

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

**Εκτίμηση Τρισδιάστατης Κίνησης και Εύρεση
Στρωμάτων Βάθους σε Ακολουθίες Εικόνων**

Νικόλαος Γ. Κομοντάκης

Μεταπτυχιακή Εργασία

Ηράκλειο, 1997

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

Εκτίμηση Τρισδιάστατης Κίνησης και Εύρεση Στρωμάτων Βάθους σε Ακολουθίες Εικόνων

Εργασία που υποβλήθηκε από τον

Νικόλαο Γ. Κομοντάκη

ως μερική εκπλήρωση των απαιτήσεων

για την απόκτηση

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Νικόλαος Γ. Κομοντάκης
Τμήμα Επιστήμης Υπολογιστών

Εισηγητική Επιτροπή:

Γεώργιος Τζιρίτας
Αναπληρωτής Καθηγητής, Επόπτης

Τραγανίτης Απόστολος
Αναπληρωτής Καθηγητής, Μέλος

Τραχανιάς Πάνος
Επικουρος Καθηγητής, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Ηράκλειο, 1997

Εκτίμηση Τρισδιάστατης Κίνησης και Εύρεση Στρωμάτων Βάθους σε Ακολουθίες Εικόνων

Νίκος Κομοντάκης

Μεταπτυχιακή εργασία

Τμήμα Επιστήμης Υπολογιστών

Πανεπιστήμιο Κρήτης

Περίληψη

Δύο είναι τα θέματα που αντιμετωπίζονται κατά την παρούσα εργασία: το πρόβλημα της εκτίμησης της τρισδιάστατης κίνησης και αυτό της εύρεσης των στρωμάτων βάθους μέσα σε μια ακολουθία εικόνων.

Η μέθοδος που ακολουθείται για την επίλυση του πρώτου προβλήματος αποτελείται από δύο στάδια. Κατά το πρώτο στάδιο προσδιορίζεται ένα δισδιάστατο πεδίο οπτικής ροής. Κατά το δεύτερο στάδιο, οι παράμετροι της τρισδιάστατης κίνησης υπολογίζονται μέσω σχέσεων που συνδέουν αυτές με την οπτική ροή που έχει ήδη υπολογισθεί. Το κύριο πρόβλημα που παρουσιάζεται κατά τη διάρκεια της δεύτερης φάσης του αλγορίθμου, έγκειται στο γεγονός ότι η οπτική ροή περιέχει ένα ποσοστό από εσφαλμένες εκτιμήσεις (outliers). Η αρνητική επίδραση των εσφαλμένων αυτών εκτιμήσεων στο τελικό αποτέλεσμα, αυξάνει με την πολυπλοκότητα του μοντέλου που χρησιμοποιείται για τη περιγραφή της κίνησης της κάμερας. Για να αντιμετωπιστεί το πρόβλημα αυτό, η προσέγγιση που υιοθετείται κατά την παρούσα εργασία, συνίσταται στη χρήση μιας ιεραρχίας μοντέλων τρισδιάστατης κίνησης. Σκοπός της ιεράρχησης αυτής είναι η εύρεση του απλούστερου δυνατού μοντέλου, το οποίο είναι ικανό να περιγράψει ικανοποιητικά τη κίνηση της κάμερας. Επιπλέον, για να μειωθεί περαιτέρω η αρνητική επίδραση των outliers γίνεται χρήση και ενός robust αλγορίθμου υπολογισμού.

Σχετικά με το δεύτερο πρόβλημα, δηλαδή την εύρεση των στρωμάτων βάθους, αυτό ανάγεται σε ένα πρόβλημα τμηματοποίησης του πεδίου της οπτικής ροής. Για την τμηματοποίηση αυτή γίνεται χρήση του επαναληπτικού αλγόριθμου χαλάρωσης HCF (Highest Confidence First). Προκειμένου να αυξηθεί η αποτελεσματικότητα και σταθερότητα του αλγορίθμου, η εξαγωγή των στρωμάτων βάθους βασίζεται σε πληροφορία που προέρχεται σε συνδυασμό με την οπτική ροή και από την φωτεινότητα.

Πειραματικά αποτελέσματα εξάγονται για πραγματικές αλλά και συνθετικές ακολουθίες εικόνων με στόχο τον έλεγχο της απόδοσης των προτεινόμενων μεθόδων.

3 - D Motion Estimation and Image Depth Layering

Nikos Komodakis

Master Thesis

Department of Computer Science

University of Crete, Greece

Abstract

In this thesis two problems are addressed: estimation of camera 3-D motion and image depth layering.

For the first problem, the method we adopt consists of two stages. During the first stage a 2-D motion field is computed. During the second stage, the 3-D motion parameters are identified by equations linking the projected 2-D motions and 3-D motions inside the image sequence. During this stage, the main problem comes from the fact that the optical flow field may contain a set of noisy and partially incorrect data (outliers). The negative effects of this set of outliers to the motion estimation, increase with the complexity of the motion model that is used to describe the camera motion. Therefore, the approach we adopt relies on the use of a hierarchy of motion models. Simpler models are first tested, and then more complex models are considered. In this way we aim at finding the simplest possible model, which can adequately describe the camera motion. Additionally, at each stage of the algorithm a robust estimation method is used to cope with the set of outliers.

The second problem, that is image depth layering, is ultimately reduced to the problem of segmenting the optical flow field. For this purpose, an iterative deterministic relaxation algorithm is used and the extraction of the image depth layers is based on both optical flow and

luminance information. The iterative algorithm employed in this thesis is the HCF (Highest Confidence First).

Experimental results regarding real and synthetic image sequences are given to demonstrate the proposed methods.

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες προς τον κ. Γ. Τζιρίτα για την πολύτιμη βοήθεια του στην προετοιμασία της παρούσας εργασίας. Η συμβολή του ήταν καθοριστική σε όλες τις φάσεις, από την αρχή μέχρι το τέλος, και δίχως την ουσιαστική και συνεχή υποστήριξη του θα ήταν αδύνατο να ολοκληρωθεί η παρούσα εργασία. Επιπλέον θέλω να τον ευχαριστήσω διότι σε όλη τη διάρκεια της συνεργασίας μας ήταν πάντα πρόθυμος να με βοηθήσει και να προτείνει λύσεις για τα διάφορα προβλήματα που προέκυπταν.

Θα ήταν παράλειψη εκ μέρους μου να μην ευχαριστήσω τους γονείς μου, Αικατερίνη και Γεώργιο Κομοντάκη, που όλα αυτά τα χρόνια μου συμπαραστάθηκαν με όλες τους τις δυνάμεις. Χωρίς την αμέριστη συμπαράσταση τους θα μου ήταν αδύνατο να πετύχω οτιδήποτε.

Ακόμη, θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες προς τα μέλη της επιτροπής Πάνο Τραχανιά και Απόστολο Τραχανίτη για τις πολύτιμες παρατηρήσεις και διορθώσεις τους. Τέλος, θα ήθελα να ευχαριστήσω το Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης για την υλικοτεχνική υποδομή που μου παρείχε καθώς και το πρόγραμμα ΠΕΝΕΔ για την οικονομική υποστήριξη της εργασίας αυτής.

Περίεχόμενα

1. ΓΕΝΙΚΗ ΕΙΣΑΓΩΓΗ	1
2. ΣΧΕΣΕΙΣ ΟΠΤΙΚΗΣ ΡΟΗΣ ΚΑΙ ΤΡΙΣΔΙΑΣΤΑΤΗΣ ΚΙΝΗΣΗΣ	4
2.1 Εισαγωγή	4
2.2 Ορισμός και διατύπωση του προβλήματος	4
2.3 Πεδίο οπτικής ροής	6
2.4 Μοντέλο κάμερας	6
2.5 Τρισδιάστατα μοντέλα κίνησης	7
2.6 Σχέσεις οπτικής ροής και τρισδιάστατης κίνησης	10
3. M-ESTIMATOR	13
3.1 Εισαγωγή	13
3.2 Robust μέθοδοι υπολογισμού	13
3.3 Περιγραφή του M-estimator	14
3.4 Συναρτήσεις κόστους και επίδρασης	17
3.5 Προβλήματα του M-estimator	19
3.6 AM-estimator	20
3.7 M-estimator και εκτίμηση της κίνησης της κάμερας	22
4. ΙΕΡΑΡΧΙΑ ΤΡΙΣΔΙΑΣΤΑΤΩΝ ΜΟΝΤΕΛΩΝ ΚΙΝΗΣΗΣ	24
4.1 Εισαγωγή	24
4.2 Ιεραρχία τρισδιάστατων μοντέλων κίνησης - Βασικές έννοιες	24
4.3 Κριτήριο αποδοχής - απόρριψης	25
4.4 Σειρά επίσκεψης μοντέλων	26
4.5 Εύρεση ιεραρχίας μοντέλων τρισδιάστατης κίνησης	27
4.6 Μεταφορά κατά τον οριζόντιο άξονα	29
4.7 Εστίαση (zooming)	31
4.8 Περιστροφή γύρω από τον οπτικό άξονα	33
4.9 Πλήρης μεταφορική κίνηση	34
4.10 Πλήρης περιστροφή	35
4.11 Πλήρης τρισδιάστατη κίνηση	36
4.12 Πειραματικά αποτελέσματα - συμπεράσματα	36
5. ΕΥΡΕΣΗ ΣΤΡΩΜΑΤΩΝ ΒΑΘΟΥΣ	45
5.1 Εισαγωγή	45
5.2 Καθορισμός του προβλήματος	45
5.3 Μεθοδολογία για τη λύση του προβλήματος	46
5.4 Αλγόριθμος τμηματοποίησης	48

5.4.1 Στάδιο 1 ^ο : Ομαδοποίηση με βάση κίνηση και φωτεινότητα	50
5.4.1.1 Προσδιορισμός ετικετών λόγω οπτικής ροής	50
5.4.1.2 Προσδιορισμός ετικετών λόγω φωτεινότητας	52
5.4.2 Στάδιο 2 ^ο : Εφαρμογή HCF αλγορίθμου	53
5.4.3 Στάδιο 3 ^ο : συνένωση περιοχών	53
5.5 Πειραματικά αποτελέσματα - συμπεράσματα	56
6. ΠΑΡΑΡΤΗΜΑ Α - HCF	60
6.1 Εισαγωγή	60
6.2 Ορισμός ενός πεδίου Markov	60
6.3 HCF και πεδία Markov	62
6.4 HCF (Highest Confidence First)	64
7. ΠΑΡΑΡΤΗΜΑ Β	68
7.1 Εκτίμηση των παραμέτρων μιας μίξης κατανομών	68
7.2 Κβαντιστής Lloyd-Max	69
8. ΒΙΒΛΙΟΓΡΑΦΙΑ	71

Κατάλογος σχημάτων

σχήμα 2.1: τα τρισδιάστατα διανύσματα μετατόπισης U_A και U_B των σημείων του χώρου A και B έχουν την ίδια προβολή u πάνω στο επίπεδο της εικόνας	5
σχήμα 2.2: μοντέλο κάμερας	7
σχήμα 3.1: Επαναλαμβανόμενη σταθμισμένη μέση τιμή	17
σχήμα 3.2: συναρτήσεις επίδρασης (αριστερά) του Huber και (δεξιά) του Tukey	19
σχήμα 3.3: AM-estimator	22
σχήμα 4.1: Ιεραρχία μοντέλων κίνησης	29
σχήμα 4.2	31
σχήμα 4.3	32
σχήμα 4.4	33
σχήμα 4.5: Frames 5 και 6 από την ακολουθία Flower Garden	37
σχήμα 4.6: (αριστερά) οπτική ροή για Flower Garden, (δεξιά) γωνίες απόκλισης για την ακολουθία Flower Garden με χρήση AM-estimator	37
σχήμα 4.7: frames 200, 201 από την ακολουθία Coastguard	38
σχήμα 4.8: (αριστερά) πεδίο ταχυτήτων για την ακολουθία Coastguard και (δεξιά) οι αντίστοιχοι inliers ως προς τη κίνηση της κάμερας.	38
σχήμα 4.9: τιμές γωνίας απόκλισης για την ακολουθία Coastguard με χρήση (αριστερά) M-estimator και (δεξιά) Least-squares.	39
σχήμα 4.10: frames 9 και 10 της ακολουθίας Kiel	39
σχήμα 4.11: (αριστερά) αραιό πεδίο οπτικής ροής για την ακολουθία Kiel, (δεξιά) τιμές γωνίας απόκλισης για ακολουθία Kiel με χρήση M-estimator.	40
σχήμα 4.12: frames 9 και 10 της ακολουθίας Nasa	40
σχήμα 4.13: (αριστερά) πεδίο ταχυτήτων για ακολουθία Nasa, (δεξιά) τιμές γωνίας απόκλισης για τα frames 0 - 20 της ακολουθίας Nasa με χρήση M-estimator.	41
σχήμα 4.14: (αριστερά) frame 9 της ακολουθίας Yosemite και (δεξιά) αντίστοιχη οπτική ροή	41
σχήμα 4.15: (αριστερά) inliers ως προς τη κίνηση της κάμερας για το frame 9 της ακολουθίας Yosemite και (δεξιά) γραφική παράσταση της συνάρτησης σφάλματος του AM-estimator για τα πρώτα 20 frames της ακολουθίας Yosemite.	42
σχήμα 4.16: (αριστερά) frame 130 από την ακολουθία Stefan και (δεξιά) οι αντίστοιχοι inliers ως προς τη κίνηση της κάμερας.	43
σχήμα 5.1: αλγόριθμος τμηματοποίησης	49
σχήμα 5.2: συνένωση περιοχών	55
σχήμα 5.3: Frames 5 και 6 από την ακολουθία Flower Garden και αντίστοιχο πεδίο ταχυτήτων	57
σχήμα 5.4: (αριστερά) ιστόγραμμα της u -συνιστώσας της οπτικής ροής και (δεξιά) προσέγγιση του ιστογράμματος αυτού με κατανομές Γκάους.	58
σχήμα 5.5: (αριστερά) ιστόγραμμα της φωτεινής έντασης και (δεξιά) κβαντισμός της εικόνας 5 του Flower Garden ως προς τη φωτεινότητα με επίπεδα απόφασης $\{52, 100, 169\}$.	58
σχήμα 5.6: (αριστερά) αποτέλεσμα του HCF αλγορίθμου, (δεξιά) τελικό αποτέλεσμα τμηματοποίησης.	59
σχήμα 5.7: Χρόνος εκτέλεσης HCF αλγορίθμου.	59
σχήμα 6.1: γειτονιά 8 σημείων και αντίστοιχες παρέες	61
σχήμα 6.2: Αλγόριθμος HCF	67

*Κ ε φ α λ α ι ο 1***1. ΓΕΝΙΚΗ ΕΙΣΑΓΩΓΗ**

Το θέμα της ανάλυσης της τρισδιάστατης κίνησης έχει προσελκύσει αρκετό ενδιαφέρον τα τελευταία χρόνια και συνιστά έναν διαρκώς εξελισσόμενο τομέα έρευνας. Αποτελεί δε αντικείμενο μελέτης για πολλές εφαρμογές που έχουν να κάνουν με το τομέα της ανάλυσης ακολουθιών εικόνων ([5], [9]) και με τον τομέα της υπολογιστικής στερεοσκοπικής δυναμικής όρασης ([6], [7], [8]). Επίσης το ίδιο πρόβλημα εμφανίζεται και στην κωδικοποίηση και μετάδοση ακολουθιών εικόνων ([1], [2], [3], [4]). Σε μια πρώτη φάση, η ανάλυση της τρισδιάστατης κίνησης απαιτεί την ανίχνευση κινήσεων μέσα σε μια ακολουθία εικόνων. Σε μια πιο προχωρημένη φάση απαιτεί την εξαγωγή των παραμέτρων της τρισδιάστατης κίνησης καθώς και την εξαγωγή των σχημάτων των κινούμενων αντικειμένων, όπου αυτό είναι δυνατό.

Κατά τη παρούσα εργασία θα ασχοληθούμε με 2 προβλήματα. Το πρώτο από αυτά αφορά τον υπολογισμό της τρισδιάστατης κίνησης της κάμερας. Σε μια ακολουθία εικόνων, είναι αρκετά συνηθισμένη η περίπτωση όπου η κάμερα δεν παραμένει στατική αλλά αντιθέτως κινείται μέσα στον τρισδιάστατο χώρο. Για την περαιτέρω ανάλυση της συγκεκριμένης ακολουθίας εικόνων είναι επομένως απαραίτητο να μπορέσουμε να υπολογίσουμε τη κίνηση αυτή της κάμερας. Επειδή η κίνηση της κάμερας μπορεί να περιέχει το πολύ έξι βαθμούς ελευθερίας, θα πρέπει στη πιο γενική περίπτωση να υπολογίσουμε μέχρι και έξι παραμέτρους προκειμένου να προσδιορίσουμε πλήρως την τρισδιάστατη αυτή κίνηση. Τρεις από τις παραμέτρους αυτές θα οφείλονται στο μεταφορικό τμήμα της κίνησης της κάμερας, ενώ οι υπόλοιπες τρεις θα οφείλονται στην περιστροφή της κάμερας.

Στην πραγματικότητα λόγω της προβολής των τρισδιάστατων αντικειμένων στο επίπεδο της εικόνας, ένα μέρος από την υπάρχουσα πληροφορία χάνεται. Αυτό έχει σαν αποτέλεσμα να μην μπορούμε να προσδιορίσουμε επακριβώς τις συνιστώσες του διάνυσματος της μεταφορικής κίνησης αλλά μόνο τους λόγους των συνιστωσών αυτών μεταξύ τους. Ευτυχώς, αυτό το πρόβλημα δεν ισχύει για το διάνυσμα που περιγράφει την περιστροφή της κάμερας, το οποίο και μπορούμε να προσδιορίσουμε πλήρως.

Για τον υπολογισμό της κίνησης της κάμερας υπάρχουν δύο γενικές μεθοδολογίες. Σύμφωνα με την πρώτη από τις προσεγγίσεις αυτές, οι παράμετροι της τρισδιάστατης κίνησης εκτιμώνται μέσω υπολογισμών οι οποίοι βασίζονται σε ένα ήδη υπάρχον πεδίο διδιάστατης οπτικής ροής ([10], [11]). Σύμφωνα με τη δεύτερη από τις υπάρχουσες προσεγγίσεις, οι παράμετροι της

τριδιάστατης κίνησης βρίσκονται απευθείας κάνοντας χρήση χωρικών και χρονικών παραγώγων της φωτεινότητας ([12], [13]).

Η μέθοδος που υιοθετείται στην συγκεκριμένη εργασία ακολουθεί την πρώτη προσέγγιση. Πιο συγκεκριμένα, ο αλγόριθμος υπολογισμού της τριδιάστατης κίνησης αποτελείται από δύο στάδια. Κατά το πρώτο στάδιο λαμβάνει χώρα ο υπολογισμός ενός διδιάστατου πεδίου οπτικής ροής. Κατά το δεύτερο στάδιο προσδιορίζονται οι παράμετροι της τριδιάστατης κίνησης. Αυτό επιτυγχάνεται μέσω εξισώσεων οι οποίες συνδέουν τις παραμέτρους αυτές με την οπτική ροή που έχει ήδη υπολογισθεί κατά το πρώτο στάδιο. Για τον υπολογισμό του πεδίου οπτικής ροής έχουν παρουσιαστεί ήδη αρκετοί αλγόριθμοι ([14]) και το θέμα αυτό έχει ήδη αποτελέσει σημείο εκτεταμένης έρευνας. Έτσι εμείς θα ασχοληθούμε στην παρούσα εργασία κυρίως με το δεύτερο στάδιο.

Το βασικό πρόβλημα που παρουσιάζεται κατά το στάδιο αυτό οφείλεται στο γεγονός ότι το πεδίο οπτικής ροής περιέχει ένα ποσοστό από εσφαλμένα δεδομένα ([15]). Τα δεδομένα αυτά ονομάζονται **outliers**. Τα εσφαλμένα δεδομένα προκύπτουν κατά την διάρκεια του υπολογισμού της οπτικής ροής, η οποία στην περίπτωση της μίας κάμερας επιτυγχάνεται συνήθως μέσω μιας διαδικασίας αντιστοίχισης σημείων από εικόνα σε εικόνα. Η διαδικασία αυτή είναι δύσκολη και αρκετά ευαίσθητη στο θόρυβο με αποτέλεσμα η πιθανότητα να υπάρχουν λάθη να είναι πολύ μεγάλη. Στην πραγματικότητα η κατάσταση δυσκολεύει ακόμη περισσότερο όταν μέσα στην ακολουθία εικόνων υπάρχουν και ανεξάρτητα κινούμενα αντικείμενα. Στην περίπτωση αυτή όλα τα σημεία τα οποία είναι σύμφωνα με την ανεξάρτητη κίνηση, μπορεί να θεωρηθούν ως outliers ως προς την κίνηση της κάμερας. Έτσι, το ποσοστό των outliers αυξάνεται ακόμη περισσότερο κάτι το οποίο έχει προφανώς αρνητικές επιπτώσεις στον σωστό υπολογισμό της κίνησης της κάμερας.

Η προσέγγιση που ακολουθείται κατά τη παρούσα εργασία για να αντιμετωπιστεί το πρόβλημα αυτό, συνίσταται στη χρησιμοποίηση μιας **ιεραρχίας** από τριδιάστατα μοντέλα κίνησης. Στη συνέχεια ορίζεται ένα κριτήριο σύμφωνα με το οποίο θα απορρίπτεται ή θα γίνεται δεκτό ένα μοντέλο κίνησης. Αρχικά, δοκιμάζονται τα απλούστερα μοντέλα κίνησης, ενώ καθώς προχωρούμε όλο και πιο πολύπλοκα τριδιάστατα μοντέλα εξετάζονται. Κατ' αυτό το τρόπο επιδιώκεται να βρεθεί το απλούστερο δυνατό μοντέλο το οποίο είναι σε θέση να περιγράψει επαρκώς τη κίνηση της κάμερας. Επιπλέον, κατά τη διάρκεια του ελέγχου του κάθε μοντέλου κίνησης, γίνεται χρήση και ενός robust αλγορίθμου για να μειωθεί περαιτέρω η επίδραση των outliers στο τελικό αποτέλεσμα. Στην περίπτωση μας, χρησιμοποιείται ένας M-estimator αλγόριθμος.

Το δεύτερο πρόβλημα που αντιμετωπίζεται στην παρούσα εργασία είναι αυτό της εύρεσης των στρωμάτων βάθους μέσα σε μια εικόνα. Για τη λύση του προβλήματος αυτού θα πρέπει πρώτα να έχει υπολογισθεί η κίνηση της κάμερας. Στη συνέχεια το πρόβλημα της εύρεσης των στρωμάτων βάθους ανάγεται στην περίπτωση όπου η κίνηση της κάμερας είναι παράλληλη προς το επίπεδο της εικόνας. Στην περίπτωση αυτή η εύρεση των στρωμάτων βάθους συνίσταται ουσιαστικά στην διαμέριση του πεδίου των δισδιάστατων μετατοπίσεων σε περιοχές όπου τα διανύσματα της οπτικής ροής παραμένουν σταθερά.

Για το σκοπό αυτό χρησιμοποιείται ένα ολικό κριτήριο τμηματοποίησης το οποίο βασίζεται σε ένα πεδίο Markov. Για την ελαχιστοποίηση της ενέργειας που προκύπτει χρησιμοποιείται ο επαναληπτικός αλγόριθμος χαλάρωσης **Highest Confidence First (HCF)**. Η είσοδος στον αλγόριθμο αυτό βασίζεται τόσο στην οπτική ροή, όσο και στη συνάρτηση της φωτεινότητας. Η χρήση της φωτεινότητας κρίθηκε αναγκαία λόγω του γεγονότος ότι οι υπολογισμοί της οπτικής ροής είναι αρκετά επιρρεπείς στα λάθη, ιδιαίτερα σε περιοχές που εμφανίζονται για πρώτη φορά μέσα στην ακολουθία εικόνων. Αυτό έχει σαν αποτέλεσμα να παρουσιάζονται σφάλματα κατά την τμηματοποίηση, κυρίως σε σχέση με τα περιγράμματα των αντικειμένων. Για τον ακριβή επομένως προσδιορισμό των περιγραμμάτων αυτών, λαμβάνεται υπόψη και πληροφορία που προέρχεται από την φωτεινότητα.

Η εργασία αυτή ακολουθεί την εξής δομή. Στο κεφάλαιο 2 παρουσιάζονται και αναλύονται οι σχέσεις που συνδέουν την οπτική ροή με τις παραμέτρους της τρισδιάστατης κίνησης. Στο κεφάλαιο 3 υπάρχει μια συνοπτική περιγραφή και επισκόπηση των M-estimator τεχνικών. Στο κεφάλαιο 4 περιέχεται η περιγραφή της ιεραρχίας των τρισδιάστατων μοντέλων κίνησης που θα χρησιμοποιηθεί για τον υπολογισμό της κίνησης της κάμερας, ενώ παράλληλα παρουσιάζεται και κάθε μοντέλο κίνησης ξεχωριστά. Στο κεφάλαιο 5 περιγράφεται ο συνολικός αλγόριθμος τμηματοποίησης που χρησιμοποιείται για την εύρεση των στρωμάτων βάθους μέσα σε μια ακολουθία εικόνων. Τέλος στο παράρτημα Α γίνεται μια σύντομη περιγραφή του HCF αλγορίθμου.

Κεφάλαιο 2

2. ΣΧΕΣΕΙΣ ΟΠΤΙΚΗΣ ΡΟΗΣ ΚΑΙ ΤΡΙΣΔΙΑΣΤΑΤΗΣ ΚΙΝΗΣΗΣ

2.1 Εισαγωγή

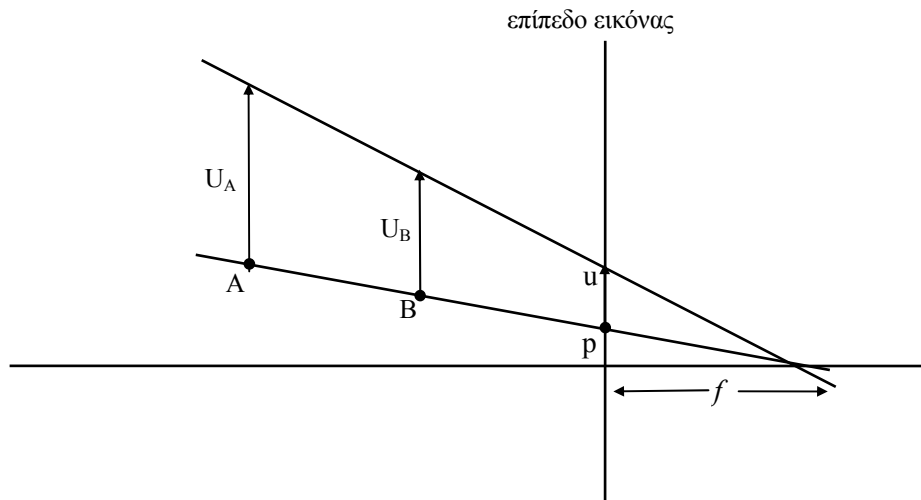
Το κεφάλαιο αυτό περιλαμβάνει κάποιες βασικές γνώσεις σχετικά με το πρόβλημα υπολογισμού της τρισδιάστατης κίνησης. Επιπλέον στην ενότητα αυτή γίνεται μια συνοπτική περιγραφή της μεθοδολογίας που θα ακολουθηθεί για την εκτίμηση της κίνησης της κάμερας στο χώρο. Έτσι, το κεφάλαιο ξεκινά με τον ορισμό και την διατύπωση του προβλήματος για να γίνει σαφές το τι ακριβώς επιδιώκεται να επιτευχθεί. Στη συνέχεια ακολουθεί μια σύντομη περιγραφή του μοντέλου της κάμερας που θα χρησιμοποιηθεί, το οποίο είναι και γνωστό ως το **μάτι της βελόνας (pinhole)**, όπου εξάγονται οι σχέσεις που συνδέουν τις συντεταγμένες των σημείων του τρισδιάστατου χώρου και τις συντεταγμένες των προοπτικών προβολών τους στην εικόνα. Ακολούθως προσδιορίζεται τι ακριβώς εννοούμε με τον όρο **τρειςδιάστατο μοντέλο κίνησης** ενώ παράλληλα περιγράφονται εν συντομία ορισμένα από τα πιο ευρέως χρησιμοποιούμενα μοντέλα κίνησης. Τέλος, εξάγονται οι σχέσεις που συνδέουν την οπτική ροή με τις παραμέτρους ενός τρισδιάστατου μοντέλου κίνησης. Στις σχέσεις αυτές θα βασιστούμε αργότερα για την εξαγωγή των τρισδιάστατων χαρακτηριστικών της κίνησης.

2.2 Ορισμός και διατύπωση του προβλήματος

Μια αρκετά συνηθισμένη περίπτωση στα προβλήματα που έχουν να κάνουν με την ανάλυση ακολουθιών εικόνων είναι αυτή της ύπαρξης δυναμικού παρατηρητή. Με άλλα λόγια, ο παρατηρητής, δηλαδή η κάμερα, δεν είναι στατικός αλλά κινείται μέσα στον τρισδιάστατο χώρο. Το πρόβλημα της εκτίμησης και του προσδιορισμού των παραμέτρων της τρισδιάστατης αυτής κίνησης αντιμετωπίζεται στην παρούσα εργασία. Η κίνηση του παρατηρητή ή της κάμερας είναι γνωστή και ως **ιδία κίνηση (egomotion)**.

Το πρόβλημα αυτό της εκτίμησης της ίδιας κίνησης είναι από μόνο του ένα αρκετά δύσκολο πρόβλημα ([15], [16], [17]). Η δυσκολία έγκειται στο γεγονός ότι αρκετή από την υπάρχουσα πληροφορία χάνεται λόγω της προβολής των σημείων του χώρου στο επίπεδο της εικόνας. Ζητάμε δηλαδή πληροφορία για τρισδιάστατη κίνηση ενώ διαθέτουμε μόνο δισδιάστατα δεδομένα.

Επιπλέον, το πρόβλημα της εκτίμησης της τρισδιάστατης κίνησης της κάμερας είναι βασικό και αποτελεί το απαραίτητο πρώτο στάδιο για την επίλυση άλλων προβλημάτων που έχουν να κάνουν με την ανάλυση ακολουθιών εικόνων. Για παράδειγμα για να εκτιμηθεί η κίνηση των ανεξάρτητα κινούμενων αντικειμένων που υπάρχουν μέσα σε μια ακολουθία εικόνων, η συνήθης πρακτική είναι να προσδιορίζεται πρώτα η κίνηση της κάμερας και αφού αυτή αφαιρεθεί από την ακολουθία εισόδου, στη συνέχεια να αντιμετωπίζεται το αρχικό πρόβλημα. Το ίδιο ισχύει και για πολλά άλλα προβλήματα όπως για παράδειγμα αυτό της ανίχνευσης των μεταβολών ([18], [19]), της εύρεσης των στρωμάτων βάθους κ.τ.λ.



σχήμα 2.1: τα τρισδιάστατα διανύσματα μετατόπισης U_A και U_B των σημείων του χώρου A και B έχουν την ίδια προβολή u πάνω στο επίπεδο της εικόνας

Στη πιο γενική περίπτωση η κίνηση της κάμερας μπορεί να περιέχει μέχρι και 6 βαθμούς ελευθερίας. Θεωρούμε δηλαδή ότι η κάμερα μπορεί ταυτόχρονα και να κινείται μεταφορικά αλλά και να εκτελεί περιστροφή. Επομένως για να προσδιορίσουμε πλήρως τη κίνηση της θα πρέπει υποχρεωτικά να υπολογίσουμε 2 διανύσματα του τρισδιάστατου χώρου: το διάνυσμα που περιγράφει τη μεταφορική κίνηση και το διάνυσμα που περιγράφει τη περιστροφική κίνηση της κάμερας.

Στην πραγματικότητα λόγω της προβολής των τρισδιάστατων αντικειμένων στο επίπεδο της εικόνας, προκύπτει το γνωστό και ως 'scale factor ambiguity' πρόβλημα (δες σχήμα 2.1). Δηλαδή, το διάνυσμα της μεταφορικής κίνησης δεν μπορεί να προσδιορισθεί ακριβώς αλλά μόνο μέχρι ένα παράγοντα κλιμάκωσης. Άρα αν T είναι το πραγματικό διάνυσμα μεταφορικής κίνησης και T_1 είναι το διάνυσμα που υπολογίζουμε, τότε θα ισχύει η σχέση:

$$T_1 = \lambda T \quad \lambda \in R$$

Ευτυχώς, το πρόβλημα αυτό δεν ισχύει για το διάνυσμα της περιστροφικής κίνησης το οποίο και μπορεί να προσδιορισθεί πλήρως. Επομένως για να υπολογίσουμε τη κίνηση της κάμερας, ο

στόχος μας θα είναι να προσδιορίσουμε, από τη μια τους λόγους των δύο συνιστωσών του διανύσματος της μετατόπισης ως προς τη τρίτη συνιστώσα, και από την άλλη τις 3 συνιστώσες του διανύσματος περιστροφής. Δηλαδή στη πιο γενική περίπτωση ψάχνουμε για 5 παραμέτρους. Η είσοδος μας θα αποτελείται από ένα ζευγάρι διαδοχικών εικόνων από μια συγκεκριμένη ακολουθία εικόνων.

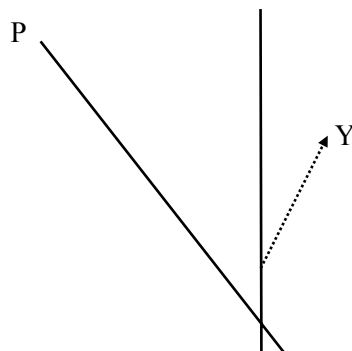
2.3 Πεδίο οπτικής ροής

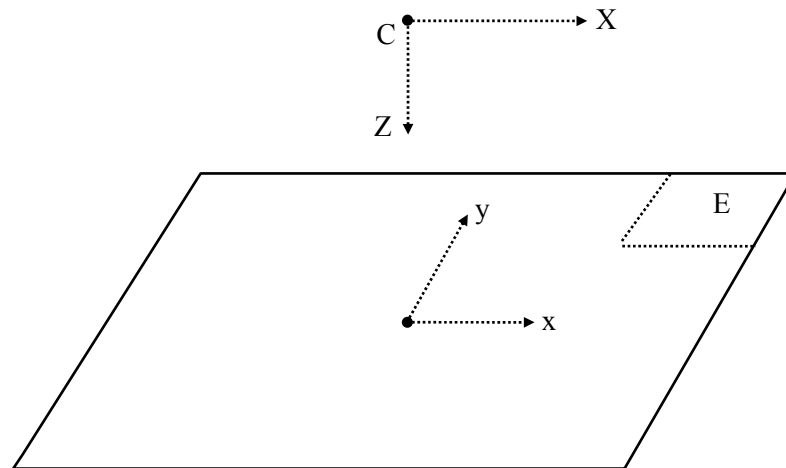
Η προσέγγιση που θα ακολουθηθεί από τη παρούσα εργασία για τον προσδιορισμό της κίνησης της κάμερας περιλαμβάνει δύο φάσεις. Κατά τη πρώτη φάση, θα υπολογιστεί ένα **πεδίο οπτικής ροής** με βάση το ζευγάρι εικόνων της εισόδου. Κατά τη δεύτερη φάση, θα γίνει η εξαγωγή των παραμέτρων της τρισδιάστατης κίνησης βασιζόμενοι κατά κύριο λόγο σε σχέσεις που συνδέουν το πεδίο οπτικής ροής με τις παραμέτρους αυτές. Το πεδίο οπτικής ροής αποτελεί στην ουσία την προβολή των τρισδιάστατων ταχυτήτων των σημείων των διαφόρων επιφανειών πάνω στο επίπεδο της εικόνας, και δείχνει την μετατόπιση των αντικειμένων από εικόνα σε εικόνα. Γι' αυτό το λόγο το πεδίο οπτικής ροής είναι γνωστό και ως **πεδίο ταχυτήτων** ή **πεδίο μετατοπίσεων**.

Ο σωστός και αξιόπιστος υπολογισμός του πεδίου ταχυτήτων έχει αποτελέσει αντικείμενο εκτεταμένης έρευνας και ήδη έχουν αναπτυχθεί μια πληθώρα αλγορίθμων προς το σκοπό αυτό ([20], [21], [22], [23], [24]). Μια σύγκριση για ορισμένους από τους πιο ευρεία διαδεδομένους αλγόριθμους μπορεί να βρει κανείς στο άρθρο των Baron, Fleet και Beauchemin [14]. Έτσι κατά τη παρούσα εργασία το ενδιαφέρον δεν θα επικεντρωθεί στον υπολογισμό της οπτικής ροής. Αντίθετα αυτή είτε θα θεωρείται δεδομένη είτε θα υπολογίζεται με βάση κάποια από τις ήδη υπάρχουσες τεχνικές για το σκοπό αυτό. Θα πρέπει επίσης να σημειωθεί ότι η παρούσα προσέγγιση δεν απαιτεί το πεδίο οπτικής ροής να είναι πυκνό. Αυτό είναι ιδιαίτερα σημαντικό δεδομένου ότι σε αρκετές περιπτώσεις (ιδιαίτερα σε πραγματικές ακολουθίες εικόνων) δεν διαθέτουμε αξιόπιστη οπτική ροή για όλα τα σημεία της εικόνας αλλά μόνο για ένα μικρό τμήμα της εικόνας.

2.4 Μοντέλο κάμερας

Κάθε κάμερα μοντελοποιείται με την οπτική της εστία C και το ορθογώνιο παραλληλόγραμμο της εικόνας E . Αυτό το μοντέλο κάμερας ονομάζεται 'το μάτι της βελόνας' (*pinhole*).





σχήμα 2.2: μοντέλο κάμερας

Ο άξονας ο οποίος διέρχεται απ' την οπτική εστία και το κέντρο της εικόνας και επιπλέον είναι κάθετος στο επίπεδο της, ονομάζεται **οπτικός άξονας**, ενώ η απόσταση της οπτικής εστίας απ' το επίπεδο της εικόνας ονομάζεται **εστιακή απόσταση**. Ορίζοντας ένα τρισδιάστατο σύστημα συντεταγμένων $\langle X, Y, Z \rangle$ όπως φαίνεται στο σχήμα 2.2 οι σχέσεις που συνδέουν τις συντεταγμένες (X, Y, Z) του σημείου P και της **προοπτικής προβολής** του (x, y) στο επίπεδο της εικόνας είναι:

$$x = f \frac{X}{Z} \quad y = f \frac{Y}{Z} \quad (2.1)$$

Στο σχήμα 2.2 ο άξονας Z ταυτίζεται με τον οπτικό άξονα και η αρχή των αξόνων με την οπτική εστία. Το f παριστάνει την εστιακή απόσταση.

2.5 Τρισδιάστατα μοντέλα κίνησης

Μια από τις κεντρικές έννοιες που θα χρησιμοποιηθούν κατά τη παρούσα εργασία είναι και αυτή του **τρειςδιάστατου μοντέλου κίνησης**. Με τον όρο αυτό θα εννοούμε ένα σύνολο παραμέτρων οι οποίες προσδιορίζουν πλήρως τη κίνηση ενός αντικειμένου (στη περίπτωση μας της κάμερας) μέσα στο τρισδιάστατο χώρο κάθε χρονική στιγμή. Το πλήθος των παραμέτρων αυτών καθορίζει και την πολυπλοκότητα του μοντέλου και κατ' επέκταση την πολυπλοκότητα της κίνησης της κάμερας. Όσο μεγαλύτερος είναι ο αριθμός αυτός, τόσο πιο περίπλοκο είναι το συγκεκριμένο μοντέλο κίνησης.

Για να οριστεί πλήρως ένα τρισδιάστατο μοντέλο κίνησης απαιτείται να καθοριστούν τα στιγμιαία διανύσματα της μετατόπισης και της περιστροφής μέσα στο τρισδιάστατο χώρο. Αν

ονομάσουμε τα διανύσματα αυτά T και R αντίστοιχα, τότε το διάνυσμα V της τρισδιάστατης κίνησης για το σημείο με συντεταγμένες (X, Y, Z) στο χώρο θα δίνεται από την εξής σχέση:

$$V^{\rho} = T^{\rho} + R^{\rho} \wedge \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2.1)$$

Παρακάτω θα αναφερθούν κάποια από τα πιο ευρέως χρησιμοποιούμενα τρισδιάστατα μοντέλα κίνησης, ορισμένα εκ των οποίων θα χρησιμοποιηθούν και κατά την παρούσα εργασία. Τα μοντέλα αυτά περιγράφουν διάφορα είδη κινήσεων οι οποίες διαφέρουν σε πολυπλοκότητα.

- **Εστίαση (zooming):** Στην περίπτωση αυτή η κάμερα εκτελεί μεταφορική κίνηση κάθετα στο επίπεδο της εικόνας. Επιπλέον, η κάμερα δεν περιστρέφεται. Άρα σύμφωνα με το συγκεκριμένο μοντέλο τα διανύσματα της μετατόπισης και της περιστροφής θα είναι:

$$T^{\rho} = \begin{bmatrix} 0 \\ 0 \\ T_z \end{bmatrix} \quad R^{\rho} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Επομένως το μοντέλο κίνησης αυτό εξαρτάται και καθορίζεται πλήρως από μία παράμετρο, το T_z . Το διάνυσμα της κίνησης θα δίνεται από τη σχέση:

$$V^{\rho} = \begin{bmatrix} 0 \\ 0 \\ T_z \end{bmatrix}$$

- **Μεταφορά κατά τον οριζόντιο άξονα:** Σύμφωνα με το μοντέλο αυτό η κάμερα εκτελεί μεταφορική κίνηση παράλληλα στο επίπεδο της εικόνας και πάνω στον άξονα X . Όπως και στην προηγούμενη περίπτωση, έτσι και εδώ, δεν έχουμε περιστροφή της κάμερας. Άρα τα διανύσματα της μετατόπισης και της περιστροφής θα δίνονται από τις σχέσεις:

$$T^{\rho} = \begin{bmatrix} T_x \\ 0 \\ 0 \end{bmatrix} \quad R^{\rho} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Η πολυπλοκότητα του μοντέλου περιορίζεται από μια παράμετρο. Το διάνυσμα της κίνησης θα δίνεται από τη σχέση:

$$V^{\rho} = \begin{bmatrix} T_x \\ 0 \\ 0 \end{bmatrix}$$

- **Περιστροφή γύρω από τον οπτικό άξονα:** Στην περίπτωση αυτή η κάμερα δεν εκτελεί μεταφορική κίνηση. Αντίθετα, εκτελεί μια απλή περιστροφή γύρω από τον οπτικό άξονα.

Επομένως τα διανύσματα της μετατόπισης και της περιστροφής θα δίνονται από τις σχέσεις:

$$\overset{\rho}{T} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \overset{\rho}{R} = \begin{bmatrix} 0 \\ 0 \\ R_z \end{bmatrix}$$

Το μοντέλο εξαρτάται μόνο από μια παράμετρο, τη συνιστώσα του διανύσματος περιστροφής γύρω από τον Z άξονα. Το διάνυσμα της κίνησης θα δίνεται από τη σχέση:

$$\overset{\rho}{V} = \begin{bmatrix} -R_z Y \\ R_z X \\ 0 \end{bmatrix}$$

- **Πλήρης μεταφορά (Full translation):** Στην περίπτωση αυτή ο μοναδικός περιορισμός που επιβάλλεται στην κίνηση της κάμερας είναι ότι δεν μπορεί να περιστραφεί. Αντίθετα δεν κάνουμε καμία υπόθεση για τη μορφή του διανύσματος της μετατόπισης. Τα διανύσματα της μετατόπισης και της περιστροφής θα είναι:

$$\overset{\rho}{T} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \quad \overset{\rho}{R} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Το συγκεκριμένο μοντέλο εξαρτάται από 3 παραμέτρους και είναι σε θέση να περιγράψει μεγαλύτερο αριθμό κινήσεων σε σχέση με τα προηγούμενα τρισδιάστατα μοντέλα κίνησης. Το διάνυσμα της κίνησης θα δίνεται από τη σχέση:

$$\overset{\rho}{V} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

- **Πλήρης περιστροφή (Full rotation):** Σύμφωνα με το μοντέλο αυτό η κάμερα δεν μπορεί να εκτελεί μεταφορική κίνηση. Αντίθετα, η περιστροφική κίνηση μπορεί να περιέχει μέχρι και 3 βαθμούς ελευθερίας. Τα διανύσματα της μετατόπισης και της περιστροφής θα δίνονται από τις σχέσεις:

$$\overset{\rho}{T} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \overset{\rho}{R} = \begin{bmatrix} R_x \\ R_y \\ R_z \end{bmatrix}$$

Το διάνυσμα που περιγράφει τη τρισδιάστατη κίνηση δίνεται από τη σχέση:

$$\mathcal{V}^{\rho} = \begin{bmatrix} R_y Z - R_z Y \\ R_z X - R_x Z \\ R_x Y - R_y X \end{bmatrix}$$

- **Πλήρης τρισδιάστατη κίνηση (Full 3-d motion):** Αυτό είναι το πιο γενικό μοντέλο κίνησης. Δεν επιβάλλει κανένα περιορισμό ούτε στη μεταφορική κίνηση αλλά ούτε και στην περιστροφική κίνηση της κάμερας. Επομένως, η τρισδιάστατη κίνηση στη περίπτωση αυτή μπορεί να περιέχει μέχρι και 6 βαθμούς ελευθερίας. Το διάνυσμα της κίνησης δίνεται από την γενική εξίσωση (2.1).

2.6 Σχέσεις οπτικής ροής και τρισδιάστατης κίνησης

Έχοντας αναφέρει κάποιες βασικές έννοιες σχετικά με την οπτική ροή αλλά και τα μοντέλα τρισδιάστατης κίνησης, θα εξάγουμε τις γεωμετρικές σχέσεις που συνδέουν τα δύο αυτά στοιχεία ([25]). Με τη βοήθεια των σχέσεων αυτών και έχοντας ήδη υπολογίσει ένα πεδίο οπτικής ροής θα προσπαθήσουμε αργότερα να εξάγουμε τα τρισδιάστατα χαρακτηριστικά της κίνησης.

Το μοντέλο κάμερας που θα χρησιμοποιήσουμε είναι αυτό που περιγράφεται στην παράγραφο 2.4. Άρα αν (X, Y, Z) είναι οι συντεταγμένες ενός σημείου P στο χώρο και (x, y) οι συντεταγμένες της προοπτικής προβολής, τότε θα ισχύουν οι σχέσεις:

$$x = f \frac{X}{Z} \quad y = f \frac{Y}{Z} \quad (2.2)$$

Χωρίς βλάβη της γενικότητας και για λόγους απλοποίησης των υπολογισμών, θα θεωρήσουμε ότι η εστιακή απόσταση f ισούται με τη μονάδα.

Στην περίπτωση αυτή, το στιγμιαίο τρισδιάστατο διάνυσμα κίνησης \mathcal{V} του σημείου P μπορεί να εκφραστεί μέσω των στιγμιαίων διανυσμάτων της μετατόπισης και της περιστροφής, T και R αντίστοιχα ([26]):

$$\mathcal{V}^{\rho} = \mathcal{T}^{\rho} + \mathcal{R}^{\rho} \wedge \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2.3)$$

Αναλύοντας το εξωτερικό γινόμενο, η παραπάνω σχέση γράφεται ισοδύναμα:

$$\mathcal{V}^{\rho} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} + \begin{bmatrix} R_y Z - R_z Y \\ R_z X - R_x Z \\ R_x Y - R_y X \end{bmatrix} \quad (2.4)$$

Με όμοιο τρόπο, μπορεί να εξαχθεί ότι οι συνιστώσες του διανύσματος της οπτικής ροής (u, v) που αντιστοιχεί στο σημείο (x, y) της εικόνας, δίνονται από τις παρακάτω αλγεβρικές εξισώσεις:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{V_X - xV_Z}{Z} \\ \frac{V_Y - yV_Z}{Z} \end{bmatrix} \quad (2.5)$$

Συνδυάζοντας τις σχέσεις (2.4) και (2.5) προκύπτει το παρακάτω ζευγάρι εξισώσεων:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{T_X - xT_Z}{Z} + R_Y - R_Z y - R_X xy + R_Y x^2 \\ \frac{T_Y - yT_Z}{Z} - R_X + R_Z x + R_Y xy - R_X y^2 \end{bmatrix} \quad (2.2)$$

Χρησιμοποιώντας τις παραπάνω εξισώσεις προκειμένου να απαλείψουμε τον άγνωστο Z (που αντιπροσωπεύει το βάθος), προκύπτει η παρακάτω γενική εξίσωση:

$$a_1 - a_2 x - a_3 y + a_4 x^2 + a_5 y^2 - a_6 xy + a_7 u - a_8 v = xv - yu \quad (2.3)$$

όπου:

$$\begin{cases} a = \frac{T_X}{T_Z} \\ b = \frac{T_Y}{T_Z} \end{cases} \quad (2.4)$$

και

$$\begin{cases} a_1 = R_X + bR_Y \\ a_2 = R_X + aR_Z \\ a_3 = R_Y + bR_Z \\ a_4 = bR_Y + R_Z \\ a_5 = aR_X + R_Z \\ a_6 = aR_Y + bR_X \\ a_7 = a \\ a_8 = b \end{cases} \quad (2.5)$$

Αν τώρα γνωρίζουμε τους συντελεστές a_i , τότε είναι φανερό ότι οι παράμετροι της τρισδιάστατης κίνησης μπορούν να προσδιορισθούν άμεσα επιλύοντας το σύστημα (2.5). Άρα το πρόβλημα υπολογισμού των παραμέτρων της τρισδιάστατης κίνησης έχει αναχθεί στο πρόβλημα υπολογισμού των 8 βασικών συντελεστών της εξίσωσης (2.3).

Θεωρητικά για να υπολογίσουμε τους συντελεστές αυτούς αρκεί να χρησιμοποιήσουμε την οπτική ροή για 8 τουλάχιστον σημεία. Στην περίπτωση αυτή με βάση την εξίσωση (2.3) θα σχηματίζαμε 8 γραμμικές σχέσεις (μία για κάθε σημείο), και θα είχαμε απλά να επιλύσουμε ένα γραμμικό σύστημα 8×8 . Ωστόσο στην πράξη τα 8 σημεία δεν αρκούν και θα πρέπει να χρησιμοποιηθούν πολύ περισσότερα. Το πρόβλημα έγκειται στο γεγονός ότι συνήθως το πεδίο οπτικής ροής που έχει υπολογισθεί, περιέχει ένα ποσοστό από εσφαλμένα δεδομένα ([27]). Τα δεδομένα αυτά (**outliers**) οφείλονται κυρίως στην ύπαρξη θορύβου καθώς και στο γεγονός ότι ο

υπολογισμός της οπτικής ροής αποτελεί από μόνος του μια αρκετά δύσκολη διαδικασία. Άρα αν προσπαθήσουμε να προσδιορίσουμε τους συντελεστές a_i επιλύοντας απλά ένα σύστημα 8×8 , τότε υπάρχει μεγάλη πιθανότητα κάποιο ή κάποια από τα 8 σημεία που θα επιλέξουμε να περιέχει σφάλματα, δηλαδή να ανήκει στο σύνολο των outliers. Το αποτέλεσμα θα είναι οι λύσεις που θα προκύψουν να διαφέρουν από τις σωστές. Το πρόβλημα γίνεται ακόμα πιο δύσκολο αν λάβουμε υπόψη μας και το γεγονός ότι είναι αρκετά συνηθισμένο φαινόμενο να υπάρχει και ένα ποσοστό ανεξάρτητης κίνησης μέσα στις πραγματικές ακολουθίες εικόνων. Στην περίπτωση αυτή, όλα τα σημεία που ανήκουν στην ανεξάρτητη κίνηση δεν ακολουθούν το μοντέλο κίνησης της κάμερας και άρα δεν ικανοποιούν την εξίσωση οπτικής ροής (2.3), αποτελούν δηλαδή και αυτά outliers.

αλγόριθμος τύπου **M-estimator**. Κύριος στόχος του αλγόριθμου αυτού θα είναι να μπορέσει να διακρίνει τα ‘καλά’ από τα ‘κακά’ δεδομένα (δηλαδή τους outliers από τα υπόλοιπα σημεία), έτσι ώστε οι υπολογισμοί για την εκτίμηση των παραμέτρων κίνησης να βασιστούν μόνο σε σημεία της πρώτης κατηγορίας. Αυτό θα επιτευχθεί με την ελαχιστοποίηση μιας συνάρτησης σφάλματος. Επιπλέον θα γίνει χρήση και μιας **ιεραρχίας τρισδιάστατων μοντέλων κίνησης**. Σκοπός της ιεραρχίας αυτής θα είναι η εύρεση του απλούστερου δυνατού μοντέλου το οποίο μπορεί να περιγράψει την κίνηση της κάμερας. Κατ’ αυτό το τρόπο θα μειωθεί στο ελάχιστο η επίδραση του θορύβου και των outliers στους τελικούς υπολογισμούς, δεδομένου ότι μειώνεται αντίστοιχα στο ελάχιστο και το πλήθος των προς εκτίμηση παραμέτρων. Η περιγραφή του M-estimator θα αποτελέσει το θέμα του επόμενου κεφαλαίου ενώ η παρουσίαση της ιεραρχίας των τρισδιάστατων μοντέλων θα γίνει στο κεφάλαιο 4.

Για την αντιμετώπιση των παραπάνω προβλημάτων θα συνδυαστούν δύο μέθοδοι κατά την παρούσα εργασία. Από τη μια θα χρησιμοποιηθεί ένας Robust αλγόριθμος υπολογισμού ([27], [28]). Στη συγκεκριμένη περίπτωση θα χρησιμοποιηθεί ένας

Κεφάλαιο 3

3. M-ESTIMATOR

3.1 Εισαγωγή

Βασικός σκοπός της ενότητας αυτής είναι η παρουσίαση του M-estimator αλγορίθμου. Ο M-estimator θα αποτελέσει κατά τη παρούσα εργασία βασικό εργαλείο για την εξαγωγή των τρισδιάστατων χαρακτηριστικών της κίνησης. Ανήκει δε στη γενική κατηγορία των **robust** μεθόδων υπολογισμού. Έτσι το κεφάλαιο αυτό ξεκινά ορίζοντας κάποιες βασικές έννοιες σχετικά με τις μεθόδους αυτές ενώ παράλληλα επεξηγείται σε τι ακριβώς εξυπηρετεί η χρήση τους. Στη συνέχεια παρουσιάζονται δύο αρκετά διαδεδομένοι M-estimators, αυτοί των Tukey και Huber καθώς και οι αντίστοιχες συναρτήσεις κόστους. Ακολουθεί η ανάλυση μιας βελτιωμένης παραλλαγής του M-estimator, ο AM-estimator. Ο αλγόριθμος αυτός προσπαθεί να εξαλείψει κάποια από τα μειονεκτήματα και τις αδυναμίες του M-estimator. Το κεφάλαιο θα κλείσει με ένα παράδειγμα εφαρμογής του M-estimator για την περίπτωση της εκτίμησης των παραμέτρων τρισδιάστατης κίνησης.

3.2 Robust μέθοδοι υπολογισμού

Οι robust μέθοδοι υπολογισμού αποτελούν εργαλεία για την επίλυση στατιστικών προβλημάτων στα οποία οι υποκείμενες υποθέσεις δεν είναι ακριβείς ([29]). Μια robust τεχνική δεν θα πρέπει να είναι ευαίσθητη σε αποκλίσεις από τις θεωρητικά ισχύουσες υποθέσεις αλλά αντίθετα θα πρέπει να είναι σε θέση να τις αντιμετωπίσει. Αυτό σημαίνει ότι ένας robust αλγόριθμος θα πρέπει να έχει καλή απόδοση ακόμα και στη περίπτωση που τα μετρηθέντα δεδομένα περιέχουν μέχρι και ένα ορισμένο ποσοστό σφαλμάτων. Θα πρέπει δηλαδή με κάποιο έξυπνο τρόπο να μπορέσει να διακρίνει τα 'καλά' δεδομένα (**inliers**) από εκείνα που περιέχουν λάθη (**outliers**) και να λάβει υπόψη του κατά τους υπολογισμούς μόνο τα πρώτα.

Δύο είναι οι βασικές έννοιες οι οποίες χρησιμοποιούνται για την αξιολόγηση ενός robust αλγορίθμου: η *σχετική αποτελεσματικότητα* (*relative efficiency*) και το *σημείο φθοράς* (*breakdown point*) που αυτός έχει.

Με τον όρο *σχετική αποτελεσματικότητα* εννοούμε το λόγο μεταξύ της μικρότερης δυνατής διασποράς που μπορεί να έχουν οι παράμετροι προς υπολογισμό και της πραγματικής διασποράς που προκύπτει κατά τη συγκεκριμένη robust μέθοδο υπολογισμού.

Με τον όρο σημείο φθοράς εννοούμε το μέγιστο ποσοστό από outliers (δηλαδή από εσφαλμένα δεδομένα) τα οποία ο αλγόριθμος μπορεί να ανεχθεί. Αν το ποσοστό των outliers είναι χαμηλότερο από το σημείο φθοράς τότε η λύση που θα δώσει ο robust αλγόριθμος θα πλησιάζει αρκετά την σωστή, ενώ αντίθετα όταν το ποσοστό των outliers υπερβεί το σημείο φθοράς τότε η σταθερότητα του αλγορίθμου αρχίζει και πέφτει σημαντικά με αποτέλεσμα οι λύσεις που προκύπτουν να μην είναι αξιόπιστες. Είναι φανερό ότι όσο μεγαλύτερο είναι το σημείο φθοράς ενός robust αλγορίθμου, τόσο πιο αποτελεσματικός είναι και ο αλγόριθμος αυτός. Για παράδειγμα το ασυμπτωτικό σημείο φθοράς του αλγορίθμου της μέσης τιμής είναι μηδέν γιατί η ύπαρξη ακόμα και ενός outlier, ο οποίος απέχει αρκετά από το υπάρχον μοντέλο, μπορεί να ‘διαφθείρει’ το τελικό αποτέλεσμα.

Έχουν προταθεί αρκετές robust μέθοδοι υπολογισμού: R-estimators, L-estimators, LMedS-estimator ([30], [31]) κ.τ.λ. Αυτές χρησιμοποιούνται ευρύτατα σε αρκετά προβλήματα ανάλυσης ακολουθιών εικόνων ([30], [34]). Μία από τις πιο διαδεδομένες τεχνικές αυτού του είδους είναι οι M-estimators. Ένα βασικό πλεονέκτημα τους, που έπαιξε σημαντικό ρόλο στην ευρεία χρήση τους, είναι το χαμηλό υπολογιστικό κόστος που διαθέτουν ενώ ταυτόχρονα μπορούν και δίνουν ικανοποιητικά αποτελέσματα. Επιπλέον παρουσιάζουν αξιοπιστία κατά τη παρουσία διαφόρων ειδών θορύβου ενώ διαθέτουν και υψηλό σημείο φθοράς. Οι M-estimators θα αποτελέσουν το κεντρικό θέμα του κεφαλαίου αυτού και στη συνέχεια θα περιγραφούν αναλυτικά.

3.3 Περιγραφή του M-estimator

Όπως έχει ήδη αναφερθεί ο M-estimator ([29]) αποτελεί μια αρκετά διαδεδομένη robust μέθοδο υπολογισμού. Το πρόβλημα του M-estimator στην πιο απλή του μορφή έχει ως εξής: Ας υποθέσουμε ότι διαθέτουμε ένα σύνολο από m δείγματα δεδομένων $r = \{r_i \mid i = 1, \dots, m\}$, τα οποία ικανοποιούν την σχέση:

$$r_i = f + n_i \quad (3.1)$$

Η παραπάνω σχέση αποτελεί το μοντέλο του προβλήματος μας. Το ζήτημα που καλούμαστε να αντιμετωπίσουμε είναι να υπολογίσουμε την άγνωστη παράμετρο f . Το πρόβλημα αυτό είναι γνωστό και ως πρόβλημα της εκτίμησης της μέσης τιμής. Ο όρος n_i στην εξίσωση παριστάνει το θόρυβο. Η κατανομή του θορύβου δεν θεωρείται γνωστή. Η μόνη υπόθεση που γίνεται είναι ότι τα n_1, \dots, n_m υπακούουν σε μια συμμετρική και ανεξάρτητη κατανομή. Ο M-estimator καλείται να αντιμετωπίσει τα προβλήματα που προέρχονται από την ύπαρξη του θορύβου. Όλα όσα θα αναφερθούν στη συνέχεια μπορεί πολύ εύκολα να αναχθούν και στην περίπτωση όπου το μοντέλο μας ταυτίζεται με το γενικό γραμμικό μοντέλο ([27]):

$$f_0 + f_1 r_{1i} + f_2 r_{2i} + \dots + f_p r_{pi} = n_i \quad (3.2)$$

θα ονομάζουμε **υπόλοιπο (residual)** ενός δείγματος, τη διαφορά μεταξύ της παρατήρησης r_i που αντιστοιχεί στο δείγμα αυτό, και της τιμής που προβλέπεται για το δείγμα από τον M-estimator. Δηλαδή αν f είναι μια εκτίμηση του M-estimator σε κάποια επανάληψη του αλγορίθμου, το υπόλοιπο για το i -οστό δείγμα θα είναι $r_i - f$. Ουσιαστικά το υπόλοιπο ενός δείγματος εκφράζει την απόκλιση του δείγματος από το μοντέλο μας. Με αυτό σαν δεδομένο, η βασική αρχή πάνω στην οποία στηρίζεται ο M-estimator είναι αρκετά απλή: προκειμένου να προσδιορισθούν οι outliers (δηλαδή τα ‘κακα’ δεδομένα), γίνεται χρήση της πληροφορίας που παρέχει το σύνολο των υπολοίπων (residual set). Ένα μεγάλο υπόλοιπο, υποδηλώνει ένα αντίστοιχα μεγάλο σφάλμα και κατά συνέπεια έναν πιθανό outlier. Άρα τα δείγματα με μεγάλα υπόλοιπα δεν θα πρέπει να ληφθούν σοβαρά υπόψη κατά τους υπολογισμούς και επομένως θα πρέπει να τους αποδοθεί μικρή σημασία.

Σύμφωνα με τον M-estimator, η λύση f^* στο παραπάνω πρόβλημα θα προκύψει ως η ελάχιστη τιμή της παρακάτω ολικής συνάρτησης λάθους:

$$f^* = \arg \min_f E(f) \quad (3.3)$$

όπου:

$$E(f) = \sum_i g(r_i - f) \quad (3.4)$$

Για την ελαχιστοποίηση της παραπάνω ποσότητας, αρκεί να επιλύσουμε την παρακάτω εξίσωση:

$$\sum_i J(r_i - f) = 0 \quad (3.5)$$

όπου:

$$J(n) = g'(n) \quad (3.6)$$

Αν θεωρήσουμε ότι η $g(n_i)$ είναι συνάρτηση του n_i^2 , τότε η συνάρτηση $J(n_i)$ παίρνει την εξής μορφή:

$$J(n_i) = 2n_i h(n_i) = 2(r_i - f)h(r_i - f) \quad (3.7)$$

όπου η συνάρτηση $h(n)$ είναι άρτια.

Στην περίπτωση αυτή, η λύση που θα δώσει ο M-estimator θα εκφράζεται με το παρακάτω άθροισμα με βάρη:

$$f^* = \frac{\sum_i h(r_i - f)r_i}{\sum_i h(r_i - f)} \quad (3.8)$$

Η εξίσωση αυτή αποτελεί τη βάση για να προχωρήσουμε στην επόμενη επανάληψη κατά τη διάρκεια του M-estimator. Κάθε φορά υπολογίζονται τα νέα σφάλματα με βάση τη τιμή f που

έχει υπολογιστεί στο προηγούμενο βήμα. Στη συνέχεια με βάση τα σφάλματα αυτά και χρησιμοποιώντας τη παραπάνω σχέση υπολογίζεται η νέα τιμή της f . Η διαδικασία αυτή συνεχίζεται μέχρι να έχουμε σύγκλιση του αλγορίθμου ή μέχρι να φτάσουμε σε ένα ορισμένο αριθμό επαναλήψεων. Κατά την εκκίνηση του αλγορίθμου θεωρούμε ότι όλα τα σημεία έχουν το ίδιο βάρος, οπότε η αρχική τιμή που παίρνει η f θα είναι:

$$f = \frac{1}{m} \sum_{i=1}^m r_i \quad (3.9)$$

Μια συνοπτική περιγραφή του M-estimator αλγορίθμου γραμμένη σε ψευδοκώδικα φαίνεται στο σχήμα 3.1.

Η συνάρτηση $g(x)$ ονομάζεται *συνάρτηση κόστους (error penalty function)* του M-estimator. Εξαρτάται δε από το υπόλοιπο ενός δείγματος. Όσο πιο μεγάλο είναι το υπόλοιπο (δηλαδή το σφάλμα) τόσο πιο μεγάλο είναι το τίμημα (κόστος) που πληρώνουμε. Αντίστοιχα η συνάρτηση $h(x)$ ονομάζεται *συνάρτηση επίδρασης (influence function)* ([35]). Ισχύει ότι:

$$h(x) = \frac{g'(x)}{2x} \quad (3.10)$$

Ο ρόλος της συνάρτησης επίδρασης είναι να μετρά τη σημασία που θα έχει ένα δείγμα στο τελικό αποτέλεσμα. Η τιμή της συνάρτησης αυτής καθορίζει το 'βάρος' ενός δείγματος και εξαρτάται από το υπόλοιπο που αντιστοιχεί στο δείγμα. Όσο πιο μεγάλο είναι το υπόλοιπο αυτό (δηλαδή το σφάλμα), τόσο πιο μικρό (κοντά στο μηδέν) θα είναι το βάρος που θα αποδοθεί στο δείγμα από τη συνάρτηση επίδρασης. Αντίθετα όταν το υπόλοιπο ενός δείγματος είναι πολύ μικρό, τότε το βάρος που θα αποδοθεί σε αυτό θα πλησιάζει τη μονάδα. Κατ' αυτό το τρόπο μειώνεται η επίδραση των outliers στο τελικό αποτέλεσμα, ενώ ταυτόχρονα οι υπολογισμοί στηρίζονται μόνο στα σωστά δεδομένα. Εδώ οφείλεται και η σταθερότητα του M-estimator καθώς και η δυνατότητα που αυτός έχει να αντιμετωπίζει τα τυχόν σφάλματα που υπάρχουν στην είσοδο.

```

/* αρχικοποίηση */
t = 1; /* πρώτη επανάληψη */
f(t) = (r1 + ... + rm) / m
/* εκκίνηση επαναληπτικής διαδικασίας */
do {
    /* υπολόγισε σφάλματα */
    for(i = 1; i ≤ m; i++)
        ni = ri - f(t)
    t++; /* αύξησε μετρητή */
    /* υπολόγισε νέα τιμή της f */
    f(t) = 
$$\frac{\sum_i h(n_i) r_i}{\sum_i h(n_i)}$$

} until (|f(t) - f(t-1)| < ε) /* σύγκλιση */
f* = f(t); /* τελική λύση */

```

σχήμα 3.1: Επαναλαμβανόμενη σταθμισμένη μέση τιμή

Είναι φανερό ότι η αποτελεσματικότητα του M-estimator εξαρτάται άμεσα από την επιλογή της συνάρτησης επίδρασης h (ή ισοδύναμα της συνάρτησης κόστους g). Για παράδειγμα στην περίπτωση των ελαχίστων τετραγώνων (Least Squares) η συνάρτηση επίδρασης είναι:

$$h(n) = 1$$

Αυτό έχει σαν συνέπεια όλα τα δεδομένα να μετρούν το ίδιο στο τελικό αποτέλεσμα. Δηλαδή οι outliers έχουν το ίδιο βάρος με τους inliers. Γι' αυτόν ακριβώς το λόγο η μέθοδος των ελαχίστων τετραγώνων είναι εξαιρετικά ευαίσθητη στα λάθη, και η ύπαρξη ακόμα και ενός outlier είναι σε θέση να 'διαφθείρει' το τελικό αποτέλεσμα. Η λύση που προκύπτει στην περίπτωση αυτή είναι:

$$f^* = \frac{1}{m} \sum_{i=1}^m r_i$$

Παρακάτω παραθέτουμε τις συναρτήσεις κόστους καθώς και τις αντίστοιχες συναρτήσεις επίδρασης για δύο από τους πιο συχνά χρησιμοποιούμενους M-estimators. Αποτελεί συνήθη πρακτική οι συναρτήσεις αυτές να ορίζονται κατά τμήματα. Οι γραφικές παραστάσεις των συναρτήσεων επίδρασης φαίνονται και στο σχήμα 3.2.

- **Tukey's M-estimator** ([36]): ο αλγόριθμος αυτός αναφέρεται συχνά και με το όνομα 'Biweight estimator'

$$h_c(x) = \begin{cases} \left(1 - \left(\frac{x}{c}\right)^2\right)^2 & |x| \leq c \\ 0 & |x| > c \end{cases}$$

$$g_c(x) = \begin{cases} \frac{1 - \left(1 - \left(\frac{x}{c}\right)^2\right)^3}{6} & |x| \leq c \\ \frac{1}{6} & |x| > c \end{cases} \quad (3.11)$$

- **Huber's M-estimator** ([29])

$$h_c(x) = \begin{cases} 1 & |x| \leq c \\ \frac{c}{|x|} & |x| > c \end{cases}$$

$$g_c(x) = \begin{cases} \left(\frac{x}{c}\right)^2 & |x| \leq c \\ 2\frac{|x|}{c} - 1 & |x| > c \end{cases} \quad (3.12)$$

Όλες οι παραπάνω συναρτήσεις εξαρτώνται από τη παράμετρο c (της οποίας η τιμή είναι πάντα θετική), η οποία είναι γνωστή και με την ονομασία *παράμετρος υπολογισμού κλίμακας*. Η επιλογή της παραμέτρου αυτής παίζει πολύ σημαντικό ρόλο στην αποτελεσματικότητα του M-estimator. Όλα τα σημεία με σφάλμα μικρότερο κατ' απόλυτη τιμή από την παράμετρο αυτή θεωρούνται **inliers** και τους αποδίδεται μεγάλο βάρος από τη συνάρτηση επίδρασης, ενώ αντίθετα τα υπόλοιπα σημεία θεωρούνται **outliers** και τους αποδίδεται μικρό βάρος. Μια κακή επομένως επιλογή της παραμέτρου υπολογισμού κλίμακας μπορεί να οδηγήσει σε λανθασμένα αποτελέσματα, γι' αυτόν ακριβώς το λόγο απαιτείται αρκετή προσοχή και μελέτη κατά την εύρεση της. Μια συνηθισμένη επιλογή για την παράμετρο c είναι η:

$$c = c_0 \cdot \text{median}|n_i|$$

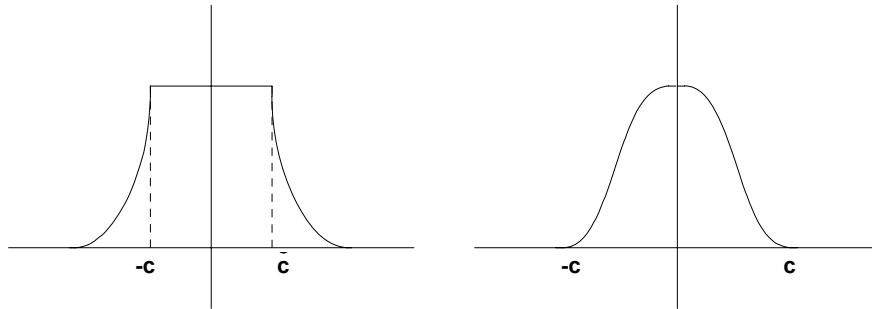
όπου c_0 είναι μια σταθερά κανονικοποίησης. Μια άλλη εξίσου συνηθισμένη επιλογή για τη παράμετρο υπολογισμού κλίμακας είναι και η παρακάτω:

$$c = c_0 S$$

$$S = \text{median}\left\{\left|n_i - \text{median}\{n_i\}\right|\right\}$$

$$c_0 = 1.4826$$

Στην παραπάνω σχέση η παράμετρος S είναι γνωστή και ως η ενδιάμεσος της απόλυτης απόκλισης (*Median of Absolute Deviation* ή για συντομία *MAD*) ([30]). Σε κάθε περίπτωση ο υπολογισμός της παραμέτρου c απαιτεί σημαντικό υπολογιστικό κόστος.



σχήμα 3.2: συναρτήσεις επίδρασης (αριστερά) του Huber και (δεξιά) του Tukey

Τέλος, αξίζει να παρατηρήσει κανείς το διαφορετικό τρόπο με τον οποίο αντιμετωπίζει τους *inliers* ο αλγόριθμος του Tukey σε σχέση με τον αλγόριθμο του Huber. Αυτό γίνεται αντιληπτό εξετάζοντας κανείς τις συναρτήσεις επίδρασης των αλγορίθμων αυτών. Έτσι ενώ ο πρώτος κάνει διάκριση μεταξύ των *inliers* αποδίδοντας στο καθένα διαφορετικό βάρος ανάλογα με το πόσο μικρό σφάλμα έχει, αντίθετα ο αλγόριθμος του Huber αντιμετωπίζει όλους τους *inliers* με την ίδια σημασία ανεξάρτητα από το σφάλμα τους. Σε αυτόν ακριβώς το λόγο οφείλεται και η καλλίτερη συμπεριφορά που παρουσιάζει ο αλγόριθμος του Tukey καθώς και το υψηλότερο σημείο φθοράς που διαθέτει.

3.5 Προβλήματα του M-estimator

Ο M-estimator αλγόριθμος, αν και αρκετά διαδεδομένος, αντιμετωπίζει σε ορισμένες περιπτώσεις κάποια προβλήματα σε σχέση με τη σταθερότητα του. Τα σημαντικότερα από αυτά είναι τα εξής:

- Είναι αρκετά ευαίσθητος στην αρχική εκτίμηση της παραμέτρου f ([39]). Μια κακή αρχική τιμή μπορεί να έχει ως αποτέλεσμα ακόμα και τη μη σύγκλιση του αλγορίθμου. Αλλά ακόμα και αν αυτό αποφευχθεί, η ανάγκη για μια καλή αρχική εκτίμηση παραμένει. Αυτό οφείλεται στο γεγονός ότι οι περισσότεροι M-estimators ορίζονται

σαν το ολικό ελάχιστο μιας εν γένει μη κυρτής συνάρτησης. Επομένως, όπως όλες οι τεχνικές που βασίζονται στο gradient, έτσι και ο M-estimator είναι δυνατόν να σταματήσει σε κάποιο τοπικό ελάχιστο της συνάρτησης αυτής, με συνέπεια να μην προκύψουν σωστά αποτελέσματα.

- Όπως είδαμε και προηγουμένως, ο M-estimator εξαρτάται από την παράμετρο κλίμακας υπολογισμού. Η παράμετρος αυτή μπορεί να αποτελέσει μια επιπλέον πηγή αστάθειας για τον αλγόριθμο. Για παράδειγμα, κατά τον υπολογισμό της παραμέτρου υπολογισμού κλίμακας χρησιμοποιούνται συνήθως είτε η ενδιάμεση τιμή είτε η ενδιάμεσος της απόλυτης απόκλισης (MAD). Και οι δύο αυτές τιμές δεν είναι αρκετά σταθερές και επηρεάζονται εύκολα από λάθη (outliers). Επιπλέον ο υπολογισμός των παραμέτρων αυτών έχει υπολογιστικό κόστος.
- Μια επιπλέον πηγή αστάθειας για τον M-estimator είναι και ο καθορισμός του κατωφλιού για την κατάταξη ενός σημείου ως inlier ή ως outlier.
- Η σύγκλιση του αλγορίθμου δεν είναι εξασφαλισμένη.

3.6 AM-estimator

Για την εξάλειψη των παραπάνω προβλημάτων υπάρχει μια βελτιωμένη παραλλαγή του M-estimator, η οποία είναι γνωστή με την ονομασία *AM-estimator* ([38]) (*Annealing M-estimator*). Το βασικό πλεονέκτημα του AM-estimator είναι το γεγονός ότι μπορεί να αντιμετωπίσει το πρόβλημα με τα τοπικά ελάχιστα και έτσι η λύση που θα δώσει δεν εξαρτάται από την αρχικοποίηση. Επιπλέον, δεν απαιτεί τον υπολογισμό ελεύθερων παραμέτρων, αποφεύγοντας έτσι την αστάθεια που συνεπάγεται η ύπαρξη τέτοιων σταθερών. Μια πολύ σημαντική συνέπεια του τελευταίου είναι και το γεγονός ότι μειώνεται σε μεγάλο βαθμό το υπολογιστικό κόστος μιας και πλέον δεν απαιτείται ο προσδιορισμός της παραμέτρου κλίμακας υπολογισμού και έτσι δεν χρειάζεται να υπολογίσουμε ενδιάμεσες τιμές.

Ο AM-estimator έχει την ίδια μορφή με τον M-estimator. Έτσι, οι εξισώσεις (3.3) έως (3.8) ισχύουν και στην περίπτωση αυτή. Η βασική διαφορά έγκειται στο γεγονός ότι η παράμετρος κλίμακας υπολογισμού c αντί να παραμένει σταθερή, αποτελεί τώρα μεταβλητή της οποίας οι τιμές προσεγγίζουν το 0^+ κατά τη διάρκεια του αλγορίθμου. Η μεταβλητή αυτή αποτελεί τώρα μια επιπλέον παράμετρο της συνάρτησης επίδρασης, η οποία πλέον θα συμβολίζεται ως h_c . Άρα για συγκεκριμένο c η λύση του AM-estimator θα είναι:

$$f_c = \frac{\sum_i h_c(n_i)r_i}{\sum_i h_c(n_i)} \quad (3.13)$$

Αρχικά το c τίθεται σε μια αρκετά μεγάλη τιμή και στη συνέχεια από επανάληψη σε επανάληψη μειώνεται προς το 0^+ . Κατ' αυτό το τρόπο προκύπτει μια ακολουθία λύσεων $\{f_c\}$ η τελευταία εκ των οποίων αποτελεί και τη τελική λύση f^* του AM-estimator. Δηλαδή ισχύει:

$$f^* = \lim_{c \rightarrow 0^+} f_c \quad (3.14)$$

Η όλη αυτή διαδικασία σαν σκοπό έχει να αυξηθεί η πιθανότητα για την αποφυγή των τοπικών ελαχίστων ([42]).

Ο λόγος για τον οποίο η παράμετρος c τίθεται αρχικά σε μια αρκετά μεγάλη τιμή $c^{(0)}$ είναι ώστε η αντίστοιχη συνάρτηση κόστους να είναι απολύτως κυρτή. Άρα στην περίπτωση αυτή θα είναι εύκολο να εντοπιστεί το μοναδικό ελάχιστο της συνάρτησης $E(f_c)$, ανεξάρτητα από την αρχική εκτίμηση. Το ελάχιστο αυτό χρησιμοποιείται στη συνέχεια σαν αρχική τιμή για την επόμενη φάση της ελαχιστοποίησης όπου αυτή τη φορά η παράμετρος c έχει μειωθεί. Καθώς η τιμή της παραμέτρου c μειώνεται, είναι πιθανό η αντίστοιχη συνάρτηση κόστους να μην είναι πλέον κυρτή και να εμφανιστούν επομένως τοπικά ελάχιστα. Ωστόσο, αν παρακολουθήσουμε τα ολικά ελάχιστα καθώς το c τείνει στο μηδέν, μπορούμε να ελπίζουμε ότι θα μπορέσουμε να προσεγγίσουμε τελικά το ολικό ελάχιστο f^* . Μια συνοπτική περιγραφή του AM-estimator σε ψευδοκώδικα φαίνεται στο σχήμα 3.3. Η αρχική τιμή που θα δοθεί στην c θα πρέπει να είναι τέτοια ώστε:

$$g_c''(n_i) > 0, \quad \forall i$$

Η συνθήκη αυτή εξασφαλίζει τη κυρτότητα της συνάρτησης κόστους g . Επιλέγοντας κατάλληλα τη συνάρτηση κόστους, η παραπάνω συνθήκη μπορεί πάντα να ικανοποιηθεί για μεγάλες τιμές του c . Για παράδειγμα η συνάρτηση Huber είναι πάντα κυρτή στο διάστημα $[-c, c]$. Άρα στη περίπτωση αυτή αρκεί να επιλέξουμε το c έτσι ώστε να ισχύει $c > |n_i|$ για όλα τα i .

```

/* αρχικοποίηση */
t = 1; /* πρώτη επανάληψη */
διάλεξε αρχική τιμή για το c;


$$f_c^{(t)} = \frac{r_1 + \dots + r_m}{m};$$


do { /* εκκίνηση επαναληπτικής διαδικασίας */
  /* υπολόγισε σφάλματα */
  for(i = 1; i ≤ m; i++)
    
$$n_i = r_i - f_c^{(t)}$$

  t = t + 1;
  c = lower(c); /* μείωσε τιμή της c */
  /* υπολόγισε νέα τιμή της f */
  
$$f_c^{(t)} = \frac{\sum_i h_c(n_i) r_i}{\sum_i h_c(n_i)}$$

} until (c < δ or  $|f_c^{(t)} - f_c^{(t-1)}| < \epsilon$ ) /* σύγκλιση */

 $f^* = f_c^{(t)}$  /* τελική λύση */

```

σχήμα 3.3: AM-estimator

Κάτω από αυτές τις συνθήκες, το σύνολο δεδομένων μας θα αποτελείται από ζευγάρια διανυσμάτων οπτικής ροής και δισδιάστατων σημείων πάνω στην εικόνα. Δηλαδή:

$$\left((u_i, v_i), (x_i, y_i) \right) \quad i = 1 \dots N$$

Τα παραπάνω ζεύγη διανυσμάτων θα αποκαλούνται στη συνέχεια **σημεία οπτικής ροής**. Θα πρέπει εδώ να σημειωθεί ότι το N στην παραπάνω σχέση δεν είναι απαραίτητο να ισούται με το πλήθος των σημείων της εικόνας. Κάτι τέτοιο εξαρτάται από το αν έχουμε στη διάθεση μας ένα πυκνό πεδίο οπτικής ροής ή όχι.

3.7 M-estimator και εκτίμηση της κίνησης της κάμερας

Με βάση αυτά τα δεδομένα, εκείνο που ψάχνουμε να βρούμε είναι οι παραμέτροι της τρισδιάστατης κίνησης. Όπως όμως έχει ήδη αναφερθεί στο προηγούμενο κεφάλαιο, κάτι τέτοιο είναι ισοδύναμο με την εύρεση των 8 βασικών συντελεστών της εξίσωσης της οπτικής ροής, η οποία έχει ως εξής:

$$a_1 - a_2x_i - a_3y_i + a_4x_i^2 + a_5y_i^2 - a_6x_iy_i + a_7u_i - a_8v_i = x_iv_i - y_iu_i \quad (3.15)$$

Από τα N σημεία οπτικής ροής που διαθέτουμε, εκείνα που κινούνται σύμφωνα με τη κίνηση της κάμερας πρέπει να ικανοποιούν την εξίσωση (3.15), ενώ αντίθετα τα υπόλοιπα πρέπει να παρουσιάζουν αποκλίσεις σε σχέση με αυτή. Με άλλα λόγια ο ρόλος της σχέσης (3.15) στη περίπτωση αυτή, είναι ανάλογος με το ρόλο της σχέσης (3.1) ή για να είμαστε απόλυτα ακριβείς με αυτόν της σχέσης (3.2). Οι συντελεστές δηλαδή a_i αντιστοιχούν στους όρους f_i ενώ οι παράμετροι r_{ji} εκφράζονται μέσω των συνιστωσών των σημείων οπτικής ροής. Επομένως οι τιμές των 8 παραμέτρων θα προκύψουν σαν οι λύσεις του M-estimator κατά την ελαχιστοποίηση της παρακάτω συνάρτησης σφάλματος:

$$E(a_1, \mathbf{K}, a_8) = \sum_{i=0}^N g(a_1 - a_2x_i - a_3y_i + a_4x_i^2 + a_5y_i^2 - a_6x_iy_i + a_7u_i - a_8v_i - (x_iv_i - y_iu_i))$$

Στη παραπάνω σχέση η συνάρτηση g αποτελεί τη συνάρτηση κόστους του M-estimator (π.χ. συνάρτηση Tukey ή συνάρτηση Huber).

Κεφάλαιο 4

4. ΙΕΡΑΡΧΙΑ ΤΡΙΣΔΙΑΣΤΑΤΩΝ ΜΟΝΤΕΛΩΝ ΚΙΝΗΣΗΣ

4.1 Εισαγωγή

Στο κεφάλαιο αυτό θα αναλυθεί ο αλγόριθμος για την εκτίμηση της κίνησης της κάμερας. Ο αλγόριθμος αυτός θα βασιστεί στη χρήση μιας **ιεραρχίας από τρισδιάστατα μοντέλα κίνησης** σε συνδυασμό με την εφαρμογή του M-estimator. Στην πρώτη παράγραφο του κεφαλαίου θα καταστεί σαφές τι εννοούμε με τον όρο ‘ιεραρχία από τρισδιάστατα μοντέλα κίνησης’. Επιπλέον θα αναλυθεί για ποιο λόγο είναι αναγκαία η χρησιμοποίηση μιας τέτοιας ιεραρχίας για τον προσδιορισμό των παραμέτρων της τρισδιάστατης κίνησης, ενώ θα παρουσιασθούν και κάποιες βασικές έννοιες σχετικά με το θέμα αυτό. Στη συνέχεια θα περιγραφεί μια συγκεκριμένη ιεραρχία από τρισδιάστατα μοντέλα κίνησης (η οποία και θα χρησιμοποιηθεί για την εκτίμηση της κίνησης της κάμερας) και κάθε ένα από τα μοντέλα της ιεραρχίας αυτής θα μελετηθεί διεξοδικά. Το κεφάλαιο θα κλείσει με την παρουσίαση κάποιων πειραματικών αποτελεσμάτων.

4.2 Ιεραρχία τρισδιάστατων μοντέλων κίνησης - Βασικές έννοιες

Όπως έχει ήδη αναφερθεί για τον προσδιορισμό των τρισδιάστατων χαρακτηριστικών της κίνησης της κάμερας θα χρησιμοποιηθεί ο M-estimator αλγόριθμος. Η ιδέα είναι ότι ο M-estimator θα προσπαθήσει να βρει εκείνες τις παραμέτρους κίνησης οι οποίες ικανοποιούν ‘καλλίτερα’ τις εξισώσεις οπτικής ροής σε κάθε σημείο. Ωστόσο μπορεί εύκολα να αποδειχθεί ότι το θεωρητικό σημείο φθοράς (breakdown point) του M-estimator αλγορίθμου είναι πάντα μικρότερο από την ποσότητα $1/(p + 1)$, όπου p παριστάνει το πλήθος των παραμέτρων που ο M-estimator πρέπει τελικά να υπολογίσει. Το άμεσο συμπέρασμα που προκύπτει από την παρατήρηση αυτή είναι ότι όσο περισσότερες είναι οι παράμετροι υπολογισμού, τόσο λιγότερο robust (σταθερός) γίνεται ο M-estimator. Επομένως, αν καταφέρουμε να μειώσουμε το πλήθος των παραμέτρων που πρέπει να εκτιμηθούν, τότε είναι λογικό να αναμένουμε καλλίτερα και πιο αξιόπιστα αποτελέσματα.

Μέχρι τώρα, οι περισσότερες από τις μελέτες που ασχολούνται με την εκτίμηση των παραμέτρων τρισδιάστατης κίνησης ([10], [15], [44], [45]), χρησιμοποιούν εκ των προτέρων ένα σταθερό τρισδιάστατο μοντέλο, χωρίς να προσπαθούν να προσαρμόσουν το μοντέλο αυτό στα χαρακτηριστικά της κίνησης που παρατηρείται στην ακολουθία εικόνων υπό εξέταση. Στην

πλειονότητα των περιπτώσεων το μοντέλο που συνήθως χρησιμοποιείται είναι το πιο γενικό μοντέλο, αυτό της πλήρους τρισδιάστατης κίνησης. Δηλαδή γινόταν η υπόθεση ότι η κάμερα περιέχει ταυτόχρονα μεταφορική αλλά και περιστροφική κίνηση και ότι κινείται με 6 βαθμούς ελευθερίας. Σε αρκετές όμως περιπτώσεις κάτι τέτοιο δεν είναι αληθές, με συνέπεια τα τελικά αποτελέσματα να μην είναι τα αναμενόμενα ή απλά να επιβαρύνεται υπολογιστικά η όλη διαδικασία χωρίς αυτό να είναι αναγκαίο.

Επομένως, αντί να χρησιμοποιούμε ένα σταθερό μοντέλο, αυτό που θα προσπαθήσουμε να επιτύχουμε είναι να βρούμε εκείνο το μοντέλο τρισδιάστατης κίνησης το οποίο μπορεί να περιγράψει πλήρως τη κίνηση της κάμερας αλλά και το οποίο επιπλέον περιέχει όσο το δυνατόν λιγότερες παραμέτρους. Με αυτό το τρόπο εκτός του ότι βελτιώνεται η ποιότητα των τελικών αποτελεσμάτων (μιας και ο M-estimator γίνεται λιγότερο ευαίσθητος στα λάθη), επιπλέον μειώνεται και η υπολογιστική πολυπλοκότητα του M-estimator (λιγότερες παράμετροι να υπολογιστούν). Από την άλλη μεριά η διαδικασία επιλογής του μοντέλου που θα χρησιμοποιηθεί απαιτεί ιδιαίτερη προσοχή. Αν για παράδειγμα χρησιμοποιήσουμε ένα πολύ απλό μοντέλο τρισδιάστατης κίνησης, ενώ στην πραγματικότητα η κίνηση της κάμερας είναι πιο πολύπλοκη, τότε είναι φανερό ότι οι παράμετροι κίνησης που θα εκτιμηθούν θα περιέχουν μεγάλα σφάλματα.

Προς το σκοπό αυτό, η προσέγγιση που θα ακολουθηθεί κατά τη παρούσα εργασία θα είναι η χρησιμοποίηση μιας **ιεραρχίας από τρισδιάστατα μοντέλα κίνησης** ([11]). Κάθε ένα από τα μοντέλα αυτά θα έχει διαφορετικό βαθμό πολυπλοκότητας, ενώ τα υψηλότερα στην ιεραρχία θα είναι σε θέση να περιγράψουν πλήρως ακόμα και τη περίπτωση της πλήρους τρισδιάστατης κίνησης. Από τη στιγμή που θα έχει καθοριστεί ποια θα είναι τα μοντέλα τρισδιάστατης κίνησης που θα περιέχονται στην ιεραρχία, τα εξής δύο θέματα μένουν να αποφασιστούν:

- το κριτήριο αποδοχής - απόρριψης με βάση το οποίο θα κρίνεται αν ένα μοντέλο πληρεί τις προϋποθέσεις για να περιγράψει τη κίνηση της κάμερας ή όχι
- και η σειρά με την οποία θα εξετασθούν τα μοντέλα της ιεραρχίας για να αποφασιστεί ποιο θα επιλεγεί τελικά

4.3 Κριτήριο αποδοχής - απόρριψης

Αρκετά συχνά το κριτήριο αποδοχής - απόρριψης συνίσταται ουσιαστικά στην ελαχιστοποίηση μιας *συνάρτησης κόστους*. Το κριτήριο αυτό αν και μπορεί να είναι διαφορετικό ανάλογα με το μοντέλο στο οποίο εφαρμόζεται, στη γενική περίπτωση ενός μοντέλου β εξαρτάται από τους παρακάτω παράγοντες:

- το σφάλμα πρόβλεψης του μοντέλου β
- τη πολυπλοκότητα του μοντέλου
- το υπολογιστικό κόστος που απαιτείται για τον υπολογισμό των παραμέτρων του συγκεκριμένου μοντέλου

Στη περίπτωση μας η συνάρτηση που θα πρέπει να ελαχιστοποιήσουμε θα αντιστοιχεί στη συνάρτηση ολικού λάθους ενός M-estimator, δεδομένου ότι η συνάρτηση αυτή μετρά το κατά πόσο οι εκτιμήσεις του M-estimator ταιριάζουν με τα δεδομένα μας. Αν καταφέρουμε να μειώσουμε τη συνάρτηση αυτή κάτω από κάποιο κατώφλι το μοντέλο γίνεται αποδεκτό αλλιώς απορρίπτεται. Βέβαια, ένα κρίσιμο σημείο κατά τη διαδικασία αυτή είναι και η σωστή επιλογή του κατωφλιού. Συγκεκριμένα παραδείγματα τέτοιων συναρτήσεων θα αναφερθούν για κάθε μοντέλο της ιεραρχίας στη συνέχεια του κεφαλαίου.

4.4 Σειρά επίσκεψης μοντέλων

Σχετικά με το δεύτερο ζήτημα, δηλαδή τη σειρά επίσκεψης των μοντέλων, δύο είναι οι πιο διαδεδομένες προσεγγίσεις:

- **Παράλληλη προσέγγιση (Parallel approach):** Σύμφωνα με τη μέθοδο αυτή όλα τα μοντέλα τρισδιάστατης κίνησης εξετάζονται παράλληλα και επιλέγεται τελικά το πιο αποτελεσματικό. Στη περίπτωση αυτή το πρόβλημα μπορεί να λυθεί χρησιμοποιώντας το μαθηματικό πλαίσιο στατιστικών μεθόδων οι οποίες βασίζονται σε στοιχεία από τη θεωρία πληροφορίας ([46]). Τέτοια κριτήρια είναι και αυτά των Akaike και Rissanen, τα οποία για ένα δεδομένο μοντέλο λαμβάνουν υπόψη και αξιολογούν την απόδοση αλλά και την πολυπλοκότητα του μοντέλου. Τα δύο αυτά κριτήρια εκφράζονται ως εξής:

- **AKAIKE κριτήριο**

$$C_b = -2\log f(y/V_b) + 2\dim(V_b) \quad (4.1)$$

- **RISSANEN κριτήριο**

$$C_b = -2\log f(y/V_b) + \dim(V_b) \times \log N \quad (4.2)$$

Στις παραπάνω σχέσεις το $f(y/V_b)$ εκφράζει την πιθανότητα του y δεδομένου του V_b , όπου το V_b παριστάνει το διάνυσμα παραμέτρων του μοντέλου b . Επίσης το N συμβολίζει το πλήθος των outliers. Οι πρώτοι όροι στις παραπάνω εκφράσεις μετρούν το βαθμό απόδοσης του μοντέλου, ενώ οι δεύτεροι όροι χρησιμεύουν ως 'ποινή' για τα πολύπλοκα

μοντέλα. Αν για παράδειγμα τα παραπάνω κριτήρια εφαρμοστούν σε δύο μοντέλα a και b , τότε θα επιλεγεί το μοντέλο a αν και μόνο αν ισχύει ότι: $C_a < C_b$.

- **Σειριακή προσέγγιση (Sequential approach):** Σύμφωνα με τη μέθοδο αυτή τα μοντέλα τρισδιάστατης κίνησης εξετάζονται με βάση κάποια προκαθορισμένη διάταξη, το ένα μετά το άλλο. Η προκαθορισμένη αυτή διάταξη αντιστοιχεί ουσιαστικά σε μια συγκεκριμένη διάσχιση της ιεραρχίας. Μόνο αν το υπό εξέταση μοντέλο αποτύχει να περιγράψει ικανοποιητικά την κίνηση της κάμερας (ή αν είναι 'χειρότερο' από το επόμενο μοντέλο στη σειρά), τότε μόνο εξετάζεται το επόμενο μοντέλο. Σχετικά με το πως θα γίνει η διάσχιση της ιεραρχίας, 3 είναι οι δυνατές λύσεις:
 - ξεκινάμε από το πιο απλό μοντέλο και προχωράμε προς το πιο πολύπλοκο
 - ξεκινάμε από το πιο γενικό μοντέλο και πάμε προς το απλούστερο μοντέλο
 - ή ξεκινάμε από ένα ενδιάμεσο μοντέλο και προχωράμε είτε προς το πιο γενικό είτε προς το πιο απλό μοντέλο.

Σε όλες αυτές τις σειριακές προσεγγίσεις, δύο είναι οι δυνατές υποθέσεις που θα πρέπει συγκριθούν ([47]) (υπό την έννοια για παράδειγμα του κριτηρίου της μέγιστης πιθανοφάνειας):

- **Υπόθεση H_0 :** Η κίνηση της κάμερας αντιστοιχεί σε ένα μοντέλο τρισδιάστατης κίνησης β
- **Υπόθεση H_1 :** Η κίνηση της κάμερας αντιστοιχεί σε ένα ελαφρώς πιο πολύπλοκο μοντέλο $\beta+1$

Η μεθοδολογία που θα ακολουθηθεί κατά τη παρούσα εργασία θα είναι αυτή της σειριακής προσέγγισης, όπου τα μοντέλα θα εξετάζονται από το απλούστερο προς το πιο πολύπλοκο. Με αυτό το τρόπο ελπίζουμε τελικά να βρούμε το απλούστερο δυνατό μοντέλο τρισδιάστατης κίνησης, το οποίο περιγράφει τη κίνηση της κάμερας.

4.5 Εύρεση ιεραρχίας μοντέλων τρισδιάστατης κίνησης

Ο επόμενος στόχος είναι να βρούμε ποια ακριβώς θα είναι τα μοντέλα της τρισδιάστατης κίνησης τα οποία θα συμπεριλάβουμε στην ιεραρχία που θα κατασκευάσουμε. Στη γενική περίπτωση η κίνηση της κάμερας μπορεί να περιέχει μέχρι και 6 βαθμούς ελευθερίας, 3 βαθμούς λόγω μεταφορικής και 3 βαθμούς λόγω περιστροφικής κίνησης. Ωστόσο κάτω από ορισμένες προϋποθέσεις ([48], [49], [11]), η κίνηση της κάμερας μπορεί να αναχθεί σε απλούστερες

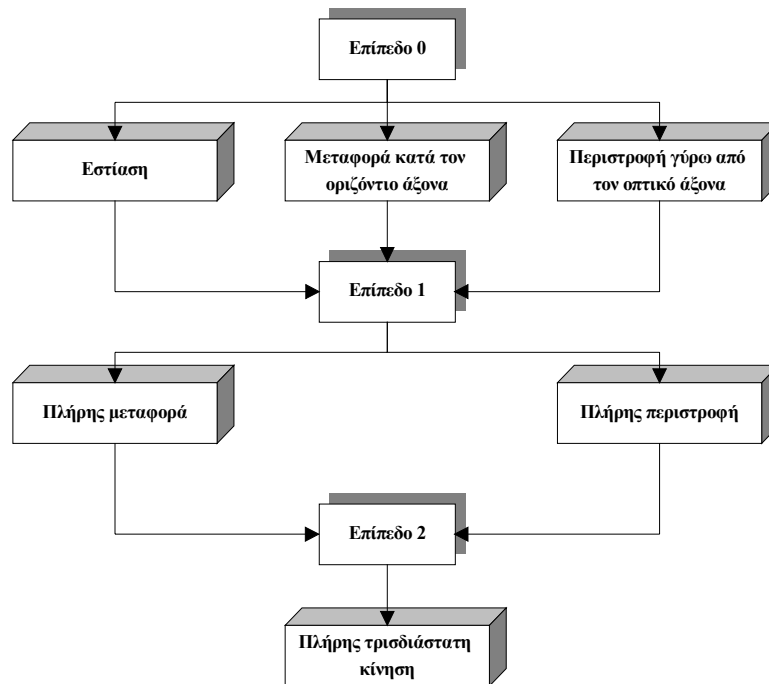
μορφές τρισδιάστατης κίνησης. Οι προϋποθέσεις που πρέπει να πληρούνται, έχουν ανάμεσα στα άλλα να κάνουν με:

- μικρές γωνίες περιστροφής της κάμερας και
- με μικρές μετατοπίσεις της κάμερας κατά το οπτικό άξονα σε σχέση με την απόσταση των υπάρχοντων αντικειμένων από την ίδια την κάμερα

Στις περιπτώσεις που ικανοποιούνται οι παραπάνω συνθήκες, τότε μπορεί να αποδειχθεί ότι η κίνηση της κάμερας είναι δυνατόν να αναχθεί σε μια από τις παρακάτω τρισδιάστατες κινήσεις:

- απλή μεταφορική κίνηση η οποία είναι παράλληλη με το επίπεδο της εικόνας
- απλή μεταφορική κίνηση η οποία είναι κάθετη στο επίπεδο της εικόνας ή ισοδύναμα μια αλλαγή στην εστιακή απόσταση (zooming)
- και τέλος απλή περιστροφική κίνηση γύρω από τον οπτικό άξονα

Με βάση τις παραπάνω παρατηρήσεις, κατασκευάστηκε η ιεραρχία μοντέλων τρισδιάστατης κίνησης που φαίνεται στο σχήμα 4.1. Όπως φαίνεται και από το σχήμα στο πρώτο επίπεδο της ιεραρχίας βρίσκονται τα απλούστερα μοντέλα κίνησης. Ενώ καθώς προχωράμε σε μεγαλύτερα επίπεδα, παρουσιάζονται όλο και πιο πολύπλοκα μοντέλα. Στο τελευταίο επίπεδο βρίσκεται το πιο γενικό μοντέλο, αυτό της πλήρους τρισδιάστατης κίνησης. Τα κριτήρια απόρριψης - αποδοχής των μοντέλων μπορούν να φανούν χρήσιμα και σε εφαρμογές κατά τις οποίες μόνο το είδος της κίνησης της κάμερας ενδιαφέρει και όχι η ακριβής εκτίμηση της. Επίσης, για όλα τα μοντέλα της ιεραρχίας θα προσδιορισθούν τα σημεία της εικόνας τα οποία είναι σύμφωνα με την κίνηση της κάμερας (inliers). Κάτι τέτοιο μπορεί να χρησιμοποιηθεί σε εφαρμογές που έχουν σχέση με προσδιορισμό ανεξάρτητης κίνησης.



σχήμα 4.1: Ιεραρχία μοντέλων κίνησης

Η διάσχιση της παραπάνω ιεραρχίας θα γίνει από το απλούστερο μοντέλο προς το πιο πολύπλοκο. Θα πρέπει σε αυτό το σημείο να σημειωθεί ότι η ιεραρχία μοντέλων που παρουσιάζεται στο σχήμα 4.1 θα μπορούσε να εμπλουτισθεί και με άλλα μοντέλα τρισδιάστατης κίνησης.

Στη συνέχεια του κεφαλαίου θα μελετηθεί κάθε ένα από τα μοντέλα που περιέχονται στη παραπάνω ιεραρχία, και θα αναφερθούν λεπτομέρειες για το πως θα γίνει η απόρριψη - αποδοχή για κάθε ένα από αυτά. Στόχος μας θα είναι να προσδιορίσουμε σε κάθε περίπτωση ένα robust κριτήριο απόρριψης - αποδοχής καθώς και μια robust μέθοδο υπολογισμού των παραμέτρων του μοντέλου όπου αυτό είναι δυνατό. Η μελέτη θα ξεκινήσει με το μοντέλο της μεταφοράς κατά τον οριζόντιο άξονα και θα καταλήξει με τη μελέτη του μοντέλου της πλήρους τρισδιάστατης κίνησης.

4.6 Μεταφορά κατά τον οριζόντιο άξονα

Στην περίπτωση αυτή η κίνηση που εκτελεί η κάμερα είναι μια απλή μεταφορική κίνηση παράλληλα προς το επίπεδο της εικόνας. Δεν υπάρχει καθόλου περιστροφή. Επομένως τα στιγμιαία διανύσματα περιστροφής και μετατόπισης, R και T , θα είναι:

$$\overset{P}{T} = \begin{bmatrix} T_x \\ 0 \\ 0 \end{bmatrix} \quad \overset{P}{R} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (4.3)$$

Χρησιμοποιώντας τις σχέσεις που συνδέουν την οπτική ροή με τις παραμέτρους της τρισδιάστατης κίνησης, η εξίσωση της οπτικής ροής ανάγεται στην παρακάτω μορφή:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{T_x}{Z} \\ 0 \end{bmatrix} \quad (4.4)$$

Δεδομένου ότι δεν γνωρίζουμε τίποτα για το βάθος Z (μιας και δεν κάνουμε καμία υπόθεση γι' αυτό), δεν μπορούμε να χρησιμοποιήσουμε την παραπάνω εξίσωση οπτικής ροής για να εκτιμήσουμε την παράμετρο T_x . Το μόνο που μπορούμε να επιτύχουμε στην περίπτωση αυτή είναι απλά να αποφασίσουμε για το αν θα αποδεχτούμε ή θα απορρίψουμε το μοντέλο. Το γεγονός ότι δεν μπορούμε να εκτιμήσουμε ακριβώς τη συνιστώσα T_x οφείλεται στο γνωστό πρόβλημα του '*scale factor ambiguity*'. Λόγω της προβολής των τρισδιάστατων αντικειμένων στο επίπεδο της εικόνας, χάνεται αρκετή από την υπάρχουσα πληροφορία. Αυτό έχει σαν αποτέλεσμα να μην είναι πλέον δυνατόν να προσδιορίσουμε επακριβώς τις συνιστώσες του διανύσματος της μεταφορικής κίνησης αλλά μόνο τους λόγους των συνιστωσών αυτών μεταξύ τους.

Το κριτήριο απόρριψης ή αποδοχής του συγκεκριμένου μοντέλου θα βασιστεί στο γεγονός ότι η v -συνιστώσα του διανύσματος οπτικής ροής πρέπει να είναι μηδέν στην περίπτωση αυτή. Άρα αν συμβολίσουμε με t την γωνία μεταξύ του διανύσματος της οπτικής ροής (u,v) και του οριζόντιου άξονα (βλέπε σχήμα 4.2), τότε η γωνία αυτή θα είναι μηδέν για όλα τα σημεία που κινούνται σύμφωνα με τη κάμερα. Δηλαδή το διάνυσμα οπτικής ροής πρέπει να είναι παράλληλο προς τον οριζόντιο άξονα. Συνεπώς για τα σημεία αυτά θα ισχύει ότι:

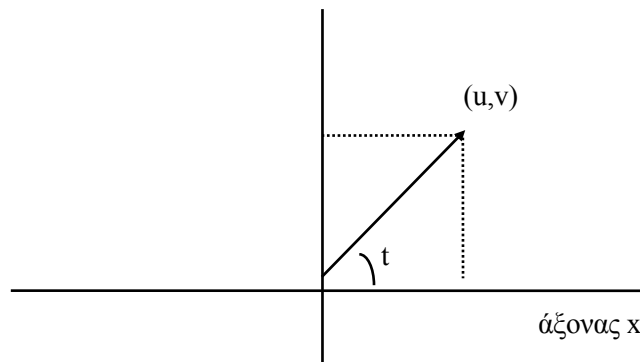
$$\arctan\left(\frac{v}{u}\right) = 0$$

Βέβαια λόγω σφαλμάτων στους υπολογισμούς, πιθανότατα το τόξο εφαπτομένης στην παραπάνω σχέση να μην είναι ακριβώς μηδέν αλλά κοντά στο μηδέν. Επομένως για να προσδιορίσουμε τα σημεία της εικόνας που κινούνται σύμφωνα με την κάμερα, αρκεί να ελαχιστοποιήσουμε ως προς a την παρακάτω ποσότητα χρησιμοποιώντας έναν M-estimator αλγόριθμο:

$$\sum_{i=0}^N g\left(a - \arctan\left(\frac{v_i}{u_i}\right)\right) \quad (4.5)$$

Η g στην παραπάνω σχέση αποτελεί τη συνάρτηση κόστους ενός M-estimator (π.χ. Tukey). Η λύση που θα δώσει ο M-estimator θα εκφράζει ουσιαστικά την απόκλιση που παρουσιάζουν τα

διανύσματα οπτικής ροής από τον οριζόντιο άξονα. Για να εκτελεί η κάμερα μεταφορά κατά τον οριζόντιο άξονα, η γωνία αυτή πρέπει να είναι είτε κοντά στις 0 μοίρες είτε κοντά στις 180 μοίρες. Άρα αν η λύση που θα προκύψει βρίσκεται κοντά στις γωνίες αυτές (η απόσταση δηλαδή είναι μικρότερη από κάποιο προκαθορισμένο κατώφλι) και υπάρχει ένας ικανοποιητικός αριθμός από inliers, τότε το μοντέλο γίνεται αποδεκτό αλλιώς απορρίπτεται. Ένα κρίσιμο σημείο είναι και η επιλογή του κατωφλιού. Κατά τη παρούσα υλοποίηση επιλέχθηκε το κατώφλι αυτό να έχει τιμή ίση με το μήκος τόξου που αντιστοιχεί σε γωνία απόκλισης 3 μοιρών, δηλαδή $6\pi/360$.



σχήμα 4.2

4.7 Εστίαση (zooming)

Στην περίπτωση της εστίασης η κάμερα μετατοπίζεται κάθετα ως προς το επίπεδο της εικόνας. Επιπλέον, η κάμερα δεν περιστρέφεται. Επομένως τα στιγμιαία διανύσματα της μετατόπισης και της περιστροφής, τα οποία περιγράφουν την τρισδιάστατη κίνηση της κάμερας, θα είναι:

$$\overset{P}{T} = \begin{bmatrix} 0 \\ 0 \\ T_z \end{bmatrix} \quad \overset{P}{R} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Άρα, βασιζόμενοι στις εξισώσεις που συνδέουν τις συνιστώσες του διανύσματος οπτικής ροής με τις παραμέτρους της τρισδιάστατης κίνησης, προκύπτει ότι:

$$\left. \begin{aligned} u &= -\frac{xT_z}{Z} \\ v &= -\frac{yT_z}{Z} \end{aligned} \right\} \Rightarrow uy - xv = 0 \quad (4.6)$$

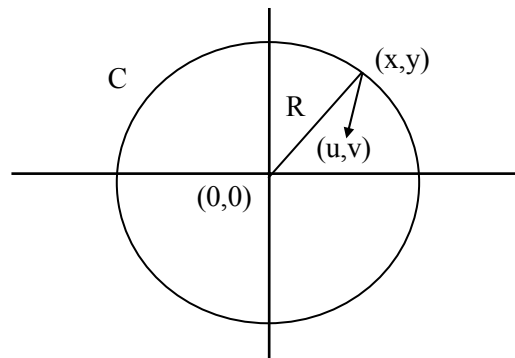
Στην παραπάνω σχέση οι συντεταγμένες x, y του σημείου είναι ως προς το κέντρο της εικόνας, το οποίο και έχει συντεταγμένες $(0, 0)$. Λόγω του ότι στην παραπάνω εξίσωση οπτικής ροής δεν συμμετέχουν καθόλου οι παράμετροι της τρισδιάστατης κίνησης, δεν μπορούμε να εκτιμήσουμε τις παραμέτρους αυτές βασιζόμενοι μόνο στην οπτική ροή. Όπως και στην προηγούμενη

περίπτωση, έτσι και εδώ, το μόνο που μπορούμε να κάνουμε είναι να αποφανθούμε για το αν η τριδιάστατη κίνηση της κάμερας είναι zooming ή όχι.

Ας ονομάσουμε C το κύκλο που διέρχεται από το σημείο με συντεταγμένες (x,y) και έχει κέντρο το οποίο ταυτίζεται με το κέντρο της εικόνας (βλέπε σχήμα 4.3). Τότε μπορεί εύκολα να αποδειχθεί ότι το ημίτονο της γωνίας μεταξύ της ακτίνας R του κύκλου και του διανύσματος της οπτικής ροής θα είναι:

$$\frac{|uy - xv|}{\sqrt{x^2 + y^2} \sqrt{u^2 + v^2}} \quad (4.7)$$

Συνδυάζοντας τις σχέσεις (4.6) και (4.7) προκύπτει ότι το ημίτονο αυτό είναι μηδέν άρα και η αντίστοιχη γωνία είναι μηδέν. Αυτό σημαίνει ότι στη περίπτωση της εστίασης τα διανύσματα οπτικής ροής συγκλίνουν προς το κέντρο της εικόνας. Το χαρακτηριστικό αυτό θα αποτελέσει τη βάση για την αποδοχή ή απόρριψη του μοντέλου.



σχήμα 4.3

Για να εντοπίσουμε δηλαδή τα σημεία της εικόνας των οποίων η κίνηση είναι σύμφωνη με την κίνηση της κάμερας, θα ελαχιστοποιήσουμε ως προς a τη παρακάτω ποσότητα:

$$\sum_{i=0}^N g \left(a - \arcsin \left(\frac{u_i y_i - x_i v_i}{\sqrt{x_i^2 + y_i^2} \sqrt{u_i^2 + v_i^2}} \right) \right) \quad (4.8)$$

Η συνάρτηση g παριστάνει τη συνάρτηση κόστους ενός M-estimator αλγορίθμου (π.χ. Tukey estimator). Η λύση a που θα δώσει ο M-estimator θα εκφράζει την απόκλιση που παρουσιάζουν τα διανύσματα οπτικής ροής από την ακτινική διεύθυνση. Αν το μήκος τόξου a είναι κοντά στο 0 ή στο π (δηλαδή η απόσταση είναι κάτω από ένα ορισμένο κατώφλι), τότε το μοντέλο της εστίασης γίνεται αποδεκτό αλλιώς απορρίπτεται. Όπως και στην περίπτωση της μεταφοράς κατά τον οριζόντιο άξονα, έτσι και εδώ, το κατώφλι αυτό θα επιλεγεί να είναι ίσο με $6\pi/360$.

Θα πρέπει εδώ να σημειωθεί ότι κατά την εφαρμογή του κριτηρίου αποδοχής ή απόρριψης, επιλέξαμε να χρησιμοποιήσουμε τη ποσότητα (4.8) αντί της απλούστερης:

$$\sum_{i=0}^N g(a - (u_i y_i - x_i v_i))$$

λόγω του ότι η τελευταία είναι περισσότερη ευαίσθητη στο θόρυβο με αποτέλεσμα οι εκτιμήσεις που θα προέκυπταν να μην ήταν αξιόπιστες στην περίπτωση ύπαρξης outliers.

4.8 Περιστροφή γύρω από τον οπτικό άξονα

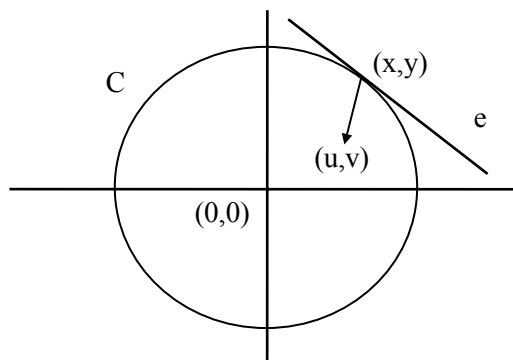
Στην περίπτωση αυτή δεν έχουμε καθόλου μεταφορική κίνηση. Αντίθετα, η κάμερα εκτελεί απλή περιστροφική κίνηση γύρω από τον οπτικό άξονα. Άρα τα διανύσματα της περιστροφής και της μετατόπισης, R και T , θα έχουν την εξής μορφή:

$$P = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad R = \begin{bmatrix} 0 \\ 0 \\ R_z \end{bmatrix}$$

Επομένως, αντικαθιστώντας τα διανύσματα αυτά στις εξισώσεις μεταξύ οπτικής ροής και τρισδιάστατης κίνησης, προκύπτει η παρακάτω εξίσωση οπτικής ροής:

$$\left. \begin{array}{l} u = -R_z y \\ v = R_z x \end{array} \right\} \Rightarrow ux + vy = 0 \quad (4.9)$$

(Οι συντεταγμένες του σημείου (x, y) λαμβάνονται ως προς το κέντρο της εικόνας). Θα βασιστούμε σε αυτή τη σχέση για να απορρίψουμε ή να αποδεχτούμε το μοντέλο της περιστροφής περί τον οπτικό άξονα. Στη περίπτωση δε που το μοντέλο γίνει αποδεκτό, στη συνέχεια θα γίνει και εκτίμηση της παραμέτρου R_z .



σχήμα 4.4

Ας είναι C ο κύκλος ο οποίος διέρχεται από το σημείο με συντεταγμένες (x,y) και έχει κέντρο το οποίο ταυτίζεται με το κέντρο της εικόνας (βλέπε σχήμα 4.4). Έστω e η ευθεία που εφάπτεται του κύκλου στο σημείο (x,y) . Είναι εύκολο να αποδειχθεί ότι το ημίτονο της γωνίας μεταξύ της εφαπτομένης e και του διανύσματος της οπτικής ροής θα είναι:

$$\frac{ux + vy}{\sqrt{x^2 + y^2} \sqrt{u^2 + v^2}}$$

Λόγω της σχέσης (4.9) το ημίτονο αυτό (καθώς και η αντίστοιχη γωνία) θα είναι μηδέν, για όλα τα σημεία που κινούνται σύμφωνα με την κάμερα. Άρα για το τεστ αποδοχής - απόρριψης θα ελαχιστοποιήσουμε ως προς a την παρακάτω ποσότητα, με τη βοήθεια μιας M-estimator τεχνικής:

$$\sum_{i=0}^N g \left(a - \arcsin \left(\frac{u_i x_i + y_i v_i}{\sqrt{x_i^2 + y_i^2} \sqrt{u^2 + v^2}} \right) \right) \quad (4.10)$$

Το μοντέλο θα γίνει αποδεκτό αν και μόνο αν η λύση που θα δώσει ο M-estimator βρίσκεται κοντά στο 0 ή στο π . Η τιμή του κατωφλιού που θα χρησιμοποιηθεί για το σκοπό αυτό θα είναι όπως και στις προηγούμενες περιπτώσεις ίση με $6\pi/360$.

Αφού έχουμε επιβεβαιώσει ότι η κάμερα εκτελεί περιστροφή περί του οπτικού άξονα, θα προσπαθήσουμε να εκτιμήσουμε την παράμετρο R_z . Θα βασιστούμε στις σχέσεις:

$$\begin{aligned} u &= -R_z y \\ v &= R_z x \end{aligned}$$

Η εκτίμηση της παραμέτρου R_z θα προκύψει σαν η λύση που θα δώσει ο M-estimator κατά την ελαχιστοποίηση της παρακάτω συνάρτησης:

$$\sum_{i=0}^N \left[g \left(R_z - \frac{v_i}{x_i} \right) + g \left(R_z + \frac{u_i}{y_i} \right) \right] \quad (4.11)$$

Αν και το παραπάνω άθροισμα λαμβάνεται για όλα τα σημεία της εικόνας, στην περίπτωση αυτή θα μπορούσε να περιέχει μόνο τα σημεία της εικόνας τα οποία επιλέχθηκαν ως *inliers* κατά το τεστ αποδοχής/απόρριψης που προηγήθηκε. Κι αυτό γιατί τα σημεία αυτά (και μόνο αυτά) είναι που κινούνται σύμφωνα με το μοντέλο της κάμερας. Αυτό θα καθιστούσε την εκτίμηση που θα προέκυπτε πιο αξιόπιστη.

4.9 Πλήρης μεταφορική κίνηση

Σε αυτή τη περίπτωση του τρισδιάστατου μοντέλου κίνησης δεν υπάρχει καθόλου περιστροφή της κάμερας. Αντίθετα, δεν κάνουμε καμία υπόθεση σχετικά με την μεταφορική κίνηση της. Επομένως τα διανύσματα περιστροφής και μετατόπισης θα είναι:

$$\overset{P}{T} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \quad \overset{P}{R} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Άρα ισχύουν οι εξής σχέσεις:

$$u = \frac{T_x - xT_z}{Z}$$

$$v = \frac{T_y - yT_z}{Z}$$

Αν συνδυάσουμε τις 2 παραπάνω εξισώσεις, προκειμένου να απαλείψουμε τον άγνωστο Z (που παριστάνει το βάθος), προκύπτει ότι:

$$u(T_y - yT_z) = v(T_x - xT_z) \Rightarrow av - bu = xv - yu \quad (4.12)$$

όπου:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{T_x}{T_z} \\ \frac{T_y}{T_z} \end{bmatrix}$$

Με βάση τη παραπάνω σχέση (4.12) είναι δυνατόν να εκτιμήσουμε μόνο τις παραμέτρους a και b . Μπορούμε δηλαδή να υπολογίσουμε μόνο τους λόγους των συνιστωσών του διανύσματος της μετατόπισης και όχι ακριβώς τις συνιστώσες αυτές. Το γνωστό δηλαδή πρόβλημα ‘scale factor ambiguity’.

Για την εκτίμηση των παραμέτρων a και b θα βασιστούμε στην εξίσωση οπτικής ροής (4.12) σε συνδυασμό με ένα M-estimator αλγόριθμο. Η ποσότητα που θα ελαχιστοποιηθεί θα είναι:

$$\sum_{i=0}^N g(av - bu - (xv - yu))$$

Το μοντέλο θα γίνει αποδεκτό αν και μόνο αν η παραπάνω ποσότητα είναι τελικά πολύ κοντά στο μηδέν, δηλαδή είναι μικρότερη από κάποιο προκαθορισμένο κατώφλι. Θα πρέπει εδώ να σημειωθεί ότι κατά την διάρκεια της παραπάνω ελαχιστοποίησης, τα σημεία στα οποία το μέτρο της οπτικής ροής είναι πολύ μικρό θα πρέπει να μη λαμβάνονται υπόψη κατά τους υπολογισμούς (ή εναλλακτικά να τους αποδίδεται πολύ μικρό βάρος από τη συνάρτηση επίδρασης του M-estimator). Αυτό είναι αναγκαίο, διότι κατά κανόνα οι υπολογισμοί της οπτικής ροής στα σημεία αυτά περιέχουν σφάλματα εξαιτίας του μικρού μεγέθους των u και v .

4.10 Πλήρης περιστροφή

Στην περίπτωση αυτού του μοντέλου η κάμερα εκτελεί μόνο περιστροφική κίνηση. Άρα τα στιγμιαία διανύσματα της τρισεδιάστατης κίνησης της κάμερας θα είναι:

$$\overset{P}{T} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \overset{P}{R} = \begin{bmatrix} R_x \\ R_y \\ R_z \end{bmatrix}$$

Οι σχέσεις οπτικής ροής και τρισεδιάστατης κίνησης θα είναι:

$$\begin{aligned} u &= R_Y - R_Z y - R_X xy + R_Y x^2 \\ v &= -R_X + R_Z x + R_Y xy - R_X y^2 \end{aligned} \quad (4.13)$$

Δεδομένου ότι στις παραπάνω σχέσεις δεν περιέχεται ο άγνωστος Z (το βάθος), οι συνιστώσες του διανύσματος της περιστροφής είναι δυνατόν να υπολογισθούν ακριβώς. Η εκτίμηση των παραμέτρων θα προκύψει κατά την ελαχιστοποίηση της παρακάτω ποσότητας:

$$\sum_{i=0}^N \left[g\left(u_i - \left(R_Y - R_Z y_i - R_X x_i y_i + R_Y x_i^2\right)\right) + g\left(v_i - \left(-R_X + R_Z x_i + R_Y x_i y_i - R_X y_i^2\right)\right) \right]$$

Αν η παραπάνω ποσότητα είναι τελικά κοντά στο μηδέν, τότε το μοντέλο της πλήρους περιστροφής θα γίνει αποδεκτό, αλλιώς θα απορριφθεί.

4.11 Πλήρης τρισδιάστατη κίνηση

Στην περίπτωση που κανένα από τα προηγούμενα μοντέλα δεν είναι ικανό να περιγράψει την κίνηση της κάμερας, θα καταφύγουμε στη περίπτωση του μοντέλου της πλήρους τρισδιάστατης κίνησης. Είναι το πιο γενικό μοντέλο όπου η κάμερα μπορεί να κινείται μέχρι και 6 βαθμούς ελευθερίας. Εδώ δεν τίθεται θέμα αποδοχής ή απόρριψης. Απλά θα πρέπει να γίνει σωστή εκτίμηση των παραμέτρων της τρισδιάστατης κίνησης κάνοντας χρήση ενός M-estimator αλγορίθμου. Οι λεπτομέρειες σχετικά με τις εξίσωση οπτικής ροής στην περίπτωση αυτή έχουν δοθεί σε προηγούμενο κεφάλαιο.

4.12 Πειραματικά αποτελέσματα - συμπεράσματα

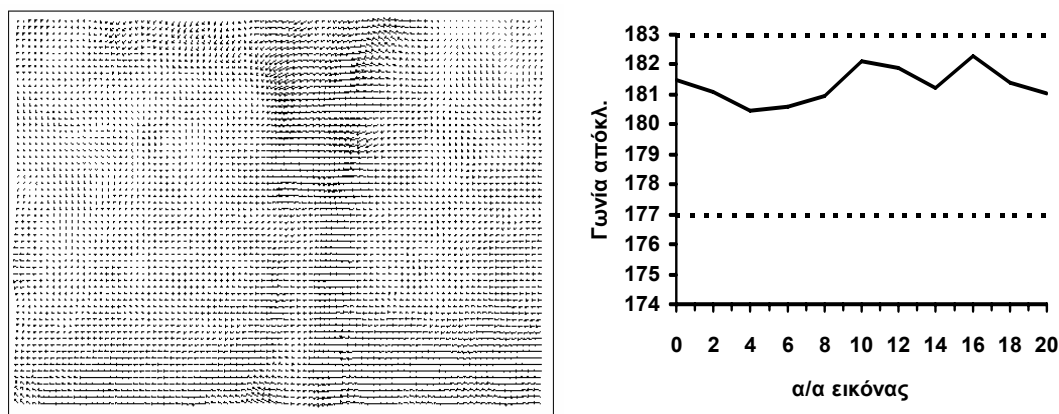
Ο αλγόριθμος εκτίμησης της τρισδιάστατης κίνησης δοκιμάστηκε σε αρκετές ακολουθίες εικόνων. Οι ακολουθίες αυτές περιείχαν είτε πραγματικά είτε συνθετικά δεδομένα. Επιπλέον κατά την διάρκεια των ακολουθιών αυτών η κάμερα παρουσίαζε μια ποικιλία τρισδιάστατων κινήσεων. Για κάθε ζευγάρι εικόνων στο οποίο εφαρμόστηκε η μέθοδος, υπολογίστηκε και ένα πεδίο οπτικής ροής. Στη συνέχεια το πεδίο αυτό της οπτικής ροής αποτέλεσε την είσοδο για τον αλγόριθμο. Θα πρέπει εδώ να σημειωθεί ότι το πεδίο οπτικής ροής δεν είναι απαραίτητο κατ' ανάγκη να είναι πυκνό. Αυτό είναι μια σημαντική ιδιότητα δεδομένου ότι σε πραγματικές ακολουθίες εικόνων είναι αρκετά πιθανό να μη διαθέτουμε αξιόπιστη οπτική ροή για όλα τα σημεία της εικόνας. Για τον υπολογισμό του πεδίου των ταχυτήτων χρησιμοποιήθηκαν διάφορες γνωστές τεχνικές όπως αυτές των Lucas, Uras, Fleet κ.τ.λ. ([14]). Σε κάθε περίπτωση χρησιμοποιήθηκε η μέθοδος που έδωσε τα καλλίτερα αποτελέσματα.

Στο σχήμα 4.5 φαίνεται ένα ζευγάρι εικόνων από την ακολουθία εικόνων Flower Garden ενώ στο σχήμα 4.6 απεικονίζεται η αντίστοιχη οπτική ροή που υπολογίστηκε. Στη ακολουθία αυτή η κάμερα εκτελεί κατά προσέγγιση μια απλή μεταφορική κίνηση παράλληλη προς το επίπεδο

της εικόνας. Δεν υπάρχει καθόλου ανεξάρτητη κίνηση. Κατά την εφαρμογή του αλγορίθμου στην ακολουθία αυτή, το μοντέλο στο οποίο σταμάτησε η διάσχιση της ιεραρχίας των τρισδιάστατων μοντέλων κίνησης ήταν αυτό της μεταφοράς κατά τον οριζόντιο άξονα ενώ όλα σχεδόν τα σημεία της εικόνας επιλέχθηκαν ως inliers σε σχέση με το συγκεκριμένο μοντέλο. Στην περίπτωση του μοντέλου αυτού εκείνο που μετράει για το αν αυτό θα γίνει αποδεκτό ή όχι είναι η γωνία μεταξύ του διανύσματος της οπτικής ροής και του οριζόντιου άξονα. Στο σχήμα 4.6 φαίνεται μια γραφική παράσταση των τιμών της γωνίας αυτής (σε μοίρες) για τα πρώτα 20 frames της ακολουθίας. Ο υπολογισμός έγινε με βάση τη σχέση (4.5) χρησιμοποιώντας την AM-estimator τεχνική και τα αποτελέσματα που προέκυψαν μετατράπηκαν σε μοίρες. Όπως είναι φανερό όλες οι τιμές στο παραπάνω διάγραμμα βρίσκονται σε απόσταση ≈ 3 μοίρες από τη γωνία 180 μοιρών, με αποτέλεσμα το μοντέλο να γίνεται αποδεκτό σε κάθε περίπτωση.



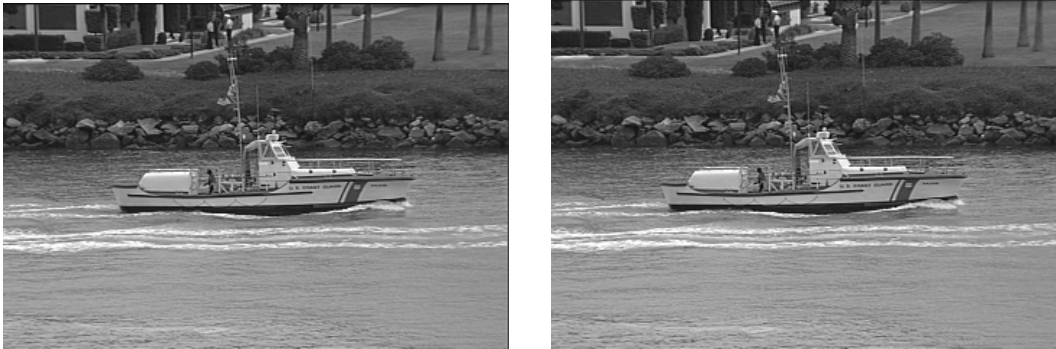
σχήμα 4.5: Frames 5 και 6 από την ακολουθία Flower Garden



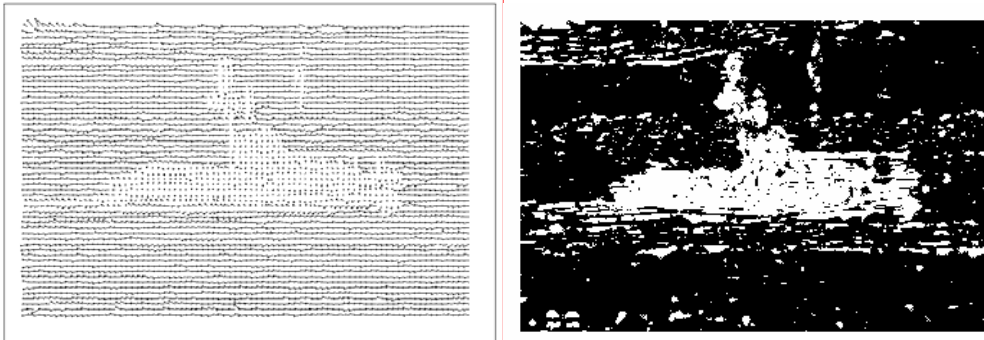
σχήμα 4.6: (αριστερά) οπτική ροή για Flower Garden, (δεξιά) γωνίες απόκλισης για την ακολουθία Flower Garden με χρήση AM-estimator

Στην ακολουθία Coastguard (σχήμα 4.7) έχουμε μια ακόμη περίπτωση μεταφοράς κατά τον οριζόντιο άξονα. Αυτή τη φορά όμως η επιπλέον δυσκολία έγκειται στο γεγονός ότι υπάρχει και ανεξάρτητη κίνηση. Στο σχήμα 4.8 φαίνεται το εκτιμηθέν πεδίο οπτικής ροής ενώ στο ίδιο

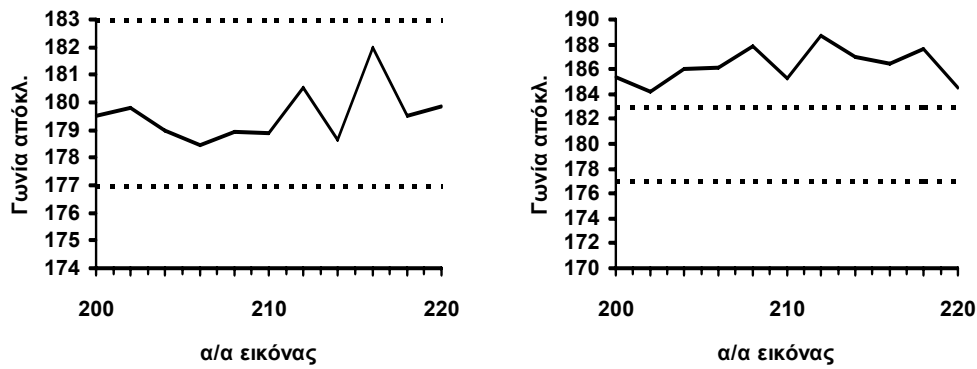
σχήμα απεικονίζονται οι inliers/outliers ως προς την κίνηση της κάμερας. Όλα τα σημεία της μικρής βάρκας, η οποία κινείται ανεξάρτητα, έχουν χαρακτηριστεί σωστά ως outliers και δεν λαμβάνονται υπόψη κατά τους υπολογισμούς. Στο σχήμα 4.9 απεικονίζονται οι τιμές της γωνίας που υπολογίστηκαν από τον αλγόριθμο για τα frames 200-220 της ακολουθίας Coastguard. Η επιτυχία στην ανίχνευση της κίνησης, οφείλεται στην περίπτωση αυτή στον σωστό προσδιορισμό των outliers από τον M-estimator. Στο ίδιο σχήμα απεικονίζονται οι τιμές της γωνίας που θα προέκυπταν αν χρησιμοποιούσαμε για την ίδια ακολουθία τη μέθοδο των ελαχίστων τετραγώνων, μια δηλαδή όχι robust τεχνική. Όπως φαίνεται, οι τιμές της



σχήμα 4.7: frames 200, 201 από την ακολουθία Coastguard



σχήμα 4.8: (αριστερά) πεδίο ταχυτήτων για την ακολουθία Coastguard και (δεξιά) οι αντίστοιχοι inliers ως προς τη κίνηση της κάμερας.



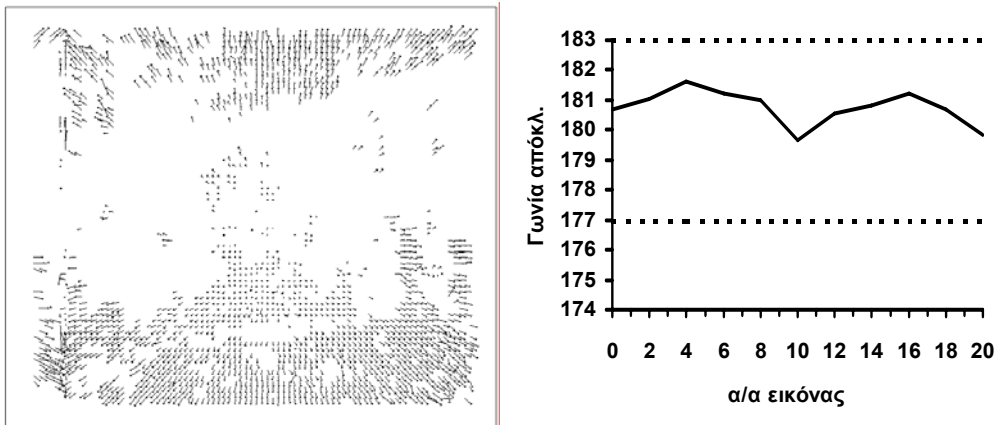
σχήμα 4.9: τιμές γωνίας απόκλισης για την ακολουθία Coastguard με χρήση (αριστερά) M-estimator και (δεξιά) Least-squares.

γωνίας που προκύπτουν είναι πολύ μεγαλύτερες από τις πραγματικές, λόγω της ύπαρξης των outliers, με αποτέλεσμα η ανίχνευση της μεταφοράς κατά τον οριζόντιο άξονα να μην είναι δυνατή.

Ένα διαφορετικό είδος κίνησης για την κάμερα έχουμε στη ακολουθία Kiel (σχήμα 4.10). Στην περίπτωση αυτή η κάμερα εστιάζει προς το κέντρο της εικόνας.

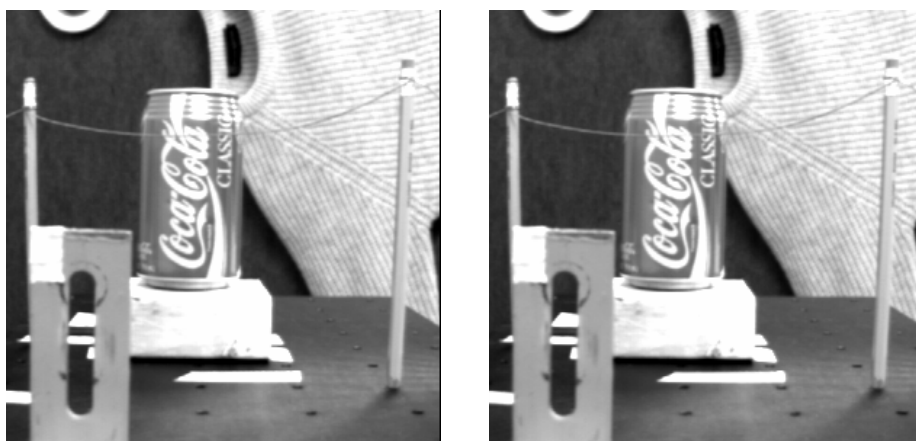


σχήμα 4.10: frames 9 και 10 της ακολουθίας Kiel

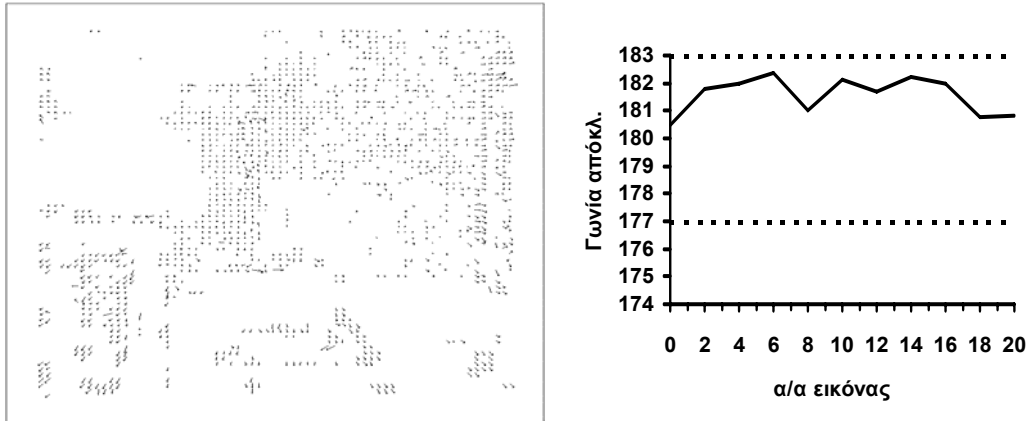


σχήμα 4.11: (αριστερά) αραιό πεδίο οπτικής ροής για την ακολουθία Kiel, (δεξιά) τιμές γωνίας απόκλισης για ακολουθία Kiel με χρήση M-estimator.

Το πεδίο οπτικής ροής που υπολογίστηκε (σχήμα 4.11) δεν είναι πυκνό αλλά διαθέτει οπτική ροή μόνο για το 32.3% των σημείων της εικόνας. Κατά την εφαρμογή του αλγορίθμου το τρισδιάστατο μοντέλο που επιλέχθηκε ήταν αυτό της εστίασης (zooming). Για την αποδοχή ή απόρριψη της εστίασης σημασία έχει η γωνία μεταξύ διανύσματος οπτικής ροής και της ευθείας που συνδέει το σημείο με το κέντρο της εικόνας. Οι τιμές της γωνίας αυτής (σε μοίρες) που υπολογίστηκαν από τον αλγόριθμο με βάση τη σχέση (4.8) για τα πρώτα 20 frames της ακολουθίας, απεικονίζονται στη γραφική παράσταση στο σχήμα 4.11. Δεδομένου ότι όλες οι τιμές αυτές κυμαίνονται σε απόσταση ≈ 3 από τις 180 μοίρες, το μοντέλο εστίασης κρίθηκε ότι ταιριάζει με την κίνηση της κάμερας.



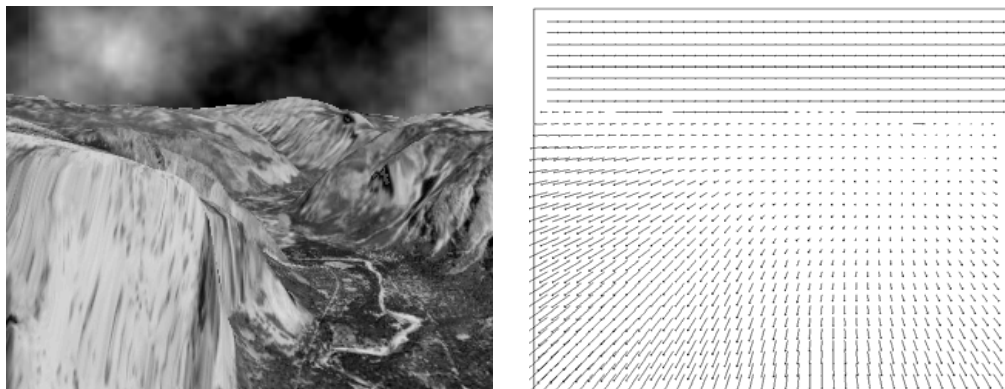
σχήμα 4.12: frames 9 και 10 της ακολουθίας Nasa



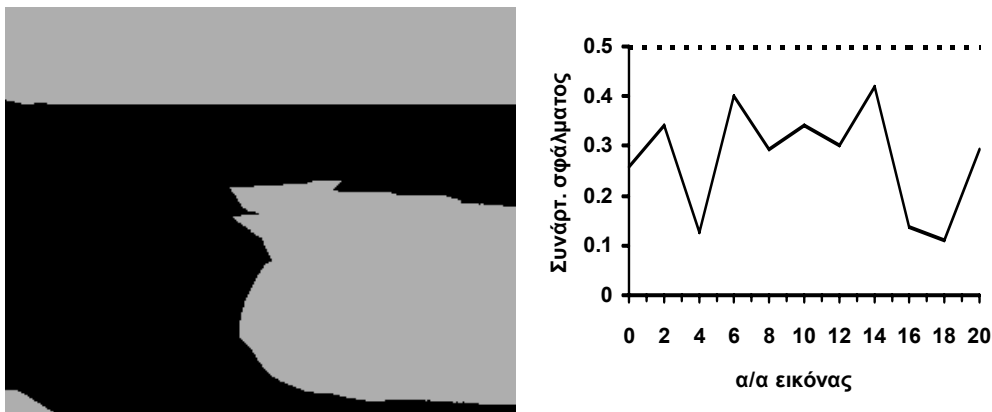
σχήμα 4.13: (αριστερά) πεδίο ταχυτήτων για ακολουθία Nasa, (δεξιά) τιμές γωνίας απόκλισης για τα frames 0 - 20 της ακολουθίας Nasa με χρήση M-estimator.

Μια ακόμη περίπτωση κίνησης εστίασης (zooming) έχουμε και στην ακολουθία Nasa (σχήμα 4.12). Εδώ η κάμερα κινείται κάθετα προς το επίπεδο της εικόνας ενώ επιπλέον δεν υπάρχει ανεξάρτητη κίνηση. Και στην περίπτωση αυτή το εκτιμηθέν πεδίο ταχυτήτων είναι αραιό (σχήμα 4.13). Οι γωνίες που εκτιμήθηκαν από τον αλγόριθμο για τα πρώτα 20 frames της ακολουθίας φαίνονται στο σχήμα 4.13.

Μια πιο περίπλοκη περίπτωση κίνησης παρουσιάζεται στην συνθετική ακολουθία Yosemite (σχήμα 4.14). Η κάμερα εκτελεί πλήρη μεταφορική κίνηση ενώ τα σύννεφα στο πάνω μέρος της εικόνας εκτελούν ανεξάρτητη μεταφορική κίνηση παράλληλα προς το επίπεδο της εικόνας. Αυτή είναι η πρώτη περίπτωση που εξετάζεται όπου εκτός από το να επιβεβαιώσουμε την κίνηση θα πρέπει επιπλέον να υπολογίσουμε και τιμές για παραμέτρους της τρισδιάστατης κίνησης.



σχήμα 4.14: (αριστερά) frame 9 της ακολουθίας Yosemite και (δεξιά) αντίστοιχη οπτική ροή



σχήμα 4.15: (αριστερά) inliers ως προς τη κίνηση της κάμερας για το frame 9 της ακολουθίας Yosemite και (δεξιά) γραφική παράσταση της συνάρτησης σφάλματος του AM-estimator για τα πρώτα 20 frames της ακολουθίας Yosemite.

Κατά την εφαρμογή του αλγορίθμου στις εικόνες 9 και 10 της ακολουθίας Yosemite, απορρίφθηκαν τα μοντέλα της μεταφοράς κατά τον οριζόντιο άξονα, της εστίασης και της περιστροφής γύρω από τον οπτικό άξονα ενώ αντίθετα επιλέχθηκε το μοντέλο της πλήρους μεταφορικής κίνησης. Οι τιμές που υπολογίστηκαν από τον αλγόριθμο για το διάνυσμα μετατόπισης είναι:

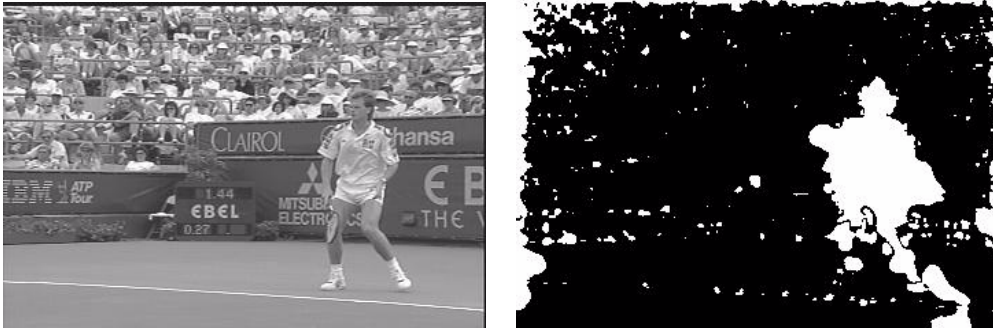
$$(T_x/T_z, T_y/T_z) = (29.582361, 57.725810)$$

Οι αντίστοιχοι inliers απεικονίζονται στο σχήμα 4.15. Τα σημεία που ανήκουν στον ουρανό κατατάχθηκαν σωστά στους outliers και δεν χρησιμοποιήθηκαν κατά τον υπολογισμό των συνιστωσών του διανύσματος της μετατόπισης. Επιπλέον, στους outliers κατατάχθηκε και ένα σύνολο σημείων τα οποία βρίσκονται κοντά στο FOE (Focus Of Expansion), μιας και στο τμήμα αυτό της εικόνας υπάρχει θόρυβος και τα διανύσματα οπτικής ροής περιέχουν σφάλματα

Τέλος, ο αλγόριθμος δοκιμάστηκε στην ακολουθία Stefan (σχήμα 4.16). Στο μεγαλύτερο τμήμα της ακολουθίας αυτής η κάμερα εκτελεί μεταφορική κίνηση κατά τον X και Z άξονα, ενώ σε ορισμένα τμήματα της ακολουθίας η κάμερα εκτελεί επιπλέον και μικρή περιστροφική κίνηση γύρω από τον Y άξονα. Ταυτόχρονα υπάρχει και η ανεξάρτητη κίνηση του τενίστα. Στο σχήμα 4.16 φαίνονται οι inliers/outliers ως προς την κίνηση της κάμερας για τη εικόνα 130 της ακολουθίας, όπως αυτοί εκτιμήθηκαν από τον AM-estimator. Το μοντέλο της ιεραρχίας που επιλέχθηκε για το συγκεκριμένο παράδειγμα ήταν αυτό της πλήρους τρισδιάστατης κίνησης. Οι παράμετροι κίνησης που υπολογίστηκαν ήταν:

$$(T_x/T_z, T_y/T_z) = (-70.833, 2.574) \text{ και } (R_x, R_y, R_z) = (0, -0.138, 0).$$

Όπως μπορεί να παρατηρήσει κανείς, τα σημεία που ανήκουν στο σώμα του τενίστα κατατάχθηκαν στους outliers.



σχήμα 4.16: (αριστερά) frame 130 από την ακολουθία Stefan και (δεξιά) οι αντίστοιχοι inliers ως προς τη κίνηση της κάμερας.

Ένα συμπέρασμα που προέκυψε από τη δοκιμή της μεθόδου στις διάφορες ακολουθίες εικόνων, είναι ότι ο ακριβής υπολογισμός της οπτικής ροής παίζει καθοριστικό ρόλο για την επιτυχία του αλγορίθμου. Αν η οπτική ροή περιέχει ένα πολύ μεγάλο ποσοστό σφαλμάτων, τότε ο αλγόριθμος προσδιορισμού της κίνησης της κάμερας είναι πολύ πιθανό να αποτύχει. Για παράδειγμα κάτι τέτοιο συμβαίνει σε ορισμένα τμήματα της ακολουθίας Stefan. Στην ακολουθία αυτή, τα σφάλματα στην οπτική ροή λαμβάνουν χώρα κυρίως στο πάνω μέρος της εικόνας εκεί όπου υπάρχουν οι κερκίδες με τους θεατές και οφείλονται στη μεγάλη ανομοιομορφία που παρουσιάζεται στο συγκεκριμένο τμήμα της εικόνας. Στις περιπτώσεις αυτές ο αλγόριθμος αδυνατεί να προσδιορίσει επιτυχώς την κίνηση της κάμερας. Υπό την έννοια αυτή ενδιαφέρον θα είχε η επέκταση της μεθόδου, έτσι ώστε ο υπολογισμός των παραμέτρων να μην βασίζεται μόνο στην οπτική ροή, αλλά και σε επιπλέον πληροφορία.

Επιπλέον, ένα ακόμη γεγονός που παρατηρήθηκε ήταν και το ότι σε ορισμένες περιπτώσεις η μεγάλη μεταφορική κίνηση επικρατούσε και 'έκρυβε' την περιστροφική κίνηση της κάμερας ή και το αντίστροφο. Αποτέλεσμα ήταν στις περιπτώσεις αυτές να επιλέγεται λάθος μοντέλο από τον αλγόριθμο. Αυτό οφείλεται στο γεγονός ότι σε αυτές τις περιπτώσεις το πεδίο ταχυτήτων περιέχει συνήθως μεγάλο ποσοστό σφαλμάτων και τα διανύσματα οπτικής ροής εμπεριέχουν στην πλειονότητα τους μόνο το μεγάλο κομμάτι της κίνησης. Ίσως, το πρόβλημα αυτό να εξαλειφθεί αν η ιεραρχία που χρησιμοποιείται εμπλουτισθεί και με άλλα μοντέλα κίνησης. Η εύρεση μιας τέτοιας εμπλουτισμένης ιεραρχίας θα παρουσίαζε ενδιαφέρον.

Επίσης σημαντικό ρόλο στην επιτυχία του αλγορίθμου παίζει και ο σωστός καθορισμός των κατώφλιων που χρησιμοποιούνται για την αποδοχή ή απόρριψη ενός τρισδιάστατου μοντέλου κίνησης. Κατά την παρούσα υλοποίηση τα κατώφλια αυτά υπολογίζονται κατά τρόπο εμπειρικό. Για τις περιπτώσεις των μοντέλων της μεταφοράς κατά τον οριζόντιο άξονα και της εστίασης, όπως φαίνεται και από τα αντίστοιχα διαγράμματα που παρουσιάστηκαν, μια γωνία-

κατώφλι κοντά στις 3 μοίρες φαίνεται να δουλεύει σωστά. Για τις υπόλοιπες περιπτώσεις επιλέχθηκε ένα κατώφλι ίσο με 0.5. Στο σχήμα 4.15 απεικονίζεται η γραφική παράσταση της συνάρτησης σφάλματος του AM-estimator σε σχέση με το κατώφλι αυτό για τα πρώτα 20 frames της ακολουθίας Yosemite. Μια επέκταση της μεθόδου θα ήταν και ο αυτόματος προσδιορισμός των κατωφλιών με βάση τα χαρακτηριστικά του εκάστοτε πεδίου οπτικής ροής.

σε περιπτώσεις όπου η αρχική εκτίμηση και η τελική εκτίμηση απείχαν αισθητά (π.χ. ακολουθία Yosemite), τότε ο M-estimator παρουσίασε αποκλίσεις από τις εκτιμήσεις του AM-estimator. Επιπλέον, ο M-estimator παρουσίασε ευαισθησία και ως προς την παράμετρο κλίμακας υπολογισμού. Όπως έχει ήδη αναφερθεί η παράμετρος αυτή δίνεται από τη σχέση:

$$c = c_0 \cdot \text{median}|n_i|$$

Στις περιπτώσεις των ακολουθιών Flower Garden, Coastguard, Kiel και Nasa η σταθερά c_0 στην παραπάνω σχέση έπρεπε να είναι κοντά στο 1. Αντίθετα για τις υπόλοιπες περιπτώσεις αυτή η σταθερά έπρεπε να βρίσκεται κοντά στη τιμή 3. Αν η c_0 είχε μεγαλύτερη τιμή από ό,τι έπρεπε, τότε ο M-estimator λάμβανε υπόψη του κατά τους υπολογισμούς και ένα μέρος από το σύνολο των outliers, με συνέπεια τα τελικά αποτελέσματα να παρουσιάζουν κάποια σφάλματα.

Τέλος, σε σχέση με τους M-estimator και AM-estimator αξίζει να σημειωθεί ότι σε περιπτώσεις ακολουθιών εικόνων όπου η τελική λύση βρισκόταν κοντά στην αρχική εκτίμηση, οι δύο αυτοί αλγόριθμοι δεν παρουσίασαν διαφορές στα αποτελέσματα που έδωσαν. Τέτοιες ακολουθίες ήταν για παράδειγμα αυτές των Nasa και Kiel. Αντίθετα

Κεφάλαιο 5

5. ΕΥΡΕΣΗ ΣΤΡΩΜΑΤΩΝ ΒΑΘΟΥΣ**5.1 Εισαγωγή**

Το κεφάλαιο αυτό θα ασχοληθεί με το δεύτερο πρόβλημα που αντιμετωπίζεται στην παρούσα εργασία, αυτό δηλαδή της εύρεσης των στρωμάτων βάθους τα οποία παρατηρούνται μέσα σε μια ακολουθία εικόνων. Η εύρεση στρωμάτων βάθους προϋποθέτει ότι έχει ήδη εκτιμηθεί η κίνηση της κάμερας. Το κεφάλαιο θα ξεκινήσει με μια σύντομη περιγραφή του προβλήματος, ενώ παράλληλα θα αναφερθούν κάποιες βασικές έννοιες που έχουν να κάνουν με το συγκεκριμένο ζήτημα. Στη συνέχεια θα εξετάσουμε μια συγκεκριμένη περίπτωση του προβλήματος της εύρεσης των στρωμάτων βάθους, αυτή κατά την οποία η κάμερα εκτελεί απλή μεταφορική κίνηση παράλληλα με το επίπεδο της εικόνας. Στην περίπτωση αυτή θα δείξουμε ότι η εύρεση των στρωμάτων βάθους ανάγεται ουσιαστικά στην τμηματοποίηση του πεδίου των ταχυτήτων σε περιοχές με σταθερή οπτική ροή. Κάθε μια τέτοια περιοχή θα αντιστοιχεί και σε ένα διαφορετικό στρώμα βάθους. Για να πετύχουμε τη τμηματοποίηση της οπτικής ροής, θα χρησιμοποιήσουμε μια μέθοδο τμηματοποίησης η οποία βασίζεται στον *HCF* αλγόριθμο και η οποία αποτελείται από 3 στάδια. Κάθε ένα από τα 3 αυτά στάδια περιγράφεται στη συνέχεια αναλυτικά. Το κεφάλαιο θα κλείσει με την παρουσίαση κάποιων πειραματικών αποτελεσμάτων.

5.2 Καθορισμός του προβλήματος

Με τον όρο *στρώμα βάθους* (*depth layer*) θα εννοούμε ένα σύνολο από σημεία του τρισδιάστατου χώρου τα οποία βρίσκονται σε περίπου σταθερό βάθος από τον παρατηρητή δηλαδή στην περίπτωση μας από την κάμερα. Η εύρεση των στρωμάτων βάθους αποτελεί ένα δύσκολο και σημαντικό πρόβλημα του τομέα της μηχανικής όρασης. Έχει δε προσελκύσει αρκετό ενδιαφέρον τα τελευταία χρόνια. Για τον προσδιορισμό του βάθους μέσα σε μια ακολουθία εικόνων έχουν προταθεί αρκετές τεχνικές. Σε γενικές γραμμές δύο είναι οι πιο διαδεδομένες προσεγγίσεις σχετικά με το ζήτημα αυτό.

Σύμφωνα με την πρώτη από τις προσεγγίσεις αυτές ([24]), για τον προσδιορισμό του βάθους γίνεται χρήση ενός πυκνού πεδίου οπτικής ροής. Στη συνέχεια επιδιώκεται να εξαχθεί πληροφορία για την τρισδιάστατη δομή (3-D structure) με βάση τις εξισώσεις που συνδέουν την οπτική ροή με το βάθος. Το πρόβλημα με τη μέθοδο αυτή έγκειται στο γεγονός ότι αρκετά συχνά οι υπολογισμοί της οπτικής ροής περιέχουν σφάλματα, λόγω του γνωστού 'aperture

problem', με συνέπεια σε ορισμένες περιπτώσεις τα παραγόμενα αποτελέσματα να μην είναι ικανοποιητικά.

Σύμφωνα με τη δεύτερη από τις προσεγγίσεις ([57], [55], [17]), ο προσδιορισμός του βάθους βασίζεται σε *απευθείας (direct)* μεθόδους οι οποίες χρησιμοποιούν πληροφορία που έχει να κάνει αποκλειστικά και μόνο με την φωτεινή ένταση. Οι αλγόριθμοι της κατηγορίας αυτής δεν απαιτούν το ενδιάμεσο στάδιο του υπολογισμού της οπτικής ροής. Βασίζονται δε στην υπόθεση ότι η φωτεινότητα που έχει ένα κινούμενο σημείο από εικόνα σε εικόνα παραμένει σταθερή. Δηλαδή:

$$I(p + u, t + 1) = I(p, t)$$

όπου το $I(p, t)$ συμβολίζει τη φωτεινότητα του σημείου $p=(x, y)$ την χρονική στιγμή t ενώ το u συμβολίζει το διάνυσμα οπτικής ροής για το σημείο p . Το βασικό μειονέκτημα αυτής της κατηγορίας των αλγορίθμων είναι ότι είναι υπερβολικά ευαίσθητες στο θόρυβο και δεν είναι επομένως αρκετά σταθερές (robust).

Υπάρχει και μια τρίτη κατηγορία αλγορίθμων ([56]) οι οποίοι συνδυάζουν στοιχεία και από τις δύο παραπάνω κατηγορίες. Οι αλγόριθμοι αυτοί βασίζονται στο φίλτρο Kalman για τον προσδιορισμό του βάθους, κάνοντας χρήση στοχαστικών δυναμικών μοντέλων. Τα στοχαστικά αυτά μοντέλα εξαρτώνται από την οπτική ροή, και επομένως και οι αλγόριθμοι αυτής της κατηγορίας απαιτούν ένα ενδιάμεσο στάδιο υπολογισμού της οπτικής ροής. Ένα μειονέκτημα των αλγορίθμων της κατηγορίας αυτής είναι το μεγάλο υπολογιστικό κόστος που έχουν.

Η προσέγγιση που θα ακολουθηθεί κατά την παρούσα εργασία θα ακολουθεί το σκεπτικό της πρώτης κατηγορίας αλγορίθμων. Ο αλγόριθμος δηλαδή θα απαιτεί τον υπολογισμό ενός πεδίου οπτικής ροής. Επιπλέον όμως, η εύρεση των στρωμάτων βάθους θα βασιστεί εκτός από την οπτική ροή και στη φωτεινότητα. Με αυτό το τρόπο επιδιώκεται ο αλγόριθμος να γίνει πιο σταθερός και λιγότερο ευαίσθητος στο θόρυβο.

5.3 Μεθοδολογία για τη λύση του προβλήματος

Για να μπορέσουμε να προσεγγίσουμε το πρόβλημα της εύρεσης των στρωμάτων βάθους βασιζόμενοι κυρίως στην οπτική ροή, θα πρέπει πρώτα να γίνει εκτίμηση της τρισδιάστατης κίνησης της κάμερας. Για την εκτίμηση της κίνησης αυτής θα βασιστούμε στην μεθοδολογία που αναπτύχθηκε κατά τα προηγούμενα κεφάλαια της παρούσας εργασίας. Στη συνέχεια, μπορούμε να προσδιορίσουμε τα στρώματα βάθους για όλα εκείνα τα σημεία της εικόνας τα οποία κινούνται σύμφωνα με τη κίνηση της κάμερας ([60], [59]).

Το αν μπορούμε να εντοπίσουμε τα στρώματα βάθους, εξαρτάται όμως και από το είδος της κίνησης της κάμερας μέσα στο χώρο. Για παράδειγμα αν η κάμερα εκτελεί μονάχα περιστροφική κίνηση (και καθόλου μεταφορική), τότε είναι αδύνατο να εξάγουμε οποιοδήποτε συμπέρασμα για το βάθος βασιζόμενοι μονάχα στην οπτική ροή και χωρίς καμία επιπλέον πληροφορία. Στην περίπτωση αυτή οι σχέσεις που συνδέουν την οπτική ροή με τις συντεταγμένες των σημείων της εικόνας, θεωρώντας γνωστές τις παραμέτρους της κίνησης της κάμερας, ανάγονται στην παρακάτω μορφή:

$$\begin{aligned} u &= R_Y - R_Z y - R_X xy + R_Y x^2 \\ v &= -R_X + R_Z x + R_Y xy - R_X y^2 \end{aligned} \quad (5.1)$$

Δεδομένου ότι στις παραπάνω σχέσεις δεν εμπεριέχεται το βάθος, είναι φανερό ότι η εύρεση των στρωμάτων βάθους είναι αδύνατη.

Επομένως, θα θεωρήσουμε ως αληθή την υπόθεση ότι η μεταφορική κίνηση της κάμερας δεν είναι μηδενική. Στη γενική περίπτωση όπου τα διανύσματα περιστροφής και μετατόπισης της κάμερας είναι $R=(R_X, R_Y, R_Z)$ και $T=(T_X, T_Y, T_Z)$, τότε θα ισχύουν οι παρακάτω εξισώσεις:

$$\begin{aligned} u &= \frac{T_X - xT_Z}{Z} + R_Y - R_Z y - R_X xy + R_Y x^2 \\ v &= \frac{T_Y - yT_Z}{Z} - R_X + R_Z x + R_Y xy - R_X y^2 \end{aligned} \quad (5.6)$$

Ή ισοδύναμα:

$$\begin{aligned} u - (R_Y - R_Z y - R_X xy + R_Y x^2) &= \frac{T_X - xT_Z}{Z} \\ v - (-R_X + R_Z x + R_Y xy - R_X y^2) &= \frac{T_Y - yT_Z}{Z} \end{aligned} \quad (5.2)$$

Με δεδομένο ότι είναι γνωστές οι παράμετροι της τρισδιάστατης κίνησης, προκύπτει άμεσα ότι τα αριστερά μέλη των παραπάνω εξισώσεων είναι γνωστά. Επομένως οι παραπάνω σχέσεις θα μπορούσαν να θεωρηθούν σαν οι εξισώσεις οπτικής ροής κατά τη περίπτωση όπου τα διανύσματα κίνησης της κάμερας θα ήταν: $T=(T_X, T_Y, T_Z)$ και $R=(0, 0, 0)$. Άρα, ο προσδιορισμός των στρωμάτων βάθους στη γενική περίπτωση μπορεί να αναχθεί στη περίπτωση όπου έχουμε μόνο μεταφορική κίνηση.

Στην περίπτωση αυτή ας θεωρήσουμε τη μία από τις 2 συνιστώσες του διανύσματος της οπτικής ροής, έστω τη u . Ισχύει ότι:

$$u = \frac{T_x - x T_z}{Z} \Leftrightarrow u = \frac{\frac{T_x}{T_z} - x}{\frac{Z}{T_z}} \Leftrightarrow \frac{u}{\frac{T_x}{T_z} - x} = \frac{T_z}{Z} \Leftrightarrow \frac{u}{a - x} = \frac{T_z}{Z} \quad (5.3)$$

Από τη παραπάνω σχέση προκύπτει ότι η ποσότητα $u/(a-x)$ παραμένει σταθερή αν και μόνο αν το βάθος παραμένει σταθερό. Επομένως για να βρούμε τα στρώματα βάθους αρκεί να διαμερίσουμε την οπτική ροή σε περιοχές όπου η παραπάνω ποσότητα έχει σταθερή τιμή. Άρα το πρόβλημα έχει αναχθεί σε ένα πρόβλημα τμηματοποίησης.

Για λόγους απλότητας θα θεωρήσουμε τη περίπτωση όπου η κάμερα εκτελεί καθαρή μεταφορική κίνηση παράλληλα προς το επίπεδο της εικόνας (με βάση την εξίσωση (5.3) οι άλλες περιπτώσεις μπορούν να αντιμετωπιστούν με παρόμοιο τρόπο). Τότε θα ισχύει ότι:

$$u = \text{σταθερή} \diamond Z = \text{σταθερό}$$

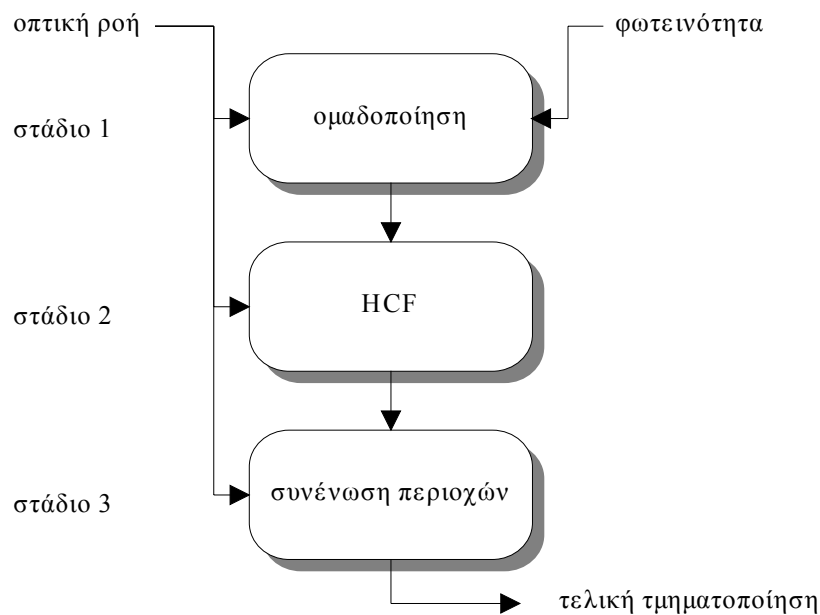
Επομένως, στη περίπτωση αυτή το πρόβλημα της εύρεσης των στρωμάτων βάθους έχει ουσιαστικά αναχθεί στο πρόβλημα της διαμέρισης της οπτικής ροής σε περιοχές όπου η u -συνιστώσα είναι σταθερή. Ο τρόπος εύρεσης των περιοχών αυτών θα αποτελέσει το αντικείμενο μελέτης της επόμενης παραγράφου.

5.4 Αλγόριθμος τμηματοποίησης

Για την διαμέριση του πεδίου της οπτικής ροής σε περιοχές με σταθερή u -συνιστώσα, θα χρησιμοποιηθεί ένας αλγόριθμος τμηματοποίησης. Μια σχηματική αναπαράσταση του αλγόριθμου αυτού φαίνεται στο σχήμα 5.1. Όπως μπορεί κανείς να παρατηρήσει από το σχήμα αυτό, ο αλγόριθμος αποτελείται ουσιαστικά από 3 στάδια:

- ομαδοποίηση με βάση την κίνηση και την φωτεινότητα
- εφαρμογή HCF αλγορίθμου
- και τέλος συνένωση περιοχών

Κατά το πρώτο στάδιο θα δημιουργηθεί ένα αρχικό σύνολο ετικετών. Σε κάθε ετικέτα θα αντιστοιχηθεί αργότερα μια περιοχή μέσα στο πεδίο οπτικής ροής. Αυτό που θέλουμε να επιτύχουμε είναι να επιλέξουμε με τέτοιο τρόπο τις ετικέτες, έτσι ώστε οι περιοχές που θα σχηματιστούν στη συνέχεια να αντιστοιχούν σε διαφορετικά στρώματα βάθους. Κατ' αυτό το τρόπο θα έχουμε ουσιαστικά ανάγει το πρόβλημα της εύρεσης των στρωμάτων βάθους σε ένα **πρόβλημα επιλογής ετικετών**.



σχήμα 5.1: αλγόριθμος τμηματοποίησης

Το σύνολο από ετικέτες θα δημιουργηθεί χρησιμοποιώντας πληροφορία που βασίζεται εκτός από την οπτική ροή και στην φωτεινότητα. Οι λόγοι για τους οποίους συμβαίνει αυτό θα εξηγηθούν αναλυτικότερα στη συνέχεια.

Το δεύτερο στάδιο του αλγόριθμου τμηματοποίησης αποτελεί και το σημαντικότερο κομμάτι του αλγορίθμου. Κατά τη διάρκεια του, θα γίνει προσπάθεια να βρεθούν οι περιοχές που αντιστοιχούν σε κάθε μια από τις ετικέτες οι οποίες δημιουργήθηκαν κατά τη προηγούμενη φάση. Για το σκοπό αυτό θα χρησιμοποιηθεί ο επαναληπτικός αλγόριθμος **HCF (Highest Confidence First)**.

Από τις περιοχές που θα προκύψουν μετά την εφαρμογή του *HCF* αλγόριθμου, είναι αρκετά πιθανό να υπάρχουν περισσότερες περιοχές από όσα είναι τα στρώματα βάθους στην πραγματικότητα. Οι περιοχές αυτές που πλεονάζουν θα εξαιρεθούν κατά το τελευταίο στάδιο του αλγόριθμου. Η εξάλειψη θα γίνει με τη συνένωση όλων των περιοχών που αντιστοιχούν στο ίδιο στρώμα βάθους. Οι περιοχές που θα προκύψουν ύστερα από τη συνένωση αυτή θα αποτελούν και την τελική τμηματοποίηση του πεδίου της οπτικής ροής. Στη συνέχεια ακολουθεί μια πιο αναλυτική περιγραφή για κάθε ένα από τα 3 στάδια που απαρτίζουν τον αλγόριθμο τμηματοποίησης.

5.4.1 Στάδιο 1^ο: Ομαδοποίηση με βάση κίνηση και φωτεινότητα

Κατά το στάδιο αυτό, όπως έχει ήδη προαναφερθεί, θα σχηματιστεί ένα σύνολο από ετικέτες. Κάθε ετικέτα από το σύνολο αυτό θα αντιστοιχεί και σε ένα στρώμα βάθους μέσα στην εικόνα, δηλαδή σε μια περιοχή του πεδίου οπτικής ροής με σταθερή u -συνιστώσα. Για την επιλογή του συνόλου των ετικετών θα χρησιμοποιηθεί κατ' αρχήν πληροφορία που βασίζεται στην οπτική ροή.

Ωστόσο ο ακριβής και αξιόπιστος υπολογισμός της οπτικής ροής αποτελεί ένα πολύ δύσκολο έργο. Έτσι στην πλειονότητα των περιπτώσεων που έχουμε να κάνουμε με πραγματικές ακολουθίες εικόνων, το πεδίο της οπτικής ροής περιέχει και ένα ποσοστό σφαλμάτων, το οποίο σε ορισμένες περιπτώσεις είναι και αρκετά μεγάλο. Τα σφάλματα αυτά μπορεί να οφείλονται σε πολλούς λόγους, όπως για παράδειγμα στο θόρυβο που έχουν τα δεδομένα, στις αποεπικαλύψεις των αντικειμένων που μετέχουν στην υπό εξέταση ακολουθία κτλ.

Επομένως για να αντιμετωπιστεί το παραπάνω πρόβλημα, κατά την δημιουργία του συνόλου των ετικετών θα χρησιμοποιηθεί και πληροφορία που προέρχεται από την φωτεινότητα ([59]). Αν και αυτό κάνει τον αλγόριθμο τμηματοποίησης πιο σταθερό σε σχέση με τις τελικές εκτιμήσεις, από την άλλη θα έχει ως συνέπεια να δημιουργηθούν πλεονάζουσες περιοχές. Οι περιοχές αυτές θα εξαλειφθούν σε μετέπειτα στάδιο του αλγορίθμου.

Έστω L_0 και L_1 τα δύο σύνολα ετικετών λόγω οπτικής ροής και φωτεινότητας αντίστοιχα. Τότε το τελικό σύνολο ετικετών L που θα δημιουργηθεί θα είναι το καρτεσιανό γινόμενο των παραπάνω 2 συνόλων. Δηλαδή:

$$L = L_0 \times L_1$$

Επομένως κάθε ετικέτα από το σύνολο αυτό θα είναι ουσιαστικά ένα διάνυσμα με 2 συνιστώσες και θα περιέχει πληροφορία και λόγω οπτικής ροής αλλά και λόγω φωτεινότητας. Ορίζουμε ως ενέργεια που θα έχει μια τέτοια ετικέτα, το άθροισμα των ενεργειών των δύο συνιστωσών της, αφού προηγουμένως η ενέργεια της δεύτερης συνιστώσας έχει πολλαπλασιαστεί με ένα παράγοντα λ . Ο παράγοντας αυτός καθορίζει τη σημασία που θα έχει η φωτεινότητα κατά τη τμηματοποίηση και στη περίπτωση μας θα είναι πάντα μικρότερος της μονάδας. Άρα ο επόμενος στόχος μας θα είναι να προσδιορίσουμε τα δύο αυτά σύνολα από ετικέτες: τις ετικέτες λόγω οπτικής ροής και τις ετικέτες λόγω φωτεινότητας.

5.4.1.1 Προσδιορισμός ετικετών λόγω οπτικής ροής

Για τον προσδιορισμό των ετικετών λόγω οπτικής ροής θα προσεγγίσουμε τη u -συνιστώσα της οπτικής ροής με μια μίξη Γκαουσιανών κατανομών. Στη περίπτωση αυτή η συνάρτηση

πυκνότητας πιθανότητας για τις τιμές της μ -συνιστώσας της οπτικής ροής θα δίνεται από την σχέση:

$$\sum_{i=0}^n \frac{P_i}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \quad (5.4)$$

όπου $N(\mu_i, \sigma_i^2)$ είναι η i -οστή κατανομή Γκάους και P_i συμβολίζει την εκ των προτέρων πιθανότητα της i -οστής κατανομής.

Διαθέτοντας τώρα την μίξη αυτή των Γκαουσιανών κατανομών, θα αντιστοιχίσουμε μία ετικέτα σε κάθε μία από τις κατανομές αυτές. Άρα το πρόβλημα της εύρεσης των ετικετών λόγω οπτικής ροής ανάγεται ουσιαστικά στο πρόβλημα του προσδιορισμού των παραμέτρων που καθορίζουν την μίξη κατανομών. Ο προσδιορισμός αυτός θα επιτευχθεί έχοντας σαν κύριο εργαλείο τη στατιστική ανάλυση που θα γίνει πάνω στο ιστόγραμμα της μ -συνιστώσας της οπτικής ροής.

Αρχική προϋπόθεση για την λύση του προβλήματος είναι ο καθορισμός του αριθμού των Γκαουσιανών κατανομών που μετέχουν στην μίξη. Κατά την παρούσα υλοποίηση του αλγορίθμου ο αριθμός αυτός προσδιορίζεται εμπειρικά και όχι με κάποια αυτόματη διαδικασία. Πιο συγκεκριμένα, η τιμή που θα έχει, θα καθοριστεί με βάση το παρατηρούμενο ιστόγραμμα οπτικής ροής. Θα πρέπει εδώ να σημειωθεί ότι αρχικά εφαρμόζεται μια διαδικασία λείανσης πάνω στο ιστόγραμμα της οπτικής ροής, η οποία αποβλέπει στην μείωση της επίδρασης του θορύβου στην όλη διαδικασία.

Έτσι στο ιστόγραμμα αυτό αναζητούμε τοπικά μέγιστα τα οποία μπορούμε να υποθέσουμε ότι αρχικά αντιστοιχούν σε μέσες τιμές διαφορετικών κατανομών Γκάους. Η προσπέλαση των τοπικών μεγίστων γίνεται ιεραρχικά και επιπλέον απαιτούμε να υπάρχει ελάχιστο κάτω φράγμα στην απόσταση μεταξύ κάθε υποψήφιας τιμής και των ήδη επιλεχθέντων μεγίστων. Με αυτό το τρόπο αποφεύγεται η διαίρεση ίδιων πληθυσμών που παρουσιάζουν μεγάλη πιθανότητα. Στην συνέχεια με τη χρήση κριτηρίου Μέγιστης Πιθανοφάνειας για την εκτίμηση των παραμέτρων αυτών, μπορούμε να εκτιμήσουμε τις άγνωστες παραμέτρους (P_i, σ_i, μ_i) για κάθε κατανομή (δες παράρτημα Β).

Αφού εφαρμοστεί η παραπάνω διαδικασία και έχουν προσδιορισθεί πλήρως οι παράμετροι της μίξης κατανομών Γκάους, η ενέργεια που θα έχει η ετικέτα (μ_i, σ_i) που αντιστοιχεί στην Γκαουσιανή κατανομή με μέση τιμή μ_i και διασπορά σ_i θα είναι:

$$f_i(x) = \frac{(x - \mu_i)^2}{\sigma_i^2} + \ln \sigma_i^2 \quad (5.5)$$

Η ενέργεια αυτή θα χρησιμοποιηθεί κατά το επόμενο στάδιο του αλγορίθμου τμηματοποίησης.

5.4.1.2 Προσδιορισμός ετικετών λόγω φωτεινότητας

Για τον προσδιορισμό των ετικετών λόγω φωτεινής έντασης θα ακολουθηθεί μια μέθοδος παρόμοια με αυτή της προηγούμενης παραγράφου. Θα βασιστούμε δηλαδή και πάλι στη στατιστική ανάλυση του ιστογράμματος της φωτεινότητας. Αυτή τη φορά όμως προκειμένου να αποφύγουμε το μεγάλο υπολογιστικό κόστος που έχει ο προσδιορισμός των παραμέτρων της μίξης κατανομών Γκάους, θα προσπαθήσουμε να προσεγγίσουμε την φωτεινότητα με τμηματικά ομοιόμορφες συναρτήσεις, πραγματοποιώντας ουσιαστικά ένα κβαντισμό της φωτεινής έντασης. Θα χρησιμοποιηθεί δηλαδή ένα ομοιόμορφο μοντέλο με σταθερές τιμές στα διάφορα κβαντισμένα διαστήματα για τις συναρτήσεις πυκνότητας πιθανότητας.

Για τον πλήρη προσδιορισμό του κβαντισμού της φωτεινότητας απαιτείται να υπολογισθούν δύο στοιχεία:

- το πλήθος των σημείων απόφασης (του άξονα των πραγματικών αριθμών)
- καθώς και τα ίδια τα σημεία αυτά.

Στη συνέχεια θα αντιστοιχίσουμε μια ετικέτα σε κάθε ένα σημείο απόφασης.

Για τον κβαντισμό της φωτεινής έντασης θα χρησιμοποιήσουμε τον αλγόριθμο των Lloyd-Max (δες παράρτημα Β). Η όλη διαδικασία έχει ως εξής: αρχικά επιλέγεται το πλήθος των επιπέδων απόφασης με βάση πιθανά μέγιστα που παρατηρούνται στο ιστόγραμμα της φωτεινότητας. Στη συνέχεια εφαρμόζεται ο αλγόριθμος Lloyd-Max μέχρις ότου εξασφαλιστεί η σύγκλιση. Αν το συνολικό μέσο τετραγωνικό σφάλμα είναι πάνω από κάποιο εμπειρικά καθοριζόμενο κατώφλι, το επίπεδο με το μεγαλύτερο τοπικό μέσο τετραγωνικό σφάλμα υποδιαιρείται και ο αλγόριθμος Lloyd-Max εφαρμόζεται με ένα επίπεδο κβαντισμού παραπάνω. Η διαδικασία αυτή επαναλαμβάνεται μέχρι το συνολικό μέσο τετραγωνικό σφάλμα να πάρει μια τιμή κάτω από το δοσμένο κατώφλι.

Αν είναι $L_i = \{k_0, \dots, k_m\}$ τα επίπεδα κβαντισμού που τελικά υπολογίστηκαν, τότε η ετικέτα που αντιστοιχεί στο i -οστό από τα επίπεδα αυτά θα έχει ενέργεια:

$$f_i(x) = |x - k_i| \quad (5.7)$$

Είναι φανερό ότι η παραπάνω ενέργεια εκφράζει την απόσταση της ετικέτας από το αντίστοιχο σημείο κβαντισμού. Θα πρέπει εδώ να σημειωθεί ότι προτού εφαρμοστεί η διαδικασία του κβαντισμού, πραγματοποιείται μια διαδικασία λείανσης της φωτεινής έντασης της εικόνας.

5.4.2 Στάδιο 2^ο: Εφαρμογή HCF αλγορίθμου

Κατά το στάδιο αυτό θα λάβει χώρα η πραγματική τμηματοποίηση της εικόνας με βάση το βάθος. Υπό αυτή την έννοια αποτελεί το σημαντικότερο στάδιο της όλης διαδικασίας. Για το σκοπό αυτό, ένα ολικό κριτήριο τμηματοποίησης θα χρησιμοποιηθεί το οποίο θα βασίζεται στο μοντέλο ενός τυχαίου πεδίου Markov. Για την ελαχιστοποίηση της συνάρτησης ενέργειας που θα προκύψει, θα χρησιμοποιηθεί ένας επαναληπτικός ντετερμινιστικός αλγόριθμος χαλάρωσης, ο οποίος είναι γνωστός και με το όνομα *HCF (Highest Confidence First)* ([53]). Μια πιο λεπτομερής περιγραφή του HCF αλγορίθμου περιέχεται στο παράρτημα Α.

Η είσοδος που δέχεται ο HCF αλγόριθμος είναι ένα σύνολο από ετικέτες μαζί με μια συνάρτηση ενέργειας για κάθε μία από τις ετικέτες αυτές. Και η έξοδος του αλγορίθμου είναι ένα σύνολο περιοχών, μια περιοχή για κάθε ετικέτα. Στη συγκεκριμένη περίπτωση σαν ετικέτες εισόδου θα χρησιμοποιηθεί το σύνολο από ετικέτες που κατασκευάστηκε κατά το προηγούμενο στάδιο του αλγορίθμου.

Κατά τη διάρκεια του HCF αλγορίθμου τα σημεία της εικόνας εξετάζονται με βάση κάποια καθορισμένη σειρά επίσκεψης. Η σειρά αυτή επίσκεψης παίζει σημαντικό ρόλο στη σύγκλιση του αλγορίθμου αλλά και στα τελικά αποτελέσματα που θα δώσει. Σε κάθε σημείο επιλέγεται εκείνη η ετικέτα (από το σύνολο των αρχικών ετικετών), η οποία προσφέρει τη μεγαλύτερη τοπική μείωση της ολικής συνάρτησης ενέργειας. Επιπλέον όμως κατά την επιλογή της ετικέτας για ένα συγκεκριμένο σημείο, λαμβάνεται υπόψη και το είδος των ετικετών που έχει επιλεγεί για τα γειτονικά σημεία. Με αυτό το τρόπο εξασφαλίζεται η ομοιογένεια της τελικής τμηματοποίησης. Αρχικά τα σημεία της εικόνας δεν διαθέτουν κάποια ετικέτα. Στο τέλος όλα τα σημεία με μια συγκεκριμένη ετικέτα θα αποτελούν μια περιοχή.

Ένα μειονέκτημα του HCF αλγορίθμου είναι ότι θεωρητικά μπορεί να συγκλίνει σε κάποιο τοπικό ελάχιστο, κάτι το οποίο άλλωστε αποτελεί χαρακτηριστικό όλων των επαναληπτικών αλγορίθμων χαλάρωσης. Από την άλλη ένα σημαντικό πλεονέκτημα του αλγορίθμου είναι και το γεγονός ότι η πλειονότητα των υπολογισμών που απαιτούνται κατά τη διάρκεια του είναι αρκετά τοπικοί. Αυτό σημαίνει ότι ένα μεγάλο τμήμα του HCF αλγορίθμου παρουσιάζεται ιδιαίτερα πρόσφορο για παραλληλοποίηση, κάτι το οποίο θα έχει ως συνέπεια μια σημαντική επιτάχυνση της όλης διαδικασίας. Οι περιοχές της εικόνας που θα δημιουργήσει ο HCF αλγόριθμος θα δοθούν σαν είσοδο για την επόμενη φάση του αλγορίθμου τμηματοποίησης.

5.4.3 Στάδιο 3^ο: συνένωση περιοχών

Κατά τη διάρκεια της φάσης αυτής του αλγορίθμου θα εξαλειφθούν όλες οι περιοχές που ο HCF δημιούργησε και οι οποίες δεν αντιστοιχούν σε πραγματικά στρώματα βάθους. Τέτοιες περιοχές

είναι δυνατό να υπάρχουν λόγω του γεγονότος ότι μέχρι αυτό το σημείο η τμηματοποίηση δεν βασίστηκε μόνο στην οπτική ροή αλλά επιπλέον και στην φωτεινότητα. Για την εξάλειψη των περιοχών αυτών είναι αναγκαία μια διαδικασία συνένωσης. Όλες οι περιοχές που αντιστοιχούν στο ίδιο στρώμα βάθους πρέπει να ενωθούν σε μία περιοχή. Αυτή τη φορά το κριτήριο με το οποίο θα πραγματοποιηθεί η συνένωση δύο περιοχών θα πρέπει να βασίζεται μονάχα στην οπτική ροή. Πιο συγκεκριμένα η διαδικασία που θα ακολουθηθεί είναι η εξής:

Θα ορίσουμε μια μετρική απόσταση μεταξύ δύο περιοχών. Η απόσταση αυτή θα βασίζεται αποκλειστικά στην οπτική ροή και θα εκφράζει την ομοιότητα δύο περιοχών σε σχέση με την οπτική ροή που περιέχουν. Όσο μικρότερη είναι η απόσταση αυτή, τόσο μεγαλύτερη είναι η ομοιότητα λόγω οπτικής ροής μεταξύ των δύο περιοχών και άρα τόσο μεγαλύτερη είναι η πιθανότητα οι δύο αυτές περιοχές να αντιστοιχούν στο ίδιο στρώμα βάθους.

Στη συνέχεια ταξινομούμε όλα τα δυνατά ζεύγη περιοχών σε μια στοίβα με βάση την απόσταση που έχουν μεταξύ τους. Τα ζευγάρια με τη μικρότερη απόσταση τοποθετούνται στη κορυφή της στοίβας ενώ τα ζευγάρια με τη μεγαλύτερη απόσταση τοποθετούνται στο τέλος της στοίβας. Σε κάθε επανάληψη του αλγορίθμου ενώνουμε το ζευγάρι περιοχών που βρίσκεται στην κορυφή, υπό την προϋπόθεση ότι η απόσταση του ζευγαριού αυτού θα είναι μικρότερη από κάποιο δοσμένο κατώφλι. Η διαδικασία αυτή συνεχίζεται μέχρι να μην μπορούμε να ενώσουμε άλλες περιοχές. Οι περιοχές που θα μείνουν θα αποτελούν τις τελικές περιοχές της τμηματοποίησης. Μια περιγραφή του αλγορίθμου συνένωσης περιοχών σε ψευδοκώδικα φαίνεται στο σχήμα 5.2.

Σημαντικό ρόλο και στην όλη διαδικασία θα παίζει και η επιλογή της μετρικής απόστασης. Έστω R_1 και R_2 δύο περιοχές. Και ας συμβολίσουμε με n_1 και n_2 τα μεγέθη των περιοχών αυτών (σε σημεία) και με m_1, m_2, s_1, s_2 τις μέσες τιμές και τις διασπορές της u -συνιστώσας της οπτικής ροής στις περιοχές αυτές. Έστω επίσης ότι $R = R_1 \parallel R_2$ είναι η περιοχή που προκύπτει ύστερα από τη συνένωση των δύο περιοχών και ότι m και s είναι η μέση τιμή και η διασπορά της οπτικής ροής αντίστοιχα στην νέα αυτή περιοχή. Στην περίπτωση αυτή οι παρακάτω αποστάσεις μπορούν να ορισθούν:

```

/* υπολόγισε αποστάσεις για όλα τα δυνατά ζευγάρια
   και βάλε τις στη στοίβα */
for(i = 0; i < n; i++)
    for(j = 0; j < n; j++)
        insert_region_pair_in_stack(region(i), region(j));

/* ξεκίνησε επαναληπτική διαδικασία */
top_dist = stack.dist;
while (top_dist < threshold)
{
    /* δημιούργησε τη περιοχή που προκύπτει από τη
       συνένωση */
    new_region = merge(stack.region0, stack.region1);

    /* αφαίρεσε από τη στοίβα όλα τα ζεύγη που
       περιέχουν περιοχές που συνενώθηκαν */
    remove_region_pairs(stack.region0, stack.region1);

    /* βάλε στη στοίβα αποστάσεις για τη νέα περιοχή */
    for(j = 0; j < n; j++)
        insert_region_pair_in_stack(new_region, region(j));
}

```

σχήμα 5.2: συνένωση περιοχών

- $d = |m_1 - m_2|$

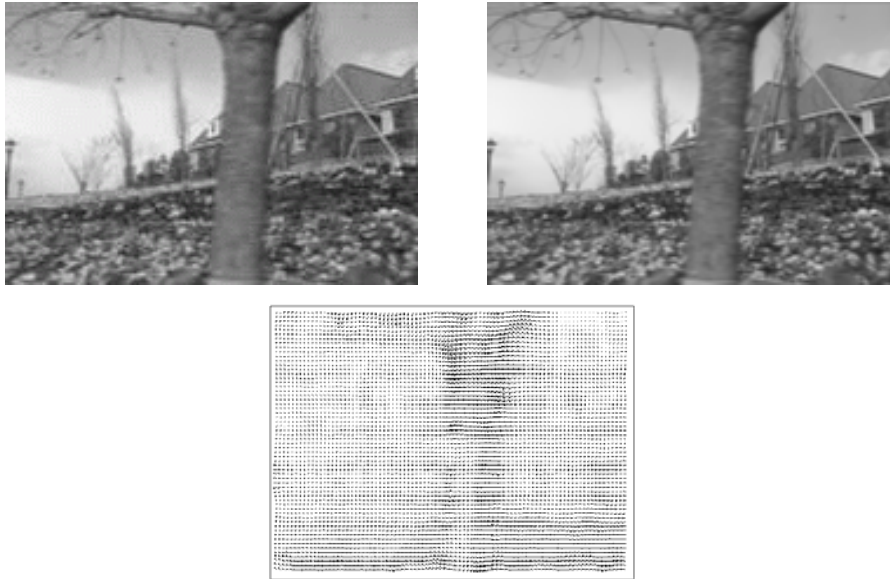
- $d = \frac{(n_1 + n_2)(m_1 - m_2)^2}{n_1 s_1^2 + n_2 s_2^2} = \frac{ns^2}{n_1 s_1^2 + n_2 s_2^2} - 1$ (Fisher distance)

Το υπολογιστικό κόστος του τμήματος αυτού του αλγορίθμου εστιάζεται στον υπολογισμό των αποστάσεων για όλα τα δυνατά ζευγάρια περιοχών. Δηλαδή το κόστος αυτό είναι $O(n^2)$, όπου το n συμβολίζει το πλήθος των περιοχών. Δεδομένου ότι στην πλειονότητα των περιπτώσεων το n αυτό δεν είναι μεγάλο, αυτό έχει σαν συνέπεια και η χρονική πολυπλοκότητα του τμήματος αυτού να μην είναι μεγάλη.

5.5 Πειραματικά αποτελέσματα - συμπεράσματα

Η μέθοδος της εύρεσης των στρωμάτων βάθους εφαρμόστηκε στην ακολουθία Flower Garden. Στο σχήμα 5.3 φαίνονται τα frames 5 και 6 από την ακολουθία αυτή καθώς και το αντίστοιχο πεδίο οπτικής ροής. Όλα τα παραπάνω θα αποτελέσουν την είσοδο στον αλγόριθμο μας. Στην συγκεκριμένη ακολουθία η κάμερα εκτελεί απλή μεταφορική κίνηση παράλληλη προς το επίπεδο προβολής. Επιπλέον, δεν υπάρχει ανεξάρτητη κίνηση μέσα στην εικόνα. Η ανίχνευση της κίνησης της κάμερας θα γίνει με τη βοήθεια της μεθόδου που αναπτύχθηκε στα προηγούμενα κεφάλαια της παρούσας εργασίας. Αφού ανιχνευθεί η κίνηση αυτή, στη συνέχεια μπορούμε να επιχειρήσουμε να εξάγουμε τα στρώματα βάθους εφαρμόζοντας τον αλγόριθμο τμηματοποίησης.

Όπως μπορεί κανείς να παρατηρήσει, 3 είναι τα στρώματα βάθους που διακρίνονται στην εικόνα: το στρώμα του ουρανού, το στρώμα που περιέχει τα σπίτια και το στρώμα που περιέχει το πιο κοντινό στη κάμερα δέντρο. Για την εφαρμογή του αλγορίθμου τμηματοποίησης θα πρέπει αρχικά να προσδιορισθούν οι ετικέτες που θα χρησιμοποιηθούν.

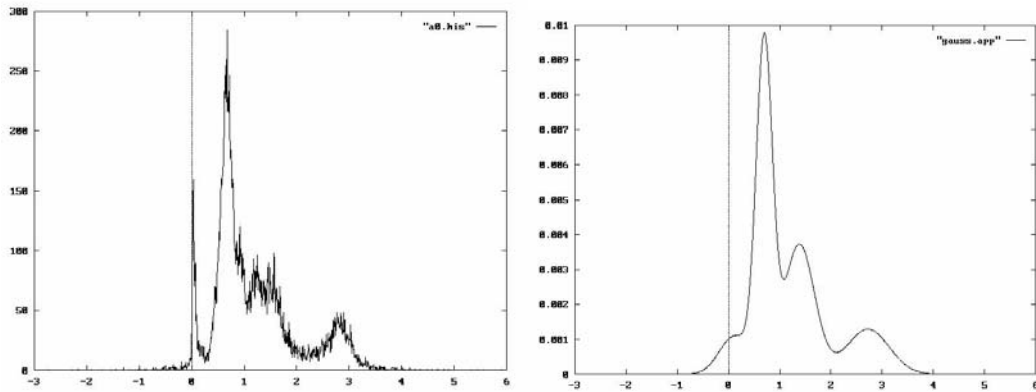


σχήμα 5.3: Frames 5 και 6 από την ακολουθία Flower Garden και αντίστοιχο πεδίο ταχυτήτων

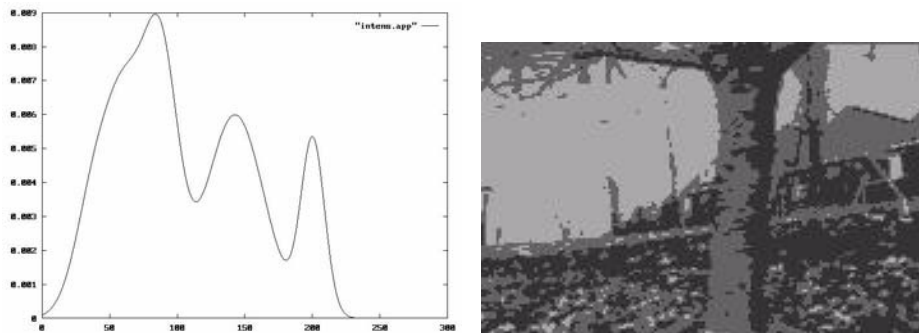
Στο σχήμα 5.4 φαίνεται το ιστογράμμα της u -συνιστώσας της οπτικής ροής ενώ στο ίδιο σχήμα περιέχεται η προσέγγιση του ιστογράμματος αυτού, κάνοντας χρήση κατανομών Γκάους. Κατά τη προσέγγιση αυτή χρησιμοποιήθηκαν 4 Γκαουσιανές κατανομές, οι παράμετροι των οποίων είναι οι εξής:

ΜΕΣΗ ΤΙΜΗ	ΔΙΑΣΠΟΡΑ
0.11	0.08
0.69	0.02
1.38	0.08
2.72	0.17

Άρα θα έχουμε 4 ετικέτες λόγω οπτικής ροής. Όμοια, για το κβαντισμό του ιστογράμματος της φωτεινότητας (σχήμα 5.5), προέκυψαν τα παρακάτω επίπεδα απόφασης ύστερα από την εφαρμογή του αλγόριθμου των Lloyd-Max: {52, 100, 169}. Οι ετικέτες λόγω φωτεινότητας θα είναι επομένως 3. Άρα ο συνολικός αριθμός ετικετών θα είναι $3 \times 4 = 12$.



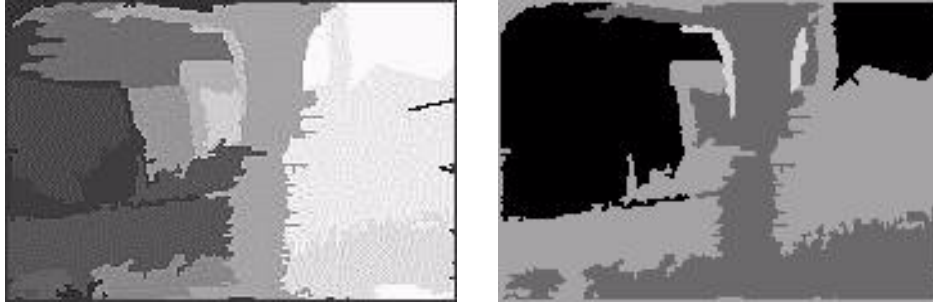
σχήμα 5.4: (αριστερά) ιστογράμμο της u-συνιστώσας της οπτικής ροής και (δεξιά) προσέγγιση του ιστογράμματος αυτού με κατανομές Γκάους.



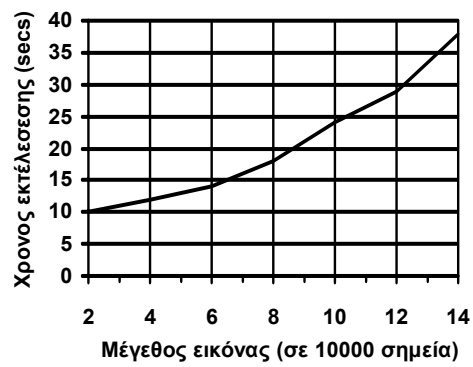
σχήμα 5.5: (αριστερά) ιστογράμμο της φωτεινής έντασης και (δεξιά) κβαντισμός της εικόνας 5 του Flower Garden ως προς τη φωτεινότητα με επίπεδα απόφασης {52, 100, 169}.

Με είσοδο τις ετικέτες αυτές, το αποτέλεσμα που προκύπτει μετά την εφαρμογή του αλγορίθμου HCF φαίνεται στο σχήμα 5.6. Είναι φανερό ότι κατά το στάδιο αυτό οι περιοχές που προέκυψαν από την τμηματοποίηση είναι πολύ περισσότερες από τα πραγματικά στρώματα βάθους. Όλες οι πλεονάζουσες περιοχές θα εξαλειφθούν κατά το τελευταίο στάδιο, δηλαδή τη συνένωση περιοχών. Το αποτέλεσμα από το στάδιο αυτό (το οποίο αποτελεί και το τελικό αποτέλεσμα) περιέχεται στο ίδιο σχήμα. Το πλήθος των περιοχών τώρα έχει μειωθεί και κάθε περιοχή αντιστοιχεί και σε ένα πραγματικό στρώμα βάθους.

Το κύριο τμήμα του υπολογιστικού κόστους για την εξαγωγή των στρωμάτων βάθους οφείλεται στην εφαρμογή του HCF αλγορίθμου. Στο σχήμα 5.7 περιέχεται μια γραφική παράσταση του χρόνου εκτέλεσης του HCF. Τέλος, ενδιαφέρον θα είχε να εξετάζαμε τα αποτελέσματα που θα προέκυπταν, αν κατά την προσέγγιση της οπτικής ροής (αλλά και της φωτεινότητας), γινόταν χρήση κατανομών Laplace αντί των κατανομών Γκάους.



σχήμα 5.6: (αριστερά) αποτέλεσμα του HCF αλγορίθμου, (δεξιά) τελικό αποτέλεσμα τμηματοποίησης.



σχήμα 5.7: Χρόνος εκτέλεσης HCF αλγορίθμου.

6. ΠΑΡΑΡΤΗΜΑ Α - HCF

6.1 Εισαγωγή

Στο κεφάλαιο αυτό θα ασχοληθούμε με την περιγραφή του HCF αλγορίθμου. Ο αλγόριθμος αυτός χρησιμοποιήθηκε κατά την παρούσα εργασία για την εύρεση των στρωμάτων βάθους, αφού προηγουμένως το συγκεκριμένο πρόβλημα είχε αναχθεί σε ένα πρόβλημα επιλογής ετικετών. Ο HCF αλγόριθμος στηρίζεται άμεσα στη θεωρία των πεδίων Markov. Για τον λόγο αυτό το κεφάλαιο θα ξεκινήσει με την παρουσίαση κάποιων βασικών εννοιών γύρω από τα πεδία αυτά, καθώς και με μια περιγραφή για το πως τα πεδία Markov μπορούν να χρησιμοποιηθούν για την μοντελοποίηση ενός προβλήματος ετικετών. Στη συνέχεια θα αναλυθεί ο τρόπος με τον οποίο ένα τέτοιο πρόβλημα είναι δυνατόν να επιλυθεί με τη βοήθεια του HCF καθώς και η σχέση του τελευταίου με τα πεδία Markov. Η ενότητα θα κλείσει με την ανάλυση του HCF.

6.2 Ορισμός ενός πεδίου Markov

Το *πρόβλημα των ετικετών* καθορίζεται πλήρως αν διαθέτουμε ένα σύνολο από αντικείμενα και ένα σύνολο από ετικέτες. Έστω $S = \{s_1, \dots, s_m\}$ ένα σύνολο από m διακριτά αντικείμενα και έστω $L = \{l_1, \dots, l_M\}$ ένα σύνολο από M ετικέτες. Το πρόβλημα των ετικετών ανάγεται ουσιαστικά στην επιλογή μιας ετικέτας από το σύνολο L για κάθε ένα από τα αντικείμενα του συνόλου S . Πολλά από τα προβλήματα του τομέα της ανάλυσης ακολουθιών εικόνων (π.χ. προβλήματα τμηματοποίησης) μπορούν να αναχθούν τελικά σε ένα πρόβλημα επιλογής ετικετών. Για παράδειγμα το πρόβλημα της εύρεσης ακμών συνίσταται ουσιαστικά στην απονομή μιας ετικέτας από το σύνολο $L = \{\text{ακμή}, \text{όχι-ακμή}\}$ σε κάθε σημείο της εικόνας.

Στη περίπτωση που υπάρχουν συσχετίσεις μεταξύ των αντικειμένων καθώς και κάποιοι περιορισμοί, για την μαθηματική μοντελοποίηση του προβλήματος της επιλογής ετικετών χρησιμοποιείται η θεωρία των **πεδίων Markov** ([50], [51], [52]). Ένα πεδίο Markov καθορίζεται πάντα σε σχέση με ένα σύστημα γειτνίασης. Το σύστημα γειτνίασης αποτελεί ουσιαστικά την μαθηματική περιγραφή των περιορισμών οι οποίοι διέπουν τα γειτονικά σημεία. Έτσι πριν προχωρήσουμε είναι αναγκαίο να δοθούν κάποιοι ορισμοί σχετικά με το θέμα αυτό.

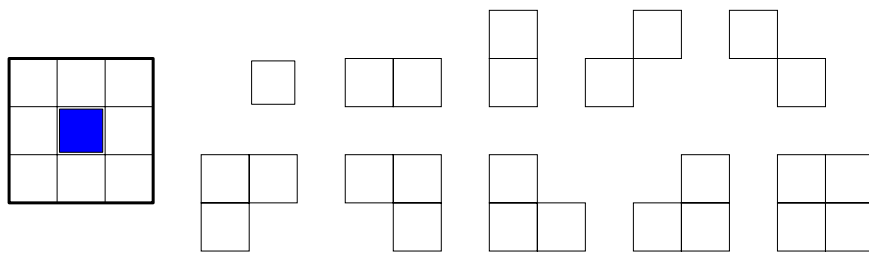
Ένα σύστημα γειτνίασης N για το σύνολο σημείων S ορίζεται ως:

$$N = \{ N_s \mid \textcircled{+} s \in S \}$$

όπου το N_s συμβολίζει το σύνολο από γειτονικά σημεία του αντικειμένου s . Ένα σύστημα γειτνίασης πρέπει να έχει τις παρακάτω ιδιότητες:

- ένα σημείο δεν αποτελεί γείτονα του εαυτού του:
 $s \not\in N_s$
- η σχέση γειτνίασης είναι συμμετρική:
 $s \in N_t \iff t \in N_s$

Το ζευγάρι (S, N) αποτελεί στην ουσία ένα γράφο όπου οι κόμβοι συμβολίζουν σημεία (αντικείμενα), και οι ακμές του γράφου (με βάρη) συμβολίζουν τους περιορισμούς που υπάρχουν για τις αποφάσεις που λαμβάνονται για γειτονικά σημεία. Θα ονομάζουμε *παρέα* του (S, N) κάθε υποσύνολο του συνόλου S του οποίου τα στοιχεία είναι γειτονικά ανά δύο. Με άλλα λόγια παρέα είναι κάθε ολικά συνδεδεμένος υπογράφος του γράφου (S, N) . Στο σχήμα 6.1 φαίνεται μια γειτονιά 8 σημείων καθώς και οι παρέες που σχετίζονται με τη γειτονιά αυτή.



σχήμα 6.1: γειτονιά 8 σημείων και αντίστοιχες παρέες

Διαθέτοντας τώρα όλα τα παραπάνω στοιχεία, δηλαδή:

- ένα σύνολο σημείων-αντικειμένων S
- ένα σύνολο πιθανών ετικετών-αποφάσεων L
- και ένα σύστημα γειτνίασης N

το νέο πρόβλημα το οποίο καλούμαστε να αντιμετωπίσουμε είναι να επιλέξουμε μια ετικέτα από το L για κάθε σημείο του συνόλου S με τέτοιο τρόπο ώστε να ικανοποιούνται οι περιορισμοί που επιβάλλονται από το σύστημα γειτνίασης N .

Με βάση τα παραπάνω προκύπτει ο παρακάτω ορισμός για τα πεδία Markov: Έστω X_s τυχαία μεταβλητή η οποία σχετίζεται με το σημείο (αντικείμενο) s και η οποία παίρνει τιμές στο σύνολο ετικετών L . Ο χάρτης αποφάσεων ω είναι το σύνολο τυχαίων μεταβλητών $X = \{X_s, s \in N\}$

$S\}$. Έστω $\omega_s = \omega(X_s)$ η απόφαση (ετικέτα) η οποία επιλέγεται για το σημείο s σύμφωνα με τον χάρτη αποφάσεων ω και έστω Ω το σύνολο όλων των δυνατών χαρτών απόφασης. Το σύνολο X θα αποτελεί ένα **πεδίο Markov** σε σχέση με το σύστημα γειτνίασης N και με συνάρτηση πυκνότητας πιθανότητας P , αν και μόνο αν ικανοποιούνται οι παρακάτω 2 συνθήκες:

- θετικότητα (Positivity)

$$P(x = \omega) > 0 \quad \forall \omega \in \Omega \quad (6.1)$$

- Μαρκοβιότητα (Markovianity)

$$P(X_s = \omega_s | X_r = \omega_r, r \in S, r \neq s) = P(X_s = \omega_s | X_r = \omega_r, r \in N_s) \quad (6.8)$$

Η ιδιότητα της Μαρκοβιότητας αναδεικνύει τα *τοπικά χαρακτηριστικά* του πεδίου X . Μια ετικέτα συσχετίζεται μόνο με γειτονικές ετικέτες. Με άλλα λόγια, οι επιδράσεις των μελών ενός πεδίου Markov περιορίζονται σε αλληλεπιδράσεις μεταξύ των γειτονικών σημείων. Είναι πάντοτε δυνατό οι γειτονιές N_s να επιλεγούν αρκετά μεγάλες ώστε να ισχύει η συνθήκη της Μαρκοβιότητας. Για παράδειγμα η μεγαλύτερη δυνατή γειτονιά είναι αυτή που αποτελείται από όλα τα υπόλοιπα σημεία. Κάθε πεδίο X αποτελεί ένα πεδίο Markov στην περίπτωση μιας τέτοιας γειτονιάς.

6.3 HCF και πεδία Markov

Η χρησιμότητα των πεδίων Markov σε προβλήματα που έχουν να κάνουν με επιλογή ετικετών, έγκειται στο γεγονός ότι η εκ των προτέρων γνώση που έχουμε για τις τοπικές εξαρτήσεις μεταξύ διαφορετικών οντοτήτων της εικόνας, μπορεί να μοντελοποιηθεί ικανοποιητικά με την τεχνική του περιορισμού των αποφάσεων σε γειτονικά σημεία. Το παρακάτω θεώρημα απλοποιεί περαιτέρω τον ορισμό των πεδίων Markov.

Θεώρημα HAMMERSLEY-CLIFFORD: Ένα τυχαίο πεδίο X αποτελεί πεδίο Markov σε σχέση με το σύστημα γειτνίασης N , αν και μόνο αν, υπάρχει μια συνάρτηση V_c τέτοια ώστε:

$$P(\omega) = \frac{e^{-\frac{1}{T}U(\omega)}}{Z} \quad \forall \omega \in \Omega \quad (6.9)$$

όπου T και Z σταθερές και

$$U(\omega) = \sum_{c \in C} V_c(\omega) \quad (6.10)$$

Το C παριστάνει το σύνολο των ολικά συνδεδεμένων υπογράφων (παρέες) σε σχέση με το σύστημα γειτνίασης N .

Η κατανομή πιθανότητας που ορίζεται από τις παραπάνω σχέσεις ονομάζεται *κατανομή Gibbs* σε σχέση με το σύστημα γειτνίασης N . Στις παραπάνω εξισώσεις η συνάρτηση U συμβολίζει την συνάρτηση ενέργειας (κόστους). Επίσης, οι συναρτήσεις V_c παριστάνουν τα δυναμικά τα οποία προστίθενται στην συνολική ενέργεια λόγω των τοπικών αλληλεπιδράσεων μεταξύ των σημείων μιας παρέας c . Οι συναρτήσεις δυναμικού ορίζονται ως εξής:

$$V_c(\omega_s, \omega_t) = \begin{cases} -\xi, & \text{αν } \omega_s = \omega_t \\ +\xi, & \text{αν } \omega_s \neq \omega_t \end{cases} \quad (6.11)$$

Η ισοδυναμία μεταξύ πεδίων Markov και κατανομών Gibbs δεν συσχετίζει απλώς τις τοπικά δεσμευμένες συναρτήσεις πυκνότητας πιθανότητας με την ολική συνάρτηση πυκνότητας πιθανότητας, αλλά μας δίνει και τη δυνατότητα εύκολου προσδιορισμού τους κάνοντας απλά χρήση των συναρτήσεων δυναμικού. Έτσι βασιζόμενοι στο θεώρημα των HAMMERLSEY-CLIFFORD και στην εξίσωση (6.8), τα τοπικά χαρακτηριστικά μπορούν να υπολογιστούν μέσω της συνάρτησης δυναμικού και της παρακάτω σχέσης:

$$P(X_s = \omega_s | X_r = \omega_r, r \neq s) = \frac{e^{-\frac{1}{T} \sum_{c \in C_s} V_c(\omega)}}{\sum_{\omega'} e^{-\frac{1}{T} \sum_{c \in C_s} V_c(\omega')}} \quad (6.12)$$

όπου C_s είναι οι δυνατές παρέες που περιέχουν το σημείο s ενώ ω' είναι ένας χάρτης αποφάσεων ο οποίος συμφωνεί με τον ω σε όλα τα σημεία εκτός ίσως από το s .

Με τον ορισμό που έχει δοθεί μέχρι τώρα για τα πεδία Markov και την ισοδυναμία τους με τις κατανομές Gibbs, έχουν οριστεί αλληλεπιδράσεις μεταξύ γειτονικών σημείων, αλλά δεν υπάρχει η δυνατότητα να μοντελοποιήσουμε την εκ των προτέρων γνώση που έχουμε για συμβατότητα μεταξύ των αποφάσεων και των διαθέσιμων παρατηρήσεων. Κάτι τέτοιο μπορεί να αποδειχθεί ότι είναι δυνατό να γίνει ([53]). Ως αποτέλεσμα στην περίπτωση αυτή προκύπτει μια νέα συνάρτηση ενέργειας της μορφής:

$$U(\omega, d) = \sum_{c \in C_s} V_c(\omega) - T \sum_{s \in S} \ln[\delta(\omega_s, d_s)] \approx U_1(\omega) + U_2(\omega, d) \quad (6.13)$$

όπου η συνάρτηση $\delta(\omega_s, d_s)$ αποτελεί ένα μέτρο της ποιότητας της λύσης σε σχέση με τις διαθέσιμες παρατηρήσεις d_s . Με όμοιο τρόπο και αλλάζοντας τις συναρτήσεις U_1 και U_2 μπορούν να υπολογιστούν και πιο πολύπλοκες μορφές για την συνάρτηση ενέργειας. Συνήθως η λύση του προβλήματος προκύπτει με την εφαρμογή του κριτηρίου της *μεγιστοποίησης της εκ των υστέρων πιθανότητας* ([54]), κάτι το οποίο ισοδυναμεί με την ελαχιστοποίηση της παραπάνω συνάρτησης $U(\omega, d)$. Στην περίπτωση δηλαδή αυτή η συνήθης συνολική διαδικασία που ακολουθείται είναι η εξής:

- Κάνε αναγωγή του προβλήματος σε ένα πρόβλημα επιλογής ετικετών και επέλεξε ένα κατάλληλο πεδίο Markov για την μοντελοποίηση του νέου προβλήματος
- Υπολόγισε την εκ των υστέρων ενέργεια
- βρες τη λύση σύμφωνα με το κριτήριο της *μεγιστοποίησης της εκ των υστέρων πιθανότητας*, ελαχιστοποιώντας την συνάρτηση ενέργειας που υπολογίστηκε στο προηγούμενο βήμα.

Ένας τέτοιος αλγόριθμος ελαχιστοποίησης είναι ο **HCF**. Βασικό πλεονέκτημα του HCF αποτελεί το γεγονός ότι έχει σχετικά μικρό υπολογιστικό κόστος. Επιπλέον παρουσιάζει ιδιαίτερη σταθερότητα (οι εκτιμήσεις του δεν επηρεάζονται σημαντικά από τη παρουσία θορύβου), ενώ δεν απαιτεί εκ μέρους του χρήστη τον καθορισμό κρίσιμων παραμέτρων (π.χ. σειρά επίσκεψης σημείων). Ακολουθεί η περιγραφή του HCF.

6.4 HCF (Highest Confidence First)

Στην παράγραφο αυτή θα παρουσιαστεί ένας αρκετά διαδεδομένος αλγόριθμος για την ελαχιστοποίηση της ενέργειας ενός πεδίου Markov. Ο συγκεκριμένος αλγόριθμος είναι γνωστός με το όνομα HCF (Highest Confidence First) ([53]). Έστω $L = \{l_1, \dots, l_M\}$ ένα σύνολο από αμοιβαία αποκλειόμενες αποφάσεις (ετικέτες) σε σχέση με τις οποίες ορίζεται ένα συγκεκριμένο πρόβλημα επιλογής ετικετών. Και έστω Ω ο χώρος όλων των πιθανών λύσεων. Η εκ των υστέρων γνώση σχετικά με τις αποφάσεις αναπαρίσταται με κατανομή Gibbs και άρα η εκ των υστέρων πιθανότητα για κάποια συγκεκριμένη λύση ω δίνεται από τη σχέση:

$$P(\omega|d) = \frac{e^{-\frac{1}{T} \left\{ \sum_{c \in C_s} V_c(\omega) - T \sum_{s \in S} \ln [P_s(\omega_s|d_s)] \right\}}}{Z} \quad (6.14)$$

όπου ο όρος $P_s(\omega_s|d_s)$ εκφράζει την ποιότητα της λύσης σε σχέση με τις διαθέσιμες παρατηρήσεις. Ας συμβολίσουμε με $L_1 = L \setminus \{l_0\}$ ένα νέο διευρυμένο σύνολο ετικετών όπου η ετικέτα l_0 αντιστοιχεί στην ‘μη λήψη απόφασης’. Έστω επίσης $\Omega_1 = \{\omega = (\omega_1, \dots, \omega_N) | \omega_s \in L_1 \text{ για όλα τα } s \in S\}$ ο νέος διευρυμένος χώρος λύσεων. Η βασική ιδέα πάνω στην οποία στηρίζεται ο HCF αλγόριθμος είναι η κατασκευή μιας ακολουθίας λύσεων $\omega^0, \omega^1, \dots, \omega^f$ οι οποίες ανήκουν στον Ω_1 χώρο και για τις οποίες ισχύει ότι: η αρχική λύση της ακολουθίας θα είναι $\omega^0 = \{l_0, \dots, l_0\}$ ενώ η τελική λύση ω^f θα αποτελεί τοπικό ελάχιστο της συνάρτησης κόστους-ενέργειας σε σχέση με το Ω .

Θα λέμε ότι ένα σημείο-αντικείμενο θα είναι *δεσμευμένο* στην ετικέτα $l \in L$ όταν κατά το t -οστό βήμα του αλγορίθμου ισχύει ότι $\omega_s^t = l$, ενώ θα λέμε ότι είναι *διαθέσιμο* όταν αντίστοιχα ισχύει

$\omega_s^l = l_0$. Επιπλέον ορίζουμε ένα κανόνα σύμφωνα με τον οποίο όταν κάποιο σημείο γίνει δεσμευμένο, τότε το σημείο αυτό δεν μπορεί πλέον να ξαναγίνει διαθέσιμο αλλά μπορεί απλά να αλλάξει την ετικέτα ως προς την οποία είναι δεσμευμένο. Η λογική του κανόνα αυτού θα φανεί στη συνέχεια.

Ας ορίσουμε τώρα την *διευρυμένη εκ των υστέρων τοπική ενέργεια* της ετικέτας $l \in L$ στο σημείο $s \in S$ σε σχέση με μια συγκεκριμένη λύση $\omega \in \Omega$, ως εξής:

$$E_s(l) = \sum_{c \in C_s} V_c'(\omega_c) - T \ln P(d_s | l) \quad (6.15)$$

όπου $\omega \in \Omega$ είναι εκείνη η λύση η οποία συμφωνεί παντού με τη λύση ω εκτός από το σημείο s όπου $\omega_s = l$. Επίσης για κάθε σημείο r στη γειτονιά του s ισχύει ότι:

$$V_c'(\omega_r) = \begin{cases} 0 & , \omega_r = l_0 \\ V_c(\omega_r) & , \omega_r \neq l_0 \end{cases} \quad (6.16)$$

Η τιμή της παραπάνω συνάρτησης E_s αποτελεί ένα μέτρο για το πόσο 'καλή' είναι η απόφαση (ετικέτα) l για το σημείο s σε σχέση με τις ετικέτες που έχουν επιλεγεί για τους γείτονες του σημείου κατά τη τρέχουσα λύση. Θα πρέπει εδώ να παρατηρήσουμε ότι μόνο τα δεσμευμένα γειτονικά σημεία συνεισφέρουν στην ποσότητα E_s , με αποτέλεσμα τα μη δεσμευμένα σημεία να μην επηρεάζουν τις αποφάσεις άλλων σημείων. Ωστόσο ένα διαθέσιμο σημείο πάντα λαμβάνει υπόψη του τις ετικέτες των ενεργών γειτόνων του κατά την διάρκεια της επιλογής της δικής του ετικέτας.

Προκειμένου τώρα να εξασφαλίσουμε την ποιότητα της τελικής εκτίμησης ω^f που θα προκύψει, θεωρούμε ότι ισχύει ο παρακάτω κανόνας ο οποίος και καθορίζει την σειρά επίσκεψης των διαφόρων σημείων της εικόνας:

Σε κάθε βήμα του αλγορίθμου μόνο στο πιο ασταθές σημείο της εικόνας επιτρέπεται να μεταβάλλει την απόφαση του ή να περάσει από την κατάσταση 'διαθέσιμο' στην κατάσταση 'δεσμευμένο'.

Η *σταθερότητα* ενός σημείου της εικόνας με βάση τη τρέχουσα λύση ω καθορίζεται ως εξής:

$$G_s(\omega) = \begin{cases} \min_{k \in L, k \neq \omega_s} \{ \Delta E_s(k, \omega_s) \}, & \text{αν } \omega_s \in L \\ - \min_{k \in L, k \neq j} \{ \Delta E_s(k, j) \}, & \text{αν } \omega_s = l_0 \end{cases} \quad (6.17)$$

όπου:

$$j \in L \text{ και } E_s(j) = \min_{k \in L} E_s(k) \quad (6.18)$$

$$\Delta E_s(j, k) = E_s(j) - E_s(k)$$

Επομένως από τις παραπάνω σχέσεις προκύπτει ότι η 'σταθερότητα' μιας απόφασης αποτελεί στην ουσία ένα συνδυασμένο μέτρο το οποίο βασίζεται από τη μια στις αποφάσεις που

παίρνονται από βήμα σε βήμα και από την άλλη στην εκ των προτέρων γνώση που έχουμε σχετικά με τις προτιμήσεις της παρούσας κατάστασης έναντι των άλλων καταστάσεων. Μια αρνητική τιμή για το G φανερώνει ότι μια πιο σταθερή (με μικρότερη ενέργεια δηλαδή) κατάσταση θα προκύψει από μια διαφορετική επιλογή αποφάσεων. Σύμφωνα με την εξίσωση (6.17) όλα τα 'διαθέσιμα' σημεία έχουν αρνητική ενέργεια G . Το μέγεθος του G αντιστοιχεί στην ποσότητα της ενέργειας που θα χάναμε ή θα κερδίζαμε αν αλλάζαμε την τρέχουσα κατάσταση. Όσο μικρότερη (αρνητική) είναι η τιμή του G , τόσο μεγαλύτερη είναι η βεβαιότητα για την αλλαγή κατάστασης.

Δεδομένου ότι ένα σημείο της εικόνας δεν μπορεί να επηρεάσει τις αποφάσεις των γειτονικών του σημείων αν αυτό δεν είναι 'δεσμευμένο', τα σημεία τα οποία παρουσιάζουν υψηλή πιθανότητα για μια απόφαση έναντι των άλλων αποφάσεων δέχονται επίσκεψη νωρίς κατά τη διάρκεια της κατασκευής. Αντίθετα στα σημεία εκείνα για τα οποία δεν υπάρχει αρκετή γνώση με βάση τις διαθέσιμες παρατηρήσεις, η κατάσταση των γειτονικών τους σημείων θα ληφθεί υπόψη κατά την επιλογή της ετικέτας τους. Έτσι μια πρόωρη απόφαση για κάποιο σημείο θα τροποποιηθεί αν οι αποφάσεις που θα ληφθούν στη συνέχεια για τα γειτονικά του σημεία συνηγορούν ενάντια στην αρχική αυτή απόφαση. Άρα μια αρχικά λανθασμένη απόφαση είναι δυνατόν να διορθωθεί στη συνέχεια.

Ο HCF αλγόριθμος μπορεί να υλοποιηθεί είτε σειριακά είτε παράλληλα. Κατά τη σειριακή υλοποίηση θα γίνει χρήση ενός σωρού (ουρά προτεραιότητας) στον οποίο θα διατηρείται η σειρά επίσκεψης των σημείων της εικόνας με βάση τις τιμές της G . Στη κορυφή του σωρού θα βρίσκεται εκείνο το σημείο που έχει την μικρότερη τιμή για την G . Μεταβάλλοντας την απόφαση για το σημείο της κορυφής του σωρού, αυτό θα έχει σαν αποτέλεσμα και την αλλαγή της τιμής της G για τα γειτονικά σημεία. Άρα στην περίπτωση αυτή θα υπάρξει και αλλαγή στην δομή του ίδιου του σωρού. Στο σχήμα 6.2 φαίνεται μια περιγραφή του HCF αλγόριθμου σε ψευδο-κώδικα.

Η ρουτίνα $\text{Change_State}(\omega_s)$ μεταβάλλει την τρέχουσα κατάσταση του σημείου s από ω_s στην κατάσταση l έτσι ώστε:

$$\Delta E_s(l, \omega_s) = \min_{k \in L, k \neq \omega_s} \Delta E_s(k, \omega_s), \text{ αν } \omega_s \in L$$

ή

$$E_s(l) = \min_{k \in L} E_s(k), \text{ αν } \omega_s = l_0$$

Κατά τη διάρκεια αυτής της μεταβολής, η σταθερότητα του σημείου s γίνεται θετική. Η ρουτίνα Update_G καλείται για όλα τα σημεία τα οποία επηρεάζονται από αυτή την αλλαγή, συγκεκριμένα οι γείτονες του s , έτσι ώστε αυτά να ενημερώσουν τα μέτρα σταθερότητας τους

σύμφωνα με τις νέες αποφάσεις. Τέλος η *Adjust_Heap(r)* διατηρεί τη σωστή διάταξη του σωρού με το να μετακινεί το στοιχείο *r* πάνω ή κάτω ανάλογα με τη νέα τιμή του *G* που αυτό έχει.

Ο αλγόριθμος συγκλίνει σε πεπερασμένο χρόνο. Κατά τη υλοποίηση του απαιτούνται $O(N)$ συγκρίσεις για τη δημιουργία του σωρού ενώ απαιτούνται και $O(\log N)$ πράξεις για να διατηρηθεί η σωστή διάταξη του σωρού κατά τη διάρκεια της επίσκεψης ενός σημείου (υπό τη προϋπόθεση πάντα ότι το μέγεθος της γειτονιάς είναι μικρό σε σχέση με το συνολικό πλήθος σημείων N). Το κόστος της διατήρησης του σωρού αντισταθμίζεται και με το παραπάνω από το γεγονός ότι χάρη σε αυτή σημειώνεται αρκετή πρόοδος σε κάθε επανάληψη. Στην αντίθετη περίπτωση θα είχαμε μόνο λίγες αλλαγές σε κάθε επανάληψη με αποτέλεσμα να ήταν απαραίτητες περίπου N επανάληψεις. Το l_0 κλειόνεκτημα αυτής της μεθόδου γίνεται φανερό καθώς το πλήθος των σημείων N μεγαλώνει.

```

top = Create_Heap( $\omega$ );
while (Gtop < 0) {
    s = top;
    Change_State( $\omega$ s);
    Update_G(Gs);
    Adjust_Heap(s);
    foreach( $r \in Ns$ ) {
        Update_G(Gr);
        Adjust_Heap(r);
    }
}
return( $\omega$ );

```

σχήμα 6.2: Αλγόριθμος HCF

7. ΠΑΡΑΡΤΗΜΑ Β

7.1 Εκτίμηση των παραμέτρων μιας μίξης κατανομών

Ένα αρκετά συνηθισμένο στατιστικό πρόβλημα είναι και αυτό της ανάλυσης κάποιου δείγματος σε ένα ορισμένο αριθμό κλάσεων. Το πλήθος των κλάσεων είναι γνωστό εκ των προτέρων ενώ κάθε κλάση περιγράφεται από μια συνάρτηση πυκνότητας πιθανότητας, της οποίας τις παραμέτρους ψάχνουμε. Έστω c το γνωστό πλήθος των κλάσεων, P_j η εκ των προτέρων πιθανότητα της κλάσης j και $p(x|\theta_j)$ η συνάρτηση πυκνότητας πιθανότητας για την δοσμένη κλάση, όπου το διάνυσμα θ_j συμβολίζει το σύνολο των άγνωστων παραμέτρων που καθορίζουν την κλάση j . Η συνάρτηση πυκνότητας πιθανότητας για την μίξη των παραπάνω κλάσεων θα είναι:

$$p(x|\phi) = \sum_{j=1}^c P_j p(x|\theta_j) \quad (7.1)$$

$$\sum_{j=1}^c P_j = 1$$

όπου το διάνυσμα ϕ αποτελείται από όλες τις άγνωστες παραμέτρους δηλαδή τις $\{\theta_j\}$ και $\{P_j\}$. Το ζητούμενο είναι ο υπολογισμός του διανύσματος ϕ .

Αυτό μπορεί να επιτευχθεί με τη χρήση του κριτηρίου της μέγιστης πιθανοφάνειας ([33], [40], [41]). Η εκ των υστέρων πιθανότητα της κλάσης i δεδομένου της παρατήρησης x θα είναι:

$$P_i(x|\phi) = \frac{P_i p(x|\theta_i)}{\sum_{j=1}^c P_j p(x|\theta_j)} \quad (7.2)$$

Αν $\{x_1, \dots, x_N\}$ είναι τα δείγματα που διαθέτουμε, τότε οι εκ των προτέρων πιθανότητες καθώς και οι παράμετροι των συναρτήσεων πυκνότητας πιθανότητας πρέπει να ικανοποιούν τις παρακάτω σχέσεις:

$$P_i = \frac{\sum_{n=1}^N P_i(x_n|\phi)}{N} \quad (7.3)$$

$$\sum_{n=1}^N P_i(x_n|\phi) \nabla_{\theta_i} \log p(x_n|\theta_i) = 0$$

όπου:

$$P_i(x_n|\varphi) = \frac{P_i p(x_n|\theta_i)}{\sum_{j=1}^c P_j p(x_n|\theta_j)} \quad (7.4)$$

Πιο συγκεκριμένα στην περίπτωση της μίξης κατανομών Γκάους, οι σχέσεις που μας δίνουν τις μέσες τιμές και τις διασπορές σε κάθε βήμα είναι:

$$\mu(j+1) = \frac{\sum_{k=1}^n P(\omega_i|x_k, \varphi(j))x_k}{\sum_{k=1}^n P(\omega_i|x_k, \varphi(j))} \quad (7.5)$$

$$\sigma_i^2(j+1) = \frac{\sum_{n=1}^N P(\omega_i|x_k, \varphi(j))(x_k - \mu_i)^2}{\sum_{n=1}^N P(\omega_i|x_k, \varphi(j))} \quad (7.6)$$

7.2 Κβαντιστής Lloyd-Max

Το πρόβλημα κβαντισμού συνίσταται ουσιαστικά στον διαχωρισμό του άξονα των πραγματικών αριθμών σε ένα σύνολο από μη επικαλυπτόμενα διαστήματα έτσι ώστε:

$$\mathfrak{R} = (-\infty, d_1] \cup (d_1, d_2] \cup \dots \cup (d_{N-1}, \infty)$$

Κάθε διάστημα αναπαρίσταται από μια τιμή r_i η οποία ανήκει σε αυτό το διάστημα. Έστω X τυχαία μεταβλητή, της οποίας η συνάρτηση πυκνότητας πιθανότητας είναι $p(x)$. Ζητείται ο προσδιορισμός τόσο των επιπέδων κβαντισμού, όσο και των σημείων απόφασης για δοσμένο αριθμό διαστημάτων και επιπέδων κβαντισμού N , έτσι ώστε να ελαχιστοποιείται το παρακάτω μέσο τετραγωνικό σφάλμα:

καλύτερη αντιπροσώπευση με (δοσμένα τα d_i) ουσία
$$D = E\{(X - Q(X))^2\} = \sum_{i=1}^N \int_{d_{i-1}}^{d_i} (x - r_i)^2 p(x) dx \quad (7.7)$$

όπου Q είναι ο τελεστής κβαντισμού και $(d_0, d_N) = (-\infty, +\infty)$.
$$r_i = E\{X|X \in (d_{i-1}, d_i)\} = \frac{\int_{d_{i-1}}^{d_i} x p(x) dx}{\int_{d_{i-1}}^{d_i} p(x) dx}, \quad i = 1, \dots, N \quad (7.8)$$

Οι Lloyd-Max ([43], [58]) έδωσαν τις αναγκαίες συνθήκες για την ελαχιστοποίηση της D . Το δεύτερο δίνει την καλύτερη διαμέριση όταν είναι γνωστά τα επίπεδα αντιπροσώπευσης: ποσότητας D , που συνίστανται σε δύο συστήματα εξισώσεων. Το πρώτο δίνει την

$$d_i = \frac{r_i + r_{i+1}}{2}, \quad i = 1, K, N - 1 \quad (7.9)$$

Στη γενική περίπτωση το παραπάνω σύστημα εξισώσεων μπορεί να λυθεί μόνο με αριθμητικές επαναληπτικές μεθόδους. Δίνονται αρχικές τιμές στα επίπεδα αντιπροσώπευσης και στη συνέχεια εφαρμόζονται διαδοχικά οι παραπάνω εξισώσεις. Θα πρέπει να σημειωθεί ότι οι παραπάνω συνθήκες είναι αναγκαίες αλλά όχι ικανές. Επομένως η επαναληπτική λύση που θα προκύψει μπορεί να αποτελεί και τοπικό ελάχιστο.

8. ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] K. Aizawam H. Harashima and T. Saito, 'Model-based analysis synthesis image coding (MBASIC) system for a person's face', *Signal Processing: Image Communication*, Vol. 1, pp. 139-152, 1989.
- [2] R. Forcheimer and O. Fahlander, 'Low bit rate coding through animation', *Picture Coding Symposium, PCS'83*, Davis, March 1983
- [3] M.Kanado, A. Koike and Y. Hatori, 'Codings with knowledge-based analysis of motion pictures', *Picture Coding Symposium, PCS'87*, Stockholm, June 1987.
- [4] H.-G. Musmann, M. Hotter and J. Ostermann, 'Object-oriented analysis-synthesis coding of moving images', *Signal Processing: Image Communication*, Vol. 1, pp. 117-138, 1989.
- [5] M. K. Leung and T. S. Huang, 'An integrated approach to 3D motion analysis and object recognition', *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. PAMI-13, No. 10, pp. 1075-1084, Oct 1991.
- [6] A.N. Netravali, T. S. Huang et al, 'Algebraic methods in 3D Motion Estimation from two-view point correspondences', *Int. Journal of Imaging Systems and Technology*, Vol. 1, pp. 78-99, 1989.
- [7] A. Mitiche and P. Bouthemy, 'Tracking modelled objects using binocular images', *Computer Vision, Graphics and Image Processing*, Vol. 32, pp. 384-396, 1985.
- [8] Y. C. Kim and J. K. Aggarwal, 'Determining object motion in a sequence of stereo images', *IEEE Journal of Robotics and Automation*, Vol. 3., No. 6, pp. 599-614, Dec. 1987.
- [9] J.L. Barron, A.D.Jepson, and J.K. Tsotsos, 'The feasibility of motion and structure from noisy time-varying image velocity information', *Int. Journal of Computer Vision*, pp. 239-269, 1990.
- [10] G. Adiv. 'Determining three-dimensional motion and structure from optical flow generated by several moving objects', *IEEE Trans. On Pattern Analysis Machine Intelligence*, 7:384-401, 1985.

- [11] H. Nicolas. 'Hierarchie de modeles de mouvement et methodes d'estimation associees.' Ph.D. Thesis, Universite de Rennes I, 1992.
- [12] A. N. Netravali and J. Salz. 'Algorithms for estimation of three-dimensional motion' *ATT Technical Journal*, 64, 1985.
- [13] J. L. Dugelay and B. Choquet, 'A 3D image analysis algorithm and stereoscopic television', *Proc. Of Festival Int. des images 3D*, Paris, Sept. 1991.
- [14] J.L.Barron, D.J. Fleet and S.S. Beauchemin, 'Performance of optical flow techniques' *Int. Journal of Computer Vision*, 12:1, 43-77, 1994.
- [15] G. Adiv, 'Inherent ambiguities in recovering 3D motion and structure from a noisy flow field', *IEEE Trans. On Pattern Analysis and Machine Intelligence*, Vol. PAMI-11, pp. 477-489, May 1989.
- [16] Nelson RC. 'Qualitative Detection of Motion by a Moving Observer', *Int. Journal of Computer Vision*, 7(1):33-46, 1991.
- [17] B.K.P. Horn and J.R. Weldon, 'Direct methods for recovering motion', *Int. Journal of Computer Vision*, Vol. 2, pp. 51-76, 1988.
- [18] N. Paragios and G. Tziritas 'Detection and location of moving objects using deterministic relaxation algorithms', *ICPR*, 1996 (Vienne).
- [19] N. Paragios, P. Perez, G. Tziritas, C. Labit and P. Boutheymy 'Adaptive detection of moving objects, using multiscale techniques', *ICPR*, 1996 (Laussane).
- [20] B.K.P. Horn and B. schunk, 'Determining optical flow', *Artificial Intelligence*, Vol. 17, pp. 185-203, 1981
- [21] P. Anandan, 'Measuring Visual Motion from Image Sequences', Ph. D. disseration, COINS TR 87-21, Univ. of Massachusetts, Amherst, MA., 1987.
- [22] P. Anandan, 'A computational framework and an algorithm for the measurement of visual motion', *Int. Journal of Computer Vision*, Vol. 2, pp. 283-310, 1989.
- [23] E. Hildreth, 'The measurement of visual motion', MIT Press, 1983.

- [24] J. K. Aggarwal and N. Nandhakumar, 'On the computation of motion from sequences of image-a review', *Proc. IEEE*, Vol. 76, no. 8, pp. 917-935, 1988.
- [25] G. Tziritas and C. Labit, 'Motion analysis for image sequence coding', Elsevier, 1994.
- [26] H. C. Longuet- Higgins, 'A computer algorithm for reconstructing a scene from two projections', *Nature*, Vol. 293, pp. 133-135, Sept. 1981.
- [27] Y. Huang *et al.* 'Optic flow field segmentation and motion estimation using a robust genetic partitioning algorithm', *IEEE Trans. On Pattern Analysis Machine Intelligence*, 17:1177-1190, Dec. 1995.
- [28] P. J. Rousseeuw and A.M. Leroy, 'Robust regression and outlier detection', New York: J. Wiley & Sons, 1987.
- [29] P. Huber, 'Robust statistics', Wiley, 1981.
- [30] P. Meer, D. Mintz, A. Rosenfield, and D.Y. Kim, 'Robust regression methods for computer vision: A review', *Int. Journal of Computer Vision*, 6:59-70, 1991.
- [31] P. J. Rousseeuw, 'Least median of squares regression', *J. Amer. Statist. Assoc.* 79(388), Dec. 1984, 871-880.
- [32] N. Komodakis, G. Tziritas, 'Robust 3-D motion estimation and depth layering', *Proceedings of 13th International Conference on Digital Signal Processing*, 1997.
- [33] R. Duda and P. Hart. 'Pattern classification and scene analysis', New York: Willey-Interscience, 1973.
- [34] X. Zhuang, T. Wang and P. Zhang. 'A highly robust estimator through partially likelihood function modeling and its application in computer vision', *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 14, 1992.
- [35] F. R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, 'Robust statistics: the approach based on influence functions', New York: J. Wiley & Sons, 1986.
- [36] J. W. Tukey. 'Explortary Data Analysis', Addison-Wisle, Reading, MA., 1977.
- [37] S. Z. Li. 'On discontinuity-adaptive smoothness priors in computer vision', *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 1995 accepted for publication

[38] S. Z. Li. 'Robustizing Robust M Estimation Using Deterministic Annealing', *Pattern Recognition*, Vol. 29, No. 1, pp. 159-166, 1996.

[39] R. M. Haralick, H. Joo, C.N. Lee, X. Zhuang, V.G. Vaidya and M.B. Kim. 'Pose estimation from corresponding point data', *IEEE Transactions On Systems, Man and Cybernetics*, 19:1426-1446, 1989.

[40] G. McLachlan, D. Peel and W. Whiten, 'Maximum likelihood clustering via normal mixture model', *Signal Processing: Image Communication*, 8:105-111, 1996.

[41] R. Schalkoff. 'Pattern Recognition: statistical, structural and neural approaches', J. Wiley and sons, 1992.

[42] S. Kirkpatrick, C. D. Gellat, and M. P. Vecchi. 'Optimization by simulated annealing', *Science*, 220:671-680, 1983.

[43] S.P. Lloyd, 'Least squares quantization in PCM', *IEEE Transactions on Information Theory*, Vol. IT-28, pp. 129-136, Mar. 1982

[44] M. Subbarao and A.M. Waxman, 'On the uniqueness of image flow solutions for planar surfaces in motion', *Computer Vision, Graphics and Image Processing*, Vol. 36, pp. 208-220, 1986.

[45] A.M. Waxman and K. Wahn, 'Image flow theory: a framework for 3D inference from time-varying imagery', *Chapter 3 in Advances in Computer Vision*, Erlbaum Associates Ed., London, pp. 164-224, 1988.

[46] J. Rissanen, 'Modeling by shortest data description', *Automatica*, Vol. 14, pp. 465-472, 1986.

[47] E. Francois, 'Interpretation qualitative du mouvement a partir d'une sequence d'images', Ph-D thesis, Universite de Rennes-I, June 1991.

[49] Y.T. Tse and R. Baker, 'Global zoom/pan estimation and compensation for video compression', *Proc. Of Int. Cong. On Speech, Acoustics and Signal Processing, ICASSP 91*, Vol. 4, pp. 2725-2728, May 1991.

[48] M. Hotter, 'Differential estimation of the global motion parameters zoom and pan', *Signal Processing*, Vol. 16, pp. 249-265, 1989.

- [50] R. Kindermann and J.L. Snell, 'Markov random fields and their applications', In *Contemporary Mathematics*, vol. 1, American Mathematical Society, 1980.
- [51] S. Geman and C. Graffigne, 'Markov random field image models and their applications to computer vision', *Proc. Intern. Cong. Mathematicians*, Berkeley, CA, pp. 1496-1517, 1986.
- [52] S. Z. Li 'Markov random field modeling in computer vision', *Springer-Verlag*, New York, 1995.
- [53] P.B. CHOU and C.M. Brown, 'The theory and practice of image labeling', *Int. Journal of Computer Vision*, Vol. 4, pp. 185-210, 1990.
- [54] S. Geman and D. Geman, 'Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images', *IEEE Trans. PAMI* 9(2):220-228, 1987.
- [55] S. Negahdaripour and B.K.P. Horn, 'Direct passive navigation', *IEEE Trans. Pattern Anal. Mach. Intelligence*, Vol. 9, 1987, pp. 168-176.
- [56] L.H. Matthies, R. Szeliski and T. Kanade, 'Kalman filter-based algorithms for estimating depth from image sequences', *Int. Journal of Computer Vision*, Vol. 3, pp. 209-236, 1989.
- [57] E.J. Weldon, Jr. and H. Liu, 'How accurately can direct motion vision determine depth?', *Proceedings of 1991 IEEE Conf. On Computer Vision and Pattern Recognition*, pp. 613-618, Hawaii, 1991.
- [58] J. Max, 'Quantizing for minimum distortion', *IRE Transactions on Information Theory*, Vol. IT-6, pp. 7-12, 1960.
- [59] F. Pedersini, A. Sarti, and S. Tubaro. 'Combined motion and edge analysis for a layer-based representation of image sequences', *Proc. IEEE Conf. On Image Processing*, pp. 921-924, Sept. 1996.
- [60] J. Wang and E. Adelson, 'Representing moving images with layers', *IEEE Trans. On Image Processing*, 3:625-638, 1994.