ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών

Εκτίμηση Τρισδιάστατης Κίνησης Κάμερας και Βάθους από Ακολουθίες Εικόνων

Διπλωματική Εργασία

Κατσαΐτης Δημήτρης

Χανιά, 2008

Πολυτεχνειο Κρητης

Τμήμα Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών

Διπλωματική Εργασία

Συγγραφέας: Κατσαΐτης Δημήτρης

Επιβλέπων: Επίκουρος Καθηγητής Μιχαήλ Γ. Λαγουδάκης

Εξεταστική επιτροπή: Επίκουρος Καθηγητής Μιχαήλ Γ. Λαγουδάκης

Αναπληρωτής Καθηγητής Ευριπίδης Πετράκης

Επίκουρη Καθηγήτρια Κατερίνα Μάνια

Πρόλογος

Για την εκπόνηση της παρούσας διπλωματικής εργασίας οφείλω να ευχαριστήσω τον Επίκουρο Καθηγητή του τμήματος Ηλεκτρονικών Μηχανικών και Μηχανικών Η/Υ κύριο Μιχαήλ Λαγουδάκη για την πολύτιμη βοήθεια και καθοδήγησή του καθ’ όλη τη διάρκειά της.

Ακόμα θα ήταν παράλειψη εκ μέρους μου να μην ευχαριστήσω τους γονείς μου Γεώργιο και Αθανασία Κατσαΐτη για την συνεχή και αμέριστη υποστήριξή τους όλα αυτά τα χρόνια.

Περίληψη

Η εκτίμηση της κίνησης μιας κάμερας αποτελεί τα τελευταία χρόνια ένα σημαντικό και διαρκώς εξελισσόμενο τομέα έρευνας της μηχανικής όρασης. Αποτελεί δε βασικό αντικείμενο μελέτης για πολλές εφαρμογές που σχετίζονται με την ανάλυση ακολουθιών εικόνων (π.χ. συμπίεση βίντεο), την οπτική αντίληψη (π.χ. εκτίμηση απόστασης, διαστάσεων, σχήματος αντικειμένων) και τη ρομποτική (π.χ. παθητική πλοήγηση οχημάτων). Οι λύσεις που έχουν προταθεί στη βιβλιογραφία ποικίλλουν ως προς την ποιότητα των αποτελεσμάτων τους και ως προς τις περιπτώσεις εφαρμογής τους και γι' αυτό το λόγο δεν υπάρχει μία καθολικά αποδεκτή μεθοδολογία αντιμετώπισης του προβλήματος. Σ' αυτή τη διπλωματική εργασία ερευνήσαμε το πρόβλημα της ανάκτησης της κίνησης μιας κάμερας στον τρισδιάστατο χώρο από ακολουθίες εικόνων, γνωστό και ως egomotion, και προτείνουμε μια ολοκληρωμένη λύση συνθέτοντας τα καλύτερα στοιχεία των υπαρχουσών μεθόδων και συμπληρώνοντας με νέες ιδέες τα διάφορα στάδια επεξεργασίας. Στόχος μας ήταν η υλοποίηση παραμετροποιήσιμου λογισμικού για την παραγωγή αξιόπιστων αποτελεσμάτων σε όλο το φάσμα συνθετικών και πραγματικών ακολουθιών εικόνων. Βάσει της συνηθέστερης πρακτικής, χωρίσαμε το βασικό πρόβλημα σε δυο επιμέρους υποπροβλήματα: εύρεση της στιγμιαίας ταχύτητας χαρακτηριστικών σημείων της ακολουθίας εικόνων στο δισδιάστατο επίπεδο της εικόνας (οπτική ροή) και επεξεργασία της οπτικής ροής για την ανάκτηση του βάθους (απόσταση από τα αντικείμενα) και της κίνησης της κάμερας στο χώρο. Ως κίνηση στο χώρο ορίζουμε όλες τις πιθανές μεταφορές (τρεις βαθμοί ελευθερίας - Καρτεσιανοί άξονες) και περιστροφές (τρεις βαθμοί ελευθερίας - γωνίες Euler) που δύναται να πραγματοποιήσει η κάμερα στον τρισδιάστατο χώρο σε κάθε χρονική στιγμή. Εκτεταμένα πειράματα σε διάφορα σενάρια τόσο με τεχνητές συνθέσεις όσο και με πραγματικές ακολουθίες εικόνων αποκαλύπτουν την αξιοπιστία, αλλά και τις αδυναμίες της προτεινόμενης μεθόδου. Τα ποικίλα συμπεράσματα της εργασίας μπορούν να βοηθήσουν τον κάθε ενδιαφερόμενο να κατανοήσει σε βάθος τις ιδιαιτερότητες αλλά και τις προοπτικές και τα όρια του προβλήματος εκτίμησης της κίνησης μιας κάμερας.

Περιεχόμενα

ΠΡΟΛΟΓΟΣ 3

ΠΕΡΙΛΗΨΗ 4

ΠΕΡΙΕΧΟΜΕΝΑ 5

ΕΙΣΑΓΩΓΗ 9

ΟΠΤΙΚΗ ΡΟΗ 13

2.1 ΕΙΣΑΓΩΓΗ 13

2.2 ΤΕΧΝΙΚΕΣ ΔΙΑΦΟΡΙΣΗΣ 14

2.2.1 Τεχνικές διαφόρισης πρώτης τάξης 14

2.2.2 Τοπικές τεχνικές διαφόρισης πρώτης τάξης 16

2.3 ΠΑΡΑΜΕΤΡΟΙ ΠΡΟΒΛΗΜΑΤΟΣ ΚΑΙ ΑΙΤΙΕΣ ΣΦΑΛΜΑΤΩΝ 18

2.3.1 Φαινόμενο διαφράγματος (Aperture problem) 18

2.3.2 Ανεξάρτητη κίνηση – θόρυβος 19

2.3.3 Μέγιστη ταχύτητα κίνησης 21

2.3.4 Συμπεράσματα: Μέγεθος «Παραθύρου» 22

2.4 ΥΠΟΛΟΓΙΣΜΟΣ ΠΑΡΑΓΩΓΩΝ 23

2.5 ΕΞΟΜΑΛΥΝΣΗ 24

2.6 ΜΕΘΟΔΟΣ LUCAS-KANADE ΚΑΙ ΠΡΟΣΘΗΚΕΣ 26

2.6.1 Επαναληπτική πολλαπλή ανάλυση 32

2.6.2 Υπολογισμός Gaussian πυραμίδας 34

2.7 ΑΛΓΟΡΙΘΜΟΣ 35

2.8 ΑΝΙΧΝΕΥΣΗ ΣΦΑΛΜΑΤΩΝ 37

2.9 ΠΕΙΡΑΜΑΤΙΣΜΟΙ 41

ΟΠΤΙΚΟ ΜΟΝΤΕΛΟ ΚΑΙ ΕΞΙΣΩΣΕΙΣ ΚΙΝΗΣΗΣ 44

3.1 ΕΙΣΑΓΩΓΗ 44

3.2 ΟΠΤΙΚΟ ΜΟΝΤΕΛΟ 45

3.3 ΣΧΕΣΗ ΔΥΟ (2D) ΚΑΙ ΤΡΙΩΝ (3D) ΔΙΑΣΤΑΣΕΩΝ 46

3.4 ΣΧΕΣΗ ΕΣΤΙΑΚΗΣ ΑΠΟΣΤΑΣΗΣ, ΟΠΤΙΚΟΥ ΠΕΔΙΟΥ ΚΑΙ ΟΠΤΙΚΟΥ ΠΛΑΝΟΥ 48

3.5 ΣΧΕΣΗ ΟΠΤΙΚΟΥ ΜΟΝΤΕΛΟΥ ΚΑΙ ΤΡΙΣΔΙΑΣΤΑΤΗΣ ΚΙΝΗΣΗΣ 50

3.6 ΣΧΕΣΗ ΟΠΤΙΚΗΣ ΡΟΗΣ ΚΑΙ ΤΡΙΣΔΙΑΣΤΑΤΗΣ ΚΙΝΗΣΗΣ 53

ΕΚΤΙΜΗΣΗ ΤΗΣ ΚΙΝΗΣΗΣ ΤΗΣ ΚΑΜΕΡΑΣ 55

4.1 Εισαγωγη ... 55
4.2 Ορισμος και διατυπωση του προβληματος ... 55
4.2.1 Προϋποθέσεις και παράμετροι ... 55
4.2.2 Μη-γραμμικό μοντέλο και ελαχιστοποίηση ... 57
4.2.3 Μετασχηματισμός μη-γραμμικού συστήματος και νόρμας ... 59
4.3 Σχετικες εργασιες ... 61
4.4 Μεθοδολογια για την λυση του προβληματος ... 63
4.4.1 Διαδικασία αναδρομικής γραμμικοποίησης Gauss-Newton ... 64
4.4.2 Επίλυση με Gauss-Newton ... 64
4.4.3 Επίλυση με διαχωρισμένο Gauss-Newton ... 67
4.4.4 Επίλυση με σταθμισμένο διαχωρισμένο Gauss-Newton ... 68
4.5 Μειωση ελαχιστων συναρτησης κοστους ... 73
4.6 Αλγοριθμος ... 76

ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ ... 79

5.1 Εισαγωγη ... 79
5.2 Συνθεση οπτικης ροης ... 80
5.2.1 Εξίσωση σύνθεσης και Οπτικό μοντέλο ... 80
5.2.2 Μοντέλο θορύβου ... 81
5.2.3 Ταχύτητα κίνησης ... 83
5.3 Δοκιμες με συνθετικη οπτικη ροη ... 84
5.3.1 Μεθοδολογία μέτρησης σφαλμάτων ... 84
5.3.2 Δοκιμές – Αποτελέσματα ... 85
5.4 Δοκιμες με πραγματικες ακολουθιες ... 113
5.4.1 Εισαγωγή ... 113
5.4.2 Τρισδιάστατη μηχανή ... 114
5.4.3 Επεξεργασία με πραγματικές ακολουθίες ... 115
5.4.4 Παρατηρήσεις ... 121

ΣΥΜΠΕΡΑΣΜΑΤΑ – ΣΥΝΟΨΗ ... 123

6.1 Συνοψη ... 123
6.2 Συμπερασματα ... 125
6.3 Μελλοντικες προεκτασεις ... 127

ΒΙΒΛΙΟΓΡΑΦΙΑ ... 129

Κεφάλαιο 1

Εισαγωγή

Πολλές φορές σε μια ακολουθία εικόνων είναι αρκετά συνηθισμένη η περίπτωση όπου η κάμερα από την οποία έχουν ληφθεί οι εικόνες δεν παραμένει σταθερή αλλά κινείται μέσα στο χώρο. Για πολλές εφαρμογές όπως π.χ. στην παθητική πλοήγηση ενός οχήματος αυτή η γνώση της κίνησης είναι καθοριστικής σημασίας. Στην πράξη, για λόγους αξιοπιστίας και οικονομίας υπολογιστικών πόρων η εκτίμηση της κίνησης πραγματοποιείται χρησιμοποιώντας μεθόδους που βασίζονται σε πληροφορία από **επιπλέον εξοπλισμό** όπως ανιχνευτές μεταβολής φωτισμού, υπέρυθρες ακτίνες, ανίχνευση και παρακολούθηση πρότυπων αντικειμένων. Θέμα αυτής της διπλωματικής εργασίας είναι να εξετάσει κατά πόσο μια μέθοδος **εκτίμησης της κίνησης της κάμερας (ΕΚΚ)** με μοναδική πληροφορία τις ίδιες τις ακολουθίες εικόνων είναι σε θέση να παραγάγει εξίσου αξιόπιστα αποτελέσματα για την κίνηση και κατά συνέπεια για τη θέση της κάμερας στο χώρο.

Δυο είναι οι βασικές προσεγγίσεις για την ΕΚΚ. Η πρώτη από αυτές βασίζεται στην επεξεργασία της δισδιάστατης στιγμιαίας ταχύτητας γνωστής και ως **οπτική ροή** (optical flow). Η δεύτερη προσέγγιση βασίζεται στην επεξεργασία προκαθορισμένων χαρακτηριστικών σημείων της εικόνας τα οποία ανιχνεύονται και παρακολουθούνται ως προς την κίνησή τους μέσα στην ακολουθία. Στην παρούσα εργασία έχουμε υιοθετήσει την πρώτη προσέγγιση. Σύμφωνα με αυτήν, το πρόβλημα της ΕΚΚ γνωστό και ως **egomotion** χωρίζεται σε δυο βασικά υποπροβλήματα: αυτό της εύρεσης της οπτικής ροής και στη συνέχεια της περαιτέρω επεξεργασίας της προκειμένου να υπολογιστούν οι παράμετροι κίνησης δηλαδή η πλήρης **μεταφορά** (translation) και **περιστροφή** (rotation) της κάμερας στο χώρο.

Ένα αξιόλογο στοιχείο είναι ότι στην πραγματικότητα λόγω της προβολής του τρισδιάστατου περιβάλλοντος πάνω στο δισδιάστατο επίπεδο της εικόνας, ένα μέρος από την τρισδιάστατη πληροφορία, που αφορά κυρίως το βάθος (depth), χάνεται (για την ακρίβεια ενσωματώνεται στις άλλες δύο διαστάσεις). Ακόμα πιο απογοητευτικό, όπως θα δούμε, είναι το γεγονός ότι στην **προοπτική προβολή** (το **οπτικό μοντέλο** που θεωρούμε για την κάμερα μας) το βάθος και η μεταφορική κίνηση είναι μεγέθη ανάλογα και έτσι στην πραγματικότητα αυτό που έχουμε την δυνατότητα να υπολογίσουμε είναι μόνο ο λόγος μεταφοράς – βάθους και όχι οι απόλυτες τιμές τους, κάτι που όμως δημιουργεί μια παράπλευρη απώλεια: Με γνώση μόνο του λόγου μεταφοράς – βάθους είναι αδύνατο να προσδιορίσουμε την **ταχύτητα της κάμερας** παρά μόνο την σχετική ως προς το βάθος. Για να αναχαιτίσουμε την ιδιομορφία αυτή, συνήθης τακτική είναι να θεωρείται ότι η ταχύτητα της μεταφορικής κίνησης στην διάρκεια ενός καρέ έχει πάντα μια σταθερή τιμή. Υποθέτουμε ότι αυτή η σύμβαση πιθανώς να εγείρει πλήθος ερωτημάτων που όμως θα δούμε αργότερα. Προς το παρόν θα αρκεστούμε στο γεγονός ότι σε πολλές εφαρμογές (όπως την παθητική πλοήγηση) το ζητούμενο δεν είναι οι απόλυτες τιμές αλλά ο λόγος μεταφοράς – βάθους.

Διάφοροι ερευνητές στο παρελθόν προκειμένου να άρουν τις δυσκολίες από ιδιομορφίες όπως η παραπάνω επιχειρούσαν να απλοποιήσουν το πρόβλημα και τις εξισώσεις του π.χ. εξαλείφοντας το βάθος. Στην συνέχεια τις εξισώσεις αυτές τις χρησιμοποιούσαν σαν βάση για τις μεθόδους εκτίμησης που ανέπτυσσαν. Το θέμα ήταν ότι οι μετασχηματισμοί αυτοί πολλές φορές αλλοίωναν την γεωμετρία του οπτικού μοντέλου. Έτσι αρκετά συχνά διάφορες μέθοδοι ΕΚΚ που παρουσιάστηκαν κατά καιρούς στηριζόταν σε εξ’ αρχής λανθασμένα μοντέλα. Το αποτέλεσμα ήταν πόλωση δηλαδή «τάση» των εκτιμήσεων προς συγκεκριμένες κατευθύνσεις καθώς και αδικαιολόγητες αποκλίσεις που επιδεινωνόταν παρουσία θορύβου. Εδώ να τονίσουμε ότι η **πόλωση** (bias) και η **απόκλιση** (variance) αποτελούν σημαντικές μετρικές αποδοτικότητας των μεθόδων αυτών.

Ιδανικά, όπως θα δούμε και αργότερα, έξι μόνο διανύσματα οπτικής ροής αρκούν για να εξάγουμε τις παραμέτρους κίνησης. Όμως ο θόρυβος διαφθείρει με τέτοιο τρόπο το τελικό αποτέλεσμα ώστε η χρήση εκατονταπλάσιου αριθμού διανυσμάτων ροής κρίνεται απαραίτητη. Θα διαπιστώσουμε δε ότι μικρά σφάλματα στα διανύσματα ροής συνήθως δημιουργούν δυσανάλογα μεγάλα σφάλματα και αποκλίσεις στην ΕΚΚ. Επειδή ένα ποσοστό διανυσμάτων ροής για λόγους που θα αναλύσουμε στην συνέχεια παρουσιάζει ακραίες τιμές θορύβου, η χρήση μεγάλου αριθμού διανυσμάτων εξασφαλίζει ότι σε ένα βαθμό αυτές οι ακραίες τιμές θα εξομαλυνθούν. Αιτία των εσφαλμένων μετρήσεων είναι ότι οι αλγόριθμοι εύρεσης οπτικής ροής επιλύουν ένα από μόνο του αρκετά δύσκολο πρόβλημα και είναι ευαίσθητοι σε έναν αριθμό από καταστάσεις που παρουσιάζονται στις πραγματικές ακολουθίες εικόνων όπως η ανεξάρτητη κίνηση και ο θόρυβος των CCD's – κωδικοποίησης που έχει σαν αποτέλεσμα την διακύμανση της έντασης των εικονοστοιχείων από καρέ σε καρέ. Το αποτέλεσμα είναι μεγάλος αριθμός από εσφαλμένα δεδομένα τα οποία ονομάζονται και **outliers**. Αυτονόητη είναι ίσως η υπόθεση ότι η κάμερα θα πρέπει να εστιάζει σε **αμετάβλητο πλάνο** (rigid motion) δηλ τα αντικείμενα επιβάλλεται να έχουν σταθερό σχήμα, να μην κινούνται προς ανεξάρτητες κατευθύνσεις και η μόνη κίνηση των αντικειμένων στα καρέ της ακολουθίας να προκαλείται από την κίνηση της κάμερας. Υπό αυτούς τους όρους έχουμε την δυνατότητα να χειριστούμε καλύτερα τις θορυβώδεις μετρήσεις οπτικής ροής.

Στην μέθοδο μας έχουμε επιχειρήσει να ελέγξουμε την ύπαρξη των outliers χρησιμοποιώντας δύο επίπεδα ελέγχου: ανίχνευση και απομάκρυνση των διανυσμάτων με ακραίες τιμές (ανεξαρτήτως μοντέλου κίνησης) και σε δεύτερο στάδιο ενσωμάτωση **robust τεχνικής στάθμισης** στην μέθοδο ΕΚΚ. Παρά την ενίσχυση της «αντοχής» της μεθόδου εκτίμησης ώστε να ανέχεται ένα σεβαστό ποσοστό θορύβου, όπως θα αποδειχθεί και στην συνέχεια, το επίπεδο εσφαλμένων δεδομένων που παράγεται από τους πλέον διαδεδομένους αλγορίθμους οπτικής ροής [20] πολλές φορές είναι αρκετά υψηλό ώστε να «αλλοιώνει» το τελικό αποτέλεσμα.

Σειρά από εκτεταμένες δοκιμές με συνθετικές ακολουθίες σε ελεγχόμενο περιβάλλον θορύβου αποκαλύπτουν πολλά και ενδιαφέροντα στοιχεία για την αξιοπιστία αλλά και τις προσδοκίες που θα πρέπει να τρέφουμε για την μέθοδο που αναπτύξαμε. Δυστυχώς πρόβλημα για την αξιολόγηση της μεθόδου αποτελεί η έλλειψη αντικειμενικών μετρήσεων για πραγματικές ακολουθίες εικόνων. Παρά ταύτα η επεξεργασία και ανάλυση πραγματικών ακολουθιών δίνει μεν μια χονδρική αλλά παράλληλα και αξιόλογη αίσθηση της αξιοπιστίας της μεθόδου. Απόπειρα να ενισχυθεί η αίσθηση αυτή πραγματοποιήθηκε με την ανάπτυξη και ενσωμάτωση μιας μηχανής παραγωγής κινούμενων γραφικών σημείων που ακολουθούν την κίνηση του περιβάλλοντος.

Η δομή της εργασίας έχει την εξής: Στο **κεφάλαιο 2** πραγματευόμαστε το ζήτημα της οπτικής ροής και αναλύουμε την μέθοδο που υιοθετήσαμε. Στο **κεφάλαιο 3** αναλύεται το οπτικό μοντέλο και οι εξισώσεις κίνησης που το διέπουν. Εκεί θα αναφερθούμε στις εξισώσεις που συνδέουν τον «πραγματικό κόσμο» με την προβολή του στο επίπεδο των εικόνων. Στην συνέχεια στο **κεφάλαιο 4** αναλύεται ο αλγόριθμος ΕΚΚ όπου και αξιολογείται μέσω εκτεταμένων προσομοιώσεων στο **κεφάλαιο 5**. Τέλος παρουσιάζουμε τα συμπεράσματα μας και τις ιδέες μας για μελλοντικές προεκτάσεις στο **κεφάλαιο 6**.

Κεφάλαιο 2

Οπτική ροή

2.1 Εισαγωγή

Όπως είναι γνωστό οι μηχανές κινηματογράφησης συλλαμβάνουν 25 καρέ το δευτερόλεπτο (frames per second, FPS) και οι διαφορές μεταξύ αυτών των καρέ μπορούν να αποτελέσουν σημαντική πηγή δεδομένων. Αν η κάμερα κινείται σε σχέση με την τρισδιάστατη σκηνή που λαμβάνει, η προκύπτουσα φαινομενική κίνηση στα διαφορά σημεία της εικόνας ονομάζεται **οπτική ροή**. Η οπτική ροή ουσιαστικά περιγράφει την σχετική κίνηση μεταξύ της κάμερας και του περιβάλλοντος λήψης στα διάφορα σημεία της εικόνας. Αυτή η σχετική κίνηση από καρέ σε καρέ εμπεριέχει εν δυνάμει πληθώρα πληροφοριών οι οποίες μπορούν να εξαχθούν με κατάλληλη επεξεργασία. Π.χ. κατά την κινηματογράφηση από ένα κινούμενο όχημα τα μακρινά αντικείμενα φαίνεται να κινούνται αργά σε σχέση με τα κοντινότερα, κατά συνέπεια η μέτρηση του ρυθμού της κίνησης μπορεί να μας δώσει στοιχεία για την απόσταση της κάμερας από τα αντικείμενα. Συνοψίζοντας θα μπορούσαμε να πούμε ότι γενικώς η οπτική ροή «κωδικοποιεί» μεγάλη ποσότητα χρήσιμης πληροφορίας που σχετίζεται με την δομή της σκηνής.

Πρακτικά η οπτική ροή είναι ένα σύνολο **δισδιάστατων διανυσμάτων** στα διάφορα σημεία της εικόνας. Καθένα από αυτά τα διανύσματα δεν είναι τίποτα άλλο παρά η **προβολή** στο επίπεδο της εικόνας του αντίστοιχου «τρισδιάστατου διανύσματος» κίνησης ενός αντικειμένου στον χώρο. Συνεπώς η οπτική ροή είναι ένα σύνολο διανυσμάτων $\boldsymbol{u} = (u_1, u_2)^T$ καθένα από τα οποία μας δείχνει την κίνηση των προβαλλόμενων αντικειμένων στην θέση $\boldsymbol{x} = (x_1, x_2)^T$ της εικόνας.

Χωρίς αμφιβολία ο αξιόπιστος υπολογισμός της οπτικής ροής έχει αποτελέσει αντικείμενο εκτεταμένης έρευνας και πολλές εναλλακτικές μέθοδοι έχουν αναπτυχθεί για τον σκοπό αυτό με ίσως πιο διαδεδομένες τις τεχνικές που βασίζονται στην ισχύ της ιδιότητας ότι η φωτεινότητα των εικονοστοιχείων σε διαδοχικά καρέ δεν αλλάζει αλλά μετακινείται σε νέες θέσεις. Με άλλα λόγια κάθε διάνυσμα του πεδίου μπορεί να προκύψει αντιστοιχίζοντας σημεία με παρόμοια φωτεινότητα σε διαδοχικά καρέ. Οι τεχνικές αυτές είναι γνωστές ως **διαφορικές** και παρακάτω θα δούμε τον τρόπο με τον οποίο κατορθώνουν να εκμεταλλευτούν το χαρακτηριστικό αυτό. Σημειώνουμε ότι από εδώ και στο εξής θα αναφερόμαστε στη φωτεινότητα ως **ένταση εικονοστοιχείου**.

Οι **τεχνικές διαφόρισης** θέτουν ως βασική υπόθεση ότι η ένταση $I(\mathbf{x},t)$ σε ένα συγκεκριμένο εικονοστοιχείο με θέση $\mathbf{x}$ στο καρέ της χρονικής στιγμής t, παραμένει σταθερή και στα επόμενα καρέ της ακολουθίας ανεξάρτητα από το αν έχει μεταφερθεί σε νέα θέση $\mathbf{x}'$. Κάτι τέτοιο θα μπορούσαμε να πούμε ότι ισχύει μόνο υπό συνθήκες, μιας και είναι περισσότερο σύνηθες η ένταση στα διαφορά σημεία να αλλάζει ανάλογα με την γωνία λήψης, τον θόρυβο λήψης – συμπίεσης κ.α. Γι' αυτό θεωρείται αυτονόητο ότι τα καρέ θα πρέπει να επεξεργαστούν με κατάλληλο τρόπο (φίλτρο συχνοτήτων) ώστε να απομακρυνθεί μέρος του θορύβου πριν δοθούν ως είσοδος στον αλγόριθμο οπτικής ροής. Όταν δεν παραβιάζεται η συνθήκη «αμετάβλητης έντασης» (και επιπλέον η εικόνα έχει πλούσιο συχνοτικό περιεχόμενο) οι αλγόριθμοι αυτοί αποδίδουν άριστα αποτελέσματα. Όλα αυτά θα τα δούμε αναλυτικά στις επόμενες ενότητες.

2.2 Τεχνικές διαφόρισης

2.2.1 Τεχνικές διαφόρισης πρώτης τάξης

Όπως προδίδει και το όνομα τους οι τεχνικές διαφόρισης αξιοποιούν την πληροφορία που προκύπτει από την παραγώγιση των εικόνων στο χώρο και στο χρόνο. Χωρίζονται δε σε δυο βασικές κατηγορίες ανάλογα με το αν γίνεται χρήση της πρώτης ή και της δεύτερης παραγώγου. Στην περίπτωση της πρώτης παραγώγου η οπτική ροή μεταξύ δύο καρέ στους χρόνους 0 και t είναι το διάνυσμα $\mathbf{u} = (u_1, u_2)^T$ το οποίο ικανοποιεί την σχέση:

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

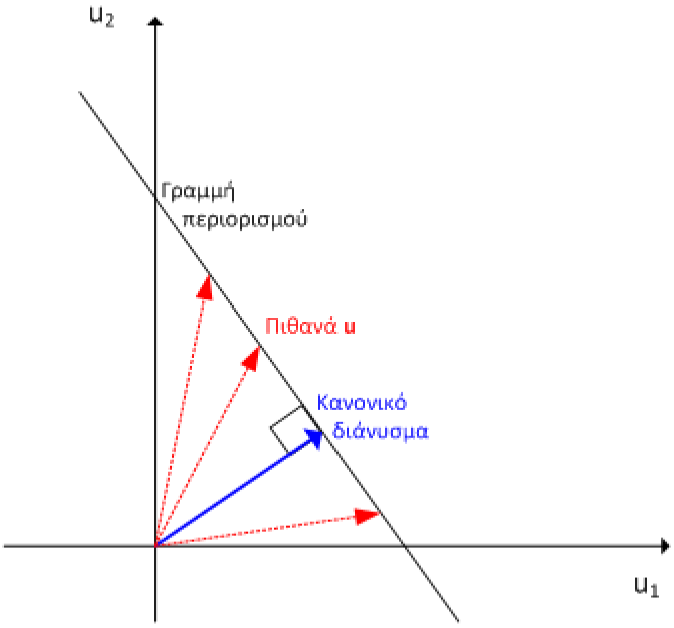
Η (1) εκφράζει την κύρια υπόθεση των τεχνικών διαφόρισης περί «αμετάβλητης έντασης» και είναι αναγκαίο να ικανοποιείται. Εννοεί δε ότι τα δυο εικονοστοιχεία πριν και μετά την μετατόπιση ***u*** έχουν ακριβώς την ίδια ένταση. Χρησιμοποιώντας την επέκταση Taylor (ή ισοδύναμα την εξίσωση περιορισμού κλίσης $\frac{dI(\mathbf{x},t)}{dt}=0$) έχουμε:

$$\nabla I(\mathbf{x},t)\cdot\mathbf{u} + I_t(\mathbf{x},t) = 0 \tag{2}$$

Όπου $I_t(\mathbf{x},t)$ είναι η χρονική παράγωγος του $I(\mathbf{x},t)$ και $\nabla I(\mathbf{x},t) = (I_x(\mathbf{x},t), I_y(\mathbf{x},t))^T$ με $I_x(\mathbf{x},t)$, $I_y(\mathbf{x},t)$ τις μερικές χωρικές παραγώγους για τον *x* και *y* άξονα της εικόνας αντίστοιχα. Η (2) είναι μια γραμμική εξίσωση με αγνώστους τα στοιχεία του διανύσματος ***u***. Αρκεί έτσι να υπολογίσουμε τις μερικές χωροχρονικές παραγώγους $I_x(\mathbf{x},t)$, $I_y(\mathbf{x},t)$, $I_t(\mathbf{x},t)$ για να έχουμε την μία από τις δύο εξισώσεις που χρειαζόμαστε για την επίλυση συστήματος με αγνώστους τα u_1, u_2. Επειδή τα πιθανά ***u*** που ικανοποιούν την εξίσωση βρίσκονται σε διάταξη ευθείας η (2) ονομάζεται και **ευθεία περιορισμού** (constraint line). Το διάνυσμα δε το οποίο είναι κάθετο στην ευθεία περιορισμού ονομάζεται **κανονικό** (normal flow) και είναι ιδιαίτερα εύκολο να υπολογιστεί (Σχήμα 2-1). Αν υποθέσουμε το κανονικό διάνυσμα $\mathbf{u}_n = s\mathbf{n}$ όπου *s* η ταχύτητα και ***n*** η διεύθυνση τότε:

$$s(\mathbf{x},t) = \frac{-I_t(\mathbf{x},t)}{\left\|\nabla I(\mathbf{x},t)\right\|}, \quad \mathbf{n}(\mathbf{x},t) = \frac{\nabla I(\mathbf{x},t)}{\left\|\nabla I(\mathbf{x},t)\right\|} \tag{3}$$

Με την (3) μπορούμε να υπολογίσουμε ένα πεδίο κανονικών διανυσμάτων ωστόσο αυτό δεν καλύπτει της ανάγκες της εφαρμογής που εξετάζουμε εδώ αφού η ΕΚΚ απαιτεί ακριβές πεδίο οπτικής ροής αποτελούμενο από πλήρη δισδιάστατα ***u***. Είναι σαφές πως για να καθορίσουμε επακριβώς τις τιμές του ***u*** είναι απαραίτητες επιπρόσθετες εξισώσεις – περιορισμοί.



Σχήμα 2-1: «*Ευθεία περιορισμού» διανύσματος οπτικής ροής*

Μία λύση για την εύρεση του ***u*** απαντάται στις τεχνικές δεύτερης τάξεως οι οποίες ουσιαστικά χρησιμοποιούν επιπλέον εξισώσεις που προέρχονται από την παραγώγιγη της (2). Σε πολλές των περιπτώσεων οι τεχνικές αυτές δεν παράγουν ικανοποιητικά αποτελέσματα [18]. Μια άλλη ιδέα που εξασφαλίζει τις επιπλέον εξισώσεις που απαιτούνται προέρχεται από την εκμετάλλευση της κοινής συμπεριφοράς που χαρακτηρίζει γειτονικά εικονοστοιχεία κατά την κίνηση τους [18,19,24,25].

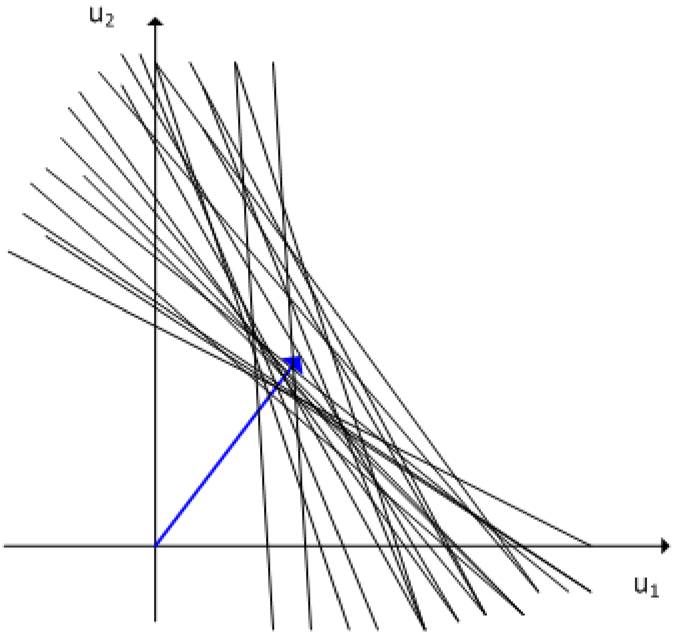
2.2.2 Τοπικές τεχνικές διαφόρισης πρώτης τάξης

Πλήθος μεθόδων που έχουν ως αφετηρία την (2) είναι διαθέσιμο [20]. Οι μέθοδοι αυτές ταξινομούνται σε δυο κύριες κατηγορίες, τις *καθολικές* και *τοπικές* τεχνικές, ανάλογα με τις επιπλέον εξισώσεις – περιορισμούς που υιοθετούν:

- Οι **καθολικές τεχνικές** λαμβάνουν υπόψη τους ολόκληρη την εικόνα αποδίδοντας οπτική ροή ακόμα και σε σημεία που το περιεχόμενο της εικόνας αντενδείκνυνται για επεξεργασία. Τυπικά παραδείγματα περιοχών εικόνας από τις οποίες δεν παρέχεται αξιόπιστη πληροφορία οπτικής ροής είναι π.χ. οι μονοχρωματικές επιφάνειες (καθαρός ουρανός, επιφάνειες στο παρασκήνιο – τοίχοι) όπου οι παράγωγοι I_x, I_y, I_t τείνουν στον μηδέν. Χρησιμοποιείτε δε σε εφαρμογές που είναι απαραίτητη **πυκνή οπτική ροή** αφού ιδανικά παράγει μετρήσεις ταχύτητας για κάθε εικονοστοιχείο (έστω και χαμηλής αξιοπιστίας!). Ο αλγόριθμος ΕΚΚ δεν απαιτεί πυκνή οπτική ροή και χονδρικά θα μπορούσαμε να πούμε ότι χρειαζόμαστε πληροφορία από λίγα και ακριβή διανύσματα. Για αυτό το λόγο στραφήκαμε στις τοπικές τεχνικές.

- Οι **τοπικές τεχνικές** λαμβάνουν υπόψη τους μικρά τμήματα της εικόνας για κάθε διάνυσμα που παράγουν. Κάθε διάνυσμα είναι εντελώς ανεξάρτητο από τα υπόλοιπα. Χρησιμοποιούν δε πληροφορία από διαδοχικά εικονοστοιχεία που βρίσκονται εντός κάποιου μικρού **παραθύρου** εκμεταλλευόμενες κοινές ιδιότητες που παρουσιάζονται κατά την κίνηση. Γενικότερα θεωρούνται ακριβείς αλλά πάσχουν από κάποιες αδυναμίες που θα πρέπει να εξετάζονται και αντιμετωπίζονται ανάλογα με την εφαρμογή και την ανοχή αυτής στα σφάλματα. Αυτές τις αδυναμίες και τα σφάλματα που δημιουργούν στις εκτιμήσεις θα εξετάσουμε στο κεφάλαιο 2.3.

Η βασική ιδέα στην οποία στηρίζονται οι τοπικές τεχνικές προέρχεται από το γεγονός ότι είναι αρκετά πιθανό γειτονικά εικονοστοιχεία να συμμετέχουν στην ίδια κίνηση. Εκμεταλλευόμενοι το θέμα της **τοπικότητας** στην κίνηση μπορούμε να υπολογίσουμε ένα πλήρες δισδιάστατο διάνυσμα ροής συνδυάζοντας τις εκτιμήσεις από γειτονικά εικονοστοιχεία: Όπως αναλύσαμε πριν, μέσω της (2) είναι δυνατός ο υπολογισμός της ευθείας γραμμής που «συνδέει» τα δύο στοιχεία του διανύσματος ***u*** μεταξύ τους. Αν λοιπόν σχηματίσουμε τέτοιες ευθείες για όλα τα γειτονικά εικονοστοιχεία τότε η τελική λύση είναι η τομή τους ή για να ακριβολογούμε (παρουσία θορύβου) το πλησιέστερο προς την τομή τους σημείο (Σχήμα 2-2). Έτσι αφού πρώτα υπολογίσουμε τις χωροχρονικές παραγώγους για την επιθυμητή θέση ***x*** και τις γειτονικές της, στην συνέχεια λαμβάνουμε την λύση επιλύοντας το υπερκαθορισμένο σύστημα γραμμικών εξισώσεων που προκύπτει. Οι εξισώσεις του συστήματος είναι όσες και ο αριθμός των εικονοστοιχείων που περιλαμβάνονται στο μπλοκ γειτόνων. Μια συνηθισμένη επιλογή είναι η συμμετοχή των δυο προηγούμενων και δυο επόμενων διαδοχικών εικονοστοιχείων για καθέναν από τους *x-y* άξονες, το οποίο αντιστοιχεί σε παράθυρο 5x5. Σε αυτή την περίπτωση το σύστημα μας περιέχει 25 γραμμικές εξισώσεις της μορφής (2).



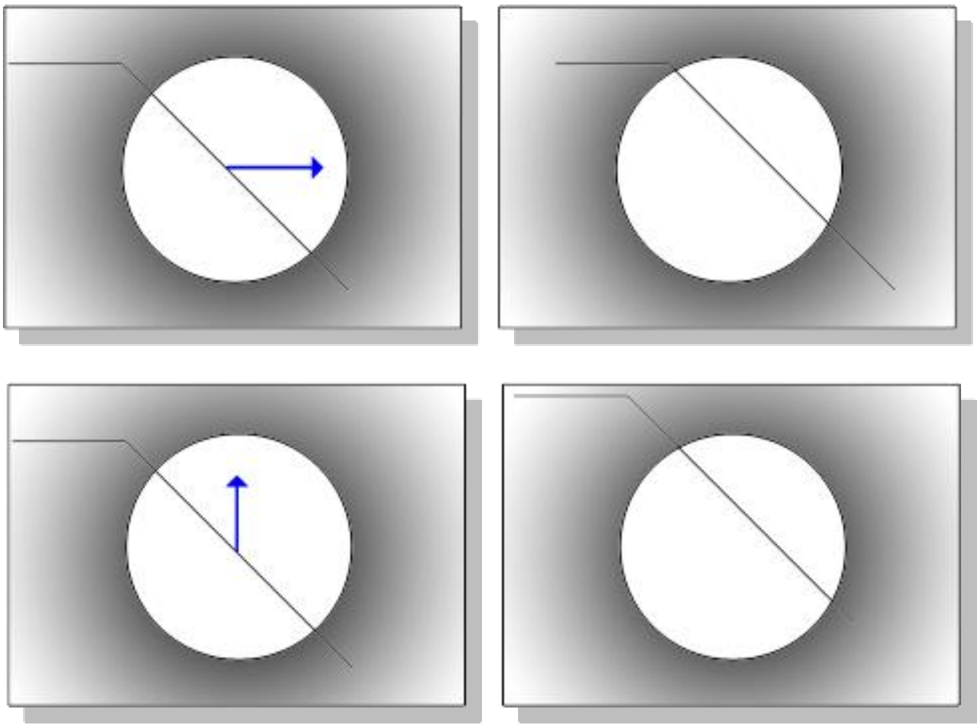
Σχήμα 2-2: *Υπολογισμός δισδιάστατου διανύσματος οπτικής ροής μέσω ευθειών περιορισμού γειτονικών σημείων.*

2.3 Παράμετροι προβλήματος και αιτίες σφαλμάτων

Όπως θα δούμε αργότερα οι τοπικές τεχνικές πρώτης τάξης παρά το ότι θεωρούνται από τις πλέον ακριβείς τα αποτελέσματα τους εμπεριέχουν ένα αξιοσημείωτο ποσοστό σφαλμάτων. Εκτός από τις προφανείς αιτίες σφαλμάτων όπως λάθη στους υπολογισμούς των παραγώγων κάποιες επιπλέον προβληματικές περιπτώσεις επηρεάζουν την ποιότητα των αποτελεσμάτων. Από εδώ και στο εξής τα σφάλματα στο πεδίο οπτικής ροής θα τα αναφέρουμε ως **θόρυβο**.

2.3.1 Φαινόμενο διαφράγματος (Aperture problem)

Είναι το φαινόμενο κατά το οποίο δεν υπάρχει σαφήνεια για την κίνηση ενός αντικειμένου αν το παρατηρείς ξέχωρα από το υπόλοιπο περιβάλλον μέσα από ένα μικρό παράθυρο. Γενικότερα ένα μειονέκτημα των τοπικών μεθόδων σε σχέση με τις καθολικές απορρέει από το γεγονός ότι δεν εκμεταλλεύονται την γενικότερη κίνηση σε όλο το μήκος και πλάτος της εικόνας αλλά «παρακολουθούν» αποσπασματικά μικρά παραθυράκια κίνησης (Σχήμα 2-3). Το φαινόμενο δε παρουσιάζεται συχνά σε περιπτώσεις ακμών που εκτείνονται κατά μήκος ευθείας διάταξης μπροστά από το παράθυρο.



Σχήμα 2-3: *Ένα αντικείμενο που κινείται προς διαφορετικές κατευθύνσεις (αριστερή στήλη) όμως η τελική θέση (δεξιά στήλη) δείχνει ίδια όταν «παρατηρούμε» μέσα από μια μικρή οπή.*

Μια λύση στο πρόβλημα αυτό είναι η αύξηση του μεγέθους του παραθύρου δηλαδή της αύξησης των γειτονικών εικονοστοιχείων που συμμετέχουν στην δημιουργία του βασικού συστήματος. Κάτι τέτοιο ναι μεν βελτιώνει τις εσφαλμένες εκτιμήσεις λόγω του φαινομένου του διαφράγματος αλλά έχει αρνητικές συνέπειες σε μια άλλη κρίσιμη παράμετρο καθώς με την αύξηση του παραθύρου αυξάνεται και η πιθανότητα ανεξάρτητης κίνησης στα γειτονικά εικονοστοιχεία. Οι ανεξάρτητες κινήσεις και η επίδρασή τους στο αποτέλεσμα συζητώνται αμέσως μετά.

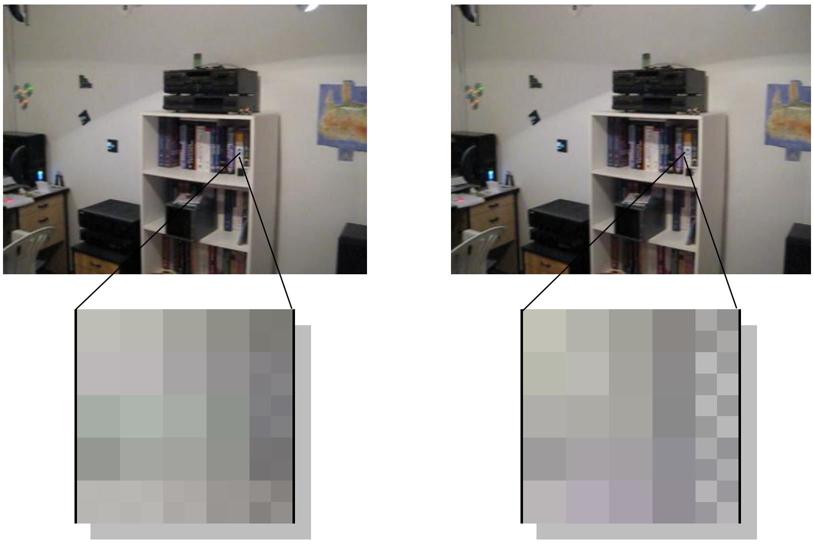
2.3.2 Ανεξάρτητη κίνηση – θόρυβος

Όπως είδαμε βασική προϋπόθεση για ακριβείς μετρήσεις ταχύτητας είναι η σκηνή να είναι αμετάβλητη στο χρόνο. Αν έχουμε ένα σχετικά μικρό παράθυρο με *ν* γειτονικά εικονοστοιχεία έντασης $I(\mathbf{x}_{1\ldots\nu},t)$ την χρονική στιγμή *t*… την επόμενη στιγμή *t+1* αναμένουμε $I(\mathbf{u}+\mathbf{x}_{1\ldots\nu},t+1) = I(\mathbf{x}_{1\ldots\nu},t)$. Δηλαδή παρά την μετατοπισμένη θέση του παραθύρου κάθε ένα από τα εικονοστοιχεία διατηρεί αντίστοιχα την ένταση του αναλλοίωτη. Ουσιαστικά τίποτα περισσότερο ή λιγότερο από την ικανοποίηση της θεμελιώδους σχέσης (1) που ορίσαμε προηγούμενα για όλα τα εικονοστοιχεία του συνόλου *ν*. Κάτι τέτοιο όπως ίσως θα μπορούσε να μαντέψει κάποιος δεν είναι απολύτως δυνατό να επιτευχθεί κυρίως για δυο λόγους: Την ανεξάρτητη κίνηση και το θόρυβο.

Μία σημαντική διαφορά ανάμεσα στον τρόπο που επιδρά ο θόρυβος σε σχέση με τις ανεξάρτητες κινήσεις είναι ότι ο θόρυβος προκαλεί σφάλματα στον υπολογισμό των παραγώγων, ενώ οι ανεξάρτητες κινήσεις ακυρώνουν την βασική προϋπόθεση των τοπικών τεχνικών περί αμετάβλητου περιβάλλοντος λήψης (*rigid scene*). Ωστόσο κατά μιαν ευρύτερη έννοια ο θόρυβος φέρει και τις ιδιότητες των ανεξάρτητων κινήσεων.

Όταν μία σκηνή περιέχει ανεξάρτητες κινήσεις τότε είναι απλά θέμα «τύχης» το παράθυρο να βρίσκεται ακριβώς πάνω στην «ακμή κίνησης» των αντικειμένων. Τα διανύσματα που υπολογίζονται από αυτές τις περιοχές έχουν σχεδόν τυχαία κατεύθυνση. Σε περιπτώσεις που γνωρίζουμε εκ των προτέρων ότι η σκηνή περιέχει ανεξάρτητες κινήσεις είναι απαραίτητη η χρήση μικρότερου παραθύρου, το οποίο όπως αναλύθηκε πριν επιδεινώνει το φαινόμενο διαφράγματος, καθώς και μειώνει την ικανότητα της ανίχνευσης μεγάλων ταχυτήτων, όπως θα περιγράψουμε αργότερα.

Ο θόρυβος στις εικόνες επιφέρει παρόμοια προβλήματα αντίστοιχα των επιπτώσεων από τις ανεξάρτητες κινήσεις. Ωστόσο οι ανεξάρτητες κινήσεις συμβαίνουν σε μεμονωμένα σημεία, ενώ αντίθετα ο θόρυβος λήψης-συμπίεσης αφορά ολόκληρη την επιφάνεια της εικόνας και η επίδραση του είναι αδιάλειπτη. Επιπλέον στοιχείο που δυσχεραίνει τυχόν προσπάθειες αντιμετώπισής του είναι η πρότερη άγνοια μας για το μέγεθος και τον τύπο του θορύβου που περιέχεται σε μια εικόνα αφού αυτά εξαρτώνται από το υλικό, δηλαδή την μηχανή κινηματογράφησης, τους φακούς, την φωτοευαίσθητη επιφάνεια, το μέσο αποθήκευσης, την συμπίεση κ.α (Σχήμα 2-4).



Σχήμα 2-4: *Διαδοχικά καρέ από ακολουθία (πάνω). Παρατηρούμε τις εντάσεις των εικονοστοιχείων σε παράθυρο 5x5 για το ίδιο αντικείμενο, Ιδανικά τα δυο παράθυρα θα έπρεπε να είναι ίδια (κάτω).*

Ωστόσο παρότι είναι δύσκολο να γνωρίζουμε τον τύπο και την ποσότητα του θορύβου που θα συναντήσουμε κατά την διάρκεια μιας ακολουθίας εικόνων θα πρέπει να θεωρήσουμε κάποια χαρακτηριστικά που στις περισσότερες των περιπτώσεων θα καλύπτουν τις συνθήκες που συναντάμε στην πράξη. Για τις ανάγκες της παρούσας ανάλυσης θεωρήσαμε ότι ο θόρυβος αφορά μόνο την ένταση των εικονοστοιχείων, δεν επηρεάζει τα σχήματα δημιουργώντας υφές, είναι ανεξάρτητος και ομοιόμορφα κατανεμημένος με μηδενική μέση τιμή. Τέλος όσο πιο «έντονος» είναι ο θόρυβος τόσο πιο «ψηλά» βρίσκεται το συχνοτικό του περιεχόμενο. Παρακάτω δείχνουμε με την εφαρμογή βαθυπερατού φίλτρου στις εικόνες πριν την επεξεργασία εξομαλύνουμε σε μεγάλο βαθμό τον θόρυβο κάτι που δρα ευεργετικά στην ποιότητα του τελικού αποτελέσματος.

2.3.3 Μέγιστη ταχύτητα κίνησης

Ως ταχύτητα κίνησης ορίζουμε το μέτρο των διανυσμάτων προς το χρόνο ενός καρέ. Επειδή το χρονικό διάστημα ανάμεσα στα καρέ της ακολουθίας είναι σταθερό συνηθίζεται να μετράμε την ταχύτητα και ως εικονοστοιχεία ανά καρέ. Ποια είναι όμως η μέγιστη μετρίσιμη ταχύτητα;

Το άνω όριο στην ταχύτητα αποδεικνύεται πως σχετίζεται με το μέγεθος του παραθύρου και με την ανάλυση (resolution) των εικόνων της ακολουθίας. Ως γενικός κανόνας ισχύει ότι όσο μεγαλύτερο είναι το παράθυρο τόσο μεγαλύτερη ταχύτητα μπορεί να μετρηθεί. Όμως η αύξηση του μεγέθους του παραθύρου έχει αρνητικές επιπτώσεις σε μια σειρά από άλλες παραμέτρους, μερικές από τις οποίες είδαμε πριν, και γι' αυτό δεν συνιστάτε. Μία λύση για την μέτρηση μεγάλων ταχυτήτων απαντάται στις **τεχνικές πολλαπλής ανάλυσης** γνωστές και ως *coarse-to-fine*, θέμα το οποίο θα δούμε αργότερα στην ενότητα 2.6.1. Τυπικά ταχύτητες μεγαλύτερες από 2 εικονοστοιχεία ανά καρέ θεωρούνται μεγάλες και δημιουργούν μειωμένης ακρίβειας πρώτες παραγώγους (*aliasing* στον άξονα του χρόνου). Σε αντιστοιχία με το παράθυρο ο ρόλος της ανάλυσης είναι φανερός: ένα παράθυρο μεγέθους 8x8, ενδεχομένως θεωρείται μεγάλο αν χρησιμοποιηθεί σε εικόνες ανάλυσης 640x480*p* παρέχοντας μας την δυνατότητα να υπολογίζουμε μεγάλες μετατοπίσεις, ενώ αντίθετα θα θεωρούνται μικρό αν οι διαστάσεις της εικόνας ήταν μεγαλύτερες π.χ. 1920x1080*p*.

2.3.4 Συμπεράσματα: Μέγεθος «Παραθύρου»

Γίνεται σαφές ότι το μέγεθος του παραθύρου είναι μια σημαντική παράμετρος που επηρεάζει αντιφατικά μια σειρά από θέματα κρίσιμα για την ακρίβεια των εκτιμήσεων. Παρατηρούμε ότι δεν υπάρχει «χρυσός» κανόνας που να καθορίζει το ιδανικό μέγεθος παραθύρου αλλά αυτό θα πρέπει να επιλέγεται κατά περίπτωση. Ο πίνακας 2-1 δείχνει πως επηρεάζονται οι παράμετροι που είδαμε στις προηγούμενες ενότητες από το μέγεθος του παραθύρου.

Μέγεθος παραθύρου / ***Επενέργεια***	**Αύξηση μεγέθους παραθύρου**	**Μείωση μεγέθους παραθύρου**
Πυκνότητα οπτικής ροής	-	+
Φαινόμενο διαφράγματος	+	-
Ανεξάρτητη κίνηση, θόρυβος	-	+
Ταχύτητα κίνησης	+	-

Σημείωση: + = βελτίωση
- = επιδείνωση

Πίνακας 2-1: *Το μέγεθος παραθύρου επηρεάζει ποικιλοτρόπως τις μετρήσεις.*

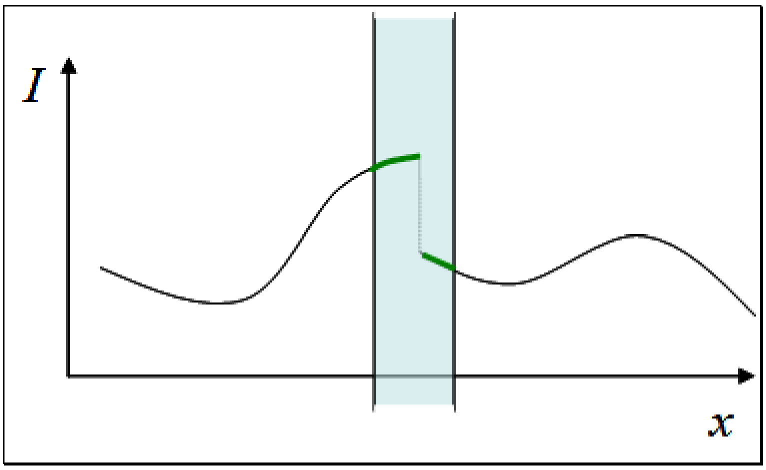
Ως γενικό συμπέρασμα θα πρέπει να σημειώσουμε ότι οι προβληματικές καταστάσεις που περιγράψαμε στις προηγούμενες ενότητες άλλοτε επηρεάζουν τον αξιόπιστο υπολογισμό των παραγώγων και άλλοτε το μοντέλο διαφορικής επίλυσης. Σε κάθε περίπτωση μια «ισορροπημένη» επιλογή για το μέγεθος του παραθύρου μετριάζει την αρνητική συνεισφορά των παραπάνω παραμέτρων στο αποτέλεσμα και αυξάνει την πιθανότητα για επιτυχείς εκτιμήσεις.

2.4 Υπολογισμός παραγώγων

Η θεωρία γύρω από την αριθμητική διαφόριση είναι αρκετά περίπλοκη και εμπεριέχει θέματα δειγματοληψίας στον χώρο και στον χρόνο. Τυπικές αιτίες προβλημάτων που δρουν αρνητικά στην ποιότητα της παραγώγισης είναι ο θόρυβος (2.3.2) και κάποια ιδιαίτερα χαρακτηριστικά των εικόνων που αντενδείκνυνται για την διαδικασία.

Ουσιαστικά η χωροχρονική διαφόριση ισοδυναμεί με την συνέλιξη της εικόνας με γραμμικό υψιπερατό φίλτρο και ως εκ τούτου η ακρίβεια εξαρτάται από τον τύπο και το μέγεθος της μάσκας του φίλτρου. Η επιλογή της μάσκας του φίλτρου αποτελεί ζήτημα προς διερεύνηση αφού τίθεται θέμα παραχωρήσεων που πρέπει να γίνουν ανάμεσα στην ακρίβεια των παραγώγων και τον χρόνο απόκρισης του αλγορίθμου. Σε αυτή την εργασία χρησιμοποιήσαμε ένα φίλτρο *τεσσάρων σημείων κεντρικής διαφοράς (4-point central differences)* με συντελεστές $\frac{1}{12}(-1,8,0,-8,1)$.

Σημαντική για την επιτυχία της παραγώγιγης είναι η διαδικασία της εξομάλυνσης αφού (εκτός από την μείωση του θορύβου) βοήθα στο να μετριαστούν κάποιες δυσμενείς για την διαφόριση περιπτώσεις όπως τα μεγάλα σφάλματα που συνήθως παρουσιάζονται σε περιοχές των εικόνων με πολύ υψηλές συχνότητες και ασυνέχειες (Σχήμα 2-5).



Σχήμα 2-5: *Ασυνέχειες στην ένταση κατά τον άξονα x λόγω ύπαρξης ακμής. Το ίδιο μπορεί να συμβαίνει κατά τον άξονα του y ή ακόμα του χρόνου t.*

2.5 Εξομάλυνση

Συχνά τα δεδομένα που καταγράφονται από κάποιο μέσο εκτός από την ωφέλιμη πληροφορία εμφανίζουν διάφορες τυχαίες διακυμάνσεις από δείγμα σε δείγμα, τον θόρυβο. Σκοπός της εξομάλυνσης είναι να εξαλείψει σε κάποιο βαθμό την τυχαία αυτή μεταβολή που αλλοιώνει την πληροφορία.

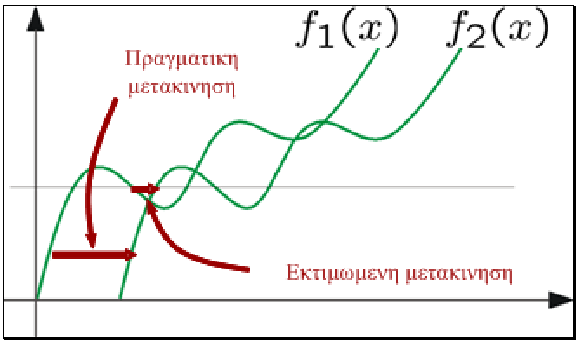
Για τα δεδομένα που καταγράφονται από μία κάμερα, τις ακολουθίες εικόνων, ως εξομάλυνση θα μπορούσαμε να ορίσουμε την πρόβλεψη της τιμής ενός εικονοστοιχείου συνυπολογίζοντας πληροφορίες από τα περιβάλλοντα εικονοστοιχεία ή ακόμα την πρόβλεψη τιμής ενός εικονοστοιχείου σε κάποιο καρέ *t* με δεδομένες μαρτυρίες από προηγούμενες και επόμενες τιμές του στον χρόνο.

Ένας τρόπος για να εξομαλύνουμε τις υψηλές συχνότητες στις οποίες εμπεριέχεται και ο θόρυβος είναι να αποδώσουμε σε κάθε εικονοστοιχείο το μέσο όρο των γειτονικών του. Όμως το ερώτημα είναι πόσα από τα γειτονικά εικονοστοιχεία θα πρέπει να ληφθούν υπόψη. Μια λύση που είναι γνωστό ότι αποδίδει ικανοποιητικά αποτελέσματα, ειδικά σε περίπτωση που το μοντέλο θορύβου της εικόνας είναι Gaussian, είναι ο σταθμισμένος μέσος όρος με χρήση ***Gaussian φίλτρου***. Με την εφαρμογή του Gaussian φίλτρου ουσιαστικά αντικαθιστούμε την ένταση $I(\mathbf{x},t)$ με την νέα ένταση $I(\mathbf{x},t)*G$ όπου '*' δηλώνει την συνέλιξη της ακολουθίας με την κατανομή G. Με την σειρά της η G μπορεί να είναι η δισδιάστατη Gaussian κατανομή G_{2D} που επιδρά αποκλειστικά στα εικονοστοιχεία της εικόνας ξέχωρα από τις μεταβολές στην ένταση που εμφανίζονται κατά την διάρκεια της ακολουθίας ή η τρισδιάστατη Gaussian κατανομή G_{3D} σε περίπτωση που επιθυμούμε κατά την εξομάλυνση να λαμβάνονται υπόψη και οι μεταβολές στην ένταση κατά τον άξονα του χρόνου:

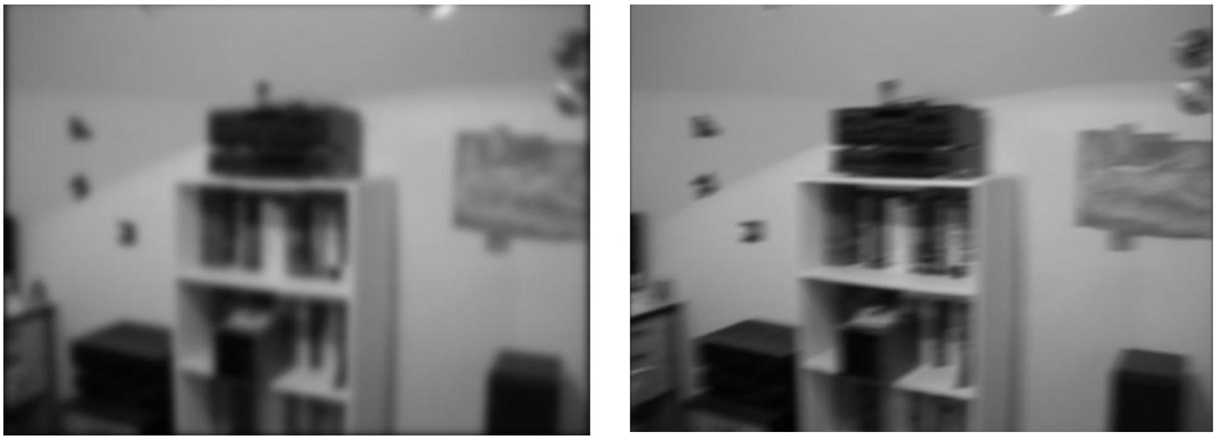
$$G_{2D}(\mathbf{x},\sigma_s)=\frac{1}{2\pi\sigma_s^{\,2}}e^{\frac{-\|\mathbf{x}\|^2}{2\sigma_s^{\,2}}},\quad G_{3D}(\mathbf{x},t,\sigma_s,\sigma_\tau)=\frac{1}{\sqrt{(2\pi)^3}\sigma_s^{\,2}\sigma_\tau}e^{-\frac{\|\mathbf{x}\|^2}{2\sigma_s^{\,2}}-\frac{t^2}{2\sigma_\tau^{\,2}}} \tag{4}$$

Στην περίπτωση της τρισδιάστατης εξομάλυνσης ολόκληρη η ακολουθία συνελίσσεται με την τρισδιάστατη μάσκα G_{3D} έτσι ώστε να συνυπολογίζεται πληροφορία από την τρίτη διάσταση, τον χρόνο. Η χρησιμότητα από κάτι τέτοιο είναι η μείωση του χρονικού aliasing, το οποίο προκαλείτε μεταξύ άλλων από τις γρήγορες κινήσεις της κάμερας (Σχήμα 2-6). Το χρονικό aliasing είναι άλλη μία παράμετρος που επιδρά αρνητικά στην ποιότητα της παραγώγισης, αν και όχι καθοριστικά. Σε κάθε περίπτωση θα πρέπει να σταθμίζουμε την βελτίωση που προκαλεί η χρήση του 3D φίλτρου στην εφαρμογή καθώς το υπολογιστικό κόστος από την συνέλιξη της ακολουθίας με την τρισδιάστατη μάσκα του φίλτρου είναι σημαντικό και δύσκολο να αγνοηθεί.

Τέλος σκοπός της εξομάλυνσης δεν είναι αποκλειστικά η μείωση του θορύβου αλλά γενικότερα η απομάκρυνση των υψηλών συχνοτήτων της εικόνας. Κάτι τέτοιο είναι απαραίτητο λόγω του ότι οι υπολογισμοί στις διαφορικές τοπικές μεθόδους βασίζονται στην χρήση του αναπτύγματος πρώτης επέκτασης *σειρών Taylor* από την προέρχεται η ευθεία περιορισμού (σχέση 2). Με την εξομάλυνση ουσιαστικά επιτυγχάνουμε την μείωση των δεύτερων (και μεγαλύτερων) παραγώγων που αντιπροσωπεύουν τις υψηλότερες συχνότητες στην εικόνα και οι οποίες δρουν αρνητικά στα αποτελέσματα των μεθόδων που βασίζονται στις παραγώγους πρώτης τάξεως. Προκειμένου να εξαλείψουμε τις υψηλές συχνότητες αποτελεσματικά το σ_s επιβάλλεται να λαμβάνει σχετικά μεγάλες τιμές (Σχήμα 2-7).



Σχήμα 2-6: *Το χρονικό aliasing προκαλεί αλλοιώσεις στις μετρήσεις μιας και μέσα σε μια εικόνα υπάρχουν πολλά εικονοστοιχεία με την ίδια ένταση.*



Σχήμα 2-7: *Gaussian βαθυπερατό χωρικό «φιλτράρισμα» με σ_s = 5 εικονοστοιχεία (αριστερά) και 3D χωροχρονικό «φιλτράρισμα» με σ_s = 1,5 εικονοστοιχεία και σ_τ = 1,5 καρέ (δεξιά).*

2.6 Μέθοδος Lucas-Kanade και προσθήκες

Σε αυτή την ενότητα θα δείξουμε μια μέθοδο επίλυσης του προβλήματος της οπτικής ροής που συνδυάζει τα καλύτερα στοιχεία διάφορων μεθόδων [18, 21] που βασίζονται στην τεχνική των *Lucas-Kanade (L-K)* [20]. Επιπλέον η μέθοδος που προτείνουμε φέρει αρκετές δικές μας βελτιωτικές καινοτομίες που εξυπηρετούν τις ιδιαίτερες ανάγκες του προβλήματος της ΕΚΚ που εξετάζουμε.

Η μέθοδος των L-K βασίζεται στην δημιουργία και επίλυση ενός συστήματος *n* γραμμικών εξισώσεων που αποτελείτε από τόσες σχέσεις της μορφής που έχει η εξίσωση (2), όσες είναι και τα γειτονικά εικονοστοιχεία που συμμετέχουν στους υπολογισμούς (Σχήμα 2-8). Τελικά το πρόβλημα εύρεσης της οπτικής ροής προκύπτει ως ελαχιστοποίηση της:

$$\sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x})\left[\nabla I(\mathbf{x},t)\cdot\mathbf{u} + I_t(\mathbf{x},t)\right]^2 \qquad \textbf{(5)}$$

όπου *Ω* είναι το σύνολο των γειτονικών εικονοστοιχείων εντός ενός σχετικά μικρού παραθύρου και *W(**x**)* υποδηλώνει ένα πίνακα με στάθμες ο οποίος αποδίδει βάρη στις εξισώσεις (σχέση 11). Ο *W(**x**)* μετριάζει προληπτικά την επίδραση των εξισώσεων που αντιστοιχούν σε εικονοστοιχεία που βρίσκονται κοντά στα άκρα του παραθύρου αφού εκεί είναι πιο πιθανό να παρουσιαστούν ανεξάρτητες κινήσεις. Τελικώς η ελαχιστοποίηση της (5) επιτυγχάνεται με επίλυση με την μέθοδο των *ελαχίστων τετραγώνων με στάθμες (weighted LS)* ως εξής:

$$A^T W^2 A \cdot \mathbf{u} = A^T W^2 b \qquad \textbf{(6)}$$

Όπου για n σημεία με θέσεις $\mathbf{x}_i \in \Omega$ και για ένα καρέ την χρονική στιγμή t οι πίνακες $A, W, \mathbf{b}$ είναι:

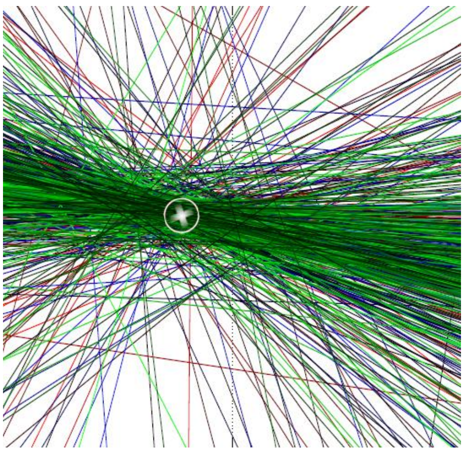
$$\begin{aligned} A &= \left[\nabla I(\mathbf{x_1}), \ldots, \nabla I(\mathbf{x_n})\right]^T, \\ W &= \mathrm{diag}\left[W(\mathbf{x_1}), \ldots, W(\mathbf{x_n})\right], \\ \mathbf{b} &= -\left[I_t(\mathbf{x_1}), \ldots, I_t(\mathbf{x_n})\right]^T. \end{aligned} \qquad \textbf{(7)}$$

Μέσω της (6) μπορούμε τελικά να βρούμε το διάνυσμα κίνησης ***u*** ως:

$$\mathbf{u} = \left[A^T W^2 A\right]^{-1} A^T W^2 \mathbf{b} \qquad \textbf{(8)}$$

Η (8) λύνεται εύκολα με διάφορες μεθόδους επίλυσης γραμμικών συστημάτων υπό την προϋπόθεση ότι ο πίνακας $A^T W^2 A$ αντιστρέφεται (δηλαδή δεν είναι ιδιόμορφος). Το $A^T W^2 A$ σε κλειστή μορφή δίνεται από τον 2x2 πίνακα:

$$A^T W^2 A = \begin{bmatrix} \sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x}) I_x^2(\mathbf{x}) & \sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x}) I_x(\mathbf{x}) I_y(\mathbf{x}) \\ \sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x}) I_x(\mathbf{x}) I_y(\mathbf{x}) & \sum_{\mathbf{x}\in\Omega} W^2(\mathbf{x}) I_y^2(\mathbf{x}) \end{bmatrix} \qquad \textbf{(9)}$$



Σχήμα 2-8: *Ευθείες περιορισμού για παράθυρο 10x10 από δυο διαδοχικά καρέ και η λύση ελαχίστων τετραγώνων.*

Ο πίνακας βαρών *W(**x**)* για ένα παράθυρο μεγέθους 5x5 και μάσκα $\frac{1}{16}\left(1,4,6,4,1\right)$ είναι:

$$W^2(\mathbf{x}) = \begin{bmatrix} 0.0039 & 0.0156 & 0.0234 & 0.0156 & 0.0039 \\ 0.0156 & 0.0625 & 0.0937 & 0.0625 & 0.0156 \\ 0.0234 & 0.0937 & 0.1406 & 0.0937 & 0.0234 \\ 0.0156 & 0.0625 & 0.0937 & 0.0625 & 0.0156 \\ 0.0039 & 0.0156 & 0.0234 & 0.0156 & 0.0039 \end{bmatrix} \quad \textbf{(10)}$$

Ενώ οι χωρικοί παράγωγοι υπολογίζονται από τις:

$$\begin{aligned} I_{\mathrm{x}}(\mathbf{x},t) &= \frac{-I(x_1-2,t) + 8I(x_1-1,t) - 8I(x_1+1,t) + I(x_1+2,t)}{12} \\ I_{\mathrm{y}}(\mathbf{x},t) &= \frac{-I(x_2-2,t) + 8I(x_2-1,t) - 8I(x_2+1,t) + I(x_2+2,t)}{12} \end{aligned} \quad \textbf{(11)}$$

Ενδεικτικές τιμές / *Παράμετροι*	**Τιμή**
Συντελεστές υψιπερατού	$\frac{1}{12}(-1,8,0,-8,1)$
Gaussian $\sigma_{s,t}$	(3D) $1.5_{\text{εικονοστοιχεία}}$, $1.5_{\text{καρέ}}$ **ή** (2D) $5_{\text{εικονοστοιχεία}}$
Μέγεθος παραθύρου	5x5 εικονοστοιχεία
Κατώφλι ιδιοτιμών $\lambda_{1,2}$	5.0
Βήμα	10 εικονοστοιχεία
Μάσκα βαρών *W(x)*	$\frac{1}{16}\left(1,4,6,4,1\right)$ **ή** σ = 1.05

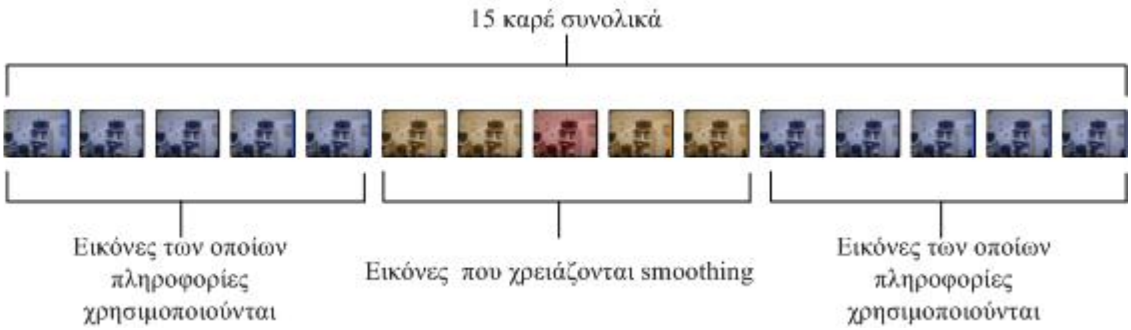
Πίνακας 2-2: *Ενδεικτικές τιμές παραμέτρων για την υλοποίηση με Gaussian3D presmoothing*

Πριν τον υπολογισμό των χωροχρονικών παραγώγων οι εικόνες επιβάλλεται να εξομαλυνθούν όπως περιγράψαμε στην ενότητα 2.5. Για την επιλογή της διαδικασίας εξομάλυνσης (G_{2D} ή G_{3D}) κριτήριο, εκτός από την βελτίωση της ακρίβειας των αποτελεσμάτων, θα πρέπει να είναι οι απαιτήσεις των φίλτρων σε υπολογιστικούς πόρους όπως και η επιβάρυνση που προκαλούν αυτά στον συνολικό χρόνο απόκρισης της μεθόδου.

π.χ. έστω ότι θέλουμε να εξομαλύνουμε την ακολουθία με την τρισδιάστατη μάσκα G_{3D}. Τότε για να υπολογίσουμε τις χρονικές παραγώγους σε ένα καρέ t_1 έχουμε:

$$I_t(\mathbf{x},t_1) \;=\; \frac{-I(\mathbf{x},t_1-2)+8I(\mathbf{x},t_1-1)-8I(\mathbf{x},t_1+1)+I(\mathbf{x},t_1+2)}{12} \qquad \textbf{(12)}$$

Όπως βλέπουμε στην (12) για τον υπολογισμό της $I_t(\boldsymbol{x},t_1)$ χρειαζόμαστε τα δυο προηγούμενα και δυο επόμενα από το τρέχον καρέ. Όμως για να λάβουμε την $I_t(\boldsymbol{x},t_1)$ θα πρέπει να εξομαλύνουμε πρώτα τις 5 εικόνες που εμπλέκονται στον υπολογισμό. Με την G_{3D} κάθε μια από αυτές τις εικόνες εξομαλύνεται στον χώρο και στον χρόνο και ως εκ τούτου απαιτεί πληροφορία από επιπλέον εικόνες εκτός των 5 εμπλεκομένων. Το επακόλουθο από αυτό είναι η διαδικασία να επιβαρύνεται υπολογιστικά καθώς και να καταλαμβάνονται επιπλέον πόροι μνήμης.



Σχήμα 2-9: *Αριθμός από εικόνες που πρέπει να επεξεργαστούν για να προκύψει η χρονική παραγωγός για ένα καρέ με χρονικό «φιλτράρισμα»* $\sigma_t = 1.5$ *καρέ.*

Το σχήμα 2-9 παρουσιάζει τον αριθμό των εικόνων που πρέπει να επεξεργαστούν για να παραχθεί το I_t μιας εικόνας (με $\sigma_t = 1.5$ καρέ). Είναι προφανές ότι η εφαρμογή της G_{3D} απαιτεί περισσότερους υπολογιστικούς πόρους σε σχέση με την αντίστοιχη δισδιάστατη G_{2D}, ενώ η βελτίωση που προκαλεί στα αποτελέσματα είναι δυσανάλογα μικρή και επιτυγχάνεται μόνο υπό συνθήκες (όπως αποδεικνύεται από τα πειράματα - ενότητα 2.9). Έτσι λόγω της φύσης της εφαρμογής που εξετάζουμε και τον σημαντικό ρόλο που διαδραματίζει για αυτήν ο χρόνος απόκρισης οδηγούμαστε στην χρήση του δισδιάστατου χωρικού φίλτρου για τους υπολογισμούς της οπτικής ροής σε αυτή την εργασία.

Συνεχίζοντας την ανάλυση για την διαδικασία επίλυσης ένα αξιόλογο ερώτημα που προκύπτει είναι το κατά πόσο η (6) έχει (μοναδική!) λύση (δηλαδή ο πίνακας $A^T W^2 A$ είναι αντιστρέψιμος). Αποδεικνύεται πως ο πίνακας $A^T W^2 A$ συχνά είναι ιδιόμορφος με αποτέλεσμα η (6) είτε να είναι αδύνατη, είτε να μην έχει μοναδική λύση. Το γεγονός αυτό παρατηρείτε συνήθως σε περιοχές των εικόνων που φέρουν κάποια συγκεκριμένα χαρακτηριστικά όπως στα σημεία με:

- Περιορισμένη υφή. (Οι χωρικοί παράγωγοι-συντελεστές του συστήματος τείνουν στο μηδέν).
- Οξείες ακμές. (Εμφανίζεται το φαινόμενο του διαφράγματος. Οι ευθείες περιορισμού παρατάσσονται σχεδόν σε παράλληλη διάταξη με αποτέλεσμα να μην σχηματίζουν μοναδικό κόμβο. Η ακρίβεια των παραγώγων είναι περιορισμένη λόγω των «ασυνεχειών» στην ένταση των εικονοστοιχείων).

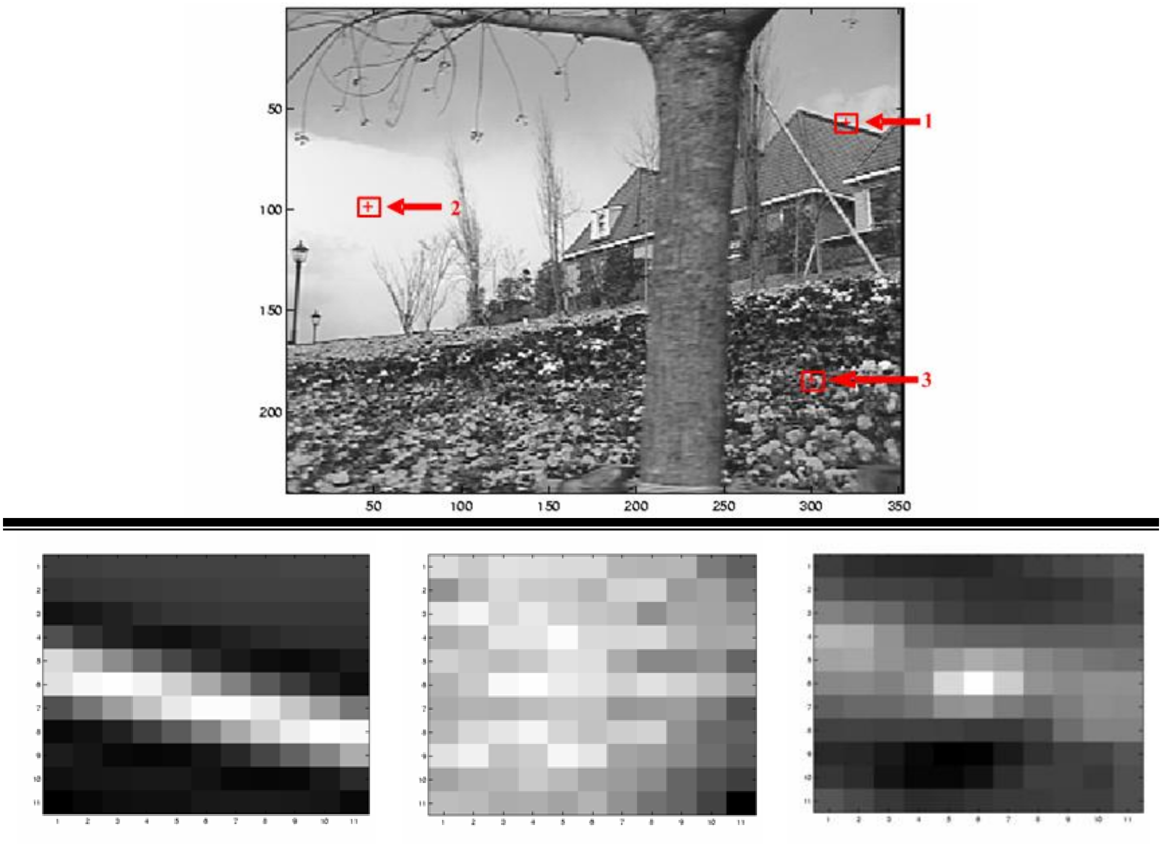
Οι παρατηρήσεις αυτές δημιουργούν μία ενδιαφέρουσα δυνατότητα χειρισμού του σφάλματος που ενδεχομένως προκύπτει μαζί με το αποτέλεσμα. Είναι φανερό πως αν επιτύχουμε την απομόνωση των υπολογισμών αποκλειστικά σε περιοχές που δεν φέρουν τα παραπάνω χαρακτηριστικά θα έχουμε αποφύγει ένα τμήμα των δεδομένων του πεδίου οπτικής ροής που είναι αρκετά πιθανό να περιέχει σφάλματα.

Για να εκμεταλλευτούμε τα παραπάνω αρκεί να καθορίσουμε κάποιο κριτήριο που να εκφράζει ποσοτικά τα ιδιαίτερα χαρακτηριστικά των περιοχών μιας εικόνας. Αυτή την πληροφορία αποδεικνύεται [22,23] ότι εμπεριέχουν οι ιδιοτιμές $\lambda_{1,2}$ του πίνακα $A^T W^2 A$ οι οποίες εξαρτώνται από το μέγεθος των χωρικών παραγώγων και τον προσανατολισμό τους. Αν υποθέσουμε $\lambda_1 \geq \lambda_2$ και με 🠉 συμβολίσουμε τις σχετικά μεγάλες τιμές ενώ με 🠋 τις σχετικά μικρές τιμές έχουμε:

- **λ_1 🠉, λ_2 🠋:** Περιπτώσεις έντονων ακμών με μεγάλες κλίσεις σε γεωμετρική διάταξη (π.χ. σε ευθεία). Το σύστημα εξισώσεων που παράγεται από πληροφορίες τέτοιων περιοχών πολλές φορές είναι ιδιόμορφο ή παράγει αναξιόπιστα – μειωμένης ακρίβειας αποτελέσματα.

- **$\lambda_1 \downarrow$, $\lambda_2 \downarrow$:** Δεν παρουσιάζεται υφή, μικρές κλίσεις (π.χ. σε μονοχρωματικό φόντο). Αυτές οι περιοχές δεν εμπεριέχουν χρήσιμες πληροφορίες για το πρόβλημα που εξετάζουμε.
- **$\lambda_1 \uparrow$, $\lambda_2 \uparrow$:** Περιοχές με πρότυπη χαρακτηριστική υφή (texture). Από αυτές τις περιοχές παράγονται κυρίως αξιόπιστα αποτελέσματα.

Αυτή η γνώση παρέχει ένα δείκτη αξιοπιστίας των μετρήσεων ανάλογα με τα χαρακτηριστικά της περιοχής που εξετάζουμε (σχήμα 2-10). Δεχόμενοι ότι για την εφαρμογή μας δεν είναι απαραίτητη πυκνή οπτική ροή αλλά προέχει η ακρίβεια των εκτιμήσεων, έχουμε το περιθώριο να είμαστε ιδιαίτερα επιλεκτικοί, απορρίπτοντας όλες εκείνες τις εκτιμήσεις που προέρχονται από περιοχές χωρίς χαρακτηριστικές υφές.



$\lambda_1 \uparrow$ $\lambda_2 \downarrow$ **1.** *Έντονη ακμή*

$\lambda_1 \downarrow$ $\lambda_2 \downarrow$ **2.** *Χωρίς χαρακτηριστική υφή*

$\lambda_1 \uparrow$ $\lambda_2 \uparrow$ **3.** *Δομή υφής*

Σχήμα 2-10: *Οι ιδιοτιμές του πίνακα $A^T W^2 A$ μας παρέχουν χρήσιμες πληροφορίες για τα ιδιαίτερα χαρακτηριστικά μιας περιοχής της εικόνας. (καρέ από την δημοφιλή ακολουθία garden).*

Για να αξιοποιήσουμε την πληροφορία που μας παρέχουν οι ιδιοτιμές ορίζουμε ένα κατώφλι *k*. Με $\lambda_1 \geq \lambda_2$, όταν ισχύει $\lambda_1 < k$ η περιοχή απορρίπτεται, ενώ όταν $\lambda_1 > k$ επιλύουμε το πρόβλημα με χρήση της (5). Στην υλοποίηση μας έχουμε χρησιμοποιήσει για το κατώφλι *k* την τιμή 5.

2.6.1 Επαναληπτική πολλαπλή ανάλυση

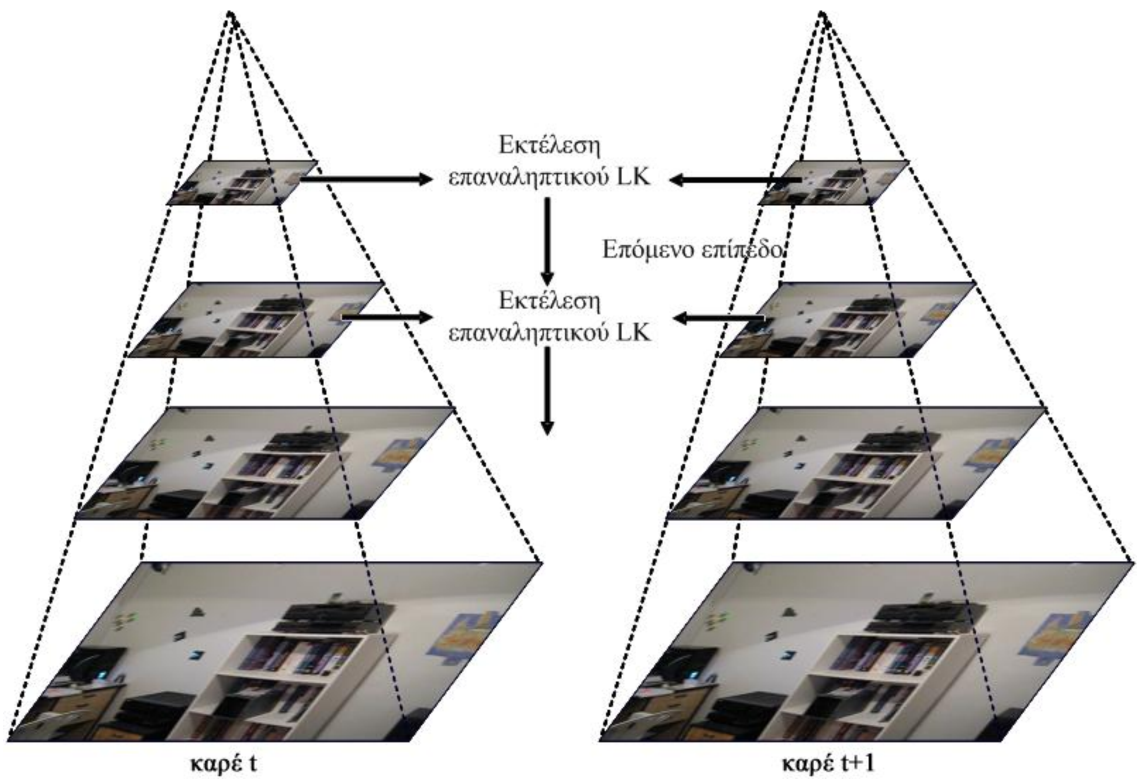
Η επαναληπτική πολλαπλή ανάλυση (iterative multi-resolution method) είναι μια μέθοδος που σκοπό έχει να εξαλείψει το πρόβλημα που προκύπτει με την **μεγάλη ταχύτητα κίνησης** όπως παρουσιάστηκε στην ενότητα 2.3.3. Ανακεφαλαιώνοντας θα λέγαμε πως τρεις είναι οι κύριοι παράγοντες για σφάλματα στις εκτιμήσεις με την μέθοδο L-K που περιγράψαμε:

- Η μη ικανοποίηση της υπόθεσης αμετάβλητης έντασης (… λόγω θορύβου).
- Η ανεξάρτητη κίνηση γειτονικών εικονοστοιχείων.
- Η μεγάλη ταχύτητα κίνησης, μεγαλύτερη από αυτή που δύναται να καταγραφεί με ένα παράθυρο συγκεκριμένου μεγέθους.

Οι δυο πρώτες καταστάσεις αποφεύγονται χάρη στην εξομάλυνση και το κριτήριο των ιδιοτιμών. Όμως το ζήτημα της ταχύτητας κίνησης δεν αντιμετωπίζεται με αυτές τις επεμβάσεις. Η αύξηση του μεγέθους του παραθύρου βελτιώνει την δυνατότητα μέτρησης «υψηλών» ταχυτήτων αλλά δεν αποτελεί την καλύτερη λύση καθώς αυτό έχει αρνητικές συνέπειες σε μια σειρά από άλλα ζητήματα που αναφέρθηκαν πριν. Μια λύση που μπορεί να αντεπεξέλθει στο πρόβλημα της ταχύτητας χωρίς να επηρεάζει αρνητικά τις υπόλοιπες παραμέτρους προέρχεται από μια τεχνική γνωστή ως ***coarse-to-fine***.

Με την *coarse-to-fine* τεχνική εφαρμόζουμε τον αλγόριθμο οπτικής ροής σε έναν αριθμό από L^m εικόνες διαφορετικού μεγέθους. Κάθε μία από αυτές είναι υποδιπλάσια σε μέγεθος σε σχέση με την προηγούμενη της. Συνεπώς η πρώτη εικόνα (η L^1) είναι η αυθεντική και η τελευταία (η L^m) αυτή με τις μικρότερες διαστάσεις. Με αυτή την αλληλουχία εικόνων διαφορετικών διαστάσεων ξεκινάμε τον υπολογισμό της οπτικής ροής από το πιο υψηλό επίπεδο L^i ($=L^m$) όπου βρίσκεται η μικρότερη εικόνα. Οι εκτιμήσεις που παράγονται εκεί χρησιμοποιούνται για αρχικοποίηση στο επόμενο επίπεδο L^{i-1} κ.ο.κ. Η ίδια διαδικασία επαναλαμβάνεται σε κάθε επίπεδο ως ότου επιστρέψουμε στην αυθεντική εικόνα με το τελικό αποτέλεσμα (Σχήμα 2-11). Έτσι ουσιαστικά με αυτό τον τρόπο σε κάθε επανάληψη «διορθώνεται» η αρχική χονδρική εκτίμηση που παράγεται από την μικρότερη εικόνα του αμέσως υψηλότερου επιπέδου.

Το κέρδος από αυτό είναι ότι μπορούν να μετρηθούν μεγαλύτερες μετατοπίσεις αφού το μέγεθος του παραθύρου παραμένει το ίδιο για κάθε επίπεδο. Με απλά λόγια ένα παράθυρο 5x5 θεωρείται σχετικά μεγάλο για μια εικόνα 160x120*p* ενώ σχετικά μικρό για μια εικόνα 640x480*p*. Υποθέτοντας λοιπόν ότι η αυθεντική εικόνα είναι μεγέθους L^0 = 640x480 τότε στο επίπεδο L^1 θα γίνει 320x240, στο L^2 = 160x120 κ.ο.κ. Αν στο L^0 η μέγιστη μετατόπιση που μπορεί να μετρηθεί είναι d_{max} τότε με L_m επίπεδα η νέα μέγιστη μετρήσιμη μετατόπιση γίνεται $d'_{\max} = \left(2^{m+1} - 1\right) d_{\max}$. Στο παράδειγμα τριών επιπέδων, με αρχική ανάλυση 640x480 και παράθυρο 5x5, έχουμε το αξιοσημείωτο κέρδος επί 15 σε σχέση με την αρχική δυνατότητα μέτρησης.



Σχήμα 2-11: *Αφού υπολογίσουμε την Gaussian πυραμίδα για διαδοχικά καρέ ξεκινάμε την εκτέλεση του αλγορίθμου από το επίπεδο με την μικρότερη ανάλυση. Τα αποτελέσματα κάθε επιπέδου τα χρησιμοποιούμε για αρχικοποίηση τιμών στο επόμενο επίπεδο.*

Γενικότερα κάθε υποδιπλασιασμός της εικόνας σημαίνει κέρδος κατά ένα συντελεστή δυο. Στην επόμενη ενότητα δείχνουμε πως μπορούμε να δημιουργήσουμε επίπεδα εικόνων διαφορετικής ανάλυσης με σταδιακή σμίκρυνση.

2.6.2 Υπολογισμός Gaussian πυραμίδας

Ένας τρόπος για να σμικρύνουμε τις εικόνες χωρίς την δημιουργία *aliasing* είναι με την χρήση κάποιας μεθόδου **παρεμβολής** (*interpolation*). Αν στο επίπεδο L^0 βρίσκεται η αυθεντική εικόνα και σε κάθε επόμενο επίπεδο οι διαστάσεις της εικόνας υποδιπλασιάζονται, τότε η εικόνα κάθε επιπέδου προκύπτει από την προηγούμενη ως:

$$\begin{aligned} I^L(x,y) = {} & \frac{1}{4} I^{L-1}(2x,2y) + \\ & \frac{1}{8}\left(I^{L-1}(2x-1,2y) + I^{L-1}(2x+1,2y) + I^{L-1}(2x,2y-1) + I^{L-1}(2x,2y+1) \right) + \\ & \frac{1}{16}\left(I^{L-1}(2x-1,2y-1) + I^{L-1}(2x+1,2y+1) + I^{L-1}(2x-1,2y+1) + I^{L-1}(2x+1,2y-1) \right) \end{aligned} \quad \textbf{(13)}$$

Ωστόσο ο παραπάνω διγραμμικός μετασχηματισμός (*bilinear interpolation*) ισοδυναμεί με τον κατά πολύ απλούστερο $I^L(x,y) = I^{L-1}(2x,2y)$ αν πριν την υποδειγματοληψία εφαρμόσουμε στην εικόνα βαθυπερατό φίλτρο με συντελεστές $\begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$ ανά διάσταση. Εντούτοις στην πράξη έχουμε την δυνατότητα να χρησιμοποιήσουμε μεγαλύτερη μάσκα με κέρδος την περαιτέρω μείωση του *aliasing* και κατ' επέκταση την βελτίωση της ποιότητας της παραγώγισης. Στην υλοποίηση μας εφαρμόσαμε την μάσκα με συντελεστές $\begin{bmatrix} \frac{1}{16} \frac{1}{4} \frac{3}{8} \frac{1}{4} \frac{1}{16} \end{bmatrix}$ για m = 3 επίπεδα.

Σχήμα 2-12: *Βήματα υπολογισμού της οπτικής ροής.*

2.7 Αλγόριθμος

Σε αυτή την ενότητα παρουσιάζουμε τον αλγόριθμο για την μέθοδο L-K με *επαναληπτική πολλαπλή ανάλυση*. Η διαδικασία ξεκινά αφού πρώτα εξομαλύνουμε τις αρχικές εικόνες και υπολογίσουμε τις Gaussian πυραμίδες για δυο διαδοχικά καρέ (Σχήμα 2-12). Τότε ξεκινάμε την επεξεργασία από την μικρότερη εικόνα του υψηλότερου επιπέδου και επαναλαμβάνουμε ως ότου φτάσουμε στην αρχική εικόνα της οποίας την οπτική ροή θέλουμε.

Κώδικας για την μέθοδο L-K επαναληπτικής πολλαπλής ανάλυσης

→ Υπολογισμός διανύσματος οπτικής ροής ***u*** στην θέση ***x***. (Για πλήρες πεδίο θα πρέπει να εφαρμοστεί για κάθε ***x***).

Είσοδος: Πυραμίδα εξομαλυμένης εικόνας *I*, Πυραμίδα εξομαλυμένης εικόνας *J*, θέση υπολογισμού ***x***, παράθυρο ***w***, επίπεδα πυραμίδας *m*.
Έξοδος: Διάνυσμα οπτικής ροής ***u***.

→ **Δημιουργία** πυραμίδας *m* επιπέδων για τις εικόνες *I* και *J*.

→ **αρχικοποίηση** (ενδιάμεσης τιμής) διανύσματος οπτικής ροής στο επίπεδο *m*: $\boldsymbol{g}^m = [0\ 0]^T$

→ **αρχικοποίηση** δείκτη επιπέδου πυραμίδας στην μικρότερη εικόνα $i = m$

→ **Μέχρις ότου** το *i* γίνει 0 κάνε

→ **Υπολόγισε** την αντίστοιχη θέση ***x*** στην εικόνα του επιπέδου *i*: $\boldsymbol{x}^i = [p_1, p_2] = \dfrac{\mathbf{x}}{2^L}$

→ **Υπολογισμός** των χωρικών παραγώγων $I_x(\boldsymbol{x}^i)$ και $I_y(\boldsymbol{x}^i)$ για την εικόνα I^i (εικόνα *I* στο επίπεδο *i*):

$$I_x(\mathbf{x}^i) = \frac{-I^i(x_1-2,x_2)+8I^i(x_1-1,x_2)-8I^i(x_1+1,x_2)+I^i(x_1+2,x_2)}{12}$$

$$I_y(\mathbf{x}^i) = \frac{-I^i(x_1,x_2-2)+8I^i(x_1,x_2-1)-8I^i(x_1,x_2+1)+I^i(x_1,x_2+2)}{12}$$

→ **Υπολογισμός** του πίνακα κλίσεων:

$$G = \sum_{x_1=p_1-w_1}^{p_1+w_1} \sum_{x_2=p_2-w_2}^{p_2+w_2} \begin{bmatrix} W^2(\mathbf{x}^i)I_x^2(\mathbf{x}^i) & W^2(\mathbf{x}^i)I_x(\mathbf{x}^i)I_y(\mathbf{x}^i) \\ W^2(\mathbf{x}^i)I_x(\mathbf{x}^i)I_y(\mathbf{x}^i) & W^2(\mathbf{x}^i)I_y^2(\mathbf{x}^i) \end{bmatrix}$$

→ **Αρχικοποίηση** της εκτίμησης του αλγορίθμου:
$\boldsymbol{v}^0 = [0\ 0]^T$

→ **Μέχρις ότου** η μεταβλητή $\left\|\boldsymbol{\eta}^k\right\| < 0.03$ ή $k <$ MAX_ITERATIONS

→ **Υπολογισμός** χρονικής παραγώγου
$I_t(\mathbf{x}^i) = I^i(\mathbf{x}^i) - J^i(x_1 + g_1^i + v_1^{k-1}, x_2 + g_2^i + v_2^{k-1})$

→ **Δημιουργία** διανύσματος χρονικών διαφορών

$$\mathbf{b}^k = \sum_{x_1=p_1-w_1}^{p_1+w_1} \sum_{x_2=p_2-w_2}^{p_2+w_2} \begin{bmatrix} W^2(\mathbf{x}^i) I_t(\mathbf{x}^i) I_x(\mathbf{x}^i) \\ W^2(\mathbf{x}^i) I_t(\mathbf{x}^i) I_y(\mathbf{x}^i) \end{bmatrix}$$

→ **Υπολογισμός** διανύσματος οπτικής ροής $\boldsymbol{\eta}^k = G^{-1}\mathbf{b}_k$

→ **Αποθήκευση** αποτελέσματος για την επόμενη επανάληψη: $\mathbf{v}^k = \mathbf{v}^{(k-1)} + \boldsymbol{\eta}^k$

→ **Τέλος επανάληψης** *k*

→ **Αποθήκευση** τελικού αποτελέσματος οπτικής ροής για το επίπεδο *i*: $\mathbf{d}^i = \mathbf{v}^k$

→ **Διαβάθμιση** του διανύσματος ***d*** για επεξεργασία στο επόμενο επίπεδο *i – 1*
$\mathbf{g}^{i-1} = 2(\mathbf{g}^i + \mathbf{d}^i)$

→ **Τέλος επανάληψης** για τα επίπεδα *m*

→ **Αποθήκευση** τελικού διανύσματος οπτικής ροής: $\mathbf{u} = \mathbf{g}^0 + \mathbf{d}^0$

Ο αλγόριθμος θα πρέπει να εφαρμοστεί για κάθε θέση ***x*** της εικόνας (ή με κάποιο βήμα) προκειμένου να λάβουμε ολοκληρωμένο το πεδίο οπτικής ροής.

Για κάποια σημεία ***x*** ενδέχεται ο αλγόριθμος να μην συγκλίνει ή το παράθυρο να μετακινείτε σε περιοχές εκτός των διαστάσεων της εικόνας. Τα σημεία αυτά θα πρέπει να απορρίπτονται κατά την διάρκεια της εκτέλεσης. Επίσης θα πρέπει να απορρίπτονται τα σημεία που δίνουν μικρές ιδιοτιμές σε οποιοδήποτε από τα L^m επίπεδα.

Σχετικά με τους υπολογισμούς θα πρέπει να γνωρίζουμε ότι ενώ οι συντεταγμένες ***x*** μιας εικόνας είναι θετικοί ακέραιοι στα επόμενα επίπεδα ενδέχεται το ***x*** διαιρούμενο με την δύναμη του δυο να προκύψει δεκαδικός. Διαχειριζόμαστε το θέμα αυτό κάνοντας υπολογισμούς «υπό-εικονοστοιχείων» με την διγραμμική παρεμβολή (14).

$$\begin{aligned} I^i(x_1, x_2) = (1-a_1)(1-a_2) I^i(x_1', x_2') + a_1(1-a_2) I^i(x_1'+1, x_2') + \\ (1-a_1) a_2 I^i(x_1', x_2'+1) + a_1 a_2 I^i(x_1'+1, x_2'+1) \end{aligned} \quad \textbf{(14)}$$

όπου $x_1^{'}$ και α_1 το ακέραιο και δεκαδικό τμήμα του x_1 αντίστοιχα, ενώ $x_2^{'}$ και α_2 το ακέραιο και δεκαδικό τμήμα του x_2 αντίστοιχα.

Σημειώνουμε ότι τα πειράματα με την προτεινόμενη μέθοδο έδειξαν ότι παράγονται χαμηλότερης αξιοπιστίας εκτιμήσεις όταν χρησιμοποιείτε το τρισδιάστατο φίλτρο για την εξομάλυνση σε σχέση με αυτές που παράγονται με χρήση του αντίστοιχου δισδιάστατου. Πιθανώς η αιτία για αυτό είναι ότι στην μέθοδο που παρουσιάζουμε οι χρονικές παράγωγοι δεν υπολογίζονται από την (12) αλλά σαν απλή διαφορά μεταξύ των εντάσεων των παραθύρων. Έτσι επιθυμούμε τα εικονοστοιχεία του παραθύρου στο πρώτο καρέ να έχουν όσο το δυνατόν την ίδια ένταση με αυτά του μετατοπισμένου παραθύρου στο επόμενο καρέ. Αντίθετα η εφαρμογή του τρισδιάστατου φίλτρου επιφέρει «ίχνη κίνησης» στις εικόνες (αποτυπώνοντας με αυτό τον τρόπο την εξομάλυνση κατά τον άξονα του χρόνου) κάτι που αλλοιώνει την αντιστοιχία μεταξύ των παραθύρων αποτρέποντας την «πλήρη» ταύτισή τους.

2.8 Ανίχνευση σφαλμάτων

Παρά τις τεχνικές που χρησιμοποιήσαμε προκειμένου να περιορίσουμε τα σφάλματα μοιραία ένα ποσοστό από τα διανύσματα που παράγεται θα περιέχει ακραίες τιμές θορύβου. Όπως θα δούμε στα επόμενα κεφάλαια απουσία θορύβου έξι μόνο διανύσματα αρκούν για την ΕΚΚ. Εντούτοις οι αλγόριθμοι οπτικής ροής παράγουν τυπικά εκατοντάδες διανύσματα ως έξοδο. Είναι προφανές ότι ο μεγάλος αριθμός διανυσμάτων που παράγονται σε αντίθεση με τον μικρό αριθμό που απαιτείται ως είσοδος στον ΕΚΚ δημιουργεί μεγάλες δυνατότητες περαιτέρω διαλογής των διανυσμάτων.

Ένα άλλο ζήτημα προς διερεύνηση είναι το κατά πόσο επηρεάζεται η ΕΚΚ από ακραία σφάλματα στο πεδίο οπτική ροής. Προς το παρόν ας δεχτούμε ότι καθένα από τα διανύσματα που εισάγονται για επεξεργασία συνεισφέρει στην λύση το ίδιο: π.χ. γνωρίζουμε ότι στον μέσο όρο ή στην λύση ελαχίστων τετραγώνων ενός συστήματος εξισώσεων έστω και ένα ακραία λανθασμένο δείγμα έχει την δυναμική να αλλοιώσει το αποτέλεσμα. Αργότερα θα δούμε ότι το ίδιο συμβαίνει και στην περίπτωση της επίλυσης για την εύρεση της κίνησης. Έτσι εγείρεται το ερώτημα αν μπορούμε να χειριστούμε τα σφάλματα εκ των υστέρων δηλαδή μετά τον υπολογισμό της οπτικής ροής: Η απάντηση είναι πως υπάρχει η δυνατότητα να το κατορθώσουμε με την χρήση **robust** τεχνικής στατιστικού ελέγχου των σφαλμάτων.

Αυτές οι τεχνικές δύνανται να χρησιμοποιηθούν σε μια μεγάλη γκάμα προβλημάτων και λειτουργούν αξιοποιώντας στατιστικά δεδομένα του εκάστοτε προβλήματος: π.χ. στο πρόβλημα που εξετάζουμε εδώ, τα διανύσματα των οποίων η κατεύθυνση ή το μέτρο αποκλίνει σε σχέση με την πλειονότητα είναι αρκετά πιθανό να είναι εσφαλμένα. Τότε μέσω μιας διαδικασίας στάθμισης η ισχύς των εν λόγω διανυσμάτων υποβαθμίζεται έτσι ώστε να μην επιδρούν στην λύση όσο τα υπόλοιπα. Αργότερα θα δούμε ότι και στην ΕΚΚ χρησιμοποιούμε αντίστοιχη robust τεχνική κατά την οποία αποδίδονται μικρότερα βάρη στα διανύσματα που αποκλίνουν από το μοντέλο κίνησης που υποδεικνύει η πλειονότητα των δειγμάτων.

Ανεξάρτητα από το μοντέλο κίνησης, μια απλή ιδέα που μπορεί εύκολα να εφαρμοστεί για να εξαλείψουμε ένα μέρος από τα εσφαλμένα διανύσματα είναι η εύρεση ενός «μέσου διανύσματος» και στην συνέχεια η απόρριψη εκείνων των δειγμάτων που αποκλίνουν περισσότερο από ένα όριο. Βέβαια η αποτελεσματικότητα της μεθόδου εξαρτάται από το κατά πόσο μπορεί να υπολογιστεί ένα μέσο διάνυσμα ανεπηρέαστο από τα θορυβώδη αυτά δείγματα. Το μέσο διάνυσμα δεν είναι άλλο από τον μέσο όρο των στοιχείων u_1, u_2 των διανυσμάτων ανά άξονα. Συνεπώς η εξεύρεση του μέσου διανύσματος μπορεί απλά να αναχθεί σε έναν υπολογισμό μέσων όρων ανά διάσταση.

Σημειώνουμε ότι ο μέσος όρος ενδεχόμενα θα μπορούσε να αφορά τον υπολογισμό του μέσου μέτρου των διανυσμάτων αλλά με αυτό τον τρόπο δεν θα δινόταν σημασία στην κατεύθυνση. Από την άλλη μεριά το μέσο διάνυσμα εμπεριέχει και το θέμα των αποκλίσεων στο μέτρο. Το ερώτημα που μένει να απαντηθεί είναι το πώς θα υπολογιστεί ένας ανθεκτικός στο θόρυβο μέσος όρος.

Ο ανθεκτικός στο θόρυβο μέσος όρος υπολογίζεται εφαρμόζοντας μια αναδρομική διαδικασία κατά την εκτέλεση της οποίας η τιμή του βελτιώνεται με τέτοιο τρόπο ώστε να αγνοεί τα εσφαλμένα δεδομένα. Αν ορίσουμε σαν υπόλοιπο r_i (*residual*) τη διαφορά μεταξύ της παρατήρησης – δείγματος u_i και του εκτιμώμενου μέσου όρου f_{mean} τότε η εύρεση του μέσου όρου ισοδυναμεί με την ελαχιστοποίηση της συνάρτησης $\sum_i f(r_i)$ όπου $f(x)$ συνάρτηση κόστους. Στον «κοινό» μέσο όρο η ελαχιστοποίηση ισοδυναμεί με την επίλυση με ελάχιστα τετράγωνα δηλαδή με $f(x) = x^2$.

Ουσιαστικά το υπόλοιπο r_i εκφράζει την απόκλιση ενός δείγματος u_i από την εκτίμηση της μέσης τιμής για την τρέχουσα επανάληψη. Μεγάλο υπόλοιπο υποδηλώνει ότι το συγκεκριμένο δείγμα αποκλίνει από την μέση τιμή που προκύπτει από την πλειονότητα των δειγμάτων. Με άλλα λόγια το υπόλοιπο είναι μια σημαντική πηγή ένδειξης σφαλμάτων στα δείγματα και όσο μεγαλύτερο είναι τόσο μικρότερη σημασία πρέπει να δίνεται στο εν λόγω δείγμα που το προκαλεί. Αν τώρα υποθέσουμε ότι ο βαθμός επίδρασης του κάθε δείγματος συναρτήσει του υπολοίπου του είναι w_i τότε η εύρεση του ανθεκτικού στο θόρυβο μέσου όρου ισοδυναμεί με την ελαχιστοποίηση της συνάρτησης ολικού κόστους $\sum_i w_i f(r_i)$. Το βάρος w_i λαμβάνει τιμές από μια συνάρτηση επίδρασης και συνήθως παίρνει τιμές στο διάστημα *(0, 1]*. Επειδή επιθυμούμε η πλειονότητα των δειγμάτων να έχει τον ίδιο βαθμό επίδρασης αλλά ένα μικρό ποσοστό αυτών που αποκλίνει εντόνως να έχει μειωμένο βαθμό επίδρασης ως συνάρτηση κόστους θα μπορούσε να χρησιμοποιηθεί η γνωστή ***Huber's weight*** η οποία ορίζεται κατά τμήματα. Αν ορίσουμε σαν κατώφλι την τιμή *c* τότε τα δείγματα των οποίων τα υπόλοιπα r_i βρίσκονται εντός του εύρους τιμών *(-c,c)* θα έχουν ισάξια συμμετοχή στον υπολογισμό του μέσου όρου. Αντίθετα μειωμένο βάρος επιβάλλεται στα δείγματα με r_i εκτός διαστήματος.

Μια διαδεδομένη τιμή για το κατώφλι είναι $c = c_o \cdot MAD(r_1 \ldots r_m)$, όπου c_o = 1.4826 και *MAD* ο *median* της απόλυτης απόκλισης. Το ζήτημα της σταθμισμένης ελαχιστοποίησης θα αναλυθεί περαιτέρω στην ενότητα 4.4.4. Ο κώδικας εύρεσης του σταθμισμένου μέσου όρου από *m* αριθμό δειγμάτων είναι:

- **<u>Κώδικας εύρεσης σταθμισμένου μέσου όρου</u>**

→ **Είσοδος:** u_i: δείγματα, *m*: αριθμός δειγμάτων, *ε*: ανοχή, c_0 : σταθερά.
Έξοδος: u_{mean}: μέσος όρος.

→ **Αρχικοποίησε** $u_{est} = \dfrac{(u_1 + \ldots + u_m)}{m}$

→ **Επανέλαβε** μέχρις $u_{est} - u_{est\_prev} < \varepsilon$

→ **Για κάθε** *i = 1 … m*

→ $r_i = u_i - u_{est}$

→ **τέλος επαναλήψεων**

→ $c = c_0\, MAD(r_1 \ldots r_m)$

→ **Για κάθε** *i = 1 … m*

→ **Υπολόγισε** στάθμες $\begin{cases} \text{Αν } |r_i| \le c, & w_i = 1 \\ \text{Αν } |r_i| > c, & w_i = \dfrac{c}{|r_i|} \end{cases}$

→ **τέλος επαναλήψεων**

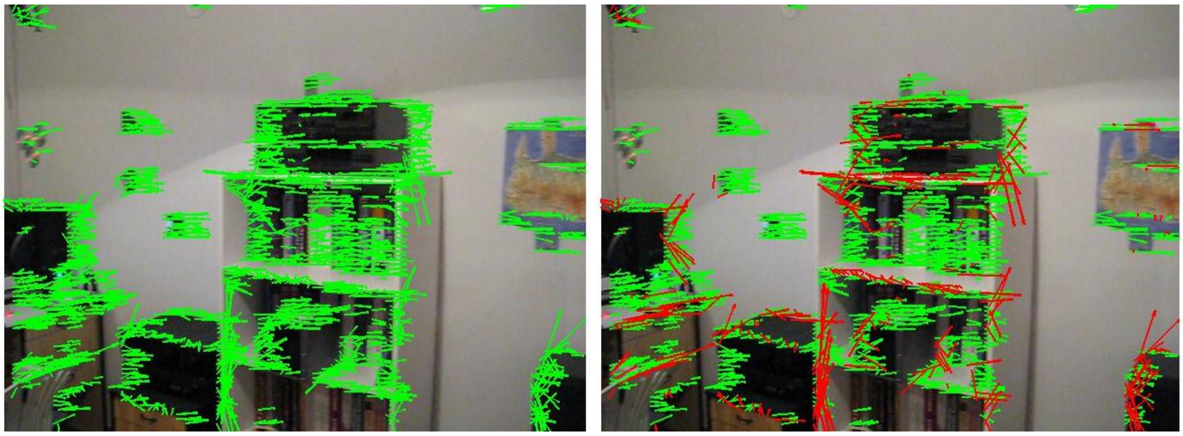
→ **Υπολόγισε** νέο μέσο όρο: $u_{est} = \dfrac{\sum_i w_i r_i}{\sum_i w_i}$

→ **τέλος επανάληψης** σύγκλισης

→ **επιστροφή** με $u_{mean} = u_{est}$

Με την μέθοδο αυτή ο νέος μέσος όρος u_{mean} που προκύπτει είναι λιγότερο επηρεασμένος από δείγματα που πιθανώς είναι εσφαλμένα. Εφόσον τα διανύσματα της οπτικής ροής είναι δύο διαστάσεων, για να υπολογίσουμε το «μέσο διάνυσμα» εκτελούμε την διαδικασία αυτή δυο φορές ώστε να βρούμε τους επιμέρους μέσους όρους ανά διάσταση *x-y*. Μετά τον υπολογισμό του μέσου διανύσματος θέτουμε το όριο μέγιστης απόκλισης των διανυσμάτων από αυτό. Για την εφαρμογή μας το όριο αυτό εμπειρικά επιλέχθηκε να είναι 3.2 εικονοστοιχεία (σχήμα 2-13).

Η κατεύθυνση και το μέτρο καθενός από τα διανύσματα που περιέχονται σε ένα πεδίο ροής «δεσμεύονται» με συγκεκριμένο τρόπο από το μοντέλο κίνησης της κάμερας. Η απλή στατιστική μέθοδος που αναπτύσσουμε εδώ δεν συνυπολογίζει την δέσμευση αυτή. Η διερεύνηση του ζητήματος αυτού δηλαδή της απόρριψης των διανυσμάτων που δεν εναρμονίζονται με το μοντέλο κίνησης γίνεται στην ενότητα 4.4.4. Οι συνέπειες από την εδώ υπεραπλούστευση είναι ότι πολλές φορές απορρίπτονται διανύσματα που δεν είναι λανθασμένα αλλά το μοντέλο κίνησης που τα ορίζει επιτάσσει να έχουν αρκετά διαφορετικές τιμές σε σχέση με την πλειονότητα. Ένα αντιπροσωπευτικό παράδειγμα είναι ότι συνήθως απορρίπτονται τα διανύσματα που οφείλονται σε κινήσεις αντικειμένων που βρίσκονται πολύ κοντά στην κάμερα. Ο λόγος είναι ότι το μέγεθος των διανυσμάτων είναι αντιστρόφως ανάλογο της απόστασης. Έτσι μικρή απόσταση Z δημιουργεί εξαιρετικά μεγάλα διανύσματα αφού $\lim_{Z \to 0^+} \frac{1}{Z} \to \infty$. Αντίστοιχα το ίδιο συμβαίνει και στις περιστροφές κατά τον οπτικό άξονα όπου τα διανύσματα που βρίσκονται μακριά από την εστία επέκτασης (ενότητα 3.5) έχουν δυσανάλογα μεγάλο μέγεθος σε σχέση με την πλειονότητα. Συγκεκριμένα το μέγεθος τους αυξάνει εκθετικά όσο απομακρυνόμαστε από το κέντρο της εικόνας. Κάποια από αυτά τα διανύσματα ενδέχεται να απορριφθούν αδίκως. Ωστόσο το τίμημα που επιφέρει η αγνόηση εν δυνάμει χρήσιμης πληροφορίας είναι δυσανάλογα μικρή σε σχέση με την αλλοίωση που επιφέρουν τα λανθασμένα διανύσματα στην ΕΚΚ.



Σχήμα 2-13: *Θορυβώδης οπτική ροή όπως παράγεται από τον L-K (αριστερά), μετά την ανίχνευση εσφαλμένων δεδομένων τα διανύσματα κόκκινου χρώματος απορρίπτονται (δεξιά).*

2.9 Πειραματισμοί

Η συμπεριφορά της L-K [20] και των παραλλαγών της [18,21,24,25] είναι ιδιαιτέρως γνωστή μιας και πρόκειται για μια διαδεδομένη μέθοδο που επανειλημμένα έχει δοκιμαστεί σε πληθώρα εφαρμογών. Κατά μια έννοια γνωρίζουμε την εξαιρετική της απόδοση αλλά και τις αδυναμίες της. Δικό μας κριτήριο είναι το κατά πόσο μπορεί η L-K σε συνδυασμό με τις προσθήκες που αναπτύξαμε να παράγει αξιόπιστες εκτιμήσεις οπτικής ροής, τέτοιες που να επιτρέπουν στην ΕΚΚ να εξάγει όσο το δυνατόν ορθότερα τρισδιάστατα χαρακτηριστικά. Επομένως οι δόκιμες που παρουσιάζουμε εδώ δεν είναι εκτενείς αλλά ενδεικτικές.

Σχήμα 2-14: *Πεδίο ροής οφειλόμενο σε αριστερή μεταφορική κίνηση της κάμερας από την μέθοδο Lucas Kanade (αριστερά) και επαναληπτική πολλαπλής ανάλυσης Lucas Kanade (δεξιά).*

Το σχήμα 2-14 παρουσιάζει ένα καρέ ανάλυσης 640x480*p* από μια ακολουθία όπου η κάμερα πραγματοποιεί (σχεδόν!) αριστερή πλάγιο-πλευρική μεταφορική κίνηση με ταχύτητα στο κέντρο 2.2 εικονοστοιχεία. Γενικά κινήσεις της κάμερας που δεν περιέχουν περιστροφές ή μετακίνηση κατά τον Z άξονα (zoom-in/out) παράγουν συνευθειακά διανύσματα (προς το παρόν ας αγνοήσουμε τα γεωμετρικά χαρακτηριστικά της σκηνής και το οπτικό μοντέλο). Αν δε τα αντικείμενα βρίσκονται στην ίδια περίπου απόσταση από την μηχανή λήψης αναμένουμε τα διανύσματα να έχουν και το ίδιο μέτρο. Εκμεταλλευόμενοι αυτά τα χαρακτηριστικά καταστρώσαμε μια απλή δόκιμη ώστε να διαπιστώσουμε χονδρικά την ακρίβεια των μεθόδων: Αφού τα διανύσματα αναμένεται να είναι συνευθειακά μετρήσαμε το σφάλμα απόκλισης ως την γωνία που σχηματίζει καθένα από τα διανύσματα του πεδίου σε σχέση με το πραγματικό διάνυσμα που θεωρούμε βάσει της κίνησης της κάμερας. Ωστόσο μετρώντας το σφάλμα ως γωνία μεταξύ διανυσμάτων αγνοούμε τις αποκλίσεις του μέτρου. Η αιτία είναι η ενδιαφέρουσα παρατήρηση πως τα διανύσματα με σωστή κατεύθυνση είχαν πάντα και το αναμενόμενο μέτρο. Αντίθετα όταν το μέτρο λάμβανε εσφαλμένες τιμές, η φορά της κατεύθυνσης ήταν σχεδόν τυχαία. Άρα η μέτρηση της γωνίας σφάλματος είναι επαρκής για να σχηματίσουμε πλήρη εικόνα.

Μέγεθος / *Στατιστικά*	L-K	L-K με πολλαπλή ανάλυση
Πυκνότητα οπτικής ροής (με βήμα = 10*p*)	**49%**	**41%**
Απόκλιση (Standard Deviation)	**11,26°**	**1,81°**
Χρόνος εκτέλεσης	**4,9 sec**	**1,2 sec**

Πίνακας 2-3: *Οπτική πυκνότητα και απόκλιση για τις δυο μεθόδους. Σαν είσοδος χρησιμοποιήθηκαν διαδοχικές εικόνες με μεγάλη μετατόπιση.*

Αφού επεξεργαστήκαμε τις εικόνες και με τις δυο μεθόδους (LK με και χωρίς επαναληπτική πολλαπλή ανάλυση) παρατηρήσαμε ότι η L-K παρήγαγε μεγάλο αριθμό εσφαλμένων διανυσμάτων (σχήμα 2-13). Η αιτία των αποκλίσεων της L-K σε σχέση με την L-K πολλαπλής ανάλυσης οφειλόταν στην μεγάλη ταχύτητα κίνησης. Όπως είχαμε τονίσει η L-K πολλαπλής ανάλυσης αντιμετωπίζει επιτυχημένα τις μεγάλες ταχύτητες κίνησης σε σχέση με την τυπική L-K. Σαν γενικό συμπέρασμα μπορούμε να πούμε ότι η L-K πολλαπλής ανάλυσης έδινε πάντα καλύτερα αποτελέσματα όταν οι εικόνες περιείχαν μεγάλες μετατοπίσεις ενώ αποτελέσματα πανομοιότυπης ακρίβειας για μικρότερες ταχύτητες.

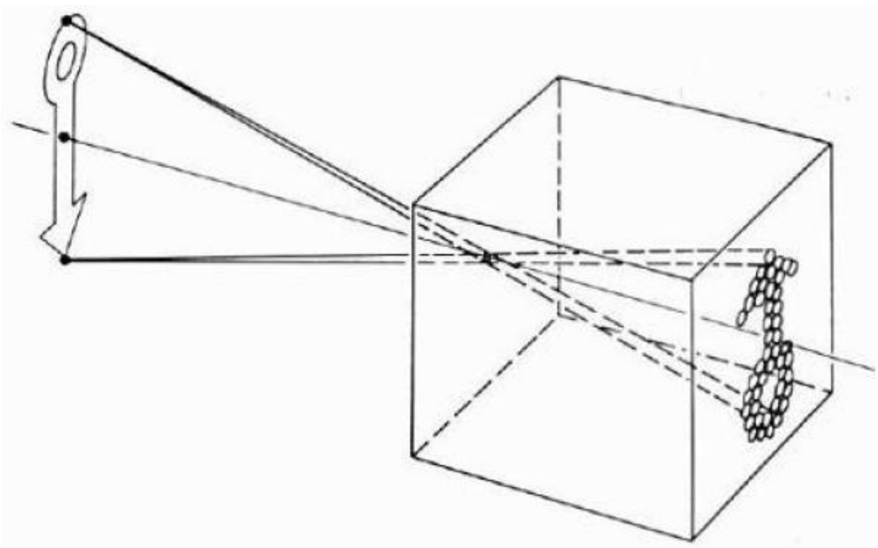
Ακόμα σημαντική είναι η διαφορά στο υπολογιστικό κόστος των δυο αλγορίθμων. Ο αυξημένος χρόνος εκτέλεσης του L-K οφείλεται στην επιβάρυνση που επιφέρει το τρισδιάστατο βαθυπερατό φίλτρο στην μέθοδο L-K σε σχέση με το αντίστοιχο δισδιάστατο φίλτρο της L-K πολλαπλής ανάλυσης. Ο πίνακας 2-3 παρουσιάζει τον ενδεικτικό χρόνο εκτέλεσης για δοκιμές με εικόνες μεγέθους 640x480 και τις ρυθμίσεις του πίνακα 2-2. Τελικώς ο L-K πολλαπλής ανάλυσης συγκεντρώνει αντίστοιχη ή καλύτερη ακρίβεια στα αποτελέσματα του, καθώς και κάποια επιπλέον δελεαστικά πλεονεκτήματα. Επομένως για τις ανάγκες της εφαρμογής επιλέξαμε αυτό τον αλγόριθμο για την εκτίμηση της οπτικής ροής.

Κεφάλαιο 3

Οπτικό μοντέλο και εξισώσεις κίνησης

3.1 Εισαγωγή

Σε αυτή την ενότητα θα δούμε τον τρόπο μετάβασης από την δισδιάστατη πληροφορία της οπτικής ροής στην τρισδιάστατη αναπαράσταση της σκηνής και το αντίθετο. Θα δούμε τις βασικές αρχές ενός ιδιαίτερα δημοφιλούς οπτικού μοντέλου καθώς και τις εξισώσεις που το διέπουν. Αυτή η διεπαφή ανάμεσα στον πραγματικό τρισδιάστατο κόσμο και την δισδιάστατη απεικόνιση του σε μια εικόνα είναι κρίσιμης σημασίας αφού ουσιαστικά είναι η βάση πάνω στην οποία θα στηριχτούμε για την ανάλυση του προβλήματος ΕΚΚ.



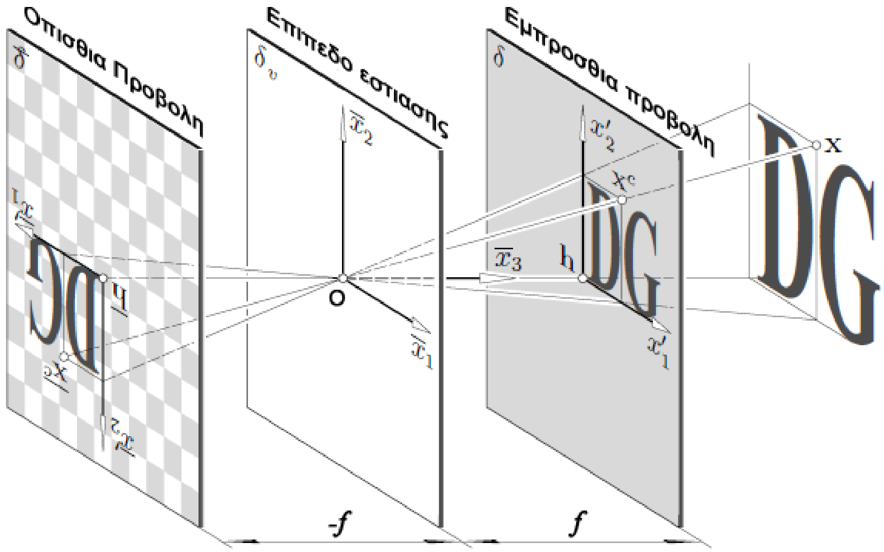
Σχήμα 3-1: *Το οπτικό μοντέλο σημειακής οπής μπορεί να αναπαρασταθεί σαν ένα κουτί που στο εμπρόσθιο μέρος έχει μια μικρή οπή και στο πίσω βρίσκεται το οπτικό πλάνο στο οποίο ανακλάται το αντικείμενο ανεστραμμένο.*

Όταν η κάμερα λήψης κινείται σε ένα χώρο ο τρόπος με τον οποίο απεικονίζονται τα αντικείμενα στο **οπτικό πλάνο** διαμορφώνεται από το ισχύον **οπτικό μοντέλο**. Ο πιο διαδεδομένος τρόπος αυτής της «μετατροπής» ονομάζεται **προοπτική προβολή**. Σύμφωνα με την προοπτική προβολή ένα σημείο του τρισδιάστατου κόσμου προβάλλεται στην εικόνα αφού «περάσει» μέσα από τον φακό της κάμερας κατά μήκος μια νοητής ευθείας γραμμής. Όσες συντεταγμένες του πραγματικού κόσμου βρίσκονται πάνω σε αυτή την ευθεία γραμμή προβάλλονται στο ίδιο σημείο στην εικόνα. Σε αυτές τις σχέσεις μετατροπής συνταγμένων θα βασιστούμε αργότερα για την εξαγωγή και ερμηνεία των τρισδιάστατων χαρακτηριστικών.

3.2 Οπτικό μοντέλο

Το οπτικό μοντέλο είναι η μαθηματική περιγραφή των στοιχείων που συνεργούν για την απεικόνιση του πραγματικού κόσμου σε μία εικόνα. Αυτή η «σύζευξη» μεταξύ των τριών διαστάσεων του πραγματικού κόσμου και των δύο διαστάσεων της εικόνας συντελείτε από το σύστημα λήψης μιας κάμερας σύμφωνα με τα χαρακτηριστικά και τις ιδιότητες του οπτικού μοντέλου που το διέπει. Το πλέον απλό αλλά ενδιαφέρον μοντέλο απεικόνισης του περιβάλλοντος είναι αυτό που ισχύει στην **κάμερα σημειακής οπής** (Pinhole camera model).

Η *κάμερα σημειακής οπής* αναπαριστάτε ως ένα κλειστό κυβικό κουτί με ένα στενό άνοιγμα στην μπροστινή του πλευρά (Σχήμα 3-1). Σε αυτό το απλό σύστημα η φωτεινή ανάκλαση κάθε σημείου $P\ (X,\ Y,\ Z)^T$ του πραγματικού κόσμου, διερχόμενη μέσα από το στενό άνοιγμα $O\ (0,\ 0,\ 0)$, προβάλλεται στο οπτικό πλάνο της συσκευής στο πίσω μέρος του κουτιού στην θέση $p\ (x,\ y,\ z)^T$. Αυτή η σχέση απεικόνισης ανάμεσα στα P και p καθορίζεται από το οπτικό μοντέλο, τις φυσικές διαστάσεις του συστήματος καθώς και κάποιες οπτικές παραμέτρους, όπως την **εστιακή απόσταση**.



Σχήμα 3-2: *Η αναπαράσταση εμπρόσθιας προβολής προτιμάται προκείμενου να απαλειφθεί το αρνητικό πρόσημο από ένα φυσικό μέγεθος όπως η εστιακή απόσταση.*

Η εστιακή απόσταση είναι το διάστημα ανάμεσα στην οπή *O* και το κέντρο του επιπέδου προβολής της εικόνα. Η νοητή ευθεία που ενώνει τα δύο αυτά σημεία αποτελεί τον **οπτικό άξονα** της κάμερας. Αν υποθέσουμε ότι η εστιακή απόσταση είναι f και ότι ο οπτικός άξονας ταυτίζεται με τον άξονα του *Z* του πραγματικού κόσμου, οι εξισώσεις που μετασχηματίζουν τις συντεταγμένες του *P* σε *p* είναι οι:

$$x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z} \tag{15}$$

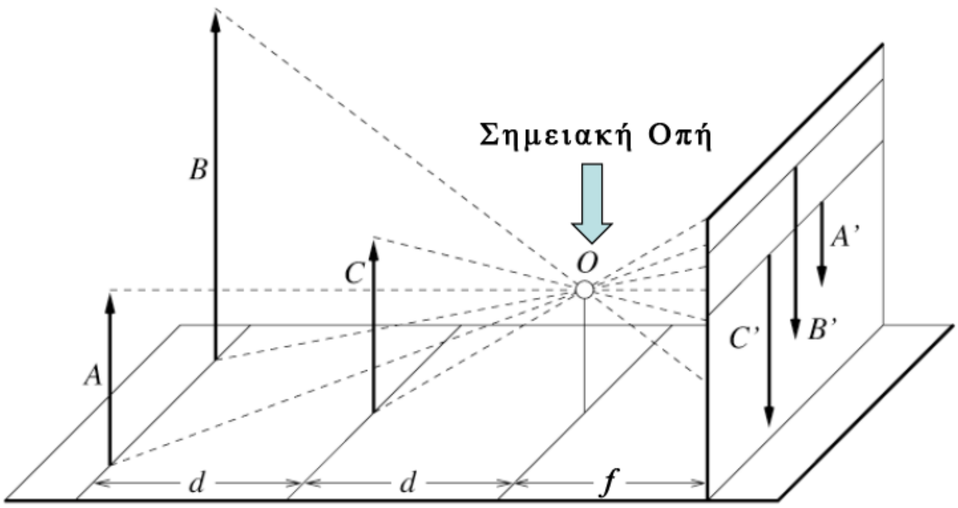
Αυτές οι εξισώσεις ορίζουν μια διαδικασία σχηματισμού των εικόνων που είναι γνωστή ως **εμπρόσθια προοπτική προβολή** (σχήμα 3-2). Το *Z* (depth) σημαίνει την απόσταση του *P* από την κάμερα. Ως λειτουργικότητα το *Z* είναι ο παράγοντας που καθορίζει, αυτό που γνωρίζουμε από την καθημερινή μας εμπειρία, ότι όσο πιο μακριά βρίσκεται ένα αντικείμενο τόσο μικρότερη θα είναι η εικόνα του.

3.3 Σχέση δύο (2D) και τριών (3D) διαστάσεων

Οι εξισώσεις μετασχηματισμού (15) μπορούν να περιγράψουν με απόλυτο τρόπο τις τρισδιάστατες συντεταγμένες σε δισδιάστατες συντεταγμένες, ωστόσο η αντίθετη διαδικασία μπορεί να γίνει μόνο σχετικά αφού οι συντεταγμένες του πραγματικού κόσμου ενσωματώνονται σε ένα λόγο απόστασης - θέσης. Αυτό αποτελεί μία βασική πηγή δυσχερειών για κάθε μέθοδο εξόρυξης τρισδιάστατων πληροφοριών από εικόνες.

Εξετάζοντας την (15) παρατηρούμε ότι ένα σημείο στην εικόνα δύναται να αντιστοιχεί σε διάφορα σημεία του τρισδιάστατου κόσμου. Για την ακρίβεια υποψήφια είναι όλα τα σημεία του κόσμου που βρίσκονται κατά μήκος της νοητής ευθείας «ακτίνας» που σχηματίζεται από την ένωση του εν λόγω σημείου της εικόνας και της σημειακής οπής *O* (σχήμα 3-3). Με αυτό το δεδομένο, το ερώτημα είναι αν υπάρχει η δυνατότητα να ανακτήσουμε μέρος από την τρισδιάστατη πληροφορία που χάθηκε.

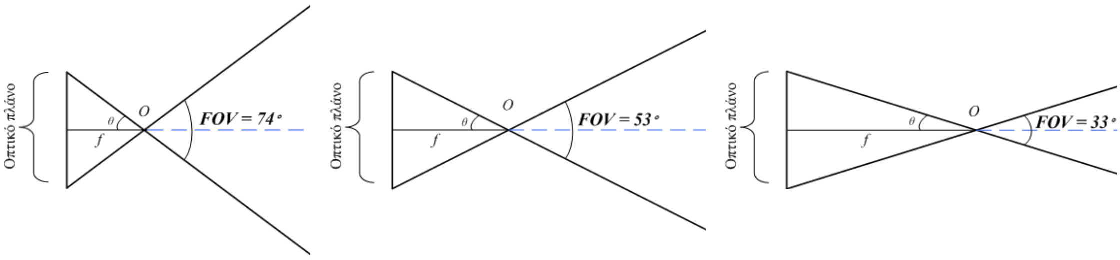
Ένας τρόπος για να επιτευχθεί αυτό είναι με την χρήση δύο ή περισσότερων καμερών διαμοιρασμένων στον χώρο έτσι ώστε να δημιουργείτε *ανομοιότητα*. Αυτό απαιτεί επιπλέον υλικό (κάμερες) ενώ δεν μπορεί να εφαρμοστεί σε απλές εικόνες που προϋπάρχουν. Ένας άλλο τρόπος χειρισμού του θέματος είναι η αναγωγή στον χρόνο δηλαδή ο συνδυασμός πληροφοριών από μια ακολουθία εικόνων που έχουν ληφθεί με κίνηση της κάμερας.



Σχήμα 3-3: *Η πληροφορία του βάθους Z χάνεται μετά την «αποτύπωση» των αντικειμένων στο οπτικό πλάνο. Επειδή το μέγεθος των αντικειμένων είναι σχετικό ως προς την απόσταση, χωρίς την πληροφορία του βάθους είναι αδύνατο να καθοριστεί το απόλυτο μέγεθος τους.*

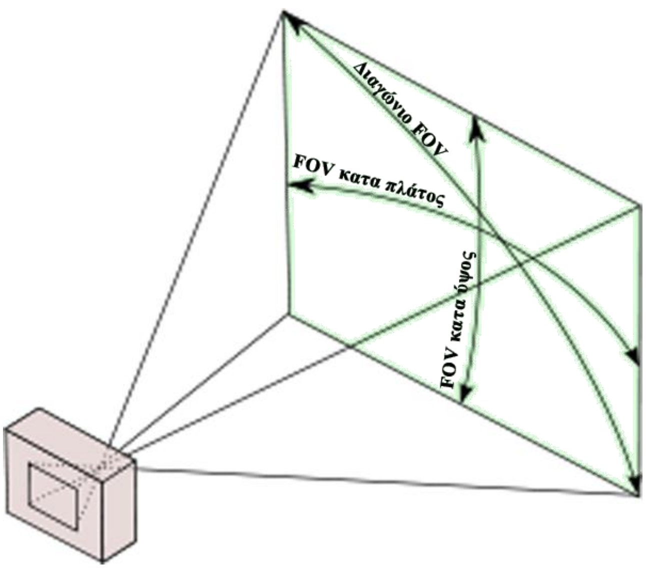
3.4 Σχέση εστιακής απόστασης, οπτικού πεδίου και οπτικού πλάνου

Μία παράμετρος που εμπλέκεται στον μετασχηματισμό (15) είναι η εστιακή απόσταση *f*. Η εστιακή απόσταση ουσιαστικά καθορίζει το μέγεθος του τμήματος του κόσμου που απεικονίζεται στο οπτικό πλάνο (Σχήμα 3-4).



Σχήμα 3-4: *Αυξάνοντας την εστιακή απόσταση διατηρώντας σταθερό το μέγεθος του πλάνου, μειώνεται το εύρος του οπτικού πεδίου.*

Αν υποθέσουμε ότι το μέγεθος του οπτικού πλάνου στην μια διάσταση είναι *d* και η εστιακή απόσταση *f* τότε το εύρος του τμήματος του κόσμου που θα φαίνεται στο πλάνο καθορίζεται από την παράμετρο του **οπτικού πεδίου** (field of view, *FOV*), όπου *FOV* είναι η γωνία που σχηματίζεται από τις ευθείες που ενώνουν την σημειακή οπή *O* με το πλάτος ή το ύψος ή την κύρια διαγώνιο του οπτικού πλάνου (Σχήμα 3-5).



Σχήμα 3-5: *Το οπτικό πεδίο ενδέχεται να μετρηθεί κατά πλάτος / ύψος ή κατά την κύρια διαγώνιο.*

Το μέγεθος του οπτικού πλάνου και της εστιακής απόστασης είναι σαφώς καθορισμένα μεγέθη για κάθε μηχανή κινηματογράφησης και συνηθίζεται να μετρώνται στις ίδιες μονάδες. Με γνωστά τα *d* και *f* (ή και χωρίς να γνωρίζουμε τις απόλυτες τιμές τους αλλά την αναλογία τους), μπορούμε να υπολογίσουμε το οπτικό πεδίο με χρήση των βασικών τριγωνομετρικών εξισώσεων που ισχύουν για τα όμοια τρίγωνα:

$$FOV = 2\tan^{-1}\left(\frac{d}{2f}\right) \qquad \textbf{(16)}$$

Ένα χαρακτηριστικό της εστιακής απόστασης *f* είναι η συχνή εμφάνισή της στις σχέσεις διαφόρων προβλημάτων κυρίως από τους τομείς της μηχανικής όρασης και των γραφικών. Όπως θα διαπιστώσουμε στις επόμενες ενότητες η εστιακή απόσταση υπεισέρχεται και στην θεμελιώδη για το πρόβλημα της ΕΚΚ σχέση που συνδέει την τρισδιάστατη κίνηση με το οπτικό μοντέλο.

Η εμπλοκή οπτικών παραμέτρων σε προβλήματα που δεν σχετίζονται με την κινηματογράφηση των εικόνων αλλά την επεξεργασία τους δυσχεραίνει τους υπολογισμούς χωρίς να λαμβάνεται κάποιο άλλο όφελος. Μία συνηθισμένη πρακτική που αντιμετωπίζει την δυσκολία αυτή είναι το «κλείδωμα» των οπτικών παραμέτρων σε μια σταθερή τιμή που εξυπηρετεί την ευκολία των υπολογισμών. Όπως θα δούμε αργότερα για το πρόβλημα της ΕΚΚ αυτή η τιμή είναι $f = 1$. Έτσι μπορούμε να χειριστούμε το πρόβλημα εκφράζοντας την εστιακή απόσταση και το οπτικό πεδίο αποκλειστικά μέσω του μεγέθους του οπτικού πλάνου *d* ως:

$$d = 2f\tan\left(\frac{FOV}{2}\right) \qquad \textbf{(17)}$$

Για να μπορεί να λαμβάνει συνεχώς η εστιακή απόσταση την τιμή της μονάδας θα πρέπει κάθε άλλη παράμετρος που συμμετέχει στους υπολογισμούς να διατιμηθεί βάσει της νέας *f*. Αυτό μπορεί να πραγματοποιηθεί μετατρέποντας όλα τα μεγέθη του προβλήματος σε μια νέα μονάδα μέτρησης που να βασίζεται στο *f*, και θα μπορούσε να ονομαστεί *«μονάδα εστιακής απόστασης»* (*focal length unit*) ή *FLU*.

Παράμετροι \ *Ενδεικτικές τιμές*	**Τιμή**
Αναλογίες πλάνου	4:3 $(640\text{x}480)_{\text{pixels}}$
FOV (κατά πλάτος)	50°
f	1
Ανάλυση εικόνων (σε FLU)	0.931x0.697

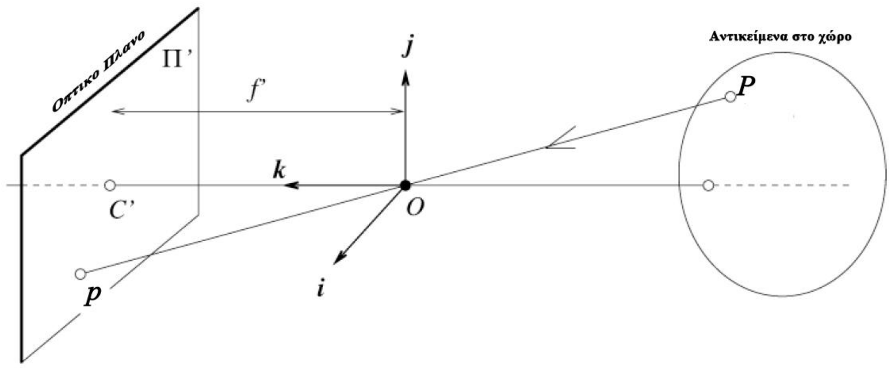
Πίνακας 3-1: *Τυπικές τιμές παραμέτρων που χρησιμοποιήθηκαν στις προσομοιώσεις.*

Όλα τα μεγέθη που συμμετέχουν στους υπολογισμούς (όπως τα διανύσματα οπτικής ροής, οι διαστάσεις των εικόνων, η μεταφορά της κάμερας *T*, το βάθος *Z* κ.α.) θα πρέπει να μετατραπούν σε *«μονάδες εστιακής απόστασης»* προκειμένου να ακολουθούν την νέα σύμβαση που εισάγαμε. Προφανές πλεονέκτημα από την χρήση της νέας μονάδας:

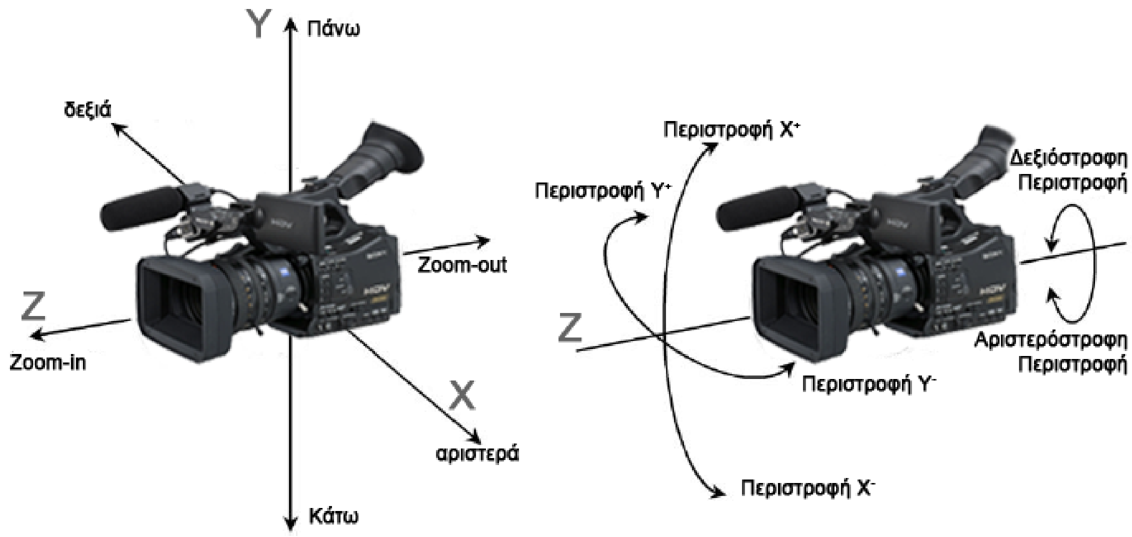
- Απλοποίηση των εξισώσεων κίνησης αφού η εστιακή απόσταση δεν λαμβάνεται υπόψη (είναι πάντα μονάδα).
- Αποδέσμευση από την ανάλυση (resolution) των εικόνων και τις φυσικές διαστάσεις του CCD της κάμερας.
- Μόνη οπτική παράμετρος που εμπλέκεται είναι το οπτικό πεδίο *FOV*, το οποίο μετριέται απλά σε μοίρες και είναι ενιαίο χαρακτηριστικό για όλους τους τύπους καμερών.

3.5 Σχέση οπτικού μοντέλου και τρισδιάστατης κίνησης

Συμπεριλαμβάνοντας και την έννοια της θέσης της κάμερας στον χώρο θα λέγαμε ότι η (15) μετασχηματίζει το σημείο *P* του πραγματικού κόσμου στην θέση *p* της εικόνας όταν ο οπτικός άξονας της κάμερας «κοιτάει» προς την κατεύθυνση του άξονα *Z* (σχήμα 3-6). Το ερώτημα είναι τι θα συμβεί αν η κάμερα μετακινηθεί σε νέα θέση στον χώρο και σε ποιο σημείο της εικόνας θα απεικονίζεται πλέον το *P*. Για να απαντήσουμε σε αυτό θα πρέπει πρώτα να ορίσουμε τι σημαίνει κίνηση της κάμερας στον χώρο.



Σχήμα 3-6: *Κάθε σημείο στο χώρο αντιστοιχίζεται σε συγκεκριμένη θέση στο οπτικό πλάνο.*



Σχήμα 3-7: *Πιθανές κινήσεις της κάμερας για 6 βαθμούς ελευθερίας (DOF6).*

Ως *κίνηση της κάμερας* ορίζουμε το σύνολο των παραμέτρων οι οποίες προσδιορίζουν πλήρως την *θέση* της κάμερας στο χώρο και την *κατεύθυνση* του οπτικού της άξονα σε κάθε χρονική στιγμή. Το πλήθος των παραμέτρων αυτών καθορίζει την πολυπλοκότητα του μοντέλου και κατ' επέκταση την πολυπλοκότητα της κίνησης της κάμερας. Το μοντέλο κίνησης που υιοθετούμε για τις ανάγκες της ΕΚΚ περιλαμβάνει κίνηση που περιγράφεται με 6 παραμέτρους, όσες ακριβώς χρειάζονται για να ορίσουμε μια πλήρως αδέσμευτη κίνηση προς κάθε *θέση* και κάθε *κατεύθυνση* στο χώρο (σχήμα 3-7). Ομαδοποιώντας τις παραμέτρους ανά τρεις τελικώς ορίζουμε το διάνυσμα $T = (t_1,t_2,t_3)^T$ που υποδηλώνει την θέση της κάμερας (σε σχέση με ένα σημείο $C = (0,0,0)^T$) και το διάνυσμα $R = (R_1,R_2,R_3)^T$ που υποδηλώνει την γωνία που σχηματίζει ο οπτικός άξονας με το σύστημα αξόνων *XYZ*. Ως $R = (0,0,0)^T$ ορίζεται η κατεύθυνση όπου ο οπτικός άξονας συμπίπτει με το *Z*. Για αυτονόητους λόγους από και εδώ και στο εξής θα αναφερόμαστε στο *T* ως **μεταφορά** και στο *R* ως **περιστροφή**.

Γίνεται φανερό πως η (15) σε αυτή την μορφή δεν μπορεί να καλύψει την νέα αυτή σημαντική επέκταση που περιλαμβάνει και κίνηση. Αν εκφράσουμε τα *p* και *P* σαν $(x,y,1)^T$ και $(X,Y,Z,1)^T$, απεικόνιση γνωστή και ως ομοιογενείς συντεταγμένες (*homogeneous coordinates*), και ορίσουμε ως πίνακα μετασχηματισμού τον *C*, λαμβάνουμε την ισοδύναμη της (15):

$$Zp = CP \Leftrightarrow z\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{18}$$

Τότε ο μετασχηματισμός για μια κάμερα που έχει την δυνατότητα να μεταφέρεται αλλά όχι να περιστρέφεται μπορεί εύκολα να προκύψει αν αντικαταστήσουμε την θέση του αντικειμένου *P* με τη σχετική του θέση ως προς την κάμερα *P'*:

$$P' = P + T = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ 0 \end{bmatrix} \tag{19}$$

Αν τώρα υποθέσουμε ότι η κάμερα εκτελεί ταυτόχρονα και περιστροφικές κινήσεις τότε το *P'* πολλαπλασιάζεται με τον *πίνακα περιστροφής*:

$$R_{matrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{20}$$

Τα στοιχεία του πίνακα περιστροφής $r_{1\ldots3,1\ldots3}$ προκύπτουν μέσω σχέσεων όπου λαμβάνονται υπόψη το διάνυσμα *R*, οι άξονες περιστροφής και η σειρά επενέργειας των περιστροφών. Σημειώνουμε ότι εδώ έχουμε υιοθετήσει το πρότυπο *περιστροφής του τηλεσκοπίου*. Τελικά για απλότητα και σαφήνεια παρουσιάζουμε σε συμπτυγμένη μορφή τον πίνακα μετακίνησης - περιστροφής *RT*. Επομένως ο μετασχηματισμός αντικειμένων *P* σε συντεταγμένες της εικόνας *p* για κινούμενη κάμερα υπολογίζεται ως:

$$Zp = C \cdot RT \cdot P \Leftrightarrow z\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{21}$$

3.6 Σχέση οπτικής ροής και τρισδιάστατης κίνησης

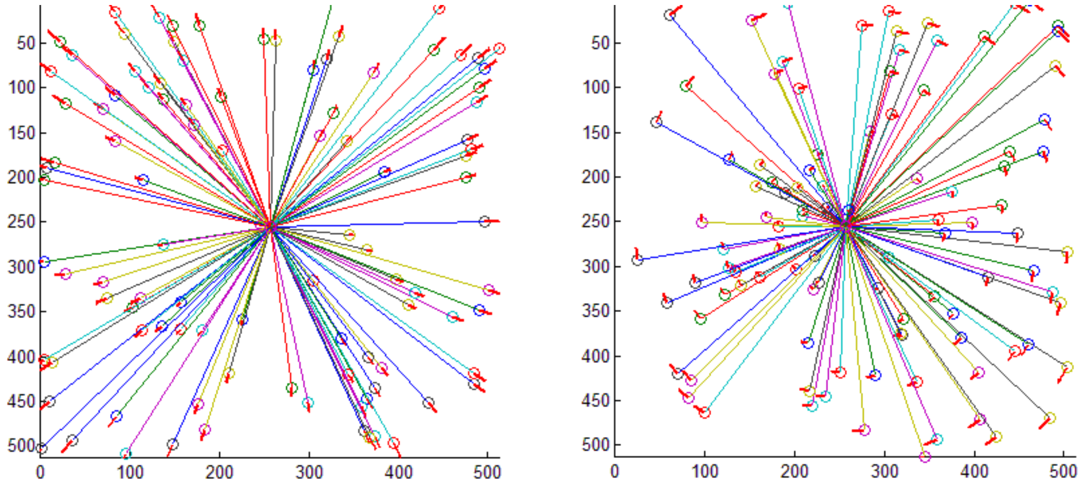
Έστω κάμερα με εστιακή απόσταση *f* = 1 που κινείται. Αν υποθέσουμε ότι την χρονική στιγμή *t* το σημείο *P* του πραγματικού κόσμου μετασχηματίζεται στο p_t και την αμέσως επόμενη χρονική στιγμή *t+1* το ίδιο σημείο *P* μετασχηματίζεται στο p_{t+1} τότε ουσιαστικά η ένωση των p_t και p_{t+1} αντιστοιχεί σε ένα διάνυσμα οπτικής ροής. Με χρήση τριγωνομετρίας λαμβάνουμε την εξίσωση που συσχετίζει την κίνηση της κάμερας με το διάνυσμα κίνησης σε ένα επιλεγμένο σημείο *(x,y)* της εικόνας ως:

$$u(x,y)=\begin{bmatrix} -\frac{T_x}{Z(x,y)}-R_y+R_z y-x\left(-\frac{T_z}{Z(x,y)}-R_x y+R_y x\right) \\ -\frac{T_y}{Z(x,y)}-R_z x+R_x-y\left(-\frac{T_z}{Z(x,y)}-R_x y+R_y x\right) \end{bmatrix} \quad \textbf{(22)}$$

όπου *Z(x,y)* η απόσταση της κάμερας από το αντικείμενο που απεικονίζεται στην θέση *(x,y)*, με την αρχή του επιπέδου *(x,y) = (0,0)* στην πάνω-αριστερή γωνία των εικόνων. Τώρα αν υποθέσουμε πως έχουμε μηδενική περιστροφή η (22) γίνεται:

$$u(x,y)=\begin{bmatrix} \frac{-T_x+xT_z}{Z(x,y)} \\ \frac{-T_y+xT_z}{Z(x,y)} \end{bmatrix} \quad \textbf{(23)}$$

Τότε παρατηρούμε ότι το διάνυσμα *u* μηδενίζεται στο σημείο $(x,y)=\left(\frac{T_x}{T_z},\frac{T_y}{T_z}\right)$. Το σημείο αυτό λέγεται **εστία επέκτασης** (*focus of expansion, FOE*) και μιλώντας πρακτικά θα λέγαμε ότι είναι εκείνο το σημείο όπου τα διανύσματα οπτικής ροής «δείχνουν» όταν πραγματοποιείται οπίσθια μεταφορική κίνηση κάθετη στο επίπεδο της εικόνας (zoom-out) και περιστρέφονται γύρω από αυτό όταν έχουμε περιστροφική κίνηση κατά τον οπτικό άξονα (clock - anticlockwise rotate) (σχήμα 3-8). Δίνοντας μια ρεαλιστική χροιά θα λέγαμε ότι η εστία επέκτασης είναι το σημείο σύγκλισης των κάθετων στο οπτικό πλάνο ευθειών. (Το σημείο «συνάντησης» των ράγων τρένου όταν παρατηρούνται από την θέση της τροχιάς τους).



Σχήμα 3-8: *Το zoom-in/out και η περιστροφή γύρω από τον οπτικό άξονα αποκαλύπτει την θέση της εστίας επέκτασης, στις περισσότερες των περιπτώσεων στο κέντρο της εικόνας.*

Αν υποθέσουμε ότι η αρχή του επιπέδου της εικόνας *(x,y)* δεν βρίσκεται πια στην πάνω-αριστερή γωνία αλλά στην θέση της εστίας επέκτασης τότε η (23) απλοποιείται ακόμα περισσότερο, και για την μεταφορική κίνηση γίνεται:

$$u(x,y)=\begin{bmatrix}\dfrac{xT_z}{Z(x,y)}\\[2ex]\dfrac{yT_z}{Z(x,y)}\end{bmatrix} \qquad \textbf{(24)}$$

Η θέση της εστίας επέκτασης εξαρτάται κυρίως από το υλικό όπως το CCD της κάμερας λήψης και τον φακό. Ωστόσο στις περισσότερες των περιπτώσεων αυτό βρίσκεται στο κέντρο των εικόνων. Έτσι επιλέγοντας σύστημα αξόνων *x-y* που έχει ως αρχή το κέντρο των εικόνων μπορούμε να απλοποιήσουμε και συμπτύξουμε την (22) στην τελική της μορφή ως:

$$u(x,y)=\begin{bmatrix}1 & 0 & -x\\ 0 & 1 & -y\end{bmatrix}\left(\frac{T}{Z(x,y)}+R\times\begin{bmatrix}x\\ y\\ 1\end{bmatrix}\right) \qquad \textbf{(25)}$$

Η (25) θα διαδραματίσει σημαντικό ρόλο στην περαιτέρω διερεύνηση του προβλήματος καθώς αποτελεί το μαθηματικό μοντέλο στο οποίο θα στηριχθεί η ανάλυση της ΕΚΚ.

Κεφάλαιο 4

Εκτίμηση της κίνησης της κάμερας

4.1 Εισαγωγή

Στα προηγούμενα κεφάλαια μιλήσαμε για την εν δυνάμει ποσότητα «τρισδιάστατης πληροφορίας» που εμπεριέχεται σε μια ακολουθία εικόνων που έχει ληφθεί με κίνηση της κάμερας. Στόχος μας σε αυτό ο κεφάλαιο είναι η ανάκτηση αυτή της πληροφορίας και η ποσοτικοποίηση της προκειμένου να μπορεί να αξιοποιηθεί από κάποια εφαρμογή. Μάλιστα η διαδικασία της ανάκτησης θα αποκτούσε ιδιαίτερη πρακτική αξία εφόσον μπορούσαν να εξαχθούν αξιόπιστα αποτελέσματα παρουσία θορύβου στα δεδομένα εισόδου. Στις ενότητες που ακολουθούν θα δούμε την μαθηματική περιγραφή του προβλήματος της ΕΚΚ, την προτεινόμενη μέθοδο επίλυσης καθώς και μία σειρά από ειδικούς χειρισμούς που βελτιώνουν την ακρίβεια των εκτιμήσεων.

4.2 Ορισμός και διατύπωση του προβλήματος

4.2.1 Προϋποθέσεις και παράμετροι

Η μέθοδος της ΕΚΚ που προτείνουμε εδώ δεν προϋποθέτει κάποιες ειδικές συνθήκες για τα δεδομένα εισόδου, πλην τριών που θα πρέπει να ικανοποιούνται υποχρεωτικά και ως ένα βαθμό αφορούν την εκτίμηση της οπτικής ροής. Οι συνθήκες που θα πρέπει να ισχύουν είναι οι εξής:

- Το περιβάλλον λήψης πρέπει να παραμένει αμετάβλητο (*rigid scene*).

- Η οπτική ροή που προκαλείται από την κίνηση της κάμερας επιβάλλεται να είναι συμβατή με το οπτικό μοντέλο που ορίζει η (25).
- Τα διανύσματα του πεδίου οπτικής ροής πρέπει να είναι ομοιόμορφα κατανεμημένα στο οπτικό πλάνο.

Εφόσον τηρούνται οι συνθήκες αυτές, μπορούμε να εφαρμόσουμε την προτεινόμενη μέθοδο ΕΚΚ σε οποιαδήποτε ακολουθία εικόνων.

Λίγο πριν εμβαθύνουμε στην μέθοδο της ΕΚΚ ορίζουμε τις παραμέτρους του προβλήματος ως εξής:

- f η εστιακή απόσταση.
- $\boldsymbol{x} = (x_1,x_2,1)^T$ μια θέση στην εικόνα διαβαθμισμένη σε *FLU* σύμφωνα με το οπτικό πεδίο (*FOV*) και το σημείο επέκτασης του οπτικού πλάνου.
- $\boldsymbol{X} = (X_1,X_2,X_3)^T$ η αντίστοιχη θέση της $\boldsymbol{x}$ στον χώρο.
- $u(\boldsymbol{x})$ το διάνυσμα οπτικής ροής στην θέση $\boldsymbol{x}$ και $U(\boldsymbol{X})$ το αντίστοιχο τρισδιάστατο διάνυσμα ταχύτητας στον χώρο.
- $\boldsymbol{t}$ το διάνυσμα μεταφοράς σε *FLU*.
- $\boldsymbol{\omega}$ το διάνυσμα περιστροφής σε *ακτίνια*.
- και $Z(\boldsymbol{x})$ η κάθετη απόσταση από το κέντρο του οπτικού πλάνου στο $\boldsymbol{X}$ σημείο του χώρου.

Με αντικατάσταση αυτών των παραμέτρων στην (25) αυτή γίνεται:

$$u(\mathbf{x}) = A(\mathbf{x})\left(\frac{t}{Z(\mathbf{x})} + \omega \times \mathbf{x}\right) \qquad \textbf{(26)}$$

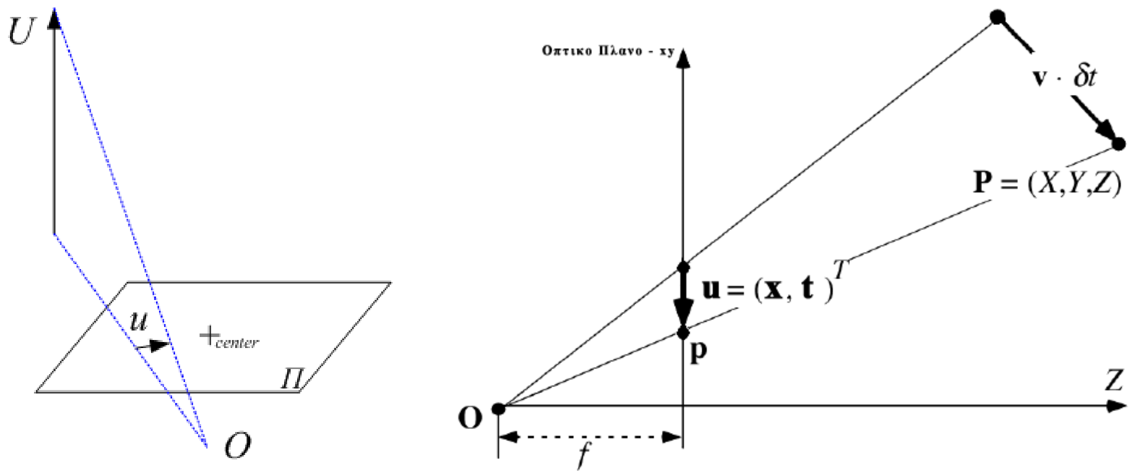
Όπου $A(\boldsymbol{x})$ ο πίνακας:

$$A(\mathbf{x}) = \begin{bmatrix} 1 & 0 & -x_1 \\ 0 & 1 & -x_2 \end{bmatrix} \qquad \textbf{(27)}$$

Το πρόβλημα της ΕΚΚ ισοδυναμεί με τον υπολογισμό των παραμέτρων $t, \omega, Z(\boldsymbol{x})$ με δεδομένα εισόδου ένα πλήθος διανυσμάτων $u(\boldsymbol{x})$ σε $\boldsymbol{x}$ διαθέσιμα σημεία.

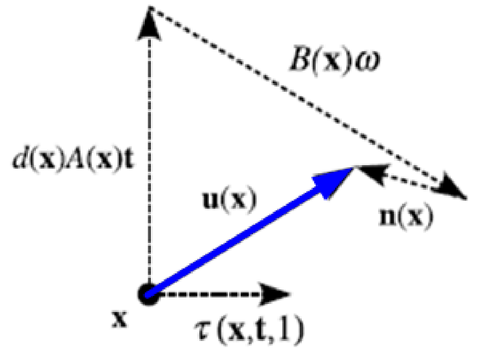
4.2.2 Μη-γραμμικό μοντέλο και ελαχιστοποίηση

Όπως παρατηρούμε στην (26) η τελευταία στήλη του *A(**x**)* ρυθμίζει το μέγεθος των διανυσμάτων *u(**x**)* που οφείλονται στην τρίτη συνιστώσα του *U(**X**)*. Ένα εμπειρικό παράδειγμα που βοηθάει να κατανοήσουμε τον ρόλο του *A(**x**)* είναι ότι κατά την εμπρόσθια μεταφορά zoom-in, (όπου λαμβάνει τιμή μόνο η τρίτη συνιστώσα του *U(**X**)*) τα διανύσματα *u(**x**)* που βρίσκονται κοντύτερα στο κέντρο της εικόνας έχουν μικρότερο μέγεθος σε σχέση με αυτά που βρίσκονται περιμετρικά. Ουσιαστικά ο *A(**x**)* είναι ο πίνακας προβολής του *U(**X**)* στο οπτικό πλάνο (Σχήμα 4-1).



Σχήμα 4-1: *Η προβολή του τρισδιάστατου διανύσματος U πάνω στο οπτικό πλάνο «σχηματίζει» το διάνυσμα οπτικής ροής υ.*

Παρατηρώντας την (26) βλέπουμε ότι η μεταφορά *t* και το βάθος *Z(**x**)* είναι μεγέθη ανάλογα και όπως επισημάναμε στην ενότητα 3.3 τα απόλυτα μεγέθη των παραμέτρων αυτών δεν είναι δυνατό να υπολογιστούν. Αυτό δημιουργεί την αναγκαιότητα «δέσμευσης» του κλάσματος *t/Z(**x**)* βάσει του ενός εκ των δύο μεγεθών. Επειδή το *t* είναι κοινή παράμετρος για κάθε *Z(**x**)* είναι λογική συνέπεια η «δέσμευση» να γίνει ως προς αυτό. Έτσι, για την μέθοδο μας, εισάγουμε τον περιορισμό του μέτρου του *t* ως $\|t\|=1$, με προφανή συνέπεια ότι η μεταφορά δεν μπορεί πλέον να είναι μηδενική. Στο κεφάλαιο 5 θα δούμε πως επηρεάζονται οι εκτιμήσεις της μεθόδου στην περίπτωση παραβίασης αυτής της συνθήκης.



Σχήμα 4-2: *Κάθε διάνυσμα οπτικής ροής μπορεί να αναλυθεί σε ένα συνδυασμό διανυσμάτων οφειλόμενων στην μεταφορά A(x)t και την περιστροφή B(x)ω καθώς και ένα κομμάτι n(x) που αντιστοιχεί σε θόρυβο. Το μοναδιαίο διάνυσμα τ είναι κάθετο στο A(x)t.*

Συνεχίζοντας την ανάλυση του προβλήματος, η (26) για λόγους απλότητας μπορεί να γραφεί ως:

$$u(\mathbf{x}) = d(\mathbf{x})A(\mathbf{x})t + B(\mathbf{x})\omega + n(\mathbf{x}) \tag{28}$$

Όπου *d(**x**)* το αντίστροφο βάθος *1/Z(**x**)* και *n(**x**)* = $(n_x, n_y)^T$ διάνυσμα του οποίου τα στοιχεία λαμβάνουν *ανεξάρτητο και ομοιόμορφα κατανεμημένο με μηδενική μέση τιμή* Gaussian θόρυβο. *B(**x**)* είναι ο πίνακας προβολής των διανυσμάτων *U(**X**)* που οφείλονται στην περιστροφή:

$$B(\mathbf{x}) = \begin{bmatrix} -x_1x_2 & 1+x_1^2 & -x_2 \\ -1-x_2^2 & x_1x_2 & x_1 \end{bmatrix} \tag{29}$$

Αν ορίσουμε ως **r**(**x**) το υπόλοιπο (residual) για κάθε σημείο ***x*** μεταξύ του «πραγματικού» διανύσματος $u_*(\mathbf{x})$ και του *u(**x**)* εκτιμώμενου διανύσματος που προκύπτει μέσω των παραμέτρων *t,ω,d(**x**)* της (28) έχουμε:

$$\mathbf{r}(\mathbf{x}) = \mathbf{u}_*(\mathbf{x}) - d(\mathbf{x})A(\mathbf{x})t - B(\mathbf{x})\omega \tag{30}$$

τότε η λύση του προβλήματος της ΕΚΚ [1,2] προκύπτει ως ελαχιστοποίηση της μη-γραμμικής (30) ως:

$$\left(\hat{t}, \omega, d\right) = \arg\min_{t,\omega,d} \frac{1}{m} \sum_{\{\mathbf{x}\}} f\left(\mathbf{r}(\mathbf{x})\right) \tag{31}$$

με *{**x**}* τις συντεταγμένες των *m* διανυσμάτων *u(**x**)* και *f(**x**)* οποιαδήποτε κυρτή, περιστροφικά συμμετρική συνάρτηση. Από την στιγμή που η **r**(**x**) είναι διάνυσμα δυο στοιχείων στην πράξη η συνάρτηση κόστους θα πρέπει να είναι μία *p*-νόρμα $f(x) = \|x\|_p^p$ με $1 \le p \le 2$.

4.2.3 Μετασχηματισμός μη-γραμμικού συστήματος και νόρμας

Γενικώς το υπόλοιπο δυο στοιχείων δεν παρουσιάζεται τόσο συχνά σε προβλήματα και είναι φανερό πως περιορίζει τις πιθανές συναρτήσεις κόστους σε *p*-νόρμες. Προκειμένου να αποφύγουμε το ζήτημα αυτό καθώς και την ακριβή υπολογιστικά νόρμα μπορούμε να χειριστούμε το θέμα ως εξής: *(βασισμένο στην μέθοδο Zhang-Tomasi [1,2])*

Αν υποθέσουμε ότι έχουμε ένα τυχαίο διάνυσμα $\alpha = [\alpha_1, \alpha_2]^T$ τότε το *κάθετο μοναδιαίο διάνυσμα* του είναι το $Q(\alpha) = \frac{[\alpha_2 \quad -\alpha_1]^T}{\|a\|}$. Έστω *e* το μοναδιαίο διάνυσμα προβολής της μεταφοράς *A(**x**)t* στο οπτικό πλάνο:

$$e(x,t,1) = \frac{A(\mathbf{x})t}{\|A(\mathbf{x})t\|} \tag{32}$$

και *τ* το κάθετο μοναδιαίο διάνυσμα της δισδιάστατης προβολής της μεταφοράς *A(**x**)t*:

$$\tau(\mathbf{x},t,1) = Q\big(A(\mathbf{x})t\big) = \frac{\big([A(\mathbf{x})t]_2, -[A(\mathbf{x})t]_1\big)^T}{\|A(\mathbf{x})t\|} \tag{33}$$

Το σχήμα 4-2 δείχνει την σημασία του *e*, *τ*. Τώρα με πολλαπλασιασμό του *e* στην (30) έχουμε:

$$\begin{gathered} e^T\mathbf{r}(\mathbf{x}) = e^T\mathbf{u}_*(\mathbf{x}) - e^T d(\mathbf{x})A(\mathbf{x})t - e^T B(\mathbf{x})\omega \overset{(32)}{\Leftrightarrow} \\ e^T\mathbf{r}(\mathbf{x}) = e^T\mathbf{u}_*(\mathbf{x}) - d(\mathbf{x})\|A(\mathbf{x})\| - e^T B(\mathbf{x})\omega \end{gathered} \tag{34}$$

Λύνοντας ως προς το αντίστροφο βάθος *d(**x**)* λαμβάνουμε:

$$d(\mathbf{x}) = \frac{e^T\left(\mathbf{u}_*(\mathbf{x}) - B(\mathbf{x})\omega\right)}{\|A(\mathbf{x})t\|} = \frac{\left(A(\mathbf{x})t\right)^T\left(\mathbf{u}_*(\mathbf{x}) - B(\mathbf{x})\omega\right)}{\|A(\mathbf{x})t\|^2} \tag{35}$$

Αντίστοιχα πολλαπλασιάζοντας την (30) με το τ εξαλείφουμε το αντίστροφο βάθος *d(**x**)* και τελικώς παίρνουμε:

$$\left\|A(\mathbf{x})t\right\| \tau(\mathbf{x},t,1)^T \left(\mathbf{u}_*(\mathbf{x}) - B(\mathbf{x})\omega\right) = 0 \qquad \textbf{(36)}$$

Η (36) είναι ανεξάρτητη από το βάθος *d(**x**)* και δεδομένης της μεταφοράς *t* εξαρτάται μόνο από την περιστροφή *ω*. Παρατηρώντας την (36) διακρίνουμε ότι είναι μια σταθμισμένη κατά $\left\|A(\mathbf{x})t\right\|$ γραμμική εξίσωση. Όπως θα δούμε αργότερα ο συντελεστής $\left\|A(\mathbf{x})t\right\|$ απλοποιεί σε ένα βαθμό τους υπολογισμούς αλλά όπως αποδεικνύεται και από άλλες μεθόδους [10] όσες χρησιμοποιούν σταθμισμένες εξισώσεις παρόμοιες με την (36) παράγουν εσφαλμένες εκτιμήσεις. Ο λόγος είναι ότι η ο $\left\|A(\mathbf{x})t\right\|$ παίζει το ρόλο «βάρους» στο υπόλοιπο $\tau\left(\mathbf{x},t,1\right)^T \left(u(\mathbf{x}) - B(\mathbf{x})\omega\right)$ ενώ εξαρτάται από τις συντεταγμένες ***x***. Εντούτοις δεν υπάρχει κανένα ισχυρό επιχείρημα που να μας προτρέπει να χρησιμοποιήσουμε διαφορετικά βάρη στις εξισώσεις ανάλογα με την θέση ***x*** των διανυσμάτων στην εικόνα. Αντιθέτως μάλιστα όλες οι εξισώσεις θα πρέπει να έχουν την ίδια επίδραση ανεξάρτητα από την θέση αυτή. Συνεπώς η ελαχιστοποίηση θα πρέπει να γίνεται χωρίς τον συντελεστή $\left\|A(\mathbf{x})t\right\|$. Αργότερα θα διαπιστώσουμε πως ο συντελεστής $\left\|A(\mathbf{x})t\right\|$ σχετίζεται με τα ελάχιστα στην συνάρτηση κόστους και ότι αυτά δύνανται να μειωθούν ρυθμίζοντας σταδιακά την επίδραση του βάρους αυτού. Τελικά το νέο προς ελαχιστοποίηση (μονοδιάστατο) υπόλοιπο γράφεται ως:

$$r(\mathbf{x}) = \tau(\mathbf{x},t,1)^T \mathbf{r}(\mathbf{x}) = \tau(\mathbf{x},t,1)^T \left(u(\mathbf{x}) - B(\mathbf{x})\omega\right) \qquad \textbf{(37)}$$

Με συνάρτηση κόστους την *f(x) = x^2*, η εκτίμηση των παραμέτρων *t,ω* προκύπτει από:

$$\left(\hat{t},\omega\right) = \arg\min_{\omega} \frac{1}{m} \sum_{\{\mathbf{x}\}} \left[\tau(\mathbf{x},t,1)^T \left(u(\mathbf{x}) - B(\mathbf{x})\omega\right)\right]^2 \qquad \textbf{(38)}$$

Η (38) είναι σχέση διγραμμικού περιορισμού (bilinear constraint) και ελαχιστοποιεί το ολικό σφάλμα που προκύπτει από την προβολή των διανυσμάτων που οφείλονται στην περιστροφική κίνηση.

Οι εξισώσεις που είδαμε σε αυτή την ενότητα μας δίνουν την δυνατότητα να διαχειριστούμε το πρόβλημα πιο αποτελεσματικά (σε σχέση με την απευθείας επίλυση του μεγάλου μη-γραμμικού συστήματος (30)) διαχωρίζοντας το σε επιμέρους μικρότερα προβλήματα προς λύση. Στις ενότητες που ακολουθούν θα δούμε πως αξιοποιούνται οι μαθηματικοί χειρισμοί που επιτελέσαμε εδώ καθώς και τα οφέλη στην ακρίβεια των εκτιμήσεων και το υπολογιστικό κόστος. Γενικότερα θα πρέπει να γνωρίζουμε ότι όσες μέθοδοι βασίζονται σε σχέσεις ίδιες ή παρόμοιες με τις παραπάνω έχουν κοινό χαρακτηριστικό ότι παράγουν ακριβείς εκτιμήσεις με χαμηλά επίπεδα *πόλωσης* και *αποκλίσεων*, γι' αυτό και συχνά αναφέρονται ως «βέλτιστες» (*optimal*).

4.3 Σχετικές εργασίες

Τα τελευταία χρόνια έχουν παρουσιαστεί διάφορες μέθοδοι ΕΚΚ. Κάθε μια από τις μεθόδους αυτές έχει να αναδείξει τα δικά της πλεονεκτήματα και χονδρικά θα λέγαμε ότι χωρίζονται σε δυο βασικές κατηγορίες ανάλογα με την αποτελεσματικότητα τους: Τις παλαιότερες προσεγγίσεις που εφάρμοζαν υπολογιστικές απλοποιήσεις χάριν ευκολίας αλλά τελικώς έπασχαν από πόλωση και έντονες αποκλίσεις στις εκτιμήσεις τους και τις πιο πρόσφατες προσεγγίσεις που βασίζονται σε μετασχηματισμούς και εξισώσεις όπως αυτές που περιγράψαμε στην προηγούμενη ενότητα. Έπεται μια σύντομη περιγραφή μερικών τυπικών προσεγγίσεων του προβλήματος.

Μια από τις πρώτες δημοφιλείς μεθόδους [10] χρησιμοποιεί σαν αφετηρία τις ίδιες εξισώσεις κίνησης που ορίσαμε πριν άλλα στη συνέχεια αντικαθιστά την αυθεντική συνάρτηση υπολοίπου (30) με την:

$$\mathbf{r}'(\mathbf{x}) = \mathbf{r}(\mathbf{x}) \left\| A(\mathbf{x}) t \right\| \quad \textbf{(39)}$$

Το πλεονέκτημα από αυτή την αντικατάσταση είναι ότι εμμέσως εξαλείφεται το βάθος. Έτσι η ελαχιστοποίηση γίνεται μόνο με τις έξι παραμέτρους που αφορούν την περιστροφή και την μεταφορά με προφανές κέρδος την απλοποίηση των υπολογισμών. Το όφελος είναι ακόμα μεγαλύτερο επειδή η (39) έχει απλό κοίλο σχήμα και παρουσιάζει μικρότερο αριθμό ελαχίστων σε σχέση με την αυθεντική, πράγμα που την καθιστά ευκολότερη στην ελαχιστοποίηση.

Παρατηρούμε ότι η (39) είναι παρόμοια με την (36). Όπως εξηγήσαμε προηγούμενα η (36) πρέπει να λυθεί χωρίς τον συντελεστή $\left\| A(\mathbf{x}) t \right\|$ ειδάλλως οδηγούμαστε σε εσφαλμένες εκτιμήσεις αφού αλλοιώνεται το σχήμα της συνάρτησης κόστους. Επειδή η $\left\| A(\mathbf{x}) t \right\|$ εξαρτάται από το ***x***, σταθμίζει το **r**(**x**) με βάση την θέση των διανυσμάτων στην εικόνα, κάτι που δεν δικαιολογείται. Πρακτικά το αποτέλεσμα από την χρήση της (39) είναι ότι όταν έχουμε μεγάλο πλήθος διανυσμάτων με συντεταγμένες μεγαλύτερες της μονάδας δημιουργείται τάση για χαμηλότερες τιμές στο τρίτο στοιχείο t_3 της μεταφοράς αφού ο αλγόριθμος προσπαθεί να εξισορροπήσει τα μεγαλύτερα **r**(**x**) που προκύπτουν λόγω των μεγαλύτερων $\|A(\mathbf{x})t\|$. Το αντίθετο συμβαίνει όταν πολλά διανύσματα έχουν συντεταγμένες μικρότερες της μονάδας όπου δημιουργείτε τάση επαύξησης του t_3.

Οι θέσεις των διανυσμάτων εξαρτώνται και από το οπτικό πεδίο. Αν το οπτικό πεδίο είναι μικρό τότε και το οπτικό πλάνο θα είναι μικρό και έτσι μοιραία θα υπάρχουν πολλά διανύσματα με συντεταγμένες μικρότερες του 1 με συνέπεια την τάση για μεγαλύτερο t_3. Το αντίθετο ισχύει σε μεγαλύτερο οπτικό πεδίο. Συνοψίζοντας, με μικρό οπτικό πεδίο έχουμε πόλωση της μεταφοράς προς την κατεύθυνση zoom-in/out, με μεγαλύτερο οπτικό πεδίο έχουμε πόλωση της μεταφοράς προς πλευρικές κινήσεις. Επειδή και οι άλλες μέθοδοι που περιγράφουμε παρακάτω χρησιμοποιούν σαν αφετηρία σχέσεις όπως η (39) σε κάποιο βαθμό πάσχουν από παρόμοια προβλήματα.

Μια άλλη μέθοδος [11] εφαρμόζοντας ένα μετασχηματισμό στην (39) δημιουργεί μια νέα σχέση με αγνώστους το βάθος και την περιστροφή συναρτήσει της μεταφοράς την οποία επιλύει με ελάχιστα τετράγωνα. Στην συνέχεια θέτει τα αποτελέσματα για το βάθος και την περιστροφή στην (39) ώστε να υπολογίσει τον μοναδικό άγνωστο πλέον, την μεταφορά. Από την στιγμή που και αυτή η μέθοδος χρησιμοποιεί την (39) πάσχει από πόλωση αντίστοιχης αιτιότητας όπως η μέθοδος *Horn & Bruss*.

Οι δυο προηγούμενες μέθοδοι χρησιμοποιούν επαναληπτικές διαδικασίες για την επίλυση των μη γραμμικών συστημάτων τους. Μια εναλλακτική προσέγγιση [12, 13, 14] υπολογίζει την μεταφορά χωρίς επαναληπτική διαδικασία. Στην γραμμική μέθοδο υποχώρου (linear subspace method) υποθέτουμε ότι από το σύνολο *N* διανυσμάτων οπτικής ροής ένας αριθμός *N-6* ανήκει σε μια ομάδα «διανυσμάτων-περιορισμού» που είναι (όσο το δυνατόν) κάθετα στην διεύθυνση μεταφοράς και ανεξάρτητα από την περιστροφή. Στην συνέχεια κατασκευάζουμε ένα πίνακα με κατάλληλη δομή για τα *N–6* «διανύσματα-περιορισμού». Σύμφωνα με την μέθοδο η μεταφορά *t* είναι το ιδιοδιάνυσμα που αντιστοιχεί στην μικρότερη ιδιοτιμή του παραπάνω πίνακα. Δυστυχώς και αυτή η μέθοδος όπως οι προηγούμενες πάσχει από πόλωση συν ότι δεν αξιοποιεί όλη τη διαθέσιμη πληροφορία *N*. Μια πιθανή χρήση μεθόδων όπως αυτή είναι η παραγωγή εκτιμήσεων που θα μπορούσαν να χρησιμοποιηθούν για αρχικοποίηση στις επαναληπτικές μεθόδους. Εμπειρικά έχει αποδειχθεί ότι πολλές φορές οι εκτιμήσεις τους δεν ενδείκνυνται ούτε για αυτό το σκοπό.

Γενικώς οι περισσότερες μέθοδοι συστηματικά αποδίδουν αποτελέσματα με τάσεις προς συγκεκριμένες κατευθύνσεις. Μια άλλη πρόταση [15] επιχειρεί να μετρήσει την τάση αυτή και να την άρει από το τελικό αποτέλεσμα. Ωστόσο ούτε αυτή η πρόταση αποδίδει καλά αποτελέσματα γιατί βελτιώνει την πόλωση αλλά δημιουργεί μεγαλύτερη ευαισθησία στον θόρυβο (αποκλίσεις). Εκτός αυτού θεωρεί ως υπόθεση μικρό οπτικό πεδίο κάτι ιδιαίτερα περιοριστικό στην πράξη.

Συμπερασματικά είναι φανερό ότι η πόλωση και οι αποκλίσεις ενισχύονται από ακατάλληλους μετασχηματισμούς του προβλήματος μέσω γραμμικοποίησεων ή άλλων υποκατάστατων των βασικών εξισώσεων του μοντέλου. Οι περισσότερες μέθοδοι ήταν εκτεθειμένες σε αυτούς τους κινδύνους αφού περιείχαν τέτοιες μετατροπές με αποτέλεσμα η πόλωση και η αποκλίσεις να είναι μεγαλύτερες από όσο θα' πρεπε. Γενικώς οι εκτιμήσεις των μεθόδων ΕΚΚ σε ένα βαθμό ήταν απογοητευτικές με αποτέλεσμα πολλοί ερευνητές να πιστεύουν ότι το πρόβλημα αυτό δεν έχει συνεπή λύση. Ειδικά στην περίπτωση μικρών οπτικών πεδίων, η πόλωση και οι αποκλίσεις ήταν τόσο αυξημένες που θα λέγαμε ότι τα αποτελέσματα ήταν … ανώφελα! Ωστόσο σε αυτή την εργασία υλοποιούμε έναν αλγόριθμο ΕΚΚ με χαμηλά επίπεδα πόλωσης και αποκλίσεων του οποίου οι εκτιμήσεις συγκλίνουν στις πραγματικές τιμές όσο αυξάνεται η ποσότητα πληροφορίας που εισάγουμε.

4.4 Μεθοδολογία για την λύση του προβλήματος

Σε αυτή την ενότητα θα δούμε την επαναληπτική μέθοδο επίλυσης μη-γραμμικών συστημάτων Gauss-Newton και στην συνέχεια την εφαρμογή της σε μια απλή εκδοχή του προβλήματος (ενότητα 4.4.2). Στην συνέχεια εκμεταλλευόμενοι τις σχέσεις που παράγαμε στην ενότητα 4.2.3 θα δείξουμε μια ταχύτερη εκδοχή του αλγορίθμου καθώς και δυο ακόμα που επιχειρούν να «θωρακίσουν» την διαδικασία ενάντια στο θόρυβο. Η υλοποίηση βασίζεται στην μέθοδο *Zhang-Tomasi* [1,2].

4.4.1 Διαδικασία αναδρομικής γραμμικοποίησης Gauss-Newton

Αν υποθέσουμε ότι έχουμε το υπόλοιπο *R(ϑ)*, με αγνώστους *ϑ = (t,ω,d)*, σύμφωνα με την Gauss-Newton *διαδικασία επίλυσης μη-γραμμικών συστημάτων* μπορούμε να υπολογίζουμε σε κάθε επανάληψη ένα βήμα *Δϑ* το οποίο προστιθέμενο στο *ϑ* «κατευθύνει» το αποτέλεσμα προς το πλησιέστερο ελάχιστο – λύση της αντικειμενικής συνάρτησης. Μέσω της ανάπτυξης σειρών Taylor αναλύουμε το υπόλοιπο *R(ϑ)* σε:

$$R(\theta_\kappa + \Delta\theta_\kappa) \approx R(\theta_\kappa) + g_k^T + \frac{1}{2}\Delta\theta_\kappa H_k \Delta\theta_\kappa \qquad \textbf{(40)}$$

Όπου $g_k = J_k \rho_k$, $H_k = J_k^T J_k + \sum_{i=1}^{2m} \rho_i H_{ik}$ με H_κ τον πίνακα *Hessian*, J_k τον *Ιακωβιανό* πίνακα, *ρ(ϑ)* το διάνυσμα υπολοίπων $[r_1, \ldots, r_m]^{\mathrm{T}}$ και H_{ik} τον *Hessian* πίνακα του *ρ(ϑ)*. Με παραγώγιση της (40) κατά $\Delta\vartheta_k$ και υπό την προϋπόθεση ότι τα υπόλοιπα ρ_i λαμβάνουν σχετικά μικρές τιμές έχουμε:

$$J_k \Delta\theta_k = -\rho_k \qquad \textbf{(41)}$$

Όπου η σχέση αυτή είναι μια απλή γραμμική εξίσωση που λύνεται εύκολα με οποιαδήποτε από τις γνωστές μεθόδους.

4.4.2 Επίλυση με Gauss-Newton

Όπως επισημάναμε πριν μπορούμε να χειριστούμε το δισδιάστατο υπόλοιπο r(**x**) της (30) σαν οποιαδήποτε άλλη βαθμωτή συνάρτηση υπολοίπου εφόσον χρησιμοποιήσουμε για την ελαχιστοποίηση μια συνάρτηση κόστους τύπου *p*-νόρμα $\|\mathrm{x}\|_p^p$. Αν *p* = 2 (δηλαδή $f(\mathrm{x}) = \|\mathrm{x}\|_2^2$) ουσιαστικά έχουμε την πλέον συνηθισμένη περίπτωση των ελαχίστων τετραγώνων:

$$\left(\hat{t}, \omega, d\right) = \arg\min_{t,\omega,d} \frac{1}{m} \sum_{\{\mathbf{x}\}} \|\mathbf{r}(\mathbf{x})\|_2^2 \qquad \textbf{(42)}$$

Αυτό μπορεί να επιτευχθεί υπολογίζοντας τον Ιακωβιανό πίνακα της (28) και στην συνέχεια επιλύοντας επαναληπτικά με την (41) ως την σύγκλιση του $\Delta\vartheta$. Ο Ιακωβιανός πίνακας της (28) για m = 6 (για οικονομία χώρου!) είναι:

$$J_k = \begin{bmatrix}
\frac{dt_1}{dx_1} & \frac{dt_2}{dx_1} & \frac{dt_3}{dx_1} & \frac{dd_1}{dx_1} & 0 & 0 & 0 & 0 & 0 & \frac{d\omega_1}{dx_1} & \frac{d\omega_2}{dx_1} & \frac{d\omega_3}{dx_1} \\
\frac{dt_1}{dy_1} & \frac{dt_2}{dy_1} & \frac{dt_3}{dy_1} & \frac{dd_1}{dy_1} & 0 & 0 & 0 & 0 & 0 & \frac{d\omega_1}{dy_1} & \frac{d\omega_2}{dy_1} & \frac{d\omega_3}{dy_1} \\
\frac{dt_1}{dx_2} & \frac{dt_2}{dx_2} & \frac{dt_3}{dx_2} & 0 & \frac{dd_2}{dx_2} & 0 & 0 & 0 & 0 & \frac{d\omega_1}{dx_2} & \frac{d\omega_2}{dx_2} & \frac{d\omega_3}{dx_2} \\
\frac{dt_1}{dy_2} & \frac{dt_2}{dy_2} & \frac{dt_3}{dy_2} & 0 & \frac{dd_2}{dy_2} & 0 & 0 & 0 & 0 & \frac{d\omega_1}{dy_2} & \frac{d\omega_2}{dy_2} & \frac{d\omega_3}{dy_2} \\
\frac{dt_1}{dx_3} & \frac{dt_2}{dx_3} & \frac{dt_3}{dx_3} & 0 & 0 & \frac{dd_3}{dx_3} & 0 & 0 & 0 & \frac{d\omega_1}{dx_3} & \frac{d\omega_2}{dx_3} & \frac{d\omega_3}{dx_3} \\
\frac{dt_1}{dy_3} & \frac{dt_2}{dy_3} & \frac{dt_3}{dy_3} & 0 & 0 & \frac{dd_3}{dy_3} & 0 & 0 & 0 & \frac{d\omega_1}{dy_3} & \frac{d\omega_2}{dy_3} & \frac{d\omega_3}{dy_3} \\
\frac{dt_1}{dx_4} & \frac{dt_2}{dx_4} & \frac{dt_3}{dx_4} & 0 & 0 & 0 & \frac{dd_4}{dx_4} & 0 & 0 & \frac{d\omega_1}{dx_4} & \frac{d\omega_2}{dx_4} & \frac{d\omega_3}{dx_4} \\
\frac{dt_1}{dy_4} & \frac{dt_2}{dy_4} & \frac{dt_3}{dy_4} & 0 & 0 & 0 & \frac{dd_4}{dy_4} & 0 & 0 & \frac{d\omega_1}{dy_4} & \frac{d\omega_2}{dy_4} & \frac{d\omega_3}{dy_4} \\
\frac{dt_1}{dx_5} & \frac{dt_2}{dx_5} & \frac{dt_3}{dx_5} & 0 & 0 & 0 & 0 & \frac{dd_5}{dx_5} & 0 & \frac{d\omega_1}{dx_5} & \frac{d\omega_2}{dx_5} & \frac{d\omega_3}{dx_5} \\
\frac{dt_1}{dy_5} & \frac{dt_2}{dy_5} & \frac{dt_3}{dy_5} & 0 & 0 & 0 & 0 & \frac{dd_5}{dy_5} & 0 & \frac{d\omega_1}{dy_5} & \frac{d\omega_2}{dy_5} & \frac{d\omega_3}{dy_5} \\
\frac{dt_1}{dx_6} & \frac{dt_2}{dx_6} & \frac{dt_3}{dx_6} & 0 & 0 & 0 & 0 & 0 & \frac{dd_6}{dx_6} & \frac{d\omega_1}{dx_6} & \frac{d\omega_2}{dx_6} & \frac{d\omega_3}{dx_6} \\
\frac{dt_1}{dy_6} & \frac{dt_2}{dy_6} & \frac{dt_3}{dy_6} & 0 & 0 & 0 & 0 & 0 & \frac{dd_6}{dy_6} & \frac{d\omega_1}{dy_6} & \frac{d\omega_2}{dy_6} & \frac{d\omega_3}{dy_6}
\end{bmatrix}$$

όπου σε κάθε στήλη τοποθετούμε τις μερικές παραγώγους των $t,\omega,d(\mathbf{x})$ ενώ ανά δύο γραμμές τις μερικές παραγώγους των $\boldsymbol{x}_{1,2}$. (Για λόγους σαφήνειας προσωρινά αναφερόμαστε στο x_2 ως y_2). Με αντικατάσταση των μερικών παραγώγων λαμβάνουμε:

$$J_k = \begin{bmatrix}
d_1 & 0 & -x_1d_1 & t_1-x_1t_3 & 0 & 0 & 0 & 0 & 0 & -x_1y_1 & 1+x_1^2 & -y_1 \\
0 & d_1 & -y_1d_1 & t_2-y_1t_3 & 0 & 0 & 0 & 0 & 0 & -1-y_1^2 & x_1y_1 & x_1 \\
d_2 & 0 & -x_2d_2 & 0 & t_1-x_2t_3 & 0 & 0 & 0 & 0 & -x_2y_2 & 1+x_2^2 & -y_2 \\
0 & d_2 & -y_2d_2 & 0 & t_2-y_2t_3 & 0 & 0 & 0 & 0 & -1-y_2^2 & x_2y_2 & x_2 \\
d_3 & 0 & -x_3d_3 & 0 & 0 & t_1-x_3t_3 & 0 & 0 & 0 & -x_3y_3 & 1+x_3^2 & -y_3 \\
0 & d_3 & -y_3d_3 & 0 & 0 & t_2-y_3t_3 & 0 & 0 & 0 & -1-y_3^2 & x_3y_3 & x_3 \\
d_4 & 0 & -x_4d_4 & 0 & 0 & 0 & t_1-x_4t_3 & 0 & 0 & -x_4y_4 & 1+x_4^2 & -y_4 \\
0 & d_4 & -y_4d_4 & 0 & 0 & 0 & t_2-y_4t_3 & 0 & 0 & -1-y_4^2 & x_4y_4 & x_4 \\
d_5 & 0 & -x_5d_5 & 0 & 0 & 0 & 0 & t_1-x_5t_3 & 0 & -x_5y_5 & 1+x_5^2 & -y_5 \\
0 & d_5 & -y_5d_5 & 0 & 0 & 0 & 0 & t_2-y_5t_3 & 0 & -1-y_5^2 & x_5y_5 & x_5 \\
d_6 & 0 & -x_6d_6 & 0 & 0 & 0 & 0 & 0 & t_1-x_6t_3 & -x_6y_6 & 1+x_6^2 & -y_6 \\
0 & d_6 & -y_6d_6 & 0 & 0 & 0 & 0 & 0 & t_2-y_6t_3 & -1-y_6^2 & x_6y_6 & x_6
\end{bmatrix}$$

Τελικά σε συμπτυγμένη μορφή η (42) ελαχιστοποιείται μέσω της αναδρομικής επίλυσης του γραμμικού συστήματος:

$$d_k A(\mathbf{x})\Delta t_k + A(\mathbf{x})t_k\Delta d_k + B(\mathbf{x})\Delta\omega_k = \mathbf{r}_k(\mathbf{x}) \quad \textbf{(43)}$$

όπου ο $\mathbf{r}_k(\mathbf{x})$ είναι ο πίνακας στήλη *1*x*2m*:

$$\mathbf{r}_k(\mathbf{x}) = \begin{bmatrix}
u_{x1} - (t_1 - x_1t_3)d_1 + x_1y_1\omega_1 - (1+x_1^2)\omega_2 + y_1\omega_3 \\
u_{y1} - (t_2 - y_1t_3)d_1 + (1+y_1^2)\omega_1 - x_1y_1\omega_2 - x_1\omega_3 \\
\vdots \\
u_{xm} - (t_1 - x_mt_3)d_m + x_my_m\omega_1 - (1+x_m^2)\omega_2 + y_m\omega_3 \\
u_{ym} - (t_2 - y_mt_3)d_m + (1+y_m^2)\omega_1 - x_my_m\omega_2 - x_m\omega_3
\end{bmatrix} \quad \textbf{(44)}$$

Η (43) είναι ένα υπερκαθορισμένο γραμμικό σύστημα που λύνεται εύκολα με οποιαδήποτε από τις γνωστές μεθόδους. Ωστόσο παρά το ότι το σύστημα έχει *m+6* αγνώστους και *2m* εξισώσεις ο βαθμός του Ιακωβιανού πίνακα είναι *m+5* δηλαδή το σύστημα είναι ιδιόμορφο με μια ελεύθερη μεταβλητή. Αυτή η ιδιομορφία μπορεί να αντιμετωπιστεί αξιοποιώντας τον περιορισμό στο μέτρο του *t* που θέσαμε στην ενότητα 4.2.2. Για να διατηρείτε ο περιορισμός $\|\mathrm{t}\| = 1$ κατά την διάρκεια των επαναλήψεων θα πρέπει οι ανανεώσεις Δt_k να λαμβάνουν τιμές που να συμβάλλουν στην διατήρηση του. Θεωρώντας ότι το $\|\Delta t\|$ λαμβάνει γενικά πολύ μικρότερες τιμές από το $\|\mathrm{t}\|$, το παραπάνω μπορεί να επιτευχθεί όταν οι ανανεώσεις Δt είναι κατά προσέγγιση κάθετες στο διάνυσμα *t* σε κάθε επανάληψη. Από αυτό τον συλλογισμό λαμβάνουμε την νέα σχέση περιορισμού $t_k^T \Delta t_k = 0$.

Για την επίλυση του συστήματος αρχικοποιούμε τα *t,ω,d* με τυχαίες τιμές και λύνουμε επαναληπτικά έως ότου το Δt_k μεταβάλλεται λιγότερο από το κατώφλι σύγκλισης (που ορίσαμε ως 10^{-13}). Στην συνέχεια εκτελούμε και πάλι την ίδια διαδικασία με νέες τυχαίες αρχικές τιμές κ.ο.κ. Τελικά επιλέγουμε ως λύση τον βρόχο που μας έδωσε το μικρότερο σφάλμα (και με μεγάλη πιθανότητα αντιστοιχεί στο ολικό ελάχιστο της συνάρτησης κόστους). Οι πολλαπλές επιλύσεις με διαφορετικές αρχικοποιήσεις επιβάλλονται λόγω των πολλαπλών τοπικών ελαχίστων.

4.4.3 Επίλυση με διαχωρισμένο Gauss-Newton

Ένα ενδιαφέρον στοιχείο που προέκυψε από τους μετασχηματισμούς που πραγματοποιήσαμε στην ενότητα 4.2.3 είναι ότι το πρόβλημα μπορεί να «διασπαστεί» με τέτοιο τρόπο ώστε να υπολογίζουμε τις παραμέτρους *t,ω,d* ξεχωριστά. Πιο συγκεκριμένα παρατηρώντας τις (35) και (38) διαπιστώνουμε ότι δύναται να υπολογίσουμε την περιστροφή και το βάθος συναρτήσει της μεταφοράς. Εκμεταλλευόμενοι το χαρακτηριστικό της διαχωρισιμότητας των παραμέτρων μας δίνεται η δυνατότητα να εφαρμόσουμε τον Gauss-Newton μόνο στο *t*, ξεχωριστά από τις υπόλοιπες παραμέτρους, αφού στην συνέχεια θα μπορούμε να τις υπολογίζουμε συναρτήσει του. Το πλεονέκτημα από κάτι τέτοιο είναι προφανές: πραγματοποιούμε Gauss-Newton σε τρεις αντί για *m+6* μεταβλητές και με μόλις 15 (ενδεικτική εμπειρική τιμή) ομοιόμορφα κατανεμημένες τυχαίες αρχικοποιήσεις μπορούμε εύκολα να εντοπίσουμε το ολικό ελάχιστο. Όπως θα αποδειχθεί αργότερα και στις δόκιμες αυτός ο τρόπος είναι γρήγορος, σταθερός και συγκλίνει στο 93,3% των περιπτώσεων, κατά μέσο όρο σε 35 επαναλήψεις. Για να επωφεληθούμε από το χαρακτηριστικό της διαχωρισιμότητας εργαζόμαστε ως εξής:

Αν στα πλαίσια μιας επανάληψης *k* «γνωρίζουμε» την μεταφορά t_k τότε από τις (32) και (33) μπορούμε να υπολογίσουμε τα e_k και τ_k αντίστοιχα. Η (38) εξαρτάται από τα ω_k και τ_k ενώ είναι ανεξάρτητη από το βάθος d_k. Επομένως μοναδικός άγνωστος είναι το ω_k και μπορούμε να τον υπολογίσουμε επιλύοντας το γραμμικό σύστημα (38). Από τη στιγμή που τα t_k και ω_k είναι διαθέσιμα το d_k υπολογίζεται από την (35). Κάθε κύκλος ανανέωσης ολοκληρώνεται με νέες τιμές για τα t_k, ω_k, d_k. Με αυτές τις νέες τιμές εφαρμόζουμε Gauss-Newton ανανεώσεις για το t μέσω της:

$$\left(\Delta t_k, \omega_k\right) = \arg\min_{\Delta t, \omega} \sum_{\{\mathbf{x}\}} \left[\tau(\mathbf{x}, t, 1)^T \left(u(\mathbf{x}) - d_k A(\mathbf{x}) \Delta t - B(\mathbf{x}) \omega \right) \right]^2 \qquad \textbf{(45)}$$

Από την (45) λαμβάνουμε τελικά το t_{k+1} ($t_{k+1} = t_k + \Delta t_k$) με το οποίο υπολογίζουμε τα ω_{k+1}, d_{k+1} κ.ο.κ.

Αυτός είναι ο κύκλος επεξεργασίας ο οποίος επαναλαμβάνεται ως ότου υπάρξει σύγκλιση σε κάποιο ελάχιστο. Η διαδικασία ξεκινάει με τυχαία αρχικοποίηση του t_0 ενώ σταματάει όταν $\|\Delta t_\kappa\| < 10^{-13}$. Για να εντοπίσουμε το ολικό ελάχιστο η διαδικασία αρχικοποιείται και επαναλαμβάνεται 15 φορές. Τελικά η λύση επιλέγεται από την εκτέλεση εκείνη που απέδωσε το μικρότερο κόστος. Για τον υπολογισμό του ολικού κόστους μπορεί να χρησιμοποιηθεί οποιαδήποτε από τις παρακάτω (ισάξιες) σχέσεις:

$$\begin{aligned} r_{\Delta t}(\mathbf{x}) &= \tau(\mathbf{x}, t, 1)^T \left(u(\mathbf{x}) - d_k A(\mathbf{x}) \Delta t - B(\mathbf{x}) \omega \right) \\ r_{\omega}(\mathbf{x}) &= \tau(\mathbf{x}, t, 1)^T \left(u(\mathbf{x}) - B(\mathbf{x}) \omega \right) \\ \mathbf{r}(\mathbf{x}) &= \left\| \mathbf{u}_*(\mathbf{x}) - d(\mathbf{x}) A(\mathbf{x}) t - B(\mathbf{x}) \omega \right\|_2^2 \end{aligned} \qquad \textbf{(46)}$$

Συνολικά η διαδικασία είναι εξαιρετικά γρήγορη και στην υλοποίηση μας μια επανάληψη «κοστίζει» περίπου 0.003 δευτερόλεπτα (σε *AMD Athlon64 x2*). Ο συνολικός χρόνος απόκρισης σχετίζεται με τον αριθμό των διανυσμάτων που εισάγονται και την ταχύτητα σύγκλισης ενώ η τελευταία εξαρτάται κυρίως από τον θόρυβο στην οπτική ροή.

4.4.4 Επίλυση με σταθμισμένο διαχωρισμένο Gauss-Newton

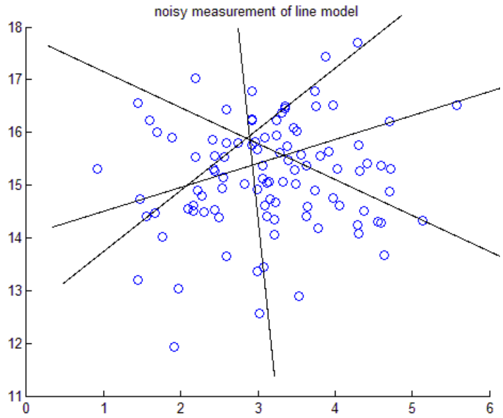
Ίσως το μεγαλύτερο πρόβλημα στην διαδικασία που μελετούμε εδώ είναι ότι η ακρίβεια των εκτιμήσεων είναι σε μεγάλο βαθμό εξαρτώμενη από το θόρυβο. Η αιτία είναι πως μικρά λάθη στις εκτιμήσεις οπτικής ροής δημιουργούν μεγάλα σφάλματα στην ΕΚΚ.

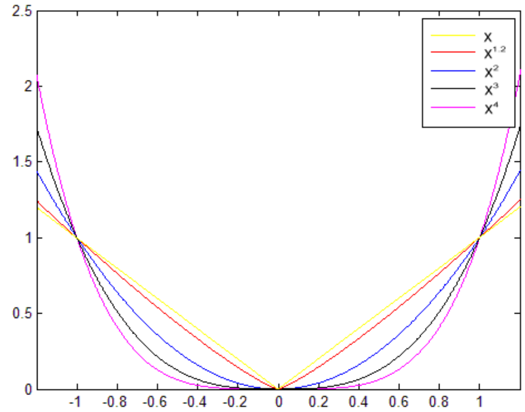
Η εισαγωγή μεγάλου αριθμού διανυσμάτων εξομαλύνει σε ένα βαθμό τις επιπτώσεις του θορύβου. Ωστόσο είναι πολύπλοκο να βρεθεί αναλυτικά ο ενδεδειγμένος αριθμός διανυσμάτων για το σκοπό αυτό. Επίσης η επεξεργασία μεγάλης ποσότητας διανυσμάτων θα είχε και ανάλογο κόστος στον χρόνο απόκρισης του αλγορίθμου. Μια καλύτερη λύση για το πρόβλημα του θορύβου είναι η επίλυση με συναρτήσεις κόστους που αυξάνουν πιο αργά από την κλασική x^2.

Αν θεωρήσουμε ως ένδειξη σφάλματος το υπόλοιπο r(**x**) ενός δείγματος u(**x**), τότε στην επίλυση με ελάχιστα τετράγωνα το κόστος των δειγμάτων με μεγάλο **r**(**x**) μεγεθύνεται λόγω της x^2. Έτσι η ελαχιστοποίηση επιχειρώντας να μειώσει το μεγάλο αυτό κόστος (και τη συνεισφορά του στο ολικό κόστος) παρασύρει την λύση προς την «κατεύθυνση» του σφάλματος. Η ιδέα είναι ότι χρησιμοποιώντας μια συνάρτηση κόστους που αυξάνει πιο αργά από την x^2 υποσκελίζουμε το κόστος των διανυσμάτων με μεγάλο υπόλοιπο και συνεπώς τους δίνουμε λιγότερη σημασία. Παρατηρούμε ότι η βασική ιδέα είναι πανομοιότυπη με αυτή που εφαρμόσαμε στην ενότητα 2.8.

Γενικώς αν η πλειονότητα της πληροφορίας ενός μοντέλου «κατευθύνει» την λύση προς ένα αποτέλεσμα αλλά ένα μικρό κομμάτι της πληροφορίας αποκλίνει από την συμφωνία αυτή τότε προφανώς αυτό το μικρό κομμάτι αποτελεί εσφαλμένη πληροφορία. Φυσικά ήδη παρουσιάσαμε την βασική υπόθεση ότι το εσφαλμένο κομμάτι της πληροφορίας είναι μικρό σε σχέση με το σύνολο της πληροφορίας. Την σημασία αυτής της υπόθεσης μπορούμε να την κατανοήσουμε με ένα απλό παράδειγμα:

Ας υποθέσουμε ότι θέλουμε να επιλύσουμε το απλό πρόβλημα εύρεσης των συντελεστών μιας ευθείας από τα θορυβώδη δείγματα της. Αν τα σημεία (δείγματα) δεν είναι στοιχισμένα σε διάταξη ώστε να διαφαίνεται ο σχηματισμός ευθείας, όπως ίσως θα αναμενόταν, αλλά αντίθετα σχηματίζουν ένα νέφος από τυχαίες θέσεις τότε η ανακατασκευή της ευθείας φαίνεται μάλλον απίθανη. Μάλιστα οι επαναληπτικές μέθοδοι επίλυσης «προκαταβάλλονται» από τις αρχικές τιμές. Στο σχήμα 4-3 (αριστερά) βλέπουμε ένα τέτοιο περιστατικό όπου ο θόρυβος έχει παραμορφώσει τόσο πολύ τις αυθεντικές θέσεις των σημείων που είναι πλέον αδύνατη η ανακατασκευή της ευθείας. Είναι σαφές ότι σε κάθε μαθηματικό μοντέλο αναλόγως των συνθηκών υπάρχει κάποιο όριο στο θόρυβο που μπορεί να ανεχτεί.





Σχήμα 4-3: *Ο θόρυβος ενδέχεται να παραμορφώσει τα δείγματα σε τέτοιο βαθμό ώστε να είναι αδύνατη η ανακατασκευή της ευθείας ή να «χωράνε» πολλές ερμηνείες για την πιθανή της θέση (αριστερά). Οι συναρτήσεις κόστους με μεγάλη πρώτη παράγωγο αποδίδουν μεγαλύτερο κόστος στα outliers. Έτσι η λύση οδηγείται μοιραία στην «κατεύθυνση» που επιτάσσουν αυτά στην προσπάθεια να μειωθεί το μεγάλο τους κόστος (δεξιά).*

Στην μέθοδο που δείξαμε στην ενότητα 4.4.3 η συνάρτηση κόστους που χρησιμοποιήσαμε είναι η $f(x) = x^2$. Ένα χαρακτηριστικό της επίλυσης με ελάχιστα τετράγωνα είναι πως έστω και ένα «ακραία» λανθασμένο δείγμα μπορεί να πλήξει το αποτέλεσμα. Υπό μία έννοια η επίλυση με ελάχιστα τετράγωνα αντιστοιχεί στην εύρεση ενός μέσου όρου αφού όλες οι εξισώσεις που συμμετέχουν έχουν κοινή επίδραση στην λύση. Εντούτοις είναι γνωστό ότι ο μέσος όρος δεν διαχειρίζεται επιτυχημένα την ύπαρξη εσφαλμένων δεδομένων με μεγάλες αποκλίσεις. Μια λύση που έχει εφαρμογή σε διάφορα προβλήματα και προέρχεται από τον τομέα της στατιστικής είναι η χρήση μιας **robust μεθόδου στατιστικής επεξεργασίας**. Οι μέθοδοι robust είναι ανθεκτικές σε αποκλίσεις δειγμάτων που δεν εναρμονίζονται με το μαθηματικό μοντέλο του προβλήματος και διατηρούν την απόδοση τους παρουσιάζοντας ανοχή σε ένα ορισμένο ποσοστό σφαλμάτων. Οι τεχνικές αυτές στηρίζονται στην στάθμιση της επίδρασης των δειγμάτων δηλαδή εμμέσως στην μετατροπή της αυθεντικής συνάρτησης κόστους x^2 ώστε να διαχειρίζεται πιο «αδιάφορα» τα εσφαλμένα δεδομένα. Επειδή η επίλυση συστημάτων με άλλες συναρτήσεις (εκτός της x^2) δεν είναι συνηθισμένη και ίσως περίπλοκη, το όφελος με τις *robust* τεχνικές είναι ότι επιδρούν μέσω κατάλληλων σταθμών w_i ενώ συνεχίζουμε να επιλύουμε με ελάχιστα τετράγωνα.

Μια εναλλακτική συνάρτηση κόστους εκτός της x^2 είναι η $f(x) = |x|$. Η επίλυση με την *«ελάχιστη απόλυτη διαφορά»* αντιμετωπίζει καλύτερα τα εσφαλμένα δείγματα και παράγει εγκυρότερα αποτελέσματα. Ωστόσο η έλλειψη κυρτότητας στο μηδέν δημιουργεί δυσκολία στον χειρισμό των δειγμάτων με μηδενικό υπόλοιπο. Το πρόβλημα αυτό διορθώνεται με μια άλλη εξίσου αποτελεσματική έναντι στον θόρυβο συνάρτηση, την $f(x) = |x|^{1.2}$.

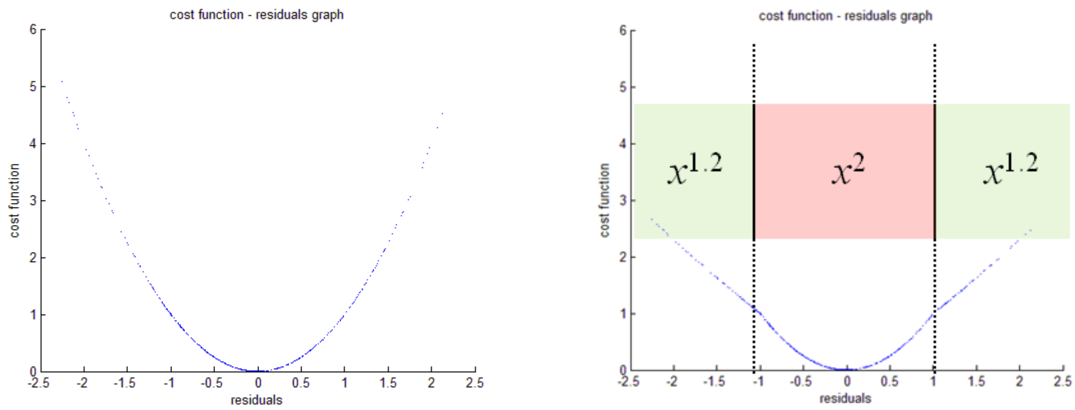
Για να ελαχιστοποιήσουμε τις (38), (45) χρησιμοποιώντας σαν συνάρτηση κόστους κάποια από τις $f(x) = |x|^q$ με $1<q<2$ δεν έχουμε παρά να πολλαπλασιάσουμε την x^2 με την κατάλληλη **συνάρτηση επίδρασης**. Για $f(x) = |x|^{1.2}$ παρουσιάζουμε την συνάρτηση επίδρασης:

$$w_i(x_i) = \frac{1}{|x_i|^{0.8}} \qquad \textbf{(47)}$$

Η συνάρτηση επίδρασης αποδίδει την «βαρύτητα»-συνεισφορά κάθε δείγματος στο τελικό αποτέλεσμα. Έτσι η «ανθεκτική» στον θόρυβο ΕΚΚ μπορεί τελικά να υπολογιστεί ελαχιστοποιώντας την:

$$\left(\Delta t, \omega\right) = \arg\min_{t,\omega,d} \frac{1}{m} \sum_{\{\mathbf{x}\}} w\left(r_{\Delta t}(\mathbf{x})\right)^2 \left|r_{\Delta t}(\mathbf{x})\right|^2 \qquad \textbf{(48)}$$

Μια ιδιομορφία της (47) είναι ότι δεν ορίζεται στο μηδέν. Γι' αυτό αλλά και επειδή είναι προτιμητέο να αντιμετωπίζουμε διαφορετικά τα μεγάλα **r**(**x**) σε σχέση με τα μικρότερα, η (47) θα πρέπει να οριστεί κατά τμήματα. Η επιλογή κάθε τμήματος γίνεται με βάση ένα κατώφλι *c* (*scale factor*) το οποίο επιλέγεται έτσι ώστε να υπάρχει η επιθυμητή αναλογία ανάμεσα στις μικρές και μεγάλες αποκλίσεις (σχήμα 4-4). Φυσικά είναι αυτονόητο ότι μια άστοχη επιλογή για την *c* ενδέχεται να οδηγήσει σε περαιτέρω σφάλματα και αποτυχία σύγκλισης, γι' αυτό απαιτείται ιδιαίτερη προσοχή για τον προσδιορισμό της κατάλληλης τιμής.



Σχήμα 4-4: *Τα υπόλοιπα σαν συνάρτηση του κόστους για την περίπτωση των ελαχίστων τετραγώνων L_2 (αριστερά) και της ελάχιστης απόλυτης δύναμης $L_{1.2}$ (δεξιά). Στην $L_{1.2}$ περίπτωση τα μεγάλα υπόλοιπα έχουν μικρότερη επίδραση - κόστος. Στο παράδειγμα τα υπόλοιπα έχουν διαβαθμιθεί γύρω από την μονάδα.*

Μια ιδιαιτέρως συνηθισμένη επιλογή για το *c* (όπως είδαμε και στην ενότητα 2.8) είναι η:

$$\begin{aligned} c_0 &= 1.4826 \\ c &= c_0 MAD(r_1 ... r_m) \end{aligned} \quad \textbf{(49)}$$

όπου c_0 σταθερά κανονικοποίησης και *MAD* ο *median* της απόλυτης απόκλισης από το *median* υπόλοιπο $median\left(\left|r_i - median\left(r_i\right)\right|\right)$. Δυστυχώς λόγω της *MAD* ο υπολογισμός της *c* απαιτεί σημαντικό υπολογιστικό κόστος.

Τελικά η συνάρτηση επίδρασης με $w_i : \Re \rightarrow (0,1]$ προκύπτει ως:

$$w_i(x) = \begin{cases} 1 & ,|x| \le c \\ \left|\dfrac{c}{x}\right|^{0.8} & ,|x| > c \end{cases} \quad \textbf{(50)}$$

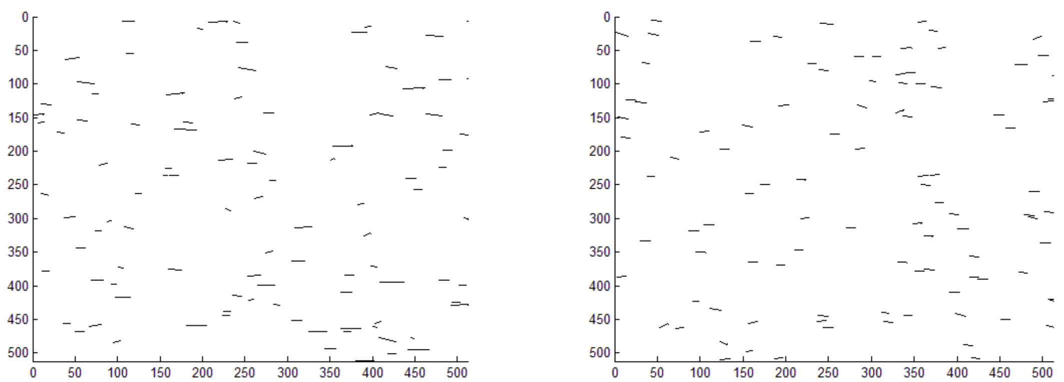
Για να υπολογίσουμε τις παραμέτρους *t,ω,d* ακολουθούμε τα βήματα που περιγράψαμε στην ενότητα 4.4.3 με μόνη διαφορά ότι χρησιμοποιούμε τις σταθμισμένες εκδόσεις των (45), (38) (όπως είναι η 48). Η διαδικασία ξεκινά με τις στάθμες w_i αρχικοποιημένες στην μονάδα. Στην αρχή κάθε κύκλου υπολογίζουμε τα νέα **r**(**x**), το κατώφλι *c* από την (49) και τις στάθμες *w*(**r**(**x**)) μέσω της (50).

Στις δοκιμές θα δούμε ότι ο σταθμισμένος Gauss-Newton συγκλίνει περίπου 5-15% πιο αργά και παράγει καλύτερα αποτελέσματα μόνο όταν ο θόρυβος είναι ανομοιόμορφα κατανεμημένος, κάτι που όμως ισχύει στην οπτική ροή που παράγεται από τους περισσότερους αλγόριθμους, όπως στον L-K.

4.5 Μείωση ελαχίστων συνάρτησης κόστους

Η μέθοδος που αναπτύξαμε επιλύει το πρόβλημα ΕΚΚ ελαχιστοποιώντας την $f(\mathbf{r}(\mathbf{x}))$ μέσω μιας επαναληπτικής διαδικασίας βελτιστοποίησης. Με άλλα λόγια θα λέγαμε πως «εξερευνούμε» την αντικειμενική συνάρτηση προκειμένου να ανακαλύψουμε το ολικό ελάχιστο στο οποίο βρίσκεται η λύση. Σημαντικές δυσκολίες στην διαδικασία αυτή είναι ότι ο θόρυβος «μετακινεί» το ολικό ελάχιστο από την πρέπουσα θέση και ότι η αντικειμενική συνάρτηση όντας περίπλοκη και εξαρτώμενη από πολλές μεταβλητές παρουσιάζει ιδιομορφίες και αρκετά τοπικά ελάχιστα.

Τα τοπικά αυτά ελάχιστα πολλές φορές αποδίδουν παρόμοιο κόστος με αποτέλεσμα να είναι αμφίβολο αν το *κριτήριο μικρότερου κόστους* θα μας οδηγήσει στην σωστή λύση. Αιτία για αυτή την δυσχέρεια είναι είτε ο θόρυβος, είτε η αμφισημία της κίνησης αφού ανόμοιες κινήσεις της κάμερας προκαλούν παρόμοιο πεδίο ροής. Π.χ. η πλαγιοπλευρική μεταφορά κατά τον *X* άξονα παράγει πανομοιότυπη οπτική ροή με την περιστροφή κατά τον *Y* άξονα (Σχήμα 4-5).



Σχήμα 4-5: *Προσομοίωση πεδίου αραιής οπτικής ροής (100 διανύσματα σε εικόνα 512x512 με οπτικό πεδίο 50^o και θόρυβο 0,7$_{pixels}$) που οφείλεται σε μεταφορική κίνηση X (αριστερά) και περιστροφική κατά Y άξονα (δεξιά).*

Η ύπαρξη του μεγάλου αριθμού ελαχίστων είναι πρόβλημα δύσκολο στην αντιμετώπιση του καθώς σχετίζεται με το σχήμα της αντικειμενικής συνάρτησης το οποίο με την σειρά του εξαρτάται από διάφορους παράγοντες όπως η διάταξη των διανυσμάτων οπτικής ροής στην εικόνα, η κίνηση της κάμερας αυτή καθ' εαυτή, καθώς και το οπτικό πεδίο.

Πιο συγκεκριμένα ο αριθμός των ελάχιστων αυξάνεται αναλόγως του πλήθους των διανυσμάτων που έχουν θέσεις $\boldsymbol{x}$ ανάλογες του t_k, με το φαινόμενο να εντείνεται όταν τα διανύσματα δεν είναι ομοιόμορφα κατανεμημένα στο οπτικό πλάνο αλλά αποσπασματικά κατακερματισμένα σε περιοχές της εικόνας. Φυσικά η κατάτμηση των διανυσμάτων σε περιοχές είναι συχνό φαινόμενο καθώς οι αλγόριθμοι οπτικής ροής επιτυγχάνουν εκτιμήσεις σε σημεία που έχουν κάποια ιδιαίτερα χαρακτηριστικά, σημεία με υφή ή ακμές, με αποτέλεσμα τα διανύσματα συχνά να ακολουθούν την κατάτμηση της εικόνας.

Τελικά λόγω του πλήθους των ελαχίστων η διαδικασία είναι πολύ πιθανό να εγκλωβιστεί σε κάποιο από αυτά. Αντίθετα δεν συμβαίνει το ίδιο με την σταθμισμένη (κατά $\|A(\mathbf{x})t\|$) έκδοση της (38), την (36) η οποία έχει λιγότερα ελάχιστα, τυπικά δυο, που έχουν σχέση μόνο με το σχήμα της που είναι απλό και κοίλο. Βέβαια όπως αναλύσαμε πριν η (36) οδηγεί σε λάθος εκτιμήσεις και δεν πρέπει να χρησιμοποιείτε.

Σχετικά με τα ελάχιστα θα πρέπει να γνωρίζουμε ότι ανάμεσα τους υπάρχουν πάντα δύο που θα τα αποκαλούμε ως *«ισχυρά»* και οφείλουν την ύπαρξη τους στο σχήμα της αντικειμενικής συνάρτησης. Απουσία θορύβου τα ελάχιστα αυτά αποδίδουν αισθητά μικρότερο κόστος σε σχέση με τα υπόλοιπα και ένα από τα δυο είναι το ολικό δηλαδή η λύση.

Έτσι ένας τρόπος [3] για να μετριαστεί ο αριθμός των ανεπιθύμητων ελαχίστων θα μπορούσε να προέλθει από τον συνδυασμό των ιδιοτήτων της (36) και της εγκυρότητας της (38). Εκμεταλλευόμενοι δε το γεγονός ότι οι δύο σχέσεις διαφέρουν μόνο κατά ένα πολλαπλασιαστή μπορούμε να δημιουργήσουμε μια συνάρτηση κόστους της οποίας η περιπλοκότητα (από άποψη σχήματος) θα μπορεί να ρυθμίζεται κατά περίπτωση, ώστε στην αρχή της βελτιστοποίησης να είναι απλή και κοίλη για να ξεπερνιέται ο κίνδυνος εγκλωβισμού και στην συνέχεια των επαναλήψεων καθώς «μετακινούμαστε» προς το κοίλο μέρος η περιπλοκότητα να αυξάνει διαδοχικά ώστε να φανερώνονται τα ισχυρά ελάχιστα.

Ισοδύναμα μπορούμε να ξεκινήσουμε την επίλυση με την (36) και σταδιακά να κάνουμε εναλλαγή στην (38). Αν υποθέσουμε *ρ* την παράμετρο ελέγχου αυτής της μετάβασης όπου στο μηδέν γίνεται ελαχιστοποίηση της (36) και στο ένα της (38) τότε μπορούμε να ορίσουμε την προοδευτική μετάβαση από το μηδέν στο ένα συναρτήσει των επαναλήψεων *k* και του μεγέθους της ανανέωσης $\|\Delta t\|$ ως: *(βασισμένο στην μέθοδο Pauwels-Hulle [3])*

$$\rho_k = \min\left(1, \rho_{k-1} + \lambda \max\left(0, \left[\frac{\log_{10}\|\Delta t_{k-1}\|}{\log_{10} e}\right]\right)\right) \qquad \textbf{(51)}$$

Όπου *e* η τιμή σύγκλισης (10^{-13}) και *λ* (= 1/4) μια σταθερά που καθορίζει την ταχύτητα προσαρμογής, δηλαδή το ρυθμό αύξησης του *ρ*. Η αρχική τιμή της *ρ* είναι μηδέν και διαδοχικά αυξάνει προς το ένα όπου είναι και η μέγιστη τιμή που μπορεί να λάβει. Είναι σημαντικό η μετάβαση από την (36) στην (38) να γίνεται σταδιακά και ομαλά κατά την διάρκεια της βελτιστοποίησης. Η (51) εγγυάται την σταδιακή προσαρμογή λαμβάνοντας υπόψη και το μέγεθος της ανανέωσης $\|\Delta t\|$. Όταν το $\|\Delta t\|$ λαμβάνει μεγάλες τιμές σημαίνει ότι απέχουμε από την σύγκλιση και τότε το *ρ* δεν θα πρέπει να αυξάνεται με μεγάλο ρυθμό. Αντίθετα όταν το $\|\Delta t\|$ λαμβάνει μικρές τιμές σημαίνει ότι πιθανότατα η διαδικασία οδεύει προς την σύγκλιση, οπότε και το *ρ* πρέπει να αυξάνεται με μεγαλύτερη ταχύτητα ώστε να φτάσει σύντομα την μονάδα και οι τελευταίες επαναλήψεις του Gauss-Newton να γίνουν με χρήση της (38) για να επιτύχουμε ακριβή αποτελέσματα. Όταν το *ρ* φτάσει στην μονάδα παραμένει εκεί μέχρι να επέλθει η σύγκλιση. Από την στιγμή που η (51) μπορεί μόνο να αυξάνεται και είναι άνω φραγμένη η σύγκλιση είναι εγγυημένη.

Συνδυάζοντας την συνάρτηση μετάβασης με την (38) λαμβάνουμε την νέα συνάρτηση:

$$\omega_k = \arg\min_{\omega} \frac{1}{m} \sum_{\{\mathbf{x}\}} \left[\tau\left(\mathbf{x}, t_k, \rho_k\right)^T \left(u(\mathbf{x}) - B(\mathbf{x})\omega\right) \right]^2 \qquad \textbf{(52)}$$

όπου *τ(x,t,ρ)*:

$$\tau(\mathbf{x}, t, \rho) = \frac{\left(\left[A(\mathbf{x})t\right]_2, -\left[A(\mathbf{x})t\right]_1\right)^T}{\|A(\mathbf{x})t\|^{\rho}} \qquad \textbf{(53)}$$

Αντίστοιχα και για την (48) λαμβάνουμε:

$$\left(\Delta t_k, \omega_k\right) = \arg\min_{\Delta t, \omega} \sum_{\{\mathbf{x}\}} \left[\tau\left(\mathbf{x}, t_k, \rho_k\right)^T \left(u(\mathbf{x}) - d_k A(\mathbf{x}) \Delta t_k - B(\mathbf{x}) \omega_k\right) \right]^2 \qquad \textbf{(54)}$$

Ενώ ο υπολογιστικός φόρτος αυτής της εκδοχής διαφέρει αμελητέα σε σχέση με αυτές των ενοτήτων 4.4.3-4 συνολικά η διαδικασία της μετάβασης επιφέρει κάποια επιβάρυνση από την άποψη ότι καθυστερείται η σύγκλιση. Όπως θα διαπιστώσουμε στις δόκιμες τα αποτελέσματα που παράγονται με αυτή τη μέθοδο είναι το ίδιο ακριβή (όχι καλύτερα!), τα ελάχιστα μειώνονται κατά μέσο όρο σε τέσσερα και ο αριθμός των επαναλήψεων αυξάνεται περίπου κατά 15 κάτι πιθανώς ανεκτό από άποψη χρόνου εκτέλεσης.

4.6 Αλγόριθμος

Σε αυτή την ενότητα θα περιγράψουμε τα βήματα του αλγορίθμου για τις εκδοχές που περιγράψαμε στις ενότητες 4.4.3-4, 4.5.

Ενδεικτικές τιμές / *Παράμετροι*	**Τιμή**
Μέγιστος αριθμός επαναλήψεων (ανά αρχικοποίηση)	300
Αριθμός αρχικοποιήσεων	15
Οπτικό πεδίο	50°
Αριθμός διανυσμάτων	100
Μέσος αριθμός επαναλήψεων	36
Μέσος χρόνος απόκρισης	2,4 sec

Πίνακας 4-1: *Τυπικές τιμές παραμετροποίησης και απόκρισης του αλγορίθμου εκτίμησης της κίνησης.*

Να σημειώσουμε ότι πριν ξεκινήσει η διαδικασία θα πρέπει να γίνει διαβάθμιση όλων των συντεταγμένων και διανυσμάτων σε μονάδες *FLU* ώστε να συνυπολογιστεί το *οπτικό πεδίο* και να αποκτήσει νόημα η σύμβαση ότι η *εστιακή απόσταση* είναι μονάδα. Ακόμα θα πρέπει να μετατραπούν οι συντεταγμένες ώστε το κέντρο να συμπίπτει με το *σημείο επέκτασης*. Γενικότερα πρέπει να είμαστε συμβατοί με το μοντέλο οπτικών εξισώσεων που περιγράψαμε στις ενότητες 3.3 και 3.5 ώστε να ισχύει η (26) που είναι η εξίσωση πάνω στην οποία βασιστήκαμε για τους μετασχηματισμούς αυτού του κεφαλαίου. Ο πίνακας 4-1 περιέχει τυπικές τιμές παραμετροποίησης του αλγορίθμου.

- **Κώδικας σταθμισμένου διαχωρισμένου Gauss-Newton**

→ **Είσοδος:** πεδίο οπτικής ροής *u*, οπτικό πεδίο *FOV*, αριθμός αρχικοποιήσεων *max_trials*, μέγιστος αριθμός επαναλήψεων *max_iterations*.
Έξοδος: *t*, *ω*, *d(**x**)*

→ **Αρχικοποίηση** βασικών μεταβλητών όπως FOV, max_iterations, max_trials, converge_threshold

→ **Μετατροπή** συντεταγμένων και διαβάθμιση διανυσμάτων βάσει του FOV και FOE

→ **Μέχρις ότου** το trials γίνει max_trials **κάνε**

→ **Τυχαία Αρχικοποίηση** t_0

→ **Αρχικοποίηση** $\rho_0 = 0$ και **υπολογισμός** των e_0, τ_0 από (33), (32): $e(x,t,1) = \dfrac{A(\mathbf{x})t}{\|A(\mathbf{x})t\|}$

ή (53)

→ **υπολογισμός** του ω_0 από (38): $\hat{\omega} = \arg\min_{\omega} \dfrac{1}{m} \sum_{\{\mathbf{x}\}} \left[\tau(\mathbf{x},t,1)^T (u(\mathbf{x}) - B(\mathbf{x})\omega) \right]^2$

ή (52)

→ **Υπολογισμούς** του d_0 από (35)

→ **Αρχικοποίηση** Σταθμών w_i σε μονάδα

→ **Μέχρις ότου** το iterations γίνει max_ iterations **ή** το || Δt_k || < converge_threshold **κάνε**

→ **υπολογισμός** του Δt_k από (48):
$(\Delta t_k, \omega_k) = \arg\min_{\Delta t,\omega} \sum_{\{\mathbf{x}\}} w(r_{\Delta t}(\mathbf{x})) \left[\tau(\mathbf{x},t,1)^T (u(\mathbf{x}) - d_k A(\mathbf{x})\Delta t - B(\mathbf{x})\omega) \right]^2$ ή

(45) ή (54)

→ **Ανανέωση** $t_k = t_{k-1} + \Delta t_k$, $\omega_k = \omega_{k\_new}$

→ **υπολογισμός** του ρ_k από (51) και e_k, τ_k με βάση το νέο t_k από (32), (33):
$\tau(\mathbf{x},t,1) = \dfrac{\left([A(\mathbf{x})t]_2, -[A(\mathbf{x})t]_1\right)^T}{\|A(\mathbf{x})t\|}$ ή (53)

→ **Επανεκτίμηση** του ω_k από (48):
$\hat{\omega} = \arg\min_{\omega} \dfrac{1}{m} \sum_{\{\mathbf{x}\}} w(r_{\omega}(\mathbf{x})) \left[\tau(\mathbf{x},t,1)^T (u(\mathbf{x}) - B(\mathbf{x})\omega) \right]^2$ ή (38) ή (52)

→ **υπολογισμός** του d_k από (35): $d(\mathbf{x}) = \dfrac{e^T(\mathbf{u}_*(\mathbf{x}) - B(\mathbf{x})\omega)}{\|A(\mathbf{x})t\|}$

→ **υπολογισμός** επιμέρους κόστους r από (46):
$\mathbf{r}(\mathbf{x}) = \mathbf{u}_*(\mathbf{x}) - d(\mathbf{x})A(\mathbf{x})t - B(\mathbf{x})\omega$

και ολικού κόστους από (42):

$$\text{residual\_cost} = \frac{1}{m}\sum_{\{\mathbf{x}\}}\|\mathbf{r}(\mathbf{x})\|_2^2$$ ή (48)

→ **υπολογισμός** των Σταθμών w_i από (50): $w_i(x) = \begin{cases} 1 & ,|x| \le c \\ \left|\frac{c}{x}\right|^{0.8} & ,|x| > c \end{cases}$

→ **Τέλος επανάληψης** iterations

→ **Αποθήκευση** των t_{k_trials}, ω_{k_trials}, d_{k_trials}, $\text{residual\_cost}_{trials}$

→ **Τέλος επανάληψης** trials

→ **Επιλογή** τελικής λύσης t, ω, d αυτής με το μικρότερο κόστος residual_cost

→ **Κανονικοποίηση** t και **διόρθωση** αναλογίας d

→ **Διόρθωση** φοράς t ώστε να εξασφαλίζεται θετικό βάθος

→ **Επιστροφή** t, ω, d

= Σταθμισμένος Gauss-Newton
= Διαχωρισμένος Gauss-Newton
= Διαχωρισμένος Gauss-Newton με διαχείριση ελαχίστων

Αυτή είναι η βασική διάρθρωση του αλγόριθμου με κάποιες προγραμματιστικές λεπτομέρειες να αποκρύπτονται για οικονομία χώρου. Ο κώδικας παρουσιάζει την εκδοχή του σταθμισμένου Gauss-Newton της ενότητας 4.4.4. Ωστόσο η δομή παραμένει ίδια και για τις εκδοχές των ενοτήτων 4.4.3 και 4.5 οι οποίες προκύπτουν με αντικατάσταση των σχέσεων που εμφανίζονται στον αλγόριθμο με αυτές των αντίστοιχων χρωμάτων. Ο αλγόριθμος για απευθείας επίλυση της (43) με Gauss-Newton (ενότητα 4.4.2) έχει διαφορετική δομή και δεν παρουσιάζεται εδώ αφού δεν χρησιμοποιείται στην πράξη.

Κεφάλαιο 5

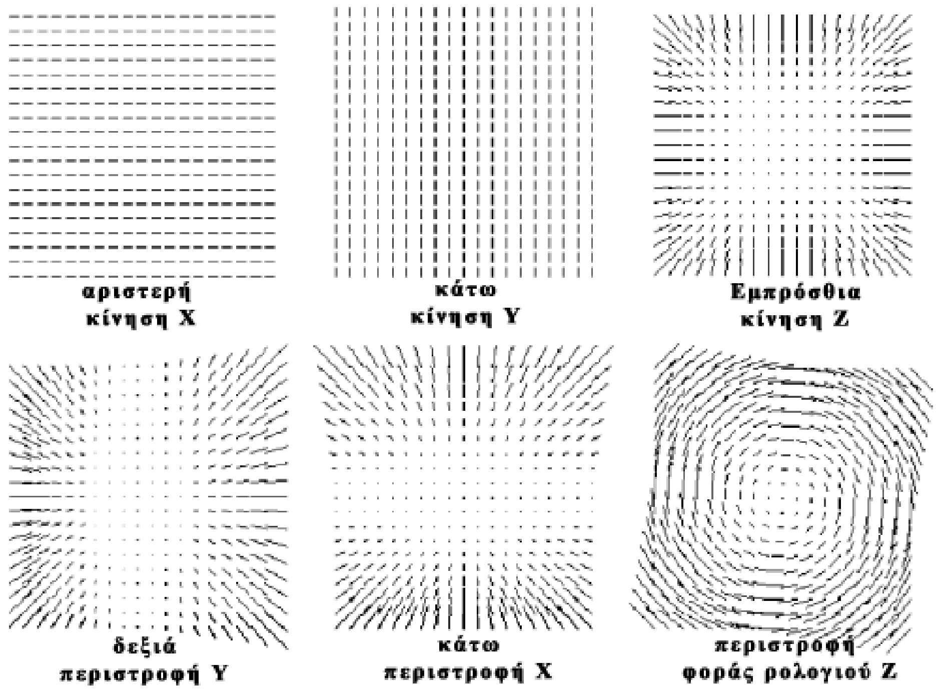
Πειραματικά αποτελέσματα

5.1 Εισαγωγή

Στο κεφάλαιο αυτό θα εξετάσουμε την απόδοση των αλγορίθμων τόσο σε επίπεδο προσομοίωσης με συνθετική οπτική ροή όσο και με πραγματικές ακολουθίες εικόνων. Η προσομοίωση με συνθετική οπτική ροή είναι επιβεβλημένη για προφανείς λόγους όπως το ελεγχόμενο περιβάλλον, η επαναληψιμότητα των μετρήσεων αλλά κυρίως η έλλειψη αντικειμενικών μετρήσεων για την κίνηση της κάμερας σε πραγματικές ακολουθίες. Ένα ακόμα προσόν των συνθετικών δεδομένων είναι ότι μπορούν να προσομοιώσουν τις οπτικές παραμέτρους κάτι που ως επί το πλείστον εξαρτάται από το υλικό. Π.χ. για να πραγματοποιήσουμε μετρήσεις με κυμαινόμενες τιμές οπτικού πεδίου θα χρειαζόμασταν διαφορετικού τύπου φακούς ή κάμερες όσες θα ήταν και οι διάφορες προς έλεγχο τιμές. Πρακτικά κάτι τέτοιο θα ήταν περιοριστικό και ασύμφορο, αν όχι αδύνατο, κάτι που ξεπερνάται όμως στην περίπτωση της προσομοίωσης με συνθετική οπτική ροή.

Ωστόσο και οι δοκιμές με πραγματικές ακολουθίες εικόνων προσφέρουν μια χονδρική αλλά αξιόλογη ένδειξη της απόδοσης και της ακρίβειας των αποτελεσμάτων. Άλλωστε αυτό που ενδιαφέρει στην πράξη είναι η απόδοση σε πραγματικές ακολουθίες υπό ρεαλιστικές συνθήκες. Έτσι θα δούμε αν κάποιες από τις θεωρητικές παρατηρήσεις που υποστηρίξαμε τελικά επιβεβαιώνονται καθώς και κατά πόσο θα μπορούσε τελικά να γίνει χρήση της μεθόδου σε πρακτικές εφαρμογές.

Ακόμα θα μιλήσουμε για την διασύνδεση των διάφορων στοιχείων που αναπτύξαμε στα προηγούμενα κεφάλαια και την ορθή παραμετροποίηση τους. Ιδιαίτερη σημασία δόθηκε στην μέτρηση εκείνων των χαρακτηριστικών που είναι συνυφασμένα με την απόδοση των αλγορίθμων ΕΚΚ όπως είναι η *πόλωση* και οι *αποκλίσεις*. Με αυτές τις αρχές θα διερευνήσουμε την απόδοση και των τριών εκδοχών του αλγόριθμου που περιγράψαμε βλέποντας τα πλεονεκτήματά τους και ποτέ ενδείκνυται η χρήση καθενός από αυτούς.

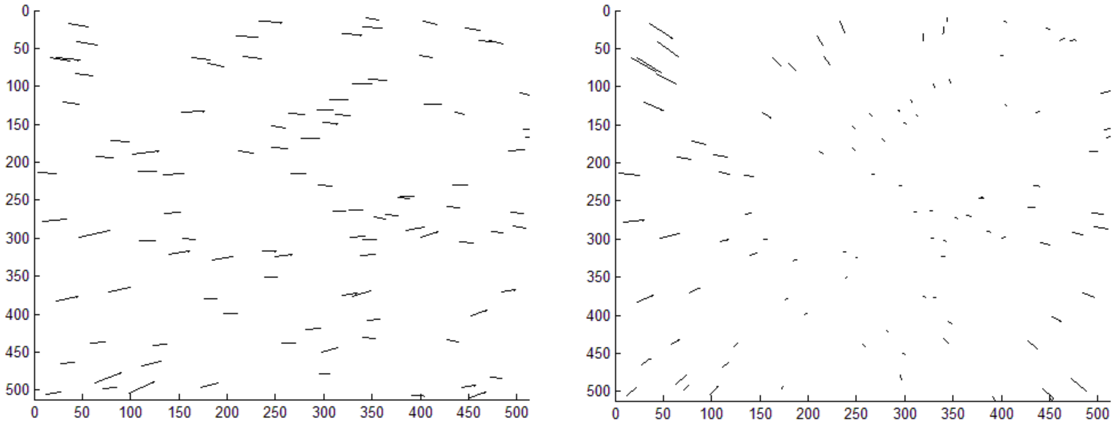


Σχήμα 5-1: *Πυκνή οπτική ροή που παράγεται από τις βασικές κινήσεις της κάμερας [9].*

5.2 Σύνθεση οπτικής ροής

5.2.1 Εξίσωση σύνθεσης και Οπτικό μοντέλο

Η οπτική ροή είναι αλληλένδετη και πλήρως συνυφασμένη με το οπτικό μοντέλο που υιοθετείται. Ήδη στο κεφάλαιο 3 περιγράψαμε τις εξισώσεις που συνδέουν την κίνηση της κάμερας με την προβολή διανυσμάτων κίνησης στο οπτικό πλάνο (σχήμα 5-1). Το ζήτημα που εγείρεται είναι π.χ. η σχέση (25) μπορεί να χρησιμοποιηθεί για να παράγουμε τα (συνθετικά) δεδομένα που χρειαζόμαστε. Η απάντηση είναι πως μπορούμε και ο λόγος είναι πως η (28) που χρησιμοποιήσαμε σαν βασική εξίσωση στην ΕΚΚ προέρχεται από την (25). Αυτό έχει την βαρύτητα του καθώς κατοχυρώνουμε ότι μια κίνηση της κάμερας αντιστοιχεί σε συγκεκριμένο πεδίο οπτικής ροής και ταυτόχρονα ότι αυτό το πεδίο ροής είναι συμβατό με τις κινήσεις που «αντιλαμβάνεται» ο αλγόριθμος ΕΚΚ. Πράγματι αυτό επιβεβαιώθηκε πειραματικά καθώς συνθέτοντας οπτική ροή με την (25) και διοχετεύοντας τα διανύσματα που παράχθηκαν στον αλγόριθμο ΕΚΚ αυτός υπολόγιζε πάντα σωστά τις παραμέτρους που την προκαλούσαν.



Σχήμα 5-2: *Προσομοίωση οπτικής ροής για κίνηση της κάμερας με μεταφορική κατεύθυνση (4 -3 5)^T και περιστροφική κατεύθυνση (-1 2 0.5)^T για οπτικό πεδίο 50° (αριστερά) και 150° (δεξιά).*

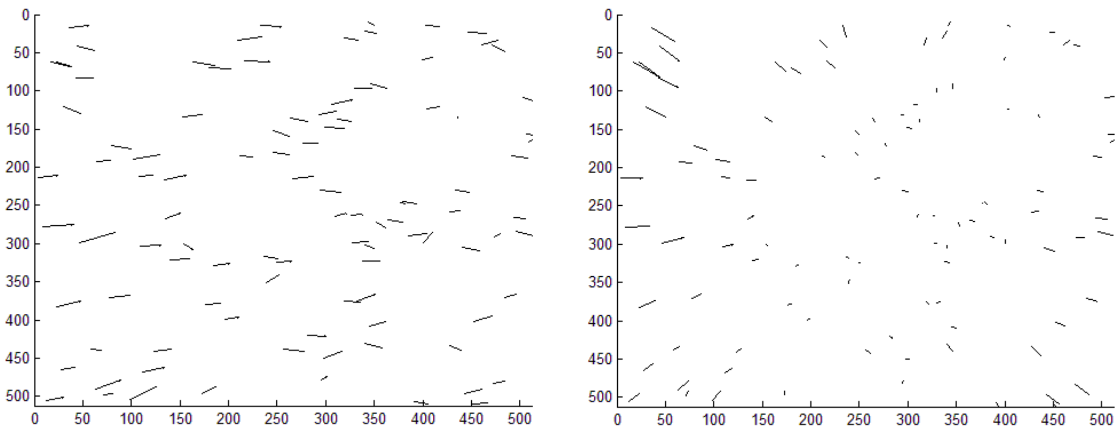
Έτσι με χρήση των *t, ω, Z,* την εστιακή απόσταση *f* και το οπτικό πεδίο *FOV* μπορούμε να προσομοιώσουμε οποιαδήποτε κίνηση της κάμερας (σχήμα 5-2). Για λόγους συνάφειας αλλά και την αποφυγή της χρήσης παραμέτρων όπως η ανάλυση των εικόνων (*resolution*) ή το μέγεθος του CCD (*mm*) όλες οι μονάδες των παραπάνω μετατράπηκαν σε *FLU*. Μετά την κατασκευή του πεδίου ροής το οπτικό πλάνο προβλήθηκε σε εικόνες ανάλυσης 512x512*p* και προωθήθηκε για περαιτέρω επεξεργασία. Στις ενότητες που ακολουθούν θέτουμε το ζήτημα του θορύβου και της ταχύτητας κίνησης.

5.2.2 Μοντέλο θορύβου

Το να αναγνωρίζονται κινήσεις που συμπίπτουν ακριβώς με το πεδίο ροής που τις παρήγαγε μικρή χρησιμότητα έχει από την στιγμή που κάτι τέτοιο δεν συναντάται στην πράξη. Αντιθέτως αλγόριθμοι όπως ο L-K συνήθως παράγουν «ακραίο» θόρυβο σε ένα ποσοστό εκ του συνόλου των διανυσμάτων. Το μοντέλο θορύβου του L-K καθώς και γενικότερα αυτών των αλγορίθμων είναι περίπλοκο, ωστόσο αν εξαιρέσουμε τις ακραίες αποκλίσεις κάποιων μεμονωμένων διανυσμάτων γενικότερα θα μπορούσαμε να προσομοιώσουμε τα λάθη στις εκτιμήσεις προσθέτοντας *ανεξάρτητο και μηδενικής μέσης τιμής* Gaussian θόρυβο ξεχωριστά σε καθένα από τα στοιχεία των δισδιάστατων διανυσμάτων (σχήμα 5-3). Παρά την παραδοχή αυτή θα πρέπει να επισημάνουμε πως το μοντέλο θορύβου των περισσότερων αλγορίθμων εύρεσης οπτικής ροής όπως και του L-K δεν είναι Gaussian.

Ένα τμήμα του θορύβου των αλγορίθμων που βασίζονται στις χωροχρονικές παραγώγους προκαλείται λόγω κάποιων ιδιαίτερων χαρακτηριστικών που φέρουν διάφορες περιοχές των εικόνων (ενότητες 2.5, 2.6). Τα διανύσματα δε που βρίσκονται σε τέτοιες περιοχές ναι μεν έχουν λανθασμένες κατευθύνσεις αλλά συνήθως μεταξύ τους είναι συνευθειακά. Αυτό φυσικά κάνει πιο δύσκολη την εξομάλυνση των λαθών σε σχέση με την περίπτωση όπου όλα τα διανύσματα του πεδίου έχουν ανεξάρτητο και μηδενικής μέσης τιμής Gaussian θόρυβο.

Γενικά το μοντέλο θορύβου του L-K είναι μάλλον δύσκολο να προσομοιωθεί επακριβώς, πράγμα που άλλωστε ξεφεύγει από τα πλαίσια αυτής της εργασίας. Ωστόσο η εφαρμογή Gaussian θορύβου με μεγάλες τυπικές αποκλίσεις και άρα μικρές τιμές **λόγου σήματος προς θόρυβο** (*signal to noise ratio, SNR)* θεωρούμε ότι δίνει μια καλή ένδειξη για την απόκριση του αλγορίθμου σε πραγματικές συνθήκες. Ο τρόπος με τον οποίο θα προσθέσουμε τον θόρυβο και το μέγεθος του είναι μείζον ζήτημα από την στιγμή που η μεγαλύτερη αιτία σφαλμάτων στην ΕΚΚ είναι ο θόρυβος στην οπτική ροή.



Σχήμα 5-3: *Οπτική ροή για μεταφορά* $(4\ -3\ 5)^T$ *, περιστροφή* $(-1\ 2\ 0.5)^T$ *και θόρυβο* $\sigma_n = 0{,}5_{pixels}$ *ανά διάσταση. Για FOV = 50° ο λόγος σήματος προς θόρυβο είναι 5.59 (δεξιά) ενώ για FOV = 150° η τιμή του γίνεται περίπου 3.8 (αριστερά). Τα μεγαλύτερα οπτικά πεδία είναι πιο ευάλωτα στο θόρυβο.*

Υπενθυμίζουμε ότι ως θόρυβο ορίζουμε κάθε απόκλιση της κατεύθυνσης των διανυσμάτων από την πρέπουσα θέση τους. Αυτή η απόκλιση μπορεί να προκύψει προσθέτοντας Gaussian θόρυβο σ_n σε κάθε ένα από τα στοιχεία των διανυσμάτων ξεχωριστά. Βέβαια προσθέτοντας Gaussian θόρυβο σ_n σε κάθε στοιχείο των διανυσμάτων το μέτρο του θορύβου $\|\sigma_n\|$ είναι μεγαλύτερο από σ_n. Για να ξεπεράσουμε την ασάφεια του πόσος είναι ο θόρυβος που επιδρά θεωρούμε ως κριτήριο το *SNR* το οποίο ορίζουμε ως την αναλογία του *μέσου μέτρου των διανυσμάτων* προς το *μέσο μέτρο του θορύβου*: $\sqrt{E\left\{\|u_i\|_2^2\right\} : E\left\{\|n_i\|_2^2\right\}}$. Το επιθυμητό *SNR* μπορεί να προκύψει απλά αν ο θόρυβος που προσθέτουμε σε κάθε διάνυσμα είναι ανάλογος του μεγέθους του. Το μέσο μέτρο των διανυσμάτων στις δοκιμές που πραγματοποιήσαμε κυμαίνεται από 2.2 έως 3.5 εικονοστοιχεία οδηγώντας αντίστοιχα σε *SNR* από 3 έως 6.6 για σ_n = 0.5 εικονοστοιχεία.

Ένα απτό παράδειγμα για να κατανοήσουμε την χρησιμότητα της διαβάθμισης του θορύβου με βάση το *SNR* είναι το εξής: όταν προσθέτουμε μία σταθερή ποσότητα θορύβου τα μικρά διανύσματα επηρεάζονται περισσότερο από τα μεγαλύτερα. Έτσι η οπτική ροή που προέρχεται από μικρές κινήσεις της κάμερας (και παράγει μικρότερα διανύσματα) θα καταλήγει σχετικά πιο θορυβώδης, κάτι όχι ιδιαίτερα αντιπροσωπευτικό για τα ισχύοντα στην πράξη. Το *SNR* λειτουργεί ως κριτήριο ώστε να ρυθμίζουμε τον θόρυβο που επιθυμούμε να επιδρά ανά περίσταση.

5.2.3 Ταχύτητα κίνησης

Τα διανύσματα που θα παραχθούν με την σύνθεση θα πρέπει όσο το δυνατόν να τηρούν κάποιες αναλογίες σε σχέση με τα μεγέθη που συναντώνται στην πραγματική οπτική ροή. Π.χ. τυπικές τιμές για την περιστροφή κυμαίνονται από 0.1 ως 0.4° ανά καρέ. Σε μια δεξιόστροφη κατά τον Y-άξονα κίνηση (και με οπτικό πεδίο 50°) παράγονται διανύσματα από 0.0017 έως 0.0069 *FLU*. Αν υποθέσουμε ότι η ανάλυση των εικόνων είναι 512x512*p* τελικά αυτό αντιστοιχεί σε διανύσματα με μέγεθος 0.81 έως 3.28*p*.

Φυσικά αντίστοιχα εύλογες τιμές θα πρέπει να έχουν τα διανύσματα που οφείλονται σε μεταφορικές κινήσεις. Ωστόσο τα πράγματα στις μεταφορές είναι πιο περίπλοκα αφού το μέγεθος των διανυσμάτων δεν εξαρτάται μόνο από την ταχύτητα της κάμερας αλλά και από την απόσταση της από τα αντικείμενα. Έτσι θα πρέπει να φροντίσουμε ώστε ο λόγος ταχύτητας - απόστασης να δίνει διανύσματα με λογικό μέτρο. Αν υποθέσουμε ότι οι αποστάσεις των αντικειμένων λαμβάνουν τυχαίες τιμές (εντός κάποιου εύρους) τότε θα πρέπει να επινοήσουμε μια ταχύτητα αναφοράς για τα διανύσματα που προκαλούνται από μεταφορικές κινήσεις.

Για να το πετύχουμε αυτό μια ιδέα είναι ότι θα μπορούσαμε να «δεσμεύσουμε» την ταχύτητα μεταφοράς ως προς την ταχύτητα περιστροφής με τρόπο ώστε τα διανύσματα που βρίσκονται στο *κέντρο της εικόνας* και προκαλούνται από αντικείμενα που βρίσκονται στην *μέση απόσταση από την κάμερα* να έχουν το ίδιο μέγεθος με τα διανύσματα που οφείλονται σε περιστροφές.

Ένα αριθμητικό παράδειγμα για να γίνει κατανοητό αυτό είναι πως αν έχουμε περιστροφή με γωνιακή ταχύτητα 0.37° ανά καρέ και οι αποστάσεις *Z* της κάμερας από τα διάφορα αντικείμενα κυμαίνονται από 2 έως 8 *FLU* τότε η ταχύτητα μεταφοράς θα προκύψει ώστε στην μέση απόσταση Z_{mid} = 5 το μέτρο των διανυσμάτων να είναι 0.37° ή 0.0065*rad* (ή 3,61*p* για ανάλυση 512x512*p*). Παρατηρώντας την (28) για να διατηρηθεί αυτή η αντιστοιχία θα πρέπει η ταχύτητα μεταφοράς να είναι $t_{speed} = Z_{mid} \cdot R_{speed}$ = 0.0325. Έτσι η σύνθεση της οπτικής ροής θα πραγματοποιείται επιλέγοντας μεταξύ άλλων και την επιθυμητή *γωνιακή ταχύτητα* ενώ η *μεταφορική ταχύτητα* θα προκύπτει από την παραπάνω σχέση.

5.3 Δοκιμές με συνθετική οπτική ροή

Αφού παρουσιάσαμε τον τρόπο δημιουργίας της συνθετικής οπτικής ροής σε αυτή την ενότητα θα δούμε μια σειρά από δοκιμές προκειμένου να εξετάσουμε την συμπεριφορά του αλγορίθμου σε διάφορα ζητήματα όπως το πως επηρεάζεται από τον θόρυβο, το οπτικό πεδίο κ.α. Σε καθένα από τα σενάρια που εξετάζουμε ιδιαίτερη βάση δίνουμε σε δυο κρίσιμες παραμέτρους που αποτελούν μέτρο απόδοσης για τους αλγόριθμους ΕΚΚ: την *πόλωση* και την *απόκλιση* από τις πραγματικές τιμές. Η πόλωση θα μπορούσαμε να πούμε πως είναι η «τάση» του αλγορίθμου να δίνει αποτελέσματα προς συγκεκριμένες κατευθύνσεις και η απόκλιση η ευαισθησία των μετρήσεων στο θόρυβο.

5.3.1 Μεθοδολογία μέτρησης σφαλμάτων

Αν υποθέσουμε ως σφάλμα μεταφοράς την γωνία που σχηματίζουν μεταξύ τους το εκτιμώμενο διάνυσμα t_{est} με το διάνυσμα της πραγματικής μεταφοράς t_{true} έχουμε:

$$t_{error} = \cos^{-1}\left(\frac{t_{est}^{T} \cdot t_{true}}{\|t_{est}\|\|t_{true}\|}\right) \quad \textbf{(55)}$$

Τότε η πόλωση και απόκλιση ισοδυναμεί αντίστοιχα με την *μέση τιμή* και *στάνταρτ απόκλιση* του t_{error}:

$$t_{error\_mean} = \frac{\sum_{m=1}^{M} \cos^{-1}\left(\frac{t_{est}^{T} \cdot t_{true}}{\|t_{est}\|\|t_{true}\|}\right)}{M}, \quad t_{error\_std} = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left(t_{est} - t_{error\_mean}\right)} \quad \textbf{(56)}$$

Ενώ αν σφάλμα περιστροφής θεωρήσουμε την νόρμα της διαφοράς του εκτιμώμενου διανύσματος ω_{est} σε σχέση με την πραγματική περιστροφή ω_{true} έχουμε:

$$\omega_{error} = \|\omega_{est} - \omega_{true}\|_2 \quad \textbf{(57)}$$

Τότε η πόλωση και απόκλιση περιστροφής υπολογίζεται ως:

$$\omega_{error\_mean} = \frac{\sum_{m=1}^{M} \|\omega_{est} - \omega_{true}\|_2}{M}, \quad \omega_{error\_std} = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left(\omega_{est} - \omega_{error\_mean}\right)} \quad \textbf{(58)}$$

5.3.2 Δοκιμές – Αποτελέσματα

Στις δοκιμές συμπεριλάβαμε τις τρεις εκδοχές του αλγορίθμου των ενοτήτων 4.4.3, 4.4.4, 4.5 ώστε να διερευνήσουμε τις διαφορές τους και τα πλεονεκτήματα του καθενός.

· *Σύνθεση δεδομένων*

Το πεδίο οπτικής ροής που συνθέσαμε αποτελείτε από 100 διανύσματα τα οποία κατανείμαμε ομοιόμορφα στην περιοχή του οπτικού πλάνου. Η αναλογία διαστάσεων μήκους-πλάτους του οπτικού πλάνου και η εστιακή απόσταση για όλες τις προσομοιώσεις τέθηκε στην μονάδα. Το οπτικό πεδίο έλαβε τιμές 50° και 150° το οποίο απέδωσε διαστάσεις οπτικού πλάνου 0.93x0.93 και 4.1x4.1 *FLU* αντίστοιχα. Σκοπός μας είναι να δείξουμε τι συμβαίνει στις ακραίες περιπτώσεις όπου το οπτικό πλάνο είναι μικρό και αρκετά (ως υπερβολικά!) μεγάλο. Στην συνέχεια το οπτικό πλάνο προβλήθηκε σε εικόνες ανάλυσης 512x512*p*. Τα διανύσματα προκλήθηκαν από κίνηση αντικειμένων σε τυχαία απόσταση από την κάμερα εντός του εύρους 1 έως 4 *FLU*. Στα στοιχεία καθενός από τα 100 διανύσματα προσθέσαμε *ανεξάρτητο και μηδενικής μέσης τιμής* Gaussian θόρυβο ο οποίος έλαβε τιμές έτσι ώστε το *SNR* να διατηρείτε μεταξύ του 5 και 10.

Παράμετροι	**Ενδεικτικές τιμές**
Βάθος	1 έως 4
FOV	50° έως 150°
Ταχύτητα περιστροφής	0.239° έως 0,391° (ανά καρέ)
Ανάλυση (εικονοστοιχεία)	512x512
Ανάλυση πλάνου	0.93x0.93

Πίνακας 5-1: *Τυπικές τιμές παραμετροποίησης του αλγορίθμου δημιουργίας συνθετικής οπτικής ροής.*

Η ταχύτητα μεταφοράς «δεσμεύτηκε» (στην μέση απόσταση Z_{mid} = 2.5) ως προς ταχύτητα περιστροφής με τιμή 0.239° ανά καρέ. Αυτό οδηγεί σε ταχύτητα μεταφοράς περίπου 0.004 *FLU* ή 2.22*p*. Τέλος η κατεύθυνση μεταφοράς καθορίστηκε σε $(4,-3,5)^T$ και η περιστροφή $(-1,2,0.5)^T$ ακτίνια.

· *Σενάριο Πειράματος 1 – για FOV, είσοδο αριθμού διανυσμάτων και ενδεικτικές διαφορές τριών εκδοχών αλγορίθμου*

Τα αποτελέσματα για 4 αντιπροσωπευτικές περιπτώσεις και για τις τρεις εκδοχές του αλγορίθμου παρουσιάζονται στο σχήμα 5-4. Για κάθε μια από τις προσομοιώσεις που περιγράψαμε πραγματοποιήσαμε 100 δοκιμές (trials). Χρησιμοποιήθηκαν δε 15 αρχικοποιήσεις και επανεκκινήσεις στον Gauss-Newton και το κατώφλι σύγκλισης καθορίστηκε σε 10^{-13}. Κάθε εκτίμηση του αλγορίθμου για την μεταφορά παρουσιάζεται στο γράφημα με τελείες ροζ χρώματος και οι εκτιμήσεις για την περιστροφή με τελείες μαύρου χρώματος. Η πραγματική μεταφορά t_{true} και περιστροφή ω_{true} σημειώνεται ως '+' και '$\triangledown$' αντίστοιχα. Καταγράφονται ακόμα η μέση μεταφορά t_{avr} και μέση περιστροφή ω_{avr} για τις 100 δοκιμές, η πόλωση και απόκλιση, το μέσο μέτρο του διανύσματος περιστροφής $\|\omega\|_{avr}$ καθώς και το *SNR*.

Οι δοκιμές έγιναν αφού εφαρμόστηκε στα διανύσματα θόρυβος σ_n = 0.5 εικονοστοιχεία. Επειδή το σ_n προστέθηκε σε κάθε στοιχείο των διανυσμάτων ξεχωριστά, ο θόρυβος που επιδρά είναι $\sqrt{2\sigma_{\mathrm{n}}^{2}}$ δηλαδή 0.71 εικονοστοιχεία. Σε κάποιες δοκιμές όπου το μέσο μέγεθος των διανυσμάτων προέκυψε αρκετά μικρό (λόγω *FOV* ή της φύσης της κίνησης που προσομοιώθηκε) πραγματοποιήσαμε διορθώσεις στο σ_n ώστε το *SNR* να βρίσκεται πάντα εντός του διαστήματος (5,10).

Εκτός από το παραπάνω σενάριο που ισχύει σε όλα τα γραφήματα του σχήματος 5-4 εφαρμόστηκαν οι εξής επιπρόσθετες παραμετροποιήσεις :

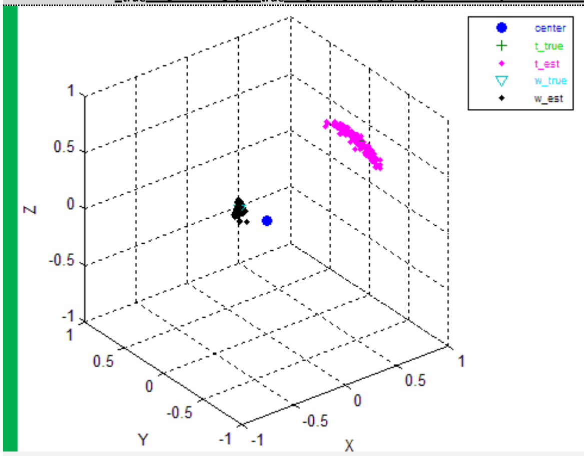
Τα πειράματα 1,2,3 (ένα για κάθε εκδοχή του αλγορίθμου) πραγματοποιήθηκαν με οπτικό πεδίο 50° και θόρυβο $\sigma_n = 0.5p$ ανά διάσταση.

Στο πείραμα 4 βλέπουμε την συμπεριφορά του αλγορίθμου (διαχωρισμένου Gauss-Newton) αυξάνοντας το οπτικό πεδίο στην τιμή των 150° ενώ ο θόρυβος παραμένει στο ίδιο επίπεδο. Στις λήψεις με τόσο μεγάλο οπτικό πεδίο γίνεται χρήση ευρυγώνιων φακών που προκαλούν μια κοίλη παραμόρφωση περιμετρικά των εικόνων, φαινόμενο γνωστό και ως μάτι του ψαριού (*fisheye lens*). Η τιμή των 150° είναι μια ακραία περίπτωση που όμως προσφέρει διαφωτιστικά στοιχεία για το πως επηρεάζεται η απόδοση του αλγορίθμου από το οπτικό πεδίο.

Στην συνέχεια (πειράματα 5,6) επιχειρούμε να συνθέσουμε ένα πιο ρεαλιστικό μοντέλο θορύβου χρησιμοποιώντας μεικτό Gaussian θόρυβο σε αναλογίες 85%-15% με σ_{n1} = 0.5 και σ_{n2} > 3, το οποίο οδηγεί σε *SNR* 5 και < 1 αντίστοιχα. Με αυτό τον τρόπο τα 85 από τα 100 διανύσματα έχουν τον θόρυβο της πρώτης δοκιμής και τα υπόλοιπα 15 έχουν μεγάλα ποσά θορύβου έτσι ώστε να αλλοιώνεται εντελώς η κατεύθυνση τους. Στα γραφήματα 5,6 βλέπουμε τις διαφορές στην ακρίβεια των εκτιμήσεων μεταξύ της σταθμισμένη και μη σταθμισμένης εκδοχής του διαχωρισμένου Gauss-Newton.

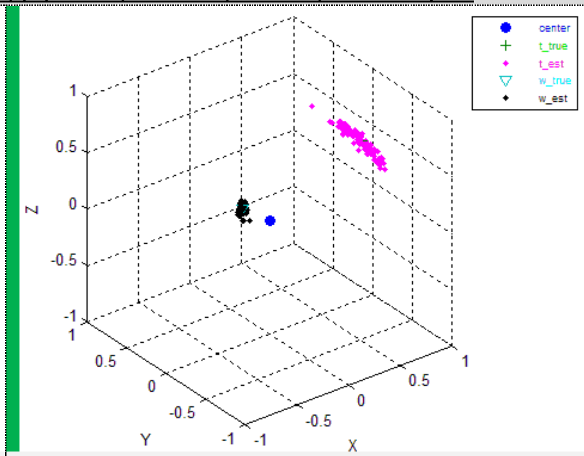
Στο τελευταίο πείραμα 7 βλέπουμε την συμπεριφορά του αλγορίθμου (διαχωρισμένου Gauss-Newton) όταν αυξάνουμε τον αριθμό των εισαγόμενων προς επεξεργασία διανυσμάτων από 100 σε 1000.

t_{true} = [4 -3 5]^T, ω_{true} = [-1 2 0.5]^T, sigma = 0.5, fov = 50°, |R| = 0.239, SNR = 5.56, 100 trials, 100 διανύσματα



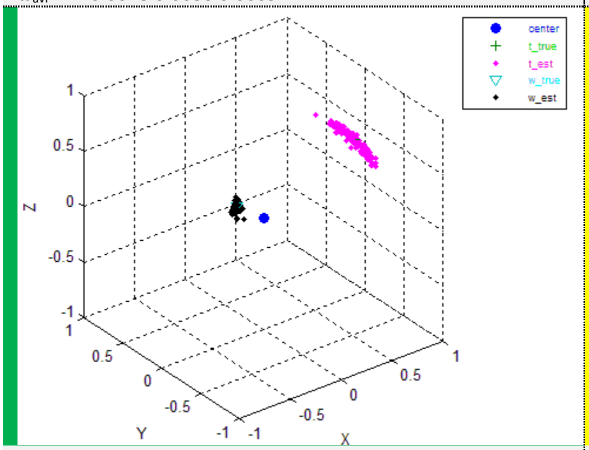
t_{error} = 7.8453 ± 4.2009
ω_{error} = 0.052621 ± 0.031962
$|R_{est}|$ = 0.237
t_{avr} = 0.5628 -0.4137 0.7156
ω_{avr} = -0.0018 0.0036 0.0009

Πείραμα 1
Επιπλέον ιδιότητες:
-



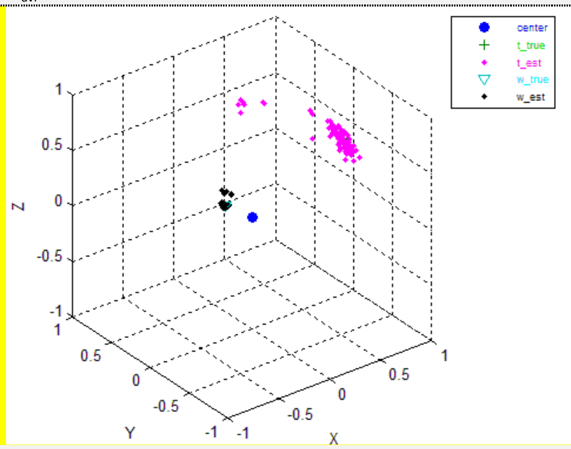
t_{error} = 7.8399 ± 4.5582
ω_{error} = 0.052435 ± 0.033243
$|R_{est}|$ = 0.236
t_{avr} = 0.5633 -0.4122 0.7161
ω_{avr} = -0.0018 0.0036 0.0009

Πείραμα 2
Επιπλέον ιδιότητες:
<u>Με στάθμιση</u>



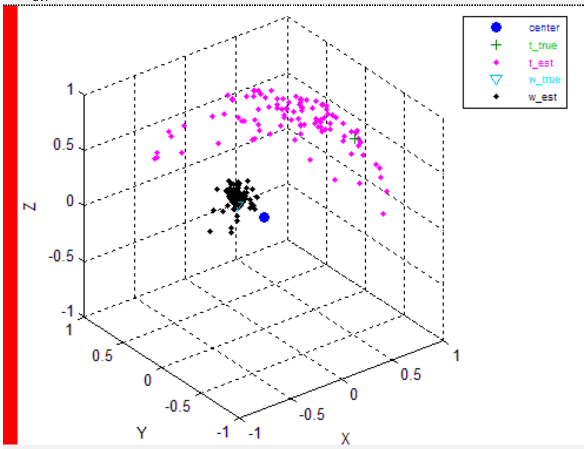
t_{error} = 7.6511 ± 4.0994
ω_{error} = 0.052017 ± 0.032191
$|R_{est}|$ = 0.236
t_{avr} = 0.5653 -0.4143 0.7133
ω_{avr} = -0.0018 0.0036 0.0009

Πείραμα 3
Επιπλέον ιδιότητες:
<u>Με μείωση ελαχίστων</u>



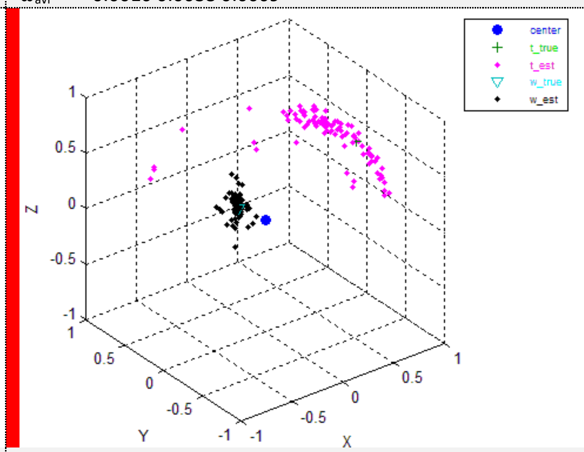
t_{error} = 9.157 ± 11.1454
ω_{error} = 0.032833 ± 0.052768
$|R_{est}|$ = 0.245
t_{avr} = 0.5398 -0.4004 0.7405
ω_{avr} = -0.0016 0.0038 0.0009

Πείραμα 4
Επιπλέον ιδιότητες:
fov = 150°, sigma = 0.2,
SNR = 10



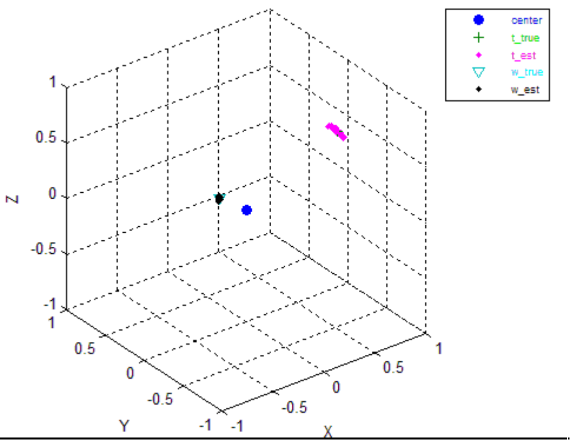
t_{error} = 37.121 ± 22.1623
ω_{error} = 0.17382 ± 0.091168
$|R_{est}|$ = 0.310
t_{avr} = 0.2171 -0.1325 0.9671
ω_{avr} = -0.0005 0.0053 0.0009

Πείραμα 5
Επιπλέον ιδιότητες:
Ποσοστό θορύβου 85-15%
sigma = 0.5-3, SNR = 5.56-1



t_{error} = 22.9176 ± 19.1367
ω_{error} = 0.12687 ± 0.095243
$|R_{est}|$ = 0.267
t_{avr} = 0.4157 -0.3370 0.8448
ω_{avr} = -0.0014 0.0043 0.0009

Πείραμα 6
Επιπλέον ιδιότητες:
Ποσοστό θορύβου 85-15%
sigma = 0.5-3, SNR = 5.56-1
<u>Με στάθμιση</u>



Πείραμα 7
Επιπλέον ιδιότητες:
Αριθμός διανυσμάτων m = 1000

Χρονική απόκριση: 5.35 sec ανά αρχικοποίηση
t_{error} = 2.0725 ± 1.154
ω_{error} = 0.013739 ± 0.0075329
$|R_{est}|$ = 0.238
t_{avr} = 0.5692 -0.4225 0.7053
ω_{avr} = -0.0018 0.0036 0.0009

Σχήμα 5-4 (+Προηγούμενη σελίδα): *Δόκιμες για 4 χαρακτηριστικές περιπτώσεις και για τις 3εις εκδοχές του αλγορίθμου εκτίμησης της κίνησης που παρουσιάσαμε.*

· *Οπτικό πεδίο*

Σαν πρώτο συμπέρασμα παρατηρούμε ότι αύξηση του οπτικού πεδίου συνεπάγεται αύξηση των σφαλμάτων (σύγκριση πειραμάτων 1 και 4). Η αιτία για αυτό είναι ότι όσο αυξάνεται το *FOV* τόσο μειώνεται το μέσο μέτρο των διανυσμάτων. Ακριβολογώντας το μέσο μέτρο των διανυσμάτων είναι αντιστρόφως ανάλογο του *FOV*, διότι αύξηση του *FOV* ισοδυναμεί με αύξηση του οπτικού πλάνου. Έτσι είσαι σαν να μεταφέρουμε τα (ίδια) διανύσματα σε μεγαλύτερη επιφάνεια. Αυτό διαφαίνεται και ποσοτικά (πίνακας 5-2) όπου ενώ το μέσο μέτρο των διανυσμάτων του πειράματος 1 (*FOV* = 50°) είναι 3.82*p*, στο πείραμα 4 με *FOV* = 150° καταλήγει σε 2.57*p*. Συνεπώς με σ_n = 0.5 λαμβάνουμε *SNR* = 5.95 για *FOV* = 50° ενώ *SNR* = 3.98 για *FOV* = 150°.

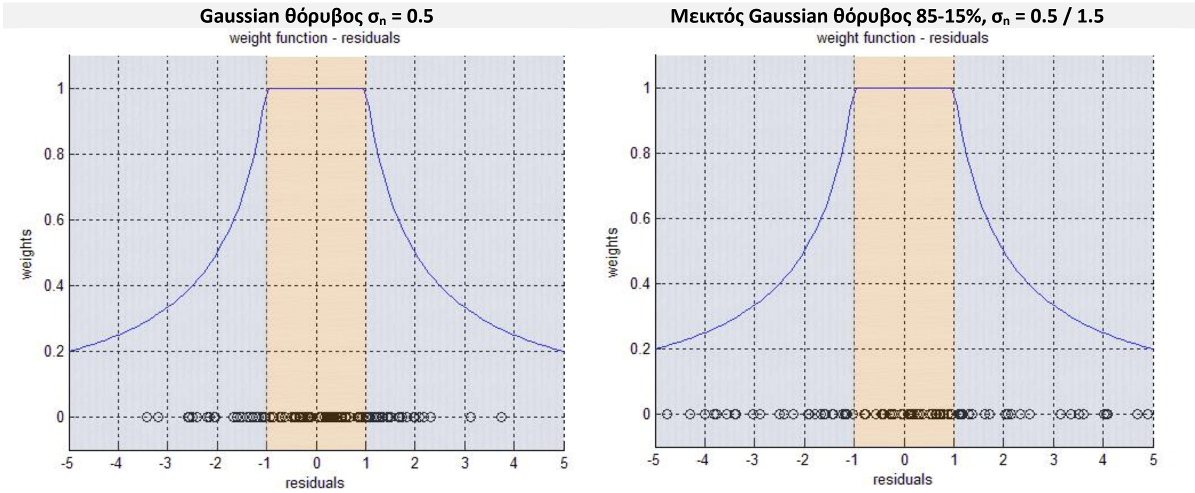
Επιπρόσθετα στα μεγάλα οπτικά πεδία τα διανύσματα που βρίσκονται κοντά στο κέντρο του πλάνου έχουν γενικώς μικρότερο μέτρο σε σχέση με αυτά που βρίσκονται περιμετρικά της εικόνας. Το αποτέλεσμα είναι να εμφανίζονται λιγότερο εσφαλμένα διανύσματα στα άκρα των εικόνων ενώ διανύσματα με περισσότερα σφάλματα προς το κέντρο. Όπως βλέπουμε στα πειράματα αυτός ο ανομοιόμορφα κατανεμημένος θόρυβος είναι δυσκολότερο να εξομαλυνθεί ανάμεσα στα δείγματα με αποτέλεσμα λιγότερο αξιόπιστες εκτιμήσεις. Στο σχήμα 5-3 διακρίνουμε την μείωση του μεγέθους των διανυσμάτων κοντά στο κέντρο του πλάνου για μεγάλο οπτικό πεδίο.

Θόρυβος:σ_n=**0.46** Μέσο μέτρο θορύβου:0.6460 Μέτρο στάνταρτ απόκλισης:0.6442 ***Παράμετροι*** ***SNR***		**Τιμές**	
t = **ω =** **FOV =**	**4 -3 5** **-1 2 0.5** **50°**	Μέσο μέτρο σήματος: SNR:	3.8296 5.9512
t = **ω =** **FOV =**	**4 -3 5** **-1 2 0.5** **150°**	Μέσο μέτρο σήματος: SNR:	2.5704 3.9867
t = **ω =** **FOV =**	**1 2 0** **0 0 0** **50°**	Μέσο μέτρο σήματος: SNR:	2.8701 4.4584
t = **ω =** **FOV =**	**0 0 1** **0 0 0** **50°**	Μέσο μέτρο σήματος: SNR:	1.0908 1.6930

Πίνακας 5-2: *Ο λόγος σήματος προς θόρυβο δεν εξαρτάται μόνο από την στάνταρτ απόκλιση του θορύβου αλλά και από άλλους παράγοντες όπως την κατεύθυνση αυτή καθεαυτή και το οπτικό πεδίο.*

· *Στάθμιση*

Σχετικά με την στάθμιση παρατηρούμε ότι η σταθμισμένη (robust) εκδοχή παράγει ελαφρώς χειρότερα αποτελέσματα σε σχέση με την μη-σταθμισμένη όπως διαπιστώνουμε από την σύγκριση των πειραμάτων 1,2. Η αιτία για αυτό είναι ότι ενώ ο μέσος όρος του θορύβου είναι μηδενικός, η εκδοχή του σταθμισμένου Gauss-Newton θεωρεί ως υπόθεση ότι ένα ποσοστό της πληροφορίας «ξεχωρίζει» έχοντας σχετικά μεγαλύτερο σφάλμα. Έτσι μοιραία θα αποδοθεί μικρότερο βάρος και θα αγνοηθεί ένα ποσοστό διανυσμάτων με εν δυνάμει χρήσιμη πληροφορία συμπαρασύροντας την λύση προς την κατεύθυνση που υποδεικνύουν τα υπόλοιπα διανύσματα (σχήμα 5-5). Ωστόσο τα αντίθετα ισχύουν όταν εφαρμόζεται *ανάμεικτος* Gaussian θόρυβος (σύγκριση πειραμάτων 5,6) με την σταθμισμένη εκδοχή να παράγει καλύτερα αποτελέσματα.



Σχήμα 5-5: *Όταν ο θόρυβος είναι κοινός για όλα τα δείγματα, η πλειονότητα τους παράγει υπόλοιπα της ίδιας τάξης. Ωστόσο ποσοστιαία θα δοθεί μικρότερο βάρος σε ένα κομμάτι ίσως ακόμα και χρήσιμης πληροφορίας (αριστερά). Η στάθμιση λειτουργεί καλύτερα με τον μεικτό θόρυβο όπου πράγματι ένα μέρος των δειγμάτων «ξεχωρίζει» παράγοντας μεγαλύτερα υπόλοιπα (δεξιά). Εδώ κατανοούμε την σπουδαιότητα της σωστής επιλογής για το κατώφλι c. Στο σχήμα τα υπόλοιπα έχουν κανονικοποιηθεί στην μονάδα βάσει του κατωφλιού.*

Εντούτοις το ποσοστό βελτίωσης στον σταθμισμένο Gauss-Newton έναντι των άλλων εκδοχών είναι μάλλον δύσκολο να προσδιοριστεί επακριβώς και θα πρέπει να εξετάζεται κατά περίπτωση. Ένα μέρος της δυσκολίας ανάλυσης της απόδοσης των διαφορετικών εκδοχών του αλγορίθμου προέρχεται από την ίδια την φύση του προβλήματος δηλαδή από το γεγονός του πλήθους των προς επίλυση αγνώστων και της πολυπλοκότητας του μη-γραμμικού μοντέλου. Στον σταθμισμένο Gauss-Newton το ζήτημα αυτό γίνεται ακόμα πιο πολυσύνθετο αφού υπεισέρχονται επιπρόσθετοι παράμετροι όπως η επενέργεια της συνάρτηση επίδρασης στο μαθηματικό μοντέλο, ο καθορισμός του κατωφλιού επιλογής μικρών-μεγάλων αποκλίσεων και η πόλωση (που δημιουργείτε στις λύσεις) από τις αρχικές τιμές. Μερικά χαρακτηριστικά που διαπιστώθηκαν στις δοκιμές είναι ότι στην επίλυση με την σταθμισμένη εκδοχή παρουσιάζονται περισσότερα ελάχιστα και η σύγκλιση πραγματοποιείτε πιο αργά σε σχέση με τις άλλες υλοποιήσεις.

· Απόκριση με είσοδο μεγάλου αριθμού διανυσμάτων

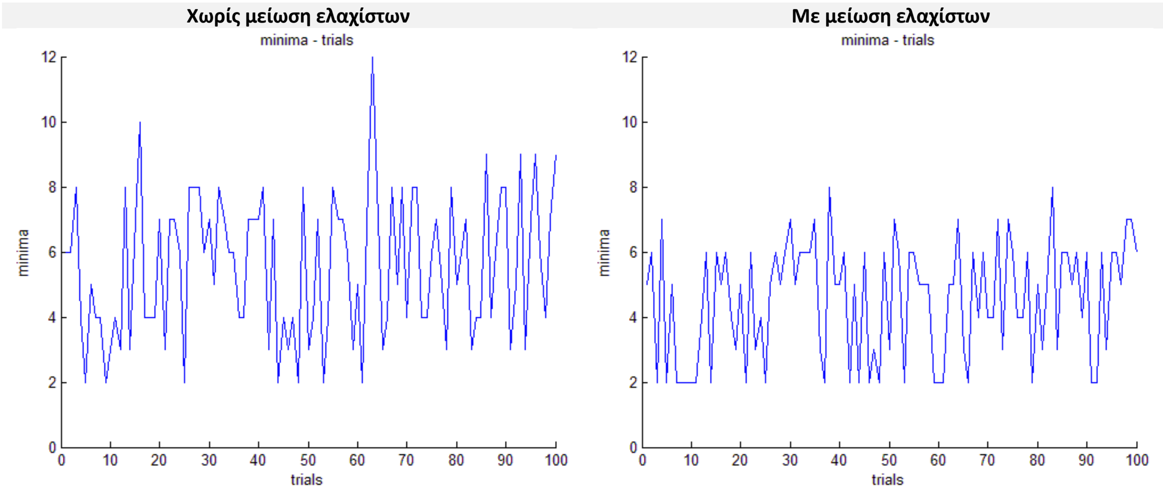
Στο πείραμα 7 επιχειρήσαμε να δούμε την συμπεριφορά του αλγορίθμου (για τον διαχωρισμένο Gauss-Newton) σε είσοδο μεγάλου αριθμού δειγμάτων (1000 διανύσματα). Το πείραμα αυτό απέδωσε τα πλέον ακριβή αποτελέσματα. Φυσικά το ανάλογο τίμημα ήταν στον χρόνο απόκρισης, χαρακτηριστικό που δεν περιλαμβάνουμε στην αξιολόγησή μας αλλά καταγράφουμε ενδεικτικά μόνο για αυτή την περίπτωση.

· Μέθοδος μείωσης ελαχίστων

Όπως παρατηρούμε στο πείραμα 3 η «μείωση ελαχίστων» δίνει αποτελέσματα πανομοιότυπης εγκυρότητας με την μέθοδο διαχωρισμένου Gauss-Newton του πειράματος 1. Συγκρίνοντας τα ελάχιστα των πειραμάτων 1 και 3 ανά trial αποδεικνύεται πως η προτεινόμενη μέθοδος καταφέρνει την μείωση των ελαχίστων σε ένα αξιόλογο βαθμό (σχήμα 5-6). Ωστόσο διαπιστώνουμε ότι αυτό δεν βελτιώνει την ακρίβεια των αποτελεσμάτων γιατί στην πράξη ο αλγόριθμος χωρίς την «μείωση ελαχίστων» δεν «επέλεγε» ποτέ ως τελική λύση κάποιο ελάχιστο από αυτά που η «μείωση ελαχίστων» εξάλειφε.

Με άλλα λόγια η ακρίβεια των εκτιμήσεων βελτιώνεται από την «μείωση ελαχίστων» μόνο όταν εξαλείφεται κάποιο από τα ανεπιθύμητα ελάχιστα που οι άλλες υλοποιήσεις θα επέλεγαν ως λύση, κάτι που συμβαίνει σπάνια. Ενδεχομένως το κέρδος από αυτή τη μέθοδο θα μπορούσε να είναι η μείωση του αριθμού των αρχικοποιήσεων (κατ' επέκταση βελτίωση του χρόνου απόκρισης).

Προς το παρόν το τίμημα που πληρούμε για τον περιορισμό των ελαχίστων είναι η καθυστέρηση της σύγκλισης. Σαν τελικό συμπέρασμα θεωρούμε ότι η ακρίβεια των εκτιμήσεων που παράγεται με την μέθοδο «μείωσης ελαχίστων» είναι ίδια με αυτή των εκτιμήσεων του διαχωρισμένου Gauss-Newton.



Σχήμα 5-6: *Η μέθοδος «μείωσης ελαχίστων» περιορίζει τα ελάχιστα αλλά δεν τα εξαλείφει. Το γράφημα παρουσιάζει το πλήθος των ελαχίστων για κάθε μία από τις 100 δοκιμές, με και χωρίς «μείωση ελαχίστων».*

· *Σχέση κατευθύνσεων και σφαλμάτων*

Ένα ενδιαφέρον στοιχείο που παρατηρήθηκε στις δοκιμές είναι οι διακυμάνσεις στα σφάλματα αναλόγως των κινήσεων που πραγματοποιεί η κάμερα. Καθώς φαίνεται οι θέσεις των διανυσμάτων και η κατεύθυνση της μεταφοράς - περιστροφής αυτή καθ' εαυτή σχετίζονται με την ακρίβεια των εκτιμήσεων. Διαφαίνεται ότι κάποιες κατευθύνσεις είναι πιο εύκολο να εκτιμηθούν σωστά με λιγότερο σφάλμα ενώ κάποιες άλλες όχι. Μερικά συμπεράσματα που προέκυψαν από τα πειράματα είναι τα εξής:

- Οι θέσεις των διανυσμάτων στην εικόνα προκαλούν ιδιομορφίες στην συνάρτηση κόστους και επηρεάζουν τον αριθμό και τις θέσεις των ελαχίστων της.
- Η κατεύθυνση της μεταφοράς και περιστροφής (ανεξάρτητα από τις εκάστοτε ταχύτητες τους) επηρεάζει το μέσο μέτρο των διανυσμάτων $E\left\{\|u_i\|_2^2\right\}$ άρα και το *SNR*. Ο πίνακας 5-2 δείχνει ακριβώς αυτή την συσχέτιση μεταξύ των κατευθύνσεων και του *SNR*.
- Σε πολλές περιπτώσεις πλαγιοπλευρικών κινήσεων ο αλγόριθμος εκτιμά λαθεμένα την κίνηση ως περιστροφική παρά μεταφορική όπως θα' πρεπε. Η αιτία είναι ότι ανόμοιες κινήσεις της κάμερας παράγουν παρόμοιο πεδίο οπτική ροής. Π.χ. Η αμιγώς μεταφορική κίνηση κατά τον *X*-άξονα παράγει πανομοιότυπο πεδίο με την αμιγώς περιστροφική κίνηση κατά τον *Y*-άξονα. Σε αυτές τις περιπτώσεις

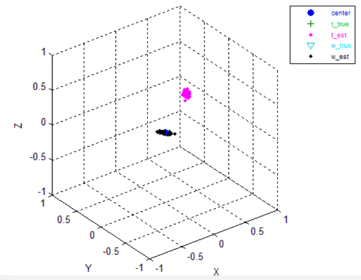
εμφανίζονται δυο ελάχιστα στην συνάρτηση κόστους με εξίσου χαμηλή τιμή ολικού κόστους. Το ένα ευθύνεται για την περιστροφή κατά *Y* και το άλλο για την μεταφορά κατά *X*. Απουσία θορύβου η διαφορά στο κόστος μεταξύ των δυο είναι αρκετά μεγάλη ώστε να είναι ξεκάθαρη η λύση – δηλαδή ποια κίνηση πραγματοποιήθηκε. Όμως παρουσία θορύβου η διαφορά περιορίζεται και τελικώς ενδέχεται μια μεταφορική κίνηση να εκτιμηθεί λάθος ως περιστροφική και το αντίθετο. Το φαινόμενο αυτό εντείνεται όσο μειώνεται το οπτικό πεδίο και γενικότερα η σύγχυση αυτή φαίνεται να εξαλείφεται όσο πιο περίπλοκη είναι η κίνηση της κάμερας. Παρακάτω θα δούμε ότι το μοντέλο κίνησης (28) είναι αρκετά «ελαστικό» ώστε να «εκφράζει» ένα πεδίο οπτικής ροής με πλήθος κινήσεων.

· *Σφάλματα συναρτήσει του θορύβου*

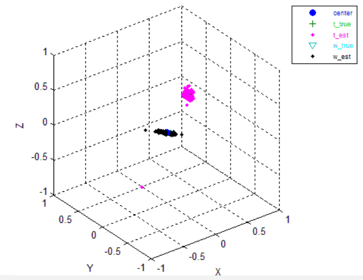
Ομολογουμένως οι πλαγιοπλευρικές μεταφορές φαίνεται να εμπεριέχουν κάποια δυσκολία στον υπολογισμό τους γι’ αυτό και πραγματοποιήσαμε μια σειρά από πειράματα με πλευρική μεταφορική κατεύθυνση $[2\ 1.5\ 0]^T$ και τον διαχωρισμένο Gauss-Newton. Στην δόκιμη αυτή (σχήμα 5-7) βλέπουμε πως κυμαίνονται η *πόλωση* και *απόκλιση* συναρτήσει του θορύβου ο οποίος λαμβάνει τιμές από 0.2 έως 0.8 εικονοστοιχεία.

Στο πείραμα φαίνεται ξεκάθαρα ότι οι επιπτώσεις του θορύβου επεκτείνονται σε μια σειρά από θέματα όπως λ.χ. η αύξηση των ελαχίστων και του χρόνου σύγκλισης. Επίσης παρατηρούμε ότι επηρεάζεται και το μέτρο της περιστροφής το οποίο λαμβάνει συνεχώς μεγαλύτερες τιμές. Ο *SNR* στις δοκιμές κυμαίνεται από 2.5 ως 10.5. Το σχήμα 5-8 συγκεντρώνει τα αποτελέσματα.

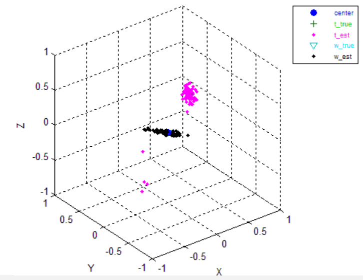
t_{true} = [2 1.5 0]^T ω_{true} = [0 0 0]^T fov = 50°, 100 trials, 100 διανύσματα



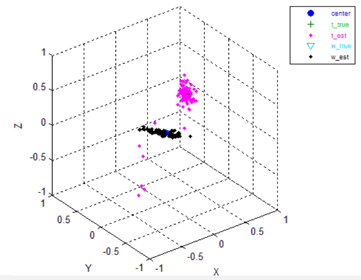
Sigma = 0.2 ⬄ SNR = 10.2543
t_{error} = 1.6861 ± 0.949
ω_{error} = 0.038759 ± 0.02529
$|R_{est}|$ = 0.002
Μ.Ο. επαναλήψεων: 18.6913
Μέσος αριθμός ελαχίστων: 1.4700



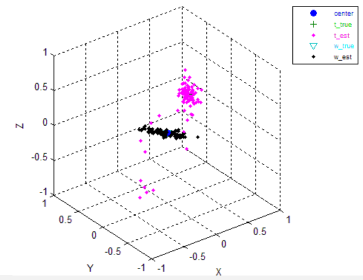
sigma = 0.3 ⬄ SNR = 6.8362
t_{error} = 4.2312 ± 16.3494
ω_{error} = 0.060119 ± 0.043534
$|R_{est}|$ = 0.001
Μ.Ο. επαναλήψεων: 21.5433
Μέσος αριθμός ελαχίστων: 1.6400



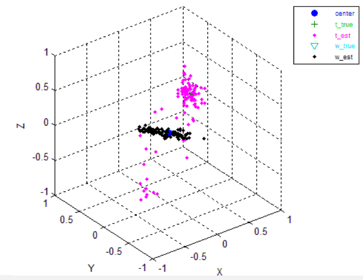
sigma = 0.4 ⬄ SNR = 5.1271
t_{error} = 10.2247 ± 32.1153
ω_{error} = 0.083945 ± 0.06286
$|R_{est}|$ = 0.004
Μ.Ο. επαναλήψεων: 24.6920
Μέσος αριθμός ελαχίστων: 1.8000



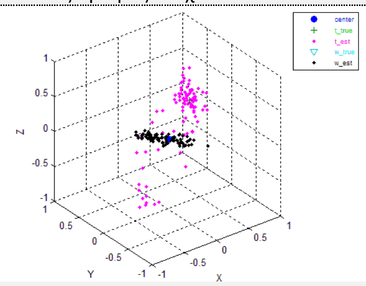
sigma = 0.5 ⬄ SNR = 4.1017
t_{error} = 14.373 ± 37.1679
ω_{error} = 0.10602 ± 0.074578
$|R_{est}|$ = 0.010
Μ.Ο. επαναλήψεων: 29.6047
Μέσος αριθμός ελαχίστων: 2.0300



sigma = 0.6 ⬄ SNR = 3.4181
t_{error} = 20.311 ± 43.8146
ω_{error} = 0.12939 ± 0.086132
$|R_{est}|$ = 0.017
Μ.Ο. επαναλήψεων: 34.1560
Μέσος αριθμός ελαχίστων: 2.3900



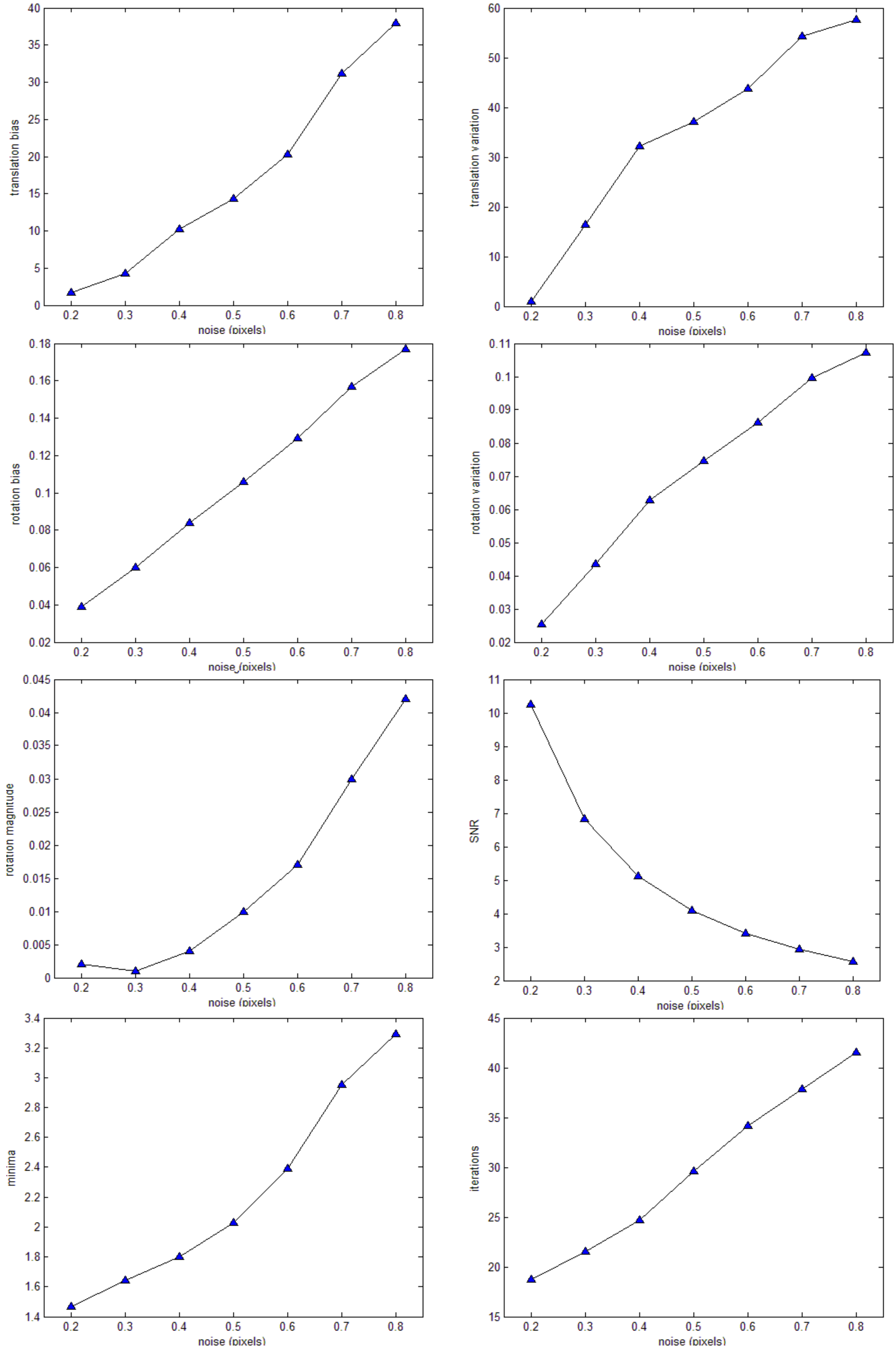
sigma = 0.7 ⬄ SNR = 2.9298
t_{error} = 31.1209 ± 54.3157
ω_{error} = 0.15665 ± 0.099736
$|R_{est}|$ = 0.030
Μ.Ο. επαναλήψεων: 37.8393
Μέσος αριθμός ελαχίστων: 2.9500



sigma = 0.8 ⬄ SNR = 2.5636
t_{error} = 37.9558 ± 57.7149
ω_{error} = 0.17713 ± 0.10727
$|R_{est}|$ = 0.042
Μ.Ο. επαναλήψεων: 41.5520
Μέσος αριθμός ελαχίστων: 3.2900

Σχήμα 5-7: *Δόκιμες πλαγιοπλευρικών κινήσεων συναρτήσει του θορύβου.*

Μετρικές απόδοσης συναρτήσει του θορύβου



Σχήμα 5-8 (Προηγούμενη σελίδα): *Η πόλωση και οι αποκλίσεις μεταφοράς-περιστροφής, το μέγεθος περιστροφής, ο λόγος σήματος προς θόρυβος, τα ελάχιστα της συνάρτησης κόστους και ο μέσος αριθμός επαναλήψεων συναρτήσει του θορύβου. Οι τιμές της στάνταρτ απόκλισης του κυμαίνονται από 0.2 έως 0.8 εικονοστοιχεία.*

· *Αρχικές τιμές και μη περιορισμός θετικού βάθους*

Στις δόκιμες χρησιμοποιήθηκαν διάφοροι τρόποι αρχικοποίησης των τιμών t_{init}. Θα πρέπει να επισημάνουμε ότι η αναλογία των *t,d* στην (28) εκτός του ότι προκαλεί αδυναμία εύρεσης των απόλυτων τιμών τους, δημιουργεί και ένα ζήτημα σχετικά με την φορά του διανύσματος μεταφοράς αφού $t \cdot d = (-t) \cdot (-d)$. Αυτό επηρεάζει την διαδικασία επίλυσης με δύο τρόπους:

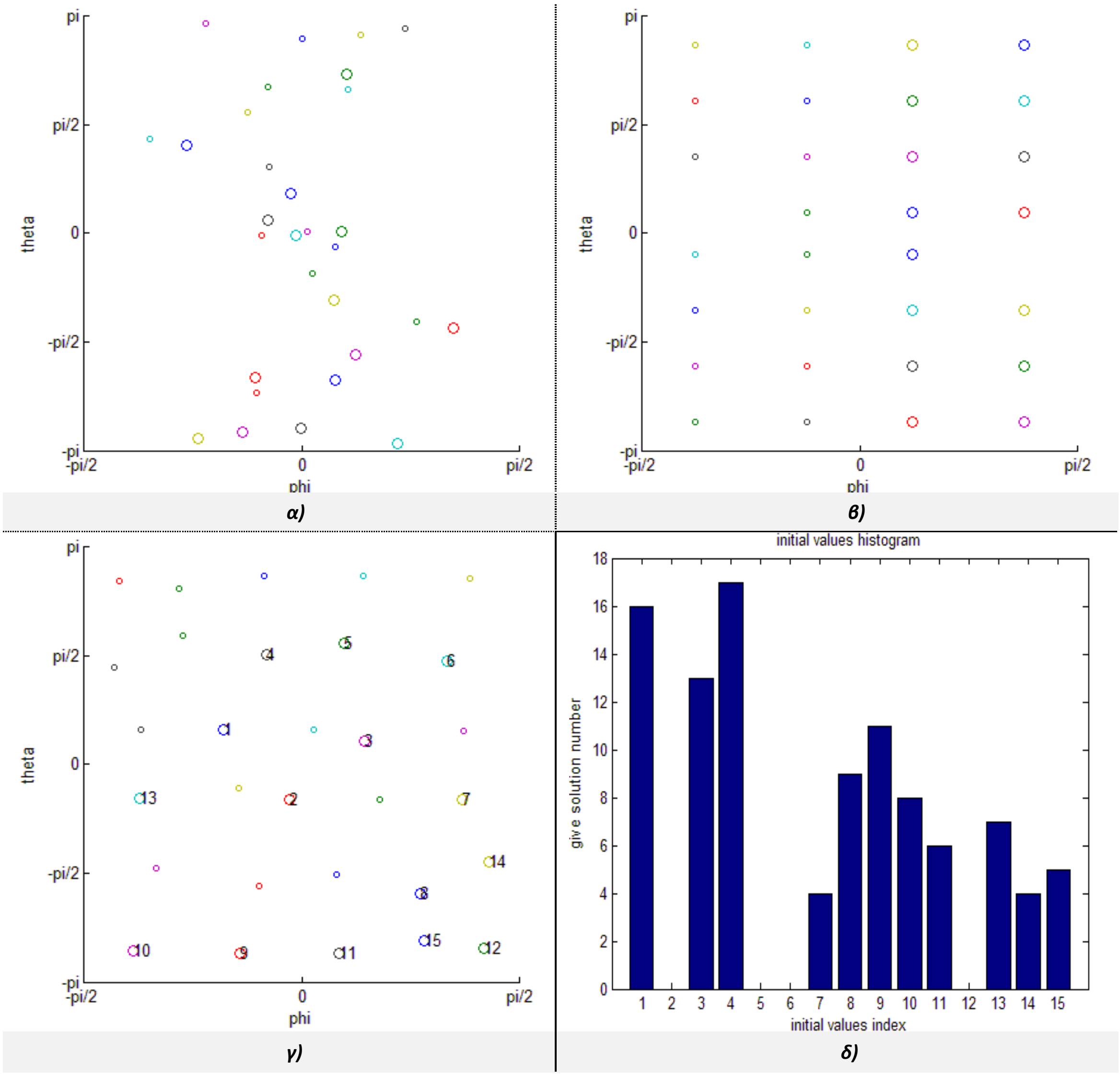
- Αντίθετες τιμές αρχικοποίησης t_{init} οδηγούν στην ίδια (αντίθετη) λύση με πλήρη ταύτιση των ενδιάμεσων (αντίθετων) τιμών σε όλα τα στάδια της βελτιστοποίησης μέχρι την σύγκλιση. Επομένως με 15 αρχικοποιήσεις ελέγχουμε το αποτέλεσμα για 30, αυτές των t_{init} και των $-t_{init}$. Με $\|t_{init}\| = 1$ μπορούμε να σκεφτούμε τις αρχικές τιμές σαν σημεία μιας σφαίρας με ακτίνα $R = 1$. Θα πρέπει λοιπόν να εκμεταλλευτούμε το γεγονός ότι αντιδιαμετρικά t_{init} οδηγούν στα ίδια αποτελέσματα φροντίζοντας να λάβουμε δείγματα από την «σφαίρα» αυτή με κατάλληλο τρόπο ώστε να καλύψουμε όσο το δυνατόν πιο ομοιόμορφα το πεδίο αρχικών τιμών.

Στο σχήμα 5-9*α)* παρατηρούμε ότι η τυχαία αρχικοποίηση δεν αξιοποιεί το γεγονός ότι αντιδιαμετρικά t_{init} θεωρούνται ισοδύναμα. Ένας τρόπος για να εξασφαλίσουμε ομαλότερα t_{init} είναι η δειγματοληψία της μισής «σφαίρα αρχικών τιμών» αφού η άλλη πλευρά «καλύπτεται» από τα αντίθετα διανύσματα της πρώτης. Το σχήμα 5-9*β)* παρουσιάζει την δειγματοληψία μισής σφαίρας με τις γωνίες διαδοχικών t_{init} ανά δύο σταθερές. Τέλος στο σχήμα 5-9*γ)* παρουσιάζουμε αρχικές τιμές έτσι κατανεμημένες ώστε να υπάρχει τουλάχιστον μια αρχική τιμή κοντά σε όλες τις βασικές μεταφορικές κινήσεις όπως λ.χ. οι αμιγώς πλαγιοπλευρικές, εμπρόσθια (zoom-in/out) κ.α. Όλα τα πειράματα σε αυτό το κεφάλαιο πραγματοποιήθηκαν με τις αρχικές τιμές του *γ).* Το γράφημα 5-9*δ)* παρουσιάζει το ιστόγραμμα των αρχικών τιμών από τα οποία προήλθε η τελική λύση για την περίπτωση του πειράματος του σχήματος 5-4 πείραμα 1.

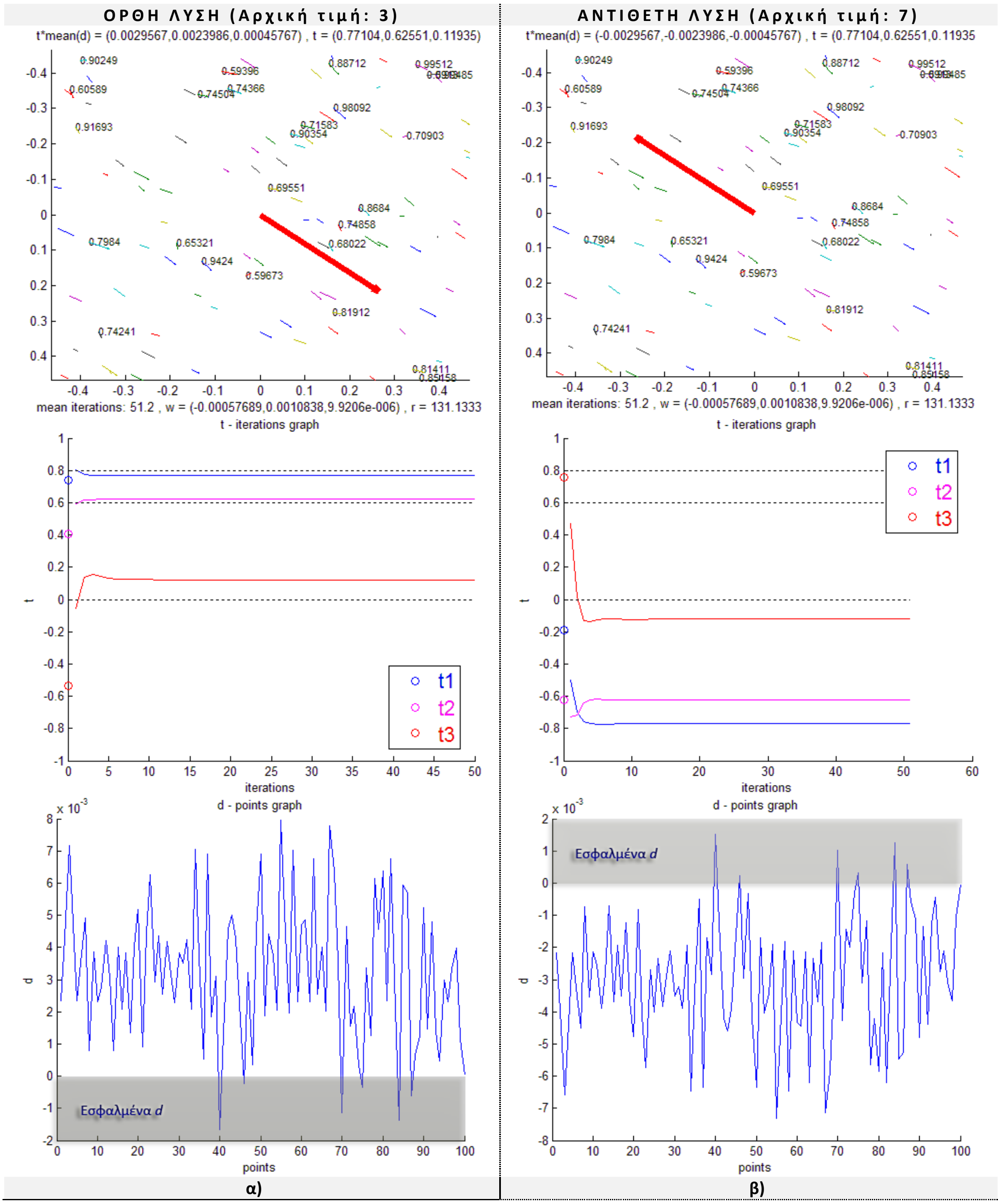
- Ο αλγόριθμος αναλόγως του t_{init} απ' το οποίο ξεκινά τις ανανεώσεις μπορεί να καταλήξει στην $t \cdot d$ ή $(-t)\cdot(-d)$ αφού θεωρούνται λύσεις εφάμιλλες. Το σχήμα 5-10 παρουσιάζει τα αποτελέσματα της ΕΚΚ σε είσοδο οπτικής ροής που προσομοιώνει την κίνηση t_{true} = *[0.8 0.6 0]^T*. Κατά την επίλυση εντοπίστηκαν 4 ελάχιστα (συνολικά) ενώ από αυτά η λύση (το ολικό ελάχιστο) εντοπίστηκε στις 7 από τις 15 αρχικοποιήσεις. Από τις 7 εκφάνσεις του ολικού ελάχιστου οι τρεις παρουσιάστηκαν με *t* και οι υπόλοιπες 4 με *–t*. Το ερώτημα είναι πως μπορούμε να καθορίσουμε την σωστή *t* κατεύθυνση.

Ένας τρόπος για να βγάλουμε συμπεράσματα για την φορά του *t* είναι από το *d* το οποίο αντιπροσωπεύει φυσικό μέγεθος και οφείλει να είναι θετικό. Αφού δεν έχουμε εξαναγκάσει κάποιον περιορισμό θετικού *d* κατά την βελτιστοποίηση, θα πρέπει μετά την σύγκλιση να επιλέξουμε την φορά των *t / d* έτσι ώστε το *d* να λαμβάνει θετικές τιμές. Στην πράξη λόγω του θορύβου και των ανεξάρτητων κινήσεων τα *d* δεν έχουν κοινό πρόσημο σε όλα τα σημεία. Ωστόσο επιλέγουμε εκείνη την φορά του *t* που δίνει τις περισσότερες θετικές *d* ενώ παράλληλα αγνοούμε τυχόν αρνητικές τιμές που παραβιάζουν την έννοια του φυσικού μεγέθους (Σχήμα 5-10).

Συχνά παρουσία έντονου θορύβου εμφανίζονται ανάμεικτα θετικά / αρνητικά *d* με αποτέλεσμα να δημιουργείται αβεβαιότητα για το ποια τελικά είναι η σωστή φορά για το *t*. Αναλογιζόμενοι το γεγονός ότι η αστοχία εκτίμησης της φοράς οδηγεί σε αποκλίσεις 180° από την σωστή λύση μπορούμε να φανταστούμε τις οδυνηρές επιπτώσεις λόγω αυτού στο συνολικό σφάλμα. Στις δόκιμες πλαγιοπλευρικών μετακινήσεων που είδαμε πριν η φορά εκτιμήθηκε λάθος στα 8 από τα 100 trials, το οποίο ήταν αρκετό για να επιφέρει ραγδαία αύξηση στο σφάλμα (σχήμα 5-7 για σ_n = 0.5).



Σχήμα 5-9: *Το t_{init} σε σφαιρικές συντεταγμένες. Οι μεγαλύτεροι κύκλοι συμβολίζουν τα t_{init} ενώ οι μικρότεροι τα αντιδιαμετρικά τους. α) Τυχαίες αρχικές τιμές. Η τυχαία αρχικοποίηση δεν αξιοποιεί το γεγονός ότι αντιδιαμετρικές τιμές οδηγούν στην ίδια λύση, β) Δειγματοληψία ίσων διαστημάτων στην «σφαίρα αρχικών τιμών», γ) Ομοιόμορφη αρχικοποίηση με χρήση γραφικής διεπαφής, δ) Ιστόγραμμα αρχικών τιμών απ' τις οποίες προήλθε η λύση για την $1^η$ δόκιμη των 100 trials.*



Σχήμα 5-10: *Λύσεις με αντίθετα t-d θεωρούνται εφάμιλλες και αντιστοιχούν στο ίδιο ελάχιστο της συνάρτησης κόστους.*

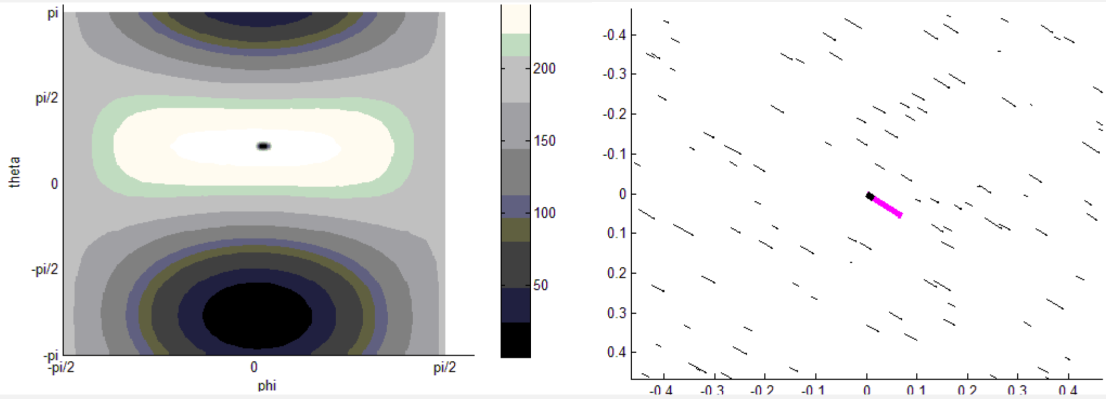
· *«Ελαστικότητα» μοντέλου*

Όπως θα δούμε παρακάτω ένα στιγμιότυπο πεδίου ροής μπορεί να «συμβιβάσει» πολλές «ερμηνείες» για την κίνηση και καθένα από τα ελάχιστα που προκύπτει από την βελτιστοποίηση αποτελεί μια πιθανή από αυτές. Το σχήμα 5-11 παρουσιάζει την ανασύνθεση του πεδίου ροής (μέσω της 28) για καθένα από τα 4 συνολικά ελάχιστα που παρήγαγε η ΕΚΚ με είσοδο το στιγμιότυπο πεδίου ροής του σχήματος 5-10. Η ομοιότητα των παραγόμενων οπτικών ροών είναι φανερή. Θα μπορούσαμε να πούμε ότι το μοντέλο (28) είναι τόσο «ελαστικό» ώστε να «δικαιολογεί» ένα εισαχθέν πεδίο ροής με ένα πλήθος από κινήσεις της κάμερας, όσες είναι και τα ελάχιστα.

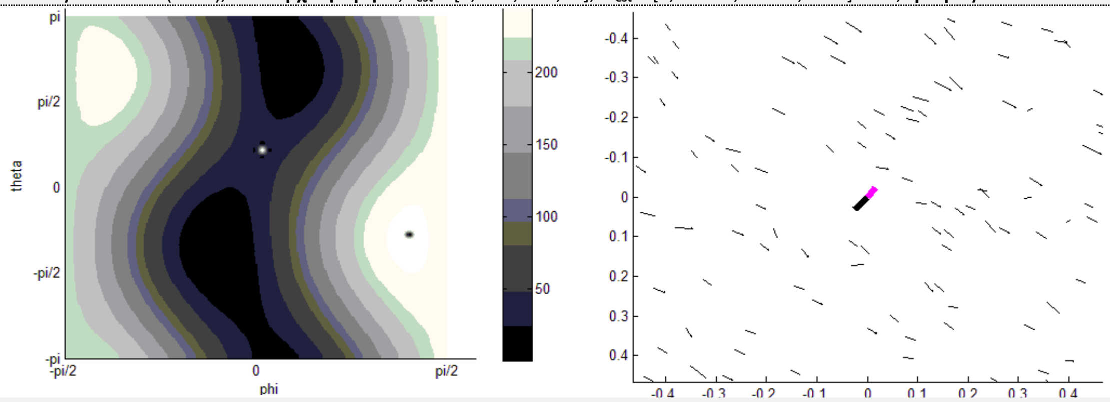
Παρατηρώντας το σχήμα 5-11 βλέπουμε ότι κάποια ελάχιστα, μετά την ανασύνθεση τους, φαίνεται να δίδουν πεδίο ροής που παραδόξως ταυτίζεται πλήρως με το θορυβώδες στιγμιότυπο απ' το οποίο προήλθαν, αναπαριστώντας πιστά ακόμα και τα εσφαλμένα διανύσματα. Η «ελαστικότητα» του μοντέλου καθώς και άλλες αιτίες όπως η απουσία εξαναγκασμού για θετικά d συμβάλλουν σε αυτό.

t_{true} = [0.8 0.6 0]^T ω_{true} = [0 0 0]^T fov = 50°, Αριθμός Αρχικοποιήσεων: 15

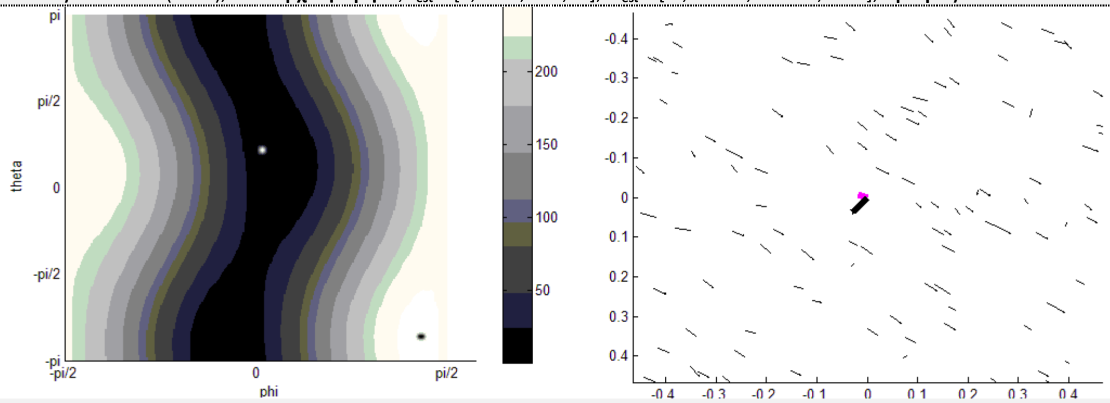
ΣΥΝΑΡΤΗΣΗ ΚΟΣΤΟΥΣ | **ΑΝΑΣΥΝΘΕΣΗ ΟΠΤΙΚΗΣ ΡΟΗΣ**



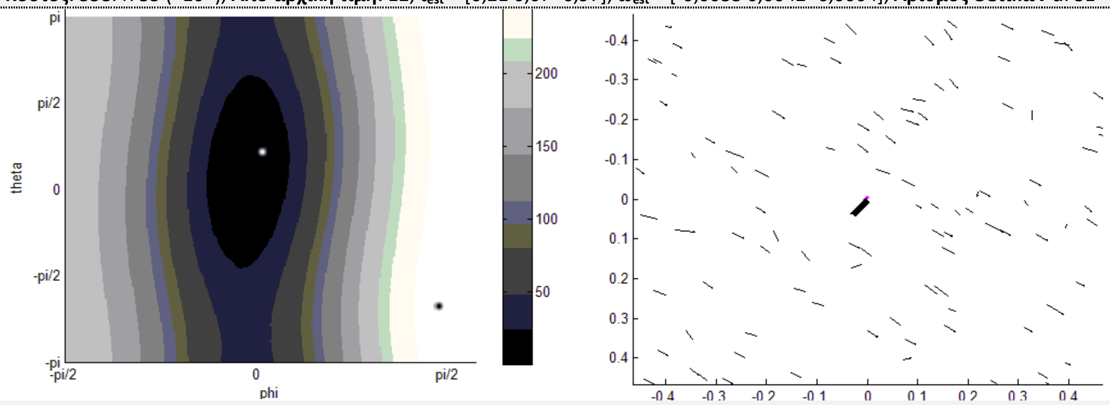
ΕΛΑΧΙΣΤΟ 1 (ΟΛΙΚΟ):
Κόστος: 131.1333 ($\times 10^{-8}$), **Από αρχική τιμή:** 1, **t_{est}** = [0,77 0,62 0,11], **ω_{est}** = [0,1491 0,1408 0,0472]$\times 10^{-3}$,**Αριθμός Θετικών d**: 98



ΕΛΑΧΙΣΤΟ 2:
Κόστος: 373.1571 ($\times 10^{-8}$), **Από αρχική τιμή:** 7, **t_{est}** = [0,17 -0,24 0,95], **ω_{est}** = [-0,0028 0,0036 -0,0002], **Αριθμός Θετικών d**: 50



ΕΛΑΧΙΣΤΟ 3:
Κόστος: 358.4735 ($\times 10^{-8}$), **Από αρχική τιμή:** 12, **t_{est}** = [0,21 0,07 -0,97], **ω_{est}** = [-0,0033 0,0042 -0,0004], **Αριθμός Θετικών d**: 51



ΕΛΑΧΙΣΤΟ 4:
Κόστος: 362.2232 ($\times 10^{-8}$), **Από αρχική τιμή:** 14, **t_{est}** = [-0,03 -0,05 0,99], **ω_{est}** = [-0,0031 0,0043 0,0003], **Αριθμός Θετικών d**: 51

Σχήμα 5-11(Προηγούμενη σελίδα): *Ανασυνθέτοντας την οπτική ροή με χρήση των παραμέτρων t,ω,d καθενός απ' τα ελάχιστα παρατηρούμε ότι αυτές προσεγγίζουν με χαρακτηριστικό τρόπο το αυθεντικό πεδίο οπτικής ροής της εισόδου (δεξιά στήλη). Στο σχήμα της συνάρτησης κόστους με μαύρο στίγμα συμβολίζουμε την εκτίμηση* t_{est} *και με λευκό την πραγματική* t_{true} *. Το πεδίο τιμών του t αποδίδεται σε σφαιρικές συντεταγμένες προκειμένου να αναπαρασταθεί στο δισδιάστατο επίπεδο του σχήματος. (αριστερή στήλη).*

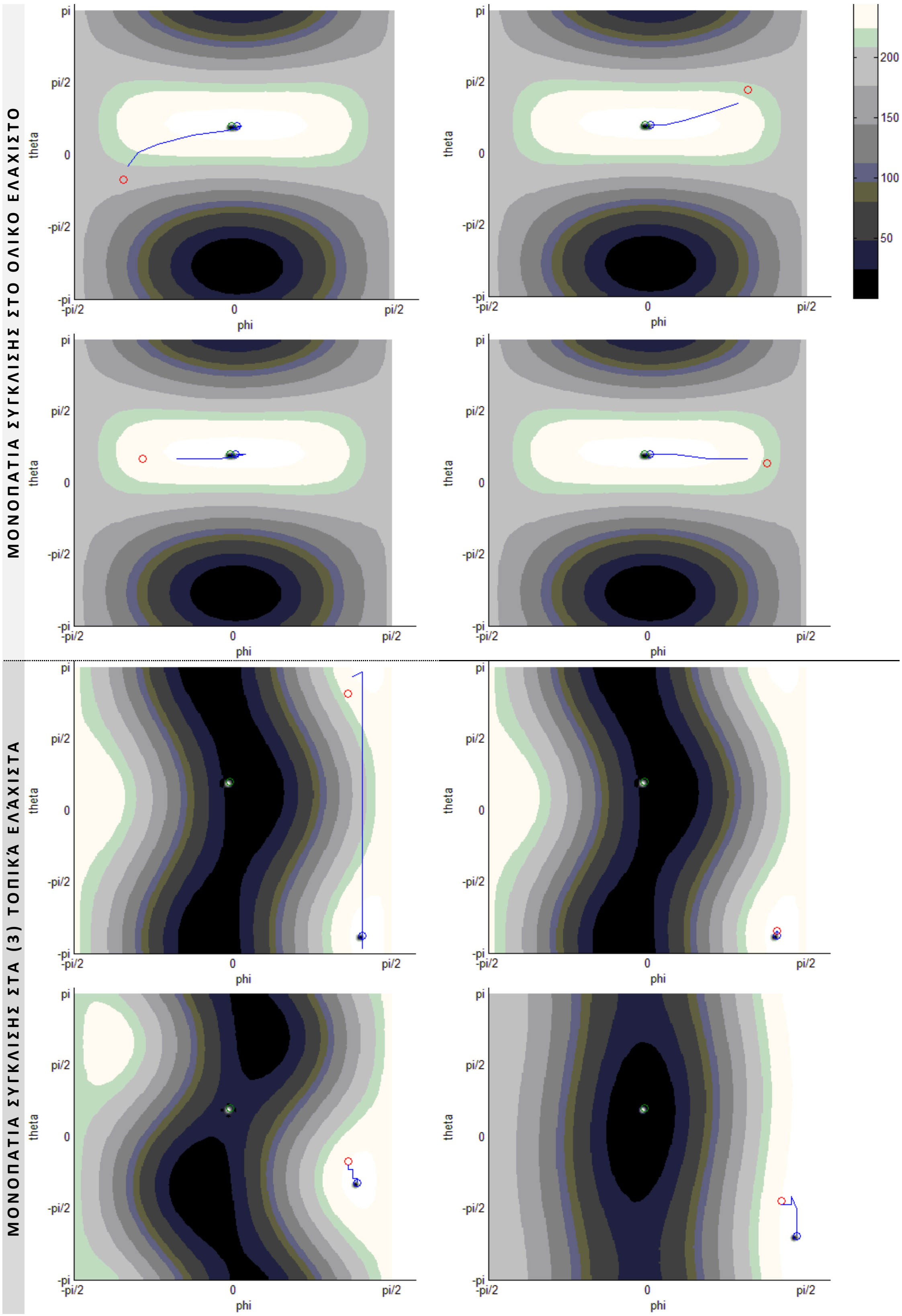
· Σχήμα συνάρτησης κόστους – Διαδικασία σύγκλισης Gauss-Newton

Στο σχήμα 5-11 βλέπουμε το σχήμα της συνάρτησης κόστους συναρτήσει του *t* για καθένα από τα ελάχιστα που προκύπτουν από την βελτιστοποίηση με είσοδο την οπτική ροή του σχήματος 5-10. Καθώς μόνο το *t* είναι ελεύθερο να μεταβάλλεται, για τα *ω* και *d* χρησιμοποιούμε τις τελικές τιμές που υπολόγισε η ΕΚΚ για καθένα από τα ελάχιστα. Το σφάλμα έχει διατιμηθεί από 0 (μαύρο) έως 255 (λευκό) για την *μέγιστη* και *ελάχιστη* τιμή του αντίστοιχα. Συνεπώς οι «λευκές περιοχές» συμβολίζουν τις περιοχές στις οποίες η διαδικασία βελτιστοποίησης θα πρέπει να συγκλίνει. Το λευκό στίγμα που βρίσκεται (περίπου) στο κέντρο (όλων) των γραφημάτων αντιστοιχεί στην πραγματική λύση t_{true}. Τα μαύρα στίγματα που συναντώνται περίπου στο κέντρο βάρους των λευκών περιοχών είναι οι εκτιμώμενες t_{est}. Το πρώτο γράφημα είναι το ολικό ελάχιστο όπου τα δυο στίγματα συμπίπτουν.

Εμβαθύνοντας στην διαδικασία σύγκλισης, το σχήμα 5-12 παρουσιάζει 8 από τα 15 «μονοπάτια» σύγκλισης που προέκυψαν στο σύνολο των τεσσάρων ελαχίστων του προηγούμενου παραδείγματος. Οι t_{init} σημειώνονται με κόκκινους κύκλους και οι t_{est} με μπλε κύκλους. Παρατηρούμε δε μία τυπική ιδιότητα των βελτιστοποιήσεων: ο αλγόριθμος συγκλίνει σε ελάχιστα που βρίσκονται κοντά στις αρχικές τιμές του, γι' αυτό και θα πρέπει να δίνεται έμφαση στην ομοιόμορφη κατανομή των t_{init}. Αδιαμφισβήτητα ο Gauss-Newton φτάνει εξαιρετικά γρήγορα κοντά στην σύγκλιση σε περίπου 5 έως 15 επαναλήψεις. Για να επιτύχουμε ακριβείς εκτιμήσεις αλλά και να διευκολυνθούμε στην ομαδοποίηση των ελαχίστων χρησιμοποιήσαμε ως κατώφλι σύγκλισης την ιδιαιτέρως αυστηρή συνθήκη $\|\Delta t\| < 10^{-13}$. Με αυτό το κατώφλι ο μέσος όρος επαναλήψεων ήταν 105 για μέτρια επίπεδα θορύβου.

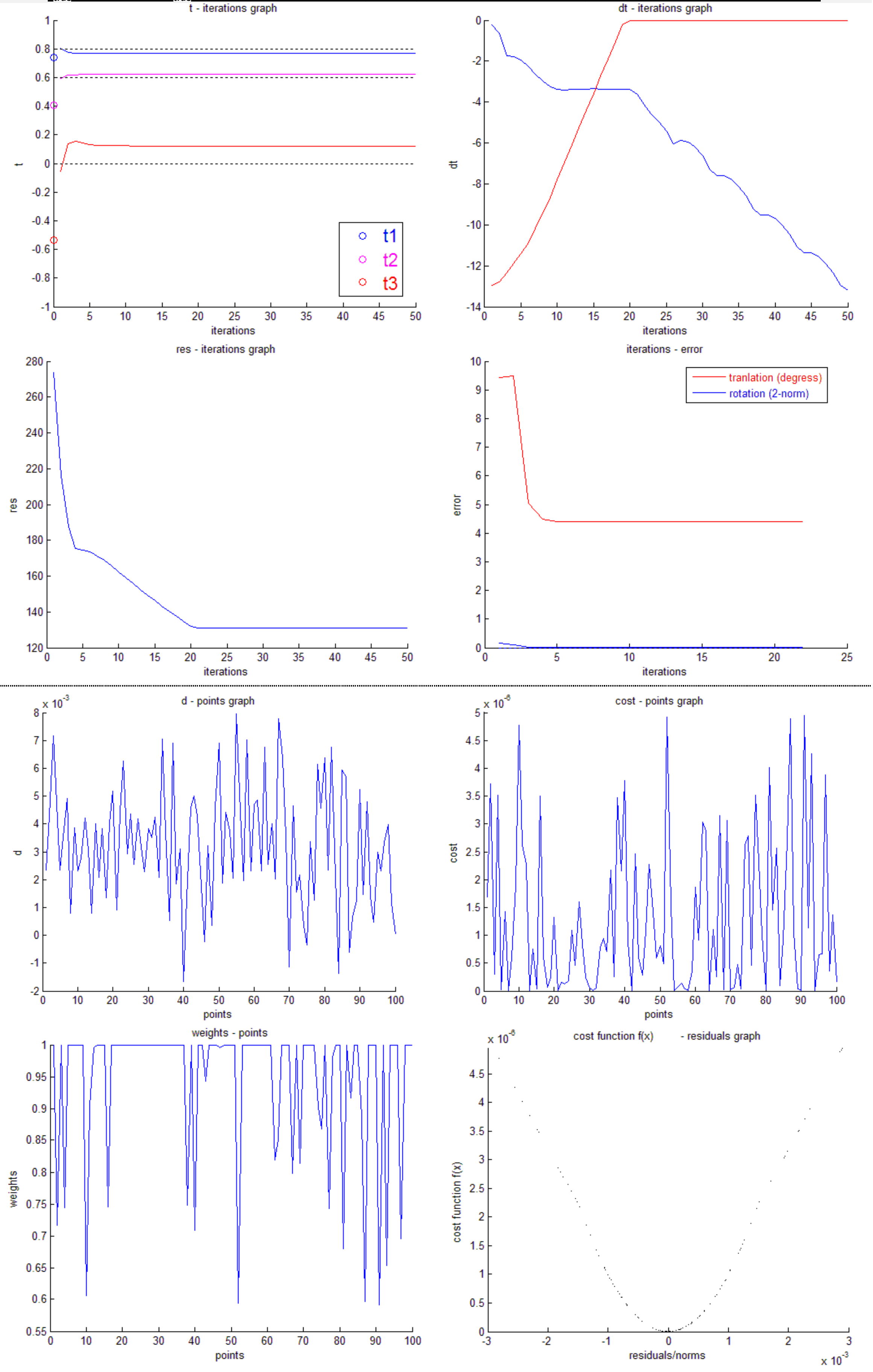
Το σχήμα 5-13 παρουσιάζει τις τιμές που λαμβάνουν διάφορες παράμετροι κατά την βελτιστοποίηση για την περίπτωση που αντιστοιχεί στο πάνω αριστερό γράφημα του σχήματος 5-12. Οι υπολογισμοί πραγματοποιήθηκαν με την *σταθμισμένη Gauss-Newton* με *«μείωση ελαχίστων»*.

Στο γράφημα των ανανεώσεων Δ*t* η κόκκινη γραμμή παρουσιάζει την παράμετρο ρ που περιγράψαμε στην ενότητα 4.5 και είναι υπεύθυνη για τον έλεγχο του αριθμού των ελαχίστων. Μετά την οριστική σταθεροποίηση της ρ παρατηρούμε πως η Δ*t* συνεχίζει να «ταλαντεύεται» (κάτι που δεν παρατηρείται στην μη-σταθμισμένη υλοποίηση) πιθανώς επειδή η «πορεία» προς την σύγκλιση επηρεάζεται από τις αλλαγές στα βάρη των διανυσμάτων. Φαίνεται ότι αναλόγως της τρέχουσας στάθμισης, αλλάζει ελαφρώς η εκτιμώμενη θέση για το ολικό ελάχιστο και επομένως μεταβάλλονται οι ανανεώσεις Δ*t* που οδηγούν σε αυτό. Ως συμπέρασμα για τον συνδυασμό «μείωσης ελαχίστων» και στάθμισης μάλλον θα πρέπει να αποφεύγεται αφού δεν προσφέρει ιδιαίτερο κέρδος συν ότι καθυστερεί την σύγκλιση.

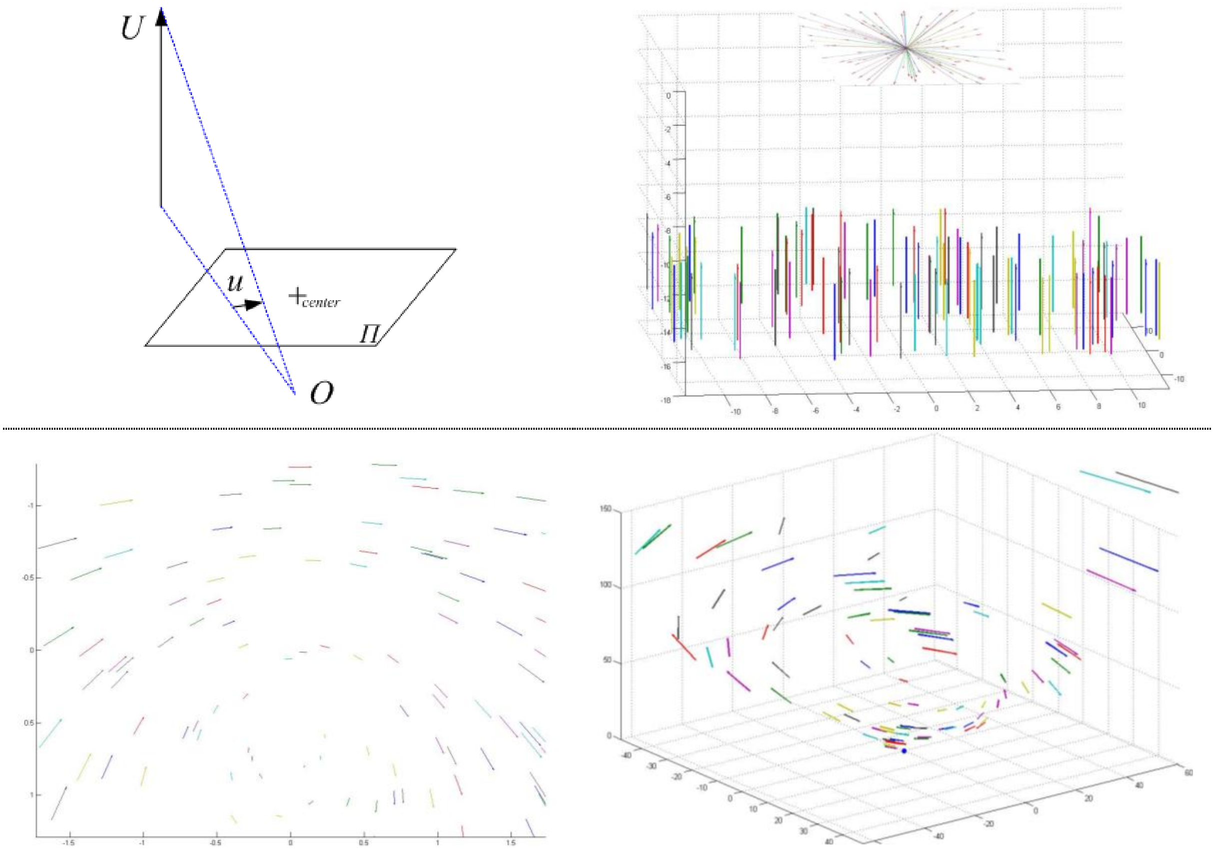


Σχήμα 5-12: *«Ανάγλυφο» κόστους συναρτήσει του t για καθένα από τα 4 ελάχιστα. Οι γραμμές δείχνουν το «μονοπάτι» από την αρχική τιμή (κόκκινος κύκλος) ως την σύγκλιση (μπλε κύκλος). Το λευκό στίγμα κοντά στην μέση των γραφημάτων είναι η πραγματική λύση* t_{true} .

t_{true} = [0.8 0.6 0]^T ω_{true} = [0 0 0]^T fov = 50°, Από αρχική τιμή: t = [0,76 0,41 -0,55]^T, με στάθμιση και μείωση ελαχίστων



Σχήμα 5-13 (Προηγούμενη σελίδα): *Η μεταφορά t, οι ανανεώσεις Δt, το ολικό κόστος και τα ω_{error}-t_{error} συναρτήσει των επαναλήψεων (4 πάνω γραφήματα) και το αντίστροφο βάθος d, οι στάθμες w_i και το κόστος για καθένα από τα 100 διανύσματα εισόδου (4 κάτω γραφήματα). Μετά την σύγκλιση βλέπουμε το κόστος κάθε διανύσματος συναρτήσει του υπολοίπου – της απόκλισης από το μοντέλο.*



Σχήμα 5-14: *Κάθε διάνυσμα οπτικής ροής προέρχεται από την προβολή των διανυσμάτων τρισδιάστατης κίνησης των αντικειμένων στο επίπεδο της εικόνας. (πάνω) Από οποιοδήποτε πεδίο οπτικής ροής μετά την εκτίμηση των t,ω,d μπορούμε και πάλι να συνθέσουμε την τρισδιάστατη κίνηση των αντικειμένων που το προκάλεσε. (κάτω)*

· *Τρισδιάστατη αναπαράσταση οπτικής ροής*

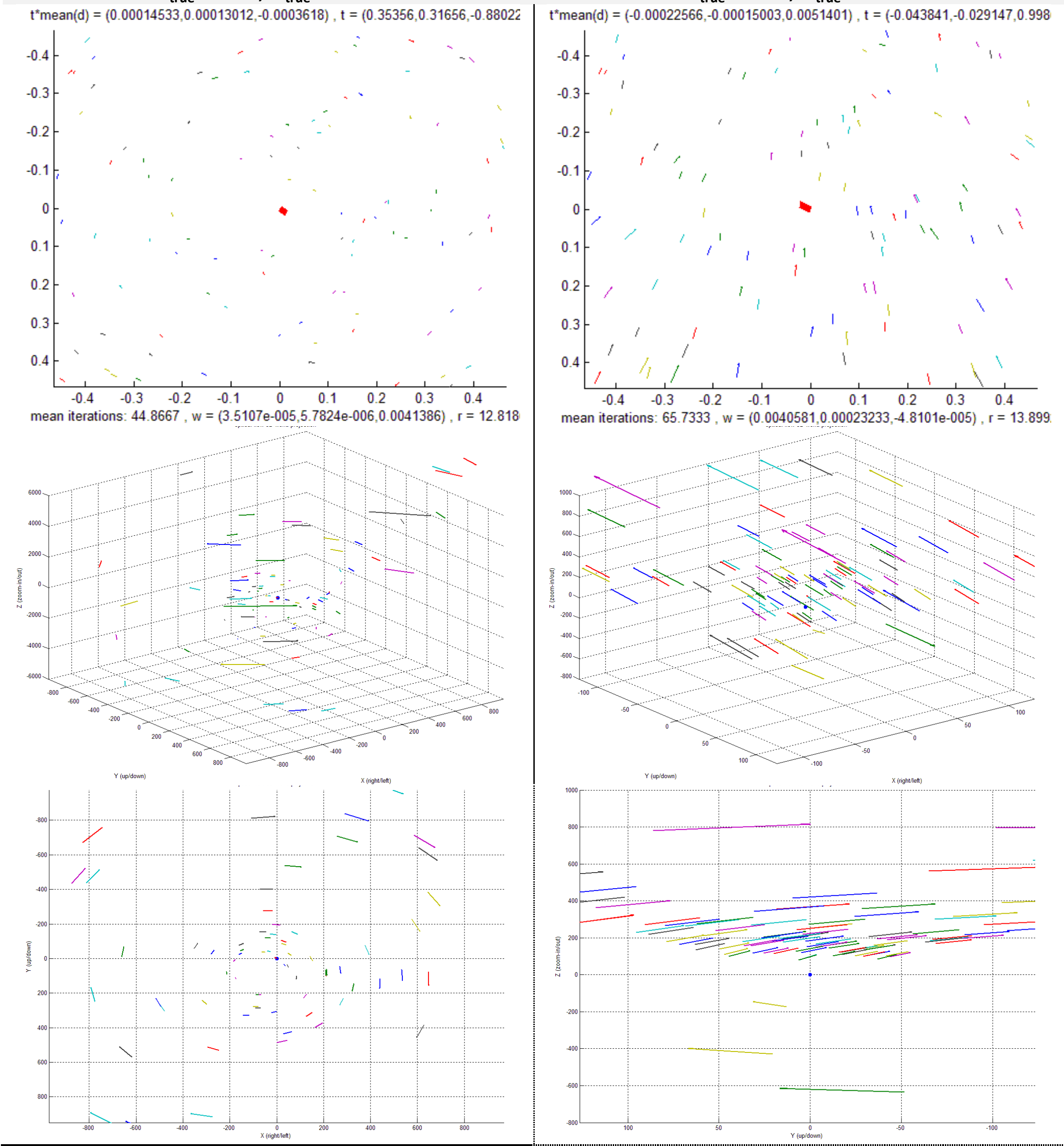
Όπως σημειώσαμε στο κεφάλαιο 3 το πεδίο ροής προκαλείται από την προβολή των διανυσμάτων τρισδιάστατης κίνησης των αντικειμένων πάνω στο οπτικό πλάνο. Έτσι γνωρίζοντας τα *t,ω,d* μπορούμε όχι μόνο να τυπώσουμε το πεδίο ροής αλλά να ανασυνθέσουμε τα πραγματικά τρισδιάστατα διανύσματα κίνησης που το προκάλεσαν (σχήμα 5-14).

Στο σχήμα 5-15 (και 5-15b) βλέπουμε την (σχετική ως προς την κάμερα) τρισδιάστατη κίνηση των αντικειμένων με καθένα απ’ τα διανύσματα i να απέχει Z_i από την κάμερα. Σκοπός της τρισδιάστατης απεικόνισης των διανυσμάτων είναι να δούμε παραστατικά την αλλοίωση που επιφέρει ο θόρυβος στα οπτικά χαρακτηριστικά του μοντέλου. Παρατηρούμε ότι κάποια διανύσματα τοποθετούνται πίσω από το οπτικό πλάνο ή ακόμα και πίσω από την σημειακή οπή (που στο σχήμα αναπαριστάται ως μπλε στίγμα). Η αιτία είναι τα αρνητικά d που προκύπτουν λόγω του θορύβου και των ανεξάρτητων κινήσεων όπως περιγράψαμε πριν.

Μια ακόμα διαταραχή που επιφέρει ο θόρυβος είναι η διαστρέβλωση του σημείου επέκτασης. Κάτι που παρατηρείται ιδιαιτέρως στις μεταφορές και περιστροφές κατά τον οπτικό άξονα Z όπου τα διανύσματα δεν είναι διατεταγμένα περιμετρικά γύρω από το κέντρο του πλάνου, όπως οφείλεται, αλλά αποκλίνουν με τυχαίο τρόπο που καθορίζεται από την θέση των εσφαλμένων διανυσμάτων.

ΤΡΙΣΔΙΑΣΤΑΤΗ ΣΥΝΘΕΣΗ ΟΠΤΙΚΗΣ ΡΟΗΣ

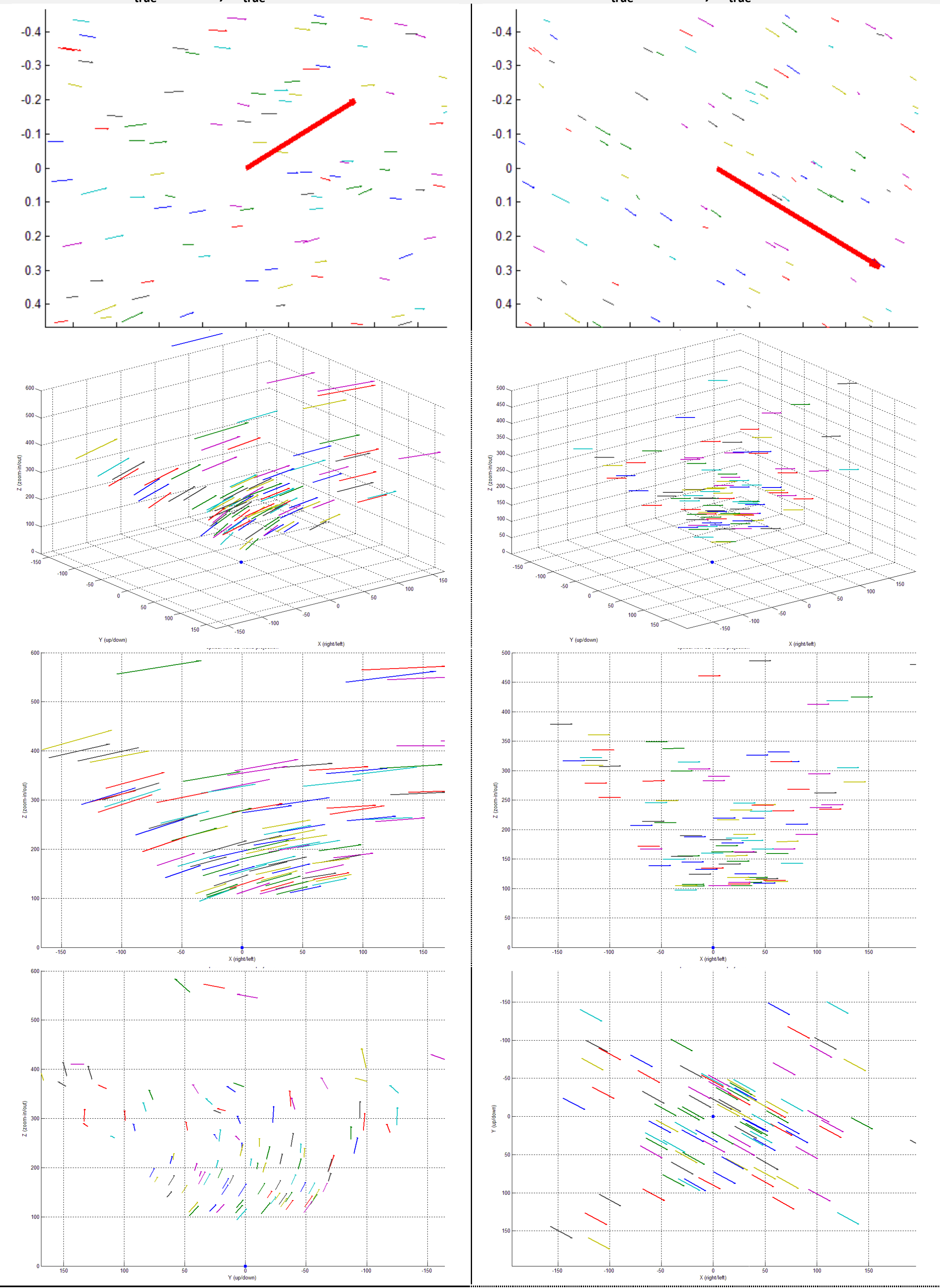
t_{true} = 0 0 0, ω_{true} = 0 0 1	t_{true} = 0 0 1, ω_{true} = 1 0 0



Σχήμα 5-15: *Στις «τέλειες» περιστροφές κατά τον οπτικό άξονα τα διανύσματα τοποθετούνται περιμετρικά γύρω από το σημείο επέκτασης (κέντρο της εικόνας). Η προβολή στο επίπεδο X-Y αποκαλύπτει ότι αυτό έχει μετακινηθεί από το κέντρο (αριστερή στήλη). Μερικά από τα διανύσματα, λόγω αρνητικών τιμών, τοποθετούνται πίσω από το οπτικό πλάνο όπως φαίνεται από την προβολή στο οριζόντιο επίπεδο X-Z (δεξιά στήλη).*

ΤΡΙΣΔΙΑΣΤΑΤΗ ΣΥΝΘΕΣΗ ΟΠΤΙΚΗΣ ΡΟΗΣ (συνέχεια)

t_{true} = 4 -3 5, ω_{true} = -1 2 0.5	t_{true} = 2 1.5 0, ω_{true} = 0 0 0



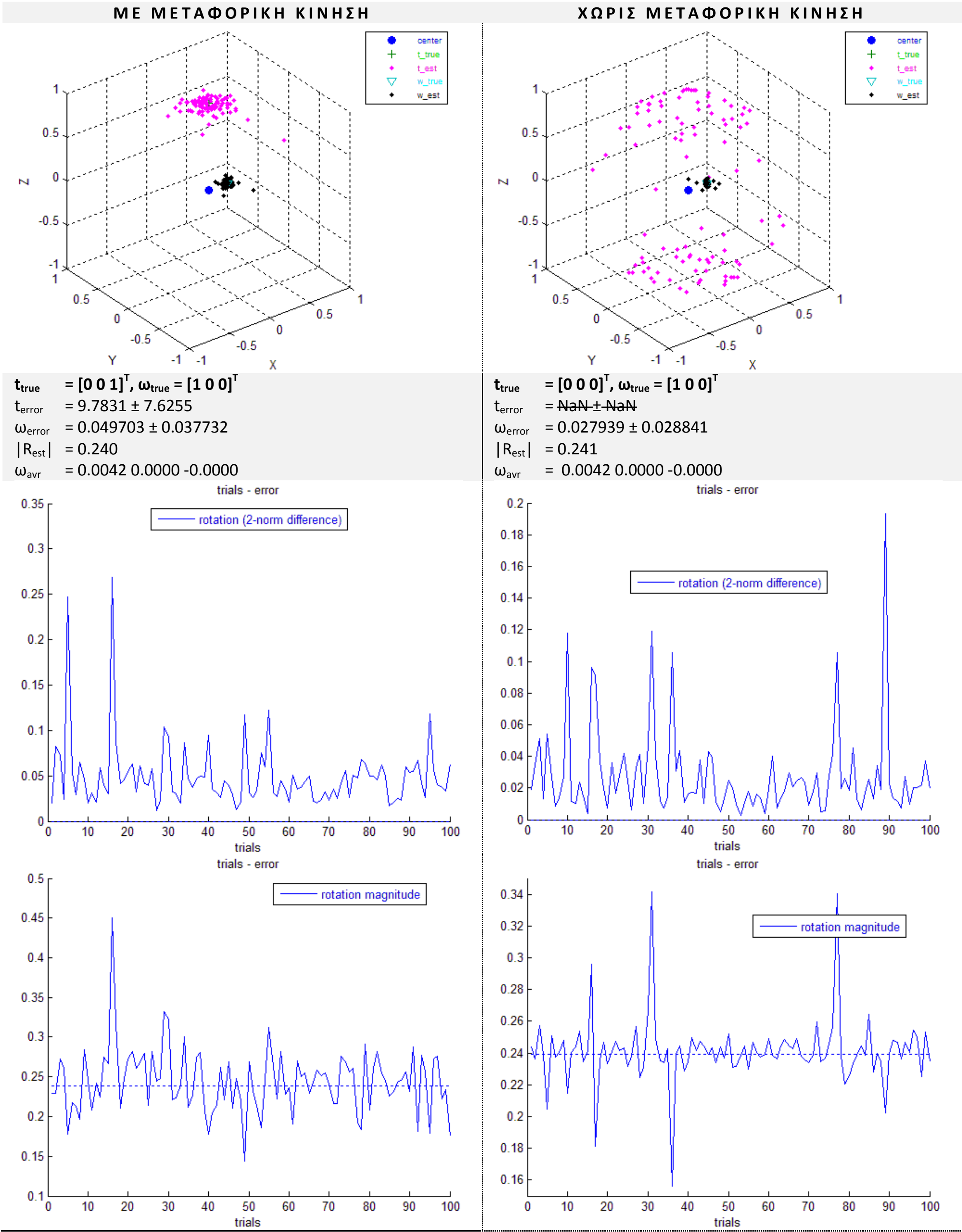
Σχήμα 5-15b: *Στις πραγματικές ακολουθίες οι κατευθύνσεις γειτονικών διανυσμάτων ενδέχεται να σχετίζονται επειδή π.χ. αναφέρονται στο ίδιο αντικείμενο. Τα τρισδιάστατα διανύσματα εδώ συγκροτούν ένα πολύπλοκο πεδίο γιατί τα διανύσματα που παράγονται στην συνθετική οπτική ροή είναι ανεξάρτητα.*

· *Ταχύτητα μεταφορικής κίνησης και περίπτωση μηδενικής μεταφοράς*

Επανειλημμένως σχολιάσαμε την αναγκαιότητα θέσπισης περιορισμού για το διάνυσμα μεταφοράς σε $\|t\| = 1$. Εντούτοις αυτό συνεπάγεται πως οι εκτιμήσεις περιλαμβάνουν υποχρεωτικά μεταφορική κίνηση και μάλιστα σταθερής ταχύτητας. Για το θέμα της ταχύτητας ίσως εκ των υστερών μπορούν να γίνουν κάποιες εκτιμήσεις αν συνδυάσουμε επιπρόσθετα στοιχεία όπως την ταχύτητα περιστροφής ή το μέσο αντίστροφο βάθος d_{mid}. Για παράδειγμα μικρό μέσο αντίστροφο βάθος ενδέχεται να σημαίνει μικρή ταχύτητα μεταφοράς. Ακόμα η ταχύτητα περιστροφής θα μπορούσε να καταστεί χρήσιμη με τον ίδιο τρόπο όπως αξιοποιήθηκε στην σύνθεση οπτικής ροής (ενότητα 5.2.3).

Ωστόσο παρά τις ενδείξεις που προσφέρουν αυτά τα στοιχεία σε κάθε περίπτωση μπορεί να έχουν «διπλή (ή τριπλή!) ανάγνωση» και οποιαδήποτε συμπεράσματα είναι επισφαλή: Π.χ. στην αμιγώς περιστροφική κίνηση (απουσία μεταφορικής κίνησης) το d_{mid} είναι μηδενικό κάτι που όμως ισχύει επίσης όταν τα αντικείμενα στο περιβάλλον λήψης είναι μακριά ή όταν λόγω θορύβου παράγονται πολλά αρνητικά d. Με αυτά τα δεδομένα δεν έχουμε την ευχέρεια να κρίνουμε αν μία μικρή τιμή d_{mid} οφείλεται σε απουσία μεταφορικής κίνησης ή σε θόρυβο και θα ήταν ριψοκίνδυνο να βγάλουμε συμπεράσματα. Διαφαίνεται ότι δεν υπάρχει η δυνατότητα να προβλέψουμε την απουσία μεταφοράς, ωστόσο η επόμενη δοκιμή αποδεικνύει ότι τουλάχιστον η περιστροφή δεν επηρεάζεται από το γεγονός αυτό. Ο έλεγχος περιλαμβάνει 2 δοκιμές (των 100 trials) με τον διαχωρισμένο Gauss-Newton. Η πρώτη δοκιμή αναπαριστά την κίνηση $t = (0\ 0\ 1)^T\ /\ \omega = (1\ 0\ 0)^T$ και η δεύτερη την $t = (0\ 0\ 0)^T\ /\ \omega = (1\ 0\ 0)^T$ (σχήμα 5-16).

Οι οπτικές παράμετροι των δοκιμών είναι όμοιες με αυτές του σχήματος 5-4 πείραμα 1. Απουσία μεταφοράς το εκτιμώμενο t_{est} λαμβάνει απρόβλεπτες τιμές γι' αυτό αγνοούμε το σφάλμα του και εστιάζουμε στην περιστροφή. Παρατηρούμε ότι η πόλωση, οι αποκλίσεις και το μέτρο της περιστροφής εμφανίζουν μικρές διαφορές ανάμεσα στις δύο δοκιμές με και χωρίς μεταφορά (σχήμα 5-16).

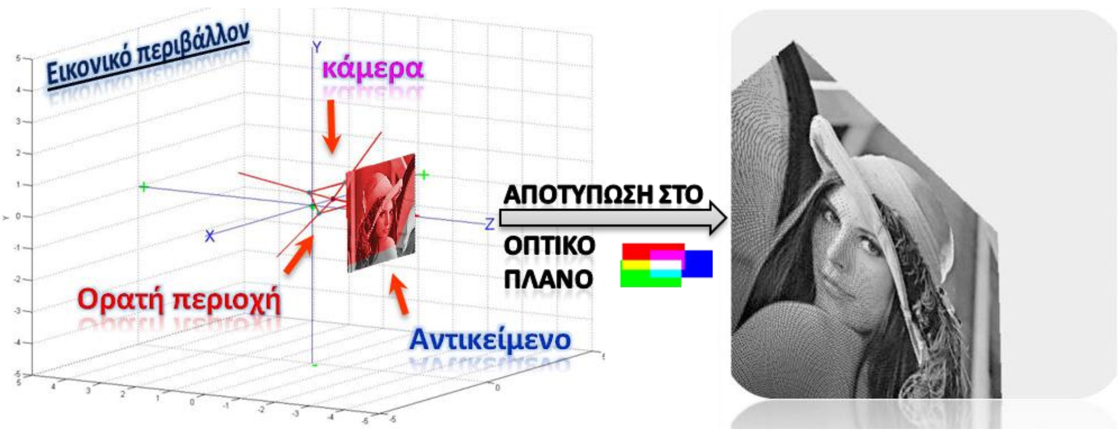


Σχήμα 5-16: *Η περιστροφική κίνηση επηρεάζεται ανεπαίσθητα από την ύπαρξη ή όχι μεταφοράς.*

5.4 Δοκιμές με πραγματικές ακολουθίες

5.4.1 Εισαγωγή

Στις δοκιμές με πραγματικές ακολουθίες εικόνων ενδεχομένως να λείπει το στοιχείο της αντικειμενικότητας αφού δεν διαθέτουμε καταγεγραμμένες μετρήσεις για την κίνηση ώστε να έχουμε δυνατότητα συγκρίσεων και αξιολόγησης. Ωστόσο επιστρατεύοντας τις θαυμαστές δυνατότητες της ανθρώπινης αντίληψης στην ερμηνεία της κίνησης και του περιβάλλοντος μπορούμε να σχηματίσουμε μια χονδρική πλην όμως αξιόλογη αίσθηση του αποτελέσματος της ΕΚΚ από την παρακολούθηση των ακολουθιών εικόνων, ειδικά αν αυτές απεικονίζουν προσφιλή μας αντικείμενα ή φέρουν κάποιο χαρακτηριστικό γνώρισμα-ένδειξη. Για να ενισχύσουμε την αίσθηση αυτή δημιουργήσαμε και ενσωματώσαμε στην υλοποίηση μας μια «μηχανή» δημιουργίας «σημείων» αναφοράς που ακολουθούν την κίνηση του περιβάλλοντος. Η **τρισδιάστατη μηχανή** δέχεται ως είσοδο τις παραμέτρους *t, ω, d* που παράγει η ΕΚΚ και τυπώνει τα σημεία αυτά υπερθετικά στις εικόνες της ακολουθίας. Η σχετική κίνηση των σημείων ως προς το περιβάλλον είναι ενδεικτικό της ακρίβειας των μετρήσεων που παράγονται.



Σχήμα 5-18: *Η τρισδιάστατη μηχανή προσομοιώνει τις κινήσεις της φυσικής κάμερας. Εδώ η εικονική κάμερα έχει στραφεί προς την γωνία λήψης του στιγμιοτύπου. Η κόκκινη απόχρωση στο σχέδιο συμβολίζει το περιεχόμενο που είναι ορατό στο πλάνο (αριστερά). Η εικόνα όπως προβάλλεται στο οπτικό πλάνο της κάμερας (δεξιά).*

5.4.2 Τρισδιάστατη μηχανή

Στο τέλος κάθε κύκλου επεξεργασίας δυο ιδιαιτέρως διαφωτιστικά γραφήματα οπτικοποιούν την κίνηση της κάμερας μέσω ενός απλού μοντέλου γραφικών το οποίο αποτελείται από το τρισδιάστατο *εικονικό περιβάλλον* με τους άξονες *XYZ* διατιμημένους σε *FLU*, την *εικονική κάμερα* που μπορεί να κινηθεί στο περιβάλλον προς κάθε δυνατή κατεύθυνση και *εικονικά σημεία* στον χώρο (σχήμα 5-18).

Τα εικονικά σημεία είναι τοποθετημένα σε διάταξη σφαίρας με ακτίνα 5 *FLU* ενώ ανά δυο διαδοχικά σχηματίζουν γωνία 10°. Η εικονική κάμερα, η οποία αρχικά τοποθετείται στο κέντρο της σφαίρας και έχει στραμμένο τον οπτικό της άξονα στο Z^+, έχει την δυνατότητα να «κινηματογραφεί» τα εικονικά σημεία και να τα αποτυπώνει στο οπτικό της πλάνο.

Με αυτή την δομή η τρισδιάστατη μηχανή προσφέρει ενδιαφέρουσες λεπτομέρειες για την εκτιμώμενη κίνηση, επιτρέποντας μας να σχηματίσουμε μια ολοκληρωμένη εικόνα μέσω της αναπαραγωγής της από δυο οπτικές γωνίες. Στην πρώτη από αυτές, την *«πρώτου προσώπου»*, το οπτικό πλάνο της εικονικής κάμερας ενσωματώνεται στις πραγματικές ακολουθίες εικόνων έτσι ώστε τα εικονικά σημεία να ακολουθούν χονδρικά σημεία του πραγματικού κόσμου. Στις λήψεις αυτές τα σημεία αναπαριστώνται ως αριθμοί που αντιπροσωπεύουν τις σφαιρικές συντεταγμένες των ιδίων. Για να κατορθώσουμε επιτυχημένο συντονισμό μεταξύ της φυσικής και της εικονικής κάμερας φροντίζουμε ώστε οι οπτικές παράμετροι τους, όπως το οπτικό πεδίο, να είναι συμβατοί. Το δεύτερο γράφημα παρουσιάζει ολόκληρο το εικονικό περιβάλλον από οπτική «τρίτου προσώπου». Τα σημεία που είναι ορατά στο οπτικό πεδίο της κάμερας παρουσιάζονται ως κόκκινες τελείες ενώ τα υπόλοιπα ως μικρότερες τελείες μωβ χρώματος. Στο σχήμα 5-22 βλέπουμε ένα στιγμιότυπο από την έξοδο της μηχανής τρισδιάστατων γραφικών που υλοποιήσαμε σε ***Matlab 7.1*** στα πλαίσια αυτής της εργασίας.

5.4.3 Επεξεργασία με πραγματικές ακολουθίες

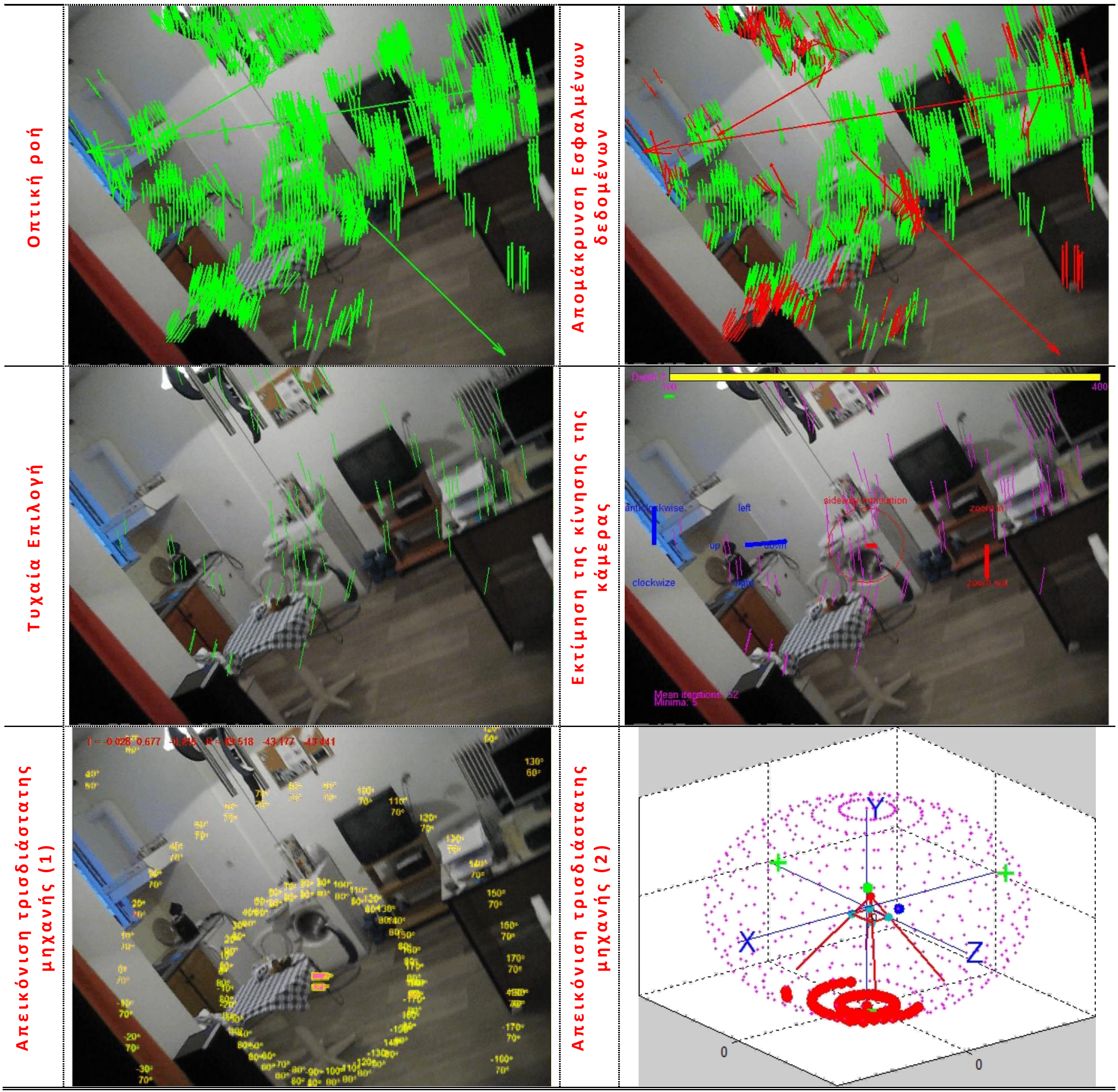
Για την επεξεργασία πραγματικών ακολουθιών το πρόβλημα της ΕΚΚ χωρίστηκε σε πέντε κύρια στάδια επεξεργασίας. Το σχήμα 5-19 παρουσιάζει την σειρά διασύνδεσης των επιμέρους στοιχείων προκειμένου να επεξεργαστούμε πραγματικές ακολουθίες εικόνων. Οι υλοποιήσεις των επιμέρους στοιχείων καθώς και οι κώδικες για τις δόκιμες αναπτύχθηκαν στο περιβάλλον ανάπτυξης της ***Matlab 7.1***.

Σχήμα 5-19: *Η διαδικασία για την επεξεργασία πραγματικών ακολουθιών εικόνων.*

Το σχήμα 5-20a,b,c παρουσιάζει ενδεικτικά στιγμιότυπα από την επεξεργασία για τρεις ακολουθίες εικόνων. Οι παράμετροι που παράγει η ΕΚΚ απεικονίζονται με την μορφή που παρουσιάζει το σχήμα 5-23. Οι ακολουθίες εικόνων δημιουργήθηκαν με γνώμονα να συμπεριλάβουμε στις δοκιμές μια σειρά από βασικές περιπτώσεις όπως «καθαρές» μεταφορές – περιστροφές, συνδυασμό τους και πλήρη τρισδιάστατη κίνηση. Ο χώρος λήψης ήταν ένας τετράγωνος χώρος δωματίου 30m^2 ενώ οι ακολουθίες ελήφθησαν με την κάμερα *Canon digital Ixus 75* σε ανάλυση 640x480*p* και κωδικοποίηση βίντεο MPEG1. Ο θόρυβος που παράγεται από την κωδικοποίηση MPEG1 έχει την μορφή ψηφιακών τεχνουργημάτων (*artifacts*) κάτι που ίσως κάνει δυσκολότερη την εξομάλυνση του (έχει προκληθεί από μη-αντιστρέψιμη συμπίεση πληροφορίας). Οι διαστάσεις του CCD είναι 6.6x4.4*mm* και η απόσταση από τον φακό 1.2*mm* οδηγώντας σε οπτικό πεδίο 56°. Οι παράμετροι αυτοί εισήχθησαν στον αλγόριθμο ΕΚΚ και στην *τρισδιάστατη μηχανή* ώστε να γίνουν σωστά οι μετασχηματισμοί σε *FLU*. Στις δόκιμες συμπεριλάβαμε και μια ακολουθία γραφικών *VRML* που καταγράψαμε με τον *Octaga Player v2.1* (σχήμα 5-21). Στην περίπτωση των γραφικών τα αποτελέσματα ήταν απογοητευτικά λόγω των τεχνουργημάτων (artifacts) και των «ζωηρών» τετραγωνισμένων ακμών. Το Gaussian φίλτρο δεν έχει την δυνατότητα να ομαλοποιήσει επαρκώς τις ατέλειες αυτές, με αποτέλεσμα εξαιρετικά θορυβώδη οπτική ροή και ό,τι αυτό συνεπάγεται για την ΕΚΚ.

Πλήρης τρισδιάστατη κίνηση

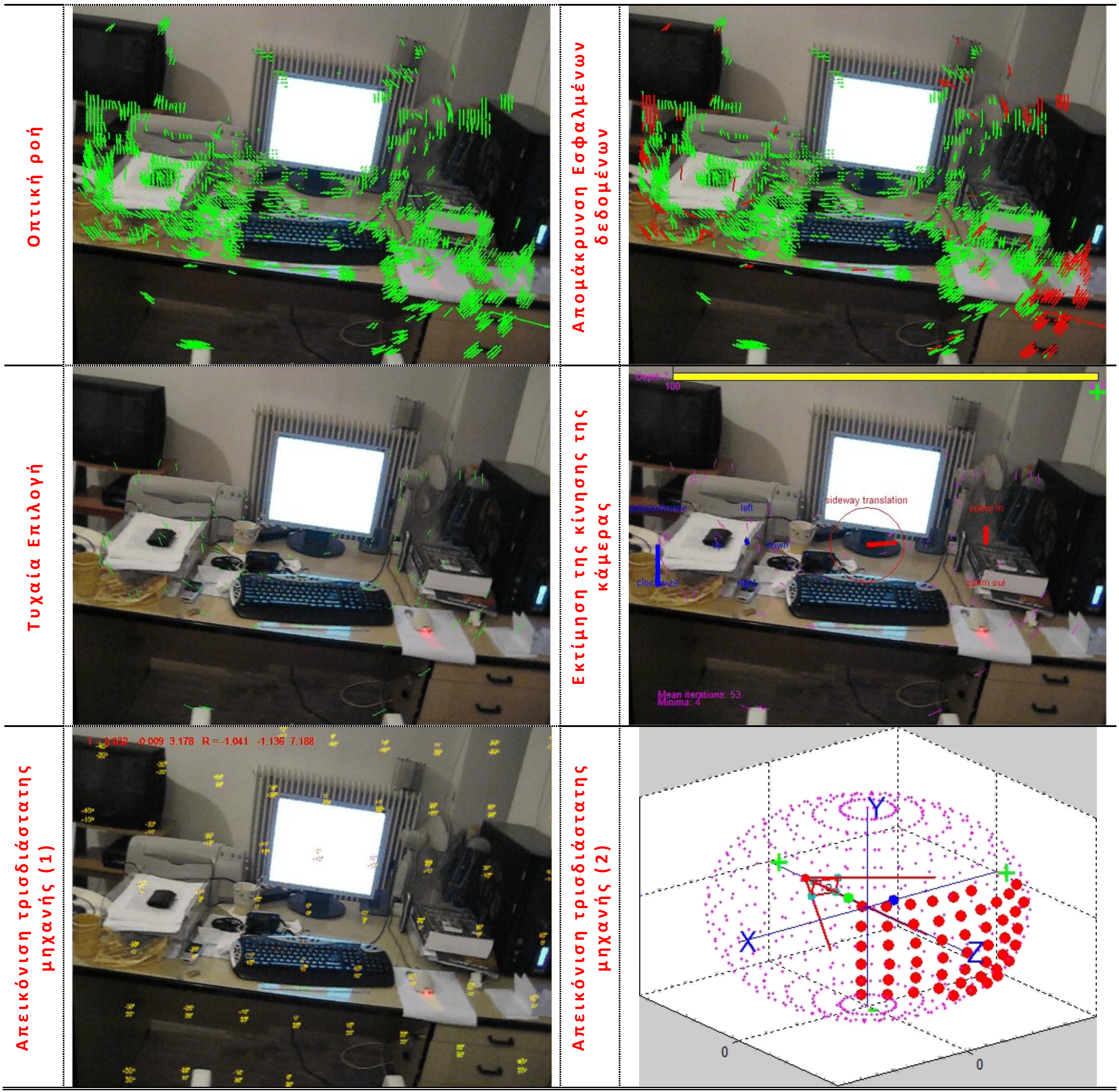
Ακολουθία No.= 04, **Τρέχων καρέ** = 525, **Συνολική διάρκεια** = 0:21 sec **ή** 632 καρέ



Σχήμα 5-20α: *Ενδεικτικά στιγμιότυπα από την επεξεργασία 5 σταδίων.*

Περιστροφική Z κίνηση

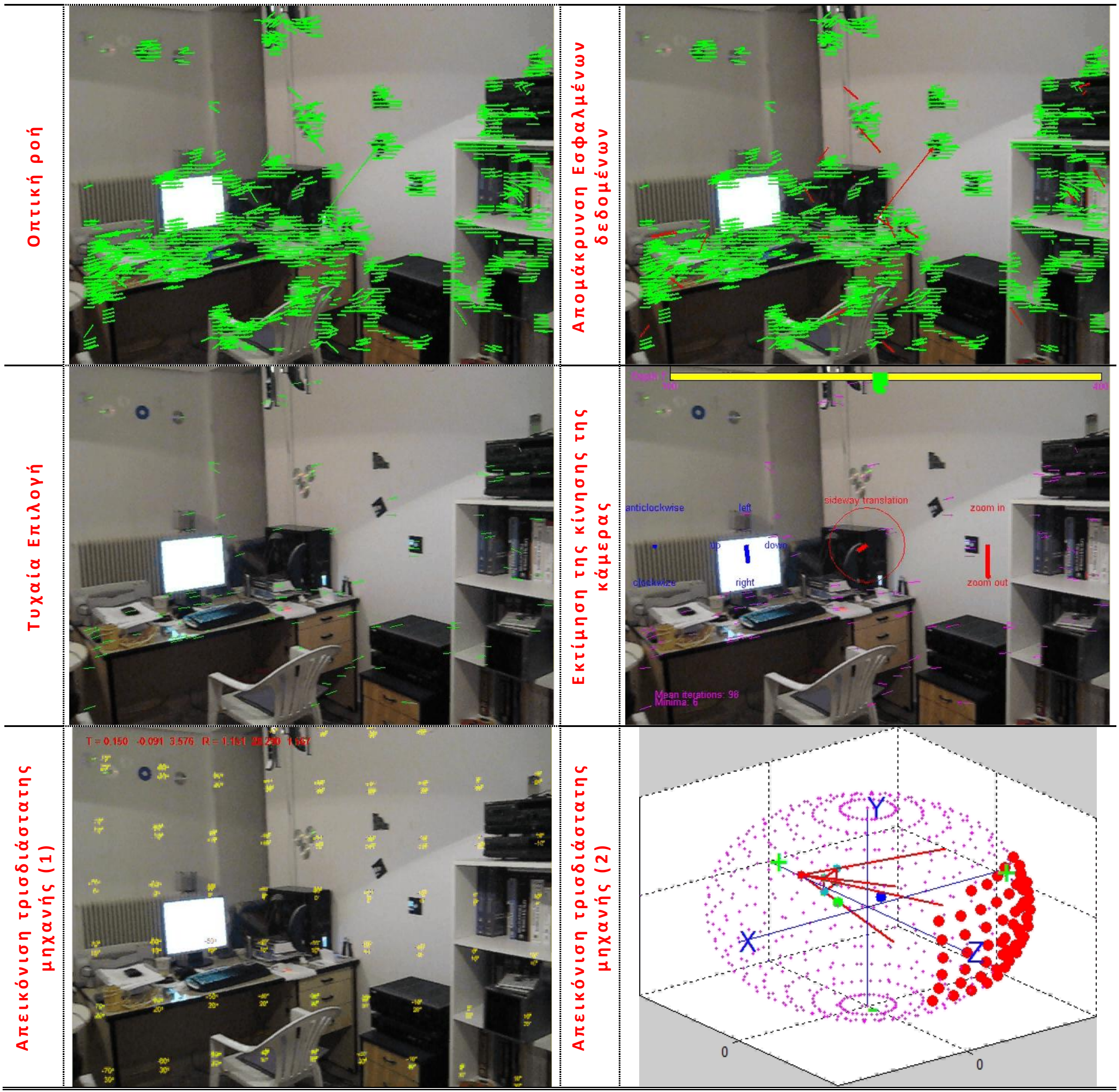
Ακολουθία No.= 42, **Τρέχων καρέ** = 17, **Συνολική διάρκεια** = 0:18 sec **ή** 469 καρέ



Σχήμα 5-20b: *Ενδεικτικά στιγμιότυπα από την επεξεργασία 5 σταδίων.*

Κυρίως περιστροφική Y κίνηση και ήπια μεταφορική X

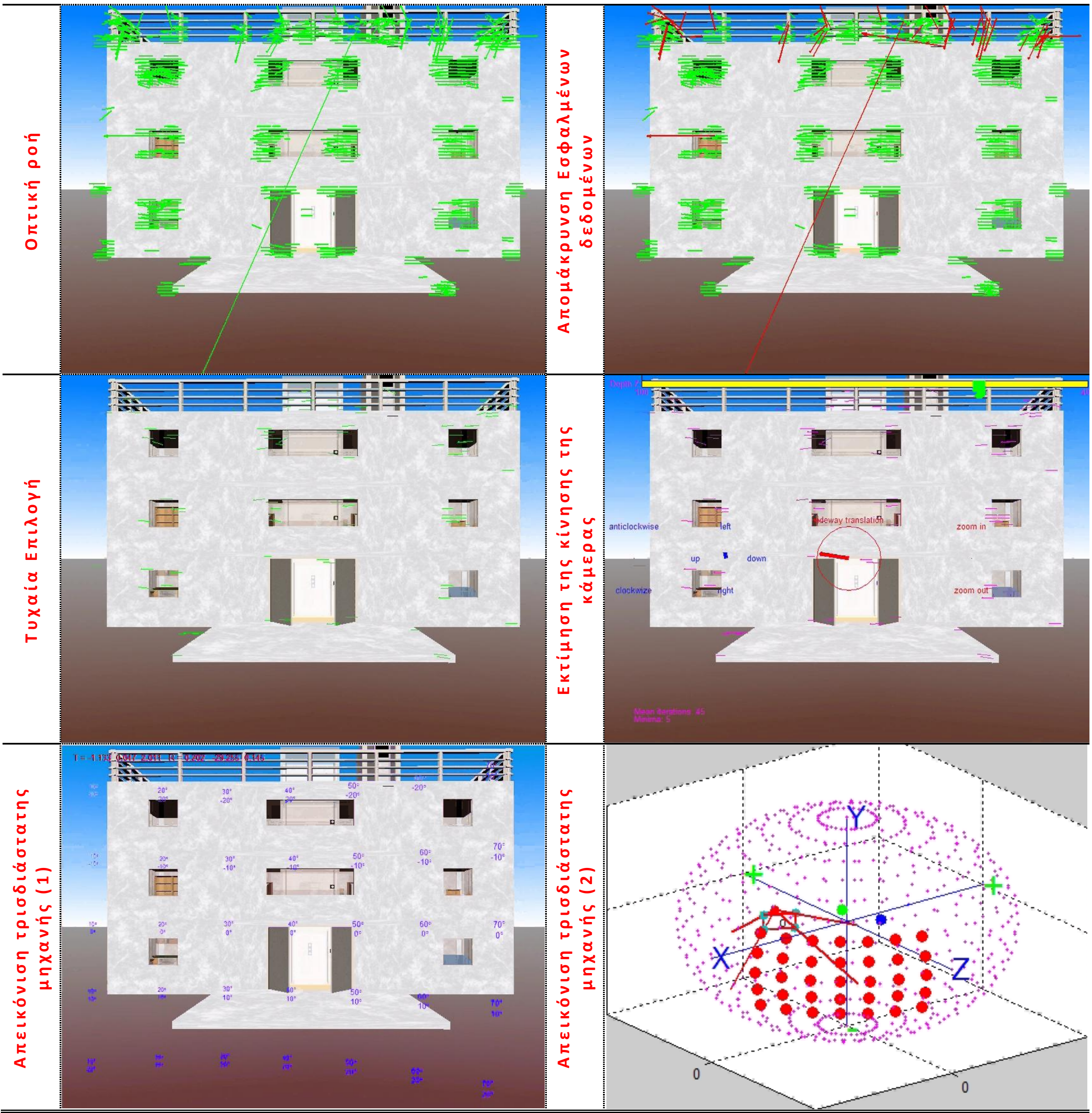
Ακολουθία No.= 43, **Τρέχων καρέ** = 90, **Συνολική διάρκεια** = 0:10 sec **ή** 258 καρέ



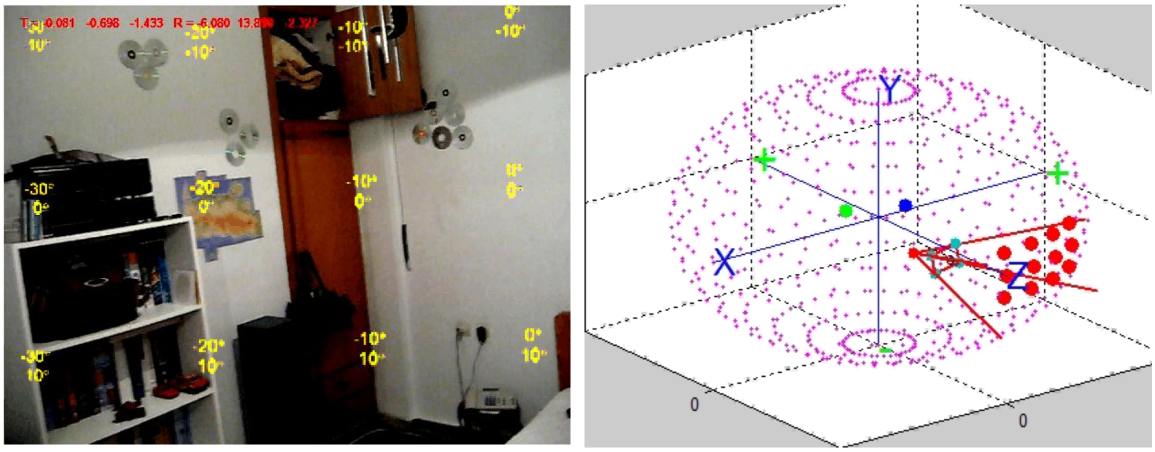
Σχήμα 5-20c: *Ενδεικτικά στιγμιότυπα από την επεξεργασία 5 σταδίων.*

Περιστροφική Υ κίνηση

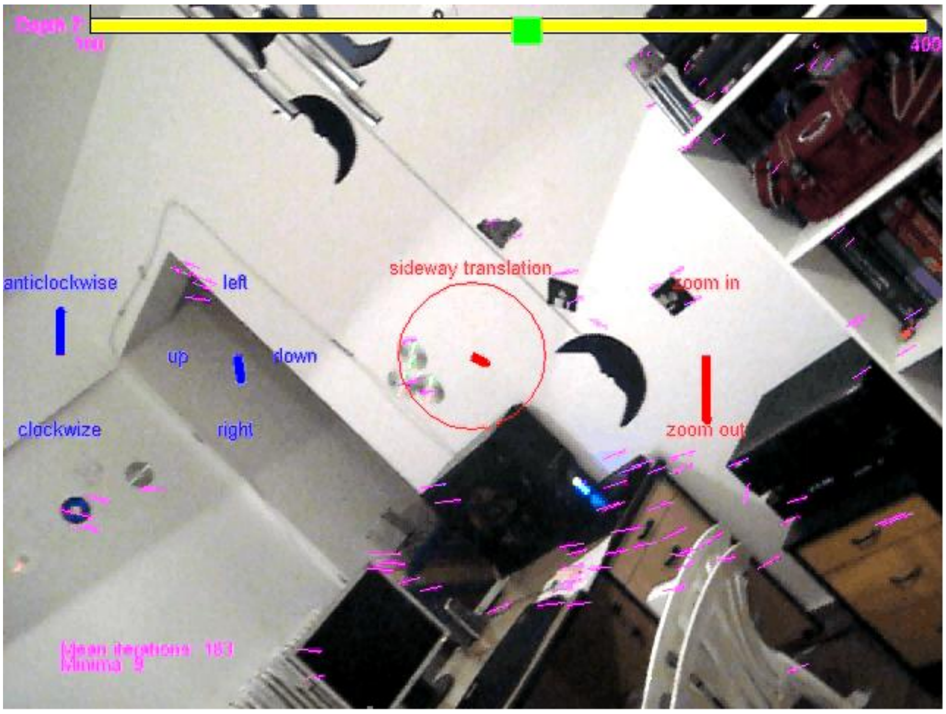
Ακολουθία. = VRML, **Τρέχων καρέ** = 148, **Συνολική διάρκεια** = 0:16 sec ή 364 καρέ



Σχήμα 5-21: *Ενδεικτικά στιγμιότυπα από την επεξεργασία με συνθετική ακολουθία VRML.*



Σχήμα 5-22: *Η τρισδιάστατη μηχανή αναπαριστά την κίνηση της κάμερας από δυο οπτικές γωνίες. Στην πρώτη τα εικονικά σημεία τυπώνονται στο οπτικό πλάνο της εικονικής κάμερας και στην συνέχεια «επικολλούνται» στις πραγματικές εικόνες (αριστερά). Στην οπτική «τρίτου προσώπου» τα σημεία αυτά εμφανίζονται ως κόκκινες τελείες ενώ μπορούμε ακόμα να δούμε την εικονική κάμερα και τα σημεία σε διάταξη σφαίρας (δεξιά).*



Σχήμα 5-23: *Αποτελέσματα από τον αλγόριθμο εκτίμησης της κίνησης. Διακρίνουμε τις παραμέτρους κίνησης: την μεταφορά (κόκκινα διανύσματα), την περιστροφή (μπλε διανύσματα), το μέσο βάθος (πράσινο ίχνος στην διατιμημένη κιτρίνη μπάρα), καθώς και τον μέσο αριθμό επαναλήψεων και τα ελάχιστα.*

Σχήμα 5-24: *Ο L-K όπως και οι περισσότεροι αλγόριθμοι εύρεσης οπτικής ροής παράγουν διανύσματα σε περιοχές της εικόνας με υφές και ακμές (αριστερά). Το αποτέλεσμα είναι η συγκέντρωση του κύριου όγκου των διανυσμάτων γύρω από αυτές (δεξιά).*

5.4.4 Παρατηρήσεις

· *Κατανομή διανυσμάτων*

Μια αιτία δυσχέρειας στην επεξεργασία με πραγματικές ακολουθίες εικόνων (σε σχέση με την συνθετική προσομοίωση της προηγούμενης ενότητας) είναι η μη ομοιόμορφη κατανομή των διανυσμάτων στην περιοχή της εικόνας και αντιθέτως η συγκέντρωση τους (*clustering*) σε συγκεκριμένες περιοχές γύρω από ακμές και υφές (σχήμα 5-24). Η αίσθηση που αποκομίσαμε είναι ότι όταν πολλά διανύσματα βρίσκονται περιφερειακά της εικόνας δημιουργείται μια ήπια «τάση» του αλγορίθμου να αποδίδει κινήσεις κατά τον οπτικό άξονα. Παραδόξως η τυχαία επιλογή που εφαρμόσαμε για την μείωση των διανυσμάτων (σε 100) είχε θετική επίδραση στο θέμα της κατάτμησης των διανυσμάτων σε περιοχές. Αιτία της βελτίωσης είναι ότι η τυχαία «εκκαθάριση» προκαλούσε «αραίωση» στην συγκέντρωση των δειγμάτων σε εκείνα τα σημεία που περιείχαν τον «κύριο όγκο» διανυσμάτων προκαλώντας κάποια ισοστάθμιση στην κατανομή.

· *Ταχύτητα μεταφορικής κίνησης*

Ένα βασικό μειονέκτημα της ΕΚΚ είναι η αδυναμία εκτίμησης της *ταχύτητας μεταφορικής κίνησης*. Ο συνυπολογισμός δεδομένων από την ταχύτητα περιστροφής ή το βάθος που εξετάσαμε στο προηγούμενο κεφάλαιο ως τρόπους αποκατάστασης της χαμένης πληροφορίας επισημάνθηκε ότι προσφέρει ενδείξεις και όχι αξιόπιστα αποτελέσματα. Πραγματοποιώντας δοκιμές με τους δυο αυτούς τρόπους το αποτέλεσμα ήταν δυσανάλογα μεγάλες διακυμάνσεις στις εκτιμήσεις της ταχύτητας μόλις διαδοχικών καρέ. Ως τελική έξοδο του ΕΚΚ αποφασίσαμε να παράγουμε μία σταθερή διαβάθμιση του *t* χωρίς να υπολογίζουμε την ταχύτητα.

Η ταχύτητα μεταφοράς αποτελεί θέμα διεξοδικότερης διερεύνησης και μια ιδέα που θα μπορούσε να εξεταστεί είναι η υιοθέτηση κάποιου προσαρμοστικού (*adaptive*) κανόνα αυξομείωσης ώστε να συνυπολογίζεται στις εκτιμήσεις το θέμα της τοπικότητας (των διαδοχικών καρέ) και παράλληλα να διατηρείται μια «ομαλότητα» κατά τις επιταχύνσεις / επιβραδύνσεις της κάμερας.

· *Γενικό συμπέρασμα*

Μια γενική αποτίμηση από τις δοκιμές είναι ότι όσο πιο πολύπλοκη ήταν η κίνηση της κάμερας τόσο καλύτερες φάνηκαν οι εκτιμήσεις, ενώ μάλλον απογοητευτικά ήταν τα αποτελέσματα στις λήψεις με μεγάλο *βάθος πεδίου* (πλήθος αντικειμένων κοντά στην κάμερα μπροστά από μακρινό φόντο) όπου το μέτρο της περιστροφής παρουσίαζε μεγάλες διακυμάνσεις κάτι που στην τρισδιάστατη απεικόνιση εμφανίσθηκε ως «τρέμουλο». Ωστόσο παρά τις όποιες ατέλειες στις περισσότερες ακολουθίες τα αποτελέσματα παρουσιάζουν εξαιρετική ακεραιότητα, εδραιώνοντας την πεποίθησή μας ότι η προτεινόμενη μέθοδος έχει μεγάλες δυνατότητες πρακτικής αξιοποίησης.

Κεφάλαιο 6

Συμπεράσματα – Σύνοψη

6.1 Σύνοψη

Η εξαγωγή τρισδιάστατων χαρακτηριστικών και ειδικότερα η ΕΚΚ από ακολουθίες εικόνων είναι ένα περίπλοκο πρόβλημα μηχανικής όρασης για το οποίο τα τελευταία χρόνια υπάρχει ιδιαίτερο ενδιαφέρον. Διάφορες προσεγγίσεις του προβλήματος που παρουσιάστηκαν στο παρελθόν παρήγαγαν απογοητευτικά αποτελέσματα έτσι ώστε πολλοί ερευνητές να πιστεύουν ότι είναι αδύνατη η χρήση της ΕΚΚ σε πρακτικές εφαρμογές. Η πλειονότητα των προσεγγίσεων αυτών ξεκινούσαν από κοινή βάση χρησιμοποιώντας τις ίδιες εξισώσεις που όμως μετασχημάτιζαν ώστε να διευκολύνουν τους υπολογισμούς, εξαλείφοντας π.χ. το βάθος ή υπολογίζοντας μια παράμετρο βάσει μιας άλλης π.χ. την μεταφορά βάσει της περιστροφής. Οι χειρισμοί αυτοί διαστρέβλωναν την συνάρτηση κόστους παράγοντας ασυνεπή μοντέλα εξισώσεων κίνησης. Όπως διαπιστώνεται κοινό χαρακτηριστικό όλων αυτών των προσεγγίσεων ήταν εκτιμήσεις με πόλωση και αποκλίσεις. Ακόμα χειρότερα αποτελέσματα παρουσίαζαν οι μέθοδοι που δεν βασίζονταν σε βελτιστοποίηση αλλά σε κλειστού τύπου (*closed form*) απευθείας υπολογισμούς με σχέσεις που προέκυπταν μέσω γραμμικοποιήσεων και άλλων μετασχηματισμών. Οι λύσεις των μεθόδων αυτών υπέφεραν από μεγάλα επίπεδα πόλωσης και δεν συνιστούνταν ούτε σαν αρχικοποιήσεις για τους αλγορίθμους βελτιστοποίησης.

Σε αυτήν την εργασία δείξαμε ότι η ανάκτηση της κίνησης της κάμερας από μέτρια θορυβώδη οπτική ροή είναι δυνατή και υπό προϋποθέσεις μπορεί να είναι αξιόπιστη, συνεπής, με αμελητέα σφάλματα. Σε σχέση με παλαιότερες προσεγγίσεις επιχειρήσαμε απευθείας ελαχιστοποίηση με τις εξισώσεις του μοντέλου χωρίς περιττούς μετασχηματισμούς, χρησιμοποιήσαμε συνάρτηση κόστους που αυξάνει πιο αργά σε σχέση με την κλασική περίπτωση των ελαχίστων τετραγώνων μειώνοντας έτσι την επίδραση των εσφαλμένων δεδομένων. Επιπρόσθετα εφαρμόσαμε μία μέθοδο που επιτυγχάνει μείωση των ανεπιθύμητων ελαχίστων αυξάνοντας σταδιακά την πολυπλοκότητα του σχήματος της συνάρτησης κόστους κατά την διάρκεια της βελτιστοποίησης. Το υπολογιστικό κόστος για τα παραπάνω δεν είναι μεγάλο και η επίλυση με την ταχύτατη Gauss-Newton επαναληπτική μέθοδο διατηρεί τον χρόνο απόκρισης σε αποδεκτά επίπεδα. Η παρατήρηση ότι η διαδικασία βελτιστοποίησης μπορεί να διαχωριστεί ώστε η περιστροφή και το βάθος να υπολογίζονται βάσει της μεταφοράς βελτιώνει εκ νέου την απόδοση αφού μας δίνει την δυνατότητα να πραγματοποιούμε Gauss-Newton ανανεώσεις μόνο στην μεταφορά t δηλαδή σε 3 αντί σε $m+6$ αγνώστους. Έτσι με 15 αρχικοποιήσεις και επανεκκινήσεις μπορούμε πρακτικά να εντοπίζουμε το ολικό ελάχιστο της συνάρτησης κόστους.

Ωστόσο η ακρίβεια των αποτελεσμάτων εξαρτάται σε μεγάλο βαθμό από τις εκτιμήσεις της οπτικής ροής. Ο αλγόριθμος L-K φημίζεται για την ακρίβεια των μετρήσεων του (τουλάχιστον για την πλειονότητα των δειγμάτων που παράγει) την απλότητα και την ταχύτητά του. Με στόχο την παραγωγή όσο το δυνατόν πιο αξιόπιστων διανυσμάτων εφαρμόσαμε κάποιες τροποποιήσεις στον αυθεντικό L-K με επεμβάσεις ώστε να ενισχύσουμε τα χαρακτηριστικά που επιθυμούμε για τις ανάγκες της εφαρμογής. Μία από τις προσθήκες, η επεξεργασία πολλαπλών αναλύσεων *coarse-to-fine*, παρουσίασε παρόμοια ακρίβεια με τον αυθεντικό L-K με το επιπλέον προσόν να υπολογίζει επιτυχημένα μεγαλύτερες μετατοπίσεις. Αξιοποιώντας δε το γεγονός ότι η ακρίβεια των μετρήσεων εξαρτάται από τα ιδιαίτερα χαρακτηριστικά τού περιεχομένου της εικόνας, ποσοτικοποιήσαμε τα τελευταία με ένα δείκτη-τιμή και με την χρήση ενός κατωφλιού εξαιρέσαμε τις ακατάλληλες για μετρήσεις περιοχές. Ακόμα και με τις προσθήκες αυτές ένα μικρό ποσοστό διανυσμάτων περιείχε ακραίες τιμές θορύβου τις οποίες αντιμετωπίσαμε με ένα επιπλέον στάδιο ελέγχου. Η εφαρμογή μιας *robust* τεχνικής στατιστικού ελέγχου στον L-K αλλά και στην ΕΚΚ φρόντισε να ανιχνεύσει και εξαιρέσει από την επεξεργασία τα εσφαλμένα αυτά δείγματα.

Ενδελεχείς δοκιμές για όλες τις εκδοχές του αλγορίθμου έγιναν με συνθετική οπτική ροή, συνθετικές ακολουθίες εικόνων καθώς και πραγματικές ακολουθίες εικόνων στις οποίες εξετάστηκαν μια σειρά από ακραία και συνηθισμένα σενάρια κινήσεων της κάμερας. Οι δοκιμές τεκμηρίωσαν τις θεωρητικές υποθέσεις που κάναμε προηγουμένως. Τελικά υπό προϋποθέσεις το σύστημα που αναπτύξαμε θα μπορούσε να έχει πρόσφορη χρήση σε κάποια πρακτική εφαρμογή. Ωστόσο δεν θα πρέπει να αγνοούμε ότι λόγω της πολυπλοκότητας του προβλήματος και της πληθώρας παραγόντων που το επηρεάζουν θα πρέπει να είμαστε προετοιμασμένοι να ανεχτούμε ένα αξιοσημείωτο ποσοστό σφάλματος στις εκτιμήσεις.

6.2 Συμπεράσματα

Επιγραμματικά κάποια αξιόλογα συμπεράσματα που αναδεικνύονται από την διερεύνηση του προβλήματος ΕΚΚ είναι τα ακόλουθα:

- Θα πρέπει να θεωρείται εξ’ αρχής δεδομένο ότι η οπτική ροή περιέχει ένα «αξιοσέβαστο» ποσό θορύβου.
- Υπάρχει περιθώριο να είμαστε ιδιαίτερα επιλεκτικοί με τα διανύσματα απορρίπτοντας όλα εκείνα για τα οποία υπάρχουν ενδείξεις ότι είναι λαθεμένα.
- Το ποσοστό των σφαλμάτων στην ΕΚΚ δεν εξαρτάται μόνο από το θόρυβο αλλά και από άλλους παράγοντες όπως το οπτικό πεδίο. Η αύξηση του οπτικού πεδίου επιδρά αρνητικά.
- Η ακρίβεια των αποτελεσμάτων βελτιώνεται όσο αυξάνεται ο αριθμός των διανυσμάτων που δίνονται ως είσοδος στον αλγόριθμο ΕΚΚ.
- Η πόλωση και οι αποκλίσεις είναι ανάλογες του θορύβου.
- Κάποιες κατευθύνσεις μεταφοράς-περιστροφής είναι πιο ευάλωτες στον θόρυβο της οπτικής ροής σε σχέση με άλλες στις ίδιες συνθήκες (σ_n του θορύβου κοινό-σταθερό).
- Κάποιες κατευθύνσεις μεταφοράς-περιστροφής έχουν περισσότερες πιθανότητες να εκτιμηθούν εσφαλμένα από τον ΕΚΚ σε σχέση με άλλες στις ίδιες συνθήκες.
- Το πεδίο οπτικής ροής που παράγεται από κάποιες μεταφορές είναι παρόμοιο με το πεδίο που παράγεται από κάποιες περιστροφές. Παρουσία θορύβου είναι πιθανό μια μεταφορά να εκληφθεί ως περιστροφή και το αντίθετο. Το φαινόμενο παρουσιάζεται συνήθως σε πλαγιοπλευρικές κινήσεις και εντείνεται όσο μικραίνει το οπτικό πεδίο.
- Κατά την προοπτική προβολή το βάθος και η μεταφορά «ενσωματώνονται» σε ένα λόγο και επομένως μόνο η εύρεση αυτού είναι δυνατή. Αυτό δημιουργεί την αναγκαιότητα περιορισμού του μέτρου μεταφοράς σε μία σταθερή τιμή. Αυτό έχει σαν συνέπεια να μην προβλέπεται η περίπτωση απουσίας μεταφοράς.

- Ακόμα και στην περίπτωση απουσίας μεταφοράς (όπου το *t* λαμβάνει τυχαίες τιμές) οι εκτιμήσεις για την περιστροφή παραμένουν ανεπηρέαστες και υπολογίζονται κανονικά χωρίς σφάλματα.
- Αντίθετες αρχικές τιμές οδηγούν σε ίδια αλλά αντίθετα αποτελέσματα μεταφοράς - βάθους. Οι αντίθετες εκτιμήσεις για την μεταφορά - βάθος θεωρούνται εφάμιλλες και αντιστοιχούν στο ίδιο ελάχιστο. Εκ των υστερών προκρίνεται η φορά της μεταφοράς που μας εξασφαλίζει τα περισσότερα θετικά βάθη.
- Το βάθος για κάποια σημεία ενδέχεται να λάβει αρνητικές τιμές.
- Τυχόν εσφαλμένη εκτίμηση της ορθής φοράς οδηγεί σε απόκλιση 180^o από το σωστό.
- Οι εκτιμήσεις των μεταφορών φαίνεται να είναι πιο επιρρεπείς σε λάθη και η «συνεισφορά» τους στο συνολικό σφάλμα είναι μεγαλύτερη σε σχέση με τις περιστροφές.
- Η συνάρτηση κόστους έχει τουλάχιστον δυο «ισχυρά» ελάχιστα.
- Ο αριθμός των ελαχίστων σχετίζεται με την διάταξη των διανυσμάτων στο οπτικό πλάνο. Ελάχιστα δημιουργούνται σε θέσεις όπου οι συντεταγμένες των διανυσμάτων είναι ανάλογες της μεταφοράς.
- Μπορούμε ελέγχοντας το σχήμα της συνάρτησης κόστους να εξωθήσουμε την βελτιστοποίηση να αποφύγει την σύγκλιση σε ένα αριθμό από ανεπιθύμητα ελάχιστα.
- Ο σταθμισμένος Gauss-Newton αποδίδει χειρότερα αποτελέσματα (έναντι του μη-σταθμισμένου) όταν ο θόρυβος είναι ενιαίος-κοινός για όλα τα δείγματα. Ωστόσο δίδει ελαφρώς καλύτερα αποτελέσματα όταν μεταχειρίζεται δείγματα με ανάμεικτο θόρυβο διαφόρων επιπέδων.
- Η βελτιστοποίηση συγκλίνει συνήθως στο ελάχιστο που είναι πιο κοντά στην αρχική τιμή απ' όπου ξεκινάει η διαδικασία επίλυσης. Κατ' επέκταση έχει σημασία η ομοιόμορφη κατανομή των αρχικών τιμών.
- Η βελτιστοποίηση συγκλίνει σχετικά γρήγορα και ο χρόνος απόκρισης της μεθόδου είναι ανεκτός.

6.3 Μελλοντικές προεκτάσεις

Ένα ερώτημα που εγείρεται είναι το κατά πόσο η μέθοδος που αναπτύξαμε σε αυτή την εργασία είναι δυνατό να αξιοποιηθεί πρακτικά (π.χ. στην παθητική πλοήγηση μιας μηχανής). Για να απαντήσουμε στο ερώτημα αυτό είναι απαραίτητο να γίνουν σχολαστικές δοκιμές και έλεγχοι με πραγματικές ακολουθίες εικόνων. Δυστυχώς η απουσία μιας βάσης δεδομένων με πληροφορίες κινήσεων κάμερας για πραγματικές ακολουθίες αλλά και γενικά η δυσκολία υλοποίησης ενός συστήματος καταγραφής των κινήσεων μιας κάμερας, μας εξαναγκάζουν στην χρήση δοκιμών με συνθετική οπτική ροή που όμως δεν μπορεί να είναι απόλυτα ενδεικτική για τα ισχύοντα στην πράξη.

Κάποιες ιδέες που πιστεύουμε ότι θα αποτελούσαν την βασική αρχή και έναυσμα για την λύση στο πρόβλημα της καταγραφής των πραγματικών κινήσεων είναι είτε μέσω της χρήσης συσκευής GPS ακριβείας που θα εγκαθίσταται στην κάμερα, είτε μέσω ανίχνευσης και παρακολούθησης πρότυπου αντικειμένου αναφοράς το οποίο θα τοποθετείται στον χώρο λήψης. Έτσι η κίνηση θα μπορεί να υπολογιστεί από το αντικείμενο αναφοράς με *calibration* (διαδικασία παρόμοια με αυτή που χρησιμοποιείται για την διόρθωση παραμορφώσεων και ατελειών στο σύστημα λήψης των καμερών). Μια ακόμα πιθανή ιδέα είναι η χρήση ρομποτικού βραχίονα που θα μετακινεί την κάμερα στις επιθυμητές θέσεις ενώ παράλληλα θα καταγράφει τις κινήσεις δειγματοληπτώντας σε χρόνο καρέ.

Ωστόσο η πρακτική χρησιμότητα της μεθόδου κρίνεται πρωτίστως από έναν άλλο παράγοντα, τον χρόνο απόκρισης. Η αξία του εγχειρήματος θα αυξάνονταν αν κατορθώναμε να πετύχουμε λειτουργία πραγματικού χρόνου (ή για να ακριβολογούμε χρόνο απόκρισης στην διάρκεια ενός καρέ, δηλαδή σε 1/25 *sec*). Έτσι ο αλγόριθμος θα πρέπει να διερευνηθεί και τροποποιηθεί ώστε να μπορεί να λειτουργεί σε πραγματικό χρόνο. Μερικές ιδέες που πιστεύουμε ότι θα οδηγούσαν προς αυτή την κατεύθυνση και θα μείωναν δραστικά το χρόνο απόκρισης είναι η επίλυση της οπτικής ροής με κάποια *multigrid* [25] μέθοδο και η συγγραφή του πηγαίου κώδικα όλων των αλγορίθμων σε *C++* με χρήση βιβλιοθηκών επεξεργασίας εικόνας (όπως η *openCV*) που χρησιμοποιούν σετ εντολών *MMX-SSE2*.

Επίσης σημαντική επιβάρυνση στον χρόνο απόκρισης προκαλεί κατά την βελτιστοποίηση η αναγκαιότητα πολλαπλών αρχικοποιήσεων και επανεκκινήσεων προκειμένου να εντοπιστεί το ολικό ελάχιστο ανάμεσα από πλήθος (ανεπιθύμητων) ελαχίστων. Στην εργασία αυτή είδαμε μια μέθοδο μείωσης ελαχίστων η οποία κατόρθωσε να μειώσει αλλά όχι να εξαλείψει όλα τα ανεπιθύμητα ελάχιστα. Αν ωστόσο η μέθοδος αυτή βελτιωθεί ώστε να εξαλείφει όλα τα ανεπιθύμητα ελάχιστα, νέες προοπτικές δημιουργούνται αφού θα μπορούμε να εντοπίζουμε το ολικό ελάχιστο δηλαδή την λύση με ένα μόνο σετ επαναλήψεων. Το υπολογιστικό κέρδος από κάτι τέτοιο είναι προφανές. Εναλλακτικά ακόμα και αν ο αριθμός των ελάχιστων δεν εξαλειφθεί εντελώς αλλά μειωθεί σε ένα εκ των προτέρων γνωστό εύρος τιμών θα έχουμε την δυνατότητα να εκτελούμε μόνο όσα σετ αρχικοποιήσεων και επανεκκινήσεων απαιτούνται ως ότου εμφανιστεί ο αναμενόμενος αριθμός ελάχιστων. Ακόμα πιθανολογούμε ότι είναι δυνατή η ανάπτυξη μιας τεχνικής που βασιζόμενη στην γνώση του σχήματος-«ανάγλυφου» της συνάρτησης κόστους και των ιδιομορφιών της θα μπορεί να εντοπίζει άμεσα ένα ελάχιστο λαμβάνοντας υπόψη άλλα ήδη γνωστά.

Τέλος θεωρούμε ότι καθένα από τα ζητήματα - προβλήματα που αντιμετωπίσαμε σε αυτή την εργασία έχει την δράση του και τον δικό του αντίκτυπο στα σφάλματα. Διατηρούμε την πεποίθηση ότι μελλοντικές επεμβάσεις για βελτίωση ή επίλυση οποιουδήποτε από αυτά τα προβλήματα θα συνεισέφερε σε μεγάλο βαθμό στην βελτίωση του τελικού αποτελέσματος.

Βιβλιογραφία

[1] T. Zhang, C. Tomasi, On the consistency of instantaneous rigid motion estimation, International Journal of Computer Vision 46 (2002) 51–79.

[2] T. Zhang, C. Tomasi, 1999. Fast, robust, and consistent camera motion estimation. Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.

[3] K Pauwels, M. V. Hulle, Optimal instantaneous rigid motion estimation insensitive to local minima, Journal Computer Vision and Image Understanding archive Volume 104 Issue 1, 2006.

[4] Y. Tian, C. Tomasi, D. Heeger, Comparison of approaches to egomotion computation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 1996, pp. 315–320.

[5] N. Komodakis, G. Tziritas, Robust 3-D motion estimation and depth layering, Proceedings of 13th International Conference on Digital Signal Processing, 1997.

[6] A. Chiuso, R. Brockett, S. Soatto, Optimal structure from motion: local ambiguities and global estimates, International Journal of Computer Vision 39 (3) (2000) 195–228.

[7] S. Soatto, P. Perona, Robust and Efficient Recovery of Rigid Motion from Subspace Constraints Solved using Recursive Identification of Nonlinear Implicit Systems, Technical Report. California Institute of Technology, Pasadena, CA 1994.

[8] C. Tomasi, T. Kanade. Shape and motion without depth. Technical Report CMU-CS-90-128, Carnegie Mellon University, May 1990.

[9] S. Park, H. Lee, S. Lee, Qualitative estimation of camera motion parameters from the linear composition of optical flow, PR(37), 4:767� 779, 2004

[10] A. Bruss, B. Horn, Passive navigation, Computer Graphics and Image Processing 21 (1983) 3–20

[11] W. J. MacLean, A. D. Jepson, and R. C. Frecker. Recovery of egomotion and segmentation of independent object motion using the EM algorithm. In 5th British Machine Vision Conference, pages 13-1 6, 1994

[12] K. Prazdny. On the information in optical flows. CGIP, 221239-259, 1983.

[13] D.J. Heeger, A.D. Jepson, 1992. Subspace methods for recovering rigid motion I: Algorithm and implementation. International Journal of Computer Vision, 7(2):95–118.

[14] D.J. Heeger, A.D. Jepson. 1993. Linear subspace methods for recovering translation direction. In Spatial Vision in Humans and Robots, L. Harris and M. Jenkin (Eds.). Cambridge University Press: New York, pp. 39–62

[15] K. Kanatani, 1993. 3-d interpretation of optical flow by renormalization. International Journal of Computer Vision, 11(3):267–282.

[16] E. Trucco, A. Verri, 1998. Introductory Techniques for 3-D Computer Vision. Prentice Hall: Upper Saddle River, NJ.

[17] H. Stachel, Descriptive Geometry Meets Computer Vision - The Geometry of Two Images, Journal for Geometry and Graphics Volume 10 (2006), No. 2, 137–153

[18] J. Barron, D. Fleet, S. Beauchemin, Performance of optical flow techniques, International Journal of Computer Vision 12 (1) (1994) 43–77.

[19] B. Horn, B. Schunk. Determining optical flow. Artificial Intelligence, Vol. 17, pp. 185-203, 1981

[20] B. D. Lucas, T. Kanade. An iterative image registration technique with an application to stereo vision. In 7th International Joint Conference on Artificial Intelligence: Springer-Verlag, New York, New York, p. 674–679 1981.

[21] J. Y. Bouguet, Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm, Intel Corporation Microprocessor Research Labs, 2000

[22] J. Shi, C. Tomasi. 1994. Good features to track. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR94), Seattle, WA, pp. 593–600.

[23] C. Tomasi, T. Kanade. Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University, April 1991.

[24] Bruhn, J. Weickert, and C. Schnorr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. International Journal of Computer Vision, 61(3):211–231, 2005

[25] Bruhn, A. and Weickert, J. 2005. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In Proc. Tenth International Conference on Computer Vision, vol. 1, pages 749–755, Beijing, China, IEEE Computer Society Press.