Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Πανοραμική σύνθεση όψεων από μια ακολουθία εικόνων

Μαρία Δ. Τράκα

Μεταπτυχιακή Εργασία

Ηράκλειο, Οκτώβριος 1999

© Copyright 1999 by Maria D. Traka
All Rights Reserved

Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Πανοραμική σύνθεση όψεων από μια ακολουθία εικόνων

Εργασία που υποβλήθηκε από τον
Μαρία Δ. Τράκα
ως μερική εκπλήρωση των απαιτήσεων για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Μαρία Δ. Τράκα
Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Εισηγητική Επιτροπή:

Γιώργος Τζιρίτας, Αναπληρωτής Καθηγητής, Επόπτης

Πάνος Τραχανιάς, Αναπληρωτής Καθηγητής, Μέλος

Απόστολος Τραγανίτης, Αναπληρωτής Καθηγητής, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Οκτώβριος 1999

Αφιερώνεται στους γονείς μου
και στα αδέλφια μου,
Ελένη και Γιώργο

Πανοραμική σύνθεση όψεων από μια ακολουθία εικόνων

Μαρία Δ. Τράκα

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Περίληψη

Η βαθμιαία εξάπλωση του video ως μέσο μετάδοσης πληροφορίας δημιούργησε νέες ανάγκες στον τομέα της αποθήκευσης, μετάδοσης και επεξεργασίας της πληροφορίας που εμπεριέχεται σε αυτό. Η έλλειψη όμως εργαλείων αναπαράστασης, διαχείρισης και τροποποίησης μιας video ακολουθίας δεν επιτρέπει την πλήρη εκμετάλλευση της πληροφορίας της. Μια λύση αποτελεί η κατασκευή και χρήση μιας εικόνας ή ορισμένων εικόνων που αναπαριστούν το σταθερό σκηνικό που απεικονίζεται σε μια video ακολουθία. Οι εικόνες αυτές ονομάζονται πανοραμικές όψεις και είναι γνωστές με τον όρο ”μωσαϊκά”. Αντικείμενο της εργασίας αποτελεί η κατασκευή τέτοιων εικόνων.

Τα στάδια της διαδικασίας κατασκευής πανοραμικών όψεων είναι η ευθυγράμμιση των στάσεων που επιτυγχάνεται με την εκτίμηση της κίνησης της κάμερας και η σύνθεση των στάσεων της ακολουθίας. Η κίνηση της κάμερας περιγράφεται από τα γνωστά παραμετρικά μοντέλα: προβολικό, αφινικό και μοντέλο εστίασης και μεταφοράς. Η εκτίμηση των παραμέτρων κίνησης επιτυγχάνεται με τη χρήση σθεναρών μεθόδων όπως ο αλγόριθμος M-estimator ή με τη χρήση αναλλοίωτων χαρακτηριστικών όπως οι σχέσεις ροπών δύο επιφανειών. Απαραίτητος για την εφαρμογή των μεθόδων ήταν ο προσδιορισμός του δισδιάστατου πεδίου οπτικής ροής, που προέκυψε εναλλακτικά από τις μεθόδους ταιριάσματος περιοχών και αντιστοίχισης γωνιών. Παραλλαγές των χαλαρωτικών μεθόδων καθώς και μεθόδων απότομης κλίσης χρησιμοποιήθηκαν για την διόρθωση των τελικών παραμέτρων κίνησης. Η σύνθεση των στάσεων της ακολουθίας επιτυγχάνεται με την απλή τοποθέτηση των στάσεων ή με τη χρήση φίλτρων όπως της μέσης και ενδιάμεσης τιμής. Το πρόβλημα των κακών ευθυγραμμίσεων, που παρουσιάζεται στην κατασκευή ”μωσαϊκών”, λύθηκε με τη χρήση της τοπολογίας των στάσεων.

Ενδεικτικά πειραματικά αποτελέσματα παρουσιάζονται για τον έλεγχο της απόδοσης και σύγκρισης των προτεινόμενων μεθόδων.

Επόπτης: Γιώργος Τζιρίτας
Αναπληρωτής Καθηγητής Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Panoramic views composition from an image sequence

Maria D. Traka

Master of Science Thesis

Computer Science Department
University of Crete

Abstract

The gradual development of video as mean of information transmition has created new needs for storing, transmitting and processing of its content. The lack of representation, manipulation and editing tools of video sequence doesn't allow the full exploitation of its content. This can be achieved by the construction and use of an image or a number of images that represent the static background of scene. These images are the panoramic views of the scene. They are also known as mosaics. In our work, we propose a method of constructing these images.

The basic steps of the construction of panoramic views are the frame alignment, which is based on the camera motion estimation, and the frame integration of the sequences. The camera motion is described by the well known parametric models: projective, affine, ‘‘zooming and transmitting" models. The estimation of motion parameters is based on the use of robust methods such as the M-estimator algorithm or on affine moment invariants. The implementation of these methods requires the determination of the optical flow field that can be resulted by the block matching or corner correspondence. Variation of steepest descent and relaxation methods were used for the correction of motion parameters. The frame integration can be achieved by frame overlapping or by using mean and median value filters. The problem of false alignments of the construction of mosaics was solved by by frame topology.

Experimental results are presented in order to indicate the efficiency of the above

methods.

Supervisor: George Tziritas
Professor of Computer Science
University of Crete

Ευχαριστίες

Στο σημείο αυτό θα ήθελα να ευχαριστήσω θερμά τον επόπτη καθηγητή μου κ. Γιώργο Τζιρίτα για την πολύτιμη βοήθεια και καθοδήγησή του καθώς και για τις γνώσεις που απέκτησα κοντά του τα τρία χρόνια της συνεργασίας μας. Η συμβολή του, η συνεχής υποστήριξή του και το αμείωτο ενδιαφέρον του ήταν καθοριστικό για την ολοκλήρωση της παρούσας εργασίας.

Επίσης θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες προς τα μέλη της επιτροπής Πάνο Τραχανιά και Απόστολο Τραγανίτη για τις εύστοχες παρατηρήσεις και διορθώσεις τους στην εργασία αυτή.

Θα πρέπει να ευχαριστήσω το Ινστιτούτο Πληροφορικής ΙΤΕ για την υλικοτεχνική υποστήριξη που μου παρείχε και το Τμήμα Επιστήμης Υπολογιστών για το σύνολο των γνώσεων που αποκόμησα.

Ακόμη θα ήθελα να ευχαριστήσω τους φίλους μου για την κατανόηση και συμπαράσταση που μου έδειξαν. Ιδιαίτερα όμως θα ήθελα να ευχαριστήσω τον Γιάννη Βελεγράκη για την αδιάκοπη υποστήριξη και την πολύτιμη βοήθεια του κατά τη διάρκεια υλοποίησης αυτής της εργασίας.

Το μεγαλύτερο όμως ευχαριστώ το χρωστώ στους γονείς μου που όλα αυτά τα χρόνια αγωνίστηκαν θυσιάζοντας ουσιαστικά πράγματα από τη ζωή τους για την μόρφωση και καλλιέργειά μου. Αφιερώνω σε αυτούς την εργασία μου ως ελάχιστη ανταμοιβή για τους κόπους και αγωνίες τους τα τελευταία χρόνια. Ελπίζω να φάνηκα άξια των προσδοκιών τους.

Μαρία Δ. Τράκα

Περιεχόμενα

Περίληψη **i**

Abstract **iii**

Ευχαριστίες **v**

Περιεχόμενα **vii**

Κατάλογος Σχημάτων **xi**

Κατάλογος Πινάκων **xv**

1 Εισαγωγή **1**

1.1 Περιγραφή του προβλήματος 1

1.2 Αντικείμενο Εργασίας 3

1.3 Οργάνωση εργασίας 5

2 Το πρόβλημα της κατασκευής του "μωσαϊκού" **7**

2.1 Ορισμός του προβλήματος 7

2.2 Προηγούμενες Προσεγγίσεις 10

3 Εκτίμηση της κίνησης της κάμερας **13**

3.1 Εισαγωγή 13

3.2 Ορισμός και περιγραφή του προβλήματος 13

3.2.1 Τρισδιάστατο Μοντέλο Κίνησης 15

3.2.2 Προβολή του τρισδιάστατου χώρου στο επίπεδο 16

3.3 Οπτική Ροή και Δισδιάστατο Πεδίο Ταχυτήτων 18

3.3.1 Γεωμετρικό Μοντέλο 19

3.4 Παραμετρικά μοντέλα κίνησης 20

3.4.1 Περιγραφή και σύγκριση παραμετρικών μοντέλων κίνησης 21
3.4.2 Εκτίμηση των παραμετρικών μοντέλων κίνησης 25
3.5 Εκτίμηση της οπτικής ροής 27
3.5.1 Ταίριασμα περιοχών 27
3.5.1.1 Κριτήρια εκτίμησης της οπτική ροής 29
3.5.2 Αντιστοίχιση Χαρακτηριστικών της Εικόνας 33
3.5.2.1 Εξαγωγή γωνιών 34
3.5.2.2 Αντιστοίχιση Γωνιών 36
3.6 Μέθοδοι σθεναρής εκτίμησης 41
3.6.1 Περιγραφή των M-estimators 42
3.6.2 Συναρτήσεις κόστους και επίδρασης 44
3.6.3 Μειονεκτήματα της Χρήσης του M-estimator 47
3.6.4 Εκτίμηση των παραμετρικών μοντέλων με τη χρήση του M-estimator 48
3.6.5 Μέθοδος Marquardt-Levemberg για την εκτίμηση των παραμέτρων 51
3.6.6 Πειραματικά αποτελέσματα 54
3.7 Χρήση αναλλοίωτων αφινικών ροπών για την εκτίμηση της κίνησης . . . 57
3.7.1 Κατασκευή περιβλημάτων από ένα σύνολο αντίστοιχων σημείων . 59
3.7.1.1 Επιλογή σημείων για την κατασκευή του πολυγώνου . . . 59
3.7.1.2 Η κατασκευή του μη-κυρτού περιβλήματος 61
3.7.2 Πειραματικά αποτελέσματα 62
3.8 Βελτιστοποίηση της εκτίμησης των παραμέτρων του μοντέλου κίνησης . 64
3.8.1 Μέθοδος της Απότομης Κλίσης 65
3.8.2 Χρήση χαλαρωτικών μεθόδων 68

4 Κατασκευή του "μωσαϊκού" 71
4.1 Εισαγωγή 71
4.2 Τρόποι αναπαράστασης ”μωσαϊκού” στο επίπεδο 72
4.2.1 Στατικό ”μωσαϊκό” (Static mosaic) 72
4.2.2 Δυναμικό ”μωσαϊκό” (Dynamic mosaic) 73
4.2.3 Χρονική πυραμίδα (Temporal pyramid) 74
4.2.4 Πολλαπλής ανάλυσης ”μωσαϊκό” (Multiresolution mosaic) 74
4.3 Η ευθυγράμμιση των στάσεων της ακολουθίας 76
4.3.1 Ευθυγράμμιση στάση - στάση 76
4.3.1.1 Ευθυγράμμιση στάσεων με χρήση του αφινικού μοντέλου κίνησης 77

4.3.1.2 Ευθυγράμμιση στάσεων με χρήση του προβολικού μοντέλου κίνησης . . . 80
4.3.2 Ευθυγράμμιση στάσης με "μωσαϊκό" . . . 81
4.3.3 Ευθυγράμμιση "μωσαϊκού" με στάση . . . 82
4.4 Σύνθεση στάσεων για την κατασκευή του "μωσαϊκού" . . . 82
4.5 Το πρόβλημα της κακής ευθυγράμμισης των στάσεων . . . 85
4.6 Τοπική διόρθωση της ευθυγράμμισης των στάσεων . . . 87

5 Συγκρίσεις μεθόδων και πειραματικά αποτελέσματα 91
5.1 Εισαγωγή . . . 91
5.2 Σύγκριση μεθόδων εκτίμησης της κίνησης της κάμερας . . . 92
5.3 Σύγκριση παραμετρικών μοντέλων κίνησης . . . 97
5.4 Υπολογιστικό κόστος . . . 103

6 Επίλογος 117
6.1 Συμπεράσματα . . . 117
6.2 Εφαρμογές "μωσαϊκού" . . . 118
6.3 Μελλοντικές κατευθύνσεις . . . 121

Α Καθορισμός παραμέτρων (Χαλαρωτικές μέθοδοι) 123

Β Κατασκευή κυρτού περιβλήματος 125

Γ Καθορισμός παραμέτρων (Κριτήριο μέτρησης της ισχύς μιας αντιστοίχισης) 127

Δ Ευρετήριο Όρων 129

Βιβλιογραφία 131

Κατάλογος Σχημάτων

1.1 Ευθυγράμμιση των στάσεων του στιγμιοτύπου 4

2.1 Ακολουθία εικόνων 7
2.2 Κατασκευή ”μωσαϊκού” 9
2.3 Κατασκευή ”μωσαϊκού ” από μια πραγματική video ακολουθία 10

3.1 Scale factor ambiguity 15
3.2 Τρισδιάστατο μοντέλο κίνησης 16
3.3 Μοντέλο κάμερας, το μάτι της βελόνας(pinhole) 17
3.4 Οι επιδράσεις των μετασχηματισμών σε μια εικόνα 24
3.5 Το φαινόμενο ”chirping” 25
3.6 Αναζήτηση με Αυξανόμενη Ακρίβεια 29
3.7 Αλγόριθμος ταιριάσματος Περιοχών (block-matching) 30
3.8 Παράδειγμα αντιστοίχισης περιοχών (μπλοκς) 33
3.9 Συσχέτιση Γωνιών 37
3.10 Το πρόβλημα της μη-συμμετρίας στη μέτρηση της υποστήριξης μιας αντιστοίχισης 40
3.11 Αλγόριθμος αντιστοίχισης γωνιών 40
3.12 Παράδειγμα αντιστοίχισης γωνιών 41
3.13 Δύο στάσεις της ακολουθίας του Stefan (Stefan070-Stefan075). 55
3.14 Το ”μωσαϊκό” που κατασκευάστηκε από δύο στάσεις (Stefan070-Stefan075) της ακολουθίας του Stefan 55
3.15 Δύο στάσεις της ακολουθίας του Stefan(Stefan240-Stefan242) 56
3.16 Το ”μωσαϊκό” που κατασκευάστηκε από δύο στάσεις της ακολουθίας Stefan (Stefan240-Stefan242). 57
3.17 Ο συνολικός αλγόριθμος εκτίμησης των παραμέτρων κίνησης με την εφαρμογή της μεθόδου των αναλλοίωτων σχέσεων αφινικών ροπών 60

3.18 Δύο στάσεις της ακολουθίας του Stefan (Stefan244-Stefan245) 63
3.19 Τα αντίστοιχα περιβλήματα των δύο στάσεων της ακολουθίας Stefan (Stefan244 και Stefan245) 64
3.20 Το "μωσαϊκό" δύο στάσεων της ακολουθίας του Stefan (Stefan244-Stefan245) 64
3.21 Δύο στάσεις της ακολουθίας Barbara1 65
3.22 Τα περιβλήματα δύο στάσεων της ακολουθίας Barbara1 65
3.23 Το "μωσαϊκό" δύο στάσεων της ακολουθίας Barbara1 που κατασκευάστηκε με χρήση της μεθόδου των ροπών 66
3.24 Μέθοδος της απότομης κλίσης 67

4.1 Χρονική πυραμίδα 75
4.2 Ευθυγράμμιση των στάσεων της ακολουθίας ως προς το σύστημα συντεταγμένων του "μωσαϊκού" 77
4.3 Ευθυγράμμιση στάσεων με χρήση του αφινικού μοντέλου κίνησης 78
4.4 Ευθυγράμμιση στάσεων με χρήση του προβολικού μοντέλου κίνησης . . . 80
4.5 Η σύνθεση των στάσεων προκύπτει με την απλή τοποθέτηση των στάσεων στο "μωσαϊκό" 83
4.6 Η σύνθεση των στάσεων με τη χρήση του φίλτρου της μέσης τιμής 84
4.7 Η σύνθεση των στάσεων με τη χρήση του φίλτρου της ενδιάμεσης τιμής . 84
4.8 Η σύνθεση των στάσεων με τη χρήση του φίλτρου της μέσης τιμής σε μια ζώνη διάστασης K του περιγράμματος των στάσεων 85
4.9 Τοπολογία μιας ακολουθίας εικόνων 88

5.1 Τελευταίο ζευγάρι στάσεων της συνθετικής ακολουθίας Barbara1. 93
5.2 Το "μωσαϊκό" της συνθετικής ακολουθίας Barbara1 με τη μέθοδο ταιριάσματος περιοχών. 94
5.3 Το "μωσαϊκό" της συνθετικής ακολουθίας Barbara1 με τη μέθοδο αντιστοίχισης γωνιών. 95
5.4 Τρεις στάσεις από την ακολουθία του Stefan. 97
5.5 Δύο στάσεις της ακολουθίας του Stefan (Stefan260-Stefan262) 97
5.6 Τα αντίστοιχα περιβλήματα δύο στάσεων της ακολουθίας Stefan (Stefan260-Stefan262) 98
5.7 Το "μωσαϊκό" δύο στάσεων της ακολουθίας του Stefan (Stefan260-Stefan262) 99
5.8 Το "μωσαϊκό" όλης της ακολουθίας του Stefan. 104
5.9 Το "μωσαϊκό" της συνθετικής ακολουθίας Barbara1 105

5.10 Δύο στάσεις της συνθετικής ακολουθίας Barbara2 105
5.11 Το ”μωσαϊκό” της συνθετικής ακολουθίας Barbara2 όπως κατασκευάστηκε με χρήση του αφινικού μοντέλου κίνησης 106
5.12 Το ”μωσαϊκό” της συνθετικής ακολουθίας Barbara2 όπως κατασκευάστηκε με χρήση του προβολικού μοντέλου κίνησης 106
5.13 Η πρώτη (a) και τελευταία (b) στάση της ακολουθίας almaden 107
5.14 Το ”μωσαϊκό” της ακολουθίας almaden 108
5.15 Η πρώτη (a) και τελευταία (b) στάση της ακολουθίας princeton 109
5.16 Το ”μωσαϊκό” της ακολουθίας princeton 110
5.17 Η πρώτη (a) και τελευταία (b) στάση της ακολουθίας yosemite 111
5.18 Το ”μωσαϊκό” της ακολουθίας yosemite 112
5.19 Το ”μωσαϊκό” της ακολουθίας του Stefan που κατασκευάστηκε με τη χρήση του αφινικού μετασχηματισμού 113
5.20 Το ”μωσαϊκό” της ακολουθίας του Stefan που κατασκευάστηκε με τη χρήση του προβολικού μετασχηματισμού 114
5.21 Το ”μωσαϊκό” της ακολουθίας του Stefan όπως προέκυψε με τη χρήση της μεθόδου διόρθωσης 115

6.1 Αποθήκευση-συμπίεση και ανάκτηση μιας video ακολουθίας 120

.

Κατάλογος Πινάκων

3.1 Μετασχηματισμοί Συντεταγμένων 21
3.2 Συναρτήσεις Κόστους και Επίδρασης 45

.

Κεφάλαιο 1

Εισαγωγή

1.1 Περιγραφή του προβλήματος

Κατά τη διάρκεια των δεκαετιών 70-80 παρουσιάστηκαν μεγάλες εξελίξεις στον τομέα της τηλεόρασης. Νέες υπηρεσίες εμφανίστηκαν και καθιερώθηκαν στον καθημερινό βίο εκτελώντας επικοινωνιακό και συμβουλευτικό ρόλο. Μέσα για την επίτευξη αυτών των ρόλων αποτέλεσαν πηγές με μεγάλη και πλούσια πληροφορία όπως οι εικόνες και κατ' επέκταση οι video ακολουθίες.

Η επανάσταση αυτή στον τομέα της τηλεόρασης οδήγησε στη μελέτη και εκμετάλλευση των μέσων αυτών ειδικότερα όμως των video ακολουθιών. Η κλίση προς τη κατεύθυνση αυτή προέκυψε λόγω των βασικών πλεονεκτημάτων του video έναντι των εικόνων. Δύο από τα βασικότερα πλεονεκτήματα του είναι η δυνατότητά του να περιγράφει μια συνεχώς μεταβαλλόμενη άποψη της σκηνής, καθώς και η δυνατότητά του να αναπαριστά τη χρονική εξέλιξη των γεγονότων.

Αποτέλεσμα των δυνατοτήτων αυτών αποτελεί ο μεγάλος αριθμός εφαρμογών που προσφάτως έχει εμφανισθεί. Οι εφαρμογές αυτές περιλαμβάνουν επεξεργασία, εκμετάλλευση και χρήση της video πληροφορίας όπως οι ψηφιακές βιβλιοθήκες και άλλες. Επίσης, η μεγάλη έκρηξη της διάδοσης της πληροφορίας μέσω του παγκόσμιου ιστού οδήγησε στην ανάγκη για απευθείας (on-line) μετάδοση εικόνων και video ακολουθιών κωδικοποιημένες σε ψηφιακή μορφή [1]. Παρόλη τη αναγκαιότητα της από πλευράς χρηστών, η επίτευξη της μετάδοσης αυτής βρίσκεται ακόμα σε αρχικό στάδιο. Ανασταλτικοί παράγοντες για την ανάπτυξή της αποτέλεσαν το κόστος αποθήκευσης μιας video ακολουθίας καθώς και το σημαντικό χρονικό κόστος μετάδοσης της. Αιτία όλων αυτών είναι ο μεγάλος αριθμός στάσεων (frames), από τον οποίο αποτελείται το video, καθώς

και το μέγεθος της καθεμιάς στάσης.

Η έλλειψη επίσης εργαλείων δεικτοδότησης, αναζήτησης και ανάκτησης αντικειμένων με βάση το περιεχόμενο τους σε μια video ακολουθία οδηγεί στη μερική εκμετάλλευση της πληροφορίας της. Για τον καθορισμό τέτοιων κατάλληλων εργαλείων καλούμαστε να αντιμετωπίσουμε δύο βασικά προβλήματα. Το ένα από αυτά - και το ουσιαστικότερο - είναι η εύρεση ενός αποδοτικού τρόπου αναπαράστασης της σκηνής που περιγράφει ένα στιγμιότυπο (shot). Το δεύτερο αφορά την εύρεση αποδοτικών μεθόδων ανάκτησης και ανάλυσης της video πληροφορίας. Ο τρόπος αναπαράστασης της σκηνής πρέπει να διευκολύνει τη διαδικασία της δεικτοδότησης, της αναζήτησης και τη διαδικασία των επερωτήσεων που βασίζεται στο περιεχόμενο της video ακολουθίας. Τέτοια συστήματα αναπαράστασης θα πρέπει μελλοντικά να απαντούν σε επερωτήσεις που αναζητούν συγκεκριμένη πληροφορία σε μια video ακολουθία, όπως "τα στιγμιότυπα της Golden Gate γέφυρας, με ένα πλοίο να περνάει από κάτω της". Χρησιμοποιώντας την ιδέα των ήδη εφαρμοσμένων συστημάτων επερωτήσεων, όπως το QBIC (Query By Image Content)[2] και το Photobook[3], η ερώτηση αυτή για να απαντηθεί πρέπει να μετασχηματιστεί σε μία περιγραφή της σκηνής στην οποία θα ήταν εμφανή διάφορα χαρακτηριστικά της όπως το χρώμα[4], τυχόν κινήσεις αντικειμένων[5] ακόμα και ένα απλό δομικό σχέδιο της σκηνής[6]. Όμως παρά την απλοποίηση της αναζήτησης με την ανεύρεση μόνο ορισμένων χαρακτηριστικών της σκηνής δεν αποφεύγουμε τη διαδικασία της αναζήτησης σε όλες της στάσεις της video ακολουθίας.

Ένας από τους απλούστερους τρόπους αναπαράστασης μιας video ακολουθίας σύμφωνα με το σύστημα QBIC (Query By Image Content) είναι μέσω των στάσεων-κλειδιών (key-frames). Λέγοντας στάση-κλειδί εννοούμε μία αντιπροσωπευτική στάση ενός στιγμιοτύπου που τυπικά μπορεί να είναι είτε η πρώτη, είτε η μεσαία, είτε η τελευταία στάση του ή ακόμα και ένας συνδυασμός αυτών. Οι επερωτήσεις για αναζήτηση και ανάκτηση αντικειμένων εκτελούνται μόνο πάνω στις στάσεις-κλειδιά[7]. Ωστόσο μία τέτοια αναπαράσταση είναι περιοριστική και αρκετά φτωχή αφού μια στάση αναπαριστά χιλιάδες ή και εκατομύρια στάσεις ενός ολόκληρου στιγμιότυπου.

Λύσεις στα παραπάνω θα προταθούν σίγουρα πολλές στο μέλλον. Η έλλειψη όμως ενός αποδοτικού τρόπου αναπαράστασης της πραγματικότητας που περιγράφουν οι video ακολουθίες, καθώς και ο ανεπαρκής άμεσος χειρισμός (manipulation) και τροποποίηση τους, αποτελεί βασικό μειονέκτημά τους και σε αυτό ακριβώς οφείλεται η μερική εκμετάλλευση της πληροφορίας τους. Στις μέρες μας η τροποποίηση και ο χειρισμός μιας video ακολουθίας ανάγεται στην τροποποίηση και χειρισμό του συνόλου

των στάσεων από τις οποίες αποτελείται. Η τροποποίηση περιλαμβάνει διαδικασίες όπως η εγγραφή και η αφαίρεση αντικειμένων της ακολουθίας. Το συνολικό κόστος επίτευξης των διαδικασιών αυτών συχνά εξαρτάται από τον αριθμό των στάσεων της ακολουθίας, αν η τροποποίηση που επιθυμούμε εφαρμόζεται σε κάθε στάση της. Η ανάγκη εύρεσης ενός αποδοτικού τρόπου αναπαράστασης της σκηνής, με την προϋπόθεση ότι οι εγγραφές και γενικότερα οι τροποποιήσεις της σκηνής πραγματοποιούνται πάνω σε αυτή, είναι άμεση. Η αναπαράσταση αυτή θα αποτελέσει σημαντικό βήμα στην ευρεία χρήση και εκμετάλλευση των video ακολουθιών.

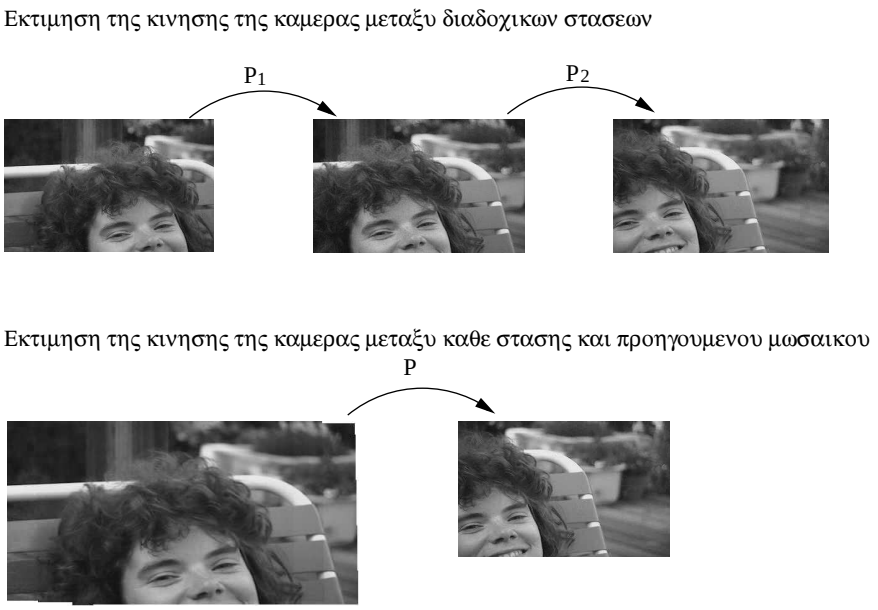
Αποδοτική αναπαράσταση, για το σκοπό αυτό, αποτελεί η κατασκευή μιας εικόνας ή ορισμένων εικόνων που αναπαριστούν όλο το σταθερό σκηνικό (background) ενός στιγμιότυπου μιας video ακολουθίας. Παράλληλα στην αναπαράσταση αυτή προστίθονται κατάλληλες δομές που περιέχουν πληροφορίες για τα κινούμενα αντικείμενα του κάθε στιγμιότυπου. Οι εικόνες που αναπαριστούν το σταθερό σκηνικό και για τις οποίες εκδηλώνεται ιδιαίτερο ενδιαφέρον στις μέρες μας, ονομάζονται πανοραμικές όψεις και είναι γνωστές με τον όρο "μωσαϊκά" (mosaic). Οι εικόνες αυτές μπορούν να χρησιμοποιηθούν για να απαντήσουμε σε ερωτήσεις που βασίζονται στα χαρακτηριστικά της στατικής σκηνής, όπως το χρώμα, η υφή [8], το σχήμα των επιφανειών που απεικονίζονται σε αυτές[2],[3]. Η αναπαράσταση αυτή επιτρέπει επιπλέον και την άμεση διαχείριση και τροποποίηση μιας video ακολουθίας. Η εγγραφή και η αφαίρεση αντικειμένων πραγματοποιείται πάνω στα "μωσαϊκά" και όχι στο σύνολο των στάσεων της ακολουθίας. Επομένως αυξάνεται η απόδοση και μειώνεται ο χρόνος επίτευξης των διαδικασιών αυτών. Επιπρόσθετα, τέτοιες αναπαραστάσεις αποτελούν αποδοτικά εργαλεία για πλοήγηση σε μια ακολουθία video .

1.2 Αντικείμενο Εργασίας

Η παρούσα εργασία ασχολείται με το πρόβλημα της κατασκευής του "μωσαϊκού". Το "μωσαϊκό" περιέχει τη στατική σκηνή ενός στιγμιοτύπου, δηλαδή τη σταθερή σκηνή που απεικονίζεται σε ένα στιγμιότυπο και δεν περιέχει αντικείμενα των στάσεων για τα οποία εντοπίστηκε ανεξάρτητη κίνηση, κίνηση δηλαδή που δεν οφείλεται σε κίνηση της κάμερας. Το "μωσαϊκό" κατασκευάζεται από όλες τις στάσεις ενός στιγμιοτύπου της ακολουθίας, δίνοντας έτσι την πανοραμική όψη της σκηνής. Η αρχή πάνω στην οποία στηρίζεται η δημιουργία του είναι η μικρή διαφορά στην πληροφορία που έχουν δύο διαδοχικές στάσεις του video.

Η παραγωγή του ”μωσαϊκού” αποτελείται από δύο στάδια:

- την ευθυγράμμιση των στάσεων του στιγμιοτύπου. Η ευθυγράμμιση αναλύεται στον υπολογισμό της κίνησης είτε μεταξύ διαδοχικών στάσεων του στιγμιοτύπου είτε μεταξύ κάθε στάσης και του ”μωσαϊκού”, που κατασκευάστηκε από τις προηγούμενες στάσεις, (σχήμα 1.1).



Σχήμα 1.1: Ευθυγράμμιση των στάσεων του στιγμιοτύπου

- τη σύνθεση των στάσεων για την κατασκευή του ”μωσαϊκού”.

Για την ευθυγράμμιση των στάσεων απαιτείται ο υπολογισμός της κίνησης της κάμερας κατά τη διάρκεια του στιγμιοτύπου. Για τον υπολογισμό της κίνησης χρησιμοποιήθηκαν διάφορα παραμετρικά μοντέλα το καθένα από τα οποία προσομοιώνει και διαφορετική κίνηση της κάμερας. Γενικά θεωρήθηκε ότι η κάμερα κινείται σε τρείς διαστάσεις, δηλαδή τόσο σε επίπεδο παράλληλο στο επίπεδο του σκηνικού, όσο και σε κάθετο. Με την προβολή όμως της κίνησης στο επίπεδο της εικόνας διαθέτουμε πληροφορία για την εκτίμηση της κίνησης σε χώρο δύο διαστάσεων. Αποτέλεσμα αυτού ήταν η χρήση παραμετρικών μοντέλων που παρέχουν την κίνηση της κάμερας σε χώρο δύο διαστάσεων. Τα δισδιάστα παραμετρικά μοντέλα δεν υπονοούν ότι η κίνηση της κάμερας είναι καθαρά κίνηση σε δύο διαστάσεις, αλλά προσωμειώνουν γενικά την τρισδιάστατη κίνηση της κάμερας όπως προβάλλεται στο επίπεδο της εικόνας. Σε μια μεγάλη κλάση πραγματικών video ακολουθιών η κάμερα κάνει πανοραμική λήψη, κλίνει προς κάποια

κατεύθυνση, αλλάζει το μήκος εστίασης της. Η πληροφορία της κίνησης χρησιμοποιείται για το μετασχηματισμό των στάσεων του στιγμιοτύπου σύμφωνα με επιλεγμένο σύστημα συντεταγμένων. Η επιλογή του συστήματος συντεταγμένων καθορίζει έναν διαφορετικό τύπο "μωσαϊκού", που καθορίζεται από την εφαρμογή στην οποία θα χρησιμοποιηθεί. Με τη συνένωση των στάσεων του στιγμιοτύπου σύμφωνα με το εκάστοτε σύστημα συντεταγμένων κατασκευάζεται το "μωσαϊκό".

1.3 Οργάνωση εργασίας

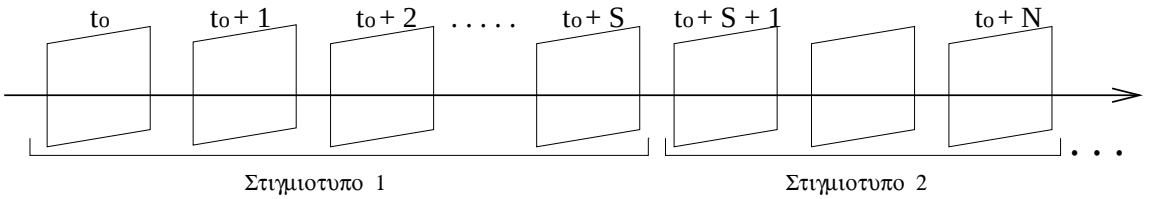
Στο κεφάλαιο 2 ορίζεται η έννοια του "μωσαϊκού" και παρουσιάζεται ο τρόπος κατασκευής του. Στη συνέχεια περιγράφονται προσεγγίσεις των επιμέρους προβλημάτων που αντιμετωπίζονται για την κατασκευή του "μωσαϊκού". Τα προβλήματα αυτά συνοψίζονται στο πρόβλημα εκτίμησης της κίνησης της κάμερας και στο πρόβλημα καθορισμού του τρόπου σύνθεσης των στάσεων της ακολουθίας για την κατασκευή του "μωσαϊκού". Στο κεφάλαιο 3 περιγράφεται αναλυτικά το πρόβλημα εκτίμησης της κίνησης και παρουσιάζονται κατάλληλα παραμετρικά μοντέλα για την περιγραφή της. Παράλληλα παρουσιάζονται διάφοροι μέθοδοι εκτίμησης της κίνησης όπως ο αλγόριθμος M-estimator και η μέθοδος των ροπών. Επιπλέον αναφέρονται και σχολιάζονται τα πειραματικά τους αποτελέσματα. Στο κεφάλαιο 4 αναφέρονται τρόποι αναπαράστασης του "μωσαϊκού" στο επίπεδο και επιπλέον περιγράφονται τρόποι ευθυγράμμισης και σύνθεσης των στάσεων της ακολουθίας για την κατασκευή του "μωσαϊκού". Τέλος, παρουσιάζεται το πρόβλημα των κακών ευθυγραμμίσεων και προτείνεται η λύση του, με τη χρήση της τοπολογίας των στάσεων. Στο κεφάλαιο 5 συγκρίνονται οι μέθοδοι εκτίμησης της κίνησης και τα παραμετρικά μοντέλα που χρησιμοποιήθηκαν. Παράλληλα παρουσιάζονται ενδεικτικά αποτελέσματα "μωσαϊκών" που κατασκευάστηκαν στην παρούσα εργασία. Το κεφάλαιο 6 αποτελεί τον επίλογο της εργασίας. Στο κεφάλαιο αυτό γίνεται μια γενική ανασκόπηση της εργασίας και αναφέρονται τα συμπεράσματα στα οποία καταλήξαμε. Επιπλέον αναφέρονται ορισμένες εφαρμογές και μελλοντικές κατευθύνσεις ή θέματα που θα πρέπει να μελετηθούν γενικότερα.

Κεφάλαιο 2

Το πρόβλημα της κατασκευής του "μωσαϊκού"

2.1 Ορισμός του προβλήματος

Με τον όρο ακολουθία video νοείται μια ακολουθία από διαδοχικές εικόνες του χώρου, δειγματοληπτημένες στο χρόνο από μία κάμερα με μεγάλη συχνότητα, της τάξης των 30 δειγμάτων το δευτερόλεπτο (σχήμα 2.1). Τα δείγματα αυτά συγκροτούν το σύνολο



Σχήμα 2.1: Ακολουθία εικόνων

Ακολουθία εικόνων είναι το σύνολο των διαδοχικών λήψεων της κάμερας με κάποια συγκεκριμένη συχνότητα δειγματοληψίας. Μία ακολουθία εικόνων αποτελείται από ένα σύνολο στιγμιοτύπων. Το κάθε στιγμιότυπο περιέχει διαφορετικό αριθμό στάσεων.

των στάσεων του video. Παρά τη μεγάλη συχνότητα δειγματοληψίας η διαφορά στην πληροφορία που περιγράφεται μεταξύ των διαδοχικών στάσεων είναι μικρή. Το μέγεθος της video ακολουθίας είναι πολύ μεγάλο αφού μερικά δευτερόλεπτα περιγράφονται από μερικές εκατοντάδες ή και χιλιάδες στάσεις. Γενικά μια video ακολουθία περιγράφει ένα σημαντικό χρονικό διάστημα. Για την καλύτερη και ουσιαστικότερη ανάλυση της χωρίζεται σε ένα αριθμό στιγμιότυπων. Ένα στιγμιότυπο είναι ένα σύνολο διαδοχικών στάσεων που απεικονίζουν μία σκηνή. Λόγω της μεγάλης χρονικής διάρκειας των λήψεων για την συγκρότηση μιας ακολουθίας, οι σκηνές και γενικότερα η πληροφορία

που απεικονίζεται αλλάζει σημαντικά. Αποτέλεσμα αυτού είναι η απεικόνιση ενός σημαντικού αριθμού διαφορετικών σκηνών μέσα στην ακολουθία. Εμφανίζεται τότε η ανάγκη του χωρισμού της ακολουθίας σε υποσύνολα, που περιέχουν στάσεις με σχετική πληροφορία. Τα υποσύνολα αυτά αποτελούν ακριβώς τα στιγμιότυπα της ακολουθίας.

Για την κατασκευή μιας video ακολουθίας η κάμερα κινείται στο χώρο. Η κίνηση της αποτελείται γενικά από μεταφορική κίνηση στις 3 διαστάσεις του χώρου (X, Y, Z) και περιστροφική κίνηση γύρω από τους άξονες X, Y και Z. Ένας αριθμός υποθέσεων λαμβάνεται με σκοπό την απλοποίηση αυτής της κίνησης. Η πρώτη υπόθεση θεωρεί ότι η σκηνή είναι σταθερή, οι μεταβολές δηλαδή στο περιεχόμενο και στο φωτισμό της είναι μικρές μεταξύ των στάσεων. Επίσης η σχετική απόσταση όλων των αντικειμένων της σκηνής ως προς την κάμερα είναι ίδια. Θεωρούμε δηλαδή ότι το βάθος των αντικειμένων παραμένει σταθερό κατά την διάρκεια της λήψης. Η τρίτη υπόθεση θεωρεί ότι οι γωνίες περιστροφής της κάμερας γύρω από τον άξονά της είναι μικρές. Θεωρούμε επίσης την ύπαρξη μη-ισοτροπικής αλλαγής στο μήκος εστίασης της κάμερας. Ισοτροπική θεωρείται η εστίαση όταν η μεγέθυνση ή η σμίκρυνση προς τις δύο κατευθύνσεις x και y είναι ίδια. Υποθέτουμε ότι η αλλαγή αυτή της κλίμακας ρυθμίζεται από τον χρήστη ενώ ο αλγόριθμος εκτίμησης της κίνησης δεν προϋποθέτει την γνώση της. Η κίνηση τότε της κάμερας ανάγεται σε μία από τις παρακάτω κινήσεις ή συνδυασμό αυτών.

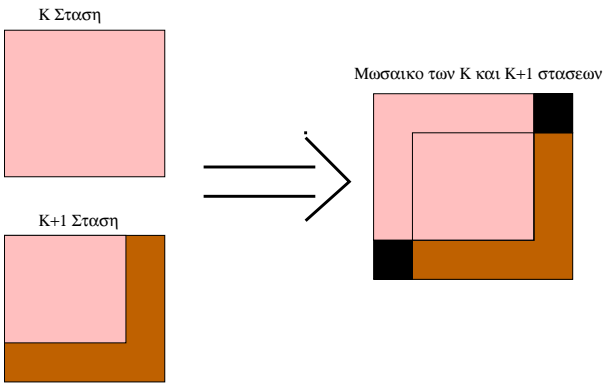
- Μεταφορική κίνηση παράλληλα στο επίπεδο της σκηνής.
- Μεταφορική κίνηση κάθετα στο επίπεδο της σκηνής. Η κίνηση αυτή ισοδυναμεί με την αλλαγή της απόστασης εστίασης της κάμερας (zoom).
- Περιστροφή γύρω από τον οπτικό άξονα της κάμερας και τους άξονες X, Y, Z.

Συλλέγοντας διαδοχικές λήψεις της κάμερας ενώ κινείται στον χώρο, με κάποια συγκεκριμένη συχνότητα δειγματοληψίας, προκύπτει η ακολουθία video.

Παρόλη την πλούσια πληροφορία που διαθέτει μια video ακολουθία η άμεση χρήση, εκμετάλλευση και ανάλυση αυτής της πληροφορίας δεν έχει κατορθωθεί. Αιτία, όπως ειπώθηκε, το μέγεθος της ακολουθίας, που από τη μια δημιουργεί πρόβλημα στην αποθήκευση της και από την άλλη κάνει χρονικά ανέφικτη την ανάλυση και εκμετάλλευση της πληροφορίας της, αφού απαιτείται πρόσβαση σε όλες τις στάσεις της. Το ”μωσαϊκό” αναιρεί τα παραπάνω προβλήματα, αφού μία μόνο εικόνα αναπαριστά όλη την πληροφορία της video ακολουθίας. Η πληροφορία αυτή, που βρίσκεται διασκορπισμένη στο σύνολο των στάσεων, συγκροτείται και απεικονίζεται σε μια εικόνα. Η

ανάκτηση και ανάλυση της πληροφορίας πραγματοποιείται επιτυχώς με την πρόσβαση στο ”μωσαϊκό”.

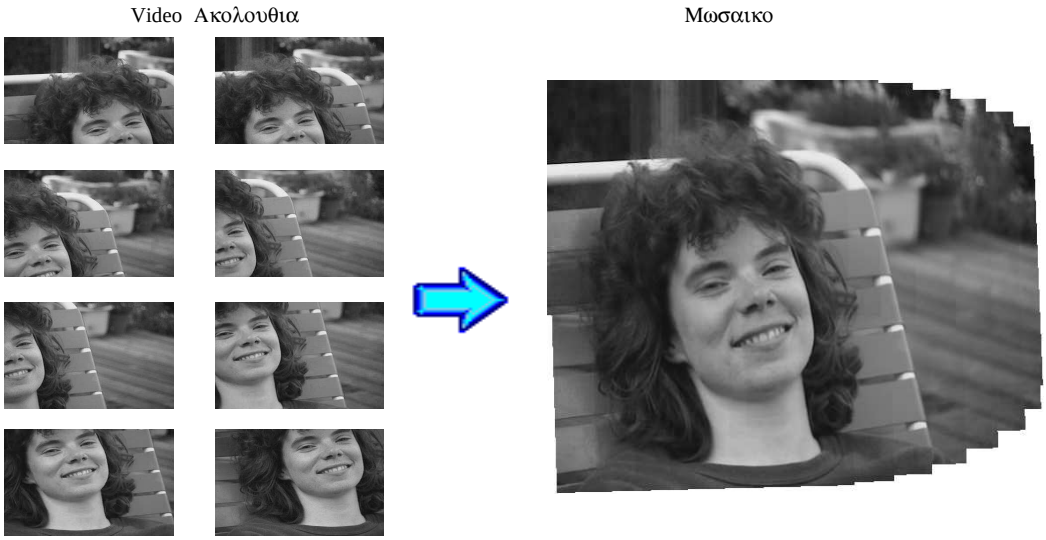
Λόγω της μικρής συχνότητας δειγματοληψίας που χρησιμοποιείται στην κατασκευή μιας video ακολουθίας οι διαδοχικές λήψεις της κάμερας περιέχουν μεγάλο ποσοστό ίδιας πληροφορίας του σκηνικού. Αποτέλεσμα αυτού είναι διαδοχικές στάσεις της video ακολουθίας να επικαλύπτονται μερικώς σε μεγάλο βαθμό. Το γεγονός αυτό εκμεταλλευόμαστε για την παραγωγή του ”μωσαϊκού” που προκύπτει από την κοινή περιοχή των στάσεων και από την επιπλέον πληροφορία του σκηνικού που το καθένα διαθέτει.



Σχήμα 2.2: Κατασκευή ”μωσαϊκού”

Κατασκευή του ”μωσαϊκού” από τις διαδοχικές στάσεις Κ και Κ+ 1 ενός στιγμιότυπου

Το σχήμα 2.2 απεικονίζει τον τρόπο σύνθεσης δύο διαδοχικών στάσεων ενός στιγμιοτύπου για την κατασκευή του ”μωσαϊκού”. Η κάμερα στην περίπτωση αυτή έχει κινηθεί προς τα κάτω και δεξιά, αφού η νέα πληροφορία στην Κ+1 στάση βρίσκεται προς την κατεύθυνση αυτή. Η κοινή περιοχή των στάσεων απεικονίζεται στο σχήμα με το ίδιο σχέδιο (pattern). Το ”μωσαϊκό” γενικά προκύπτει από την κοινή περιοχή των στάσεων καθώς και από την επιπλέον πληροφορία του καθενός. Στο σχήμα 2.2 το ”μωσαϊκό”, που κατασκευάζεται από τις στάσεις Κ και Κ+1, είναι ακριβώς το τετράπλευρο που φαίνεται, εκτός από την μαύρη περιοχή. Η περιοχή αυτή δηλώνει ότι δεν υπάρχει πληροφορία για το σκηνικό στα σημεία εκείνα, αφού η κάμερα δεν πραγματοποίησε λήψη προς την κατεύθυνση των σημείων αυτών. Ένα παράδειγμα κατασκευής ”μωσαϊκού” από μια πραγματική ακολουθία video φαίνεται στο σχήμα 2.3.



Σχήμα 2.3: Κατασκευή ”μωσαϊκού ” από μια πραγματική video ακολουθία

Αριστερά του σχήματος απεικονίζεται η video ακολουθία και δεξιά το ”μωσαϊκό” που κατασκευάστηκε από αυτή.

2.2 Προηγούμενες Προσεγγίσεις

Τα τελευταία χρόνια παρουσιάζεται ιδιαίτερο ενδιαφέρον από την πλευρά των ερευνητών για την κατασκευή ”μωσαϊκών” [9],[10],[11],[12],[13],[14],[15], [16],[17],[18],[19]. Πολλές φορές μεγάλα αντικείμενα δεν μπορούν να ληφθούν σε μία μόνο εικόνα όπως είναι η περίπτωση των αεροφωτογραφιών[16][17][20]. Χρησιμοποιώντας όμως φακούς με ευρύ πεδίο όψης όπως οι φακοί με ”μάτια ψαριού (fish-eye)” λύνεται το πρόβλημα με την κατασκευή μιας εικόνας που απεικονίζει την πανοραμική όψη της σκηνής [20]. Οι εικόνες όμως που λαμβάνονται με τέτοιους φακούς παρουσιάζουν ουσιαστική παραμόρφωση, και περιγράφουν την σκηνή με περιορισμένη ανάλυση.

Απλούστερα ”μωσαϊκά” κατασκευάστηκαν από ένα σύνολο εικόνων που προκύπτουν από τη μεταφορική κίνηση της κάμερας. Στο σύνολο των εικόνων αυτών ανήκουν οι δορυφορικές εικόνες. Η μεταφορική κίνηση της κάμερας ανιχνεύεται είτε χειρωνακτικά δείχνοντας τα αντίστοιχα σημεία, είτε με μεθόδους συσχέτισης. Άλλα, απλά ”μωσαϊκά” κατασκευάστηκαν με τη χρήση ειδικών μηχανισμών, περιστρέφοντας την κάμερα γύρω από το οπτικό της κέντρο. Οι μηχανισμοί αυτοί δημιουργούν την πανοραμική εικόνα που αναπαριστά την προβολή της σκηνής σε μια κυλινδρική επιφάνεια [21].

Σε γενικότερες κινήσεις της κάμερας που περιλαμβάνουν μεταφορά, μικρή περιστροφή, αλλαγή στο μήκος εστίασης, χρησιμοποιούνται και γενικότεροι μετασχηματι-

σμοί. Σε όλες τις περιπτώσεις όμως η διαδικασία κατασκευής του ”μωσαϊκού” χωρίζεται σε δύο βήματα. Στο πρώτο βήμα υπολογίζονται οι παράμετροι της κίνησης και στο δεύτερο βήμα κατασκευάζεται το ”μωσαϊκό”. Το πρόβλημα της ανίχνευσης της κίνησης χρονολογείται μακρύτερα από αυτό της κατασκευής του ”μωσαϊκού”. Ακολουθεί η παρουσίαση προηγούμενων προσεγγίσεων επίλυσης των δύο προβλημάτων χωριστά.

Για την ανίχνευση της κίνησης χρησιμοποιούνται συνήθως παραμετρικές μέθοδοι [9],[10],[11],[12],[13],[22]. Οι μέθοδοι αυτοί υποθέτουν ότι η κίνηση της κάμερας προσεγγίζεται με ένα παραμετρικό μοντέλο και προσπαθούν στην συνέχεια να εκτιμήσουν τις παραμέτρους του μοντέλου. Ένα από τα πιο γενικά μοντέλα κίνησης που χρησιμοποιήθηκαν στην παρούσα εργασία είναι το προβολικό μοντέλο. Η πολυπλοκότητα που εμφανίζει η χρήση του, στρέφει πολλούς ερευνητές στην χρήση απλούστερων μοντέλων όπως το αφινικό. Το προβολικό μοντέλο χρησιμοποιήθηκε κυρίως για την εκτίμηση μόνο τρισδιάστατης κίνησης (3D). Οι Faugeras και Lustman [23], οι Shashua και Navab [24], και ο Sawhney [25] μελέτησαν το πρόβλημα της εκτίμησης των προβολικών παραμέτρων ενώ υπολόγιζαν την κίνηση ενός επιμέρους κομματιού της εικόνας σαν μέρος του μεγαλύτερου προβλήματος της εύρεσης της τρισδιάστατης κίνησης. Ο Kumar et al.[26] καθώς και ο Zhang [27] προτείνουν την εκτίμηση των προβολικών παραμέτρων υπολογίζοντας την ροή κατά μήκος των επιπολικών γραμμών. Η χρήση όμως μόνο της γεωμετρικής πληροφορίας οδηγεί συχνά σε λάθη εκτίμησης. Ιδανικότερη λύση αποτελεί ο συνδυασμός των πληροφοριών από τη μια της γεωμετρικής και από την άλλη της πληροφορίας που προκύπτει από τις φωτεινότητες.

Τα τελευταία χρόνια πραγματοποιείται στροφή των μεθόδων εκτίμησης της κίνησης στην χρήση σθεναρών αλγορίθμων αφού τα δεδομένα είναι αναπόφευκτα επιρρεπή σε λάθη [24],[22],[28]. Ωστόσο οι περισσότερες από τις μέθοδους σθεναρής παλινδρόμησης που έχουν προταθεί επηρεάζονται από την ύπαρξη κακών δεδομένων και οδηγούν σε λανθασμένη εκτίμηση. Η πιο διαδεδομένη μέθοδος σθεναρής εκτίμησης είναι ο αλγόριθμος M-estimator, που χρησιμοποιήθηκε και στην παρούσα εργασία. Οι Black και Anandan [29] εφάρμοσαν τον αλγόριθμο αυτόν για τον υπολογισμό της κυρίαρχης κίνησης με χρήση δισδιάστατου παραμετρικού μοντέλου. Ωστόσο το πρόβλημα της εκτίμησης της παραμέτρου κλίμακας, που εμφανίζεται στις συναρτήσεις κόστους κατά την εφαρμογή του M-estimator, δεν λύθηκε. Στα πειράματά τους χρησιμοποίησαν προκαθορισμένη αρχική τιμή για την παράμετρο αυτή καθώς και προ-αποφασισμένο σχέδιο μείωσης αυτής σε κάθε επανάληψη. Υπολογιστικά ακριβή, αλλά πιο σθεναρή, εναλλακτική του M-estimator είναι η χρήση της LMedS (Least Median of Square) προσέγγισης, λόγω

της ικανότητάς της να απορρίπτει υψηλό ποσοστό ”κακών” δεδομένων (outliers). Την προσέγγιση αυτή χρησιμοποίησε ο Ayer [30]. Ο Szeliski [12] υπολόγισε τις παραμέτρους του προβολικού μοντέλου χρησιμοποιώντας τον αλγόριθμο M-estimator με την μέθοδο των ελαχίστων τετραγώνων. Η μέθοδος όμως αυτή είναι επιρρεπής σε λάθη αφού δεν διακρίνει την ύπαρξη ”κακών” δεδομένων (outliers).

Για την κατασκευή του ”μωσαϊκού” έχουν προταθεί πολλές μέθοδοι που βασίζονται στην σύνθεση των στάσεων με χρήση φιλτραρίσματος στο χρόνο [9],[12],[1]. Ο Tonomura et al. [31] χρησιμοποίησε μοντέλο κίνησης με τρεις παραμέτρους που περιγράφει μεταφορά στις δύο διαστάσεις και αλλαγή στο μήκος εστίασης. Την εικόνα, που δημιούργησε με την σύνθεση των στάσεων της video ακολουθίας, την ονόμασε VideoSpaceIcon. Ο Nicolas [13] χρησιμοποίησε το ίδιο παραμετρικό μοντέλο για την εκτίμηση της κίνησης μεταξύ κάθε στάσης και του εκάστοτε ”μωσαϊκού”. Θεώρησε όμως ότι είχε προηγηθεί εξαγωγή των ανεξάρτητων κινούμενων αντικειμένων σε όλη την ακολουθία. Στην παρούσα εργασία επιτυγχάνεται η άμεση εκτίμηση της κυρίαρχης κίνησης χωρίς να προηγείται εξαγωγή περιοχών με κινούμενα αντικείμενα. Κατάλληλες συναρτήσεις επιδρούν και απορρίπτουν αυτές τις περιοχές.

Η διόρθωση όμως των κακών ευθυγραμμίσεων που εμφανίζονται συχνά στο τελικό ”μωσαϊκό” δεν αποτέλεσε θέμα προς έρευνα από τους προηγούμενους. Ο Sawhney, Hsu και Kumar [10] πρότειναν την ύπαρξη της τοπολογίας των στάσεων της ακολουθία για την αποφυγή των κακών ευθυγραμμίσεων. Μετά την κατασκευή της τοπολογίας όλων των στάσεων της ακολουθίας, λαμβάνει μέρος η αναζήτηση γειτονικών μη-διαδοχικών στάσεων που διαφέρουν στο χρόνο λήψης τους. Υπολογίζεται κατόπιν η κίνηση μεταξύ των γειτονικών στάσεων και κατασκευάζεται ο γράφος ελαχίστης διαδρομής που συνδέει όλες τις στάσεις της ακολουθίας. Ο τελικός γράφος χρησιμοποιείται για την ευθυγράμμιση των στάσεων και την κατασκευή του ”μωσαϊκού”. Ο διαχωρισμός όμως διαδοχικών στάσεων στον γράφο οδηγεί στη λανθασμένη ευθυγράμμιση αυτών που είναι εμφανή κατά την τοποθέτησή τους στο ”μωσαϊκό”. Διαφορετική προσέγγιση με χρήση της τοπολογίας παρουσιάζεται στην παρούσα εργασία.

Κεφάλαιο 3

Εκτίμηση της κίνησης της κάμερας

3.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται οι μέθοδοι εκτίμησης της κίνησης μιας video ακολουθίας που υλοποιήθηκαν στην παρούσα εργασία. Στην αρχή όμως της παρουσίασης είναι σημαντική η περιγραφή του προβλήματος που καλούμαστε να αντιμετωπίσουμε. Για την καλύτερη μελέτη του προβλήματος απαραίτητη είναι η αναφορά στις πιο γνωστές κινήσεις της κάμερας στον τρισδιάστατο χώρο. Πληροφορία για την εκτίμηση των κινήσεων αυτών αποτελούν οι στάσεις της ακολουθίας. Χρησιμοποιώντας το γεωμετρικό μοντέλο κίνησης προκύπτει ένα σύνολο παραμετρικών μοντέλων που περιγράφουν τις κινήσεις αυτές. Η εκτίμηση των κινήσεων της κάμερας επιτυγχάνεται με τον υπολογισμό των παραμέτρων του εκάστοτε μοντέλου. Ο υπολογισμός αυτός πραγματοποιείται με τη χρήση σθεναρών μεθόδων, όπως ο αλγόριθμος M-estimator, ή με τη χρήση αναλλοίωτων χαρακτηριστικών των στάσεων, όπως οι σχέσεις ροπών μιας επιφάνειας. Ακολουθεί η παρουσίαση μεθόδων διόρθωσης των παραμέτρων που υπολογίστηκαν από τις μεθόδους εκτίμησης της κίνησης. Οι μέθοδοι διόρθωσης που χρησιμοποιήθηκαν αποτελούν παραλλαγές της μεθόδου της απότομης κλίσης και της μεθόδου της παλινδρόμησης.

3.2 Ορισμός και περιγραφή του προβλήματος

Ένα από τα βασικότερα προβλήματα που εμφανίζονται στον τομέα της ανάλυσης ακολουθιών εικόνων είναι η ύπαρξη του δυναμικού παρατηρητή ή αλλιώς της κάμερας. Ο δυναμικός παρατηρητής δεν παραμένει σταθερός σε κάποιο σημείο αλλά κινείται χωρίς

περιορισμό στον τρισδιάστατο χώρο. Η κίνηση αυτή του παρατηρητή είναι γνωστή ως **ιδία κίνηση(egomotion)**. Αντικείμενο αυτού του κεφαλαίου αποτελεί η εκτίμηση της κίνησης του παρατηρητή (κάμερα).

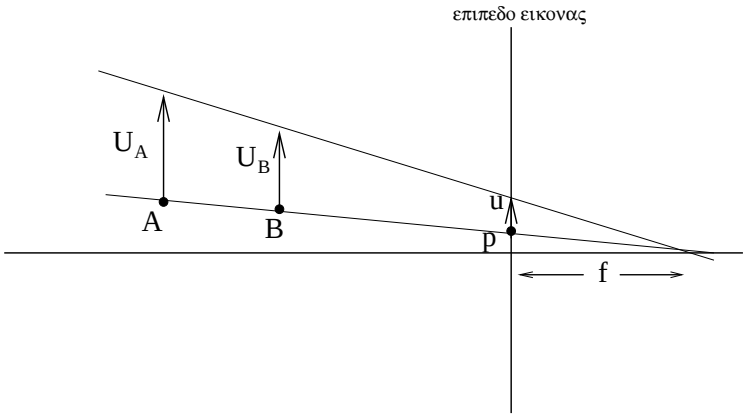
Η εκτίμηση της κίνησης αυτής αποτελεί ένα από τα δυσκολότερα προβλήματα στο τομέα της ανάλυσης ακολουθιών εικόνων. Η δυσκολία οφείλεται στην ελλειπή πληροφορία που διαθέτουμε. Πληροφορία μας είναι το επίπεδο της εικόνας που αντιστοιχεί στην προβολή της κίνησης της κάμερας σε χώρο δύο διαστάσεων. Με τη σύγκριση των εντάσεων των διαδοχικών εικόνων εκτιμάται η κίνηση της κάμερας. Στην πραγματικότητα εκτιμάται η προβολή της κίνησης στο επίπεδο της εικόνας και όχι η ιδία κίνηση της κάμερας στον τρισδιάστατο χώρο. Ζητάμε δηλαδή πληροφορία για τρισδιάστατη κίνηση ενώ διαθέτουμε μόνο δισδιάστατα δεδομένα.

Παρόλη τη δυσκολία του προβλήματος η αναγκαιότητα της λύσης του είναι μεγάλη, αφού στις περισσότερες εφαρμογές που χρησιμοποιείται η πληροφορία της video ακολουθίας, απαιτείται και η εκτίμηση της κίνησης της κάμερας. Για την κατασκευή, για παράδειγμα, του "μωσαϊκού" απαραίτητη είναι η ευθυγράμμιση των στάσεων της ακολουθίας, που επιτυγχάνεται με την εκτίμηση της κίνησης μεταξύ διαδοχικών στάσεων.

Γενικά η κίνηση της κάμερας αναλύεται στις τρεις συνιστώσες της μεταφορικής κίνησης μία για κάθε άξονα (X, Y, Z) του τρισδιάστατου χώρου και στις τρεις επιπλέον συνιστώσες της περιστροφής της γύρω από τους ίδιους άξονες. Άρα στη γενική περίπτωση η κίνηση της κάμερας έχει έξι βαθμούς ελευθερίας. Η εκτίμηση της επομένως έγκειται στον υπολογισμό των βαθμών αυτών. Με την προβολή όμως της τρισδιάστατης σκηνής στο επίπεδο της εικόνας, παρουσιάζεται το γνωστό πρόβλημα του "scale factor ambiguity", που σαν αποτέλεσμα έχει τη μείωση των βαθμών ελευθερίας. Σύμφωνα με το πρόβλημα αυτό διαφορετικά σημεία στον τρισδιάστατο χώρο με διαφορετικό βάθος (διαφορετική δηλαδή συνιστώσα μεταφορικής κίνησης στον άξονα Z) (σχήμα 3.1) προβάλλονται στο ίδιο σημείο στο επίπεδο της εικόνας. Το αποτέλεσμα αυτής της προβολής είναι ο μη ακριβής προσδιορισμός του διανύσματος της μεταφορικής κίνησης για κάθε σημείο του τρισδιάστατου χώρου. Το διάνυσμα της μεταφορικής κίνησης προσεγγίζεται μόνο μέχρι κάποιο παράγοντα κλιμάκωσης. Έτσι αν T είναι το διάνυσμα της πραγματικής μεταφορικής κίνησης και T_1 το διάνυσμα που υπολογίζουμε, ισχύει η ακόλουθη σχέση:

$$\vec{T_1} = \lambda \vec{T} \quad \text{όπου} \quad \lambda \epsilon \Re$$

Αντίθετα η περιστροφική κίνηση που υπολογίζεται είναι ακριβώς η πραγματική. Επομένως η εκτίμηση της κίνησης της κάμερας επιτυγχάνεται με τον υπολογισμό των



Σχήμα 3.1: Scale factor ambiguity

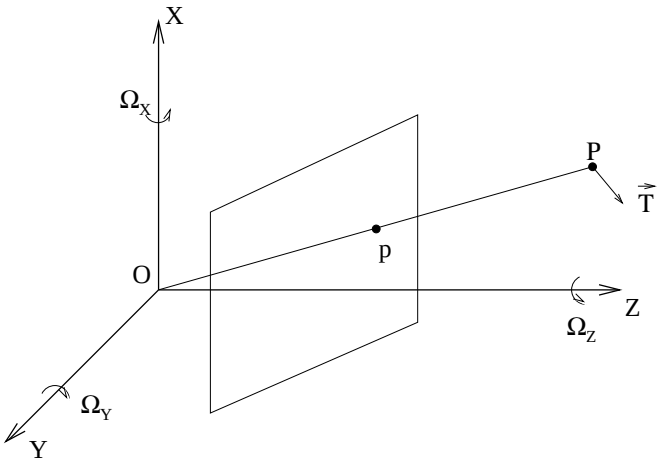
τριών συνιστωσών της περιστροφικής κίνησης και του λόγου των δύο συνιστωσών της μεταφορικής κίνησης ως προς την τρίτη συνιστώσα (αυτή του άξονα Z που δηλώνει το βάθος). Οι έξι τότε βαθμοί ελευθερίας για την εκτίμηση της κίνησης της κάμερας ανάγονται σε πέντε. Παρακάτω θα αναφερθούμε με περισσότερη λεπτομέρεια στην τρισδιάστατη κίνηση της κάμερας και την αντιστοιχία της με την κίνηση που απεικονίζεται στο επίπεδο της εικόνας.

3.2.1 Τρισδιάστατο Μοντέλο Κίνησης

Θεωρώντας ότι η κάμερα κινείται στον τρισδιάστατο χώρο χωρίς περιορισμούς, το γενικότερο μοντέλο κίνησης που περιγράφει την κίνηση της είναι το τρισδιάστατο μοντέλο (σχήμα 3.2). Για να οριστεί πλήρως ένα τρισδιάστατο μοντέλο κίνησης απαιτείται να καθοριστούν τα στιγμιαία διανύσματα της μετατόπισης και της περιστροφής μέσα στο τρισδιάστατο χώρο. Ονομάζοντας τα διανύσματα αυτά, T και Ω αντίστοιχα, το διάνυσμα V της τρισδιάστατης κίνησης για το σημείο P με συντεταγμένες (X, Y, Z) στο χώρο, δίνεται από τη σχέση:

$$V = T + \Omega \wedge \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.1}$$

Υποθέτοντας ότι Ω_X, Ω_Y, Ω_Z και T_X, T_Y, T_Z είναι οι συνιστώσες της περιστροφικής και της μεταφορικής κίνησης, η σχέση 3.1 αναλύεται στην εξής:



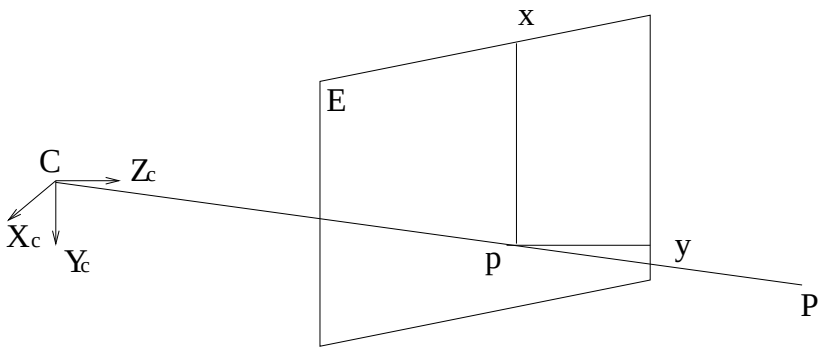
Σχήμα 3.2: Τρισδιάστατο μοντέλο κίνησης

$$\vec{V} = \begin{bmatrix} \frac{dX}{dt} \\ \frac{dY}{dt} \\ \frac{dZ}{dt} \end{bmatrix} = \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix} + \begin{bmatrix} \Omega_Y Z - \Omega_Z Y \\ \Omega_Z X - \Omega_X Z \\ \Omega_X Y - \Omega_Y X \end{bmatrix} \tag{3.2}$$

Η δυσκολία εκτίμησης της πλήρης τρισδιάστατης κίνησης ώθησε τους ερευνητές στη λήψη υποθέσεων για την απλοποίησή της. Λαμβάνοντας υπόψη το σύνολο των video ακολουθιών και μελετώντας την κίνηση της κάμερας σε αυτές, κατέληξαν στο συμπέρασμα ότι οι πιο συνηθισμένες μορφές κινήσεων είναι από τη μια η μεταφορική κίνηση στις τρεις διαστάσεις και από την άλλη η περιστροφική κίνηση γύρω από τον οπτικό άξονα. Οι μορφές αυτές των κινήσεων της κάμερας μελετήθηκαν στην παρούσα εργασία. Η υπόθεση που λαμβάνεται για τον περιορισμό της κίνησης απαιτεί τα αντικείμενα του σκηνικού να βρίσκονται στο ίδιο βάθος ή καλύτερα η απόσταση τους από την κάμερα να είναι ίδια. Η μελέτη της προβολής του τρισδιάστατου σκηνικού στο επίπεδο είναι απαραίτητη για την εκτίμηση των μορφών αυτών κίνησης.

3.2.2 Προβολή του τρισδιάστατου χώρου στο επίπεδο

Το οπτικό μας σύστημα καθώς και κάθε άλλος αισθητήρας όρασης αναπαριστά το φυσικό περιβάλλον σε χώρο τριών διαστάσεων. Αισθητήρας όρασης θεωρείται και η κάμερα η οποία απεικονίζει τον τρισδιάστατο χώρο σε ένα επίπεδο. Κάθε κάμερα μοντελοποιείται με την οπτική της εστία C και το επίπεδο E της εικόνας. Το πιο γνωστό μοντέλο κάμερας καλείται *"μάτι της βελόνας" (pinhole).*



Σχήμα 3.3: Μοντέλο κάμερας, το μάτι της βελόνας(pinhole)

Ο άξονας που διέρχεται απ' την οπτική εστία και το κέντρο της εικόνας και επιπλέον είναι κάθετος στο επίπεδό της, ονομάζεται **οπτικός άξονας**, ενώ η απόσταση της οπτικής εστίας απ' το επίπεδο της εικόνας ονομάζεται **εστιακή απόσταση**. Ορίζοντας ένα τρισδιάστατο σύστημα συντεταγμένων <X, Y, Z> όπως το σχήμα 3.3, οι σχέσεις που συνδέουν τις συντεταγμένες (X, Y, Z) ενός σημείου P και της προοπτικής προβολής του, (x, y), στο επίπεδο της εικόνας είναι [32]:

$$x = F_x \frac{X}{Z} + K_d F_x \frac{X}{Z} \left(\frac{X^2}{Z^2} + \frac{Y^2}{Z^2}\right)$$

και

$$y = F_y \frac{Y}{Z} + K_d F_y \frac{X}{Z} \left(\frac{X^2}{Z^2} + \frac{Y^2}{Z^2}\right)$$

Στο σχήμα 3.3 ο άξονας Z_c ταυτίζεται με τον οπτικό άξονα και η αρχή των αξόνων με την οπτική εστία C.

- F_x, F_y είναι οι λόγοι της εστιακής παραμέτρου f ως προς το οριζόντιο ή κάθετο μήκος του δειγματοληπτικού pixel αντίστοιχα. Το pixel (picture element) είναι η μονάδα μέτρησης μιας εικόνας.

  $$F_x = \frac{f}{l_x} \qquad F_y = \frac{f}{l_y}$$

  όπου l_x και l_y είναι συντελεστές, ανάλογοι του μεγέθους του στοιχειώδους pixel.

- K_d είναι η ακτινική παραμόρφωση της κάμερας.

Συνήθως όμως σε ένα μεγάλο φάσμα ερευνών, η ακτινική παραμόρφωση αγνοείται και η εστιακή απόσταση κανονικοποιείται ($F_x = 1$, $F_y = 1$). Έτσι οι παραπάνω εξισώσεις απλοποιούνται στις:

$$x = \frac{X}{Z} \qquad \text{και} \qquad y = \frac{Y}{Z} \tag{3.3}$$

3.3 Οπτική Ροή και Δισδιάστατο Πεδίο Ταχυτήτων

Η κίνηση που γίνεται αντιληπτή από την πληροφορία των φωτεινοτήτων των εικόνων ονομάζεται φαινόμενη κίνηση ή οπτική ροή. Η φαινόμενη κίνηση δεν είναι ακριβώς η άμεση προβολή του τρισδιάστατου πεδίου ταχυτήτων. Η κίνηση αυτή απορρέει από μία πολύπλοκη διεργασία που λαμβάνει υπόψη εκτός από την πραγματική κίνηση, το φωτισμό της σκηνής, τις ανταναακλαστικές ιδιότητες των αντικειμένων που συνθέτουν τη σκηνή, το σχήμα των επιφανειών, καθώς και τα χαρακτηριστικά του παρατηρητή που χρησιμοποιείται για την απόκτηση των εικόνων. Η οπτική ροή μπορεί να αποκτηθεί παρατηρώντας τις χωρο-χρονικές μεταβολές της έντασης. Η αρχή, πάνω στην οποία στηρίζεται η χρήση της για την εκτίμηση της κίνησης, είναι η διατήρηση της έντασης της φωτεινότητας στο χρόνο. Αν υποθέσουμε ότι $(\Delta_x,\ \Delta_y)$ είναι το διάνυσμα της ταχύτητας, που εκφράζεται από τη μετατόπιση των σημείων της εικόνας, η εξίσωση που εκφράζει τη διατήρηση της φωτεινότητας κατά τη διάρκεια του χρόνου είναι:

$$I(x,\ y;t)\ = I(x-\Delta_x,\ y-\Delta_y;t-\Delta_t) \tag{3.4}$$

Αν το διάστημα Δ_t είναι μικρό, η πρώτη σειρά Taylor της εξίσωσης 3.4 δίνει τη γνωστή εξίσωση της οπτικής ροής:

$$I_x(x,\ y;t)\ u\ +\ I_y(x,\ y;t)\ v\ +\ I_t(x,\ y;t)\ = 0 \tag{3.5}$$

όπου $I_x,\ I_y,\ I_t$ είναι οι χωρο-χρονικές παράγωγοι της έντασης στην εικόνα και (u, v) είναι το διάνυσμα της οπτικής ροής.

Αντίθετα η κίνηση που προκύπτει με προβολή της τρισδιάστατης πραγματικής κίνησης στο επίπεδο της εικόνας καλείται δισδιάστατο πεδίο ταχυτήτων. Η οπτική ροή χρησιμοποιείται συχνά για την εκτίμηση του πεδίου αυτού. Η εξίσωση της οπτικής ροής καθώς και η σχέση που εκφράζει τη διατήρηση της φωτεινότητας στο χρόνο αποτελούν κριτήρια για την εκτίμηση της οπτικής ροής σε κάθε σημείο της εικόνας. Ένα από τα παραδείγματα αποτυχίας της οπτικής ροής να εκτιμήσει την κίνηση της κάμερας είναι η περίπτωση της σφαίρας που περιστρέφεται γύρω από τον άξονά της και που η ένταση της φωτεινότητας είναι ομοιόμορφη σε κάθε σημείο της. Εξαιτίας της μη-μεταβολής της φωτεινότητας το κριτήριο της οπτικής ροής δεν καταφέρνει να ανιχνεύσει την περιστροφική κίνηση. Για την εκτίμηση του δισδιάστατου πεδίου ταχυτήτων απαραίτητη είναι η μελέτη του γεωμετρικού μοντέλου κίνησης που ακολουθεί. Το μοντέλο αυτό συνδέει την κίνηση της κάμερας με τη μετατόπιση των συντεταγμένων των σημείων της εικόνας.

3.3.1 Γεωμετρικό Μοντέλο

Το γεωμετρικό μοντέλο κίνησης παρέχει σε κάθε περίπτωση την προβολή της πραγματικής κίνησης πάνω στο επίπεδο της εικόνας, την οποία και ονομάσαμε δισδιάστατο πεδίο ταχυτήτων [32]. Ακολουθεί η εκτίμηση του πεδίου αυτού στην περίπτωση της προοπτικής και παράλληλης προβολής. Η σχέση που συνδέει τις συντεταγμένες της εικόνας (x, y) με τις αντίστοιχες συντεταγμένες του τρισδιάστατου χώρου (X, Y, Z) είναι,

- στην προοπτική προβολή:

$$x = \frac{X}{Z} \qquad \text{και} \qquad y = \frac{Y}{Z}$$

- ενώ στην παράλληλη προβολή:

$$x = X \qquad \text{και} \qquad y = Y$$

Υποθέτουμε ότι $\vec{V} = (\frac{dX}{dt}, \frac{dY}{dt}, \frac{dZ}{dt})$ είναι το διάνυσμα της ταχύτητας του σημείου (X, Y, Z) στον τρισδιάστατο χώρο, το διάνυσμα της ταχύτητας $(\frac{dx}{dt}, \frac{dy}{dt})$ σε κάθε σημείο της εικόνας (x, y) εξαρτάται από τη μορφή προβολής. Για την περίπτωση της προοπτικής προβολής είναι:

$$\begin{aligned} \frac{dx}{dt} &= \frac{1}{Z}\frac{dX}{dt} - \frac{X}{Z^2}\frac{dZ}{dt} \\ \frac{dy}{dt} &= \frac{1}{Z}\frac{dY}{dt} - \frac{Y}{Z^2}\frac{dZ}{dt} \end{aligned} \tag{3.6}$$

ενώ στην περίπτωση της παράλληλης ισχύει:

$$\begin{aligned} \frac{dx}{dt} &= \frac{dX}{dt} \\ \frac{dy}{dt} &= \frac{dY}{dt} \end{aligned} \tag{3.7}$$

Θεωρείται ότι η επιφάνεια που λαμβάνει η κάμερα τοπικά προσεγγίζεται από ένα επίπεδο του οποίου η εξίσωση δίνεται από τη σχέση:

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

Το επίπεδο αυτό δεν κόβει την αρχή των αξόνων. Στην αντίθετη περίπτωση το επίπεδο προβάλλεται σε μία ευθεία γραμμή της εικόνας. Εφαρμόζοντας την εξίσωση 3.3 στην εξίσωση του επιπέδου, προκύπτει:

$$\frac{1}{Z} = n_x x + n_y y + n_z \tag{3.8}$$

Το διάνυσμα τότε της ταχύτητας σε κάθε σημείο της εικόνας, στην περίπτωση της προοπτικής προβολής προκύπτει με εφαρμογή των σχέσεων 3.2 και 3.8 στη σχέση 3.6, και

είναι το ακόλουθο:

$$\begin{aligned}\frac{dx}{dt} &= (n_Z\,T_X + \Omega_Y) + (n_X\,T_X - n_Z\,T_Z)\,x + (n_Y\,T_X - \Omega_Z)\,y \\ &\quad + (\Omega_Y - n_X\,T_Z)\,x^2 - (n_Y\,T_Z + \Omega_X)\,xy \\ \frac{dy}{dt} &= (n_Z\,T_Y - \Omega_X) + (n_X\,T_Y + \Omega_Z)\,x + (n_Y\,T_Y - n_Z\,T_Z)\,y \\ &\quad + (\Omega_Y - n_X\,T_Z)\,xy - (n_Y\,T_Z + \Omega_X)\,y^2\end{aligned} \tag{3.9}$$

ενώ στην περίπτωση της παράλληλης προβολής με εφαρμογή των σχέσεων 3.2 και 3.8 στη σχέση 3.7, είναι:

$$\begin{aligned}\frac{dx}{dt} &= T_X + \frac{\Omega_Y}{n_Z} - \frac{n_X}{n_Z}\,\Omega_Y\,x - (\frac{n_Y}{n_z}\,\Omega_Y + \Omega_Z)\,y \\ \frac{dy}{dt} &= T_Y + \frac{\Omega_X}{n_Z} + (\frac{n_X}{n_Z}\,\Omega_X + \Omega_Z)\,x + \frac{n_Y}{n_z}\,\Omega_X\,y\end{aligned} \tag{3.10}$$

Η εξίσωση 3.9 ορίζει ένα πολύπλοκο μη-γραμμικό μοντέλο κίνησης με οκτώ βαθμούς ελευθερίας. Αντίθετα η εξίσωση 3.10 ορίζει ένα απλούστερο γραμμικό μοντέλο κίνησης. Αντικαθιστώντας στις εξισώσεις 3.9 και 3.10 τους συντελεστές των συντεταγμένων x, y ή τους συνδυασμούς αυτών με πραγματικές παραμέτρους προκύπτει ένα σύνολο παραμετρικών μοντέλων κίνησης. Κάθε μοντέλο περιγράφει ένα διαφορετικό είδος κίνησης της κάμερας.

3.4 Παραμετρικά μοντέλα κίνησης

Μεγάλο ενδιαφέρον παρουσιάζουν για τους ερευνητές οι μέθοδοι εκτίμησης της κίνησης με χρήση παραμετρικών μοντέλων. Τα μοντέλα αυτά κίνησης περιγράφουν συνολικά την κίνηση της κάμερας όπως προβάλλεται στο επίπεδο της εικόνας παρέχοντας το δισδιάστατο πεδίο ταχυτήτων. Η εκτίμηση της κίνησης που επιτυγχάνεται με αυτές τις μεθόδους τείνει σε μεγάλο βαθμό στην πραγματική κίνηση, κάτω από την προϋπόθεση ότι η επιλογή του παραμετρικού μοντέλου είναι ορθή, δηλαδή ότι το επιλεγμένο μοντέλο περιγράφει ικανοποιητικά την κίνηση της κάμερας. Η αποτελεσματικότητα των μεθόδων αυτών οδήγησε στη δημιουργία ενός σημαντικού αριθμού παραμετρικών μοντέλων.

Το σύνολο των γνωστών παραμετρικών μοντέλων κίνησης προκύπτει από τις εξισώσεις 3.9 και 3.10, με την αντικατάσταση, των συντελεστών των συντεταγμένων των σημείων (x, y) της εικόνας ή τους συνδυασμούς αυτών (xy, x^2, y^2) με πραγματικές παραμέτρους. Οι υποθέσεις που λαμβάνονται για τον προσδιορισμό της κίνησης της κάμερας ορίζουν και τις τιμές των παραμέτρων. Παράμετροι που εκφράζουν διαφορετική κίνηση από αυτή που προσδιορίζεται μηδενίζονται. Επομένως από τη μια το σύνολο των παραμέτρων ενός μοντέλου και από την άλλη η πολυπλοκότητα του μοντέλου εξαρτάται

Πίνακας 3.1: Μετασχηματισμοί Συντεταγμένων

Μοντέλο	**Μετασχηματισμός Συντεταγμένων από το x στο x'**	Παράμετροι
Μεταφορά	$\mathbf{x'} = \mathbf{x} + \mathbf{b}$	$\mathbf{b} \in \Re^2$
Εστίαση και Μεταφορά	$\mathbf{x'} = \mathbf{q}\,\mathbf{x} + \mathbf{b}$	$q \in \Re$, $b \in \Re^{2x2}$
Αφινικό	$\mathbf{x'} = \mathbf{A}\mathbf{x} + \mathbf{b}$	$\mathbf{A} \in \Re^{2\times 2}$, $\mathbf{b} \in \Re^2$
Bilinear	$x' = q_{x'xy}xy + q_{x'x}x + q_{x'y}y + q_{x'}$ $y' = q_{y'xy}xy + q_{y'x}x + q_{y'y}y + q_{y'}$	$q_* \in \Re$
Projective	$\mathbf{x'} = \frac{\mathbf{A}\,\mathbf{x} + \mathbf{b}}{\mathbf{c^T}\,\mathbf{x} + 1}$	$\mathbf{A} \in \Re^{\mathbf{2\times 2}}$, $\mathbf{b}, \mathbf{c} \in \Re^{\mathbf{2}}$
Relative projective	$\mathbf{x'} = \frac{\mathbf{A}\,\mathbf{x} + \mathbf{b}}{\mathbf{c^T}\,\mathbf{x} + 1} + \mathbf{x}$	$\mathbf{A} \in \Re^{\mathbf{2\times 2}}$, $\mathbf{b}, \mathbf{c} \in \Re^{\mathbf{2}}$
Pseudo perspective	$x' = q_{x'x}x + q_{x'y}y + q_{x'} + q_\alpha x^2 + q_\beta xy$ $y' = q_{y'x}x + q_{y'y}y + q_{y'} + q_\alpha xy + q_\beta y^2$	$q_* \in \Re$
Biquadratic	$x' = q_{x'x^2}x^2 + q_{x'xy}xy + q_{x'y^2}y^2 + q_{x'x}x + q_{x'y}y + q_{x'}$ $y' = q_{y'x^2}x^2 + q_{y'xy}xy + q_{y'y^2}y^2 + q_{y'x}x + q_{y'y}y + q_{y'}$	$q_* \in \Re$

από την κίνηση που αυτό καλείται να περιγράψει. Το σύνολο των γνωστών κινήσεων ορίζει και ένα σύνολο παραμετρικών μοντέλων.

Τα παραμετρικά μοντέλα εκφράζουν ένα μετασχηματισμό συντεταγμένων. Ο μετασχηματισμός αυτός απεικονίζει τις συντεταγμένες των σημείων μιας εικόνας $\mathbf{x} = [x,\ y]^T$ σε ένα νέο σύνολο συντεταγμένων $\mathbf{x'} = [x', y']^T$. Στις μεθόδους εκτίμησης της κίνησης μεταξύ δύο εικόνων, το σύνολο των νέων συντεταγμένων $[x',\ y']$ ανήκει στα αντίστοιχα σημεία της δεύτερης εικόνας που προέκυψε από την κίνηση της κάμερας στην ίδια σκηνή. Το δισδιάστατο διάνυσμα της ταχύτητας $(\frac{dx}{dt},\ \frac{dy}{dt})$ που φαίνεται στις εξισώσεις 3.9 και 3.10, εκφράζει τη μετακίνηση του σημείου $[x,\ y]$ της μιας εικόνας στο αντίστοιχο σημείο $[x',\ y']$ της δεύτερης λόγω της κίνησης της κάμερας, δηλαδή: $\begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \begin{bmatrix} x' - x \\ y' - y \end{bmatrix}$ Το σύνολο των γνωστών μετασχηματισμών παρουσιάζεται στον πίνακα 3.1 [11].

Είναι φανερό ότι η εκτίμηση της κίνησης έγκειται στον υπολογισμό των παραμέτρων ενός από τους μετασχηματισμούς του πίνακα 3.1. Ο αριθμός των παραμέτρων ποικίλει από δύο ως δώδεκα εξαρτώμενος από την πολυπλοκότητα της κίνησης που περιγράφουν.

3.4.1 Περιγραφή και σύγκριση παραμετρικών μοντέλων κίνησης

Η περιγραφή και η σύγκριση των παραμετρικών μοντέλων και κατ' επέκταση των μετασχηματισμών, είναι απαραίτητη για την επιλογή του καταλληλότερου μοντέλου για

κάθε μορφή κίνησης που παρουσιάζεται. Το σύνολο των πιο γνωστών μετασχηματισμών όπως αναφέραμε απεικονίζεται στον πίνακα 3.1. Παρακάτω ακολουθεί περιγραφή και σύγκριση των μετασχηματισμών αυτών.

Παρατηρώντας το σύνολο των ακολουθιών συμπεραίνουμε ότι η πιο συχνά εμφανιζόμενη κίνηση της κάμερας είναι η μεταφορική. Η κίνηση αυτή, που εκφράζεται ως η μετατόπιση των συντεταγμένων μιας εικόνας κατά σταθερό παράγοντα, περιγράφεται από τον πρώτο μετασχηματισμό του πίνακα και καλείται *μοντέλο μεταφοράς*. Παρόλο που το μοντέλο της μεταφοράς δεν απαιτεί περιορισμούς και αποτελεί τον ευκολότερο ως προς την εφαρμογή μετασχηματισμό παρουσιάζεται ακατάλληλο για την περιγραφή μεταβολών που οφείλονται για παράδειγμα στην αλλαγή της εστιακής απόστασης της κάμερας, στην περιστροφή και κλίση της κάμερας προς μία κατεύθυνση. Ο μετασχηματισμός αυτός προκύπτει από την εξίσωση 3.10, μηδενίζοντας τις συνιστώσες της περιστροφικής κίνησης. Η σχέση που προκύπτει τότε είναι:

$$\frac{dx}{dt} = T_x \qquad \frac{dy}{dt} = T_y$$

Οι συνιστώσες T_x και T_y της μεταφορικής κίνησης ορίζουν το διάνυσμα της παραμέτρου **b** του πίνακα 3.1.

Το επόμενο σε εμφάνιση και πολυπλοκότητα μοντέλο είναι αυτό *της εστίασης και μεταφοράς* που περιγράφει την αλλαγή της εστιακής απόστασης της κάμερας καθώς και τη μεταφορά της. Συγκεκριμένα περιγράφει ισοτροπική εστίαση της κάμερας δηλαδή ίση μεταβολή στο μήκος εστίασης και προς τις δύο κατευθύνσεις x, y και μεταφορά της κάμερας επίσης στις δύο κατευθύνσεις. Η ισοτροπική εστίαση περιγράφεται από την παράμετρο **q** ενώ η μεταφορά από το διάνυσμα της παραμέτρου **b**. Ο μετασχηματισμός αυτός χρησιμοποιήθηκε από τον H. Nicolas ([13]) για την κατασκευή του "μωσαϊκού". Προκύπτει από την εξίσωση 3.9 μηδενίζοντας τους συντελεστές του y στην κατεύθυνση του x και αντίστροφα καθώς και τους συντελεστές του x^2, y^2 και xy. Παρατηρείται ότι τα x και y στις δύο εξισώσεις πολλαπλασιάζονται με τον παράγοντα T_z που εκφράζει τη μεταφορική κίνηση της κάμερας ως προς τον άξονα Z και είναι αυτή ακριβώς η συνιστώσα της ταχύτητας που περιγράφει την αλλαγή στην απόσταση εστίασης.

Γενικότερη περίπτωση μετασχηματισμού από τους προηγούμενους εκφράζει το *αφινικό μοντέλο* που αποτελεί τον πιο ευρύ χρησιμοποιούμενο μετασχηματισμό του πίνακα 3.1. Ο αφινικός μετασχηματισμός περιέχει έξι παραμέτρους καθεμία από τις οποίες περιγράφει και διαφορετική κίνηση. Στο σύνολο των κινήσεων που περιγράφει περιλαμβάνεται μεταφορά της κάμερας προς τις δύο κατευθύνσεις, περιστροφή γύρω από τον οπτικό της άξονα, αλλαγή στο μήκος εστίασης της και μικρή κλίση προς

την οριζόντια ή την κατακόρυφη κατεύθυνση που εκφράζεται ως παράγοντας shear. Αποτέλεσμα της εφαρμογής του shear σε μία εικόνα είναι το τέντωμα της εικόνας ως προς τις διαγώνιες της. Το αφινικό μοντέλο αδυνατεί να περιγράψει πανοραμική λήψη της κάμερας και κλίση της κάμερας προς οποιαδήποτε κατεύθυνση. Το αφινικό μοντέλο προκύπτει από την εξίσωση 3.9 μηδενίζοντας τους συντελεστές των x^2, xy και y^2 και αντικαθιστώντας του υπόλοιπους συντελεστές από τις έξι παραμέτρους του. Η παράμετρος του x ως προς την κατεύθυνση του x και η παράμετρος του y ως προς την κατεύθυνση του y εκφράζουν την αλλαγή στην απόσταση εστίασης της κάμερας και είναι ανάλογες της μεταφορικής της κίνησης T_z. Ενώ η παράμετρος του x και y ως προς την αντίστροφη τους κατεύθυνση περιγράφουν τον παράγοντα shear που εκφράζει την περιστροφική κίνηση της κάμερας ως προς τον άξονα Z.

Το πιο κατάλληλο και αποδοτικότερο μοντέλο για την περιγραφή των κινήσεων της κάμερας είναι το *προβολικό (projective)* του πίνακα 3.1. Το μοντέλο αυτό διαθέτει τις επιθυμητές οκτώ παραμέτρους που είναι επαρκείς για την περιγραφή όλων των πιθανών κινήσεων της κάμερας. Αποτελεί επέκταση του αφινικού μοντέλου επομένως εκτός από το σύνολο των κινήσεων που περιγράφει το αφινικό μοντέλο, περιγράφει και περιστροφή της κάμερας ως προς τους άξονες X, Y. Ο μόνος περιορισμός που υφίσταται για το συγκεκριμένο μοντέλο θέλει όλα τα αντικείμενα της σκηνής να έχουν το ίδιο βάθος ως προς τη θέση της κάμερας. Εξαιτίας όμως της μορφής του μετασχηματισμού παρουσιάζεται δυσκολία στον υπολογισμό των παραμέτρων του. Αποτέλεσμα αυτού είναι πολλοί ερευνητές να αποφεύγουν τη χρήση του και να ανατρέχουν στη χρήση απλούστερων μοντέλων όπως το αφινικό ή σε άλλες προσεγγίσεις του προβολικού.

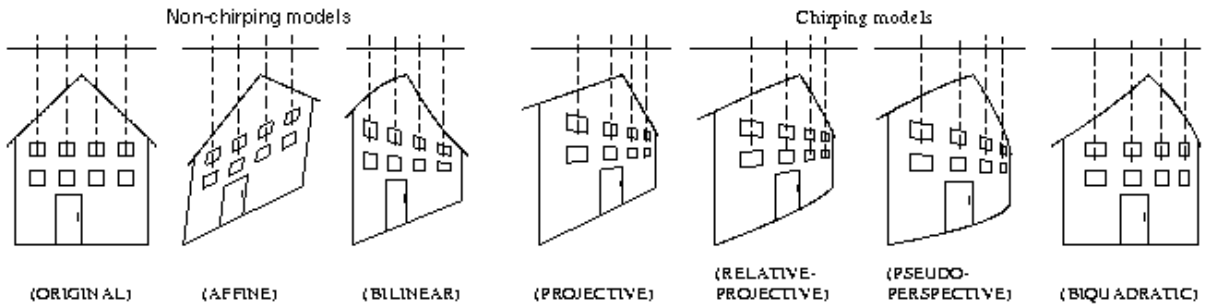
Το μεγαλύτερο και πολυπλοκότερο μοντέλο κίνησης του πίνακα 3.1 καλείται *biquadratic*. Η διαφορά του με τους προβολικούς μετασχηματισμούς έγκειται στο διαφορετικό αριθμό παραμέτρων που περιέχει. Το biquadratic αποτελεί το μεγαλύτερο μοντέλο κίνησης που χρησιμοποιείται από τους ερευνητές και είναι ικανό να περιγράψει το σύνολο των πιθανών κινήσεων της κάμερας. Λόγω όμως του μεγάλου πλήθους παραμέτρων του, παρουσιάζει μεγαλύτερη ευασθησία σε σφάλματα σε σύγκριση με τα άλλα μοντέλα κίνησης. Επομένως η χρήση του μοντέλου αυτού περιορίζεται μόνο στις ακολουθίες που εμφανίζεται σημαντική και πολύπλοκη η κίνηση της κάμερας.

Το *bilinear* μοντέλο αποτελεί επίσης μοντέλο οκτώ παραμέτρων και χρησιμοποιείται ευρύτατα στον τομέα της επεξεργασίας εικόνων. Το μοντέλο αυτό προκύπτει εύκολα από το biquadratic αφαιρώντας τους τέσσερις παράγοντες x^2 και y^2. Περιγράφει έναν αρκετά μεγάλο αριθμό πιθανών κινήσεων της κάμερας. Αποτυγχάνει όμως στην πλήρη

περιγραφή της περιστροφικής κίνησης της κάμερας.

Τέλος, το *pseudo-perspective* και το *relative-projective* μοντέλο ανήκουν και τα δύο στην κατηγορία των μοντέλων με τις οκτώ παραμέτρους. Το μοντέλο pseudo-perspective προκύπτει από το μοντέλο των δέκα παραμέτρων αφαιρώντας τους παράγοντες x^2 και y^2 και περιορίζοντας με δύο βαθμούς ελευθερίας τις υπόλοιπες τέσσερις τετραγωνικές παραμέτρους. Αποτελούν και τα δύο ικανοποιητικά μοντέλα κίνησης αφού περιγράφουν αρκετά καλά τις πιο συχνά εμφανιζόμενες κινήσεις της κάμερας στο σύνολο των ακολουθιών.

Για τη σύγκριση των μοντέλων του πίνακα 3.1 ο Steve Mann και Rosalind W. Picard [11] εισήγαγαν τους όρους "keystoning" και "chirping" που περιγράφουν φαινόμενα του πραγματικού κόσμου. Ο όρος "keystoning" εκφράζει ακριβώς την προοπτική των γραμμών, στην οποία η δήλωση του βάθους σε μία εικόνα παρουσιάζεται με τη σύγκλιση παράλληλων γραμμών. Με τον όρο "chirping" εννοούμε την επίδραση της αύξησης ή της μείωσης της χωρικής συχνότητας βάση της χωρικής θέσης (όπως φαίνεται σχήμα 3.4 και 3.5). Παρατηρώντας το σχήμα 3.4 ο Steve Mann και η Rosalind W. Picard συμπέραναν

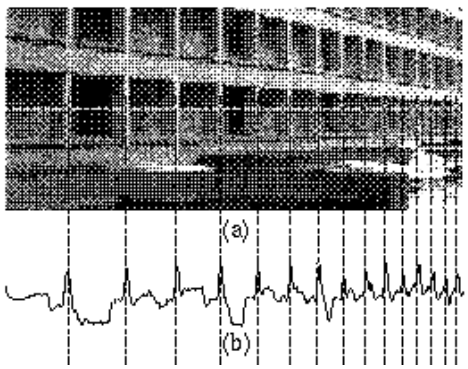


Σχήμα 3.4: Οι επιδράσεις των μετασχηματισμών σε μια εικόνα

Η μεταφορά δεν αλλάζει την αρχική (original) εικόνα απλά τη μετακινεί. Μόνο οι τέσσερις δεξιότεροι μετασχηματισμοί επιδρούν στην περιοδικότητα του μεγέθους των παραθύρων (επιφέροντας το επιθυμητό "chirping" που αντιστοιχεί ακριβώς σε αυτό που βλέπουμε στο πραγματικό κόσμο). Από αυτούς τους τέσσερις μετασχηματισμούς μόνο ο προβολικός (projective) διατηρεί τις ευθείες γραμμές.

ότι το ιδανικότερο μοντέλο κίνησης είναι το προβολικό μοντέλο. Το προβολικό από τη μια περιγράφει τα φαινόμενα "chirping" και "keystoning" και από την άλλη διατηρεί τις ευθείες γραμμές σε αντίθεση με τα υπόλοιπα μοντέλα που απλά περιγράφουν τα παραπάνω φαινόμενα χωρίς σίγουρη επιτυχία κάθε φορά.

Παρόλη την επιτυχή περιγραφή των φαινομένων "chirping" και "keystoning", το προβολικό μοντέλο δε χρησιμοποιείται ευρύτατα από τους ερευνητές. Ανασταλτικό



Σχήμα 3.5: Το φαινόμενο ”chirping”

(a) αντικείμενο του πραγματικού κόσμου που εμφανίζει περιοδικότητα. Το φαινόμενο ”chirping” εκφράζει την περιοδικότητα στην προοπτική. Εμφανίζεται με την αύξηση ή μείωση της χωρικής συχνότητας βάση της τοπικής θέσης. (b) περιοδικότητα στην εικόνα

παράγοντα της χρήσης του αποτελεί η μορφή του που από τη μια δυσκολεύει τον υπολογισμό των παραμέτρων του και από την άλλη η ύπαρξη παραμέτρων στον παρανομαστή κάνει το μοντέλο ευαίσθητο σε μικρά λάθη εκτίμησης της κίνησης, ιδιαίτερα αν η σκηνή χαρακτηρίζεται ως περίπλοκη. Στην παρούσα εργασία χρησιμοποιήθηκαν τα μοντέλα εστίασης και μεταφοράς, το αφινικό και το προβολικό. Το συμπέρασμα στο οποίο καταλήξαμε είναι ότι ένα πολύπλοκο μοντέλο κίνησης, με μεγάλο δηλαδή αριθμό παραμέτρων, από τη μια είναι ικανό να περιγράψει πολύπλοκες κινήσεις της κάμερας, από τη άλλη όμως σε απλές κινήσεις είναι δυνατό να εμφανίζει μικρά σφάλματα στις παραμέτρους του, περιγράφοντας μορφές κινήσεων που δεν είναι ορατές. Επομένως αναγκαία είναι η ιεραρχία των μοντέλων και η επιλογή σε κάθε περίπτωση του μικρότερου μοντέλου που θεωρητικά είναι κατάλληλο για την περιγραφή της κίνησης της κάμερας.

3.4.2 Εκτίμηση των παραμετρικών μοντέλων κίνησης

Η εκτίμηση ενός παραμετρικού μοντέλου κίνησης ανάγεται στον υπολογισμό των παραμέτρων που περιέχει. Ο αριθμός των παραμέτρων αυτών ποικίλει σε κάθε μοντέλο, ξεκινώντας από δύο παραμέτρους, που περιλαμβάνει το μοντέλο της μεταφοράς, και καλήγοντας σε δέκα που περιέχει το biquadratic μοντέλο. Για την εκτίμηση των παραμέτρων του μοντέλου της μεταφοράς, αρκεί ο υπολογισμός της οπτικής ροής για ένα σημείο. Για το αφινικό μοντέλο, που αποτελεί γενικότερο μοντέλο κίνησης από τα προηγούμενα, αρκεί τουλάχιστον οπτική ροή για τρία σημεία, ενώ για το προβολικό ή οποιοδήποτε άλλο μοντέλο με οκτώ παραμέτρους αρκεί οπτική ροή για τέσσερα σημεία. Το biquadratic μοντέλο, που περιέχει το μεγαλύτερο αριθμό παραμέτρων, χρειάζεται

τουλάχιστον οπτική ροή για πέντε σημεία. Η εκτίμηση τότε των παραμέτρων του κάθε μοντέλου ανάγεται στη λύση ενός συστήματος. Για το προβολικό και το "relative projective" μοντέλο το σύστημα αποτελείται από 8 μη-γραμμικές εξισώσεις όσες δηλαδή και οι παράμετροί τους. Αντίθετα για τα υπόλοιπα μοντέλα κίνησης το σύστημα που προκύπτει αποτελείται από τόσες γραμμικές εξισώσεις όσες και οι παράμετροι του εκάστοτε μοντέλου.

Στην πραγματικότητα όμως για την ορθή εκτίμηση των παραμέτρων δεν αρκεί ούτε ο μεγαλύτερος αριθμός σημείων που αναφέρθηκε, απαιτούνται πολύ περισσότερα. Το πρόβλημα οφείλεται στο γεγονός ότι το υπολογισμένο πεδίο της οπτικής ροής συχνά παρουσιάζει ένα ποσοστό εσφαλμένων δεδομένων. Τα εσφαλμένα δεδομένα της οπτικής ροής είναι γνωστά σαν **outliers**, είναι δηλαδή σημεία που δε συμφωνούν με την κυρίαρχη κίνηση της κάμερας. Τα σημεία αυτά οφείλονται είτε στην ύπαρξη θορύβου στην εικόνα είτε πιο συχνά στην ύπαρξη ανεξάρτητων κινούμενων αντικειμένων στη σκηνή. Η χρήση των σημείων αυτών στην εκτίμηση των παραμέτρων οδηγεί επίσης σε λανθασμένο υπολογισμό των παραμέτρων του εκάστοτε μοντέλου και κατ' επέκταση σε λανθασμένο υπολογισμό της κίνησης της κάμερας. Απαραίτητος είναι επομένως ο διαχωρισμός τους από το σύνολο των σωστών εκτιμήσεων της οπτικής ροής.

Ο μεγάλος αριθμός εκτιμήσεων της οπτικής ροής παρέχει εμπιστοσύνη στην τελική εκτίμηση των παραμέτρων, αφού ένα μεγάλο ποσοστό της σκηνής λαμβάνει μέρος στον υπολογισμό τους. Οι παράμετροι τότε του μοντέλου περιγράφουν συνολικά την κίνηση της κάμερας. Αυξάνεται επίσης η αναγκαιότητα του μεγάλου αριθμού εκτιμήσεων της οπτικής ροής στις περιπτώσεις που η σκηνή αποτελείται από διαφορετικά επίπεδα κίνησης. Απαιτείται τότε ένα σημαντικό ποσοστό εκτιμήσεων για κάθε επίπεδο της σκηνής. Για την επίτευξη μεγαλύτερης εμπιστοσύνης στις τελικές εκτιμημένες παραμέτρους, τα σημεία εκτίμησης της οπτικής ροής οφείλουν να είναι διασκορπισμένα πάνω στη σκηνή καλύπτοντας ένα μεγάλο ποσοστό της. Η εκτίμηση τότε των παραμέτρων του κάθε μοντέλου θα τείνει στην πραγματική.

Για την αντιμετώπιση του προβλήματος της εκτίμησης της οπτικής ροής στο σύνολο των σημείων της εικόνας υλοποιήθηκαν δύο μέθοδοι με ταυτόχρονη απόρριψη των "κακών δεδομένων" (outliers). Η εκτίμηση των παραμέτρων του εκάστοτε μοντέλου επιτεύχθηκε, από τη μια με τη χρήση ενός αλγορίθμου σθεναρότητας και από την άλλη με τη χρήση αναλλοίωτων χαρακτηριστικών όπως οι σχέσεις ροπών μιας επιφάνειας. Τέλος αναπτύχθηκαν μέθοδοι για τη διόρθωση των εκτιμημένων παραμέτρων κίνησης.

3.5 Εκτίμηση της οπτικής ροής

Στην ενότητα αυτή παρουσιάζονται δύο μέθοδοι για τον υπολογισμό της οπτικής ροής που βασίζονται στις χωρικές μεταβολές. Δεν είναι απαραίτητη η εκτίμηση της οπτικής ροής πάνω σε όλα τα σημεία της εικόνας. Αναγκαία όμως είναι η ύπαρξη ενός αρκετά ικανοποιητικού αριθμού σημείων, στα οποία η εκτίμηση της οπτικής ροής συμφωνεί με την κίνηση της κάμερας. Συνεπώς απαραίτητος είναι ο διαχωρισμός των "κακών δεδομένων" (outliers) από το σύνολο των εκτιμήσεων της οπτικής ροής. Το σύνολο των τελικών σημείων που θα χρησιμοποιηθεί για την εκτίμηση των παραμέτρων του μοντέλου κίνησης καλείται σύνολο "καλών δεδομένων" ή **inliers**. Παρακάτω παρουσιάζονται οι μέθοδοι υπολογισμού της οπτικής ροής και το σύνολο των κριτηρίων που εφαρμόζονται σε κάθε περίπτωση για την αφαίρεση των "κακών δεδομένων" (outliers).

3.5.1 Ταίριασμα περιοχών

Μία από τις μεθόδους που χρησιμοποιούνται για τον υπολογισμό της οπτικής ροής μεταξύ δύο εικόνων μιας ακολουθίας είναι η μέθοδος του ταιριάσματος των περιοχών ή μπλοκς (block matching). Τα μπλόκς που χρησιμοποιούνται στη μέθοδο αυτή είναι γενικά ορθογώνιες περιοχές της εικόνας με μέγεθος εκ των προτέρων καθορισμένο και σταθερό. Υποθέτουμε ότι κατά την κίνηση της κάμερας κάθε τέτοιο μπλοκ μετακινείται εκτελώντας απλά μεταφορική κίνηση προς τις δύο κατευθύνσεις. Σκοπός της μεθόδου είναι η εύρεση του διανύσματος της ταχύτητας (u, v) του μπλοκ. Το διάνυσμα αυτό ουσιαστικά περιγράφει το αντίστοιχο μπλοκ στη δεύτερη εικόνα.

Γενικά η μέθοδος αναζητά το μπλοκ της πρώτης εικόνας σε μια περιοχή στη δεύτερη. Το εύρος της περιοχής αναζήτησης ορίζεται βάση του διανύσματος της ταχύτητας και είναι:

$$| u | \leq \Delta_h \qquad \text{και} \qquad | v | \leq \Delta_v \tag{3.11}$$

Οι αποδεκτές συνιστώσες του διανύσματος της ταχύτητας είναι πολλαπλάσιοι ακέραιοι της χωρικής περιόδου δειγματοληψίας ή της μισής αυτής περιόδου. Αν το μέγεθος του μπλοκ είναι K γραμμές και L στήλες, το μέγεθος της περιοχής αναζήτησης, στην περίπτωση της ακέραιας ταχύτητας, είναι $(K + 2\Delta_v)$ $(L + 2\Delta_h)$. Ο αριθμός των πιθανών τιμών της ταχύτητας και επομένως ο αριθμός των δυνατών θέσεων του μπλοκ πάνω στη δεύτερη εικόνα είναι $(2\Delta_h + 1)(2\Delta_v + 1)$. Για παράδειγμα, για οριζόντια και κάθετη ταχύτητα των 7 pixel, ο αριθμός των δυνατών θέσεων στη δεύτερη εικόνα είναι 225. Αφού

ορίστηκε η διάσταση της περιοχής αναζήτησης στη δεύτερη εικόνα, το κριτήριο που πρέπει να ελαχιστοποιηθεί για την εκτίμηση της ταχύτητας (u, v) βασίζεται στην αρχή διατήρησης της φωτεινότητας στο χρόνο και είναι:

$$(\vec{u}, \vec{v}) = \min_{(u,v)\epsilon\mathcal{Z}} \sum_{i=0}^{K-1} \sum_{j=0}^{L-1} \varrho(I_{DFD}(i,j;u,v)) \tag{3.12}$$

Οι περιορισμοί που ισχύουν για τις συνιστώσες της ταχύτητας είναι: $|u| \leq \Delta_h,\ |v| \leq \Delta_v$, ενώ η συνάρτηση ϱ συνήθως παρουσιάζεται με μία από τις ακόλουθες μορφές:

$$\varrho(x) = \mid x \mid \qquad \text{ή} \qquad \varrho(x) = (x)^2 \tag{3.13}$$

Η μεταβλητή I_{DFD} εκφράζει τη διαφορά των εντάσεων του μπλοκ της πρώτης εικόνας από το μπλοκ της δεύτερης μετατοπισμένη κατά παράγοντα (u, v) στις δύο κατευθύνσεις.

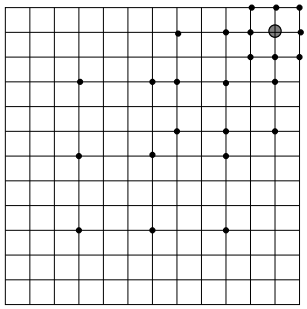
$$I_{DFD} = I(i,j;t) - I(i-u, j-v; t+1) \tag{3.14}$$

Αν και οι λειτουργίες που είναι αναγκαίες για την εκτίμηση του διανύσματος της ταχύτητας είναι απλές, η πρακτική εφαρμογή τους απαιτεί μεγάλο υπολογιστικό κόστος. Υποθέτοντας ότι η μέγιστη ταχύτητα είναι 7, οριζόντια και κάθετα, ο συνολικός αριθμός λειτουργιών ανά δευτερόλεπτο είναι 684.288.000. Οι λειτουργίες αυτές περιλαμβάνουν αφαίρεση, εκτίμηση της απόλυτης τιμής (αν $\varrho(x) = \mid x \mid$) και άθροιση των αποκλίσεων που εκφράζονται από τη μεταβλητή I_{DFD}.

Αρκετοί αλγόριθμοι σήμερα κατορθώνουν γρήγορη αναζήτηση και ανεύρεση της περιοχής στη δεύτερη εικόνα με την ελάχιστη τιμή του κριτηρίου. Οι καλύτεροι από αυτούς είναι η *λογαριθμική αναζήτηση δύο διαστάσεων (2-Dlogarithmic search)*[33], η *αναζήτηση με αυξανόμενη ακρίβεια (increasing accuracy search)*[34] και η *αναζήτηση προς την κατεύθυνση της κλίσης (conjugate direction search)*[35]. Συγκρίνοντας τους αλγορίθμους αυτούς ο Santillana Rivero [36] κατέληξε στο συμπέρασμα ότι ο αλγόριθμος που υπερέχει, έχει δηλαδή το μικρότερο υπολογιστικό κόστος, είναι αυτός της αναζήτησης με αυξανόμενη ακρίβεια. Βρίσκει τη θέση με την ελάχιστη τιμή του κριτηρίου σε πολύ μικρό αριθμό βημάτων.

Στην παρούσα εργασία χρησιμοποιήθηκε ο αλγόριθμος της αναζήτησης με αυξανόμενη ακρίβεια. Σύμφωνα με τον αλγόριθμο αυτό, στο πρώτο βήμα ελέγχεται το διάνυσμα της ταχύτητας πάνω σε περιοχή αναζήτησης μικρής ακρίβειας. Για τον καθορισμό του διανύσματος της ταχύτητας αρχικά εκτιμάται η τιμή του κριτηρίου σε εννέα θέσεις και τέλος επιλέγεται η θέση εκείνη με την ελάχιστη τιμή του κριτηρίου. Κατόπιν σε περιοχή αναζήτησης που ορίζεται γύρω από την περιοχή που ανέδειξε το ελάχιστο στο

προηγούμενο βήμα και με μέγεθος μικρότερο της προηγούμενης περιοχής αναζητάται το νέο ελάχιστο από τις νέες εννέα θέσεις. Η διαδικασία συνεχίζεται έτσι αυξάνοντας την ακρίβεια αναζήτησης μέχρι τη λήψη ακρίβειας ενός pixel. Παράδειγμα εφαρμογής του αλγορίθμου φαίνεται στο σχήμα 3.6. Στο παράδειγμα αυτό το μέγιστο διάνυσμα της ταχύτητας είναι (6, 6) και το διάνυσμα της ταχύτητας που τελικά προκύπτει (5, 5). Στην παρούσα εργασία τελική ακρίβεια για τη σύγκλιση του αλγορίθμου θεωρήθηκε η



Σχήμα 3.6: Αναζήτηση με Αυξανόμενη Ακρίβεια

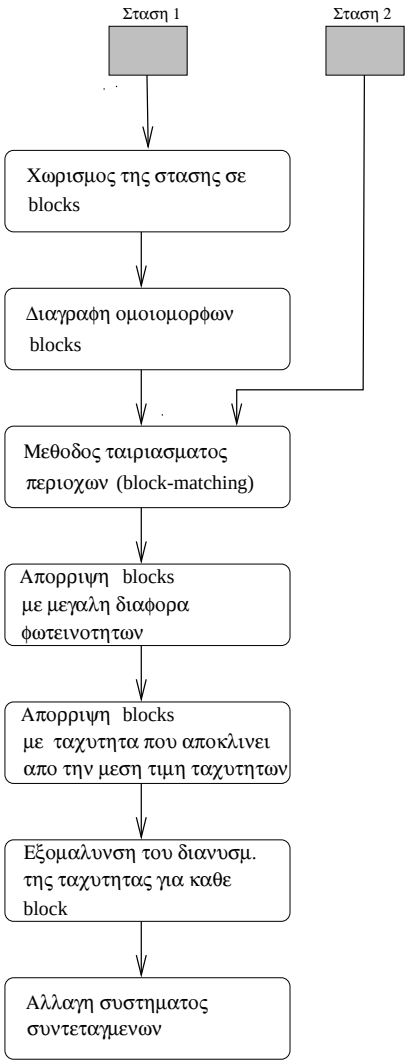
ακρίβεια μισού pixel. Η παρέμβαση αυτή έγινε για την εύρεση της ακριβής θέσης του αντίστοιχου μπλοκ στη δεύτερη εικόνα. Η ακρίβεια όμως αυτή έχει σαν αποτέλεσμα την πρόσβαση σε δεκαδικές συντεταγμένες της εικόνας. Στην περίπτωση αυτή ο υπολογισμός της εκάστοτε έντασης της εικόνας προκύπτει από την παρεμβολή στις γειτονικές ακέραιες συντεταγμένες. Για το σκοπό αυτό χρησιμοποιείται η μέθοδος της *δι-γραμμικής παρεμβολής*. Στην ακόλουθη εξίσωση δίνεται η έκφραση της έντασης της εικόνας σε δεκαδική θέση, όπως προκύπτει με την παρεμβολή στα τέσσερα γειτονικά σημεία της.

$$\begin{aligned} I(i-u, j-v) \;=\; & (1-d_v)\,(1-d_u)\;\, I(i-\lfloor u \rfloor, j-\lfloor v \rfloor) \\ & +\; (1-d_v)\, d_u\, I(i-\lfloor u \rfloor - 1, j-\lfloor v \rfloor) \\ & +\; d_v (1-d_u)\;\, I(i-\lfloor u \rfloor, j-\lfloor v \rfloor - 1) \\ & +\; d_u d_v\, I(i-\lfloor u \rfloor - 1, j-\lfloor v \rfloor - 1) \end{aligned} \tag{3.15}$$

όπου u = $\lfloor u \rfloor$ + d_u, v = $\lfloor v \rfloor$ + d_v, και $0 \geq d_u \geq 1$, $0 \geq d_v \geq 1$.

3.5.1.1 Κριτήρια εκτίμησης της οπτική ροής

Για την εκτίμηση της οπτικής ροής μεταξύ δύο στάσεων μιας video ακολουθίας χρησιμοποιήθηκε ο αλγόριθμος που φαίνεται στο σχήμα 3.7. Παρακάτω θα αναλύσουμε τα βήματα του αλγορίθμου και το σύνολο των κριτηρίων που χρησιμοποιήθηκαν.



Σχήμα 3.7: Αλγόριθμος ταιριάσματος Περιοχών (block-matching)

1. *Χωρισμός της εικόνας σε μπλοκς σταθερού μεγέθους.* Υφίσταται χωρισμός της μιας από τις δύο στάσεις σε μπλοκς μεγέθους 16x16. Ο λόγος που γίνεται μια τέτοια διαμέριση είναι για την καλύτερη (μεγαλύτερη ακρίβεια) και ευκολότερη ανίχνευση της κίνησης. Το μικρό μέγεθος μπλοκ διευκολύνει τη διαδικασία της απόρριψης μπλοκς που δεν αντιστοιχούν στη συνολική κίνηση της κάμερας. Τέτοια μπλοκς, όπως έχει ήδη αναφερθεί, αντιστοιχούν σε θόρυβο ή πιο συχνά σε κινούμενα αντικείμενα. Από την άλλη, το μέγεθος του μπλοκ δεν πρέπει να είναι ιδιαίτερα μικρό, γιατί έτσι η πληροφορία που διαθέτει δεν είναι σημαντική και χάνεται η δυνατότητα σωστής ανίχνευσης της κίνησης. Τέλος θα πρέπει να αναφερθεί ότι ο χωρισμός της στάσης σε μπλοκς δεν αρχίζει από το περίγραμμα της.

Μια μικρή περιοχή γύρω από τη στάση δεν υφίσταται διαμέριση. Αυτό οφείλεται στην απαίτηση, που θέλει η πληροφορία που αναπαριστά το κάθε μπλοκ να υπάρχει και στη δεύτερη στάση. Το πλάτος της περιοχής που δεν υφίσταται διαμέριση είναι ανάλογο της μέγιστης μετατόπισης της κάμερας μεταξύ των στάσεων και προς τις δύο κατευθύνσεις (οριζόντια και κάθετα στο επίπεδο του σκηνικού).

2. *Διαγραφή ομοιόμορφων μπλοκς.* Απορρίπτονται τα μπλοκς που εμφανίζουν ομοιομορφία στη φωτεινότητα. Η εκτίμηση της κίνησης στα μπλοκς αυτά δεν είναι ορθή, αφού λόγω της ομοιομορφίας δεν εμφανίζεται διαφορά στη φωτεινότητα μεταξύ των μπλοκς που συγκρίνονται για την ανίχνευση της κίνησης. Η συμμετοχή τέτοιων μπλοκς στην τελική εκτίμηση της κυρίαρχης κίνησης οδηγεί σε λάθος αποτελέσματα.

3. *Εκτίμηση της κίνησης σε κάθε μπλοκ.* Ο υπολογισμός της κίνησης σε κάθε περιοχή επιτυγχάνεται με χρήση της μεθόδου ταιριάσματος περιοχών (block-matching), που ήδη αναφέραμε. Για γρήγορη και αποτελεσματικότερη αναζήτηση χρησιμοποιείται ο αλγόριθμος της αναζήτησης με αυξανόμενη ακρίβεια (Increasing accuracy search) που φτάνει μέχρι την ακρίβεια μισού pixel. Στην περίπτωση που έχει ήδη εκτιμηθεί το διάνυσμα της ταχύτητας μιας γειτονικής περιοχής, χρησιμοποιείται αυτό σαν αρχική τιμή στον υπολογισμό της ταχύτητας του εκάστοτε μπλοκ. Η τιμή αυτή υφίσταται διόρθωση που επιτυγχάνεται με εφαρμογή της μεθόδου ταιριάσματος περιοχών σε μικρότερο εύρος περιοχής αναζήτησης της τάξης του ενός pixel διαφορά ($\pm$ 1).

4. *Απόρριψη μπλοκς με μεγάλη διαφορά φωτεινότητας.* Μετά την εκτίμηση του διανύσματος της ταχύτητας για κάθε μπλοκ, απαιτείται η εφαρμογή κάποιου κριτηρίου που θα ανιχνεύσει και θα απορρίψει περιοχές με λανθασμένη εκτίμηση της κίνησης. Λανθασμένο θεωρείται το διάνυσμα της ταχύτητας που δε συμφωνεί με τη συνολική κίνηση της κάμερας. Το κριτήριο αυτό απορρίπτει τα μπλοκς που εμφανίζουν συνολικά μεγάλη διαφορά ως προς τις φωτεινότητες με τα αντίστοιχα μπλοκς που προκύπτουν βάση του υπολογιζόμενου διανύσματος της ταχύτητας στην άλλη εικόνα. Το όριο της μέγιστης διαφοράς καθορίζεται από το χρήστη. Η διαφορά των φωτεινοτήτων των αντίστοιχων μπλοκς εκφράζεται από την σχέση:

$$D = \sum \sum I_{DFD}^2 = \sum_{x \,\epsilon\, R} \sum_{y \,\epsilon\, R} (I(x, y, t-1) \; - I(x-u, y-v, t)\;)^2 \qquad (3.16)$$

όπου (u, v) είναι το διάνυσμα της ταχύτητας του εκάστοτε μπλοκ στην πρώτη στάση που εκτιμήθηκε βάση της μεθόδου ταιριάσματος περιοχών.

5. *Απόρριψη μπλοκς με διάνυσμα ταχύτητας που αποκλίνει από τη μέση τιμή των ταχυτήτων.* Το δεύτερο κριτήριο που εφαρμόζεται απορρίπτει τα μπλοκς με διάνυσμα ταχύτητας που αποκλίνει σημαντικά από τη μέση τιμή των ταχυτήτων όλων των μπλοκς της στάσης που έχουν απομείνει. Συνήθως αυτά τα διανύσματα ταχύτητας αντιστοιχούν σε "κακά δεδομένα" (outliers), επομένως απορρίπτονται ορθά.

6. *Εξομάλυνση του διανύσματος της ταχύτητας κάθε μπλοκ.* Για τη διόρθωση της ήδη εκτιμημένης ταχύτητας του κάθε μπλοκ εξομαλύνονται οι διαφορές που παρουσιάζει το διάνυσμα της ταχύτητας αυτής με τα διανύσματα των ταχυτήτων των γειτονικών σε αυτό μπλοκς. Η εξομάλυνση αυτή επιτυγχάνεται με εφαρμογή ενός σταθμισμένου φίλτρου πάνω στο διάνυσμα της ταχύτητας κάθε μπλοκ που έχει απομείνει. Το φίλτρο αυτό εφαρμόζεται σε γειτονιά των οκτώ μπλοκς. Η απόκριση του φίλτρου είναι:

$$y(m,n) = \frac{\sum_{k=-x}^{x}\sum_{l=-y}^{y} \Phi(I(m-k,n-l) - I(m,n))\, g(|k| + |l|)\, I(m-k,n-l)}{\sum_{k=-x}^{x}\sum_{l=-y}^{y} \Phi(I(m-k,n-l) - I(m,n))\, g(|k| + |l|)} \quad (3.17)$$

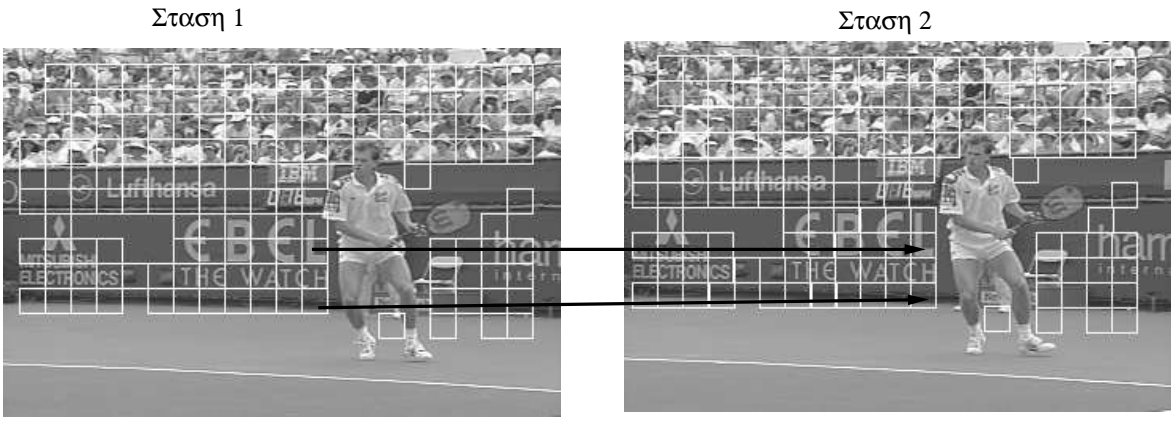
όπου x και y είναι οι διαστάσεις του μπλοκ, ενώ Φ η ακόλουθη συνάρτηση:

$$\Phi(\psi) = \begin{cases} 0 & \text{αν δεν υπάρχει μπλοκ με μεγαλύτερη} \\ & \text{από } (\pm 1)\ pixel \text{ διαφορά ταχύτητας} \\ 1 & \text{αντίθετα} \end{cases} \quad (3.18)$$

Τέλος για τη μεταβλητή g ισχύουν τα εξής: g(0) = 4, g(1) = 2 και g(2) = 1.

7. *Κανονικοποίηση των συντεταγμένων των μπλοκς.* Πριν την εκτίμηση του εκάστοτε παραμετρικού μοντέλου κίνησης εφαρμόζεται αλλαγή στο σύστημα συντεταγμένων της κάθε περιοχής της στάσης. Νέο τώρα σύστημα συντεταγμένων θεωρείται το κέντρο της στάσης. Η αλλαγή αυτή εφαρμόζεται για κανονικοποίηση των συντεταγμένων. Έτσι προκύπτουν παράμετροι κίνησης με μικρότερες τιμές, πράγμα που κάνει ευκολότερους τους υπολογισμούς και χειρισμούς τους.

Ένα παράδειγμα εφαρμογής του αλγορίθμου εκτίμησης της οπτικής ροής φαίνεται στο σχήμα 3.8. Ομοιόμορφες περιοχές όπως αυτή του γηπέδου καθώς και περιοχές που ανήκουν σε κινούμενα αντικείμενα όπως η περιοχή που καταλαμβάνει ο τενίστας του σχήματος, έχουν απορριφθεί.



Σχήμα 3.8: Παράδειγμα αντιστοίχισης περιοχών (μπλοκς)

3.5.2 Αντιστοίχιση Χαρακτηριστικών της Εικόνας

Μια από τις μεθόδους που συχνά χρησιμοποιούνται για τον υπολογισμό της οπτικής ροής σε ορισμένα σημεία της εικόνας είναι η μέθοδος της αντιστοίχισης χαρακτηριστικών. Τα βήματα που ακολουθούνται στη μέθοδο αυτή είναι αρχικά η εξαγωγή των χαρακτηριστικών και τέλος η αντιστοίχιση τους. Χαρακτηριστικά εξάγονται και από τις δύο στάσεις των οποίων επιθυμούμε να εκτιμήσουμε την κίνηση. Η διαδικασία της αντιστοίχισης θεωρεί δεδομένα τα χαρακτηριστικά της μιας στάσης και σκοπεύει στην εύρεση των αντίστοιχών τους στην άλλη στάση.

Τί ακριβώς όμως ονομάζουμε χαρακτηριστικό μιας εικόνας στον τομέα της επεξεργασίας εικόνων[37],[38]. Η έννοια χαρακτηριστικό αναφέρεται σε δύο οντότητες:

1. σε μια γενική ιδιότητα της εικόνας ή κάποιου μέρους αυτής, για παράδειγμα η μέση τιμή των εντάσεων των pixel της (**ολικό χαρακτηριστικό**).

2. ή σε ένα μέρος της εικόνας με ιδιαίτερες ιδιότητες, για παράδειγμα σε ένα κύκλο, μια γραμμή ή μια γωνία (**τοπικό χαρακτηριστικό**).

Τα ολικά χαρακτηριστικά σε αντίθεση με τα τοπικά δε χρησιμοποιούνται για τη λύση προβλημάτων όπως ο υπολογισμός της οπτικής ροής. Παρακάτω η έννοια χαρακτηριστικό θα αντιστοιχεί στην έννοια των τοπικών χαρακτηριστικών μιας εικόνας. Επομένως χαρακτηριστικά μιας εικόνας είναι αυτά που παρουσιάζουν σημαντική ιδιαιτερότητα, είναι τοπικά και ανιχνεύσιμα μέρη της. Το ότι παρουσιάζουν σημαντική ιδιαιτερότητα σημαίνει ότι συνδέονται με σημαντικές ιδιότητες των επιμέρους κομματιών της εικόνας. Για παράδειγμα, το περίγραμμα των αντικειμένων της σκηνής,

που είναι τοπικό χαρακτηριστικό της εικόνας συνδέεται με τις έντονες αλλαγές στη φωτεινότητα στα σημεία του περιγράμματος. Το ότι χαρακτηρίζονται ως ανιχνεύσιμα σημαίνει ότι υπάρχει αλγόριθμος εύρεσης τους, αλλιώς δεν είναι άξια ύπαρξης.

Το σύνολο των χαρακτηριστικών που υπάρχουν χωρίζονται σε ορισμένες κατηγορίες όπως:

- *χωρικά χαρακτηριστικά.* Στην κατηγορία αυτή ανήκουν τα χαρακτηριστικά του ιστογράμματος, όπως η διασπορά, η μέση τιμή κ.α.

- *χαρακτηριστικά σημείων.* Στην κατηγορία αυτή ανήκουν τα χαρακτηριστικά που εξάγονται από ένα μόνο σημείο. Τέτοια χαρακτηριστικά είναι οι γωνίες για τις οποίες θα αναφέρουμε περισσότερα παρακάτω.

- *ακμές,* όπως είναι φανερό στην κατηγορία αυτή ανήκουν οι ακμές, που λόγω της σημασίας τους στον τομέα της επεξεργασίας εικόνων ορίζουν μόνες τους κατηγορία.

- *χαρακτηριστικά σχήματος.* Τέτοια χαρακτηριστικά είναι το περίγραμμα, οι ροπές, οι γραμμές και οι καμπύλες. Όλα αυτά τα χαρακτηριστικά είναι ικανά να περιγράψουν το σχήμα ενός αντικειμένου.

- *χαρακτηριστικά επιφανειών.* Στην κατηγορία αυτή ανήκουν και τα χαρακτηριστικά του σχήματος αφού μια επιφάνεια περιγράφεται και μόνο από το σχήμα της. Στην κατηγορία αυτή ανήκουν επιπλέον η υφή και το χρώμα της επιφάνειας.

Για τον υπολογισμό της οπτικής ροής σε σημεία της εικόνας απαιτείται η εξαγωγή των χαρακτηριστικών που σημασιοδοτούν τα σημεία αυτά. Επομένως η κατηγορία που εμφανίζει ενδιαφέρον στην περίπτωση της εκτίμησης της οπτικής ροής είναι τα χαρακτηριστικά σημείων. Στην κατηγορία αυτή ανήκουν οι γωνίες που εκτός του ότι είναι εύκολα ανιχνεύσιμες, χαρακτηρίζουν πλήρως ένα σημείο. Παρακάτω θα αναφέρουμε τον αλγόριθμο εξαγωγής γωνιών καθώς και τα κριτήρια που χρησιμοποιήθηκαν για την αντιστοίχιση των γωνιών.

3.5.2.1 Εξαγωγή γωνιών

Οι γωνίες αποτελούν χαρακτηριστικό σημείων και δεν αντιστοιχούν οπωσδήποτε σε κάποια γεωμετρική οντότητα της παρατηρούμενης σκηνής. Μεταφράζονται σαν γωνίες όχι όμως μόνο με την έννοια της τομής των γραμμών της εικόνας αλλά και με βάση

των συνδυασμών ή της δομής των εντάσεων στο χώρο. Αλγόριθμοι για την ανίχνευση γωνιών προτάθηκαν στις εργασίες [27], [39]. Στην παρούσα εργασία χρησιμοποιήθηκε ο αλγόριθμος που παρουσιάζεται στο [37].

Για την ανίχνευση των γωνιών θεωρούμε το διάνυσμα των χωρικών παραγώγων, $[I_x, I_y]^T$, ένα τυχαίο σημείο p της εικόνας, τη γειτονιά Q του σημείου p, και τον πίνακα C, που ορίζεται ως:

$$C = \begin{bmatrix} \sum {I_x}^2 & \sum I_x I_y \\ \sum I_x I_y & \sum {I_y}^2 \end{bmatrix} \tag{3.19}$$

τα αθροίσματα στον πίνακα C εφαρμόζονται στην γειτονιά Q. Ο πίνακας αυτός χαρακτηρίζει τη δομή των εντάσεων στη γειτονιά.

Δεδομένου ότι ο πίνακας C είναι συμμετρικός, μπορεί να διαγωνιοποιηθεί. Χωρίς περιορισμό της γενικότητας θεωρούμε το διαγώνιο πίνακα C της ακόλουθης εξίσωσης:

$$C = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{3.20}$$

Οι δύο ιδιοτιμές, λ_1 και λ_2, που εμφανίζονται στον πίνακα C, είναι θετικές. Ας υποθέσουμε ότι $\lambda_1 \geq \lambda_2$. Η γεωμετρική σημασία των λ_1 και λ_2 μπορεί να γίνει κατανοητή μέσω μερικών ειδικών περιπτώσεων. Στην πρώτη περίπτωση, θεωρούμε μια ομοιόμορφη γειτονιά σημείων, Q. Η χωρική παράγωγος μηδενίζεται παντού στη γειτονιά, ο πίνακας C γίνεται μηδενικός και τότε οι ιδιοτιμές: λ_1=λ_2=0. Στη δεύτερη περίπτωση υποθέτουμε ότι η γειτονιά Q περιέχει ένα ιδανικό μαύρο και άσπρο βήμα ακμής, τότε οι ιδιοτιμές: λ_2 = 0 και $\lambda_1 > 0$. Το ιδιοδιάνυσμα που συνδέεται με την ιδιοτιμή λ_1 είναι παράλληλο στο διάνυσμα των χωρικών παραγώγων. Στην τρίτη περίπτωση, υποθέτουμε ότι η γειτονιά Q περιέχει τη γωνία ενός μαύρου τετραγώνου πάνω σε άσπρο φόντο. Λόγω της ύπαρξης των δύο κύριων κατευθύνσεων στη γειτονιά Q, αναμένουμε $\lambda_1 \geq \lambda_2 > 0$. Όσο πιο μεγάλες είναι οι ιδιοτιμές λ_1 και λ_2 τόσο πιο έντονες εμφανίζονται οι αντίστοιχες γραμμές που αυτές περιγράφουν στην εικόνα. Τα ιδιοδιανύσματα γενικά κωδικοποιούν την κατεύθυνση της ακμής, ενώ οι ιδιοτιμές την ισχύ των ακμών. Μια γωνία κατασκευάζεται από δύο έντονες ακμές. Επομένως αφού $\lambda_1 \geq \lambda_2$, *γωνία* παρουσιάζεται στη θέση όπου η μικρότερη ιδιοτιμή λ_2 είναι αρκετά μεγάλη. Γενικότερα στα σημεία γωνιών η επιφάνεια της έντασης έχει δύο χαρακτηριστικές κατευθύνσεις που συνδέονται με τις ιδιοτιμές του C. Και οι δύο αυτές ιδιοτιμές είναι σημαντικά μεγαλύτερες του μηδενός.

Ας θεωρήσουμε I την εικόνα στην οποία επιθυμούμε να ανιχνεύσουμε τις γωνίες. τ είναι το κατώτερο όριο για την ιδιοτιμή λ_2. Πάνω από το όριο τ θεωρούμε ότι ανιχεύθηκε

γωνία. (2N+1)x(2N+1) pixels είναι το μέγεθος του τετραγωνικού παραθύρου που εκφράζει τη γειτονιά Q. Τα βήματα τότε του αλγορίθμου ανίχνευσης γωνιών είναι:

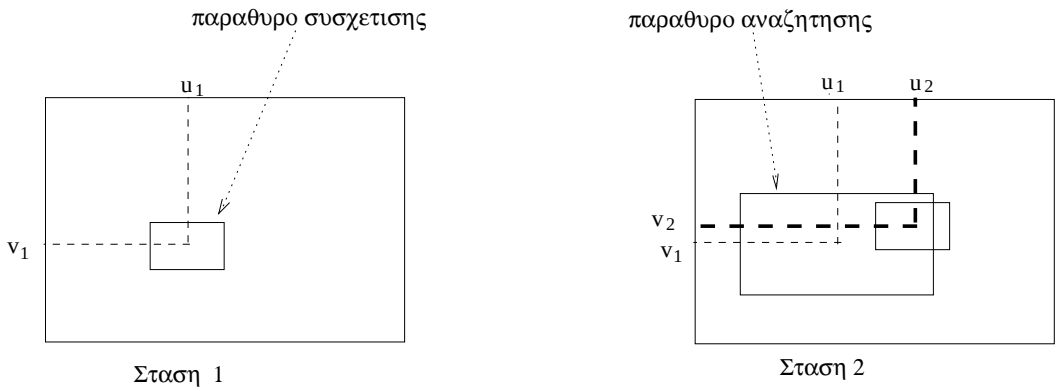
1. Υπολογίζεται το διάνυσμα των χωρικών παραγώγων πάνω στην εικόνα I (I_x, I_y)
2. Για κάθε σημείο της εικόνας p:
   - υπολογίζεται ο πίνακας C από την εξίσωση 3.19 πάνω στη (2N+1) x (2N+1) γειτονιά Q του σημείου p.
   - υπολογίζεται η μικρότερη ιδιοτιμή λ_2 του πίνακα C.
   - αν η $\lambda_2 >$ τ τότε αποθηκεύονται οι συντεταγμένες του σημείου p στη λίστα L.
3. Ταξινομείτε η λίστα L με φθίνουσα σειρά ως προς την ιδιοτιμή λ_2.
4. Ελέγχεται η ταξινομημένη λίστα L των υποψήφιων γωνιών ώστε για κάθε τρέχον σημείο p της λίστας αφαιρούνται τα σημεία της γειτονιάς του, που θεωρούνται υποψήφια για γωνία και ανήκουν στη λίστα L.

Τέλος το αποτέλεσμα του αλγορίθμου είναι μία λίστα με χαρακτηριστικά σημεία για τα οποία η ιδιοτιμή $\lambda_2 >$ τ. Τα σημεία αυτά θεωρούνται σημεία γωνιών της εικόνας. Το κατώτερο όριο τ και το μέγεθος της γειτονιάς καθορίζονται από το χρήστη.

3.5.2.2 Αντιστοίχιση Γωνιών

Για την εφαρμογή του αλγορίθμου της αντιστοίχισης των γωνιών θεωρούμε ότι έχει προηγηθεί η εξαγωγή των γωνιών στις δύο εικόνες, που επιθυμούμε να υπολογίσουμε την οπτική ροή. Ο αλγόριθμος της αντιστοίχισης αποτελεί παραλλαγή του αλγορίθμου [27]. Τα βήματα που ακολουθούνται για την επίτευξη σωστής αντιστοίχισης είναι:

1. *Αντιστοίχιση των γωνιών των δύο εικόνων, μέσω της μεθόδου της συσχέτισης.* Για κάθε σημείο m_1 της εικόνας 1 στο οποίο ανιχνεύτηκε γωνία ορίζεται μέγεθος παραθύρου (2n+1) x (2m+1) με κέντρο το σημείο m_1. Επιλέγεται επίσης ορθογώνια περιοχή αναζήτησης μεγέθους ($2d_u$+1) x ($2d_v$+1) γύρω από αυτό το σημείο στην εικόνα 2. Εκτελείται στη συνέχεια η μέθοδος της συσχέτισης μεταξύ του παραθύρου που ορίζεται από το σημείο m_1 στην πρώτη εικόνα με όλα σημεία m_2 στα οποία ανιχνεύτηκε γωνία και βρίσκονται στην περιοχή αναζήτησης στη δεύτερη εικόνα. Η περιοχή αναζήτησης αντανακλά προηγούμενη γνώση για τη μέγιστη τιμή απομάκρυνσης των αντίστοιχων γωνιών των δύο εικόνων. Η γνώση αυτή χρησιμοποιείται στη μείωση του μεγέθους της περιοχής αναζήτησης για την εύρεση



Σχήμα 3.9: Συσχέτιση Γωνιών

του αντίστοιχου σημείου, από ολόκληρη την εικόνα σε ένα δεδομένο παράθυρο αναζήτησης. Τα σημεία, στα οποία ανιχνεύθηκαν γωνίες, καλούνται χαρακτηριστικά σημεία. Το κριτήριο συσχέτισης που εφαρμόζεται είναι η τετραγωνική απόκλιση των εντάσεων των αντίστοιχων χαρακτηριστικών σημείων στα οποία οποία ανιχνεύθηκαν γωνίες στις δύο εικόνες.

$$D(m_1, m_2) = \sum_{i=-n}^{n} \sum_{j=-m}^{m} (I_1(u_1 + i, v_1 + j) - I_2(u_2 + i, v_2 + j))^2 \tag{3.21}$$

Αν η ποσότητα D(m_1, m_2) δεν ξεπερνά ένα άνω όριο τότε οι γωνίες που βρίσκονται στα σημεία m_1 και m_2 θεωρούνται υποψήφιες για αντιστοίχιση. Το κριτήριο όμως συσχέτισης εφαρμόζεται στο συνόλο των γωνιών που ανιχνεύτηκαν και βρίσκονται στο παράθυρο αναζήτησης της δεύτερης εικόνας, οπότε η τιμή του είναι πιθανό να εμφανίζεται μικρότερη του άνω ορίου για περισσότερες της μιας υποψήφιες αντίστοιχες γωνίες του σημείου m_1 της πρώτης εικόνας. Επομένως είναι πιθανό κάθε χαρακτηριστικό σημείο της πρώτης εικόνας να έχει ένα σύνολο από υποψήφια αντίστοιχα χαρακτηριστικά σημεία της δεύτερης και το αντίστροφο. Το σύνολο των υποψήφιων αντιστοιχίσεων είναι πιθανό να είναι κενό. Με την εφαρμογή όμως κατάλληλων κριτηρίων επιτυγχάνεται σωστή και ένα προς ένα αντιστοίχιση χαρακτηριστικών σημείων, από την πρώτη εικόνα στη δεύτερη εικόνα και το αντίστροφο.

Στην εφαρμογή του αλγορίθμου επιλέχθηκαν για το μέγεθος του παραθύρου συσχέτισης n=m=7 ενώ για το μέγεθος της περιοχής αναζήτησης τα d_u και d_v θεωρήθηκαν ίσα με το ένα τέταρτο του μεγέθους της εικόνας.

2. *Αφαίρεση των ασαφών υποψήφιων αντιστοιχίσεων των γωνιών των δύο εικόνων.* Το αποτέλεσμα της προηγούμενης μεθόδου όπως ειπώθηκε είναι ότι κάθε χαρακτηριστικό σημείο στην πρώτη εικόνα έχει αντιστοιχιθεί σε ένα σύνολο χαρακτηριστικών σημείων στη δεύτερη και το αντίστροφο. Η μέθοδος που χρησιμοποιείται για την αφαίρεση των ασαφών αντιστοιχίσεων ανήκει στην γνωστή κατηγορία χαλαρωτικών μεθόδων (relaxation technique).

   Για την εφαρμογή της μεθόδου, θεωρούμε ένα υποψήφιο για αντιστοίχιση διάνυσμα χαρακτηριστικών (m_{1i}, m_{2j}), όπου m_{1i} είναι ένα χαρακτηριστικό σημείο στην πρώτη εικόνα και m_{2j} είναι το αντίστοιχο χαρακτηριστικό του σημείο της δεύτερη εικόνας. Έστω N(m_{1i}) και N(m_{2j}) οι γειτονιές του m_{1i} και m_{2j} αντίστοιχα, που απεικονίζονται με κύκλο ακτίνας R. Αν το υποψήφιο διάνυσμα (m_{1i}, m_{2j}) αποτελεί ορθή αντιστοίχιση χαρακτηριστικών τότε ένας σημαντικός αριθμός υποψήφιων αντιστοιχίσεων παρουσιάζονται στη γειτονιά τους με τη μορφή (n_{1k}, n_{2l}) όπου $n_{1k} \epsilon N(m_{1i})$ και το $n_{2l} \epsilon N(m_{2j})$. Η απόσταση του n_{1k} από το m_{1i} πρέπει να είναι ίδια με την απόσταση του n_{2l} από το m_{2j}. Από την άλλη, αν το (m_{1i}, m_{2j}) θεωρείται λανθασμένη αντιστοίχιση τότε ένας μικρός αριθμός αντιστοιχίσεων παρουσιάζεται στη γειτονιά των m_{1i} και m_{2j}.

   Τυπικότερα ορίζεται η ποσότητα μέτρησης της υποστήριξης της κάθε αντιστοίχισης, που καλείται *ισχύ της αντιστοίχισης (S_M).*

$$S_M(m_{1i}, m_{2j}) = c_{ij} \sum_{n_{1k} \epsilon N(m_{1i})} [\max_{n_{2l} \epsilon N(m_{2j})} \frac{c_{kl}\delta(m_{1i}, m_{2j}; n_{1k}, n_{2l})}{1 + dist(m_{1i}, m_{2j}; n_{1k}, n_{2l})}] \qquad (3.22)$$

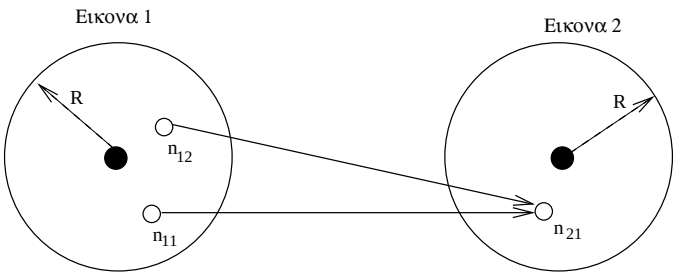
   όπου τα c_{ij} και c_{kl} εκφράζουν τη συσχέτιση της αντιστοίχισης (m_{1i}, m_{2j}) και της γειτονικής αντιστοίχισης (n_{1k}, n_{2l}) αντίστοιχα. Ενώ η συνάρτηση δ ενδυναμώνει τη συνεισφορά μιας γειτονικής αντιστοίχισης όταν η σχετική της απόσταση από την τρέχουσα αντιστοίχιση είναι μικρή. Οι παράμετροι δ, *dist* και c_{kl} ορίζονται στο Παράρτημα Γ. Γενικά υποψήφιες γωνίες για αντιστοίχιση με μικρή τιμή στην ισχύ αντιστοίχισης (3.22) απορρίπτονται.

   Το κριτήριο μέτρησης της υποστήριξης μιας αντιστοίχισης παρουσιάστηκε αποδοτικό για την ορθή αντιστοίχιση χαρακτηριστικών σημείων. Παρατηρήθηκε ότι στη γειτονιά μιας ορθής υποψήφιας αντιστοίχισης παρουσιάζονται πολλές άλλες αντιστοιχίσεις. Το κριτήριο τότε διαμορφώθηκε ώστε να ελέγχει το πλήθος των γειτονικών υποψήφιων αντιστοιχίσεων. Ο έλεγχος αυτός εφαρμόζεται αθροίζοντας τη συνεισφορά κάθε αντιστοίχισης. Επομένως τόσο πιο ορθή είναι η αντιστοίχιση όσο πιο πολλές αντιστοιχίσεις υπάρχουν στη γειτονιά της, άρα όσο πιο μεγάλη

είναι η τιμή της ισχύς της (S_M). Για τον ίδιο λόγο, όσο πιο μικρή είναι η απόσταση των γειτονικών αντιστοιχίσεων από την τρέχουσα αντιστοίχιση τόσο πιο ορθή είναι η τρέχουσα αντιστοίχιση. Έτσι στο κριτήριο της ισχύς χρησιμοποιήθηκε η σχετική απόσταση κάθε γειτονικής αντιστοίχισης από την τρέχουσα. Η σχετική απόσταση παίζει το ρόλο βάρους στη συνεισφορά των γειτονικών υποψήφιων αντιστοιχίσεων. Στην εξίσωση 3.22 η σχετική απόσταση ορίζεται από την παράμετρο r. Ενώ ο ρόλος της καθορίζεται από τη συνάρτηση δ. Για μεγάλη σχετική απόσταση η τιμή της συνάρτησης δ τείνει στο μηδέν και αντίστροφα. Επιπλέον πολλές φορές εμφανίζονται χαρακτηριστικά σημεία με περισσότερες της μιας υποψήφιες αντιστοιχίσεις. Το κριτήριο (3.22) μεριμνά ακόμα και αυτές τις περιπτώσεις με την εισαγωγή της συνάρτησης max. Έτσι, αν ένα χαρακτηριστικό σημείο γειτονικό του m_{1i} στην αριστερή εικόνα έχει ένα σύνολο υποψήφιων χαρακτηριστικών σημείων για αντιστοίχιση στη γειτονιά του m_{2j} στη δεξιά εικόνα, τότε μόνο το ένα διάνυσμα αντιστοίχισης με τη μικρότερη διαφορά απόστασης λαμβάνεται υπόψη στην ισχύ της αντιστοίχισης (m_{1i}, m_{2j}). Το κριτήριο ισχυροποιείται με την εισαγωγή της μέσης απόστασης κάθε γειτονικής αντιστοίχισης από την τρέχουσα αντιστοίχιση. Έτσι η συνεισφορά κάθε υποψηφίου για αντιστοίχιση διανύσματος της γειτονιάς λαμβάνει επιπλέον βάρος ίσο με τη μέση απόσταση του από την τρέχουσα αντιστοίχιση (m_{1i}, m_{2j}). Η πρόσθεση του παράγοντα '1' έγινε για να εμποδίσει τα πολύ μεγάλα βάρη στα πολύ κοντινά σημεία. Με άλλα λόγια, ένα κοντινό υποψήφιο προς αντιστοίχιση διάνυσμα παρέχει μεγαλύτερη υποστήριξη στην τρέχουσα αντιστοίχιση (m_{1i}, m_{2j}) από ότι ένα απομακρυσμένο.

Η μέτρηση όμως της υποστήριξης της αντιστοίχισης που ορίστηκε στην εξίσωση 3.22 δεν είναι συχνά συμμετρική. Με άλλα λόγια η ισχύς μιας αντιστοίχισης δεν είναι συχνά ίδια αν αντιστραφούν οι ρόλοι των δύο εικόνων, δηλαδή πιθανόν να ισχύει $S_M(m_{1i}, m_{2j}) \neq S_M(m_{2j}, m_{1i})$. Αυτό συμβαίνει όταν περισσότερα του ενός χαρακτηριστικά σημεία $n_{1k} \in \mathcal{N}\ (m_{1i})$ είναι υποψήφια για αντιστοίχιση με ένα μόνο χαρακτηριστικό σημείο $n_{2l} \in \mathcal{N}\ (m_{2j})$. Στο σχήμα 3.10, το χαρακτηριστικό σημείο n_{11} και το χαρακτηριστικό σημείο n_{12} μοιράζονται το ίδιο σημείο n_{21} σαν υποψήφιο για αντιστοίχιση.

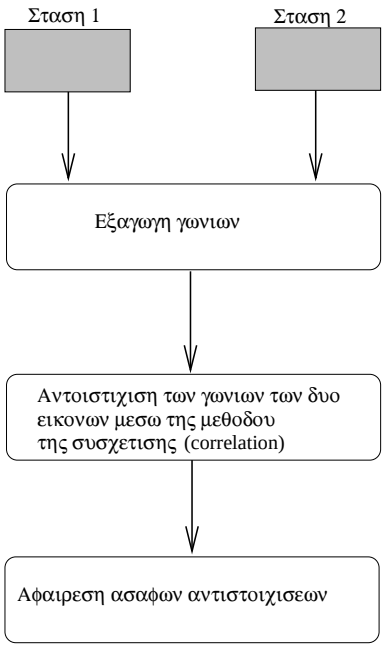
Στην εργασία μας για την επίτευξη της συμμετρίας εφαρμόστηκε η ακόλουθη τροποποίηση: Πριν τον υπολογισμό του αθροίσματος στην εξίσωση 3.22, εξετάζεται αν υπάρχουν χαρακτηριστικά σημεία $n_{1k} \in \mathcal{N}\ (m_{1i})$ που έχουν σαν υποψήφιο αντίστοιχο το ίδιο σημείο $n_{2l} \in \mathcal{N}\ (m_{2j})$, επιλέγεται το χαρακτηριστικό σημείο που



Σχήμα 3.10: Το πρόβλημα της μη-συμμετρίας στη μέτρηση της υποστήριξης μιας αντιστοίχισης

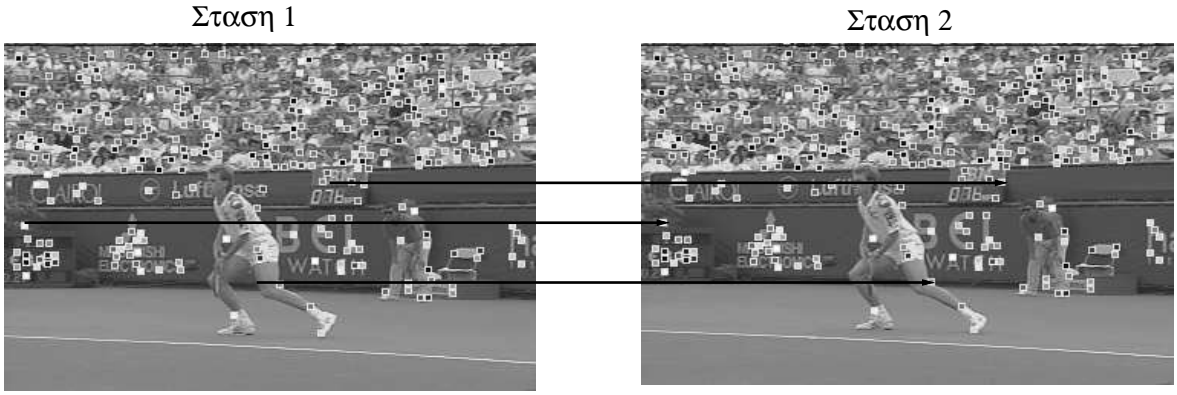
δίνει τη μεγαλύτερη τιμή συσχέτισης (c, σχέση Γ.1) με τον υποψήφιο αντίστοιχο n_{2l}. Με τον τρόπο αυτό βεβαιώνεται ότι το ίδιο ζευγάρι θα επιλεγεί ακόμα και αν αντιστραφούν οι ρόλοι των δύο εικόνων.

Η αντιστοίχιση που προκύπτει από τα παραπάνω κριτήρια είναι ικανοποιητική. Πιθανά σφάλματα αντιστοίχισης αφαιρούνται κατά την εκτίμηση των παραμέτρων του εκάστοτε μοντέλου κίνησης. Τα βήματα του αλγορίθμου αντιστοίχισης γωνιών φαίνονται στο σχήμα 3.11 Ένα παράδειγμα του αλγορίθμου αντιστοίχισης γωνιών



Σχήμα 3.11: Αλγόριθμος αντιστοίχισης γωνιών

φαίνεται στο σχήμα 3.12



Σχήμα 3.12: Παράδειγμα αντιστοίχισης γωνιών

3.6 Μέθοδοι σθεναρής εκτίμησης

Οι μέθοδοι σθεναρής εκτίμησης [40],[41] χρησιμοποιούνται για την εξαγωγή στατιστικών συμπερασμάτων. Τα στατιστικά συμπεράσματα, όπως είναι γνωστό, εξάγονται μέσω των παρατηρήσεων και των υποθέσεων που γίνονται για την εκάστοτε δεδομένη κατάσταση. Οι υποθέσεις αυτές τείνουν να περιγράψουν την πραγματική κατάσταση, παρουσιάζοντας όμως λάθη απόκλισης από αυτήν. Ο όρος *"σθεναρότητα"* δηλώνει τη μη-ευαισθησία σε μικρές αποκλίσεις από τις θεωρητικά ισχύουσες υποθέσεις. Επομένως οι μέθοδοι σθεναρής εκτίμησης είναι αποδοτικοί ακόμα και στην ύπαρξη ενός σημαντικού αριθμού μικρών αποκλίσεων. Η απόδοση τους οφείλεται στην ικανότητά τους να διακρίνουν από το σύνολο των δεδομένων τα "ορθά δεδομένα" (inliers) από εκείνα που περιέχουν λάθη (outliers), και να λαμβάνουν υπόψη τους στους υπολογισμός τους μόνο τα "ορθά δεδομένα" (inliers). Ο Hampel et al σημείωσε ότι ο βασικός σκοπός των robust τεχνικών είναι να ανακτήσουν τη δομή που ταιριάζει καλύτερα στην πλειοψηφία των δεδομένων, ενώ ταυτόχρονα να αναγνωρίζουν και να απορρίπτουν τα "κακά δεδομένα".

Δύο από τις έννοιες που χρησιμοποιούνται [42] για την αξιολόγηση ενός robust αλγορίθμου είναι *η σχετική του αποτελεσματικότητα (relative efficiency)* και *το σημείο κατάρρευσης του (breakdown point). Η σχετική αποτελεσματικότητα* εκφράζει το λόγο μεταξύ της μικρότερης δυνατής διασποράς που μπορεί να έχουν οι παράμετροι προς υπολογισμό και της πραγματικής διασποράς που προκύπτει κατά την εφαρμογή της robust μεθόδου υπολογισμού. Ενώ *το σημείο κατάρρευσης* εκφράζει το μέγιστο ποσοστό "κακών δεδομένων" (outliers) που ο αλγόριθμος μπορεί να ανεχθεί. Η απόδοση της μεθόδου τείνει να είναι η βέλτιστη όταν ο αριθμός των "κακών δεδομένων" είναι πολύ μικρότερος του σημείου κατάρρευσης. Διαφορετικά η λύση που προκύπτει από τη

μέθοδο δε είναι αξιόπιστη. Επομένως το μεγάλο σημείο κατάρρευσης μιας robust μεθόδου σημασιοδοτεί την αποτελεσματικότητά της. Η πιθανότητα, η λύση που θα δοθεί να είναι αξιόπιστη και να τείνει στην πραγματική, είναι μεγάλη.

Πολλές είναι οι μέθοδοι σθεναρής εκτίμησης που έχουν προταθεί. Οι περισσότερες από αυτές μπορούν να ομαδοποιηθούν σε τρεις κατηγορίες στους M-estimators, L-estimators και R-estimators [40], [43]. Παρακάτω θα παρουσιασθούν οι M-estimators. Η κατηγορία αυτή χρησιμοποιείται συχνά για την προσσέγγιση παραμετρικών μοντέλων κίνησης. Αποτελούν αξιόπιστες μέθοδοι σθεναρότητας με μεγάλη ανεχτικότητα στα "κακά δεδομένα".

3.6.1 Περιγραφή των M-estimators

Οι M-estimators θεωρούνται μία από τις πιο διαδεδομένες μεθόδους σθεναρής εκτίμησης. Το πρόβλημα, που αντιμετωπίζουν, στη γενική του μορφή είναι το εξής: Υποθέτουμε ότι υπάρχει ένα σύνολο από N δείγματα δεδομένων $\{x_i | i\epsilon\{1..N\}\}$. Οι μέθοδοι αυτοί προσπαθούν να προσεγγίσουν τα δείγματα αυτά κάνοντας κάποιες υποθέσεις που μπορεί να μην είναι ακριβώς οι πραγματικές. Βάση αυτών, τα δείγματα μοντελοποιούνται από την ακόλουθη σχέση:

$$x_i = T_n + n_i$$

Η παράμετρος T_n εκφράζει την προσέγγιση της μεθόδου του M-estimator στα δείγματα ενώ η n_i παριστάνει το σφάλμα της προσέγγισης T_n στο i-οστό δείγμα (x_i). Η παράμετρος n_i είναι γνωστή ως υπόλοιπο (residual). Όπως είναι φανερό το υπόλοιπο n_i=$x_i - T_n$ εκφράζει την απόκλιση του δείγματος από το μοντέλο T_n που προσεγγίζει η μέθοδος. "Κακά δεδομένα" (outliers) της μεθόδου είναι τα δείγματα που παρουσιάζουν μεγάλη απόκλιση από το μοντέλο T_n. Επομένως για να προσδιορισθούν οι outliers γίνεται χρήση της πληροφορίας που παρέχει το σύνολο των υπολοίπων (residuals). Ένα μεγάλο υπόλοιπο, υποδηλώνει ένα αντίστοιχα μεγάλο σφάλμα και κατά συνέπεια ένα πιθανό "κακό δεδομένο" (outlier). Άρα τα δείγματα με μεγάλα υπόλοιπα δεν πρέπει να λαμβάνονται σοβαρά υπόψη κατά τους υπολογισμούς και επομένως θα πρέπει να τους αποδοθεί μικρή σημασία.

Σύμφωνα με τον M-estimator η προσέγγιση T_n στο παραπάνω πρόβλημα προκύπτει από την ελαχιστοποίηση της παρακάτω συνάρτησης λάθους:

$$\min_{T_n} \sum \rho(x_i - T_n) \tag{3.23}$$

Για την ελαχιστοποίηση της παραπάνω συνάρτησης, αρκεί η επίλυση της εξίσωσης:

$$\sum \psi(x_i - T_n) = 0 \tag{3.24}$$

Η συνάρτηση $\psi(x;\theta) = \frac{\vartheta}{\vartheta\theta}\rho(x;\theta)$ καλείται M-προσέγγιση. Η επιλογή $\rho(x;\theta) = -logf(x;\theta)$ δίνει τη γνήσια ML-προσέγγιση. Η τελευταία εξίσωση μπορεί να πάρει την εξής μορφή:

$$\sum w(x_i - T_n)\ (x_i - T_n) = 0 \tag{3.25}$$

$$\text{όπου } w(x_i - T_n) = \frac{\psi(x_i - T_n)}{x_i - T_n} \tag{3.26}$$

Η λύση T_n που δίνει ο M-estimator εκφράζεται από την παρακάτω μέση τιμή με βάρη:

$$T_n = \frac{\sum w(x_i - T_n)\ x_i}{\sum w(x_i - T_n)} \tag{3.27}$$

τα βάρη εξαρτώνται κάθε φορά από το τρέχον δείγμα. Η μέθοδος του M-estimator είναι επαναληπτική και σκοπεύει στην καλύτερη προσέγγιση των "ορθών δεδομένων". Σε κάθε βήμα της επανάληψης υπολογίζονται τα νέα σφάλματα $x_i - T_n$ (residuals) με βάση τη τιμή της προσέγγισης T_n που υπολογίστηκε στο προηγούμενο βήμα. Κατόπιν υπολογίζεται η νέα τιμή της προσέγγισης T_n. Η διαδικασία συνεχίζεται μέχρι τη σύγκλιση του αλγορίθμου ή για ένα καθορισμένο αριθμό επαναλήψεων. Κατά την εκκίνηση του αλγορίθμου θεωρούμε ότι όλα τα σημεία έχουν το ίδιο βάρος. Η αρχική τιμή της προσέγγισης T_n είναι:

$$T_n = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.28}$$

Αναλυτικότερα η μέθοδος σε μορφή ψευδοκώδικα έχει ως εξής:

1. **Αρχικοποίηση**

   - t = 1; /* πρώτη επανάληψη (t είναι ο αριθμός των επαναλήψεων) */
   - $T_n^{(t)} = \frac{1}{N}\sum_{i=1}^{N} x_i$

2. **Επαναληπτική διαδικασία**{

   - /*υπολογισμός σφαλμάτων n_i */

     for(i=1; i≤ N; i++)

     $r_i = x_i - T_n^{(t)}$

   - t ++; /* αύξηση του αριθμού των επαναλήψεων */

- /* υπολογισμός της νέας τιμής της προσσέγγισης T_n */
  $T_n = \frac{\sum w(x_i - T_n)\ x_i}{\sum w(x_i - T_n)}$

}**until**($|T_n^{(t)} - T_n^{(t-1)}| < \epsilon$) /* σύγκλιση αλγορίθμου (ή (t > K) όπου K ο μέγιστος αριθμός επαναλήψεων) */

3. Η $T_n^{(t)}$ εκφράζει την τελική προσέγγιση της μεθόδου στο σύνολο των δειγμάτων.

Η συνάρτηση $\rho(x)$ ονομάζεται *συνάρτηση κόστους (error penalty function).* Το όρισμα της συνάρτησης σε robust τεχνικές είναι το υπόλοιπο r_i (residual). Όσο πιο μεγάλο είναι το υπόλοιπο ενός δείγματος τόσο μεγαλύτερο είναι το κόστος που πληρώνουμε. Η συνάρτηση w ονομάζεται *συνάρτηση επίδρασης (influence function).* Η συνάρτηση επίδρασης είναι ανάλογη της συνάρτησης ψ που εκφράζει την κλίση της συνάρτησης κόστους ρ. Ο ρόλος της συνάρτησης w είναι να καθορίζει τη σημασία του κάθε δείγματος στο τελικό αποτέλεσμα αποδίδοντας ένα βάρος σε κάθε υπόλοιπο ανάλογα με την τιμή του. Όσο πιο μεγάλο υπόλοιπο έχουν τα δείγματα τόσο πιο μικρό βάρος τους αποδίδεται. Αντίθετα όταν το υπόλοιπο είναι πολύ μικρό το βάρος που αποδίδεται πλησιάζει τη μονάδα. Επομένως με την επιλογή κατάλληλων συναρτήσεων ρ ή w μειώνεται η επίδραση των "κακών δεδομένων" (outliers) στο τελικό αποτέλεσμα, ενώ οι υπολογισμοί στηρίζονται μόνο στα σωστά δεδομένα. Παρακάτω θα παρουσιασθούν κάποιες συναρτήσεις που παίζουν το ρόλο των συναρτήσεων ρ και w.

3.6.2 Συναρτήσεις κόστους και επίδρασης

Για την εφαρμογή της τεχνικής του M-estimator απαιτείται ο ορισμός της συνάρτησης κόστους (ρ) ή της συνάρτησης επίδρασης (w). Δεδομένου όμως ότι η σταθερότητα και η δυνατότητα του M-estimator να διακρίνει τα "κακά δεδομένα" οφείλεται στις συναρτήσεις αυτές, η επιλογή τους απαιτείται ιδιαίτερης προσοχής.

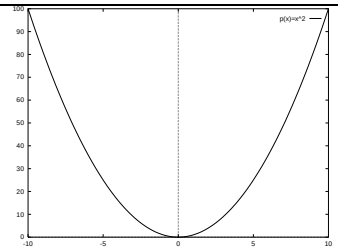
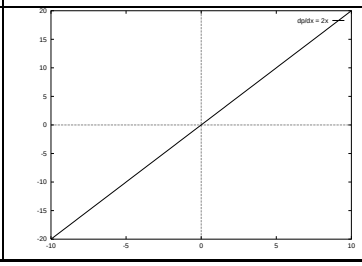
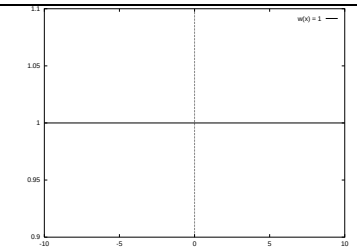
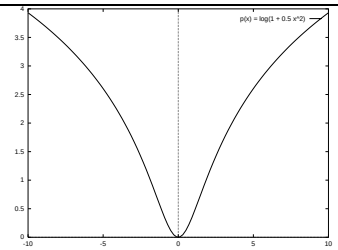
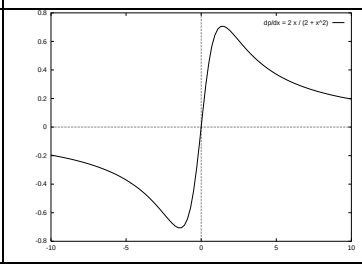
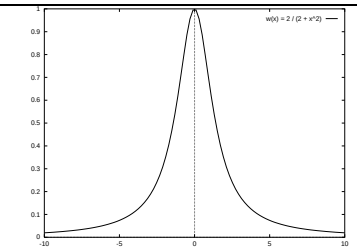
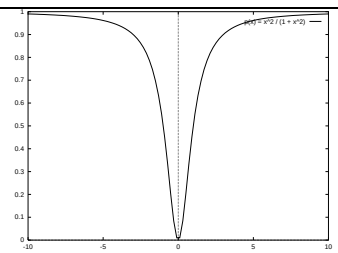
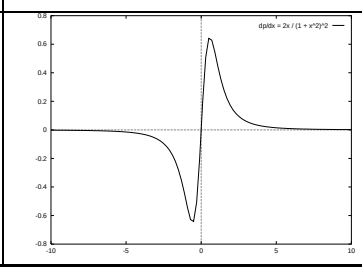
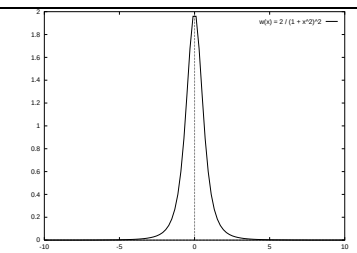
Οι συναρτήσεις κόστους και κατ' επέκταση οι συναρτήσεις επίδρασης που επιλέχθηκαν παραθέτονται παρακάτω. Οι γραφικές παραστάσεις των συναρτήσεων κόστους (ρ) και επίδρασης (w) (με σ=1) φαίνονται στο πίνακα 3.2

- **Τετραγωνική συνάρτηση**.

$$\begin{aligned} \rho(x) &= x^2 \\ \psi(x) &= 2x \\ w(x) &= 1 \end{aligned} \tag{3.29}$$

Η τετραγωνική συνάρτηση αποτελεί τη συνάρτηση κόστους της μεθόδου των *ελαχίστων τετραγώνων.* Στη μέθοδο αυτή η συνάρτηση επίδρασης w ισούται με τη

Πίνακας 3.2: Συναρτήσεις Κόστους και Επίδρασης

Τετραγωνική Συνάρτηση		
$\rho(x) = x^2$	$\psi(x) = 2x$	$w(x) = 1$
		
Lorentzian Συνάρτηση		
$\rho(x,\sigma) = log(1 + \frac{1}{2}(\frac{x}{\sigma})^2)$	$\psi(x,\sigma) = \frac{2x}{2\sigma^2 + x^2}$	$w(x,\sigma) = \frac{2}{2\sigma^2 + x^2}$
		
Geman & Mc-Clure Συνάρτηση		
$\rho(x,\sigma) = \frac{\frac{x^2}{\sigma^2}}{1 + \frac{x^2}{\sigma^2}}$	$\psi(x,\sigma) = \frac{\frac{2x}{\sigma^2}}{\left(1 + \frac{x^2}{\sigma^2}\right)^2}$	$w(x,\sigma) = \frac{2\sigma^2}{(\sigma^2 + x^2)^2}$
		

μονάδα για κάθε δείγμα ”καλό” ή ”κακό”. Επομένως όλα τα δεδομένα έχουν το ίδιο βάρος στο τελικό αποτέλεσμα. Αυτό έχει σαν αποτέλεσμα η μέθοδος να είναι εξαιρετικά ευαίσθητη στα λάθη, ενώ η ύπαρξη ακόμα και ενός ”κακού δεδομένου” είναι σε θέση να οδηγήσει σε λανθασμένη εκτίμηση του τελικού αποτελέσματος. Μια παραλλαγή της τετραγωνικής συνάρτησης εκφράζεται με τη χρήση κατωφλιού που διακρίνει τα ”καλά” από τα ”κακά” δεδομένα. Η παραλλαγή αυτή εκφράζεται από τις συναρτήσεις:

$$\begin{aligned}
\rho(x) &= \begin{cases} x^2 & : \quad x < P \\ A & : \quad \text{αλλιώς} \end{cases} \\
\psi(x) &= \begin{cases} 2x & : \quad x < P \\ 0 & : \quad \text{αλλιώς} \end{cases} \\
w(x) &= \begin{cases} 1 & : \quad x < P \\ 0 & : \quad \text{αλλιώς} \end{cases}
\end{aligned} \tag{3.30}$$

Η παράμετρος P εκφράζει το κατώφλι που διακρίνει το σύνολο των δειγμάτων σαν ”καλά” (inliers) ή ”κακά” (outliers) δεδομένα.

- **Lorentzian συνάρτηση**.

$$\begin{aligned}
\rho(x,\sigma) &= log(1 + \tfrac{1}{2}(\tfrac{x}{\sigma})^2) \\
\psi(x,\sigma) &= \tfrac{2x}{2\sigma^2 + x^2} \\
w(x,\sigma) &= \tfrac{2}{2\sigma^2 + x^2}
\end{aligned} \tag{3.31}$$

Είναι φανερό από τη γραφική παράσταση της συνάρτησης της επίδρασης ότι στα δείγματα, που το υπόλοιπό τους τείνει στο μηδέν, αποδίδεται βάρος επίδρασης στο αποτέλεσμα περίπου ίσο με τη μονάδα. Αντίθετα η επίδραση των δειγμάτων με μεγάλο σφάλμα τείνει στο μηδέν. Στις συναρτήσεις αυτές σημαντικό ρόλο παίζει η παράμετρος σ που καλείται *παράμετρος κλίμακας*. Η παράμετρος σ που ελέγχει τη θέση του σημείου κλίσης της καμπύλης, σημασιοδοτεί το σημείο πέρα από το οποίο υπάρχει μείωση στο βάρος επίδρασης των δειγμάτων.

- **Geman & McClure συνάρτηση**.

$$\begin{aligned}
\rho(x,\sigma) &= \frac{\frac{x^2}{\sigma^2}}{1 + \frac{x^2}{\sigma^2}} \\
\psi(x,\sigma) &= \frac{\frac{2x}{\sigma^2}}{\left(1 + \frac{x^2}{\sigma^2}\right)^2} \\
w(x,\sigma) &= \frac{2\sigma^2}{(\sigma^2 + x^2)^2}
\end{aligned} \tag{3.32}$$

Συγκρίνοντας τις γραφικές παραστάσεις των δύο τελευταίων συναρτήσεων επίδρασης συμπεραίνεται ότι η συνάρτηση Geman & McClure θεωρεί μεγαλύτερο ποσοστό

δειγμάτων ως ”κακά δεδομένα”. Επομένως στο τελικό αποτέλεσμα συμμετέχουν μόνο τα πιο ”καλά” δείγματα. Σε αυτόν ακριβώς το λόγο οφείλεται και η καλύτερη συμπεριφορά που παρουσιάζει η συνάρτηση αυτή. Και σε αυτή τη συνάρτηση ο ρόλος της παραμέτρου κλίμακας (σ) είναι ουσιαστικός.

Η επιλογή της παραμέτρου κλίμακας (σ) παίζει σημαντικό ρόλο στη συμπεριφορά και αποτελεσματικότητα του M-estimator. Η παράμετρος αυτή ρυθμίζει το σημείο κλίσης της συνάρτησης της επίδρασης. Καθορίζει δηλαδή το πλήθος των δειγμάτων με τη μέγιστη επίδραση (inliers) στο τελικό αποτέλεσμα. Μια κακή επιλογή της παραμέτρου κλίμακας μπορεί να οδηγήσει σε λανθασμένα αποτελέσματα. Δύο από τις πιο συνηθισμένες επιλογές είναι:

$$\sigma = \sqrt{\frac{1}{N} \sum (r_i^2)} \tag{3.33}$$

και

$$\begin{aligned} &\sigma = c_0 \, median\,(|r_i|) \\ &c_0 = 1.4826 \end{aligned} \tag{3.34}$$

Η τελευταία επιλογή παρουσιάζει καλύτερα αποτελέσματα. Η παράμετρος c_0 εκφράζει μία σταθερά κανονικοποίησης. Η εκτίμηση βάση της ενδιάμεσης τιμής ανέχεται την ύπαρξη ”κακών δεδομένων” με ποσοστό ανεκτικότητας 50%. Ο υπολογισμός της παραμέτρου κλίμακας λαμβάνει μέρος σε κάθε επανάληψη του M-estimator χρησιμοποιώντας την τρέχουσα τιμή της προσέγγισης $\mathrm{T}_n^{(t)}$.

3.6.3 Μειονεκτήματα της Χρήσης του M-estimator

Παρόλο της διαδεδομένης χρήσης του ο αλγόριθμος M-estimator παρουσιάζει ένα αριθμό προβλημάτων που η γνώση τους οδηγεί στη καλύτερη χρήση του. Τα σημαντικότερα από αυτά είναι:

- Ο αλγόριθμος αυτός είναι αρκετά ευαίσθητος στην αρχική εκτίμηση των παραμέτρων T_n (3.28). Μια κακή αρχική εκτίμηση μπορεί να οδηγήσει ακόμα και στη μη-σύγκλιση του αλγορίθμου. Ακόμα όμως και με την αποφυγή της μη-σύγκλισης, η αστάθεια του αλγορίθμου παραμένει. Δεδομένου ότι ανήκει στην κατηγορία των αλγορίθμων που εξαρτώνται από την κλίση (gradient) χαρακτηρίζεται από την αδυναμία εύρεσης του ελαχίστου. Συνήθως η ύπαρξη τοπικών ελαχίστων δυσκολεύει τις μεθόδους αυτές στην εύρεση του ολικού ελαχίστου. Η σύγκλιση σε τοπικό

ελάχιστο οδηγεί τον αλγόριθμο σε λανθασμένα αποτελέσματα. Επομένως για την αποτελεσματικότητα του αλγορίθμου απαιτείται καλή αρχική εκτίμηση.

- Όπως αναφέρθηκε και προηγούμενα η αποτελεσματικότητα του M-estimator εξαρτάται από την παράμετρο κλίμακας. Μια κακή εκτίμηση της παραμέτρου οδηγεί σε αστάθεια του αλγορίθμου. Αναγκαία είναι επομένως η ορθή επιλογή της.

- Ο καθορισμός του κατωφλιού για την κατάταξη ενός σημείου ως "καλό" ή "κακό" δεδομένο αποτελεί άλλη μια περίπτωση αστάθειας του αλγορίθμου.

3.6.4 Εκτίμηση των παραμετρικών μοντέλων με τη χρήση του M-estimator

Μια από τις πιο διαδεδομένες χρήσεις του αλγορίθμου M-estimator είναι η εκτίμηση των παραμέτρων ενός μοντέλου κίνησης. Στην περίπτωση αυτή θεωρείται ότι το σύνολο των δεδομένων που διαθέτουμε μοντελοποιείται με ένα από τα μοντέλα κίνησης που μελετήθηκαν προηγούμενα και παρουσιάστηκαν στον πίνακα 3.1.

Το σύνολο των δεδομένων είναι το αποτέλεσμα της εκτίμησης της οπτικής ροής. Επομένως τα δεδομένα αποτελούνται από ζευγάρια διανυσμάτων οπτικής ροής και δισδιάστατων σημείων του επιπέδου της εικόνας ή από τα ζευγάρια των αντίστοιχων δισδιάστατων σημείων των επιπέδων των δύο εικόνων, στις οποίες εκτιμάται η κίνηση της κάμερας. Δηλαδή:

$$\begin{aligned} &((u_i,\ v_i)\ ,\ (x_i,\ y_i)\ ) \qquad i\text{=}1...N \\ &\text{ή} \\ &((x_i',\ y_i')\ ,\ (x_i,\ y_i)\ ) = (p_i',\ p_i) \qquad i\text{=}1...N \end{aligned} \tag{3.35}$$

όπου N είναι το σύνολο των δειγμάτων και άρα των εκτιμήσεων που απέμειναν μετά την εφαρμογή ενός από τους αλγορίθμους εκτίμησης της οπτικής ροής που αναφέραμε.

Έστω ότι ζητείται η εκτίμηση της κίνησης μεταξύ δύο εικόνων I_1 και I_2. Εφαρμόζοντας κάποιον από τους αλγορίθμους εκτίμησης της οπτικής ροής που αναφέραμε διαθέτουμε ένα σύνολο αντιστοιχίσεων της μορφής $(p_i,\ p_i')$. Το p_i εκφράζει ένα δισδιάστατο σημείο της εικόνας I_1, ενώ το p_i' το αντίστοιχο σημείο του p_i στην εικόνα I_2. Υποθέτουμε ότι η κίνηση της κάμερας μοντελοποιείται βάση ενός μετασχηματισμού M με παραμέτρους θ. Η αντιστοίχιση των σημείων στις δύο εικόνες βάση του μετασχηματισμού είναι:

$$p_i'(\theta)\ = M\,(p_i) \tag{3.36}$$

όπου $p_i'(\theta)$ είναι το αντίστοιχο σημείο του p_i στην εικόνα I_2 όπως προκύπτει από το μετασχηματισμό Μ. Ενώ το p_i' είναι το αντίστοιχο σημείο στην εικόνα I_2 όπως όμως υπολογίσθηκε αρχικά από την οπτική ροή. Το σφάλμα μεταξύ των δειγμάτων και των αντιστοιχίσεων, που προκύπτουν από τις παραμέτρους του μοντέλου, έχει μία από τις ακόλουθες μορφές:

$$r_i = p_i'(\theta) - p_i' \tag{3.37}$$

ή

$$r_i = I_2(p_i'(\theta)) - I_1(p_i) \tag{3.38}$$

Το πρώτο κριτήριο στηρίζεται καθαρά στη γεωμετρία των σημείων στο επίπεδο, ενώ στο δεύτερο λαμβάνονται υπόψη μόνο οι φωτεινότητες. Το πρώτο κριτήριο υπερτερεί έναντι του δεύτερου υπό την προϋπόθεση ότι οι αντιστοιχίσεις των σημείων στις δύο εικόνες είναι ακριβείς. Ο λόγος που υπερέχει είναι ότι τυχόν αλλαγές στο φωτισμό των δύο εικόνων δεν το επηρεάζουν αφού δεν εξαρτάται από τις φωτεινότητες. Με την ελαχιστοποίηση μιας από τις συναρτήσεις κόστους που αναφέραμε, με όρισμα ένα από τα σφάλματα r_i, υπολογίζονται οι παράμετροι θ του μετασχηματισμού Μ. Για παράδειγμα, υποθέτουμε ότι ο μετασχηματισμός μεταξύ των εικόνων I_1 και I_2 περιγράφεται από το γνωστό αφινικό μοντέλο:

$$\begin{aligned} x'(a) &= a_1 x + a_2 y + a_3 \\ y'(a) &= a_4 x + a_5 y + a_6 \end{aligned} \tag{3.39}$$

όπου $(x'(a),\ y'(a))$ και $(x,\ y)$ είναι οι συντεταγμένες των σημείων $p_i'(a)$ (όπως υπολογίζεται από τις παραμέτρους a του μοντέλου) και p_i αντίστοιχα. $a_1 ... a_6$ είναι το σύνολο των παραμέτρων του μοντέλου που θα προσεγγιστούν από τον M-estimator. Συγκεκριμένα, οι τιμές των παραμέτρων θα προκύψουν σαν λύσεις του M-estimator κατά την ελαχιστοποίηση της συνάρτησης σφάλματος:

$$E(a_1, ..., a_6) = \min_{a_i} \sum_{i=1}^{N} \varrho(p_i'(a) - p_i') \tag{3.40}$$

Στην περίπτωση των ελαχίστων τετραγώνων και με σφάλμα r_i αυτό της εξίσωσης 3.37, η σχέση 3.40 μετατρέπεται στην εξής:

$$\min_{a_i} \sum_{i=1}^{N} (x' - a_1 x - a_2 y - a_3)^2 + (y' - a_4 x - a_5 y - a_6)^2 \tag{3.41}$$

Στην παρούσα εργασία κατά την εφαρμογή αυτή του αλγορίθμου, η αρχική εκτίμηση των παραμέτρων του συνόλου των μοντέλων του πίνακα 3.1 προκύπτει από τη μέθοδο των ελαχίστων τετραγώνων. Όλες οι αντιστοιχίσεις που υπολογίστηκαν με τις μεθόδους εκτίμησης της οπτικής ροής συμμετέχουν στον υπολογισμό των παραμέτρων με βάρος επίδρασης ίσο με τη μονάδα. Στην περίπτωση όμως του προβολικού μοντέλου η αρχική εκτίμηση των παραμέτρων εφαρμόζεται στην ακόλουθη παραλλαγή του.

$$\begin{cases} x' = \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1} \\ y' = \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1} \end{cases} \Longrightarrow \begin{cases} a_7x'x + a_8x'y + x' = a_1x + a_2y + a_3 \\ a_7y'x + a_8y'y + y' = a_4x + a_5y + a_6 \end{cases} \tag{3.42}$$

Η σχέση τότε των ελαχίστων τετραγώνων που ελαχιστοποιείται είναι η:

$$\begin{aligned} \min_{a_i} \sum_{i=1}^{N} (a_7x'x + a_8x'y + x' - a_1x - a_2y - a_3)^2 \\ +(a_7y'x + a_8y'y + y' - a_4x - a_5y - a_6)^2 \end{aligned} \tag{3.43}$$

αντί της σχέσης:

$$\min_{a_i} \sum_{i=1}^{N} \left(x' - \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1} \right)^2 + \left(y' - \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1} \right)^2 \tag{3.44}$$

όπου τα $(x',\ y')$ και $(x,\ y)$ αποτελούν τις αντιστοιχίσεις σημείων μεταξύ των δύο εικόνων, που προήλθαν από την εκτίμηση της οπτικής ροής.

Ο λόγος αυτής της παραλλαγής είναι για την επίτευξη γρηγορότερης αρχικής εκτίμησης των παραμέτρων. Συγκεκριμένα, η μορφή του προβολικού μετασχηματισμού οδηγεί για τον υπολογισμό των παραμέτρων του, στη λύση ενός μη-γραμμικού συστήματος που ως γνωστό αποτελεί χρονοβόρα διαδικασία. Η ελαχιστοποίηση όμως της εξίσωσης 3.43 οδηγεί σε τιμές των παραμέτρων που αποκλίνουν από τις πραγματικές αφού αποτελούν λύση μιας άλλης παραλλαγμένης εξίσωσης. Το γεγονός όμως ότι αποτελούν αρχικές εκτιμήσεις των παραμέτρων δεν οδηγεί σε περαιτέρω προβλήματα. Οι πραγματικές τιμές των παραμέτρων θα υπολογισθούν κατά την επανάληψη του αλγορίθμου με την ελαχιστοποίηση της σχέσης 3.44. Επίσης σε κάθε επανάληψη του αλγορίθμου M-estimator απορρίπτονται σημεία με τιμή κριτηρίου μεγαλύτερη του $2 * \sigma$. Με την απόρριψη αυτή αυξάνεται η απόδοση του αλγορίθμου.

Είναι φανερό ότι η ελαχιστοποίηση μιας από τις συνάρτησεις κόστους ρ, που αναφέραμε, πλην όμως της τετραγωνική συνάρτησης με κατώφλι ή χωρίς, ή ακόμα και η χρήση του προβολικού μετασχηματισμού, οδηγεί στη λύση ενός μη-γραμμικού συστήματος. Για την αποφυγή της λύσης ενός τέτοιου συστήματος χρησιμοποιήθηκε η επαναληπτική μέθοδος Levemberg-Marquard που αποτελεί περίπτωση της γενικής μεθόδου ελαχιστοποίησης Gauss-Newton.

Λόγω των προβλημάτων που εμφανίζει γενικότερα ο αλγόριθμος M-estimator είναι δυνατό οι τιμές των παραμέτρων του εκάστοτε μοντέλου κίνησης να μην είναι εντελώς ακριβείς. Το γεγονός αυτό αντιμετωπίζεται με τη διόρθωση των τελικών εκτιμημένων παραμέτρων. Για τη διόρθωση χρησιμοποιούνται μέθοδοι που στηρίζονται στο διάνυσμα της κλίσης (gradient).

3.6.5 Μέθοδος Marquardt-Levemberg για την εκτίμηση των παραμέτρων

Η μέθοδος Marquardt-Levemberg (ML) [43] για την εκτίμηση των παραμέτρων αποτελεί προσέγγιση της γενικής μεθόδου του Newton για προβλήματα που περιλαμβάνουν ελαχιστοποίηση του αθροίσματος ορισμένων συναρτήσεων, ως προς κάποιο αριθμό άγνωστων παραμέτρων. Θα παρουσιαστεί αρχικά η γενική μέθοδος του Newton και κατόπιν θα αναφέρουμε τις αλλαγές που παρουσιάζει η μέθοδος ML. Στη μέθοδο Newton χρησιμοποιείται ο αντίστροφος πίνακας Hessian (πίνακας των δευτέρων μερικών παραγώγων) σαν βήμα για τη διόρθωση των παραμέτρων. Για κάθε στοιχείο του πίνακα H_{kl} ισχύει $H_{kl} = \sum_i (\frac{\vartheta^2 \rho}{\vartheta r_i^2} \frac{\vartheta r_i}{\vartheta \theta_l} \frac{\vartheta r_i}{\vartheta \theta_k} + \frac{\vartheta \rho}{\vartheta r_i} \frac{\vartheta^2 r_i}{\vartheta \theta_k \vartheta \theta_l})$. Η μέθοδος αυτή μοιάζει με τη γνωστή μέθοδο της Απότομης Κλίσης και είναι επίσης επαναληπτική. Σε κάθε βήμα της επανάληψης οι παράμετροι υφίστανται διόρθωση. Ο δεύτερος όρος του πίνακα Hessian της συνάρτησης του κόστους παραλείπεται, αφού απαιτείται τα στοιχεία του Hessian πίνακα να είναι παντού θετικά. Η παραδοχή αυτή λαμβάνεται επειδή τα στοιχεία αυτά εκφράζουν την κατεύθυνση του βήματος της διόρθωσης που πρέπει να είναι φθίνουσα. Κάθε στοιχείο του πίνακα τότε έχει τη μορφή: $H_{kl} = \sum_i \frac{\vartheta^2 \rho}{\vartheta r_i^2} \frac{\vartheta r_i}{\vartheta \theta_l} \frac{\vartheta r_i}{\vartheta \theta_k}$. Η μέθοδος ML παραλείπει επίσης το δεύτερο όρο προσθέτοντας ταυτόχρονα στα διαγώνια στοιχεία του πίνακα Hessian μια παράμετρος για την απόδοση καλύτερων αποτελέσματων.

Η διόρθωση των παραμέτρων $\delta\theta^{(m)}$ στο m+1 βήμα της επανάληψης, δεδομένης μιας λύσης $\theta^{(m)}$, δίνεται από την εξίσωση:

$$\delta\theta^{(m)} = -H^{-1}(\theta^{(m)})\, g(\theta^{(m)}) \tag{3.45}$$

και

$$\theta^{(m+1)} = \theta^{(m)} + \alpha\, \delta\theta^{(m)} \tag{3.46}$$

για θετικό α. Το $\mathbf{H}(\theta^{(\mathbf{m})})$ είναι η προσέγγιση που αναφέρθηκε για τον πίνακα Hessian της συνάρτησης κόστους (ρ), που περιλαμβάνει δηλαδή μόνο τις πρώτες παραγώγους της παραμέτρου r_i, που εκφράζει το υπόλοιπο (residual). Ενώ η παράμετρος $g(\theta^{(m)})$ εκφράζει το διάνυσμα των πρώτων παραγώγων της συνάρτησης κόστους ως προς τις τρέχουσες

παραμέτρους $\theta^{(m)}$. Οι συνιστώσες του διανύσματος g και τα στοιχεία του πίνακα H δίνονται από τις σχέσεις:

$$g_k = \sum_i \frac{\vartheta\rho}{\vartheta r_i}\frac{\vartheta r_i}{\vartheta\theta_k} \qquad H_{kl} = \sum_i \frac{\vartheta^2\rho}{\vartheta r_i^2}\frac{\vartheta r_i}{\vartheta\theta_l}\frac{\vartheta r_i}{\vartheta\theta_k} \tag{3.47}$$

όπου g_k το κ-οστό στοιχείο του g και H_{kl} το στοιχείο k-γραμμής και l-στήλης του πίνακα Hessian. Η διόρθωση $\delta\theta$ των παραμέτρων προκύπτει από τη λύση ενός γραμμικού συστήματος K εξισώσεων που δίνεται από τη σχέση:

$$\sum_i H_{kl}\delta\theta_l = -g_k, \qquad k, l=1...K, i=1...N \tag{3.48}$$

Στην περίπτωση της μεθόδου των *ελαχίστων τετραγώνων(SS)*, $\left(\frac{\vartheta^2\rho}{\vartheta r_i^2} = \frac{1}{\sigma^2} \text{ και } \frac{\vartheta\rho}{\vartheta r_i} = \frac{r_i}{\sigma^2}\right)$ η παραπάνω εξίσωση μετατρέπεται στην:

$$\sum_l\sum_i \frac{\vartheta r_i}{\vartheta\theta_k}\frac{\vartheta r_i}{\vartheta\theta_l}\delta\theta_l = -r_i\frac{\vartheta r_i}{\vartheta\theta_k}, \quad k, l=1...K, i=1...N.$$

Ωστόσο για μη-τετραγωνικές συναρτήσεις ρ, το διάνυσμα των δευτέρων παραγώγων τους $\left(\frac{\vartheta^2\rho}{\vartheta r_i^2}\right)$ είναι πιθανό να είναι αρνητικό. Η διόρθωση τότε κατά την επανάληψη της μεθόδου είναι δυνατό να μην υπολογίζεται κατά τη φθίνουσα κατεύθυνση της συνάρτησης κόστους. Για παράδειγμα, ο παράγοντας $\frac{\vartheta^2\rho}{\vartheta r_i^2}$ για τις συναρτήσεις κόστους ρ_{LO} (Lorentzian) και ρ_{GM}(Geman & McClure), είναι αντίστοιχα:

$$\frac{4\sigma^2 - 2r^2}{(2\sigma^2 + r^2)^2} \quad \& \quad \frac{2\sigma^2(\sigma^2 - 3r^2)}{(\sigma^2 + r^2)^3}$$

Προσεγγίζοντας το $\ddot{\rho}(r_i) = \frac{\vartheta^2\rho}{\vartheta r_i^2}$ με τον παράγοντα $\frac{\hat{\rho}(r_i)}{r_i} = \frac{\frac{\vartheta\rho(r_i)}{\vartheta r_i}}{r_i}$ που είναι παντού θετικός, η εξίσωση 3.48 μετατρέπεται στην εξής:

$$\sum_l\sum_i \frac{\hat{\rho}(r_i)}{r_i}\frac{\vartheta r_i}{\vartheta\theta_k}\frac{\vartheta r_i}{\vartheta\theta_l}\delta\theta_l = -\frac{\hat{\rho}(r_i)}{r_i}r_i\frac{\vartheta r_i}{\vartheta\theta_k}, \quad k, l=1...K, i=1...N. \tag{3.49}$$

τα στοιχεία τότε του πίνακα Hessian έχουν τη μορφή:

$$H_{kl} = \sum_i \frac{\hat{\rho}(r_i)}{r_i}\frac{\vartheta r_i}{\vartheta\theta_k}\frac{\vartheta r_i}{\vartheta\theta_l}. \tag{3.50}$$

Η διαφορά της μεθόδου ML από την μέθοδο Newton έγκειται ακριβώς στη διαφορετική προσέγγιση που υποστηρίζει για τον πίνακα Hessian. Η ML ορίζει έναν καινούργιο πίνακα ο οποίος διαφέρει από τον Hessian στα διαγώνια στοιχεία του. Ο λόγος αυτής της διαφοροποίησης είναι η καλύτερη προσέγγιση του βήματος της διόρθωσης. Το g_k όπως ήδη ειπώθηκε είναι το διάνυσμα των πρώτων παραγώγων και εκφράζει την κλίση της

συνάρτησης του σφάλματος και όχι πόσο μακριά θα επεκταθεί η κλίση. Το πλάτος της κλίσης εκφράζεται από τον πίνακα Hessian. Η ML αποτελεί περίπτωση της μεθόδου της Απότομης Κλίσης: $\delta\theta_k = -constant\ x\ g_k$, θέτοντας στη μεταβλητή constant συγκεκριμένη τιμή από τον πίνακα Hessian. Η διάσταση της μεταβλητής αυτής είναι θ_k^2 αφού $\frac{1}{\theta_k}$ είναι η διάσταση του g_k και θ_k η διάσταση του $\delta\theta_k$. Παρατηρώντας τα στοιχεία του πίνακα Hessian [H] μόνο μια ποσότητα υπάρχει με αυτήν τη διάσταση, και αυτή είναι τα διαγώνια στοιχεία του πίνακα $\frac{1}{\mathrm{H}_{kk}}$. Θέτουμε τότε τη μεταβλητή constant που εκφράζει το βήμα της διόρθωσης ίση με $\frac{1}{\mathrm{H}_{kk}}$. Επειδή όμως το βήμα αυτό μπορεί να είναι αρκετά μεγάλο από μόνο του, διαιρούμε τη μεταβλητή αυτή με μία σταθερά $\lambda \gg 1$ για τη μείωση του βήματος. Η εξίσωση της απότομης κλίσης μετατρέπεται στην εξής:

$$\delta\theta_l = -\frac{1}{\lambda\,\mathrm{H}_{ll}} g_l \quad \text{ή} \quad \lambda\,\mathrm{H}_{ll}\,\delta\theta_l = -g_l \tag{3.51}$$

Είναι απαραίτητο το στοιχείο H_{ll} να είναι θετικό, το οποίο εγγυάται η εξίσωση 3.50.

Με βάση τα παραπάνω η μέθοδος ML ορίζει το νέο πίνακα H′ με στοιχεία:

$$\begin{aligned} H'_{kk} &\equiv H_{kk}(1 + \lambda) \\ H'_{kl} &\equiv H_{kl} \quad (k \neq l) \end{aligned} \tag{3.52}$$

Αντικαθιστώντας το νέο πίνακα στην εξίσωση 3.48 προκύπτει η ακόλουθη:

$$\sum_i H'_{kl}\,\delta\theta_l = -g_k, \quad k,l{=}1...K,\ i{=}1...N \tag{3.53}$$

Όταν το λ είναι πολύ μεγάλο, ο πίνακας H' παρουσιάζεται να κυριαρχεί στη διαγώνιο. Τότε η εξίσωση 3.53 ισοδυναμεί με την εξίσωση:

$$\lambda H_{kk}\,\delta\theta_k = -g_k \tag{3.54}$$

Από την άλλη αν το λ προσεγγίζει το μηδέν η εξίσωση 3.53 ισοδυναμεί με την εξίσωση 3.48. Δεδομένης κάποιας αρχικής εκτίμησης των παραμέτρων θ τα βήματα που ακολουθούνται σύμφωνα με τη μέθοδο LM είναι:

- Υπολογισμός του σφάλματος $f(\theta) = \sum_i \rho(r_i, \sigma; \theta)$.
- Επιλογή της αρχικής τιμής του λ (λ=0.001).
- (†) Επίλυση του γραμμικού συστήματος των εξισώσεων 3.53 ως προς $\delta\theta$. Κατόπιν ακολουθεί η εκτίμηση της συνάρτησης του σφάλματος με τις νέες παραμέτρους $f(\theta + \delta\theta) = \sum_i \rho(r_i, \sigma; \theta + \delta\theta)$.

- Αν το $f(\theta + \delta\theta) \geq f(\theta)$ τότε το λ αυξάνεται κατά παράγοντα 10. Κατόπιν επιστρέφουμε στο βήμα (†).

- Αν το $f(\theta + \delta\theta) < f(\theta)$, τότε το λ μειώνεται κατά παράγοντα 10 και ενημερώνεται η τιμή των παραμέτρων $\theta \leftarrow \theta + \delta\theta$. Κατόπιν επιστρέφουμε στο βήμα (†).

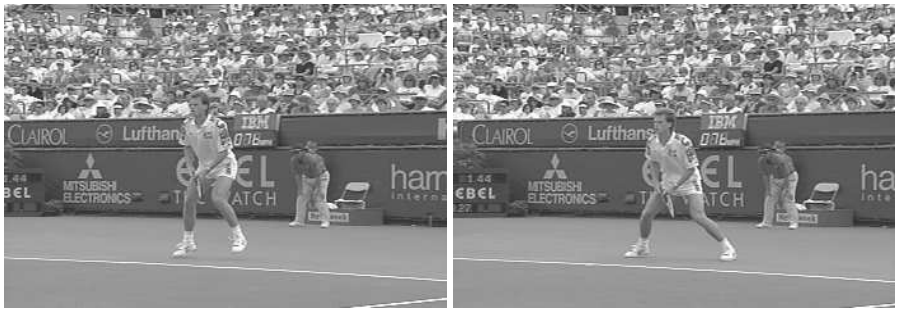
Η επανάληψη συνεχίζεται έως ότου η διαφορά του σφάλματος από επανάληψη σε επανάληψη είναι μικρότερη από καθορισμένο όριο της τάξης του 0.001.

3.6.6 Πειραματικά αποτελέσματα

Ο αλγόριθμος M-estimator για την εκτίμηση της κίνησης δοκιμάστηκε σε αρκετές ακολουθίες εικόνων. Οι ακολουθίες αυτές περιείχαν είτε πραγματικά είτε συνθετικά δεδομένα. Κατά τη διάρκεια των ακολουθιών αυτών η κάμερα παρουσίαζε μια ποικιλία τρισδιάστατων κινήσεων. Για κάθε ζευγάρι εικόνων υπολογίστηκε η οπτική ροή των σημείων της εικόνας. Οι μέθοδοι εκτίμησης της οπτικής ροής είναι από τη μια το ταίριασμα περιοχών της εικόνας και από την άλλη η αντιστοίχιση γωνιών. Με την εφαρμογή κατάλληλων κριτηρίων απορρίφθηκαν σημεία ή περιοχές που ανήκαν σε outliers. Η υπολογιζόμενη οπτική ροή χρησιμοποιήθηκε για την εκτίμηση των παραμέτρων ενός μοντέλου κίνησης, που έχει προεπιλεγεί, ώστε να περιγράφει την κίνηση της κάμερας στην σκηνή.

Ο αλγόριθμος εφαρμόστηκε στην ακολουθία του Stefan. Στο σχήμα 3.13 απεικονίζονται δύο στάσεις της ακολουθίας. Η κίνηση της κάμερας στις στάσεις αυτές είναι κυρίως μεταφορική προς την οριζόντια κατεύθυνση. Εμφανίζεται όμως και μικρή περιστροφή ως προς τον κατακόρυφο άξονα και αλλαγή στην απόσταση εστίασης της κάμερας. Η κίνηση αυτή περιγράφηκε από το γνωστό αφινικό μοντέλο κίνησης. Η οπτική ροή εκτιμήθηκε με τη μέθοδο ταιριάσματος περιοχών. Υποθέσαμε ότι η μέγιστη απομάκρυνση της κάμερας στις στάσεις αυτές είναι 8 pixel. Πρέπει να σημειωθεί ότι ο αλγόριθμος εκτίμησης της οπτικής ροής υπολογίζει αυτόματα την απομάκρυνση αυτή. Ξεκινά με μια ελάχιστη τιμή απομάκρυνσης που δίνεται στην είσοδο. Αν ο αριθμός των μπλοκς που απομένουν με τη χρήση του αλγορίθμου M-estimator είναι μικρός τότε διπλασιάζεται η μέγιστη απομάκρυνση και υπολογίζονται εκ νέου οι παράμετροι του μοντέλου κίνησης. Ο αλγόριθμος συνεχίζει μέχρι να απομείνει ένας σημαντικός αριθμός μπλοκς ικανός στο να υπολογίσει την κίνηση της κάμερας στο σκηνικό. Η μέγιστη τιμή του ορίου, που χρησιμοποιήθηκε στο κριτήριο αντιστοίχισης μπλοκς, το οποίο ορίζεται ως η ελάχιστη διαφορά φωτεινοτήτων δύο υποψήφιων μπλοκς, ήταν 25. Στον

αλγόριθμο M-estimator χρησιμοποιήθηκε η μέθοδος των ελαχίστων τετραγώνων που περιγράφεται ακριβώς από την εξίσωση 3.41. Σε κάθε επανάληψη του αλγορίθμου απορρίπτονται μπλοκς που εμφάνιζαν μεγάλη τιμή κριτηρίου της τάξης του 1 pixel. Για τον έλεγχο των παραμέτρων κίνησης που υπολογίστηκαν κατασκευάστηκε το "μωσαϊκό" των δύο στάσεων του σχήματος 3.13. Οι στάσεις ευθυγραμμίστηκαν χρησιμοποιώντας τις παραμέτρους κίνησης που υπολογίστηκαν από τη μέθοδο. Το "μωσαϊκό" που κατασκευάστηκε φαίνεται στο σχήμα 3.14. Το "μωσαϊκό" δεν εμφανίζει προβλήματα κακής



(a) (b)

Σχήμα 3.13: Δύο στάσεις της ακολουθίας του Stefan (Stefan070-Stefan075).

Στις στάσεις αυτές η κάμερα εκτελεί μεταφορική κίνηση, μικρή περιστροφή γύρω από τον κατακόρυφο άξονα και επιπλέον αλλάζει η εστιακή της απόσταση

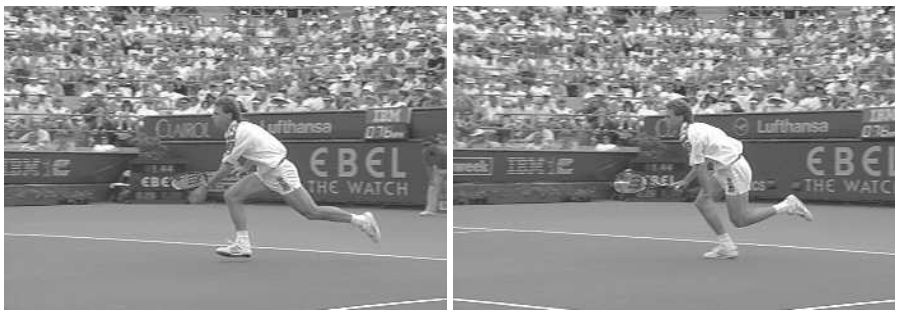
Σχήμα 3.14: Το "μωσαϊκό" που κατασκευάστηκε από δύο στάσεις (Stefan070-Stefan075) της ακολουθίας του Stefan

Χρησιμοποιήθηκε ο αλγόριθμος M-estimator για την εκτίμηση της κίνησης της κάμερας. Η κίνηση αυτή περιγράφηκε από το αφινικό μοντέλο

ευθυγράμμισης των στάσεων, επομένως η κίνηση της κάμερας υπολογίστηκε σωστά.

Ο αλγόριθμος M-estimator εφαρμόστηκε επίσης στις στάσεις του σχήματος 3.15. Στις στάσεις αυτές η κάμερα εκτελεί μεταφορική κίνηση και περιστρέφεται γύρω από τον

κατακόρυφο άξονα. Χρησιμοποιήθηκε το προβολικό μοντέλο για την περιγραφή αυτής της κίνησης. Η εκτίμηση της οπτικής ροής έγινε με χρήση της μεθόδου αντιστοίχισης γωνιών. Το μέγεθος του παραθύρου αναζήτησης για την αντιστοίχιση γωνιών ήταν 80x80. Ενώ το παράθυρο συσχέτισης είχε μέγεθος 7x7. Το κάτω όριο στον αλγόριθμο εξαγωγής γωνιών ήταν 3000. Στον αλγόριθμο M-estimator χρησιμοποιήθηκε η συνάρτηση κόστους Geman & Mc-Clure. Το κριτήριο ελαχιστοποίησης που εφαρμόστηκε ήταν η ελάχιστη γεωμετρική διαφορά των αντίστοιχων γωνιών (3.37). Σε κάθε βήμα της επανάληψης υπολογιζόταν η παράμετρος κλίμακας σ. Η επιλογή της της παραμέτρου, ώστε να εκφράζει την ενδιάμεση τιμή των υπολοίπων, υπήρξε καθοριστική για την απόδοση του αλγορίθμου. Το "μωσαϊκό", που κατασκευάστηκε ευθυγραμμίζοντας τις στάσεις βάσει τις παραμέτρους κίνησης που εκτιμήθηκαν από τον αλγορίθμο, φαίνεται στο σχήμα 3.16. Προβλήματα κακής ευθυγράμμισης των στάσεων δεν είναι ορατά, επομένως υπήρξε ορθή εκτίμηση της περιστροφικής κίνησης της κάμερας.



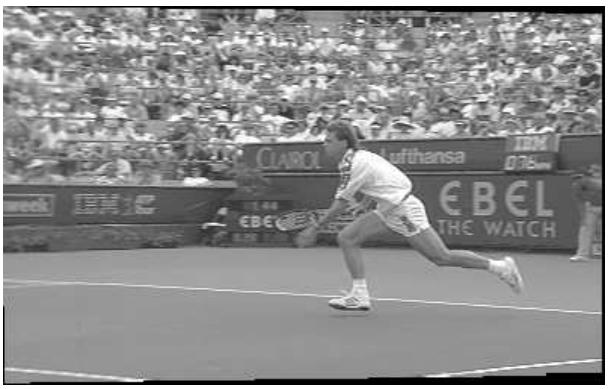
(a) (b)

Σχήμα 3.15: Δύο στάσεις της ακολουθίας του Stefan(Stefan240-Stefan242)

Η κίνηση της κάμερας στις στάσεις αυτές είναι μεταφορική και περιστροφική γύρω από τον κατακόρυφο άξονα

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 μεγάλο ποσοστό σφαλμάτων, τότε ο αλγόριθμος προσδιορισμού της κίνησης είναι πολύ πιθανό να αποτύχει.

Επιπλέον παρατηρήθηκε ότι η επιλογή της συνάρτησης κόστους Geman & Mc-Clure ήταν καθοριστική για την ορθή εκτίμηση της κίνησης της κάμερας. Παράλληλα η εφαρμογή του κριτηρίου της γεωμετρικής αντιστοίχισης χαρακτηριστικών σημείων (3.37) αποδείχτηκε κατάλληλο υπό την προϋπόθεση ορθής εκτίμησης της οπτικής ροής.



Σχήμα 3.16: Το "μωσαϊκό" που κατασκευάστηκε από δύο στάσεις της ακολουθίας Stefan (Stefan240-Stefan242).

Για την εκτίμηση της κίνησης χρησιμοποιήθηκε ο αλγόριθμος M-estimator. Το μοντέλο, που περιέγραψε την κίνηση της κάμερας στις στάσεις, ήταν το προβολικό.

3.7 Χρήση αναλλοίωτων αφινικών ροπών για την εκτίμηση της κίνησης

Στην ενότητα αυτή θα επικεντρωθούμε στη χρήση της έννοιας των ροπών [44] για την εκτίμηση των παραμέτρων ενός μοντέλου κίνησης. Οι ροπές είναι γενικά χαρακτηριστικά που περιγράφουν αποτελεσματικά το σχήμα ενός αντικειμένου ή μιας επιφάνειας. Γενικά η ροπή (p+q) τάξης μιας δισδιάστατης συνάρτησης f(x, y) ορίζεται από τη σχέση:

$$m_{pq} = \int \int_{\mathcal{R}} x^p y^q f\,(x, y)\; dxdy \tag{3.55}$$

Η ροπή υπολογίζεται μόνο πάνω σε σημεία που ανήκουν σε ένα αντικείμενο ή μια επιφάνεια. Επομένως, η f(x, y) = 1 για σημεία που ανήκουν στο αντικείμενο ή στην επιφάνεια και f(x, y) = 0 για σημεία που δεν ανήκουν.

Για την εφαρμογή των ροπών στη μέθοδο εκτίμησης του μετασχηματισμού μεταξύ δύο στάσεων, αναγκαία είναι η αντιστοίχιση των επιφανειών αυτών των στάσεων. Μια επιφάνεια σε μια στάση ορίζεται από ένα σύνολο σημείων και η αντίστοιχή της στην επόμενη από το ίδιο σύνολο των αντίστοιχων σημείων της πρώτης. Η μέθοδος αυτή λαμβάνοντας υπόψη την αντιστοιχία των επιφανειών στις δύο στάσεις, υπολογίζει την κίνηση της κάμερας με τη χρήση των ροπών που εκτιμούνται πάνω στις επιφάνειες δυο στάσεων. Όπως είναι φανερό για την ορθή εκτίμηση του μετασχηματισμού των δύο στάσεων απαιτείται σωστή αντιστοίχιση των επιφανειών και κατ' επέκταση σωστή αντιστοίχιση των σημείων από τα οποία σχηματίζονται. Για την εύρεση των αντίστοιχων

σημείων χρησιμοποιήθηκαν οι μέθοδοι που παρουσιάστηκαν για την εκτίμηση της οπτικής ροής.

Αναλυτικότερα μετά τον αλγόριθμο ταιριάσματος περιοχών (block matching) ή αντιστοίχισης γωνιών και των κριτηρίων που αναφέραμε, απομένει ένας σημαντικός αριθμός μπλοκς ή γωνιών αντίστοιχα. Λόγω όμως της αυστηρά ορθής αντιστοίχισης σημείων που απαιτούμε, ο αριθμός των σημείων (γωνίες ή τα κέντρα των μπλοκς) που απομένει υφίσταται επιπλέον εξέταση της ορθότητας των αντιστοιχίσεων του. Η ορθότητα αυτή ελέγχεται με χρήση της τεχνικής των M-estimators. Η τεχνική αυτή χρησιμοποιείται λόγω της ανάγκης εύρεσης αντίστοιχων σημείων και άρα επιφανειών που να συμφωνούν ως προς την κίνηση της κάμερας. Υποθέτουμε ότι η αντιστοιχία των επιφάνειων στις δύο στάσεις στη μέθοδο των ροπών είναι αποτέλεσμα ενός αφινικού μετασχηματισμού. Εφαρμόζοντας την τεχνική των M-estimators για αφινικό μετασχηματισμό υπολογίζονται οι παράμετροι του μετασχηματισμού όπως προκύπτουν από τις δεδομένες αντιστοιχίες των σημείων, απορρίπτοντας σημεία που ανήκουν σε περιοχές με διαφορετική κίνηση ως προς τη συνολική κίνηση της κάμερας.

Από το σύνολο των αντίστοιχων σημείων επιλέγονται ορισμένα, με τη χρήση κριτηρίων που εξαρτώνται από την τοπολογική θέση των σημείων αυτών στις στάσεις. Κατόπιν από το σύνολο των επιλεγμένων σημείων κατασκευάζεται το περίβλημα των σημείων της πρώτης στάσης. Το περίβλημα της δεύτερης στάσης κατασκευάζεται από τα αντίστοιχα των επιλεγμένων σημείων της πρώτης. Η κατασκευή του περιβλήματος καθώς και τα κριτήρια για την επιλογή των σημείων θα παρουσιασθούν αναλυτικότερα σε παρακάτω ενότητα.

Τα περιβλήματα που προκύπτουν περιγράφονται στη συνέχεια με χρήση των ροπών. Οι σχέσεις ορισμένων ροπών αποτελούν αναλλοίωτα χαρακτηριστικά (invariants) για τον αφινικό μετασχηματισμό. Τα αναλλοίωτα χαρακτηριστικά είναι γενικά ιδιότητες γεωμετρικών σχηματισμών οι οποίες μένουν αμετάβλητες κάτω από συγκεκριμένους μετασχηματισμούς. Σύμφωνα με τον Hu[1962] υπάρχει ένα σύνολο σχέσεων μεταξύ καθορισμένων ροπών που είναι αναλλοίωτες στη μεταφορική, στην περιστροφική κίνηση και στην αλλαγή του μήκους εστίασης της κάμερας. Η πιο διαδεδομένη μέθοδος για τον υπολογισμό των αναλλοίωτων σχέσεων χαρακτηριστικών και συγκεκριμένα αναλλοίωτων σχέσεων ροπών είναι η μέθοδος της **κανονικοποίησης** ([45]). Το κύριο πλεονέκτημα της κανονικοποίησης είναι η εύρεση της **κανονικής θέσης** (standard position) για κάθε δισδιάστατο αντικείμενο η οποία είναι σταθερή για όλους τους μετασχηματισμούς του αντικειμένου. Σύμφωνα με τη μέθοδο αυτή υπολογίζονται οι παράμετροι του αφινικού

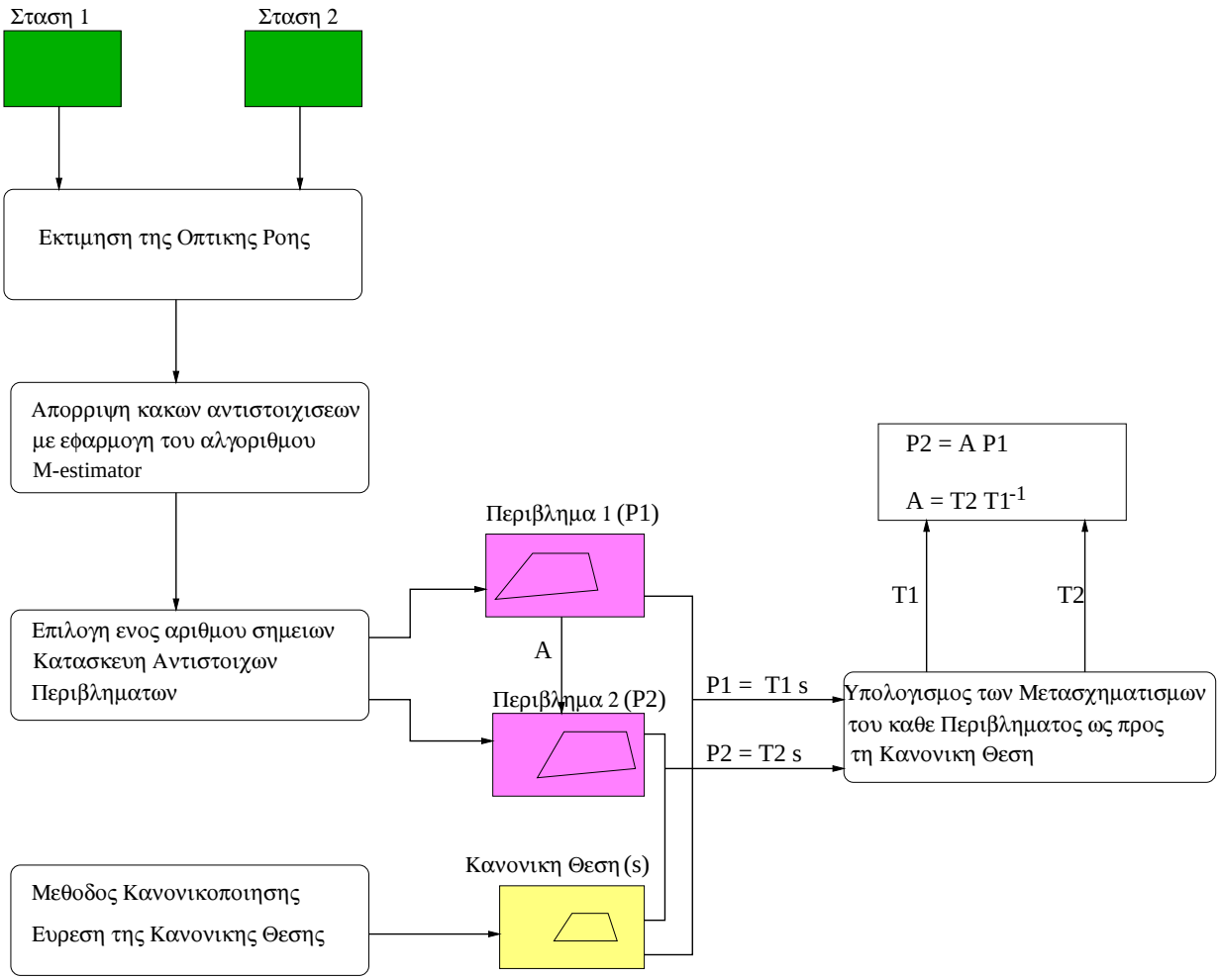
μετασχηματισμού κάθε επιφάνειας ή αντικειμένου με την κανονική της θέση. Για την εύρεση των παραμέτρων του μετασχηματισμού μεταξύ των επιφανειών συνδυάζονται οι δύο μετασχηματισμοί των επιφανειών ως προς την κανονική τους θέση. Επίσης λόγω των μεγάλων τάξεων ροπών που χρησιμοποιούνται για την εύρεση των παραμέτρων κάθε επιφάνειας με την κανονική θέση, απαιτείται η λύση ενός μη-γραμμικού συστήματος. Για την αποφυγή της χρονοβόρου διαδικασίας της λύσης ενός τέτοιου συστήματος καταφεύγουμε στο χωρισμό κάθε αφινικού μετασχηματισμού σε επιμέρους διαδοχικούς μετασχηματισμούς. Υπολογίζοντας έτσι αναλλοίωτες σχέσεις ροπών σε κάθε στάδιο του συνολικού μετασχηματισμού εκτιμούνται οι επιμέρους παράμετροι των διαδοχικών μετασχηματισμών και τελικά ο συνολικός μετασχηματισμός μεταξύ των δύο επιφανειών ([46]). Ο συνολικός αλγόριθμος εκτίμησης των παραμέτρων ενός μετασχηματισμού με τη χρήση αναλλοίωτων σχέσεων ροπών φαίνεται στο σχήμα 3.17.

3.7.1 Κατασκευή περιβλημάτων από ένα σύνολο αντίστοιχων σημείων

Στην ενότητα αυτή θα περιγραφεί η κατασκευή των δυο περιβλημάτων που ορίζουν τις επιφάνειες των δύο στάσεων. Όπως αναφέρθηκε προηγούμενα για την κατασκευή των περιβλημάτων είναι ήδη γνωστές οι αντιστοιχίες σημείων της πρώτης στάσης σε σημεία της δεύτερης. Επομένως απαραίτητη είναι μόνο η επιλογή των σημείων της πρώτης στάσης για την κατασκευή του περιβλήματος αφού το περίβλημα της δεύτερης στάσης προκύπτει απλά από τα αντίστοιχα των επιλεγμένων σημείων της πρώτης. Συχνά ο συνολικός αριθμός των σημείων σε κάθε στάση είναι πολύ μεγάλος. Το γεγονός αυτό αυξάνει το κόστος κατασκευής του περιβλήματος. Απαραίτητη είναι επομένως η μείωση του αριθμού των σημείων για κάθε στάση. Τα κριτήρια επιλογής των σημείων και η κατασκευή του περιβλήματος παρουσιάζονται στις επόμενες ενότητες.

3.7.1.1 Επιλογή σημείων για την κατασκευή του πολυγώνου

Η επιλογή των σημείων γίνεται με τρόπο ώστε η επιφάνεια του περιβλήματος, που θα προκύψει, να είναι αρκετά μεγάλη ώστε να καλύπτει ένα μεγάλο ποσοστό της σκηνής. Η μέθοδος τότε των ροπών θα υπολογίσει τη συνολική κίνηση της κάμερας στη σκηνή. Ο αλγόριθμος που χρησιμοποιείται για το σκοπό αυτό επιλέγει τα πιο απομακρυσμένα σημεία. Όταν επιλεγούν δύο σημεία του περιβλήματος τότε λαμβάνει χώρα η διαδικασία της απόρριψης των γειτονικών τους, ώστε τα τελικά σημεία που θα επιλεγούν να είναι τα πιο απομακρυσμένα και διασκορπισμένα σημεία του αρχικού συνόλου σημείων. Ο αλγόριθμος συγκλίνει σε έναν αριθμό Κ σημείων. Ο αριθμός αυτός πρέπει να είναι



Σχήμα 3.17: Ο συνολικός αλγόριθμος εκτίμησης των παραμέτρων κίνησης με την εφαρμογή της μεθόδου των αναλλοίωτων σχέσεων αφινικών ροπών

μεγαλύτερος του δύο (K > 2) για την ύπαρξη κλειστής επιφάνειας. Συνήθως όμως απαιτείται να είναι μεγαλύτερος του πέντε. Όταν ο αριθμός των αντίστοιχων σημείων είναι μεγάλος επιλέγονται περίπου τα 20 πιο απομακρυσμένα σημεία. Τα βήματα του αλγορίθμου που ακολουθήθηκαν περιγράφονται αναλυτικότερα παρακάτω:

- Αν N είναι ο συνολικός αριθμός σημείων, $\frac{N(N-1)}{2}$ είναι ο αριθμός των δυνατών ζευγαριών. Για κάθε ζεύγος (P1, P2) υπολογίζεται η ευκλείδια απόσταση D των άκρων του.

- (†)Επιλέγεται το ζευγάρι των σημείων (P1, P2) με τη μεγαλύτερη απόσταση από το σύνολο των δυνατών ζευγαριών. Τα σημεία αυτά εισάγονται στη λίστα των επιλεγμένων σημείων L.

- Αφαιρούνται από το σύνολο των αρχικών ζευγών, τα ζεύγη με άκρο κάποιο από τα γειτονικά των P1 και P2 καθώς και το ζεύγος (P1, P2) αφού έχει ήδη επιλεγεί. Ως γειτονιά ορίζεται μια τετραγωνική περιοχή διάστασης 10 pixel με κέντρο πρώτα το σημείο P1 και έπειτα το σημείο P2.

- Αν ο αριθμός των επιλεγμένων σημείων της λίστας L είναι μικρότερος του K τότε επιστρέφουμε στο βήμα (†) αλλιώς ο αλγόριθμος συγκλίνει.

Η επιλογή των σημείων έγινε με σκοπό τη μείωση του συνολικού αριθμού σημείων που συμμετέχουν στην κατασκευή του περιβλήματος. Μια από τις πιο προφανής λύσεις θα ήταν η κατασκευή του κυρτού περιβλήματος που προκύπτει από το σύνολο αυτών των σημείων. Η ύπαρξη όμως της πιθανότητας το σχήμα του κυρτού περιβλήματος να είναι ομοιόμορφο έχει σαν αποτέλεσμα την εκτίμηση της κίνηση της κάμερας με μικρή ακρίβεια. Η χρήση όμως του μη-κυρτού περιβλήματος οδηγεί σε ακριβέστερα αποτελέσματα της εκτίμησης της κίνησης, αφού το περίπλοκο σχήμα περιέχει μεγαλύτερη πληροφορία για την κίνηση στη σκηνή.

3.7.1.2 Η κατασκευή του μη-κυρτού περιβλήματος

Η κατασκευή του μη-κυρτού περιβλήματος που προκύπτει από τα επιλεγμένα σημεία συμπίπτει με τη λύση του γνωστού ευκλείδιου προβλήματος του *Πλανόδιου Πωλητή* (TSP : Traveling Salesman Problem). Το πρόβλημα αυτό έγκειται στην επιλογή της μικρότερης διαδρομής που θα ακολουθήσει ένας πλανόδιος για να επισκεφθεί N πόλεις. Στην περίπτωση μας για την κατασκευή του μη-κυρτού περιβλήματος, απαιτείται να βρεθεί η ελάχιστη διαδρομή για να επισκεφτούμε όλα τα σημεία με την προϋπόθεση να μην υπάρχουν τομές ακμών που ενώνουν δύο σημεία. Το πρόβλημα αυτό (TSP) λέγεται ευκλείδιο TSP αφού οι θέσεις των σημείων που θα επισκεφτούμε βρίσκονται όλες στο ίδιο επίπεδο και το κόστος του ταξιδιού μεταξύ ενός ζευγαριού θέσεων υπολογίζεται βάση της ευκλείδιας απόστασης. Για τη λύση του προβλήματος δεν υπάρχει ταχύς ακριβής αλγόριθμος. Η ύπαρξη ορισμένων ευριστικών αλγορίθμων οδηγούν σε κάποια λύση του. Ένας από τους ευριστικούς αλγορίθμους που ακολουθήθηκε ανήκει στην κατηγορία των διεργασιών κατασκευής της διαδρομής (tour construction procedures). Στην κατηγορία αυτή των διεργασιών για την εκτέλεση του αλγορίθμου απαιτείται να επιλεγούν:

- μια αρχική διαδρομή (αρχικοποίηση)

- το κριτήριο επιλογής μιας ακμής σαν ακμή της συνολικής διαδρομής

- το κριτήριο που καθορίζει τη θέση εισαγωγής της επιλεγμένης ακμής στις ακμές της συνολικής διαδρομής.

Ο ευριστικός αυτός αλγόριθμος ονομάζεται *διαδικασία εισαγωγής κυρτού περιβλήματος (Convex Hull Insertion Procedure)* και προτάθηκε από το Stewart το 1977. Τα βήματα του αλγορίθμου είναι:

- *Βήμα 1.* Κατασκευή του κυρτού περιβλήματος του συνόλου των σημείων. Το κυρτό περίβλημα παρέχει μια αρχική εκτίμηση της διαδρομής.

- *Βήμα 2.* (Εισαγωγή). Για κάθε σημείο k που δεν ανήκει στην τρέχουσα διαδρομή, αποφασίζεται μεταξύ ποιας ακμής (i, j) της διαδρομής θα εισαχθεί το σημείο k. Για τη λήψη αυτής της απόφασης αναζητείται η ακμή i, j της διαδρομής στην οποία το κριτήριο c_{ik} + $c_{kj} - c_{ij}$ αποκτά την ελάχιστη τιμή. Η μεταβλητή c_{kl} εκφράζει την ευκλείδια απόσταση του σημείου k από το σημείο l.

- *Βήμα 3.*(Επιλογή). Για όλα τα (i, k, j) που βρέθηκαν στο βήμα 2, επιλέγεται το (i^*, k^*, j^*) ώστε το $\frac{c_{i^*k^*}+c_{k^*j^*}}{c_{i^*j^*}}$ να είναι το ελάχιστο.

- *Βήμα 4.* Εισάγεται το σημείο k^* στη διαδρομή μεταξύ των σημείων i^* και j^*.

- *Βήμα 5.* Επαναλαμβάνουμε τα βήματα 2 ως 4 μέχρι την κατασκευή ενός Hamiltonian κύκλου (ο κύκλος που περνά από όλα τα σημεία της διαδρομής).

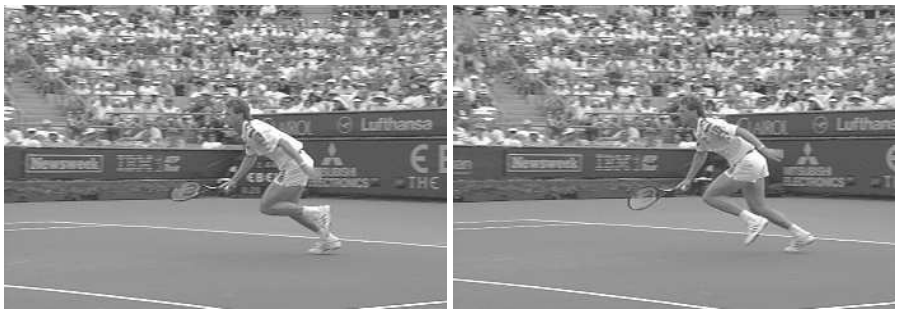
Για την κατασκευή του κυρτού περιβλήματος χρησιμοποιήθηκε ο αλγόριθμος του Graham-Scan. Ο αλγόριθμος αυτός περιγράφεται αναλυτικά στο Παράρτημα Β.

3.7.2 Πειραματικά αποτελέσματα

Ο αλγόριθμος εκτίμησης της κίνησης με τη χρήση αναλλοίωτων σχέσεων αφινικών ροπών προϋποθέτει την εξαγωγή του περιβλήματος των επιφανειών των δύο εικόνων, που θα συγκριθούν για τον υπολογισμό της κίνησης της κάμερας. Για την δημιουργία των περιγραμμάτων επελέγησαν σημεία με τη χρήση του αλγορίθμου M-estimator.

Η μέθοδος των ροπών εφαρμόστηκε σε ένα σημαντικό αριθμό ακολουθιών και τα αποτελέσματα εκτίμησης της κίνησης ήταν ικανοποιητικά, με την προϋπόθεση ότι η αντιστοιχία των επιφανειών είναι ορθή. Μια από τις δυσκολότερες ακολουθίες που εφαρμόστηκε ήταν η ακολουθία του Stefan. Δύο από τις στάσεις της ακολουθίας απεικονίζονται στο σχήμα 3.18. Η κάμερα στις στάσεις αυτές περιστρέφεται γύρω από τον κατακόρυφο άξονα και εκτελεί και μεταφορική κίνηση προς την οριζόντια κατεύθυνση.

Η οπτική ροή εκτιμήθηκε με τη χρήση της μεθόδου ταιριάσματος περιοχών. Η αλγόριθμος M-estimator χρησιμοποιεί την υπολογισμένη οπτική ροή και επιλέγει έναν αριθμό σημείων για τη δημιουργία των αντίστοιχων περιβλημάτων στις δύο εικόνες. Τα περιβλήματα που που κατασκευάστηκαν απεικονίζονται στο σχήμα 3.19. Η εκτίμηση των παραμέτρων του αφινικού μετασχηματισμού πραγματοποιήθηκε υπολογίζοντας διάφορες σχέσεις των ροπών αυτών των περιβλημάτων. Το "μωσαϊκό", που κατασκευάστηκε χρησιμοποιώντας τις παραμέτρους κίνησης που υπολογίστηκαν από τη μέθοδο, φαίνεται στο σχήμα 3.20. Παρόλη την πολυπλοκότητα της κίνησης η μέθοδος υπολόγισε σωστά τις παραμέτρους του αφινικού μετασχηματισμού. Γεγονός που συμπεραίνεται από την κατασκευή ενός ικανοποιητικού "μωσαϊκού".



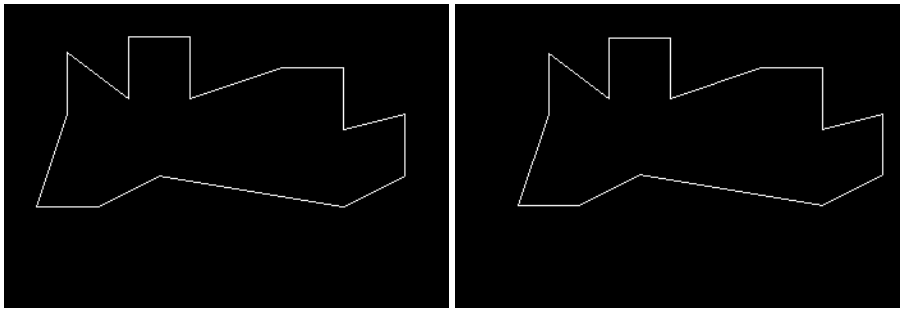
(a) (b)

Σχήμα 3.18: Δύο στάσεις της ακολουθίας του Stefan (Stefan244-Stefan245)

Στις στάσεις αυτές η κάμερα περιστρέφεται γύρω από τον κατακόρυφο άξονα και εκτελεί και μεταφορική κίνηση προς την οριζόντια κατεύθυνση

Η μέθοδος των ροπών εφαρμόστηκε επίσης σε ακολουθία με συνθετικά δεδομένα. Δύο στάσεις της ακολουθίας φαίνονται στο σχήμα 3.21. Η κίνηση της κάμερας στις στάσεις αυτές είναι κυρίως μεταφορική ως προς την οριζόντια κατεύθυνση. Τα περιβλήματα (3.22) κατασκευάστηκαν από το σύνολο των αντίστοιχων σημείων που προήλθαν από τη χρήση του αλγορίθμου M-estimator. Η εκτίμηση της οπτικής ροής πραγματοποιήθηκε με χρήση της μεθόδου αντιστοίχισης γωνιών. Το "μωσαϊκό", που κατασκευάστηκε με τη χρήση των παραμέτρων κίνησης που υπολογίστηκαν από τη μέθοδο, απεικονίζεται στο σχήμα 3.23. Η μέθοδος των ροπών ανίχνευσε σωστά την κίνηση της κάμερας αφού το "μωσαϊκό" που προέκυψε παρουσιάζεται ικανοποιητικό.

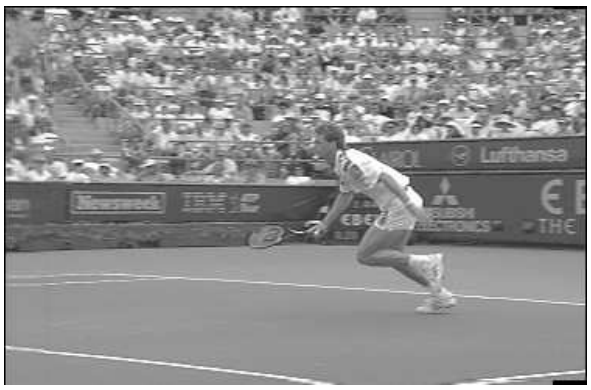
Ένα από τα συμπεράσματα που καταλήξαμε ήταν ότι η μέθοδος των ροπών είναι αποδοτική στις περιπτώσεις μόνο που υπάρχει σωστή αντιστοίχιση των περιβλημάτων των δύο εικόνων. Παράλληλα, η χρήση μόνο της γεωμετρικής πληροφορίας αυξάνει



(a) (b)

Σχήμα 3.19: Τα αντίστοιχα περιβλήματα των δύο στάσεων της ακολουθίας Stefan (Stefan244 και Stefan245)

Τα περιβλήματα αυτά θα χρησιμοποιηθούν για την εκτίμηση της κίνησης της κάμερας με την μέθοδο των ροπών



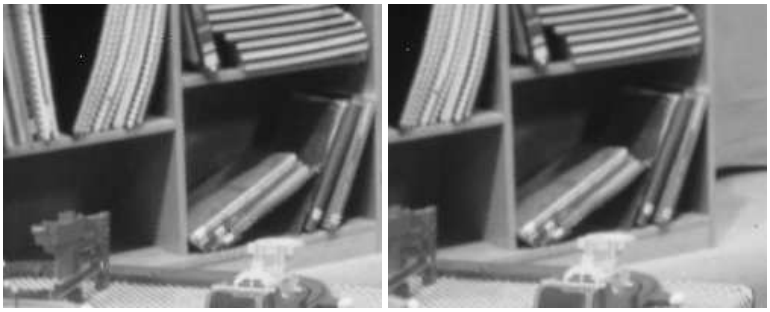
Σχήμα 3.20: Το ”μωσαϊκό” δύο στάσεων της ακολουθίας του Stefan (Stefan244-Stefan245)

Το ”μωσαϊκό” κατασκευάστηκε με τη χρήση της μεθόδου των αναλλοίωτων σχέσεων αφινικών ροπών

την απόδοση της μεθόδου αφού τυχόν αλλαγές στο φωτισμό ή η ύπαρξη ομοιόμορφων περιοχών στη σκηνή δεν επηρεάζει τα τελικά αποτελέσματα της.

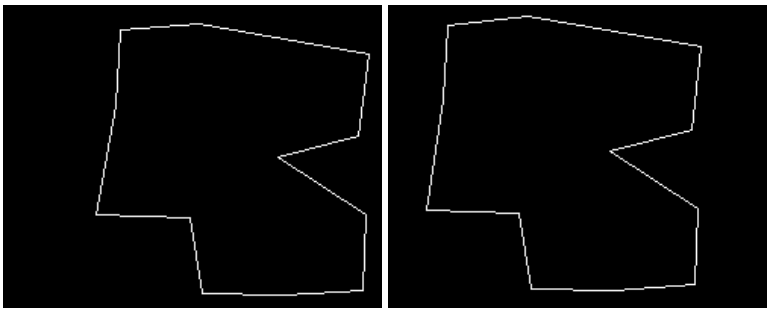
3.8 Βελτιστοποίηση της εκτίμησης των παραμέτρων του μοντέλου κίνησης

Η εκτίμηση της κίνησης που πραγματοποιήθηκε με τις προηγούμενες μεθόδους είναι ορθή. Λόγω όμως των μειονεκτημάτων που παρουσιάζει από τη μια ο αλγόριθμος M-estimator και από την άλλη η μη-ακριβής αντιστοίχιση των σημείων για τον προσδιορισμό της επιφάνειας στην οποία θα εφαρμοστούν οι ροπές, είναι αναγκαία κάποιες φορές η διόρθωση των εκτιμημένων παραμέτρων του μοντέλου κίνησης. Οι μέθοδοι διόρθωσης



(a) (b)

Σχήμα 3.21: Δύο στάσεις της ακολουθίας Barbara1



(a) (b)

Σχήμα 3.22: Τα περιβλήματα δύο στάσεων της ακολουθίας Barbara1

Τα περιβλήματα αυτά θα χρησιμοποιηθούν για την εκτίμηση των παραμέτρων του αφινικού μετασχηματισμού με τη χρήση της μεθόδου των ροπών.

που υλοποιήθηκαν στην εργασία μας βασίζονται σε παραλλαγές της γνωστής μεθόδου της *Απότομης Κλίσης (Steepest descent)*. Οι αλγόριθμοι διόρθωσης που εφαρμόστηκαν για τη βελτιστοποίηση των παραμέτρων του αφινικού μοντέλου κίνησης. Επίσης η περιοχή $\mathcal{R}$ στην οποία εφαρμόστηκαν τα κριτήρια προς ελαχιστοποίηση δεν ήταν ολόκληρη η εικόνα αλλά το σύνολο των σημείων της εικόνας που πήραν μέρος στην εκτίμηση των τελικών παραμέτρων του εκάστοτε μοντέλου.

3.8.1 Μέθοδος της Απότομης Κλίσης

Σκοπός της μεθόδου είναι η ελαχιστοποίηση ενός κριτηρίου και ο υπολογισμός των παραμέτρων από τις οποίες αυτό εξαρτάται [32]. Στη συγκεκριμένη περίπτωση το κριτήριο προς ελαχιστοποίηση στηρίζεται στην αρχή διατήρησης των δεδομένων στο

Σχήμα 3.23: Το ”μωσαϊκό” δύο στάσεων της ακολουθίας Barbara1 που κατασκευάστηκε με χρήση της μεθόδου των ροπών

χρόνο και προκύπτει απλά, χρησιμοποιώντας τη μέθοδο των ελαχίστων τετραγώνων σε μια περιοχή $\mathcal{R}$, δηλαδή:

$$\mathrm{E}_D(u(a)\,, v(a)\,) = \sum_{(x,y)\epsilon\mathcal{R}} [I(x,y;t) - I(x + u(a)\,\delta t, y + v(a)\,\delta t, t + \delta t)\,]^2 \qquad (3.56)$$

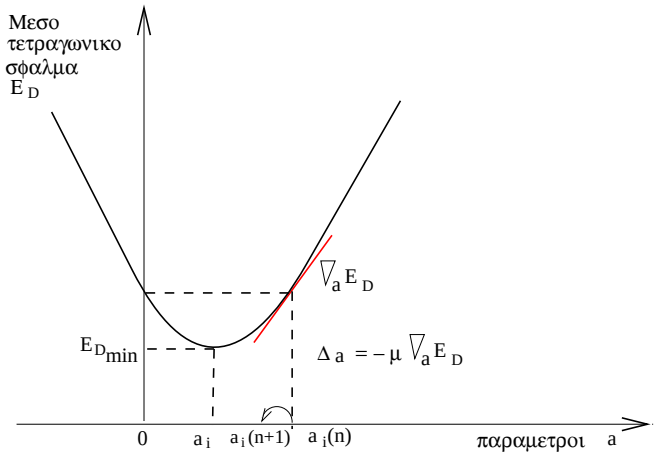
όπου a είναι οι παράμετροι του μοντέλου κίνησης που προσδιορίζουν το διάνυσμα της ταχύτητας.

Σύμφωνα με τη μέθοδο αυτή, οι παράμετροι a_i μεταβάλλονται στο χρόνο και οι τιμές τους ρυθμίζονται με μια επαναληπτική διαδικασία, κατά μήκος της επιφάνειας του σφάλματος με σκοπό την προοδευτική μετακίνηση τους προς τη βέλτιστη λύση. Η επιφάνεια του σφάλματος περιγράφεται από την εξίσωση 3.56 που εκφράζει το κριτήριο ελαχιστοποίησης. Η μέθοδος της απότομης κλίσης αναζητά συνεχώς το ελάχιστο σημείο της επιφάνειας του σφάλματος. Οι διαδοχικές ρυθμίσεις που εφαρμόζονται στις παραμέτρους είναι προς την κατεύθυνση της απότομης κλίσης δηλαδή προς την αντίθετη κατεύθυνση του διανύσματος των χωρικών παραγώγων που ορίζεται από το $\bigtriangledown_{a_i} E_D$. Μια τέτοια ρύθμιση φαίνεται στο σχήμα 3.24.

Αν υποθέσουμε ότι $a_i(n)$ είναι η παράμετρος a_i του μοντέλου στο n βήμα της επανάληψης και $\Delta a_i(n)$ είναι η διόρθωση της παραμέτρου a_i στο ίδιο βήμα, σύμφωνα με τη μέθοδο της απότομης κλίσης η διόρθωση αυτή θα ισούται με:

$$\Delta a_i(n) = -\mu \bigtriangledown_{a_i} E_D(n) \qquad (3.57)$$

Γενικά, αν α είναι το διάνυσμα των παραμέτρων του μοντέλου και $\Delta a(n)$ είναι το διάνυσμα που περιέχει τη ρύθμιση στο βήμα n του συνόλου των παραμέτρων a_i, η



Σχήμα 3.24: Μέθοδος της απότομης κλίσης

παραπάνω σχέση μετατρέπεται στην εξής:

$$\Delta a(n) = -\mu \ \nabla_a E_D(n) \tag{3.58}$$

Αναλύοντας τη μεταβλητή που εκφράζει τη διόρθωση $\Delta a(n)$ ως προς τους δύο όρους $a(n)$ και $a(n+1)$ και αντικαθιστώντας $\nabla_a E_D(n)$ με την πρώτη παράγωγο του σφάλματος E_D ως προς a, ισχύει:

$$a(n+1) = a(n) - \mu \sum_{(i,j)\epsilon R} \phi_n(i,j) \ (I(i,j;t) - I(i+u_n,\ j+v_n;\ t+\delta t)\ ) \tag{3.59}$$

όπου $(u_n,\ v_n)$ το διάνυσμα της ταχύτητας για το σημείο (i, j) όπως προκύπτει βάση του διανύσματος α των παραμέτρων στο βήμα n της επανάληψης. $\phi_n(i,j)$ είναι το διάνυσμα που εξαρτάται από τις χωρικές παραγώγους της έντασης. Η γενική μορφή του διανύσματος της κλίσης για το αφινικό μοντέλο είναι:

$$\phi(i,j) = \begin{bmatrix} I_x(i+u, j+v, t) \\ iI_x(i+u, j+v, t) \\ jI_x(i+u, j+v, t) \\ I_y(i+u, j+v, t) \\ iI_y(i+u, j+v, t) \\ jI_y(i+u, j+v, t) \end{bmatrix} \tag{3.60}$$

Η παράμετρος μ εκφράζει το βήμα της διόρθωσης και πρέπει να ικανοποιεί την ανισότητα $\mu < \frac{2}{\lambda_{max}}$ όπου λ_{max} είναι η μεγαλύτερη ιδιοτιμή του πίνακα $\sum_{(i,j)\epsilon R} \phi(i,j)\ \phi^T(i,j)$.

Μια άλλη προσέγγιση της μεθόδου είναι αυτή του κανονικοποιημένου διανύσματος της κλίσης. Στη διάρκεια των επαναλήψεων η διαφορά των εντάσεων των δύο εικόνων προσεγγίζεται γραμμικά με την εξίσωση:

$$\begin{aligned} &I(i,j;t) \; - I(i+u, j+v; t+\delta t) \; = \\ &I(i,j;t) \; - I(i+u_n, j+v_n; t+\delta t) \; + (a-a_n)^T \, \phi_n(i,j) \end{aligned} \tag{3.61}$$

Κάτω από αυτές τις συνθήκες, η ελαχιστοποίηση του παραπάνω κριτηρίου δίνει την εξής διόρθωση των παραμέτρων:

$$\begin{aligned} a(n+1) \; = a(n) \; - \; &\left(\textstyle\sum_{(i,j)\epsilon R} \phi_n(i,j) \; \phi_n^T(i,j) \right)^{-1} \\ &\textstyle\sum_{(i,j)\epsilon R} \phi_n(i,j) \; (I(i,j;t) \; - I(i+u_n, j+v_n; t+\delta t) \;) \end{aligned} \tag{3.62}$$

Ο Dugelay και ο Sanson πρότειναν για την αποφυγή της εύρεσης του αντιστρόφου πίνακα, μια απλοποιημένη μορφή της εξίσωσης 3.62, αντικαθιστώντας τον πίνακα $\sum_{(i,j)\epsilon R} \phi_n(i,j) \; \phi^T(i,j)$ με το μοναδιαίο πίνακα πολλαπλασιασμένο με το ίχνος του πίνακα της κλίσης. Η διόρθωση των παραμέτρων a_i του μοντέλου δίνονται τότε από τον τύπο:

$$a(n+1) \; = a(n) \; - \; \frac{\sum_{(i,j)\epsilon R} \phi_n(i,j) \; (I(i,j;t) \; - I(i+u, j+v; t+\delta t) \;)}{\sum_{(i,j)\epsilon R} \phi_n^T(i,j) \; \phi_n(i,j)} \tag{3.63}$$

Στην παρούσα εργασία χρησιμοποιήθηκαν οι δύο από τις παραπάνω εξισώσεις διόρθωσης 3.59 και 3.63. Στην κατάσταση της αρχικοποίησης το διάνυσμα των παραμέτρων έχει την τιμή που υπολογίστηκε με τη μέθοδο των M-estimators ή με τη χρήση αναλλοίωτων σχέσεων των χαρακτηριστικών μιας εικόνας. Ο αλγόριθμος που χρησιμοποιείται είναι επαναληπτικός. Σε κάθε βήμα της επανάληψης υπολογίζονται οι νέες παράμετροι καθώς και το σφάλμα που ορίζεται από την αρχή διατήρησης των φωτεινοτήτων στο χρόνο. Αν η τιμή του τρέχοντος σφάλματος υπερβεί το σφάλμα που υπολογίστηκε στην προηγούμενη επανάληψη τότε ο αλγόριθμος τερματίζει. Οι τελικές τιμές των παραμέτρων του μοντέλου κίνησης είναι οι τιμές που υπολογίστηκαν στο προηγούμενο βήμα της επανάληψης.

3.8.2 Χρήση χαλαρωτικών μεθόδων

Παραλλαγή των μεθόδων απότομης κλίσης αποτελούν και οι χαλαρωτικές μέθοδοι [29]. Το κριτήριο ελαχιστοποίησης δεν είναι όμως η τετραγωνική διαφορά των εντάσεων των διαδοχικών εικόνων αλλά μια παραλλαγή αυτού. Η παραλλαγή αυτή προκύπτει από τη γνωστή εξίσωση της οπτικής ροής ($I_x(x,y;t) \; u + I_y(x,y;t) \; v + I_t(x,y;t)$ =0 εξίσωση 3.5). Το σφάλμα, που αποτελεί και το νέο κριτήριο προς ελαχιστοποίηση, έχει την εξής

μορφή:

$$\mathrm{E}_D(u,v) = \sum_{\mathcal{R}} \left((\bigtriangledown I)^T \mathbf{u} + I_t \right)^2$$

όπου το $\bigtriangledown I$ είναι το διάνυσμα των χωρο-χρονικών παραγώγων στο σημείο (x,y). Το διάνυσμα της ταχύτητας εξαρτάται από το σύνολο των παραμέτρων a_i του μοντέλου κίνησης. Επίσης για τη βελτίωση της σθεναρότητας του αλγορίθμου στην έκφραση του σφάλματος χρησιμοποιείται κατάλληλη συνάρτηση ρ. Το σφάλμα τότε παίρνει την εξής μορφή:

$$\mathrm{E}_D(u,v) = \sum_{\mathcal{R}} \rho((\bigtriangledown I)^T \mathbf{u}(\alpha) + I_t) \tag{3.64}$$

Η ελαχιστοποίηση του σφάλματος προκύπτει με τη χρήση της μεθόδου της *ταυτόχρονης χαλάρωσης(Simultaneous Over-Relaxation).*

Η μέθοδος της ταυτόχρονης χαλάρωσης ανήκει στην οικογένεια των χαλαρωτικών τεχνικών όπως η μέθοδος Jacobi και η μέθοδος του Gauss-Seidel. Στην περίπτωσή μας η μέθοδος της ταυτόχρονης χαλάρωσης χρησιμοποιείται για την ελαχιστοποίηση του σφάλματος $\mathrm{E}_D(u,v)$ ως προς τις ταχύτητες u, v ή τις παραμέτρους του μοντέλου. Η μέθοδος αυτή είναι επαναληπτική και ενημερώνει σε κάθε βήμα της επανάληψης τις παραμέτρους του μοντέλου.

$$a_{(n+1)\,i} = a_{(n)\,i} - \omega \frac{1}{T_{a_i}} \frac{\theta E_D}{\theta a_i} \tag{3.65}$$

όπου το ω ονομάζεται παράμετρος χαλάρωσης και χρησιμοποιείται για να διορθώσει το διάνυσμα $a_{(n+1)}$ των παραμέτρων στο βήμα (n+1). Ο παράγοντας $\mathrm{T}(a_i)$ αποτελεί το άνω όριο της δεύτερης παραγώγου του σφάλματος,

$$T_{a_i} \geq \frac{\theta^2 E_D}{\theta {a_i}^2}$$

Ενώ η συνάρτηση ρ που χρησιμοποιήθηκε είναι η γνωστή συνάρτηση Geman-McClure. Λεπτομέρειες για τους παράγοντες T_{a_i} και $\frac{\theta E_D}{\theta a_i}$ παρουσιάζονται στο παράρτημα Α.

Κεφάλαιο 4

Κατασκευή του "μωσαϊκού"

4.1 Εισαγωγή

Το ”μωσαϊκό” κατασκευάζεται από το σύνολο των στάσεων της ακολουθίας παρέχοντας την πανοραμική όψη του σκηνικού που έλαβε η κάμερα. Οι τρόποι αναπαράστασης του ποικίλουν. Η επιλογή της αναπαράστασης εξαρτάται από την εφαρμογή που τελικά θα χρησιμοποιηθεί το ”μωσαϊκό”. Το ”μωσαϊκό” μπορεί να αναπαρασταθεί σε επίπεδη, σε κυλινδρική, ή σε σφαιρική επιφάνεια. Η αναπαράσταση του σε επίπεδη επιφάνεια δεν απαιτεί κανένα επιπλέον κόστος απεικόνισης αφού κατά την κατασκευή του το ”μωσαϊκό” προβάλλεται στο επίπεδο. Αντίθετα η αναπαράσταση σε κυλινδρική ή σφαιρική επιφάνεια [14] απαιτεί επιπλέον το μετασχηματισμό των συντεταγμένων του ”μωσαϊκού” σε κυλινδρικές ή σφαιρικές συντεταγμένες αντίστοιχα. Στην παρούσα εργασία ασχοληθήκαμε με την αναπαράσταση του ”μωσαϊκού” σε επίπεδη επιφάνεια. Εξετάζοντας τον τρόπο αυτό αναπαράστασης του προκύπτει ένα σύνολο διαφορετικών προσεγγίσεων, που ορίζει έναν αριθμό διαφορετικών αναπαραστάσεων.

Στο κεφάλαιο αυτό θα αναλύσουμε τα δύο στάδια κατασκευής του ”μωσαϊκού”. Το πρώτο στάδιο κατασκευής καλείται **ευθυγράμμιση των στάσεων**. Η χρήση των μεθόδων εκτίμησης της κίνησης του προηγούμενου κεφαλαίου παρέχει το μετασχηματισμό κάθε ζεύγους των διαδοχικών στάσεων. Το σύνολο των μετασχηματισμών όλων των διαδοχικών στάσεων χρησιμοποιείται για την ευθυγράμμιση των στάσεων της ακολουθίας σε κάποιο καθορισμένο σύστημα συντεταγμένων. Στο δεύτερο στάδιο, που καλείται **σύνθεση των στάσεων**, χρησιμοποιείται η ευθυγράμμιση του προηγούμενου σταδίου για τη σύνθεση των στάσεων της ακολουθίας. Και τα δυο παραπάνω στάδια εξαρτώνται από τον τρόπο αναπαράστασης του ”μωσαϊκού” στο επίπεδο. Ορισμένοι από τους τρόπους

αναπαράστασης του ”μωσαϊκού” παρουσιάζονται στην ενότητα που ακολουθεί.

4.2 Τρόποι αναπαράστασης "μωσαϊκού" στο επίπεδο

Η κατασκευή του ”μωσαϊκού” επιτυγχάνεται, χρησιμοποιώντας τις στάσεις ολόκληρης της ακολουθίας ή ενός στιγμιοτύπου, αν η ακολουθία αποτελείται από περισσότερα. Αν και η ιδέα του ”μωσαϊκού” είναι απλή μια καλύτερη ματιά αποκαλύπτει μικρές λεπτομέρειες, που οδηγούν σε διαφορετικές προσεγγίσεις του προβλήματος. Για παράδειγμα, όταν δύο στάσεις που περιλαμβάνονται στο ”μωσαϊκό” επικαλύπτονται αλλά έχουν ληφθεί σε διαφορετικές χρονικές στιγμές, η τιμή της έντασης του ίδιου pixel στο ”μωσαϊκό” μπορεί να είναι διαφορετική στις δύο στάσεις. Το κριτήριο επιλογής της έντασης ορίζει τότε δυο διαφορετικές αναπαραστάσεις. Παρόμοια οι διαφορές στην ανάλυση δύο στάσεων οδηγεί στο δίλημμα της επιλογής της ανάλυσης του ”μωσαϊκού”. Οι διαφορετικές αυτές επιλογές έχουν σαν αποτέλεσμα διαφορετικές προσεγγίσεις στον τρόπο κατασκευής του ”μωσαϊκού”. Οι επιλογές είναι συνάρτηση των εφαρμογών, στις οποίες θα χρησιμοποιηθεί τελικά το ”μωσαϊκό”. Στην παρούσα εργασία ασχοληθήκαμε με την κατασκευή του στατικού και δυναμικού ”μωσαϊκού” περιορισμένο σε ”μωσαϊκό” δύο διαστάσεων.

4.2.1 Στατικό "μωσαϊκό" (Static mosaic)

Μια video ακολουθία αποτελείται συχνά από συνεχή στιγμιότυπα σκηνών. Το στατικό ”μωσαϊκό” [9] κατασκευάζεται για κάθε στιγμιότυπο και παρέχει την όψη του σκηνικού που απεικονίζεται στο στιγμιότυπο. Η κατασκευή του επιτυγχάνεται ευθυγραμμίζοντας όλες τις στάσεις του στιγμιοτύπου σε ένα καθορισμένο σύστημα συντεταγμένων, που επιλέγεται από το χρήστη και συνήθως είναι το σύστημα συντεταγμένων της πρώτης στάσης. Οι ευθυγραμμισμένες στάσεις κατόπιν συνθέτονται με φιλτράρισμα στο χρόνο ή απλά κρατώντας την πληροφορία μιας στάσης και προσθέτοντας την επιπλέον πληροφορία των υπολοίπων που προέκυψαν από την κίνηση της κάμερας. Με το φιλτράρισμα στο χρόνο τα κινούμενα αντικείμενα είτε εξαφανίζονται είτε εμφανίζεται θολή η περιοχή που βρίσκονται. Με την προσέγγιση του στατικού ”μωσαϊκού” λαμβάνεται η στατική σκηνή που έλαβε η κάμερα σε όλο το στιγμιότυπο χωρίς τα κινούμενα αντικείμενα, όταν η σύνθεση των στάσεων επιτυγχάνεται με φιλτράρισμα στο χρόνο, ή με τα κινούμενα αντικείμενα σε μία θέση, όταν η σύνθεση των στάσεων προκύπτει προσθέτοντας την επιπλέον πληροφορία της καθεμιάς. Τα σημα-

ντικότερα υπολείμματα ή υπόλοιπα (residuals) υπολογίζονται μεταξύ κάθε στάσης και του στατικού ”μωσαϊκού”. Τα υπολείμματα εκφράζουν συνήθως ανεξάρτητα κινούμενα αντικείμενα ή αλλαγές στο σκηνικό. Το ”μωσαϊκό” που προκύπτει μαζί με τα σημαντικότερα υπολείμματα αποτελεί αποτελεσματική αναπαράσταση των video ακολουθιών. Το στατικό ”μωσαϊκό” είναι ιδανικό για την αποθήκευση video και για γρήγορη πλοήγηση σε μεγάλες ψηφιακές βιβλιοθήκες. Επίσης παρέχει αποδοτική και άμεση πρόσβαση στις ανεξάρτητες στάσεις της ακολουθίας και αυξάνει την απόδοση της δεικτοδότησης βάση του περιεχομένου σε μια video ακολουθία.

4.2.2 Δυναμικό "μωσαϊκό" (Dynamic mosaic)

Το στατικό ”μωσαϊκό” λόγω του τρόπου κατασκευής του δεν αναπαριστά τη δυναμική εξέλιξη των γεγονότων της video ακολουθίας. Ο τρόπος αναπαράστασης που κατορθώνει τη δυναμική αναπαράσταση των γεγονότων καλείται δυναμικό ”μωσαϊκό”. Το δυναμικό ”μωσαϊκό” [9] είναι μια ακολουθία εξελισσόμενων ”μωσαϊκών”. Το περιεχόμενο κάθε ”μωσαϊκού” της ακολουθίας κατασκευάζεται από την πρόσφατη πληροφορία που απεικονίζεται στη νεότερη στάση και την επιπλέον πληροφορία των παλαιότερων στάσεων. Η αναπαράσταση της χρονικής εξέλιξης των γεγονότων επιτυγχάνεται ακριβώς με τη διαδικασία της ενημέρωσης του ”μωσαϊκού” με την πιο πρόσφατη πληροφορία.

Το δυναμικό ”μωσαϊκό” κατασκευάζεται με δύο τρόπους που εξαρτώνται από την επιλογή του συστήματος συντεταγμένων. Όταν το σύστημα συντεταγμένων ή αλλιώς το σύστημα αναφοράς, στο οποίο θα κατασκευαστεί το ”μωσαϊκό”, είναι καθορισμένο και σταθερό, τότε κάθε νέα στάση αντιστοιχίζεται στο προηγούμενο δυναμικό ”μωσαϊκό”, που κατασκευάστηκε βάση του σταθερού συστήματος αναφοράς. Από την άλλη, όταν το σύστημα συντεταγμένων αλλάζει δυναμικά και αντιστοιχεί στο σύστημα συντεταγμένων της τρέχουσας στάσης, τότε το προηγούμενο δυναμικό ”μωσαϊκό” απεικονίζεται στο σύστημα αναφοράς της τρέχουσας στάσης. Πρέπει να σημειωθεί ότι τα ανεξάρτητα κινούμενα αντικείμενα δεν είναι θολά όπως στην περίπτωση του στατικού ”μωσαϊκού” που κατασκευάστηκε με φιλτράρισμα στο χρόνο αλλά η θέση τους συνεχώς ενημερώνεται. Αυτή η προσέγγιση παρέχει τη στατική σκηνή και τα κινούμενα αντικείμενα στην πιο πρόσφατη θέση τους.

Η διαφορά στο περιεχόμενο του στατικού από το δυναμικό ”μωσαϊκό” είναι το συνόλο των υπολειμμάτων που δεν απεικονίζονται στο ”μωσαϊκό”. Τα υπολείμματα του δυναμικού ”μωσαϊκού” αντανακλούν μόνο τις αλλαγές στη σκηνή που συνέβησαν στο χρόνο που πέρασε καθώς και τα επιπλέον μέρη της σκηνής που αποκαλύφτηκαν για

πρώτη φορά στην κάμερα. Τα υπολείμματα όμως του στατικού "μωσαϊκού" περιλαμβάνουν και τα κινούμενα αντικείμεναν, αν υπάρχουν. Είναι φανερό ότι το σύνολο των υπολειμμάτων στο δυναμικό "μωσαϊκό" είναι μικρότερο αφού οι αλλαγές στο σκηνικό μεταξύ διαδοχικών στάσεων είναι μικρές. Επομένως το δυναμικό "μωσαϊκό" παρέχει αποδοτικότερη αναπαράσταση της σκηνής από το στατικό "μωσαϊκό". Στερείται όμως της δυνατότητας της άμεσης πρόσβασης σε ανεξάρτητες στάσεις της ακολουθίας που είναι ουσιαστική για το χειρισμό (manipulation) και εγγραφή (editing) video.

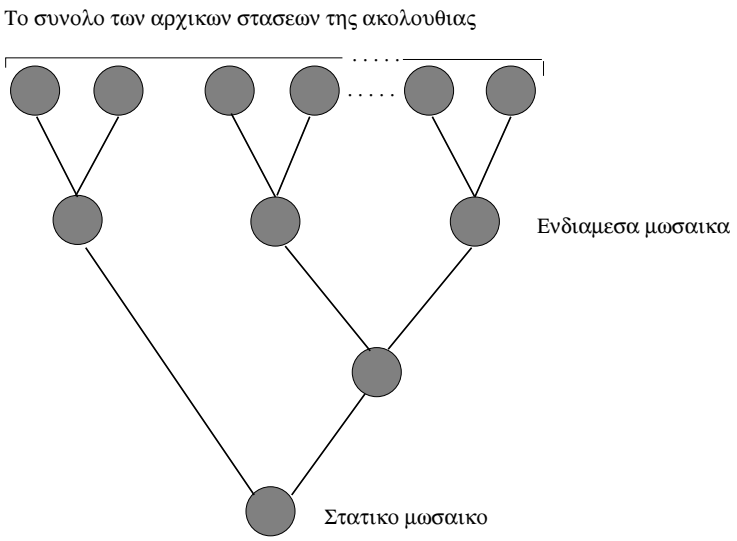
4.2.3 Χρονική πυραμίδα (Temporal pyramid)

Το στατικό και δυναμικό "μωσαϊκό" είναι δύο ακραίες προσεγγίσεις. Στο στατικό "μωσαϊκό" εμφανίζεται μεγάλη διαφορά στην πληροφορία της κάθε στάσης της ακολουθίας με την πληροφορία του "μωσαϊκού". Αντίθετα στο δυναμικό "μωσαϊκό" η διαφορά αυτή είναι περίπου μηδενική, αφού με την πληροφορία της τρέχουσας στάσης (πιο πρόσφατη πληροφορία) ενημερώνεται το "μωσαϊκό". Αποτέλεσμα αυτών είναι το δυναμικό "μωσαϊκό" να είναι αποδοτικότερο. Η αδυναμία του όμως να επιτύχει, σε αντίθεση με το στατικό "μωσαϊκό", άμεση πρόσβαση σε ανεξάρτητες στάσεις είναι ουσιαστική. Για την γεφύρωση του χάσματος μεταξύ των δύο ακραίων προσεγγίσεων χρησιμοποιείται η χρονική πυραμίδα [9].

Η χρονική πυραμίδα είναι η πυραμίδα που το τελευταίο της επίπεδο αποτελείται από τις στάσεις της ακολουθίας. Κάθε κόμβος της πυραμίδας είναι και ένα "μωσαϊκό" που προκύπτει από δύο κόμβους του προηγούμενου επιπέδου. Η δειγματοληψία πάνω στην πυραμίδα μειώνεται καθώς κατεβαίνουμε τα επίπεδα της πυραμίδας με κατεύθυνση προς το πρώτο. Το πρώτο επίπεδο περιλαμβάνει ένα μόνο "μωσαϊκό", που παρουσιάζει ομοιότητα με το στατικό "μωσαϊκό". Τα διαδοχικά επίπεδα της αναπαριστούν "μωσαϊκά" με τα σημαντικότερα υπολείμματα (residuals) σε διάφορα χρονικά διαστήματα. Η ανακατασκευή των στάσεων κατορθώνεται συνδέοντας ιεραρχικά το στατικό "μωσαϊκό", με τα "μωσαϊκά" των διαφόρων επιπέδων που περιέχουν τα υπολείμματα (residuals) των στάσεων. Στο σχήμα 4.1 φαίνεται η μορφή της χρονικής πυραμίδας.

4.2.4 Πολλαπλής ανάλυσης "μωσαϊκό" (Multiresolution mosaic)

Λόγω της αλλαγής της απόστασης της εστίασης της κάμερας κατά τη διάρκεια λήψης της, πολλές στάσεις της ακολουθίας εμφανίζονται με διαφορετική ανάλυση. Κατασκευάζοντας το "μωσαϊκό" στη χαμηλότερη ανάλυση, προκύπτουν απώλειες της



Σχήμα 4.1: Χρονική πυραμίδα

πληροφορίας σε περιοχές με υψηλή συχνότητα. Αντίθετα κατασκευάζοντας το ”μωσαϊκό” με την υψηλότερη ανιχνευόμενη ανάλυση μέσα στην ακολουθία παρουσιάζεται πρόβλημα υπερδειγματοληψίας (oversampling) των περιοχών με χαμηλή ανάλυση. Με την εφαρμογή ενός από των δύο παραπάνω τρόπων υπάρχει περίπτωση απώλειας ή παραμόρφωσης της πληροφορίας. Ένας καλύτερος τρόπος προσέγγισης είναι η χρήση της πυραμίδας πολλαπλών αναλύσεων [9]. Κάθε επίπεδο της πυραμίδας είναι και ένα επίπεδο ανάλυσης. Η πληροφορία της κάθε στάσης προβάλλεται στο κοντινότερο αντίστοιχο επίπεδο ανάλυσης της πυραμίδας. Η πυραμίδα αυτή είναι σποραδική αφού τα επίπεδα ανάλυσης δεν περιέχουν ολόκληρη την πληροφορία του ”μωσαϊκού”. Αυτό είναι συνέπεια του γεγονότος ότι ορισμένες περιοχές του ”μωσαϊκού” αναπαρίστανται σε υψηλή ανάλυση και άλλες μόνο σε χαμηλή. Κατά την ανακατασκευή κάθε στάσης από την πυραμίδα επιλέγεται η υψηλότερη ανάλυση που υπάρχει για τη δεδομένη πληροφορία της στάσης μέσα στην πυραμίδα.

Το πολλαπλής ανάλυσης ”μωσαϊκό” είναι τελείως διαφορετική αναπαράσταση από τη χρονική πυραμίδα, παρόλο που και οι δύο χρησιμοποιούν την δομή δεδομένων της πυραμίδα με φορά από το ευρύτερο προς το στενότερο επίπεδο (coarse-to-fine). Το βασικό δομικό στοιχείο κάθε επιπέδου του πολλαπλής ανάλυσης ”μωσαϊκού” είναι pixel ενώ στη χρονική πυραμίδα είναι ”μωσαϊκό”. Το πολλαπλής ανάλυση ”μωσαϊκό” μπορεί να εφαρμοστεί στη στατική, στη δυναμική και στη χρονική πυραμίδα αναπαράστασης ”μωσαϊκού”.

4.3 Η ευθυγράμμιση των στάσεων της ακολουθίας

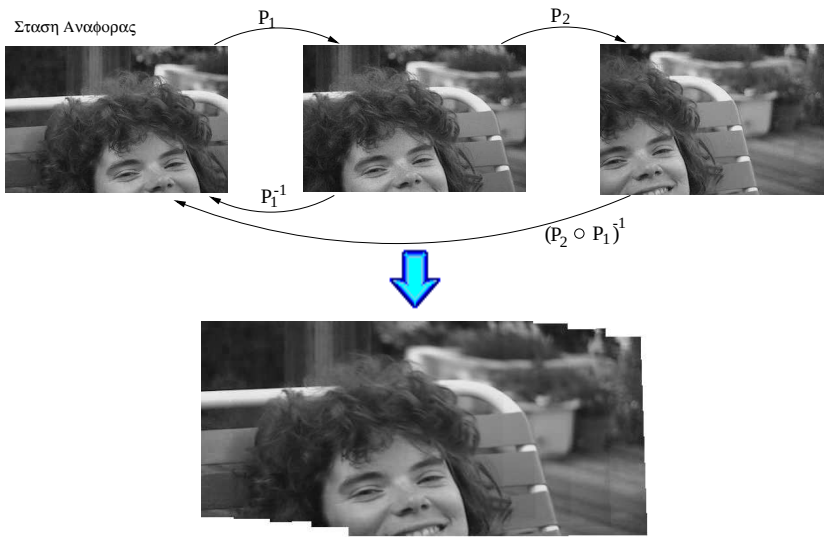
Η ευθυγράμμιση των στάσεων μιας ακολουθίας γενικά επιτυγχάνεται με την εύρεση της κίνησης της κάμερας στο σκηνικό, που περιλαμβάνει η ακολουθία. Ειδικότερα όμως, οι τρόποι ευθυγράμμισης των στάσεων ποικίλουν και εξαρτώνται από την επιλογή του συστήματος αναφοράς, και από την επιλογή των εικόνων ως προς τις οποίες θα εκτιμηθεί η κίνηση. Κάθε διαφορετικό σύστημα αναφοράς που επιλέγεται ορίζει και μία διαφορετική αναπαράσταση του "μωσαϊκού" που απαιτεί και διαφορετική ευθυγράμμιση. Τέτοιες αναπαράστασεις αναφέρθηκαν στην προηγούμενη ενότητα. Αντίθετα η επιλογή των εικόνων πραγματοποιείται με σκοπό την καλύτερη εκτίμηση της κίνησης και επομένως τη δημιουργία "μωσαϊκού" με λιγότερα λάθη στην ευθυγράμμιση των στάσεων. Οι τρεις πιο γνωστοί τρόποι ευθυγράμμισης καθώς τα πλεονεκτήματα ή τα μειονεκτήματα του καθενός παρουσιάζονται στις ακόλουθες ενότητες.

4.3.1 Ευθυγράμμιση στάση - στάση

Η πρώτη προσέγγιση που χρησιμοποιήθηκε από τους ερευνητές για την επίτευξη της ευθυγράμμισης των στάσεων της ακολουθίας με σκοπό την κατασκευή του "μωσαϊκού" είναι η ευθυγράμμιση στάση-στάση. Στην περίπτωση αυτή υπολογίζεται η κίνηση της κάμερας μεταξύ των διαδοχικών στάσεων όλης της ακολουθίας. Οι παράμετροι κίνησης που προκύπτουν συνθέτονται για την εκτίμηση των παράμετρων ευθυγράμμισης της κάθε στάσης. Οι παράμετροι ευθυγράμμισης καθορίζουν την ακριβή θέση της στάσης μέσα στο "μωσαϊκό".

Το στατικό "μωσαϊκό" που αναφέραμε κατασκευάζεται με το συγκεκριμένο τρόπο ευθυγράμμισης των στάσεων. Κατά την κατασκευή του "μωσαϊκού" όλες οι στάσεις της ακολουθίας ευθυγραμμίζονται σε ένα καθορισμένο σύστημα αναφοράς. Ένα διαφορετικό σύστημα αναφοράς ορίζει και μια διαφορετική αναπαράσταση του "μωσαϊκού". Αν το σύστημα αναφοράς είναι το σύστημα συντεταγμένων μιας καθορισμένης στάσης, όλες οι στάσεις της ακολουθίας ευθυγραμμίζονται ως προς αυτήν. Στην περίπτωση όμως επιλογής ενός εικονικού συστήματος αναφοράς, απαιτείται και ο μετασχηματισμός μεταξύ του εικονικού συστήματος συντεταγμένων και μιας από τις στάσεις της ακολουθίας, που καλείται **στάση αναφοράς**. Οι παράμετροι ευθυγράμμισης κάθε στάσης προκύπτουν από τη σύνθεση των παραμέτρων του μετασχηματισμού, εξαιτίας του εικονικού συστήματος, και των παραμέτρων της ευθυγράμμισης της στάσης ως προς τη στάση αναφοράς.

Η διαδικασία της ευθυγράμμισης όλης της ακολουθίας απαιτεί ένα μόνο πέρασμα υπολογισμών κατά μήκος της ακολουθίας. Κατά το πέρασμα αυτό εκτιμούνται οι μετασχηματισμοί ευθυγράμμισης των διαδοχικών στάσεων. Ακολουθιακά οι μετασχηματισμοί αυτοί συνθέτονται για την εκτίμηση του μετασχηματισμού της κάθε στάσης ως προς το σύστημα αναφοράς (σχήμα 4.2).



Σχήμα 4.2: Ευθυγράμμιση των στάσεων της ακολουθίας ως προς το σύστημα συντεταγμένων του ”μωσαϊκού”

Το σύστημα συντεταγμένων της στάσης αναφοράς είναι το σύστημα συντεταγμένων πάνω στο οποίο θα κατασκευαστεί το ”μωσαϊκό”

Στην παρούσα εργασία υλοποιήθηκε η ευθυγράμμιση στάση-στάση. Μετά την εκτίμηση των παραμέτρων, μεταξύ των διαδοχικών στάσεων, ενός από τα μοντέλα κίνησης, εστίασης και μεταφοράς, αφινικού και προβολικού, ακολουθεί η σύνθεση των παραμέτρων του μοντέλου, που επιλέγεται, με σκοπό την ευθυγράμμιση όλων των στάσεων ως προς το σύστημα αναφοράς. Παρακάτω παρουσιάζεται η σύνθεση των παραμέτρων για το σύνολο των μοντέλων που υλοποιήθηκαν στην εργασία μας.

4.3.1.1 Ευθυγράμμιση στάσεων με χρήση του αφινικού μοντέλου κίνησης

Η σύνθεση των παραμέτρων του αφινικού, του μοντέλου εστίασης και μεταφοράς και του μοντέλου απλής μεταφοράς του πίνακα 3.1, παρουσιάζει ομοιότητα αφού οι μετασχηματισμοί αυτοί μπορούν να χωριστούν σε έναν πίνακα A_{2x2} που εκφράζει το μετασχηματισμό των συντεταγμένων (x, y) της κάθε στάσης και σε έναν πίνακα B_{2x1} που εκφράζει τη μετατόπιση της κάμερας στις δύο διαστάσεις. Οι μετασχηματισμοί αυτοί

περιγράφονται συνοπτικότερα από την εξίσωση:

$$P' = A_{2x2}\ P\ +\ B_{2x1} \tag{4.1}$$

η μεταβλητή P εκφράζει τις συντεταγμένες $(x,\ y)$ κάθε σημείου της στάσης ενώ η P′ τις αντίστοιχες συντεταγμένες $(x',\ y')$ που προκύπτουν εφαρμόζοντας το μετασχηματισμό. Ο πίνακας A_{2x2} διαφέρει για καθένα από τους τρεις πρώτους μετασχηματισμούς του πίνακα 3.1. Για παράδειγμα, για τον αφινικό μετασχηματισμό που αποτελεί το γενικότερο μοντέλο κίνησης από τα τρία πρώτα μοντέλα του πίνακα 3.1, οι πίνακες A_{2x2} και B_{2x1} δίνονται από τις ακόλουθες εξισώσεις:

$$A_{2x2}=\begin{bmatrix} a_1 & a_2 \\ a_4 & a_5 \end{bmatrix} \quad B_{2x1}=\begin{bmatrix} a_3 \\ a_6 \end{bmatrix} \tag{4.2}$$

όπου οι μεταβλητές $a_1, a_2, ...a_6$ εκφράζουν τις παραμέτρους του μοντέλου που η ανάλυση του στις δύο διαστάσεις είναι:

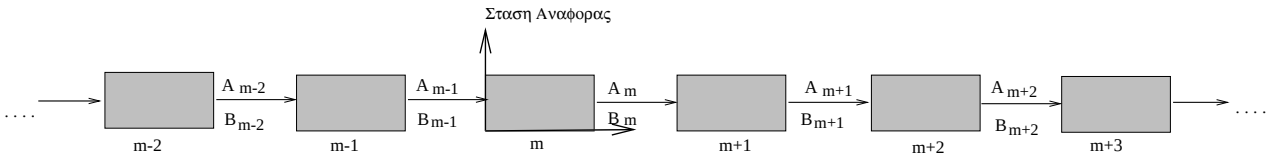
$$\begin{aligned} x'&=a_1x\ +\ a_2y\ +\ a_3 \\ y'&=a_4x\ +\ a_5y\ +\ a_6 \end{aligned} \tag{4.3}$$

Αντίθετα στην περίπτωση του μοντέλου κίνησης της απλής μεταφοράς ο πίνακας A_{2x2} είναι μοναδιαίος και ο πίνακας B_{2x1} συμπίπτει με το διάνυσμα της παραμέτρου b που εκφράζει την μεταφορά της κάμερας στον πίνακα 3.1.

Στην περίπτωση αυτής της ευθυγράμμισης, όπως ήδη ειπώθηκε, έχει εκτιμηθεί η κίνηση της κάμερας και επομένως οι παράμετροι του μετασχηματισμού μεταξύ των διαδοχικών στάσεων. Η σχέση που συνδέει τις δύο διαδοχικές στάσεις, k και k+1, όταν είναι γνωστός ο μετασχηματισμός τους, είναι:

$$P_{k+1} = A_k\ P_k\ +\ B_k \tag{4.4}$$

όπου A_k και B_k είναι ο μετασχηματισμός και η μετατόπιση αντίστοιχα που υφίσταται η k-οστή στάση για να απεικονιστεί στο σύστημα συντεταγμένων της k+1.



Σχήμα 4.3: Ευθυγράμμιση στάσεων με χρήση του αφινικού μοντέλου κίνησης

Έστω ότι επιλέγουμε σύστημα αναφοράς, βάση του οποίου θα αναπαρασταθεί το "μωσαϊκό", το σύστημα συντεταγμένων της m στάσης. Η στάση αυτή μπορεί να είναι οποιαδήποτε στάση του στιγμιοτύπου. Στο σχήμα 4.3 παριστάνεται η ακολουθία των στάσεων. A_i και B_i (i=0...N) είναι οι πίνακες που εκφράζουν το μετασχηματισμό κάθε στάσης ως προς τη διαδοχικής της. Ο μετασχηματισμός αυτός υπολογίζεται με τις μεθόδους εκτίμησης της κίνησης που αναφέραμε στο προηγούμενο κεφάλαιο. Για παράδειγμα, η στάση m λαμβάνεται από τη στάση m-1 μέσω της σχέσης: $P_m = A_{m-1}P_{m-1} + B_{m-1}$. Επομένως η ευθυγράμμιση κάθε στάσης $\lambda\epsilon\{0...N\}$ ως προς το σύστημα αναφοράς της m δίνεται από τις παρακάτω σχέσεις:

$$P_m = \left(\prod_{\substack{i=m-1\\ \lambda<m-1}}^{\lambda} A_i\right) P_\lambda + \left(\sum_{\substack{i=m-1\\ \lambda+1<m-1}} \left(\prod_{\substack{j=\lambda\\ i-1>\lambda}}^{i-1} A_j\right) B_i\right) + B_\lambda \qquad \text{αν το } \lambda < m \tag{4.5}$$

$$P_m = \left(\prod_{\substack{i=\lambda-1\\ m<\lambda-1}}^{m} A_i\right)^{-1} \left(P_\lambda - \left(\sum_{\substack{i=m\\ \lambda-2>m}}^{\lambda-2} \left(\prod_{\substack{j=\lambda-1\\ i+1<\lambda-1}}^{i+1} A_j\right) B_i\right) - B_{\lambda-1}\right) \qquad \text{αν το } \lambda > m \tag{4.6}$$

όπου P_m οι συντεταγμένες της στάσης λ ως προς το σύστημα αναφοράς της m. Ο δεύτερος δείκτης στο άθροισμα και στο γινόμενο καθορίζει το βήμα της μεταβλητής i ή j του αθροίσματος ή γινομένου. Όταν ο εκθέτης του γινομένου ή του αθροίσματος είναι μεγαλύτερος από τον πρώτο δείκτη, τότε το βήμα της μεταβλητής του γινομένου ή του αθροίσματος είναι +1 αλλιώς είναι -1. Για παράδειγμα ο δείκτης λ < m-1 στο πρώτο γινόμενο της σχέσης 4.5 εκφράζει ότι το βήμα της μεταβλητής i του γινομένου είναι -1. Όμοια ο δείκτης i-1 > λ στο δεύτερο γινόμενο της ίδιας σχέσης εκφράζει ότι το βήμα της μεταβλητής j του γινομένου είναι +1. Η ευθυγράμμιση που παρουσιάστηκε ισχύει για κάθε μετασχηματισμό του πίνακα 3.1 που μπορεί να γραφεί σύμφωνα με την εξίσωση 4.4.

Στην περίπτωση της επιλογής ενός εικονικού συστήματος συντεταγμένων ως σύστημα αναφοράς, η ευθυγράμμιση της κάθε στάσης ακολουθεί πάλι τις εξίσωσεις 4.5 και 4.6, θεωρώντας τη στάση m σαν στάση αναφοράς. Επιπλέον όμως λαμβάνεται υπόψη και ο μετασχηματισμός της m στάσης ($A_{\Sigma A}$, $B_{\Sigma A}$) ως προς το εικονικό σύστημα συντεταγμένων. Επομένως, οι συντεταγμένες κάθε στάσης λ ως προς το εικονικό σύστημα αναφοράς δίνονται από τη σχέση:

$$P_{\Sigma A} = A_{\Sigma A}\, P_m + B_{\Sigma A} \tag{4.7}$$

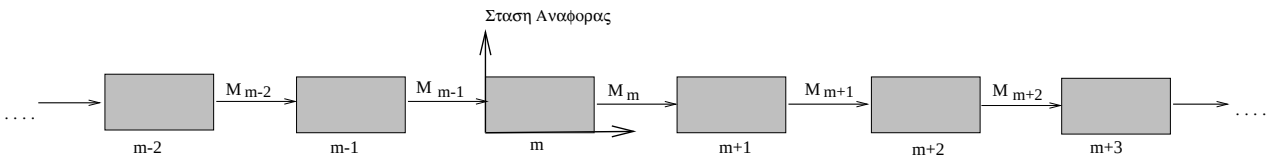
όπου το P_m προκύπτει από τις εξισώσεις 4.5 και 4.6.

4.3.1.2 Ευθυγράμμιση στάσεων με χρήση του προβολικού μοντέλου κίνησης

Στην ενότητα αυτή θα περιγραφεί η ευθυγράμμιση όλων των στάσεων της ακολουθίας υποθέτοντας ότι το προβολικό μοντέλο που απεικονίζεται στον πίνακα 3.1 περιγράφει το μετασχηματισμό μεταξύ δύο οποιονδήποτε διαδοχικών στάσεων της ακολουθίας. Παριστάνοντας αναλυτικότερα τον προβολικό μετασχηματισμό στις δύο διαστάσεις, προκύπτουν οι ακόλουθες σχέσεις:

$$\begin{aligned} x' &= \frac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + 1} \\ y' &= \frac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + 1} \end{aligned} \tag{4.8}$$

Ο προβολικός μετασχηματισμός, λόγω της περίπλοκης έκφρασης της εξίσωσης, που



Σχήμα 4.4: Ευθυγράμμιση στάσεων με χρήση του προβολικού μοντέλου κίνησης

τον περιγράφει, δε μπορεί να απεικονιστεί απλά σε έναν πίνακα. Θεωρούμε όμως ότι εκφράζεται από τον πίνακα $M_k = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix}$. Οι αλγεβρικές πράξεις τότε που χρησιμοποιούν τον πίνακα απαιτούν ιδιαίτερη προσοχή. Για παράδειγμα ο αντίστροφος μετασχηματισμός του προβολικού μετασχηματισμού που ορίζεται από τον πίνακα M_k, δεν ισοδυναμεί με τον αριθμητικό αντίστροφο του πίνακα αλλά με τον πίνακα του μερικών οριζουσών του, δηλαδή:

$$IM_k = adj\, M_k = |M_k|\ M_k^{-1} \tag{4.9}$$

όπου $|M_k|$ είναι η ορίζουσα του πίνακα M_k και το M_k^{-1} είναι ο αριθμητικός αντίστροφος του πίνακα M_k. Η σχέση που συνδέει τις διαδοχικές στάσεις k και k+1, όταν είναι γνωστός ο μετασχηματισμός τους είναι:

$$P_{k+1} = M_k\ P_k \tag{4.10}$$

όπου τα P_k και P_{k+1} απεικονίζουν τις συντεταγμένες των σημείων των k και k+1 στάσεων αντίστοιχα.

Για την ευθυγράμμιση των διαδοχικών στάσεων της ακολουθίας (4.4) επιλέγουμε αρχικά το σύστημα αναφοράς, ως προς το οποίο θα αναπαρασταθεί το "μωσαϊκό". Έστω ότι το σύστημα αναφοράς είναι το σύστημα συντεταγμένων της m στάσης. Η σχέση που συνδέει τα σημεία των δύο διαδοχικών στάσεων k και k+1 απεικονίζεται στην εξίσωση 4.10. Η ευθυγράμμιση κάθε στάσης λ ως προς το σύστημα συντεταγμένων της m στάσης προκύπτει με μία από τις ακόλουθες σχέσεις:

$$P_m = \left(\prod_{\substack{i=m-1 \\ \lambda<m-1}}^{\lambda} M_i \right) P_\lambda \qquad \text{αν το } \lambda < m \tag{4.11}$$

$$P_m = \left(\prod_{\substack{i=\lambda-1 \\ m<\lambda-1}}^{m} M_i \right)^{-1} P_\lambda \qquad \text{αν το } \lambda > m \tag{4.12}$$

Η σχέση του δεύτερου δείκτη στην αύξηση ή στη μείωση του βήματος του γινόμενου ή του αθροίσματος, ακολουθεί τη λογική που εφαρμόστηκε στην ευθυγράμμιση των στάσεων με χρήση του αφινικού μοντέλου κίνησης. Το γινόμενο των πινάκων M_i στην παραπάνω εξίσωση εκφράζει τη σύνθεση προβολικών μετασχηματισμών και αντιστοιχεί επίσης σε ένα προβολικό μετασχηματισμό (4.8): $M_{prod} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix}$. Ο αντίστροφος του γινομένου των προβολικών μετασχηματισμών (IM_{prod}) που απεικονίζεται στην εξίσωση 4.12, υπολογίζεται από την εξίσωση 4.9 ($IM_{prod} = |M_{prod}|\ M_{prod}^{-1}$).

Στην περίπτωση επιλογής ενός εικονικού συστήματος συντεταγμένων, ως σύστημα αναφοράς, η ευθυγράμμιση της κάθε στάσης ακολουθεί τις εξίσωσεις 4.11 και 4.12, θεωρώντας τη στάση m σαν στάση αναφοράς. Επιπλέον όμως λαμβάνεται υπόψη και ο μετασχηματισμός της m στάσης ($M_{\Sigma A}$) ως προς το εικονικό σύστημα αναφοράς. Οι συντεταγμένες κάθε στάσης λ ως προς το εικονικό σύστημα αναφοράς δίνονται από τη σχέση:

$$P_{\Sigma A} = M_{\Sigma A}\ P_m \tag{4.13}$$

όπου το P_m προκύπτει από τις εξισώσεις 4.11 και 4.12.

4.3.2 Ευθυγράμμιση στάσης με "μωσαϊκό"

Η ευθυγράμμιση αυτή υπονοεί την εκτίμηση της κίνησης της κάμερας όχι μεταξύ των διαδοχικών στάσεων, αλλά μεταξύ της κάθε στάσης και του τρέχοντος "μωσαϊκού".

Το τρέχον ”μωσαϊκό” κατασκευάζεται από το σύνολο των προηγούμενων στάσεων της νέας στάσης. Για το χειρισμό του προβλήματος των μεγάλων ταχυτήτων μεταξύ του τρέχοντος ”μωσαϊκού” και της κάθε νέας στάσης, οι παράμετροι του μετασχηματισμού που υπολογίστηκαν μεταξύ της προηγούμενης στάσης και του προηγούμενου ”μωσαϊκού”, χρησιμοποιούνται σαν αρχική εκτίμηση των νέων παραμέτρων μετασχηματισμού.

4.3.3 Ευθυγράμμιση "μωσαϊκού" με στάση

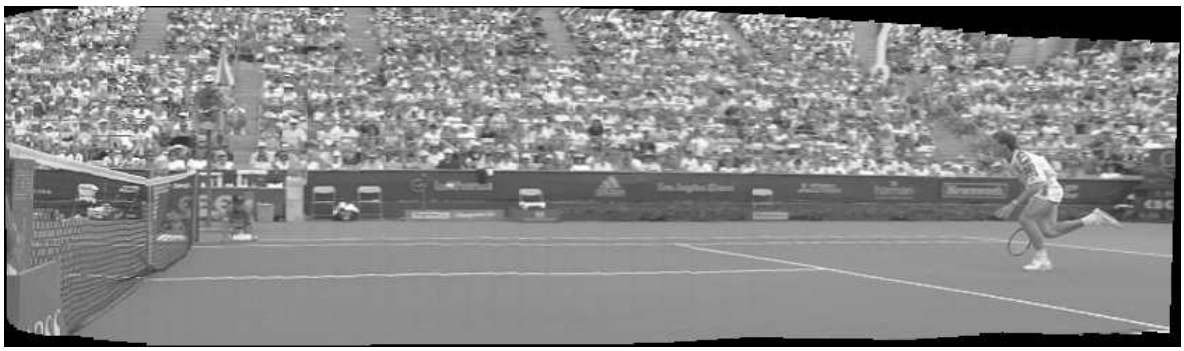
Σε ορισμένες δυναμικές εφαρμογές όπως στη μετάδοση video σε πραγματικό χρόνο είναι σημαντικό να διατηρούνται οι εικόνες στο δικό τους αρχικό σύστημα συντεταγμένων. Σε αυτή την περίπτωση, είναι περισσότερο χρήσιμη η ευθυγράμμιση του ”μωσαϊκού” στο σύστημα συντεταγμένων της τρέχουσας στάσης. Ο μετασχηματισμός τότε μεταξύ του τρέχοντος ”μωσαϊκού” και της τρέχουσας στάσης είναι όμοιος με το μετασχηματισμό μεταξύ της προηγούμενης και της νέας στάσης.

4.4 Σύνθεση στάσεων για την κατασκευή του "μωσαϊκού"

Το τελικό βήμα για την κατασκευή του ”μωσαϊκού” είναι οι σύνθεση των στάσεων σύμφωνα με το είδος της ευθυγράμμισης που επιλέχθηκε. Η σύνθεση των στάσεων στην εργασία μας επιτυγχάνεται με τους παρακάτω τρόπους.

1. Μια από τις πιο απλές μεθόδους σύνθεσης είναι η απλή τοποθέτηση των στάσεων στη θέση που τους αντιστοιχεί στο ”μωσαϊκό”, όπως προέκυψε από τη διαδικασία της ευθυγράμμισης τους. Στην περίπτωση αυτή η τοποθέτηση μπορεί να γίνει με δύο τρόπους γεγονός που βασίζεται στην εφαρμογή που θα χρησιμοποιηθεί το ”μωσαϊκό”. Αν θέλουμε το ”μωσαϊκό” που θα κατασκευαστεί να περιέχει την παλαιότερη πληροφορία του σκηνικού, η τοποθέτηση των στάσεων πραγματοποιείται ξεκινώντας από την πιο πρόσφατη στάση και καταλήγοντας στην τοποθέτηση της πιο παλιάς. Έτσι η πληροφορία των παλαιότερων στάσεων τοποθετείται πάνω από τις πρόσφατες στο υπό κατασκευή ”μωσαϊκό”. Παράδειγμα τέτοιας σύνθεσης φαίνεται στο σχήμα 4.5. Αντίθετα αν επιθυμείται το ”μωσαϊκό” να περιλαμβάνει την πιο πρόσφατη πληροφορία του σκηνικού τότε η τοποθέτηση έχει αντίθετη φορά, αρχίζοντας από την παλαιότερη στάση και καταλήγοντας στην πιο πρόσφατη. Και με τους δύο τρόπους λαμβάνεται το σταθερό σκηνικό που έλαβε η κάμερα καθώς και τα ανεξάρτητα κινούμενα αντικείμενα, αν υπάρχουν σε κάποια θέση. Στην πρώτη περίπτωση η θέση των κινούμενων αντικειμένων συμπίπτει με τη θέση τους

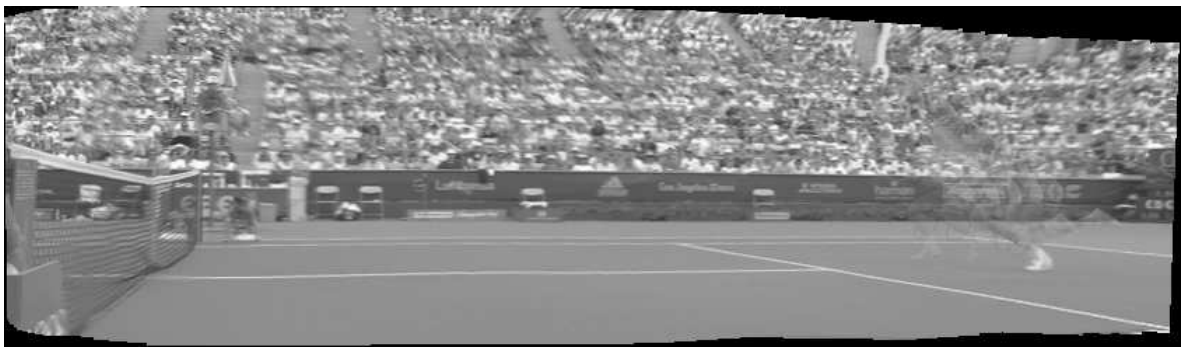
στη στάση αναφοράς, ενώ στην άλλη περίπτωση συμπίπτει με τη θέση τους στην τελευταία πιο πρόσφατη στάση του στιγμιοτύπου. Στο παράδειγμα του σχήματος 4.5 ο τενίστας που αποτελεί το κινούμενο αντικείμενο της ακολουθίας βρίσκεται στη θέση που κατέχει στη στάση αναφοράς. Η σύνθεση με τους τρόπους αυτούς χρησιμοποιείται συνήθως όταν η εκτίμηση της κίνησης μεταξύ των στάσεων είναι ακριβής. Σε διαφορετική περίπτωση τα λάθη της ευθυγράμμισης των στάσεων στο ”μωσαϊκό” είναι ορατά.



Σχήμα 4.5: Η σύνθεση των στάσεων προκύπτει με την απλή τοποθέτηση των στάσεων στο ”μωσαϊκό”

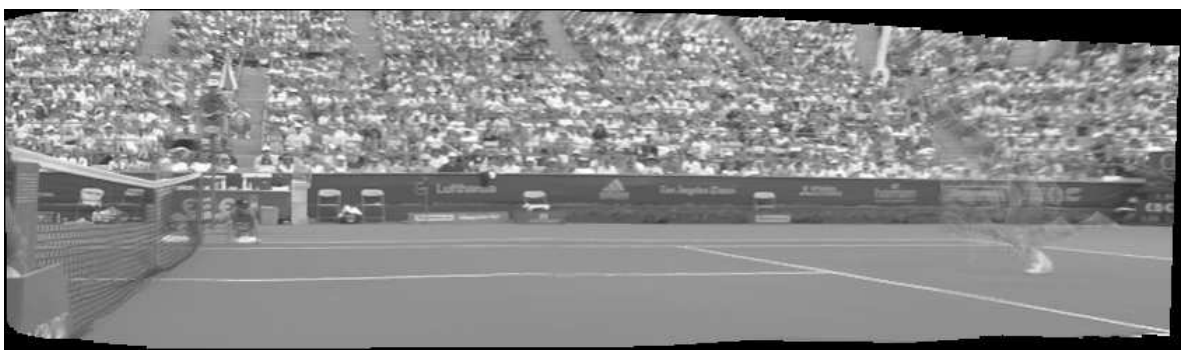
Το ”μωσαϊκό” κατασκευάζεται ξεκινώντας από την πιο πρόσφατη και καταλήγοντας στην πιο παλιά. Ο τενίστας αποτελεί το κινούμενο αντικείμενο της ακολουθίας. Η θέση του στο ”μωσαϊκό” συμπίπτει με τη θέση του στη στάση αναφοράς

2. Εκτός από την απλή τοποθέτηση των στάσεων ενδιαφέρον παρουσιάζει και η σύνθεση με φιλτράρισμα στο χρόνο. Ένα από τα πιο διαδεδομένα φίλτρα που χρησιμοποιούνται είναι αυτό της μέσης τιμής. Η τιμή της έντασης του κάθε pixel του ”μωσαϊκού” προκύπτει από τη μέση τιμή των εντάσεων των pixel των στάσεων, που σύμφωνα με τη διαδικασία της ευθυγράμμισης τοποθετούνται στη θέση αυτή του ”μωσαϊκού”. Η χρήση του φίλτρου αυτού στις ακολουθίες με ανεξάρτητα κινούμενα αντικείμενα έχει σαν αποτέλεσμα είτε την εξαφάνιση των αντικειμένων είτε την εμφάνιση θολών περιοχών στις θέσεις που αυτά κατέχουν μέσα στο ”μωσαϊκό” (ghostlike). Παράδειγμα τέτοιας σύνθεσης φαίνεται στο σχήμα 4.6. Το φίλτρο εφαρμόστηκε στις πιο πρόσφατες στάσεις που έχουν πληροφορία για κάθε ένα σημείο του ”μωσαϊκού”. Η προσέγγιση αυτή πραγματοποιήθηκε για τη μείωση του βαθμού θολότητας που εμφανίζεται στη σύνθεση αυτή. Με τη μέθοδο αυτή λαμβάνεται το σταθερό σκηνικό που έλαβε η κάμερα. Επίσης λόγω του φίλτρου αυτού, τα λάθη της κακής ευθυγράμμισης των στάσεων απορροφούνται και δεν είναι εμφανή στο τελικό ”μωσαϊκό”.



Σχήμα 4.6: Η σύνθεση των στάσεων με τη χρήση του φίλτρου της μέσης τιμής

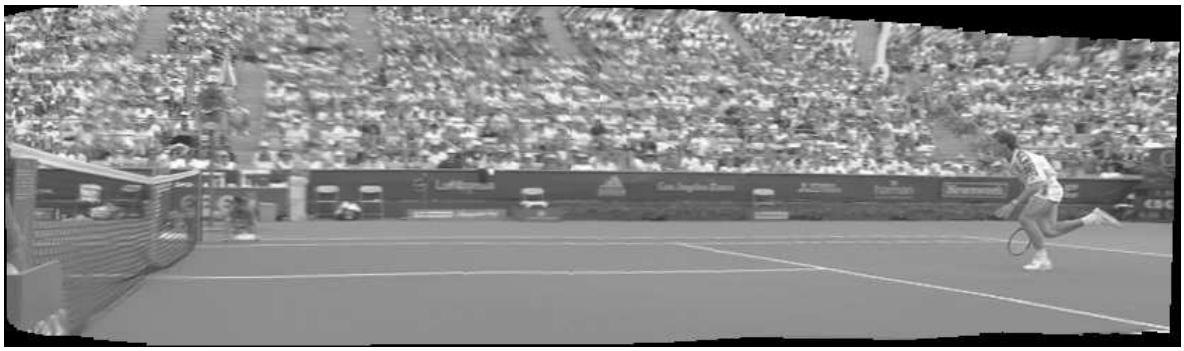
3. Ένα ακόμα φίλτρο με ευρεία χρήση, στη σύνθεση στάσεων για την κατασκευή του ”μωσαϊκού”, είναι το φίλτρο της ενδιάμεσης τιμής. Όπως και το φίλτρο της μέσης τιμής εξαφανίζει ή παρουσιάζει θολές τις περιοχές που βρίσκονται μέσα στο ”μωσαϊκό” τα ανεξάρτητα κινούμενα αντικείμενα του στιγμιοτύπου. Το ”μωσαϊκό”, που λαμβάνουμε με το φίλτρο αυτό, περιέχει επίσης το σταθερό σκηνικό που έλαβε η κάμερα. Περιπτώσεις λαθών λόγω των κακών ευθυγραμμίσεων διορθώνονται όπως και στην περίπτωση του φίλτρου της μέσης τιμής. Η διαφορά του με το φίλτρο της μέσης τιμή είναι ότι το ”μωσαϊκό” που κατασκευάζεται είναι λιγότερο θολό από το ”μωσαϊκό” που προκύπτει από το φίλτρο της μέσης τιμής. Ένα παράδειγμα τέτοιας σύνθεσης φαίνεται στο σχήμα 4.7. Η περιοχή που βρίσκεται ο τενίστας που αποτελεί το κινούμενο αντικείμενο της ακολουθίας εμφανίζεται ιδιαίτερα θολή.



Σχήμα 4.7: Η σύνθεση των στάσεων με τη χρήση του φίλτρου της ενδιάμεσης τιμής

4. Ενδιαφέρον παρουσιάζει και μία άλλη προσέγγιση σύνθεσης στάσεων η οποία συνδυάζει τους δύο πρώτους τρόπους σύνθεσης (1, 2). Για τη σύνθεση των

στάσεων χρησιμοποιείται η πρώτη μέθοδος δηλαδή η κάθε στάση τοποθετείται στο ”μωσαϊκό” στη θέση που της καθορίστηκε κατά την ευθυγράμμιση. Κατά την τοποθέτηση όμως κάθε στάσης τα σημεία που ανήκουν σε ζώνη του περιγράμματος της, διάστασης K, δεν τοποθετούνται απλά στο ”μωσαϊκό”. Η τιμή των εντάσεων του ”μωσαϊκού” στις θέσης αυτές προκύπτει με την εφαρμογή του φίλτρου της μέσης τιμής στα ακριανά αυτά σημεία της στάσης και στα σημεία των άλλων στάσεων που αντιστοιχούν στις θέσεις αυτές του ”μωσαϊκού”. Με τον τρόπο αυτό κακές ευθυγραμμίσεις στα όρια των στάσεων δεν είναι εμφανείς στο τελικό ”μωσαϊκό”. Ένα παράδειγμα εφαρμογής αυτής της προσέγγισης φαίνεται στο σχήμα 4.8. Ο τενίστας που αποτελεί το κινούμενο αντικείμενο της ακολουθίας βρίσκεται στη θέση που κατέχει στη στάση αναφοράς.



Σχήμα 4.8: Η σύνθεση των στάσεων με τη χρήση του φίλτρου της μέσης τιμής σε μια ζώνη διάστασης K του περιγράμματος των στάσεων

4.5 Το πρόβλημα της κακής ευθυγράμμισης των στάσεων

Μια από τις πρώτες προσεγγίσεις για την ευθυγράμμιση των στάσεων στην κατασκευή του ”μωσαϊκού” ήταν όπως αναφέραμε η ευθυγράμμιση στάση-στάση. Στην προσέγγιση αυτή οι παράμετροι του μετασχηματισμού μεταξύ διαδοχικών στάσεων συνθέτονται για να προκύψει η θέση της κάθε στάσης στο ”μωσαϊκό”. Το βασικό όμως πρόβλημα σε αυτή την προσέγγιση είναι ότι οι παράμετροι, που λαμβάνονται κατά την ευθυγράμμιση των στάσεων δεν είναι αρκετά ακριβείς, για την κατασκευή ενός καλά ευθυγραμισμένου ”μωσαϊκού”. Μικρά λάθη στον υπολογισμό των παραμέτρων ευθυγράμμισης μεταξύ των διαδοχικών στάσεων, ή μικρές αποκλίσεις του μοντέλου κίνησης από το πραγματικό, συσσωρεύονται στην εκτίμηση των τελικών παραμέτρων για την

ευθυγράμμιση της κάθε στάσης στο ”μωσαϊκό”. Τα σφάλματα συσσώρευσης εξαρτόνται σημαντικά από το μήκος της video ακολουθίας. Σε ένα video μεγάλου μήκους ο αριθμός των σφαλμάτων που εμφανίζονται σε επιμέρους μετασχηματισμούς των στάσεων είναι μεγαλύτερος αφού μεγαλύτερος είναι και ο αριθμός των στάσεων. Τα σφάλματα αυτά πολλαπλασιάζονται κατά την ευθυγράμμιση των στάσεων σε ένα καθορισμένο σύστημα αναφοράς και αυξάνονται βαθμιαία.

Για την αποφυγή του παραπάνω προβλήματος χρησιμοποιήθηκε η προσέγγιση της ευθυγράμμισης της κάθε στάσης με το ”μωσαϊκό”. Τα αποτελέσματα ήταν καλύτερα κάτω όμως από ορισμένες προϋποθέσεις. Μια από τις προϋποθέσεις ήταν η κατασκευή του ”μωσαϊκού” από τη στάση αναφοράς και ολόκληρη την πληροφορία της νέας στάσης. Αντίθετα, κατά την κατασκευή δηλαδή του ”μωσαϊκού” με την παλαιότερη πληροφορία, προβλήματα κακής εκτίμησης της κίνησης εμφανίζονται. Τα προβλήματα οφείλονται στη μεγάλη διαφορά πληροφορίας που παρουσιάζουν η νέα στάση και η περιοχή του ”μωσαϊκού” που αντιστοιχεί στη θέση της νέας στάσης. Η διαφορά αυτή εμφανίζεται λόγω των αλλαγών της σκηνής στις περιοχές αυτές και οφείλεται στην εμφάνιση ή εξαφάνιση κινούμενων αντικειμένων ή στην αλλαγή του φωτισμού. Επίσης, με τον ίδιο τρόπο κατασκευής του ”μωσαϊκού”, προβλήματα προκαλεί και η τυχόν κακή ευθυγράμμιση της προηγούμενης στάσης, αφού η περιοχή του ”μωσαϊκού” με την κακή ευθυγράμμιση λαμβάνει μέρος στην εκτίμηση της κίνησης με τη νέα στάση. Παράλληλα, προβλήματα παρουσιάζει και η διαφορά στην ανάλυση που παρουσιάζουν το εκάστοτε ”μωσαϊκό” και η νέα στάση, ιδιαίτερα αν είναι μεγάλη. Η μεγάλη διαφορά στην ανάλυση μπορεί να οδηγήσει σε λανθασμένη εκτίμηση της κίνησης αν το παραμετρικό μοντέλο που χρησιμοποιείται δεν είναι ικανό να την καλύψει. Επομένως ιδιαίτερη προσοχή πρέπει να λαμβάνεται στην επιλογή του συστήματος αναφοράς, αφού εν μέρη αυτό ρυθμίζει την ανάλυση.

Άμεση λύση για τα παραπάνω προβλήματα δεν έχει προταθεί μέχρι τώρα. Επομένως για την κατασκευή ενός ορθού ”μωσαϊκού” αποδεχόμαστε τις παραπάνω προσεγγίσεις και αναζητούμε τρόπους διόρθωσης της εκτίμησης της κίνησης μεταξύ των διαδοχικών στάσεων ή μεταξύ κάθε στάσης και του αντίστοιχου ”μωσαϊκού”. Για το σκοπό αυτό στην παρούσα εργασία προτείνουμε την τοπική διόρθωση της ήδη υπολογιζόμενης κίνησης μεταξύ των διαδοχικών στάσεων.

4.6 Τοπική διόρθωση της ευθυγράμμισης των στάσεων

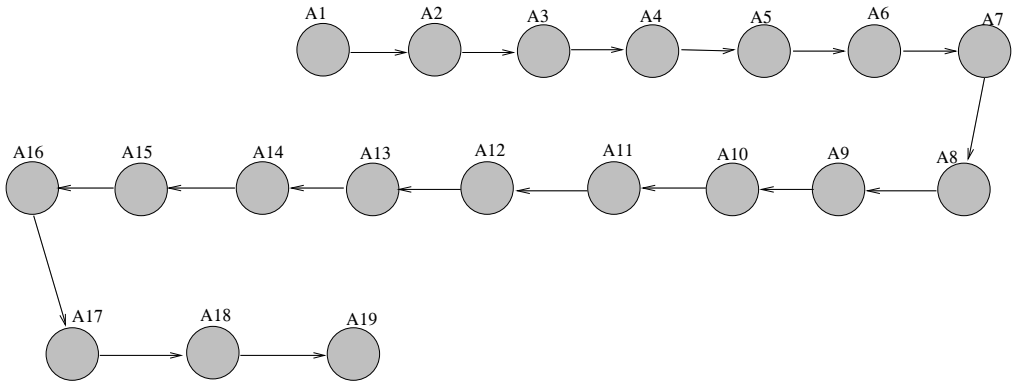
Για την κατασκευή ενός καλά ευθυγραμμισμένου ”μωσαϊκού” καταφεύγουμε στη διόρθωση των παραμέτρων του μοντέλου κίνησης που έχουν ήδη υπολογιστεί με τις μεθόδους εκτίμησης της κίνησης που παρουσιάστηκαν. Πριν την εφαρμογή του αλγορίθμου της τοπική διόρθωσης θεωρούμε ότι έχει πραγματοποιηθεί η διαδικασία της ευθυγράμμισης όλων των στάσεων του στιγμιοτύπου ως προς το ”μωσαϊκό”. Η ευθυγράμμιση των στάσεων που εφαρμόζεται είναι ακριβώς η μέθοδος της ευθυγράμμισης στάσης-στάσης. Ο αλγόριθμος της διόρθωσης παίρνει μέρος κατά τη διαδικασία κατασκευής του ”μωσαϊκού” , και είναι ένας από τους αλγορίθμους εκτίμησης της κίνησης που αναφέραμε σε προηγούμενες ενότητες. Στη διαδικασία κατασκευής του ”μωσαϊκού” όλες οι στάσεις του στιγμιοτύπου μετασχηματίζονται χρησιμοποιώντας τις παραμέτρους κίνησης που υπολογίστηκαν κατά την ευθυγράμμιση των στάσεων. Ο μετασχηματισμός των στάσεων πραγματοποιείται με τη χρήση της μεθόδου της δι-γραμμικής παρεμβολής που ήδη αναφέραμε (εξίσωση 3.15). Ο αλγόριθμος της διόρθωσης εφαρμόζεται πάνω σε επιλεγμένες μετασχηματισμένες στάσεις. Οι νέες παράμετροι κίνησης που προκύπτουν αποτελούν τη διόρθωση της κίνησης. Συνθέτοντας τις παραμέτρους της διόρθωσης με τις παραμέτρους της ευθυγράμμισης της στάσης προκύπτουν οι νέες παράμετροι ευθυγράμμισης της στάσης ως προς το ”μωσαϊκό”. Οι νέες παράμετροι περιγράφουν μια κίνηση που προσεγγίζει καλύτερα την πραγματική. Για την εφαρμογή όμως του αλγορίθμου απαραίτητοι είναι:

1. ο καθορισμός της συχνότητας με την οποία θα γίνεται η διόρθωση
2. και η επιλογή των εικόνων που θα λάβουν μέρος στη διαδικασία της διόρθωσης

Τοπική διόρθωση θα ήταν δυνατό να γίνεται για κάθε ζευγάρι διαδοχικών στάσεων μέσα στο ”μωσαϊκό”. Λόγω όμως της αύξησης του κόστους εκτέλεσης του αλγορίθμου, αποφεύγεται. Η συχνότητα διόρθωσης είναι αντιστρόφως ανάλογη της ποιότητας του ”μωσαϊκού”, αλλά όμως και του κόστους κατασκευής. Απαιτείται επομένως η επιλογή μιας ενδιάμεσης τιμής για τη συχνότητα της διόρθωσης.

Στην παρούσα εργασία προτείνεται η χρήση της τοπολογίας των στάσεων του στιγμιοτύπου για τον καθορισμό της συχνότητας διόρθωσης και την επιλογή των εικόνων που θα λάβουν μέρος σε αυτή. Η τοπολογία εκφράζει τη θέση της κάθε στάσης στο ”μωσαϊκό” και εξαρτάται από την κατεύθυνση της κίνησης της κάμερας. Για παράδειγμα, το σχήμα 4.9 απεικονίζει μια ακολουθία στάσεων στην οποία η κάμερα κινείται πρώτα αριστερά στο σκηνικό και στη συνέχεια επανέρχεται δεξιά λαμβάνοντας

μεγαλύτερο ποσοστό της πληροφορίας του σκηνικού συγκρινόμενο με το ποσοστό της πληροφορίας που είχε ήδη λάβει κατά την προηγούμενη κίνηση της. Μελετώντας



Σχήμα 4.9: Τοπολογία μιας ακολουθίας εικόνων

την τοπολογία ολόκληρης της ακολουθίας επιλέγεται και η συχνότητα διόρθωσης. Από πειραματικά αποτελέσματα έχει διαπιστωθεί ότι στις περιπτώσεις που η τοπολογία της ακολουθίας είναι παρόμοια με αυτή του σχήματος 4.9, όταν δηλαδή η κάμερα επανέρχεται σε περιοχές που έχει ήδη λάβει το σκηνικό τους, τα λάθη στην ευθυγράμμιση των στάσεων στο ”μωσαϊκό” εμφανίζονται έντονα. Οι στάσεις που προέρχονται κατά την επαναφορά της κάμερας τοποθετούνται σε λανθασμένη θέση στο ”μωσαϊκό” σε σχέση με τις ήδη προηγούμενα τοποθετημένες στάσεις δημιουργώντας έτσι λάθη ευθυγράμμισης. Επομένως στις ακολουθίες που η κάμερα εμφανίζεται να αλλάζει κατεύθυνση κίνησης προτείνεται η εφαρμογή του αλγορίθμου της διόρθωσης σε κάποια από τις στάσεις της νέας κατεύθυνσης.

Διόρθωση όμως απαιτείται και στις στάσεις που τοπολογικά είναι γειτονικές της στάσης αναφοράς. Όπως έχει ήδη ειπωθεί όλες οι στάσεις της ακολουθίας ευθυγραμμίζονται βάση της στάσης αναφοράς. Τα λάθη που εμφανίζονται στην ευθυγράμμιση των στάσεων μέσα στο ”μωσαϊκό” εκφράζουν ακριβώς την απόκλιση κάθε μετασχηματισμένης στάσης ως προς το σύστημα αναφοράς. Στην ιδανικότερη περίπτωση, στην κατασκευή δηλαδή ενός ορθού ”μωσαϊκού”, η απόκλιση αυτή θα έτεινε στο μηδέν. Επομένως απαραίτητη είναι η διόρθωση της ευθυγράμμισης μεταξύ της κάθε στάσης με τη στάση αναφοράς. Η διαφορετική όμως πληροφορία που περιέχει κάθε στάση συγκρινόμενη με αυτή της στάσης αναφοράς, κάνει αδύνατη τη διόρθωση της κακής ευθυγράμμισης. Διόρθωση υφίστανται μόνο οι γειτονικές στάσεις της στάσης αναφοράς. Οι γειτονικές στάσεις προκύπτουν από την τοπολογία της ακολουθίας και δεν είναι μόνο

οι χρονικά διαδοχικές ή κοντινές στάσεις. Στο παράδειγμα του σχήματος γειτονικές, για παράδειγμα, της στάσης Α1 θεωρούνται οι στάσεις Α13 και Α14 καθώς και η στάση Α2 ως διαδοχική της στάσης Α1. Όπως επίσης έχουμε αναφέρει το πρόβλημα των λανθασμένων ευθυγραμμίσεων οφείλεται στη συσσώρευση των λαθών που εμφανίζονται κατά την εκτίμηση της κίνησης μεταξύ των διαδοχικών στάσεων. Για τη μείωση της συσσώρευση λαθών εφαρμόζεται ο αλγόριθμος της διόρθωσης κάθε φορά που η κάμερα έχει μεταφερθεί κατά κάποια απόσταση Α. Στις περιπτώσεις που κάμερα κινείται σε μία κατεύθυνση ο αλγόριθμος εφαρμόζεται μόνο αφού η κάμερα μετακινηθεί κατά απόσταση Α.

Συνοπτικότερα η συχνότητα της διόρθωσης εξαρτάται από την τοπολογία της ακολουθίας. Συγκεκριμένα στην εργασία αυτή προτείνεται η εφαρμογή του αλγορίθμου της τοπικής διόρθωσης:

- στις στάσεις που αλλάζει κατεύθυνση η κίνηση της κάμερας
- στις στάσεις που είναι τοπολογικά γειτονικές της στάσης αναφοράς
- όταν η κάμερα έχει μεταφερθεί κατά Α pixels.

Βάση επίσης της τοπολογίας καθορίζεται το ζεύγος των στάσεων που θα πάρουν μέρος στον αλγόριθμο της διόρθωσης. Η μία από τις δύο στάσεις (Γ) του ζεύγους, καθορίζεται από τη συχνότητα της διόρθωσης. Η δεύτερη στάση (Δ) επιλέγεται βάση των ακόλουθων κριτηρίων:

- πρέπει να ανήκει στο σύνολο των γειτονικών στάσεων της στάσης Γ. Η γειτονιά καθορίζεται από την τοπολογία. Σαν γειτονιά ορίζεται μια τετραγωνική περιοχή διάστασης Κ γύρω από τη θέση της στάσης Γ στο "μωσαϊκό". Όλες οι στάσεις που τοποθετούνται στην περιοχή αυτή του "μωσαϊκού" είναι γειτονικές της στάσης Γ. Η στάση Δ επιλέγεται να ανήκει στη γειτονιά της Γ ώστε να περιέχουν και οι δύο ένα μεγάλο ποσοστό της ίδιας πληροφορίας για την ακριβέστερη διόρθωση της κίνησης μεταξύ τους.
- πρέπει να είναι η παλαιότερη χρονικά στάση της γειτονιάς της Γ. Η επιλογή της παλαιότερης οφείλεται στο γεγονός που απαιτεί τα λάθη συσσώρευσης ως τη στάση Δ να είναι τα ελάχιστα δυνατά.

Κεφάλαιο 5

Συγκρίσεις μεθόδων και πειραματικά αποτελέσματα

5.1 Εισαγωγή

Η κατασκευή της πανοραμικής όψης μιας σκηνής απαιτεί, όπως ήδη αναφέραμε, την εκτίμηση της κίνησης της κάμερας και τη σύνθεση των στάσεων της ακολουθίας που αναπαριστά τη σκηνή. Στην παρούσα εργασία προτάθηκαν διάφοροι μέθοδοι εκτίμησης της κίνησης καθεμιά από τις οποίες παρουσιάζει τόσο έναν αριθμό πλεονεκτημάτων όσο και μειονεκτημάτων. Οι μέθοδοι αυτοί ανήκουν στη γνωστή κατηγορία των παραμετρικών μεθόδων και μελετούν την κίνηση της κάμερας περιγράφοντας τη με τη χρήση κατάλληλων μοντέλων κίνησης. Τα γνωστότερα μοντέλα κίνησης παρουσιάστηκαν στο τρίτο κεφάλαιο. Οι μέθοδοι αυτοί καθώς και τα γνωστότερα παραμετρικά μοντέλα δοκιμάστηκαν σε αρκετές ακολουθίες εικόνων. Οι ακολουθίες αυτές περιείχαν είτε πραγματικά είτε συνθετικά δεδομένα. Η κάμερα κατά την διάρκεια αυτών των ακολουθιών παρουσίαζε ποικιλία τρισδιάστατων κινήσεων. Μελετώντας το σύνολο των παραμετρικών αποτελεσμάτων ένας αριθμός συμπερασμάτων προέκυψε για την απόδοση της κάθε μεθόδου. Στην ενότητα που ακολουθεί παρουσιάζονται τα μειονεκτήματα και πλεονεκτήματα της κάθε μεθόδου καθώς και ένα σύνολο ενδεικτικών πειραματικών αποτελεσμάτων.

5.2 Σύγκριση μεθόδων εκτίμησης της κίνησης της κάμερας

Οι μέθοδοι εκτίμησης της κίνησης που παρουσιάστηκαν στην εργασία χωρίζονται σε δύο βασικές κατηγορίες, στις μεθόδους εκτίμησης της κίνησης με χρήση του αλγορίθμου M-estimator και στις μεθόδους εκτίμησης της κίνησης με τη χρήση αναλλοίωτων σχέσεων ροπών. Η εφαρμογή του αλγορίθμου M-estimator καθώς και η διαδικασία κατασκευής των επιφανειών που συγκρίνονται για τον υπολογισμό της κίνησης της κάμερας επιτυγχάνεται με δύο μεθόδους, με τη μέθοδο ταιριάσματος περιοχών (μπλοκς) στις δύο εικόνες και με τη μέθοδο αντιστοίχισης χαρακτηριστικών που εξάγονται από τις εικόνες όπως για παράδειγμα οι γωνίες. Οι μέθοδοι εκτίμησης της κίνησης εφαρμόζονται σε κάθε διαδοχικό ζευγάρι εικόνων της ακολουθίας.

Συγκρίνοντας αρχικά τις μεθόδους εκτίμησης της οπτικής ροής συμπεραίνουμε ότι η αντιστοίχιση γωνιών είναι γενικά αποδοτικότερη μέθοδος υπό την προϋπόθεση ύπαρξης ανομοιομορφίας στην πληροφορία της σκηνής. Απαιτείται δηλαδή το σκηνικό να έχει κατάλληλη υφή ώστε να υπάρχουν σημεία που χαρακτηρίζονται σαν γωνίες. Η διαφορά των μεθόδων εκτίμησης της οπτικής ροής που χρησιμοποιήθηκαν στην παρούσα εργασία έγκειται στην μορφή της "οντότητας" που χρησιμοποιούν κατά την διαδικασία της αναζήτησης του αντίστοιχου χαρακτηριστικού. Για την μέθοδο ταιριάσματος περιοχών η "οντότητα" αυτή είναι ένα ορθογώνιο μπλοκ διάστασης 16x16 ενώ για την μέθοδο αντιστοίχισης γωνιών είναι ένα σημείο με συγκεκριμένες ιδιότητες. Παρατηρήθηκε ότι σε περιπτώσεις σημαντικής αλλαγής της απόστασης εστίασης της κάμερας, που είναι ορατή σαν σμύκρινση ή μεγέθυνση στην εικόνα, ή ακόμα γενικά σε περιπτώσεις έντονων κινήσεων της κάμερας ο αλγόριθμος ταιριάσματος περιοχών αδυνατεί να περιγράψει την κίνηση. Το γεγονός αυτό οφείλεται στο ότι το ίδιο μέγεθος μπλοκ που προκύπτει μετά την αλλαγή κλίμακας εμφανίζεται με διαφορετική πληροφορία, μικρότερη ή μεγαλύτερη, από την αρχική πληροφορία του μπλοκ εξαιτίας της μεγέθυνσης ή σμύκρινσης που υφίσταται η πληροφορία. Αντίθετα σε μικρές αλλαγές στην κίνηση της κάμερας και σε περιοχές με ανομοιομορφία ο αλγόριθμος ταιριάσματος περιοχών καταφέρνει επιτυχώς να υπολογίσει την κίνηση της κάμερας. Από την άλλη η μέθοδος αντιστοίχισης γωνιών καταφέρνει να αντιστοιχίσει σωστά τα χαρακτηριστικά σημεία. Σφάλματα στην περίπτωση αυτή εμφανίζονται λόγω της αποτυχίας εύρεσης της ακριβής θέσης των γωνιών στις δύο εικόνες. Κατά την αντιστοίχιση τότε των όμοιων γωνιών που στην πραγματικότητα δεν αντιστοιχούν στο ίδιο σημείο τα σφάλματα γίνονται ορατά και επηρεάζουν το τελικό αποτέλεσμα. Επιπλέον μειονέκτημα της μεθόδου αντιστοίχισης

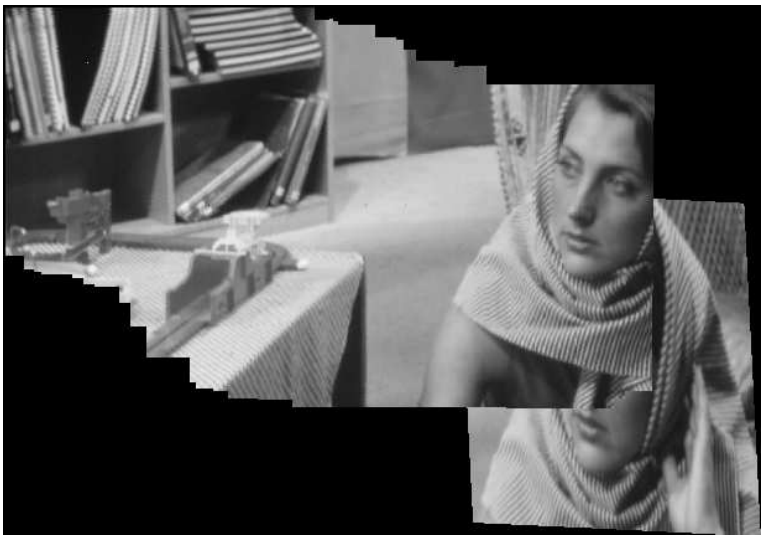
γωνιών είναι το μεγάλο επιπρόσθετο υπολογιστικό κόστος που απαιτεί για την εξαγωγή των χαρακτηριστικών σημείων των δύο εικόνων.

Για το σκοπό της σύγκρισης των μεθόδων εκτίμησης της οπτικής ροής χρησιμοποιήθηκε συνθετική ακολουθία 32 στάσεων. Η κάμερα στην ακολουθία των εικόνων εκτελεί εκτός από μεταφορική κίνηση στις δύο διαστάσεις, αλλαγή στην απόσταση εστίασης και κάποια πολύ μικρή περιστροφή γύρω από τον κατακόρυφο άξονα, σε ορισμένες μόνο εικόνες. Γενικά η κίνηση της χαρακτηρίζεται μικρή από εικόνα σε εικόνα. Στο τελευταίο ζευγάρι όμως της ακολουθίας εμφανίζεται μεγάλη αλλαγή στην κλίμακα της δεύτερη εικόνας (σχήμα 5.1 (b)) και μεταφορική κίνηση στην οριζόντια κατεύθυνση. Η μέθοδος ταιριάσματος περιοχών απέτυχε να εκτιμήσει την κίνηση, αντίθετα η εκτίμηση της μεθόδου αντιστοίχισης γωνιών ήταν επιτυχής. Η πανοραμική εικόνα, που κατασκευάστηκε χρησιμοποιώντας τη κίνηση που εκτιμήθηκε κατά τη μέθοδο ταιριάσματος περιοχών, παρουσιάζει έντονο σφάλμα στην ευθυγράμμιση των τελευταίων εικόνων και απεικονίζεται στο σχήμα 5.2. Αντίθετα το "μωσαϊκό" που κατασκευάζεται με χρήση της μεθόδου αντιστοίχισης γωνιών παριστάνεται στο σχήμα 5.3.

(a) (b)

Σχήμα 5.1: Τελευταίο ζευγάρι στάσεων της συνθετικής ακολουθίας Barbara1.
Στο ζευγάρι αυτό παρουσιάζεται εστίαση και μεταφορά της κάμερας.

Από το σύνολο των πραγματικών και συνθετικών ακολουθιών που εφαρμόσαμε τις μεθόδους αυτές συμπεραίνουμε ότι αποδοτικότερη εκ των δύο, σε ιδιαίτερα σημαντικές κινήσεις της κάμερας, είναι η μέθοδος αντιστοίχισης γωνιών. Σε ακολουθίες όμως μικρών κινήσεων η μέθοδος ταιριάσματος περιοχών είναι λιγότερο ευαίσθητη σε σφάλματα αφού η αντιστοίχιση των περιοχών επιτυγχάνεται με μεγάλη ακρίβεια της τάξης του μισού pixel. Αντίθετα η μέθοδος αντιστοίχισης γωνιών παρουσιάζεται αστοθής λόγω της αποτυχίας εντοπισμού της ακριβής θέσης των γωνιών. Η απόκλιση από την πραγματική θέση είναι της τάξης του pixel. Παράλληλα, εξαιτίας του μεγαλύτερου



Σχήμα 5.2: Το "μωσαϊκό" της συνθετικής ακολουθίας Barbara1 με τη μέθοδο ταιριάσματος περιοχών.

Η μέθοδος ταιριάσματος περιοχών χρησιμοποιείται για την εκτίμηση της οπτικής ροής. Οι παράμετροι του αφινικού μοντέλου που χρησιμοποιήθηκε για να περιγράψει την κίνηση της κάμερας στην ακολουθία εκτιμήθηκαν από τον αλγόριθμο M-estimator.

υπολογιστικού κόστους που απαιτεί, σε αντίθεση με τη μέθοδο ταιριάσματος περιοχών που παρουσιάζεται γρηγορότερη, η χρήση της περιορίζεται.

Συγκρίνοντας τον αλγόριθμο M-estimator και τη μέθοδο εκτίμησης της κίνησης με τη χρήση αναλλοίωτων σχέσεων αφινικών ροπών συμπεραίνουμε ότι η δεύτερη είναι καλύτερη αλλά περισσότερο ευαίσθητη σε λάθη εκτίμησης της οπτικής ροής. Επομένως τις περισσότερες φορές η χρήση του M-estimator είναι προτιμότερη. Η μέθοδος των ροπών εξαρτάται σημαντικά από την αντιστοίχιση των επιφανειών στις δύο εικόνες. Στην περίπτωση λανθασμένης αντιστοίχισης των επιφανειών η μέθοδος αποτυγχάνει να υπολογίσει την κίνηση της κάμερας μεταξύ δύο εικόνων. Αντίθετα ο αλγόριθμος M-estimator εκτιμά την κίνηση της κάμερας ακόμα και με λάθη κακής αντιστοίχισης των χαρακτηριστικών. Η σθεναρότητα του αλγορίθμου M-estimator στα λάθη αυξάνεται με τη χρήση κατάλληλης συνάρτησης κόστους. Αποδοτικότερη συνάρτηση αποδείχτηκε η Geman & Mc-Clure αφού από τη μια παρουσιάζει ανεχτικότητα σε λανθασμένες εκτιμήσεις και από την άλλη απορρίπτει τις εκτιμήσεις που δεν συμφωνούν εξ' ολοκλήρου με την κίνηση της κάμερας. Ελέγχοντας το υπολογιστικό κόστος της καθεμιάς παρατηρείται ότι η μέθοδος των ροπών είναι γρηγορότερη όταν η αντιστοιχία των επιφανειών είναι γνωστή. Εξαιτίας των επαναλήψεων ο αλγόριθμος M-estimator παρουσιάζει μεγαλύτερο υπολογιστικό κόστος.

Σχήμα 5.3: Το ”μωσαϊκό” της συνθετικής ακολουθίας Barbara1 με τη μέθοδο αντιστοίχισης γωνιών.

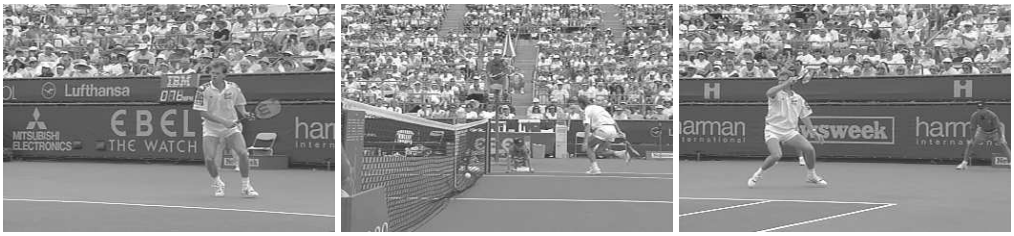
Η μέθοδος αυτή χρησιμοποιείται για την εκτίμηση της οπτικής ροής. Ο αλγόριθμος M-estimator εφαρμόστηκε για την εκτίμηση των παραμέτρων του αφινικού μοντέλου κίνησης που περιέγραψε την ακολουθία.

Οι μέθοδοι εφαρμόστηκαν στην ακολουθία του Stefan που αποτελείται από 300 στάσεις. Στην ακολουθία αυτή η κάμερα συνδυάζει διάφορες μορφές κινήσεων. Σημαντικότερη εμφανίζεται η μεταφορική κίνηση στη οριζόντια κατεύθυνση ενώ παράλληλα εκτελεί μικρή μεταφορική κίνηση και προς την κατακόρυφη κατεύθυνση. Στις τελευταίες στάσεις της ακολουθίας η μεταφορική κίνηση στην οριζόντια κατεύθυνση παρουσιάζεται ιδιαίτερη σημαντική αφού ξεπερνά την διάσταση των 20 pixels. Χαρακτηριστική είναι η αλλαγή στην απόσταση εστίασης της κάμερας που εμφανίζεται κατά μήκος της ακολουθίας, η αλλαγή αυτή μεταφράζεται σε μη-ισοτροπική αλλαγή της κλίμακας στις δύο διαστάσεις των στάσεων της ακολουθίας. Επίσης σε ένα υποσύνολο στάσεων η κάμερα εκτελεί περιστροφική κίνηση γύρω από τον κατακόρυφο άξονα Δυσκολία στην εκτίμηση της κίνησης παρουσιάζουν τα τρία διαφορετικά επίπεδα της σκηνής καθώς και η ομοιομορφία που παρουσιάζει το επίπεδο του γηπέδου. Χαρακτηριστικές στάσεις της ακολουθίας εμφανίζονται στο σχήμα 5.4. Οι μέθοδοι εκτίμησης της κίνησης που αναφέραμε αδυνατούν να υπολογίσουν την κίνηση της κάμερας στην ακολουθία αυτή. Το γεγονός αυτό οφείλεται στην περίπλοκη μορφή του σκηνικού. Πληροφορία για την εκτίμηση της κίνησης παρουσιάζεται στο επίπεδο της κερκίδας και στις διαφημιστικές πινακίδες που ορίζουν διαφορετικό επίπεδο κίνησης. Η εφαρμογή των αλγορίθμων υπολογισμού της οπτικής ροής που χρησιμοποιούνται και από τις δύο μεθόδους εκτίμησης της κίνησης θεωρούν inliers τα σημεία ή τις περιοχές που αντιστοιχούν κυρίως στην κερ-

κίδα του σκηνικού. Επομένως οι παράμετροι του μοντέλου κίνησης που χρησιμοποιείται, περιγράφουν ουσιαστικά την κίνηση της κάμερας όπως απεικονίζεται στο επίπεδο της κερκίδας και μόνο. Αποτέλεσμα αυτού, είναι η κατασκευή της πανοραμικής όψης του γηπέδου με αρκετά λάθη ευθυγράμμισης των στάσεων. Το "μωσαϊκό", που κατασκευάστηκε εφαρμόζοντας τον αλγόριθμο που βασίζεται στις αναλλοίωτες σχέσεις ροπών για την εκτίμηση της κίνησης, αναπαρίσταται στο σχήμα 5.8. Η οπτική ροή των σημείων εκτιμήθηκε με την μέθοδο ταιριάσματος περιοχών. Οι αντίστοιχες επιφάνειες που ο αλγόριθμος επέλεξε στο σύνολο των στάσεων καλύπτουν μόνο το επίπεδο της κερκίδας. Επομένως η κίνηση που εκτιμάται αποκλίνει από την πραγματική αφού δεν εκφράζει την κίνηση στο σύνολο της σκηνής αλλά σε ένα μόνο επίπεδο της. Η τμηματοποίηση της σκηνής στα επίπεδα που αποτελείται και η εκτίμηση της κίνησης σε κάθε επίπεδο χωριστά ίσως υπολογίσει καλύτερα την κίνηση σε ολόκληρη την ακολουθία. Η προσέγγιση αυτή αποτελεί μελλοντική επέκταση της εργασίας. Αντίθετα όταν οι αντίστοιχες επιφάνειες που χρησιμοποιεί ο αλγόριθμος των ροπών καθοριστούν από τον χρήστη και ληφθούν ώστε να καλύπτουν ένα μεγάλο μέρος της σκηνής, τα αποτελέσματα είναι σαφώς καλύτερα. Στο σχήμα 5.5 απεικονίζονται δύο στάσεις της ακολουθίας. Η κίνηση της κάμερας στις στάσεις αυτές είναι μεταφορική προς την οριζόντια κατεύθυνση και επιπλέον μεταβάλλεται η απόσταση εστίασης της. Οι επιφάνειες που επέλεξε ο χρήστης για την εκτίμηση της κίνησης φαίνονται στο σχήμα 5.6. Η κίνηση που ανιχνεύτηκε με τη μέθοδο των ροπών είναι:

$$A_{2x2}=\begin{bmatrix} 0.960 & 0.00 \\ 0.00 & 0.966 \end{bmatrix} \quad B_{2x1}=\begin{bmatrix} 36.2 \\ 0.6 \end{bmatrix} \tag{5.1}$$

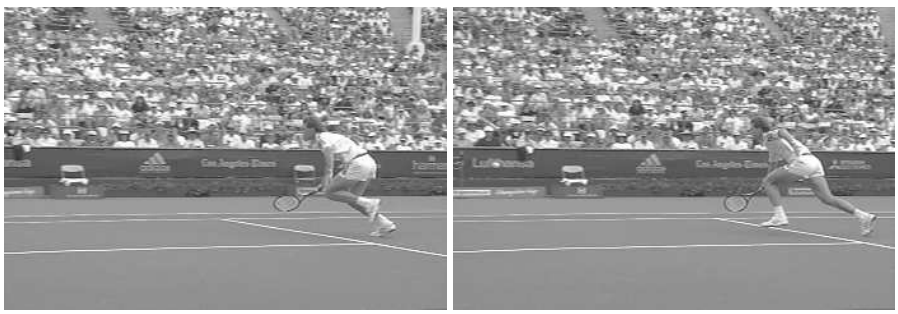
Στην παραπάνω σχέση ο πίνακας A_{2x2} εκφράζει την αλλαγή στην απόσταση εστίασης της κάμερας και την πιθανή περιστροφή της ενώ ο πίνακας B_{2x1} περιγράφει τη μεταφορική κίνηση της κάμερας, που στο παράδειγμα αυτό στην οριζόντια κατεύθυνση είναι ίση με 36.2 pixels. Το "μωσαϊκό" που κατασκευάζεται από την συγκεκριμένη εκτίμηση φαίνεται στο σχήμα 5.7. Η μέθοδος εκτίμησης της κίνησης με τη χρήση αναλλοίωτων σχέσεων αφινικών ροπών εφαρμόστηκε και στη συνθετική ακολουθία Barbara1. Στην ακολουθία αυτή που παρουσιάζεται κυρίως μεταφορική κίνηση και αλλαγή στην απόσταση εστίασης της κάμερας η κίνηση εκτιμήθηκε σωστά. Τα περιγράμματα που εξάγονται καλύπτουν το μεγαλύτερο μέρος της εικόνας, επομένως δεν εμφανίζονται λάθη εκτίμησης της κίνησης όπως στην περίπτωση της ακολουθίας Stefan. Η οπτική ροή υπολογίστηκε με τη μέθοδο ταιριάσματος περιοχών. Το "μωσαϊκό" που κατασκευάστηκε από τις εκτιμήσεις αυτές των παραμέτρων κίνησης φαίνεται στο σχήμα 5.9



(a) (b) (c)

Σχήμα 5.4: Τρεις στάσεις από την ακολουθία του Stefan.

Η στάση (a) είναι η πρώτη στάση της ακολουθίας, η (b) στάση αναπαριστά την πληροφορία που απεικονίζεται στο δεξί άκρο του ”μωσαϊκού”, ενώ η (c) αναπαριστά την πληροφορία του αριστερότερου άκρου του ”μωσαϊκού” στην ακολουθία του Stefan.



(a) (b)

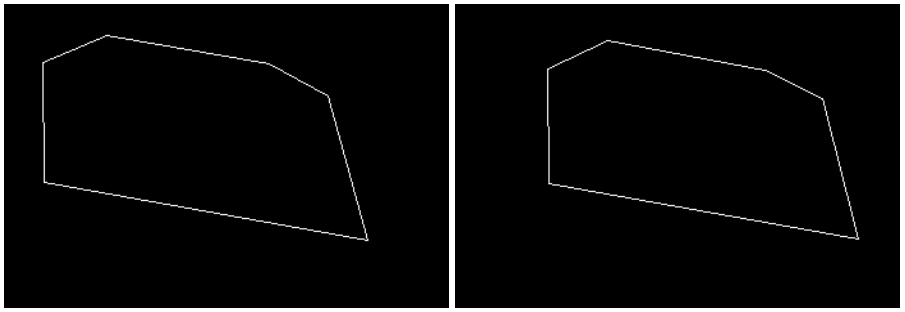
Σχήμα 5.5: Δύο στάσεις της ακολουθίας του Stefan (Stefan260-Stefan262)

Στις στάσεις αυτές παρουσιάζεται έντονη η μεταφορική κίνηση της κάμερας

Η χρήση των ροπών για την εκτίμηση των παραμέτρων μεγαλύτερων μοντέλων κίνησης όπως του προβολικού δεν είναι εφικτή. Η μορφή της συνάρτησης των ροπών δεν διευκολύνει τον ορισμό αναλλοίωτων για άλλους πολυπλοκότερους μετασχηματισμούς. Επομένως οι σχέσεις των ροπών αποτελούν αναλλοίωτες μόνο για αφινικούς μετασχηματισμούς αφού δεν μπορούν να εφαρμοστούν σε άλλους μετασχηματισμούς. Οι πιο συχνά χρησιμοποιούμενες αναλλοίωτες για προβολικούς μετασχηματισμούς είναι λόγοι ”cross ratios”.

5.3 Σύγκριση παραμετρικών μοντέλων κίνησης

Στο τρίτο κεφάλαιο παρουσιάστηκε το σύνολο των γνωστότερων παραμετρικών μοντέλων κίνησης. Από το σύνολο των μοντέλων αυτών στην παρούσα εργασία χρησι-

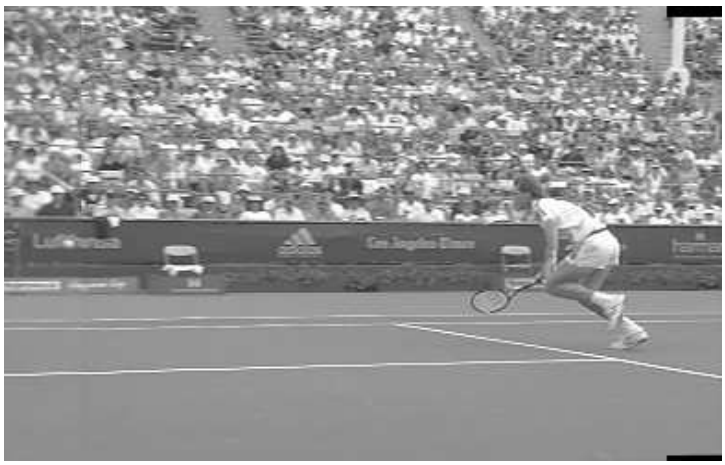


(a) (b)

Σχήμα 5.6: Τα αντίστοιχα περιβλήματα δύο στάσεων της ακολουθίας Stefan (Stefan260-Stefan262)

Τα περιβλήματα αυτά επελέγησαν από τον χρήστη για την εκτίμηση της κίνησης με την μέθοδο των ροπών.

μοποιήθηκαν το μοντέλο εστίασης και μεταφοράς, το αφινικό και το προβολικό μοντέλο. Διάκριση στην ποιότητα των μοντέλων αυτών κίνησης δεν πραγματοποιείται. Το κάθε μοντέλο κίνησης περιγράφει συνολικά μια διαφορετική μορφή κίνησης της κάμερας. Το μοντέλο εστίασης και μεταφοράς περιγράφει ουσιαστικά μεταφορική κίνηση στις τρεις διαστάσεις. Η μεταφορική κίνηση στην κατεύθυνση του άξονα Z εκφράζεται ως αλλαγή στην απόσταση εστίασης της κάμερας. Ο περιορισμός του μοντέλου αυτού είναι ότι περιγράφει μόνο ισοτροπική αλλαγή κλίμακας. Αντίθετα το αφινικό μοντέλο κίνησης περιγράφει ακριβώς ότι και το μοντέλο εστίασης και μεταφοράς χωρίς τον περιορισμό όμως της ισοτροπικής αλλαγής της κλίμακας και επιπλέον την περιστροφή της κάμερας γύρω από τον άξονα Z. Το επόμενο σε πολυπλοκότητα μοντέλο κίνησης, το προβολικό, εκφράζει το σύνολο των τρισδιάστατων κινήσεων της κάμερας υπό την προϋπόθεση ότι τα αντικείμενα του σκηνικού βρίσκονται στο ίδιο βάθος. Τα μοντέλα αυτά χρησιμοποιήθηκαν σε αρκετές ακολουθίες εικόνων συνθετικές και πραγματικές για την περιγραφή της κίνησης της κάμερας σε αυτές. Από το σύνολο των πειραματικών αποτελεσμάτων καταλήξαμε στην προτίμηση του απλούστερου κάθε φορά μοντέλου κίνησης που είναι ικανό να περιγράψει την κίνηση της κάμερας. Η προτίμηση του μεγαλύτερου μοντέλου κίνησης, για παράδειγμα του προβολικού, δεν αποδεικνύεται πάντα ορθή. Στις περιπτώσεις που η κίνηση της κάμερας περιγράφεται ικανοποιητικά από ένα μικρότερο μοντέλο, η προτίμηση μας για ένα μεγαλύτερο από τη μια αυξάνει το υπολογιστικό κόστος εφαρμογής του αλγορίθμου αφού απαιτείται η εκτίμηση μεγαλύτερου αριθμού παραμέτρων και από την άλλη τα μικρά σφάλματα του αλγορίθμου ή ακόμα και η πιθανή ύπαρξη του θορύβου είναι δυνατό να παρουσιάζουν κίνηση που στην



Σχήμα 5.7: Το ”μωσαϊκό” δύο στάσεων της ακολουθίας του Stefan (Stefan260-Stefan262)

Το ”μωσαϊκό” κατασκευάστηκε χρησιμοποιώντας την εκτίμηση της κίνησης που υπολογίστηκε από τη μέθοδο των ροπών.

πραγματικότητα δεν υπάρχει. Σφάλματα του αλγορίθμου είναι για παράδειγμα είτε εξ’ ολοκλήρου κακές αντιστοιχίσεις χαρακτηριστικών ή μπλοκς είτε ακόμα και μικρά λάθη στον εντοπισμό της ακριβής θέσης του αντίστοιχου χαρακτηριστικού. Τα σφάλματα αυτά ακόμα και όταν είναι πολύ μικρά επηρεάζουν την εκτίμηση των παραμέτρων του εκάστοτε μοντέλου κίνησης. Πολλές φορές εμφανίζονται παράμετροι που περιγράφουν διαφορετικό είδος κίνησης με μικρές τιμές.

Η σύγκριση του αφινικού και προβολικού μοντέλου επιτεύχθηκε με τη χρήση συνθετικής ακολουθίας, 26 στάσεων. Η ακολουθία αυτή κατασκευάστηκε ώστε να περιγράφει διάφορες μορφές κινήσεων, περιστροφή ως προς τους άξονες, μεταφορική κίνηση και αλλαγή στην εστίαση της κάμερας. Πολλές από τις κινήσεις που απεικονίστηκαν στις στάσεις της ακολουθίας δεν περιγράφονται από το αφινικό μοντέλο. Αντίθετα όμως το προβολικό ως πολυπλοκότερο μοντέλο διαθέτει παραμέτρους που εκφράζουν ακριβώς το σύνολο αυτών των κινήσεων. Στην περίπτωση αυτή το προβολικό μοντέλο απέδωσε ικανοποιητικά αποτελέσματα αφού η κίνηση της ακολουθίας είναι ιδιαίτερα περίπλοκη. Ιδιαίτερο ενδιαφέρον παρουσιάζει το τελευταίο ζευγάρι εικόνων της ακολουθίας που απεικονίζεται στο σχήμα 5.10. Η κίνηση της κάμερας στην περίπτωση αυτή είναι μεταφορική καθώς και περιστροφική ως προς τον κατακόρυφο και οριζόντιο άξονα. Ο αλγόριθμος M-estimator χρησιμοποιήθηκε για την εκτίμηση της κίνησης σε όλη την ακολουθία των εικόνων και για τα δύο μοντέλα κίνησης. Η οπτική ροή εκτιμήθηκε με τη χρήση της μεθόδου αντιστοίχισης γωνιών, ενώ η συνάρτηση κόστους στην εφαρμογή του M-estimator ήταν η Geman & Mc-Clure. Η παράμετρος κλίμακας αντιστοιχεί στην

ενδιάμεση τιμή των υπολοίπων (residuals), ενώ σε κάθε επανάληψη του αλγορίθμου απορρίπτονται τα "αντίστοιχα χαρακτηριστικά" στα οποία η τιμή του κριτηρίου ελαχιστοποίησης παρουσιάζεται μεγαλύτερη κατά διπλάσια τιμή της παραμέτρου κλίμακας. Θεωρήθηκε αποτελεσματική η αφαίρεση αυτή των χαρακτηριστικών παρά τη σθεναρότητα του αλγορίθμου. Το κριτήριο ελαχιστοποίησης που χρησιμοποιήθηκε εκφράζεται από την εξίσωση 3.37. Ορίζεται ως η διαφορά στη τιμή αντιστοίχισης που καθορίζει το μοντέλο από την αντιστοίχιση που εκτιμήθηκε κατά τη διαδικασία υπολογισμού της οπτικής ροής. Το "μωσαϊκό" που κατασκευάστηκε με τη χρήση του αφινικού μοντέλου κίνησης απεικονίζεται στο σχήμα 5.11. Ορατές είναι οι κακές ευθυγραμμίσεις που περιέχει καθώς και η αποτυχία εκτίμησης της κίνησης μεταξύ του τελευταίου ζεύγους. Αντίθετα το "μωσαϊκό" που δημιουργήθηκε με τη χρήση του προβολικού μοντέλου, που αποτελεί γενικότερο και καταλληλότερο μοντέλο για την περιγραφή αυτών των κινήσεων, είναι ικανοποιητικό και παριστάνεται στο σχήμα 5.12.

Το μοντέλο εστίασης και μεταφοράς χρησιμοποιήθηκε για να περιγράψει απλούστερες κινήσεις της κάμερας. Η ακολουθία almaden αποτελεί παράδειγμα μεγάλης ακολουθίας στην οποία η κίνηση της κάμερας είναι μεταφορική κυρίως προς την οριζόντια κατεύθυνση ενώ παράλληλα εμφανίζεται και μικρή αλλαγή στην εστίαση της κάμερας. Εξαιτίας της ομοιόμορφης πληροφορίας που απεικονίζεται σε αυτή, ο αλγόριθμος εξαγωγής χαρακτηριστικών αποτυγχάνει στην εξαγωγή ενός σημαντικού αριθμού γωνιών. Επομένως για το σκοπό αυτό χρησιμοποιείται ο αλγόριθμος M-estimator με χρήση της μεθόδου ταιριάσματος περιοχών για την εκτίμηση της οπτικής ροής. Η ακολουθία αυτή αποτελείται από ένα μεγάλο αριθμό στάσεων (1000), γεγονός που κάνει ευαίσθητη την κατασκευή του "μωσαϊκού" αφού μικρά λάθη στην εκτίμηση της κίνησης μεταξύ των διαδοχικών στάσεων συσσωρεύονται και εμφανίζονται λάθη κακής ευθυγράμμισης στο τελικό "μωσαϊκό". Στην ακολουθία αυτή η σύνθεση των στάσεων επιτεύχθηκε με τη χρήση ευθυγράμμισης μωσαϊκού-στάσης και το "μωσαϊκό" που κατασκευάστηκε απεικονίζεται στο σχήμα 5.14, ενώ η πρώτη και τελευταία στάση της ακολουθίας παριστάνεται στο σχήμα 5.13. Το "μωσαϊκό" με τη χρήση της ευθυγράμμισης στάσης-στάσης ήταν επίσης ικανοποιητικό.

Το αφινικό μοντέλο κίνησης χρησιμοποιήθηκε για την εκτίμηση της κίνησης στην ακολουθία princeton που αποτελείται από 350 στάσεις. Το αφινικό μοντέλο περιγράφει ακριβώς το σύνολο των κινήσεων της κάμερας στη συγκεκριμένη ακολουθία. Η κάμερα εκτελεί μεταφορική κίνηση, αλλάζει το μήκος εστίασης της και περιστρέφεται γύρω από τον τον κατακόρυφο άξονα. Το "μωσαϊκό" κατασκευάστηκε με τη χρήση

ευθυγράμμισης στάσης-στάσης. Χρησιμοποιήθηκε ο αλγόριθμος M-estimator, με τη μέθοδο αντιστοίχισης γωνιών για την εκτίμηση της οπτικής ροής. Επίσης λόγω της απλής κίνησης που παρουσιάζεται επιλέχθηκε σαν συνάρτηση κόστους η τετραγωνική συνάρτηση με κατώφλι που αναφέρθηκε στο τρίτο κεφάλαιο. Το "μωσαϊκό" που κατασκευάστηκε απεικονίζεται στο σχήμα 5.16, ενώ η πρώτη και τελευταία στάση της ακολουθίας παριστάνονται στο σχήμα 5.15. Εφαρμόζοντας το προβολικό μοντέλο κίνησης στη συγκεκριμένη ακολουθία και με τη χρήση του αλγορίθμου M-estimator παρατηρήθηκαν μικρά σφάλματα στις παραμέτρους σε ορισμένες στάσεις. Ακόμα όμως και τα πολύ μικρά λάθη όπως έχουμε ήδη τονίσει μπορεί να μεγαλώσουν σε σημαντικό βαθμό κατά την κατασκευή του "μωσαϊκού". Στην περίπτωση αυτή η συσσώρευση των σφαλμάτων δεν ήταν ιδιαίτερα σημαντική.

Παράδειγμα εκτίμησης της κίνησης με τη χρήση του προβολικού μοντέλου παρουσιάζεται στην ακολουθία yosemite. Η κίνηση της κάμερας είναι μεταφορική ενώ σε ορισμένα σημεία αλλάζει η εστίαση της. Η κάμερα επιπλέον περιστρέφεται γύρω από τον κατακόρυφο άξονα. Η ακολουθία αυτή αποτελείται από 1000 στάσεις. Για την εκτίμηση της κίνησης χρησιμοποιήθηκε ο αλγόριθμος M-estimator, με τη χρήση της μεθόδου αντιστοίχισης γωνιών για την εκτίμηση της οπτικής ροής. Χρησιμοποιήθηκε επίσης η συνάρτηση κόστους Geman & McClure. Λόγω της απλότητας της κίνησης και της αποτελεσματικότητας του μοντέλου, η κίνηση δεν εκτιμήθηκε για κάθε διαδοχικό ζευγάρι στάσεων αλλά ανά τρείς στάσεις. Το "μωσαϊκό" που κατασκευάστηκε φαίνεται στο σχήμα 5.18, ενώ η πρώτη και τελευταία στάση της ακολουθίας απεικονίζονται στο σχήμα 5.17.

Πολυπλοκότερη χαρακτηρίζεται η ακολουθία Stefan (350 στάσεις) που ήδη αναφέραμε. Ο αλγόριθμος M-estimator εφαρμόστηκε στο αφινικό και στο προβολικό μοντέλο κίνησης. Χρησιμοποιήθηκε επίσης, ο αλγόριθμος αντιστοίχισης γωνιών για τον υπολογισμό της οπτικής ροής και η συνάρτηση επιρροής Geman & McClure. Το "μωσαϊκό" που κατασκευάστηκε με χρήση του αφινικού μοντέλου κίνησης φαίνεται στο σχήμα 5.19, ενώ το "μωσαϊκό" με χρήση του προβολικού μοντέλου απεικονίζεται στο σχήμα 5.20. Στο σημείο αυτό πρέπει να σημειωθεί ότι η κατά την κατασκευή του "μωσαικού" με τη χρήση του προβολικού μοντέλου, εφαρμόστηκε robust τεχνική και μηδενίστηκαν οι παράμετροι κίνησης με πολύ μικρές τιμές. Ο λόγος αυτής της παρέμβασης έγινε για την μείωση των συσσωρευτικών σφαλμάτων που εμφανίζονται στη κατασκευή του "μωσαϊκού". Τα σφάλματα αυτά προέρχονται από την κακή επιλογή σημείων για την εκτίμηση της κίνησης. Τα σημεία που επιλέγονται από τον αλγόριθμο είναι σημεία που

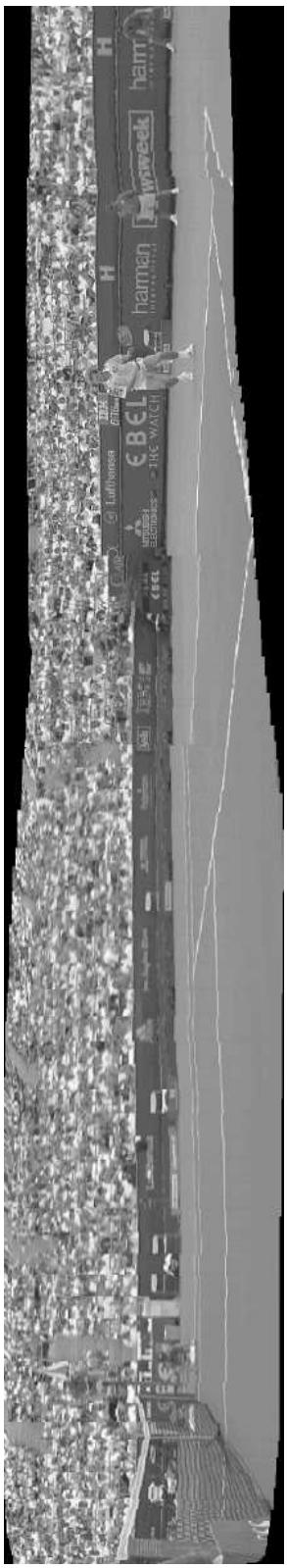
προέρχονται μόνο από το ένα κυρίως επίπεδο της σκηνής. Το γεγονός αυτό οφείλεται στη πληροφορία του σκηνικού που παρουσιάζεται ομοιόμορφη στο επίπεδο του γηπέδου και με ιδιαίτερη υφή στην κερκίδα. Από τη μορφή των "μωσαϊκών" παρατηρείται ότι παρόλη τη πολυπλοκότητα της σκηνής το προβολικό μοντέλο καταφέρνει να απεικονίσει το σκηνικό αντίθετα το αφινικό παρουσιάζει μεγαλύτερα σφάλματα εκτίμησης της κίνησης. Παρατηρώντας το "μωσαϊκό" που κατασκευάστηκε με τη χρήση του αφινικού μοντέλου συμπεραίνουμε ότι εκτιμάται κλίση της κάμερας στις τελευταίες στάσεις της ακολουθίας. Για την κατασκευή των "μωσαϊκών" οι στάσεις της ακολουθίας ευθυγραμμίστηκαν στο σύστημα συντεταγμένων της πρώτης στάσης. Το πρόβλημα των κακών ευθυγραμμίσεων που παρουσιάζεται στο "μωσαϊκό" που κατασκευάστηκε με τη χρήση του προβολικού μοντέλου κίνησης οφείλεται όπως ήδη αναφέραμε στη συσσώρευση μικρών λαθών της εκτίμησης της κίνησης. Τα "μωσαϊκά" που κατασκευάστηκαν με μικρότερο αριθμό στάσεων ήταν ικανοποιητικά.

Για τη διόρθωση των κακών ευθυγραμμίσεων χρησιμοποιήθηκε η μέθοδος διόρθωσης με τη χρήση τοπολογίας. Η μέθοδος της διόρθωσης εφαρμόστηκε στο αποτέλεσμα του προβολικού μετασχηματισμού και πραγματοποιήθηκε με τη χρήση του αλγορίθμου M-estimator πάνω στις μετασχηματισμένες στάσεις της ακολουθίας. Η οπτική ροή υπολογίστηκε με τη μέθοδο αντιστοίχισης γωνιών. Η διόρθωση έλαβε μέρος στις περιπτώσεις αλλαγής κατεύθυνσης της κίνησης και μετά από μεταφορά της κάμερας ανά 40 pixels. Γειτονική θεωρήθηκε η στάση που το κέντρο της απήχε απόσταση 40 pixels από το κέντρο της τρέχουσας στάσης. Το "μωσαϊκό" που κατασκευάστηκε από τη μέθοδο αυτή απεικονίζεται στο σχήμα 5.21. Παρατηρώντας το τελικό "μωσαϊκό" συμπεραίνουμε ότι διορθώθηκε το μεγαλύτερο μέρος κακών εκτιμήσεων της κίνησης.

Στο σύνολο των αποτελεσμάτων παρουσιάστηκαν ποικίλες μορφές κινήσεων της κάμερας, μεταφορική στις δύο διαστάσεις, μη-ισοτροπική εστίαση επίσης στις δύο διαστάσεις, περιστροφή ως προς τον κατακόρυφο άξονα και μικρή κλίση της κάμερας. Οι μέθοδοι εκτίμησης της κίνησης που χρησιμοποιήθηκαν εμφανίστηκαν αποδοτικές στον υπολογισμό αυτών των κινήσεων. Προϋπόθεση της ορθής αυτής εκτίμησης είναι ο σωστός υπολογισμός της οπτικής ροής και η χρήση του καταλληλότερου μοντέλου κίνησης. Απαιτείται επίσης το σύνολο των σημείων που απομένουν από τη διαδικασία εκτίμησης της οπτικής ροής να καλύπτουν το μεγαλύτερο μέρος της σκηνής ώστε η τελική κίνηση που εκτιμάται να προσεγγίζει την πραγματική.

5.4 Υπολογιστικό κόστος

Το υπολογιστικό κόστος της διαδικασίας κατασκευής του ”μωσαϊκού” είναι σημαντικό. Αιτία αυτού είναι ο μεγάλος αριθμός στάσεων που περιέχουν τα στιγμιότυπα μιας ακολουθίας που αναπαριστάνονται στο ”μωσαϊκό”. Η διαδικασία κατασκευής του ”μωσαϊκού” όπως ήδη αναφέραμε χωρίζεται σε δύο στάδια, στην εκτίμηση της κίνησης και στη σύνθεση των στάσεων. Οι αλγόριθμοι εκτίμησης της κίνησης εκτελέστηκαν σε μηχάνημα Sun sparc Ultra-Enterprise, με λειτουργικό SunOS 5.6. Ο χρόνος εκτέλεσης του αλγορίθμου M-estimator που χρησιμοποιεί τη μέθοδο ταιρίασματος περιοχών για την εκτίμηση της οπτικής ροής ήταν 19.24 δευτερόλεπτα με κατανάλωση 20% της CPU. Στον αλγόριθμο αυτό χρησιμοποιήθηκε η μέθοδος των ελαχίστων τετραγώνων. Ο χρόνος εκτέλεσης του ίδιου αλγορίθμου με τη χρήση της αντιστοίχισης γωνιών για την εκτίμηση της οπτικής ροής και τη χρήση της συνάρτησης κόστους Geman & Mc-Clure ήταν 170 δευτερόλεπτα ενώ κατανάλωσε το 23% της CPU. Το μεγάλο υπολογιστικό κόστος της προσέγγισης αυτής οφείλεται στην εξαγωγή των χαρακτηριστικών σημείων και στην χρήση της επαναληπτικής μεθόδου Marquardt-Levemberg για την επίλυση του μη-γραμμικού συστήματος που παρουσιάζεται. Ο χρόνος εκτέλεσης της μεθόδου των ροπών με τη χρήση της μεθόδου ταιριάσματος περιοχών για την εκτίμηση της οπτικής ροής ήταν 6.22 δευτερόλεπτα καταναλώνοντας 9% CPU. Ενώ ο χρόνος εκτέλεσης της ίδιας μεθόδου εκτίμησης της κίνησης με τη χρήση της αντιστοίχισης γωνιών, για την εκτίμηση της οπτικής ροής, ήταν 130.7 δευτερόλεπτα καταναλώνοντας 23%CPU. Ο αλγόριθμος σύνθεσης των στάσεων εκτελέστηκε στο ίδιο μηχάνημα μεταξύ δύο στάσεων μιας ακολουθίας. Ο χρόνος εκτέλεσής του ήταν 1.33 δευτερόλεπτα καταναλώνοντας 2.5% CPU. Ο συνολικός χρόνος εκτέλεσης για την κατασκευή του ”μωσαϊκού” εξαρτάται από τον αριθμό των στάσεων της ακολουθίας.



Σχήμα 5.8: Το ”μωσαϊκό” όλης της ακολουθίας του Stefan.

Η εκτίμηση της κίνησης πραγματοποιήθηκε με τη μέθοδο αναλλοίωτων σχέσεων ροπών

Σχήμα 5.9: Το ”μωσαϊκό” της συνθετικής ακολουθίας Barbara1

Η εκτίμηση της κίνησης έγινε με τη χρήση της μεθόδου των ροπών.

(a) (b)

Σχήμα 5.10: Δύο στάσεις της συνθετικής ακολουθίας Barbara2

Στις στάσεις αυτές εμφανίζεται έντονη η περιστροφή της κάμερας ως προς τον κατακόρυφο και οριζόντιο άξονα

Σχήμα 5.11: Το ”μωσαϊκό” της συνθετικής ακολουθίας Barbara2 όπως κατασκευάστηκε με χρήση του αφινικού μοντέλου κίνησης

Σχήμα 5.12: Το ”μωσαϊκό” της συνθετικής ακολουθίας Barbara2 όπως κατασκευάστηκε με χρήση του προβολικού μοντέλου κίνησης

(a) (b)

Σχήμα 5.13: Η πρώτη (a) και τελευταία (b) στάση της ακολουθίας almaden

Σχήμα 5.14: Το ”μωσαϊκό” της ακολουθίας almaden

Το ”μωσαϊκό” κατασκευάστηκε με χρήση του μοντέλου εστίασης και μεταφοράς

(a) (b)

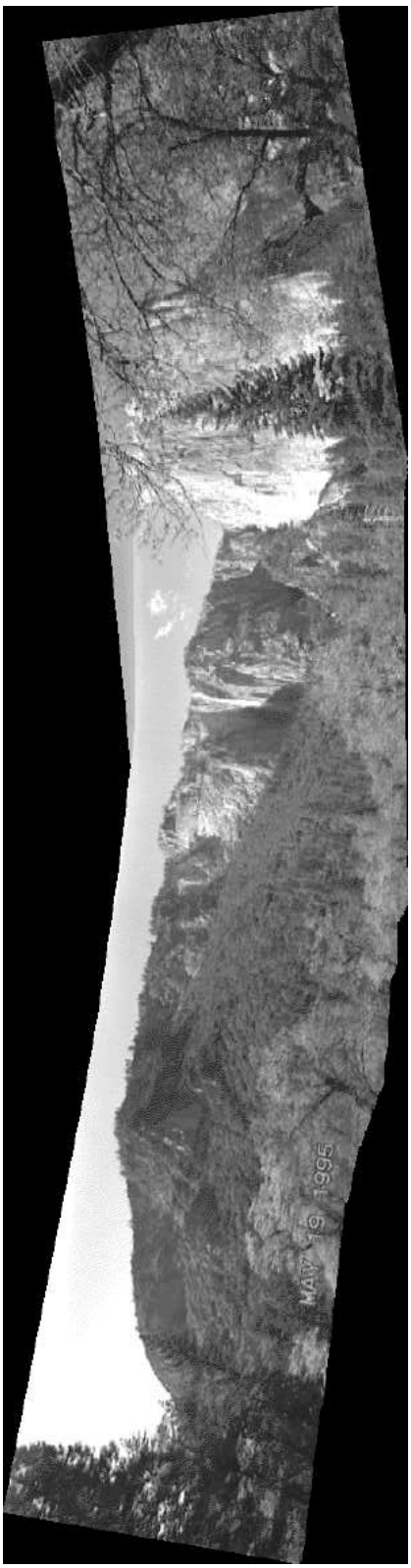
Σχήμα 5.15: Η πρώτη (a) και τελευταία (b) στάση της ακολουθίας princeton

Σχήμα 5.16: Το ”μωσαϊκό” της ακολουθίας princeton

Το ”μωσαϊκό” κατασκευάστηκε με τη χρήση του αφινικού μοντέλου κίνησης

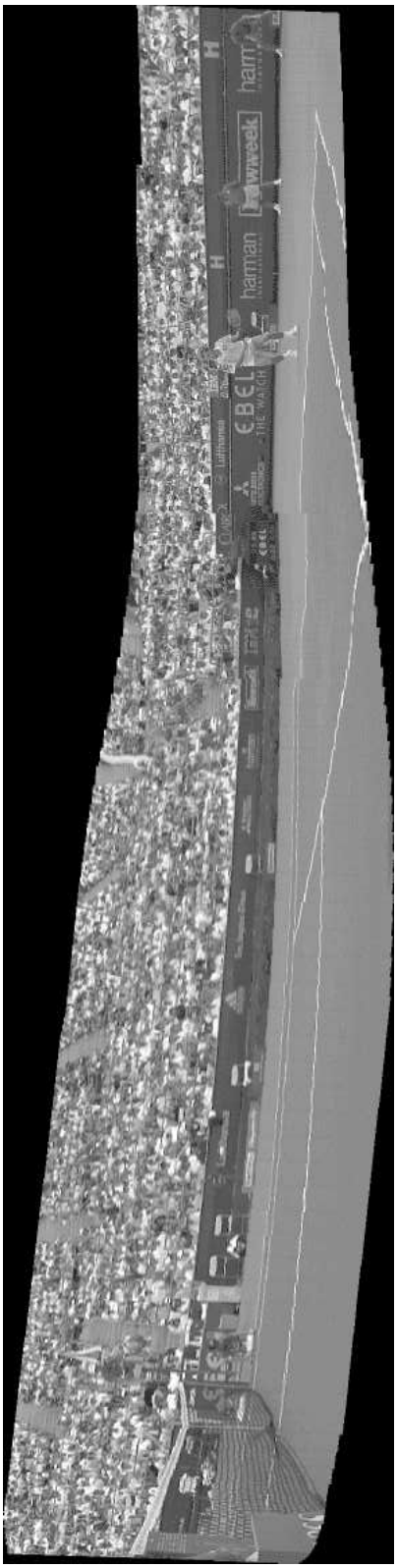
(a) (b)

Σχήμα 5.17: Η πρώτη (a) και τελευταία (b) στάση της ακολουθίας yosemite



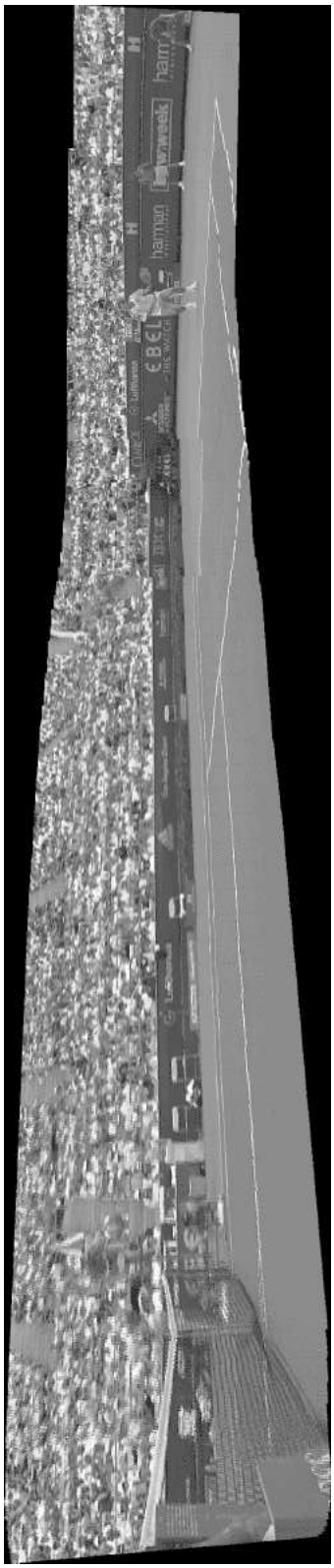
Σχήμα 5.18: Το ”μωσαϊκό” της ακολουθίας yosemite

Το ”μωσαϊκό” κατασκευάστηκε με τη χρήση του προβολικού μετασχηματισμού.

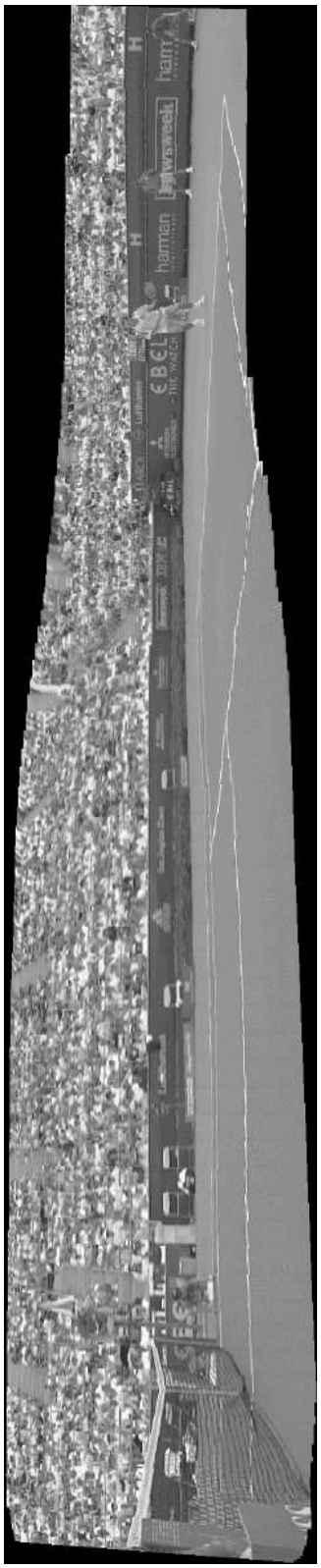


Σχήμα 5.19: Το ”μωσαϊκό” της ακολουθίας του Stefan που κατασκευάστηκε με τη χρήση του αφινικού μετασχηματισμού

Στο ”μωσαϊκό” αυτό παρουσιάζεται κακή αναπαράσταση της σκηνής που οφείλεται σε κακή εκτίμηση της κίνησης
Μαρία Δ. Τράκα



Σχήμα 5.20: Το ”μωσαϊκό” της ακολουθίας του Stefan που κατασκευάστηκε με τη χρήση του προβολικού μετασχηματισμού



Σχήμα 5.21: Το ”μωσαϊκό” της ακολουθίας του Stefan όπως προέκυψε με τη χρήση της μεθόδου διόρθωσης

Κεφάλαιο 6

Επίλογος

6.1 Συμπεράσματα

Το θέμα της εργασίας ήταν η κατασκευή της πανοραμικής όψης μιας σκηνής όπως απεικονίζεται σε μία ακολουθία εικόνων. Παρουσιάστηκαν και αναλύθηκαν τα στάδια που απαιτούνται για την κατασκευή της πανοραμικής όψης.

Στο κεφάλαιο 2 ορίστηκε η έννοια του ”μωσαϊκού” και περιγράφηκαν τα στάδια κατασκευής του.

Στο κεφάλαιο 3 παρουσιάστηκαν αναλυτικά δύο μέθοδοι εκτίμησης της κίνησης της κάμερας. Στην πρώτη η κίνηση υπολογίζεται από τον αλγόριθμο M-estimator που θεωρείται ο πιο διαδεδομένος αλγόριθμος σθεναρής εκτίμησης. Στη δεύτερη χρησιμοποιήθηκαν αναλλοίωτες σχέσεις ροπών για τον υπολογισμό των παραμέτρων του αφινικού μοντέλου κίνησης.

Στο κεφάλαιο 4 παρουσιάστηκαν τρόποι σύνθεσης των στάσεων της ακολουθίας για την κατασκευή του ”μωσαϊκού”. Παράλληλα παρουσιάστηκαν τα μειονεκτήματα των γνωστών τρόπων ευθυγράμμισης που προκαλούν το πρόβλημα των κακών ευθυγραμμίσεων, που εμφανίζεται συχνά στην κατασκευή των ”μωσαϊκών”. Για τη λύση των κακών ευθυγραμμίσεων προτάθηκε η μέθοδος διόρθωσης που βασίζεται στην τοπολογία των στάσεων της ακολουθίας.

Στο κεφάλαιο 5 συγκρίθηκαν οι μέθοδοι εκτίμησης της κίνησης, οι μέθοδοι υπολογισμού της οπτικής ροής και τα παραμετρικά μοντέλα που υλοποιήθηκαν στην παρούσα εργασία. Παράλληλα παρουσιάστηκαν ενδεικτικά αποτελέσματα ”μωσαϊκών” που κατασκευάστηκαν με τη χρήση των προηγούμενων μεθόδων και μοντέλων κίνησης.

Ένας αριθμός συμπερασμάτων προέκυψε κατά την σύγκριση των μεθόδων και

μοντέλων στο κεφάλαιο 5. Έτσι, καλύτερη και αποδοτικότερη προσέγγιση θεωρείται η μέθοδος της χρήσης αναλλοίωτων αφινικών ροπών. Η μέθοδος όμως αυτή εξαρτάται από την αντιστοίχιση των επιφανειών των δύο εικόνων στις οποίες εκτιμούνται οι σχέσεις των ροπών. Αν η αντιστοιχία είναι πάντα ορθή τότε η μέθοδος των ροπών υπερτερεί. Στην περίπτωση όμως εσφαλμένης αντιστοίχισης η μέθοδος αποτυγχάνει ολοκληρωτικά να εκτιμήσει την κίνηση της κάμερας. Πιο σθεναρός στα λάθη αντιστοίχισης εμφανίζεται ο αλγόριθμος M-estimator. Πολλές φορές εμφανίζεται καταλληλότερος αφού το σύνολο των δεδομένων που διαθέτουμε είναι πολύ πιθανό να περιέχει μικρά σφάλματα στα οποία η μέθοδος των ροπών είναι περισσότερο ευαίσθητη.

Συγκρίνοντας τα παραμετρικά μοντέλα κίνησης που υλοποιήθηκαν παρατηρήσαμε ότι η επιλογή του μεγαλύτερου μοντέλου για την περιγραφή της κίνησης της κάμερας δεν είναι πάντα ορθή. Η προτίμηση του μικρότερου, ικανού όμως να περιγράψει την κίνηση της κάμερας αποτελεί ιδανικότερη λύση. Για το σκοπό αυτό καταλήξαμε στην προσέγγιση της ιεράρχισης των παραμετρικών μοντέλων. Δεδομένης μιας ακολουθίας εικόνων εξετάζεται σε κάθε περίπτωση κάθε μοντέλο κίνησης ξεκινώντας από τη βάση της ιεραρχίας. Κατά την διαδικασία αυτή το πρώτο μοντέλο που εμφανίζεται να περιγράφει την κίνηση της κάμερας επιλέγεται να χρησιμοποιηθεί στη μέθοδο εκτίμησης της κίνησης για την κατασκευή του ”μωσαϊκού”. Ο έλεγχος του καταλληλότερου μοντέλου κίνησης μπορεί να επιτευχθεί εφαρμόζοντας το κριτήριο MDL (Minimum Description Length). Αν και λόγω έλλειψης χρόνου δεν υλοποιήθηκε η παραπάνω προσέγγιση, εντούτοις φαίνεται αρκετά αποδοτική.

6.2 Εφαρμογές "μωσαϊκού"

Οι εφαρμογές που χρησιμοποιείται η αναπαράσταση της video ακολουθίας με τη μορφή του ”μωσαϊκού” είναι πολλές. Οι πιο σημαντικές από αυτές περιγράφονται παρακάτω.

- *Αποθήκευση και ανακατασκευή των video ακολουθιών με χρήση του ”μωσαϊκού”.* Το στατικό ”μωσαϊκό” είναι κατάλληλο για την αποθήκευση στιγμιοτύπων. Μια video ακολουθία αποθηκεύεται σε ένα σύνολο ”μωσαϊκών”, το μέγεθος του οποίου εξαρτάται από τον αριθμό των στιγμιοτύπων της ακολουθίας. Πληροφορία όμως της ακολουθίας αποτελούν και τα ανεξάρτητα κινούμενα αντικείμενα της σκηνής καθώς και οι τυχόν μεταβολές της κατά τη διάρκεια του χρόνου λήψης της ακολουθίας. Για την πλήρη αποθήκευση της ακολουθίας απαιτείται επιπλέον η

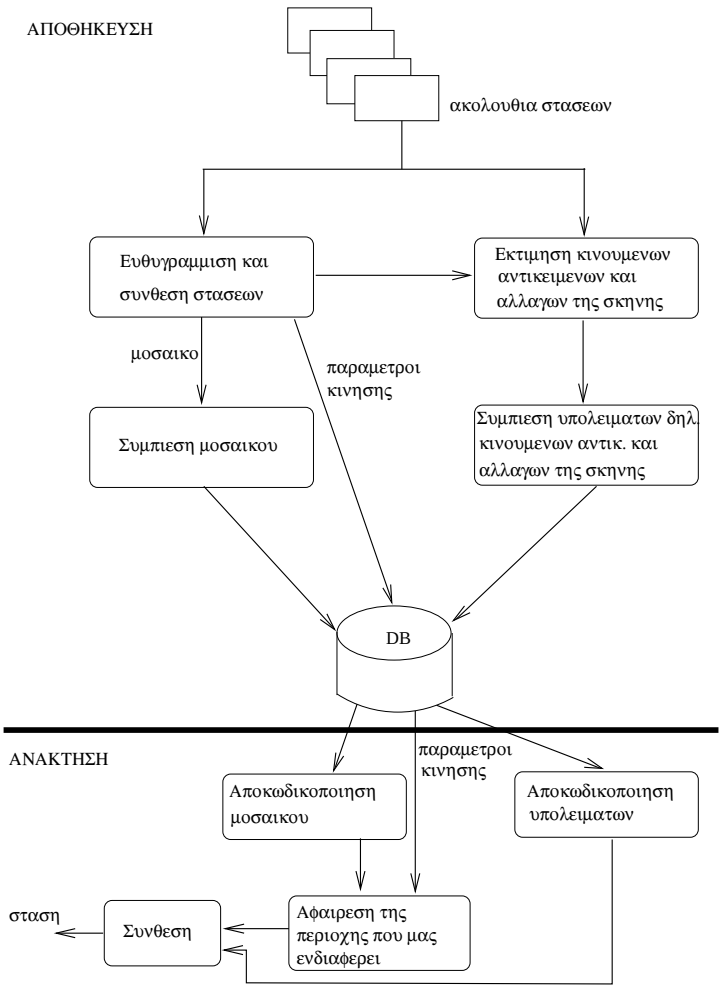
δημιουργία κατάλληλων δομών που περιέχουν πληροφορίες για τα ανεξάρτητα κινούμενα αντικείμενα και τις μεταβολές της σκηνής. Οι πληροφορίες αυτές λαμβάνονται ως η διαφορά κάθε στάσης του στιγμιοτύπου με το "μωσαϊκό" που το αναπαριστά. Για την ανακατασκευή των στάσεων εκτός των άλλων απαραίτητη είναι και η αποθήκευση των παραμέτρων ευθυγράμμισης της κάθε στάσης της ακολουθίας.

- *Συμπίεση των video ακολουθιών με χρήση του "μωσαϊκού".* Η συμπίεση μιας video ακολουθίας προκύπτει απλά με τη συμπίεση του συνόλου των "μωσαϊκών" που την αναπαριστά και του συνόλου των δομών των ανεξάρτητων κινούμενων αντικειμένων. Συμπιέζοντας παράλληλα και τις παραμέτρους ευθυγράμμισης επιτυγχάνεται η ανακατασκευή όλων των στάσεων της ακολουθίας στο αρχικό τους σύστημα συντεταγμένων. Η διαδικασία αποθήκευσης, συμπίεσης και ανάκτησης μιας video ακολουθίας απεικονίζεται στο σχήμα 6.1.

- *Γρήγορη μετάδοση των video ακολουθιών.* Ο χρόνος μετάδοσης της πληροφορίας που περιέχουν οι video ακολουθίες είναι πολύ πιο μικρός όταν μεταδίδονται τα "μωσαϊκά" που την αναπαριστούν παρά οι εκατοντάδες ή και χιλιάδες στάσεις που περιέχει. Ο χρόνος αυτός μετάδοσης αυξάνεται σε μικρό όμως βαθμό με τη μετάδοση και των δομών των ανεξάρτητα κινούμενων αντικειμένων και των αλλαγών της σκηνής, που εμπεριέχονται στην πληροφορία μιας video ακολουθίας.

- *"Μωσαϊκό" σύνοψης.* Παρουσιάζεται πολλές φορές επιθυμητή η σύνοψη των γεγονότων που λαμβάνουν μέρος στη video ακολουθία. Το "μωσαϊκό" όπως αναφέραμε απεικονίζει τη στατική σκηνή που είδε η κάμερα. Προβάλλοντας το σύνολο των υπολειμμάτων (residuals) της κίνησης όπως τα ανεξάρτητα κινούμενα αντικείμενα σε όλες τις θέσεις τους μέσα στο "μωσαϊκό" λαμβάνεται η σύνοψη των γεγονότων κατά τη διάρκεια λήψης της ακολουθίας.

- *Εγγραφή και χειρισμός της video ακολουθίας.* Η εγγραφή (editing) και ο χειρισμός (manipulation) του video αποτελούν εφαρμογές στις οποίες απαραίτητη είναι η συμμετοχή του χρήστη. Ορισμένες από τις πιο συχνές λειτουργίες είναι από τη μια η εισαγωγή αντικειμένων στο video και από την άλλη η αφαίρεση αντικειμένων του video. Χωρίς τη χρήση του "μωσαϊκού" οι λειτουργίες αυτές λαμβάνουν μέρος σε κάθε στάση της video ακολουθίας. Με τη χρήση όμως του "μωσαϊκού" η εισαγωγή και η αφαίρεση πραγματοποιείται μόνο πάνω σε αυτό και αυτόματα εφαρμόζεται σε κάθε στάση της ακολουθίας κατά τη διάρκεια της ανακατασκευής της από



Σχήμα 6.1: Αποθήκευση-συμπίεση και ανάκτηση μιας video ακολουθίας

το ”μωσαϊκό”. Αποτέλεσμα της χρήσης του ”μωσαϊκού” αποτελεί η αύξηση της απόδοσης και η μείωση του χρονικού κόστους εκτέλεσης της διαδικασίας εγγραφής και χειρισμού του video.

- *Αναζήτηση και πλοήγηση σε μια video ακολουθία.* Το ”μωσαϊκό” αποτελεί ένα από τα βασικότερα εργαλεία για λειτουργίες πάνω σε βάσεις δεδομένων από ακολουθίες εικόνων. Με τη χρήση του επιτυγχάνεται από τη μια γρήγορη πλοήγηση σε κάθε σημείο της ακολουθίας και από την άλλη διευκολύνεται η διαδικασία των επερωτήσεων μέσα στη βάση. Παράλληλα αποτελούν αξιόπιστα εργαλεία ανεύρεσης και ανάκτησης αντικειμένων που περιέχει το video. Η αναζήτηση στο ”μωσαϊκό” είναι σαφώς πιο γρήγορη από την αναζήτηση σε όλες τις στάσεις της ακολουθίας.

6.3 Μελλοντικές κατευθύνσεις

Για την πλήρη ολοκλήρωση της παρούσας εργασίας, ένας αριθμός επιπρόσθετων επεκτάσεων προτείνεται να πραγματοποιηθεί στο μέλλον.

Μία από τις επιδιώξεις μας είναι η αυτόματη επιλογή του παραμετρικού μοντέλου που περιγράφει την κίνηση της κάμερας στην ακολουθία. Η λύση που προτείνουμε είναι χρήση μιας ιεραρχίας μοντέλων κίνησης [42] και η επιλογή του μικρότερου καταλληλότερου μοντέλου για την περιγραφή της κίνησης.

Επιπλέον, αν και η μέθοδος της τοπολογίας διορθώνει σε κάποιο ποσοστό την κακή ευθυγράμμιση ορισμένων στάσεων που εμφανίζονται στο τελικό "μωσαϊκό", εντούτοις υπάρχουν ακόμα ατέλειες που θα έπρεπε να αφαιρεθούν διορθώνοντας την εκτίμηση της κίνησης. Ο χωρισμός της σκηνής στα διαφορετικά επίπεδα που περιέχει και η εκτίμηση της κίνησης σε κάθε επίπεδο χωριστά ίσως να διόρθωνε σε σημαντικό βαθμό τις ατέλειες που εμφανίζονται στο "μωσαϊκό", ιδιαίτερα στις ακολουθίες με πολλαπλά επίπεδα κίνησης.

Η ανίχνευση και η εξαγωγή των ανεξάρτητα κινούμενων αντικειμένων της ακολουθίας αποτελεί επίσης μελλοντική επέκταση της εργασίας. Για την εξαγωγή τους αρκεί η σύγκριση κάθε στάσης της ακολουθίας με το στατικό "μωσαϊκό" που περιέχει την πληροφορία του σταθερού σκηνικού. Η διαφορά που προκύπτει εκφράζει τα υπολείμματα της σκηνής (residuals) που είναι είτε κινούμενα αντικείμενα είτε τυχόν αλλαγές στο φωτισμό. Το στατικό όμως "μωσαϊκό" περιέχει τα κινούμενα αντικείμενα στην θέση που κατέχουν στην στάση αναφοράς. Για το σκοπό αυτό κατάλληλες μέθοδοι ανίχνευσης κινούμενων αντικειμένων πρέπει να εφαρμοστούν.

Τέλος προτείνεται η αναπαράσταση των "μωσαϊκών" σε σφαιρική ή κυλινδρική επιφάνεια. Οι αναπαραστάσεις αυτές παράγουν τα γνωστά περιβάλλοντα εικονικής πραγματικότητας. Η δημιουργία αυτών των αναπαραστάσεων απαιτεί τον μετασχηματισμό των επίπεδων συντεταγμένων σε σφαιρικές ή κυλινδρικές ανάλογα με την τελική μορφή αναπαράστασης που επιθυμούμε.

Παράρτημα Α

Καθορισμός παραμέτρων (Χαλαρωτικές μέθοδοι)

Θεωρώντας το αφινικό μοντέλο κίνησης με εξισώσεις:

$$u(a) = \begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix} = \begin{bmatrix} a_0 + a_1 x + a_2 y \\ a_3 + a_4 x + a_5 y \end{bmatrix}$$

Οι μερικές παράγωγοι της εξίσωσης 3.64 της σθεναρής παλινδρόμησης για το αφινικό μοντέλο κίνησης είναι:

$$\frac{\vartheta E_D}{\vartheta a_0} = \sum_{\mathcal{R}} I_x \psi((\bigtriangledown I)^T u(a) + I_t, \sigma)$$

$$\frac{\vartheta E_D}{\vartheta a_1} = \sum_{\mathcal{R}} I_x x \psi((\bigtriangledown I)^T u(a) + I_t, \sigma)$$

$$\frac{\vartheta E_D}{\vartheta a_2} = \sum_{\mathcal{R}} I_x y \psi((\bigtriangledown I)^T u(a) + I_t, \sigma)$$

$$\frac{\vartheta E_D}{\vartheta a_3} = \sum_{\mathcal{R}} I_y \psi((\bigtriangledown I)^T u(a) + I_t, \sigma)$$

$$\frac{\vartheta E_D}{\vartheta a_4} = \sum_{\mathcal{R}} I_y x \psi((\bigtriangledown I)^T u(a) + I_t, \sigma)$$

$$\frac{\vartheta E_D}{\vartheta a_5} = \sum_{\mathcal{R}} I_y y \psi((\bigtriangledown I)^T u(a) + I_t, \sigma)$$

Οι παράγοντες αλλαγής κλίμακας T_{a_i} που χρησιμοποιούνται στην ελαχιστοποίηση είναι:

$$T_{a_0} = {I_x}^2 \max_x \ddot{\rho}(x) \qquad T_{a_3} = {I_y}^2 \max_x \ddot{\rho}(x)$$

$$T_{a_1} = {I_x}^2 x^2 \max_x \ddot{\rho}(x) \qquad T_{a_2} = {I_x}^2 y^2 \max_x \ddot{\rho}(x)$$

$$T_{a_4} = {I_y}^2 x^2 \max_x \ddot{\rho}(x) \qquad T_{a_5} = {I_y}^2 y^2 \max_x \ddot{\rho}(x)$$

για τη συνάρτηση σθεναρότητας (ρ) Geman-McClure

$$\psi(x,\sigma) = \frac{2x\sigma}{(\sigma + x^2)^2} \quad \text{και} \quad \max_x \ddot{\rho}(x,\sigma) = \frac{2}{\sigma^2}$$

για τη συνάρτηση σθεναρότητας (ρ) Lorentzian

$$\psi(x,\sigma) = \frac{2x}{2\sigma^2 + x^2} \quad \text{και} \quad \max_x \ddot{\rho}(x,\sigma) = \frac{1}{\sigma^2}$$

Παράρτημα Β

Κατασκευή κυρτού περιβλήματος

Στο παράρτημα αυτό αναφέρεται ο αλγόριθμος του Graham-Scan για την κατασκευή του κυρτού πολυγώνου από ένα σύνολο σημείων Ν. Παρακάτω ακολουθεί ψευδοκώδικας για την δημιουργία ενός κυρτού περιβλήματος από ένα σύνολο Ν σημείων του επιπέδου.

Διαδικασία Convex(P, ΚΠ:ΠίνακαςΣημείων; N, K: Ακέραιοι);

Επιστρέφει τα (ΚΠ, Κ);

Μεταβλητές:

Ο, Α, Β, Γ: *Σημεία*;

Σ : *Στοίβα Από Σημεία*;

Ι : *Ακέραιος*;

Συνάρτηση Angle(O, P: *Σημεία*):*Πραγματικός*;

/* Angle := γωνία OP με x-άξονα, (εάν O = P τότε Angle = 0.0) */

Συνάρτηση Length(P, Q: *Σημεία*):*Πραγματικός*;

/* Length := το μήκος |*Q*P| */

Συνάρτηση IsLess(O, P, Q:*Σημεία*):*Αληθές?*;

/* Συγκρίνει τα P, Q ως προς το O. IsLess := Angle(OP) < Angle(OQ) αν ίσες τότε το Length(OP) < Length(OQ) */

Διαδικασία Sort(P: *ΠίνακαςΣημείων*; N:*Ακέραιος*; O:*Σημείο*);

Επιστρέφει(P);

/* Ταξινόμησε P(1..N) ως προς την IsLess(O, A, B) */

Συνάρτηση OnLeftOf(A, B, Γ:*Σημεία*):*Ναι/Οχι*

/* Εξετάζει εάν το Β είναι "αριστερά" της ΓΑ ευθείας ή όχι */

```
{
        Εάν N ≤ 3 τότε {ΚΠ:=P K:=3}
        Αλλιώς
        {
            Επιλογή του Ο από τα Ρ σημεία με μεγαλύτερη x-συνιστώσα και εκ των
            ίσων ως εκείνο με μεγαλύτερη y-συνιστώσα.
            PointSort(P, N, O)
            Με το Σ {Push(P(1)); Push(P(2)); Push(P(3)) }
            Β := Σ.Pop;                 /προ-προ-ηγούμενο σημείο
            Γ := Σ.Pop                  /προ-ηγούμενο σημείο
            Δια I=4 Έως N+1
            {
                A := T(I mod N)         /ηγούμενο σημείο
                Εφόσον OnLeftOf(A, B, Γ) Τότε
                {
                    B := Γ              /Το Β αγνοείται πλέον
                    Γ := Σ.Pop          /το νέο Γ εκ της στοίβας
                }
                Σ.Push(Γ)               /το Γ επιστρέφεται στη στοίβα
                Γ := B
                B := A
            }
            K := το μέγεθος της στοίβας Σ
            ΚΠ := το περιεχόμενο της στοίβας Σ
        }
}
```

Παράρτημα Γ

Καθορισμός παραμέτρων (Κριτήριο μέτρησης της ισχύς μιας αντιστοίχισης)

Η συσχέτιση c_{m_1,m_2} που παρουσιάζεται στην εξίσωση 3.22 ορίζεται ως εξής:

$$c_{m_1,m_2} = \frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m}[I_1(u_1+i, v_1+j) - \overline{I_1(u_1,v_1)}]\,[I_2(u_2+i, v_2+j) - \overline{I_2(u_2,v_2)}]}{(2n+1)\,(2m+1)\sqrt{\sigma^2(I_1)\,\sigma^2(I_2)}} \quad (\Gamma.1)$$

όπου $\overline{I_k(u,v)} = \sum_{i=-n}^{n}\sum_{j=-m}^{m} I_k(u+i, v+j) \,/\, [(2n+1)\,(2m+1)]$ είναι η μέση τιμή στο σημείο (u, v) της εικόνας I_k (k = 1, 2), και $\sigma(I_k)$ είναι το μέτρο της διασποράς της εικόνα I_k στη γειτονιά (2n+1)x(2m+1) του (u, v), που δίνεται από την παρακάτω σχέση:

$$\sigma(I_k) = \sqrt{\frac{\sum_{i=-n}^{n}\sum_{j=-m}^{m} I_k(u,v)^2}{(2n+1)\,(2m+1)} - \overline{I_k(u,v)}^2} \quad (\Gamma.2)$$

Η μεταβλητή *dist*$(m_{1i}, m_{2j}; n_{1k}, n_{2l})$ της εξίσωσης 3.22 εκφράζει τη μέση απόσταση των δύο ζευγαριών, ως:

$$dist(m_{1i}, m_{2j}; n_{1k}, n_{2l}) = [d(m_{1i}, n_{1k}) + d(m_{2j}, n_{2l})] \,/ 2$$

όπου d(m, n) = $|m - n|$ είναι η ευκλείδια απόσταση μεταξύ των m και n και

$$\delta(m_{1i}, m_{2j}; n_{1k}, n_{2l}) = \begin{cases} e^{-r/\varepsilon_r} & \text{αν } (n_{1k}, n_{2l}) \text{ είναι υποψήφιογιααντιστοίχιση και } r < \varepsilon_r \\ 0 & \text{αλλιώς} \end{cases}$$

όπου r είναι η διαφορά των σχετικών αποστάσεων και δίνεται από την παρακάτω σχέση:

$$r = \frac{|d(m_{1i}, n_{1k}) - d(m_{2j}, n_{2l})|}{dist(m_{1i}, m_{2j}; n_{1k}, n_{2l})}$$

ενώ το ε_r είναι το κατώτερο όριο της διαφοράς των σχετικών αποστάσεων.

Παράρτημα Δ

Ευρετήριο Όρων

Αναλλοίωτος	Invariant
Αφινικό	Affine
Διγραμμική Παρεμβολή	Bilinear Interpolation
Εγγραφή	Editing
Ευθυγράμμιση	Allignment
Κακά δεδομένα	Outliers
Καλά δεδομένα	Inliers
Κλίση	Gradient
Μέθοδος Απότομης Κλίσης	Steepest Descent Method
Οπτική Ροή	Optical Flow
Πανοραμική σύνθεση όψεων	Mosaic
Παράλληλη Προβολή	Parallel Projection
Πλοήγηση	Browsing
Προβολικός	Projective
Προοπτική προβολή	Perspective Projection
Ροπή	Moment
Σημείο Κατάρρευσης	Breakdown Point
Σθεναρός	Robust
Στάση	Frame

Στιγμιότυπο	Shot
Συνάρτηση Επίδρασης	Influence Function
Συνάρτηση Κόστους	Error Penalty Function
Σχετική Αποτελεσματικότητα	Relative Efficiency
Υπόλοιπο-Υπόλειμμα	Residual
Χαλαρωτικές Μέθοδοι	Relaxation Methods
Χειρισμός	Manipulation

Βιβλιογραφία

[1] S. Harpreet Sawhney and Serge Ayer. Compact representation of videos through dominant and multiple motion estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8):814--830, August 1996.

[2] Wayne Niblack, R. Barber, W. Equitz, Myron Flickner, E. Glasman, Dragutin Petkovic, Peter Yanker, Christos Faloutsos, and G. Taubin. The QBIC project: querying images by content using color, texture and shape. Research Report, IBM Research Division, Almaden Research Center, San Jose, California, February 1993.

[3] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Tools for content-base manipulation of image databases. Technical Report 255, MIT Media Lab Vismod, 1993.

[4] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by image and video content: The QBIC system. *Computer*, 28(9):23--32, September 1995.

[5] J-K. Hong H. Hashihara, J. Tokahashi. Scene retrieval method for motion image databases. technical report, IBM Research Laboratory In Tokyo, 1991.

[6] K. Hirata and T. Kato. Rough sketch-based image information retrieval. In *Proc. NEC RD*, volume 34, pages 263--273, Apr 1993.

[7] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by image and video content: The QBIC system. *Computer*, 28(9):23--32, September 1995.

[8] N. Alvertos S. Liapis and G. Tziritas. Maximum liklihood texture classification and bayesian texture segmentation using discrete wavelet frames. In *International Conference in Digital Signal Processing DSP*, volume 2, pages 1107--1110, July 1997.

[9] Jim Bergen Michal Irani, P. Anandan, Rakesh Kumar, and Steve Hsu. Efficient representations of video sequences and their applications. *Signal Processing Image Communication*, 8:327--351, 1996.

[10] H. S. Sawhney, S.Hsu, and R. Kumar. Robust video mosaicing through topology inference and local to global alignment. *Lecture Notes in Computer Science*, 1407:103--118, Oct 1998.

[11] Steve Mann and Rosalind W. Picard. Video orbits of the projective group: A simple approach to featureless estimation of parameters. *IEEE Transaction On Image Processing*, 6(2):12, Sept 1997.

[12] Richard Szeliski. Video mosaics for virtual environments. *IEEE Computer Graphics and Applications*, pages 22--30, March 1996.

[13] H. Nicolas. Video object manipulation based on mosaic representation for post-production applications. In *proceedings of IX European Signal Processing Conference*, volume II, page 937, Sept 1998.

[14] Heung-Yeung Shum and Richard Szeliski. Panoramic image mosaics. Technical Report MSR-TR-97-23, Microsoft, Redmond , WA 98052. http://www.research.microsoft.com/.

[15] Harpreet S. Sawhney and Rakesh Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 21(3):235--243, March 1999.

[16] I. Finci B. Rousso, S.Peleg and A. Rav-Acha. Universal mosaicing using pipe projection. In *ICCV*, pages 945--952, 1998.

[17] S. Peleg and J. Herman. Panoramic mosaics by manifold projection. In *Proc. Conf. Computer Vision and Pattern Recognition, CVPR*, pages 338--343, June 1997.

[18] S. Peleg and J. Herman. Panoramic mosaicing with videobrush. In *DAPRA Image Understanding Workshop*, pages 261--264, May 1997.

[19] S. Peleg B. Rousso and I. Finci. Mosaicing with generalized strips. In *DAPRA Image Understanding Workshop*, pages 255--260, May 1997.

[20] Yalin Xiong and Ken Turkowski. Creating image-based vr using a self-calibrating fisheye lens. In *Proc. Conf. Computer Vision and Pattern Recognition, CVPR*, pages 237--243, San Juan, Puerto Rico, June 1997.

[21] J. Meehan. Panoramic photography. Watson-Guptill, 1990.

[22] Harpreet S. Sawhney, Serge Ayer, and M. Gorkani. Motion based 2D?3D dominant motion estimation for mosaicing and video representation. In Eric Grimson, editor, *International Conference on Computer Vision*, pages 583--590. IEEE, June 1995.

[23] O. D. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *International Journal of Pattern Recognition and Artificial Intelligence*, 2:485--508, 1988.

[24] A. Shashua and N. Navab. Relative affine structure: Theory and application to 3D reconstruction from perspective views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 483--489, Los Alamitos, CA, USA, June 1994. IEEE Computer Society Press.

[25] H. Sawhney. Simplifying motion and structure analysis using planar parallax and image warping. In *In Proc. ICPR*, volume 1, pages 403--408, Oct 1994.

[26] R Kumar, P Anandan, and K Hanna. Shape recovery from multiple views: a parallax based approach. In *ARPA Image Understanding Workshop, Monterey, CA, Nov. 1994*, 2929 Campus Drive, Suite 260, San Mateo, California 94403o, November 1994. ARPA, Image Understanding, Morgan Kauffmann Publishers.

[27] Zhengyou Zhang Rachid Deriche, Olivier Faugeras, and Quang-Tuan Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report RR-2273, Inria, Institut National de Recherche en Informatique et en Automatique, May 1994.

[28] Jean-Marc Odobez and Patrick Bouthemy. Robust multiresolution estimation of parametric motion models applied to complex scenes. Publication interne no 788, Inria, Institut de Recherche en Informatique et Systemes Aleatoires, Campus de Beaulieu - 35042 Rennes Cedex - France, Jan 1994.

[29] Michael J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75--104, January 1996.

[30] P. Schroeter S. Ayer and J. Bigun. Segmentation of moving objects by robust motion parameter estimation over multiple frames. In *In ECCV*, Stockholm, Sweden, May 1994.

[31] Yoshinobu Tonomura, Akihito Akutsu, Kiyotaka Otsuji, and Toru Sadakata. VideoMAP and videospaceicon: Tools for anatomizing video content. In *Proceedings of ACM INTERCHI'93 Conference on Human Factors in Computing Systems*, Structuring Images for Interaction, pages 131--136, 1993.

[32] G. Tziritas and C. Labit. *Motion Analysis for Image Sequence Coding*. Elsevier, 1994.

[33] Jaswant R. Jain and Anil K. Jain. Displacement measurement and its application in interframe image coding. *IEEE Trans. Commun.*, COM-29(12):1799--1808, Dec 1981.

[34] T. Koga, Kazumoto Iinuma, Yukihiko Iijima, A. Hirano, and Tatsuo Ishiguro. Motion-compensated interframe coding for video conferencing. In *Proc. NTC*, pages G5.3.1--G5.3.5, 1981.

[35] R. Srinivasan and K. R. Rao. Predictive coding based on efficient motion estimation. *IEEE Transactions on Communications*, 33:888--896, 1985.

[36] J. Santillana Rivero. *Etude du codage MIC differentiel adaptatif de sequences d' image de television avec compensation du movements.* PhD thesis, Ingenieur, INSA de Rennes, 1985.

[37] Emanuele Trucco and Alessandro Verri. *Introductory Techniques for 3-D Computer Vision.* Prentice Hall, 1998.

[38] A.K. Jain. *Fundaments of Digital Image Processing.* Prentice Hall, 1989.

[39] S. M. Smith and J.M. Brady. Susan - a new approach to low image processing. Technical Report TR95SMS1, Department of Clinical Neurology, Oxford University, Radcliffe Infirmary, Oxford OX2 6HE, UK, 1995.

[40] Peter J. Huber. *Robust Statistics.* Willey, 1981.

[41] F. Hampel. Lecture notes in statistics. In H. Rieder, editor, *Robust Statistics, Data Analysis, and Computer Intensive Methods; In Honor of Peter Huber's 60th Birthday.* Springer, 1996.

[42] N. Komodakis. Robust 3d motion estimation and depth layering. Master's thesis, Computer Science Department - University of Crete, September 1997. In proceedings of 13th International Conf. on Digital Signal Processing.

[43] William T. Ueterling William H. Press, Saul A. Teukolsy and Brian P. Flannary. *Numerical Recipes In C.* Cambridge University Press.

[44] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing.* Addison Wesley, 1993.

[45] Herbert Susse Irene Rothe and Klaus Voss. The method of normalization to determine invariants. *IEEE Transanctions On Pattern Analysis And Machine Intelligence*, 18(4):366--375, April 1996.

[46] M. Traka G. Tzanetakis and G. Tziritas. Motion estimation based on affine moment invariants. In *In proceedings of IX European Signal Processing Conference*, volume II, page 925, Sept 1998.