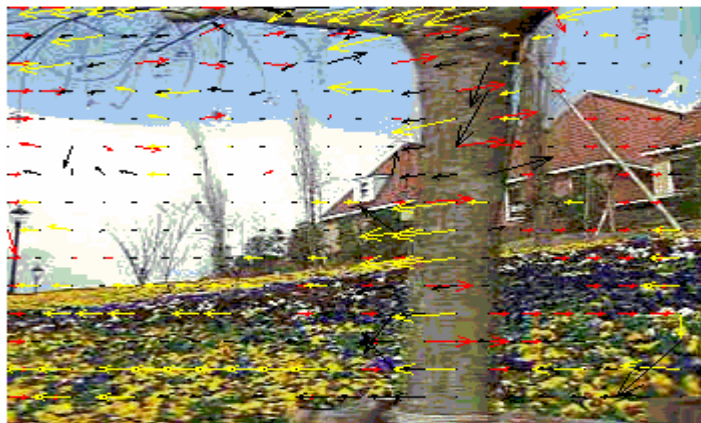
Η εφαρμογή των μεθόδων πάνω σε DC εικόνες όπως γίνεται πλέον ξεκάθαρο είναι μια πολύ γρηγορότερη διαδικασία. Πέραν όμως από αυτό το όφελος οι DC εικόνες δίνουν καλύτερα αποτελέσματα γιατί κατά κάποια έννοια έχουν μειώσει κατά πολύ τον θόρυβο λόγω της ψηφιοποίησης και κίνησης της κάμερας. Αυτό οφείλεται στο γεγονός ότι η τιμή που τελικά δίνει την πληροφορία είναι μια μέση τιμή των τιμών έντασης του μπλοκ. Από εδώ και πέρα λοιπόν σε όλο το στάδιο της παρουσίασης των μεθόδων η αναφορά σε ένταση εικονοπλαισίου θα σημαίνει την αντίστοιχη ένταση που δίνει η DC εικόνα του

2.9 Αντιστάθμιση και κανονικοποίηση της κίνησης “Motion compensation and motion vector normalization”

Μια βασική πληροφορία που χρειάζεται να εξαχθεί για την επεξεργασία του ψηφιακού συμπιεσμένου βίντεο είναι η πληροφορία κίνησης. Τα διανύσματα κίνησης όμως όπως περιγράφονται κατά την διαδικασία υπολογισμού τους δεν μας δίνουν μια ενιαία μορφή πληροφορίας και αυτό γιατί δεν έχουν όλα τα εικονοπλαισίου της ίδιας κατηγορίας διανύσματα τα οποία θα έχουν το ίδιο σημείο αναφοράς. Κάθε macroblock (MB) μπορεί να έχει από μηδέν, ένα ή δυο διανύσματα κίνησης ανάλογα με την κατηγορία εικονοπλαισίου στο οποίο ανήκει (I,P,B). Ακόμα διαφέρει και το γεγονός ότι το ένα MB μπορεί να είναι ενδοπλαισιακά κωδικοποιημένο (intracoded), με προς τα εμπρός ή προς τα πίσω πρόβλεψη (forward- ή backward predicted) ή τέλος με διπλής κατεύθυνσης πρόβλεψη (bidirectional predicted) κωδικοποιημένο. Επιπλέον, αυτά τα διανύσματα κίνησης μπορεί να είναι “forward predicted” ή “backward predicted” σε σχέση με κάποιο εικονοπλαίσιο αναφοράς το οποίο δεν είναι απαραίτητο να είναι γειτονικό με εκείνο το εικονοπλαίσιο στο οποίο ανήκει το MB. Το πρόβλημα που προκύπτει είναι όταν πρέπει να συγκριθεί το διάνυσμα του MB ενός I εικονοπλαισίου που δεν έχει διανύσματα κίνησης με ενός P εικονοπλαισίου που έχει ένα διάνυσμα κίνησης ή με ενός B εικονοπλαίσιο που έχει δυο διανύσματα κίνησης. Επομένως απαιτείται μια ενιαία μορφή αναπαράστασης του συνόλου των διανυσμάτων ανεξάρτητα την κατηγορία του εικονοπλαισίου και το είδος της πρόβλεψης του διανύσματος. Οι τρεις κατηγορίες εικονοπλαισίων με τα αντίστοιχα διανύσματα τους αναπαριστώνται στις παρακάτω εικόνες.



Εικόνα (2.9.1): Η εικονοπλαίσιο δεν παρουσιάζονται διανύσματα κίνησης,



Εικόνα (2.9.2): Η εικονοπλαίσιο, με κίτρινο χρώμα : "backward predicted", κόκκινο: "forward predicted", μαύρο: "bidirectional predicted",



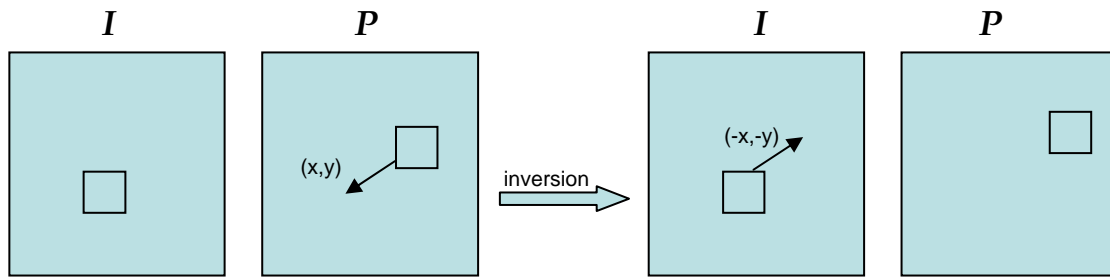
Εικόνα (2.9.3): Η εικονοπλαίσιο, μόνο forward predicted διανύσματα κίνησης

Στην προσέγγιση που έχει υλοποιηθεί [22] αναπαριστάται κάθε εικονοπλαίσιο με διάνυσμα κίνησης σε σχέση με ένα επόμενο εικονοπλαίσιο ανεξάρτητα από το

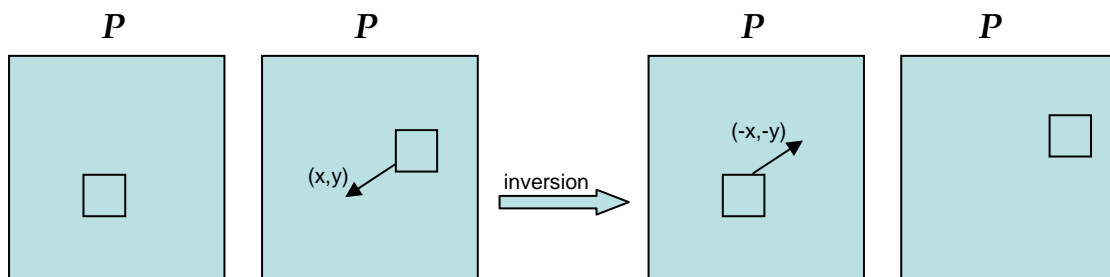
είδος τους. Έτσι κάθε εικονοπλαίσιο έχει διανύσματα των MB που δείχνουν την απόκλιση τους σε σχέση τα MB του επόμενου εικονοπλαισίου ανεξάρτητα αυτά πως είχαν κωδικοποιηθεί.

2.9.1 Υπολογισμός κανονικοποιημένων διανυσμάτων μεταξύ I και P εικονοπλαισίων

Στην προσέγγιση στην οποία αναφερόμαστε παραπάνω αυτό που χρειάζεται να γίνει είναι να υπολογιστεί ένα διάνυσμα για κάθε MB εικονοπλαισίου που θα δείχνει την απόκλιση του από το αντίστοιχο MB του επόμενου εικονοπλαισίου ανεξάρτητα από το είδος τους και το πώς αυτά σχετίζονται μεταξύ τους στην διαδικασία της κωδικοποίησης. Σε μια ακολουθία εικονοπλαισίων όπου υπάρχουν μονάχα P εικονοπλαίσια μεταξύ των I, μπορεί για κάθε MB εικονοπλαισίου να εξαχθεί ένα διάνυσμα “backward predicted” σε σχέση με το επόμενο του. Αυτό μπορεί να συμβεί για όλα τα MB των I και P εικονοπλαισίων πέραν του τελευταίου P εικονοπλαισίου το οποίο δεν έχει καμία σχέση με το επόμενο του I, αυτή η περίπτωση θα αναλυθεί στην επόμενη ενότητα όπου έχουμε και B εικονοπλαίσια. Έτσι μια πρώτη περίπτωση είναι όταν υπάρχει ένα I εικονοπλαίσιο που ακολουθείται από ένα P, τότε σαν διάνυσμα “backward predicted” του MB του I σε σχέση με του P είναι το “forward predicted” διάνυσμα του P σε σχέση με το I το οποίο έχει υπολογιστεί κατά την κωδικοποίηση, αντεστραμμένο. Συνεπώς αν σε ένα MB στο P εικονοπλαίσιο περιγράφεται η απόκλιση του από το αντίστοιχο στο προηγούμενο I με το διάνυσμα (x,y) είναι λογικό το ίδιο MB στο I εικονοπλαίσιο να έχει διάνυσμα “backward predicted” που δείχνει την απόκλιση του από το επόμενο P το αντίστροφο διάνυσμα $(-x,-y)$, σχήμα 2.9.1.1. Η ίδια ακριβώς διαδικασία ακολουθείται μεταξύ MB διαδοχικών P εικονοπλαισίων όπως φαίνεται και στο σχήμα.



Σχήμα 2.9.1.1: Ροή των διανυσμάτων μεταξύ του I εικονοπλαισίου και του επόμενου P εικονοπλαισίου



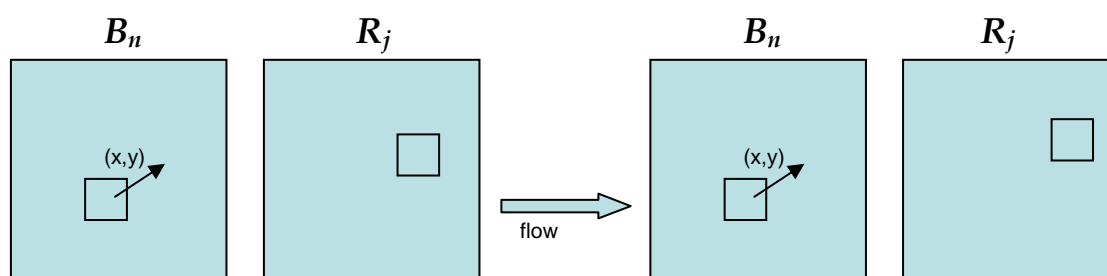
Σχήμα 2.9.1.2: Ροή των διανυσμάτων μεταξύ του P εικονοπλαισίου και του επόμενου P εικονοπλαισίου

2.9.2 Υπολογισμός κανονικοποιημένων διανυσμάτων μεταξύ B εικονοπλαισίων

Η προσπάθεια κανονικοποίησης των διανυσμάτων στην περίπτωση των B εικονοπλαισίων είναι αρκετά πιο περίπλοκη από την διαδικασία που περιγράφηκε στην προηγούμενη ενότητα [22].

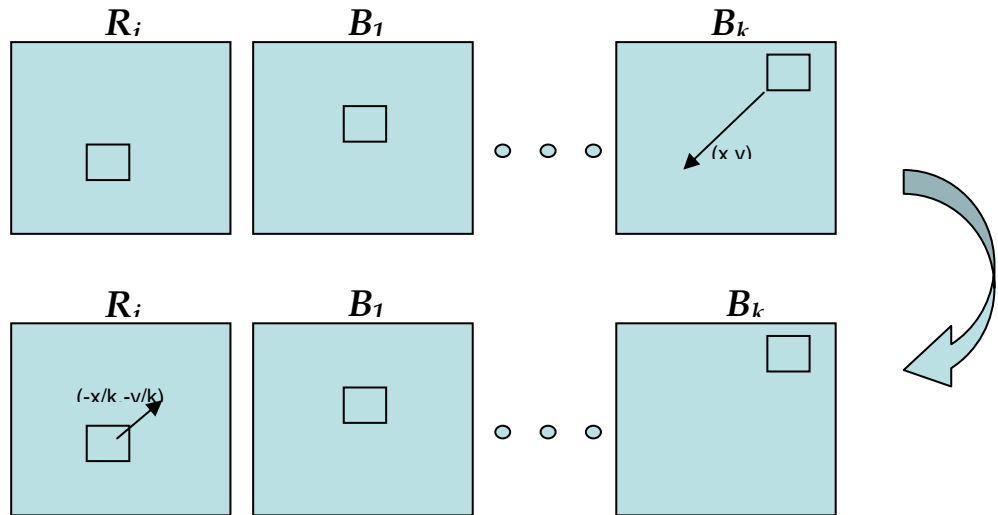
Ας θεωρήσουμε δυο εικονοπλάγια αναφοράς R_i και R_j τα οποία μπορεί να είναι I ή P και B τα εικονοπλάγια μεταξύ τους B_1, \dots, B_n , όπου n είναι ο αριθμός των B εικονοπλαισίων (τυπικά $n=2$).

Αρχικά εξάγονται τα διανύσματα μεταξύ του πρώτου εικονοπλαισίου αναφοράς R_i και του επόμενου B_l , χρησιμοποιώντας τα “forward predicted” διανύσματα κίνησης του B_l σε σχέση με το R_i . Η περίπτωση αυτή είναι όμοια με τα I-P εικονοπλαίσια που αναλύθηκε παραπάνω. Η αντιστροφή των “forward predicted” διανυσμάτων κίνησης των MB του B_l σε σχέση με το R_i είναι τα κανονικοποιημένα διανύσματα του R_i σε σχέση με το B_l . Ομοίως χρησιμοποιούνται τα “backward predicted” διανύσματα κίνησης του B_n εικονοπλαισίου σε σχέση με το R_j , τα οποία εδώ δεν χρειάζεται να αντιστραφούν για να ορίσουν τη ροή των διανυσμάτων από το ένα στο επόμενο εικονοπλαίσιο, σχήμα 2.9.2.1.



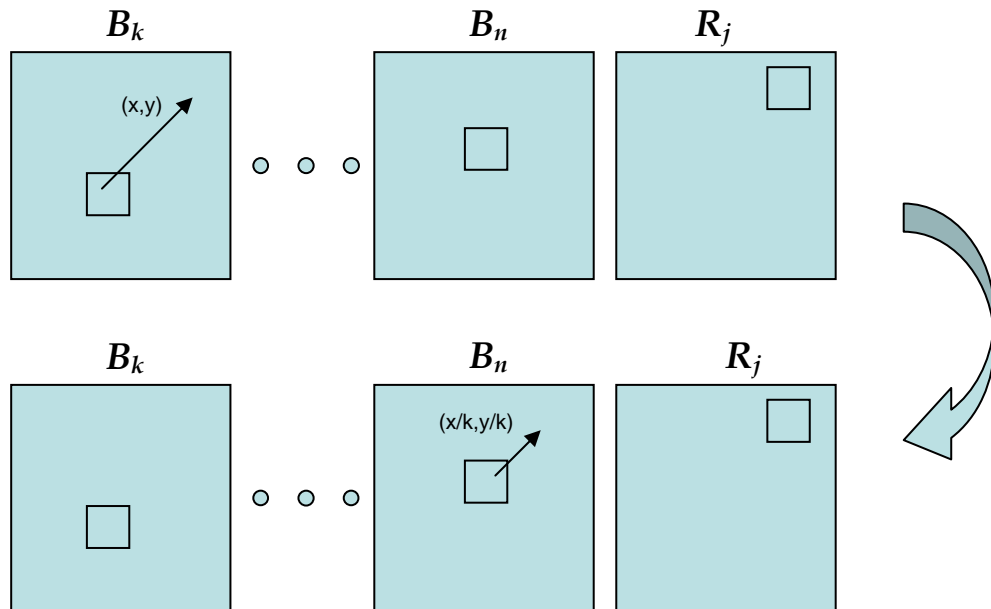
Σχήμα 2.9.2.1: Τα backward predicted διανύσματα του B_n χρησιμοποιούνται για να ορίσουν την ροή των διανυσμάτων του

Στον παραπάνω συλλογισμό δεν έχουμε θεωρήσει τις περιπτώσεις στις οποίες ένα MB στο B_l εικονοπλαίσιο δεν έχει “forward predicted” διάνυσμα κίνησης σε σχέση με το αντίστοιχο R_i , ή την περίπτωση όπου ένα MB στο B_n εικονοπλαίσιο δεν έχει “backward predicted” διάνυσμα κίνησης σε σχέση με το R_j . Στην πρώτη περίπτωση ψάχνουμε ένα επόμενο εικονοπλαίσιο πέραν του B_l , έστω το B_k , όπου το αντίστοιχο MB έχει “forward predicted” διάνυσμα κίνησης σε σχέση με το R_i και παίρνουμε το αντίστροφο διάνυσμα αυτού. Ανάλογα με το πόσα εικονοπλαίσια μπροστά βρέθηκε το διάνυσμα αυτό, έστω k , διαιρούμε το διάνυσμα με αυτόν τον παράγοντα k . **σχήμα 2.9.2.2** Αν δεν μπορούμε να βρούμε τέτοιο B εικονοπλαίσιο θεωρούμε ότι η ροή των διανυσμάτων κίνησης σε σχέση με το επόμενο εικονοπλαίσιο είναι μη καθορίσιμη.



Σχήμα 2.9.2.2: Η ροή των διανυσμάτων του R_i εξάγεται από τα forward predicted διανύσματα του B_k

Ομοίως και στην δεύτερη περίπτωση ψάχνουμε στα εικονοπλάισια που προηγούνται του B_k μέχρι να βρεθεί το αντίστοιχο MB το οποίο θα έχει “backward predicted” διάνυσμα κίνησης σε σχέση με το R_j . Στην συνέχεια διαιρείται το διάνυσμα αυτό με τον αριθμό που δείχνει ποσά εικονοπλάισια κινηθήκαμε πίσω για να βρεθεί “backward predicted” διάνυσμα κίνησης, σχήμα 2.9.2.3.



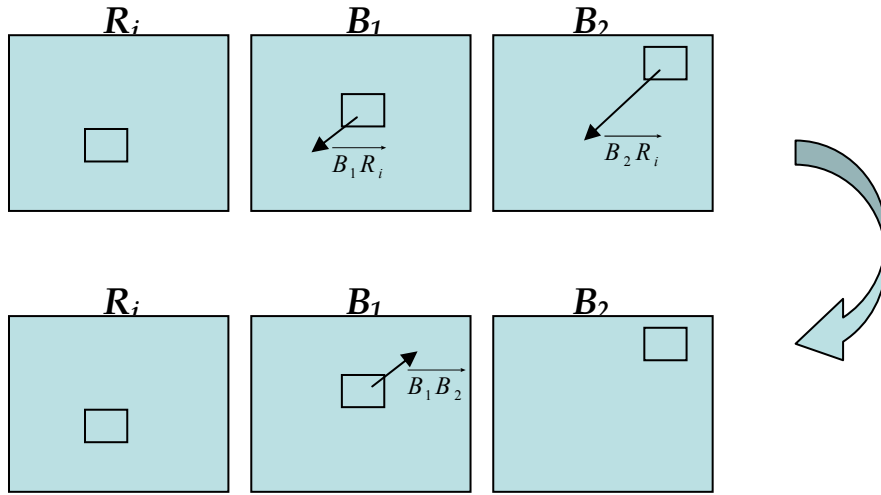
Σχήμα 2.9.2.3: Η ροή των διανυσμάτων του B_n εξάγεται από τα backward predicted διανύσματα του B_k

Το επόμενο κομμάτι στην διαδικασία της κανονικοποίησης είναι να οριστεί η ροή των διανυσμάτων κίνησης μεταξύ διαδοχικών B εικονοπλαισίων. Θεωρώντας ότι κάθε MB σε ένα B εικονοπλαίσιο μπορεί να έχει κωδικοποιηθεί με βάση μια από τις τρεις παρακάτω κατηγορίες “forward predicted” (F), “backward predicted” (B) ή “bidirectionally predicted” (D), υπάρχουν εννέα πιθανοί συνδυασμοί για το πώς εμφανίζεται ένα διάνυσμα σε ένα MB από το B_1 στο επόμενο B_2 εικονοπλαίσιο. Οι εννέα αυτοί συνδυασμοί παρουσιάζονται σε ζευγάρια FF, FB, FD, BF, BB, BD, DF, DB, και DD.

Ας θεωρήσουμε ότι το “forward predicted” διάνυσμα κίνησης ενός συγκεκριμένου MB στο B_1 εικονοπλαίσιο σε σχέση με το εικονοπλαίσιο αναφοράς R_i είναι το $\overrightarrow{B_1 R_i}$, και το αντίστοιχο στο B_2 είναι το $\overrightarrow{B_2 R_i}$. Στην συνέχεια ορίζουμε το διάνυσμα του MB στο B_1 σε σχέση με το B_2 το $\overrightarrow{B_1 B_2}$, η σχέση που προκύπτει είναι

$$\overrightarrow{B_1 B_2} = -(\overrightarrow{B_2 R_i} - \overrightarrow{B_1 R_i})$$

Για τον υπολογισμό αυτής της ποσότητας τα μόνα που χρειάζονται είναι τα “forward predicted” διανύσματα κίνησης $\overrightarrow{B_1 R_i}$ και $\overrightarrow{B_2 R_i}$, τα οποία μπορούν να εξαχθούν στις περιπτώσεις στις οποίες τα MB έχουν τους εξής συνδυασμούς FF, FD, DF or DD σχήμα 2.9.2.4

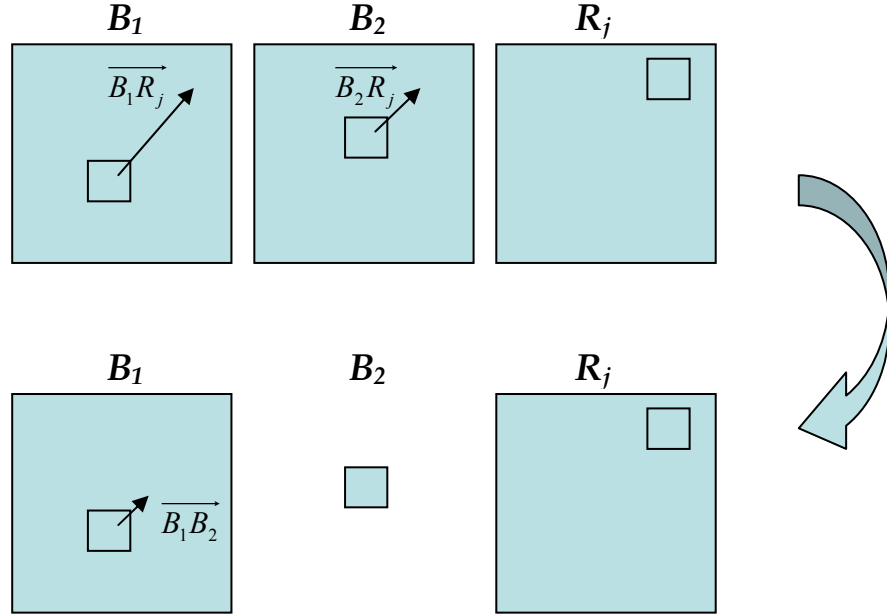


Σχήμα 2.9.2.4: Η ροή των διανυσμάτων του B_1 εξάγεται από τα *forward predicted* διανύσματα του B_1 και B_2

Αν τα ζευγάρια των MB έχουν τους συνδυασμούς BB, BD ή DB, μπορούμε να υπολογίσουμε πάλι το $\overrightarrow{B_1B_2}$ χρησιμοποιώντας τα “backward predicted” διανύσματα κίνησης του B_1 και του B_2 σε σχέση με το R_j . Θεωρώντας τώρα τα “backward predicted” διανύσματα κίνησης του MB στο B_1 το $\overrightarrow{B_1R_j}$ και αντίστοιχα του B_2 το $\overrightarrow{B_2R_j}$. Τότε έχουμε

$$\overrightarrow{B_1B_2} = \overrightarrow{B_1R_j} - \overrightarrow{B_2R_j}$$

Όπου το $\overrightarrow{B_1B_2}$ μπορεί να υπολογιστεί εύκολα όπως φαίνεται και στο σχήμα **2.9.2.5**



Σχήμα 2.9.2.5: Η ροή των διανυσμάτων του B_1 εξάγεται από τα *backward predicted* διανύσματα του B_1 και B_2

Οι μόνες υποπεριπτώσεις που δεν καλύφθηκαν είναι όταν έχουμε FB και BF. Στην περίπτωση του FB, το διάνυσμα $\overrightarrow{B_1B_2}$ είναι μη καθορίσιμο, γιατί αυτός ο συνδυασμός δηλώνει ότι μεταξύ των δυο B εικονοπλαισίων υπάρχει ένα Cut, αφού το MB στο B_1 κωδικοποιείται με βάση μόνο ένα προηγούμενο εικονοπλαίσιο αναφοράς ενώ στο αμέσως επόμενο εικονοπλαίσιο B_2 το MB κωδικοποιείται με βάση μόνο ένα επόμενο εικονοπλαίσιο αναφοράς, άρα δεν υπάρχει και συσχέτιση μεταξύ τους

Στην υποπερίπτωση BF, θα χρειαστεί να υπολογιστούν τα διανύσματα $\overrightarrow{R_iB_1}$ $\overrightarrow{B_2R_i}$ και χρησιμοποιώντας την παρακάτω σχέση τους να υπολογιστεί το

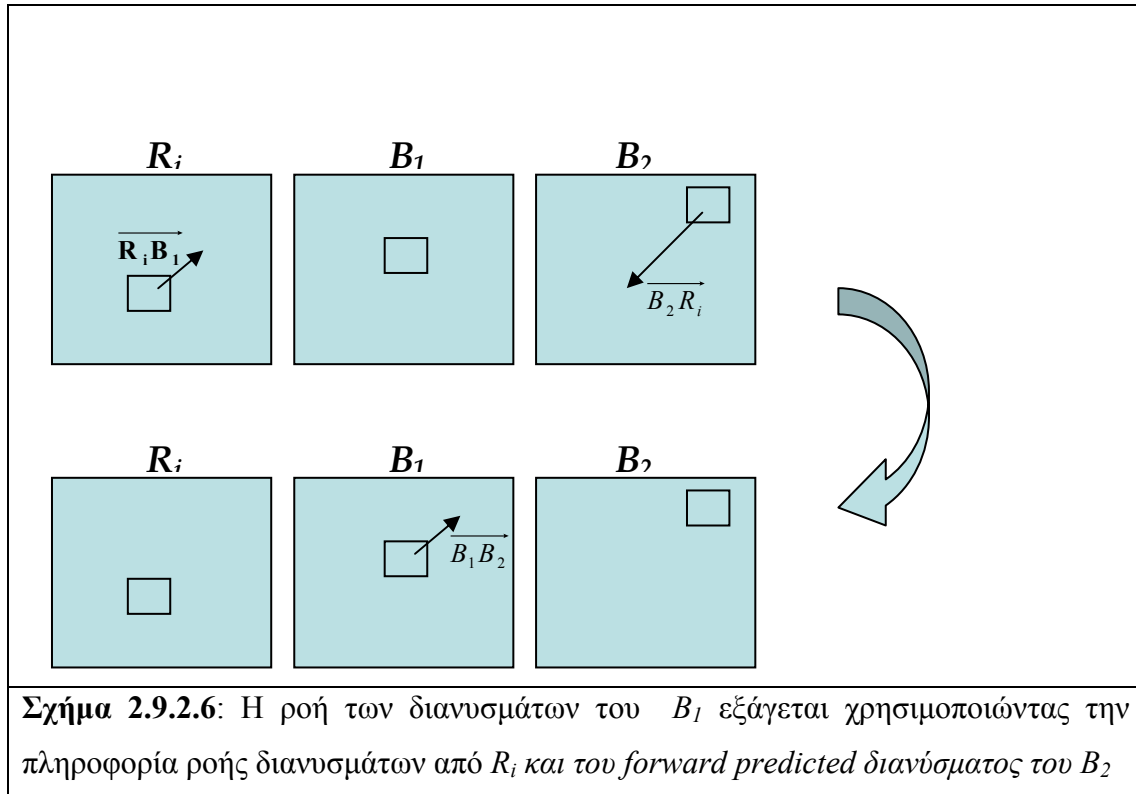
$$\overrightarrow{B_1B_2} = -(\overrightarrow{R_iB_1} + \overrightarrow{B_2R_i})$$

Αρχικά ένα MB στο B_1 εικονοπλαίσιο κωδικοποιείται σε σχέση με το επόμενο εικονοπλαίσιο αναφοράς το R_i και προκύπτει το διάνυσμα $\overrightarrow{B_1R_i}$ που δεν μας βοηθάει στον υπολογισμό των παραπάνω διανυσμάτων. Για τον υπολογισμό του $\overrightarrow{R_iB_1}$ διανύσματος χρειάζεται να εφαρμοστεί η τεχνική στην οποία ψάχνουμε για ένα επόμενο εικονοπλαίσιο του B_1 όπου βρίσκεται το συγκεκριμένο MB και έχει “forward predicted” διάνυσμα, το οποίο το αντιστρέφουμε και το διορθούμε με τον συντελεστή κ όπως παρουσιάστηκε και παραπάνω. Στην συνέχεια χρησιμοποιούμε

το “forward predicted” διάνυσμα $\overrightarrow{B_2 R_i}$ του MB του B_2 σε σχέση με το R_i , και τέλος υπολογίζεται το

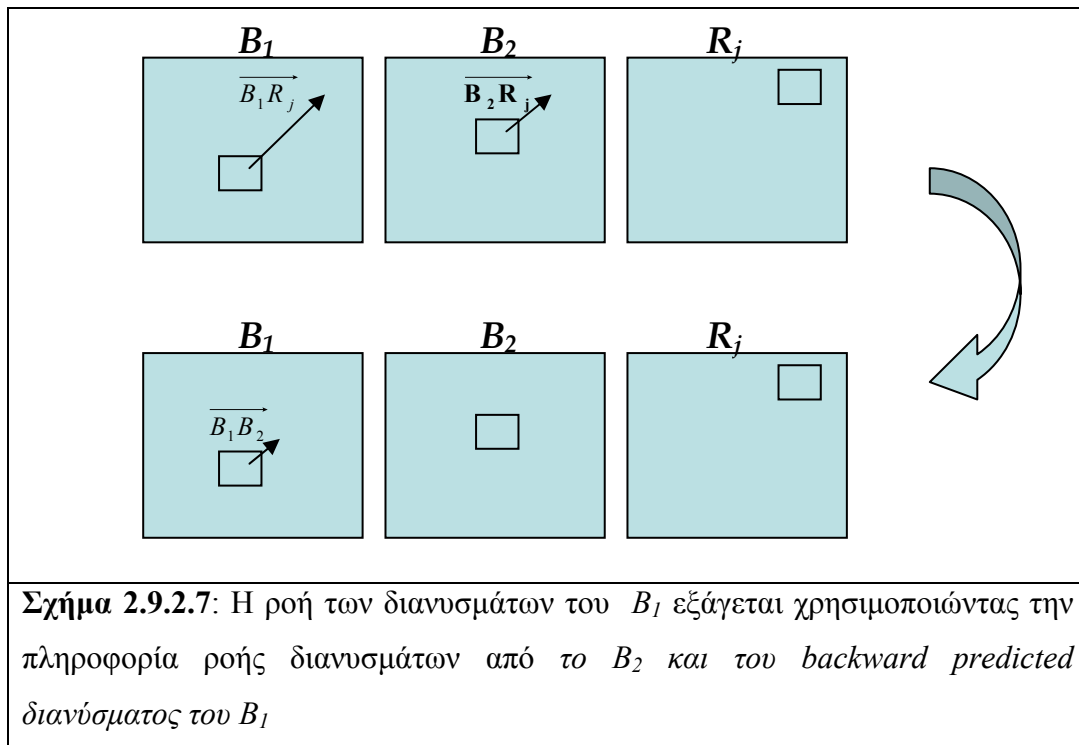
$$\overrightarrow{B_1 B_2} = -(\overrightarrow{R_i B_1} + \overrightarrow{B_2 R_i})$$

Η διαδικασία φαίνεται στο σχήμα 2.9.2.6

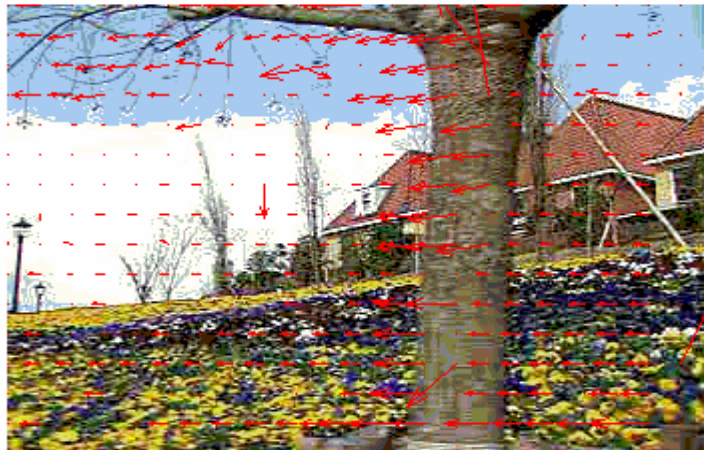


Ομοίως στην υποπερίπτωση BF το διάνυσμα $\overrightarrow{B_2 R_j}$ προσδιορίζεται με την ίδια τεχνική που υπολογίστηκε στην προηγούμενη περίπτωση το $\overrightarrow{R_i B_1}$. Για τον υπολογισμό του $\overrightarrow{B_1 R_j}$ χρησιμοποιούμε το “backward predicted” διάνυσμα του MB του B_1 σε σχέση με το R_j , και τέλος υπολογίζεται το $\overrightarrow{B_1 B_2}$ από την παρακάτω σχέση.

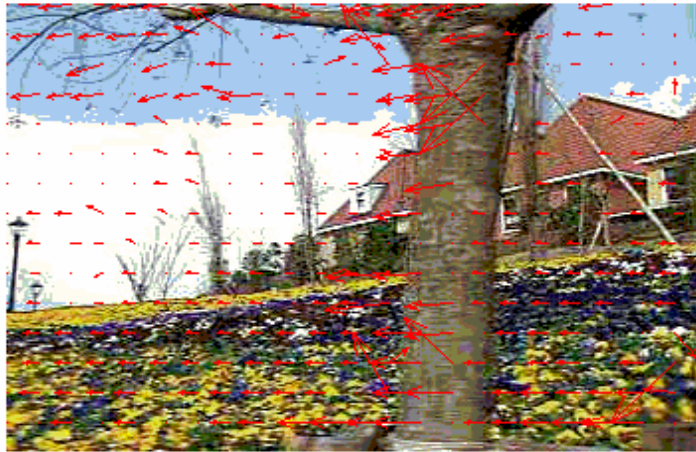
$$\overrightarrow{B_1 B_2} = \overrightarrow{B_1 R_j} - \overrightarrow{B_2 R_j}$$



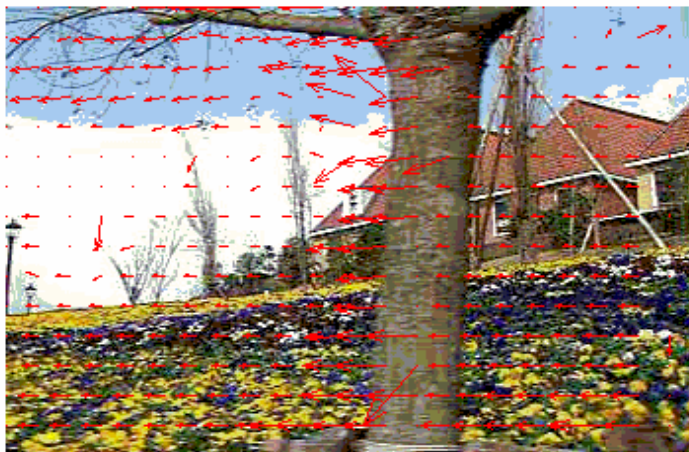
Με την εφαρμογή των μεθόδων κανονικοποίησης των διανυσμάτων η πληροφορία από την κίνηση αναπαριστάται με μια ενιαία μορφή για κάθε εικονοπλαίσιο. Παρακάτω εμφανίζονται τα τρία διαφορετικά εικονοπλαίσια (I,P,B) με τα κανονικοποιημένα τους διανύσματα.



Εικόνα (2.9.2.1): Αποτελέσματα των κανονικοποιημένων διανυσμάτων I εικονοπλαισίου



Εικόνα (2.9.2.2): Αποτελέσματα των κανονικοποιημένων διανυσμάτων B εικονοπλαισίου



Εικόνα (2.9.2.3): Αποτελέσματα των κανονικοποιημένων διανυσμάτων P εικονοπλαισίου

Κεφάλαιο 3

Μελέτη Μεθόδων Κατωφλίωσης

3.1 Εισαγωγή στην μελέτη μεθόδων κατωφλίωσης

Ένα από τα πρωταρχικά θέματα μελέτης σε μεθόδους τμηματοποίησης είναι η διαδικασία σωστού υπολογισμού κατωφλίου. Οι μέθοδοι υπολογισμού κατωφλίου έχουν σκοπό να υπολογίσουν κάποιες τιμές όρια με τις οποίες θα μπορούμε να διαχωρίσουμε τυχόν μεταβολές στο εξεταζόμενο σήμα. Σε μια τέτοια διαδικασία υπεισέρχονται αρκετά λανθασμένα αποτελέσματα κυρίως σε σήματα με ακανόνιστες και συνεχώς μεταβαλλόμενες τιμές. Στις περισσότερες μελέτες τμηματοποίησης ένα μεγάλο ποσοστό απώλειας σωστών αποτελεσμάτων οφείλεται στην αδυναμία των μεθόδων να υπολογίσουν την κατάλληλη τιμή κατωφλίου (thresholding).

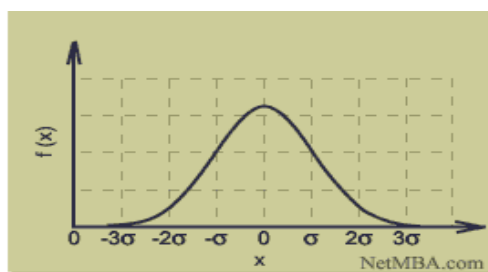
Η βασική διαδικασία με την οποία αναλύεται και τμηματοποιείται ένα βίντεο στην συγκεκριμένη διπλωματική εργασία εξαρτάται άμεσα από την ικανότητα των μεθόδων κατωφλίωσης να προσαρμόζονται στο εκάστοτε σήμα το οποίο περιγράφει κάποιο χαρακτηριστικό του βίντεο, όπως για παράδειγμα η διάφορα ιστογραμμάτων έντασης δυο διαδοχικών εικονοπλαισίων ή αντίστοιχα η διασπορά των διανυσμάτων κίνησης τους. Οι μέθοδοι αυτοί δίνουν ένα σύνολο τιμών (τιμές κατωφλίου) για ένα συγκεκριμένο σήμα και κάνοντας σύγκριση του σήματος με την αντίστοιχη τιμή κατωφλίου εξάγονται κάποια συμπεράσματα για τις μεταβολές στο περιεχόμενο του βίντεο. Συνεπώς είναι απαραίτητη η μελέτη των μεθόδων αυτών επικεντρώνοντας στα πλεονεκτήματα αλλά και τις αδυναμίες τους. Οι μέθοδοι αυτοί επιγραμματικά είναι :

- | | |
|--|---------------------------------|
| α) Μέθοδος Διπλής Σύγκρισης | (Twin Comparison) |
| β) Μέθοδος Διπλής Σύγκρισης σε Παράθυρο | (Twin Comparison Window) |
| γ) Μέθοδος Παραθύρου Ολίσθησης | (Sliding Window) |
| δ) Προσαρμοστική Μέθοδος | (Adaptive) |

Πριν την λεπτομερή περιγραφή και ανάλυση των παραπάνω μεθόδων είναι σημαντικό για την κατανόηση της λειτουργίας τους μια μικρή αναφορά στην έννοια της τιμής κατωφλίου στο συγκεκριμένο πρόβλημα .

Στην προσπάθεια ανίχνευσης μεταβολών σε μια σειρά εικονοπλαισίων παρατηρούνται μεταβολές σε χαρακτηριστικά αυτών όπως για παράδειγμα το ιστόγραμμα της έντασης του εικονοπλαισίου το οποίο μεταβάλλεται κατά πολύ σε σύγκριση με το διαδοχικό επόμενο του ή η διασπορά των γωνιών των διανυσμάτων κίνησης η οποία έχει αντίστοιχα υποστεί κάποια σημαντική μεταβολή. Οι μεταβολές αυτές όμως δεν είναι αποτέλεσμα μόνο της αλλαγής κίνησης της κάμερας ή κάποιου οπτικού εφέ που υπάρχει στο βίντεο αλλά και του θορύβου που υπεισέρχεται. Οι κύριες μορφές θορύβου στο βίντεο είναι ο θόρυβος από τη συσκευή σύλληψης του βίντεο και από την ψηφιοποίηση του αναλογικού σήματος του βίντεο. Συνεπώς το πρόβλημα που θέτεται είναι να βρεθεί μια τιμή κατωφλίου που δεν θα επηρεάζεται από το θόρυβο θεωρώντας τον μεταβολή και εν συνεχεία να διαχωρίζει την απότομη μεταβολή από την σταδιακή μεταβολή.

Θεωρώντας ότι όλες οι παραπάνω μορφές θορύβου περιγράφονται σαν Gaussian θόρυβος και ότι τα τυχαία δείγματα μεταβάλλονται ακολουθώντας την κανονική κατανομή (ή Gaussian distribution), ισχύει ότι αν σ είναι η τυπική απόκλιση και μ είναι η μέση τιμή των τιμών του δειγματικού χώρου (χαρακτηριστικά του βίντεο: διαφορά ιστογραμμάτων έντασης, διασπορά γωνιών διανυσμάτων κίνησης, μεταβολή του πλήθους των intracoded macroblocks και αλλά)



Σχήμα 3.1.1: Καμπύλη κανονικής κατανομής , $f(x)$ συχνότητα εμφάνισης

Τότε γνωρίζουμε ότι

- Το 68% των δεδομένων έχουν τιμή κάτω από το άθροισμα $\mu + \sigma$
- Το 95% των δεδομένων έχουν τιμή κάτω από το άθροισμα $\mu + 2\sigma$
- Σχεδόν όλα τα δεδομένα (99,7%) έχουν τιμή κάτω από το άθροισμα $\mu + 3\sigma$

Για να ανιχνεύσουμε τις μεταβολές θα πρέπει να αναζητήσουμε ανάμεσα στο πολύ μικρό ποσοστό που διαφοροποιείται από το σύνολο των τιμών.

Συνεπώς ,στο συγκεκριμένο πρόβλημα , αν θέσουμε τιμή κατωφλιού την τιμή

$$T_a = \mu + \alpha\sigma \quad (3.1.1)$$

Όπου α = σταθερά τιμή οριζόμενη από τον χρηστή και παίρνει τιμή από 2.5 έως 4

σ = τυπική απόκλιση

μ = μέση τιμή

οποιαδήποτε τιμή ξεπερνά αυτό το όριο θεωρείται μεταβολή του περιεχομένου του βίντεο.

3.2 Μέθοδος Διπλής Σύγκρισης - *Twin Comparison method*

Η μέθοδος “Διπλής Σύγκρισης” (Twin Comparison) είναι μια μέθοδος που προτάθηκε από τον **Zhang** [7] και βασίζεται στον υπολογισμό δυο τιμών κατωφλίων μιας υψηλής τιμής T_b , που πέραν αυτής ανιχνεύονται απότομες μεταβολές του σήματος και μια μικρότερη τιμή κατωφλίου T_s , έτσι ώστε μεταξύ των δυο αυτών τιμών να ανιχνεύονται οι σταδιακές μεταβολές.

Αντίστοιχα οι τιμές κατωφλίων υπολογίζονται από τους τύπους

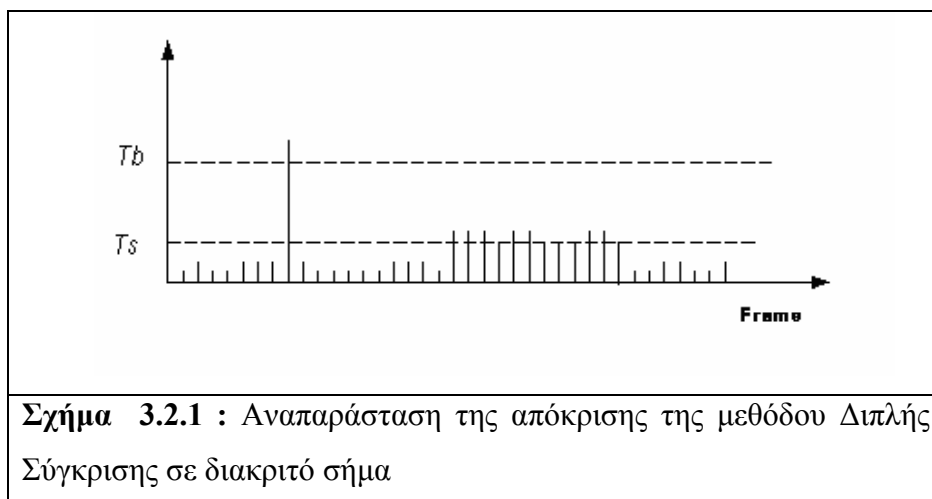
$$T_b = \mu + a_t \cdot \sigma \quad (3.2.1)$$

$$\text{και} \quad T_s = \beta \cdot \mu \quad (3.2.2)$$

η μεταβλητή β ορίζεται από τον χρήστη και παίρνει την τιμή από 1.5 έως 2. Ενώ οι υπόλοιπες μεταβλητές περιγράφηκαν παραπάνω στο *Κεφάλαιο 3.1*.

Η βασικότερη διάφορα της μεθόδου Διπλής Σύγκρισης από τις μεθόδους που θα περιγραφούν παρακάτω, είναι ότι η πρώτη χρειάζεται προεπεξεργασία ολόκληρου του σήματος για να δώσει τις δυο τιμές κατωφλίων. Αυτό έχει σαν αποτέλεσμα να καθυστερεί να δώσει κάποιο αποτέλεσμα και δεν μπορεί να χρησιμοποιηθεί σε εφαρμογές πραγματικού χρόνου. Παρόλα αυτά δίνει πολύ καλά αποτελέσματα κυρίως στην ανίχνευση απότομων μεταβολών.

Παρακάτω φαίνεται πως λειτουργεί η μέθοδος στο σήμα



Αναφορά στα βασικά βήματα της διαδικασίας

Ο υπολογισμός των κατωφλίων γίνεται αρκεί να υπολογιστούν οι τιμές της μέσης τιμής του σήματος μ και της τυπικής απόκλισης σ , ενώ στη συνέχεια υπολογίζονται οι δυο τιμές κατωφλίων $T_b = \mu + a_t \cdot \sigma$ και $T_s = \beta \cdot \mu$

Οι τιμές του σήματος πέραν του T_b ορίζονται σαν απότομες μεταβολές τους σήματος ενώ οι τιμές του σήματος μεταξύ των T_s και T_b ορίζονται σαν σταδιακές.

Παρακάτω παρουσιάζονται κάποιες γραφικές παραστάσεις που απεικονίζουν την απόκριση της μεθόδου Διπλής Σύγκρισης στο σήμα της διαφοράς των ιστογραμμάτων έντασης διαδοχικών εικονοπλαισίων.

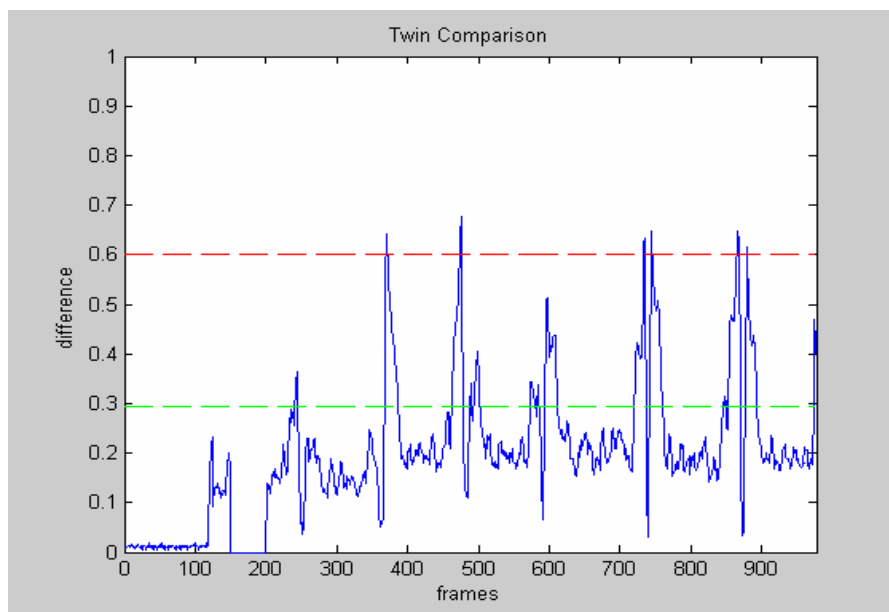


Figure (3.2.1)

Οι τιμές κατωφλίων T_b (κόκκινο), T_s (πράσινο) στην εφαρμογή της μεθόδου Διπλής Σύγκρισης σε σήμα διαφοράς ιστογραμμάτων έντασης.

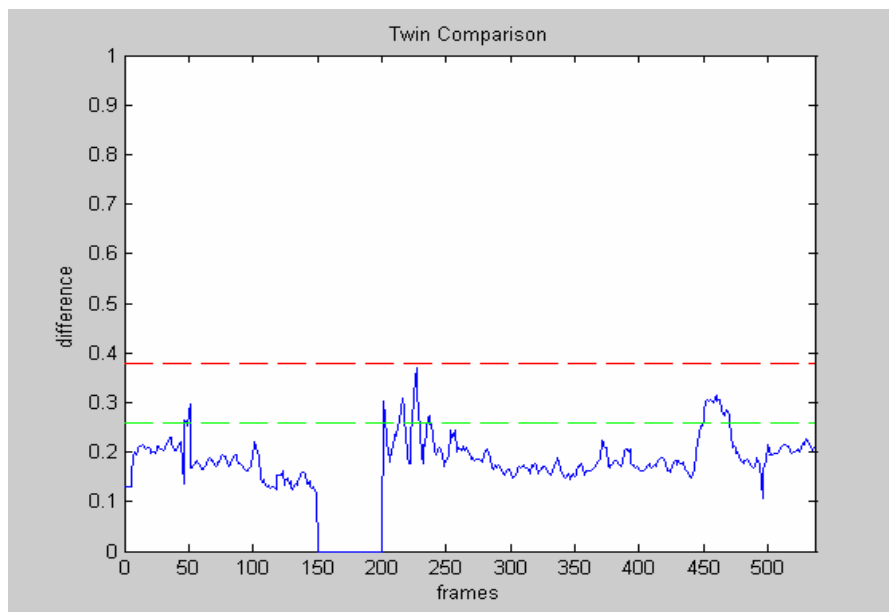


Figure (3.2.2)

Οι τιμές κατωφλίων T_b (κόκκινο), T_s (πράσινο) στην εφαρμογή της μεθόδου Διπλής Σύγκρισης σε σήμα διαφοράς ιστογραμμάτων έντασης.

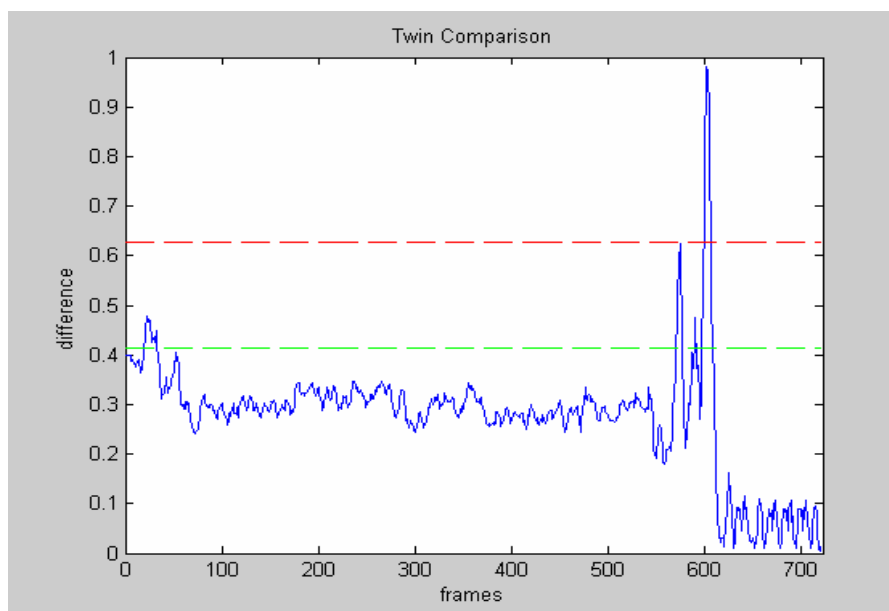


Figure (3.2.3)

Οι τιμές κατωφλίων T_b (κόκκινο), T_s (πράσινο) στην εφαρμογή της μεθόδου Διπλής Σύγκρισης σε σήμα διαφοράς ιστογραμμάτων έντασης

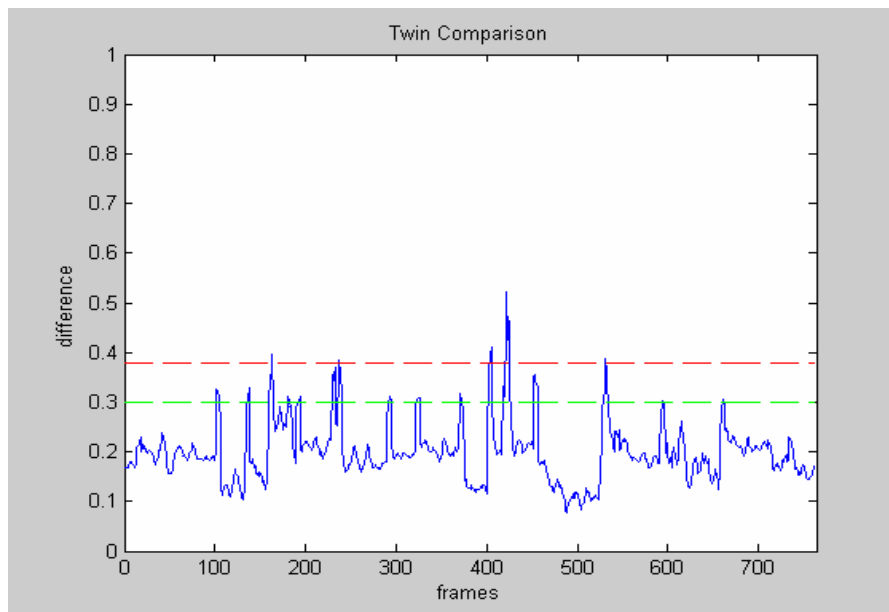


Figure (3.2.4)

Οι τιμές κατωφλίων T_b (κόκκινο), T_s (πράσινο) στην εφαρμογή της μεθόδου Διπλής Σύγκρισης σε σήμα διαφοράς ιστογραμμάτων έντασης

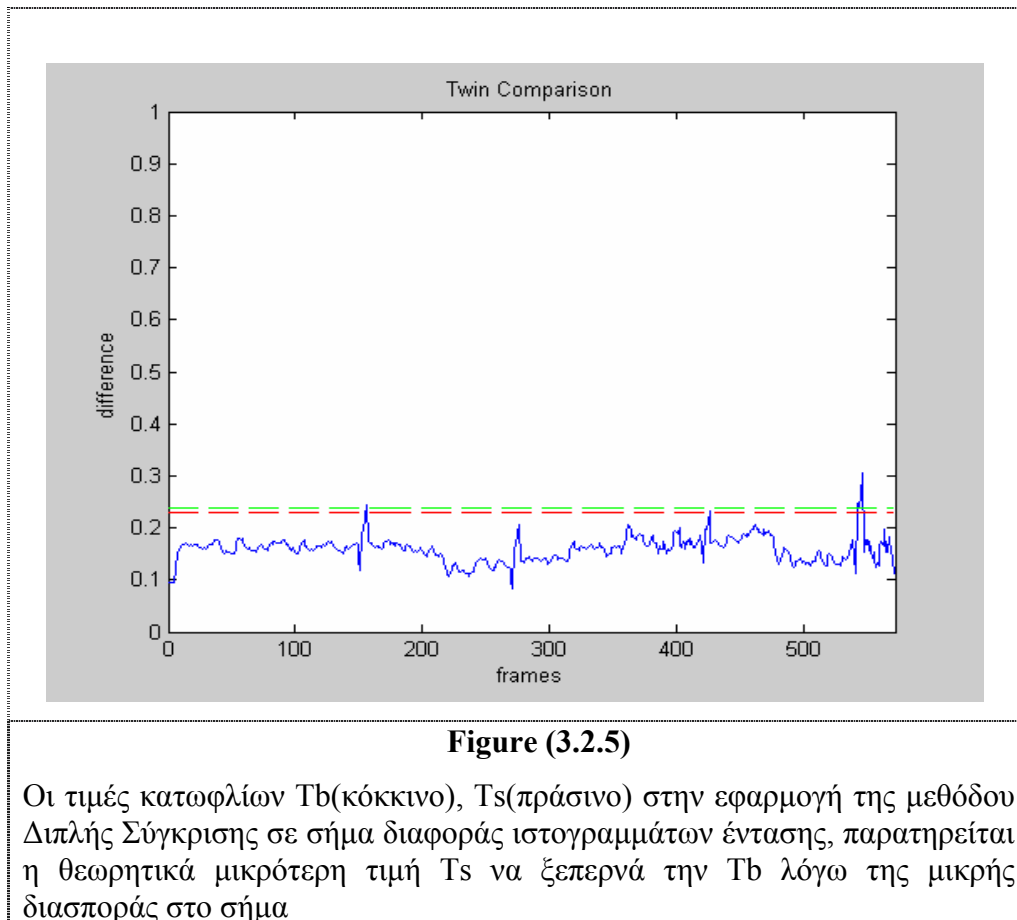
Από τις παραπάνω γραφικές φαίνεται πως η μέθοδος Twin Comparison λειτουργεί πολύ αποτελεσματικά στην πλειοψηφία των σημάτων. Πολύ πιθανόν βέβαια σε ένα αρκετά μεγαλύτερο βίντεο, (πχ 10.000 frames) όπου εναλλάσσεται κατά πολύ η δράση του περιεχομένου του, να μην δίνει τόσο αντικειμενικά αποτελέσματα αφού η τιμή του κατωφλίου θα παραμένει σταθερή σε ένα σύνολο από διαφορετικής δράσης και κινητικότητας βίντεο. Ένα ακόμα πρόβλημα υπάρχει όταν το σήμα έχει πολύ μικρή διασπορά, δηλαδή όχι πολλές και μεγάλες αλλαγές, τότε η χαμηλή τιμή του κατωφλίου $T_s = \beta \cdot \mu$ ξεπερνά την μεγαλύτερη τιμή κατωφλίου

$$T_b = \mu + \alpha \cdot \sigma$$

$$\text{Δηλαδή όταν } Ts > Tb \rightarrow \beta \mu > \mu + \alpha \sigma \rightarrow (\beta - 1) \mu > \alpha \sigma$$

$$\beta = 1.5, \alpha = 4 \text{ σταθερά. Άρα όταν } 0.5 \mu > 4 \sigma \rightarrow \mu > 8 \sigma$$

Έτσι προκαλείται μια σύγχυση για το ποια τιμή θα είναι το όριο σύγκρισης για τις απότομες και αντίστοιχα τις σταδιακές μεταβολές. Πιο κάτω φαίνεται μια τέτοια περίπτωση, όπου το όριο T_s ξεπερνά το T_b



Συμπερασματικά η μέθοδος της Διπλής Σύγκρισης :

1. Λειτουργεί πιο αποτελεσματικά από όλες τις μεθόδους κατωφλίωσης για την ανίχνευση μεταβολών σε συνεχές σήμα
2. Χρειάζεται προεπεξεργασία ολόκληρου του σήματος για των υπολογισμό των ποσοτήτων της μέσης τιμής του σήματος μ και της τυπικής απόκλισης σ , συνεπώς δεν μπορεί να λειτουργήσει σε εφαρμογές πραγματικού χρόνου αφού χρειάζεται να ολοκληρωθεί η χρονοσειρά εικόνων (βίντεο) για να μπορούν να εξαχθούν αποτελέσματα
3. Είναι πολύ αποδοτικό κυρίως στην ανίχνευση απότομων μεταβολών του σήματος
4. Σε σήματα μικρής διασποράς παρατηρείται, η χαμηλή τιμή κατωφλίου $T_s = \beta \cdot \mu$ να ξεπερνάει την θεωρητικά υψηλότερη τιμή κατωφλίου $T_b = \mu + a_t \cdot \sigma$

3.3 Μέθοδος Διπλής Σύγκρισης σε παράθυρο - Twin Comparison Window method

Η Μέθοδος Διπλής Σύγκρισης σε παράθυρο ([20] Zhnag) είναι μια μέθοδος η οποία συνδυάζει τον τρόπο λειτουργίας της Διπλής Σύγκρισης και Παραθύρου Ολίσθησης μεθόδου και αποσκοπεί στην υψηλή αποδοτικότητα της μεθόδου Διπλής Σύγκρισης και την προσαρμοστικότητα στις μεταβολές του σήματος που δίνουν οι μέθοδοι με παράθυρο.

Πιο συγκεκριμένα εφαρμόζεται η μέθοδος της Διπλής Σύγκρισης σε ένα παράθυρο με 15 τιμές και εκεί υπολογίζονται η μέση τιμή του σήματος μ και η τυπική απόκλιση σ στη συνέχεια υπολογίζονται οι δυο τιμές κατωφλίου $T_b = \mu + a_t \cdot \sigma$ και $T_s = \beta \cdot \mu$

για τις τιμές που βρίσκονται στο παράθυρο, με $a=3$ και $\beta=1,5$. Το παράθυρο μετακινείται μια θέση δεξιά βάζοντας στο παράθυρο την νέα τιμή σήματος και ξαναυπολογίζονται οι τιμές κατωφλίων, αυτή η διαδικασία συνεχίζει μέχρι το τέλος του σήματος.

Περιγραφικά ο αλγόριθμος συνοψίζεται στα εξής βήματα :

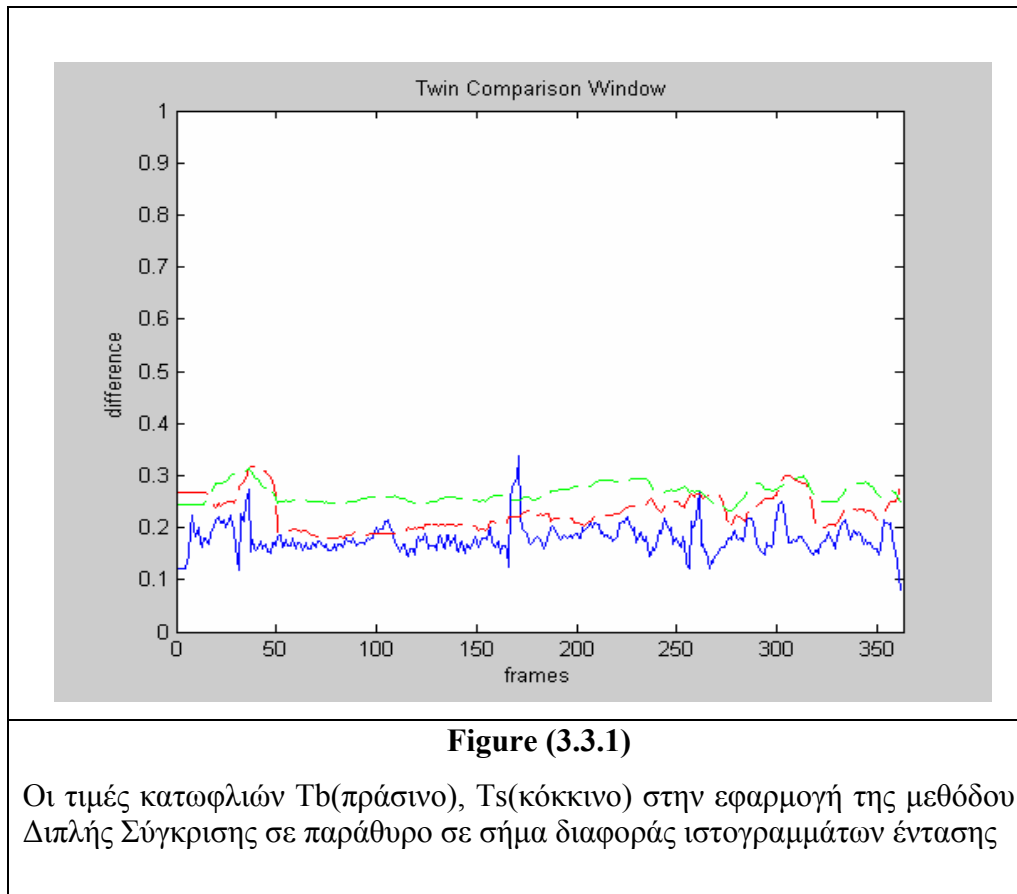
1. Κατασκευάζεται το παράθυρο με X (ενδεικτικά 15) τιμές του δείγματος
2. Υπολογίζονται οι ποσότητες μέση τιμή του σήματος μ και η τυπική απόκλιση σ
3. Με τις τιμές από το βήμα 2 υπολογίζονται οι δυο τιμές κατωφλίων

$$T_b = \mu + a_t \cdot \sigma \quad \text{και} \quad T_s = \beta \cdot \mu \quad \text{όπου } a=3 \text{ και } \beta=1.5$$

ομοίως οι τιμές του σήματος πέραν του T_b ορίζονται σαν απότομες μεταβολές τους σήματος ενώ οι τιμές του σήματος μεταξύ των T_s και T_b ορίζονται σαν σταδιακές

4. Μετακινείται το παράθυρο μια θέση δεξιά και υπολογίζονται οι νέες τιμές των βημάτων 2 και 3 ,επαναλαμβάνεται το ίδιο μέχρι το τέλος του σήματος.

Πιο κάτω φαίνονται κάποιες γραφικές παραστάσεις της απόκρισης της μεθόδου
Διπλής Σύγκρισης σε παράθυρο σε διαφορετικά σήματα



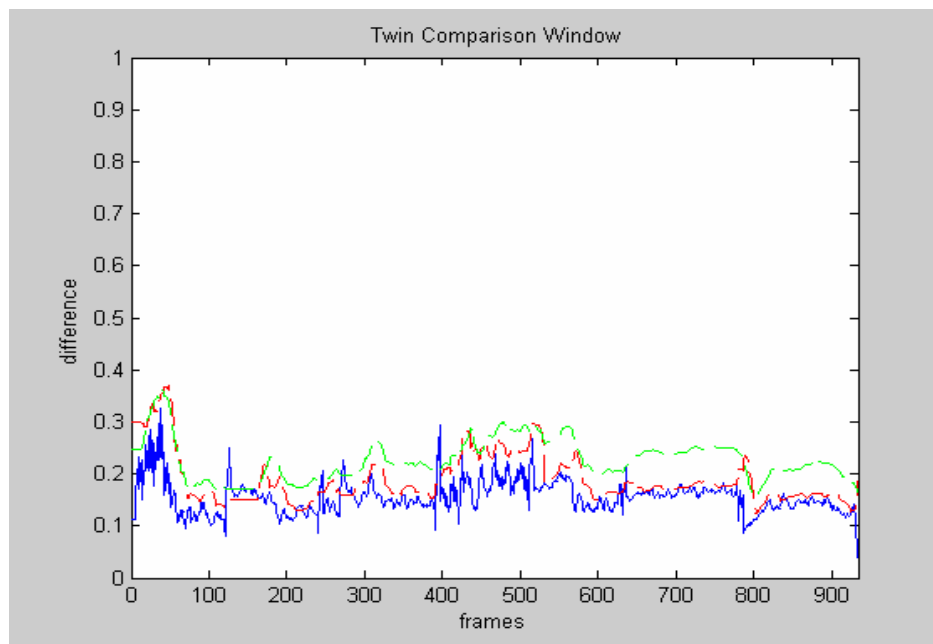


Figure (3.3.2)

Οι τιμές κατωφλιών T_b (πράσινο), T_s (κόκκινο) στην εφαρμογή της μεθόδου Διπλής Σύγκρισης σε παράθυρο σε σήμα διαφοράς ιστογραμμάτων έντασης

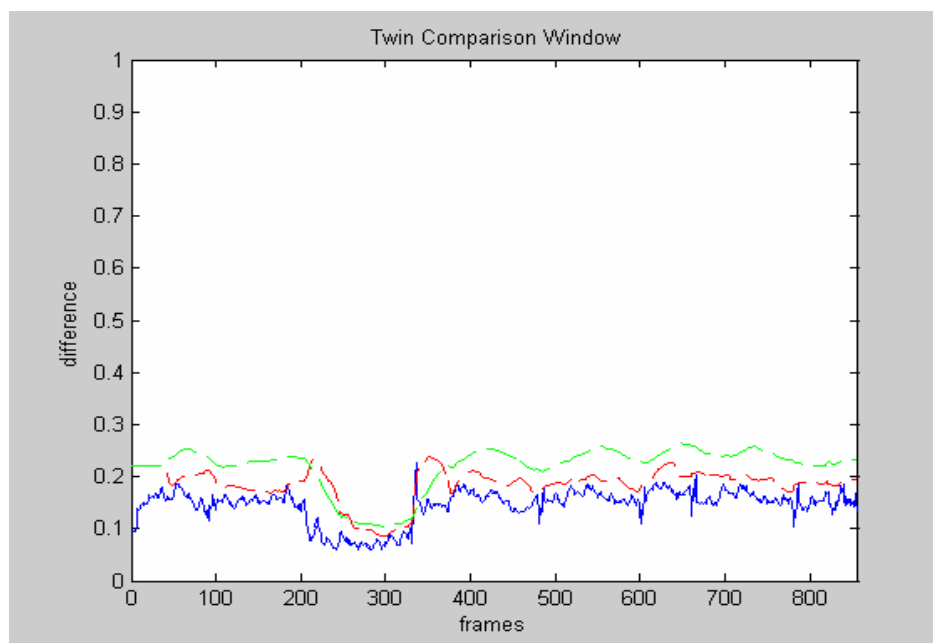
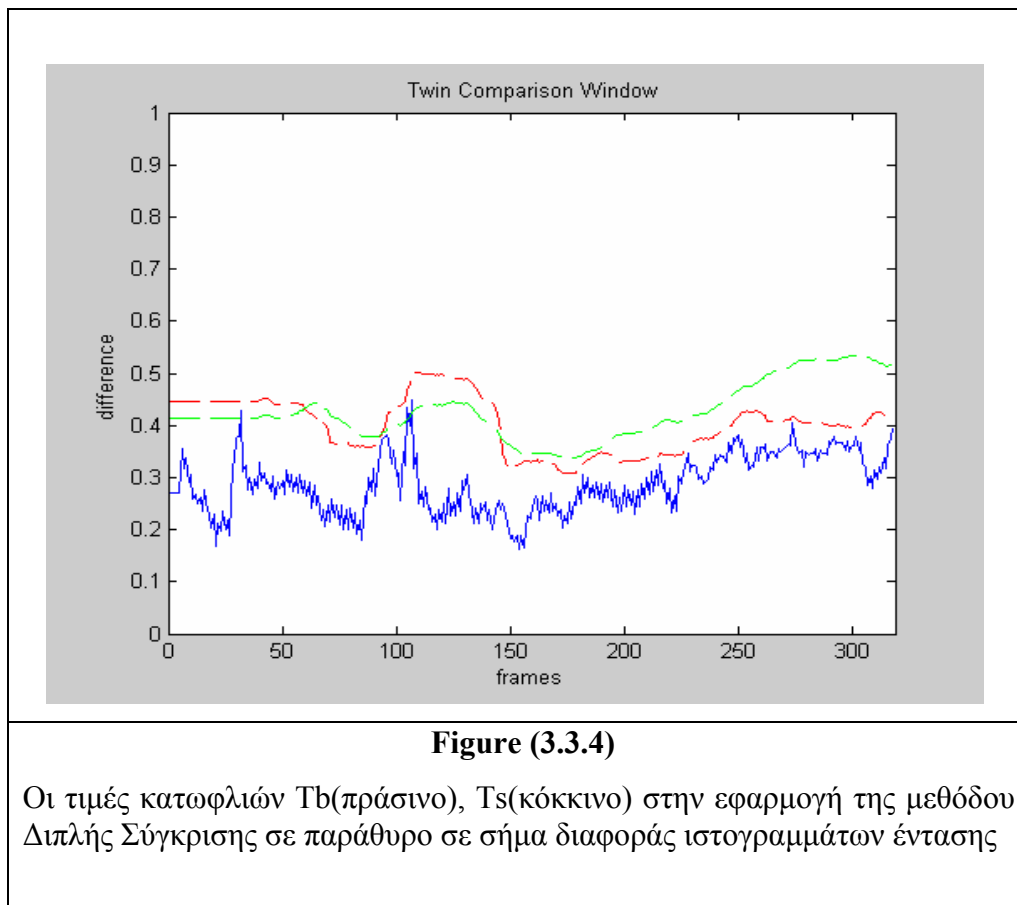


Figure (3.3.3)

Οι τιμές κατωφλιών T_b (πράσινο), T_s (κόκκινο) στην εφαρμογή της μεθόδου Διπλής Σύγκρισης σε παράθυρο σε σήμα διαφοράς ιστογραμμάτων έντασης



Συμπερασματικά η μέθοδος Διπλής Σύγκρισης σε παράθυρο:

- 1 Είναι μια συνδυαστική μέθοδο Διπλής Σύγκρισης και Παραθύρου Ολίσθησης
2. Συναγωνίζεται σε απόδοση την μέθοδο Διπλής Σύγκρισης και πλεονεκτεί αυτής διότι δεν χρειάζεται προεπεξεργασία ολόκληρου του σήματος συνεπώς μπορεί να χρησιμοποιηθεί και σε εφαρμογές πραγματικού χρόνου. Ενώ συγχρόνως δίνει πιο αντικειμενικά αποτελέσματα και σε σήματα μεγάλης χρονικής διάρκειας αφού δεν αντιλαμβάνεται το σήμα σαν ένα ενιαίο σύνολο δίνοντας μια σταθερή τιμή για ολόκληρο το σήμα με την οποία θα γίνει ανίχνευση μεταβολών ,πράγμα που κάνει η μέθοδο Διπλής Σύγκρισης
3. Το μέγεθος του παραθύρου συμπεριλαμβάνει 15 τιμές οι οποίες δεν είναι αρκετές για να θεωρηθεί ότι είναι ένας δειγματικός χώρος που μεταβάλλεται ακολουθώντας την κανονική κατανομή. Με βάση όμως αυτήν την υπόθεση υπολογίζονται οι τιμές

κατωφλιών συνεπώς ίσως θα ήταν τυπικά σωστότερο να αυξηθεί το μέγεθος του παραθύρου σε X τιμές με $X \gg 15$.

4. Προσαρμόζεται σε τοπικές μεταβολές του σήματος που μπορεί να οφείλονται και σε μεταβολή περιεχομένου. Αναμένεται να αποδώσει καλύτερα σε βίντεο με αρκετές αλλαγές περιεχομένου (animations, ειδήσεις και διαφημίσεις).

3.4 Μέθοδος Παραθύρου Ολίσθησης - *Sliding Window method*

Οι μέθοδοι με παράθυρο γενικά χρησιμοποιούν ένα υποσύνολο διαδοχικών τιμών του σήματος που περικλείονται σε ένα 'παράθυρο'. Εκεί υπολογίζεται μια τιμή κατωφλίου και κάθε φορά υπεισέρχεται η επόμενη τιμή σήματος η οποία θα συνεισφέρει ή όχι στον υπολογισμό μια νέας τιμής κατωφλίου. Αυτό έχει σαν αποτέλεσμα η τιμή κατωφλίου να προσαρμόζεται άμεσα σε τυχόν μεταβολές του σήματος.

Συγκεκριμένα κατασκευάζεται ένα παράθυρο με μέγεθος 15 τιμών. Το μέγεθος προέκυψε από την πειραματική μελέτη ότι η τιμή 15 δίνει αρκετά ικανοποιητικά αποτελέσματα στο σύνολο των δειγμάτων. Έτσι το κατώφλι $Tsl_w(i)$, η μέση τιμή $\mu sl_w(i)$ και η τυπική απόκλιση $\sigma sl_w(i)$ υπολογίζονται στο σύνολο των τιμών που βρίσκονται στο παράθυρο. Η παράμετρος i αντιπροσωπεύει τον αριθμό του εικονοπλαισίου που εισάγεται στο παράθυρο κάθε φορά και παίρνει τιμές από 15 έως το συνολικό αριθμό εικονοπλαισίων του βίντεο. Τοποθετείται λοιπόν το παράθυρο στην αρχή του συνόλου των δεδομένων και στην συνέχεια υπολογίζονται οι τιμές $\mu sl_w(i)$, $\sigma sl_w(i)$ σύμφωνα με τον τύπο (3.2.1)

$$Tsl_w(i) = \mu sl_w(i) + \alpha \sigma sl_w(i) \quad (3.4.1)$$

για τις πρώτες 15 τιμές, ενώ η μεταβλητή α παίρνει την τιμή $\alpha=3$ η οποία και πάλι προέκυψε από την πειραματική μελέτη ότι δίνει ικανοποιητικά αποτελέσματα.

Στη συνέχεια συγκρίνεται η τιμή $Tsl_w(i)$ με την αμέσως επόμενη τιμή έξω από το παράθυρο τιμών δηλαδή την τιμή που δίνει το $(i+15)$ εικονοπλαίσιο (δεκατοέκτο). Αν ισχύει ότι $D(i+1) > Tsl_w(i)$ τότε η τιμή του κατωφλίου παραμένει στην προηγούμενη τιμή της και το παράθυρο μετακινείται μια θέση προς τα δεξιά και επαναλαμβάνει την ίδια σύγκριση με την επόμενη τιμή εικονοπλαισίου.

Αν όμως ισχύει $D(i+1) < Tsl_w(i)$ τότε υπολογίζονται ξανά οι τιμές :

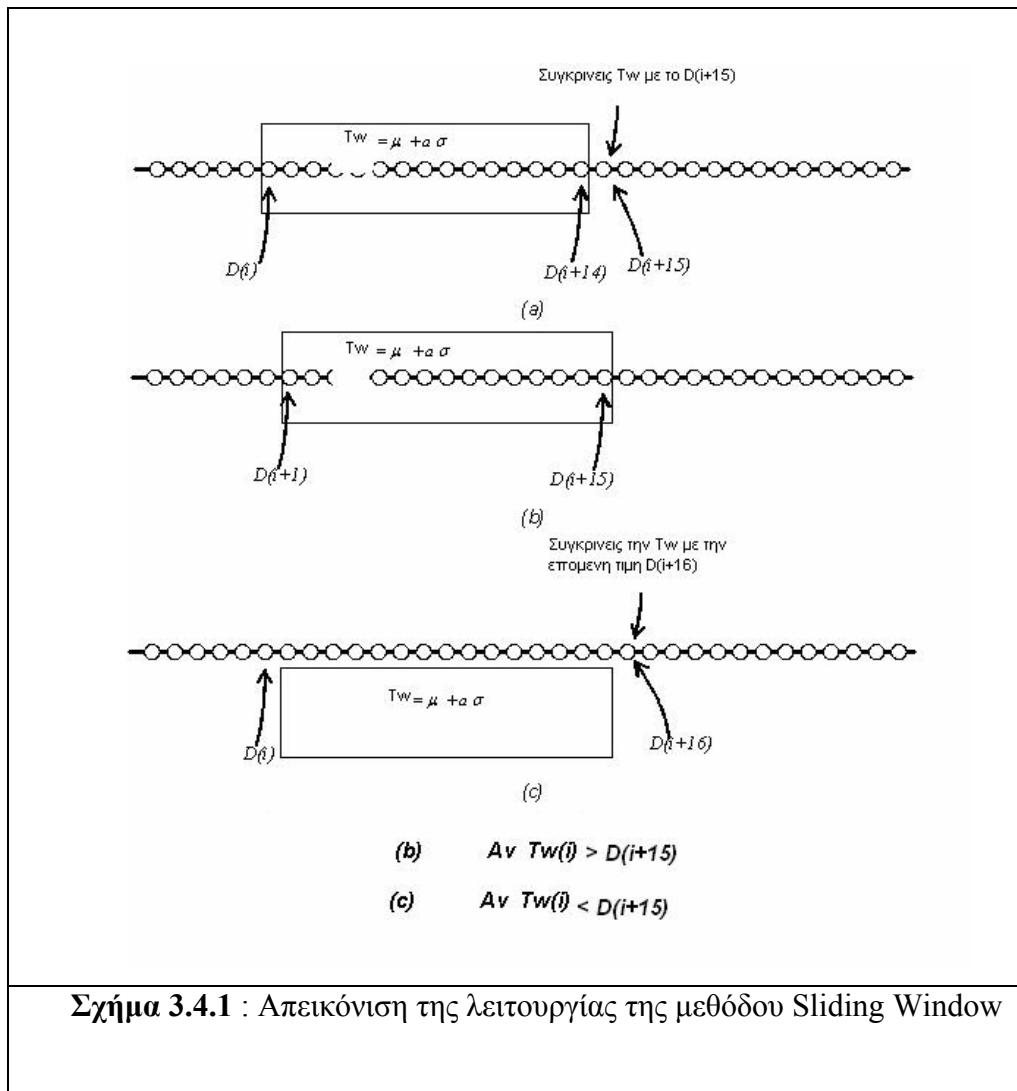
$$\mu sl_w(i), \quad \sigma sl_w(i) \text{ και τέλος η τιμή } Tsl_w(i) = \mu sl_w(i) + \alpha \sigma sl_w(i)$$

και μετακινείται πάλι το παράθυρο μια θέση προς τα δεξιά για να επαναλάβει την ίδια σύγκριση με την επόμενη τιμή εικονοστοιχείου αλλά με το καινούργιο $Tsl_w(i)$.

Πρέπει να αναφερθεί ότι στις πρώτες 15 τιμές δεν γίνονται οι συγκρίσεις όπως περιγράφονται από την παραπάνω διαδικασία αλλά μονάχα υπολογίζονται οι τιμές $\mu sl_w(i)$, $\sigma sl_w(i)$ για να πάρει αρχική τιμή το $Tsl_w(i)$.

Αναφορά στα βασικά βήματα της διαδικασίας :

1. Κατασκευάζεται το παράθυρο με τις 15 αρχικές τιμές του δείγματος
2. Υπολογίζονται οι ποσότητες μέσης τιμής $\mu sl_w(i)$ και τυπικής απόκλισης $\sigma sl_w(i)$.
3. Με τις τιμές από το βήμα 2 υπολογίζεται το $Tsl_w(i)$ βάσει του τύπου (3.4.1)
4. Γίνεται η σύγκριση του $Tsl_w(i)$ με την επόμενη τιμή έξω από το παράθυρο $D(i+1)$
5. Αν $Tsl_w(i) > D(i+1)$ τότε προσθέτεται στο σύνολο των στοιχείων του παραθύρου η τιμή $D(i+1)$ ενώ παραλείπει το πρώτο στοιχείο του παραθύρου έτσι ώστε το σύνολο να παραμένει 15. Στη συνέχεια υπολογίζονται οι νέες ποσότητες $\mu sl_w(i)$, $\sigma sl_w(i)$ και $Tsl_w(i) = \mu sl_w(i) + \alpha \sigma sl_w(i)$
6. Αν $Tsl_w(i) < D(i+1)$ τότε η νέα τιμή δεν εισέρχεται στο σύνολο των στοιχείων του παραθύρου και συνεπώς η τιμή κατωφλίου παραμένει η προηγούμενη. Το παράθυρο μετακινείται προς τα δεξιά μέχρι να βρει $D(i+1)$ ώστε να ικανοποιεί την συνθήκη του βήματος 5. Στο σημείο αυτό ορίζουμε ότι ξεκινάει μια μεταβολή. Ανάλογα τώρα την διάρκεια της, συμπεραίνουμε αν είναι απότομη ή σταδιακή μεταβολή. Για παράδειγμα αν η συνθήκη 5 ισχύει για το πολύ 3-4 συνεχόμενα εικονοπλαίσια τότε έχουμε απότομη μεταβολή αν η διάρκεια είναι μεγαλύτερη, δηλαδή περισσότερα από πέντε συνεχόμενα εικονοπλαίσια τότε έχουμε σταδιακή μεταβολή.
7. Επαναλαμβάνεται αυτή η διαδικασία για όλα τα εικονοπλαίσια του βίντεο



Πιο κάτω παραθέτονται κάποιες γραφικές παραστάσεις που δείχνουν την συμπεριφορά της μεθόδου σε σήματα διαφοράς ιστογραμμάτων έντασης διαδοχικών εικονοπλαισίων. Στις δυο πρώτες γραφικές παραστάσεις παρουσιάζεται η απόκριση της μεθόδου σε σήματα με ομαλές και μικρές μεταβολές .

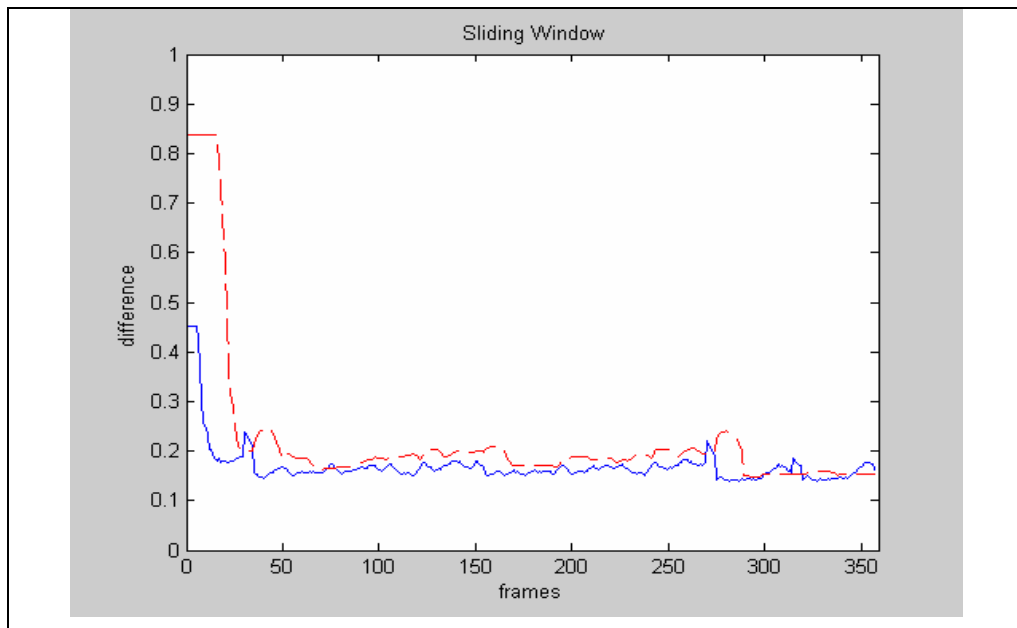


Figure (3.4.1)

Η τιμή κατωφλίου T (κόκκινο) στην εφαρμογή της μεθόδου Παραθύρου Ολίσθησης σε σήμα διαφοράς ιστογραμμάτων έντασης

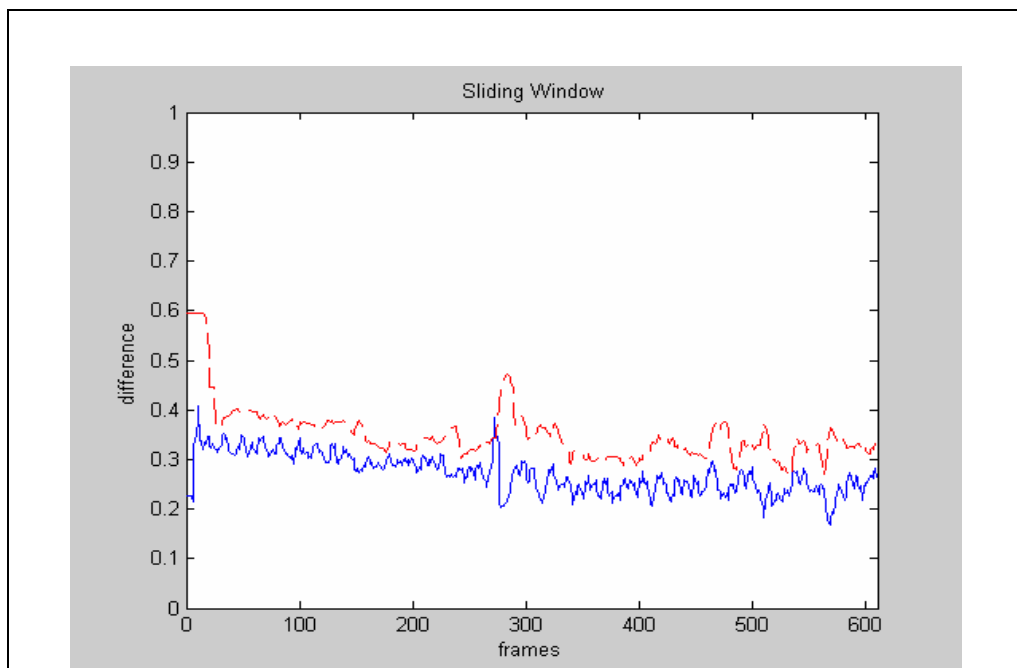


Figure (3.4.2)

Η τιμή κατωφλίου T (κόκκινο) στην εφαρμογή της μεθόδου Παραθύρου Ολίσθησης σε σήμα διαφοράς ιστογραμμάτων έντασης

Πρόβλημα προκύπτει στη μέθοδο Παραθύρου Ολίσθησης όταν το σήμα παίρνει πολύ μικρές τιμές ή μηδενικές τιμές. Τότε η τιμή του κατωφλίου σταδιακά γίνεται πολύ μικρή και δεν μπορεί να επανέλθει σε μια υψηλότερη τιμή την οποία θα είχε σε οποιαδήποτε άλλη κατάσταση. Αν παρατηρήσουμε στο βήμα 6 της μεθόδου γίνεται ένας έλεγχος της τιμής κατωφλίου με την τιμή του σήματος “ $Tsl_w(i) < D(i+1)$ ”, συνεπώς αν η τιμή κατωφλίου πέσει πολύ χαμηλά ($Tsl_w(i)$ κοντά στο μηδέν) η ανισότητα θα επαληθεύεται συνέχεια και δεν θα μπορέσει να ξεφύγει από εκείνη την τιμή. Οι δυο παρακάτω γραφικές επαληθεύουν αυτήν την παρατήρηση. Στο Figure (3) και Figure(4) υπάρχουν μηδενικές τιμές στο σήμα με αποτέλεσμα να μηδενίζεται το κατώφλι να μην επανέρχεται στις κανονικές τιμές..

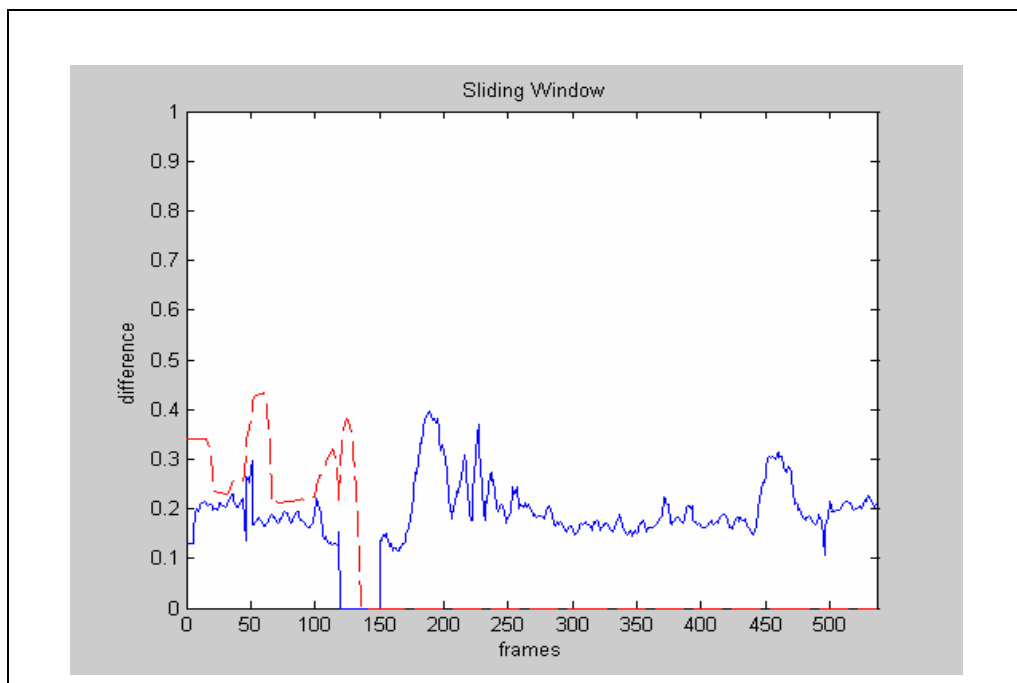


Figure (3.4.3)

Η τιμή κατωφλίου T(κόκκινο) στην εφαρμογή της μεθόδου Παραθύρου Ολίσθησης σε σήμα διαφοράς ιστογραμμάτων έντασης. Το σήμα έχει υποστεί τεχνικά το μηδενισμό του στο διάστημα 120-150 frames με αποτέλεσμα να παρατηρηθεί το πρόβλημα με την τιμή κατωφλίου.

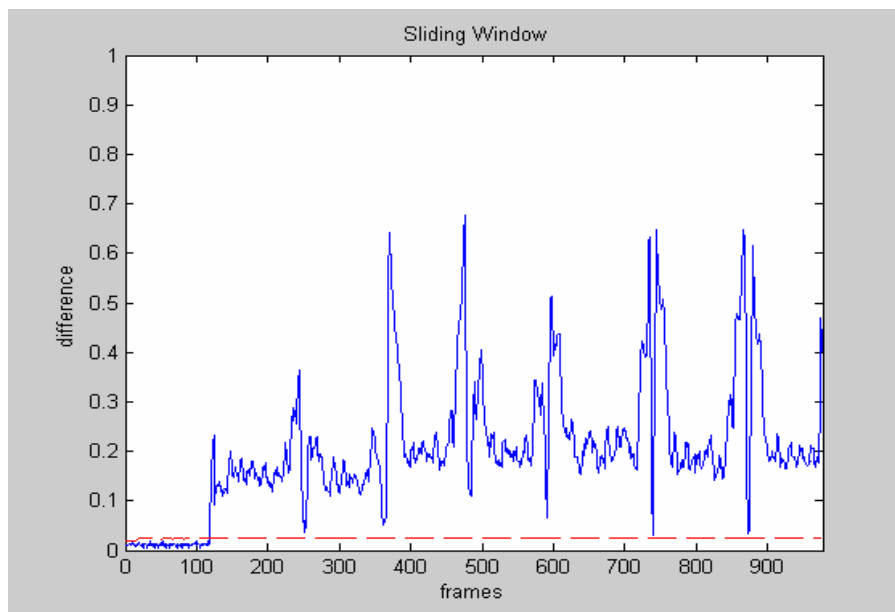


Figure (3.4.4)

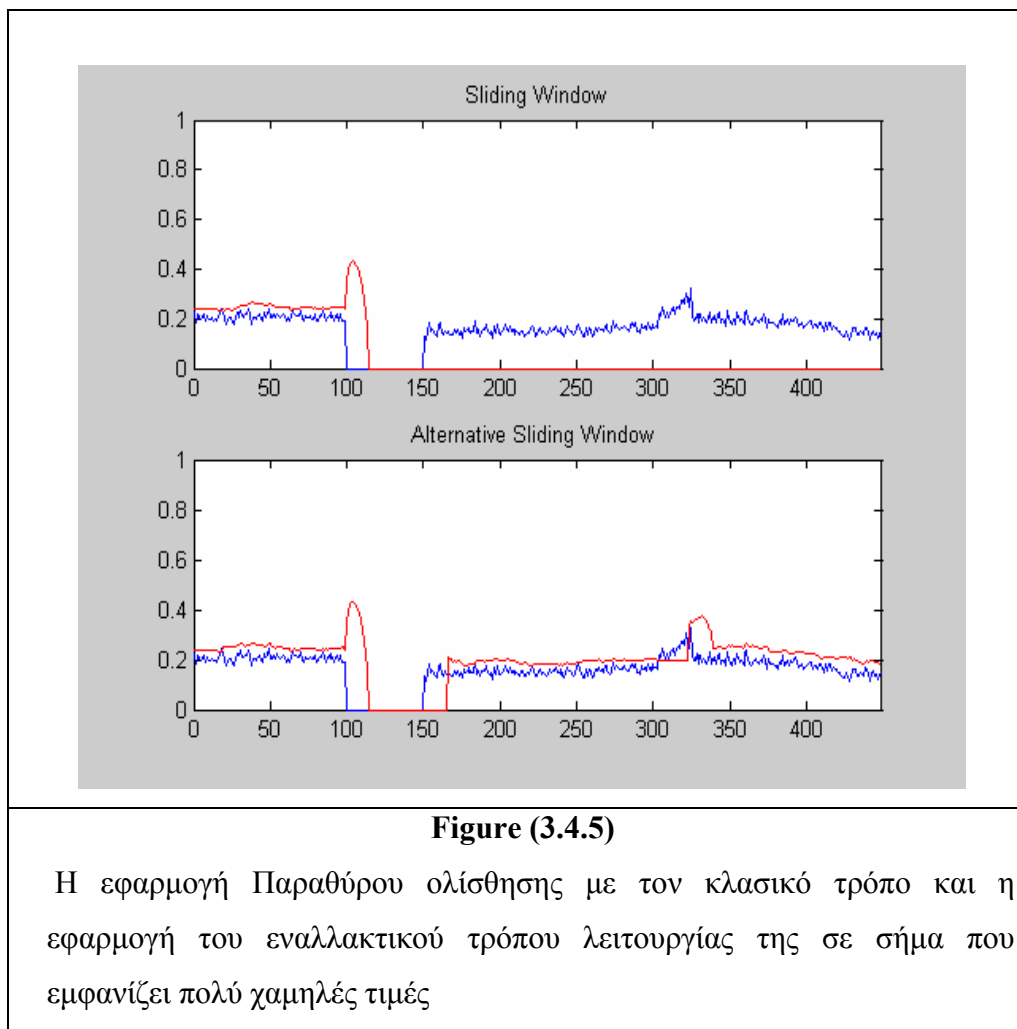
Η τιμή κατωφλίου T (κόκκινο) στην εφαρμογή της μεθόδου Παραθύρου Ολίσθησης σε σήμα διαφοράς ιστογραμμάτων έντασης. Το σήμα έχει αρχικά πολύ χαμηλές τιμές με αποτέλεσμα να παρατηρηθεί το πρόβλημα με την τιμή κατωφλίου

Το πρόβλημα του εγκλωβισμού της τιμής κατωφλίου σε τοπικά ελάχιστα αντιμετωπίζεται με ένα εναλλακτικό τρόπο λειτουργίας της μεθόδου Παραθύρου Ολίσθησης. Αρχικά στη μέθοδο στο στάδιο 6 όπου το σήμα γίνεται μεγαλύτερο από το κατώφλι, η τιμή κατωφλίου παραμένει η ίδια μέχρι να βρεθεί τιμή σήματος που θα είναι μικρότερη από αυτή και τότε θα αρχίσουν πάλι να υπολογίζονται οι τιμές κατωφλίου. Όταν η μέθοδος βρεθεί σε αυτή την κατάσταση η εναλλακτική μέθοδος προτείνει να παραμένει η τιμή κατωφλίου η ίδια απλά στο διάστημα που βρίσκεται μέσα στην μεταβολή να υπολογίζεται μια νέα τιμή $T_{internal} = \mu$, δηλαδή η μέση τιμή των τιμών του παραθύρου. Παίρνουμε την τιμή της μέσης τιμής γιατί μέσα στην μεταβολή αλλάζει έντονα η διασπορά και εμείς θέλουμε μια τιμή που θα κινείται κάτω από τις τιμές που ανοδικά μεταβάλλονται και όχι πάνω όπως γίνεται σε όλη την άλλη διαδικασία. Στην συνέχεια θα βγαίνει από την συνθήκη όταν βρεθεί τιμή σήματος που θα είναι μικρότερη από την $T_{internal}$ ή μικρότερη από την T_{sl_w} που είχε αρχικά σταματήσει να υπολογίζεται. Έξω από την μεταβολή η τιμή $T_{internal}$ έχει την ίδια τιμή T_{sl_w} . Συνεπώς η συνθήκη επαναφοράς, στάδιο 5,

γίνεται :”Αν $Tsl_w(i) > D(i+1)$ η $T_{internal} > D(i+1)$ τότε προσθέτεται στο σύνολο των στοιχείων του παραθύρου η τιμή $D(i+1)$ ενώ παραλείπει το πρώτο στοιχείο του παραθύρου έτσι ώστε το σύνολο να παραμένει 15. Στη συνέχεια υπολογίζονται οι νέες ποσότητες

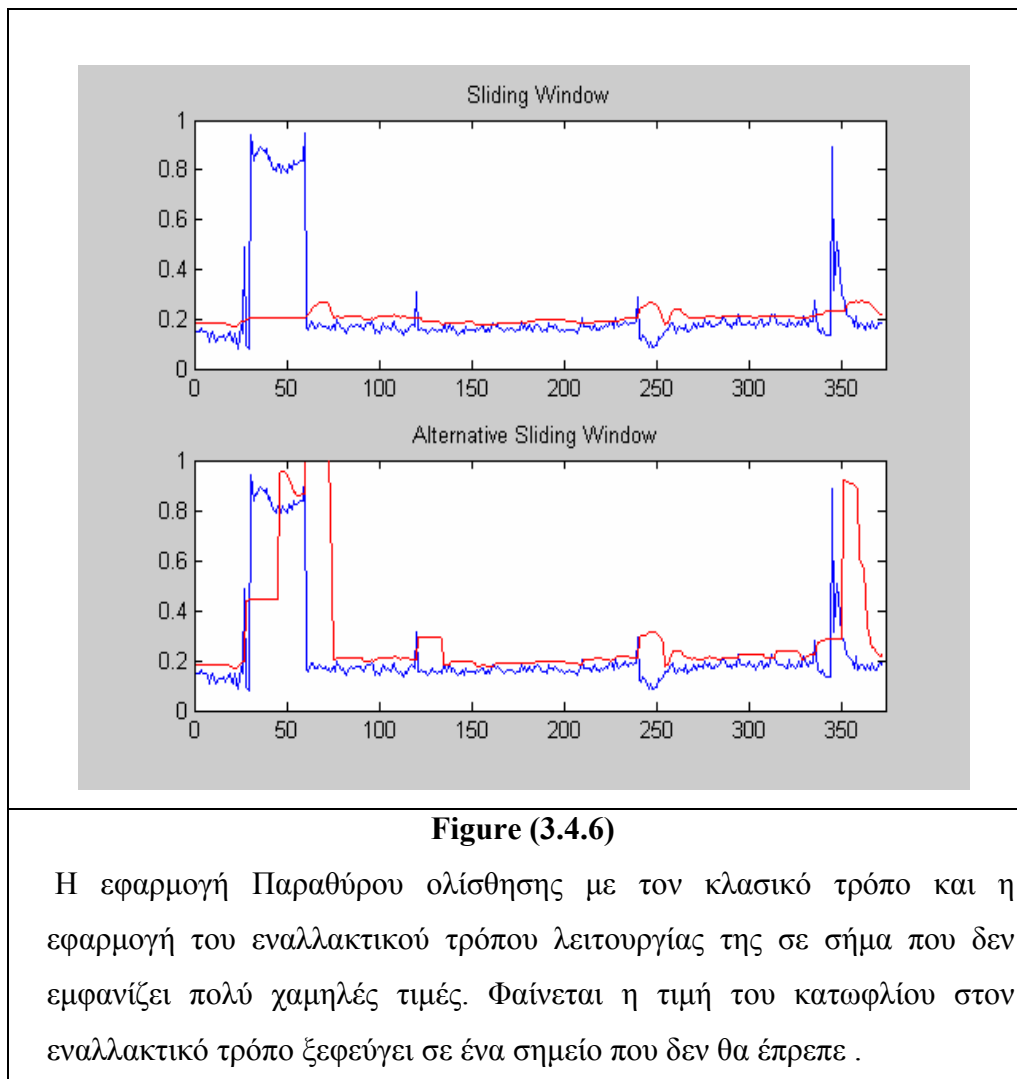
$$\mu sl_w(i) , \sigma sl_w(i) \text{ και } Tsl_w(i) = \mu sl_w(i) + \alpha \sigma sl_w(i) “$$

Πιο κάτω φαίνεται ένα παράδειγμα πως λειτουργούσε αρχικά η μέθοδος Παραθύρου Ολίσθησης σε πολύ χαμηλές τιμές και πως αντιμετωπίστηκε με την εναλλακτική μέθοδο Παραθύρου Ολίσθησης.



Η αλήθεια είναι ότι ο εναλλακτικός τρόπος λειτουργίας της μεθόδου Παραθύρου Ολίσθησης δεν λειτουργεί με μεγαλύτερη απόδοση από τον κλασικό στο σύνολο των

δειγμάτων. Αυτό οφείλεται με βάση την συνθήκη που προσθέσαμε ότι σε σήματα που δεν εμφανίζουν τόσο χαμηλές τιμές η τιμή κατωφλίου κάποιες φορές ξεφεύγει σε σημείο που δεν θα έπρεπε.



Συμπερασματικά η μέθοδος Παραθύρου Ολίσθησης :

1. Λειτουργεί αποτελεσματικά για να ανιχνεύσει μεταβολές σε ένα συνεχές σήμα και προσαρμόζεται σε τοπικές αλλαγές
2. Δεν χρειάζεται προεπεξεργασία ολόκληρου του σήματος παρά μόνο των 15 τιμών, στην συγκεκριμένη μελέτη, ή αντίστοιχα x τιμών που βρίσκονται κάθε φορά στο παράθυρο (window).

3. Δεν λειτουργεί σωστά όταν οι τιμές του σήματος έχουν πολύ χαμηλές τιμές σε σχέση με το υπόλοιπο σύνολο για μεγάλο χρονικό διάστημα (περισσότερο από 15 συνεχόμενες τιμές)
4. Χρησιμοποιεί μόνο μια παράμετρο αντί για δυο που χρειάζεται η μέθοδος Διπλής Σύγκρισης
5. Προσαρμόζεται σε τοπικές μεταβολές του σήματος που μπορεί να οφείλονται και σε μεταβολή περιεχομένου. Αναμένεται να αποδώσει καλύτερα σε βίντεο με αρκετές αλλαγές περιεχομένου(animations,ειδήσεις και διαφημίσεις).

3.5 Προσαρμοστική Μέθοδος - Adaptive method

Η Προσαρμοστική Μέθοδος (Adaptive method) όπως και η μέθοδος Παραθύρου Ολίσθησης υπολογίζει μια τιμή κατωφλίου η οποία δεν έχει κάποια σταθερή τιμή για όλο το σήμα αλλά μεταβάλλεται και προσαρμόζεται στις πιθανές αλλαγές του σήματος. Η διαδικασία υπολογισμού του κατωφλίου είναι λίγο διαφορετική από την προηγούμενη μέθοδο αλλά παρόλα αυτά διατηρεί την ίδια λογική. Δηλαδή κάθε φορά εισέρχεται μια νέα τιμή, υπολογίζονται κάποια μεγέθη, γίνεται κάποια σύγκριση και τέλος υπολογίζεται το κατωφλίου.

Στην Προσαρμοστική μέθοδο δεν υπολογίζονται οι τιμές μέσα σε ένα παράθυρο αλλά οι τιμές εμπεριέχουν αναδρομικά και τις παλιότερες. Πιο συγκεκριμένα υπολογίζονται κάθε φορά τρεις ποσότητες οι μ_a , λ_a και σ_a

όπου μ_a το προσαρμοστική μέση τιμή (adaptive mean)

λ_a η προσαρμοστική στατιστική απόκλιση (adaptive second moment)

και σ_a προσαρμοστική τυπική απόκλιση (adaptive standard deviation)

Οι τιμές των παραπάνω ποσοτήτων προκύπτουν από τους εξής τύπους

$$\mu_a(i) = \mu_a(i-1) - c \cdot (\mu_a(i-1) - D(i)) \quad (3.5.1)$$

$$\lambda_a(i) = \lambda_a(i-1) - c \cdot (\lambda_a(i-1) - (D(i))^2) \quad (3.5.2)$$

$$\sigma_a(i) = \sqrt{|\mu_a(i)^2 - \lambda_a(i)|} \quad (3.5.3)$$

$$\mu_a(1) = D(1)$$

$$\lambda_a(1) = (D(1))^2$$

$$\sigma_a(1) = \sqrt{\mu_a(1)^2 - \lambda_a(1)} = 0$$

$$\text{Τέλος η τιμή του κατωφλίου } T_a(i) = \mu_a(i) + a_a \cdot \sigma_a(i) \quad (3.5.4)$$

Η μεταβλητή c που εμφανίζεται στον τύπο 3.5.1 και 3.5.2 είναι ένας συντελεστής που δηλώνει την ευαισθησία της τιμής του κατωφλίου στις αλλαγές του σήματος και παίρνει την τιμή $c=0.05$. Ενώ η τιμή a_a είναι η ίδια σταθερά που αναφέρεται και στην μέθοδο Παραθύρου Ολίσθησης και εδώ παίρνει την τιμή $a=3$

Αρχικά για τις πρώτες 15 τιμές του σήματος υπολογίζεται το κατώφλι βάση των παραπάνω τύπων χωρίς να γίνεται σύγκριση της τιμής του κατωφλίου με την επόμενη τιμή του σήματος. Συνεπώς αφού τελειώσει ο υπολογισμός των 15 τιμών γίνεται ένας έλεγχος παρομοίως με της μεθόδου Παραθύρου Ολίσθησης, δηλαδή ελέγχεται κάθε φορά αν η τιμή κατωφλίου $T_a(i)$ είναι μικρότερη ή μεγαλύτερη από την επόμενη τιμή του σήματος $D(i+1)$.

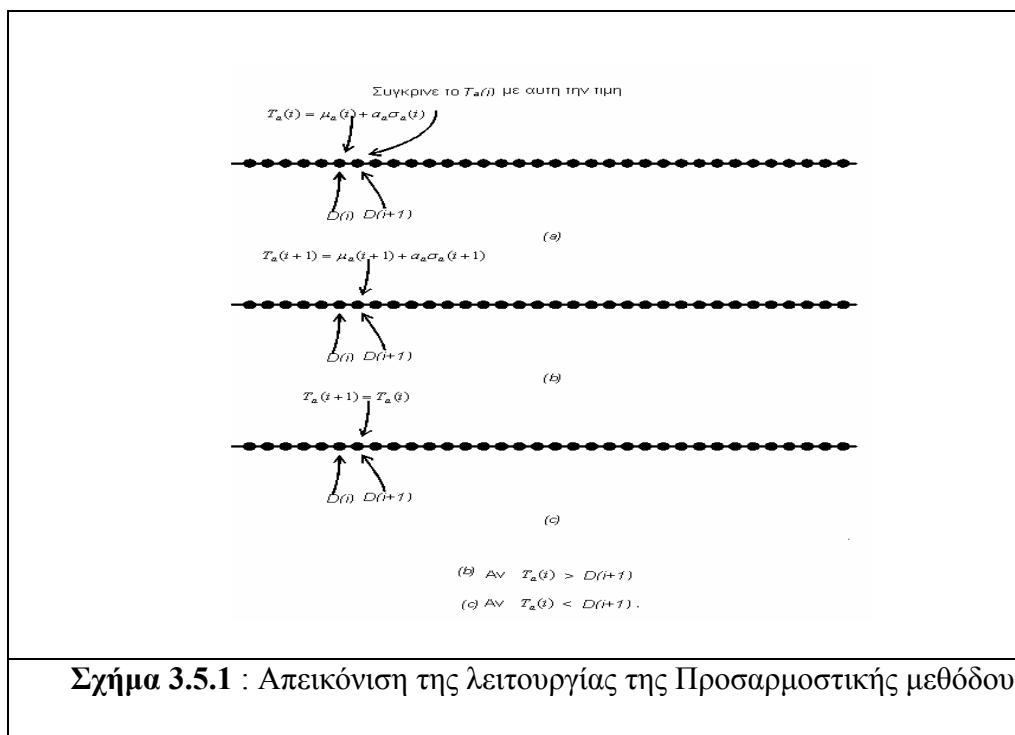
Αν είναι μικρότερη $T_a(i) < D(i+1)$ τότε η τιμή του κατωφλίου παραμένει όπως είχε η προηγούμενη τιμή της

$$T_a(i+1) = T_a(i)$$

μέχρι να βρεθεί τιμή μεγαλύτερη.

Ενώ αντίθετα αν είναι μεγαλύτερη $T_a(i) > D(i+1)$ τότε υπολογίζονται οι ποσότητες προσθέτοντας την νέα τιμή

$$T_a(i) = \mu_a(i) + a_a \cdot \sigma_a(i)$$



Αναφορά στα βασικά βήματα της διαδικασίας :

1. Υπολογίζονται οι τρεις ποσότητες οι μ_a , λ_a και σ_a για τις πρώτες 15 τιμές του σήματος
2. Αντίστοιχα με τις τιμές από το βήμα 1 υπολογίζονται οι 15 πρώτες τιμές του κατωφλίου $T_{ad_w}(i)$ με βάση τους τύπους (3.5.1) (3.5.2) (3.5.3) (3.5.4)
3. Γίνεται η σύγκριση του $T_{ad_w}(i)$ με την επόμενη τιμή $D(i+1)$
4. Αν $T_{ad_w}(i) > D(i+1)$ τότε υπολογίζονται οι νέες τιμές μ_a , λ_a και σ_a και $T_{ad_w}(i) = \mu_a(i) + a_a \cdot \sigma_a(i)$
5. Αν $T_{ad_w}(i) < D(i+1)$ τότε δεν υπολογίζονται ξανά νέες τιμές αλλά διατηρείται η προηγούμενη τιμή κατωφλίου. Στο σημείο αυτό ορίζουμε ότι ξεκινάει μια μεταβολή. Ανάλογα τώρα την διάρκεια της, συμπεραίνουμε αν είναι απότομη ή σταδιακή μεταβολή. Για παράδειγμα αν η συνθήκη 5 ισχύει για το πολύ 3-4 συνεχόμενα εικονοπλαίσια τότε έχουμε απότομη μεταβολή αν η διάρκεια είναι μεγαλύτερη, δηλαδή περισσότερα από 7 συνεχόμενα εικονοπλαίσια τότε έχουμε σταδιακή μεταβολή.
Επιστρέφει στο βήμα 3 για να γίνει σύγκριση με την επόμενη τιμή
6. Επαναλαμβάνεται αυτή η διαδικασία για όλα τα εικονοπλαίσια του

Ένα βασικό χαρακτηριστικό της Προσαρμοστικής μεθόδου είναι ότι δεν είναι τόσο ευαίσθητη στις μεταβολές του σήματος όπως η Sliding Window, δηλαδή αργεί να αντιδράσει. Αυτό είναι λογικό αφού η τιμή του κατωφλίου δεν εξαρτάται μονάχα από τις 15 προηγούμενες τιμές αλλά από όλες τις προηγούμενες επομένως καθυστερεί την εμφάνιση μιας έντονης μεταβολής. Η συμπεριφορά αυτή πολλές φορές όμως είναι χρήσιμη στον υπολογισμό του κατωφλίου αφού το πρόβλημα που αναφέρθηκε με τις χαμηλές τιμές του σήματος στη μέθοδο Παραθύρου Ολίσθησης εδώ δεν επηρεάζει το αποτέλεσμα. Στην παρακάτω γραφική παράσταση φαίνεται η απόκριση της μεθόδου η οποία δεν επηρεάζεται από τις χαμηλές τιμές στο διάστημα 150 με 200.

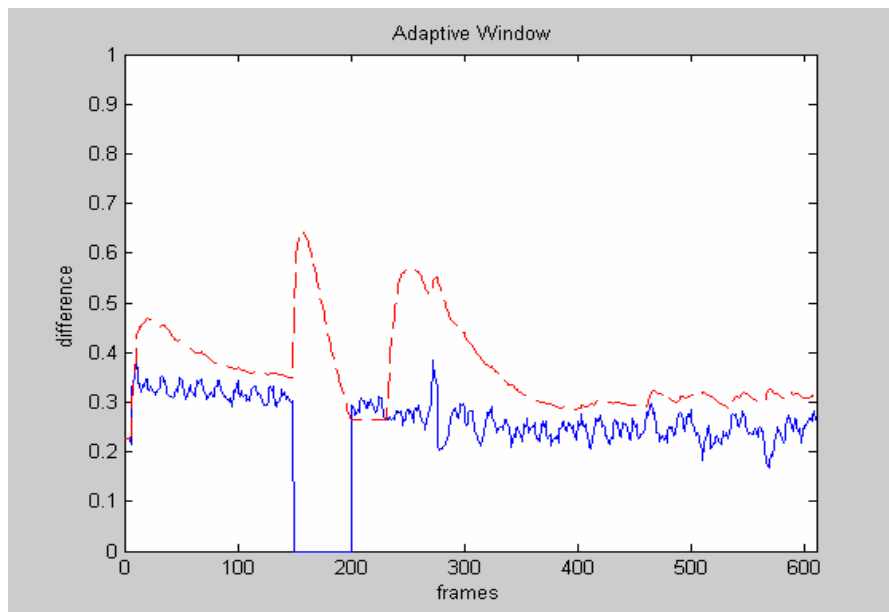


Figure (3.5.1)

Η τιμή κατωφλίου T (κόκκινο) στην εφαρμογή της Προσαρμοστικής μεθόδου σε σήμα διαφοράς ιστογραμμάτων έντασης. Μάλιστα παρατηρείται ότι όταν το σήμα πέφτει απότομα σε πολύ μικρές τιμές αντί να πέφτει και η τιμή κατωφλίου να συμβαίνει κάτι αντίθετο και να ανεβαίνει. Αυτό συμβαίνει γιατί η τιμή κατωφλίου εξαρτάται από την διασπορά που σε μια τέτοια περίπτωση είναι αρκετά μεγάλη.

Στη συνέχεια παρουσιάζονται κάποιες γραφικές παραστάσεις που απεικονίζουν την απόκριση της Προσαρμοστικής μεθόδου στο σήμα της διαφοράς των ιστογραμμάτων έντασης διαδοχικών frames .

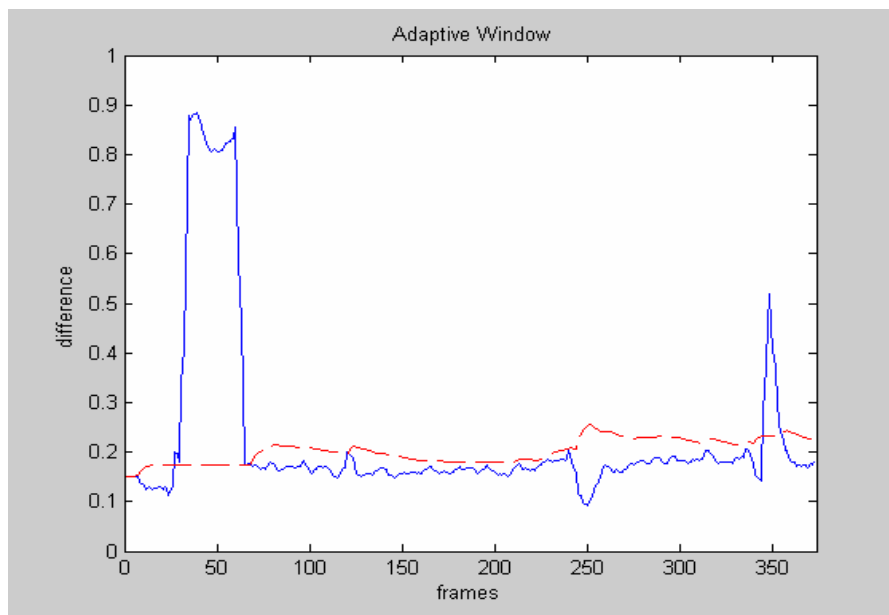


Figure (3.5.2)

Η τιμή κατωφλίου T(κόκκινο) στην εφαρμογή της Προσαρμοστικής μεθόδου σε σήμα διαφοράς ιστογραμμάτων έντασης

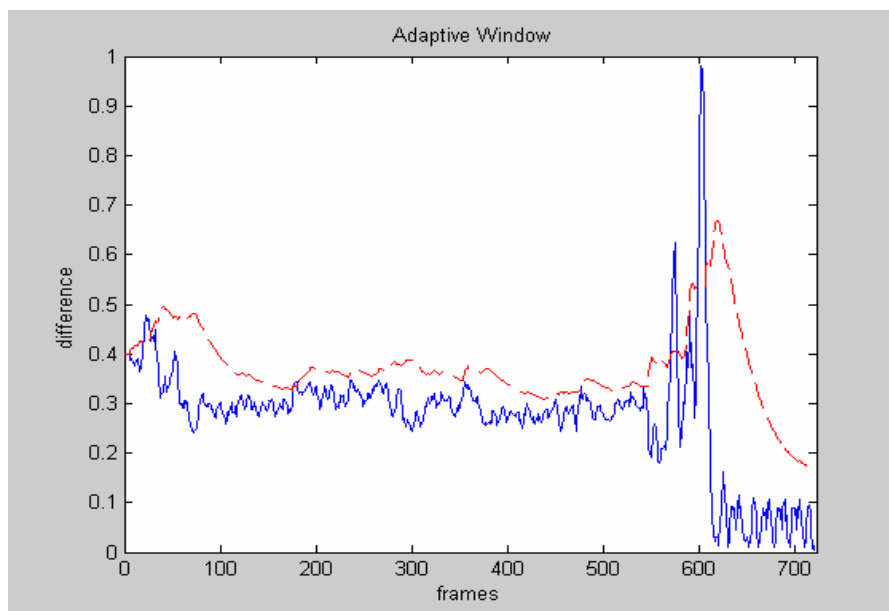


Figure (3.5.3)

Η τιμή κατωφλίου T(κόκκινο) στην εφαρμογή της Προσαρμοστικής μεθόδου σε σήμα διαφοράς ιστογραμμάτων έντασης .

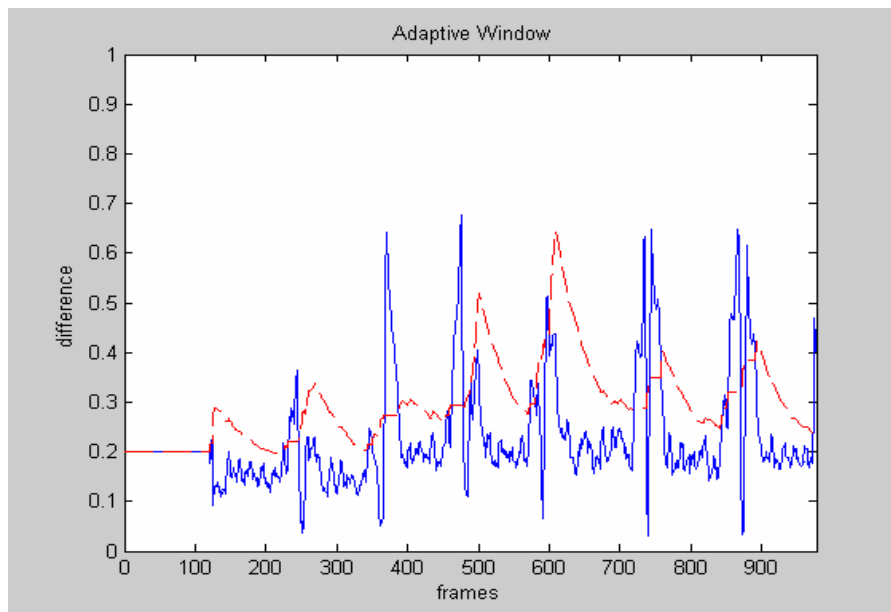


Figure (3.5.4)

Η τιμή κατωφλίου T (κόκκινο) στην εφαρμογή της Προσαρμοστικής μεθόδου σε σήμα διαφοράς ιστογραμμάτων έντασης

Συμπερασματικά η Προσαρμοστική μέθοδος :

1. Λειτουργεί αποτελεσματικά για να ανιχνεύσει μεταβολές σε ένα συνεχές σήμα και προσαρμόζεται σε τοπικές αλλαγές
2. Δεν χρειάζεται προεπεξεργασία ολόκληρου του σήματος για την εξαγωγή της τιμής κατωφλίου, άρα μπορεί να εφαρμοστεί και σε εφαρμογές πραγματικού χρόνου
3. Δεν είναι τόσο ευαίσθητη στις μεταβολές όσο η μέθοδος Παραθύρου Ολίσθησης επειδή η τιμή κατωφλίου εξαρτάται και από όλες τις προηγούμενες τιμές και όχι μόνο από τις 14 προηγούμενες τιμές που γίνεται στις μεθόδους με παράθυρο. Αυτό έχει σαν αποτέλεσμα να χάνει κάποιες μικρές μεταβολές του σήματος, αλλά από την άλλη μεριά αντιδρά σωστότερα σε μεταβολές μεγάλης διάρκειας όπου το σήμα έχει πολύ μικρές (μηδενικές) τιμές .

4.Χρησιμοποιεί μόνο μια παράμετρο αντί για δυο που χρειάζεται η μέθοδος Διπλής Σύγκρισης.

5. Προσαρμόζεται σε τοπικές μεταβολές του σήματος που μπορεί να οφείλονται και σε μεταβολή περιεχομένου. Αναμένεται να αποδώσει καλύτερα σε βίντεο με αρκετές αλλαγές περιεχομένου(animations,ειδήσεις και διαφημίσεις).

Κεφάλαιο 4

Κατάτμηση του βίντεο και χαρακτηρισμός της κίνησης

4.1 Η βασική ιδέα κατάτμησης και ανάλυση των προβλημάτων εφαρμογής της

Στο σημείο αυτό είναι απαραίτητο να αναφέρουμε λεπτομερέστερα ποιες είναι οι βασικές μεταβολές στο βίντεο τις οποίες προσπαθούμε να ανιχνεύσουμε με σκοπό να καθορίσουμε τα όρια κάθε στιγμιότυπου (shot boundary). Οι μεταβολές που τυπικά θεωρούμε ότι συμβαίνουν σε μια χρονοσειρά εικόνων είναι οι απότομες μεταβολές : Cut, Flash και οι σταδιακές μεταβολές : Dissolve, Zoom, Pan και Tilt. Με τον όρο Dissolve στην συγκεκριμένη εργασία θα συμπεριλαμβάνονται και οι μεταβολές Fade in , Fade out και Wipe οι οποίες σε άλλες εργασίες εξετάζονται ξεχωριστά.

Από τις παραπάνω μεταβολές μονάχα τα Cuts και τα Dissolves αποτελούν όρια ενός στιγμιότυπου, όλες οι υπόλοιπες μεταβολές δεν αποτελούν όρια. Παρόλα αυτά παρουσιάζουν αρκετά κοινά χαρακτηριστικά με τις πρώτες με αποτέλεσμα τις περισσότερες φορές να συγχέονται στη προσπάθεια ανίχνευσης τους με μια μόνο μέθοδο. Συνεπώς η βασική ιδέα κατάτμησης είναι αρχικά να ανιχνευτούν το σύνολο των μεταβολών και στη συνέχεια να καθοριστούν σαν όρια στιγμιότυπου τα σημεία στα οποία εμφανίζονται Cuts και όχι Flashes όπως επίσης και τα σημεία στα οποία υπάρχουν Dissolves και όχι Zoom, Pan ή Tilt. Συμπεραίνουμε λοιπόν ότι ο χαρακτηρισμός και η ανάλυση των μεταβολών του βίντεο από την μια μεριά μας δίνει πληροφορία για το περιεχόμενο του και από την άλλη μας βοηθάει στις τεχνικές κατάτμησης στιγμιότυπων να εξαλείψουμε τυχόν λάθος όρια “false positives” (Flash, Zoom, Pan και Tilt).

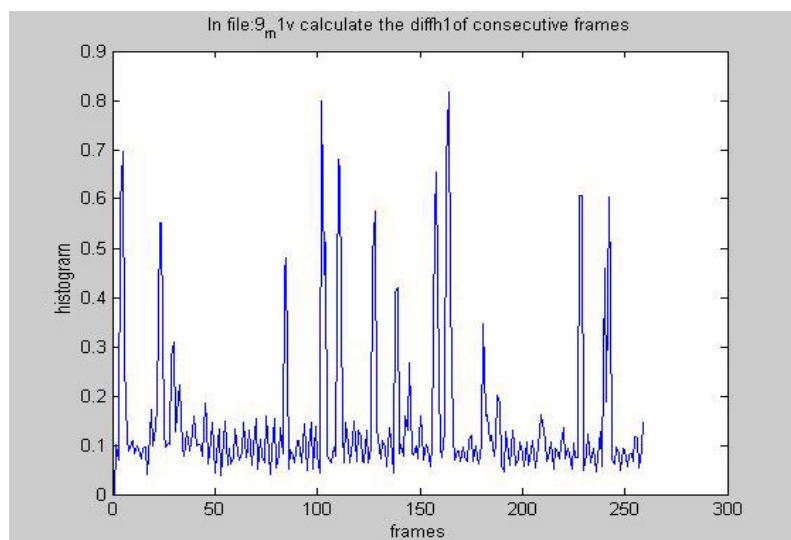
Όσον αφορά τώρα την πληροφορία από τα βίντεο, την οποία θα χρησιμοποιήσουμε στις τεχνικές ανίχνευσης μεταβολών, είναι δυο μορφών : τιμές έντασης και διανύσματα κίνησης. Όπως αναλύθηκε στο Κεφαλαίο 2.8 και 2.9 για κάθε εικονοπλαιο υπάρχει ένας πίνακας με τιμές έντασης της αντίστοιχης DC εικόνας του και για κάθε block του εικονοπλαισίου ένα ζευγάρι τιμών διανυσμάτων

(u, v). Όπου u είναι η οριζόντια και v η κατακόρυφη συνιστώσα του διανύσματος κίνησης του. Οι μεταβολές αυτών των ποσοτήτων μεταξύ των διαδοχικών εικονοπλαισίων καθορίζουν και τα κριτήρια των τεχνικών ανίχνευσης μεταβολών που θα παρουσιάσουμε στα παρακάτω κεφάλαια.

4.2 Μέθοδος ανίχνευσης φλας (flash model)

Η ανίχνευση απότομων μεταβολών, που κυρίως είναι τα cuts, είναι μια σχετικά εύκολη διαδικασία χρησιμοποιώντας την πληροφορία από τις μεταβολές της έντασης. Παρόλα αυτά πολλές φορές υπεισέρχονται κάποια λάθη πέραν από την δυσκολία υπολογισμού threshold, όπως για παράδειγμα η εμφάνιση ενός flash (απότομη λάμψη) στην διάρκεια ενός βίντεο. Το φαινόμενο αυτό προκαλεί δυο απότομες υψηλές τιμές στο σήμα $D(j)$ σαν δυο πολύ κοντινά cuts στην αρχή και στο τέλος του flash και έτσι ξεγελάει το σύστημα ανιχνεύοντας το σαν δυο cuts και κατ' επέκταση θέτει ανύπαρκτα όρια στιγμιότυπου (shot boundary).

Το μοντέλο που προτείνεται στο άρθρο των Dong Zhang, Wei Qi, Hong Jiang [20] έχει σκοπό την διόρθωση αυτών των false positives που προκαλούνται από το flash. Ένα flash διαρκεί περισσότερα από ένα εικονοπλαίσια και εμφανίζεται σαν μια απότομη λάμψη με μια μικρή παύση στην κίνηση της εικόνας για όσο διάστημα διαρκεί. Κατά την διάρκεια ενός flash παρατηρείται ότι μέση τιμή της έντασης αυξάνει λόγω της υψηλής φωτεινότητας και συνεπώς μοιάζει πολύ με Cut



Σχήμα 4.2.1 : Στο σχήμα παρουσιάζεται η διαφορά ιστογραμμάτων έντασης σε ένα βίντεο με flashes. Στα σημεία που εμφανίζονται πολύ κοντινές κορυφές είναι η αρχή και το τέλος ενός flash και όχι δυο συνεχόμενα Cuts.

Ο τύπος που μας δίνει την **μέση τιμή έντασης** σε ένα i εικονοπλαίσιο δίνεται από τον παρακάτω τύπο.

$$AI_i = \frac{\sum_{j=1}^{Bins} j \cdot H_i(j)}{\sum_{j=1}^{Bins} H_i(j)} \quad (4.2.1)$$

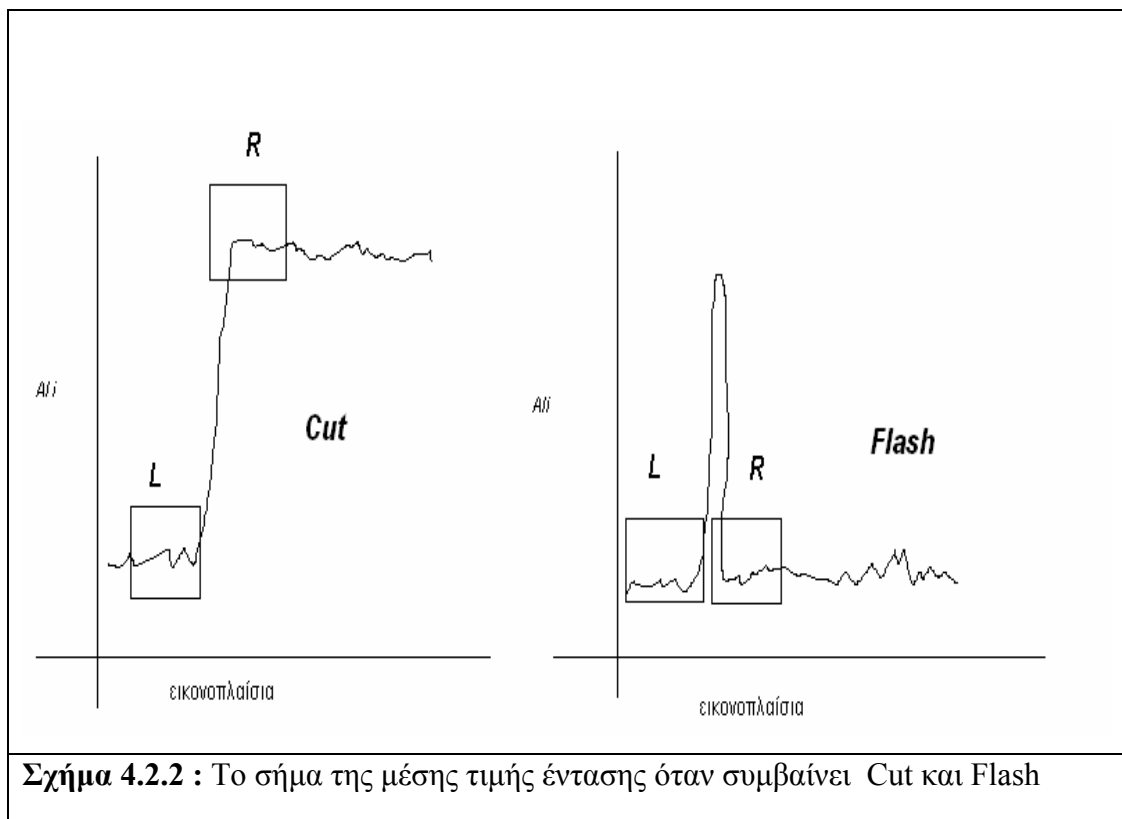
Ενώ η **διάφορα της μέσης τιμής έντασης** δυο διαδοχικών εικονοπλαισίων i και $i-1$ είναι

$$AID_i = AI_i - AI_{i-1} \quad (4.2.2)$$

Αρχικά στη μέθοδο αυτή ορίζονται δυο ποσότητες

H1: Η διαφορά της μέσης τιμής της έντασης μεταξύ του τρέχοντος και του προηγούμενου εικονοπλαισίου.

H2: Η διαφορά της μέσης τιμής έντασης μεταξύ των εικονοπλαισίων του αριστερού παραθύρου (περιλαμβάνει 6-7 εικονοπλαίσια) **L** που προηγείται του τρέχοντος εικονοπλαισίου και αντίστοιχα των εικονοπλαισίων του δεξιού παραθύρου (περιλαμβάνει 6-7 εικονοπλαίσια) **R** που ακολουθεί του τρέχοντος εικονοπλαισίου. Ο έλεγχος που γίνεται με την βοήθεια των παραθύρων στηρίζεται στο ότι στα γειτονικά εικονοπλαίσια πριν και μετά το flash οι μέσες τιμές των εντάσεων έχουν σχεδόν τις ίδιες τιμές, όπως φαίνεται και από το σχήμα 4.2.2



Σχήμα 4.2.2 : Το σήμα της μέσης τιμής έντασης όταν συμβαίνει Cut και Flash

Όταν συμβαίνει flash η ποσότητα $H2$ είναι σχεδόν μηδέν. Όπως φαίνεται στο σχήμα 4.2.2 όταν συμβαίνει flash προσωρινά εμφανίζεται κάποια λάμψη που λόγω έντονης φωτεινότητας δίνει υψηλή μέση τιμή έντασης και στην συνέχεια επανέρχεται πάλι στην προηγούμενη ροή. Άρα αν υπολογίσουμε τις μέσες τιμές των AI_i σε κάθε παράθυρο (**L** και **R**) θα δούμε ότι είναι σχεδόν ίσες και κατ' επέκταση η διάφορα τους είναι μηδενική ή παρά πολύ μικρή. Το $H1$ σε αυτή την περίπτωση έχει μια αρκετά μεγαλύτερη τιμή.

Από την άλλη όταν συμβαίνει Cut η τιμή $H2$ είναι ίδια με την $H1$. Αρχικά το $H1$ όπως αναφέρθηκε και νωρίτερα είναι η διάφορα της μέσης τιμής του τρέχοντος εικονοπλαισίου σε σχέση με την μέση τιμή του προηγούμενου, υπολογίζεται από την σχέση (4.3.2), η διαφορά αυτή θεωρητικά είναι αρκετά μεγάλη. Το $H2$ όμως στην περίπτωση του Cut δείχνει ότι τα εικονοπλαισια που υπάρχουν στο παράθυρο πριν το Cut (**L**) έχουν μια συγκεκριμένη μέση τιμή έντασης ενώ μετά το Cut (**R**) οι μέσες τιμές έντασης βρίσκονται σε ένα άλλο επίπεδο (Στο σχήμα 4.2.2 φαίνεται πως μετά το Cut το επίπεδο των AI_i ανέβηκε θα μπορούσε βέβαια να συμβαίνει και το αντίστροφο). Άρα η διαφορά αυτών των δυο επιπέδων είναι η τιμή που δίνει και η διάφορα $H1$.

Μέθοδος 1 : Διαχωρισμός Cut από Flash

Αν χρησιμοποιηθεί ο λόγος των δυο αυτών ποσοτήτων το μοντέλο γίνεται ως εξής

$$Ratio = \frac{H2}{H1} \quad (4.2.3)$$

Συνεπώς αν Ratio 0 ή τιμή κοντά στο 0 τότε στο εικονοπλαιο αυτό υπάρχει Flash

Αντίθετα αν Ratio 1 ή τιμή κοντά στο 1 τότε στο εικονοπλαιο αυτό υπάρχει Cut.

Τα δυο παραπάνω κριτήρια μεταφράζονται στην σχέση (4.2.4), όπου το T είναι ένα κατώφλι με τιμή σχεδόν 0,5

$$Ratio = \begin{cases} > T & \text{Cut Model} \\ < T & \text{Flash Model} \end{cases} \quad (4.2.4)$$

Η πειραματική μελέτη της μεθόδου έδειξε ότι το μοντέλο αυτό λειτουργεί αρκετά ικανοποιητικά στον διαχωρισμό του πραγματικού Cut από το flash. Πολλές φορές όμως σαν flashes ανιχνεύονται και κάποια άσπρα εικονοπλάισια που παρεμβάλλονται με μια μικρή παύση κατά την διάρκεια του βίντεο ή πολύ γρήγορα fade in και fade out μικρής διάρκειας (4-6 εικονοπλαισίων). Τέτοια εικονοπλάισια εμφανίζονται συνήθως σε animations που υπάρχει γρήγορη εναλλαγή του σκηνικού.

Ακολουθώντας λοιπόν τα στάδια της Μεθόδου 1 γίνεται ένα πρώτο ξεκαθάρισμα και ελάττωση των false positives που δημιουργούνται από την συνύπαρξη Cuts και Flashes σε ένα βίντεο.

4.3 Πληροφορία κίνησης για ανίχνευση zoom, pan και tilt

Μια από τις βασικές πηγες πληροφορίας η οποία εξάγεται από συμπιεσμένο βίντεο σε μορφή MPEG είναι τα διανύσματα κίνησης. Όπως αναφέρθηκε και νωρίτερα η συμπιεσμένη μορφή βίντεο δίνει περισσότερη αξιοποιήσιμη πληροφορία για το περιεχόμενο του. Το μέτρο τους, η γωνία που σχηματίζουν με το οριζόντιο επίπεδο ακόμα και το πλήθος τους είναι ποσότητες που καθορίζουν αν υπάρχει κάποια κίνηση της κάμερας ή κάποιο εφε στο βίντεο. Έτσι μεταβολές όπως το pan, zoom και tilt γίνονται ανιχνεύσιμες αν ικανοποιούν κάποια από τα κριτήρια που ορίζουν οι μέθοδοι που θα περιγράψουν στο κεφάλαιο 4.3.2.

Αρχικά αφού γίνει κανονικοποίηση των διανυσμάτων κίνησης όπως περιγράφεται στο Κεφάλαιο 2.9 δημιουργείται ένα διάνυσμα κίνησης για κάθε εικονοπλαίσιο ανεξάρτητα από το είδος του εικονοπλαισίου (I,P,B). Το κάθε διάνυσμα αντιπροσωπεύεται από δυο τιμές (u , v), την οριζόντια και κατακόρυφη απομάκρυνση του macroblock του τρέχοντος εικονοπλαισίου από το επόμενο. Ουσιαστικά υπάρχουν δυο πίνακες για κάθε εικονοπλαίσιο, ένας με τιμές που αντιπροσωπεύουν την οριζόντια κίνηση προς τα δεξιά του κάθε macroblock και ο άλλος την κατακόρυφη κίνηση προς τα κάτω. Από τους δυο αυτούς πίνακες ελέγχονται σε ποια κατηγορία ανήκει το κάθε macroblock και ανάλογα εξάγεται το μέτρο και η γωνία του διανύσματος του.

Τα macroblocks τα οποία δεν μπορούν να ανιχνευτούν στο επόμενο εικονοπλαίσιο ονομάζονται **Intracoded** και κωδικοποιούνται με ενδοεικονοπλαισιακή κωδικοποίηση όπως τα I εικονοπλαίσια. Μάλιστα κατά την διαδικασία της κανονικοποίησης της κίνησης σε τέτοιου είδους macroblocks δίνεται σαν τιμή διανύσματος κίνησης προς τα δεξιά τιμή έντασης = 1000 και στην συνέχεια με έναν έλεγχο διαχωρίζεται από τα υπόλοιπα. Επίσης μια άλλη κατηγορία macroblocks είναι τα λεγόμενα **Static** macroblocks τα οποία έχουν πολύ μικρές τιμές έντασης των διανυσμάτων κίνησης σχεδόν ασήμαντες για να θεωρηθεί ότι αυτά μετακινήθηκαν από το ένα στο επόμενο εικονοπλαίσιο. Τα static macroblocks τα ορίζουμε εμείς για να κατηγοριοποιήσουμε καλύτερα μια ομάδα από macroblocks των οποίων οι τιμές είναι πολύ μικρές και πρέπει να τα αγνοήσουμε κατά τον υπολογισμό κίνησης. Μια υποκατηγορία των static macroblocks είναι τα **Skipped** macroblocks τα οποία δεν

έχουν μετακινηθεί καθόλου δηλαδή οι τιμές των διανυσμάτων κίνησης είναι μηδενικές. Τέλος όλα τα υπόλοιπα macroblocks είναι εκείνα τα οποία έχουν μετακινηθεί ως προς το επόμενο εικονοπλαίσιο και μπορεί να προβλεφθεί η κίνηση τους, σε αυτά υπολογίζονται το μέτρο και η γωνία τους. Η πρώτη και η τελευταία κατηγορία είναι και εκείνες οι οποίες συνεισφέρουν ουσιαστικά στην πληροφορία κίνησης.

Κάνοντας αυτό τον διαχωρισμό στις παραπάνω κατηγορίες γίνεται ξεκάθαρο ότι η πληροφορία που στην συνέχεια θα χρειαστεί να επεξεργαστεί δεν είναι σταθερή ποσότητα αφού ο αριθμός των macroblocks σε κάθε κατηγορία διαφέρουν από εικονοπλαίσιο σε εικονοπλαίσιο. Επομένως το πρώτο βήμα είναι μια κανονικοποίηση αυτής της πληροφορίας. Πιο συγκεκριμένα οι ποσότητες που θα υπολογίζονται ανά εικονοπλαίσιο θα πολλαπλασιάζονται και με ένα συντελεστή που θα δείχνει το ποσοστό της πληροφορίας που δίνει το κάθε εικονοπλαίσιο. Ο συντελεστής αυτός ονομάζεται γ και είναι ίσος με

$$\gamma = 1 - \frac{IntracodedVectors + StaticVectors}{TotalMacroblocks\ in\ Frame} \quad (4.3.1)$$

4.3.1 Ιστόγραμμα γωνιών διανυσμάτων κίνησης

Όπως προαναφέρθηκε ένα διάνυσμα κίνησης περιγράφεται σαν ένα ζευγάρι τιμών (u, v) , που συμβολίζουν την οριζόντια και κατακόρυφη απομάκρυνση του macroblock του τρέχοντος εικονοπλαισίου από το επόμενο. Από αυτές τις τιμές άμεσα προκύπτει και η γωνία που σχηματίζει το διάνυσμα με το οριζόντιο επίπεδο σύμφωνα με τον τύπο

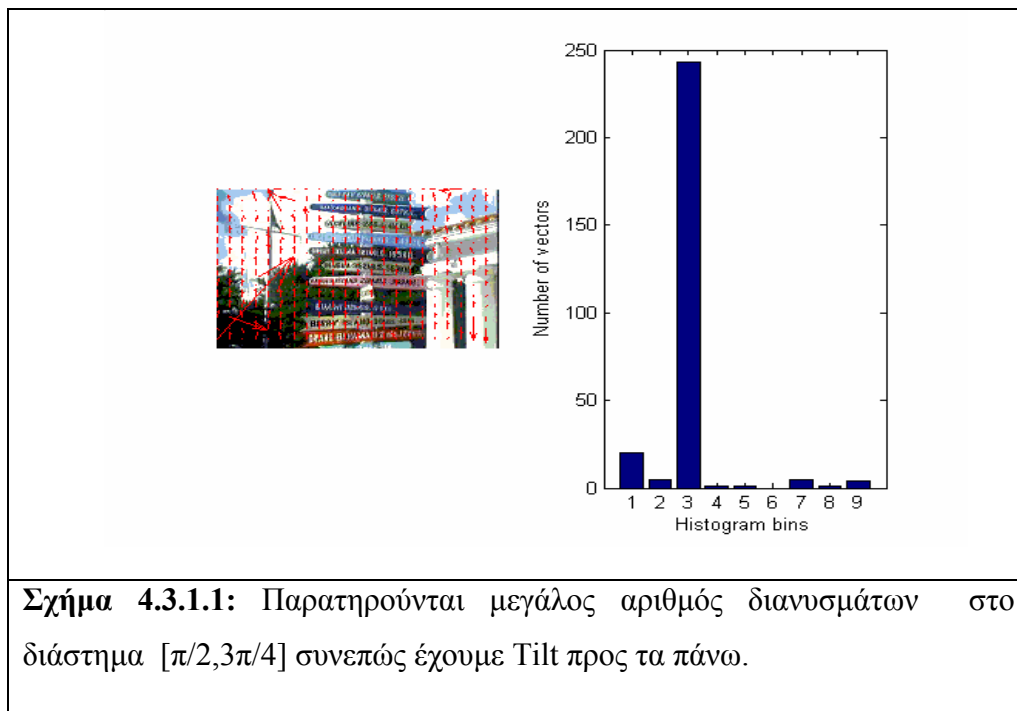
$$\theta = \tan^{-1}(v/u) \quad ITT \quad [0, 2\pi] \quad (4.3.1.1)$$

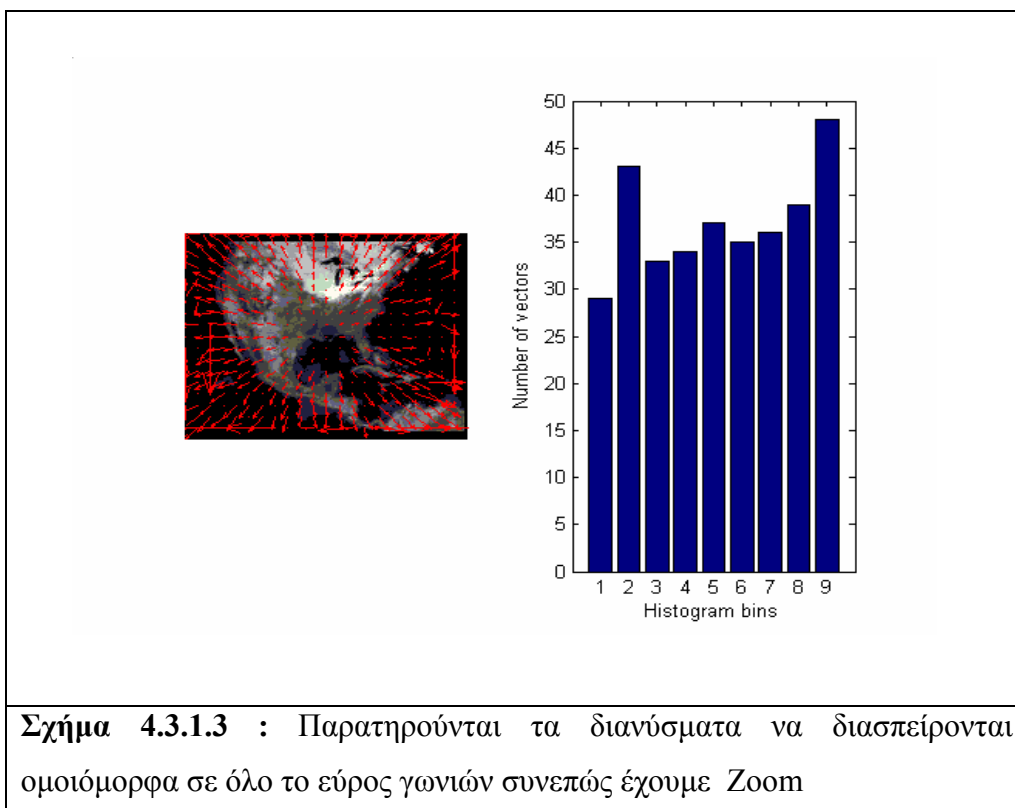
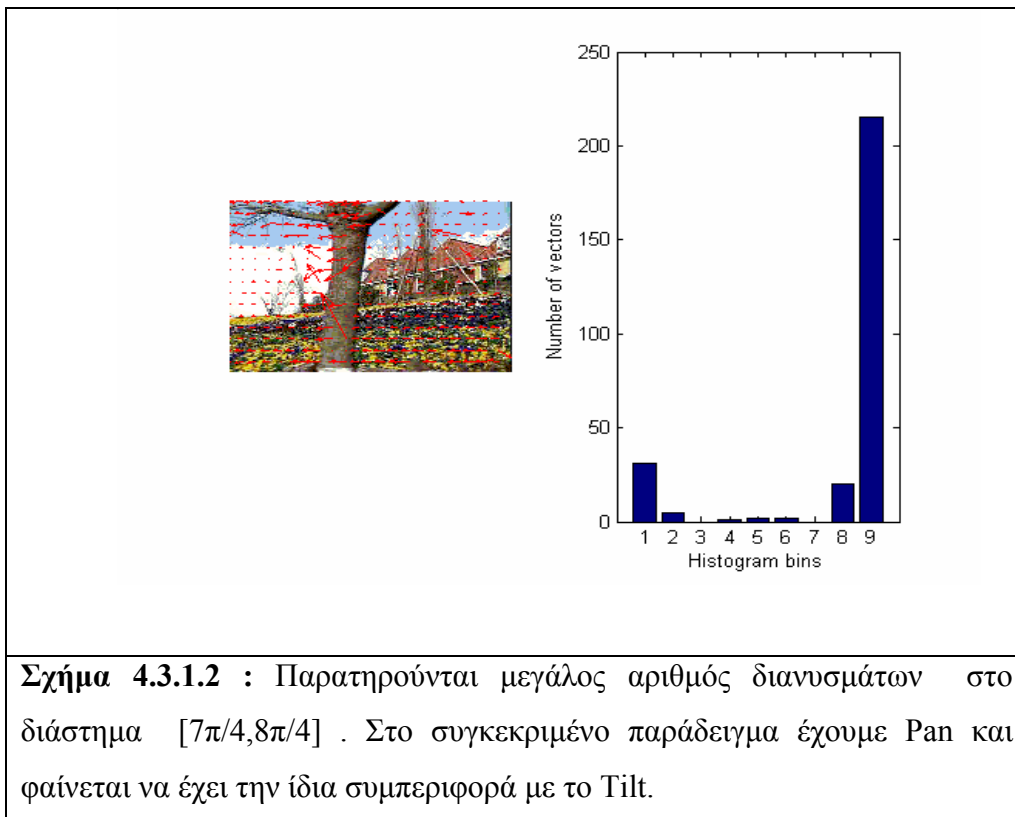
Οι γωνίες αυτές μπορούν να πάρουν οποιαδήποτε τιμή στο διάστημα $[0, 2\pi]$ συνεπώς χρειάζεται να γίνει κάποια κβαντοποίηση. Οι γωνίες θ των διανυσμάτων

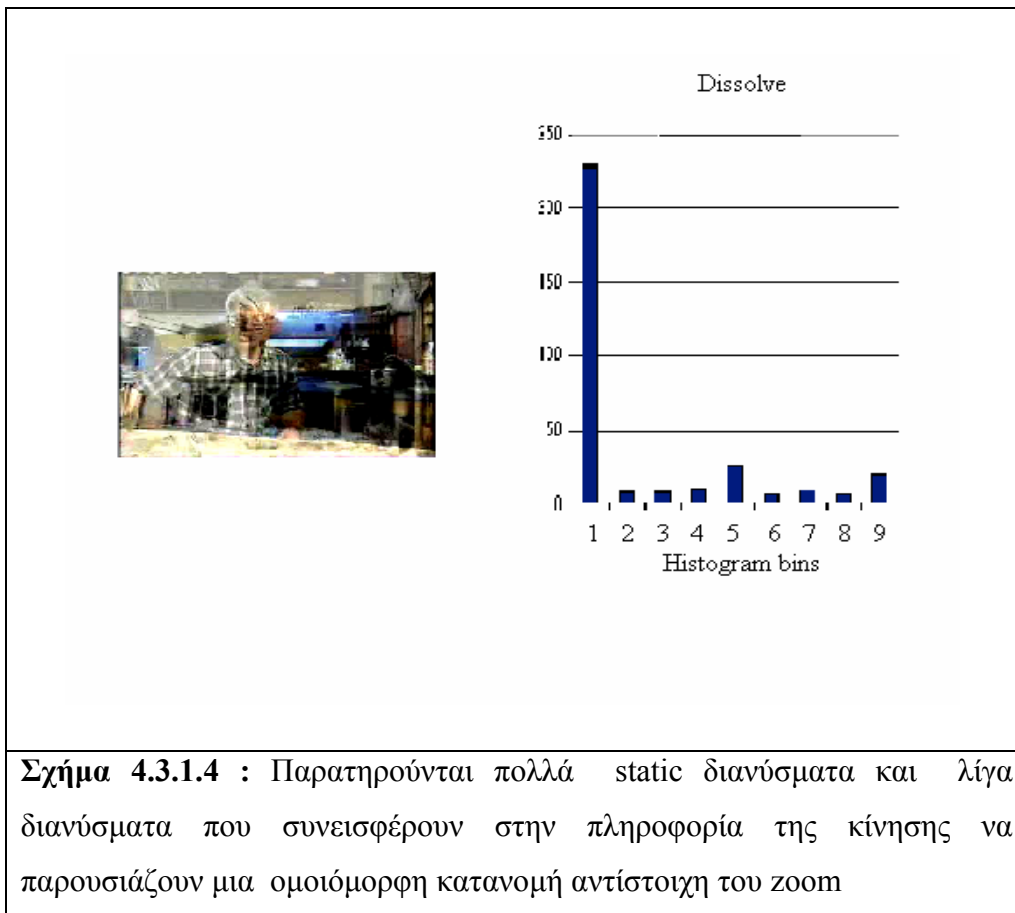
χωρίζονται σε 8 κατηγορίες ($[0, \pi/4]$ $[\pi/4, \pi/2]$ $[\pi/2, 3\pi/4]$... $[7\pi/4, 2\pi]$) που ορίζουν και την κατεύθυνση τους.

Άρα το ιστόγραμμα που ορίζεται είναι ένα ιστόγραμμα με εννιά τιμές που αντιπροσωπεύουν η πρώτη τιμή τον αριθμό των static διανυσμάτων και οι επόμενες οχτώ τιμές τον αριθμό των διανυσμάτων που ανήκουν στις οχτώ κατευθύνσεις που αντίστοιχα έχουν ομαδοποιηθεί.

Στα παρακάτω σχήματα φαίνονται η μορφή τέτοιων ιστογραμμάτων







Η μελέτη αυτών των ιστογραμμάτων δίνει πολύ χρήσιμα συμπεράσματα όσον αφορά την ανάλυση του βίντεο και κυρίως όσον αφορά την ανίχνευση συγκεκριμένων μεταβολών όπως το Zoom, Pan και Tilt.

Ξεκινώντας από τα δυο πρώτα σχήματα στα οποία παρατηρείται Tilt και Pan, γίνεται αντιληπτό από την μορφή του ιστογράμματος ότι στο πρώτο σχήμα οι κατευθύνσεις των διανυσμάτων στην πλειοψηφία τους ανήκουν σε τιμές γωνιών $[\pi/4, \pi/2]$ που σημαίνει ότι τα macroblocks έχουν μια τάση να κινηθούν προς τα πάνω. Άρα υπάρχει μια κίνηση της κάμερας προς εκείνη την κατεύθυνση. Στο δεύτερο σχήμα αντίστοιχα η πλειοψηφία των τιμών των γωνιών είναι στο πεδίο $[7\pi/4, 2\pi]$ που σημαίνει ότι υπάρχει μια κίνηση προς τα αριστερά. Οι δυο καταστάσεις δεν μπορούν να διαχωριστούν συνεπώς η ανίχνευση θα γίνεται σαν ανίχνευση του συνόλου Pan και Tilt, άλλωστε το Pan μπορεί να θεωρηθεί σαν ένα Tilt που αλλάζει απότομα τη φορά του. Στη συνέχεια στο τρίτο σχήμα φαίνεται χαρακτηριστικά ένα Zoom όπου οι γωνίες των διανυσμάτων διασπείρονται ομοιόμορφα σε όλο το πεδίο τιμών $[0, 2\pi]$. Ενώ κάτι αντίστοιχο φαίνεται να

συμβαίνει και στο τέταρτο σχήμα που παρουσιάζει ένα Dissolve με πολλά static διανύσματα και τα υπόλοιπα διανύσματα να διασπείρονται ομοιόμορφα σαν την περίπτωση που έχουμε Zoom.

Συμπερασματικά

1. Όταν οι τιμές των γωνιών των διανυσμάτων κίνησης διασπείρονται ομοιόμορφα στον άξονα των γωνιών και καλύπτουν όλο το πεδίο τιμών $[0, 2\pi]$, τότε υπάρχει **Zoom** ή **Dissolve**.
2. Όταν οι τιμές των γωνιών των διανυσμάτων κίνησης συγκεντρώνονται σε μια κατεύθυνση, τότε υπάρχει **Tilt** ή **Pan** προς εκείνη την κατεύθυνση.
3. Όταν οι τιμές των γωνιών των διανυσμάτων κίνησης συγκεντρώνονται στην πρώτη τιμή του ιστογράμματος που αντιπροσωπεύει τον αριθμό των static διανυσμάτων, τότε δεν υπάρχει κάποια μεταβολή ή έχουμε **Zoom**.

4.3.2 Μέθοδοι ανίχνευσης zoom, pan, tilt

Η προσπάθεια υλοποίησης μεθόδων ανίχνευσης μεταβολών όπως το Zoom, Pan και Tilt μετατίθεται σε μελέτη των παραπάνω ιστογραμμάτων. Μια σειρά από διαδικασίες και υπολογισμούς ποσοτήτων που αναδεικνύονται από την πληροφορία που δίνουν τα ιστογράμματα είναι και η βασική πηγή εξαγωγής συμπερασμάτων. Αρχικά υπολογίζεται η διασπορά των οχτώ τιμών των διευθύνσεων του ιστογράμματος σύμφωνα με τον τύπο:

$$\sigma_h^2 = \sum_{i=2}^{K=9} p_i (i - \bar{i})^2 \quad (4.3.2.1)$$

$$p_i = \frac{h_i}{\sum_{i=2}^{K=9} h_i} \quad (4.3.2) \quad \text{με} \quad \bar{i} = \sum_{i=2}^{K=9} p_i i \quad (4.3.2.3)$$

Στην συνέχεια εξαιτίας του γεγονότος ότι κάθε φορά ο αριθμός των διανυσμάτων που συνεισφέρουν στην πληροφορία της κίνησης αλλάζει από εικονοπλαίσιο σε εικονοπλαίσιο, θα πρέπει να κανονικοποιηθούν πολλαπλασιάζοντας αυτές με τον συντελεστή γ που περιγράφηκε παραπάνω

$$\sigma_{h,norm}^2 = \gamma \sigma_h^2 \quad (4.3.2.4)$$

Έτσι προκύπτουν **πολύ μεγάλες τιμές διασποράς όταν έχουμε zoom ή dissolve** (διασπείρονται ομοιόμορφα οι τιμές των γωνιών). Ενώ αντίθετα **πολύ μικρές τιμές διασποράς όταν υπάρχει Tilt και Pan**. Στην κανονικοποιημένη μορφή της διασποράς επειδή πολλαπλασιάζεται με τον συντελεστή γ μπορεί να αλλάξει έντονα η τιμή της διασποράς. Στην περίπτωση που έχουμε Dissolve υπάρχουν πολλά static διανύσματα και συνεπώς γίνεται πολύ μικρό το γ . Έτσι παρόλο που αρχική διασπορά των τιμών του μπορεί να ήταν μεγάλη αντίστοιχη με το Zoom, κατά την κανονικοποίηση επειδή το γ είναι πολύ μικρό γίνεται και η κανονικοποιημένη διασπορά πολύ μικρή, αντιστοίχου επιπέδου με όταν έχουμε Pan ή Tilt.

Μια επίσης πολύ σημαντική παραδοχή είναι ότι για να θεωρηθεί μια τέτοια συμπεριφορά ως μεταβολή, θα πρέπει να υπάρχει μικρό ποσοστό macroblocks με static διανύσματα σε σχέση με το σύνολο τους. Άρα θα πρέπει να ισχύει πέραν του Dissolve που γνωρίζουμε ότι έχει αρκετά macroblocks με static διανύσματα ότι για να έχουμε κάποια από τις μεταβολές Zoom, Pan και Tilt θα πρέπει **το σύνολο των macroblocks με static διανύσματα να είναι “K” < 40%** των συνολικών macroblocks. Με αυτό τον τρόπο μπορούμε να διαχωρίσουμε το Dissolve από το Pan και Tilt που έχουν και τα δυο μικρή διασπορά

Μέθοδος 2 :Ανίχνευση Zoom,Pan και Tilt

1. Αν $K < 0,4$ και $\sigma_{h,norm} > T_a$, Το εικονοπλαίσιο ανήκει σε μεταβολή Zoom
2. Αν $K < 0,4$ και $\sigma_{h,norm} < T_a$, Το εικονοπλαίσιο ανήκει σε μεταβολή Tilt ή Pan
3. Αν $K > 0,4$, Το εικονοπλαίσιο δεν ανήκει σε καμία από τις παραπάνω μεταβολές ή ανήκει σε μεταβολή Dissolve

Το T_a είναι το threshold που ορίζεται κάθε φορά ανάλογα με την μέθοδο που εφαρμόζεται . Στον συγκεκριμένο αλγόριθμο εφαρμόστηκαν και οι τέσσερις μέθοδοι thresholding που έχουν παρουσιαστεί και οι αποδόσεις τους παρουσιάζονται στα πειραματικά αποτελέσματα στο κεφάλαιο 5. Στην προσπάθεια αυτή πολλές φορές ανιχνεύονται και dissolves πέραν των άλλων, κάτι που δημιουργεί προβλήματα αφού δυσχεραίνει κατά πολύ την εξάλειψη των false positive που είναι και ο αρχικός σκοπός αυτής της ανάλυσης

4.4 Ανίχνευση απότομων και σταδιακών μεταβολών

Η βασική επιδίωξη σε αυτή την ενότητα είναι η ανίχνευση απότομων και σταδιακών μεταβολών με δυο μεθόδους με τη χρήση διαφορών ιστογραμμάτων έντασης και με την χρήση διαφοράς Intracoded macroblocks. Στην συνέχεια θα οριστούν και ο αλγόριθμος καθορισμού ορίων στιγμιότυπου ο οποίος θα συμπεριλαμβάνει ένα συνδυασμό όλων των μεθόδων ανίχνευσης μεταβολών με κριτήρια για την εξάλειψη των false positives.

4.4.1 Μέθοδος διαφοράς Ιστογραμμάτων τιμών έντασης

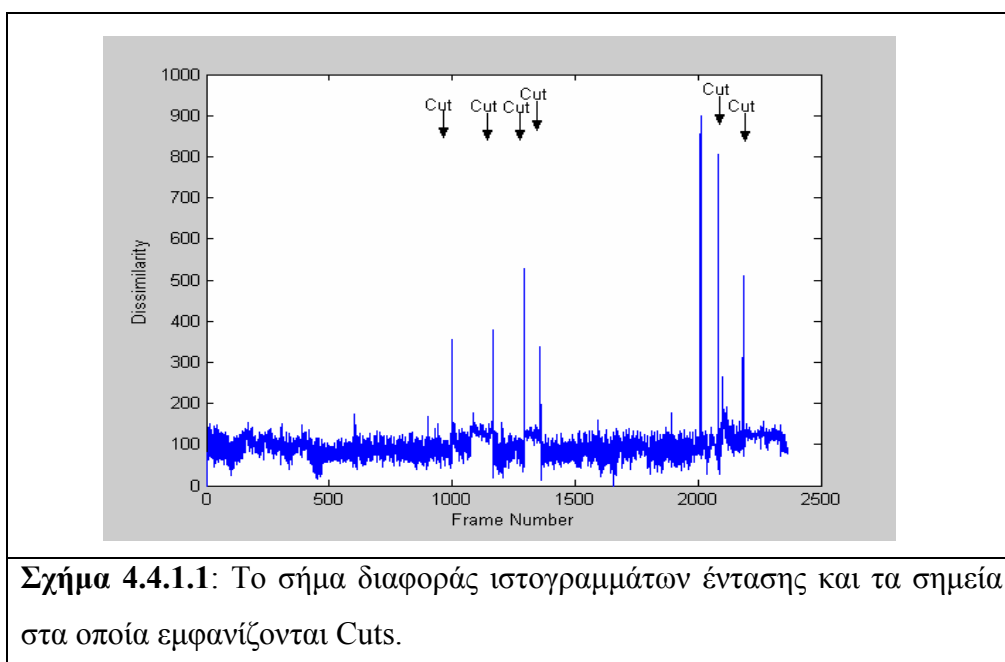
Μια από τις γνωστότερες τεχνικές στην ανίχνευση μεταβολών σε ένα βίντεο είναι η χρήση ιστογραμμάτων. Η βασική ιδέα στην μέθοδο αυτή είναι ότι τα εικονοπλάισια που ανήκουν στο ίδιο στιγμιότυπο παρουσιάζουν αρκετά όμοια ιστογράμματα τιμών έντασης. Συνεπώς αν υπολογιστούν οι διαφορές μεταξύ δυο διαδοχικών εικονοπλαισίων, όπου θα υπάρχει ένα σταθερό σκηνικό, η τιμή αυτής της διαφοράς θα είναι πολύ μικρή. Αντίστοιχα όπου παρουσιάζεται απότομα μια μεγάλη τιμή εκεί υπάρχει απότομη μεταβολή (cut), ενώ όπου παρατηρούνται σταδιακά υψηλές σημαίνει ότι υπάρχει σταδιακή μεταβολή (για να θεωρηθεί μια μεταβολή ως σταδιακή θα πρέπει να έχει διάρκεια μεγαλύτερη των πέντε διαδοχικών εικονοπλαισίων “dissolve, fade-in, fade-out, pan zoom, tilt”).

Επίσης είναι σημαντικό να κατανοηθεί ότι η πληροφορία που δίνει το ιστόγραμμα δεν είναι ικανή να ανιχνεύσει τυχόν εσωτερική κίνηση αντικειμένων, αφού η περιγραφή είναι συνολική και όχι τοπική. Ουσιαστικά το ιστόγραμμα δίνει πληροφορία για την αναλογία των εντάσεων φωτεινότητας που βρίσκονται σε μια εικόνα. Μια εσωτερική κίνηση διατηρεί σχεδόν σταθερή την αναλογία αυτή και κατ' επέκταση δεν είναι εφικτή και η ανίχνευση της. Αυτό το γεγονός μάλλον επιδρά θετικά αφού είναι χρήσιμο η πληροφορία που χρησιμοποιείται να είναι απαλλαγμένη από τέτοιου είδους μικρές αλλαγές που απλά θα αλλοίωναν το επιθυμητό αποτέλεσμα.

Μέθοδος 3 :Ανίχνευση απότομων και σταδιακών μεταβολών με την χρήση διαφορών ιστογραμμάτων έντασης.

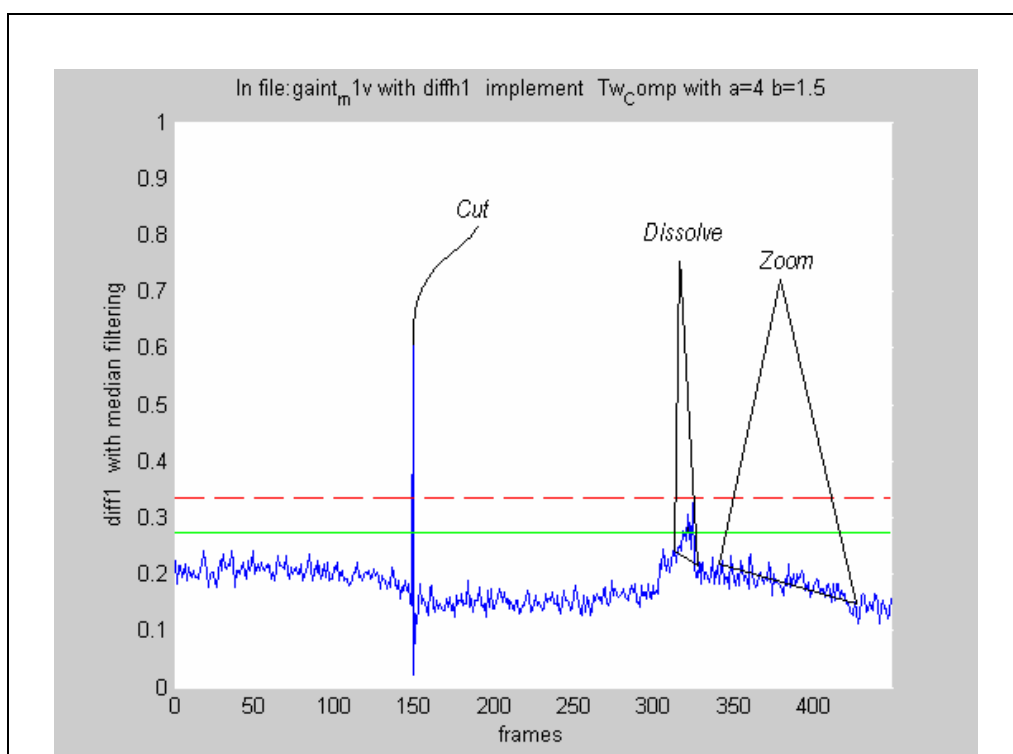
1. Υπολογίζεται το ιστόγραμμα του κάθε εικονοπλαισίου $H_k(i) = \frac{n_i}{M \times N}$
 $i=0 \dots 255$ όπου i είναι η τιμή της έντασης, n_i ο αριθμός των εικονοστοιχείων με ένταση i , και k το αντίστοιχο εικονοπλαίσιο, $1 \leq k \leq TotalFrames$
2. Υπολογίζεται η ποσότητα $D(j) = \sum_{i=0}^{255} (|H_k(i) - H_{k-1}(i)|)$
3. Εφαρμόζονται αντίστοιχα και οι τέσσερις μέθοδοι κατώφλιωσης στο σήμα $D(j)$
4. Ελέγχονται τα σημεία στα οποία το $D(j)$ ξεπερνάει τα αντίστοιχα κατώφλια
5. Αν τα σημεία αυτά ξεπερνούν λιγότερο από πέντε διαδοχικές τιμές το κατώφλι τότε υπάρχει απότομη μεταβολή, ενώ αντίστοιχα αν υπάρχουν περισσότερες από πέντε διαδοχικές τιμές που ξεπερνούν το κατώφλι τότε υπάρχει σταδιακή μεταβολή.

Η γραφική παράσταση της $D(j)$ ποσότητας φαίνεται στο παρακάτω σχήμα όπου οι υψηλές κορυφές παρατηρούνται στα σημεία με απότομη μεταβολή ενώ οι σταθερές τιμές αντιπροσωπεύουν εικονοπλαίσια με παρόμοιο σκηνικό.



Στην ποσότητα $D(j)$ εφαρμόζονται και οι τέσσερις μέθοδοι κατωφλίωσης. Πριν όμως εφαρμοστούν οι μέθοδοι είναι απαραίτητη μια προεπεξεργασία του σήματος. Το $D(j)$ είναι ένα σήμα που τις περισσότερες φορές εμφανίζει πολλές κοντινές μεταβολές με μεγάλες αυξομειώσεις και για αυτό το λόγο γίνεται μια ομαλοποίηση (smoothing) με ένα φίλτρο μέσης τιμής.

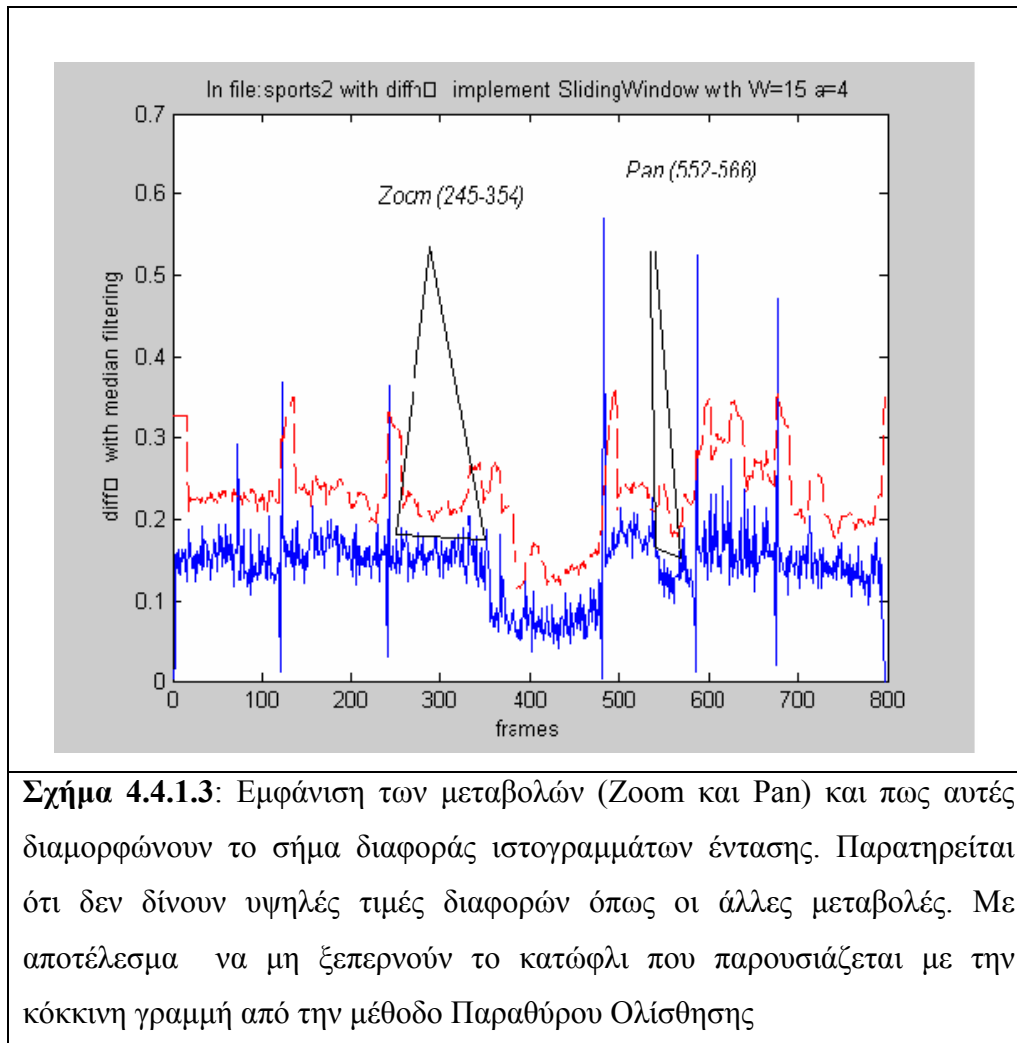
Στη συνέχεια εφαρμόζονται οι μέθοδοι στο νέο ομαλοποιημένο σήμα και ελέγχεται σε ποια σημεία αυτό ξεπερνάει το κατώφλι (threshold). Οι μεταβολές όπως το zoom, pan και tilt ουσιαστικά δεν προκαλούν σημαντική αλλαγή στην αναλογία των τιμών έντασης όπως εμφανίζονται στο ιστόγραμμα και κατ' επέκταση δεν παρουσιάζεται μεγάλη αλλαγή μεταξύ των ιστογραμμάτων των διαδοχικών εικονοπλαισίων. Το σήμα $D(j)$ δεν μεταβάλλεται αρκετά έτσι ώστε οι μέθοδοι κατωφλίωσης να το θεωρήσουν ως σταδιακή μεταβολή. Το συμπέρασμα αυτό το επαληθεύει και το παρακάτω σχήμα στο οποίο υπάρχουν και τα τρία είδη μεταβολών και φαίνεται η απόκριση τους στο σήμα $D(j)$ με την εφαρμογή της μεθόδου Διπλής Σύγκρισης.



Σχήμα 4.4.1.2: Εμφάνιση και των τριών μεταβολών (Cut,Dissolve,Zoom) και πως αυτές διαμορφώνουν το σήμα διαφοράς ιστογραμμάτων έντασης. Παρατηρείται ότι το Zoom δεν δίνει υψηλές τιμές διαφορών όπως οι άλλες μεταβολές (Cut και Dissolve).

Σε αυτό το παράδειγμα φαίνεται ένα cut που συμβαίνει στο (149) εικονοπλαίσιο, ένα Dissolve στα (309-331) ενώ στη συνέχεια στα εικονοπλάσια (328-432) υπάρχει ένα Zoom. Στο σήμα δεν παρατηρείται κάποια αλλαγή και συνεπώς οι μέθοδοι κατωφλίωσης αδυνατούν να ανιχνεύσουν κάτι τέτοιο σαν σταδιακή μεταβολή.

Ομοίως στο παρακάτω σχήμα

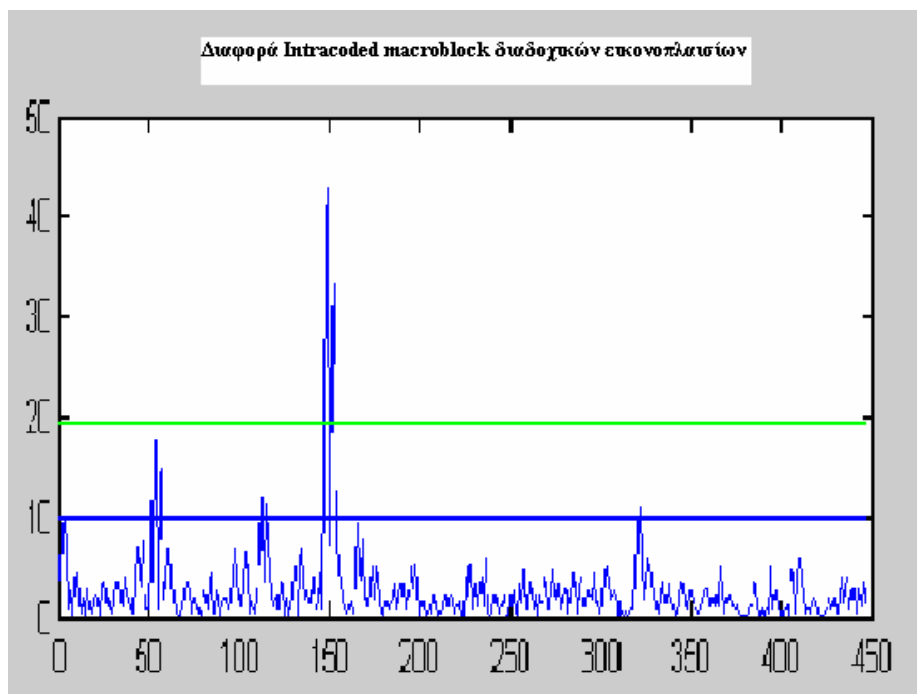


Παρόλα αυτά οι μέθοδοι για την ανίχνευση των μεταβολών Zoom, Pan και Tilt πέραν του ότι δίνουν πληροφορία για το περιεχόμενο του βίντεο θα χρησιμοποιηθούν για την εξάλειψη των false positives που υπάρχουν σε μεθόδους ανίχνευσης μεταβολών που βασίζονται σε intracoded macroblocks και σε τέτοιες μεθόδους το σήμα μας επηρεάζεται έντονα από την ύπαρξη Zoom, Pan και Tilt.

4.4.2 Μέθοδος διαφορών των Intracoded macroblocks

Από την μελέτη των διανυσμάτων κίνησης εξάγεται πληροφορία για το πλήθος των intracoded macroblocks δηλαδή των macroblocks των οποίων η κίνηση τους στο επόμενο εικονοπλαίσιο δεν μπορεί να προβλεφτεί. Αυτό συμβαίνει στις περιπτώσεις που υπάρχει ένα Cut στην ροή του βίντεο, αφού δεν υπάρχει πληροφορία που να μεταφέρεται από το εικονοπλαίσιο που προηγείται από το Cut στο επόμενο του. Το ίδιο συμβαίνει όταν εμφανίζεται Dissolve, Zoom, Pan και Tilt όπου και εκεί δεν μπορούν να προβλεφτούν τα διανύσματα κίνησης και συνεπώς υπάρχουν αρκετά intracoded macroblocks.

Για αυτό το λόγο δημιουργούνται τα διαγράμματα διαφοράς αριθμού intracoded macroblocks μεταξύ δυο διαδοχικών εικονοπλαισίων. Τα διαγράμματα αυτά με τρόπο ανάλογο της έντασης που περιγράφηκε στο Κεφάλαιο 4.4.1 παρουσιάζουν απότομες κορυφές σε πιθανά σημεία που εμφανίζονται Cuts και εξίσου ομαλές κορυφές σε σημεία που πιθανόν εμφανίζεται Dissolve, Zoom, Pan και Tilt.



Σχήμα 4.4.2.1: Η διαφορά Intracoded macroblocks διαδοχικών εικονοπλαισίων δίνει πληροφορία αντίστοιχη με εκείνη των διαφορών ιστογραμμάτων έντασης.

Μέθοδος 4 :Ανίχνευση απότομων και σταδιακών μεταβολών με την χρήση διαφορών Intracoded macroblocks.

1. Υπολογίζεται το $D_{intra}(j)$ για διαδοχικά εικονοπλαισισια
2. Εφαρμόζονται αντίστοιχα και οι τέσσερις μέθοδοι κατωφλιωσης στο σήμα $D_{intra}(j)$.
3. Ελέγχονται τα σημεία στα οποία το $D(j)$ ξεπερνάει τα αντίστοιχα κατώφλια
4. Αν τα σημεία αυτά ξεπερνούν λιγότερο από πέντε διαδοχικές τιμές το κατώφλι τότε υπάρχει απότομη μεταβολή, ενώ αντίστοιχα αν υπάρχουν περισσότερες από πέντε διαδοχικές τιμές που ξεπερνούν το κατώφλι τότε υπάρχει σταδιακή μεταβολή

$D_{intra}(j)$ = διαφορά αριθμού intracoded vectors δυο διαδοχικών εικονοπλαισίων

Το Ta και εδώ είναι το κατώφλι που ορίζεται κάθε φορά ανάλογα με την μέθοδο που εφαρμόζεται. Στην Twin Comparison μέθοδο Ta είναι το χαμηλό Threshold Ts . Στα σήματα αυτά εφαρμόζονται και οι τέσσερις μέθοδοι thresholding και δίνουν αρκετά καλά αποτελέσματα όπως θα φανεί από τις πειραματικές μετρήσεις.

Το σήμα αυτό επηρεάζεται από μεταβολές Zoom, Pan και Tilt παρουσιάζοντας παρόμοια συμπεριφορά με τα Dissolves. Για παράδειγμα στο σχήμα μας 4.4.2.1 εφαρμόζουμε την Twin Comparison μέθοδο στο σήμα διάφορων Intracoded macroblocks και συμπεραίνουμε ότι υπάρχουν τρεις σταδιακές μεταβολές και μια απότομη μεταβολή. Οι δυο πρώτες σταδιακές μεταβολές που εμφανίζονται στα εικονοπλαισισια 50-65 και 148-165 είναι Tilt ενώ αντίθετα η μεταβολή που εμφανίζεται στα εικονοπλαισισια 309-331 είναι Dissolve. Άρα από αυτές τις μεταβολές οι μόνη που μας ενδιαφέρει είναι η τελευταία. Χρησιμοποιώντας όμως μονάχα την πληροφορία από τα Intracoded macroblocks δεν μπορούμε να κάνουμε αυτή την διάκριση. Συνεπώς η μέθοδος αυτή θα χρειαστεί να συνδυαστεί με τα αποτελέσματα της Μεθόδου 2 (Ανίχνευσης Zoom, Pan και Tilt) για να απορρίψουμε τα αποτελέσματα που τελικά δεν είναι Dissolves.

4.5 Αλγόριθμος ανίχνευσης ορίων στιγμιότυπου

Ολοκληρώνοντας την περιγραφή των τεχνικών με την βοήθεια των οποίων ανιχνεύουμε το σύνολο των μεταβολών είναι απαραίτητο να ορίσουμε ένα αλγόριθμο ο οποίος θα καθορίζει τις διαδικασίες και τα κριτήρια με τα οποία ένα σημείο στην χρονοσειρά εικόνων αποτελεί όριο αλλαγής στιγμιότυπου.

Να αναφέρουμε πάλι ότι σημασιολογικά όριο ενός στιγμιότυπου (shot boundary) είναι το σημείο στο οποίο εμφανίζεται οι μεταβολές Cut και Dissolve. Θεωρητικά το Cut εμφανίζεται σαν μεταβολή που έχει διάρκεια ενός εικονοπλαισίου και μας δίνει ένα σημείο συνεπώς αυτό ισοδυναμεί με το όριο. Στο Dissolve που η διάρκεια της μεταβολής είναι περισσότερα από ένα εικονοπλαίσια έχουμε ένα διάστημα τιμών, σε αυτή την περίπτωση ορίζουμε την αρχή του διαστήματος σαν το τέλος ενός στιγμιότυπου και το τέλος του διαστήματος σαν την έναρξη ενός νέου στιγμιότυπου.

Αλγόριθμος 1 : Ανίχνευσης ορίων στιγμιότυπου

A. Εφαρμόζουμε και τις τέσσερις μεθόδους

Μέθοδος 1 : Διαχωρισμός Cut από Flash

Θεωρούμε το σύνολο των σημείων που ανιχνεύουμε $flash = \{F\}$

Μέθοδος 2 : Ανίχνευση Zoom, Pan και Tilt

Θεωρούμε τα σύνολα με $Zoom = \{Z\}$, $Pan = \{P\}$, $Tilt = \{T\}$

Μέθοδος 3 : Ανίχνευση απότομων και σταδιακών μεταβολών με την χρήση διαφορών ιστογραμμάτων έντασης

Θεωρούμε τα σύνολα με απότομες μεταβολές (Abrupt Transition) $\{AT_{diff_hist}\}$ και με σταδιακές μεταβολές (Gradual Transition) $\{GT_{diff_hist}\}$

Μέθοδος 4 :Ανίχνευση απότομων και σταδιακών μεταβολών με την χρήση

διαφορών Intracoded macroblocks

Ομοίως τα σύνολα με απότομες μεταβολές (Abrupt Transition) $\{AT_{Intra_mb}\}$ και με σταδιακές μεταβολές (Gradual Transition) $\{GT_{Intra_mb}\}$

B. Όρια στιγμιότυπου (Shot Boundaries) είναι το σύνολο $\{SB\}$ και ισούται με την ένωση Cuts και Dissolves. Τα Cuts προκύπτουν από την τομή των αποτελεσμάτων των δυο μεθόδων απότομων μεταβολών αφαιρώντας τα αποτελέσματα flashes από την μέθοδο ανίχνευσης flash. Τα flash αφαιρούνται από την τομή τους γιατί η μεταβολή flash επηρεάζει και τις διαφορές ιστογράμματος και τις διαφορές των Intracoded διανυσμάτων. Από την άλλη τα Dissolves προκύπτουν από την τομή των αποτελεσμάτων της μεθόδου ανίχνευσης σταδιακών μεταβολών με την χρήση διαφοράς ιστογραμμάτων με το σύνολο των αποτελεσμάτων από την μέθοδο ανίχνευσης σταδιακών μεταβολών με την χρήση Intra-coded macroblocks αφαιρώντας από την δεύτερη μέθοδο τα false positives zoom, pan και tilt, τα αποτελέσματα δηλαδή της μεθόδου ανίχνευσης zoom, pan και tilt με τη μέθοδο διασποράς γωνιών. Εδώ δεν αφαιρούμε τις μεταβολές zoom, pan και tilt από την τομή τους γιατί ουσιαστικά επηρεάζουν μονάχα την μέθοδο με Intracoded διανύσματα.

$$\{SB\} = \{C\} \cup \{D\} \quad (1)$$

$$\{C\} = \left(\left(\{AT_{diff_hist}\} \cap \{AT_{intra_mb}\} \right) - \{F\} \right) \quad (2)$$

$$\{D\} = \left(\{GT_{diff_hist}\} \cap \left(\{GT_{intra_mb}\} - \left(\{Z\} \cup \{P\} \cup \{T\} \right) \right) \right) \quad (3)$$

Κεφάλαιο5

Πειραματικά αποτελέσματα- Συμπεράσματα

5.1 Πειραματικά αποτελέσματα

Η μελέτη των αλγόριθμων έγινε σε ένα εκτενέστερο σύνολο δειγμάτων που αποτελείται από συνολικά 64 αρχεία βίντεο (40.741 frames 1630 sec) που ανήκουν στις παρακάτω κατηγορίες και τα οποία δημιουργήθηκαν με τέτοιο τρόπο ώστε να συμπεριλαμβάνουν όλα τα είδη των εφε που θα χρειαστεί να ανιχνευτούν (zoom, pan-tilt, dissolve, cut).

Avertissements=4, Animation=17, Documentary=15, Sports=12, News=11, Video clips=4 (Τα δείγματα αντλήθηκαν από [26-31] ιστοσελίδες)

Η προσπάθεια ένταξης όλων των παραπάνω μεταβολών περισσότερες από μια φορές την κάθε μια (πχ: 3 cuts, 2 dissolves, 3 zoom κλπ) σε αρχεία βίντεο μέσης διάρκειας 30 sec έκανε αρκετά δύσκολη την προσπάθεια διακριτοποίησης αυτών των σημάτων. Τα σήματα αυτά εμφανίζονται με πολλές κοντινές αλλαγές και διαφέρουν αρκετά από σήματα που δίνουν βίντεο με φυσική ροή λήψης.

Μετά την δημιουργία των αρχείων βίντεο χρειαζόταν και η περιγραφή του περιεχομένου τους όσον αφορά τις μεταβολές. Επομένως κάποιος εξωτερικός παρατηρητής ο οποίος παρακολούθησε όλα τα βίντεο και κατέγραψε σε ποια σημεία υπάρχουν cuts, dissolves, zooms, pans, tilts και flashes. Αυτά αποθηκεύτηκαν σε ένα αρχείο text το οποίο αποτέλεσε και το ground truth. Στην συνέχεια τα αποτελέσματα του κάθε αλγόριθμου συγκρίνονταν αυτόματα με αυτό το αρχείο για να προκύψει τελικά και η απόδοση των μεθόδων.

Το μέτρο το οποίο χρησιμοποιήθηκε για τον υπολογισμό της απόδοσης των μεθόδων είναι οι ποσότητες **precision** και **recall** που χρησιμοποιούνται ευρέως σε Information Retrieval τεχνικές.

Precision ορίζεται το κλάσμα των σωστά ανιχνεύσιμων μεταβολών προς το σύνολο των ανιχνεύσιμων μεταβολών

$$Precision = \frac{\text{Σωστες_Ανιχνετισιμες_Μεταβολες}}{\text{Συνολικες_Ανιχνευσιμες_Μεταβολες}}$$

$$0 \leq Precision \leq 1$$

Η τιμή *Precision* κοντά στο 1 δηλώνει ότι αυτό που ανιχνεύει η μέθοδος είναι σωστό και ότι δεν υπάρχουν αρκετά false positives

Recall ορίζεται το κλάσμα των σωστά ανιχνεύσιμων μεταβολών προς το σύνολο των πραγματικών μεταβολών

$$Recall = \frac{\text{Σωστες_Ανιχνευσιμες_Μεταβολες}}{\text{Πραγματικες_Μεταβολες}}$$

$$0 \leq Recall \leq 1$$

Η τιμή *Recall* κοντά στο 1 δηλώνει ότι ο αλγόριθμος ανιχνεύει όλες τις αλλαγές που συμβαίνουν στο βίντεο που μπορεί όμως μέσα σε αυτές πολλές να είναι false positives. Η ιδανική κατάσταση για ένα αλγόριθμο είναι να έχει *Precision=1* και *Recall=1*, που σημαίνει ότι ανιχνεύει όλες τις αλλαγές και είναι και σωστές .

5.2 Αποτελέσματα από την Μέθοδο 1, ανίχνευσης φλας

Ο αριθμός των δειγμάτων τα οποία εξετάστηκαν με βάση την Μέθοδο 1 ήταν πολύ μικρός και οφειλόταν στην έλλειψη βίντεο με εμφάνιση flashes. Στον παρακάτω πίνακα φαίνονται οι ποσότητες Precision και Recall για 5 εξεταζόμενα βίντεο που περιείχαν συνολικά 13 flashes. Όπως επίσης και το σύνολο των flashes που ανίχνευσε σωστά (Correct flash) η μέθοδος, όπως επίσης και το συνολικό αριθμό των flashes που υπήρχαν (Total flash) στα δείγματα μας σύμφωνα με την αρχική ανάλυση από εξωτερικό παρατηρητή.

<i>Ανίχνευση flash με τη χρήση του Μεθόδου 1</i>		
<i>5 βίντεο 3367 εικονοπλαίσια 135 sec</i>		
<i>Πίνακας 5.2.1</i>		
Μέθοδος 1	Precision Flash	Recall Flash
	0,438	0,923
	Correct flash=12 Total flash=13	

Παρόλο που τα δείγματα δεν επαρκούν στο να δώσουν ένα αντικειμενικό συμπέρασμα φαίνεται πως η Μέθοδος 1 τείνει να έχει καλή απόδοση στο να ανιχνεύει όλα τα flash που υπάρχουν στο βίντεο (υψηλό Recall) αλλά ανιχνεύει και πολλές απότομες μεταβολές που δεν είναι flashes, αρκετά false positives (χαμηλό Precision), κάτι που είχε αναφερθεί και στην αρχική περιγραφή της μεθόδου στο κεφαλαίο 4.2.

5.3 Αποτελέσματα από την Μέθοδο 2, ανίχνευσης zoom, pan και tilt

Οι ποσότητες Precision και Recall μετρήθηκαν για την Μέθοδο 2 σε ένα σύνολο δειγμάτων από 38 βίντεο. Για τον υπολογισμό κατωφλιού στην Μέθοδο 2 εφαρμόστηκαν και οι τέσσερις μέθοδοι κατωφλίωσης και σε κάθε περίπτωση μετρήθηκαν οι ποσότητες αυτές. Τα αποτελέσματα φαίνονται στον Πίνακα 5.3.1, επίσης παρουσιάζονται και οι σωστές ανιχνεύσιμες μεταβολές από την μέθοδο και οι ολικές μεταβολές όπως έχουν αναλυθεί από τον εξωτερικό παρατηρητή

<p><i>Ανίχνευση zoom-pan και tilt με τη χρήση της Μεθόδου 2</i> <i>«Διασπορά γωνιών διανυσμάτων κίνησης»</i> <i>38 βίντεο 25368 εικονοπλαίσια 1014 sec</i></p>				
Πίνακας 5.3.1				
Μέθοδος Κατωφλίωσης	<u>Prec Zoom-Pan</u>	<u>Rec Zoom-Pan</u>	<u>Prec Tilt</u>	<u>Rec Tilt</u>
Adaptive	0,499	0,578	0,714	0,682
Correct zoom-pan=37 Total zoom-pan=64			Correct-tilt=28 Total-tilt=41	
Sliding Window	0,389	0,678	0,536	0,823
Correct zoom-pan=43 Total zoom-pan=64			Correct-tilt=33 Total-tilt=41	
Twin Comparison	0,576	0,743	0,692	0,729
Correct zoom-pan=47 Total zoom-pan=64			Correct-tilt=29 Total-tilt=41	
Twin Comparison Window	0,35	0,634	0,54	0,689
Correct zoom-pan=40 Total zoom-pan=64			Correct-tilt=27 Total-tilt=41	

Τα αποτελέσματα που έχουν εξαχθεί από την εφαρμογή της Μεθόδου 2 είναι αρκετά ικανοποιητικά. Πρέπει να ληφθεί υπόψη ότι στα δείγματα συμπεριλαμβάνονται όλες οι κατηγορίες των μεταβολών με αποτέλεσμα το σήμα να μεταβάλλεται συνεχώς και να γίνεται ακόμα πιο δύσκολη την ανίχνευση τους

5.4 Αποτελέσματα από την Μέθοδο 3, ανίχνευσης απότομων και σταδιακών μεταβολών με την χρήση διαφορών ιστογραμμάτων έντασης.

Στον παρακάτω πίνακα φαίνονται οι αποδόσεις της Μεθόδου 3 στην ανεύρεση απότομων και σταδιακών μεταβολών για κάθε μια από τις τέσσερις μεθόδους κατωφλίωσης .

<p><i>Ανίχνευση απότομων και σταδιακών μεταβολών με τη χρήση της Μεθόδου 3 «Διαφορές ιστογραμμάτων έντασης» 64 βίντεο 40741 εικονοπλαίσια 1630 sec</i></p> <p><i>Πίνακας 5.4.1</i></p>				
Μέθοδος Κατωφλίωσης	<u>Prec Abrupt</u>	<u>RecAbrupt</u>	<u>Prec Gradual</u>	<u>Rec Gradual</u>
Adaptive	0,579	0,943	0,553	0,439
Correct cut=336 Total cut=357			Correct dis=69 Total dis= 157	
Sliding Window	0,517	0,935	0,483	0,401
Correct cut=334 Total cut=357			Correct dis=63 Total dis= 157	
Twin Comparison	0,742	0,844	0,611	0,324
Correct cut=301 Total cut=357			Correct dis=51 Total dis= 157	
Twin Comparison Window	0,526	1	0,667	0,375
Correct cut=357 Total cut=357			Correct dis=59 Total dis=157	

Από τα αποτελέσματα του παραπάνω πίνακα εξάγονται τα εξής συμπεράσματα. Χαρακτηριστικό των προσαρμοστικών μεθόδων κατωφλίσωσης και των μεθόδων με παράθυρο είναι ότι εμφανίζουν υψηλότερο Recall από Precision, αυτό οφείλεται στο ότι οι τιμές κατωφλίου τους βρίσκονται πιο κοντά στο σήμα και ανιχνεύουν ευκολότερα τις αλλαγές που πιθανόν συμβαίνουν. Από την άλλη όμως εμφανίζουν χαμηλότερο Precision γιατί συνυπολογίζουν πιο εύκολα και αλλαγές στο σήμα που πιθανόν έχουν προέλθει από θόρυβο ή κάποια τυχαία εσωτερική κίνηση αντικειμένων

Το αντίθετο συμβαίνει στην μέθοδο Διπλής Σύγκρισης όπου οι τιμές κατωφλίου δεν ακολουθούν τόσο έντονα το σήμα και τις μεταβολές του. Αυτό έχει σαν αποτέλεσμα να υπολογίζει αξιόπιστα τις πραγματικές μεταβολές, υψηλότερο Precision, αλλά να χάνει και πιθανές μεταβολές που λόγω της απόστασης του από το σήμα δεν μπορεί να αντιληφθεί, χαμηλότερο Recall.

Όσον αφορά την πληροφορία που παίρνουμε από την διάφορα ιστογραμμάτων έντασης, είναι αρκετά αξιόπιστη για να δώσει ικανοποιητικά αποτελέσματα στην ανίχνευση ορίων στιγμιότυπου. Μάλιστα αν μπορούσε να βελτιωθεί ακόμα η αποδοτικότητα των μεθόδων κατωφλίσωσης η πληροφορία από την διάφορα ιστογραμμάτων θα γινόταν περισσότερο αξιοποιήσιμη.

5.5 Αποτελέσματα από την Μέθοδο 4, ανίχνευσης απότομων και σταδιακών μεταβολών με την χρήση *intracoded macroblocks*

Στον παρακάτω πίνακα φαίνονται οι αποδόσεις της Μεθόδου 3 στην ανεύρεση απότομων και σταδιακών μεταβολών για κάθε μια από τις τέσσερις μεθόδους κατωφλίωσης

Ανίχνευση απότομων και σταδιακών μεταβολών με τη χρήση της μεθόδου 4 « Intracoded vectors » 64 βίντεο 40741 εικονοπλαίσια 1630 sec				
Πίνακας 5.5.1				
Μέθοδος Κατωφλίωσης	<u>Prec Abrupt</u>	<u>RecAbrupt</u>	<u>Prec Gradual</u>	<u>Rec Gradual</u>
Adaptive	0,619	0,723	0,419	0,369
Correct cut=258 Total cut=357			Correct diss=58 Total diss=157	
Sliding Window	0,671	0,800	0,407	0,414
Correct cut=286 Total cut=357			Correct diss=65 Total diss= 157	
Twin Comparison	0,700	0,885	0,476	0,388
Correct cut=316 Total cut=357			Correct diss=61 Total diss= 157	
Twin Comparison Window	0,628	0,959	0,395	0,471
Correct cut=342 Total cut=357			Correct diss=74 Total diss=157	

Παρατηρούμε ότι οι τιμές Precision και Recall για τις απότομες μεταβολές με την χρήση και των τεσσάρων μεθόδων κατωφλίωσης κυμαίνονται στα ίδια επίπεδα με τα

αποτελέσματα της Μεθόδου 3 (Πίνακας 5.4.1) .Όσον αφορά τώρα τις σταδιακές μεταβολές παρατηρείται μια μείωση των τιμών Precision και Recall. Αυτό ήταν κάτι που περιμέναμε και οφείλεται στο γεγονός ότι στην ανίχνευση σταδιακών μεταβολών με Intracoded macroblock γνωρίζουμε ότι ανιχνεύουμε και μεταβολές όπως Zoom, Pan και Tilt με αποτέλεσμα να πέφτει η απόδοση της μεθόδου. Την πληροφορία όμως αυτή στην συνέχεια θα την συνδυάσουμε τα αποτελέσματα της Μεθόδου 1 και 2 για την εξάλειψη false positives που προκαλούνται από (Flash, Zoom, Pan, Tilt) και θα παρατηρήσουμε μια ουσιαστική βελτίωση.

5.6 Αποτελέσματα από τον Αλγόριθμο 1, υπολογισμό ορίων στιγμιότυπου

Στους παρακάτω πίνακες θα παρουσιάσουμε τις αποδόσεις από του όρους που συνθέτουν το τελικό κριτήριο καθορισμού ορίου στιγμιότυπου του αλγόριθμου 1 .

$$\{SB\} = \{C\} \cup \{D\} \quad (5.6.1)$$

Ξεκινώντας από τον όρο

$$\{C\} = \left(\left(\{AT_{diff_hist}\} \cap \{AT_{intra_mb}\} \right) - \{F\} \right) \quad (5.6.2)$$

Η παραπάνω σχέση ορίζει τα Cuts σαν την τομή των αποτελεσμάτων των απότομων μεταβολών από τις Μεθόδους 3 και 4 αφαιρώντας τα αποτελέσματα Flash που δίνει η Μέθοδος 1.

$\{C\} = \left(\left(\{AT_{diff_hist}\} \cap \{AT_{intra_mb}\} \right) - \{F\} \right)$				
Πίνακας 5.6.1				
Μέθοδος	Precision Cut		Recall Cut	
Κατωφλίωσης				
Adaptive	0,670	+12 %	0,783	-6 %
Sliding Window	0,683	+ 13 %	0,798	-8 %
Twin Comparison	0,799	+ 11 %	0,812	- 6 %
Twin Comparison Window	0,675	+ 17 %	0,903	- 7 %
Πίνακας 5.6.1: Οι τιμές Precision και Recall για την εύρεση Cut σύμφωνα με τη σχέση 5.6.2 και το ποσοστό αύξησης ή μείωσης τους σε σχέση με τις τιμές που έδιναν ξεχωριστά οι Μέθοδοι 3 και 4				

Στη συνέχεια από την σχέση

$$\{D\} = \left(\{GT_{diff_hist}\} \cap \left(\{GT_{intra_mb}\} - (\{Z\} \cup \{P\} \cup \{T\}) \right) \right) \quad (5.6.3)$$

Προκύπτουν τα Dissolves σαν την τομή των αποτελεσμάτων της Μεθόδου 3 και της Μεθόδου 4 . Από τα αποτελέσματα της Μεθόδου 4 έχουν αφαιρεθεί τα false positives (Zoom, Pan, Tilt)

$\{D\} = \left(\{GT_{diff_hist}\} \cap \left(\{GT_{intra_mb}\} - (\{Z\} \cup \{P\} \cup \{T\}) \right) \right)$				
Πίνακας 5.6.2				
Μέθοδος Κατωφλιωσης	Precision Dissolve		Recall Dissolve	
Adaptive	0,523	+10 %	0,387	-4 %
Sliding Window	0,493	+ 11 %	0,378	-7 %
Twin Comparison	0,652	+ 20 %	0,341	- 4 %
Twin Comparison Window	0,643	+ 20 %	0,401	- 5 %
Πίνακας 5.6.2: Οι τιμές Precision και Recall για την εύρεση Dissolve σύμφωνα με τη σχέση 5.6.3 και το ποσοστό αύξησης ή μείωσης τους σε σχέση με τις τιμές που έδιναν ξεχωριστά οι Μέθοδοι 3 και 4				

Η μεταβολή που παρατηρείται στις τιμές και των δυο πινάκων είναι μια σημαντική αύξηση της τιμής Precision και μια μικρότερης τάξης μείωση της τιμής Recall. Αυτό ήταν κάτι που περιμέναμε να συμβεί αφού με την τομή των αποτελεσμάτων των δυο Μεθόδων αλλά και με την εξάλειψη των false positives, σε κάθε περίπτωση, τα αποτελέσματα που παίρνουμε είναι πιο ακριβή με κόστος βέβαια να χάνουμε και κάποιες καταβολές. Αυτή η συμπεριφορά μεταφράζεται με αύξηση της τιμής Precision και μείωση της τιμής Recall.

5.7 Συμπεράσματα –Μελλοντική εργασία

Στο τέλος αυτής της μελέτης είναι σκόπιμο να αναφερθούν συνοπτικά τα βασικά συμπεράσματα που απορρέουν από τα πειραματικά δεδομένα. Η βασική επιδίωξη στην προσπάθεια κατάτμησης του βίντεο είναι ο υπολογισμός μιας τιμής κατωφλίου η οποία προσαρμόζεται στις αλλαγές του σήματος. Οι γνωστές από την βιβλιογραφία μέθοδοι κατωφλιωσης δεν φαίνεται να εμφανίζουν σημαντικές διαφορές μεταξύ τους όσον αφορά την αποδοτικότητα τους. Η μέθοδος Παραθύρου Ολίσθησης παρουσιάζει μια σταθερότητα στην αποδοτικότητα της στο σύνολο των δειγμάτων. Αντίθετα οι άλλες τρεις μέθοδοι κατωφλιωσης με παράθυρο εμφανίζουν αυξομειώσεις στην απόδοση τους η οποία εξαρτάται από την συμπεριφορά του σήματος. Κάποιες από αυτές λειτουργούν αποδοτικότερα σε συγκεκριμένα είδη σημάτων ενώ οι υπόλοιπες σε διαφορετικά. Πέρα από τη διαδικασία υπολογισμού του κατωφλιού σημαντική είναι και η πληροφορία η οποία εξάγεται από την μελέτη των intracoded macroblocks. Με βάση την ποσότητα αυτή οι μέθοδοι μπορούν να ανιχνεύσουν ικανοποιητικά απότομες και σταδιακές μεταβολές ακόμα και αν στα αποτελέσματα συνυπάρχουν αρκετά false positives. Αυτά τα false positives που κατά κύριο λόγο είναι zoom, tilt και pan μπορούν εύκολα να ανιχνευθούν με συγκεκριμένες τεχνικές και να αφαιρεθούν στην συνέχεια από τα αρχικά αποτελέσματα.. Πρέπει να τονιστεί ότι παρόμοια διαδικασία δεν έχει νόημα να γίνει στις διαφορές ιστογραμμάτων έντασης γιατί αυτές δεν μεταβάλλονται σημαντικά σε μεταβολές όπως το zoom, pan και tilt.

Σαν μελλοντική δουλειά για την εργασία αυτή είναι η εξαγωγή εικονοπλαισίων κλειδιών (key frame) τα οποία αποτελούν και το χαρακτηριστικό σημείο για κάθε στιγμιότυπο. Συνεχίζοντας την μελέτη της κατάτμησης στην επόμενη σημασιολογική ομάδα η οποία είναι η σκηνή (scene). Μια άλλη προσέγγιση μελλοντικής δουλειάς είναι η ανίχνευση των περαιτέρω μορφών κίνησης όπως η εσωτερική κίνηση ή η ανίχνευση αντικειμένων. Επίσης θα μπορούσε να υλοποιηθεί ένας συνδυαστικός αλγόριθμος εύρεσης ορίων στιγμιότυπου χρησιμοποιώντας τα αποτελέσματα από την κάθε μέθοδο με ένα πιο σύνθετο τρόπο, αυξάνοντας την αποδοτικότητα του αλγορίθμου. Τέλος ένα επίσης ουσιαστικό κομμάτι εργασίας είναι η μελέτη της βελτίωσης των μεθόδων κατωφλιωσης που χρησιμοποιήσαμε. Μια σημαντική

βελτίωση σε αυτό το πεδίο συνεισφέρει άμεσα στην αύξηση της αποδοτικότητας των μεθόδων κατάτμησης.

1.Βιβλιογραφία

- [1] Y. Tonomura and S. Abe. Content oriented visual interface using video icons for visual database systems. *Journal of Visual Languages and Computing*, 1(2):183–198, June 1990.
- [2] A. Nagasaka and Y. Tanaka. Automatic video indexing and full-video search for object appearances. In *IFIP Working Conference on Visual Database Systems*, pages 113–127, Budapest, Hungary, October 1991
- [3] U. Gargi and R. Kasturi. An evaluation of color histogram-based methods in video indexing. In *International Workshop on Image Databases and Multimedia Search*, pages 75–82, Amsterdam, The Netherlands, August 1996
- [4] D. Pye, N.J. Hollinghurst, T.J. Mills, and K.R. Wood. Audio-visual segmentation for contentbased retrieval. In *International Conference on Spoken Language Processing*, Sydney, Australia, December 1998
- [5] C. O’Toole, A. Smeaton, N. Murphy, and S. Marlow. Evaluation of automatic shot boundary detection on a large video test suite. In *Conference on Challenge of Information Retrieval*, Newcastle, UK, February 1999
- [6] X.U. Cabedo and S.K. Bhattacharjee. Shot detection tools in digital video. In *Workshop on Non-linear Model Based Image Analysis*, pages 121–126, Glasgow, Scotland, July 1998.
- [7] H.J. Zhang, A. Kankanhalli, and S.W. Smoliar. Automatic partitioning of full-motion video. *Multimedia Systems*, 1(1):10–28, July 1993.
- [8] W. Zhao, J. Wang, D. Bhat, K. Sakiewics, N. Nandhakumar, and W. Chang. Improving colorbased video shot detection. In *IEEE International Conference on*

Multimedia Computing and Systems, volume 2, pages 752–756, Florence, Italy, June 1999

- [9] A. Dailianas, R.B. Allen, and P. England. Comparison of automatic video segmentation algorithms. In *SPIE Conference on Integration Issues in Large Commercial Media Delivery Systems*, volume 2615, pages 2–16, Philadelphia, PA, October 1995. 36 S. Lef`evre, J. Holler, N. Vincent : A Review of Real-time Segmentation of Uncompressed Video Sequences for Content-Based Search and Retrieval RFAI publication: Real Time Imaging, to appear
- [10] Y. Taniguchi, A. Akutsu, and Y. Tonomura. Panorama excerpts: Extracting and packing panoramas for video browsing. In *ACM International Conference on Multimedia*, pages 427–436, Seattle, WA, November 1997.
- [11] R. Kasturi and R.C. Jain. Dynamic vision. In R. Kasturi and R.C. Jain, editors, *Computer Vision: Principles*, pages 469–480. IEEE Computer Society Press, Washington, 1991
- [12] M.S. Lee, Y.M. Yang, and S.W. Lee. Automatic video parsing using shot boundary detection and camera operation analysis. *Pattern Recognition*, 34(3):711–719, March 2001.
- [13] W. Xiong and J.C.M. Lee. Efficient scene change detection and camera motion annotation for video classification. *Journal of Computer Vision and Image Understanding*, 71(2):166–181, August 1998.
- [14] R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying production effects. *Multimedia Systems*, 7(2):119–128, March 1999
- [15] A.F. Smeaton, G. Gormley J. Gilvarry and, 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*, pages 45–62, Dublin, Ireland, September 1999.

- [16] R. Lienhart. Comparison of automatic shot boundary detection algorithms. In *SPIE Conference on Storage and Retrieval for Image and Video Databases*, volume 3656, pages 290–301, San Jose, CA, January 1999.
- [17] A. Akutsu, Y. Tonomura, H. Hashimoto, and Y. Ohba. Video indexing using motion vectors. In *SPIE Conference on Visual Communications and Image Processing*, volume 1818, pages 1522–1530, Boston, MA, November 1992.
- [18] M. Cherfaoui and C. Bertin. Temporal segmentation of videos: a new approach. In *SPIE Conference on Digital Video Compression: Algorithms and Technologies*, volume 2419, pages 38–47, San Jose, CA, February 1995
- [19] O. Fatemi, S. Zhang, and S. Panchanathan. Optical flow based model for scene cut detection. In *Canadian Conference on Electrical and Computer Engineering*, volume 1, pages 470–473, Calgary, Canada, May 1996
- [20] Dong Zhang, Wei Qi, Hong Jiang Zhang «A New Shot Boundary Detection Algorithm» *Multimedia Systems*, Boston 2-10-1999.
- [21] Boon-Lock Yeo, “Efficient processing of compressed images and video ” IEEE Computer Society Press, Boston 1993
- [22] C. Dulaverakis*, S. Vagionitis*, M. Zervakis, E. Petrakis «Adaptive Methods for Motion Characterization and Segmentation of MPEG Compressed Frame Sequences» (Department of Electronic and Computer Engineering Technical University of Crete Chania, Crete, Greece) 2003
- [23] Δρ. Σοφία Τσεκερίδου, Λέκτορας Δ.Π.Θ. Δρ. Σοφία Τσεκερίδου, Λέκτορας Δ.Π.Θ. «Ανθρώπινο Δίκτυο Ερευνητικής Και Τεχνολογικής Επιμόρφωσης Σε Τεχνολογίες Εξόρυξης Και Διαχείρισης Γνώσης Από Πολυμέσα»
-

Πολυτεχνείο Κρήτης, Χανιά, Πολυτεχνείο Κρήτης, Χανιά, 25 25 Ιουνίου Ιουνίου 200
2004

[24] Kobla, V., *et al.* *Compressed Domain Video Indexing Techniques Using DCT and Motion Vector Information in MPEG Video*. in *Storage and Retrieval for Image and Video Databases*. 1997

[25] *Joseph Gilvarry* « Extraction of Motion Vectors from an MPEG Stream»
School of Electronic Engineering, Dublin City University , Technical report 1999

2. Ιστοσελίδες

[26] <http://www.cs.cf.ac.uk/Dave/Multimedia/node261.html>

[27] <http://www-2.cs.cmu.edu/afs/cs/project/cil/ftp/html/vision.html>

[28] <http://www.mpeg1.de/www/inmovies.htm>

[29] <http://home.in.tum.de/~paula/mpeg/index.html>

[30] http://www.tmpgenc.net/e_main2.html

[31] http://faculty.fuqua.duke.edu/~charvey/Video/Digital_video.htm