



ΑΝΑΠΤΥΞΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΑΛΓΟΡΙΘΜΟΥ ΟΠΤΙΚΗΣ ΟΔΟΜΕΤΡΙΑΣ / SLAM

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΝΑΚΗΣ ΠΑΝΑΓΙΩΤΗΣ (ΑΜ:1607)

<u>ΕΠΙΒΛΕΠΟΝΤΕΣ:</u> ΠΕΤΣΑ Ε. ΓΡΑΜΜΑΤΙΚΟΠΟΥΛΟΣ Λ. ΚΑΡΡΑΣ Γ.

ΑΘΗΝΑ, ΙΟΥΝΙΟΣ 2019

Εύχαριστιές

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω τους καθηγητές μου, κ. Λάζαρο Γραμματικόπουλο, κ. Έλλη Πέτσα και κ. Γιώργο Καρρά που μου δώσανε την ευκαιρία να ασχοληθώ με το συγκεκριμένο θέμα, αλλά κυρίως για την υπομονή και την συνεχή καθοδήγηση που μου προσφέρανε καθ' όλη τη διάρκεια της διπλωματικής εργασίας. Επίσης, θα ήθελα να ευχαριστήσω όλους όσους στάθηκαν δίπλα μου με την υποστήριξη και τις συμβουλές τους. Τέλος ένα μεγάλο ευχαριστώ στην οικογένεια μου για όσα μου έχει προσφέρει όλα αυτά τα χρόνια.

Προλογος

Η διπλωματική εργασία με τίτλο «Ανάπτυξη και Αξιολόγηση Αλγορίθμου Οπτικής Οδομετρίας / SLAM» εκπονήθηκε στο πλαίσιο του Μεταπτυχιακού Προγράμματος «Γεωχωρικές Τεχνολογίες» του Τμήματος Μηχανικών Τοπογραφίας και Γεωπληροφορικής του Πανεπιστημίου Δυτικής Αττικής.

Το τεύχος αυτό περιγράφει διαφορετικές προσεγγίσεις για την επίλυση του προβλήματος της Ταυτόχρονης Εύρεσης Θέσης και Χαρτογράφησης (SLAM) και της Οπτικής Οδομετρίας (Visual Odometry). Ο σκοπός της, εκτός από την εκτενή βιβλιογραφική έρευνα πάνω στις δύο μεθοδολογίες, ήταν και η προσπάθεια υλοποίησης ενός βασικού παραμετροποιήσιμου αλλά αυτοματοποιημένου αλγορίθμου, ο οποίος θα μπορεί να λειτουργεί σε ποικίλα περιβάλλοντα, διαφορετικής κλίμακας, και μάλιστα σε πραγματικό χρόνο. Τελικό αποτέλεσμα ήταν η δημιουργία δύο τελικά αλγορίθμων, ίδιας βασικής δομής, αλλά με χρήση διαφορετικής μεθοδολογίας ανίχνευσης θέσης και συνθετότητα χαρτογράφησης.

Οι αλγόριθμοι χρησιμοποιούν και επεξεργάζονται δεδομένα από σύστημα μονής κάμερας, για την ανάκτηση της θέσης της κάμερας και τη χαρτογράφηση του χώρου. Η δομή τους έχει χωριστεί σε τέσσερα βασικά κομμάτια. Πρώτο, η Αρχικοποίηση, στο οποίο γίνεται εύρεση της σχετικής θέσεις ενός ζεύγους εικόνων, από τον υπολογισμό του δεσμευμένου επιπολικού πίνακα Ε, με τον αλγόριθμο των 5 σημείων, καθώς και η δημιουργία του πρώτου συνόλου τρισδιάστατων σημείων από τις ομολογίες τους. Δεύτερο κομμάτι, η Ανίχνευση, όπου υπολογίζεται η τρέχουσα θέση της κάμερας με 3D-2D ομολογίες, και αποφασίζεται αν το τρέχον καρέ θα χρησιμοποιηθεί για την επέκταση του μοντέλου με τον υπολογισμό νέων τρισδιάστατων σημείων. Το τρίτο κομμάτι της Χαρτογράφησης αναλαμβάνει την τριγωνοποίηση των ομολογιών του νέου καρέ και του τελευταίου κλειδιού-καρέ, τα οποία προστίθενται στο συνολικό μοντέλο. Τελευταίο κομμάτι, η Βελτιστοποίηση, αναλαμβάνει τη συνόρθωση του μοντέλου με χρήση Τοπικής Συνόρθωσης Δέσμης (Local Bundle Adjustment), η οποία βελτιστοποιεί τα τρισδιάστατα σημεία και τις θέσεις των **n** τελευταίων καμερών. Το κομμάτι της Βελτιστοποίησης ενεργοποιείται για πρώτη φορά με την ολοκλήρωση της Αρχικοποίησης και με κάθε λήξη της διαδικασίας Χαρτογράφησης.

Η γλώσσα προγραμματισμού που χρησιμοποιήθηκε για την υλοποίηση του αλγορίθμου ήταν η C++, για τις διαδικασίες υπολογισμού θέσης και χαρτογράφησης επιλέχθηκε η βιβλιοθήκη όρασης υπολογιστών OpenCV, και για τη διαδικασία της συνόρθωσης η βιβλιοθήκη επίλυσης προβλημάτων μορφής γράφου g2o.

Για την αξιολόγηση των αλγορίθμων χρησιμοποιήθηκαν σύνολα δεδομένων οδομετρίας, ΚΙΤΤΙ καθώς και πειραματικά δεδομένα τα οποία προήλθαν από την περιμετρική βιντεοσκόπηση, ενός τραπεζιού εμπλουτισμένου με αντικείμενα.

Στα παρακάτω κεφάλαια γίνεται μια αναφορά στις βασικές αρχές των δύο μεθοδολογιών και μερικούς από τους τρόπους, που έχουν προταθεί από τη βιβλιογραφία, για την επίλυση σημαντικών εγγενών προβλημάτων αυτών. Ακολουθεί παρουσίαση και αξιολόγηση των υλοποιημένων αλγορίθμων. Και τέλος συνοψίζονται τα συμπεράσματα της αξιολόγησης, τα αποτελέσματα και οι δυσκολίες που προέκυψαν, καθώς και προτάσεις και ιδέες για περαιτέρω ανάπτυξη και έρευνα.

UNIVERSITY OF WEST ATTICA SCHOOL OF ENGINEERING DEPARTMENT OF SURVERING AND GEOINFORMATICS MASTER OF SCIENCE IN GEOSPATIAL TECHNOLOGIES

MASTER THESIS

DEVELOPMENT AND EVALUATION OF A VISUAL ODOMETRY / SLAM ALGORITHM

NAKIS PANAGIOTIS

JUNE 2019

ABSTRACT

This thesis describes different approaches to solving the Visual Odometry (VO) and Simultaneous Localization and Mapping (SLAM) problem. Its purpose, apart from an extensive bibliographic research on the two methodologies, was the attempt of implementing a basic, parameterizable but automated algorithm, which could operate in a variety of environments, scales and runs in real time. The end result was not one, but two algorithms, with the same basic structure, but using different pose estimation methodologies and mapping complexity.

Both algorithms use as input and process data from a Monocular camera system to retrieve the camera's location and the mapping of the environment. Their structure has been divided into four basic modules. The first one, Initialization, is responsible for the computation of the relative pose of a pair of frames, calculating the Essential Matrix **E**, using the *five-point algorithm*, as well as the triangulation and creation of the first set of 3D points of the environment. The second module Localization, is responsible for the computation of the current frames pose, using 3D-2D correspondences, and decides whether or not, the current frame will be used to extend the environment model by computing new 3D points. The third module, Mapping, triangulates new 3D points, with 2D-2D correspondences of the current frame and the last keyframe, which are added to the overall model. The last module, Optimization, optimizes the poses and 3D points of the last **n** number of keyframes, using Local Bundle Adjustment. This module is triggered for the first time after Initialization has been successful, and at the end of each Mapping process.

The programming language that was selected for the implementation of these algorithms was C++; for the calculation of each frames pose and the triangulation of the 3D points, the OpenCV computer vision library has been selected; and the posegraph optimization library g20 has been used for the optimization of the model.

In order to evaluate the algorithms, odometry datasets provided by KITTI have been used, as well as an experimental dataset derived from a video camera, capturing a table scene, enriched with different objects.

The following chapters present the basic principles of the two methodologies, and ways proposed by the literature of solving some important inherent problems. Then follows a presentation of the two aforementioned algorithms as well as their evaluation using real data. Finally, the conclusions of the evaluation, the results and the difficulties that have arisen, as well as suggestions and ideas for further research and development are summarized.

Keywords: Visual Odometry, Simultaneous Localization and Mapping, Monocular, Pose Estimation.

ΠΕΡΙΕΧΟΜΕΝΑ

Εικό	νες		ix		
Διαν	γράμ	ιματα	x		
Пілс	ακες		x		
1	Εισαγωγή				
1.	.1	Στόχος της Διπλωματικής Εργασίας	3		
1.	.2	Δομή Κειμένου	4		
2	Θεω	ορητικό Υπόβαθρο	5		
2	.1	Οπτική Οδομετρία	5		
	2.1.1	1 Μἑθοδοι Βἁσει Χαρακτηριστικών Σημείων	6		
	2.1.2	2 Άμεσες Μἑθοδοι	10		
	2.1.3	3 Κάμερα και Βαθμονόμηση	11		
	2.1	1.3.1 Γεωμετρία Κάμερας	11		
	2.1	1.3.2 Παραμορφώσεις Φακού	15		
	2.1	1.3.3 Βαθμονόμηση Μηχανής	16		
	2.1.4	4 Γεωμετρία του Προβλήματος και Εκτίμηση Κίνησης	18		
	2.1.5	5 RANSAC	25		
	2.1.6	6 Βελτιστοποίηση	28		
	2.1	1.6.1 Συνόρθωση Δέσμης	29		
	2.1	1.6.2 Συνόρθωση Γράφου			
2	.2	Ταυτόχρονη Οπτική Εύρεση Θέσης και Χαρτογράφηση	31		
	2.2.1	1 Μέθοδοι Βασισμένες σε Φίλτρα	31		
	2.2.2	2 Μέθοδοι Βασισμένες σε Βασικά Καρέ	33		
	2.2.3	3 Κλείσιμο Βρόχων	34		
	2.2.4	4 Επανεντοπισμός	36		
3	Пар	οουσίαση Αλγορίθμων			
3	.1	Αρχικοποίηση			
3	.2	Ανίχνευση	41		
3	.3	Χαρτογράφηση	42		
3	.4	Βελτιστοποίηση	43		
4	Αξιο	ολόγηση Αλγορίθμων	45		
5	Συμπεράσματα και Προτάσεις για Περαιτέρω Έρευνα				
6	Βιβλιογραφία61				

Εικονές

Εικόνα 2.1: Εσωτερική γεωμετρία μηχανής (Hartley and Zisserman 2003)12
Εικόνα 2.2: Σχέση των συστημάτων της εικόνας (Hartley and Zisserman 2003)
Εικόνα 2.3: Σχέση μεταξύ του συστήματος αναφοράς χώρου και του συστήματος της
κάμερας (Hartley and Zisserman 2003)14
Εικόνα 2.4: Παράδειγμα επίδρασης της παραμόρφωσης λόγω φακού16
Εικόνα 2.5: Παράδειγμα λήψης εικόνων σκακιέρας διαφορετικού προσανατολισμού
για τη διαδικασία της βαθμονόμησης (Gary Bradski and Kaehler 2008)18
Εικόνα 2.6: Εικόνα με τις σχέσεις της επιπολικής γεωμετρίας (Hartley and Zisserman
2003)19
Εικόνα 2.7: Απεικόνιση του τρόπου με τον οποίο μπορούν να ερμηνευτούν οι τέσσερις
διαφορετικές λύσεις (Hartley and Zisserman 2003)23
Εικόνα 2.8: Παράδειγμα προσαρμογής ευθείας σε δεδομένα που περιέχουν έντονο
θόρυβο (χονδροειδώς λανθασμένα δεδομένα)26
Εικόνα 2.9: Παράδειγμα αναπαράστασης ενός προβλήματος μορφής γράφου30
Εικόνα 3.1: Βασικό διάγραμμα ροής μεθοδολογιών44
Εικόνα 4.1: Δείγμα εικόνων από το σύνολο δεδομένων 05 του ΚΙΤΤΙ
Εικόνα 4.2: Δείγμα εικόνων από το σύνολο των πειραματικών δεδομένων του
Τραπεζιού που κατασκευάστηκε46
Εικόνα 4.3: Καρἑ από τα δύο σύνολα δεδομἑνων ΚΙΤΤΙ 05 (πάνω) και Τραπἑζι (κἁτω),
όπου εμφανίζονται τα χαρακτηριστικά σημεία τα οποία έχουν ανιχνευθεί επιτυχώς από
το προηγούμενο καρἑ (πράσινα), και τα νἑα υποψήφια χαρακτηριστικά προς
ανίχνευση στο επόμενο καρέ (κίτρινα)
Εικόνα 4.4: Τελικό νέφος τρισδιάστατων σημείων από το σύνολο δεδομένων ΚΙΤΤΙ 05.
Με πράσινο χρώμα παρουσιάζεται η αληθής τροχιά και με κίτρινο η υπολογιζόμενη.54
Εικόνα 4.5: Τελικό νέφος τρισδιάστατων σημείων από το σύνολο δεδομένων του
Τραπεζιού. Με πράσινο χρώμα παρουσιάζεται η αληθής τροχιά και με κίτρινο η
υπολογιζόμενη

Δ iappammata

Διάγραμμα 2.1: Διάγραμμα του αριθμού των επαναλήψεων RANSAC σε
διαφορετικούς αλγορίθμους για πιθανότητα επιτυχίας 0.99. Οι υπολογισμοί του
διαγράμματος έγιναν σύμφωνα με την εξίσωση (2.31)
Διάγραμμα 4.1: Αποτελέσματα αξιολόγησης αλληλουχίας ΚΙΤΤΙ – 00. Μέσο γωνιακό
σφάλμα: 0.003953 deg/m. Μἑσο τροχιακό σφάλμα: 9.1350 %
Διάγραμμα 4.2: Αποτελέσματα αξιολόγησης αλληλουχίας ΚΙΤΤΙ – 03. Μέσο γωνιακό
σφάλμα: 0.004927 deg/m και μέσο τροχιακό σφάλμα: 9.0104 %
Διάγραμμα 4.3: Αποτελέσματα αξιολόγησης αλληλουχίας ΚΙΤΤΙ – 04. Μέσο γωνιακό
σφάλμα: 0.003266 deg/m και μέσο τροχιακό σφάλμα: 3.8045 %50
Διάγραμμα 4.4: Αποτελέσματα αξιολόγησης αλληλουχίας ΚΙΤΤΙ – 05. Μέσο γωνιακό
σφάλμα: 0.007334 deg/m, μέσο τροχιακό σφάλμα: 10.2664 %
Διάγραμμα 4.5: Αποτελέσματα αξιολόγησης αλληλουχίας Τραπέζι – αλγόριθμος KLT.
Μέσο γωνιακό σφάλμα: 0.021834 deg/cm και μέσο τροχιακό σφάλμα: 2.8555 %51
Διάγραμμα 4.6: Αποτελέσματα αξιολόγησης αλληλουχίας Τραπέζι – αλγόριθμος ORB.
Μέσο γωνιακό σφάλμα: 0.021120 deg/cm και μέσο τροχιακό σφάλμα: 3.3387 %52

Πινάκες

Πίνακας 4.1: Περιγραφικά στοιχεία των συνόλων δεδομένων	47
Πίνακας 4.2: Χαρακτηριστικά των ολοκληρωμένων συνόλων δεδομένων	52
Πίνακας 4.3: Αποτελέσματα της αξιολόγησης των συνόλων δεδομένων	52

1 Εισαγωγή

Η Ταυτόχρονη Εύρεση Θέσης και Χαρτογράφηση (Simultaneous Localization and Mapping - SLAM) είναι μια μεθοδολογία που χρησιμοποιείται για την χαρτογράφηση ενός άγνωστου περιβάλλοντος, με τη χρήση ενός συστήματος αισθητήρων (laser range sensors, GPS, inertial measurement sensors, cameras, sonars κ.α.), και ταυτόχρονη εύρεση της θέσης του μέσα σε αυτό, χωρίς καμία προϋπάρχουσα πληροφορία. Ο όρος localization αφορά την προσπάθεια εύρεσης της θέσης ενός ρομπότ μέσα σε ένα χώρο με τη χρήση χαρακτηριστικών σημείων του χώρου αυτού, ενώ ο όρος mapping αφορά τη χαρτογράφηση του περιβάλλοντος (μοντελοποίηση) θα πρέπει να είναι γνωστή η θέση του ρομπότ και ταυτόχρονα για την εύρεση της θέσης θα πρέπει να υπάρχει ένα ακριβές μοντέλο του περιβάλλοντος. Αυτό ακριβώς το πρόβλημα προσπαθεί να λύσει το SLAM.

Σύμφωνα με τους Durrant-Whyte and Bailey (2006), η έρευνα στο αντικείμενο του SLAM χρονολογείται από το 1986, όταν συζητήθηκε για πρώτη φορά η ιδέα της χρήσης πιθανοτικών-θεωρητικών μεθόδων για την εφαρμογή σε προβλήματα χαρτογράφησης και εντοπισμού (1986 IEEE Robotics and Automation Conference, Σαν Φρανσίσκο). Το άρθρο των Smith, Self, & Cheeseman, (1990) απέδειξε ότι καθώς ένα όχημα (ρομπότ) κινείται μέσα σε ένα άγνωστο περιβάλλον λαμβάνοντας σχετικές παρατηρήσεις των «οροσήμων» (landmarks), οι εκτιμήσεις της θέσης αυτών των οροσήμων συνδέονται αναγκαστικά μεταξύ τους εξαιτίας του κοινού σφάλματος στην εκτιμώμενη τροχιά του οχήματος. Η δομή του προβλήματος και η καθιέρωση του ακρωνυμίου «SLAM» πρωτοπαρουσιάστηκε από τους Durrant-Whyte, Rye, and Nebot (1996) στο 1995 International Symposium on Robotics Research.

Η απλούστερη μορφή εντοπισμού θέσης είναι με τη χρήση μεθόδων οδομετρίας τροχών (wheel odometry), που βασίζονται σε μετρητές της περιστροφής των τροχών του ρομπότ. Σε αυτές τις μεθόδους, οι μετρήσεις περιστροφής τροχού χρησιμοποιούνται διαδοχικά σε συνδυασμό με το μοντέλο κίνησης του ρομπότ για την εύρεση της τρέχουσας θέσης του σε σχέση με ένα σύστημα συντεταγμένων. Η μέθοδος αυτή έχει ορισμένους σημαντικούς περιορισμούς. Πρώτον, περιορίζεται στα τροχοφόρα οχήματα εδάφους και δεύτερον, δεδομένου ότι ο εντοπισμός είναι διαδοχικός (βάσει της προηγούμενης εκτιμώμενης θέσης), τα λάθη μέτρησης συσσωρεύονται με την πάροδο του χρόνου και προκαλούν παρέκκλιση της τροχιάς του ρομπότ από την πραγματική της θέση. Για να ξεπεραστούν αυτά τα προβλήματα χρησιμοποιήθηκαν άλλες στρατηγικές με την εμφάνιση συνδυασμού αισθητήρων, όπως αδρανειακά συστήματα (IMUs), GPS, laser, sonar, κ.α.

Τα τελευταία χρόνια, η ανάπτυξη της τεχνολογίας και της υπολογιστικής ισχύος έκανε πρόσφορο το έδαφος για την περαιτέρω ανάπτυξη της ρομποτικής και της όρασης υπολογιστών. Μέχρι τις αρχές του 2000, η χρήση κάμερας δεν υπήρξε στο επίκεντρο της εξέλιξης του SLAM. Ο λόγος ήταν ότι θεωρήθηκε πολύ δύσκολο να μετατραπούν τα αραιά χαρακτηριστικά σημεία από μια εικόνα σε αξιόπιστους χάρτες που παράγονται σε πραγματικό χρόνο, δεδομένου μάλιστα ότι οι ρυθμοί δεδομένων που προέρχονται από μια κάμερα είναι πολύ υψηλότεροι από εκείνους που προέρχονται από άλλους αισθητήρες. Ως εκ τούτου, οι ερευνητές στην περιοχή της όρασης υπολογιστών επικεντρώθηκαν σε προβλήματα ανακατασκευής από μικρά σύνολα εικόνων, αναπτύσσοντας το πεδίο που είναι γνωστό ως Δομή από Κίνηση (Structure-from-Motion – SFM). Το γενικό πρόβλημα που προσπαθεί να λύσει το SFM είναι η τρισδιάστατη ανακατασκευή μιας σκηνής που καταγράφεται από ένα σύνολο από διαδοχικές ή και όχι εικόνες, όσο και η εύρεση της θέσης των φωτογραφικών μηχανών. Το προϊόν της ανακατασκευής και οι θέσεις των καμερών βελτιστοποιούνται σε ένα τελευταίο βήμα, με μεθόδους όπως η Συνόρθωση Δέσμης (Bundle Adjustment), του οποίοι ο χρόνος αυξάνεται με τον αριθμό των εικόνων (Fraundorfer and Scaramuzza 2011). Η Οπτική Οδομετρία (Visual Odometry) είναι μια ειδική περίπτωση SFM, η οποία προσπαθεί να εκτιμήσει σε πραγματικό χρόνο, από αλληλουχίες εικόνων, την κίνηση μιας κάμερας μέσα σε ένα περιβάλλον,.

Η έρευνα στο πεδίο της Οπτικής Οδομετρίας ξεκινάει από το 1980, με τη διδακτορική διατριβή του Hans Moravec (Moravec 1980) σχετικά με την εκτίμηση της κίνησης ενός οχήματος μόνο από οπτικά δεδομένα. Ο Moravec χρησιμοποίησε μια στερεοσκοπική προσέγγιση, με βάση μια κάμερα η οποία κινούταν σε ράγες. Η δοκιμή της μεθόδου έγινε με την προσαρμογή του συστήματος σε ένα πλανητικό όχημα, με περιήγηση σε εσωτερικό χώρο. Από το 1980 μέχρι το 2000, η περισσότερη έρευνα για την οπτική οδομετρία έγινε κυρίως από τη NASA για την προετοιμασία της διαστημικής αποστολής πλανητικού οχήματος στον Άρη, το 2004 (Lacroix et al. 1999; Olson et al. 2000). Όπως

και στο πεδίο του SLAM, το όνομα της Οπτικής Οδομετρίας ήρθε πολύ αργότερα, με άρθρο του David Nistér το 2004 (Nistér, Naroditsky, and Bergen 2004).

Ο στόχος του SLAM είναι η προσπάθεια απόκτησης μιας συνολικής συνεκτικής εκτίμησης της διαδρομής ενός ρομπότ, στην περίπτωση του Visual SLAM της τροχιάς της κάμερας. Ο ορισμός της οδομετρίας είναι η εκτίμηση των διαδοχικών αλλαγών των θέσεων των αισθητήρων ενός ρομπότ, αυτοκινήτου κ.α., με τη πάροδο του χρόνου με βάση συστήματα όπως μετρητές περιστροφής τροχών για την απόκτηση σχετικών κινήσεων των αισθητήρων.

Σε αντίθεση με το Visual SLAM, η οπτική οδομετρία στοχεύει στην εύρεση της τροχιάς σταδιακά, βελτιστοποιώντας μόνο τις τελευταίες m θέσεις της τροχιάς. Μεθοδολογίες οπτικής οδομετρίας χρησιμοποιούνται σε εφαρμογές όπου ενδιαφέρεται κανείς μόνο για την τοπική συνοχή της τροχιάς της κάμερας και το τρισδιάστατο μοντέλο χρησιμοποιείται απλώς για την ακριβέστερη εκτίμηση αυτής της τροχιάς, ενώ στις μεθοδολογίες Visual SLAM το ενδιαφέρον αφορά και την συνολική συνοχή του τρισδιάστατου μοντέλου. Παραδείγματα είναι η αυτόνομη πλοήγηση ενός αυτοκινήτου και η χαρτογράφηση ενός ορυχείου. Από τεχνικής απόψεως, το Visual SLAM και η Visual Odometry είναι παρόμοιες τεχνικές καθώς και οι δύο υπολογίζουν τις θέσεις των αισθητήρων

Τις τελευταίες δεκαετίες, ο τομέας της ρομποτικής, της όρασης υπολογιστών και των αυτόνομων συστημάτων έχει προσελκύσει σημαντική προσοχή από τους ερευνητές σε όλο τον κόσμο. Δημοσιευμένες προσεγγίσεις του Visual SLAM και του Visual Odometry χρησιμοποιούνται σε πληθώρα εφαρμογών, όπως σε αυτόνομα αυτοκίνητα, μη επανδρωμένα εναέρια οχήματα, πλανητικά οχήματα, νεοεμφανιζόμενα οικιακά ρομπότ, εφαρμογές επαυξημένης πραγματικότητας, πλοήγηση σε εσωτερικούς χώρους και ακόμη και ρομπότ νανοϊατρικών εφαρμογών.

1.1 Στοχός της Διπλωματικής Εργασίας

Σκοπός της παρούσας διπλωματικής εργασίας ήταν η κατ' αρχάς εκτενής βιβλιογραφική έρευνα σχετικά με τις μεθοδολογίες της Οπτικής Οδομετρίας και του Ταυτόχρονου Εντοπισμού Θέσης και Χαρτογράφησης, και παράλληλα η απόπειρα υλοποίησης ενός βασικού παραμετροποιήσιμου αλλά αυτοματοποιημένου αλγορίθμου, ο οποίος θα μπορεί να λειτουργεί σε ποικίλα περιβάλλοντα, διαφορετικής κλίμακας, και μάλιστα σε πραγματικό χρόνο. Μέσω των πειραματισμών για την επίλυση προβλημάτων κατά την εκπόνηση της εργασίας, υλοποιήθηκαν δύο αλγόριθμοι, ίδιας βασικής δομής αλλά με διαφορετική μεθοδολογία ανίχνευσης θέσης και συνθετότητα χαρτογράφησης. Οι αλγόριθμοι λαμβάνουν ως δεδομένα εισόδου αλληλουχία εικόνων γνωστής σειράς. Αποτέλεσμα των αλγορίθμων είναι ο υπολογισμός της τροχιάς της μηχανής, καθώς αυτή κινείται σε ένα περιβάλλον, και η ταυτόχρονη χαρτογράφησή του. Οι αλγόριθμοι υλοποιήθηκαν για να δέχονται και να επεξεργάζονται δεδομένα από σύστημα μονής κάμερας.

Ο πρώτος προβληματισμός αφορούσε την επιλογή της γλώσσας προγραμματισμού για την ανάπτυξη της εφαρμογής. Τα κύρια κριτήρια που τέθηκαν για αυτή την επιλογή ήταν η ύπαρξη πληθώρας βιβλιοθηκών υποστήριξης για την υλοποίηση των διαδικασιών που θα παρουσιαστούν στη συνέχεια, η υψηλή υπολογιστική ταχύτητα λόγω της ανάγκης για λειτουργία σε πραγματικό χρόνο και η εξοικείωση που έχει ο δημιουργός του προγράμματος με τη γλώσσα. Η γλώσσα που ικανοποιούσε τις περισσότερες απαιτήσεις και τελικά επιλέχθηκε ήταν η C++.

1.2 Δ OMH KEIMENOY

Ο σκοπός αυτού του τεύχους είναι η εξοικείωση του αναγνώστη με τις βασικές αρχές των δύο γενικών μεθοδολογιών που προαναφέρθηκεν και με μερικούς από τους τρόπους που έχουν προταθεί για την επίλυση σημαντικών προβλημάτων, καθώς και η παρουσίαση των δύο εν τέλει αλγορίθμων που αναπτύχθηκαν. Στο Κεφάλαιο 2 αναφέρονται συνοπτικά ορισμένα βασικά στοιχεία των δύο μεθοδολογιών. Αναλύονται τεχνικές αναλόγως της ρύθμισης του συστήματος, το βασικό γεωμετρικό πρόβλημα που καλούνται να επιλύσουν οι μεθοδολογίες, καθώς και οι διαφορές τους. Το Κεφάλαιο 3 περιλαμβάνει την ανάλυση των δύο αλγορίθμων που υλοποιήθηκαν κατά την εκπόνηση της διπλωματικής εργασίας. Στο Κεφάλαιο 4 παρουσιάζεται η μεθοδολογία αξιολόγησης των αλγορίθμων, και επιπλέον εκτίθενται και αξιολογούνται τα αποτελέσματα που εξήχθησαν με τη χρήση πραγματικών δεδομένων οδομετρίας όσο και ελεύθερης κίνησης σε εσωτερικό χώρο. Στο Κεφάλαιο 5 του τεύχους, τέλος, παρατίθενται τα συμπεράσματα που προέκυψαν από την αξιολόγηση των αλγορίθμων, μια ανάλυση των δυνατοτήτων και των προβλημάτων τους, αλλά και προτάσεις για ενδεχόμενη περαιτέρω μελλοντική έρευνα και ανάπτυξη. Η εργασία ολοκληρώνεται με την βιβλιογραφία που χρησιμοποιήθηκε και παραπέμπεται στην εργασία.

2 Θεώρητικο Υποβάθρο

2.1 ΟΠΤΙΚΗ ΟΔΟΜΕΤΡΙΑ

Η οπτική οδομετρία (Visual Odometry) είναι η διαδικασία προσδιορισμού της κίνησης ενός συστήματος με ανάλυση διαδοχικών δεδομένων από μια κάμερα ή ένα ζεύγος καμερών. Σύμφωνα με τη βιβλιογραφία, η οπτική οδομετρία μπορεί, ανάλογα με τον τύπο των αισθητήρων, να κατηγοριοποιηθεί ως Stereo Visual Odometry (δύο κάμερες σε γνωστή διάταξη), Monocular Visual Odometry (μονοεικονική), RGD-B Visual Odometry (συστήματα με αισθητήρες καταγραφής και βάθους). Η πρώτη κατηγορία μιμείται την ανθρώπινη όραση. Σε αντίθεση με τα συστήματα μιας κάμερας, στην δεύτερη περίπτωση η κλίμακα του μοντέλου είναι προφανώς απευθείας γνωστή. Ωστόσο, τα στερεοσκοπικά συστήματα απαιτούν αυστηρό συγχρονισμό στην καταγραφή της πληροφορίας, χωρίς τον οποίο το σφάλμα συγχρονισμού που μεταδίδεται με την πάροδο του χρόνου έχει ισχυρή επίδραση στην ακρίβεια της επίλυσης (Poddar, Kottath, and Karar 2018).

Στη στερεοσκοπική όραση, η τρισδιάστατη πληροφορία εξάγεται σε ένα χρονικό βήμα με ταυτόχρονη παρατήρηση χαρακτηριστικών σημείων στην αριστερή και τη δεξιά εικόνα, οι οποίες χωρίζονται από μια βάση γνωστής απόστασης. Η κίνηση της κάμερας εκτιμάται με την παρατήρηση χαρακτηριστικών σημείων δύο διαδοχικών χρονικών βημάτων. Στην οπτική οδομετρία με σύστημα μονής κάμερας, η τρισδιάστατη πληροφορία ανακατασκευάζεται με την παρατήρηση χαρακτηριστικών σημείων δύο διαδοχικών χρονικών βημάτων, δηλαδή σε δύο χρονικά βήματα. Σε αντίθεση με τη στερεοσκοπική όραση, η κίνηση της κάμερας, δηλαδή ο μετασχηματισμός που συνδέει τις δύο εικόνες, δεν μπορεί να υπολογιστεί πλήρως. Η κλίμακα του μετασχηματισμού είναι άγνωστη και συνήθως ορίζεται με μια προκαθορισμένη τιμή. Με εισαγωγή μιας τρίτης εικόνας στην επίλυση και εύρεση ομόλογων σημείων και στις τρεις εικόνες, η σχετική κλίμακα και η θέση της κάμερας προσδιορίζονται με γνώση τρισδιάστατης γεωμετρίας του χώρου ή με τον τριεστιακό τανυστή (trifocal tensor). Επομένως, χαρακτηριστικά σημεία θα πρέπει να παρατηρούνται σε τουλάχιστον τρεις διαδοχικές εικόνες (Yousif, Bab-Hadiashar, and Hoseinnezhad 2015).

Οι υπολογισμοί της τροχιάς της κάμερας γίνονται διαδοχικά, ανά καρέ. Τα σφάλματα, έτσι, αυτών των υπολογισμών γενικά συσσωρεύονται με το πέρασμα του χρόνου, με

2. Θεωρητικό Υπόβαθρο

αποτέλεσμα η τρέχουσα θέση της κάμερας να αποκλίνει από την πραγματική. Το φαινόμενο αυτό ονομάζεται ολίσθηση (drift). Οι μεθοδολογίες οπτικής οδομετρίας δεν λύνουν το πρόβλημα της ολίσθησης συνολικά (globally), αλλά προσπαθούν να το ελαχιστοποιήσουν με μεθόδους βελτιστοποίησης, όπως π.χ. με συνόρθωση δέσμης κινούμενου παραθύρου (sliding window bundle adjustment). Σε αυτή τη μέθοδο γίνεται συνόρθωση στις τελευταίες *m* εικόνες σε αντίθεση με τη συνηθισμένη συνόρθωση δέσμης σε ολόκληρο το μοντέλο. Η ολίσθηση μπορεί να μειωθεί με το συνδυασμό πληροφορίας από άλλους αισθητήρες, όπως αδρανειακά συστήματα και GPS.

Η οπτική οδομετρία χωρίζεται σε τρεις υποκατηγορίες ανάλογα με τον τρόπο εξαγωγής της πληροφορίας από τις εικόνες και τον τρόπο συνταύτισης: μέθοδοι βάσει χαρακτηριστικών σημείων (feature-based), άμεσες μέθοδοι (direct ή appearance-based) και στις υβριδικές (hybrid). Οι μέθοδοι χαρακτηριστικών σημείων βασίζονται σε διακριτά χαρακτηριστικά σημεία τα οποία ανιχνεύονται στις επικαλυπτόμενες εικόνες. Οι άμεσες μέθοδοι βασίζονται στην πληροφορία της έντασης των εικονοψηφίδων της εικόνας ή επιμέρους περιοχών της. Τέλος, οι υβριδικές μέθοδοι χρησιμοποιούν ένα συνδυασμό των προηγούμενων δύο (Fraundorfer and Scaramuzza 2011).

2.1.1 ΜΕΘΟΔΟΙ ΒΑΣΕΙ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΣΗΜΕΙΩΝ

Όπως προαναφέρθηκε, οι μέθοδοι αυτές βασίζονται στον εντοπισμό και την ανίχνευση δισδιάστατων σημείων σε διαδοχικές εικόνες (προβολές του ίδιου σημείου του φυσικού τρισδιάστατου χώρου). Ο μεγαλύτερος αριθμός εφαρμογών είναι βασισμένος σε αυτή τη μεθοδολογία, και τούτο λόγω της πληθώρας διαθέσιμων αλγορίθμων για τον εντοπισμό και την συνταύτιση χαρακτηριστικών σημείων μεταξύ εικόνων μεγάλης σχετικά βάσης. Σημαντικό μειονέκτημα μπορεί να θεωρηθεί η εξάρτηση των αλγορίθμων εντοπισμού και συνταύτισης από τις τιμές των παραμέτρων και κατωφλίων τους, με αποτέλεσμα την ανάγκη ρύθμισης και επιλογής τους ανάλογα με τη δομή του περιβάλλοντος στο οποίο θα λειτουργήσουν. Καθώς οι περισσότεροι αλγόριθμοι ρυθμίζονται με κύριο μέλημα την ταχύτητα εκτέλεσης παρά την ακρίβεια, είναι απαραίτητη η χρήση αξιόπιστων τεχνικών εκτίμησης και αντιμετώπισης λανθασμένων ομολογιών (Forster, Pizzoli, and Scaramuzza 2014).

Υπάρχουν δύο κύριες προσεγγίσεις για την αντιστοίχιση χαρακτηριστικών σημείων. Η πρώτη είναι με την ανίχνευση χαρακτηριστικών σημείων μιας εικόνας με τοπικές μεθόδους συσχέτισης, γνωστή στη βιβλιογραφία ως οπτική ροή (optical flow). Η δεύτερη είναι με την αντιστοίχιση ανεξάρτητα εξαγμένων χαρακτηριστικών σημείων των εικόνων βάσει κάποιας μετρικής ομοιότητας μεταξύ των περιγραφέων τους (Fraundorfer and Scaramuzza 2012).

Δεδομένου ότι τα χαρακτηριστικά σημεία είναι γενικά τα κύρια, αν όχι τα μοναδικά, δεδομένα μιας εφαρμογής, ο συνολικός αλγόριθμος του συστήματος είναι συχνά τόσο καλός όσο και ο αλγόριθμος ανίχνευσης. Επιθυμητές ιδιότητες που θα πρέπει να έχει ένας αλγόριθμος ανίχνευσης χαρακτηριστικών σημείων είναι (Fraundorfer and Scaramuzza 2012):

- Ακρίβεια εντοπισμού (Accuracy). Οι συντεταγμένες του σημείου στην εικόνα θα πρέπει να είναι ακριβείς υπό συνθήκες διαφορετικής κλίμακας.
- Διακριτότητα (Distinctness). Τα χαρακτηριστικά θα πρέπει να μπορούν να ανιχνευθούν με ακρίβεια σε πολλαπλές εικόνες.
- Ταχύτητα υπολογισμού (Computational Efficiency). Η επεξεργασία της εικόνας θα πρέπει να γίνεται σε πραγματικό χρόνο.
- Ανθεκτικότητα (Robustness). Δεν θα πρέπει να επηρεάζονται από θόρυβο στις εικόνες και θολότητα (συνήθως από απότομη κίνηση της κάμερας).
- Σταθερότητα (Invariance). Να μπορούν να εντοπισθούν υπό συνθήκες διαφορετικής φωτεινότητας, αλλαγών στην προοπτική, κλίμακας και στροφής.

Οι ανιχνευτές χαρακτηριστικών σημείων (σημειακοί τελεστές) μπορούν να διαχωριστούν σε δύο κατηγορίες, σε ανιχνευτές κορυφών (corners), όπως οι Harris (Harris and Pike 1988), FAST (Rosten and Drummond 2006), Shi-Tomasi (Shi and Tomasi 1993), ORB (Rublee et al. 2011), και σε ανιχνευτές κηλίδων (blobs), όπως οι SIFT (Lowe 2004), SURF (Bay, Tuytelaars, and Van Gool 2006), CENSURE (Agrawal, Konolige, and Blas 2008). Μια κορυφή ορίζεται ως το σημείο τομής δύο η περισσότερων ακμών. Μια κηλίδα είναι ένας σχηματισμός στην εικόνα που διαφέρει από το άμεσο περιβάλλον του κατά την ένταση, το χρώμα και την υφή· δεν είναι ακμή ούτε κορυφή (Fraundorfer and Scaramuzza 2012). Γενικά οι ανιχνευτές κορυφών είναι υπολογιστικά ταχύτεροι σε σχέση με τους ανιχνευτές κηλίδων, αλλά είναι δυσκολότερο να αντιστοιχηθούν, λόγω της ομοιομορφίας τους. Οι κηλίδες, από την άλλη μεριά, μπορούν να ανιχνευθούν πιο εύκολα σε διαφορετικές κλίμακες σε σχέση με τις κορυφές, αλλά παρέχουν μικρότερη ακρίβεια θέσης στην εικόνα (Fraundorfer and Scaramuzza 2012). Η επιλογή του κατάλληλου ανιχνευτή χαρακτηριστικών θα πρέπει να εξεταστεί προσεκτικά, ανάλογα με τις υπολογιστικές απαιτήσεις, τον τύπο περιβάλλοντος όπου θα λειτουργήσει η εφαρμογή και τη μέση απόσταση μεταξύ των θέσεων λήψης (Fraundorfer and Scaramuzza 2012). Επισκοπήσεις των βασικών αλγορίθμων καθώς και συγκρίσεις των δυνατοτήτων τους

2. Θεωρητικό Υπόβαθρο

έχουν γίνει από τους Schmidt, Kraft, and Kasiński (2010), Peng (2012), Hartmann, Klussendorff, and Maehle (2013) και Tareen and Saleem (2018).

Οι αλγόριθμοι εξαγωγής χαρακτηριστικών λειτουργούν συνήθως σε δύο βήματα. Στο πρώτο βήμα εφαρμόζεται μια συνάρτηση απόκρισης χαρακτηριστικών στην εικόνα (π.χ. corner response function για τον αλγόριθμο Harris και difference-of-Gaussian για τον SIFT). Σε δεύτερο βήμα ακολουθεί η εύρεση των χαρακτηριστικών σημείων, τα οποία είναι τα τοπικά ακρότατα της συνάρτησης του πρώτου βήματος. Προκειμένου να εξασφαλιστεί κάποια ανεξαρτησία από την εκάστοτε κλίμακα απεικόνισης, το βήμα αυτό εφαρμόζεται σε διαφορετικές αναλύσεις της ίδιας εικόνας (Fraundorfer and Scaramuzza 2012).

Ωστόσο, μόνη η θέση των χαρακτηριστικών σημείων δεν επαρκεί για την ανίχνευσή τους σε άλλες εικόνες, επομένως για την επιτυχή αντιστοίχισή τους θα πρέπει να τους προσδοθεί ένας τύπος περιγραφής. Ο πιο απλός περιγραφέας είναι η "εμφάνιση" του χαρακτηριστικού, δηλαδή οι εντάσεις των γειτονικών εικονοψηφίδων του (pixel) σε ένα προκαθορισμένο παράθυρο. Σε αυτή την περίπτωση, μπορούν να χρησιμοποιηθούν το άθροισμα των τετραγωνικών διαφορών (SSD) ή ο συντελεστής συσχέτισης (NCC) για τη σύγκριση δύο χαρακτηριστικών (Fraundorfer and Scaramuzza 2012). Στη βιβλιογραφία έχουν προταθεί αρκετές μεθοδολογίες δημιουργίας περιγραφέων χαρακτηριστικών. Ο περιγραφέας SIFT (Lowe 2004) δημιουργεί ένα διάνυσμα με 128 κανονικοποιημένους προσανατολισμούς κλίσεων γειτονικών εικονοψηφίδων του χαρακτηριστικού σημείου. Αυτό τον καθιστά ανθεκτικό σε αλλαγές φωτεινότητας, στροφής και κλίμακας, τον κάνει όμως πολύ απαιτητικό υπολογιστικά. Ο περιγραφέας BRIEF (Calonder et al. 2010) κατασκευάζει ένα διάνυσμα 128 ή 256 δυαδικών τιμών από συγκρίσεις φωτεινότητας ζευγών εικονοψηφίδων της γειτονικής περιοχής του χαρακτηριστικού σημείου. Είναι υπολογιστικά ταχύτερος από τον SIFT, παρουσιάζει υψηλή διακριτή ισχύ, αλλά αδυνατεί σε μεγάλες αλλαγές προσανατολισμού και κλίμακας. Ο τελεστής FAST (Rosten and Drummond 2006) ορίζει ένα σημείο ως χαρακτηριστικό όταν η ένταση της τιμής του διαφέρει σημαντικά από τις τιμές των 16 περιμετρικών επί κύκλου σημείων. Δεν εξάγει περιγραφείς για τα χαρακτηριστικά σημεία, είναι πιο γρήγορος υπολογιστικά από τις άλλες μεθόδους, αλλά δεν είναι ανθεκτικός σε θόρυβο. Ο περιγραφέας ORB (Rublee et al. 2011) είναι ένας συνδυασμός παραλλαγών του αλγορίθμου ανίχνευσης FAST και του περιγραφέα BRIEF, ο οποίος αντιμετωπίζει προβλήματα διαφορετικής στροφής και κλίμακας.

8

2. Θεωρητικό Υπόβαθρο

Ο απλούστερος τρόπος αντιστοίχισης χαρακτηριστικών δύο εικόνων είναι με σύγκριση όλων των χαρακτηριστικών στην πρώτη εικόνα με όλα της δεύτερης εικόνας. Ένα μειονέκτημα αυτής της εξαντλητικής αντιστοίχισης προκύπτει όσο αυξάνεται ο αριθμός των χαρακτηριστικών (Fraundorfer and Scaramuzza 2012). Μια πιο πρόσφορη προσέγγιση είναι η χρησιμοποίηση μιας δομής δεικτών, όπως ένα πολυδιάστατο δέντρο αναζήτησης (k-d free). Μια ακόμα ταχύτερη αντιστοίχιση χαρακτηριστικών γίνεται με τον περιορισμό της αναζήτησης χαρακτηριστικών σε περιοχές της δεύτερης εικόνας όπου αναμένεται ότι θα βρίσκεται η ομολογία. Αυτό μπορεί να επιτευχθεί με τη χρήση μοντέλων κίνησης από εξωτερικό αισθητήρα ή αναζήτηση αντιστοιχίας πάνω στις επιπλοκές γραμμές (βλ. ενότητα 2.1.4) (Fraundorfer and Scaramuzza 2012).

Μια εναλλακτική μεθοδολογία συνταύτισης είναι με την εύρεση χαρακτηριστικών σε μια εικόνα και αναζήτηση της πιθανής θέσης τους στις επόμενες με μεθόδους συσχέτισης. Αυτή η προσέγγιση είναι κατάλληλη όταν οι εικόνες λαμβάνονται από κοντινές θέσεις. Η μεθοδολογία είναι γνωστή ως οπτική ροή. Ως οπτική ροή χαρακτηρίζεται η φαινόμενη κίνηση ενός αντικειμένου μεταξύ δύο διαδοχικών εικόνων, η οποία προκαλείται από την κίνηση της κάμερας ή του αντικειμένου. Ο πιο διαδεδομένος αλγόριθμος οπτικής ροής είναι ο Kanade-Lucas-Tomasi (KLT) (Shi and Tomasi 1993). Αυτός υποθέτει ότι η ροή είναι ουσιαστικά σταθερή σε μια τοπική γειτονιά 3x3 της υπό εξέταση εικονοψηφίδας, και η κίνηση μπορεί να μοντελοποιηθεί ως:

$$L(x, y, t + r) = L(x - x_{disp}(x, y, t, r), y - y_{disp}(x, y, t, r))$$
(2.1)

Μια εικόνα της χρονικής στιγμής t + r μπορεί να ληφθεί με τη μετακίνηση των εικονοψηφίδων $\mathbf{x} = (\mathbf{x}, \mathbf{y})$ της εικόνας της στιγμής t κατά το διάνυσμα μετακίνησης $\boldsymbol{\delta} = (\mathbf{x}_{disp}, \mathbf{y}_{disp})$. Οι μετατοπίσεις είναι διαφορετικές στις εικονοψηφίδες μιας εικόνας, επομένως είναι πιο αντιπροσωπευτικός ο ορισμός του διανύσματος ως μιας αφινικής κίνησης:

$$\delta = Dx + d, \tag{2.2}$$

όπου

$$D = \begin{bmatrix} d_{xx} & d_{xy} \\ d_{yx} & d_{yy} \end{bmatrix},$$
 (2.3)

ο πίνακας παραμόρφωσης και d η μετάθεση του κέντρου του παραθύρου του χαρακτηριστικού. Τονίζεται ότι οι συντεταγμένες x της εξίσωσης (2.3) αναφέρονται στο κέντρο του παραθύρου. Έτσι, ένα σημείο x στην πρώτη εικόνα I μετατοπίζεται στο σημείο Ax + d στη δεύτερη εικόνα J, όπου A = I + D, με I τον2x2 μοναδιαίο πίνακα:

$$J(Ax+d) = L(x)$$
(2.4)

Το μοντέλο κίνησης της εξίσωσης (2.4) περιγράφει ένα ιδανικό μοντέλο, αλλά δεν ισχύει απόλυτα λόγω της παρουσίας θορύβου στην εικόνα. Το πρόβλημα υπολογισμού των παραμέτρων κίνησης υπολογίζεται ως:

$$\in = \iint_{W} [J(Ax+d) - L(x)]^2 w(x) dx,$$
(2.5)

όπου W είναι το δεδομένο παράθυρο του χαρακτηριστικού και w(x) η συνάρτηση βάρους. Στη απλούστερη περίπτωση η συνάρτηση βάρους ισούται με μονάδα. Εναλλακτικά, μπορεί να είναι μια συνάρτηση κανονικής κατανομής η οποία θα δίνει έμφαση στο κέντρο του παραθύρου. Στη συνέχεια γίνεται γραμμικοποίηση του ολοκληρώματος με ανάπτυγμα κατά Taylor:

$$J(Ax + d) = J(x) + g^{T}(u),$$
(2.6)

και, τέλος, ελαχιστοποίηση με την επαναληπτική μεθοδολογία Newton-Raphson. Αναλυτική παρουσίαση του αλγορίθμου βρίσκεται στις δημοσιεύεις των Shi and Tomasi (1993) και Baker and Matthews (2004). Η εφαρμογή του αλγορίθμου σε διαφορετικές αναλύσεις τις εικόνας δίνει τη δυνατότητα ανταπόκρισης σε σχετικά μεγάλες κινήσεις της κάμερας.

2.1.2 Άμεσες Μεθοδοι

Οι άμεσες μέθοδοι εκτιμούν τη δομή και την κίνηση της κάμερας απευθείας από τις τιμές έντασης των εικόνων. Σε αντίθεση με τις μεθόδους χαρακτηριστικών σημείων που ελαχιστοποιούν το σφάλμα επαναπροβολής, οι άμεσες μέθοδοι ελαχιστοποιούν ένα ραδιομετρικό σφάλμα, και δείχνουν να λειτουργούν καλύτερα σε περιοχές με φτωχή υφή ή σε περιπτώσεις αλλαγής εστίασης της κάμερας και θολότητας εξ αιτίας γρήγορης κίνησης (Forster, Pizzoli, and Scaramuzza 2014). Οι άμεσες μέθοδοι μπορούν να χρησιμοποιήσουν όλες τις εικονοψηφίδες της εικόνας (πυκνές προσεγγίσεις) (Kerl, Sturm, and Cremers 2013; Newcombe, Lovegrove, and Davison 2011), εικονοψηφίδες με μεγάλη κλίση έντασης (ημι-πυκνές προσεγγίσεις) (Engel, Schöps, and Cremers 2014; Engel, Stückler, and Cremers 2015; Engel, Sturm, and Cremers 2013; Forster, Pizzoli, and Scaramuzza 2014), ή apaia επιλεγμένες εικονοψηφίδες (apaiές προσεγγίσεις) (Engel, Koltun, and Cremers 2018). Και οι τρεις προσεγγίσεις ελαχιστοποιούν το ραδιομετρικό σφάλμα μεταξύ δύο εικόνων. Αυτό μπορεί να θεωρηθεί ως επέκταση του αλγορίθμου KLT στις τρεις διαστάσεις. Λεπτομερή επεξήγηση αυτής της επέκτασης περιγράφουν οι Irani and Anandan (1999).

2.1.3 ΚΑΜΕΡΑ ΚΑΙ ΒΑΘΜΟΝΟΜΗΣΗ

Ο κύριος αισθητήρας ενός συστήματος οπτικής οδομετρίας είναι η κάμερα. Για έναν αποτελεσματικό και λειτουργικό αλγόριθμο είναι σημαντική η γνώση του γεωμετρικού μοντέλου μιας κάμερας. Η κάμερα είναι ένα μέσο καταγραφής της ανακλώμενης ακτινοβολίας ενός τρισδιάστατου αντικειμένου σε ένα επίπεδο. Στη γενική της μορφή, το ανακλώμενο φως από το περιβάλλον προβάλλεται, μέσω του οπτικού κέντρου του φακού, σε μια φωτο-ευαίσθητη επιφάνεια η οποία το μετατρέπει σε ηλεκτρικό σήμα. Οι φωτοευαίσθητες επιφάνειες (αισθητήρες) σε μια ψηφιακή μηχανή είναι του τύπου Charge-Coupled Device (CCD) ή Complementary Metal-Oxide Semiconductor (CMOS). Και οι δύο τύποι αισθητήρων χρησιμοποιούν την ίδια αρχή μετατροπής των φωτονίων σε ηλεκτρόνια. Η διαφορά τους βρίσκεται στον τρόπο κατασκευής και στην ποιότητα των εικόνων βάσει αυτού. Πάνω στους αισθητήρες βρίσκονται διατεταγμένα έως και μερικά εκατομμύρια στοιχεία ενός ημιαγώγιμου υλικού ευαίσθητου στο φως, τα οποία μετρούν την ένταση της φωτεινότητας που προσπίπτει πάνω τους και την αποθηκεύουν ως αριθμό. Μια ψηφιακή εικόνα είναι ουσιαστικά ένας κάναβος του οποίου το κάθε κελί περιέχει μια τιμή ακτινοβολίας που αντιστοιχεί σε ένα χρώμα.

2.1.3.1 Γεωμετρία Κάμερας

Ο απλούστερος τύπος φωτογραφικής μηχανής που μπορεί να βρεθεί στον πραγματικό κόσμο είναι η κάμερα σημειακής οπής, ή αλλιώς σκοτεινός θάλαμος. Είναι ένα απλό κουτί με μια πολύ μικρή τρύπα στο μπροστινό μέρος και ένα φωτοευαίσθητο υλικό στην απέναντι εσωτερική μεριά. Το φως μιας σκηνής διέρχεται μέσα από την οπή και προβάλλει μια ανεστραμμένη εικόνα στην αντίθετη μεριά του κουτιού. Το μοντέλο λειτουργίας μια τέτοιας μηχανής είναι πολύ απλό, όπως παρουσιάζεται στο βιβλίο των Hartley και Zisserman (Hartley and Zisserman 2003).

Έστω εστιακό επίπεδο το επίπεδο στο οποίο προσπίπτουν οι ακτίνες. Το κέντρο προβολής C από το οποίο διέρχονται οι ακτίνες είναι η αρχή ενός Ευκλείδειου συστήματος

11

συντεταγμένων. Μια ευθεία με αρχή ένα σημείο στο χώρο $X = (X, Y, Z)^T$, η οποία διέρχεται από το κέντρο προβολής C, προσπίπτει στο εστιακό επίπεδο Z = f σε θέση με συντεταγμένες $x = (x, y)^T$. Ένα σημείο του τρισδιάστατου Ευκλείδειου χώρου $\mathbb{R}^3 (X, Y, Z)^T$ αντιστοιχεί σε ένα σημείο του δισδιάστατου Ευκλείδειου χώρου $\mathbb{R}^2 (fX/Z, fY/Z, f)^T$ το οποίο βρίσκεται πάνω στο εστιακό επίπεδο. Αν αγνοήσει κανείς την τελευταία συντεταγμένη της εικόνας, η σχέση γράφεται ως:

$$(X, Y, Z)^T \mapsto (fX/Z, fY/Z)^T$$
(2.7)

Η κάθετος προς το επίπεδο της εικόνας με αρχή το προβολικό κέντρο ονομάζεται κύριος άξονας ή κύρια ακτίνα της κάμερας, ενώ το σημείο ρ όπου ο κύριος άξονας συναντά το επίπεδο της εικόνας ονομάζεται πρωτεύον σημείο. Οι σχέσεις παρουσιάζονται στην Εικόνα 2.1.



Εικόνα 2.1: Εσωτερική γεωμετρία μηχανής (Hartley and Zisserman 2003).

Αν το σημείο του χώρου και το σημείο της εικόνας αντιπροσωπευθούν με ομογενείς συντεταγμένες, τότε η μεταξύ τους σχέση περιγράφεται από το παρακάτω γινόμενο πινάκων:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} = \begin{bmatrix} f & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$
(2.8)

Η σχέση (2.8) γράφεται ως x = PX, όπου x οι ομογενείς συντεταγμένες του σημείου στο προβολικό επίπεδο, X οι ομογενείς συντεταγμένες του σημείου στο χώρο και P ο προβολικός πίνακας της κάμερας (camera projection matrix).

Όπως αναφέρθηκε προηγουμένως, το εν λόγω μοντέλο περιγράφει μια απλή υπόθεση γεωμετρίας μηχανής. Για να περιγραφεί ένα πιο ρεαλιστικό μοντέλο θα πρέπει να ληφθούν υπόψη κάποιες επιπλέον παράμετροι. Στις παραπάνω εξισώσεις υπάρχει η υπόθεση ότι το πρωτεύον σημείο έχει συντεταγμένες (x = 0, y = 0), ενώ στην πραγματικότητα έχει συντεταγμένες (p_x, p_y) σε σχέση με την αρχή του συστήματος συντεταγμένων της εικόνας. Το σύστημα συντεταγμένων της εικόνας παρουσιάζεται στην Εικόνα 2.5. Η σχέση (2.8) γράφεται ως:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{bmatrix} f & p_x & 0 \\ f & p_y & 0 \\ & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$
(2.9)

όπου p_x , p_y οι συντεταγμένες του πρωτεύοντος σημείου και

$$\boldsymbol{K} = \begin{bmatrix} f & p_x \\ f & p_y \\ & 1 \end{bmatrix}$$
(2.10)

Ο πίνακας **K** ονομάζεται πίνακας μηχανής (camera matrix) και περιγράφει την εσωτερική γεωμετρία της μηχανής. Η συνοπτική μορφή της εξίσωσης (2.9) γίνεται **x** = **K**[**I** |**0**]**X**.



Εικόνα 2.2: Σχέση των συστημάτων της εικόνας (Hartley and Zisserman 2003).

Λόγω πιθανής μη τετραγωνικότητας των εικονοψηφίδων ενός αισθητήρα μηχανής (CCD, CMOS), εάν οι συντεταγμένες της εικόνας μετρούνται σε εικονοψηφίδες τότε αυτό έχει το επιπλέον αποτέλεσμα της εισαγωγής άνισων συντελεστών κλίμακας σε κάθε διεύθυνση. Η μορφή του πίνακα βαθμονόμησης **K** (2.10) γίνεται τότε:

$$K = \begin{bmatrix} a_x & x_o \\ a_y & y_o \\ & 1 \end{bmatrix}$$
(2.11)

όπου $a_x = fm_x$, $a_y = fm_y$, $x_o = m_x p_x$ και $y_o = m_y p_y$, με m_x και m_y τον αριθμό των εικονοψηφίδων ανά μονάδα απόστασης για τις διευθύνσεις x και y, αντίστοιχα.

Σε ορισμένες περιπτώσεις ο πίνακας **Κ** μπορεί να περιέχει μια επιπρόσθετη παράμετρο s, η οποία περιγράφει τη μη ορθογωνιότητα των αξόνων x, y. Το φαινόμενο αυτό ονομάζεται στρέβλωση (skewness) και προκαλεί παραμόρφωση στην προβαλλόμενη εικόνα. Ο πίνακας **Κ** θα έχει πλέον τη μορφή:

$$K = \begin{bmatrix} a_x & s & x_o \\ & a_y & y_o \\ & & 1 \end{bmatrix}$$
(2.12)

Τελευταία υπόθεση είναι ότι τα τρισδιάστατα σημεία εκφράζονται σε ένα διαφορετικό Ευκλείδειο σύστημα συντεταγμένων, το σύστημα αναφοράς χώρου (world coordinate system), το οποίο δεν συμπίπτει με εκείνο της κάμερας. Η σχέση των δύο συστημάτων εκφράζεται από έναν μετασχηματισμό στερεού σώματος, δηλαδή μια μετάθεση και μια στροφή στον χώρο. Η συνοπτική μορφή της εξίσωσης γίνεται:

$$\boldsymbol{x} = \boldsymbol{K}\boldsymbol{R}[\boldsymbol{I} \mid -\boldsymbol{C}]\boldsymbol{X} \tag{2.13}$$

όπου **R** είναι ο πίνακας στροφής διαστάσεων 3x3 και **C** είναι οι συντεταγμένες της κάμερας στο σύστημα αναφοράς του χώρου.

Οι παράμετροι που περιέχει ο πίνακας **Κ** περιγράφουν την εσωτερική γεωμετρία της μηχανής και ο πίνακας ονομάζεται επίσης πίνακας ενδογενών παραμέτρων (intrinsic). Οι παράμετροι των **R** και **C**, οι οποίοι συνδέουν το σύστημα της μηχανής με το σύστημα συντεταγμένων του χώρου, ονομάζονται εξωγενείς παράμετροι (extrinsic).



Εικόνα 2.3: Σχέση μεταξύ του συστήματος αναφοράς χώρου και του συστήματος της κάμερας (Hartley and Zisserman 2003).

2.1.3.2 ΠΑΡΑΜΟΡΦΩΣΕΙΣ ΦΑΚΟΥ

Οι παραμορφώσεις του φακού είναι φαινόμενα που στις περισσότερες φωτογραμμετρικές εφαρμογές απαγορεύουν την απευθείας εφαρμογή του απλού μοντέλου της κάμερας σημειακής οπής της προηγούμενης ενότητας (απλή κεντρική προβολή). Οι παραμορφώσεις δεν επηρεάζουν την ποιότητα της εικόνας αλλά έχουν σημαντική επίδραση στη γεωμετρία της. Θεωρητικά μπορεί να οριστεί ένας φακός ο οποίος δεν θα εισάγει παραμορφώσεις. Στην πράξη ωστόσο κανένας φακός δεν είναι τέλειος. Αυτό συμβαίνει κυρίως για κατασκευαστικούς λόγους, καθώς είναι απλούστερο να κατασκευαστεί ένας σφαιρικός φακός παρά ένας ιδανικός παραβολικός (Πέτσα 2000). Είναι επίσης δύσκολο να ευθυγραμμιστεί μηχανικά ο φακός και ο αισθητήρας της μηχανής.

Τα δύο κύρια είδη παραμόρφωσης είναι η συμμετρική ακτινική παραμόρφωση (radial symmetric distortion) και η έκκεντρη παραμόρφωση (decentering distortion). Η πρώτη περιγράφει το φαινόμενο όπου οι φωτεινές ακτίνες δεν εξέρχονται με την ίδια γωνία υπό την οποία προσπίπτουν στο φακό, με αποτέλεσμα οι ευθείες του χώρου να μην απεικονίζονται στην εικόνα ως ευθείες αλλά καμπυλωμένες. Η συμμετρική ακτινική διαστροφή εισάγει μια μεταβολή κλίμακας που είναι ενιαία για κάθε ακτινική απόσταση στην εικόνα (Πέτσα 2000). Οι αποκλίσεις που προκαλεί είναι μηδενικές στο κέντρο της εικόνας, δηλαδή στο πρωτεύον σημείο, και μεγιστοποιούνται στις τέσσερις γωνίες της. Η έκεντρη παραμόρφωση, η οποία αναλύεται στην μη συμμετρική ακτινική και την εφαπτομενική (tangential) διαστροφή, προκύπτει από τη μη παραλληλία του φακού και του επιπέδου της εικόνας.

Οι παραμορφώσεις αυτές περιγράφονται από το μοντέλο των Brown-Conrady (Brown 1966):

$$x_{u} = x_{d} + (x_{d} - x_{0})(K_{1}r^{2} + K_{2}r^{4} + \dots) + (P_{1}(r^{2} + 2(x_{d} - x_{0})^{2}) + 2P_{2}(x_{d} - x_{0})(y_{d} - y_{0}))(1 + P_{3}r^{2} + P_{4}r^{4} + \dots) y_{u} = y_{d} + (y_{d} - y_{0})(K_{1}r^{2} + K_{2}r^{4} + \dots) + (2P_{1}(x_{d} - x_{0})(y_{d} - y_{0}) + P_{2}(r^{2} + 2(y_{d} - y_{0})^{2}))(1 + P_{3}r^{2} + P_{4}r^{4} + \dots)$$

$$(2.14)$$

όπου (x_u, y_u) οι διορθωμένες συντεταγμένες του σημείου της εικόνας, (x_d, y_d) οι παραμορφωμένες συντεταγμένες του σημείου της εικόνας, (x_0, y_0) οι συντεταγμένες του πρωτεύοντος σημείου, K_n οι συντελεστές της συμμετρικής ακτινικής παραμόρφωσης, P_n οι συντελεστές της έκκεντρης παραμόρφωσης και $r = \sqrt{(x_d - x_0)^2 + (y_d - y_0)^2}$ η ακτινική απόσταση από το πρωτεύον σημείο. Οι συντελεστές της ακτινικής και της έκκεντρης παραμόρφωσης δίνονται από τον κατασκευαστή (στις λεγόμενες μετρικές μηχανές) ή υπολογίζονται μαζί με τα υπόλοιπα στοιχεία της εσωτερικής γεωμετρίας κατά τη διαδικασία βαθμονόμησης της μηχανής.



Εικόνα 2.4: Παράδειγμα επίδρασης της παραμόρφωσης λόγω φακού.

2.1.3.3 ΒΑΘΜΟΝΟΜΗΣΗ ΜΗΧΑΝΗΣ

Σε αυτή την ενότητα γίνεται μια μικρή αναφορά στη διαδικασία εύρεσης των παραμέτρων της εσωτερικής γεωμετρίας της μηχανής (πίνακας **K**) μαζί με τις πρόσθετες παραμέτρους που περιγράφουν τις παραμορφώσεις του φακού. Πιο συγκεκριμένα θα παρουσιαστεί η μεθοδολογία βαθμονόμησης του Zhang (Zhang 2000), η οποία είναι εκείνη που χρησιμοποιήθηκε στην παρούσα εργασία. Η γνώση των παραμέτρων της μηχανής είναι απαραίτητη για την πλειονότητα των αλγορίθμων φωτογραμμετρίας και όρασης υπολογιστών προκειμένου να εξαχθούν μετρικές πληροφορίες από εικόνες. Η βαθμονόμηση μηχανής θεωρείται από την φωτογραμμετρική κοινότητα ως μια από τις πιο περίπλοκες διαδικασίες. Οι τεχνικές υλοποίησης της βαθμονόμησης χωρίζονται σε δύο κατηγορίες, την τυπική φωτογραμμετρική βαθμονόμηση και την αυτοβαθμονόμηση.

- Οι τεχνικές τυπικής φωτογραμμετρικής βαθμονόμησης γίνονται με τη χρήση ενός αντικειμένου ή επιφάνειας βαθμονόμησης, ή συνδυασμού τους, των οποίων η γεωμετρία στον τρισδιάστατο χώρο είναι γνωστή με πολύ καλή ακρίβεια.
- Οι τεχνικές αυτοβαθμονόμησης δεν χρησιμοποιούν αντικείμενα βαθμονόμησης. Με καταγραφές μιας σταθερής σκηνής από μια μηχανή είναι δυνατή η ανάκτηση της εσωτερικής γεωμετρίας της καθώς και των παραμέτρων που καθιστούν δυνατή την πιστή τρισδιάστατη ανακατασκευή της σκηνής.

2. Θεωρητικό Υπόβαθρο

Η προαναφερθείσα δημοφιλής μέθοδος βαθμονόμησης της κάμερας κατά τον Ζ. Zhang χρησιμοποιεί απεικονίσεις, από διαφορετικές θέσεις και προσανατολισμούς, ενός επίπεδου σχηματισμού βαθμονόμησης, του οποίου η διάταξη και οι μετρικές διαστάσεις είναι απολύτως γνωστές. Η προσέγγιση μπορεί να θεωρηθεί ως συνδυασμός τυπικής φωτογραμμετρικής βαθμονόμησης και αυτοβαθμονόμησης καθώς χρησιμοποιούνται δισδιάστατες μετρικές πληροφορίες και όχι τρισδιάστατες. Η κίνηση της κάμερας ή του προτύπου δεν είναι απαραίτητο να είναι γνωστή. Η διαδικασία βαθμονόμησης έχει ως εξής:

- Λήψη τουλάχιστον 3 εικόνων με διαφορετική προοπτική (με μετακίνηση είτε της φωτογραφικής μηχανής είτε του προτύπου).
- 2. Εξαγωγή των χαρακτηριστικών σημείων του προτύπου σε κάθε εικόνα. Για παράδειγμα, στην συνήθη περίπτωση ενός σχηματισμού τύπου σκακιέρας, χαρακτηριστικά σημεία θα είναι οι κορυφές των κελιών. Είναι απαραίτητη η αντιστοίχιση των εικονοσημείων που έχουν εξαχθεί με τα σημεία του επίπεδου προτύπου.
- 3. Από τα παρατηρούμενα σημεία υπολογίζεται η σχέση ομογραφίας κάθε εικόνας με το επίπεδο του χώρου. Μια ομογραφία περιγράφει τον προβολικό μετασχηματισμό μεταξύ δύο επιπέδων, εν προκειμένω του επιπέδου της πραγματικής σκακιέρας και της προβολής της στη δισδιάστατη εικόνα.
- Από τις ομογραφίες γίνεται υπολογισμός των 5 παραμέτρων του εσωτερικού προσανατολισμού, αγνοώντας για την ώρα τις παραμορφώσεις της ακτινικής διαστροφής.
- 5. Με γνωστές πλέον τις εσωτερικές παραμέτρους, γίνεται υπολογισμός των εξωτερικών παραμέτρων **R**, **f** για την κάθε εικόνα.
- 6. Οι παράμετροι της ακτινικής διαστροφής *k*₀, *k*₁ μπορούν να εκτιμηθούν με μια γραμμική ελαχιστοποίηση βάσει ελαχίστων τετραγώνων.
- Τέλος, με χρήση των εκτιμώμενων τιμών των παραμέτρων ως προσεγγιστικών, όλες οι παράμετροι διορθώνονται με μη γραμμική βελτιστοποίηση.



Εικόνα 2.5: Παράδειγμα λήψης εικόνων σκακιέρας διαφορετικού προσανατολισμού για τη διαδικασία της βαθμονόμησης (Gary Bradski and Kaehler 2008).

Στην παρούσα εργασία χρησιμοποιήθηκαν τα εργαλεία που περιλαμβάνονται στην βιβλιοθήκη OpenCV, η οποία χρησιμοποιεί τη μεθοδολογία του Zhang για τη βαθμονόμηση της μηχανής και το μοντέλο των Brown-Conrady για την έκφραση των παραμέτρων παραμόρφωσης λόγω φακού.

2.1.4 Γεωμετρία του Προβληματός και Εκτίμηση Κίνησης

Ένα σύστημα οπτικής οδομετρίας κατά τα χρονικά βήματα k - 1 και k περιγράφεται με ένα μετασχηματισμό στερεού σώματος (Rigid Body Transformation) $T_{k,k-1} \in \mathbb{R}^{4 \times 4}$ της μορφής:

$$T_{k,k-1} = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ 0 & 1 \end{bmatrix},$$
 (2.15)

όπου $R_{k,k-1} \in SO(3)$ είναι πίνακας στροφής και $t_{k,k-1} \in \mathbb{R}^{3 \times 1}$ το διάνυσμα μετάθεσης. Ο μετασχηματισμός αυτός περιγράφει την κίνηση του συστήματος από τη χρονική στιγμή k - 1 στη στιγμή k. Η τρέχουσα θέση C_k του συστήματος υπολογίζεται από τη σχέση:

$$C_k = C_{k-1} T_k, (2.16)$$

όπου C_{k-1} είναι η προηγούμενη θέση του συστήματος. Αν θεωρηθεί ότι το πρώτο χρονικό βήμα, με θέση $C_0 = I_4$, είναι και η αρχή του συστήματος αναφοράς χώρου, τότε η τρέχουσα θέση υπολογίζεται με τη ενσωμάτωση όλων τον προηγούμενων μετασχηματισμών. Σκοπός λοιπόν είναι ο υπολογισμός των σχετικών μετασχηματισμών $T_{k,k-1}$ από τις εικόνες I_k και I_{k-1} και στη συνέχεια η ενσωμάτωσή τους για τον υπολογισμό της συνολικής θέσης/προσανατολισμού $C_{0:n}$ της κάμερας. Οι μεθοδολογίες εκτίμησης της κίνησης της κάμερας χωρίζονται σε τρεις κατηγορίες ανάλογα με το αν οι αντιστοιχίες των χαρακτηριστικών f_{k-1} και f_k των εικόνων I_{k-1} και I_k ορίζονται σε δύο ή τρεις διαστάσεις (Fraundorfer and Scaramuzza 2011). Στη βιβλιογραφία οι μεθοδολογίες αυτές αναφέρονται ως: 2D-to-2D, 3D-to-3D, 2D-to-3D.

• 2D-to-2D: οι αντιστοιχίες των χαρακτηριστικών f_{k-1} και f_k ορίζονται σε δυο διαστάσεις (συντεταγμένες της εικόνας). Η γεωμετρική σχέση μεταξύ δύο εικόνων I_{k-1} και I_k που καταγράφουν την ίδια σκηνή περιγράφεται από την επιπολική γεωμετρία. Η επιπολική γεωμετρία είναι ανεξάρτητη της δομής της σκηνής και εξαρτάται μόνο από τις εσωτερικές και εξωτερικές παραμέτρους των δύο φωτογραφικών μηχανών. Ένα σημείο Χ του χώρου καταγράφεται από δύο εικόνες, με κέντρα προβολής C_L, C_R για την αριστερή και για τη δεξιά, αντίστοιχα. Το τρίγωνο που σχηματίζουν τα τρία σημεία ονομάζεται επιπολικό επίπεδο π. Οι τομές του επιπέδου αυτού με τα επίπεδα της εικόνας ονομάζονται επιπολικές ευθείες **Ι**_L, **Ι**_R. Η νοητή ευθεία (βάση του στερεοζεύγους) που συνδέει τα δύο κέντρα προβολής τέμνει το επίπεδο κάθε εικόνας. Τα σημεία τομής **e**_L, **e**_R ονομάζονται πόλοι του στερεοζεύγους . Οι κεντρικές προβολές του σημείου X στο επίπεδο κάθε εικόνας είναι x_L , x_R . Όπως μπορεί να παρατηρηθεί στην Εικόνα 2.6α, τα ζευγάρια (x_L, e_L) και (x_R, e_R) ανήκουν στις επιπολικές ευθείες l_L και l_R . Από τα παραπάνω είναι φανερό ότι η αντιστοιχία του σημείου x_L στην δεξιά εικόνα, σε περίπτωση που το x_R δεν είναι γνωστό, πρέπει να αναζητηθεί μόνο κατά μήκος της επιπολικής ευθείας **ι**_R. Αυτό συνιστά την επιπολική δέσμευση.



Εικόνα 2.6: Εικόνα με τις σχέσεις της επιπολικής γεωμετρίας (Hartley and Zisserman 2003).

Η επιπολική δέσμευση περιγράφεται, κατά περίπτωση, από τον δεσμευμένο επιπολικό πίνακα (essential matrix) **E** και τον επιπολικό πίνακα (fundamental matrix) **F**. Η σχέση μεταξύ των δύο πινάκων είναι:

$$\boldsymbol{E} = \boldsymbol{K}_R^T \boldsymbol{F} \boldsymbol{K}_L, \qquad (2.17)$$

όπου K_L και K_R ο πίνακας βαθμονόμησης της αριστερής και της δεξιάς κάμερας, αντίστοιχα, όπως παρουσιάζεται στην εξίσωση (2.12). Ο E, ο οποίος προϋποθέτει βαθμονομημένες εικόνες, περιέχει τις πληροφορίες για τη μετάθεση και την στροφή, οι οποίες περιγράφουν τη σχετική θέση της δεξιάς κάμερας ως προς την αριστερή, δηλαδή τον σχετικό προσανατολισμό τους. Ο F περιέχει τις ίδιες πληροφορίες με τον E με επιπλέον πληροφορίες για την εσωτερική γεωμετρία των δύο καμερών. Οι δύο πίνακες μπορούν να υπολογιστούν από γνωστές αντιστοιχίες σημείων στις δύο εικόνες. Οι δυο ευρέως χρησιμοποιούμενοι αλγόριθμοι για τον υπολογισμό των δύο αυτών πινάκων είναι ο αλγόριθμος των 5 σημείων (5-point algorithm) (Nistér 2004) για τον E και ο αλγόριθμος των 8 σημείων (8-point algorithm) (Longuet-Higgins 1981) για τον F. Η χρήση του αλγορίθμου των 5 σημείων απαιτεί, όπως προαναφέρθηκε, γνώση του πίνακα βαθμονόμησης της μηχανής K, σε αντίθεση με τον αλγόριθμο των 8 σημείων.

Ο επιπολικός πίνακας **F** είναι διαστάσεων 3x3, τάξης 2. Εάν σημείο **X** του τρισδιάστατου χώρου με προβολές x_L στην αριστερή εικόνα και x_R στη δεξιά, τότε τα σημεία αυτά ικανοποιούν τη σχέση (Hartley and Zisserman 2003):

$$\boldsymbol{x}_R^T \boldsymbol{F} \boldsymbol{x}_L = 0 \tag{2.18}$$

Κάθε ζεύγος ομόλογων σημείων $\mathbf{x}_L = (x, y, 1)^T$ και $\mathbf{x}_R = (x', y', 1)^T$ σε ομογενείς συντεταγμένες ορίζει μια εξίσωση για την επίλυση του **F**. Η γραμμική εξίσωση είναι της μορφής:

$$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33}$$
(2.19)

Η εξίσωση (2.19) μπορεί να εκφραστεί ως:

$$(x'x, x'y, x', y'x, y'y, y', x, y, 1)f = 0,$$
(2.20)

όπου f το διάνυσμα που περιλαμβάνει τους συντελεστές f_{ij} . Για n ζεύγη σημείων, προκύπτουν γραμμικές εξισώσεις της μορφής:

$$\boldsymbol{A}\boldsymbol{f} = \begin{bmatrix} x_1'x_1 & x_1'y_1 & x_1' & y_1'x_1 & y_1'y_1 & y_1' & x_1 & y_1 & 1\\ \vdots & \vdots \\ x_n'x_n & x_n'y_n & x_n' & y_n'x_n & y_n'y_n & y_n' & x_n & y_n & 1 \end{bmatrix} \boldsymbol{f} = 0$$
(2.21)

Για την ύπαρξη μοναδικής λύσης του **F**, ο πίνακας **A** πρέπει να έχει τάξη 8, οπότε απαιτούνται κατ' ελάχιστον 8 ζεύγη ομόλογων σημείων για την επίλυση του γραμμικού συστήματος. Η ύπαρξη λανθασμένων ομολογιών, μεγαλώνουν την τάξη του πίνακα **A** σε 9. Επίλυση του προβλήματος αυτού γίνεται με ανάλυση των ιδιαζουσών τιμών (SVD) του πίνακα **A**:

$$A = UDV^T, (2.22)$$

από την οποία η τελευταία στήλη του πίνακα V αντιστοιχεί στην ελαχιστοτετραγωνική λύση του προβλήματος. Η λύση του διανύσματος **f** με αυτό τον τρόπο ελαχιστοποιεί το ||Af|| υπό τη δέσμευση ||f|| = 1 (Hartley and Zisserman 2003). Η δέσμευση μοναδιαίας νόρμας έχει το πλεονέκτημα ότι επιβάλλει τη μη μηδενικότητα της λύσης, δηλαδή "αίρει το πρόβλημα της απειρίας λύσεων του ομογενούς συστήματος (Καλησπεράκης 2010). Ο πίνακας F που προκύπτει από την ανάλυση των ιδιαζουσών τιμών έχει τάξη 8 και δεν ικανοποιεί τη δέσμευση της μηδενικής ορίζουσας, δηλαδή τη δέσμευση μη αντιστρεψιμότητας (μη ιδιόμορφος ή ομαλός πίνακας). Ο πιο βολικός τρόπος επιβολής της δέσμευσης αυτής είναι με την αντικατάσταση του πίνακα **F** με έναν πίνακα **F**' ο οποίος ελαχιστοποιεί τη νόρμα Frobenius ||F - F'|| με μηδενική ορίζουσα, $\det(F') = 0$. Η επιβολή γίνεται άμεσα από την ανάλυση ιδιαζουσών τιμών του πίνακα $F = UDV^T$, όπου ο διαγώνιος πίνακας D = diag(r, s, t) με $r \ge$ $s \ge t$. Η επιβολή της δέσμευσης γίνεται με αντικατάσταση της μικρότερης ιδιοτιμής του πίνακα **D** με μηδέν, δηλαδή t = 0. Τότε ο πίνακας $F' = U diag(r, s, 0)V^T$ ελαχιστοποιεί τη νόρμα Frobenius ||F - F'|| (Hartley and Zisserman 2003). Η μεθοδολογία αυτή αποτελεί την απλή περίπτωση του αλγορίθμου των 8 σημείων. Το αποτέλεσμα της μεθοδολογίας είναι πολύ ευαίσθητο στη παρουσία θορύβου στις μετρήσεις. Ο Hartley (1995) πρότεινε μια παραλλαγή του αλγορίθμου με κανονικοποίηση των μετρήσεων, πράγμα που βελτιώνει σημαντικά την ακρίβεια του υπολογισμού. Η κανονικοποίηση είναι μια απλή μετάθεση, έτσι ώστε το κέντρο βάρους των μετρήσεων να βρίσκεται στην αρχή του συστήματος συντεταγμένων της εικόνας, και μια αλλαγή κλίμακας έτσι ώστε η μέση απόσταση (RMS) να είναι ίση με $\sqrt{2}$. Τα βήματα της μεθοδολογίας είναι αρχικά κανονικοποίηση των μετρήσεων σε $\hat{x}_L = T x_L$ και $\hat{x}_R = T' x_R$, όπου T και T' οι μετασχηματισμοί κανονικοποίησης των μετρήσεων, εύρεση του κανονικοποιημένου επιπολικού πίνακα \widehat{F} όπως περιεγράφηκε προηγουμένως, και τέλος αποκανονικοποίηση του \hat{F} έτσι ώστε να αντιστοιχεί στις αρχικές μετρήσεις, F= T'^{T} FT (Hartley and Zisserman 2003). Ο αλγόριθμος αυτός είναι γνωστός ως κανονκοποιημένος αλγόριθμος των 8 σημείων (Hartley 1995).

Ο επιπολικός πίνακας **F** μπορεί να θεωρηθεί ως μια γενίκευση του δεσμευμένου επιπολικού πίνακα **E**. Δύο κάμερες με κανονικοποιημένους πίνακες καμερών $\mathbf{K} = [I|O]$ και $\mathbf{K}' = [I|O]$ σχετίζονται μεταξύ τους με τον δεσμευμένο επιπολικό πίνακα **E** της μορφής:

$$\boldsymbol{E} = [\boldsymbol{t}]_{\boldsymbol{x}} \boldsymbol{R} = \boldsymbol{R} [\boldsymbol{R}^T \boldsymbol{t}]_{\boldsymbol{x}}, \qquad (2.23)$$

όπου [t]_x ο αντισυμμετρικός πίνακας:

$$[\mathbf{t}]_{x} = \begin{bmatrix} 0 & -t_{3} & t_{2} \\ -t_{3} & 0 & -t_{1} \\ -t_{2} & t_{1} & 0 \end{bmatrix}$$
(2.24)

Όπως φαίνεται από την εξίσωση (2.23), ο δεσμευμένος επιπολικός πίνακας E εξαρτάται αποκλειστικά από τον πίνακα στροφής R και το διάνυσμα μετάθεσης t. Έχει 5 βαθμούς ελευθερίας, 3 για τον πίνακα στροφής και 2 για το διάνυσμα μετάθεσης εξαιτίας του προβλήματος της αβεβαιότητας της κλίμακας. Η εκτίμηση του δεσμευμένου επιπολικού πίνακα E μπορεί να γίνει με τον αλγόριθμο των 8 σημείων. Ο μειωμένος αριθμός βαθμού ελευθερίας σε σχέση με τον επιπολικό πίνακα F μεταφράζεται σε επιπλέον δεσμεύσεις. Επιπρόσθετα της δέσμευσης της μηδενικής ορίζουσας det(E) = 0, οι δύο μη μηδενικές ιδιοτιμές του πίνακα D κατά την ανάλυση ιδιαζουσών τιμών SVD θα πρέπει να είναι ίσες και συνήθως ορίζονται ίσες με τη μονάδα:

$$\boldsymbol{E} = \boldsymbol{U} diag(1,1,0) \boldsymbol{V}^{T}, \qquad (2.25)$$

Η ανάλυση ιδιαζουσών τιμών του πίνακα E για την ανάκτηση του πίνακα στροφής και της μετάθεσης δίνει τέσσερις πιθανές λύσεις. Για έναν δεσμευμένο επιπολικό πίνακα E της εξίσωσης (2.25), και με πρώτο κανονικοποιημένο πίνακα κάμερας K = [I|0], οι τέσσερις πιθανές λύσεις της δεύτερης κάμερας K' είναι οι εξής:

1)
$$[UWV^{T}|+u_{3}]$$

2) $[UWV^{T}|-u_{3}]$
3) $[UW^{T}V^{T}|+u_{3}]$
4) $[UW^{T}V^{T}|-u_{3}],$
(2.26)

όπου W:

$$\boldsymbol{W} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$
(2.27)

και u₃ η τρίτη στήλη του πίνακα U.

Από τις τέσσερις λύσεις του (2.26), μόνο ένα ζευγάρι πίνακα στροφών **R** και μετάθεσης **t** προβάλλει τα σημεία μπροστά και από τις δύο κάμερες, οπότε είναι απαραίτητη η δοκιμή τουλάχιστον ενός σημείου για την διαπίστωση του σωστού ζεύγους (Hartley and Zisserman 2003). Οι τέσσερις λύσεις παρουσιάζονται στην Εικόνα 2.7. Ο D. Nistér πρότεινε μια πιο αποτελεσματική λύση για τον προσδιορισμό του δεσμευμένου επιπολικού πίνακα με τη χρήση μόνο πέντε ζευγών σημείων, που είναι ο ελάχιστος αριθμός σημείων καθώς η επίλυση έχει πέντε βαθμούς ελευθερίας. Ο αλγόριθμος είναι γνωστός ως αλγόριθμος των πέντε σημείων (five-point algorithm) (Nistér 2004).



Εικόνα 2.7: Απεικόνιση του τρόπου με τον οποίο μπορούν να ερμηνευτούν οι τέσσερις διαφορετικές λύσεις (Hartley and Zisserman 2003).

3D-to-3D: τα σημεία f_{k-1} και f_k ορίζονται σε τρεις διαστάσεις. Σε αυτή την περίπτωση,
 ο μετασχηματισμός T_{k,k-1} των δύο θέσεων εκτιμάται με ελαχιστοποίηση της ευκλεί δειας απόστασης μεταξύ των αντίστοιχων τρισδιάστατων σημείων:

$$\arg\min_{T_{k,k-1}} \sum_{i} \|\tilde{X}_{k}^{i} - |T_{k,k-1}\tilde{X}_{k-1}^{i}\|, \qquad (2.28)$$

όπου το i δηλώνει την i-οστή ομολογία, \tilde{X}_k το τρισδιάστατο σημείο σε ομογενείς συντεταγμένες του τρέχοντος καρέ και \tilde{X}_{k-1} το αντίστοιχο ομόλογο σημείο του προηγούμενου. Ο ελάχιστος αριθμός ομόλογων σημείων για τον υπολογισμό του μετασχηματισμού είναι 3. Η χρήση περισσότερων σημείων προσφέρει καλύτερη ακρίβεια, αλλά απαιτεί περισσότερο υπολογιστικό χρόνο (Yousif, Bab-Hadiashar, and Hoseinnezhad 2015). Στην περίπτωση περισσότερων από τις ελάχιστες απαιτούμενες ομολογίες, σύμφωνα με τους Arun, Huang, and Blostein (1987) το διάνυσμα μετάθεσης $t_{k,k-1}$ του μετασχηματισμού μπορεί να υπολογιστεί από τη διαφορά των κέντρων βάρους των δύο τρισδιάστατων σημειοσυνόλων, και οι στροφές με την ανάλυση ιδιαζουσών τιμών SVD:

$$t_{k,k-1} = \bar{X}_k - R\bar{X}_{k-1}$$

$$R_{k,k-1} = VU^T$$
(2.29)

όπου - η μέση τιμή, $\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{T} = svd \left((X_{k-1} - \bar{X}_{k-1})(X_{k} - \bar{X}_{k})^{T} \right).$

Οι υπολογιζόμενοι μετασχηματισμοί έχουν ορθή κλίμακα και έτσι η τροχιά μπορεί να υπολογιστεί με απευθείας ενσωμάτωση των μετασχηματισμών (Fraundorfer and Scaramuzza 2011).

• **3D-to-2D**: σε αυτή τη περίπτωση τα χαρακτηριστικά σημεία f_{k-1} ορίζονται σε τρεις διστάσεις και τα f_k είναι η δισδιάστατη προβολή τους στην εικόνα I_k . Στη περίπτωση συστημάτων μιας κάμερας χρειάζονται ομολογίες μεταξύ τριών εικόνων. Τα σημεία f_{k-1} προκύπτουν από τριγωνισμό των χαρακτηριστικών των εικόνων I_{k-1} και I_{k-2} (2D-to-2D ομολογίες) και τα f_k είναι η προβολή τους στην εικόνα I_k . Ο μετασχηματισμός $T_{k,k-1}$ σε αυτή τη περίπτωση μπορεί να υπολογιστεί ελαχιστοποιώντας το σφάλμα επαναπροβολής (reprojection error):

$$\arg\min_{T_{k,k-1}} \sum_{i} \|p_{k}^{i} - \hat{p}_{k-1}^{i}\|^{2}, \qquad (2.30)$$

όπου p_k^i είναι το παρατηρούμενο δισδιάστατο σημείο στην τρέχουσα εικόνα I_k και \hat{p}_{k-1}^i είναι η προβολή του αντίστοιχου τρισδιάστατου σημείου X_{k-1}^i στην τρέχουσα εικόνα I_k μετά την εφαρμογή του μετασχηματισμού $T_{k,k-1}$.

Το πρόβλημα εκτίμησης της θέσης μιας βαθμονομημένης μηχανής, δεδομένου ενός συνόλου *n* από τρισδιάστατα σημεία και τις δισδιάστατες ομολογίες τους (πρόβλημα οπισθοτομίας), ονομάζεται και perspective-*n*-point (PnP). Η θέση της κάμερας έχει 6 βαθμούς ελευθερίας (6 DoF), 3 για τις στροφές και 3 για τη μετάθεση. Ο ελάχιστος αριθμός ομόλογων σημείων για την επίλυση του προβλήματος αυτού είναι 3, και στη βιβλιογραφία της όρασης υπολογιστών αναφέρεται ως perspective-

2. Θεωρητικό Υπόβαθρο

3-point (P3P). Είναι μια μη επαναληπτική μεθοδολογία που χρησιμοποιεί τον νόμο των συνημιτόνων, η οποία δίνει μέχρι και 4 λύσεις στο πρόβλημα, που στη συνέχεια μπορούν να αποσαφηνιστούν με χρήση ενός τέταρτου σημείου (Fraundorfer and Scaramuzza 2011). Ανάλυση της μεθοδολογίας και του μαθηματικού μοντέλου παρουσιάζουν οι Gao et al. (2003). Υπάρχουν πολλές παραλλαγές του προβλήματος για βαθμονομημένες και μη κάμερες, για 3 έως *n* σημεία. Οι πιο ευρέως χρησιμοποιούμενοι αλγόριθμοι είναι ο αλγόριθμος *P3P* (Gao et al. 2003), ο οποίος επιλύει το πρόβλημα στη βασική του μορφή, ο αλγόριθμος *EPnP* των Lepetit, Moreno-Noguer, and Fua (2009), που βασίζεται στην ιδέα ότι κάθε ένα από τα *n* σημεία μπορεί να εκφραστεί ως ένα σταθμισμένο άθροισμα τεσσάρων εικονικών σημείων ελέγχου, και ο αλγόριθμος *UPnP* των Penate-Sanchez, Andrade-Cetto, and Moreno-Noguer (2013), ο οποίος είναι μια επέκταση του *EPnP* για την περiπτωση των μη βαθμονομημένων καμερών και επιτρέπει την εκτίμηση της θέσης της κάμερας και της εστιακής απόστασης *f*.

Οι τεχνικές αντιστοίχισης χαρακτηριστικών σημείων πολλές φορές δίνουν εσφαλμένες ομολογίες (outliers) μεταξύ των εικόνων, οι οποίες οδηγούν σε εσφαλμένες εκτιμήσεις, ή και αδυναμία εκτίμησης, εάν συμπεριληφθούν στα προαναφερθέντα μοντέλα. Αιτίες ύπαρξης των λανθασμένων ομολογιών είναι η ύπαρξη θορύβου στην εικόνα, αποκρύψεις, θολότητα στην εικόνα λόγω γρήγορης κίνησης της κάμερας, μεγάλες αλλαγές στην φωτεινότητα ή την οπτική γωνία και αλλαγές στην κλίμακα. Η απόρριψη των λανθασμένων ομολογιών είναι ένα πολύ λεπτό βήμα, και ο χρόνος υπολογισμού του συνδέεται στενά με τον ελάχιστο αριθμό σημείων που είναι αναγκαία για την εκτίμηση της κίνησης (Fraundorfer and Scaramuzza 2011). Η συνηθέστερη μέθοδος για την απόρριψη των λανθασμένων ομολογιών είναι η Random Sample Consensus (RANSAC) (Fischler and Bolles 1987).

2.1.5 RANSAC

Ο αλγόριθμος RANSAC παρουσιάστηκε για πρώτη φορά από τους Fischler and Bolles (1987), ως μια επαναληπτική μέθοδος εύρωστης εκτίμησης των παραμέτρων ενός μαθηματικού μοντέλου από ένα σύνολο δεδομένων που περιλαμβάνουν χονδροειδώς λανθασμένα δεδομένα (outliers). Είναι ένας μη ντετερμινιστικός αλγόριθμος υπό την έννοια ότι παράγει ένα λογικό αποτέλεσμα με μια ορισμένη πιθανότητα p, η οποία αυξάνεται με την αύξηση των επαναλήψεων (Yousif, Bab-Hadiashar, and Hoseinnezhad 2015). Η βασική ιδέα πίσω από τον αλγόριθμο είναι ο υπολογισμός υποθέσεων μοντέλων από τυχαία δειγματοληπτικά σύνολα δεδομένων και στη συνέχεια η επαλήθευση
των υποθέσεων στα υπόλοιπα δεδομένα. Η υπόθεση που συγκεντρώνει την υψηλότερη συναίνεση επιλέγεται ως λύση (Fraundorfer and Scaramuzza 2012). Ο αλγόριθμος RANSAC αποτελείται από δύο επαναληπτικά βήματα (Zuliani 2009):

- Υπόθεση: επιλογή των δειγμάτων, με μεθόδους τυχαίας δειγματοληψίας από τα δεδομένα εισόδου, για τον υπολογισμό των παραμέτρων του μοντέλου. Ο αριθμός των δειγμάτων είναι ο μικρότερος δυνατός για τον ορισμό τους.
- Έλεγχος υπόθεσης: έλεγχος συμβατότητας των στοιχείων ολόκληρου του συνόλου δεδομένων με τις παραμέτρους που εκτιμήθηκαν από το προηγούμενο βήμα.

Ο αλγόριθμος τερματίζεται όταν βρεθεί το σύνολο με τον μεγαλύτερο αριθμό συναινούντων δεδομένων, και το μοντέλο υπολογίζεται ξανά μόνο με αυτά. Η λύση τείνει να σταθεροποιείται όταν ο αριθμός των επαναλήψεων αυξάνεται. Ο αριθμός *N* των επαναλήψεων που είναι απαραίτητες για την (υπό δεδομένη πιθανότητα) εξασφάλιση μιας σωστής λύσης υπολογίζεται ως:

$$N = \frac{\log(1-p)}{\log(1-(1-\epsilon)^{s})'}$$
(2.31)

όπου s ο ελάχιστος αριθμός των δεδομένων που χρειάζεται για τον υπολογισμό του μοντέλου, ϵ το εκτιμώμενο ποσοστό των χονδροειδώς λανθασμένων δεδομένων και p είναι η ζητούμενη πιθανότητα επιτυχίας (π.χ. p = 0.99) (Fischler and Bolles 1987).



Εικόνα 2.8: Παράδειγμα προσαρμογής ευθείας σε δεδομένα που περιέχουν έντονο θόρυβο (χονδροειδώς λανθασμένα δεδομένα).

Στην Εικόνα 2.8 παρουσιάζεται το απλό πρόβλημα προσαρμογής εξίσωσης ευθείας σε δεδομένα που περιέχουν θόρυβο. Στο παράδειγμα αυτό έγιναν προσαρμογές του γραμμικού μοντέλου που υπολογίστηκαν από τυχαίες δειγματοληψίες στα ελάχιστα δεδομένα (δύο σημεία). Το μοντέλο που προσαρμόζεται καλύτερα στα δεδομένα, δηλαδή περιλαμβάνει τον μεγαλύτερο αριθμό αποδεκτών σημείων (inliers) επιλέγεται ως λύση. Στην πράξη δεν υπάρχει απόλυτη εγγύηση ότι το τελικό μοντέλο θα είναι σωστό, και η πιθανότητα επιτυχίας του αλγορίθμου εξαρτάται κυρίως από το ποσοστό του θορύβου στα δεδομένα, καθώς και από την συνθετότητα του μοντέλου. Το ποσοστό χονδροειδών σφαλμάτων που μπορεί να χειριστεί ο αλγόριθμος RANSAC είναι έως περίπου 50% (Zuliani 2009).

Για την εκτίμηση της κίνησης της κάμερας μεταξύ δύο θέσεων λήψης, μοντέλο υπολογισμού είναι ο σχετικός προσανατολισμός (**R**, t), και τα δεδομένα είναι οι αντιστοιχίες των χαρακτηριστικών σημείων. Για τον αλγόριθμο των 8 σημείων που περιγράφηκε προηγουμένως για τον υπολογισμό του επιπολικού πίνακα (γεωμετρικό μοντέλο) τα ελάχιστα σημεία δειγματοληψίας είναι 8. Ο εντοπισμός των ζευγαριών χαρακτηριστικών που συναινούν με το μοντέλο μπορεί να βρεθεί με κριτήριο την απόσταση σημείου από την ομόλογη επιπολική γραμμή (Fraundorfer and Scaramuzza 2011). Η απόσταση *Sampson*, όπως ονομάζεται, είναι μια προσέγγιση πρώτης τάξης του γεωμετρικού σφάλματος (Hartley and Zisserman 2003):

$$\sum_{i} \frac{(x_i'^T F x_i)^2}{(F x_i)_1^2 + (F x_i)_2^2 + (F^T x_i')_1^2 + (F^T x_i')_2^2'}$$
(2.32)

όπου $(Fx_i)_j^2$ το τετράγωνο της *j*-οστής καταχώρισης του πίνακα Fx_i .

Όπως φαίνεται από την εξίσωση (2.31), ο αριθμός των επαναλήψεων του αλγορίθμου επηρεάζεται σημαντικά από τον αριθμό των παραμέτρων του μοντέλου, οπότε προτιμότεροι είναι αλγόριθμοι με τη μικρότερη δυνατή ανάγκη αρχικοποίησης. Παραδείγματος χάριν, όπως παρουσιάζεται και στο Διάγραμμα 2.1, για ένα ποσοστό επιτυχίας p =99% και ποσοστό outliers $\epsilon = 50\%$, ο αλγόριθμος των 8 σημείων θα χρειαστεί περίπου 1200 επαναλήψεις μέχρι να βρεθεί το σωστό μοντέλο, ενώ ο αλγόριθμος των 5 σημείων περίπου 150.



Διάγραμμα 2.1: Διάγραμμα του αριθμού των επαναλήψεων RANSAC σε διαφορετικούς αλγορίθμους για πιθανότητα επιτυχίας 0.99. Οι υπολογισμοί του διαγράμματος έγιναν σύμφωνα με την εξίσωση (2.31).

Αυτό δεν συμβαίνει όμως πάντα. Υποτίμηση του ποσοστού των χονδροειδών σφαλμάτων (outliers) στα δεδομένα μπορεί να οδηγήσει σε ανάγκη περισσότερων επαναλήψεων (Fraundorfer and Scaramuzza 2012). Για την αύξηση της ταχύτητας υπολογισμού έχουν προταθεί πολλές παραλλαγές του RANSAC. Η ανάλυση ορισμένων μεθοδολογιών καθώς και η σύγκρισή τους παρουσιάζεται από τους Raguram, Frahm, and Pollefeys (2008).

2.1.6 Βελτιστοποιήση

Όπως έχει αναφερθεί, η εύρεση της τρέχουσας θέσης της κάμερας C_k επιτυγχάνεται με ενσωμάτωση όλων των μετασχηματισμών $T_{k,k-1}$. Κάθε ένας από αυτούς τους μετασχηματισμούς έχει μια αβεβαιότητα υπολογισμού, συνεπώς η αβεβαιότητα της θέσης της κάμερας εξαρτάται από τις αβεβαιότητες των προγενέστερων μετασχηματισμών. Η αβεβαιότητα της θέσης της κάμερας αυξάνεται συνεχώς με τη ενσωμάτωση των μετασχηματισμών με αποτέλεσμα την δημιουργία του φαινομένου της ολίσθησης της τροχιάς της κάμερας. Είναι επομένως σημαντικό να παραμείνουν μικρές οι αβεβαιότητες των μεμονωμένων μετασχηματισμών (Fraundorfer and Scaramuzza 2012). Οι δύο κύριες μεθοδολογίες βελτιστοποίησης είναι η συνόρθωση δέσμης (bundle adjustment) και η συνόρθωση γράφου (pose-graph optimization).

2.1.6.1 Σύνορωση Δέσμης

Σκοπός της μεθόδου συνόρθωσης δέσμης (Bundle Adjustment) είναι η ταυτόχρονη βελτιστοποίηση των θέσεων τις κάμερας και του τρισδιάστατου μοντέλου. Χρησιμοποιείται σε περιπτώσεις όπου οι ομολογίες των χαρακτηριστικών σημείων ανιχνεύονται σε περισσότερα από δύο καρέ (Fraundorfer and Scaramuzza 2012). Μαθηματικά εκφράζεται ως η ελαχιστοποίηση του σφάλματος επαναπροβολής (reprojection error) των συντεταγμένων των παρατηρούμενων χαρακτηριστικών σημείων και των προβλεπόμενων της μεθόδου (Triggs et al. 1999).

Αν θεωρήσουμε X^i το *i*-οστό τρισδιάστατο σημείο του μοντέλου, p_k^i το αντίστοιχο δισδιάστατο χαρακτηριστικό στην εικόνα k, και $g(X^i, C_k)$ την επαναπροβολή του σύμφωνα με τη θέση της κάμερας C_k , η συνάρτηση κόστους της συνόρθωσης δέσμης είναι:

$$\underset{x^{i}, C_{k}}{\arg\min} \sum_{i, k} \left\| p_{k}^{i} - g(X^{i}, C_{k}) \right\|^{2},$$
(2.33)

Η συνόρθωση δέσμης είναι μια επαναληπτική μη γραμμική βελτιστοποίηση ελαχίστων τετραγώνων. Ο αλγόριθμος Levenberg–Marquardt (Levenberg 1944; Marquardt 1963) έχει αποδειχθεί ότι είναι ένας από τους πιο επιτυχημένους για αυτή τη διαδικασία, λόγω της ευκολίας εφαρμογής του και της χρήσης μιας αποτελεσματικής στρατηγικής απόσβεσης (damping factor), ο οποίος οδηγεί σε γρηγορότερη σύγκλιση με κάθε επανάληψη. Όντας μη γραμμικός αλγόριθμος, απαιτεί καλή εκτίμηση κοντά στο τοπικό ελάχιστο, και υπολογιστικά μπορεί να γίνει εξαιρετικά μεγάλο πρόβλημα ελαχιστοποίησης, λόγω του αριθμού των παραμέτρων που εμπλέκονται (Hartley and Zisserman 2003). Έχουν δημιουργηθεί αρκετές παραλλαγές του αλγορίθμου οι οποίοι εκμεταλλεύονται τη δομή του προβλήματος, αποφεύγοντας την αποθήκευση και τους υπολογισμούς σε μηδενικά στοιχεία, λειτουργούν δηλαδή μόνο όπου υπάρχουν συνδέσεις μεταξύ σημείων και κάμερας (Konolige and Garage 2010; Lourakis and Argyros 2009).

Η συνόρθωση δέσμης συνήθως ονομάζεται σύνολη (Global Bundle Adjustment – GBA) όταν λαμβάνονται υπόψη όλα τα καρέ, και τοπική (Local Bundle Adjustment - LBA) στην οποία γίνεται βελτιστοποίηση σε σταθερό αριθμό *n* εικόνων. Η GBA προφανώς οδηγεί σε πιο ακριβή βελτιστοποίηση σε σύγκριση με τη LBA, δεδομένου ότι λαμβάνει υπόψη όλα τα δεδομένα. Ωστόσο η βελτιστοποίηση σε μια υποπεριοχή περιορίζει τον αριθμό των παραμέτρων της επίλυσης, επομένως η LBA είναι πιο κατάλληλη για εφαρμογές που λειτουργούν σε πραγματικό χρόνο (Yousif, Bab-Hadiashar, and Hoseinnezhad 2015).

2.1.6.2 Σύνορωση Γραφού

Η συνόρθωση γράφου (Pose-Graph Optimization) είναι μια μεθοδολογία βελτιστοποίησης ενός προβλήματος το οποίο αναπαρίσταται με μορφή γράφου. Ο γράφος αποτελείται από δύο τύπων στοιχεία, κόμβους και ακμές. Στα προβλήματα Visual Odometry και Visual SLAM, οι κόμβοι αντιπροσωπεύουν τις θέσεις του συστήματος αισθητήρων και τα σημεία του χώρου, ενώ οι ακμές αντιπροσωπεύουν τους περιορισμούς μεταξύ των κόμβων, δηλαδή τις μετρήσεις και το γεωμετρικό μοντέλο. Με δεδομένο τον γράφο, η βελτιστοποίηση στοχεύει στην βέλτιστη εκτίμηση-θέση των τιμών των κόμβων ώστε να ελαχιστοποιούν τα σφάλματα των περιορισμών. Αποτελεί ένα μη γραμμικό πρόβλημα βελτιστοποίησης ελαχίστων τετραγώνων, όπου μπορούν να χρησιμοποιηθούν για την επίλυση του αλγόριθμοι όπως οι Gauss-Newton ή Levenberg–Marquardt. Το πρόβλημα μπορεί να λυθεί με την ελαχιστοποίηση συνάρτησης της μορφής:

$$F(x) = \sum_{\langle i,j \rangle \in \mathcal{C}} \underbrace{e(x_i, x_j, z_{ij})^T \Omega_{ij} e(x_i, x_j, z_{ij})}_{F_{ij}}$$

$$x^* = \operatorname{argmin} F(x),$$
(2.34)

όπου $x = (x_1^T, ..., x_n^T)^T$ ένα διάνυσμα παραμέτρων. Το κάθε x_i αντιπροσωπεύει μια παράμετρο, z_{ij} και Ω_{ij} αντιπροσωπεύουν τη μέτρηση και τον πίνακα βαρών, αντίστοιχα, ενός περιορισμού που συσχετίζει τις παραμέτρους x_i και x_j , ενώ $e(x_i, x_j, z_{ij})$ είναι η συνάρτηση κόστους που εκτιμά πόσο καλά οι παράμετροι x_i και x_j ικανοποιούν τον περιορισμό z_{ij} (Kümmerle et al. 2011).



Εικόνα 2.9: Παράδειγμα αναπαράστασης ενός προβλήματος μορφής γράφου.

Η συνόρθωση γράφου θεωρείται τεχνική αιχμής στον χώρο της ρομποτικής και της όρασης υπολογιστών, και μπορεί να επιλύσει απλά προβλήματα, όπως μια συνόρθωση δέσμης, μέχρι και σύνθετα προβλήματα όπως συστήματα με συνδυασμό διαφορετικών αισθητήρων (GPS, IMU, laser scanner κ.λπ.).

2.2 ΤΑΥΤΟΧΡΟΝΗ ΟΠΤΙΚΗ ΕΥΡΕΣΗ ΘΕΣΗΣ ΚΑΙ ΧΑΡΤΟΓΡΑΦΗΣΗ

Η ταυτόχρονη οπτική εύρεση θέσης και χαρτογράφηση (Visual Simultaneous Localization and Mapping - VSLAM) είναι μια μεθοδολογία που χρησιμοποιείται για την χαρτογράφηση ενός άγνωστου περιβάλλοντος, με τη χρήση ενός συστήματος αισθητήρων και ταυτόχρονη εύρεση της θέσης του μέσα σε αυτό, χωρίς καμία προϋπάρχουσα πληροφορία. Από τεχνικής απόψεως, τα VSLAM και VO είναι παρόμοιες τεχνικές καθώς και οι δύο υπολογίζουν τις θέσεις των αισθητήρων. Μεθοδολογίες VO χρησιμοποιούνται σε εφαρμογές όπου ενδιαφέρουν μόνο η τοπική συνοχή της τροχιάς της κάμερας και το τρισδιάστατο μοντέλο χρησιμοποιείται απλώς για την ακριβέστερη εκτίμησή της, ενώ στις μεθοδολογίες VSLAM ζητούμενο είναι και η συνολική συνοχή του τρισδιάστατου μοντέλου. Η διαφορά μεταξύ των δύο μεθοδολογιών είναι ότι ένας αλγόριθμος VSLAM χρησιμοποιεί επιπλέον τεχνικές για τη σύνολη βελτιστοποίηση του τρισδιάστατου μοντέλου. Εξ αιτίας του προβλήματος της ολίσθησης, η αναπαράσταση του χάρτη ποτἑ δεν θα ευθυγραμμίζεται απολύτως σε περιοχές τις οποίες έχει επισκεφθεί προηγουμένως το σύστημα. Για τη εξασφάλιση ότι ο χάρτης είναι συνολικά συνεκτικός, ένας αλγόριθμος VSLAM θα πρέπει να μπορεί να αναζητεί και να βρίσκει συνδέσεις μεταξύ περιοχών. Στη βιβλιογραφία η αναζήτηση τέτοιων περιοχών αντιμετωπίζεται με τεχνικές κλεισίματος βρόχων (loop closure). Οι βρόχοι προστίθενται ως περιορισμοί στις τεχνικές βελτιστοποίησης που παρουσιάστηκαν στο κεφάλαιο 2.1.6. Τότε, το σωρευτικό σφάλμα κατά τη διάρκεια κίνησης της κάμερας μπορεί να εκτιμηθεί και να χρησιμοποιηθεί για να διορθωθεί η τροχιά της μηχανής και το γενικό τρισδιάστατο μοντέλο (Taketomi, Uchiyama, and Ikeda 2017).

Στο πεδίο του VSLAM έχουν επικρατήσει δύο μεθοδολογίες: μέθοδοι βασισμένες σε φίλτρα (filter-based), οι οποίες συσσωματώνουν διαδοχικές μετρήσεις από όλες τις εικόνες χρησιμοποιώντας κατανομές πιθανοτήτων σε σχέση με τα χαρακτηριστικά σημεία και τις θέσεις της κάμερας, και μέθοδοι βασισμένες σε βασικά καρέ (keyframe-based) οι οποίες εκτιμούν το τρισδιάστατο μοντέλο χρησιμοποιώντας μόνο επιλεγμένα βασικά καρέ (keyframes) επιτρέποντας την εκτέλεση υπολογιστικά πιο απαιτητικών αλλά ακριβέστερων βελτιστοποιήσεων (συνόρθωση δέσμης, συνόρθωση γράφου).

2.2.1 ΜΕΘΟΔΟΙ ΒΑΣΙΣΜΕΝΕΣ ΣΕ ΦΙΛΤΡΑ

Οι πρώτοι αλγόριθμοι μονοεικονικού SLAM λειτουργούσαν με μεθόδους φιλτραρίσματος. Σε αυτή την προσέγγιση το κάθε καρέ το επεξεργάζεται ένα φίλτρο για τον υπολογισμό του μοντέλου του περιβάλλοντος και της τροχιάς της κάμερας. Οι μεθοδολογίες

αυτές παρουσιάζουν μειονεκτήματα όπως η σπατάλη υπολογισμών για κάθε νέο καρέ, πράγμα που γενικά παρέχει ελάχιστη νέα πληροφορία, καθώς και τα σφάλματα υπολογισμού λόγω της γραμμικοποίησης των συναρτήσεων. Ένα μεγάλο πλεονέκτημα ωστόσο των μεθόδων με χρήση φίλτρου είναι η δυνατότητα συγχώνευσης δεδομένων (data fusion) από διαφορετικούς αισθητήρες προκειμένου να δοθεί μια πιο ισχυρή περιγραφή του περιβάλλοντος ή σημείων ενδιαφέροντος, με αποτέλεσμα τη μείωση της αβεβαιότητας των πληροφοριών των αισθητήρων σε σύγκριση με την περίπτωση όπου αυτοί μοντελοποιούνται ξεχωριστά.

Από τις πρώτες αλλά και η πιο συνηθισμένη τεχνική για τις μεθόδους βάσει φίλτρων είναι η Extended Kalman Filter (EKF), η οποία είναι βασισμένη στο φίλτρο Bayes για το φιλτράρισμα και την πρόβλεψη μη γραμμικών εξισώσεων, των οποίων η γραμμικές προσεγγίσεις λαμβάνονται από σειρές Taylor πρώτης τάξης. Βασική υπόθεση του EKF-SLAM είναι πως ο θόρυβος τόσο των αισθητήρων όσο και της οδομετρίας μπορούν να περιγραφούν με μια κατανομή Gauss. Το EKF είναι χωρισμένο σε δύο στάδια: (1) το στάδιο της πρόβλεψης (prediction), στο οποίο η μελλοντική θέση της κάμερας εκτιμάται με βάση τη τρέχουσα θέση της και τις πληροφορίες που εφαρμόζονται για την αλλαγή θέσης της από τον χρόνο k στον k+1, και (2) το στάδιο της ενημέρωσης (update), στο οποίο ενημερώνεται η τρέχουσα θέση της κάμερας στο σύστημα. Το EKF-SLAM είναι απλό στην εφαρμογή, αλλά ο σημαντικός περιορισμός του βρίσκεται στη γραμμικοποίηση των μη γραμμικών εξισώσεων μέσω σειρών Taylor, με αποτέλεσμα μειωμένη ακρίβεια σε μερικές περιπτώσεις σε σχέση με άλλες μεθόδους (Yousif, Bab-Hadiashar, and Hoseinnezhad 2015).

Τα φίλτρα σωματιδίων (Particle Filters) είναι ένας διαφορετικός τρόπος προσέγγισης του προβλήματος του SLAM. Η πρώτη εφαρμογή παρουσιάστηκε το 2002 από τους Montemerlo et al. (2002). Βασικό πλεονέκτημα του εν λόγω φίλτρου είναι η δυνατότητα αντιμετώπισης περιπτώσεων στις οποίες το μοντέλο κίνησης του ρομπότ δεν περιγράφεται από κατανομές Gauss ή οι μη γραμμικές εξισώσεις του προβλήματος δεν μπορούν να προσεγγιστούν επαρκώς με ανάπτυγμα Taylor. Ουσιαστικά, η εκ των υστέρων πιθανότητα για την κατάσταση του ρομπότ και του χάρτη, αντί να περιγράφεται από μια συνεχή κατανομή, αναπαρίσταται από ένα σύνολο από δείγματα (particles). Κάθε δείγμα διατηρεί την υπόθεση μιας πιθανής θέσης του ρομπότ και του προσδίδεται βάρος, το οποίο προκύπτει από την ενσωμάτωση της μέτρησης στο μοντέλο, αποτελώντας μια εκτίμηση της ποιότητας της υπόθεσης του εκάστοτε δείγματος. Σε δεύτερο βήμα πραγματοποιείται διαδικασία επαναδειγματοληψίας βάσει των συντελεστών βαρύτητας. Το προκύπτον δείγμα κατανέμεται σύμφωνα με μια προσέγγιση προς την επιθυμητή θέση, της οποίας η προσέγγιση είναι σωστή όσο ο αριθμός της δειγματοληψίας τείνει το άπειρο (Sim et al. 2005; Thrun, Burgard, and Fox 2005).

2.2.2 ΜΕΘΟΔΟΙ ΒΑΣΙΣΜΕΝΕΣ ΣΕ ΒΑΣΙΚΑ ΚΑΡΕ

Όπως αναφέρθηκε προηγουμένως, οι πρώτοι αλγόριθμοι SLAM μονής κάμερας λειτουργούσαν με μεθόδους βασισμένες σε φίλτρα. Το κάθε καρέ το επεξεργάζεται ένα φίλτρο για τον υπολογισμό του μοντέλου του περιβάλλοντος και της τροχιάς της κάμερας. Έχει ήδη σημειωθεί ότι οι μεθοδολογίες αυτές επιβαρύνονται υπολογιστικά με κάθε νέο καρέ αν και αυτό γενικά συνεισφέρει ελάχιστη νέα πληροφορία, ενώ υπόκεινται και σε σφάλματα λόγω της γραμμικοποίησης. Οι μέθοδοι που στηρίζονται σε βασικά καρέ (keyframe-based) εκτιμούν τον χάρτη χρησιμοποιώντας μόνο επιλεγμένα καρέ, γεγονός που με την σειρά του επιτρέπει τη χρήση ακριβέστερων μεθόδων συνόρθωσης (βλ. ενότητα 2.1.6) για τη βελτιστοποίηση του χάρτη και της τροχιάς της κάμερας. Οι Strasdat, Montiel, and Davison (2012) συνέκριναν τις δύο αυτές μεθοδολογίες και, μέσω πειραμάτων, κατέληξαν στο συμπέρασμα ότι οι δεύτερες (keyframe-based) είναι πιο αποδοτικές από τις πρώτες (filter-based) όσον αφορά την ακρίβεια ανά υπολογιστικόν (feature-based) και οι άμεσες (direct-based) που παρουσιάστηκαν στις 2.1.1 και 2.1.2, αντίστοιχα, μπορούν να κατηγοριοποιηθούν και ως keyframe-based.

Τα βασικά καρέ (keyframes) είναι ένα υποσύνολο των καρέ. Η ιδέα των μεθόδων που τα αξιοποιούν παρουσιάστηκε για πρώτη φορά το 2007 από τους Klein and Murray (2007). Είναι ο πρώτος αλγόριθμος ο οποίος χώρισε τις διαδικασίες της ανίχνευσης (tracking) και της χαρτογράφησης (mapping) σε δύο διαφορετικά νήματα του επεξεργαστή (CPU threads). Αυτά τα δύο νήματα εκτελούνται παράλληλα, έτσι ώστε το υπολογιστικό κόστος της χαρτογράφησης να μην επηρεάζει την ανίχνευση. Αποτέλεσμα αυτού είναι η δυνατότητα να χρησιμοποιείται η συνόρθωση δέσμης (Bundle Adjustment), η οποία έχει μεγάλο υπολογιστικό κόστος, για τη βελτιστοποίηση του μοντέλου. Στη μεθοδολογία τους, τα βασικά καρέ προστίθενται όταν η ποιότητα ανίχνευσης θεωρείται ότι είναι καλή· όταν έχουν μεσολαβήσει πάνω από 20 καρέ από την τελευταία εισαγωγή ενός βασικού καρέ· παράλληλα, η κάμερα θα πρέπει να απέχει μια ελάχιστη απόσταση από το κοντινότερο βασικό καρέ (Klein and Murray 2007). Οι Engel, Schöps, and Cremers (2014) προσθέτουν ένα βασικό καρέ όταν το τρέχον καρέ υπερβαίνει το

κατώφλι ενός σταθμισμένου συνδυασμού σχετικής απόστασης και γωνίας από το προηγούμενο βασικό καρέ. Οι προσθέτουν ένα βασικό καρέ όταν η ευκλείδεια απόσταση του νέου καρέ σε σχέση με όλα τα βασικά καρέ υπερβαίνει το 12% του μέσου βάθους της σκηνής. Οι Mur-Artal, Montiel, and Tardos (2015) εισάγουν ένα βασικό καρέ όταν έχουν περάσει τουλάχιστον 20 καρέ από την τελευταία εισαγωγή βασικού καρέ όταν το τρέχον καρέ έχει ανιχνεύσει τουλάχιστον 50 σημεία και όταν το τρέχον καρέ ανιχνεύει λιγότερο από το 90% των σημείων ενός βασικού καρέ, επιλεγμένου βάσει ενός συντελεστή ορατότητας. Οι Engel, Koltun, and Cremers (2018) εισάγουν ένα βασικό καρέ όταν το πεδίο θέασης διαφέρει σημαντικά από το τελευταίο βασικό καρέ· όταν η μετάθεση της κάμερας δημιουργεί ή διορθώνει προβλήματα αποκρύψεων και όταν ο χρόνος έκθεσης της κάμερας αλλάξει σημαντικά. Οι Henry et al. (2010) χρησιμοποίησαν τεχνική RANSAC για να βρουν το μετασχηματισμό μεταξύ της τρέχουσας εικόνας και του τελευταίου keyframe. Η τρέχουσα εικόνα ορίζεται ως keyframe εφόσον οι λανθασμένες ομολογίες (outliers) είναι πάνω από ένα ορισμένο όριο. Σε μεταγενέστερο άρθρο τους (Henry et al. 2012) παρουσίασαν μια πιο αποδοτική μεθοδολογία, η οποία ορίζει ένα καινούριο keyframe όταν η σωρευτική στροφή ή μετάθεση των υπολογιζόμενων μετασχηματισμών υπερβαίνει ένα όριο. Αυτό επιτρέπει την προσαρμογή της πυκνότητας των βασικών καρέ στην κίνηση της κάμερας, χωρίς να απαιτείται η πρόσθετη χρήση RANSAC για κάθε νέο καρέ.

Όπως είναι προφανές, δεν υπάρχει μια συγκεκριμένη μεθοδολογία για την εισαγωγή ενός βασικού καρέ, και η επιλογή εξαρτάται καθαρά από τον τύπο του αισθητήρα του συστήματος (μονοεικονμικό, διεικονικό, RGB-D), τον τύπο της μεθοδολογίας (direct-based, feature-based) και την αρχιτεκτονική του αλγορίθμου (π.χ. αρχικά πυκνή εισαγωγή βασικών καρέ, με μετέπειτα αραίωση για τη μείωση του όγκου των δεδομένων).

2.2.3 ΚΛΕΙΣΙΜΟ ΒΡΟΧΩΝ

Πρωταρχικός σκοπός των τεχνικών κλεισίματος βρόχων (loop-closure) είναι η μείωση της ολίσθησης που συσσωρεύεται στην τροχιά του συστήματος με τον χρόνο. Η οδομετρία που βασίζεται σε αισθητήρες κίνησης καθώς και σε οπτικές πληροφορίες είναι επιρρεπής σε σφάλματα, παρασύροντας την εκτιμώμενη τροχιά από την αληθή της θέση. Επομένως, είναι απαραίτητη η χρήση τεχνικών αναζήτησης στην τροχιά του συστήματος μη γειτονικών εικόνων οι οποίες παρατηρούν την ίδια σκηνή, ώστε με τη προσθήκη ενός περιορισμού στα μοντέλα να περιορίζεται σημαντικά η συσσωρευμένη

ολίσθηση στην εκτίμηση της τροχιάς της κάμερας, και κατά συνέπεια τα σφάλματα στο τρισδιάστατο μοντέλο.

Στη βασική της μορφή, γίνεται προσπάθεια προσαρμογής της τρέχουσας εικόνας με όλες τις προηγούμενες, με τεχνικές αντιστοίχισης χαρακτηριστικών σημείων. Αυτή η προσέγγιση είναι ιδιαίτερα δαπανηρή σε υπολογιστικό κόστος λόγω του ότι ο αριθμός των εικόνων αυξάνεται με την πάροδο του χρόνου, οπότε η αντιστοίχιση της τρέχουσας εικόνας με όλες τις προηγούμενες δεν είναι κατάλληλη για εφαρμογές πραγματικού χρόνου.

Οι Henry et al. (2010) προσπαθούν για κάθε νέο βασικό καρέ να ανιχνεύσουν κλεισίματα βρόχου με όλα τα προηγούμενα βασικά καρέ. Λόγω του υψηλού υπολογιστικού κόστους της χρήσης RANSAC για όλα τα βασικά καρέ, λαμβάνουν υπόψη μόνο εκείνα που βρίσκονται σχετικά μακριά από την τρέχουσα θέση της κάμερας. Για περαιτέρω φιλτράρισμα των βασικών καρέ χρησιμοποιούν προσέγγιση αναγνώρισης θέσεων (place recognition), βασισμένη στο δέντρο λεξιλογίου (vocabulary tree) που περιγράφεται από τους Nistér and Stewenius (2006). Η τεχνική αυτή επιτρέπει το κάθε βασικό καρέ να εκπροσωπείται από ένα οπτικό λεξικό (bag-of-visual-words), βάσει των περιγραφέων του κάθε βασικού καρέ, και επιτρέπει την ταχύτατη εύρεση υποψήφιων βασικών καρέ για πιθανή εύρεση βρόχου. Τα βασικά καρέ που εν τέλει περνούν το φιλτράρισμα υπόκεινται σε RANSAC με το τρέχον βασικό καρέ και ,εφόσον υπάρχουν επαρκείς και γεωμετρικά σωστές ομολογίες, θεωρείται ως κλείσιμο βρόχου, οπότε και μπορεί να διορθωθεί το συσσωρευμένο σφάλμα στην εκτίμηση της τροχιάς της κάμερας μέχρι το τρέχον καρέ.

Σε αντίθεση με το εργαλείο bag-of-visual-words, το οποίο κατασκευάζει ένα ιστόγραμμα για κάθε εικόνα βασιζόμενο στη συχνότητα εμφάνισης των περιγραφικών λέξεων σε αυτή, η μεθοδολογία VLAD (Vector of Locally Augmented Descriptors) (Jégou et al. 2010) κατασκευάζει έναν περιγραφέα συγκεντρώνοντας τα υπόλοιπα (residuals) μεταξύ κάθε περιγραφέα εικόνας και της αντίστοιχης οπτικής λέξης, που έχει ως αποτέλεσμα μια πιο ισχυρή οπτική προσέγγιση στην αναγνώριση ομοιότητας εικόνων. Εναλλακτικά, αντί για επιλογή βασικών καρέ ή αντιστοίχιση με όλα τα προηγούμενα πλαίσια, οι Endres et al. (2012) παρουσίασαν μια μέθοδο κλεισίματος βρόχου που προσαρμόζει το τρέχον καρέ με τα 20 τελευταία, με αποτέλεσμα μια υπολογιστικά πιο αποτελεσματική προσέγγιση (Yousif, Bab-Hadiashar, and Hoseinnezhad 2015).

Οι (Cummins and Newman 2008) ανέπτυξαν μια μέθοδο για την ανίχνευση κλεισίματος βρόχου με βάση την αναγνώριση της οπτικής εμφάνισης περιοχών από όπου έχουν ήδη περάσει. Η αντιστοίχιση γίνεται με τον εντοπισμό παρουσίας ή απουσίας χαρακτηριστικών από ένα λεξικό οπτικών λέξεων. Το σύστημα λαμβάνει υπόψη τις πιθανότητες των χαρακτηριστικών που εμφανίζονται μαζί και είναι σε θέση να υπολογίσει την πιθανότητα δύο εικόνων να εμφανίζουν την ίδια περιοχή. Αυτή η μέθοδος δεν εξαρτάται από τη δημιουργία ενός μετρικού χάρτη αφού συγκρίνει μόνο τις εικόνες απευθείας (Williams et al. 2009). Βασισμένοι στην εργασία των Cummins and Newman (2008), οι Glover et al. (2012) έχουν δημιουργήσει έναν αλγόριθμο ανοιχτού κώδικα.

Οι Clemente et al. (2007) παρουσίασαν μια μεθοδολογία ανίχνευσης βρόχων βασισμένη στην αναζήτηση αντιστοιχιών μεταξύ των κοινών χαρακτηριστικών σε διαφορετικούς υπο-χάρτες. Το σύστημά τους χρησιμοποιεί τόσο την ομοιότητα της οπτικής εμφάνισης όσο και τις σχετικές αποστάσεις μεταξύ των χαρακτηριστικών για να βρεθεί το μεγαλύτερο συμβατό σύνολο κοινών χαρακτηριστικών μεταξύ δύο υπο-χαρτών. Αφού βρεθεί σταθερό σύνολο, μπορεί να προσδιοριστεί εύκολα η σχετική κλίμακα, η στροφή και η μετάθεση που απαιτούνται για την ευθυγράμμιση των δύο υπο-χαρτών (Williams et al. 2009).

Οι Lim, Lim, and Kim (2014), τέλος, παρουσίασαν μια μεθοδολογία βασισμένη σε αντιστοιχίες των τρισδιάστατων σημείων του χάρτη και των δισδιάστατων της εικόνας. Χρησιμοποίησαν δέντρο λεξιλογίου για να βρουν αντιστοιχίες μεταξύ των χαρακτηριστικών του υπό αναζήτηση καρέ με χαρακτηριστικά γνωστών τρισδιάστατων συντεταγμένων του χάρτη. Όταν συνταυτίσουν χαρακτηριστικά περισσότερα από ένα κατώφλι, τότε γίνεται επαλήθευση με την τεχνική RANSAC (με μοντέλο 3D-to-2D) για την απόρριψη των λανθασμένων ομολογιών και την εύρεση του μετασχηματισμού του βρόχου.

Είναι σημαντικό να σημειωθεί ότι ενώ σε ένα σύστημα VSLAM τα κλεισίματα βρόχων είναι πολύ χρήσιμα, η εισαγωγή εσφαλμένων βρόχων μπορεί να έχει καταστροφικές συνέπειες στην διόρθωση της τροχιάς και του χάρτη.

2.2.4 ΕΠΑΝΕΝΤΟΠΙΣΜΟΣ

Ένα σημαντικό πρόβλημα των αλγορίθμων Visual SLAM και Visual Odometry είναι η αποτυχία στον υπολογισμό της θέσης του συστήματος λόγω ταχείας κίνησης της κάμερας ή ύπαρξης παρεμβολών, όπως εμπόδια μπροστά στο φακό. Αποτυχία υπολογισμού θέσης σημαίνει αδυναμία σύνδεσης της τρέχουσας θέσης του συστήματος με

το περιβάλλον που έχει δημιουργηθεί προηγουμένως. Σε αυτή την περίπτωση είναι απαραίτητη η παύση των διαδικασιών εντοπισμού και χαρτογράφησης του συστήματος, και η αρχή ενεργειών για την αποκατάσταση του προβλήματος.

Οι τεχνικές επανεντοπισμού είναι παρόμοιες με εκείνες της εύρεσης κλειστών βρόχων. Η πιο απλή αλλά υπολογιστικά απαιτητική διαδικασία είναι η προσπάθεια αντιστοίχισης του τρέχοντος καρέ με όλα τα προηγούμενα βασικά καρέ. Για την επίλυση του προβλήματος, οι Yousif, Bab-Hadiashar, and Hoseinnezhad (2015) χρησιμοποιούν τυχαία αναζήτηση για την αντιστοίχιση ταξινομημένων χαρακτηριστικών μεταξύ του τρέχοντος καρέ και των βασικών καρέ του συστήματος. Οι Mur-Artal, Montiel, and Tardos (2015) χρησιμοποιούν τεχνική Bag-of-visual-Words για την συνταύτιση του τρέχοντος καρέ με μια βάση από προηγούμενα βασικά καρέ για την ανάκτηση της θέσης της κάμερας. Οι Kasyanov et al. (2017)παρουσίασαν έναν τρόπο ανάκτησης της θέσης συνδυάζοντας ένα καινούργιο γράφημα θέσεων, που κατασκευάστηκε μετά την αποτυχία του αλγορίθμου, με το γράφημα που είχε κατασκευαστεί προηγουμένως. Ο συνδυασμός γίνεται με εύρεση αντιστοιχιών των βασικών καρέ των δύο γραφημάτων.

3 ΠΑΡΟΥΣΙΑΣΗ ΑΛΓΟΡΙΘΜΩΝ

Σε αυτό το κεφάλαιο παρουσιάζονται δύο εφαρμογές οπτικής οδομετρίας που δημιουργήθηκαν για τη παρούσα διπλωματική εργασία. Η γλώσσα προγραμματισμού που επιλέχθηκε για την ανάπτυξή τους είναι η C++, λόγω της υψηλής υπολογιστικής ταχύτητας που παρέχει και της υποστήριξης από πληθώρα βιβλιοθηκών για την υλοποίηση διαδικασιών που παρουσιάστηκαν στο προηγούμενο κεφάλαιο. Για την ανίχνευση χαρακτηριστικών σημείων στις εικόνες, την αντιστοίχισή τους, καθώς και την υλοποίηση των απαραίτητων μαθηματικών μοντέλων για την επίλυση της μεθοδολογίας, επιλέχθηκε η βιβλιοθήκη OpenCV (G Bradski 2000). Για την βελτιστοποίηση του μοντέλου επιλέχθηκε η βιβλιοθήκη g2ο (Kümmerle et al. 2011).

Η OpenCV είναι η κορυφαία βιβλιοθήκη ανοιχτού κώδικα η οποία δημιουργήθηκε με σκοπό την παροχή υποδομής για εφαρμογές όρασης υπολογιστών, επεξεργασίας εικόνας και μηχανικής εκμάθησης. Η g2o είναι μια βιβλιοθήκη ανοιχτού κώδικα, γραμμένη σε γλώσσα C++, η οποία έχει δημιουργηθεί για την βελτιστοποίηση μη γραμμικών προβλημάτων ρομποτικής και όρασης υπολογιστών, τα οποία μπορούν να περιγραφούν με μορφή γράφου. Ο γενικός στόχος της είναι η εύρεση της κατάλληλη διαμόρφωσης των παραμέτρων του γράφου, οι οποίοι εξηγούν ένα σύνολο μετρήσεων επηρεασμένων από θόρυβο Gauss.

Και οι δύο εφαρμογές έχουν δημιουργηθεί για συστήματα μονής κάμερας. Η ανίχνευση της θέσης γίνεται με αντιστοιχίες 3D-to-2D (βλέπε ενότητα 2.1.4), οπότε είναι απαραίτητη μια αρχικοποίηση του συστήματος για τη δημιουργία των πρώτων τρισδιάστατων σημείων. Η διαφορά των εφαρμογών βρίσκεται στον τρόπο ανίχνευσης και αντιστοίχισης των χαρακτηριστικών σημείων. Η πρώτη εφαρμογή κάνει χρήση του αλγορίθμου Shi-Tomasi για την εύρεση των χαρακτηριστικών σημείων από τις εικόνες, και στη συνέχεια η ανίχνευσή τους γίνεται με τον αλγόριθμο Kanade-Lucas-Tomasi. Η μεθοδολογία αυτή δεν υπολογίζει περιγραφείς για το κάθε χαρακτηριστικό σημείο, αλλά λειτουργεί καθαρά με τη φωτεινότητα των γειτονικών τους εικονοψηφίδων για την ανίχνευσή τους. Η δεύτερη εφαρμογή ανιχνεύει χαρακτηριστικά σημεία FAST στις εικόνες και τα αντιστοιχίζει βάσει ενός περιγραφέα ORB που έχει υπολογίσει για το κάθε ένα από αυτά. Η χρήση περιγραφέων δίνει τη δυνατότητα αντιστοίχισης τρισδιάστατων σημείων σημείων σε πολλαπλά βασικά καρέ. Η μεθοδολογία των δύο εφαρμογών παρουσιάζεται σε τέσσερα μέρη: εκείνα της αρχικοποίησης, της ανίχνευσης, της χαρτογράφησης και της βελτιστοποίησης.

- Αρχικοποίηση: σε αυτό το στάδιο δημιουργείται το πρώτο σύνολο τρισδιάστατων σημείων, τα οποία θα χρησιμοποιηθούν ως βάση για τη αρχή της διαδικασίας της ανίχνευσης.
- 2. Ανίχνευση: εδώ υπολογίζεται η τρέχουσα θέση της κάμερας και αποφασίζεται αν το τρέχον καρέ θα χαρακτηριστεί ως βασικό, το οποίο θα χρησιμοποιηθεί για τον τριγωνισμό νέων τρισδιάστατων σημείων.
- **3. Χαρτογράφηση:** όταν ένα καρέ χαρακτηριστεί ως βασικό, στην φάση της χαρτογράφησης ανακατασκευάζονται τα νέα τρισδιάστατα σημεία του μοντέλου.
- 4. Βελτιστοποίηση: με την ολοκλήρωση της διαδικασίας χαρτογράφησης, το η διαδικασία της βελτιστοποίησης πραγματοποιεί συνόρθωση δέσμης στα n τελευταία βασικά καρέ και στα τρισδιάστατα σημεία που αυτά παρατηρούν.

Η διαδικασία της αρχικοποίησης του συστήματος πραγματοποιείται μόνο μια φορά, με την εκκίνηση της εφαρμογής, ενώ οι υπόλοιπες τρεις επαναλαμβάνονται με τη σειρά που αναφέρθηκαν μέχρι τον τερματισμό του αλγορίθμου.

Οι εφαρμογές χρησιμοποιούν τον αλγόριθμο των 5 σημείων (βλέπε ενότητα 2.1.4) για τον υπολογισμό του δεσμευμένου επιπολικού πίνακα, οπότε είναι απαραίτητη η προβαθμονόμηση της κάμερας.

3.1 Αρχικοποίηση

Σκοπός της αρχικοποίησης, είναι η δημιουργία ενός αρχικού συνόλου τρισδιάστατων σημείων στο οποίο βασίζεται στη συνέχεια η διαδικασία της ανίχνευσης θέσης της κάμερας.

Για την αρχικοποίηση απαιτούνται δισδιάστατες ομολογίες ενός ζεύγους εικόνων. Από της ομολογίες υπολογίζεται ο δεσμευμένος επιπολικός πίνακας *E*, με τον αλγόριθμο των 5 σημείων με χρήση RANSAC, για την εύρεση και απομόνωση των ομολογιών που προσδίδουν μεγάλα σφάλματα στο μοντέλο. Ο γνωστός πλέον δεσμευμένος επιπολικός πίνακας υπόκειται σε ανάλυση ιδιαζουσών τιμών για την εξαγωγή του πίνακα στροφής *R* και του διανύσματος μετάθεσης *t*. Από τον δεσμευμένο επιπολικό πίνακα, όπως φαίνεται στην εξίσωση (2.26), λαμβάνονται δύο πιθανοί πίνακες στροφής και δύο διανύσματα μετάθεσης μεταξύ των δύο καρέ. Ο προσδιορισμός του σωστού ζεύγους γίνεται στο ίδιο βήμα, με επιβολή του περιορισμού χειρότητας (cheirality constraint)

3. Παρουσίαση Αλγορίθμων

(Nistér, Naroditsky, and Bergen 2004). Με τον περιορισμό αυτό, τα σημεία της σκηνής πρέπει να τριγωνίζονται μπροστά από τις δύο κάμερες. Το ζευγάρι *R, t* που επιλέγεται είναι αυτό που μετά τον τριγωνισμό των δισδιάστατων ομολογιών θα εμφανίζει τα περισσότερα σημεία μπροστά και από τις δύο κάμερες¹. Τα τρισδιάστατα σημεία στη συνέχεια φιλτράρονται με ένα κατώφλι ελάχιστης παραλλακτικής γωνίας και μέγιστης απόστασης από την αρχή του συστήματος αναφοράς. Στο τελευταίο βήμα της αρχικοποίησης γίνεται βελτιστοποίηση των θέσεων των καμερών και των τρισδιάστατων σημείων με συνόρθωση δέσμης (αναλυτική επεξήγηση της μεθόδου παρουσιάζεται στην ενότητα 3.4). Τα δύο καρέ που χρησιμοποιήθηκαν για τον τριγωνισμό των τρισδιάστατων σημείων θεωρούνται πλέον τα πρώτα δύο βασικά καρέ του συστήματος.

Στις περιπτώσεις που ο αριθμός των σημείων που βρίσκονται μπροστά και από τις δύο κάμερες, μετά από τον έλεγχο χειρότητας, είναι μικρότερος από ένα ποσοστό του συνόλου των ζευγών των ομολογιών που χρησιμοποιήθηκαν για την εύρεση του δεσμευμένου επιπολικού πίνακα, ή ο αριθμός των σημείων μετά το φιλτράρισμα παραλλακτικής γωνίας και απόστασης (αλλά και οποιαδήποτε στιγμή κατά τη διαδικασία της αρχικοποίησης) πέσει κάτω από ένα ορισμένο κατώφλι, τότε το δεύτερο καρέ απορρίπτεται, τα χαρακτηριστικά σημεία ανιχνεύονται στο επόμενο καρέ και η διαδικασία επαναλαμβάνεται. Το ποσοστό την σημείων μετά τον έλεγχο χειρότητας χρησιμοποιείται για τη θεώρηση μιας δυνατής επιπολικής γεωμετρίας. Τα απομακρυσμένα σημεία και εκείνα που τριγωνίστηκαν με μικρή παραλλακτική γωνία φιλτράρονται λόγω της μεγάλης αβεβαιότητας θέσης μετά τον τριγωνισμό.

Για την εύρεση των ομολογιών η πρώτη εφαρμογή εξάγει χαρακτηριστικά σημεία Shi-Tomasi από το πρώτο καρέ και τα ανιχνεύει στο επόμενο με τον αλγόριθμο Kanade-Lucas-Tomasi.

Η δεύτερη εφαρμογή εξάγει, από το κάθε καρέ, χαρακτηριστικά σημεία FAST με περιγραφείς ORB και τα αντιστοιχεί βάσει των περιγραφέων τους με τον αλγόριθμο GMS (Bian et al. 2017). Τα χαρακτηριστικά σημεία των δύο βασικών καρέ τα οποία δεν θα αντιστοιχηθούν αποθηκεύονται για πιθανό μελλοντικό τριγωνισμό τους.

¹ Το πρώτο καρέ έχει μοναδιαίο πίνακα στροφής και μηδενικό διάνυσμα μετάθεσης. Λόγω της φύσης των συστημάτων μονής κάμερας, η κλίμακα του μοντέλου είναι άγνωστη, οπότε ο ορισμός της γίνεται με το μέτρο του διανύσματος μετάθεσης **t**, το οποίο από την ανάλυση των ιδιαζουσών τιμών ισούται πάντοτε με την μονάδα.

3.2 ΑΝΙΧΝΕΥΣΗ

Η ανίχνευση της θέσης της κάμερας γίνεται με αντιστοιχίες 3D-to-2D (προβολές των τρισδιάστατων σημείων στο τρέχον καρέ), με τον αλγόριθμο EPnP (Lepetit, Moreno-Noguer, and Fua 2009) με χρήση RANSAC, όπως αυτό παρουσιάστηκε στην ενότητα 2.1.4. Μετά την εύρεση της θέσης της κάμερας, για να θεωρηθεί το τρέχον καρέ ως βασικό, θα πρέπει να έχουν ανιχνευθεί τουλάχιστον 50 χαρακτηριστικά σημεία, και το ποσοστό των τρισδιάστατων σημείων που αντιστοιχήθηκαν επιτυχώς μετά την επίλυση με RANSAC να είναι μικρότερο από ένα προκαθορισμένο ποσοστό. Το ποσοστό αυτό εξασφαλίζει την ύπαρξη κίνησης μεταξύ του τελευταίου βασικού καρέ και του τρέχοντος. Αν, για παράδειγμα, δεν ανιχνευθεί το 30% των τρισδιάστατων σημείων του τελευταίου βασικού καρέ, τότε θεωρείται ότι η κάμερα βρίσκεται σε κίνηση. Το νέο βασικό καρέ χρησιμοποιείται για τον τριγωνισμό των νέων χαρακτηριστικών σημείων. Σε περίπτωση που ένας από τους δύο ελέγχους αποτύχει τότε το τρέχον καρέ απορρίπτεται, τα χαρακτηριστικά σημεία ανιχνεύονται στο επόμενο καρέ και η διαδικασία επαναλαμβάνεται.

Στην πρώτη εφαρμογή, πριν την αρχή της διαδικασίας ανίχνευσης, εξάγονται νέα χαρακτηριστικά σημεία Shi-Tomasi από το δεύτερο βασικό καρέ, τα οποία θα χρησιμοποιηθούν στη συνέχεια στη διαδικασία της χαρτογράφησης για τη δημιουργία νέων τρισδιάστατων σημείων. Τα χαρακτηριστικά αυτά περιορίζονται σε μέρη του δεύτερου βασικού καρέ στα οποία δεν έχουν ανιχνευθεί προηγουμένως άλλα χαρακτηριστικά σημεία. Αυτό επιτυγχάνεται με τη δημιουργία μιας δυαδικής εικόνας, η οποία περιέχει τετράγωνα διαστάσεων m * m, μηδενικών τιμών, με κέντρο τετραγώνων τις συντεταγμένες του κάθε χαρακτηριστικού σημείου του δεύτερου βασικού καρέ. Η δυαδική εικόνα έχει ίδιες διαστάσεις με τα καρέ, και χρησιμοποιείται ως μάσκα για τον περιορισμό των νέων χαρακτηριστικών σημείων. Λόγος του περιορισμού αυτού είναι η ομοιόμορφη κατανομή των χαρακτηριστικών σημείων στα καρέ.

Η διαδικασία της ανίχνευσης, στην πρώτη μεθοδολογία, ξεκινάει με ένα σύνολο χαρακτηριστικών σημείων του δεύτερου βασικού καρέ, γνωστών τρισδιάστατων συντεταγμένων από την αρχικοποίηση, και ένα σύνολο χαρακτηριστικών σημείων τα οποία θα χρησιμοποιηθούν σε επόμενο βήμα του αλγορίθμου για τη δημιουργία νέων τρισδιάστατων σημείων. Η εύρεση των δισδιάστατων ομολογιών τους στο τρέχον καρέ γίνεται με τον αλγόριθμο Kanade-Lucas-Tomasi. Στην περίπτωση της απόρριψης του τρέχοντος καρέ, η ανίχνευση των συνόλων γίνεται στο επόμενο καρέ και η διαδικασία της ανίχνευσης επαναλαμβάνεται μέχρι να θεωρηθεί το καρέ ως βασικό. Στη δεύτερη εφαρμογή η διαδικασία της ανίχνευσης ξεκινάει με ένα σύνολο χαρακτηριστικών σημείων γνωστών τρισδιάστατων συντεταγμένων του δεύτερου βασικού καρέ και ένα σύνολο χαρακτηριστικών σημείων που έχουν εξαχθεί από το τρέχον καρέ. Η εύρεση των 3D-to-2D ομολογιών γίνεται, όπως αναφέρθηκε στη διαδικασία της αρχικοποίησης, με αντιστοίχιση των περιγραφέων τους. Τα χαρακτηριστικά σημεία του τρέχοντος καρέ που δεν αντιστοιχίζονται αποθηκεύονται και χρησιμοποιούνται στο επόμενο βήμα για τον τριγωνισμό τους. Όπως και στη πρώτη εφαρμογή, σε περίπτωση απόρριψης του τρέχοντος καρέ εξάγονται νέα χαρακτηριστικά σημεία από το επόμενο καρέ και η διαδικασία επαναλαμβάνεται.

3.3 Χαρτογραφήση

Στην πρώτη εφαρμογή, μετά την εισαγωγή του νέου βασικού καρέ το δεύτερο σετ των δισδιάστατων ομολογιών χρησιμοποιείται για τον τριγωνισμό των νέων τρισδιάστατων σημείων. Αρχικά γίνεται υπολογισμός του δεσμευμένου επιπολικού πίνακα E που τα περιγράφει, με χρήση RANSAC για την απομάκρυνση τυχών λανθασμένων ομολογιών. Με γνωστές τις θέσεις των δύο τελευταίων βασικών καρέ και τις δισδιάστατες ομολογίες τους, γίνεται τριγωνισμός των νέων χαρακτηριστικών σημείων. Μετά τον τριγωνισμό τα σημεία φιλτράρονται με βάση μια ελάχιστη παραλλακτική γωνία και μια μέγιστη απόσταση από τις κάμερες. Τέλος πραγματοποιείται συνόρθωση δέσμης στα n τελευταία βασικά καρέ και τα τρισδιάστατα σημεία που αυτά παρατηρούν. Μετά τη διαδικασία της συνόρθωσης δέσμης, ανιχνεύονται καινούργια χαρακτηριστικά σημεία στο τελευταίο βασικό καρέ, όπως περιγράφηκε στο τελευταίο βήμα της αρχικοποίησης, και οι διαδικασίες της ανίχνευσης θέσης και του τριγωνισμού επαναλαμβάνονται.

Στη δεύτερη εφαρμογή τα νέα τρισδιάστατα σημεία τριγωνίζονται από ομολογίες των χαρακτηριστικών σημείων του τελευταίου βασικού καρέ τα οποία δεν είχαν αντιστοιχηθεί στο βήμα της ανίχνευσης με τα μη αντιστοιχισμένα χαρακτηριστικά των l προηγούμενων βασικών καρέ. Αν k το τελευταίο βασικό καρέ, γίνεται αντιστοίχιση χαρακτηριστικών σημείων του με τα χαρακτηριστικά σημεία του k - l βασικού καρέ. Στη συνέχεια γίνεται υπολογισμός του δεσμευμένου επιπολικού πίνακα E με RANSAC, και σε επόμενο βήμα οι σωστές ομολογίες τριγωνίζονται και φιλτράρονται με τη διαδικασία που περιγράφηκε στην πρώτη εφαρμογή. Τα τρισδιάστατα σημεία στη συνέχεια προβάλλονται στα τελευταία l βασικά καρέ, εκτός του k και του k - l, και γίνεται προσπάθεια εύρεσης ομολογιών με τα μη αντιστοιχισμένα τους χαρακτηριστικά. Η διαδικασία επαναλαμβάνεται μέχρι και το k – 1 βασικό καρέ. Όπως και στη πρώτη εφαρμογή, στο τέλος πραγματοποιείται συνόρθωση δέσμης στα n τελευταία βασικά καρέ για τη βελτιστοποίηση των δεδομένων.

3.4 Βελτιστοποίηση

Όπως έχει αναφερθεί στην ενότητα 2.1.6.2, το μοντέλο της συνόρθωσης γράφου αποτελείται από κόμβους και ακμές. Οι θέσεις των βασικών καρέ και τα τρισδιάστατα σημεία του μοντέλου προστίθενται στον γράφο ως κόμβοι, και οι παρατηρήσεις, δηλαδή οι προβολές των τρισδιάστατων σημείων στα καρέ, ως ακμές. Η επίλυση της συνόρθωσης πραγματοποιείται για τα τελευταία n βασικά καρέ, από τα οποία το πρώτο ρυθμίζεται ως σταθερό (fixed in place). Για κάθε βασικό καρέ εισάγεται ένας κόμβος, τύπου VertexSE3Expmap, ο οποίος περιέχει τον πίνακα στροφής **R** και το διάνυσμα μετάθεσης t που είναι προς βελτιστοποίηση. Τα προς βελτιστοποίηση τρισδιάστατα σημεία, τα οποία έχουν ανιχνευθεί σε τουλάχιστον 3 βασικά καρέ, εισάγονται στον γράφο ως κόμβοι, τύπου VertexSBAPointXYZ, όπως επίσης και εκείνα που έχουν τριγωνισθεί από τα τελευταία δύο βασικά καρέ, τα οποία εμφανίζονται μόνο δύο φορές αλλά χρησιμοποιούνται για την ανίχνευση της θέσης του επόμενου καρέ. Για κάθε βασικό καρέ όπου εντοπίζεται το σημείο προστίθεται μια ακμή, τύπου EdgeSE3ProjectXYZ. Οι ακμές συνδέουν τους κόμβους των βασικών καρέ και τους κόμβους των τρισδιάστατων σημείων, και περιέχουν τις συντεταγμένες των εικονοψηφίδων των παρατηρούμενων τρισδιάστατων σημείων από το σχετικό καρέ, όπως επίσης και τις παραμέτρους της εσωτερικής γεωμετρίας της μηχανής. Η συνάρτηση σφάλματος που περιέχουν οι ακμές αυτού του τύπου είναι το σφάλμα επαναπροβολής των τρισδιάστατων σημείων στις εικόνες. Σε κάθε ακμή προστίθεται ένας διαγώνιος μοναδιαίος πίνακας βαρών, διαστάσεων 2x2. Ο πίνακας αυτός αντιπροσωπεύει την αξιοπιστία της μέτρησης. Επομένως, ανάλογα την ακρίβεια της μέτρησης ή την εμπιστοσύνη ως προς αυτή, προστίθεται βάρος το οποίο επηρεάζει αντίστοιχα τη συνόρθωση. Στη συγκεκριμένη εφαρμογή οι μετρήσεις θεωρούνται ισοβαρείς.

Με το τέλος της εφαρμογής, όταν σταματήσει δηλαδή η διαδικασία της ανίχνευσης θέσης και του εντοπισμού, πραγματοποιείται σύνολη συνόρθωση των δεδομένων. Όλες οι συνορθώσεις γίνονται με 10 επαναλήψεις, με χρήση του αλγορίθμου Levenberg-Marquardt (OptimizationAlgorithmLevenberg) και επίλυση βασισμένη σε αραιή παραγοντοποίηση Cholesky (LinearSolverEigen).

43

3. Παρουσίαση Αλγορίθμων



Εικόνα 3.1: Βασικό διάγραμμα ροής μεθοδολογιών.

4 ΑΞΙΟΛΟΓΗΣΗ ΑΛΓΟΡΙΘΜΩΝ

Σε αυτό το κεφάλαιο γίνεται παρουσίαση των αποτελεσμάτων της αξιολόγησης των δύο αλγορίθμων που αναλύθηκαν στο προηγούμενο κεφάλαιο. Για την αξιολόγηση των αλγορίθμων χρησιμοποιήθηκαν σύνολα δεδομένων οδομετρίας του KITTI Vision Benchmark Suite², το οποίο είναι έργο του Τεχνολογικού Ινστιτούτου της Καρλσρούης της Γερμανίας και του Τεχνολογικού Ινστιτούτου Τογοτα του Σικάγο των Ηνωμένων Πολιτειών της Αμερικής. Τα δεδομένα αυτά δημιουργήθηκαν από την κίνηση οχήματος μέσα σε πόλη, εξοπλισμένου με δύο συστήματα καμερών υψηλής ανάλυσης και γνωστής διάταξης, έναν σαρωτή laser και ένα σύγχρονο σύστημα εντοπισμού θέσης. Από τις αλληλουχίες των δεδομένων που παρέχουν επιλέχθηκαν τέσσερις, στα οποία οι πραγματικές θέσεις των καρέ της κάθε μίας, καθώς και τα δεδομένα βαθμονόμησης των καμερών, είναι γνωστά.



Εικόνα 4.1: Δείγμα εικόνων από το σύνολο δεδομένων 05 του ΚΙΤΤΙ.

Εκτός των δεδομένων οδομετρίας ΚΙΤΤΙ, δημιουργήθηκαν πειραματικά δεδομένα τα οποία προήλθαν από την περιμετρική βιντεοσκόπηση ενός τραπεζιού εμπλουτισμένου με αντικείμενα. Η βιντεοσκόπηση έγινε με 60 καρέ το δευτερόλεπτο και ανάλυση εικόνας 1920 x 1088. Για τις ανάγκες της αξιολόγησης έγινε υποδειγματοληψία των καρέ του βίντεο, και εξήχθησαν 4 καρέ ανά δευτερόλεπτο. Οι "αληθείς" θέσεις των καρέ που χρησιμοποιηθήκαν για τη σύγκριση, καθώς και η βαθμονόμηση της κάμερας, προέκυψαν με την επίλυσή τους με το φωτογραμμετρικό λογισμικό Agisoft.

² <u>http://www.cvlibs.net/datasets/kitti/eval_odometry.php</u>



Εικόνα 4.2: Δείγμα εικόνων από το σύνολο των πειραματικών δεδομένων του Τραπεζιού που κατασκευάστηκε.

Η αξιολόγηση των αποτελεσμάτων χωρίζεται σε γωνιακό σφάλμα και τροχιακό σφάλμα, όπως παρουσιάζεται στο σχετικό άρθρο των Geiger, Lenz, and Urtasun (2012). Το αποτέλεσμα κάθε συνόλου δεδομένων ΚΙΠΙ χωρίζεται σε 8 υποπεριοχές των 100 μέτρων (100, 200,..., 800), ενώ το σύνολο δεδομένων του τραπεζιού σε 8 υποπεριοχές του ενός μέτρου (1,2,..., 8) λόγω των μικρών διαστάσεων στου αντικειμένου. Το γωνιακό σφάλμα μετρείται σε μοίρες ανά μέτρο (deg/m) και το τροχιακό σφάλμα σε ποσοστό (%). Η διαδικασία υπολογισμού των σφαλμάτων έχει ως εξής:

- Θέτοντας ως αναφορά τα αληθή δεδομένα, υπολογίζεται η αθροιστική απόσταση του κάθε καρέ από την αρχή της τροχιάς.
- Ξεκινώντας από το πρώτο καρέ n των αληθών και τον υπό αξιολόγηση δεδομένων, βρίσκεται το καρέ mλ με αθροιστική απόσταση πιο κοντά στην πρώτη υποπεριοχή λ, και υπολογίζεται ο σχετικός τους προσανατολισμός.

$$PoseDelta_{GT} = PoseGT_n^{-1} * PoseGT_{m\lambda}$$

$$PoseDelta_{SLAM} = PoseSLAM_n^{-1} * PoseSLAM_{m\lambda}$$

$$(4.1)$$

 Στη συνέχεια υπολογίζεται ο "σχετικός προσανατολισμός" των σχετικών προσανατολισμών (στην ουσία δηλαδή η διαφορά αληθούς και εκτιμηθέντος σχετικού προσανατολισμού):

$$PoseError = PoseDelta_{SLAM}^{-1} * PoseDelta_{GT}$$
(4.2)

Το γωνιακό σφάλμα προκύπτει από τον υπολογισμού της γωνίας στροφής του πίνακα στροφής που προκύπτει έτσι

$$RotationError = a\cos\left(\frac{Tr(R) - 1}{2}\right),$$
(4.3)

όπου

$$Tr(R) = PoseError_{11} + PoseError_{22} + PoseError_{33}$$

$$(4.4)$$

το ίχνος του πίνακα στροφής **R**. Το τροχιακό σφάλμα υπολογίζεται από την L2 νόρμα:

$$TranslationError = \sqrt{PoseError_{14}^{2} + PoseError_{24}^{2} + PoseError_{34}^{2}}$$
(4.5)

Τα δύο σφάλματα αυτά ανάγονται στο μέτρο, διαιρούνται δηλαδή με την απόσταση λ (σε m) της υποπεριοχής.

 Η παραπάνω διαδικασία πραγματοποιείται για όλα τα καρέ των συνόλων δεδομένων για όλες τις υποπεριοχές. Το συνολικό σφάλμα υπολογίζεται από τον μέσο όρο όλων των υπολογισμένων σφαλμάτων.

Sequence	Path Length	Frames	FPS	Resolution	HFOV	VFOV
	(m)			(pixel)	(deg)	(deg)
KITTI - 00	3724.187	4541	10	1226 x 370	81.6	29.31
KITTI - 03	560.888	801	10	1226 x 370	81.6	29.31
KITTI - 04	393.645	271	10	1226 x 370	81.6	29.31
KITTI - 05	2205.576	2761	10	1226 x 370	81.6	29.31
Τραπέζι	8.992	248	4	1920 x 1088	54.79	32.73

Πίνακας 4.1: Περιγραφικά στοιχεία των συνόλων δεδομένων.

Τα αποτελέσματα παρουσιάζονται στα επόμενα με τέσσερα διαγράμματα για το κάθε σύνολο δεδομένων. Στα πρώτα δύο διαγράμματα (δηλαδή στην πρώτη σειρά της εικόνας για κάθε σύνολο), παρουσιάζονται τα γωνιακά και τα τροχιακά σφάλματα της κάθε υποπεριοχής (μέσος όρος σφάλματος ανά υποπεριοχή αξιολόγησης). Το κάτω αριστερά διάγραμμα παρουσιάζει την τροχιά του κάθε συνόλου, όπως αυτή δόθηκε από τα αληθή δεδομένα και όπως υπολογίστηκε από τους αλγορίθμους. Τα κάτω δεξιά διαγράμματα παρουσιάζουν τις σχέσεις μεταξύ των υπολογισμένων και των αληθών συντεταγμένων και στροφών του κάθε βασικού καρέ. Από τους δύο αλγορίθμους που

παρουσιάστηκαν στο προηγούμενο κεφάλαιο, ο μόνος που κατάφερε να ολοκληρώσει όλα τα σύνολα δεδομένων ήταν αυτός που χρησιμοποιεί τη μεθοδολογία Kanade-Lucas-Tomasi για την ανίχνευση των χαρακτηριστικών σημείων (Διάγραμμα 4.1 έως Διάγραμμα 4.5). Λόγω της ιδιαίτερης δομής του περιβάλλοντος όπου κινείται η κάμερα στα δεδομένα KITTI (εμπρόσθια κίνηση και επαναλαμβανόμενα μοτίβα), ο δεύτερος αλγόριθμος, που χρησιμοποιεί περιγραφείς ORB για την αντιστοίχιση των χαρακτηριστικών σημείων, αποτύγχανε πολύ συχνά. Συνεπώς, η μετρητική του αξιολόγηση έγινε μόνο για τα πειραματικά δεδομένα του τραπεζιού, τα οποία κατάφερε να τα ολοκληρώσει επιτυχώς (Διάγραμμα 4.6).



Διάγραμμα 4.1: Αποτελέσματα αξιολόγησης αλληλουχίας ΚΙΠΙ – 00. Μέσο γωνιακό σφάλμα: 0.003953 deg/m. Μέσο τροχιακό σφάλμα: 9.1350 %.

0.008 Rotation Error [deg/] 800.0 for [deg/] 800.0 k 100.0 c + Rotation Error + Translation Error Translation Error [%] Path Length [m] Path Length [m] Ζ Ground Truth 20 [m] 20 [m] SLAM -20 Z [m] Omega Phi ---- Kappa Error [deg] -1 -50 Keyframe X [m]







Διάγραμμα 4.3: Αποτελέσματα αξιολόγησης αλληλουχίας ΚΙΠΙ – 04. Μέσο γωνιακό σφάλμα: 0.003266 deg/m και μέσο τροχιακό σφάλμα: 3.8045 %.



Διάγραμμα 4.4: Αποτελέσματα αξιολόγησης αλληλουχίας ΚΙΠΙ – 05. Μέσο γωνιακό σφάλμα: 0.007334 deg/m, μέσο τροχιακό σφάλμα: 10.2664 %.









Διάγραμμα 4.6: Αποτελέσματα αξιολόγησης αλληλουχίας Τραπέζι – αλγόριθμος ORB. Μέσο γωνιακό σφάλμα: 0.021120 deg/cm και μέσο τροχιακό σφάλμα: 3.3387 %.

Seq.	SLAM	KeyFrames	Elapsed	Global BA	Num. 3D	Avg. FPS
	Path (m)		time (sec)	time (sec)	Points	
00	3316.588	2162	481.987	8.51825	223221	14.1686
03	530.844	329	57.1701	1.48735	47964	21.3115
04	380.013	147	17.4884	0.450869	15309	22.045
05	2073.302	1442	245.602	4.95672	130541	15.4686
T-KLT	8.457	64	15.1534	0.276927	10122	19.2143
T-ORB	8.536	114	59.1368	0.741166	36100	7.34459

Пілакас 4.2:	Χαρακτηριστικά	των ολοκλ	Νρωμένων	συνόλων	δεδομένων
in anas ne	rap and pio mo	1001 0/1010		0010/1001	0000000000

Sequence	Rotation Error	Translation Error (%)
00	0.003953 (deg/m)	9.1350
03	0.004927 (deg/m)	9.0104
04	0.003266 (deg/m)	3.8045
05	0.007334 (deg/m)	10.2664
T-KLT	0.021834 (deg/cm)	2.8555
T-ORB	0.021120 (deg/cm)	3.3387

Πίνακας 4.3: Αποτελέσματα της αξιολόγησης των συνόλων δεδομένων.



Εικόνα 4.3: Καρέ από τα δύο σύνολα δεδομένων ΚΙΤΤΙ 05 (πάνω) και Τραπέζι (κάτω), όπου εμφανίζονται τα χαρακτηριστικά σημεία τα οποία έχουν ανιχνευθεί επιτυχώς από το προηγούμενο καρέ (πράσινα), και τα νέα υποψήφια χαρακτηριστικά προς ανίχνευση στο επόμενο καρέ (κίτρινα).



Εικόνα 4.4: Τελικό νέφος τρισδιάστατων σημείων από το σύνολο δεδομένων ΚΙΤΤΙ 05. Με πράσινο χρώμα παρουσιάζεται η αληθής τροχιά και με κίτρινο η υπολογιζόμενη.





Εικόνα 4.5: Τελικό νέφος τρισδιάστατων σημείων από το σύνολο δεδομένων του Τραπεζιού. Με πράσινο χρώμα παρουσιάζεται η αληθής τροχιά και με κίτρινο η υπολογιζόμενη.

5 Σύμπερασματά και Προτάσεις για Περαιτέρω Ερεύνα

Σκοπός της παρούσας διπλωματικής εργασίας ήταν η κατά το δυνατόν εκτενής βιβλιογραφική έρευνα σχετικά με τις μεθοδολογίες της Οπτικής Οδομετρίας και του Ταυτόχρονου Εντοπισμού Θέσης και Χαρτογράφησης, καθώς και μια πρώτη υλοποίηση ενός βασικού παραμετροποιήσιμου αλλά αυτοματοποιημένου αλγορίθμου, ο οποίος θα μπορεί να λειτουργεί σε ποικίλα περιβάλλοντα διαφορετικής κλίμακας. Σε αυτό το κεφάλαιο συνοψίζονται τα συμπεράσματα αξιολόγησης, τα αποτελέσματα και οι δυσκολίες που προέκυψαν κατά την υλοποίηση των τελικά δύο αλγορίθμων, ενώ επίσης κατατίθενται ορισμένες προτάσεις και ιδέες για περαιτέρω ανάπτυξη και διόρθωση των διαδικασιών.

Κατ' αρχάς, στα αποτελέσματα του πρώτου αλγορίθμου, για τα σύνολα δεδομένων KITTI, παρατηρείται ένα μέσο γωνιακό σφάλμα 0.00487 deg/m και ένα μέσο σφάλμα μετάθεσης 8.0541%, πράγμα που τον κατατάσσει στη χαμηλή δεκάδα μεταξύ των αλγορίθμων οδομετρίας που έχουν αξιολογηθεί από το ΚΙΤΠ. Η σύγκριση βέβαια του αλγορίθμου της διπλωματικής εργασίας δεν είναι απολύτως ισότιμη καθώς χρησιμοποιήθηκαν 4 από τα 21 σύνολα δεδομένων που είναι απαραίτητα για την ολοκληρωμένη αξιολόγηση. Το μέσο γωνιακό σφάλμα μπορεί να θεωρηθεί αρκετά ικανοποιητικό σε αντίθεση με το σφάλμα μετάθεσης. Το υψηλό αυτό σφάλμα, όμως, ήταν αναμενόμενο καθώς δεν έχει υλοποιηθεί τρόπος διόρθωσης της κλίμακας της τροχιάς με μεθόδους κλεισίματος βρόχων (SLAM). Όπως μπορεί να παρατηρηθεί από τα διαγράμματα τροχιάς του κάθε συνόλου δεδομένων, ειδικά στα Διάγραμμα 4.1 Διάγραμμα 4.4, το φαινόμενο της ολίσθησης της τροχιάς και κλίμακας είναι αρκετά ισχυρό, αλλά όπως αναφέρθηκε προηγουμένως, αναμενόμενο. Τεχνικές κλεισίματος βρόχων σε αυτά τα δύο σύνολα δεδομένων (00 και 05) θα είχαν περιορίσει σημαντικά το πρόβλημα του φαινομένου της ολίσθησης της κλίμακας, καθώς υπάρχουν περιοχές στις οποίες το όχημα επισκέπτεται πολλαπλές φορές. Σε περιπτώσεις όπου η κάμερα δεν επισκέπτεται προηγούμενες περιοχές (03 και 04), το φαινόμενο αυτό είναι αδύνατο να διορθωθεί, λόγω της φύσης των μονοεικονικών αλγορίθμων, εκεί ωστόσο, λόγω και της σχετικά απλής μορφής της τροχιάς, τα αποτελέσματα είναι αρκετά καλύτερα.

Από τα αριθμητικά αποτελέσματα της αξιολόγησης του πρώτου αλγορίθμου για το σύνολο δεδομένων του Τραπεζιού, μπορεί κάποιος να βγάλει συμπέρασμα ότι ο αλγόριθμος δεν λειτουργεί ικανοποιητικά σε περιβάλλοντα μικρής κλίμακας, καθώς το γωνιακό σφάλμα φαίνεται να αυξάνεται αισθητά από τις 0.00487 μοίρες/μ στις 0.021834 μοίρες/εκ. Κάτι τέτοιο όμως δεν ισχύει, καθώς από τα γραφήματα σύγκρισης των τροχιών και γωνιών, μπορεί να παρατηρηθεί το ακριβώς αντίθετο (Διάγραμμα 4.1 Διάγραμμα 4.5). Ο λόγος της μεγάλης διαφοράς σε αυτά τα δύο σφάλματα μένει άγνωστος, και αποτελεί αντικείμενο μελέτης για τον μελλοντικό ερευνητή, ο οποίος καλείται να συνυπολογίσει την διαφορά γεωμετρίας των δύο διαδρομών και την διαφορά γεωμετρίας των αντίστοιχων λήψεων, την σχέση του άξονα λήψης με την διεύθυνση της τροχιάς σε κάθε σύνολο δεδομένων, και μέσω αυτών, να κανονικοποιήσει τα σφάλματα των αλγορίθμων καθιστώντας τα συγκρίσιμα.

Ο δεύτερος αλγόριθμος αξιολογήθηκε μόνο με το σύνολο δεδομένων μικρής κλίμακας (Διάγραμμα 4.6) καθώς δεν κατάφερε να ολοκληρώσει κάποιο από τα δεδομένα ΚΙΠΙ. Ο λόγος της συχνής αποτυχίας του αλγορίθμου είναι η ιδιαίτερη δομή του αστικού περιβάλλοντος με επαναλαμβανόμενα μοτίβα, πράγμα που προκαλούσε μεγάλη δυσκολία στη συνταύτιση χαρακτηριστικών σημείων μεταξύ των καρέ. Ένας τρόπος αποφυγής αυτού του προβλήματος θα ήταν με τη χρήση ενός μοντέλου κίνησης, το οποίο θα περιόριζε τις περιοχές εύρεσης ομολογιών, και θα έκανε ακριβέστερη συνταύτιση των χαρακτηριστικών σημείων. Το γωνιακό σφάλμα του αλγορίθμου είναι ελάχιστα καλύτερο, στο ίδιο σύνολο δεδομένων, από τον πρώτο αλγόριθμο, αλλά παρουσιάζει 1.5% μεγαλύτερο τροχιακό σφάλμα.

Η διαφορά της ταχύτητας λειτουργίας των δύο αλγορίθμων είναι αρκετά μεγάλη. Ο πρώτος αλγόριθμος έχει ένα μέσο όρο 18.44 καρέ ανά δευτερόλεπτο, ενώ ο δεύτερος μόλις 7.34. Αυτό εμφανίζεται εξ αιτίας του πολυπλοκότερου τρόπου λειτουργίας, με την επαναπροβολή των χαρακτηριστικών σημείων σε προγενέστερα καρέ, και της πιο απαιτητικής μεθοδολογίας συνταύτισης χαρακτηριστικών. Οι μεθοδολογίες συνταύτισης περιγραφέων για την εύρεση ομολογιών είναι πιο εύρωστοι σε περιπτώσεις ταχείας κiνησης της κάμερας, που έχει ως αποτέλεσμα τη δημιουργία θολών καρέ. Σε αυτή τη περίπτωση, το κακής ποιότητας καρέ θα απορριπτόταν και η συνταύτιση θα γινόταν με το επόμενο, καθώς δεν είναι απαραίτητη η σύνδεση μεταξύ των καρέ όπως στον αλγόριθμο KLT, ο οποίος θα αποτύγχανε. Ο αλγόριθμος KLT από την άλλη μεριά είναι πιο ταχύς στην εύρεση ομολογιών, αλλά πιο επιρρεπής σε σφάλματα καθώς έχει την τάση να συνεχίζει να ανιχνεύει λανθασμένες ομολογίες όταν λειτουργεί σε δύσκολα περιβάλλοντα.

Πειραματισμοί με τις παραμέτρους των διαδικασιών της αρχικοποίησης έδειξαν ότι η ποιότητα της αρχικοποίησης παίζει τον σημαντικότερο ρόλο στη συνολική ακρίβεια του μοντέλου και της τροχιάς, καθώς τα δεδομένα αυτής είναι η βάση για το κύριο τμήμα

του αλγορίθμου. Εξίσου σημαντική είναι και καλή κατανομή των χαρακτηριστικών σημείων στην εικόνα. Σε περιοχές με μειωμένη υφή, οι μεθοδολογίες παρουσιάζουν βέβαια υψηλά σφάλματα λόγω της έλλειψης χαρακτηριστικών σημείων, έως και αποτυχία του αλγορίθμου από αδυναμία ανίχνευσης της κάμερας.

Το σφάλματα της ολίσθησης θα μπορούσαν να μειωθούν με την πρόσθεσης τεχνικών κλεισίματος βρόχων, καθώς και του συνδυασμού δεδομένων εξωτερικών αισθητήρων όπως αδρανειακά συστήματα και GPS. Το πρόβλημα της αποτυχίας της ανίχνευσης της κάμερας, η οποία έχει καταστρεπτική επίπτωση στη λειτουργία του αλγορίθμου, μπορεί να διορθωθεί με τη δημιουργία μεθοδολογιών ανάκτησης θέσης. Μερικές από τις τεχνικές για την επίλυση αυτών των προβλημάτων παρουσιάστηκαν στην ενότητα 2.2.3 και 2.2.4.

Ο μέσος χρόνος υπολογισμού των καρέ του πρώτου αλγορίθμου είναι αρκετά ικανοποιητικός, και ελαφρώς ταχύτερος από τον πραγματικό χρόνο καταγραφής της πληροφορίας. Ωστόσο, υπάρχουν πολλές πτυχές του κώδικα που θα μπορούσαν να βελτιωθούν για να επιτευχθεί υψηλότερος ρυθμός, ειδικά με την πρόσθεση τεχνικών κλεισiματος βρόχων, οι οποίες σε συνδυασμό με τη βελτιστοποίηση είναι από τις υπολογιστικά πιο απαιτητικές διαδικασίες. Η ταχύτητα υπολογισμού θα μπορούσε να αυξηθεί, για παράδειγμα, με τον διαχωρισμό των κύριων διαδικασιών (ανίχνευση, χαρτογράφηση και βελτιστοποίηση) να πραγματοποιούν τους υπολογισμούς τους σε ξεχωριστά νήματα του επεξεργαστή, παράλληλα, χωρίς να απαιτείται η αναμονή ολοκλήρωσης της μιας για την εκκίνηση της επόμενης. Η ανίχνευση περιορισμένου αριθμού χαρακτηριστικών σημείων, με ομοιόμορφη κατανομή σε ολόκληρη την επιφάνεια του καρέ, σε συνδυασμό με τη χρήση ενός μοντέλου κίνησης για τη συνταύτιση των ομολογιών, θα μπορούσε να επιφέρει ταχύτερο χρόνο ανίχνευσης της τροχιάς της κάμερας, καθώς θα μειωνόταν ο χρόνος συνταύτισης των χαρακτηριστικών όσο και ο αριθμός των επαναλήψεων RANSAC λόγω του μικρότερου αριθμού λανθασμένων ομολογιών.

Στους αλγορίθμους που υλοποιήθηκαν, τα δεδομένα που εισέρχονται στο μοντέλο βελτιστοποίησης παίρνουν το ίδιο βάρος. Η χρήση βαρών σε αυτά θα είχε ως αποτέλεσμα γρηγορότερη σύγκλιση του μοντέλου και καλύτερη ακρίβεια υπολογισμών, καθώς τα δεδομένα με μεγαλύτερη αβεβαιότητα θα επηρέαζαν λιγότερο το μοντέλο.

Ένα από τα πιο ενδιαφέροντα πεδία έρευνας, τέλος, είναι η αντιμετώπιση δυναμικών αντικειμένων στο περιβάλλον όπου κινείται η κάμερα. Απρόσμενες αλλαγές στο περιβάλλον, όπως η παρουσία κινούμενων ανθρώπων ή οχημάτων, θα μπορούσαν πιθανώς να υποβαθμίσουν την ποιότητα της εκτίμησης της θέσης, και ακόμη να οδηγήσουν στην αποτυχία του αλγορίθμου. Πρόκειται για ένα από τα σημαντικότερα αλλά και πιο δύσκολα προβλήματα που καλούνται να αντιμετωπίσουν οι αλγόριθμοι VO και SLAM.

6 Βιβλιογραφία

- 1. Agrawal, Motilal, Kurt Konolige, and Morten Rufus Blas. 2008. "Censure: Center Surround Extremas for Realtime Feature Detection and Matching." In Springer, 102–15.
- Arun, K Somani, Thomas S Huang, and Steven D Blostein. 1987. "Least-Squares Fitting of Two 3-D Point Sets." IEEE Transactions on Pattern Analysis & Machine Intelligence (5): 698–700.
- 3. Baker, Simon, and Iain Matthews. 2004. "Lucas-Kanade 20 Years on: A Unifying Framework." International journal of computer vision 56(3): 221–55.
- 4. Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. 2006. "Surf: Speeded up Robust Features." In Springer, 404–17.
- 5. Bian, JiaWang et al. 2017. "Gms: Grid-Based Motion Statistics for Fast, Ultra-Robust Feature Correspondence." In IEEE, 2828–37.
- 6. Bradski, G. 2000. "The OpenCV Library." Dr. Dobb's Journal of Software Tools.
- 7. Bradski, Gary, and Adrian Kaehler. 2008. Learning OpenCV: Computer Vision with the OpenCV Library. "O'Reilly Media, Inc."
- 8. Brown, Duane C. 1966. "Decentering Distortion of Lenses." *Photogrammetric Engineering and Remote Sensing*.
- Calonder, Michael, Vincent Lepetit, Christoph Strecha, and Pascal Fua. 2010. "Brief: Binary Robust Independent Elementary Features." In Springer, 778–92.
- 10. Clemente, Laura A et al. 2007. "Mapping Large Loops with a Single Hand-Held Camera."
- Cummins, Mark, and Paul Newman. 2008. "FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance." The International Journal of Robotics Research 27(6): 647–65.
- 12. Durrant-Whyte, Hugh, and Tim Bailey. 2006. "Simultaneous Localization and Mapping: Part I." *IEEE robotics & automation magazine* 13(2): 99–110.
- 13. Durrant-Whyte, Hugh, David Rye, and Eduardo Nebot. 1996. "Localization of Autonomous Guided Vehicles." In *Robotics Research*, Springer, 613–25.
- 14. Endres, Felix et al. 2012. "An Evaluation of the RGB-D SLAM System." In IEEE, 1691–96.
- 15. Engel, Jakob, Vladlen Koltun, and Daniel Cremers. 2018. "Direct Sparse Odometry." IEEE transactions on pattern analysis and machine intelligence 40(3): 611–25.
- Engel, Jakob, Thomas Schöps, and Daniel Cremers. 2014. "LSD-SLAM: Large-Scale Direct Monocular SLAM." In Springer, 834–49.
- 17. Engel, Jakob, Jörg Stückler, and Daniel Cremers. 2015. "Large-Scale Direct SLAM with
Stereo Cameras." In IEEE, 1935–42.

- 18. Engel, Jakob, Jurgen Sturm, and Daniel Cremers. 2013. "Semi-Dense Visual Odometry for a Monocular Camera." In , 1449–56.
- Fischler, Martin A, and Robert C Bolles. 1987. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." In Readings in Computer Vision, Elsevier, 726–40.
- 20. Forster, Christian, Matia Pizzoli, and Davide Scaramuzza. 2014. "SVO: Fast Semi-Direct Monocular Visual Odometry." In IEEE, 15–22.
- Fraundorfer, Friedrich, and Davide Scaramuzza. 2011. "Visual Odometry: Part i: The First 30 Years and Fundamentals." *IEEE Robotics and Automation Magazine* 18(4): 80– 92.
- 22. Fraundorfer, Friedrich, and Davide Scaramuzza . 2012. "Visual Odometry: Part Ii: Matching, Robustness, Optimization, and Applications." *IEEE Robotics & Automation Magazine* 19(2): 78–90.
- 23. Gao, Xiao-Shan, Xiao-Rong Hou, Jianliang Tang, and Hang-Fei Cheng. 2003. "Complete Solution Classification for the Perspective-Three-Point Problem." *IEEE* transactions on pattern analysis and machine intelligence 25(8): 930–43.
- 24. Geiger, Andreas, Philip Lenz, and Raquel Urtasun. 2012. "Are We Ready for Autonomous Driving? The KITTI Vision Benchmark Suite."
- 25. Glover, Arren et al. 2012. "OpenFABMAP: An Open Source Toolbox for Appearance-Based Loop Closure Detection." In IEEE, 4730–35.
- 26. Harris, Christopher G, and J M Pike. 1988. "3D Positional Integration from Image Sequences." Image and Vision Computing 6(2): 87–90.
- 27. Hartley, Richard. 1995. "In Defence of the 8-Point Algorithm." In IEEE, 1064–70.
- 28. Hartley, Richard, and Andrew Zisserman. 2003. Multiple View Geometry in Computer Vision. Cambridge university press.
- 29. Hartmann, Jan, Jan Helge Klussendorff, and Erik Maehle. 2013. "A Comparison of Feature Descriptors for Visual SLAM." In IEEE, 56–61.
- 30. Henry, Peter et al. 2010. "RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments." In Citeseer.
- Henry, Peter et al. 2012. "RGB-D Mapping: Using Kinect-Style Depth Cameras for Dense 3D Modeling of Indoor Environments." The International Journal of Robotics Research 31(5): 647–63.
- 32. Irani, Michal, and P Anandan. 1999. "About Direct Methods." In Springer, 267-77.
- 33. Jégou, Hervé, Matthijs Douze, Cordelia Schmid, and Patrick Pérez. 2010. "Aggregating

Local Descriptors into a Compact Image Representation." In IEEE, 3304–11.

- 34. Kasyanov, Anton, Francis Engelmann, Jörg Stückler, and Bastian Leibe. 2017. "Keyframe-Based Visual-Inertial Online SLAM with Relocalization." *arXiv preprint arXiv:1702.02175*.
- 35. Kerl, Christian, Jurgen Sturm, and Daniel Cremers. 2013. "Dense Visual SLAM for RGB-D Cameras." In Citeseer, 2100–2106.
- 36. Klein, Georg, and David Murray. 2007. "Parallel Tracking and Mapping for Small AR Workspaces." In IEEE, 225–34.
- 37. Konolige, Kurt, and Willow Garage. 2010. "Sparse Sparse Bundle Adjustment." In Citeseer, 101–2.
- Kümmerle, Rainer et al. 2011. "G 2 o: A General Framework for Graph Optimization." In IEEE, 3607–13.
- 39. Lacroix, Simon, Anthony Mallet, Raja Chatila, and Laurent Gallo. 1999. "Rover Self Localization in Planetary-like Environments." In , 433.
- 40. Lepetit, Vincent, Francesc Moreno-Noguer, and Pascal Fua. 2009. "Epnp: An Accurate o (n) Solution to the Pnp Problem." International journal of computer vision 81(2): 155.
- 41. Levenberg, Kenneth. 1944. "A Method for the Solution of Certain Non-Linear Problems in Least Squares." Quarterly of applied mathematics 2(2): 164–68.
- 42. Lim, Hyon, Jongwoo Lim, and H Jin Kim. 2014. "Real-Time 6-DOF Monocular Visual SLAM in a Large-Scale Environment." In IEEE, 1532–39.
- 43. Longuet-Higgins, H Christopher. 1981. "A Computer Algorithm for Reconstructing a Scene from Two Projections." *Nature* 293(5828): 133.
- 44. Lourakis, Manolis I A, and Antonis A Argyros. 2009. "SBA: A Software Package for Generic Sparse Bundle Adjustment." ACM Transactions on Mathematical Software (TOMS) 36(1): 2.
- 45. Lowe, David G. 2004. "Distinctive Image Features from Scale-Invariant Keypoints." International journal of computer vision 60(2): 91–110.
- 46. Marquardt, Donald W. 1963. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." Journal of the society for Industrial and Applied Mathematics 11(2): 431–41.
- 47. Montemerlo, Michael et al. 2002. "FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem." Aaai/iaai 593598.
- 48. Moravec, Hans. 1980. Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. Pittsburgh, PA: Carnegie Mellon University.

- 49. Mur-Artal, Raul, Jose Maria Martinez Montiel, and Juan D Tardos. 2015. "ORB-SLAM: A Versatile and Accurate Monocular SLAM System." *IEEE Transactions on Robotics* 31(5): 1147–63.
- 50. Newcombe, Richard A, Steven J Lovegrove, and Andrew J Davison. 2011. "DTAM: Dense Tracking and Mapping in Real-Time." In IEEE, 2320–27.
- 51. Nistér, David. 2004. "An Efficient Solution to the Five-Point Relative Pose Problem." *IEEE* transactions on pattern analysis and machine intelligence 26(6): 756–70.
- 52. Nistér, David, Oleg Naroditsky, and James Bergen. 2004. "Visual Odometry." In leee, I–I.
- 53. Nistér, David, and Henrik Stewenius. 2006. "Scalable Recognition with a Vocabulary Tree." In leee, 2161–68.
- 54. Olson, Clark F, Larry H Matthies, H Schoppers, and Mark W Maimone. 2000. "Robust Stereo Ego-Motion for Long Distance Navigation." In IEEE, 453–58.
- 55. Penate-Sanchez, Adrian, Juan Andrade-Cetto, and Francesc Moreno-Noguer. 2013. "Exhaustive Linearization for Robust Camera Pose and Focal Length Estimation." *IEEE* transactions on pattern analysis and machine intelligence 35(10): 2387–2400.
- 56. Peng, Zhen. 2012. "Efficient Matching of Robust Features for Embedded SLAM."
- 57. Poddar, Shashi, Rahul Kottath, and Vinod Karar. 2018. Evolution of Visual Odometry Techniques.
- 58. Raguram, Rahul, Jan-Michael Frahm, and Marc Pollefeys. 2008. "A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus." In Springer, 500–513.
- 59. Rosten, Edward, and Tom Drummond. 2006. "Machine Learning for High-Speed Corner Detection." In Springer, 430–43.
- 60. Rublee, Ethan, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. "ORB: An Efficient Alternative to SIFT or SURF." In IEEE, 2564–71.
- 61. Schmidt, Adam, Marek Kraft, and Andrzej Kasiński. 2010. "An Evaluation of Image Feature Detectors and Descriptors for Robot Navigation." In Springer, 251–59.
- 62. Shi, Jianbo, and Carlo Tomasi. 1993. Good Features to Track. Cornell University.
- 63. Sim, Robert et al. 2005. "Vision-Based SLAM Using the Rao-Blackwellised Particle Filter." In , 9–16.
- 64. Smith, Randall, Matthew Self, and Peter Cheeseman. 1990. "Estimating Uncertain Spatial Relationships in Robotics." In Autonomous Robot Vehicles, Springer, 167–93.
- 65. Strasdat, Hauke, José M M Montiel, and Andrew J Davison. 2012. "Visual SLAM: Why Filter?" Image and Vision Computing 30(2): 65–77.

- 66. Taketomi, Takafumi, Hideaki Uchiyama, and Sei Ikeda. 2017. "Visual SLAM Algorithms: A Survey from 2010 to 2016." IPSJ Transactions on Computer Vision and Applications 9(1): 16.
- 67. Tareen, Shaharyar Ahmed Khan, and Zahra Saleem. 2018. "A Comparative Analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK." In IEEE, 1–10.
- 68. Thrun, Sebastian, Wolfram Burgard, and Dieter Fox. 2005. Probabilistic Robotics. MIT press.
- 69. Triggs, Bill, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. 1999. "Bundle Adjustment—a Modern Synthesis." In Springer, 298–372.
- 70. Williams, Brian et al. 2009. "A Comparison of Loop Closing Techniques in Monocular SLAM." Robotics and Autonomous Systems 57(12): 1188–97.
- 71. Yousif, Khalid, Alireza Bab-Hadiashar, and Reza Hoseinnezhad. 2015. "An Overview to Visual Odometry and Visual Slam: Applications to Mobile Robotics." Intelligent Industrial Systems 1(4): 289–311.
- 72. Zhang, Zhengyou. 2000. "A Flexible New Technique for Camera Calibration." *IEEE Transactions on pattern analysis and machine intelligence* 22.
- 73. Zuliani, Marco. 2009. "RANSAC for Dummies." Vision Research Lab, University of California, Santa Barbara.
- 74. Καλησπεράκης, Ηλίας. 2010. "Η Επιπολική Γεωμετρία Στον Προβολικό Και Τον Ευκλείδειο Χώρο." Εθνικό Μετσόβιο Πολυτεχνείο (ΕΜΠ). Σχολή Αγρονόμων και Τοπογράφων Μηχανικών (ΣΑΤΜ).
- 75. Πέτσα, Έλλη. 2000. "Θεμελιώδεις Έννοιες Και Θεμελιώδη Προβλήματα Της Φωτογραμμετρίας." Αθήνα: ΤΕΙ Αθήνας, Τμήμα Τοπογραφίας.