

Πολυτεχνείο Κρήτης  
Τμήμα Ηλεκτρονικών Μηχανικών και  
Μηχανικών Ηλεκτρονικών Υπολογιστών



## Διπλωματική εργασία

«Συνδυασμός παραμετρικών μοντέλων και μοντέλων κίνησης  
για εντοπισμό επιπέδων με εφαρμογές στο MPEG-4 και 7»

Χανιά, Οκτώβριος 2004

Διονυσόπουλος Διονύσης

Επιβλέπων Καθηγητής: Ζερβάκης Μιχάλης (καθηγητής)

Εξεταστική Επιτροπή:

Πετράκης Ευριπίδης

(αναπληρωτής καθηγητής)

Χριστοδουλάκης Σταύρος

(καθηγητής)

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή της εργασίας μου κ.Ζερβάκη Μιχάλη για την συνεργασία και την ιδιαίτερη κατανόησή του καθόλη την διάρκεια της εργασίας αυτής.Η καθοδήγησή του ήταν πάντοτε πολυτιμότεος αρωγός στο φωτεινό μονοπάτι της έρευνας.Επίσης αφιερώνω την εργασία αυτή στους γονείς μου για όλα αυτά που μου προσέφεραν καθόλη την διάρκεια της πορείας μου στο Πολυτεχνείο.

## ΠΕΡΙΕΧΟΜΕΝΑ

Περιεχόμενα.....	3
Πίνακας σχημάτων.....	4
Περίληψη.....	5
Εισαγωγή-Ορισμος του προβλήματος.....	6
Διανύσματα κίνησης.....	7
Η χρησιμότητα των επιπέδων στα πρότυπα MPEG-4 και 7.....	9
Προηγούμενες μελέτες για την εύρεση επιπέδων.....	10
Η δομή της διπλωματικής.....	12
Υπολογισμός διανυσμάτων κίνησης με την μέθοδο Horn-Schunk.....	13
Ορισμός της οπτικής ροής.....	15
Η εξίσωση οπτικής ροής και η συνάρτηση κόστους.....	18
Υπολογισμός οπτικής ροής κατά Horn-Schunk.....	19
Παραμετρικά μοντέλα κίνησης σκοπός και λειτουργία τους.....	20
Η κίνηση της κάμερας και οι μαθηματικές σχέσεις που την διέπουν.....	24
Είδη παραμετρικών μοντέλων.....	25
Εκτίμηση παραμετρικού μοντέλου από O.F.E.....	27
Φιλτράρισμα του υπολογισμού των παραμετρικών διανυσμάτων με την χρήση βάρους.....	29
Εκτίμηση παραμετρικού μοντέλου από διανύσματα κίνησης.....	32
Multistage Affine Classification αλγόριθμος-Εισαγωγή.....	37
K-means αλγόριθμος.....	37
Η αναγκαιότητα χρήσης του multistage affine classification αλγορίθμου .....	39
Γενική περιγραφή του multistage affine classification αλγορίθμου.....	39
Διάγραμμα ροής.....	42
Περιγραφή του multistage affine classification αλγορίθμου.....	43
Υπολογισμός των αξιόπιστων (reliable) blocks.....	44
Διαδικασία ένωσης (merge procedure).....	45
Διαδικασία κατηγοριοποίησης (labeling procedure).....	47
Πειραματικά Αποτελέσματα.....	48-67
Σύγκριση των δύο μεθόδων-Συμπεράσματα-Μελλοντικές Επεκτάσεις...70	

## ΠΙΝΑΚΑΣ ΣΧΗΜΑΤΩΝ

Σχήμα 1. Αντιστοίχιση του video στους άξονες.....	6
Σχήμα 2.Πιθανή αλληλουχία καρέ μεταξύ των δύο καρέ.....	8
Σχήμα 3.Η οπτική ροή ( $\gamma$ ) μεταξύ των καρέ ( $\alpha$ ) και ( $\beta$ ).....	18
Σχήμα 4.(α)Περιστρεφόμενος κύλινδρος,(β) η οπτική ροή και ( $\gamma$ ) τα διανύσματα κίνησης.....	19
Σχήμα 5.Οι εξισώσεις των παραμετρικών μοντέλων.....	27
Σχήμα 6.Η τελική κατάσταση του k-means.....	38
Σχήμα 7.Η διαφορά των παραμετρικών διανυσμάτων από O.F.E. και motion vectors.....	41
Σχήμα 8.Διαδοχικά καρέ του “flow garden”.....	50-51
Σχήμα 9.Τα layers του “flow garden” από optical flow equation.....	52-53
Σχήμα 10. Διαδοχικά καρέ του “coast guard”.....	53-54
Σχήμα 11. Τα layers του “coast guard” από optical flow equation.....	54-55
Σχήμα 12. Διαδοχικά καρέ του “ping pong ”.....	55-56
Σχήμα 13. Τα layers του “ping pong” από optical flow equation.....	58-59
Σχήμα 14. Διαδοχικά καρέ του “flow garden.....	59
Σχήμα 15. Τα layers του “flow garden” από motion vectors.....	59-60
Σχήμα 16. Διαδοχικά καρέ του “coast guard”.....	60-61
Σχήμα 17. Τα layers του “coast guard” από motion vectors	62
Σχήμα 18. Διαδοχικά καρέ του “ping pong ”.....	63-64
Σχήμα 19. Τα layers του “ping pong” από motion vectors.....	64-65

## **ΠΕΡΙΛΗΨΗ**

Για την σωστή και αποδοτική μετετροπή του βίντεο σε MPEG-4 και 7 χρειάζεται ένας προσαρμόσιμος και ευέλικτος τρόπος εύρεσης των επιπέδων που εμφανίζονται σε αυτό. Στην εργασία αυτή προτείνονται τρόποι που εκμεταλεύονται την χωρική πληροφορία ,την κίνηση που ανιχνεύεται ανάμεσα στα καρέ του βίντεο καθώς και τα παραμετρικά μοντέλα κίνησης. Η εύρεση των επιπέδων βασίζεται σε ένα πολλαπλό βαθμίδων κατηγοριοποίησης αλγόριθμο ο οποίος συγκρίνει διανύσματα κίνησης που προκύπτουν τόσο από τα παραμετρικά μοντέλα κίνησης, όσο και από την κανονικοποίηση της εξίσωσης οπτικής ροής.

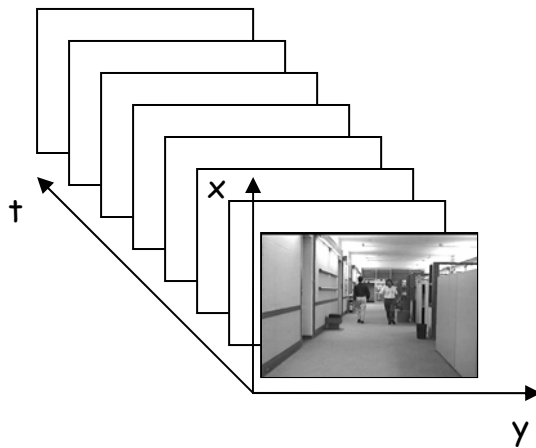
Η μέθοδος που προτείνουμε χρησιμοποιήθηκε επιτυχώς σε σκηνές με κινούμενη κάμερα γεγονός που την κάνει ευέλικτη σε φυσικές σκηνές όπου η κίνηση της κάμερας είναι δεδομένη.

# ΚΕΦΑΛΑΙΟ 1

## Εισαγωγή

### 1 α)Ορισμός του προβλήματος

Σκοπος της διπλωματικής αυτής είναι η εύρεση των επιπέδων (layers) που περιέχονται σε ένα βίντεο, από το πρότυπο συμπίεσης MPEG 4. Ως βίντεο ορίζεται μία ροή από εικόνες  $(x,y)$  στο χρόνο  $(t)$ . Οι εικόνες αυτές λέγονται καρέ του βίντεο και είναι η δισδιάστατη διακριτή αναπαράσταση του φυσικού χώρου (Σχήμα 1).



Σχήμα 1. Αντιστοίχιση του βίντεο στους άξονες

Ως layer σε αυτή την εργασία ορίζουμε μία περιοχή της εικόνας του βίντεο (καρέ) η οποία βρίσκεται κάτω από παρόμοια κίνηση. Δηλαδή τα pixels τα οποία ανήκουν στο ίδιο layer μοιράζονται κάποιο κοινό μοντέλο κίνησης. Ως μοντέλο θα μπορούσε να οριστεί κάποια μαθηματική σχέση η οποία στηρίζεται και περιγράφει την κίνηση (motion) μιας περιοχής. Θα μπορούσε να χρησιμοποιηθεί και ως μοντέλο η υφή (texture) μιας περιοχής. Συνήθως χρησιμοποιούνται μοντέλα κίνησης τα οποία είναι είτε μη παραμετρικά (κεφάλαιο 2) είτε παραμετρικά (κεφάλαιο 3).

Πιο συγκεκριμένα στην εργασία αυτή κάθε layer περιλαμβάνει περιοχές του frame στις οποίες τα διανύσματα κίνησης υπακούουν σε ένα μοντέλο. Γίνεται φανερό ότι βασικά στοιχεία της εργασίας αποτελούν οι αλγόριθμοι εύρεσης διανυσμάτων κίνησης και των παραμετρικών

μοντέλων, καθώς και η μεθοδολογία σύγκρισης της συμβατότητας διανυσμάτων κίνησης με ένα (ή περισσότερα) παραμετρικά μοντέλα.

Σαν αλγόριθμο υπολογισμού διανυσμάτων κίνησης από την ακολουθία εικόνων του βίντεο χρησιμοποιούμε αυτόν του Horn και Shunk που βασίζεται στην κανονικοποίηση της εξίσωσης οπτικής ροής.

Σαν παραμετρικό μοντέλο κίνησης χρησιμοποιούμε το affine model με έξι παραμέτρους. Υλοποιούμε δύο αλγόριθμους εύρεσης του παραμετρικού μοντέλου, μία από την ακολουθία εικόνων βάση της εξίσωσης οπτικής ροής και μία από τα (ήδη υπολογισμένα) διανύσματα κίνησης χρησιμοποιώντας ομαδοποίηση (k-means) των διανυσμάτων και weighted least squares για τον υπολογισμό των παραμέτρων που διέπουν κάθε ομάδα. Από τα παραμετρικά μοντέλα μπορούμε αμέσως να ορίσουμε το «παραμετρικό» διάνυσμα κίνησης σε κάθε pixel. Συνολικά λοιπόν μπορούμε να έχουμε δύο ειδών διανύσματα κίνησης σε κάθε pixel, ένα από την εξίσωση οπτικής ροής και ένα από το παραμετρικό μοντέλο της περιοχής.

Συγκρίνοντας αυτά τα δύο είδη διανυσμάτων μπορούμε άμεσα να ορίσουμε συμβατότητα (ή όχι) του κάθε pixel με κάθε παραμετρικό μοντέλο. Για να μειώσουμε το απαιτούμενο υπολογιστικό κόστος η σύγκριση αυτή γίνεται σε επίπεδο block και έπειτα (εάν χρειάζεται) σε επίπεδο pixel.

## **1α) Διανύσματα κίνησης (motion vectors)**

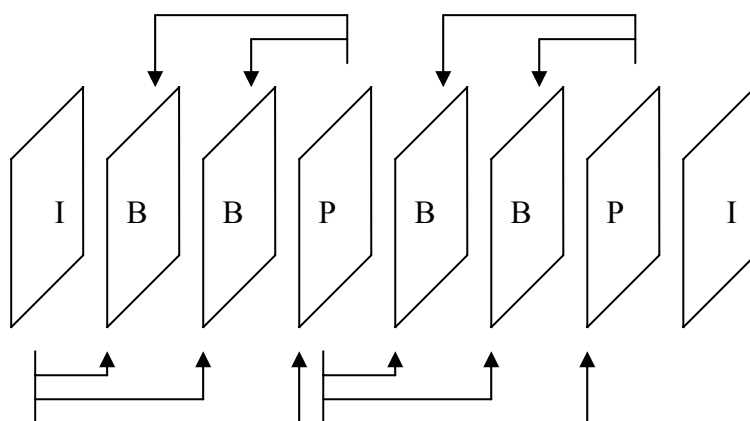
Προκειμένου να κατανοήσουμε καλύτερα την φύση του προβλήματος προχωράμε στον ορισμό των διανυσμάτων κίνησης που είναι και ο βασικός στόχος των προτύπων MPEG.

Το πρότυπο Mpeg1 καθώς και τα Mpeg2, Mpeg4 και 7 χρησιμοποιούν τριών ειδών καρέ. Τα I-καρέ, τα B-καρέ και τα P-καρέ. Το πρώτο είναι το καρέ αναφοράς και αποτελεί το πρώτο καρέ μιας σκηνής σε ένα video, το οποίο δεν εξαρτάται από τα υπόλοιπα καρέ του video. Το Mpeg1 προσφέρει χαμηλή συμπίεση και χρησιμοποιείται προκειμένου να μπορεί να γίνει γρήγορη τυχαία πρόσβαση. Για την συμπίεσή του χρησιμοποιείται ο προκαλών απώλεια ενέργειας JPEG αλγόριθμος συμπίεσης (lossy Jpeg). Μέχρι το επόμενο καρέ

αναφοράς, τα υπόλοιπα καρέ είναι τύπου B ή P. Τα P-καρέ χρησιμοποιούν πληροφορία από το προηγούμενο I-καρέ ή και τα προηγούμενα P-καρέ και προσφέρουν μέτρια συμπίεση. Τα B-καρέ είναι αυτά που προσφέρουν τη μεγαλύτερη συμπίεση και χρησιμοποιούν πληροφορία από προηγούμενα ή και μελλοντικά P-καρέ και B-καρέ (**Error! Reference source not found.**).

Είναι κατανοητό λοιπόν ότι προκειμένου να επιτευχθεί συμπίεση, χρησιμοποιούμε χρονική πληροφορία, δηλαδή την εκ των προτέρων γνώση ότι τα περισσότερα καρέ σε ένα βίντεο δεν είναι ανεξάρτητα μεταξύ τους (I-καρέ) αλλά είναι σχεδόν ίδια με τα προηγούμενα – επόμενα και μπορούμε να εξοικονομήσουμε χώρο αν δούμε και αξιοποιήσουμε αυτήν την σχέση. Δηλαδή τα I-καρέ εμφανίζονται όταν αλλάζει σκηνή στο βίντεο ή όταν οι απαιτήσεις επιβάλλουν I-καρέ για γρήγορη αναζήτηση. Τα υπόλοιπα καρέ του βίντεο είναι τύπου B και P και όσον αφορά τα αντικείμενα που αναζητάμε, τα αντιμετωπίζουμε με τον ίδιο τρόπο.

Η εύρεση των διανυσμάτων κίνησης ξεκινάει από το πρότυπο Mpeg1. Τα διανύσματα κίνησης περιγράφουν την κίνηση κάθε σημείου (pixel) ή περιοχής (block) ανάμεσα στα καρέ του video. Είναι η προσπάθεια εισαγωγής της χρονικής πληροφορίας στον αλγόριθμο συμπίεσης Mpeg. Το σύνολο των διανυσμάτων αυτών μεταξύ δυο καρέ του βίντεο καλείται οπτική ροή (optical flow).



Σχήμα 2. Πιθανή αλληλουχία μεταξύ δύο I καρέ



## **1β) Η χρησιμότητα των επιπέδων(layers) στα πρότυπα MPEG 4 και MPEG 7**

Σκοπός των προτύπων Mpeg είναι να περιγράψουν τον τρόπο με τον οποίο θα είναι αποθηκευμένη η πληροφορία του βίντεο στο μέσο αποθήκευσης μιας ψηφιακής συσκευής (Οπτικός δίσκος, Σκληρός Δίσκος κ. α.). Τα πρότυπα αυτά χρησιμοποιούνται για να επιτευχθεί συμπίεση του video. Η συμπίεση του video είναι απαραίτητη προκειμένου να υπάρχει επάρκεια χώρου στα υπάρχοντα αποθηκευτικά συστήματα (CD,DVD) και να είναι δυνατή η μετάδοση δημιουργώντας όσο το δυνατό λιγότερη κίνηση στο κανάλι επικοινωνίας. Ως κανάλι επικοινωνίας θεωρείται οποιοδήποτε δίκτυο υπολογιστών ή συσκευών δια μέσου του οποίου θα γίνει η μετάδοση του βίντεο.

Η περιγραφή του συμπιεσμένου κατά Mpeg4 και Mpeg7 βίντεο δεν επιβάλλει συγκεκριμένο τρόπο για την υλοποίηση της συμπίεσης. Ανάλογα με την φύση του βίντεο καθώς και με τις απαιτήσεις του κατασκευαστή σε χώρο, επεξεργαστική ισχύ του hardware και ευκρίνεια υπάρχουν διάφορες παράμετροι που πρέπει να προσδιοριστούν.

Γενικός στόχος της εξέλιξης νέων τρόπων και μεθόδων που υπακούουν στα παραπάνω πρότυπα είναι η μείωση του απαιτούμενου αποθηκευτικού χώρου και η ταχύτερη μετατροπή ανάμεσα στα πρότυπα ιδίως στο Mpeg4 που επικρατεί την χρονική στιγμή συγγραφής αυτής της μελέτης και στο Mpeg7 που είναι το μελλοντικό πρότυπο, όταν τόσο η τεχνολογία αλλά και η τεχνογνωσία το επιτρέψουν. Πάνω σε αυτή την κατεύθυνση μας δίνει αρκετές λύσεις και η αναπαράσταση του βίντεο σε επίπεδα.

Στόχος της μελέτης αυτής είναι η πιο αποτελεσματική αναζήτηση των επιπέδων που υπάρχουν στα καρέ του video, ούτως ώστε να επιτευχθεί μεγαλύτερη και πιο ποιοτική συμπίεση σύμφωνα με τα παραπάνω πρότυπα. Η γνώση των επιπέδων(layers) που απαρτίζουν ένα στιγμιότυπο ενός video είναι μια χρήσιμη πληροφορία. Συγκεκριμένα, μια προσέγγιση των

διαφορετικών ταχυτήτων που περιγράφουν τα επίπεδα αρκεί για σωστή συμπίεση. Η βασική ιδέα είναι ότι επίπεδο (layer) σε μια εικόνα είναι οτιδήποτε μοιράζεται κάποιο κοινό μαθηματικό μοντέλο κίνησης. Σε αυτό το σημείο θα πρέπει να διευκρινήσουμε ότι δεν χρησιμοποιούμε τις εννοιές στατικό και μη στατικό διότι τα επίπεδα έχουν διαφορετικές ταχύτητες μεταξύ τους. Εάν κάποιο επίπεδο (layer) παρουσιάζει την ίδια ταχύτητα για κάποια καρέ δεν είναι απαραίτητη η μετάδοσή του. Συνεπώς πετυχαίνουμε συμπίεση της πληροφορίας και εξοικονομούμε χώρο και χρόνο.

## **1γ) Προηγούμενες μελέτες για την εύρεση επιπέδων(layers)**

Διάφορες μελέτες έχουν λάβει χώρα για την εξαγωγή των επιπέδων (layers) βασιζόμενες κυρίως στην κίνηση. Αυτές περιλαμβάνουν μοντέλα εκτίμησης της κίνησης (motion estimation) και ομαδοποίηση pixel και περιοχών της εικόνας βασισμένη στον k-means αλγόριθμο [5]. Κάποιες άλλες μέθοδοι βασίζονται στην χρήση του Expectation-Maximization αλγορίθμου [10]. Μερικές μέθοδοι χρησιμοποιούν ένα είδος αρχικοποίησης [1] [2] [3]. Δηλαδή ορίζουν τον αριθμό των μοντέλων και το είδος της κίνησης που περιγράφουν αυτά. Χωρίς την σωστή χρήση της αρχικοποίησης οι μέθοδοι αυτοί δεν οδηγούν σε ικανοποιητικά αποτελέσματα. Όλες οι μέθοδοι

χρησιμοποιούν επαναληπτικά σχήματα (κεφ. 4 διάγραμμα ροής ) προκειμένου να καταλήξουν στην βέλτιστη λύση. Σε κάθε επανάληψη προκύπτει ένα κυρίαρχο μοντέλο που εκφράζει την κίνηση για την δεδομένη φάση [1]. Οι περισσότερες μελέτες έχουν γίνει με την χρήση χαμηλής τάξης πολυωνυμικών παραμετρικών μοντέλων (affine μοντέλο) και όχι μεγαλύτερης τάξης (biquadratic, bilinear) τα οποία δεν αποδίδουν τόσο καλή απεικόνιση των επιπέδων (layers) και κυρίως είναι ευαίσθητα σε θόρυβο, παραμορφώσεις και άλλους παράγοντες.

## **1δ) Η δομή της διπλωματικής**

### **1δ1) Εκτίμηση της οπτικής ροής**

Στο πρώτο στάδιο της εργασίας αυτής έγινε υπολογισμός της οπτικής ροής (Κεφ.2 παρ.α1). Ξεκινώντας από μία συνάρτηση κόστους και αναπτυσσοντας την κατά TAYLOR καθώς και με μεθόδους βελτιστοποίησης καταλήγουμε στην εξίσωση οπτικής ροής (Κεφ.2 παρ.α2).

Χρησιμοποιώντας την μέθοδο Horn και Schunck κανονικοποιήσαμε την εξίσωση της οπτικής ροής η οποία από μόνη της δεν μπορούσε να δώσει ακριβή αποτελέσματα. Έτσι σύμφωνα με την εξίσωση (7) παρ.2β υπολογίσαμε τα διανύσματα κίνησης (motion vectors) το σύνολο των οποίων αποτελεί την οπτική ροή.

### **1δ2) Παραμετρική κίνηση**

Στο δεύτερο σκέλος της εργασίας υπολογίσαμε την παραμετρική κίνηση [3] [4] [5]. Το παραμετρικό μοντέλο που χρησιμοποιήσαμε είναι το affine μοντέλο [3] [4] [5] [6] [7]. Ουσιαστικά μοντελοποιούμε τα οριζόντια και τα κάθετα διανύσματα κίνησης σε άθροισμα τριών παραγόντων όπου ο κάθε παράγοντας περιείχε και από μία παράμετρο. Έτσι συνολικά για κάθε ζεύγος διανυσμάτων κίνησης έχουμε έξι παραμέτρους τις οποίες και τις υπολογίζουμε με δύο διαφορετικούς τρόπους. Διευκρινίζουμε ότι η εικόνα χωρίστηκε σε τετράγωνα (μεγέθους  $16 \times 16$ ) μη επικαλυπτόμενες περιοχές όπου για καθεμία περιοχή υπολογίστηκε και μία εξάδα παραμέτρων. Στον πρώτο τρόπο ο υπολογισμός έγινε με την συνδρομή της χωρικής και χρονικής παραγώγου (Κεφ.2 παρ. 2α2-Κεφ.3 παρ. 3β). Στον δεύτερο τρόπο έγινε με την συνδρομή των υπολογισθέντων διανυσμάτων της οπτικής ροής (κεφ.2). Από τα παραμετρικά μοντέλα μπορούμε εύκολα να εξάγουμε διανύσματα

παραμετρικής κίνησης για κάθε pixel. Έτσι μετά το τέλος αυτής της διαδικασίας έχουμε στη διάθεσή μας δύο ειδών διανύσματα κίνησης, τα παραμετρικά και τα διανύσματα οπτικής ροής κατά Horn και Schunck. Η εξαγωγή των layers βασίζεται στην σύγκριση συμβατότητας αυτών των δύο τύπων διανυσμάτων.

### **1δ3) Η εξαγωγή των επιπέδων (layers)**

Όπως προαναφέραμε σαν επίπεδο ορίζεται μια περιοχή της εικόνας που βρίσκεται κάτω από παρόμοια κίνηση. Δηλαδή περιγράφεται από κάποιο κοινό μοντέλο κίνησης. Έχοντας τα παραμετρικά και τα διανύσματα κατά Horn και Schunck χρησιμοποιήσαμε σαν μέτρο σύγκρισης τις μεταξύ τους διαφορές.

Είναι φυσιολογικό να μην μπορεί να απεικονιστεί ακριβώς ένας διάνυσμα όταν αναπτύσσεται σε άθροισμα τριών παραγόντων σε κάθε κατεύθυνση. Συνεπώς τα παραμετρικά και τα διανύσματα της οπτικής ροής λίγο ως πολύ διαφέρουν μεταξύ τους. Από την μεταξύ τους διαφορά καταλήξαμε στο συμπέρασμα πως όταν σε μια περιοχή υπάρχει απλή κίνηση τα δύο ειδών διανύσματα συγκλίνουν μεταξύ τους ενώ υπάρχει μεγάλη απόκλιση μεταξύ τους σε περιοχές όπου βρίσκονται κάτω από πολλαπλή και πολλές φορές επικαλυπτόμενη κίνηση. Προφανώς μεγάλη απόκλιση σημαίνει ότι το παραμετρικό μοντέλο αδυνατεί να περιγράψει, να απεικονήσει την κίνηση που υπάρχει σε αυτού του είδους τις περιοχές.

Χρησιμοποιώντας την παραπάνω πληροφορία και την απόσταση που έχουν οι τετραγωνικές περιοχές σύμφωνα με τις παραμέτρους μεταξύ τους ομαδοποιήσαμε την εικόνα (καρέ) του βίντεο σε επίπεδα (layers). Βέβαια η ομαδοποίηση δεν έγινε μόνο με τους προαναφερθέντες παράγοντες αλλά κυρίως με την χρήση του Multistage affine classification αλγορίθμου (κεφ. 4)

μέρος του οποίου είναι και ο k-means [5] .Από τα πειράματα που διεξάγαμε (κεφ. 5) διαπιστώσαμε πως οι μέθοδοι λειτουργούν αρκετά αποτελεσματικά.

## ΚΕΦΑΛΑΙΟ 2

### Υπολογισμός διανυσμάτων κίνησης με την μέθοδο κανονικοποίησης Horn-Schunck

#### 2α) ΕΙΣΑΓΩΓΗ

Προκειμένου στο ψηφιακό βίντεο να μελετήσουμε διάφορα είδη κίνησης και να διαχωρίσουμε τα αντικείμενα που περιέχονται σε αυτό σε κινητά, ακίνητα ακόμα και να προχωρήσουμε σε λεπτομερέστερες αναλύσεις, όπως ο διαχωρισμός των ταχυτήτων μεταξύ κινούμενων αντικειμένων χρησιμοποιούμε διάφορα είδη μοντέλων κίνησης όπου κοινός σκοπός τους είναι ο υπολογισμός των διανυσμάτων κίνησης και κατά συνέπεια της κίνησης που υπάρχει στο ψηφιακό βίντεο.

Για τον υπολογισμό της κίνησης σε ένα βίντεο έχουν χρησιμοποιηθεί διάφορα μοντέλα κίνησης-μέθοδοι [1] [2] [3] [4] [7] [9]. Οι κυριότερες από αυτές είναι οι διαφορικές (differential), οι μέθοδοι βασισμένες στη συχνότητα (frequency-based) και οι μέθοδοι ταιριάσματος (block matching). Έχουν γίνει αρκετές προσπάθειες συλλογής και αξιολόγησης των μεθόδων υπολογισμού της οπτικής ροής [8] [16] [18].

Οι διαφορικές μέθοδοι υπολογίζουν την οπτική ροή από χωρικές και χρονικές παραγώγους του video. Οι μέθοδοι αυτοί χωρίζονται σε δύο κύριες κατηγορίες ανάλογα με την τάξη των παραγώγων που χρησιμοποιούν.

Οι μέθοδοι πρώτης τάξης συνήθως χρησιμοποιούν πρωτοτάξια κατωδιαβατά (first-order low-pass) ή ζωνοδιαβατά (first-order band-pass) φίλτρα. Ορίζουμε ως τη συνάρτηση των διανυσμάτων της κίνησης στο χώρο  $\mathbf{x} = (x, y)$  και στο χρόνο  $t$ . Επίσης  $u(\mathbf{x}, t)$  και  $v(\mathbf{x}, t)$  είναι οι συνιστώσες της ταχύτητας στο χώρο και  $I(\mathbf{x}, t)$  το video. Η συνάρτηση βάση της οποίας γίνεται ο υπολογισμός των διανυσμάτων κίνησης είναι η:

$$\nabla I(\mathbf{x}, t) \mathbf{v} + I_t(\mathbf{x}, t) = 0$$

όπου  $\nabla I(\mathbf{x}, t)\mathbf{v}$  είναι το εσωτερικό γινόμενο των χωρικών παραγώγων με τα διανύσματα κίνησης. Συνεπώς η παραπάνω εξίσωση μπορεί να γραφτεί και ως εξής:

$$I_x(\mathbf{x}, t)u + I_y(\mathbf{x}, t)v + I_t(\mathbf{x}, t) = 0$$

όπου  $I_x(\mathbf{x}, t), I_y(\mathbf{x}, t)$  οι χωρικές παράγωγοι κατά τον  $x$  και  $y$  άξονα αντίστοιχα.

Η παραπάνω εξίσωση παραστάνει την εξίσωση της οπτικής ροής.

Οι μέθοδοι δεύτερης τάξης χρησιμοποιούν φίλτρα δεύτερης παραγώγου (Hessian) της εικόνας. Η συνάρτηση βάση της οποίας γίνεται ο υπολογισμός των διανυσμάτων κίνησης είναι η:

$$\begin{bmatrix} I_{xx}(\mathbf{x}, t) & I_{yx}(\mathbf{x}, t) \\ I_{xy}(\mathbf{x}, t) & I_{yy}(\mathbf{x}, t) \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} I_{tx}(\mathbf{x}, t) \\ I_{ty}(\mathbf{x}, t) \end{pmatrix} = 0$$

Οι μέθοδοι βασισμένες στη συχνότητα χρησιμοποιούν τη σχέση  $I(\mathbf{x}, t)\mathbf{v} + I_t(\mathbf{x}, t) = 0$  εκφρασμένη στο πεδίο της συχνότητας, βάση της οποίας γίνεται ο υπολογισμός των διανυσμάτων κίνησης:

$$\hat{I}(\mathbf{k}, \omega) = \hat{I}_0(\mathbf{k})\delta(\omega + \mathbf{v}^T \mathbf{k})$$

όπου  $\hat{I}(\mathbf{k})$  είναι ο μετασχηματισμός Fourier της εικόνας  $I(\mathbf{x}, 0)$ ,  $\delta(\mathbf{k})$  είναι η κρουστική συνάρτηση δέλτα του Dirac,  $\omega$  είναι η συχνότητα και  $\mathbf{k} = (k_x, k_y)$  είναι η χωρική συχνότητα.

Τέλος η πιο γνωστή κατηγορία μεθόδων είναι αυτή του ταιριάσματος (block matching). Στην κατηγορία αυτή ανήκουν οι μέθοδοι βασισμένες στην περιοχή (region – based) και οι μέθοδοι χαρακτηριστικών (feature – based). Οι μέθοδοι βασίζονται στην αναζήτηση των διανυσμάτων κίνησης μονάχα σε μια γειτονιά του σημείου και η εύρεσή τους βασίζεται στην σχέση:

$$I(\mathbf{x}, t) - I(\mathbf{x} + \mathbf{u}, t + 1) = 0$$

Τα βασικά πλεονεκτήματα των μεθόδων αυτών είναι η μεγάλη ανοχή τους στο θόρυβο, στην αλλοίωση της εικόνας και στην απλότητα του υπολογισμού της οπτικής ροής με αποτέλεσμα την χρήση μικρού αριθμού καρέ. Βασικό μειονέκτημα της μεθόδου είναι η ταχύτητα.



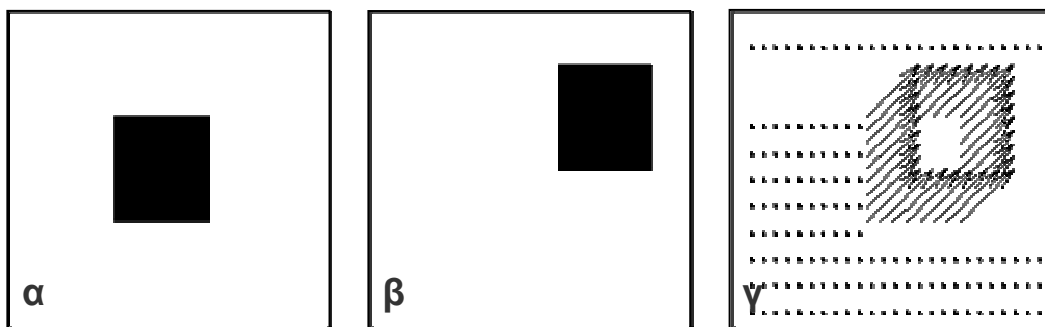
Το κυριότερο πρόβλημα αυτών των μεθόδων είναι το γεγονός ότι μελετάμε τρισδιάστατη κίνηση που λαμβάνει χώρα στον πραγματικό κόσμο με δισδιάστατες εικόνες. Συνεπώς από αυτή την μετατροπή και μόνο έχουμε χάσει πληροφορία (σχήμα 4).

Μοντελοποιούμε το βίντεο ως συνάρτηση  $I(\mathbf{x}, t)$  όπου  $\mathbf{x}=(x,y)$  είναι οι συντεταγμένες του pixel στην εικόνα και το  $t$  μας δείχνει την χρονική στιγμή της ροής του βίντεο δηλαδή τον αριθμό του καρέ στο οποίο βρισκόμαστε. Για παράδειγμα η έκφραση  $I(4,5,4)$  σημαίνει την ένταση του pixel στη θέση (4,5) της εικόνας την χρονική στιγμή 4 δηλαδή στο καρέ 4.

Παρακάτω παραθέτουμε τον ορισμό της οπτικής ροής που είναι και το μοντέλο κίνησης με το οποίο θα ασχοληθούμε και θα μας βοηθήσει στην διακπερέωση της εργασίας μας. Επίσης δίνουμε ένα παράδειγμα που εξηγεί το πρόβλημα που προκύπτει από την τρισδιάστατη απεικόνιση της κίνησης στον πραγματικό κόσμο στις δύο διαστάσεις της εικόνας. Το πρόβλημα αυτό το αντιμετωπίζουν όλα τα προαναφερθέντα μοντέλα κίνησης.

## 2α1)Ορισμός της οπτικής ροής

Ως οπτική ροή (optical flow) ορίζουμε το σύνολο των διανυσμάτων της κίνησης  $(u,v)$  μεταξύ δυο διαδοχικών εικόνων (καρέ) ενός βίντεο όπως φαίνεται και στο σχήμα 3.

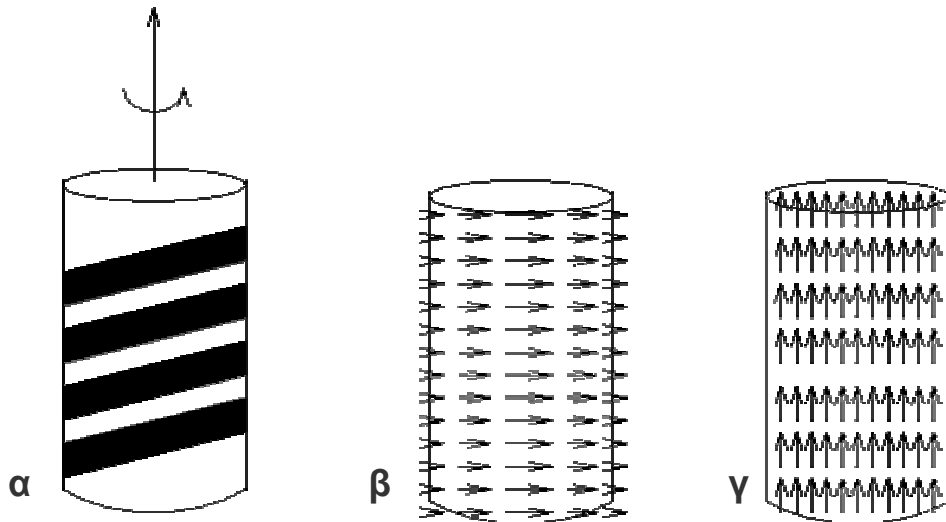


Σχήμα 3.

Ορισμός οπτικής ροής

Η οπτική ροή ( $\gamma$ ) μεταξύ των καρέ ( $\alpha$ ) και ( $\beta$ )

Τα διανύσματα κίνησης δεν αντιπροσωπεύουν πάντα την οπτική ροή καθώς για την εύρεσή τους υπάρχει μονάχα η προβολή του χώρου σε δύο διαστάσεις. Αυτό φαίνεται και στο σχήμα 4, όπου η προβολή του κυλίνδρου σε δυο διαστάσεις προκαλεί την διαφορετική εκτίμηση της κίνησης.



Σχήμα 4. Προβλήματα που παρουσιάζονται με την εξίσωση οπτικής ροής

(α)Περιστρεφόμενος κύλινδρος (β) η οπτική ροή (γ) τα διανύσματα κίνησης

## 2 α2) Η εξίσωση της οπτικής ροής και η συνάρτηση κόστους

Μία τυπική συνάρτηση κόστους που μας δείχνει το λάθος στον υπολογισμό των διανυσμάτων κίνησης είναι η

$$e(x, y, t) = ((I(x, y, t) - I(x - u, y - v, t1))^2 \quad (1)$$

όπου τα  $u$ ,  $v$  είναι τα διανύσματα κίνησης μεταξύ δύο διαδοχικών καρέ του βίντεο. Αν την εξίσωση (1) την αναπτύξουμε κατά TAYLOR [6] θα καταλήξουμε στην σχέση

$$I_x(\mathbf{x}, t)u + I_y(\mathbf{x}, t)v + I_t(\mathbf{x}, t) = 0 \quad (2)$$

Στην ίδια σχέση θα καταλήξουμε εάν εφαρμόσουμε και την μέθοδο της ελαχιστοποίησης [6] [7] [10]. Η εξίσωση (2) είναι η εξίσωση οπτικής ροής. Τα  $u$ ,  $v$  είναι τα διανύσματα κίνησης, η  $I_x(\mathbf{x}, t)$  είναι η χωρική παράγωγος ως προς τον  $x$ -άξονα. Είναι η διαφορά των εντάσεων μεταξύ δύο γειτονικών pixel κατά την κατεύθυνση του  $x$ -άξονα και ορίζεται από τον τύπο

$$I_x(\mathbf{x}, t) = I(x+1, y, t) - I(x, y, t) \quad (3)$$

Η  $I_y(\mathbf{x}, t)$  είναι η χωρική παράγωγος ως προς τον  $y$ -άξονα. Ουσιαστικά είναι η διαφορά των εντάσεων μεταξύ δύο γειτονικών pixel κατά την κατεύθυνση του  $y$ -άξονα και με την σειρά της ορίζεται από τον τύπο

$$I_y(\mathbf{x}, t) = I(x, y+1, t) - I(x, y, t) \quad (4)$$

Η  $I_t(\mathbf{x}, t)$  είναι η χρονική παράγωγος και ουσιαστικά είναι η διαφορά των εντάσεων του ίδιου pixel ανάμεσα σε δύο διαδοχικά καρέ. Ο τύπος που ορίζεται είναι

$$I_t(\mathbf{x}, t) = I(x, y, t+1) - I(x, y, t) \quad (5)$$

Η εξίσωση (2) δηλαδή η εξίσωση της οπτικής ροής παρουσιάζει και αυτή προβλήματα λόγω θορύβου παραμορφώσεων και άλλων αιτίων. Συνεπώς καταφεύγουμε σε μία μορφή κανονικοποίησης προκειμένου να ελαχιστοποιήσουμε τα λάθη αυτά και να εξάγουμε τα βέλτιστα διανύσματα κίνησης που να περιγράφουν όσο το δυνατό καλύτερα την κίνηση των αντικειμένων. Η μέθοδος που χρησιμοποιήσαμε περιγράφεται στην παρακάτω παράγραφο.

## 2β) Υπολογισμός οπτικής ροής κατά Horn-Schunck

Σύμφωνα με τον Horn και Schunck [7], η οπτική ροή ορίζεται σε σχέση με τη χωρική και τη χρονική αλλαγή της έντασης της φωτεινότητας ανάμεσα στα καρέ του video. Οπτική ροή είναι η κατανομή της φαινομενικής ταχύτητας της κίνησης του χρωματικού μοτίβου σε μια εικόνα.

Η πολυπλοκότητα του προβλήματος είναι αναπόφευκτη εφόσον δεν είναι ένα πρόβλημα πλήρως ορισμένο [8] [9]. Για αυτό το λόγο γίνονται κάποιες απαραίτητες υποθέσεις προκειμένου να γίνει μια εκτίμηση και να μπορέσουν να υπολογιστούν τα διανύσματα κίνησης. Έτσι η ταχύτητα ανάμεσα σε δύο καρέ του βίντεο, μπορεί να υπολογιστεί από το ολοκλήρωμα:

$$F_t = \int_D (\nabla I \mathbf{g} + I_t)^2 + \lambda^2 (\|\nabla u\|_2^2 + \|\nabla v\|_2^2) d\mathbf{x} \quad (6)$$

που ορίζεται στα όρια της εικόνας του καρέ.

Ιδανικά το ολοκλήρωμα αυτό είναι μηδέν. Με πραγματικά δεδομένα όμως, ο πρώτος όρος  $(\nabla I \mathbf{g} + I_t)^2$  είναι διάφορος του μηδενός λόγω θορύβου, παραμορφώσεων και άλλων φαινομένων που έχουν να κάνουν με τη φύση

του video και ο δεύτερος όρος  $\|\nabla u\|_2^2 + \|\nabla v\|_2^2$  είναι επίσης διάφορος του μηδενός, για εκτίμηση που δε συμφωνεί με τα γειτονικά διανύσματα κίνησης αφού θεωρούμε ότι ιδεατά τα διανύσματα κίνησης είναι ομαλά (smoothness constraint). Έτσι το ολοκλήρωμα αυτό είναι μηδέν για επιλογή διανυσμάτων κίνησης τέτοιων ώστε να εκφράζουν τόσο την κίνηση με κριτήριο την ένταση (πρώτος όρος) όσο και την κίνηση με κριτήριο την ομοιομορφία των διανυσμάτων στα πλαίσια της περιοχής γύρω του (δεύτερος όρος). Ο παράγοντας  $\lambda$  υπάρχει για να μπορεί να ρυθμίζει την ισχύ των δυο δεσμεύσεων που αποτελούν τη βάση του υπολογισμού των διανυσμάτων κίνησης.

Αναπτύσσοντας το ολοκλήρωμα λοιπόν έχουμε:

$$F_t = \int_D \lambda (I_x u + I_y v + I_t)^2 + (u_x^2 + u_y^2 + v_x^2 + v_y^2) d\mathbf{x}$$

Οπότε σε διακριτή μορφή το ολοκλήρωμα θα πάρει τη μορφή:

$$F_t = \sum_{i,j} \lambda (I_x u + I_y v + I_t)^2 + \frac{1}{4} ((u_{i+1,j} - u_{ij})^2 + (u_{i,j+1} - u_{ij})^2 + (v_{i+1,j} - v_{ij})^2 + (v_{i,j+1} - v_{ij})^2)$$

Έτσι προκειμένου να ελαχιστοποιήσουμε το ολοκλήρωμα αρκεί:

$$\nabla F_t = 0 \text{ ή αλλιώς } \begin{cases} \frac{\partial F_t}{\partial u} = 0 \\ \frac{\partial F_t}{\partial v} = 0 \end{cases}$$

Η μερική παράγωγος  $\frac{\partial F_t}{\partial u}$  υπολογίζεται παρακάτω:

$$\begin{aligned} \frac{\partial F_t}{\partial u} &= 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_x - \frac{1}{2}[(u_{i+1,j} - u_{ij}) + (u_{i,j+1} - u_{ij})] + \frac{1}{2}[(u_{ij} - u_{i-1,j}) + (u_{ij} - u_{i,j-1})] \Leftrightarrow \\ \frac{\partial F_t}{\partial u} &= 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_x + 2u_{ij} - 2\left(\frac{1}{4}(u_{i+1,j} + u_{i,j+1} + u_{i-1,j} + u_{i,j-1})\right) \Leftrightarrow \\ \frac{\partial F_t}{\partial u} &= 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_x + 2u_{ij} - 2\bar{u}_{ij} \end{aligned}$$

Ομοίως υπολογίζουμε και την μερική παράγωγο  $\frac{\partial F_t}{\partial v}$  οπότε έχουμε:

$$\frac{\partial F_t}{\partial v} = 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_y + 2v_{ij} - 2\bar{v}_{ij}$$

Οπότε το σύστημα που έχουμε να λύσουμε είναι:

$$\begin{cases} 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_x + 2u_{ij} - 2\bar{u}_{ij} = 0 \\ 2\lambda(I_x u_{ij} + I_y v_{ij} + I_t)I_y + 2v_{ij} - 2\bar{v}_{ij} = 0 \end{cases} \Leftrightarrow \begin{cases} (1 + \lambda I_x^2)u_{ij} + \lambda I_x I_y v_{ij} = \bar{u} - \lambda I_x I_t \\ \lambda I_x I_y u_{ij} + (1 + \lambda I_y^2)v_{ij} = \bar{v} - \lambda I_y I_t \end{cases}$$

Σε μορφή πινάκων μπορεί να γραφτεί:

$$\begin{bmatrix} 1 + \lambda I_x^2 & \lambda I_x I_y \\ \lambda I_x I_y & 1 + \lambda I_y^2 \end{bmatrix} \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} = \begin{bmatrix} \bar{u} - \lambda I_x I_t \\ \bar{v} - \lambda I_y I_t \end{bmatrix}$$

Για την επίλυση του παραπάνω συστήματος εφαρμόσαμε την επαναληπτική μέθοδο Jacobi (Jacobi iterations), όπου η λύση του συστήματος δίνεται από τον τύπο:

$$x^{(m+1)} = D^{-1}(L+U)x^{(m)} + D^{-1}b \mid m=0,1,2\dots$$

Ο πίνακας των συντελεστών των  $u, v$  ονομάζεται  $A$  και ο πίνακας των σταθερών όρων  $b$ . Από αυτούς κατασκευάζουμε τους πίνακες  $D, L$  και  $U$  ώστε να ισχύει  $A=D-L-U$  όπου  $D$  να είναι διαγώνιος,  $L$  αυστηρά κάτω τριγωνικός και  $U$  αυστηρά πάνω τριγωνικός πίνακας. Έτσι οι πίνακες του σχήματος είναι οι παρακάτω:

$$D = \begin{bmatrix} 1 + \lambda I_x^2 & 0 \\ 0 & 1 + \lambda I_y^2 \end{bmatrix} \quad U = \begin{bmatrix} 0 & -\lambda I_x I_y \\ 0 & 0 \end{bmatrix} \quad L = \begin{bmatrix} 0 & 0 \\ -\lambda I_x I_y & 0 \end{bmatrix}$$

$$b = \begin{bmatrix} \bar{u} - \lambda I_x I_t \\ \bar{v} - \lambda I_y I_t \end{bmatrix} \text{ και } D^{-1} = \begin{bmatrix} \frac{1}{1 + \lambda I_x^2} & 0 \\ 0 & \frac{1}{1 + \lambda I_y^2} \end{bmatrix}$$

Έτσι έχουμε:

$$x^{(m+1)} = D^{-1}(L+U)x^{(m)} + D^{-1}b$$

$$\begin{bmatrix} u^{(m+1)} \\ v^{(m+1)} \end{bmatrix} = \begin{bmatrix} 1/(1 + \lambda I_x^2) & 0 \\ 0 & 1/(1 + \lambda I_y^2) \end{bmatrix} \begin{bmatrix} 0 & -\lambda I_x I_y \\ -\lambda I_x I_y & 0 \end{bmatrix} \begin{bmatrix} u^{(m)} \\ v^{(m)} \end{bmatrix} + \begin{bmatrix} 1/(1 + \lambda I_x^2) & 0 \\ 0 & 1/(1 + \lambda I_y^2) \end{bmatrix} \begin{bmatrix} \bar{u} - \lambda I_x I_t \\ \bar{v} - \lambda I_y I_t \end{bmatrix}$$

$$\begin{bmatrix} u^{(m+1)} \\ v^{(m+1)} \end{bmatrix} = \begin{bmatrix} 0 & \frac{-\lambda I_x I_y}{1 + \lambda I_x^2} \\ \frac{-\lambda I_x I_y}{1 + \lambda I_y^2} & 0 \end{bmatrix} \begin{bmatrix} u^{(m)} \\ v^{(m)} \end{bmatrix} + \begin{bmatrix} \frac{\bar{u} - \lambda I_x I_t}{1 + \lambda I_x^2} \\ \frac{\bar{v} - \lambda I_y I_t}{1 + \lambda I_y^2} \end{bmatrix}$$

$$\begin{aligned}
\begin{bmatrix} u^{(m+1)} \\ v^{(m+1)} \end{bmatrix} &= \begin{bmatrix} \frac{-\lambda I_x I_y}{1 + \lambda I_x^2} v^{(m)} \\ \frac{-\lambda I_x I_y}{1 + \lambda I_y^2} u^{(m)} \end{bmatrix} + \begin{bmatrix} \frac{\bar{u} - \lambda I_x I_t}{1 + \lambda I_x^2} \\ \frac{\bar{v} - \lambda I_y I_t}{1 + \lambda I_y^2} \end{bmatrix} \\
\begin{bmatrix} u^{(m+1)} \\ v^{(m+1)} \end{bmatrix} &= \begin{bmatrix} \frac{\bar{u} - \lambda I_x I_t - \lambda I_x I_y v^{(m)}}{1 + \lambda I_x^2} \\ \frac{\bar{v} - \lambda I_y I_t - \lambda I_x I_y u^{(m)}}{1 + \lambda I_y^2} \end{bmatrix} \tag{7}
\end{aligned}$$

Για  $\chi^{(0)}$  επιλέγονται αυθαίρετες τιμές, συνήθως μηδέν. Βέβαια όταν πρόκειται για παραπάνω των δυο καρέ, στην ακολουθία του video που ζητάμε να βρούμε τα διανύσματα κίνησης, τότε μια καλή αρχικοποίηση είναι η λύση του προηγούμενου ζευγαριού. Αυτό είναι προφανές δεδομένου ότι περιμένουμε μικρές αλλαγές στην κίνηση μεταξύ δυο διαδοχικών καρέ.

Για τον υπολογισμό της εξίσωσης (7) έγινε η ακόλουθη παραδοχή:

- Ότι υπάρχει η παράγωγος της εικόνας σε κάθε σημείο ως προς  $x$  και ως προς  $y$ . Δηλαδή:

$$\frac{\partial x}{\partial t} \neq 0 \text{ και } \frac{\partial y}{\partial t} \neq 0$$

Αυτό σημαίνει ότι δεν υπολογίζεται οπτική ροή για σημεία για ομοιόμορφες περιοχές με μηδενικές παραγώγους στο χώρο. Για τα σημεία αυτά προσεγγίζουμε την οπτική ροή, υπολογίζοντάς την έμμεσα, ως το μέσο όρο της οπτικής ροής στα γειτονικά σημεία.

Συνεπώς έχουμε τα διανύσματα κίνησης τα οποία θα χρησιμεύσουν για την εύρεση του παραμετρικού μοντέλου.

## ΚΕΦΑΛΑΙΟ 3

### Παραμετρικά μοντέλα κίνησης

#### 3α) Σκοπός και λειτουργία των παραμετρικών μοντέλων

Τα παραμετρικά μοντέλα έχουν ως σκοπό να υπολογίσουν κίνηση με διαφορετικό τρόπο βέβαια από τα απλά μοντέλα κίνησης που περιγράψαμε στο δεύτερο κεφάλαιο. Η ουσιαστική διαφορά είναι η παραμετροποίηση των διανυσμάτων κίνησης (motion vectors). Αυτή η παραμετροποίηση γίνεται προκειμένου να συλλέξουμε και να ομαδοποιήσουμε pixel με όμοια κίνηση. Επιπλέον σαν αποτέλεσμα μπορούμε επίσης να ελέξουμε την συμβατότητα των διανυσμάτων κίνησης με κάποιο παραμετρικό μοντέλο. Εάν η περιοχή διέπεται από ίδια κίνηση τότε ο υπολογισμός της κίνησης είναι πιο ακριβής. Τα διανύσματα κίνησης εκφράζονται σαν άθροισμα παραγόντων που περιέχουν κάποιες παραμέτρους καθεμία από αυτές περιγράφει και ένα διαφορετικό είδος κίνησης (Κεφ.3 -παρ.3β). Έτσι λοιπόν θεωρείται ότι αυτά τα μοντέλα υπολογισμού της παραμετρικής οπτικής ροής μπορούν να αναπαραστήσουν την κίνηση σε μία περιοχή της εικόνας με ένα πολυώνυμο χαμηλής τάξης (το πολύ δευτέρου βαθμού). Σε αυτή την διπλωματική η περιοχή είναι τετραγωνική μεγέθους (16\*16) και το πολυώνυμο είναι πρώτου βαθμού (Κεφ.3-παρ.3δ). Φυσικά από κάθε παραμετρικό μοντέλο μπορούμε εύκολα να εξάγουμε διανύσματα κίνησης για κάθε σημείο της περιοχής που εξετάζουμε, τα οποία ονομάζουμε παραμετρικά διανύσματα κίνησης. Μία περιοχή με ομοιόμορφη κίνηση υποδηλώνει ένα συμπαγές αντικείμενο ενδιαφέροντος το οποίο μπορεί να οριστεί σαν ένα layer ενδιαφέροντος κατά το MPEG 4.

Ο σκοπός των παραμετρικών μοντέλων είναι η εύρεση των περιοχών με ομοιόμορφη κίνηση (layers). Βεβαίως δεν γνωρίζουμε εκ των προτέρων τα όρια των περιοχών οπότε ξεκινάμε από τα τετραγωνικά blocks. Για κάθε block υπολογίζουμε ένα παραμετρικό μοντέλο από τα δεδομένα μας που είναι το video sequence με την χρήση της εξίσωσης οπτικής ροής (optical flow equation) (Κεφ.3 παρ.3β). Όμως κάποια block μπορεί να ανήκουν στο ίδιο



layer οπότε αφού το διαπιστώσουμε μέσω σύγκρισης τα ενώνουμε (Κεφ.4 παρ.4δ2). Σε κάποιες άλλες περιπτώσεις ένα block μπορεί να βρίσκεται στα όρια δύο ή περισσότερων layers. Τότε χωρίζουμε (splitting) το block με βάση το motion vector κάθε pixel. Άρα πολλές φορές χρειάζεται να ορίσουμε παραμετρικό μοντέλο όχι από τα αρχικά δεδομένα (frames) αλλά από διανύσματα κίνησης. Οπότε σε αυτή την περίπτωση εξάγουμε παραμετρικό μοντέλο από τα διανύσματα κίνησης (motion vectors) της οπτικής ροής (Κεφ.3 παρ.3δ).

### 3β) Η κίνηση της κάμερας και οι μαθηματικές σχέσεις που την διέπουν

Προτού προχωρήσουμε στην ανάλυση των δύο μεθόδων υπολογισμού των παραμετρικών διανυσμάτων κίνησης αναφέρουμε κάποιες μαθηματικές σχέσεις με τις οποίες επιτυγχάνεται η απεικόνιση των τρισδιάστατων σκηνών που λαμβάνει η κάμερα σε δισδιάστατα διανύσματα κίνησης (motion vectors). Προτού δώσουμε αυτούς τους τύπους αναφέρουμε κάποιες μαθηματικές σχέσεις που διέπουν την κίνηση της κάμερας.

Σε γενικές γραμμές η κάμερα μπορεί να κάνει 3 είδη translational και 3 είδη περιστροφικών (rotational) κινήσεων κάθε μία σε κάθε άξονα (X, Y, Z) του τρισδιάστατου κόσμου. Το translation κατά μήκος των X, Y αξόνων ονομάζεται “track and boom”, η περιστροφή (rotation) γύρω από τους X, Y χωρίς σημαντική αλλαγή στο βάθος Z ονομάζεται “pan” και “tilt” αντιστοίχως, “roll” ονομάζεται η περιστροφή γύρω από τον Z άξονα και “zoom” το translation στον Z άξονα ή η αλλαγή του εστιακού βάθους. Συνεπώς η κάμερα έχει 6 βαθμούς ελευθερίας στις κινήσεις τις. Οι εξισώσεις που περιγράφουν την κίνησή της είναι οι εξής.

$$\bar{V} = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \end{bmatrix} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} + \begin{bmatrix} R_y Z - R_z Y \\ R_z X - R_x Z \\ R_x Y - R_y X \end{bmatrix}$$

όπου οι  $x, y, z$  συντεταγμένες περιγράφουν το επίπεδο της κάμερας και οι  $X, Y, Z$  τις τρεις διαστάσεις του χώρου.

όπου  $T_x, T_y, T_z$  είναι translation κίνηση και  $R_x, R_y, R_z$  εκφράζουν περιστροφική (rotational) κίνηση. Η πλήρης απεικόνιση της τρισδιάστατης κίνησης της κάμερας στα δισδιάστατα καρέ του βίντεο είναι ένα δύσκολο πρόβλημα καθώς δεν πρέπει να χάσουμε πληροφορία από μία πλούσια απεικόνιση σε μία πιο φτωχή. Αρκετές μέθοδοι έχουν εξελιχθεί για την δισδιάστατη εκτίμηση της κίνησης κάνοντας ειδικές υποθέσεις και εφαρμόζοντας περιορισμούς στην συμπεριφορά της κίνησης.

Έτσι από τις παρακάτω εξισώσεις εξάγεται ο σχεδιάσμος από την τρισδιάστατη στην δισδιάστατη κίνηση.

$$\frac{dx}{dt} = \frac{1}{Z} \frac{dX}{dt} - \frac{X}{Z^2} \frac{dZ}{dt}$$

$$\frac{dy}{dt} = \frac{1}{Z} \frac{dY}{dt} - \frac{Y}{Z^2} \frac{dZ}{dt}$$

Υποθέτοντας ότι οι επιφάνειες που η κάμερα λαμβάνει μπορούν να προσεγγισθούν από ένα πλάνο αυτής της εξίσωσης.

$$n_x X + n_y Y + n_z Z = 1$$

τα δισδιάστατα διανύσματα κίνησης (motion vectors) δίδονται από τις παρακάτω εξισώσεις.

$$\frac{dx}{dt} = (n_z T_x + R_y) + (n_x T_x - n_z T_z)x + (n_y T_x - R_z)y + (-n_x T_x + R_y)x^2 - (n_y T_z + R_x)xy$$

$$\frac{dy}{dt} = (n_z T_y - R_x) + (n_x T_y + R_z)x + (n_y T_y - n_z R_z)y - (n_y T_y + R_x)y^2 + (-n_x T_z + R_y)xy$$

Αυτή η εξίσωση είναι ένα μιγαδικό μη γραμμικό μοντέλο με οχτώ βαθμούς ελευθερίας. Αντικαθιστώντας τους συντελεστές των παραπάνω εξισώσεων με πραγματικές παραμέτρους προκύπτουν τα παραμετρικά μοντέλα κίνησης. Οι υποθέσεις που έγιναν για την συμπεριφορά της κίνησης έχουν να κάνουν με την εξής παραδοχή. Οι παράμετροι οι οποίες εκφράζουν ένα διαφορετικό είδος κίνησης από αυτό του είναι ικανό να περιγράψει το μοντέλο θέτονται ίσες με το μηδέν. Συνεπώς ο αριθμός των παραμέτρων και

κατά συνέπεια η πολυπλοκότητα του μοντέλου εξαρτώνται βασικά από την κίνηση την οποία καλείται το μοντέλο να περιγράψει. Στο παρακάτω σχήμα δίνονται οι εξισώσεις που περιγράφουν την λειτουργία των παραμετρικών μοντέλων [5].

Model	Transformation	Parameters
Translation	$x' = x + b$	$b \in R^2$
Affine	$x' = Ax + b$	$A \in R^2, b, c \in R^2$
Bilinear	$x' = q_{x'x}xy + q_{x'y}x + q_{x'y}y + q_{x'}$ $y' = q_{y'xy}xy + q_{y'x}x + q_{y'y}y + q_{y'}$	$q_* \in R$
Projective	$x' = \frac{Ax + b}{c^T x + 1}$	$A \in R^{2 \times 2}, b, c \in R^2$
Pseudo-Perspective	$x' = q_{x'x}x + q_{x'y}y + q_{x'} + q_a x^2 + q_b xy$ $y' = q_{x'y}x + q_{y'y}y + q_{y'} + q_a xy + q_b y^2$	$q_* \in R$
Biquadratic	$x' = q_{x'x}x^2 + q_{x'xy}xy + q_{x'yy}y^2 + q_{x'x}x + q_{x'y}y + q_{x'}$ $y' = q_{y'x}x^2 + q_{y'xy}xy + q_{y'yy}y^2 + q_{y'x}x + q_{y'y}y + q_{y'}$	$q_* \in R$

Σχήμα 5 .

Οι εξισώσεις των παραμετρικών μοντέλων

### 3γ)Είδη παραμετρικών μοντέλων

#### 3γ1)Translational μοντέλο

Το translational μοντέλο είναι το πιο απλό μοντέλο και είναι ικανό να περιγράψει μόνο translational κίνηση η οποία είναι και η πιο κοινή κίνηση της κάμερας. Η κίνηση εκφράζεται σαν μία μετακίνηση των συντεταγμένων της εικόνας από ένα σταθερό παράγοντα. Παρόλο που αυτό το μοντέλο είναι πολύ απλό και δεν έχει περιορισμούς στην χρήση του, είναι ακατάλληλο για περισσότερο πολύπλοκες κινήσεις όπως το rotation και το zooming.

### **3γ2)Affine μοντέλο**

Το affine μοντέλο είναι και το πιο συχνά χρησιμοποιούμενο για την δισδιάστατη εκτίμηση της κίνησης (motion estimation). Είναι και το παραμετρικό μοντέλο που χρησιμοποιούμε σε αυτή την διπλωματική. Χαρακτηρίζεται από έξι παραμέτρους καθεμία από τις οποίες περιγράφει και ένα διαφορετικό είδος κίνησης. Οι κινήσεις που μπορεί το affine μοντέλο να περιγράψει είναι οι translational κινήσεις, το rotation της κάμερας γύρω από τον οπτικό άξονα και το zooming.

### **3γ3)Projective μοντέλο**

Είναι το πιο κατάλληλο και επαρκές μοντέλο για την περιγραφή όλων των κινήσεων της κάμερας. Μπορεί να περιγράψει όλες τις κινήσεις που περιγράφει το affine μοντέλο συν το rotation γύρω από τους X,Y άξονες. Το μόνο περιορισμό που έχει είναι ότι όλα τα αντικείμενα της σκηνής πρέπει να είναι στο ίδιο βάθος.

### **3γ4)Biquadratic μοντέλο**

Είναι το πιο περίπλοκο μοντέλο και είναι ικανό να περιγράψει όλες τις πιθανές κινήσεις της κάμερας. Το πρόβλημα αυτού του μοντέλου είναι οι 12 παράμετροι οι οποίες το κάνουν επιρρεπές σε λάθη λόγω πολλών υπολογισμών που απαιτεί. Η χρήση του είναι εμπιστευτική όταν υπάρχουν πολλαπλές και πολύπλοκες κινήσεις. Είναι το μόνο ικανό για να περιγράψει όλα τα είδη της κίνησης.

### 3γ5)Bilinear μοντέλο

Το Bilinear είναι οχτώ παραμέτρων μοντέλο και είναι η απλοποίηση του biquadratic η οποία γίνεται με τον μηδενισμό των παραγόντων  $x^2, y^2$  του biquadratic(σχήμα 5).Το Bilinear δεν είναι ικανό να περιγράψει κινήσεις rotation γύρω από τους άξονες.

### 3δ)Εκτίμηση παραμετρικού μοντέλου από O.F.E.

Η επέκταση της οπτικής ροής που υπολογίσαμε με την κατά Horn και Schunck κανονικοποίηση,στην affine κίνηση δίνεται από τους παρακάτω τύπους.

$$u(x, y) = a_0 + a_1x + a_2y \quad (8)$$

$$v(x, y) = a_3 + a_4x + a_5y \quad (9)$$

Όπου  $u(x, y), v(x, y)$  είναι τα οριζόντια και τα κάθετα διανύσματα κίνησης(motion vectors).Οι παράμετροι  $a_0, a_1, a_2, a_3, a_4, a_5$  είναι οι έξη παράμετροι του affine μοντέλου καθένας από τους οποίους περιγράφει και ένα διαφορετικό είδος κίνησης.

Από την μορφή που έχουν οι παραπάνω εξισώσεις παρατηρούμε ότι οι παράμετροι  $a_0, a_1, a_2$  περιγράφουν την οριζόντια κίνηση στον x-άξονα και οι παράμετροι  $a_3, a_4, a_5$  περιγράφουν την κάθετη κίνηση στον y-άξονα.

Άρχικα θεωρούμε την εξίσωση της οπτικής ροής που περιγράφεται στο 2 κεφάλαιο την  $I_x(\mathbf{x},t)u + I_y(\mathbf{x},t)v + I_t(\mathbf{x},t) = 0$  και την οποία εφόσον την υψώσουμε στο τετράγωνο και αντικαταστήσουμε τις τιμές των διανυσμάτων κίνησης  $u, v$  με αυτές των εξισώσεων (8) και (9) καταλήγουμε στην εξής μορφή.

$$E = \sum_{\mathbf{x} \in B} (I_x(\mathbf{x},t)(a_0 + a_1x + a_2y) + I_y(\mathbf{x},t)(a_3 + a_4x + a_5y) + I_t(\mathbf{x},t))^2$$

όπου B συμβολίζει το τετραγωνικό block (16\*16).

Παραγοντοποιώντας την E σε σχέση με τις παραμέτρους  $a_0, a_1, a_2, a_3, a_4, a_5$  κάθε φορά και θέτοντας τα αποτελέσματα ίσα με το μηδέν καταλήγουμε σε έξι εξισώσεις με έξι αγνώστους.

Έτσι έχουμε ένα σύστημα της μορφής  $C = A^{-1}B$  όπου C είναι ένας πίνακας της διαστάσεων 6\*1 και έχει ως όρους τους την εξάδα των παραμέτρων του Affine μοντέλου που επιδιώκουμε να υπολογίσουμε για κάθε block. Ο A είναι ένας πίνακας τετραγωνικός διαστάσεων 6\*6 και έχει ως όρους γινόμενα με παράγοντες τις χωρικές παραγώγους  $I_x(\mathbf{x},t), I_y(\mathbf{x},t)$ , καθώς και τις θέσεις (x,y) των pixel της εικόνας. Ο B είναι ένας πίνακας διαστάσεων 6\*1 και έχει ως όρους γινόμενα με παράγοντες τις χωρικές παραγώγους  $I_x(\mathbf{x},t), I_y(\mathbf{x},t)$ , την χρονική παράγωγο  $I_t(\mathbf{x},t)$  και τις θέσεις x,y της εικόνας.

Αναλυτικά οι πίνακες έχουν την εξής μορφή.

$$C = \begin{bmatrix} a0 \\ a1 \\ a2 \\ a3 \\ a4 \\ a5 \end{bmatrix}$$

$$A = \begin{bmatrix} \sum_{x \in B} I^2 x & \sum_{x \in B} x I^2 x & \sum_{x \in B} y I^2 x & \sum_{x \in B} I x I y & \sum_{x \in B} x I x I y & \sum_{x \in B} y I x I y \\ \sum_{x \in B} x I^2 x & \sum_{x \in B} x^2 I^2 x & \sum_{x \in B} x y I^2 x & \sum_{x \in B} x I x I y & \sum_{x \in B} x^2 I x I y & \sum_{x \in B} x y I x I y \\ \sum_{x \in B} y I^2 x & \sum_{x \in B} x y I^2 x & \sum_{x \in B} y^2 I^2 x & \sum_{x \in B} y I x I y & \sum_{x \in B} x y I x I y & \sum_{x \in B} y^2 I x I y \\ \sum_{x \in B} I x I y & \sum_{x \in B} x I x I y & \sum_{x \in B} y I x I y & \sum_{x \in B} I^2 y & \sum_{x \in B} x I^2 y & \sum_{x \in B} y I^2 y \\ \sum_{x \in B} x I x I y & \sum_{x \in B} x^2 I x I y & \sum_{x \in B} x y I x I y & \sum_{x \in B} x I^2 y & \sum_{x \in B} y^2 I^2 x & \sum_{x \in B} x y I^2 y \\ \sum_{x \in B} y I x I y & \sum_{x \in B} x y I x I y & \sum_{x \in B} y^2 I x I y & \sum_{x \in B} y I^2 y & \sum_{x \in B} x y I^2 y & \sum_{x \in B} y^2 I^2 y \end{bmatrix}$$

$$B = \begin{bmatrix} -\sum_{x \in B} I x I t \\ -\sum_{x \in B} x I x I t \\ -\sum_{x \in B} y I y I t \\ -\sum_{x \in B} I y I t \\ -\sum_{x \in B} x I y I t \\ -\sum_{x \in B} y I x I t \end{bmatrix}$$

Έχοντας πλέον υπολογίσει τις τιμές των παραμέτρων για κάθε ένα τετραγωνικό block από τις σχέσεις (8) και (9) υπολογίζουμε τα παραμετρικά διανύσματα κίνησης του affine μοντέλου. Στο κεφάλαιο 4 θα προχωρήσουμε

στην σύγκριση των δύο ειδών διανυσμάτων που αποτελεί ένα σημαντικό βήμα για την εξαγωγή των επιπέδων(layers).

### 3ε) Φιλτράρισμα του υπολογισμού των παραμετρικών διανυσμάτων με την χρήση του βάρους

Όπως έχουμε αναφέρει τα παραμετρικά μοντέλα αντιμετωπίζουν προβλήματα σε περιοχές όπου υπάρχει πολλαπλή κίνηση στην δική μας περίπτωση στις τετραγωνικές περιοχές όπου λαμβάνουν χώρα τρεις διαφορετικές κινήσεις του δέντρου ,του κήπου και του σπιτιού.Σε αυτές τις περιπτώσεις οι τιμές των affine παραμέτρων και κατά συνέπεια τα παραμετρικά διανύσματα κίνησης αποκλίνουν σε μεγάλο βαθμό από τις προσδοκώμενες τιμές με αποτέλεσμα η μέθοδος να οδηγείται σε ανακριβή αποτελέσματα. Σε αυτή την ενότητα περιγράφουμε μία μέθοδο λύνοντας ένα παρόμοιο αλλά όχι δευτεροβάθμιο πρόβλημα.Με την χρήση αυτής της μεθόδου αποφεύγουμε την επίδραση των αποκλίνουσων τιμών (outliers)(Κεφ.3-παρ.ε) οι οποίες παρουσιάζονται κυρίως σε περιοχές πολλαπλής κίνησης.

Έτσι λοιπόν γίνεται έλεγχος σε κάθε σημείο της εικόνας βάσει του τύπου  $|I_x(\mathbf{x},t)u + I_y(\mathbf{x},t)v + I_t(\mathbf{x},t)| > \frac{\sigma}{\sqrt{3}}$  (10)

όπου η τιμή του  $\sigma \in R$  καθορίζεται από τις απαιτήσεις του προβλήματος.Όταν λοιπόν υπολογιστούν οι affine παράμετροι για κάθε μία περιοχή,αντικαθιστώντας τις εξισώσεις (8) και (9) στην (10) παραγματοποιούμε τον έλεγχο για αποκλίνουσες(outlying) μετρήσεις.Έφσον η απόλυτη τιμή της παράστασης  $x = I_x(\mathbf{x},t)u + I_y(\mathbf{x},t)v + I_t(\mathbf{x},t)$  ξεπερνάει το όριο τότε γίνεται χρήση του συντελεστή βάρους(weight) ο οποίος εξομαλύνει την αποκλίνουσα συμπεριφορά του αλγορίθμου σε αυτές τις περιπτώσεις.



Ουσιαστικά πρόκειται για χρήση της Geman-McClure συνάρτησης [8]

$$\text{η οποία ορίζεται από τον τύπο } p(x, \sigma) = \frac{x^2}{x^2 + \sigma^2}$$

Πιο αναλυτικά το βάρος (weight) μας δίνεται από την σχέση

$$w(x) = \frac{y(x)}{x} \quad (11)$$

$$\text{όπου } y(x) = \frac{dp(x, \sigma)}{dx} \quad (12)$$

$$\text{και } x = I_x(\mathbf{x}, t)u + I_y(\mathbf{x}, t)v + I_t(\mathbf{x}, t)$$

Συνεπώς από την εξίσωση (11) και (12) έχουμε  $w = \frac{2\sigma^2}{x^2 + \sigma^2}$  η οποία μας δίνει την τιμή με την οποία φιλτράρουμε τις αποκλίνουσες(outlying) τιμές.

Με αυτή την παραδοχή η εξίσωση  $C = A^{-1}B$  μετετρέπεται στην εξής μορφή και στους πίνακες  $A$  και  $B$  πολλαπλασιάζεται σε καθένα από τα στοιχεία τους ο συντελεστής βάρους (weight)  $w$ . Οπότε έχουμε τις εξής μορφές.

$$C = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}$$

$$A =$$

$$\begin{bmatrix} \sum_{x \in B} wI^2x & \sum_{x \in B} wxI^2x & \sum_{x \in B} wyI^2x & \sum_{x \in B} wIxIy & \sum_{x \in B} wxIxIy & \sum_{x \in B} wyIxIy \\ \sum_{x \in B} wxI^2x & \sum_{x \in B} wx^2I^2x & \sum_{x \in B} wxyI^2x & \sum_{x \in B} wxIxIy & \sum_{x \in B} wx^2IxIy & \sum_{x \in B} wxyIxIy \\ \sum_{x \in B} wyI^2x & \sum_{x \in B} wxyI^2x & \sum_{x \in B} wy^2I^2x & \sum_{x \in B} wyIxIy & \sum_{x \in B} wxyIxIy & \sum_{x \in B} wy^2IxIy \\ \sum_{x \in B} wIxIy & \sum_{x \in B} wxIxIy & \sum_{x \in B} wyIxIy & \sum_{x \in B} wI^2y & \sum_{x \in B} wxI^2y & \sum_{x \in B} wyI^2y \\ \sum_{x \in B} wxIxIy & \sum_{x \in B} wx^2IxIy & \sum_{x \in B} wxyIxIy & \sum_{x \in B} wxI^2y & \sum_{x \in B} wy^2I^2x & \sum_{x \in B} wxyI^2y \\ \sum_{x \in B} wyIxIy & \sum_{x \in B} wxyIxIy & \sum_{x \in B} wy^2IxIy & \sum_{x \in B} wyI^2y & \sum_{x \in B} wxyI^2y & \sum_{x \in B} wy^2I^2y \end{bmatrix}$$

$$\text{και } B = \begin{bmatrix} -\sum_{x \in B} wIxIt \\ -\sum_{x \in B} wxIxIt \\ -\sum_{x \in B} wyIyIt \\ -\sum_{x \in B} wIyIt \\ -\sum_{x \in B} wxIyIt \\ -\sum_{x \in B} wyIxIt \end{bmatrix}$$

### 3στ) Εκτίμηση παραμετρικού μοντέλου από διανύσματα κίνησης

Σε αυτή την μέθοδο επιδιώκουμε να ελαχιστοποιήσουμε την σχέση

$$\min_{a_0, a_5} \|u - u'\|^2 + \|v - v'\|^2$$

όπου  $u = u(a_0, a_1, a_2)$  και  $v = v(a_3, a_4, a_5)$  τα παραμετρικά διανύσματα κίνησης που περιγράφονται από τις εξισώσεις (8) και (9) της υποενότητας 3δ. Τα  $u', v'$  είναι τα διανύσματα της οπτικής ροής. Το πρόβλημα αυτό μπορεί να σπάσει σε δύο μέρη εξαιτίας του γεγονότος ότι μόνο οι  $a_0, a_1, a_2$  παράμετροι περιγράφουν την κίνηση στον x-άξονα και αντίστοιχα στον y-άξονα οι παράμετροι  $a_3, a_4, a_5$ .

Συνεπώς μπορούμε να εξετάσουμε ξεχωριστά τις παρακάτω σχέσεις χωρίς η μία να επηρεάζει την άλλη.

$$\min_{a_0, a_1, a_2} \|u - u'\|^2 \quad (11)$$

$$\min_{a_3, a_4, a_5} \|v - v'\|^2 \quad (12)$$

Ουσιαστικά έχουμε να λύσουμε ένα παρόμοιο πρόβλημα με αυτό του (Κεφ.3-παρ.δ) αποκλείοντας από αυτή την διαδικασία τις χωρικές παραγώγους  $I_x(\mathbf{x}, t), I_y(\mathbf{x}, t)$  και την χρονική παράγωγο  $I_t(\mathbf{x}, t)$ . Αντίθετα συμμετέχουν τα διανύσματα οπτικής ροής  $u', v'$ .

Τα βήματα που ακολουθούνται είναι ίδια με αυτά της διαφορικής affine μεθόδου(3δ). Δηλαδή παραγωγίζουμε τις σχέσεις (11) και (12) ως προς  $a_0, a_1, a_2, a_3, a_4, a_5$  και τις θέτουμε ίσες με το μηδέν. Άρα έχουμε έξι εξισώσεις με έξι αγνώστους. Όμως οι  $a_0, a_1, a_2$  είναι υπεύθυνες για την κίνηση στον x-άξονα οπότε μηδενίζουμε τα στοιχεία των πινάκων A και B (Κεφ.3-παρ.δ) στα οποία συμπεριλαμβάνεται η χωρική παράγωγος  $I_y(\mathbf{x}, t)$  διότι οι  $a_0, a_1, a_2$  δεν περιγράφουν κίνηση στον y-άξονα. Ομοίως γίνεται και για τις παραμέτρους  $a_3, a_4, a_5$  μηδενίζουμε αντίστοιχα τα στοιχεία των πινάκων A, B (Κεφ.3-παρ.δ) όπου υπάρχει η  $I_x(\mathbf{x}, t)$  χωρική παράγωγος. Σε αυτή την διαδικασία περικλείεται και η χρήση του βάρους(weight). Οπότε από την εξίσωση  $C = A^{-1}B$  παίρνουμε τις affine εξάδες για κάθε τετραγωνικό block. Πιο αναλυτικά έχουμε

$$C = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}$$

$$A = \begin{bmatrix} \sum_{x \in B} w & \sum_{x \in B} wx & \sum_{x \in B} wy & 0 & 0 & 0 \\ \sum_{x \in B} wx & \sum_{x \in B} wx^2 & \sum_{x \in B} wxy & 0 & 0 & 0 \\ \sum_{x \in B} wy & \sum_{x \in B} wxy & \sum_{x \in B} wy^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sum_{x \in B} w & \sum_{x \in B} wx & \sum_{x \in B} wy \\ 0 & 0 & 0 & \sum_{x \in B} wx & \sum_{x \in B} wy^2 & \sum_{x \in B} wxy \\ 0 & 0 & 0 & \sum_{x \in B} wy & \sum_{x \in B} wxy & \sum_{x \in B} wy^2 \end{bmatrix}$$

$$B = \begin{bmatrix} -\sum_{x \in B} wU \\ -\sum_{x \in B} wxU \\ -\sum_{x \in B} wyU \\ -\sum_{x \in B} wV \\ -\sum_{x \in B} wxV \\ -\sum_{x \in B} wyV \end{bmatrix}$$

Οπότε έχοντας υπολογίσει τις affine παραμέτρους για κάθε block αντικαθιστούμε αυτές τις τιμές στις εξισώσεις (8) και (9) (Κεφ.3-παρ.δ) εξαγάγουμε τα παραμετρικά διανύσματα με την βοήθεια των διανυσμάτων κίνησης της οπτικής ροής.

## ΚΕΦΑΛΑΙΟ 4

### Η εξαγωγή των επιπέδων (layers) με την χρήση του affine πολλαπλών βαθμίδων κατηγοριοποίησης αλγορίθμου (Multistage Affine Classification Algorithm)

#### 4 α)Εισαγωγή

Έχοντας δώσει δύο αλγορίθμους (Κεφ.3-παρ.δ-στ) για την εκτίμηση των παραμετρικών μοντέλων στις τετραγωνικές περιοχές δηλαδή στα block της εικόνας διαστάσεων  $16*16$  εξάγουμε τα παραμετρικά διανύσματα κίνησης για κάθε pixel από τις σχέσεις

$$u(x, y) = a_0 + a_1x + a_2y$$

$$v(x, y) = a_3 + a_4x + a_5y$$

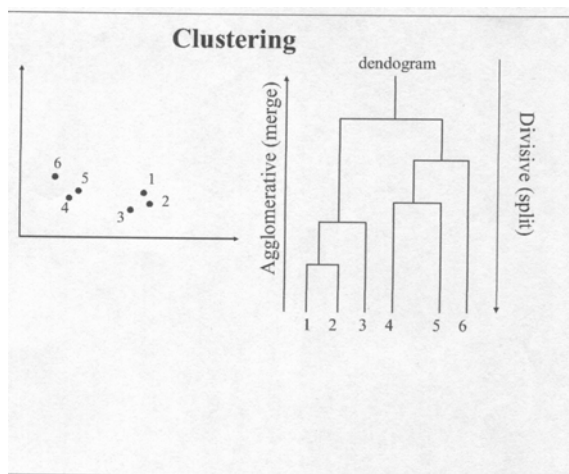
τις οποίες περιγράψαμε αναλυτικά στο προηγούμενο κεφάλαιο (παρ.3δ).Σημειώνουμε ότι οι affine παράμετροι είναι ίδιες για τα pixels του ίδιου block.Έτσι μπαίνουμε στην διαδικασία για να εξάγουμε τα επίπεδα (layers) με την χρήση του affine πολλαπλών βαθμίδων κατηγοριοποίησης αλγορίθμου (multistage affine classification algorithm).Προτού προχωρήσουμε στην περιγραφή αυτού του αλγορίθμου αναφέρουμε κάποια σημαντικά στοιχεία για τον k-means αλγόριθμο ο οποίος χρησιμοποιείται από τον multistage affine classification αλγόριθμο σε ένα από τα σταδιά του και συγκεκριμένα στην διαδικασία ένωσης (merge procedure)(Κεφ.4-παρ.δ2).

## 4 β) Ο k-means αλγόριθμος

Ο K-Means είναι η απλούστερη λύση clustering [2] [3] [4] [5] [6] [7]. Αυτός επαναληπτικά τοποθετεί κάθε block στο layer από το κέντρο του οποίου απέχει την λιγότερη απόσταση. Η απόσταση μεταξύ των blocks ως προς το πολυδιάστατο affine διάστημα (Κεφ.4-παρ.δ) πρέπει να ορίζεται. Αρκετά διαφορετικά αποτελέσματα θα πάρουμε με διαφορετικές αποστάσεις. Η σύγκλιση του αλγορίθμου εξαρτάται από το σημείο εκκίνησής του. Δηλαδή από τις περιοχές που ορίζονται αρχικά ως κέντρα των layers. Ουσιαστικά ο αλγόριθμος αυτός λειτουργεί ως εξής.

- Έστω ότι δίνεται ένα σύνολο χαρακτηριστικών  $\{X_i\}$  (στη δική μας περίπτωση τα χαρακτηριστικά είναι οι affine παράμετροι του κάθε block)
- Υπολογίζονται οι αποστάσεις μεταξύ των χαρακτηριστικών  $d(x, x')$
- Βρίσκεται από τον αλγόριθμο μία κατάτμιση τέτοια ώστε. Εάν η  $d(x_i, x_j)$  είναι «μικρή» τότε τα  $x_i, x_j$  ανήκουν στο ίδιο layer ενώ εάν είναι η  $d(x_i, x_j)$  είναι «μεγάλη» τότε τα  $x_i, x_j$  ανήκουν σε διαφορετικά layer.

Την παραπάνω λειτουργία μπορούμε να την παρατηρήσουμε και στο σχήμα 6.



Σχήμα 6. Η τελική κατάσταση του k-means

Στο σχήμα 6 βλέπουμε την κατάσταση στην οποία καταλήγει ο αλγόριθμος. Τα στοιχεία 4,5,6 ανήκουν στο ίδιο cluster και τα στοιχεία 1,2,3 ανήκουν σε διαφορετικό από ότι τα 4,5,6 διότι οι αποστάσεις των 1,2,3 ως προς τα 4,5,6 είναι «μεγάλες».

#### **4 γ) Η αναγκαιότητα χρήσης του multistage affine classification αλγορίθμου**

Κάθε block περιγράφεται από μία εξάδα affine παραμέτρων. Εάν υπάρχει απλή κίνηση σε αυτό το block τότε αυτές οι παράμετροι παρέχουν ακριβείς πληροφορίες για την κίνηση. Υπάρχει όμως και η περίπτωση σε κάποια block να παρουσιάζεται πολλαπλή κίνηση δηλαδή περισσότερα του ενός layer να συμμετέχουν σε αυτά. Σε αυτή την περίπτωση η κίνηση που περιγράφουν οι affine παράμετροι και κατά συνέπεια τα παραμετρικά διανύσματα κίνησης είναι λάθος για κάποια τουλάχιστον pixels. Επομένως χρειαζόμαστε μία μέθοδο για να εγγυάται την ποιότητα του παραμετρικού μοντέλου σε αυτά τα blocks.

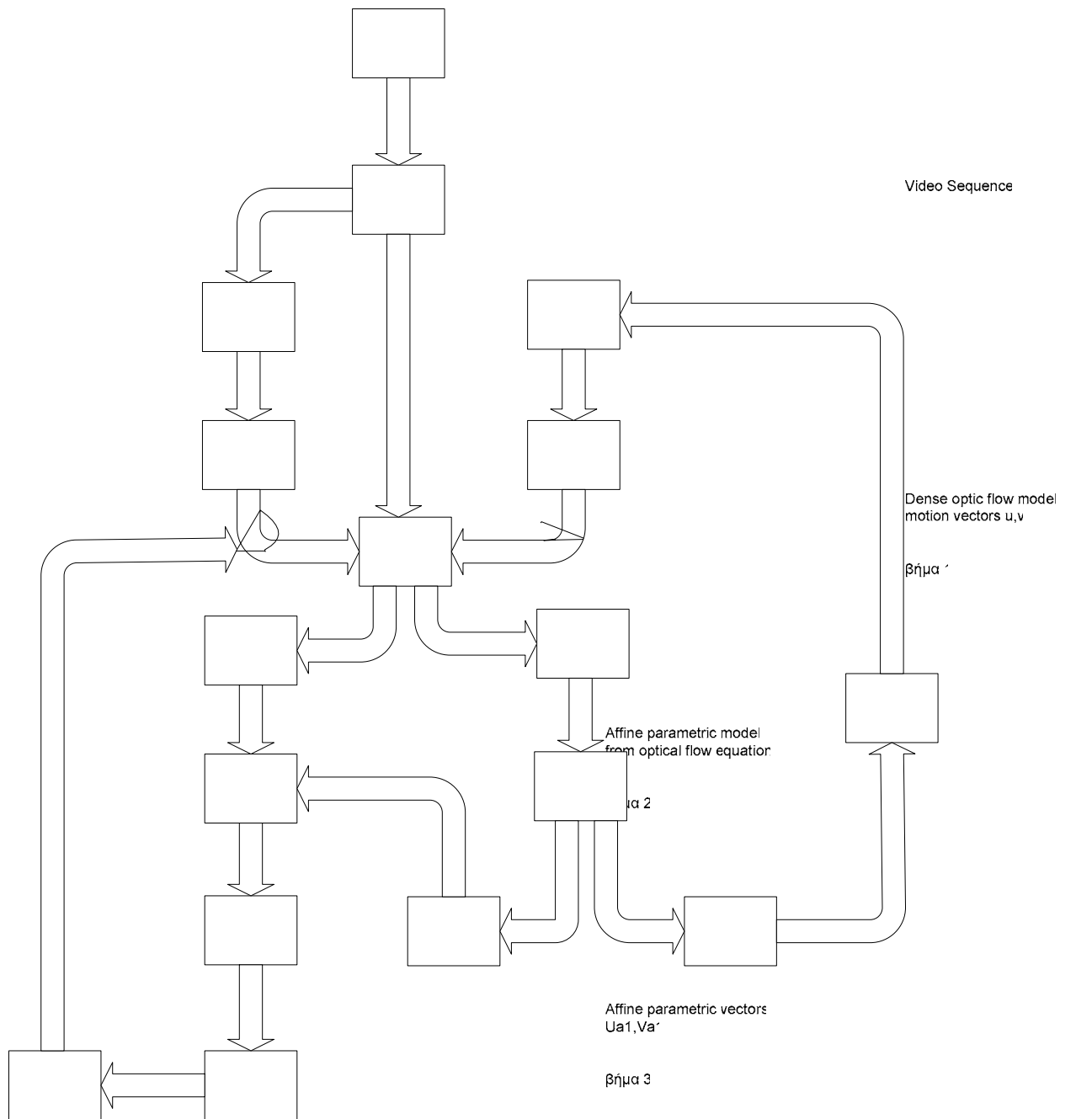
Εάν κάποιο block χαρακτηρίζεται σωστά από το παραμετρικό μοντέλο τότε μπορεί να καταταχθεί ολόκληρο σε κάποιο από τα layers. Στην αντίθετη περίπτωση «σπάμε» το block σε ομάδες pixels και θεωρούμε κάθε ομάδα ξεχωριστά.

Τέλος για όμοιες περιοχές (blocks ή κομμάτια blocks) τα οποία διέπονται από το ίδιο (παρόμοιο) παραμετρικό μοντέλο, χρειάζεται μία διαδικασία συνένωσης (merging) σε μεγαλύτερες περιοχές. Οι διαδικασίες αυτές εξετάζονται αναλυτικότερα στα επόμενα κεφάλαια.

## 4δ)Γενική περιγραφή του multistage affine classification αλγορίθμου

Ως είσοδο στον multistage affine classification αλγόριθμο έχουμε τα διανύσματα κίνησης (motion vectors) για κάθε pixel που έχουν υπολογιστεί από την κανονικοποίηση της εξίσωσης οπτικής ροής με την μέθοδο Horn-Schunk. Αρχικά υπολογίζουμε ένα παραμετρικό μοντέλο για κάθε block από την εξίσωση οπτικής ροής (Ο.Φ.Ε.-Κεφ.3-παρ.δ). Από τα παραμετρικά μοντέλα υπολογίζουμε τα παραμετρικά διανύσματα κίνησης μέσω των σχέσεων (8) και (9) του κεφ. 3 της παραγράφου 3β. Αρχικά ο αλγόριθμος συγκρίνει δύο είδη διανυσμάτων κίνησης σε όλο το block, τα παραμετρικά διανύσματα και τα διανύσματα οπτικής ροής (Κεφ.4-παρ.δ1). Αν το σφάλμα είναι μικρό τότε τα block στα οποία υπάρχει αυτή η κατάσταση κρίνονται αξιόπιστα για την περιγραφή της κίνησης και μπορούν να καταταχθούν σε κάποιο από τα (layers) μέσω της διαδικασίας ένωσης (merge procedure) (Κεφ4.-παρ.δ2). Αν το σφάλμα είναι μεγάλο τότε «σπάμε» το block (splitting) σε υποπεριοχές μέσω της διαδικασίας κατηγοριοποίησης (labeling procedure) (Κεφ.4-παρ.δ3). Σε αυτήν την περίπτωση το αρχικό παραμετρικό μοντέλο που έχουμε υπολογίσει διέπει μόνο την «κύρια» κίνηση μέσα στο block (λόγω ότι έχουμε χρησιμοποιήσει robust αλγόριθμο για τον υπολογισμό του μοντέλου αυτού). Στην συνέχεια βρίσκουμε τα pixels για τα οποία τα δύο μοντέλα κίνησης είναι παρόμοια, δηλαδή δίνουν μικρό σφάλμα ανά pixel. Αυτά τα pixels (consistent pixels) κατηγοριοποιούνται (label) με βάση το αρχικό παραμετρικό μοντέλο κίνησης που τα διέπει (το οποίο είναι ήδη υπολογισμένο) και αποτελούν τα pixels που υπακούουν στο μοντέλο της «κύριας» κίνησης μέσα στο block. Έτσι και αυτά με την σειρά τους μέσω της διαδικασίας ένωσης (merge procedure) μπορούν να καταταχθούν σε κάποιο από τα layers. Τα υπόλοιπα pixels (unlabeled pixels) χρειάζεται να ομαδοποιηθούν ξανά για να δώσουν ένα νέο παραμετρικό μοντέλο (δευτερεύουσα κίνηση μέσα στο block) είτε με την εκτίμηση από τα Horn-Schunk διανύσματα κίνησης, είτε από την εξίσωση οπτικής ροής. Ο multistage affine classification αλγόριθμος περιγράφεται από το ακόλουθο διάγραμμα ροής.





Διάγραμμα Ροής

Reliable Blocks

βήμα 5

Merge Procedure

## 4δ) Περιγραφή του multistage affine classification αλγορίθμου

### 4 δ1)Υπολογισμός των αξιόπιστων (reliable) blocks και διαδικασία διαχωρισμού (splitting procedure)

Σαν είσοδο στην παρακάτω διαδικασία block difference(σχήμα 12) έχουμε δύο ειδών διανυσμάτων κίνησης.Τα διανύσματα οπτικής ροής(dense optic flow) και τα παραμετρικά διανύσματα κίνησης τα οποία μπορούν να υπολογιστούν με δύο τρόπους ,με την χρήση της εξίσωσης οπτικής ροής  $(u_{a1}, v_{a1})$  και από τα διανύσματα κίνησης από την dense optic flow  $(u_{a2}, v_{a2})$  βήματα (7,8,9,10,11).Η εικόνα έχει χωριστεί σε 300 τετραγωνικές περιοχές μεγέθους  $(16*16)$ .Χρησιμοποιώντας την σχέση

$$\sigma^2 = \frac{1}{N} \sum_{x \in B} \sum_{y \in B} [(U(x, y) - Ua(x, y))^2 + (V(x, y) - Va(x, y))^2] \quad (13)$$

όπου ως  $(x,y)$  δηλώνεται η θέση των pixel της εικόνας , τα  $U(x, y)$  ,  $V(x, y)$  δηλώνουν τα dense optic flow vectors ,τα  $Ua(x, y)$  ,  $Va(x, y)$  δηλώνουν τα παραμετρικά διανύσματα (parametric vectors) και N είναι ο αριθμός των pixel που υπάρχουν σε ένα τετραγωνικό block μεγέθους  $16*16$ .Όπως παρατηρούμε υπολογίζεται το άθροισμα όλων των διαφορών των δύο ειδών διανυσμάτων για όλες τις θέσεις του block της εικόνας.Η τιμή που προκύπτει από αυτή την διαδικασία είναι η τιμή του σ.Τα block που έχουν σ μεγαλύτερο από ένα όριο (threshold) «σπάνε» (splitting) σε υποπεριοχές μέσω της labeling procedure (Κεφ.4-παρ.δ3) διότι οι παραμετροί τους  $(a_0 - a_5)$  και κατά συνέπεια το παραμετρικό μοντέλο κρίνεται μη επαρκές εξαιτίας πολλαπλών κινήσεων μέσα στο block.Έτσι στο τέλος αυτής της διαδικασίας λαμβάνουμε τις αξιόπιστες τετραγωνικές περιοχές (reliable blocks-σχήμα 13) και τις μη αξιόπιστες τετραγωνικές περιοχές (non-reliable blocks-σχήμα 14).

## 4 δ2) Διαδικασία ένωσης (merge procedure)

Έχοντας ως είσοδο τα αξιόπιστα block (βήμα 5) αλλά και αξιόπιστες υποπεριοχές των blocks (βήμα 7) που προκύπτουν από την διαδικασία κατηγοριοποίησης (labeling procedure) προσπαθούμε να τα ομαδοποιήσουμε σε περιοχές παρόμοιας κίνησης (layers). Αυτό επιτυγχάνεται χρησιμοποιώντας τις παραμέτρους του κάθε block και υπολογίζοντας τις μεταξύ τους αποστάσεις ως προς τον εξαδιάστατο affine χώρο βάσει του τύπου

$$D = \|a1 - a2\| = [(a1 - a2)^T (a1 - a2)] \quad (14)$$

όπου  $a1, a2$  είναι πίνακες διαστάσεων  $6 \times 1$  και περιλαμβάνουν τις affine παραμετρους του κάθε block. Ο  $(a1 - a2)$  είναι η διαφορά των πινάκων  $a1, a2$ , ο  $(a1 - a2)^T$  είναι ο ανεστραμμένος πίνακας κατα γραμμή και στήλη του πίνακα  $(a1 - a2)$  και είναι διαστάσεων  $1 \times 6$ . Συνεπώς μπορούμε να έχουμε τις αποστάσεις που έχουν μεταξύ τους όλα τα block.

Σε αυτό το σημείο εφαρμόζεται ο k-means αλγόριθμος προκειμένου να καταλήξουμε στο επιθυμητό αποτέλεσμα των περιοχών με παρόμοια κίνηση. Εκτιμούμε με την εξίσωση (14) τις αποστάσεις που έχουν όλα τα block μεταξύ τους και εφόσον δύο block ή υποπεριοχές των block (Κεφ.4-παρ.δ3) έχουν απόσταση μεταξύ τους  $0.2 < D < 1.0$  γίνονται ένα και οι παράμετροι που περιγράφουν αυτή την περιοχή είναι ο μέσος όρος των affine παραμέτρων των δύο περιοχών. Η διαδικασία αυτή (merge procedure) έχει πολλαπλά περάσματα, απαιτεί την σύγκριση όλων των συνδυασμών των blocks ή υποπεριοχών τους που συμμετέχουν σε μία δεδομένη φάση του αλγορίθμου στην διαδικασία ένωσης (merge-procedure) και σταματάει όταν δεν υπάρχουν πλέον αποστάσεις μεταξύ των block ή υποπεριοχών τους που να βρίσκονται στο διάστημα  $0.2 < D < 1.0$ .

### 4δ3) Διαδικασία κατηγοριοποίησης (labeling procedure)

Όπως προαναφέρθηκε σε όλα τα block δεν είναι δυνατή η ακριβής αναπαράσταση της κίνησης με το affine μοντέλο λόγω πολλαπλών κινήσεων που παρουσιάζονται σε μερικά από αυτά. Έτσι πρέπει να διαχωρίσουμε (splitting) το block σε περιοχές μη σταθερού μεγέθους προκειμένου να διαπιστώσουμε σε ποια layer ανήκει η καθεμία περιοχή του block. Αυτό το πρόβλημα μας το λύνει η διαδικασία κατηγοριοποίησης (labeling procedure). Όσα block θεωρούνται μη αξιόπιστα (non-reliable) υποβάλλονται στην διαδικασία κατηγοριοποίησης για να διαπιστωθεί ποια από τα pixel τους είναι αξιόπιστα. Ο έλεγχος αυτός γίνεται παρόμοια της σχέσης (13) βάσει του τύπου

$$[(U(x, y) - Ua(x, y))^2 + (V(x, y) - Va(x, y))^2] < T \quad (15)$$

όπου  $U(x, y)$ ,  $V(x, y)$  είναι τα dense optic flow vectors και τα  $Ua(x, y)$ ,  $Va(x, y)$  τα parametric vectors. Το  $T$  είναι ένα όριο (threshold) το οποίο καθορίζει την αξιοπιστία των pixel του block. Όσα έχουν τιμή πάνω από το  $T$  δεν συμμετέχουν στη (labeling procedure) αλλά μεταφέρονται στα επόμενα στάδια του αλγορίθμου. Για όσες περιοχές pixel μείνουν unlabeled, χρειάζεται η επανεκτίμηση των affine παραμέτρων τους βάσει των γνωστών μεθόδων (Κεφ.3-παρ.δ-στ) για να μπορέσουν να ομαδοποιηθούν ξανά. Οι περιοχές pixel ή υποπεριοχές του block που δεν κατηγοριοποιήθηκαν (unlabeled) χρησιμοποιούνται στα μετέπειτα στάδια του αλγορίθμου με αυξημένα όρια (thresholds) για την σχέση 15. Ο αλγόριθμος σταματά όταν όλα τα pixel γίνουν labeled.

## 4 στ) Σύντομη περιγραφή του αλγορίθμου

Ο αλγόριθμος που περιγράψαμε αποτελείται από τα εξής βήματα

- Υπολογισμός των διανυσμάτων οπτικής ροής κατά Horn και Schunck
- Υπολογισμός των παραμετρικών affine μοντέλων κίνησης μέσω της O.F.E.
- Υπολογισμός των παραμετρικών διανυσμάτων κίνησης σύμφωνα με τις σχέσεις  $u(x, y) = a_0 + a_1x + a_2y$  και  $v(x, y) = a_3 + a_4x + a_5y$

- Υπολογισμός των αξιόπιστων block σύμφωνα με την σχέση

$$\sigma^2 = \frac{1}{N} \sum_{x \in B} \sum_{y \in B} [(U(x, y) - Ua(x, y))^2 + (V(x, y) - Va(x, y))^2]$$

- Κατηγοριοποίηση των pixel των αναξιόπιστων block(labeling procedure) σύμφωνα με την σχέση  $[(U(x, y) - Ua(x, y))^2 + (V(x, y) - Va(x, y))^2] < T$ .

Προκύπτουν υποπεριοχές αυτών των blocks.Εφόσον έχουμε υποπεριοχές non-labeled επανεκτίμηση των affine παραμέτρων αυτών σύμφωνα με τις μεθόδους των παραγράφων 3δ-3στ και συμμετοχή τους στα επόμενα στάδια του αλγορίθμου . Έφασον έχουμε υποπεριοχές labeled τότε συμμετέχουν στην διαδικασία ένωσης (merge procedure).

- Διαδικασία ένωσης (merge procedure) ολόκληρων των block (reliable) ή υποπεριοχών άλλων blocks (labeled).Ένωση αυτών που απέχουν στον εξαδιάστατο affine χώρο  $0.2 < D < 1.0$  και σαν καινούριες affine παράμετροι των ενωμένων περιοχών ο μέσος όρος τους.Συνεχίζεται μέχρι να μην υπάρχουν αποστάσεις στην περιοχή  $0.2 < D < 1.0$

- Όρισε τις νέες περιοχές,block που κρίθηκαν αναξιόπιστα στο προηγούμενο πέρασμα και για τα μη κατηγοριοποιημένα pixel (unlabeled) αύξησε τα όρια

(threshold) και επέστρεψε στο βήμα 4.Επανέλαβε μέχρι τα γίνουν όλα τα pixel κατηγοριοποιημένα (labeled).

## ΚΕΦΑΛΑΙΟ 5

### Πειραματικά αποτελέσματα

#### 5α) Περιγραφή των video sequences

Εφαρμόσαμε τον multistage affine classification αλγόριθμο σε διάφορα διαδοχικά καρέ (frames) των “flow garden”, “coast guard” και “ping pong” video sequences. Ο αλγόριθμος εφαρμόστηκε δύο φορές σε κάθε video sequence για να εξετάσουμε τις δυνατότητες και τις διαφορές των δύο μεθόδων υπολογισμού παραμετρικών μοντέλων i) για το affine παραμετρικό μοντέλο κίνησης όπως αυτό προέκυψε από την εξίσωση οπτικής ροής και ii) για το affine παραμετρικό μοντέλο κίνησης που υπολογίζεται από τα διανύσματα κίνησης (motion vectors) της dense optic flow. Η διαφορά των δύο μεθόδων εντοπίζεται μόνο στο σημείο του υπολογισμού παραμέτρων της δευτερεύουσας κίνησης σε subblocks. Η πρώτη μέθοδος που βασίζεται απευθείας στην εξίσωση οπτικής ροής θεωρείται ακριβέστερη της δεύτερης που ομαδοποιεί motion vectors τα οποία έχουν ήδη υπολογιστεί με βάση την εξίσωση οπτικής ροής. Για αυτό τον λόγο χρησιμοποιείται πάντα στο πρώτο επίπεδο σύγκρισης κατά blocks. Σε επόμενο επίπεδο σύγκρισης κατά pixel, η χρήση της δεύτερης μεθόδου οργάνωσης των διανυσμάτων κίνησης που είναι ήδη διαθέσιμα επιταγχύνει τη διαδικασία υπολογισμού παραμέτρων και μειώνει το υπολογιστικό κόστος. Στην εργασία αυτή μας ενδιαφέρουν να εξετάσουμε τις δυνατότητες, ως προς την ακρίβεια και τα προβλήματα που δημιουργεί το σχήμα αυτό.

Στην “flow garden” video sequence υπάρχει translational κίνηση της κάμερας προς τα δεξιά χωρίς να υπάρχει κάποια κίνηση αντικειμένου μέσα στην σκηνή. Παρατηρούμε ότι τα αντικείμενα ενδιαφέροντος, το δέντρο, ο κήπος και το σπίτι-ουρανός (σχήμα 8) <<κινούνται>> προς τα αριστερά καθώς η κάμερα κινείται στην αντίθετη κατεύθυνση. Οι κινήσεις που κάνουν αυτές οι περιοχές είναι translational. Το δέντρο που βρίσκεται πιο κοντά στην κάμερα κινείται με μεγαλύτερη ταχύτητα προς τα αριστερά από τις υπόλοιπες περιοχές (σπίτι-ουρανός, κήπος) που βρίσκονται σε μεγαλύτερο βάθος, καθώς η κάμερα κινείται προς τα δεξιά. Τα επίπεδα (layers) που εξάγουμε και με τους

δύο διαφορετικούς τρόπους εφαρμογής του αλγορίθμου στα διαδοχικά καρέ 22-23 του video sequence (σχήματα 9,15) αντιστοιχούν στο δέντρο, τον κήπο και το επίπεδο σπίτι-ουρανός (σχήματα 9,15) και απεικονίζονται στον πίνακα υποστήριξης (support map) (σχήμα 9) όπου κάθε επίπεδο απεικονίζεται με διαφορετική ένταση.

Στην “coast guard” video sequence αρχικά παρατηρούμε τέσσερα αντικείμενα ενδιαφέροντος. Το μεγάλο πλοίο, το μικρό πλοίο, την ακτή και την θάλασσα (σχήμα 10). Το μεγάλο πλοίο κινείται προς τα δεξιά, η θάλασσα συμπεριφέρεται τυχαία και η κάμερα ακολουθεί το μικρό πλοίο που κινείται προς τα αριστερά. Οι κινήσεις που κάνουν αυτά τα αντικείμενα είναι translational. Τα επίπεδα (layers) που εξάγουμε και με τους δύο τρόπους εφαρμογής του αλγορίθμου στα διαδοχικά καρέ 8 και 9 του video sequence (σχήματα 8,9) είναι η ακτή, η πλώρη του μεγάλου πλοίου και η θάλασσα με το μικρό πλοίο μαζί (σχήματα 11,17) απεικονίζονται στον πίνακα υποστήριξης (support map). Παρατηρούμε ότι ο k-means αλγόριθμος με τα συγκεκριμένα thresholds αποστάσεων που χρησιμοποιήσαμε δημιουργεί 3 αντί για 4 επιθυμητά επίπεδα. Ο αριθμός των επιπέδων μπορεί να αλλάξει, αλλάζοντας thresholds στον k-means αλγόριθμο, αλλά δεν υπάρχει αυτοματοποιημένη μέθοδος για τον καθορισμό αυτό.

Στην “ring pong” video sequence υπάρχουν τέσσερα αντικείμενα ενδιαφέροντος (σχήμα 12). Το τραπέζι του “ring pong”, ο παίκτης, το μπαλάκι και το πίσω μέρος του παίκτη που είναι οι τοίχοι του δωματίου. Η κάμερα ακολουθεί τον παίκτη ο οποίος κάνει τυχαίες κινήσεις. Τα επίπεδα (layers) που εξάγουμε σε αυτή την περίπτωση είναι τρία και αντιστοιχούν στον παίκτη, τους τοίχους του δωματίου μαζί με το πίσω μέρος του τραπέζιου και στο μπροστινό μέρος του τραπέζιου. Ίδιες παρατηρήσεις σχετικά με τον αριθμό των επιπέδων ισχύουν και σε αυτή την περίπτωση εφαρμογής του αλγορίθμου, παρατηρούμε μικρές διαφορές στις δύο μορφές, καθώς καθέμία ξεχωριστά απεικονίζει μικρές λεπτομέρειες που υπάρχουν στην σκηνή και ιδίως σε αυτή όπου τα αντικείμενα ενδιαφέροντος βρίσκονται σε αποστάσεις πολύ μικρές το ένα με το άλλο.



## 5β)Πειραματική σύγκριση των διανυσμάτων κίνησης της dense optic flow και των παραμετρικών διανυσμάτων

Προτού δώσουμε τα αποτελέσματα του αλγορίθμου αναφέρουμε μία μέθοδο σύγκρισης των δύο ειδών διανυσμάτων κίνησης. Η σύγκριση έγινε και για τους δύο τρόπους υπολογισμού του παραμετρικού μοντέλου. Δημιουργήσαμε τρεις εικόνες βάσει των τύπων

$$I_1(x, y) = \sqrt{(U(x, y) - U_a(x, y))^2 + (V(x, y) - V_a(x, y))^2} * \frac{I_{\min}(x, y)}{I_{\max}(x, y)} * 255 \quad (16)$$

$$I_2(x, y) = \sqrt{(U_{a1}(x, y) - U_{a2}(x, y))^2 + (V_{a1}(x, y) - V_{a2}(x, y))^2} * \frac{I_{\min}(x, y)}{I_{\max}(x, y)} * 255$$

(17)

Δύο εικόνες βάσει του τύπου (16)

όπου  $U(x, y), V(x, y)$  τα διανύσματα της dense optic flow και  $U_a(x, y), V_a(x, y)$  τα παραμετρικά διανύσματα που έχουν προκύψει από δύο διαφορετικούς τρόπους εκτίμησης του παραμετρικού μοντέλου (O.F.E.-motion vectors) και μία εικόνα βάσει του τύπου (17) όπου  $U_{a1}(x, y), V_{a1}(x, y)$  τα παραμετρικά διανύσματα από O.F.E. και  $U_{a2}(x, y), V_{a2}(x, y)$  τα παραμετρικά διανύσματα από τα Horn-Schunk διανύσματα. Ο παράγοντας  $\frac{I_{\min}(x, y)}{I_{\max}(x, y)} * 255$  χρησιμοποιήθηκε προκειμένου οι τιμές των pixel των εικόνων  $I_1(x, y), I_2(x, y)$  να περιοριστούν στο διάστημα (0,255). Όπου  $I_{\min}(x, y)$  η ελάχιστη τιμή της  $\sqrt{(U(x, y) - U_a(x, y))^2 + (V(x, y) - V_a(x, y))^2}$  και  $I_{\max}(x, y)$  η μέγιστή της.

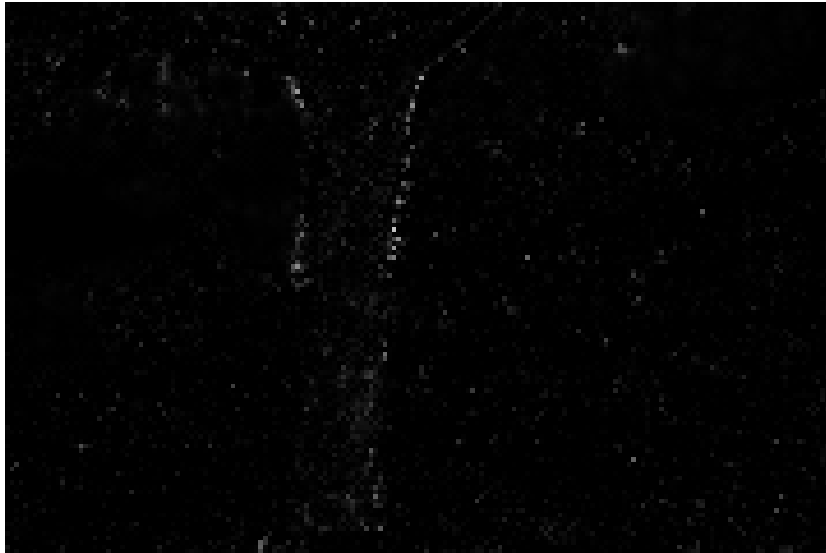
Οι τύποι αυτοί εφαρμόστηκαν στα διαδοχικά καρέ 21,22 (σχήμα 9) της “flow garden” sequence και έδωσαν τα εξής αποτελέσματα.



$\alpha)$



$\beta)$



γ)

Σχήμα 7. Τα αποτελέσματα των σχέσεων 16,17

α)για εκτίμηση απο O.F.E και β)για εκτίμηση από motion vectors

γ)η διαφορά των παραμετρικών διανυσμάτων από O.F.E. και motion vectors

Παρατηρώντας το σχήμα 7 διαπιστώνουμε ότι στις περιοχές με το μαύρο χρώμα συμπίπτουν σε μεγάλο βαθμό τα δύο ειδών διανύσματα κίνησης,δηλαδή το παραμετρικό μοντέλο απεικονίζει ακριβώς την κίνηση.Αντίθετα κάποιες περιοχές με γκριζο χρώμα οι οποίες και είναι τετραγωνικά blocks απεικονίζουν σφάλματα του παραμετρικού μοντέλου και για τις δύο περιπτώσεις.Οι περιοχές αυτές διαφέρουν στις δύο εικόνες πράγμα που πιστοποιεί τις διαφορές που υπάρχουν απο τους δύο διαφορετικούς τρόπους εκτίμησης του παραμετρικού affine μοντέλου.

Μελετώντας το σχήμα 7γ παρατηρούμε ότι η διαφορά των δύο μεθόδων υπολογισμού παραμετρικής κίνησης εμφανίζεται σε block μορφή,αν και οι διαφορές υπολογίζονται ανά pixel (για τα παραμετρικά διανύσματα κίνησης).Αυτό συμβαίνει λόγω του υπολογισμού και των δύο μεθόδων ανά block και αυτές οι block συσχετίσεις καθώς και οι block διαφοροποιήσεις διατηρούνται σε όλη την έκταση του κάθε block.

**5γ)Αποτελέσματα αλγορίθμου με την χρήση του παραμετρικού μοντέλου από Ο.Φ.Ε.**

**5γ1)Αποτελέσματα “flow garden”**



α)



β)

Σχήμα 8. Διαδοχικά καρέ του “flow garden”

α)Καρέ 21

β)Καρέ 22



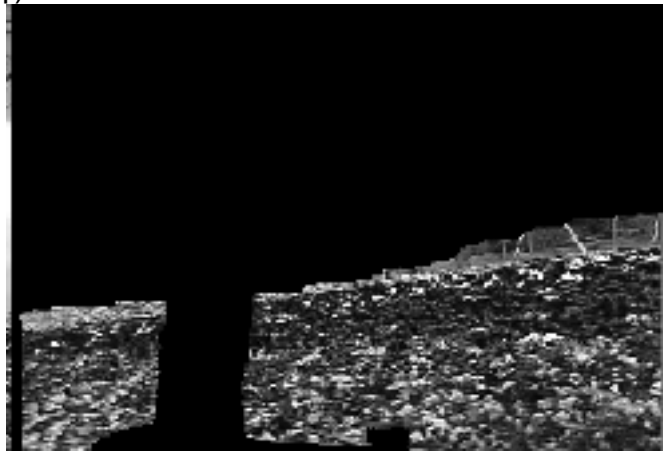
α)



β)



γ)



δ)

Σχήμα 9.

Επίπεδα “flow garden”  
 α)πίνακας-layers β) layer-δέντρο  
 γ)layer-σπίτι-ουρανός δ)layer-κήπος

## 5γ2)Αποτελέσματα “coast guard”



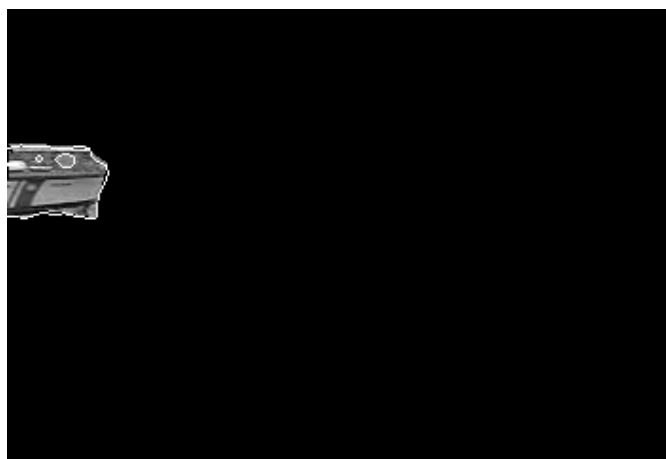
α)



β)  
Σχήμα 10. Διαδοχικά καρέ “coast guard”  
α)Καρέ 8                      β)Καρέ 9



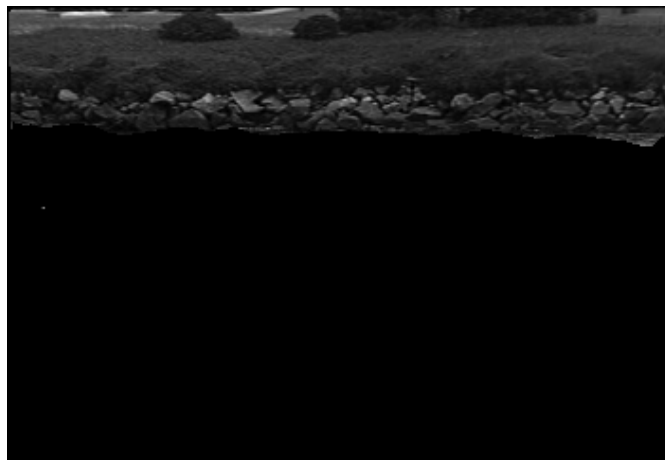
α)



β)



γ)



δ)

Σχήμα 11

Επίπεδα “Coast Guard”

α)πίνακας-layers β)layer-πλώρη

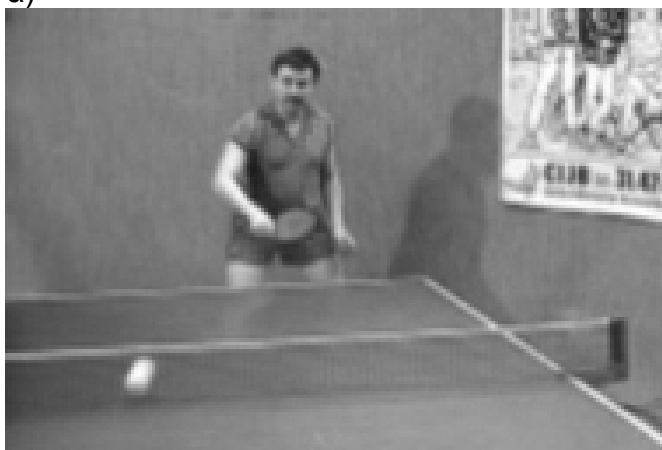
γ) layer-θάλασσα δ) layer-ακτή



### 5γ3)Αποτελέσματα “ring pong”

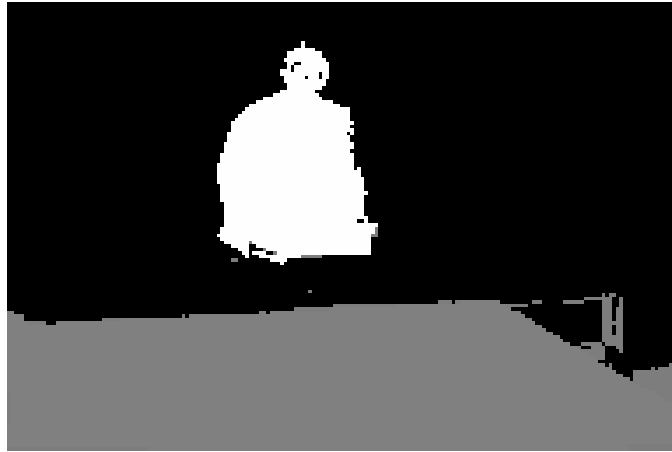


α)



β)

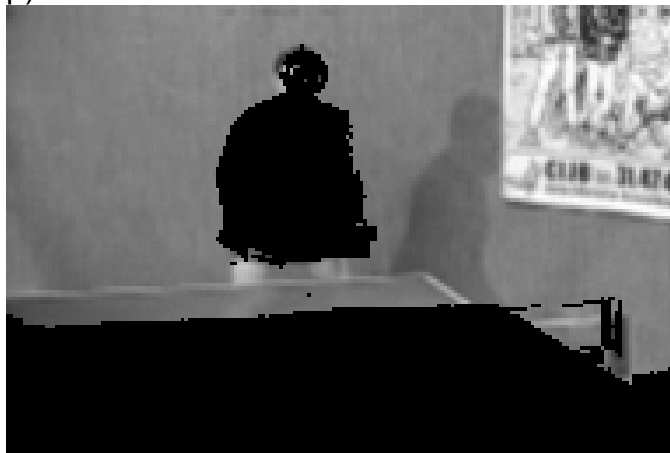
Σχήμα 12. Διαδοχικά καρέ “ring pong”  
α)Καρέ 50      β)Καρέ 51



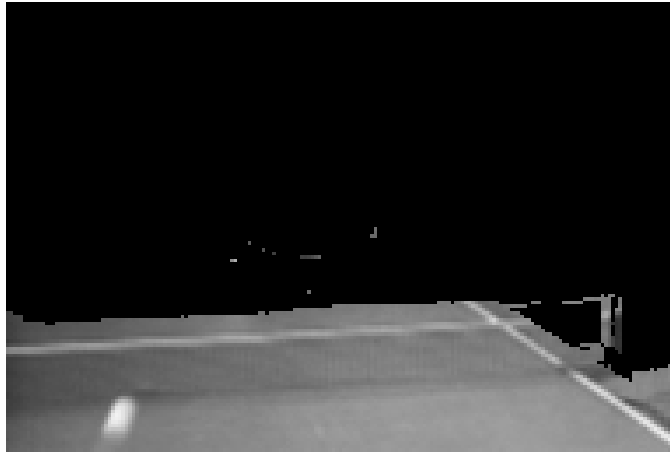
α)



β)



γ)



δ)

Σχήμα 13.

Επίπεδα “ring pong”

α)πίνακας-layers β)layer-παίκτη

γ) layer-πίσω μέρος τραπεζιού-τοίχοι

δ) layer-μπροστινό μέρος τραπεζιού

**5δ)Αποτελέσματα αλγορίθμου με τη χρήση του παραμετρικού μοντέλου από motion vectors**

**5δ1)Αποτελέσματα “flow garden”**



α)



β)

Σχήμα 14. Διαδοχικά καρέ του “flow garden”

α)Καρέ 21

β)Καρέ 22



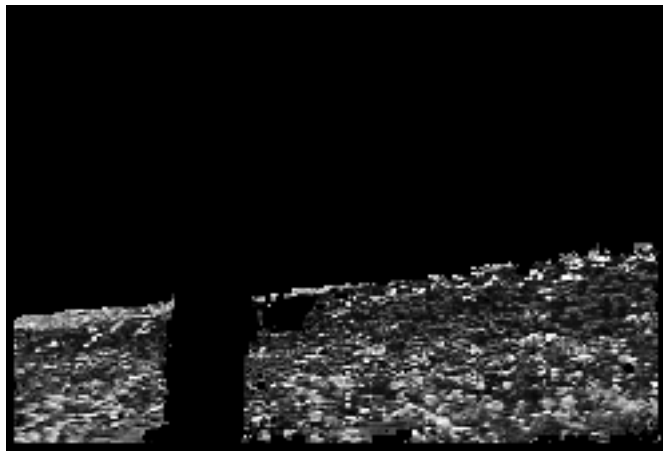
α)



β)



γ)



δ)

Σχήμα 15.

Επίπεδα “flow garden”

α)πίνακας-layers

β) layer-δέντρο

γ)layer-σπίτι-ουρανός

δ)layer-κήπος

## 5δ2)Αποτελέσματα “coast guard”

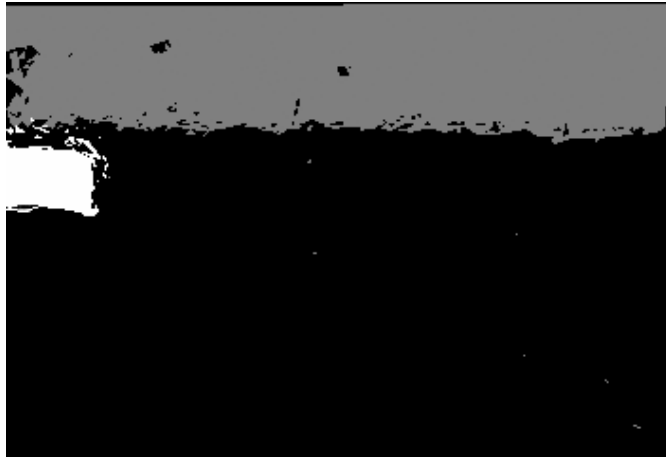


α)

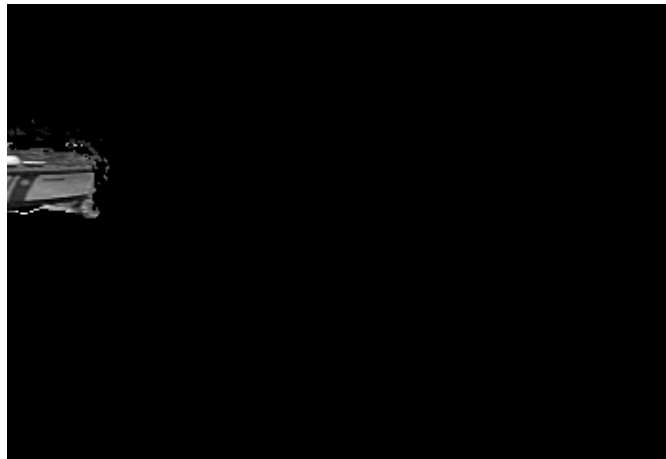


β)

Σχήμα 16. Διαδοχικά καρέ “coast guard”  
α)Καρέ 8                      β)Καρέ 9



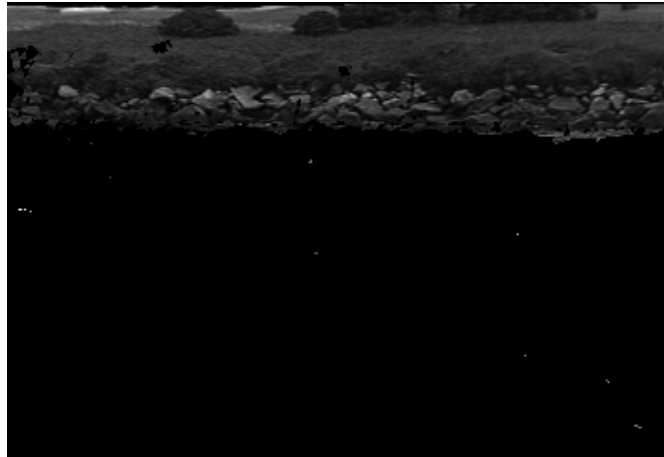
α)



β)



γ)



δ)

Σχήμα 17.

Επίπεδα “Coast Guard”

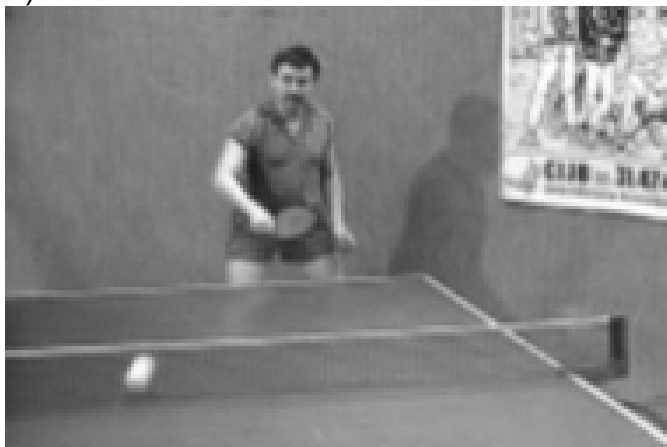
α) πίνακας-layers β) layer-πλώρη

γ) layer-θάλασσα δ) layer-ακτή

### 5δ3) Αποτελέσματα “ring pong”

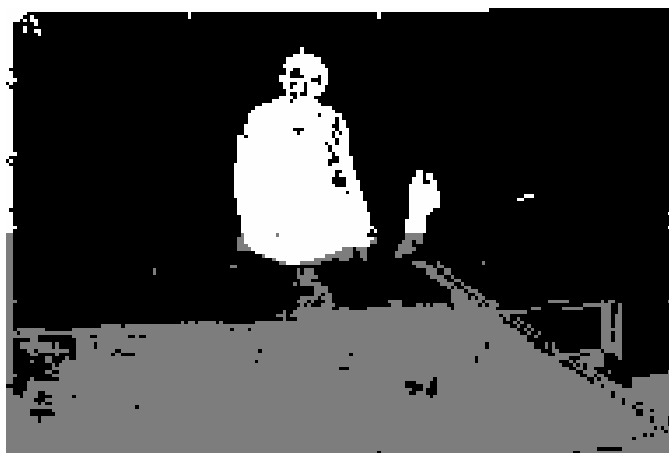


α)

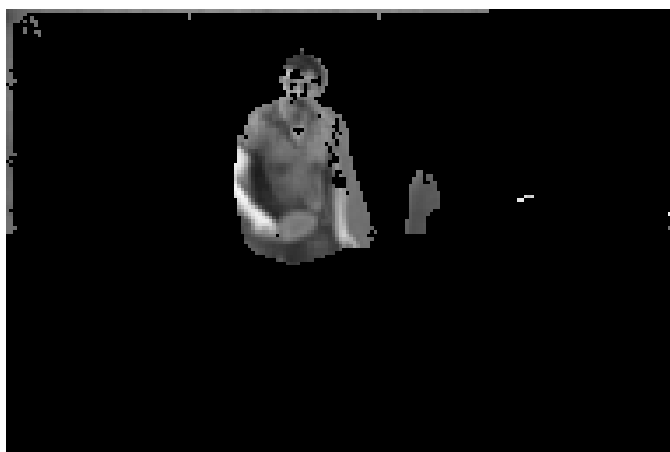




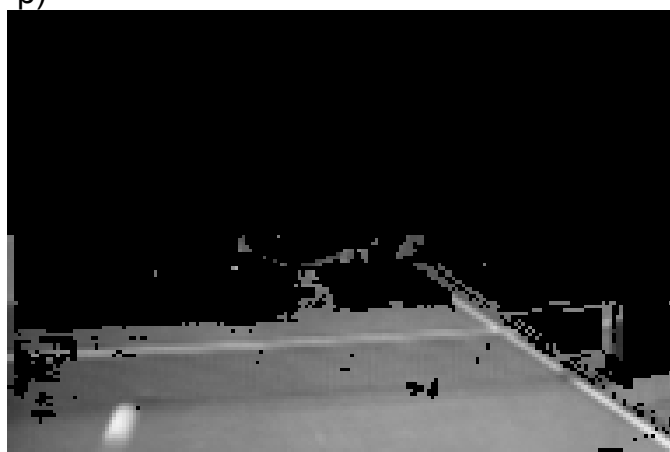
β)  
Σχήμα 18. Διαδοχικά καρέ “ping pong”  
α)Καρέ 50 β)Καρέ 51



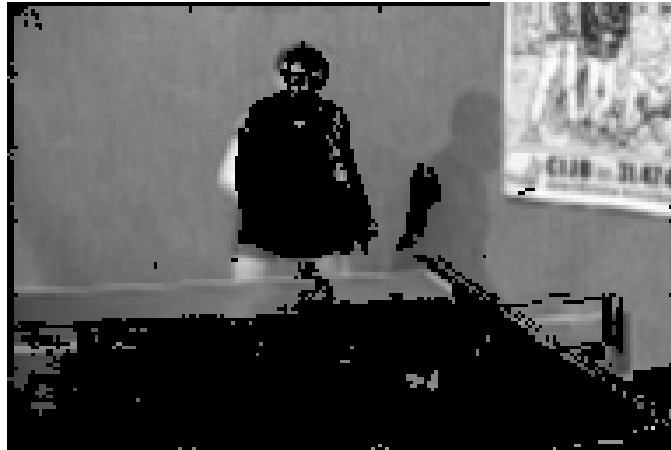
α)



β)



γ)



δ)  
Σχήμα 19.

Επίπεδα “ring pong”  
α)πίνακας-layers β)layer-παίκτη  
γ)layer-μπροστινό μέρος τραπέζιού  
δ)layer-τοίχων

## ΚΕΦΑΛΑΙΟ 6

### **Σύγκριση των δύο μεθόδων-Συμπεράσματα-Μελλοντικές Επεκτάσεις**

Και οι δύο διαφορετικές εφαρμογές του αλγορίθμου έδωσαν ικανοποιητικά αποτελέσματα στην αναπαράσταση των επιπέδων (layers). Παρουσίασαν ωστόσο και κάποια μειονεκτήματα τα οποία εξετάζουμε πιο αναλυτικά. Συγκεκριμένα ο πρώτος τρόπος εφαρμογής του αλγορίθμου (μοντέλο από O.F.E.) είναι περισσότερο ευσταθής. Παρατηρώντας τους πίνακες υποστήριξης (support maps) (σχήματα 9 α, 11 α, 13 α, 15 α, 17 α, 19 α ) βλέπουμε ότι κάθε layer είναι σαφώς διαχωρισμένο από τα υπόλοιπα και γενικότερα οι περιοχές παρουσιάζουν μία συνέχεια. Όμως αυτός ο τρόπος

εφαρμογής μειονεκτεί στο γεγονός ότι οι περιοχές κάποιου layer εισέρχονται στις περιοχές κάποιου άλλου. Αυτό το διαπιστώνουμε μελετώντας το σχήμα 9-α όπου το layer του δέντρου εισέρχεται στο layer του σπίτι-ουρανός στην περιοχή που βρίσκεται ανάμεσα στα κλαριά του δέντρου, όπως επίσης και το layer του κήπου εισέρχεται στο layer σπίτι-ουρανός (σχήμα 9-δ). Στα ίδια συμπεράσματα καταλήγουμε παρατηρώντας το layer πλώρης του μεγάλου πλοίου να εισέρχεται στο layer της θάλασσας (σχήμα 11-β) καθώς και το layer ακτής να εισέρχεται στο layer της θάλασσας (σχήμα 11-δ).

Ο δεύτερος τρόπος εφαρμογής του αλγορίθμου (εκτίμηση μοντέλου από motion vectors) είναι περισσότερο ακριβής στην περιγραφή των ορίων των layer. Παρατηρώντας τα σχήματα (9α, 11α, 13α, 15α, 17α, 19α) αντιλαμβανόμαστε ότι αποσαφηνίζει με μεγαλύτερη ακρίβεια τα όρια κάθε layer σε σχέση με την πρώτη μέθοδο. Αυτό παρατηρείται έντονα στο σχήμα 11β όπου γίνεται σαφής διαχωρισμός ανάμεσα στα κλαριά του δέντρου και τον ουρανό, όπως και στο σχήμα 17 β όπου διαχωρίζεται σαφώς η πλώρη του μεγάλου πλοίου από την θάλασσα. Όμως παρουσιάζει σφάλματα σε μικρές περιοχές των επιπέδων όπου κάποιες μικρές ομάδες ή και μεμονομένα pixels ανήκουν σε διαφορετικό layer. Αυτό το συμπέρασμα γίνεται αντιληπτό μελετώντας τα σχήματα 15, 17, 19 όπου σε καθένα από αυτά παρατηρείται η παρουσία μικρών περιοχών σε layers ξένα από αυτά. Για παράδειγμα το layer κήπου περιέχει κάποιες μικρές ομάδες pixels οι οποίες ανήκουν σε διαφορετικά layers.

Στα ίδια συμπεράσματα καταλήγουμε μελετώντας και τα αποτελέσματα της “ring pong” video sequence. Και οι δύο μέθοδοι παρουσιάζουν σφάλματα στα όρια της εικόνας. Η δεύτερη μέθοδος (εκτίμηση από motion vectors) απαιτεί λιγότερο χρόνο υπολογισμού και αυτό είναι σαφές διότι από την σχέση που χρησιμοποιείται για να εξαχθεί το παραμετρικό μοντέλο (Κεφ.3-παρ.στ) απαιτούνται προφανέστατα λιγότερες πράξεις. Επιπροσθέτως ο υπολογισμός των διανυσμάτων κίνησης οπτικής ροής απαιτεί αρκετό υπολογιστικό χρόνο.

Οπότε η επιλογή κάποιας μεθόδου εξαρτάται από τις απαιτήσεις του προβλήματος καθώς και από τις δυνατότητες που υπάρχουν σε επίπεδο hardware (υπολογιστική ισχύς, μνήμη).

Επίσης και οι δύο μέθοδοι δεν μπόρεσαν να ξεχωρίσουν την κίνηση της μικρής βάρκας (σχήμα) διότι η κίνησή της είναι πολύ μικρή και είναι δύσκολο να διαχωριστεί από την κίνηση του νερού. Επιπλέον και οι δύο μέθοδοι δεν μπόρεσαν να ξεχωρίσουν το μπαλάκι στο “ring pong” sequence το οποίο καταλαμβάνει πολύ μικρό μέρος της εικόνας.

Για την υλοποίηση του αλγορίθμου χρησιμοποιήθηκε αποκλειστικά η γλώσσα προγραμματισμού C για λόγους ευελιξίας και ταχύτητας. Η δυνατότητα αποδοτικής λειτουργίας του αλγορίθμου σε κινητή κάμερα επιτρέπει την λειτουργία του σε κάθε φυσική σκηνή όπου η κίνηση της κάμερας είναι δεδομένη.

Για την μελέτη αυτή χρησιμοποιήθηκαν εξ’ ολοκλήρου ασπρόμαυρες σκηνές (grayscale). Το χρώμα στην εικόνα περιέχει μεγάλη πληροφορία που είναι πάρα πολύ χρήσιμη και σημαντική. Μελλοντική επέκταση του αλγορίθμου αυτού, είναι η μελέτη και ο συνδυασμός της συμπεριφοράς του σε κάθε χρωματική συνιστώσα ξεχωριστά. Η χρήση του χρώματος θα βοηθήσει σε πιο λεπτομερέστερη περιγραφή των επιπέδων διότι θα έχουμε στην διάθεσή μας περισσότερη πληροφορία.

Επιπρόσθετα σαν μελλοντική επέκταση αυτής της εργασίας είναι ένας καλύτερος αλγόριθμος clustering που να βγάλει με ακρίβεια και τον αριθμό των layers

Επίσης ο συνδυασμός labeling από κίνηση και labeling από intensity segmentation θα μπορούσε να δώσει ακριβέστερη λύση στον υπολογισμό των επιπέδων (layers).

## Παραπομπές

- [1] John Wang and Edward Adelson, Representing moving images with layers, M.I.T. Media laboratory Perceptual Computing Section Technical Report No 279, 1994
- [2] Yu Huang, Dynamic Gesture Analysis and Tracking Based on Dominant Motion Estimation and Kalman Filter, Dept of Computer Science University of Erlangen-Nurnberg, 1995
- [3] Heinrich Niemann, Layers-Based Image Segmentation Incorporating Motion Estimation with Static Segmentation, University of Erlangen-Nurnberg, 1995
- [4] Sawhney H.S., Ayer S: Compac representations of video through dominant and multiple motion estimation, IEEE T-PAMI, 18(8). 1996
- [5] Black M J, Jepson A D: Estimation optical flow in segmented images using variable-order parametric model with local deformations, IEEE T-PAMI, 18(10). 1996
- [6] B. K. Horn and B. G. Schunck, Determining Optical Flow, Artificial Intelligence, Vol. 17, 1981, pp. 185- 204.
- [7] S.S. Beauchemin and J.L. Barron, The computation of optical flow, ACM Computing Surveys, Vol. 27, No.3, pp. 433-467, 1995 H.W. Haussecker

- [8]D.J.Fleet, Computing Optical Flow with Physical Models of Brightness Variation, IEEE Conference on Computer Vision and Pattern Recognition, vol. II, pp. 760-767, 2000
- [9]B. Galvin, B. McCane, K. Novins, D. Mason, S. Mills, Recovering Motion Fields: An Evaluation of Eight Optical Flow Algorithms, British Machine Vision Conference
- [10]L.S. Shapiro,Affine Analysis of Image Sequences.Cambridge University Press,1995
- [11]Andrew W.Moore,K-means and Hierarchical Clustering,School of Computer Science Carnegie Mellon University
- [12]M.J.Black and A.Jepson,Estimating multiple independent motions in segmented images using parametric models with local deformations.Workshop on motion of non-rigid and articulated objects,pp.220-227,Austin ,Nov.1994
- [13]J.L. Barron.Performance of optical flow techniques.IJCV,12(1),1994
- [14]T.Darrell and A.Pentland.Robust estimation of a multi-layer motion representation.Workshop on Visual Motion,pp. 173-178,Princeton,Oct.1991
- [15]G.J. McLachlan and K.E. Basford.Mixture Models:Inference and applications of Clustering.Marcel Dekker Inc.,N.Y.,1988
- [16]L.Wang,S.B.Kang,R.Szelinski and H.Y.Shum,Optimal texture map reconstruction from multiple views.In CVPR01,pages 1:347-354,2001
- [17]A.D. Jepson and M.J. Black.Mixture models for optical flow computation,In CVPR 93
- [18]W.Li and E.Salari,"Successive elimination algorithm for motion estimation",IEEE Trans.Image Processing,Jan 1995
- [19]J.R. Berger,P.J.Burt,Dynamic multiple-motion computation,in Artificial Intelligence and Computer Vision,The Netherlands:Elsevier,1992
- [20]A.M. Tekalp,Digital Video Processing,Englewood Cliffs,NJ:Prentice-Hall,1995