

Διπλωματική εργασία

---

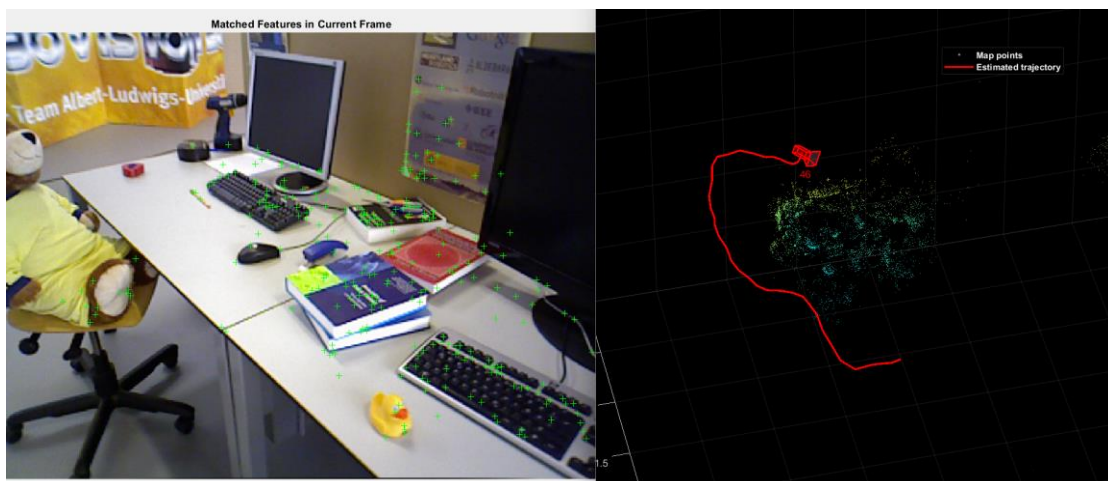
**Μελέτη και εφαρμογή αλγορίθμων Visual SLAM στην Φωτογραμμετρία  
και Όραση Υπολογιστών**

---

**Πέτσα Μαρία**  
(Αριθμός Μητρώου 17040)

Επιβλέπων καθηγητής :

Γραμματικόπουλος Λάζαρος, Αναπληρωτής καθηγητής ΠΑΔΑ



Αθήνα, Μάρτιος 2024

Diploma Thesis

---

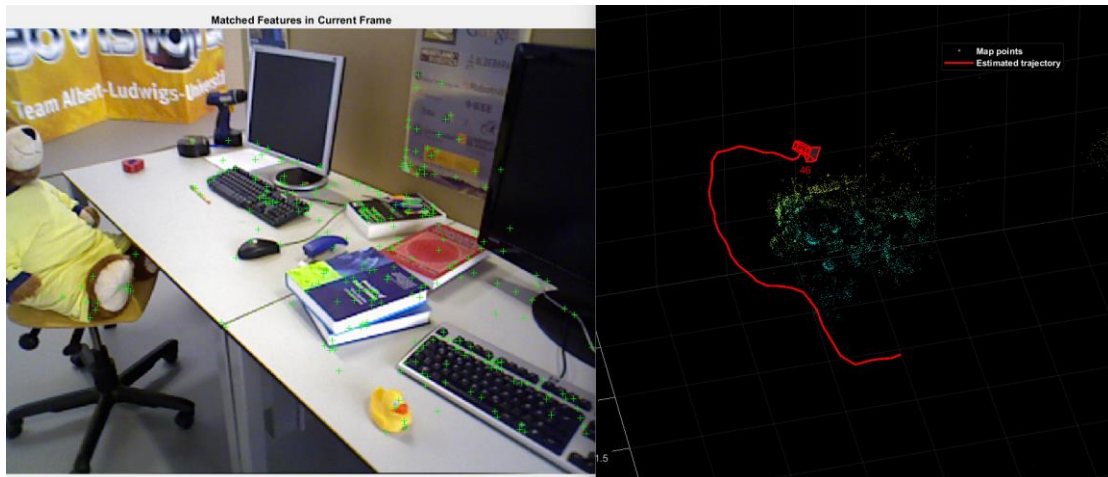
**Study and Implementation of Visual SLAM algorithms in  
Photogrammetry and Computer Vision**

---

**Petsa Maria**  
(Registration Number: 17040)

Supervisor:

Grammatikopoulos Lazaros, Associate Professor UNIWA



Athens, March 2024

## Μελέτη και Εφαρμογή Αλγορίθμων Visual SLAM στην Φωτογραμμετρία και Όραση Υπολογιστών

### Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η πτυχιακή/διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι Εξεταστική Επιτροπή:

Α/α	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΑΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1	Γραμματικόπουλος Λάζαρος (Επιβλέπων)	Αναπληρωτής Καθηγητής ΠΑΔΑ	
2	Πέτσα Ελένη	Καθηγήτρια ΠΑΔΑ	
3	Σφήκας Γεώργιος	Επίκουρος Καθηγητής ΠΑΔΑ	

## ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη Πέτσα Μαρία του Χρήστου, με αριθμό μητρώου 17040 φοιτήτρια του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Τοπογραφίας και Γεωπληροφορικής, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Η Δηλούσα

*“C’ est le temps que tu as perdu pour  
ta rose qui fait ta rose si importante.”*

**Antoine de Saint-Exupéry, Le Petit Prince**

## Ευχαριστίες

---

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κύριο Λάζαρο Γραμματικόπουλο για τις πολύτιμες γνώσεις στο αντικείμενο της Φωτογραμμετρίας και Όρασης Υπολογιστών και για την εμπιστοσύνη του στην ανάθεση του θέματος και την άμεση ανταπόκριση και καθοδήγηση καθ' όλη την διάρκεια της διπλωματικής μου εργασίας.

Ακόμα, θα ήθελα να ευχαριστήσω επιπλέον την κυρία Έλλη Πέτσα, τον κύριο Χρήστο Στεντούμη και τον κύριο Γιώργο Σφήκα για τις πολύτιμες γνώσεις Φωτογραμμετρίας.

Ευχαριστώ τον Ανδρέα Ελ Σαέρ, διδακτορικό φοιτητή του Πανεπιστημίου Δυτικής Αττικής, για την προθυμία του να με βοηθήσει καθώς και για τις πολύτιμες και χρήσιμες γνώσεις που μου έδωσε.

Ευχαριστώ ακόμα τους φίλους και συναδέλφους Μανώλη Ταουλάι και Μανώλη Παπαδημητρίου για τις πολύωρες φωτογραμμετρικές συζητήσεις οι οποίες εκτός από ενδιαφέρουσες ήταν και πολύ χρήσιμες.

Τέλος, ευχαριστώ την οικογένεια μου που με στήριξε καθ' όλη την διάρκεια των σπουδών μου.

## Περίληψη

---

Στην παρούσα διπλωματική εργασία μελετάται η διαδικασία ταυτόχρονου εντοπισμού θέσης και χαρτογράφησης χρησιμοποιώντας ως αισθητήρα μια RGB κάμερα (Monocular Visual SLAM). Αρχικά αναλύονται οι βασικές φωτογραμμετρικές διαδικασίες που είναι απαραίτητο να γνωρίζει κανείς προκειμένου να είναι ευκολότερα κατανοητή η μέθοδος Visual SLAM, ενώ στη συνέχεια πραγματοποιείται εκτενής βιβλιογραφική παρουσίαση των αλγορίθμων και των διαφορετικών μεθόδων που περιλαμβάνει. Τέλος, παρουσιάζεται ο αλγόριθμος ORB-SLAM και, μέσω του Matlab, εφαρμόζεται σε ακολουθίες βίντεο (δεδομένα αναφοράς από το Ίντερνετ). Κατά την εφαρμογή του, έγινε πειραματισμός με διαφορετικές παραμέτρους που αφορούν τα χαρακτηριστικά σημεία αλλά και με την εφαρμογή διαφορετικών αλγορίθμων εντοπισμού χαρακτηριστικών σημείων, συγκεκριμένα του SIFT και του SURF. Με αυτόν τον τρόπο μελετήθηκε διεξοδικά η επίδραση των χαρακτηριστικών σημείων στην τελική ακρίβεια του Visual SLAM.

Λέξεις κλειδιά: Visual SLAM, Monocular Visual SLAM, ORB, SURF, SIFT, εντοπισμός, χαρτογράφηση, Φωτογραμμετρία, Όραση Υπολογιστών, αναγνώριση σκηνής, bag-of-words, γωνιακό σφάλμα, τροχιακό σφάλμα.

## Abstract

---

In the present thesis, the process of simultaneous localization and mapping is studied using an RGB as a sensor (Monocular Visual SLAM). First, the basic photogrammetric procedures that one needs to know in order to understand the Visual SLAM method easier are analyzed, then an extensive bibliographic presentation of the algorithms and the different methods it includes. Finally, the ORB – SLAM algorithm is presented and, through MATLAB, applied to video sequences (reference data from the Internet). During its implementation, experimentation was done with different parameters concerning the characteristic points and also with the application of different algorithms for detecting features, specifically SIFT and SURF. In this way, the influence of feature points on the final accuracy of Visual SLAM was thoroughly studied.

Keywords: Visual SLAM, Monocular Visual SLAM, ORB, SURF, SIFT, tracking, mapping, Photogrammetry, Computer Vision, place recognition, bag-of-words, rotation error, translation error.

## Περιεχόμενα

Σκοπός της εργασίας.....	13
Δομή της εργασίας.....	13
1 Εισαγωγή.....	14
2 Βασικές φωτογραμμετρικές διαδικασίες.....	16
2.1 Εσωτερικός προσανατολισμός μηχανής – Συστήματα της εικόνας.....	16
2.1.1 Εσωτερικός προσανατολισμός – Ακτινική διαστρόφη (Radial Distortion) 17	
2.1.2 Εσωτερικός προσανατολισμός – Ασύμμετρη παραμόρφωση εκκεντρότητας (decentering distortion).....	19
2.1.3 Μη καθετότητα των pixel – Απόκλιση από την καθετότητα (skewness) 20	
2.1.4 Διαφορετική κλίμακα κατά x και κατά y (Aspect ratio).....	20
2.2 Πίνακας της μηχανής (camera matrix) και εξωτερικός προσανατολισμός...21	
2.3 Βαθμονόμηση μηχανής.....	22
2.4 Γεωμετρία λήψης και δημιουργία τρισδιάστατου μοντέλου από δύο εικόνες 28	
2.4.1 Δεσμευμένος επιπολικός πίνακας (E).....	30
2.4.2 Επιπολικός Πίνακας (F).....	31
2.5 Εντοπισμός και συνταύτιση χαρακτηριστικών σημείων.....	33
2.5.1 Ο αλγόριθμος SIFT.....	34
2.5.2 Ο αλγόριθμος SURF.....	35
2.5.3 Ο αλγόριθμος ORB.....	37
2.6 Βελτίωση αποτελεσμάτων.....	39
2.6.1 Αλγόριθμος RANSAC.....	39
2.6.2 Μέθοδος δέσμης.....	40
2.6.3 Συνόρθωση γράφου.....	40
3 Οπτικός Ταυτόχρονος Εντοπισμός θέσης και Χαρτογράφησης (Visual SLAM)42	
3.1 Διαφορά μεταξύ των μεθόδων φιλτραρίσματος και βασικών καρτέ.....	43
3.2 Τεχνικές κλεισίματος βρόχων και επανεντοπισμού.....	44
4 Περιγραφή του αλγορίθμου ORB – SLAM.....	46
5 Περιγραφή του αλγορίθμου της εφαρμογής.....	48
5.1 Αρχικοποίηση του χάρτη.....	48
5.2 Εντοπισμός θέσης και στροφών.....	49
5.3 Χαρτογράφηση.....	50
5.4 Κλείσιμο βρόχων.....	50



6	Αποτελέσματα.....	51
6.1	Αποτελέσματα εντοπισμού και εξαγωγής χαρακτηριστικών σημείων ORB	54
6.2	Αποτελέσματα εντοπισμού και εξαγωγής χαρακτηριστικών σημείων SURF	61
6.3	Αποτελέσματα εντοπισμού και εξαγωγής χαρακτηριστικών σημείων SIFT	67
6.4	Αποτέλεσμα τροχιάς και τρισδιάστατου χάρτη στο σετ δεδομένων που δημιουργήθηκε.....	71
6.5	Γενικά συμπεράσματα για τις διαφορετικές μεθόδους εντοπισμού και εξαγωγής χαρακτηριστικών σημείων .....	72
7	Συμπεράσματα .....	73
	Βιβλιογραφία .....	75

## Πίνακας εικόνων

Εικόνα 1: Σύστημα χώρου, εικόνας και εικονοσυντεταγμένων (Πηγή: Li S., Yoon H-S., 2023).	17
Εικόνα 2: Ακτινική παραμόρφωση.	18
Εικόνα 3: Μηνοειδής ακτινική παραμόρφωση (Πηγή: Neale W., Hessel D., Terpstra T., 2011).	18
Εικόνα 4: Πιθοειδής ακτινική παραμόρφωση (Πηγή: Πηγή: Neale W., Hessel D., Terpstra T., 2011).	19
Εικόνα 5: Έκκεντρη παραμόρφωση.	20
Εικόνα 6: Απόκλιση από την καθετότητα	20
Εικόνα 7: Διαφορετική κλίμακα κατά x και y.	21
Εικόνα 8: Μέρος του dataset βαθμονόμησης.	25
Εικόνα 9: Μέρος του dataset βαθμονόμησης.	25
Εικόνα 10: Μέρος του dataset βαθμονόμησης.	25
Εικόνα 11: Μέρος του dataset βαθμονόμησης.	25
Εικόνα 12: Χαρακτηριστικά σημεία πάνω στην σκακιέρα.	25
Εικόνα 13: Σφάλματα επαναπροβολής.	26
Εικόνα 14: Camera – centric.	26
Εικόνα 15: Pattern – centric.	26
Εικόνα 16: Μέσο σφάλμα επαναπροβολής.	27
Εικόνα 17: Εικόνα πριν την διόρθωση.	27
Εικόνα 18: Εικόνα μετά την διόρθωση.	27
Εικόνα 19: Εικόνα πριν την διόρθωση.	27
Εικόνα 20: Εικόνα μετά την διόρθωση.	27
Εικόνα 21: Επιπολική Γεωμετρία δυο εικόνων.	28
Εικόνα 22: Δημιουργία των διαφορετικών επιπέδων κλίμακας (Πηγή: (Lowe, 2004)).	34
Εικόνα 23: Δημιουργία του περιγραφέα (Πηγή: (Lowe, 2004)).	35
Εικόνα 24: Δημιουργία επιπέδων κλίμακας στον SURF (Πηγή: (Bay et al., 2006)).	36
Εικόνα 25: Υπολογισμός περιγραφέων στον αλγόριθμο SURF (Πηγή: (Bay et al., 2006)).	37
Εικόνα 26: Παράδειγμα προβλήματος γράφου. Οι κόμβοι παρουσιάζονται ως κύκλοι και οι ακμές ως γραμμές που ενώνουν τους κόμβους. (Πηγή: (Kummerle et al., 2011)).	41
Εικόνα 27: Περιγραφή του αλγορίθμου ORB - SLAM (Πηγή: Mur-Artal et al. (2015)).	47
Εικόνα 28: Μέρος του σετ δεδομένων Freiburg 2 desk validation του TUM ( <a href="https://cvg.cit.tum.de/data/datasets/rgbd-dataset">https://cvg.cit.tum.de/data/datasets/rgbd-dataset</a> ).	52
Εικόνα 29: Μέρος του σετ δεδομένων Freiburg 3 long office household του TUM ( <a href="https://cvg.cit.tum.de/data/datasets/rgbd-dataset">https://cvg.cit.tum.de/data/datasets/rgbd-dataset</a> ).	52
Εικόνα 30: Σετ δεδομένων που δημιουργήθηκε όπου απεικονίζονται αντικείμενα πάνω σε τραπέζι.	53
Εικόνα 31: Freiburg 2-Scale Factor παραμένει σταθερό και ίσο με 1.2 και αλλάζουν τα Num Levels.	55
Εικόνα 32: Freiburg 2-Scale Factor παραμένει σταθερό και ίσο με 1.3 και αλλάζουν τα Num Levels.	55

Εικόνα 33: Freiburg 2-Scale Factor παραμένει σταθερό και ίσο με 1.4 και αλλάζουν τα Num Levels. ....	56
Εικόνα 34: Freiburg 2-Scale Factor παραμένει σταθερό και ίσο με 1.5 και αλλάζουν τα Num Levels. ....	56
Εικόνα 35: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για scale Factor 1.3.....	57
Εικόνα 36: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για scale Factor 1.5.....	57
Εικόνα 37: Freiburg 3 - Scale Factor παραμένει σταθερό και ίσο με 1.2 και αλλάζουν τα Num Levels. ....	58
Εικόνα 38: Freiburg 3 - Scale Factor παραμένει σταθερό και ίσο με 1.3 και αλλάζουν τα Num Levels. ....	58
Εικόνα 39: Freiburg 3 - Scale Factor παραμένει σταθερό και ίσο με 1.4 και αλλάζουν τα Num Levels. ....	58
Εικόνα 40: Freiburg 3 - Scale Factor παραμένει σταθερό και ίσο με 1.5 και αλλάζουν τα Num Levels. ....	59
Εικόνα 41: Freiburg 3- Αποτελέσματα τροχιάς και σημείων για scale Factor 1.2 num Levels 3.....	60
Εικόνα 42: Freiburg 3- Αποτελέσματα τροχιάς και σημείων για scale Factor 1.2 num Levels 6.....	60
Εικόνα 43: : Freiburg 3- Αποτελέσματα τροχιάς και σημείων για scale Factor 1.3 num Levels 7.....	60
Εικόνα 44: Freiburg 2-Αλλαγή στο Metric Threshold. ....	62
Εικόνα 45: Freiburg 2-Αλλαγή στο Num Octaves. ....	62
Εικόνα 46: Freiburg 2-Αλλαγή στο Num Scale Levels. ....	62
Εικόνα 47: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Metric Threshold 500. ....	63
Εικόνα 48: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Num Octaves 4. ....	63
Εικόνα 49: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Num Scale Levels 5.....	64
Εικόνα 50: Freiburg 3-Αλλαγή στο Metric Threshold.. ....	65
Εικόνα 51: Freiburg 3-Αλλαγή στο Num Octaves. ....	65
Εικόνα 52: Freiburg 3-Αλλαγή στο Num Scale Levels. ....	65
Εικόνα 53: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Metric Threshold 500. ....	66
Εικόνα 54: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για num Octaves 4.....	66
Εικόνα 55: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Num Scale Levels 5.....	67
Εικόνα 56: Freiburg 3 – Αλλαγή στο Contrast Threshold. ....	68
Εικόνα 57: Freiburg 3 – Αλλαγή στο Edge Threshold. ....	68
Εικόνα 58: Freiburg 3 – Αλλαγή στο Num Layers In Octave. ....	69
Εικόνα 59: Freiburg 3 – Αλλαγή στο Sigma. ....	69
Εικόνα 60: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Contrast Threshold 0.008.....	70

Εικόνα 61: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Edge Threshold 20.0.....	70
Εικόνα 62: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Sigma 1.7.....	71
Εικόνα 63: Αποτέλεσμα αλγορίθμου Visual SLAM στο σετ δεδομένων που απεικονίζει τα αντικείμενα στο τραπέζι.....	71
Εικόνα 64: Ελάχιστες τιμές σφαλμάτων από όλες τις δοκιμές και στα δυο σετ δεδομένων.....	72

## Σκοπός της εργασίας

---

Σκοπός της παρούσας εργασίας είναι η μελέτη και η κατανόηση του προβλήματος του Οπτικής Πλοήγησης και Χαρτογράφησης (Visual SLAM) τόσο θεωρητικά όσο και πρακτικά. Στο θεωρητικό μέρος, παρουσιάζονται διάφορες μέθοδοι που χρησιμοποιούνται σε διαφορετικές εφαρμογές Visual SLAM. Πραγματοποιείται εκτενής ανάλυση των θεωρητικών φωτογραμμετρικών γνώσεων που είναι απαραίτητες.

Στο πρακτικό κομμάτι, εφαρμόζεται ένας αλγόριθμος Visual SLAM, βασισμένος στον αλγόριθμο ORB - SLAM με αλλαγή σετ δεδομένων και αλγορίθμων χαρακτηριστικών σημείων. Έτσι, θα μελετηθούν τα δυνατά σημεία αλλά και οι αδυναμίες του αλγορίθμου. Πραγματοποιήθηκε, ακόμα δημιουργία ενός σετ δεδομένων για τα πλαίσια της εργασίας, οπότε μελετήθηκε και η διαδικασία υπολογισμού του εσωτερικού προσανατολισμού καθώς και οι δυσκολίες που μπορεί να συναντηθούν.

## Δομή της εργασίας

---

Η παρούσα εργασία χωρίζεται στα εξής κεφάλαια: Στο πρώτο Κεφάλαιο πραγματοποιείται μια εισαγωγή στην διαδικασία του Visual SLAM μέσω παρουσίασης των βασικών προβληματισμών αλλά και μιας ιστορικής αναδρομής.

Στο δεύτερο Κεφάλαιο παρουσιάζεται το θεωρητικό υπόβαθρο, το οποίο είναι απαραίτητο. Αρχικά μελετάται το πρόβλημα του εσωτερικού προσανατολισμού και η σχέση των συστημάτων της κάμερας, της εικόνας και του χώρου. Έπειτα, παρουσιάζεται η γεωμετρική σχέση που προκύπτει από δυο εικόνες και πως, εφαρμόζοντας την μπορεί κανείς να εξάγει τρισδιάστατη πληροφορία μόνο από δυο εικόνες. Σε αυτό το σημείο συστήνονται οι έννοιες του Επιπολικού Πίνακα (Fundamental Matrix) και του Δεσμευμένου Επιπολικού Πίνακα (Essential Matrix). Μελετώνται οι αλγόριθμοι εντοπισμού και συνταύτισης χαρακτηριστικών SIFT, SURF και ORB καθώς και ο αλγόριθμος απόρριψης των ακραίων τιμών RANSAC. Στο τέλος του δεύτερου κεφαλαίου μελετώνται μέθοδοι συνόρθωσης, πιο συγκεκριμένα η μέθοδος δέσμης και η συνόρθωση γράφου, για την βελτιστοποίηση των αποτελεσμάτων.

Στο τρίτο Κεφάλαιο πραγματοποιείται μια συνοπτική περιγραφή του προβλήματος του ταυτόχρονου εντοπισμού θέσης και χαρτογράφησης. Συντελείται μια ανασκόπηση διαφορετικών αλγορίθμων στην Βιβλιογραφία.

Στο τέταρτο Κεφάλαιο παρουσιάζεται ο αλγόριθμος ORB SLAM στον οποίο βασίζεται ο κώδικας που χρησιμοποιήθηκε.

Τέλος στο πέμπτο Κεφάλαιο παρουσιάζεται ο κώδικας που χρησιμοποιήθηκε καθώς και τα αποτελέσματα που προέκυψαν.

## 1 Εισαγωγή

Στην Φωτογραμμετρία, στην Όραση Υπολογιστών αλλά και στην Ρομποτική το πρόβλημα του ταυτόχρονου εντοπισμού θέσης και χαρτογράφησης (SLAM) είναι αρκετά δημοφιλές εδώ και πολλά χρόνια. Η μέθοδος SLAM στοχεύει την ικανότητα ενός κινούμενου οχήματος (ή ρομπότ) να εντοπίζει την θέση και τον προσανατολισμό του στον χώρο, ενώ ταυτόχρονα να δημιουργεί έναν χάρτη του περιβάλλοντος.

Για την πραγματοποίηση του αλγορίθμου αυτού, χρησιμοποιούνται διάφοροι αισθητήρες. Μπορεί, για παράδειγμα, να χρησιμοποιηθεί ως μοναδικός αισθητήρας μια RGB κάμερα, έτσι ο υπολογισμός της θέσης και του προσανατολισμού αλλά και του τρισδιάστατου χάρτη βασίζονται αποκλειστικά στις διαδικασίες της Φωτογραμμετρίας και της Όρασης Υπολογιστών. Ακόμα, μπορεί να χρησιμοποιηθεί αισθητήρας RGB – D όπου μπορεί να αξιοποιηθεί ως πληροφορία και η απόσταση των αντικειμένων από τον αισθητήρα. Επιπλέον, μπορεί να εφαρμοστεί αλγόριθμος SLAM όπου να αντλεί την πληροφορία του χώρου από αισθητήρες LIDAR (Light Detection And Ranging). Φυσικά, μπορεί να χρησιμοποιηθεί συνδυασμός της οπτικής μεθόδου και των μεθόδων SLAM με LIDAR. Για τον εντοπισμό της θέσης του οχήματος ο αισθητήρας μπορεί να συνδυαστεί με αισθητήρες εντοπισμού θέσης GNSS.

Οι έρευνες πάνω στο συγκεκριμένο πρόβλημα χρονολογούνται ήδη από το 1986, όπου εκείνη την χρονιά ξεκινούσαν να εφαρμόζονται οι πιθανολογικές θεωρίες ως μέσο επίλυσης προβλημάτων τόσο στην Τεχνητή Νοημοσύνη όσο και στην Εκμάθηση Μηχανής (Durrant – Whyte & Bailey, 2006).

Πιο συγκεκριμένα, σύμφωνα με τους Durrant-Whyte & Bailey (2006), στο συνέδριο IEEE Robotics and Automation που πραγματοποιήθηκε το 1986 έγινε πρώτη φορά αναφορά σε πρόβλημα SLAM με πιθανολογικό χαρακτήρα. Το συγκεκριμένο συνέδριο αποτέλεσε ορόσημο σε ό,τι αφορά τα μοντέλα επίλυσης των SLAM προβλημάτων, καθώς πολλοί επιστήμονες προσέγγισαν τόσο το πρόβλημα του εντοπισμού όσο και της χαρτογράφησης με την επιστήμη της θεωρίας πιθανοτήτων και με πιθανολογικές μεθόδους (Durrant-Whyte & Bailey, 2006). Αυτή η προσέγγιση είχε ως αποτέλεσμα να συζητηθούν καθώς και να προταθούν διάφορες ιδέες για την εφαρμογή της θεωρίας πιθανοτήτων στο πρόβλημα του εντοπισμού θέσης και την χαρτογράφηση, και τελικά να αποδειχθεί πως η εφαρμογή πιθανοτικών θεωριών είναι μια καλή λύση στο πρόβλημα της συνεχούς χαρτογράφησης (Durrant-Whyte & Bailey, 2006). Η θεωρία πιθανοτήτων χρησιμοποιείται στα προβλήματα SLAM όταν το όχημα προβλέπει ποια θα είναι η επόμενη κατάσταση του, πριν βρεθεί σε αυτήν.

Τα επόμενα χρόνια αποδείχθηκαν αρκετά εποικοδομητικά σε ό,τι αφορά την έρευνα πάνω σε συστήματα SLAM με πιθανολογικό χαρακτήρα. Υπήρξαν σημαντικές έρευνες και δημοσιεύσεις πάνω στο πρόβλημα αυτό όπως για παράδειγμα των Ayache & Faugeras, όπου ήταν από τις αρχικές έρευνες της οπτικής πλοήγησης (Durrant-Whyte & Bailey, 2006).

Μια δημοσίευση – ορόσημο για την διαδικασία του ταυτόχρονου εντοπισμού και χαρτογράφησης πραγματοποιήθηκε από τους Smith et al. (1990) όπου πρακτικά απέδειξαν πως τα χαρακτηριστικά σημεία που μετράει ένα αυτόνομο όχημα, προκειμένου να πλοηγηθεί σε ένα άγνωστο περιβάλλον, συσχετίζονται μεταξύ τους.

Άρα μια πλήρη λύση στο πρόβλημα εντοπισμού και χαρτογράφησης απαιτεί συνδυασμό της θέσης του οχήματος και κάθε χαρακτηριστικού σημείου που έχει μετρήσει. Οι Durrant-Whyte & Bailey (2006) αναφέρουν πως στο συγκεκριμένο άρθρο δεν μελετήθηκε η σύγκλιση που θα είχε ο χάρτης και θεωρήθηκε πως τα σφάλματα δεν θα σύγκλιναν αλλά αντιθέτως θα μεγάλωναν διαρκώς. Αυτό, σε συνδυασμό και με την πολυπλοκότητα της χαρτογράφησης οδήγησαν τους ερευνητές στην προσπάθεια μείωσης ή πλήρους απαλοιφής της συσχέτισης μεταξύ των ορόσημων και της θέσης του οχήματος, και τελικά στον διαχωρισμό του εντοπισμού και της χαρτογράφησης ως ξεχωριστά προβλήματα.

Φυσικά, αυτό δεν διήρκησε πολύ, καθώς σύντομα οι ερευνητές συνειδητοποίησαν πως το πρόβλημα του εντοπισμού και της χαρτογράφησης είναι αναπόφευκτο να λυθούν ταυτόχρονα. Το 1995 στο International Symposium on Robotics Research συστήθηκε για πρώτη φορά το ακρωνύμιο SLAM μαζί με την δομή του (Durrant-Whyte & Bailey, 2006).

Σύμφωνα με τον Stachniss (2013), για να πραγματοποιηθεί μια διαδικασία χαρτογράφησης θα πρέπει να είναι γνωστές, με πολύ καλή ακρίβεια, οι θέσεις του οχήματος καθ' όλη την διάρκεια της κίνησης του στον χώρο. Αντίστοιχα, για να πραγματοποιηθεί μια διαδικασία εντοπισμού πρέπει να είναι διαθέσιμος ένας ακριβής χάρτης του χώρου στον οποίο κινείται. Και στις δύο περιπτώσεις θεωρούμε πως είναι γνωστά με τόσο καλή ακρίβεια που αντιπροσωπεύουν την πραγματικότητα.

Όπως είναι αναμενόμενο, δεν είναι εύκολο να γνωρίζει κανείς σε ποιες ακριβώς θέσεις βρέθηκε το όχημα. Αν, για παράδειγμα, χρησιμοποιεί έναν πολύ καλό αισθητήρα GNSS που μετράει τις συντεταγμένες του στον χώρο κάθε στιγμή, δεν είναι βέβαιο πως αυτές οι μετρήσεις θα είναι πλήρως απαλλαγμένες από σφάλματα. Ενδέχεται τα σφάλματα του αισθητήρα να αυξάνονται διαρκώς επηρεάζοντας το τελικό αποτέλεσμα. Αυτό αποτελεί το βασικότερο πρόβλημα της χαρτογράφησης. Ο Stachniss (2013) αναφέρει πως είναι λάθος να θεωρήσει κανείς πως γνωρίζει χωρίς σφάλμα την κίνηση του οχήματος στον χώρο και που ακριβώς είναι τα σημεία που μέτρησε.

Σύμφωνα με τους Durrant-Whyte & Bailey (2006), το πρόβλημα του ταυτόχρονου εντοπισμού θέσης και χαρτογράφησης έχει πλέον λυθεί από την θεωρητική του πλευρά, με διάφορους τρόπους. Η εφαρμογή του, όμως, εξακολουθεί να παραμένει αντικείμενο προς μελέτη αφού είναι επιθυμητή η δημιουργία ενός συστήματος SLAM/ Visual SLAM όπου να μπορεί να εφαρμοστεί σε οποιοδήποτε περιβάλλον με επιτυχία ή να μπορεί να εφαρμοστεί σε δυναμικά περιβάλλοντα.

Όπως γίνεται αντιληπτό, το πρόβλημα SLAM είναι ανοικτό στην επιστημονική κοινότητα εδώ και πολλά χρόνια. Μέχρι σήμερα έχουν αναπτυχθεί διαφορετικοί αλγόριθμοι για την επίλυσή του και εξακολουθούν να προτείνονται καινούριες μέθοδοι ή να βελτιώνονται οι ήδη υπάρχουσες.

Φυσικά μπορεί να γίνει συνδυασμός περισσότερων του ενός αισθητήρα, για παράδειγμα μπορεί να εφαρμοστεί Visual SLAM με στερεοσκοπική κάμερα και με αισθητήρα IMU και GPS. Περισσότερα για αυτά θα παρουσιαστούν στα επόμενα κεφάλαια όπου θα γίνει βιβλιογραφική ανασκόπηση στους πιο γνωστούς αλγορίθμους SLAM/ Visual SLAM.

Ο οπτικός ταυτόχρονος εντοπισμός θέσης και χαρτογράφησης (Visual SLAM) στηρίζεται στην γεωμετρία της κεντρικής προβολής καθώς και στην αυτοματοποίηση των βασικών φωτογραμμετρικών μεθόδων που επιτρέπουν την εξαγωγή γεωμετρικής πληροφορίας από εικόνες. Συνεπώς, είναι απαραίτητο να γνωρίζει κανείς τις διαδικασίες αυτές ώστε να μπορέσει να κατανοήσει εις βάθος τις διαδικασίες των αλγόριθμων Visual SLAM.

## 2 Βασικές φωτογραμμετρικές διαδικασίες

Σε αυτό το κεφάλαιο θα αναφερθεί το θεωρητικό υπόβαθρο και θα περιγραφούν οι βασικότερες φωτογραμμετρικές διαδικασίες που σχετίζονται με την μέθοδο V-SLAM.

### 2.1 Εσωτερικός προσανατολισμός μηχανής – Συστήματα της εικόνας

Το μαθηματικό μοντέλο που ορίζει την γεωμετρία της λήψης είναι η κεντρική προβολή, και το πιο γνωστό χαρακτηριστικό της κεντρικής προβολής είναι το γεγονός πως τα σημεία δεν αναπαρίστανται με μια ενιαία κλίμακα (Πέτσα Ε., 2000).

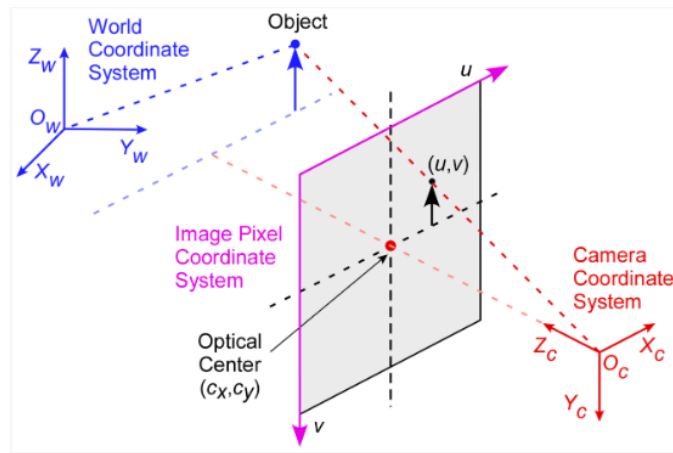
Τα σημεία του χώρου ορίζουν με το προβολικό κέντρο μια δέσμη προβολικών ακτίνων που τέμνουν το επίπεδο της εικόνας στα αντίστοιχα εικονοσημεία. Ο εσωτερικός προσανατολισμός είναι η διαδικασία της μαθηματικής αναδημιουργίας της δέσμης των προβολικών ακτίνων που δημιουργήσαν την εικόνα (Πέτσα Ε., 2000). Η εύρεση του εσωτερικού προσανατολισμού της κάμερας είναι αρκετά κρίσιμη για τις φωτογραμμετρικές διαδικασίες που έπονται αυτής. Με τον εσωτερικό προσανατολισμό ορίζεται η θέση του προβολικού κέντρου και η απόστασή του από το επίπεδο του αισθητήρα (Πέτσα Ε., 2000), δηλαδή το σύστημα της εικόνας. Σύστημα της εικόνας είναι εκεί όπου πραγματοποιούνται οι μετρήσεις, οι οποίες εκφράζονται ως προς το κέντρο προβολής.

Συχνά γίνεται σύγχυση μεταξύ του συστήματος της εικόνας και του συστήματος των εικονοσυντεταγμένων. Με απλά λόγια θα μπορούσε κανείς να τα περιγράψει ως εξής:

- Το σύστημα της εικόνας: είναι τρισδιάστατο σύστημα με αρχή το κέντρο προβολής
- Το σύστημα των εικονοσυντεταγμένων: είναι ένα δισδιάστατο σύστημα με αρχή το κέντρο της εικόνας.

Το κέντρο προβολής και το κέντρο της εικόνας δεν είναι απαραίτητο να ταυτίζονται. Η διαφορά μεταξύ των συστημάτων είναι αρκετά εμφανής, καθώς στο ένα σύστημα είναι οι μετρήσεις που κάνουμε ως προς την (ψηφιακή ή αναλογική) εικόνα, ενώ το άλλο σύστημα αναφέρεται στον αισθητήρα (δηλαδή στην κάμερα). Μια οπτική αναπαράσταση των δύο συστημάτων αλλά και του συστήματος του χώρου, παρουσιάζεται στην [Εικόνα 1](#).





Εικόνα 1: Σύστημα χώρου, εικόνας και εικονοσυντεταγμένων (Πηγή: Li S., Yoon H-S., 2023).

Γενικώς στις φωτογραμμετρικές διαδικασίες αναφερόμαστε στο σύστημα της εικόνας, διότι με βάση αυτό το σύστημα μπορεί κανείς να εφαρμόσει μετασχηματισμούς για να ορίσει την θέση της κάμερας στον χώρο (εξωτερικός προσανατολισμός).

Προφανώς, τα δυο αυτά συστήματα έχουν σχέση μεταξύ τους και μπορεί κανείς να εφαρμόσει μετασχηματισμούς από το ένα σύστημα στο άλλο.

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} - \begin{bmatrix} x_0 \\ y_0 \\ c \end{bmatrix} \quad (1)$$

Όπου  $c$  αντιπροσωπεύει την σταθερά της μηχανής, η οποία αναπαριστά την απόσταση του κέντρου του φακού από το επίπεδο του αρνητικού.

Η Εξίσωση (1) είναι η σχέση που συνδέει το σύστημα της εικόνας με το σύστημα των εικονοσυντεταγμένων. Μπορεί, κανείς, δηλαδή να υπολογίσει σε ποιο σημείο της εικόνας, σε συντεταγμένες pixel, προβάλλεται το πρωτεύον σημείο. Το πρωτεύον σημείο είναι η προβολή του σημείου λήψης στο επίπεδο της εικόνας.

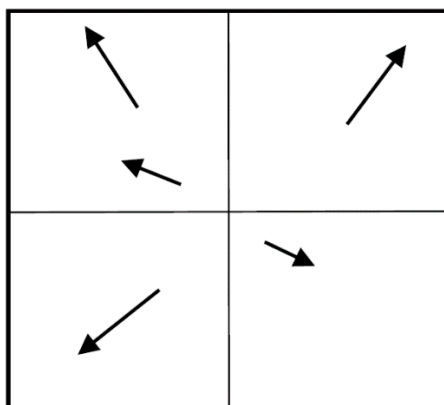
Συνδέοντας τα δύο συστήματα αναφοράς δεν συνεπάγεται και προσδιορισμός του εσωτερικού προσανατολισμού. Συχνά συναντάται το πρόβλημα της απόκλισης από την καθετότητα (skewness) και η διαφορετική κλίμακα κατά  $x$  και κατά  $y$  (aspect ratio). Ακόμα, ο φακός μιας φωτογραφικής μηχανής είναι αδύνατον να μην περιέχει σφάλματα. Τα πιο βασικά σφάλματα που επηρεάζουν τον εσωτερικό προσανατολισμό είναι η ακτινική διαστρόφη (radial distortion) και η ασύμμετρη παραμόρφωση (decentering distortion).

### 2.1.1 Εσωτερικός προσανατολισμός – Ακτινική διαστρόφη (Radial Distortion)

Είναι γνωστό πως οι φακοί δημιουργούν διαφορετικές παραμορφώσεις (σφάλματα απόκλισης από την κεντρική προβολή). Ένα από τα σημαντικότερα σφάλματα που συναντώνται στους φακούς των φωτογραφικών μηχανών είναι η ακτινική διαστρόφη.

Η ακτινική διαστρόφη έχει ως συνέπεια οι ακτίνες που δημιούργησαν την εικόνα (οι προβολικές ακτίνες) να μην εξέρχονται από τον φακό με την ίδια γωνία με την οποία προσπίπτουν (Πέτσα Ε., 2000). Η ακτινική παραμόρφωση μετατοπίζει τα σημεία της

εικόνας ακτινικά προς ή από το πρωτεύον σημείο (Πέτσα Ε., 2000). Στο ακόλουθο σχήμα παρουσιάζεται γραφικά η συγκεκριμένη παραμόρφωση.



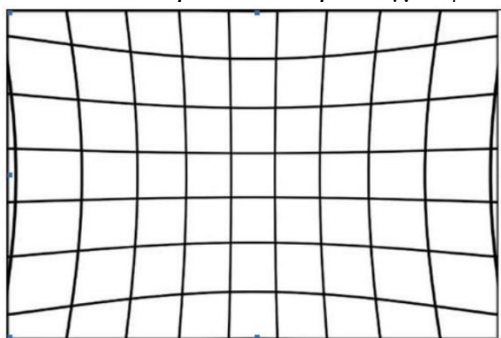
Εικόνα 2: Ακτινική παραμόρφωση.

Όπως φαίνεται και στην [Εικόνα 2](#), ένα ακόμα χαρακτηριστικό της ακτινικής διαστρόφης είναι πως δεν επιδρά με τον ίδιο τρόπο σε όλη την εικόνα. δηλαδή, σε κάποια σημεία είναι μεγαλύτερη από κάποια άλλα, ενώ σε κάποια είναι μικρότερη. Πιο συγκεκριμένα, μεγαλώνει αναλογικά με την απόσταση από το κέντρο της εικόνας (Πέτσα Ε., 2000). Συνέπεια της ακτινικής διαστρόφης είναι η διαφορετική κλίμακα που αποτελεί «*Συνάρτηση της ακτινικής απόστασης από το πρωτεύον σημείο*» (Πέτσα Ε. 2000, σελ. 28).

Είναι εμφανές ότι η καμπύλωση των ευθειών δεν συμπίπτει με τις απαιτήσεις της κεντρικής προβολής. Γι' αυτόν τον λόγο είναι αδύνατη η εύρεση σταθεράς της μηχανής που θα ορίζει σωστά το πρωτεύον σημείο στην εικόνα, και θα επαληθεύει κάθε δέσμη ακτινών (Πέτσα Ε., 2000).

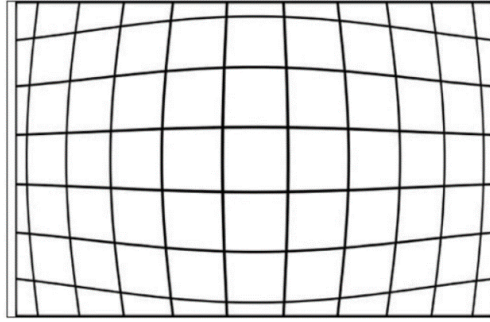
Η ακτινική διαστρόφη χωρίζεται σε δυο είδη:

1. Μηνοειδής (αρνητική): Τα σημεία στην πραγματικότητα θα έπρεπε να είναι ακτινικά πιο μακριά από το κέντρο. Ένα παράδειγμα φαίνεται στην [Εικόνα 3](#).



Εικόνα 3: Μηνοειδής ακτινική παραμόρφωση (Πηγή: Neale W., Hessel D., Terpstra T., 2011).

2. Πιθοειδής (θετική): Τα σημεία στην πραγματικότητα θα έπρεπε να είναι ακτινικά προς το κέντρο. Ένα παράδειγμα φαίνεται στην [Εικόνα 4](#).



Εικόνα 4: Πιθοειδής ακτινική παραμόρφωση (Πηγή: Πηγή: Neale W., Hessel D., Terpstra T., 2011).

Η ακτινική παραμόρφωση περιγράφεται από ένα πολυώνυμο με περιττές δυνάμεις του  $r$  (Εξίσωση ( 2 )).

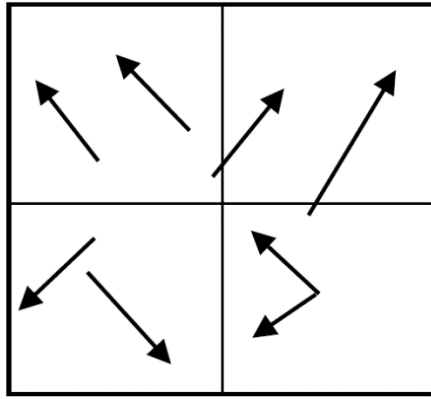
$$\Delta r = k_0 r + k_1 r^3 + k_2 r^5 + k_3 r^7 + \dots \quad (2)$$

Το μέγεθος  $\Delta r$  περιγράφεται από τους δυο, το πολύ τρεις πρώτους όρους. Ο πρώτος όρος συνήθως απορροφάται από την σταθερά της μηχανής (Πέτσα Ε., 2000). Τα  $k_1, k_2, \dots$ , είναι συντελεστές, τους οποίους πρέπει να υπολογίσει κανείς για να μπορέσει να βρει την ακτινική διαστροφή και να εξαλείψει τα σφάλματα των φακών. Όταν το  $k_1 < 0$  τότε η παραμόρφωση είναι πιθοειδής, ενώ όταν  $k_1 > 0$  είναι μηνοειδής. Είναι πιθανό να υπάρχει και συνδυασμός τους.

### 2.1.2 Εσωτερικός προσανατολισμός – Ασύμμετρη παραμόρφωση εκκεντρότητας (decentering distortion)

Η ασύμμετρη παραμόρφωση εκκεντρότητα (ή έκκεντρη παραμόρφωση) είναι αποτέλεσμα της μη παραλληλίας του φακού και του επιπέδου της εικόνας και αναλύεται σε μια εφαπτομενική και μια ασύμμετρη ακτινική συνιστώσα (Νάκης, 2019).

Η εφαπτομενική παραμόρφωση υπάρχει σε παράλληλο προσανατολισμό σε κάθε σημείο της εικόνας και μεταβάλλεται ως συνάρτηση της κατεύθυνσης από το πρωτεύον σημείο σε κάθε σημείο. Στο ακόλουθο σχήμα παρουσιάζεται με γραφικό τρόπο η ασύμμετρη παραμόρφωση εκκεντρότητας.



Εικόνα 5: Έκκεντρη παραμόρφωση.

Οι Brown & Fryer (1986) αναφέρουν πως στο μοντέλο του Conrady οι συνιστώσες της ακτινικής ( $\Delta_r$ ) και της εφαπτομενικής ( $\Delta_t$ ) παραμόρφωσης δίνονται από τις σχέσεις:

$$\Delta_r = 3P(r) \sin(\varphi - \varphi_0) \quad (3)$$

$$\Delta_t = P(r) \cos(\varphi - \varphi_0) \quad (4)$$

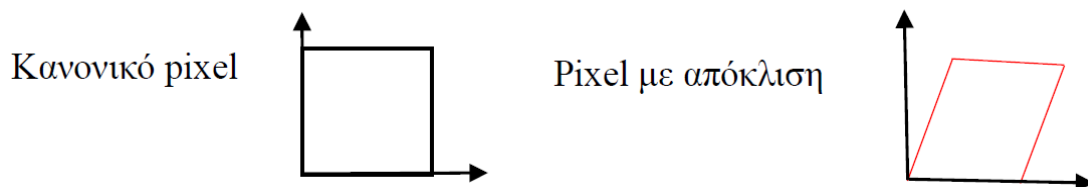
Αυτές οι σχέσεις αφορούν μια εικόνα με  $(x, y)$  συντεταγμένες που έχουν αναχθεί ως προς το πρωτεύον σημείο (Brown & Fryer, 1986). Ως προς τις συντεταγμένες  $(x, y)$  οι σχέσεις γίνονται ως εξής ((Brown & Fryer, 1986):

$$\Delta_x = [P_1(r^2 + 2x^2) + 2P_2xy][1 + P_3r^2 + \dots] \quad (5)$$

$$\Delta_y = [P_2(r^2 + 2y^2) + 2P_1xy][1 + P_3r^2 + \dots] \quad (6)$$

### 2.1.3 Μη καθετότητα των pixel – Απόκλιση από την καθετότητα (skewness)

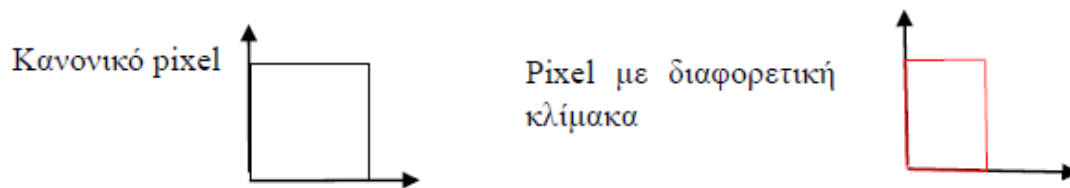
Σε αυτή την περίπτωση τα pixel της εικόνας δεν είναι ορθογώνια, αλλά σχηματίζουν γωνιά που δεν είναι ίση με  $90^\circ$ . Αυτό προκαλεί στρέβλωση στην εικόνα και, φυσικά, αν δεν ληφθεί υπόψιν επηρεάζει σημαντικά και τις ακόλουθες φωτογραμμετρικές διαδικασίες. Στο ακόλουθο σχήμα φαίνεται παράδειγμα ενός κανονικού pixel και ενός pixel με απόκλιση από την καθετότητα.



Εικόνα 6: Απόκλιση από την καθετότητα

### 2.1.4 Διαφορετική κλίμακα κατά x και κατά y (Aspect ratio)

Όπως και προηγουμένως, έτσι και σε αυτή την περίπτωση τα pixel δεν έχουν πλέον τετράγωνο σχήμα, καθώς οι πλευρές διαφέρουν ανάλογα με την διαφορετική κλίμακα. Ένα παράδειγμα φαίνεται στο ακόλουθο σχήμα.



Εικόνα 7: Διαφορετική κλίμακα κατά x και y.

Και σε αυτή την περίπτωση, αν είναι αναγκαίο να μετρηθούν εικονοσυντεταγμένες για τις ακόλουθες φωτογραμμετρικές διαδικασίες, είναι απαραίτητο να ληφθεί υπόψη η συγκεκριμένη διαφορά κλίμακας, εφόσον υπάρχει. Διαφορετικά θα προσθέσει ένα σφάλμα κλίμακας σε όλες τις μετρήσεις.

## 2.2 Πίνακας της μηχανής (camera matrix) και εξωτερικός προσανατολισμός

Όλες οι προηγούμενες παράμετροι μπορούν να περιγραφούν με τον Πίνακα Μηχανής, γνωστό και ως Camera Matrix. Ο Πίνακας της Μηχανής συμβολίζεται συνήθως στην βιβλιογραφία με το γράμμα  $\mathbf{K}$  και περιγράφει την γεωμετρία της μηχανής (Νάκης, 2019). Ο πίνακας  $\mathbf{K}$  για μηχανή με τετράγωνα pixel, δηλαδή χωρίς απόκλιση από την καθετότητα και διαφορά κλίμακας, έχει την εξής μορφή:

$$\mathbf{K} = \begin{bmatrix} -f & 0 & p_x \\ 0 & -f & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

Όπου  $f$  είναι η σταθερά της μηχανής και  $p_x, p_y$  είναι οι συντεταγμένες του πρωτεύοντος σημείου (Γραμματικόπουλος et al., 2018).

Στην περίπτωση που η μηχανή έχει βαθμονομηθεί ως προς τα γραμμικά στοιχεία, δηλαδή χωρίς την ακτινική και έκκεντρη διαστρόφη, ο πίνακας  $\mathbf{K}$  έχει την ακόλουθη μορφή (Γραμματικόπουλος et al., 2018):

$$\mathbf{K} = \begin{bmatrix} -f_x & s & p_x \\ 0 & -f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (8)$$

Φυσικά, η χρήση δυο διαφορετικών σταθερών της μηχανής δεν είναι δεσμευτική. Αντ' αυτού μπορεί να χρησιμοποιηθεί η ακόλουθη μορφή του πίνακα  $\mathbf{K}$ .

$$\mathbf{K} = \begin{bmatrix} -f & s & p_x \\ 0 & -af & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

Σε αυτή την περίπτωση το  $a$  (aspect ratio) αφορά την διαφορετική κλίμακα και το  $s$  (skewness) την απόκλιση από την καθετότητα των αξόνων  $x$  και  $y$ .

Οι Γραμματικόπουλος et al. (2018) αναφέρουν πως στις περισσότερες φωτογραμμετρικές διαδικασίες, οι μετρήσεις αναφέρονται σε ένα εξωτερικό σύστημα

αναφοράς και όχι στο σύστημα της εικόνας. Αυτό συναντάται και στους περισσότερους αλγορίθμους Visual SLAM. Επειδή η διαδικασία Visual SLAM συνήθως δεν περιέχει εξωτερική πληροφορία, το σύστημα αναφοράς δεν είναι κάποιο γεωδαιτικό αλλά ένα αυθαίρετο σύστημα. Παρόλα αυτά, το αυθαίρετο σύστημα έχει συντεταγμένες XYZ και είναι εξωτερικό σύστημα του χώρου (world coordinate). Ο πίνακας της μηχανής είναι ένα από τα στοιχεία που χρειάζεται κανείς αν θέλει να συνδέσει το σύστημα του χώρου με το σύστημα της εικόνας.

Σύμφωνα με τους Γραμματικόπουλος et al. (2018) τα συστήματα αυτά συνδέονται με το πολύ τρεις μεταθέσεις και τρεις στροφές και ο μετασχηματισμός αυτός περιγράφεται ως εξής:

$$x = KR[I|C]X = K[R|t]X \quad (10)$$

Όπου το  $x$  είναι το σημείο της εικόνας και το  $X$  είναι το αντίστοιχο σημείο στον χώρο,  $K$  είναι ο πίνακας της μηχανής,  $R$  είναι ο πίνακας στροφής του ενός συστήματος ως προς το άλλο και  $C$  είναι οι τρισδιάστατες συντεταγμένες του προβολικού κέντρου στο σύστημα αναφοράς του χώρου (Γραμματικόπουλος et al., 2018).

Οι παράμετροι της μηχανής ονομάζονται και ενδογενείς παράμετροι (intrinsic), ενώ οι παράμετροι  $C$  και  $R$  ονομάζονται εξωγενείς παράμετροι (extrinsic) (Νάκης, 2019).

### 2.3 Βαθμονόμηση μηχανής

Η βαθμονόμηση της μηχανής είναι η διαδικασία εύρεσης των παραμέτρων του πίνακα  $K$ , καθώς και των μη γραμμικών παραμέτρων που αναφέρθηκαν παραπάνω (ακτινική και έκκεντρη παραμόρφωση). Αποτελεί ένα από τα πιο περίπλοκα προβλήματα στην Φωτογραμμετρία (Νάκης, 2019) και μπορεί να πραγματοποιηθεί με δύο τρόπους:

1. Βαθμονόμηση με πεδίο ελέγχου (π.χ. σκακιέρα) και φωτογράφιση από διάφορες οπτικές γωνίες.
2. Αυτοβαθμονόμηση, όπου δεν είναι αναγκαίο να υπάρχει κάποιο μοτίβο όπως η σκακιέρα, και συνήθως πραγματοποιείται σε διαδικασίες όπως η συνόρθωση δέσμης (bundle adjustment). Δηλαδή, ταυτόχρονα με άλλες παραμέτρους (εξωτερικός προσανατολισμός) μπορούν να υπολογιστούν και οι παράμετροι του εσωτερικού προσανατολισμού.

Όπως έχει αναφερθεί παραπάνω, στην συγκεκριμένη εργασία χρησιμοποιήθηκαν κάποια έτοιμα σύνολα εικόνων (image/video dataset) από το internet αλλά λήφθηκε και ένα σύνολο εικονιστικών δεδομένων στο εργαστήριο. Για τις ανάγκες της εργασίας θεωρήθηκε σημαντικό να πραγματοποιηθεί βαθμονόμηση της μηχανής που χρησιμοποιήθηκε, ώστε να ανακτηθούν οι παράμετροι της μηχανής και να λυθεί το πρόβλημα του Visual SLAM αποφεύγοντας την αυτοβαθμονόμηση.

Για την λήψη των δεδομένων χρησιμοποιήθηκε μηχανή κινητού iPhone 6s με εγγραφή βίντεο στα 1080p HD στα 30 fps. Για την βαθμονόμηση χρησιμοποιήθηκε ένα μοτίβο τύπου σκακιέρας, όπου το κάθε τετράγωνο είχε διαστάσεις 45 χιλιοστά.

Για την πραγματοποίηση της βαθμονόμησης χρησιμοποιήθηκε εφαρμογή βαθμονόμησης “Camera Calibrator App” του MATLAB, η οποία έχει βασιστεί στις εξής δημοσιεύσεις:

- Zhang, Z. (2000)
- Heikkila, J., & O. Silven (1997)
- Scaramuzza et al. (2006)
- Urban et al. (2015)

Για την βαθμονόμηση που πραγματοποιήθηκε, χρησιμοποιήθηκε η διαδικασία του αλγορίθμου που βασίστηκε στα δυο πρώτα άρθρα, καθώς τα δυο τελευταία αφορούν φωτογραφικές μηχανές που δεν ακολουθούν το μοντέλο της κάμερας σημειακής οπής.

Ο συγκεκριμένος αλγόριθμος απαιτεί τουλάχιστον τρεις εικόνες, όμως παρέχει καλύτερα αποτελέσματα αν εισαχθούν 10 με 20 εικόνες. Στηρίζεται στον αυτόματο εντοπισμό χαρακτηριστικών σημείων πάνω στην σκακιέρα (ή εν γένη σε κάποιο δεδομένο μοτίβο) και υπολογίζει τις συντεταγμένες (x, y) των σημείων αυτών στην κάθε εικόνα. Επί της ουσίας ακολουθεί το κλασσικό μοντέλο της κάμερας σημειακής οπής (pinhole camera):

$$w[x \ y \ 1] = [X \ Y \ Z \ 1] \begin{bmatrix} R \\ t \end{bmatrix} K \quad (11)$$

Όπου [X, Y, Z] είναι οι συντεταγμένες στον χώρο, (x, y) είναι οι αντίστοιχες συντεταγμένες στην εικόνα σε pixel, R είναι ο πίνακας στροφής της μηχανής στον χώρο, t είναι η μετάθεση της κάμερας στον χώρο και K είναι ο πίνακας της μηχανής (Mathworks). Η μεταβλητή w αποτελεί έναν αυθαίρετο συντελεστή κλίμακας ομογενών συντεταγμένων (Mathworks). Στον αλγόριθμο που χρησιμοποιείται στο συγκεκριμένη εφαρμογή ο πίνακας της μηχανής έχει την εξής μορφή:

$$K = \begin{bmatrix} -f_x & 0 & 0 \\ s & -f_y & 0 \\ c_x & c_y & 1 \end{bmatrix} \quad (12)$$

Αυτό που παρατηρείται είναι πως η σχέση που χρησιμοποιεί ο αλγόριθμος ( 12 ) είναι ίδια με την Εξίσωση ( 10 ), με την μοναδική διαφορά ότι στην Εξίσωση ( 10 ) δεν υπάρχει ο όρος w, ο πολλαπλασιασμός πινάκων έχει διαφορετική σειρά. Όμως επειδή οι πίνακες είναι αντιδιαμετρικοί, για παράδειγμα τα στοιχεία στον πίνακα K της Εξίσωσης ( 10 ) είναι τα ίδια με του πίνακα K της Εξίσωσης ( 12 ) απλώς σε αντιδιαμετρική θέση. Όταν πολλαπλασιαστούν μεταξύ τους οι όροι και στις δύο εξισώσεις θα δώσουν το ίδιο αποτέλεσμα.

Σύμφωνα με το εγχειρίδιο της Mathworks, ο αλγόριθμος περιέχει τα εξής βήματα:

- Θεωρεί ότι οι παραμορφώσεις του φακού είναι ίσες με μηδέν, υπολογίζοντας τις ενδογενείς και τις εξωγενείς παραμέτρους σε κλειστή μορφή (closed form), όπως αναφέρει ο Zhang (2000). Ο τρόπος υπολογισμού των παραμέτρων του εσωτερικού προσανατολισμού ως πρόβλημα κλειστής μορφής, γλυτώνει από

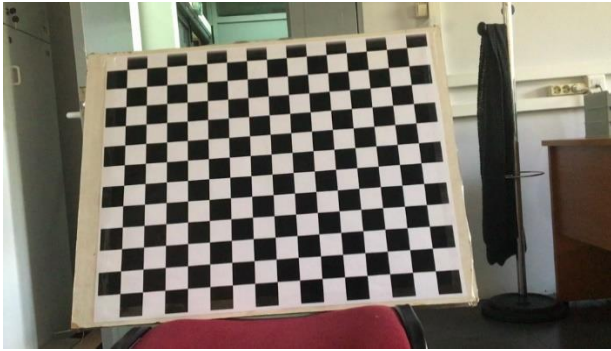
την αναγκαστική γνώση των αρχικών τιμών καθώς δεν είναι επαναληπτική μέθοδος. Η λύση που προτείνει ο Zhang, δεν απαιτεί γνώση των 3D συντεταγμένων. Βρίσκει ομολογίες στις εικόνες όπου εμφανίζεται ένα αντικείμενο με επαναλαμβανόμενο μοτίβο και για κάθε σημείο 2D στην εικόνα και 3D στον χώρο. Χρησιμοποιώντας τις σχέσεις της ομογραφίας μπορεί κανείς να βρει τις δεσμεύσεις του εσωτερικού προσανατολισμού (Zhang, 2000). Με τον υπολογισμό των περιορισμών αυτών φτάνει κανείς σε μια μοναδική λύση, δεδομένου πως χρησιμοποιεί αριθμό εικόνων μεγαλύτερο ή ίσο με 3 (Zhang, 2000). Έχοντας υπολογίσει τις παραμέτρους της κάμερας μπορεί κανείς να λύσει τις εξισώσεις ως προς τις εξωτερικές παραμέτρους και να υπολογίσει τον πίνακα στροφής και μετάθεσης. Σύμφωνα με τον Zhang (2000), επειδή τα δεδομένα δεν είναι απαλειμμένα από θόρυβο, είναι προφανές πως η λύση αυτή δεν είναι πάντοτε η βέλτιστη, όμως μπορεί να χρησιμοποιηθεί ως αρχικές παρατηρήσεις για μετέπειτα συνόρθωση.

- Έπειτα, λαμβάνοντας τα παραπάνω αποτελέσματα ως αρχικές παρατηρήσεις υπολογίζει ταυτόχρονα τις παραμέτρους (γραμμικές και μη) με τον αλγόριθμο ελαχίστων τετραγώνων για μη γραμμικές εξισώσεις Levenberg–Marquardt (Levenverg, 1944 , Marquardt, 1963).

Για τον υπολογισμό των εξωτερικών παραμέτρων, χρειάστηκε να είναι γνωστή η διάσταση του τετραγώνου της σκακιέρας. Η εφαρμογή έδινε την δυνατότητα βαθμονόμησης χωρίς να υπολογιστούν οι μη γραμμικοί παράμετροι του εσωτερικού προσανατολισμού, δηλαδή να υπολογιστούν μόνο οι συντελεστές του πίνακα της μηχανής. Όμως, στην παρούσα εφαρμογή επιλέχθηκε να υπολογιστούν και οι μη γραμμικοί συντελεστές. Το πρόβλημα του υπολογισμού του εσωτερικού προσανατολισμού μπορεί να λυθεί και με αρχικές τιμές, αν είναι γνωστές.

Όπως αναφέρθηκε και παραπάνω, για τον υπολογισμό των παραμέτρων του εσωτερικού προσανατολισμού λήφθηκε ένα βίντεο στο οποίο απεικονίζεται μια σκακιέρα. Από αυτό το βίντεο λήφθηκαν διάφορα καρέ από τις διαφορετικές θέσεις. Παρακάτω παρουσιάζονται μερικές από τις εικόνες που λήφθηκαν.





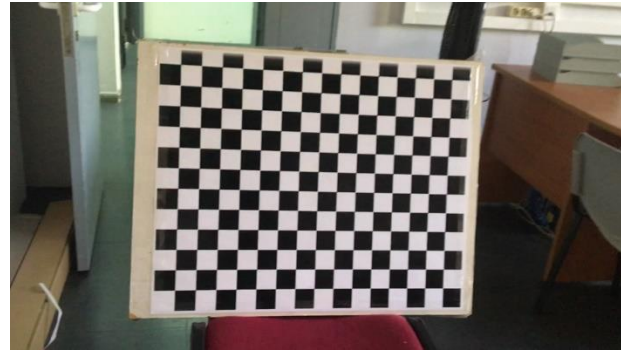
Εικόνα 8: Μέρος του dataset βαθμονόμησης.



Εικόνα 9: Μέρος του dataset βαθμονόμησης.



Εικόνα 10: Μέρος του dataset βαθμονόμησης.



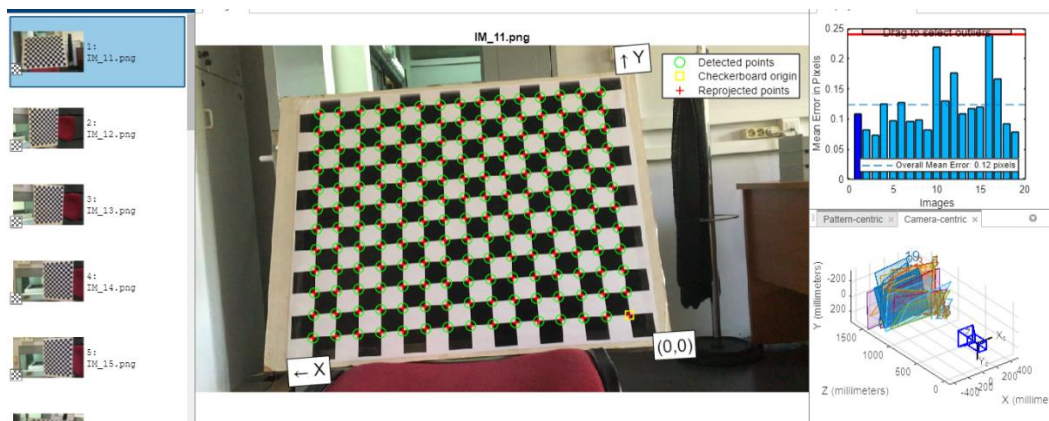
Εικόνα 11: Μέρος του dataset βαθμονόμησης.

Είναι εμφανές πως οι εικόνες είναι από διαφορετικές θέσεις και διαφορετικές οπτικές, όπως απαιτεί ο αλγόριθμος. Συνολικά χρησιμοποιήθηκαν 19 εικόνες. Ο αλγόριθμος εντοπίζει τα χαρακτηριστικά σημεία πάνω στην σκακιέρα και ορίζει ένα σύστημα αναφοράς (Εικόνα 12).



Εικόνα 12: Χαρακτηριστικά σημεία πάνω στην σκακιέρα.

Αφού ελέγχθηκε ότι όλα τα σημεία της σκακιέρας εντοπίστηκαν με επιτυχία, επιλέχθηκε να υπολογιστούν η εφαπτομενική διαστροφή και η απόκλιση από την καθετότητα, και ακολούθησε η βαθμονόμηση. Αφού πραγματοποιήθηκε η διαδικασία, όπως αναφέρθηκε παραπάνω, η εφαρμογή δείχνει τα σημεία επαναπροβολής (reprojection points) τα οποία υπολογίζονται μαζί με τις παραμέτρους του εσωτερικού και εξωτερικού προσανατολισμού (Εικόνα 13).

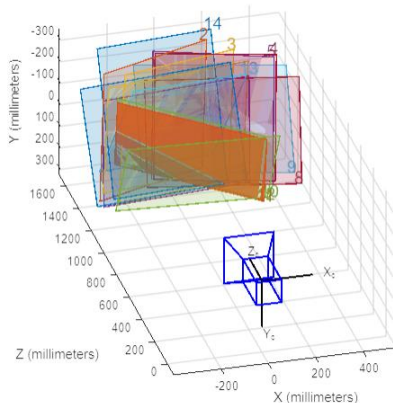


Εικόνα 13: Σφάλματα επαναπροβολής.

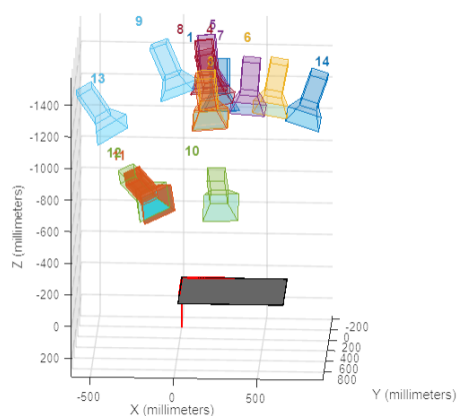
Με κόκκινο σύμβολο φαίνονται τα επαναπροβαλλόμενα σημεία. Η εφαρμογή προβάλλει τα σημεία που έχει εντοπίσει από το σύστημα του χώρου στο σύστημα της εικόνας και με αυτόν τον τρόπο υπολογίζεται το σφάλμα επαναπροβολής. Πιο αναλυτικά, εντοπίζει τα χαρακτηριστικά σημεία (κορυφές) πάνω στην σκακιέρα και βρίσκει τις 2D συντεταγμένες τους πάνω στην εικόνα συνδέοντας έτσι τις εικονοσυντεταγμένες των σημείων με τις συντεταγμένες στους στον χώρο (σκακιέρα). Μέσω της μεθόδου δέσμης, υπολογίζει με συνόρθωση τον βέλτιστο εσωτερικό και εξωτερικό προσανατολισμό για όλες τις εικόνες. Ύστερα, συγκρίνει τις 2D αρχικές εικονοσυντεταγμένες με τις 2D εικονοσυντεταγμένες μετά την επαναπροβολή τους και υπολογίζει την μεταξύ τους απόσταση, η οποία αποτελεί το σφάλμα επαναπροβολής.

Μαζί με τις παραμέτρους του εσωτερικού προσανατολισμού υπολογίζονται οι εξωτερικοί παράμετροι της κάμερας, δηλαδή οι στροφές και οι μεταθέσεις. Αφού υπολογιστούν, μπορούν να οπτικοποιηθούν με δυο τρόπους:

1. Ως προς την κάμερα (Camera Centric): δηλαδή σαν η κάμερα να μην μετακινήθηκε καθόλου αλλά να μετακινούταν το αντικείμενο (Εικόνα 14).
2. Ως προς το αντικείμενο (Pattern Centric): δηλαδή όπως έγινε στην πραγματικότητα, η κίνηση της κάμερας ως προς το αντικείμενο (Εικόνα 15).



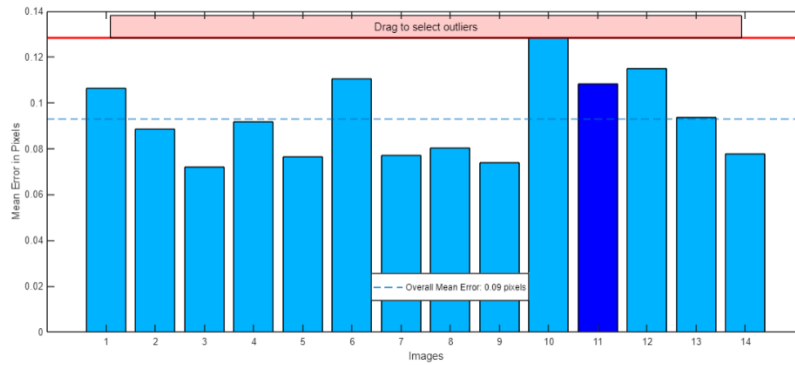
Εικόνα 14: Camera – centric.



Εικόνα 15: Pattern – centric.

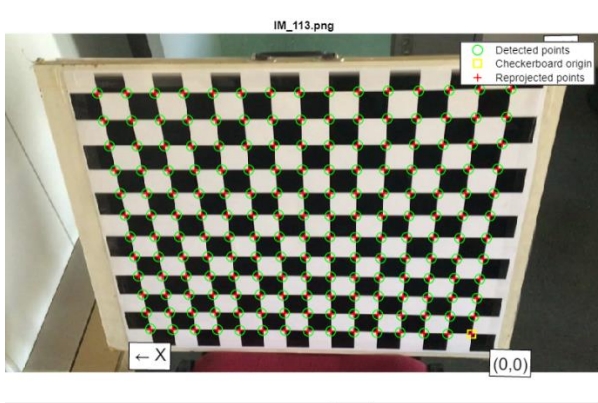
Όπως φαίνεται στην Εικόνα 13 το μέσο σφάλμα υπολογίστηκε ίσο με 0.12 pixel. Παρ'όλα αυτά παρέχεται η δυνατότητα επιλογής των εικόνων που επηρεάζουν περισσότερο στο σφάλμα αυτό ώστε να ξανά πραγματοποιηθεί βαθμονόμηση χωρίς να ληφθούν

υπόψιν. Η τελική βαθμονόμηση πραγματοποιήθηκε με 14 εικόνες και είχε μέσο σφάλμα επαναπροβολής ίσο με 0.09 pixels (Εικόνα 16).



Εικόνα 16: Μέσο σφάλμα επαναπροβολής.

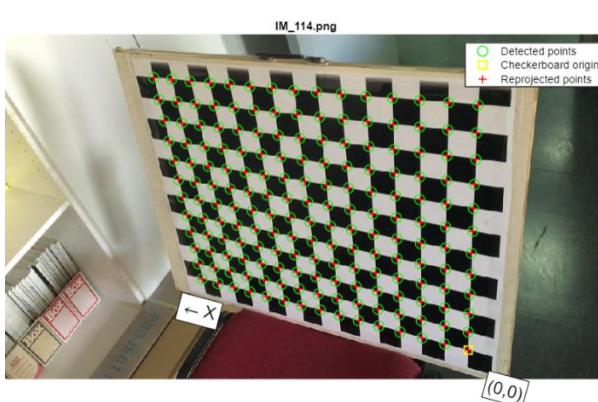
Το πρόγραμμα επιτρέπει την διόρθωση των αρχικών εικόνων από ακτινική και εφαπτομενική διαστρόφη.



Εικόνα 17: Εικόνα πριν την διόρθωση.



Εικόνα 18: Εικόνα μετά την διόρθωση.



Εικόνα 19: Εικόνα πριν την διόρθωση.



Εικόνα 20: Εικόνα μετά την διόρθωση.

Στις συγκεκριμένες εικόνες δεν παρατηρείται ιδιαίτερη διαφορά, και αυτό πιθανώς οφείλεται στο γεγονός ότι δεν υπήρχε έντονη παραμόρφωση του φακού εξ' αρχής.

Οι παράμετροι του εσωτερικού προσανατολισμού που υπολογίστηκαν παρουσιάζονται στον Πίνακα 1.

Πίνακας 1: Παράμετροι εσωτερικού προσανατολισμού.

Εστιακή απόσταση (pixels)	$950.8658 \pm 0.2107$	$957.6735 \pm 0.2111$
Πρωτεύον σημείο (pixels)	$474.3570 \pm 0.2461$	$271.8360 \pm 0.2332$
Απόκλιση από την καθετότητα (skew)	$-1.2339 \pm 0.0294$	
Ακτινική παραμόρφωση	$0.1174 \pm 0.0009$	$-0.2030 \pm 0.0052$
Εφαπτομενική παραμόρφωση	$0.0002 \pm 0.0001$	$-0.0006 \pm 0.0001$

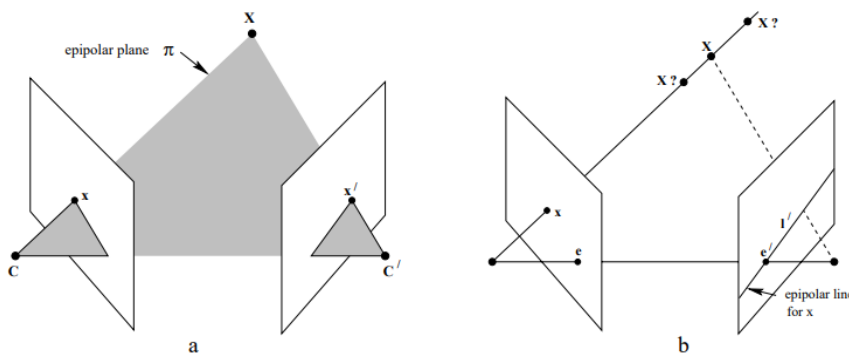
Ο πίνακας της μηχανής υπολογίστηκε ίσως με:

$$K = \begin{bmatrix} 950.8658 & -1.2339 & 474.3570 \\ 0 & 957.6735 & 271.8360 \\ 0 & 0 & 1 \end{bmatrix}$$

#### 2.4 Γεωμετρία λήψης και δημιουργία τρισδιάστατου μοντέλου από δύο εικόνες

Η διαδικασία του Visual SLAM απαιτεί “on the fly” υπολογισμό τρισδιάστατων συντεταγμένων, τόσο των σημείων του χώρου όσο και της θέσης της κάμερας. Στην Φωτογραμμετρία είναι γνωστό πως μπορούν να υπολογιστούν 3D συντεταγμένες από εικόνες του Σχετικού και του Απόλυτου προσανατολισμού, καθώς και της μεθόδου δέσμης.

Ο υπολογισμός των συντεταγμένων μπορεί να πραγματοποιηθεί λαμβάνοντας υπόψη την γεωμετρία της λήψης. Πιο συγκεκριμένα, είναι γνωστό πως κατά την λήψη μιας εικόνας, το προβολικό κέντρο  $O'$  ορίζει με ένα σημείο  $x'$  στην εικόνα την ευθεία  $O'x'$  η οποία τέμνει το επίπεδο της εικόνας στο σημείο  $x'$  (Πέτσα, 2000). Το σημείο  $x'$  είναι η προβολή του σημείου  $X$  του χώρου, στην εικόνα. Αντίστοιχα, έχοντας και μια δεύτερη εικόνα, το προβολικό κέντρο  $O''$  τέμνει την εικόνα στο σημείο  $x''$  μέσω της ευθείας  $O''x''$  (Πέτσα, 2000). Οι προεκτάσεις των δυο ευθειών τέμνουν το σημείο  $X$  στον χώρο (Εικόνα 21).



Εικόνα 21: Επιπολική Γεωμετρία δυο εικόνων.

Το επιπολικό επίπεδο  $\pi$  ορίζεται από τις ευθείες  $O'x'$  στην μια εικόνα και  $O''x''$  στην δεύτερη εικόνα, όπου η επέκτασή τους τέμνει το σημείο  $X$  στον πραγματικό χώρο, και από την βάση του στερεοζεύγους. (Πέτσα, 2000). Βάση του στερεοζεύγους αποτελεί η απόσταση του ενός προβολικού κέντρου από το άλλο, δηλαδή η ευθεία που ενώνει τα δυο προβολικά κέντρα.

Απλοποιώντας τα παραπάνω, το σημείο του χώρου  $X$ , οι προβολές του στις δυο εικόνες ( $x', x''$ ) καθώς και τα προβολικά κέντρα των δυο εικόνων κείνται στο ίδιο επίπεδο  $\pi$  (Hartley & Zisserman, 2004).

Εκμεταλλεόμενοι την επιπολική γεωμετρία, ορίζονται οι επιπολικές ευθείες. Δεδομένου ότι το δεύτερο σημείο στην εικόνα βρίσκεται στο επιπολικό επίπεδο  $\pi$ , άρα βρίσκεται στην ευθεία τομής  $I''$  του επιπολικού επιπέδου με το επίπεδο της δεύτερης εικόνας (Hartley & Zisserman, 2004). Η ευθεία αυτή ονομάζεται επιπολική ευθεία και αντίστοιχα η τομή του επιπολικού επιπέδου με το επίπεδο της πρώτης εικόνας, δημιουργεί την επιπολική ευθεία  $I'$ . Με την χρήση των επιπολικών ευθειών, η αναζήτηση για ομολογίες μεταξύ δυο εικόνων γίνεται πολύ πιο εύκολη, καθώς είναι βέβαιο πως το ομόλογο σημείο θα βρίσκεται στην ίδια γραμμή στην δεύτερη εικόνα.

Άρα, έχοντας γνωστό εσωτερικό προσανατολισμό, και με μόνη πληροφορία τα ομόλογα σημεία μπορεί κανείς να υπολογίσει τρισδιάστατες συντεταγμένες στον χώρο. Εδώ τίθεται το θέμα της γνώσης της βάσης του στερεοζεύγους, δηλαδή την μετάθεση της κάμερας από την πρώτη θέση λήψης στην δεύτερη ώστε να αποκατασταθεί η κλίμακα του μοντέλου (Πέτσα, 2000). Για τον υπολογισμό του σχετικού προσανατολισμού αρχικά εντοπίζονται ομόλογα σημεία στις δυο εικόνες, αφού είναι γνωστό πως οι ακτίνες που ορίζουν τα ομόλογα σημεία στις δυο εικόνες τέμνονται. Χρησιμοποιώντας τα ομόλογα σημεία, προσανατολίζεται η μια εικόνα ως προς την άλλη (Πέτσα, 2000). Δηλαδή, αν διατηρηθεί η πρώτη εικόνα σταθερή, υπολογίζεται η σχετική στροφή της δεύτερης εικόνας ως προς την πρώτη. Με αυτόν τον τρόπο εξασφαλίζονται οι αλληλοτομίες των ομολόγων ακτινών και προκύπτει ένα στερεομοντέλο καλής γεωμετρικής ποιότητας (Πέτσα, 2000).

Ο σχετικός προσανατολισμός, δυο βαθμονομημένων εικόνων, περιγράφεται από πέντε παραμέτρους:

→ 3 παράμετροι στροφής της δεύτερης εικόνας ως προς την πρώτη

→ Σχετικές συνιστώσες  $b_y, b_z$

Φυσικά, σχετικός προσανατολισμός μπορεί να υπολογιστεί για περισσότερες από δυο εικόνες, έχοντας όμως μια σταθερή εικόνα αναφοράς και προσανατολίζοντας μια-μια τις εικόνες ως προς την προηγούμενη, δηλαδή θα πρέπει να πραγματοποιούνται σε ζεύγη εικόνων οι υπολογισμοί.

Μια πολύ σημαντική γεωμετρική συνθήκη ώστε να μπορέσει να υπολογιστεί ο σχετικός προσανατολισμός είναι η συνθήκη συνεπιπεδότητας. Με την συνεπιπεδότητα εξασφαλίζεται πως τα διανύσματα  $X'$  και  $X''$ , τα οποία εκφράζουν τις παρατηρήσεις του σημείου  $X$  του χώρου καθώς και το διάνυσμα της βάσης είναι συνεπίπεδα (στο αντίστοιχο επιπολικό επίπεδο) όταν εκφραστούν και μετατεθούν σε κοινό σύστημα (Γραμματικόπουλος et al, 2018).

Τα διανύσματα  $X'$  και  $X''$  δεν αφορούν τα σημεία  $x'$   $x''$  της εικόνας αλλά τις ακτίνες που προερχόμενες από το κέντρο προβολής, τέμνουν τα σημεία στην εικόνα και επεκτείνονται στο σημείο  $X$  του χώρου (Γραμματικόπουλος et. al, 2018).

$$X' = K_1 \times x'$$

$$X'' = K_2 \times x''$$

Θεωρώντας πως τα διανύσματα  $B$ ,  $X'$  και  $X''$  είναι συνεπίπεδα, και έχοντας την στροφή της δεύτερης εικόνας ως προς την πρώτη ( $R^T$ ):

$$X'^T \times t_B \times R^T \times X'' = 0 \quad (13)$$

#### 2.4.1 Δεσμευμένος επιπολικός πίνακας (E)

Έχοντας ομόλογα σημεία σε δυο εικόνες μπορεί κανείς να ορίσει τον Δεσμευμένο Επιπολικό Πίνακα ώστε να υπολογίσει τον σχετικό προσανατολισμό του στερεοζεύγους:

$$x''^T \times E \times x' = 0 \quad (14)$$

Δεδομένου πως ο εσωτερικός προσανατολισμός είναι γνωστός, δηλαδή οι συντεταγμένες στην εικόνα είναι κανονικοποιημένες, δηλαδή έχουν διορθωθεί ως προς τις παραμέτρους του εσωτερικού προσανατολισμού) (Hartley & Zisserman, 2004), ο Δεσμευμένος επιπολικός πίνακας ορίζεται με τον εξής τρόπο:

$$E = t_B \times R^T = [t_B]_{\times} \times R^T \quad (15)$$

Όπου  $R^T$  ο πίνακας της σχετικής στροφής της δεύτερης εικόνας ως προς την πρώτη και  $[t_B]_{\times}$  είναι ο αντισυμμετρικός πίνακας της μετάθεσης της βάσης του στερεοζεύγους:

$$[t]_{\times} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

Ένας από τους πιο διαδεδομένους αλγορίθμους υπολογισμού του Δεσμευμένου επιπολικού πίνακα είναι ο αλγόριθμος των 5 σημείων (5-Point Algorithm) του Nistér (2004). Για την επίλυση του συγκεκριμένου αλγορίθμου είναι απαραίτητη η γνώση του εσωτερικού προσανατολισμού. Ο πίνακας αυτός έχει διαστάσεις  $3 \times 3$ , όμως πέντε ανεξάρτητες παραμέτρους (Γραμματικόπουλος et. al, 2018), εκείνες του σχετικού προσανατολισμού, ενώ υπάρχει η αβεβαιότητα της κλίμακας. Ο Nistér (2004) αναφέρει πως οι συντεταγμένες των σημείων που εντοπίζονται στις εικόνες βρίσκονται σε ομογενείς συντεταγμένες, το ίδιο και τα σημεία του χώρου. Χρησιμοποιώντας την επιπολική δέσμευση, ο Nistér (2004) αναφέρει πως μπορεί να γραφτεί και ως:

$$x^T \times E = 0 \quad (16)$$

Όπου το διάνυσμα  $q$  και ο πίνακας  $E$  ορίζεται ως:

$$x = [x_1 x'_1 \ x_2 x'_2 \ x_3 x'_3 \ x_1 x'_2 \ x_2 x'_1 \ x_3 x'_2 \ x_1 x'_3 \ x_2 x'_3 \ x_3 x'_1]$$

Με  $x$  συμβολίζονται οι συντεταγμένες του σημείου στην μια εικόνα, ενώ  $q'$  είναι οι συντεταγμένες του ομολόγου σημείου στην δεύτερη εικόνα. Προσθέτοντας και τις υπόλοιπες παρατηρήσεις, δεδομένου ότι λαμβάνονται πέντε σημεία, δημιουργείται ένας πίνακας διαστάσεων  $5 \times 9$ . Ο Nistér (2004) περιγράφει πως με μεθόδους γραμμικής άλγεβρας, όπως Ανάλυση ιδιαζουσών τιμών (Singular Value Decomposition, SVD),

QR αποσύνθεση κ.α. μπορεί να λυθεί το σύστημα των εξισώσεων και εν τέλη να υπολογιστεί ο Δεσμευμένος Επιπολικός Πίνακας (E).

Αφού υπολογιστεί ο δεσμευμένος επιπολικός πίνακας, μπορούν να βρεθούν οι παράμετροι του σχετικού προσανατολισμού R και t. Σύμφωνα με τον Nistér (2004), έστω ότι η ανάλυση ιδιαζουσών τιμών δίνει την λύση του Δεσμευμένου Επιπολικού Πίνακα:

$$E \sim U \text{diag}(1,1,0) V^T$$

Όπου U και V επιλέγονται οι πίνακες για τους οποίους ισχύει:  $\det(U) > 0$  και  $\det(V) > 0$ , τότε  $t \sim t_u = [u_{13} \ u_{23} \ u_{33}]^T$

και ο πίνακας R είναι ίσος  $R_a \equiv U D V^T$  ή  $R_b = U D^T V^T$

Ο D θεωρείται ίσος με:

$$\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ο Nistér (2004) αναφέρει πως μετρά την ανάλυση ιδιαζουσών τιμών, προκύπτει μια αβεβαιότητα ως προς την λύση του δεσμευμένου επιπολικού πίνακα E. Πιο συγκεκριμένα, αναφέρει πως προκύπτουν 4 πιθανές λύσεις:

$$\begin{aligned} P_A &= [R_a | t_u] \\ P_B &= [R_a | -t_u] \\ P_C &= [R_b | t_u] \\ P_D &= [R_b | -t_u] \end{aligned}$$

Από αυτές τις τέσσερις λύσεις, φυσικά μια είναι η αποδεκτή. Η λύση που αναφέρει ο Nistér (2004) είναι μέσω του κριτήριου χειρότητας (cheirality constraint), δηλαδή χρησιμοποιώντας εάν σημείο (ομόλογο στις δυο εικόνες) και η λύση στην οποία το σημείο εμφανίζεται μπροστά και από τις δυο κάμερες είναι η ζητούμενη. Ο Nistér (2004) αναφέρει επιπλέον πως μια από τις εναπομένουσες λύσεις περιγράφει την στροφή της βάσης κατά 180 μοίρες.

Ο Δεσμευμένος επιπολικός πίνακας, όπως έχει ήδη αναφερθεί, μπορεί να υπολογιστεί δεδομένου του εσωτερικού προσανατολισμού. Σε περιπτώσεις αγνώστου εσωτερικού προσανατολισμού είναι αναμενόμενο ο Δεσμευμένος Επιπολικός Πίνακας να μην είναι εφικτό να υπολογιστεί. Παρόλα αυτά η ανάκτηση του σχετικού προσανατολισμού μπορεί να πραγματοποιηθεί μέσω του Επιπολικού Πίνακα.

#### 2.4.2 Επιπολικός Πίνακας (F)

Σύμφωνα με τους Hartley & Zisserman (2004) ο επιπολικός πίνακας αποτελεί την αλγεβρική αναπαράσταση της επιπολικής γεωμετρίας. Παραπάνω έγινε αναφορά στην έννοια της επιπολικής ευθείας, όπου αναφέρθηκε πως η αναζήτηση των ομολόγων σημείων πλέον περιορίζεται στην ίδια γραμμή των εικόνων. Το σημείο της πρώτης εικόνας ορίζει την επιπολική ευθεία στην δεύτερη εικόνα και φυσικά δεσμεύει το ομόλογο του να βρίσκεται πάνω στην ευθεία αυτή. Η σχέση αυτή περιγράφεται με τον εξής τρόπο:

$$x_1 \rightarrow I_2 \quad (17)$$

Οι Γραμματικόπουλος et al. (2018) αναφέρουν πως η συγκριμένη σχέση αποτελεί ετερογραφία και απεικονίζει τα σημεία σε ευθείες.

Σύμφωνα με τον Καλησπεράκη (2010) η εξίσωση που ορίζει την επιπολική ευθεία ενός σημείου  $x$  στην εικόνα ορίζεται από την Ετερογραφία:

$$I'' = F \times x' \quad (18)$$

Ο πίνακας  $F$  είναι γνωστός και ως επιπολικός πίνακας και έχει 7 βαθμούς ελευθερίας.

Η δέσμευση πως το ομόλογο σημείο στην δεύτερη εικόνα θα βρίσκεται πάνω στην επιπολική ευθεία υπολογίζεται ως:

$$x'' \times F \times x' = 0$$

Με μια πρώτη ματιά φαίνεται πως η Εξίσωση (18) μοιάζει αρκετά με την Εξίσωση (14) του Δεσμευμένου Επιπολικού Πίνακα. Σύμφωνα με τον Καλησπεράκη (2010) οι δυο σχέσεις διαφέρουν καθώς η μια αναφέρεται σε μη κανονικοποιημένες συντεταγμένες, ενώ η άλλη αναφέρεται σε κανονικοποιημένες συντεταγμένες της εικόνας. Αναφέρει ακόμα, πως οι δυο αυτές σχέσεις εκφράζουν στην πραγματικότητα διαφορετικές δεσμεύσεις, πράγμα που σημαίνει ότι η διαφορά τους δεν είναι μόνο ως προς τις συντεταγμένες της εικόνας.

Ο επιπολικός πίνακας μπορεί να υπολογιστεί μόνο από τα σημεία στην εικόνα, χωρίς να χρειάζεται ο εσωτερικός προσανατολισμός. Αυτό οδηγεί σε απαίτηση περισσότερων ελάχιστων παρατηρήσεων απ' ότι ο δεσμευμένος επιπολικός πίνακας, καθώς σε αυτή την περίπτωση οι άγνωστες παράμετροι είναι περισσότεροι (παράμετροι εσωτερικού προσανατολισμού και σχετικού προσανατολισμού).

Εφόσον ο Επιπολικός πίνακας έχει 7 βαθμούς ελευθερίας, αυτό σημαίνει πως ο ελάχιστος αριθμός των ομολόγων σημείων που απαιτούνται για τον υπολογισμό του είναι 7. Οι πιο διαδεδομένοι αλγόριθμοι για τον υπολογισμό του πίνακα  $F$  είναι:

1. Ο αλγόριθμος των 7 σημείων, όπου απαιτούνται οι ελάχιστες ομολογίες, δηλαδή 7 ομόλογα σημεία και
2. Ο αλγόριθμος των 8 σημείων όπου λαμβάνονται 8 ομόλογα σημεία.

Ο επιπολικός πίνακας έχει την μορφή:

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

Γράφοντας τον πίνακα  $F$  σε διανυσματική μορφή

$$f = [f_{11} \ f_{12} \ f_{13} \ f_{21} \ f_{22} \ f_{23} \ f_{31} \ f_{32} \ f_{33}]$$

Σύμφωνα με τους Γραμματικόπουλος et al. (2019), οι περισσότεροι αλγόριθμοι υπολογισμού του πίνακα  $F$  στηρίζονται στο γραμμικό σύστημα εξισώσεων που αναλύθηκε παραπάνω.



Μια από αυτές τις λύσεις αποτελεί και ο αλγόριθμος των 8 σημείων που αναφέρθηκε παραπάνω. Έχοντας 8 ομόλογα σημεία, ο πίνακας  $A$  έχει διαστάσεις  $8 \times 9$ . Υπό ιδανικές συνθήκες, δηλαδή μη ύπαρξη σφαλμάτων στις μετρήσεις, ο πίνακας  $A$  θα ήταν τάξης 8, όμως λόγω των σφαλμάτων στην πραγματικότητα είναι τάξης 9 (Γραμματικόπουλος et al., 2018).

Το πρόβλημα βελτιστοποίησης που προκύπτει είναι το εξής:

$$f = \operatorname{argmin} \|Af\|_2; f \neq 0_{\mathbb{R}^9} \quad (19)$$

Οπού ουσιαστικά ελαχιστοποιείται η ποσότητα  $\|Af\|$ . Για την λύση του προβλήματος πραγματοποιείται με ανάλυση ιδιζουσών τιμών και έπειτα επιβάλλεται η δέσμευση της μηδενικής ορίζουσας. Η δέσμευση αυτή επιβάλλεται επειδή από την ανάλυση ιδιζουσών τιμών, το αποτέλεσμα μπορεί να μην ικανοποιεί την δέσμευση της μη αναστρεψιμότητας του επιπολικού πίνακα.

Εφαρμόζοντας τεχνικές βελτιστοποίησης όπως η χρήση της Frobenius νόρμας πίνακα για την εύρεση του κοντινότερου  $3 \times 3$  πίνακα και εφαρμόζοντας ανάλυση ιδιζουσών τιμών μπορεί να προκύψει η λύση του Επιπολικού Πίνακα.

## 2.5 Εντοπισμός και συνταύτιση χαρακτηριστικών σημείων

Στο προηγούμενο εδάφιο μελετήθηκε η γεωμετρία που ορίζεται από ένα ζεύγος εικόνων και αναλύθηκε ο υπολογισμός του σχετικού προσανατολισμού των εικόνων μέσω ομόλογων σημείων.

Τα ομόλογα σημεία στις εικόνες μπορούν είτε να μετρηθούν χειροκίνητα είτε αυτόματα. Στις περισσότερες φωτογραμμετρικές εφαρμογές, η μέτρηση ομόλογων σημείων στις εικόνες είναι αρκετά χρονοβόρα. Γι' αυτόν τον λόγο υπάρχουν στην βιβλιογραφία της Φωτογραμμετρίας και Όρασης Υπολογιστών αρκετές μέθοδοι αυτομάτου υπολογισμού ομολόγων σημείων μεταξύ εικόνων.

Στις μεθόδους Visual SLAM χρησιμοποιούνται αυτόματοι αλγόριθμοι εντοπισμού και συνταύτισης χαρακτηριστικών σημείων στις εικόνες. Οι αλγόριθμοι εντοπισμού χαρακτηριστικών σημείων πρακτικά αφορούν σημειακούς τελεστές, όπου αρκετοί από αυτούς λειτουργούν και ως σημειακοί περιγράφεις.

Αρχικά, οι αλγόριθμοι εντοπισμού και συνταύτισης σημείων ήταν μόνο σημειακοί τελεστές. Δηλαδή, εντόπιζαν τα χαρακτηριστικά σημεία και έπειτα εφάρμοζαν διάφορες τεχνικές ώστε να ελέγξουν κατά πόσο τα σημεία στις δυο εικόνες ταιριάζουν.

Για παράδειγμα, ο σημειακός τελεστής Harris (Harris & Stephens, 1988) εντοπίζει γωνίες στις εικόνες και τις θεωρεί ως χαρακτηριστικά σημεία. Γωνία αποτελεί η μεταβολή της έντασης σε όλες τις διευθύνσεις. Ο τελεστής Harris ακολουθεί την λογική του κινούμενου παραθύρου, δηλαδή ανοίγει ένα παράθυρο στην εικόνα και παίρνει τις τιμές της έντασης γύρω από το κεντρικό pixel, ενώ χρησιμοποιεί ένα κριτήριο γωνιακότητας.

Στην παρούσα εργασία μελετήθηκαν οι δυνατότητες του Visual SLAM αλγορίθμου με τους εξής αλγορίθμους εντοπισμού και συνταύτισης χαρακτηριστικών:

- SIFT
- SURF
- ORB

### 2.5.1 Ο αλγόριθμος SIFT

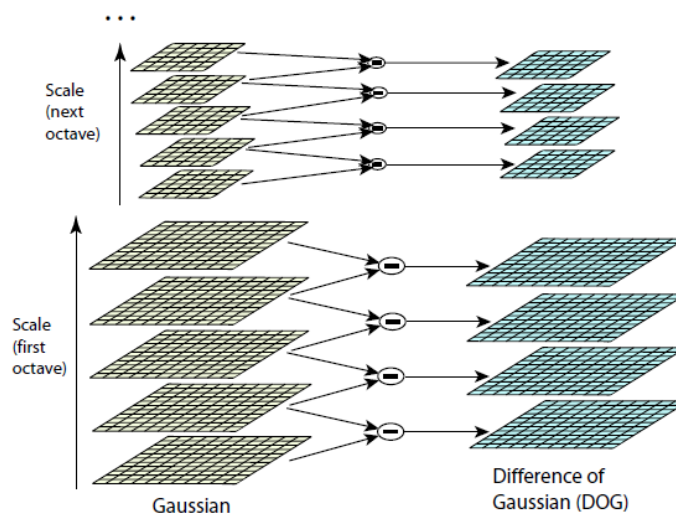
Ο SIFT δημοσιεύθηκε από τον Lowe (1999) και αποτελεί σημειακό τελεστή εντοπισμού σημείων αλλά και περιγραφέα. Βασικό χαρακτηριστικό του SIFT είναι πως τα σημεία που εντοπίζει είναι αμετάβλητα ως προς την στροφή, την κλίμακα και την μετατόπιση της εικόνας. Είναι ακόμα μερικώς αμετάβλητα στην αλλαγή του φωτισμού και επίσης στις αφινικές παραμόρφωσεις (Lowe, 1999).

Οι εικόνες φιλτράρονται με διαδοχικό τρόπο (Lowe, 1999) και δημιουργούνται πυραμίδες της αρχικής εικόνας μέσω της συνάρτησης Gauss, και σε αυτές τις πυραμίδες πραγματοποιούνται οι αναζητήσεις των χαρακτηριστικών σημείων. Έτσι δημιουργείται ο χώρος των κλιμάκων, όπου βρίσκονται οι μέγιστες και οι ελάχιστες τιμές της συνάρτησης Gauss, και εκεί αναδεικνύεται αν υπάρχει χαρακτηριστικό σημείο.

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \times e^{-\frac{x^2}{2\sigma^2}} \quad (20)$$

Η συνάρτηση Gauss της Εξίσωσης (20) αφορά μόνο την μια διάσταση, εφόσον είναι διαχωρίσιμη κατά x και y και μπορεί να εφαρμοστεί κατ' αυτόν τον τρόπο (Lowe, 1999).

Ο SIFT λαμβάνει το πρώτο επίπεδο στην πυραμίδα Gauss και παίρνει τις οκτάβες (για διαφορετικά  $\sigma$ ), αφαιρεί την πρώτη οκτάβα από την δεύτερη και δημιουργεί μια εικόνα DoG (Difference of Gaussian - Διαφορά της συνάρτησης Gauss) όπως φαίνεται στη Εξίσωση (). Έπειτα, αφαιρεί την δεύτερη από την τρίτη οκτάβα και δημιουργεί μια νέα εικόνα DoG και ούτω καθεξής (Lowe, 2004). Ακολουθώντας, προχωράει στην επόμενη ανάλυση και επανυπολογίζει τις διαφορές της συνάρτησης Gauss όπως περιεγράφηκαν μόλις (Lowe, 2004).



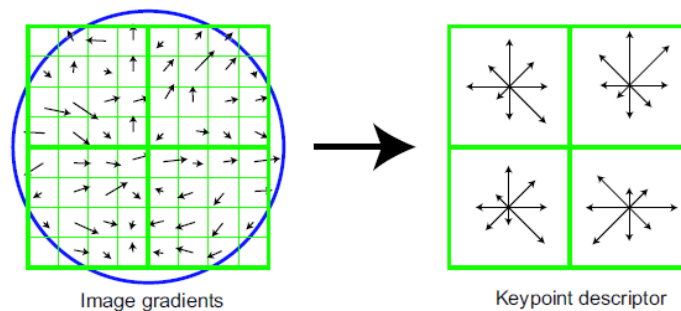
Εικόνα 22: Δημιουργία των διαφορετικών επιπέδων κλίμακας (Πηγή: (Lowe, 2004)).

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \times I(x, y) \quad (21)$$

Ο SIFT εφαρμόζει διάφορες τεχνικές βελτιστοποίησης των σημείων που έχει εντοπίζει, απορρίπτοντας τυχόν λανθασμένα σημεία. Αυτό πραγματοποιείται με μεθόδους όπως προσαρμόζοντας μια 3D τετραγωνική συνάρτηση στην κορυφή του DoG, υπολογισμός του Εσσιανού πίνακα (Hessian Matrix). (Lowe, 2004).

Μέσω του περιγραφέα μπορεί να είναι γνωστός ο προσανατολισμός του κάθε χαρακτηριστικού σημείου ώστε να είναι πλέον ανεπηρέαστα από τυχόν στροφή της εικόνας (Lowe, 2004). Αρχικά θα πρέπει να υπολογιστεί η κυρία κλίση στην περιοχή γύρω από το χαρακτηριστικό σημείο (Lowe, 2004). Αφού υπολογιστεί η κύρια κλίση, δημιουργείται ένα ιστόγραμμα κλίσεων (orientation histogram) όπου οι κορυφές του ιστογράμματος αντιπροσωπεύουν τις κύριες κλίσεις (Lowe, 2004). Σύμφωνα με τον Lowe (2004) το ιστόγραμμα των κλίσεων έχει 36 πεδία όπου καλύπτεται το εύρος των 365 μοιρών των προσανατολισμών (Lowe, 2004).

Αφού υπολογιστούν τα ιστογράμματα κλίσεων, μπορεί να βρεθεί ο περιγραφέας. Όπως έχει ήδη αναφερθεί, ο περιγράφες πρέπει να είναι ανεπηρέαστος από τις στροφές. Για να πέτυχει αυτό ο Lowe (2004), χρησιμοποιεί ένα παράθυρο διαστάσεων 16×16 ή 32×32 περί το σημείο και στρέφει ως προς τον προσανατολισμό των χαρακτηριστικών τις συντεταγμένες του περιγραφέα και τους προσανατολισμού της παραγωγού. Έπειτα, οι διαστάσεις των παραθύρων μικραίνουν για παράδειγμα παίρνουν διαστάσεις 4×4 ή 8×8, και σε κάθε υποπεριοχή υπολογίζεται το ιστόγραμμα των διευθύνσεων των κλίσεων και αποθηκεύεται σε πεδία (Lowe, 2004). Αποθηκεύονται ακόμα πληροφορίες όπως η διανομή των διευθύνσεων των κλίσεων, η τοπική υφή κ.α. Στο τέλος όλη αυτή η πληροφορία συνδυάζεται και δημιουργείται το διάνυσμα του περιγραφέα (Lowe, 2004).



Εικόνα 23: Δημιουργία του περιγραφέα (Πηγή: (Lowe, 2004)).

Από τα πειράματα που πραγματοποίησε ο Lowe (2004) αποδείχθηκε πως τα βέλτιστα αποτελέσματα λήφθηκαν με την χρήση παραθύρου 44 για τον υπολογισμό των ιστογραμμάτων με 8 τιμές κλίσης, δηλαδή έναν περιγραφέα διαστάσεων 44×8=128.

### 2.5.2 Ο αλγόριθμος SURF

Μια ακόμα μέθοδος εντοπισμού χαρακτηριστικών σημείων αλλά και δημιουργίας των περιγραφέων τους είναι ο SURF (Speeded Up Robust Features). Ο SURF χρησιμοποιεί τον Εσσιανό πίνακα (Hessian Matrix) για τον εντοπισμό σημείων (Bay Tuytelaars & Van Gool., 2006).

Όπως και ο SIFT έτσι και ο SURF αναζητεί σημεία στον χώρο των κλίμακων, ώστε να είναι ανεπηρέαστος από αλλαγές κλίμακας στην εικόνα, όμως σε αντίθεση με τον SIFT δημιουργεί εικόνες DoG, αλλά εφαρμόζει τετραγωνικά φίλτρα (box filter), διότι είναι πιο γρήγορο να υπολογιστούν (Bay et al., 2006).

Πιο αναλυτικά, αρχικά υπολογίζει μια εικόνα ολοκληρώματος σε μια θέση  $x = (x, y)^T$  (Bay Tuytelaars & Van Gool, 2006) με τον εξής τρόπο:

$$I_{\Sigma}(X) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (22)$$

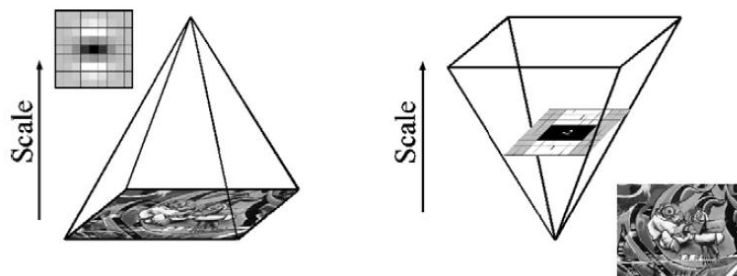
Η εικόνα αυτή αναπαριστά το άθροισμα των πίξελ στην αρχική εικόνα  $I$  μέσα σε μια τετραγωνική περιοχή (Bay Tuytelaars & Van Gool., 2006).

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (23)$$

Όπου το  $x$  είναι ένα σημείο στην εικόνα και το  $\sigma$  είναι η κλίμακα του. Το  $L_{xx}$  είναι η συνέλιξη της παραγώγου δεύτερης τάξης Gauss με την εικόνα στο σημείο  $x$  (Bay Tuytelaars & Van Gool., 2006).

Η προσέγγιση του Εσσιανού πίνακα μπορεί να γίνει αρκετά γρηγορά με την χρήση των φίλτρων μέσου όρου (box filters). Πιο συγκεκριμένα, στα σημεία όπου η ορίζουσα του Εσσιανού Πίνακα είναι μέγιστη θεωρούνται και ως σημεία ενδιαφέροντος (Bay Tuytelaars & Van Gool, 2006). Αυτά τα σημεία αποθηκεύονται σε έναν χάρτη τον οποίο ο Bay Tuytelaars & Van Gool (2006) αναφέρει ως “blob response map”, αφού θεωρείται πως σε εκείνα τα σημεία υπάρχει μεγάλη πιθανότητα εμφάνισης χαρακτηριστικού σημείου (Bay Tuytelaars & Van Gool, 2006).

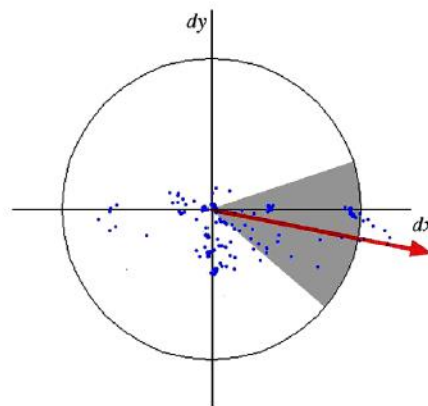
Χάρη στο φίλτρο μέσου όρου, δεν είναι απαραίτητο να ξανά φιλτράρεται η εικόνα, όπως παρουσιάστηκε στον αλγόριθμο SIFT, αλλά μπορεί να εφαρμοστεί φίλτρο μεγαλύτερων διαστάσεων στην αρχική εικόνα (Bay Tuytelaars & Van Gool, 2006). Αρά, αντί να μειώνονται οι διαστάσεις της εικόνας, αυξάνονται οι διαστάσεις του φίλτρου ώστε να δημιουργηθούν τα επίπεδα της κλίμακας (Bay Tuytelaars & Van Gool, 2006).



Εικόνα 24: Δημιουργία επιπέδων κλίμακας στον SURF (Πηγή: (Bay et al., 2006)).

Παρόμοια με τον SIFT, έτσι και στον SURF ο περιγραφέα έχει πληροφορία για τις τιμές της έντασης γύρω από το χαρακτηριστικό σημείο (Bay Tuytelaars & Van Gool, 2006).

Ο τρόπος υπολογισμού του περιγραφέα σε αυτή την περίπτωση είναι μέσω των κυματιδίων Haar πρώτου βαθμού στις διαστάσεις  $x$  και  $y$ , ενώ το διάνυσμα του είναι μόνο 64 διαστάσεων (Bay Tuytelaars & Van Gool, 2006). Υπολογίζει τις αποκρίσεις των κυματιδίων, οι οποίες εμφανίζονται ως σημεία όπου η ισχυρότερη οριζόντια απόκριση είναι στον άξονα των  $x$  ενώ αντίστοιχα η ισχυρότερη κατακόρυφη στον άξονα των  $y$  (Bay Tuytelaars & Van Gool, 2006). Η ισχυρότερη απόκριση υπολογίζεται μέσω ενός κινουμένου παραθύρου και ορίζεται ως το άθροισμα όλων των αποκρίσεων (Bay Tuytelaars & Van Gool, 2006).



Εικόνα 25: Υπολογισμός περιγραφέων στον αλγόριθμο SURF (Πηγή: (Bay et al., 2006)).

Ο προσανατολισμός του σημείου ενδιαφέροντος ορίζεται από το ξεχωριστό άθροισμα των  $\chi$  και των  $\psi$ , όπου δημιουργούν ένα τοπικό διάνυσμα προσανατολισμού (Bay Tuytelaars & Van Gool, 2006). Το μέγιστο διάνυσμα είναι και εκείνο που ορίζει τον προσανατολισμό αυτό.

Όπως και στον SIFT έτσι και στον SURF, η δημιουργία του περιγραφέα γίνεται αρχικά σε μικρές περιοχές της εικόνας και έπειτα ενοποιούνται και αποθηκεύονται στο διάνυσμα του περιγραφέα.

### 2.5.3 Ο αλγόριθμος ORB

Ένας ακόμα αρκετά γνωστός αλγόριθμος εντοπισμού χαρακτηριστικών σημείων και δημιουργία των περιγραφέων τους για την συνταύτιση είναι ο ORB (Oriented FAST and Rotated BRIEF). Ο αλγόριθμος αυτός προτάθηκε πρώτη φορά από τους Rublee et al. (2011).

Η κύρια ιδέα του συγκεκριμένου αλγορίθμου ήταν να λάβουν δυο γνωστούς αλγορίθμους, ένας για εντοπισμό σημείων και ένας για δημιουργία περιγραφέων - συνταύτιση, να στηριχτούν στα δυνατά σημεία του κάθε αλγορίθμου και να δημιουργήσουν έναν ενιαίο αλγόριθμο.

Ο αλγόριθμος εντοπισμού σημείων FAST δεν περιέχει κάποια πληροφορία για τον προσανατολισμό των σημείων που εντοπίζει. Διατηρώντας τον αλγόριθμο ως έχει, ο Rublee et al. (2011) πρόσθεσαν την αναγνώριση του προσανατολισμού.

Αρχικά, εντοπίζονται τα σημεία μέσω του αλγορίθμου FAST, δηλαδή μέσω του κατώφλιού της τιμής της έντασης μεταξύ του κεντρικού πίξελ και εκείνων που απέχουν κυκλικά γύρω από το κέντρο (Rublee et al. , 2011). Αρχικά το κατώφλι τίθεται με τέτοιο τρόπο ώστε να ληφθούν αρκετά σημεία τα οποία έπειτα φιλτράρονται μέσω του κριτήριου γωνιακότητας του τελεστή Harris. Για να είναι ανεπηρέαστα από κλίμακα τα σημεία που εντοπίζονται, δημιουργείται πυραμίδα της εικόνας όπου εντοπίζονται τα σημεία, όπως περιεγράφηκε παραπάνω, σε κάθε επίπεδο (Rublee et al. , 2011).

Ο Rublee et al. (2011) αναφέρει πως η μέθοδος που χρησιμοποίησαν για τον υπολογισμό του προσανατολισμού είναι η ένταση του κεντροειδούς (intensity centroid). Θεωρώντας πως η ένταση μιας γωνίας απέχει από το κέντρο της και μέσω του διανύσματος αυτού μπορεί να αποδοθεί ένας προσανατολισμός (Rublee et al. , 2011).

Για την βέλτιστη ανθεκτικότητα στην στροφή, υπολογίζονται μέσα σε μια κυκλική περιοχή με συγκεκριμένη ακτίνα  $r$ , η οποία είναι ίση με το μέγεθος του patch (Rublee et al. , 2011).

Ο περιγραφέας BRIEF, είναι ένα διάνυσμα το οποίο περιέχει δυαδική πληροφορία σχετικά με την ένταση της εικόνας σε υποπεριοχές (patches) έπειτα από ένα σετ δοκιμών (Rublee et al. , 2011).

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x_i, y_i) \quad (24)$$

Όπου για μια δεδομένη εξομαλυμένη υποπεριοχή της εικόνας  $p$  και δεδομένη τιμή έντασης σε ένα σημείο  $x$   $p(x)$  ένα δυαδικό τεστ είναι της μορφής:

$$\tau(p; x, y, ) := \begin{cases} 1 : p(x) < p(y) \\ 0 : p(x) \geq p(y) \end{cases} \quad (25)$$

Φυσικά, ο BRIEF επηρεάζεται από την στροφή. Γι' αυτόν τον λόγο χρησιμοποιώντας την βασική ιδέα του BRIEF απέδωσαν προσανατολισμό, δημιουργώντας τον steered BRIEF

Συγκρίνοντας τον BRIEF με τον steered BRIEF, αποδείχθηκε πως ο steered BRIEF δεν εντοπίζει μικροδιαφορές (Rublee et al. , 2011). Για να βελτιώσουν τα αποτελέσματα δημιούργησαν μια μέθοδο εκμάθησης ώστε να επιλέγεται ένα καλό υποσύνολο των δυαδικών τεστ. Ο αλγόριθμος που δημιούργησαν ο Rublee et al. (2011) αναζητεί από όλα τα πιθανά δυαδικά τεστ, εκείνα όπου εμφανίζεται υψηλή απόκλιση και είναι ασυσχέτιστα. Ο αλγόριθμος τους έχει ως σκοπό την δημιουργία ενός διανύσματος αποθήκευσης των αποτελεσμάτων, που να έχει διαστάσεις 256 (Rublee et al. , 2011). Το διάνυσμα αυτό προκύπτει έπειτα από αναζήτηση των τεστ που έχουν την επιθυμητή συσχέτιση (μεγαλύτερη από ένα κατώφλι) έναντι των υπολοίπων που έχουν δημιουργηθεί για όλα τα patches εκπαίδευσης (Rublee et al. , 2011). Η διαδικασία αυτή είναι επαναληπτική και τερματίζεται μόνο όταν έχει δημιουργηθεί το διάνυσμα με τα καλύτερα τεστ και είναι διαστάσεων 256. Αν δεν προκύψουν οι αναγκαίες διαστάσεις, αυξάνεται το κατώφλι και επαναλαμβάνεται η διαδικασία (Rublee et al. , 2011).

Οι αλγόριθμοι εντοπισμού και συνταύτισης σημείων στην βιβλιογραφία είναι πάρα πολλοί, και οι περισσότεροι από αυτούς επιφέρουν αρκετά πετυχημένα αποτελέσματα.

## 2.6 Βελτίωση αποτελεσμάτων

### 2.6.1 Αλγόριθμος RANSAC

Από τον εντοπισμό και την συνταύτιση χαρακτηριστικών σημείων με αυτόματο τρόπο λαμβάνεται ένας μεγάλος αριθμός ομολόγων. Όπως είναι αναμενόμενο η χειροκίνητη αξιολόγηση τους, για τυχόν αστοχίες στην συνταύτιση, είναι αρκετά χρονοβόρα και δύσκολη.

Ο αλγόριθμος Random Sample Consensus (RANSAC) (Fischler and Bolles, 1981), μπορεί να χρησιμοποιηθεί για τον υπολογισμό του μαθηματικού μοντέλου που έχει σημαντικό αριθμό χονδροειδών σφαλμάτων (Fischler and Bolles, 1981). Στην περίπτωση των αυτομάτων αλγορίθμων συνταύτισης, χρησιμοποιώντας τον αλγόριθμο RANSAC απορρίπτονται τα χονδροειδή σφάλματα στην συνταύτιση και τα σημεία που παραμένουν είναι πιο πιθανό να αποτελούν ομόλογα σημεία.

Ο συγκεκριμένος αλγόριθμος είναι επαναληπτικός και σε αντίθεση με άλλες μεθόδους, που αρχικοποιούνται με ένα μεγάλο σετ δεδομένων και έπειτα απορρίπτονται τιμές, μπορεί να αρχικοποιηθεί με πολύ λίγες παραμέτρους (Fischler and Bolles, 1981).

Επιλεγεί τυχαία υποσύνολα των δεδομένων και προσαρμόζει το μοντέλο σε κάθε ένα από αυτά όπου υπολογίζει τα inliers, τα οποία ορίζονται μέσω ενός κατωφλιού (Fischler and Bolles, 1981). Ύστερα από ένα συγκεκριμένο αριθμό επαναλήψεων και συνδυασμών, μπορεί να βρεθεί το βέλτιστο μοντέλο, το οποίο φυσικά είναι εκείνο με τα περισσότερα inliers (Fischler and Bolles, 1981).

Στην περίπτωση που τα σημεία δεν είναι πολλά, μπορεί να υπολογιστεί αρκετά γρηγορά η βελτιστη ευθεία. Όμως, δεν ισχύει σε κάθε περίπτωση αυτό, αφού ενδέχεται τα δεδομένα να είναι πολλά και το υπολογιστικό κόστος, για να πραγματοποιηθούν όλοι οι συνδυασμοί ελάχιστων δεδομένων, να είναι απαγορευτικό.

Γι' αυτόν τον λόγο μπορεί να βρεθεί ο μέγιστος αριθμός δοκίμων ( $N$ ) με τον εξής τρόπο (Scaramuzza, 2011):

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \varepsilon)^s)} \quad (26)$$

Οπού  $s$  ο ελάχιστος αριθμός σημείων,  $\varepsilon$  είναι το ποσοστό των outliers στα σημεία και  $p$  είναι η επιθυμητή πιθανότητα να συμπεριλαμβάνονται μη χονδροειδή σφάλματα (Scaramuzza, 2011).

Ο συγκεκριμένος αλγόριθμος προσαρμόζει δεδομένο μαθηματικό μοντέλο όσο το δυνατόν καλύτερα, όμως δεν είναι βέβαιο πως θα λειτουργεί σωστά υπό κάθε συνθήκη. Ενδέχεται να υπάρξουν περιπτώσεις όπου ο θόρυβος των δεδομένων να είναι αρκετά έντονος και να υπάρξει ως συνέπεια αστοχία του αλγορίθμου.

Ο αλγόριθμος RANSAC μπορεί να συμβάλλει στον υπολογισμό του επιπολικού πίνακα, εφόσον είναι αναγκαίος ο εντοπισμός ομολόγων σημείων ώστε να υπολογιστεί ο επιπολικός πίνακας.

### 2.6.2 Μέθοδος δέσμης

Στην πραγματικότητα, και ειδικά σε εφαρμογές Visual SLAM, ο εντοπισμός σημείων και ο υπολογισμός της θέσης και του προσανατολισμού της κάμερας για το δεδομένο καρέ, πραγματοποιείται σε περισσότερα από δυο καρέ. Παραπάνω αναφέρθηκε η εύρεση του σχετικού προσανατολισμού σε δυο εικόνες ώστε, αποκαθιστώντας την σχετική γεωμετρία μεταξύ τους, να υπολογιστούν 3D συντεταγμένες στον χώρο.

Είναι αναμενόμενο πως στην πραγματικότητα ο υπολογισμός και η επίλυση του μαθηματικού μοντέλου του σχετικού προσανατολισμού για δυο καρέ μπορεί να μην επιφέρει την βέλτιστη λύση, αφού προστίθενται νέες εικόνες στις οποίες εμφανίζεται η ίδια περιοχή. Τα σημεία που δημιουργούν το τρισδιάστατο μοντέλο εμφανίζονται σε παραπάνω από δυο εικόνες, παρόλα αυτά με την ξεχωριστή επίλυση με τον σχετικό προσανατολισμό, τα σημεία αυτά υπολογίζονται ξεχωριστά κάθε φορά.

Η μέθοδος δέσμης υπολογίζει τον σχετικό προσανατολισμό για πολλές εικόνες ταυτόχρονα, ενώ πραγματοποιεί και συνόρθωση. Αυτό πραγματοποιείται εφόσον οι εικόνες εμφανίζουν αρκετά ομόλογα. Οι παράμετροι (θέση της κάμερας και/ή η βαθμονόμηση) υπολογίζονται ταυτόχρονα μέσω ελαχιστοποίησης της συνάρτησης του κόστους (Triggs et al., 2000). Δηλαδή, η μέθοδος δέσμης ελαχιστοποιεί το σφάλμα επαναπροβολής (reprojection error). Το σφάλμα αυτό, όπως αναφέρθηκε και παραπάνω, είναι η διαφορά μεταξύ των θέσεων των σημείων που έχουν παρατηρηθεί στην εικόνα και εκείνων που έχουν προβλεφθεί προβάλλοντας τα 3D σημεία στην εικόνα χρησιμοποιώντας τις υπολογιζόμενες παραμέτρους του εσωτερικού και εξωτερικού προσανατολισμού (Lourakis & Argyros, 2009).

Το σφάλμα επαναπροβολής δίνεται από την σχέση:

$$||Project(P_i, X_k) - u_{i,k}|| \quad (27)$$

Οπού το  $P_i$  αναπαριστά τις παραμέτρους της κάμερας (θέση και προσανατολισμός, εστιακή απόσταση κ.α.) για την κάμερα  $i$ , δηλαδή την κάμερα της πρώτης εικόνας,  $X_k$  είναι η 3D συντεταγμένη του σημείου  $k$ .  $Project(P_i, X_k)$  είναι η προβολή του 3D σημείου  $X_k$  στο επίπεδο της εικόνας  $i$  χρησιμοποιώντας τις παραμέτρους του εσωτερικού προσανατολισμού  $P_i$ . Τέλος το  $u_{i,k}$  αντιπροσωπεύει την παρατηρούμενη 2D θέση του χαρακτηριστικού σημείου στην εικόνα  $i$ .

Αρά, για την ελαχιστοποίηση του:

$$\min \left( \sum_{i=1}^N \sum_{k=1}^{K_i} ||Project(P_i, X_k) - u_{i,k}|| \right) \quad (28)$$

Για τον υπολογισμό, χρησιμοποιείται η μέθοδος ελάχιστων τετράγωνων για μη γραμμικές εξισώσεις (Lourakis & Argyros, 2009). Μια πολύ καλή λύση στο πρόβλημα αυτό αποτελεί ο αλγόριθμος των Levenberg-Marquardt. Φυσικά, για να εφαρμοστεί ο αλγόριθμος αυτός χρειάζονται αρχικές παρατηρήσεις.

### 2.6.3 Συνόρθωση γράφου

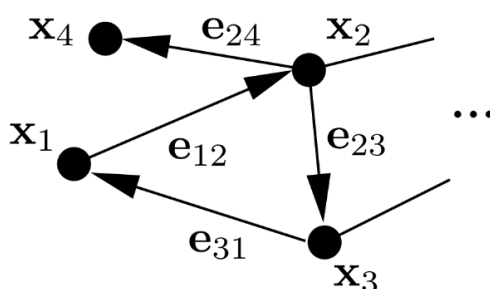
Η συνόρθωση γράφου είναι μια τεχνική που χρησιμοποιείται σε διαδικασίες όπως το Visual SLAM, η συνόρθωση δέσμης κλπ. (Kummerle et al., 2011). Ειδικότερα,



χρησιμοποιείται για την επίλυση προβλημάτων όπου ελαχιστοποιείται μια μη γραμμική συνάρτηση (Kummerle et al., 2011).

Χρησιμοποιώντας τον γράφο (ή γράφημα) μπορεί κανείς να υπολογίσει εκείνες τις μεταβλητές που περιγράφουν στο μέγιστο ένα σετ μετρήσεων οι οποίες επηρεάζονται από θόρυβο Gauss (Kummerle et al., 2011).

Ο γράφος αποτελείται από ακμές και κόμβους, όπου σύμφωνα με τους Grisetti et al. (2010) οι ακμές αντιπροσωπεύουν την θέση του αισθητήρα και των σημείων στον χώρο, και οι κόμβοι, οι οποίοι ενώνουν δυο ακμές, αποτελούν τις μετρήσεις και τους γεωμετρικούς περιορισμούς μεταξύ των αντίστοιχων ακμών.



Εικόνα 26: Παράδειγμα προβλήματος γράφου. Οι κόμβοι παρουσιάζονται ως κύκλοι και οι ακμές ως γραμμές που ενώνουν τους κόμβους. (Πηγή: (Kummerle et al., 2011)).

Μπορεί κανείς να λύσει το πρόβλημα, ελαχιστοποιώντας μια συνάρτηση  $F(x)$  η οποία ορίζεται από:

1. το διάνυσμα των παραμέτρων  $x = (x_1, \dots, x_T)^T$ , όπου το κάθε  $x$  αντιπροσωπεύει μια ακμή.
2. Το  $z_{ij}$  αποτελεί την πρόβλεψη που προκύπτει από μέτρηση και το  $\Omega_{ij}$  αποτελεί τον πίνακα των βαρών ενός περιορισμού που μεταξύ των  $x_j$  και  $x_i$  (Kummerle et al., 2011).
3. Το σφάλμα  $e(x_i, x_j, z_{ij})$  μεταξύ των παραμέτρων  $x_i, x_j$  και του γεωμετρικού περιορισμού τους  $z_{ij}$ , δηλαδή κατά ποσό οι παράμετροι  $x_i$  και  $x_j$  ικανοποιούν τον περιορισμό  $z_{ij}$ .

Αρά, το πρόβλημα λύνεται ως εξής:

$$F(x) = \sum_{(i,j) \in C} e(x_i, x_j, z_{ij})^T \Omega_{ij} e(x_i, x_j, z_{ij}) \quad (29)$$

$$\operatorname{argmin}_x F_x$$

Αρχικά, η συνόρθωση γράφου προτάθηκε από τους Lu και Milios (1997), όμως δεν έγινε αρκετά δημοφιλής για πολλά χρόνια λόγω της δυσκολίας χρήσης απλών τεχνικών ελαχιστοποίησης του σφάλματος (Grisetti et al., 2010). Πλέον, στην βιβλιογραφία υπάρχουν πολλές τεχνικές Visual SLAM οι οποίες χρησιμοποιούν την συνόρθωση γράφου.

### 3 Οπτικός Ταυτόχρονος Εντοπισμός θέσης και Χαρτογράφησης (Visual SLAM)

Η διαδικασία του Visual SLAM (Visual Simultaneous Localization and Mapping) πραγματοποιείται για την δημιουργία χάρτη ενός αγνώστου περιβάλλοντος, ενώ ταυτόχρονα υπολογίζεται η θέση/τροχιά του αισθητήρα (στην συγκεκριμένη περίπτωση κάμερα) σε κάθε χρονική στιγμή. Θα μπορούσαμε συνεπώς να πούμε πως πρόκειται για μία διαδικασία πλοήγησης και χαρτογράφησης μέσω εικόνων.

Στην βιβλιογραφία συναντώνται διάφοροι αλγόριθμοι που χρησιμοποιούν ως αισθητήρα την κάμερα. Πιο συγκεκριμένα, μπορεί να χρησιμοποιηθεί η πληροφορία από μια μόνο RGB κάμερα (Monocular), από δύο κάμερες RGB (Stereo) αλλά και από κάμερα βάθους RGB-D. Το πλεονέκτημα της χρήσης δύο καμερών ή κάμερας βάθους είναι η ταυτόχρονη εκτίμηση του βάθους της σκηνής, που πληροφορεί για το πόσο μακριά είναι τα αντικείμενα από την κάμερα. Αυτή η πληροφορία δεν είναι άμεσα διαθέσιμη στα προβλήματα ταυτόχρονου εντοπισμού και χαρτογράφησης που χρησιμοποιούν μόνο μια κάμερα.

Εκτός από το Visual SLAM, υπάρχει άλλη μια αντίστοιχη μέθοδος στη βιβλιογραφία που εντοπίζει την θέση της κάμερας στον χώρο και δημιουργεί έναν τοπικό χάρτη. Η μέθοδος αυτή ονομάζεται Οπτική Οδομετρία (VO, Visual Odometry). Η κύρια διαφορά μεταξύ του SLAM και της Οδομετρίας είναι πως η δεύτερη επικεντρώνεται περισσότερο για την συνέπεια των υπολογισμών της τροχιάς (Scaramuzza & Fraundonfer, 2011), ενώ στο SLAM είναι βασικό να υπάρχει συνέπεια στον συνολικό χάρτη (Scaramuzza & Fraundonfer, 2011).

Οι αλγόριθμοι μονοεικονικής Οπτικής Οδομετρίας (Monocular VO) μπορούν να κατηγοριοποιηθούν τόσο ως προς την μέθοδο εντοπισμού χαρακτηριστικών σημείων, ενώ οι αλγόριθμοι Visual SLAM ως προς την μέθοδο βελτιστοποίησης (Scaramuzza & Fraundonfer, 2011).

Σχετικά με τις μεθόδους Οπτικής Οδομετρίας, χαρακτηριστικά σημεία μπορούν να εντοπιστούν είτε με αυτόματο τρόπο, όπως οι αλγόριθμοι όπως SIFT και ORB που αναφέρθηκαν σε προηγούμενο κεφάλαιο, είτε με άμεσες μεθόδους. Οι άμεσες μέθοδοι πρακτικά αναλύουν τις τιμές των pixel απευθείας (Scaramuzza & Fraundonfer, 2011). Για παράδειγμα, άμεση μέθοδο αποτελούν αλγόριθμοι που λαμβάνουν την πληροφορία του βάθους από τις εικόνες, έτσι γνωρίζοντας την απόσταση του κάθε σημείου από την εικόνα πραγματοποιούν συνταύτιση.

Ως προς τους αλγόριθμους Visual SLAM, αναλύονται σε μεθόδους βασισμένες σε φίλτρα και μεθόδους μη βασισμένες σε φίλτρα, γνωστές και ως μέθοδοι βασικών καρέ (keyframes) (Scaramuzza & Fraundonfer, 2011).

Οι Cheng et al. (2022) χωρίζουν τα βασικά μέρη ενός αλγορίθμου Visual SLAM σε Front - end, Back – end, κλείσιμο βρόγχου (loop closure) και χαρτογράφηση.

Ως front – end παρουσιάζουν τις διαδικασίες οπτικής οδομετρίας, δηλαδή εντοπισμός χαρακτηριστικών σημείων και υπολογισμός της κίνησης της κάμερας ανάμεσα σε δυο γειτονικά καρέ (Cheng et al., 2022).

Χρησιμοποιώντας τα δεδομένα του front – end, ο αλγόριθμος προχωράει στο back – end όπου βελτιστοποιείται η θέση της κάμερας και λαμβάνονται με συνεπή τρόπο τόσο η τροχιά που ακολούθησε η κάμερα όσο και ο συνολικός χάρτης σημείων (Cheng et al., 2022). Σε αυτό το σημείο πραγματοποιούνται είτε μέθοδοι φιλτραρίσματος, είτε μέθοδοι βασικών καρτέ.

Η επιμέρους διαδικασία που αφορά τον κλείσιμο των βρόχων έχει ως σκοπό την αναγνώριση των περιοχών που ο αισθητήρας έχει ξανά επισκεφθεί. Η αναγνώριση των ήδη επισκεπτόμενων περιοχών συμβάλλει στην διόρθωση των συσσωρευμένων σφαλμάτων (φαινόμενο ολίσθησης) που ενδέχεται να προκύψουν από μια διαδικασία Visual SLAM (Cheng et al., 2022).

Όπως έχει ήδη αναφερθεί ένα από τα κύρια συστατικά του Visual SLAM είναι η δυνατότητα της χαρτογράφησης των σημείων του χώρου, καθώς και της θέσης της κάμερας στον χώρο. Αναφερόμενοι σε αυτόνομη πλοήγηση, ο χάρτης που λαμβάνεται από την διαδικασία του αυτόματου εντοπισμού και χαρτογράφησης έχει ως σκοπό τον εντοπισμό του οχήματος στον εν λόγω χάρτη βοηθώντας στην αποφυγή εμποδίων (Cheng et al., 2022).

### **3.1 Διαφορά μεταξύ των μεθόδων φιλτραρίσματος και βασικών καρτέ.**

Όπως αναφέρθηκε παραπάνω, οι αλγόριθμοι Visual SLAM χωρίζονται, κατά κύριο λόγο, ανάλογα με τον τρόπο που επεξεργάζονται τα δεδομένα σε μέθοδοι φιλτραρίσματος και μέθοδοι βασικών καρτέ.

Οι μέθοδοι φιλτραρίσματος χρησιμοποιούν φίλτρα με σκοπό να υπολογίσουν τις άγνωστες παραμέτρους. Πιο συγκεκριμένα, οι μέθοδοι αυτοί εφαρμόζουν κατανομή πιθανοτήτων ώστε να συνοψίσουν την πληροφορία που έχει υπολογιστεί κατά την διάρκεια του αλγορίθμου Visual SLAM (Strasdat et al., 2012).

Μερικά από τα φίλτρα που χρησιμοποιούνται είναι το φίλτρο Bayes, Kalman ή Extended Kalman (Kazerouni et al., 2022). Η διαφορά μεταξύ του φίλτρου Kalman και του Extended Kalman είναι πως το πρώτο θεωρεί πως τα δεδομένα είναι επηρεασμένα από θόρυβο Gauss και επιλύει σύστημα γραμμικών εξισώσεων (Kazerouni et al., 2022). Το Extended Kalman πρακτικά γραμμικοποιεί συστήματα, όπως είναι το Visual SLAM, τα οποία δεν είναι γραμμικά (Kazerouni et al., 2022).

Οι Leonard & Durrant – Whyte (1991) χρησιμοποίησαν το EKF (Extended Kalman Filter) σε μια εφαρμογή Visual SLAM όπου το περιβάλλον ήταν ήδη γνωστό εκ των προτέρων. Γενικώς, υπάρχουν αρκετές εφαρμογές του EKF σε VSALM προβλήματα, όμως οι Castellanos et al. (2004) αναφέρουν πως στις περισσότερες περιπτώσεις εφαρμογής του συγκεκριμένου φίλτρου, τα σφάλματα στην γραμμικοποίηση των εξισώσεων μπορεί να επιφέρουν ασυνέπεια στα αποτελέσματα της εν λόγω εφαρμογής.

Μια ακόμα εφαρμογή Visual SLAM βασισμένη σε τεχνικές φιλτραρίσματος προτάθηκε από τους Montemerlo et al. (2002), γνωστή και ως FastSLAM. Σε αυτόν τον αλγόριθμο χρησιμοποιούνται δύο τρόποι φιλτραρίσματος, Kalman και φίλτρο σωματιδίων, ακολουθεί, δηλαδή, Bayesian μεθοδολογία (Montemerlo et al., 2002).

Τα φίλτρα σωματιδίων (particle filters) αποτελούν ακόμα μια μέθοδο επεξεργασίας των δεδομένων που λαμβάνονται κατά τον υπολογισμό της θέσης της κάμερας και του

τρισδιάστατου μοντέλου. Ανήκουν στην οικογένεια των αλγορίθμων Monte Carlo (Murphy, 1999).

Ένας αρκετά γνωστός αλγόριθμος φίλτρου σωματιδίων είναι ο Rao-Blackwellized (Grisetti et al., 2007). Ένα από τα αρνητικά του αλγορίθμου αυτού είναι πως αν τα σωματίδια που απαιτούνται για τον χάρτη είναι πάρα πολλά, η πολυπλοκότητα του θα είναι αυξημένη (Grisetti et al., 2007).

Οι μέθοδοι των βασικών καρτέ δεν υπολογίζουν το τρισδιάστατο μοντέλο και την θέση της κάμερα λαμβάνοντας υπόψη όλα τα καρτέ, αλλά μερικά τα οποία πληρούν συγκεκριμένες προϋποθέσεις. Οι μέθοδοι των βασικών καρτέ χρησιμοποιούν μέθοδο δέσμης (Strasdat et al., 2012).

Ένας γνωστός αλγόριθμος που εφαρμόζει την τεχνική των βασικών καρτέ είναι ο αλγόριθμος PTAM (Parallel Tracking And Mapping) των Klein και Murray (2007), στον οποίο έγινε η πρόταση διαχωρισμού των διαδικασιών εντοπισμού και χαρτογράφησης ως ξεχωριστές (Klein & Murray, 2007). Η εισαγωγή βασικού καρτέ πρέπει να ακολουθεί τους εξής περιορισμούς: θα πρέπει να έχουν περάσει τουλάχιστον 20 βασικά καρτέ από την εισαγωγή του προηγούμενου και η κάμερα δεν απέχει από το κοντινότερο σημείο του χάρτη πάνω από μια συγκεκριμένη απόσταση (Klein & Murray, 2007).

Οι Engel et al. (2014) δημοσίευσαν τον αλγόριθμο LSD – SLAM (Large Scale Direct monocular SLAM). Ο αλγόριθμος αυτός εντοπίζει σημεία με άμεσες μεθόδους. Για την εισαγωγή ενός καινούργιου βασικού καρτέ υπολογίζουν την απόσταση της κάμερας από τον χάρτη, και εάν η απόσταση αυτή είναι σημαντική τότε προστίθεται νέο βασικό καρτέ (Engel et al., 2014). Για την δημιουργία τιμής κατωφλίου συνδυάζουν την σχετική γωνία και απόσταση στο τρέχον καρτέ (Engel et al., 2014).

Οι Forstner et al. (2014) εισήγαγαν τον αλγόριθμο SVO (Semi-Direct Visual Odometry). Ο αλγόριθμος αυτός εντοπίζει ομόλογα σημεία με άμεση μέθοδο, ενώ χρησιμοποιεί μέθοδο βασισμένη σε χαρακτηριστικά σημεία μόνο στο αρχικό βασικό καρτέ όπου και αρχικοποιείται ο τρισδιάστατος χάρτης (Forstner et al., 2014). Η επιλογή του βασικού καρτέ πραγματοποιείται εάν η Ευκλείδεια απόσταση του τρέχοντος καρτέ σε σχέση με τα βασικά καρτέ είναι μεγαλύτερη από το 12% του μέσου βάθους σκηνής (Forstner et al., 2014).

Μια ακόμα μέθοδος βασικών καρτέ προτάθηκε από τους Mur-Artal et al. (2015) και είναι γνωστή ως ORB – SLAM. Ο αλγόριθμος αυτός χρησιμοποιεί μεθόδους αυτόματου εντοπισμού χαρακτηριστικών σημείων. Η εισαγωγή νέου βασικού καρτέ γίνεται αφού έχουν περάσει περισσότερα από 20 καρτέ από τον προηγούμενο επανεντοπισμό ή από την τελευταία εισαγωγή βασικού καρτέ (Mur-Artal et al., 2015). Αν στο τρέχον καρτέ έχουν εντοπιστεί τουλάχιστον 50 σημεία και αν στο τρέχον καρτέ έχουν εντοπιστεί λιγότερα από το 90% των σημείων που εμφανίζονται στο βασικό καρτέ αναφοράς. Ο αλγόριθμος ORB – SLAM θα αναλυθεί περισσότερο παρακάτω, αφού σε εκείνον βασίστηκε και ο αλγόριθμος που χρησιμοποιήθηκε στην παρούσα εργασία.

### 3.2 Τεχνικές κλεισίματος βρόχων και επανεντοπισμού

Το κλείσιμο των βρόχων σε τεχνικές Visual SLAM είναι αρκετά χρήσιμο σε περιπτώσεις όπου στα δεδομένα επανεμφανίζεται η ίδια περιοχή. Με την μέθοδο αυτή

αναγνωρίζονται οι περιοχές που έχει ήδη επισκεφθεί το όχημα ή το ρομπότ, με αποτέλεσμα να μπορούν εκ των υστέρων να διορθωθούν σφάλματα στην θέση και τον προσανατολισμό της κάμερας, και συνεπώς σε όλη τη διαδρομή που έχει ακολουθήσει η κάμερα.

Οι Williams et al. (2009) αναφέρουν πως οι αλγόριθμοι εντοπισμού κλεισίματος βρόχων μπορούν να χωριστούν σε τρεις κατηγορίες ανάλογα με τον τρόπο που αναζητούν τις ομολογίες:

- Map-to-map: όπου τα ομόλογα σημεία αναζητούνται μεταξύ δυο υπό-χαρτών και αξιοποιείται τόσο η εμφάνιση όσο η σχετική τους θέση.
- Image-to-image: Η αναζήτηση γίνεται ανάμεσα στο τρέχον καρέ με όλα τα προηγούμενα.
- Image-to-map: Η αναζήτηση πραγματοποιείται μεταξύ του τρέχοντος καρέ και των σημείων στον χάρτη.

Οι Williams et al. (2009) αναφέρουν πως για περιοχές που εμφανίζουν λίγα χαρακτηριστικά σημεία λειτουργεί καλύτερη η μέθοδος map-to-image, ενώ για περιβάλλοντα με περισσότερα σημεία έχει καλύτερα αποτελέσματα η μέθοδος image-to-image.

Μια αρκετά γνωστή μέθοδος που χρησιμοποιείται στο κλείσιμο των βρόχων ονομάζεται Bag-of-words. Η μέθοδος αυτή έχει ως αποτέλεσμα την δημιουργία οπτικών λέξεων και οπτικού λεξιλογίου ώστε να είναι ευκολότερη η ανάκτηση των εικόνων. Πιο συγκεκριμένα η δημιουργία των οπτικών λέξεων πραγματοποιείται μετά τον υπολογισμό των περιγραφέων των χαρακτηριστικών σημείων, οι οποίοι ομαδοποιούνται ανάλογα με την ομοιότητα τους με αλγορίθμους όπως ο K – Means (Sivic & Zisserman, 2003). Οπτικό λεξιλόγιο αποτελεί η κατανομή των ομαδοποιημένων, πλέον, περιγραφέων σε ιστόγραμμα (Sivic & Zisserman, 2003).

Με την τεχνική Bag-of-words, μπορεί κανείς να βελτιώσει την διαδικασία ανάκτησης των εικόνων αναπαριστώντας το οπτικό λεξιλόγιο σε δένδρική μορφή όπως προτάθηκε από τους Nister & Stewenius (2006). Μέσω της συγκεκριμένης αναπαράστασης, το οπτικό λεξιλόγιο αποκτά μια ιεραρχική μορφή.

Στην βιβλιογραφία υπάρχουν αρκετές τεχνικές Visual SLAM που χρησιμοποιούν την τεχνική bag-of-words για να πραγματοποιήσουν κλείσιμο βρόχων. Για παράδειγμα οι Mur-Artal et al. (2015) χρησιμοποιούν το bag-of-words για να αναζητήσουν πιθανούς βρόχους. Οι Cummings & Newman (2008) δημιουργούν bag-of-words από τις μετρήσεις του αισθητήρα με σκοπό να υπολογίσουν την πιθανότητα να εμφανίζεται η ίδια περιοχή σε δυο εικόνες.

Μια διαφορετική μέθοδος, που όμως έχει βασιστεί στις μεθόδους bag-of-words και Fisher kernel ονομάζεται VLAD (Vector of Locally Aggregated Descriptors) (Jégou et al. 2010). Όπως και στην μέθοδο bag-of-words, υπολογίζονται οι περιγραφείς των σημείων και ποσοτικοποιούνται με αλγορίθμους ταξινόμησης όπως ο K – Means στο κοντινότερο κεντροειδές (Jégou et al. 2010). Η διαφορά του συγκεκριμένου αλγορίθμου βρίσκεται στο γεγονός ότι υπολογίζεται η κατανομή των περιγραφέων γύρω από το κεντροειδές τους και η πληροφορία αυτή αποθηκεύεται σε ένα διάνυσμα (Jégou et al. 2010). Αυτό πραγματοποιείται με το άθροισμα της διαφοράς των

περιγραφών με τις αντίστοιχες οπτικές λέξεις, δηλαδή από το κεντροειδές τους (Jégou et al. 2010).

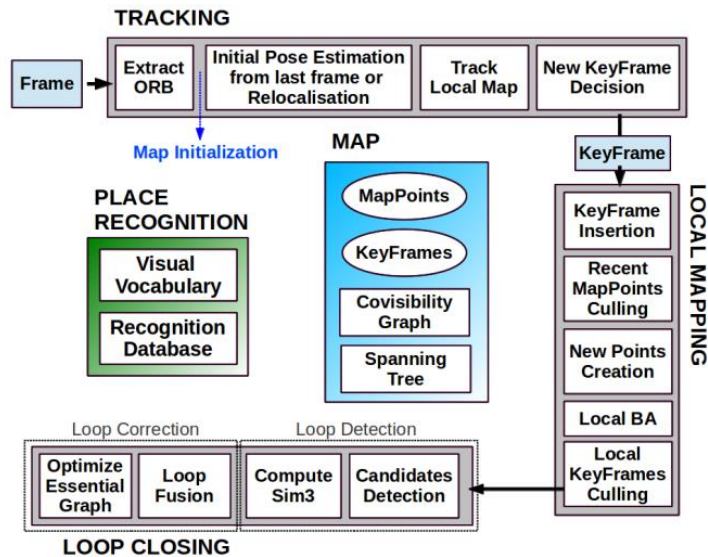
Ο επανεντοπισμός της θέσης είναι μια αρκετά σημαντική διαδικασία στα προβλήματα Visual SLAM. Το πρόβλημα του επανεντοπισμού βοηθάει στις περιπτώσεις όπου η θέση και ο προσανατολισμός της κάμερας έχουν χαθεί. Ο υπολογισμός της θέσης και του προσανατολισμού της κάμερας μπορεί να αποτύχουν σε αρκετές περιπτώσεις, όπως για παράδειγμα όταν συμβαίνει γρήγορη κίνηση της κάμερας και δεν είναι εφικτή η αντιστοίχιση των χαρακτηριστικών σημείων που έχουν εντοπιστεί, με το προηγούμενο καρέ.

Ο επανεντοπισμός μπορεί να πραγματοποιηθεί με αλγορίθμους αναγνώρισης περιοχών όπως το bag-of-words. Αυτή την μέθοδο ακολούθησαν οι Mur-Artal et al. (2015) στον αλγόριθμο ORB-SLAM, όπου όταν χαθεί ο εντοπισμός, αναζητείται στην βάση αναγνώρισης το βασικό καρέ με το οποίο εμφανίζονται τα περισσότερα κοινά σημεία. Εάν μετά γεωμετρικό έλεγχο το βασικό καρέ εξακολουθεί να εμφανίζει αρκετά ομόλογα με το τρέχον, πραγματοποιείται συνολικός επανεντοπισμός. Οι Klein & Murray (2008) χρησιμοποιούν βασικά καρέ για την πραγματοποίηση του επανεντοπισμού. Μειώνουν την ανάλυση των βασικών καρέ και πραγματοποιούν το ίδιο στο τρέχον καρέ όταν χαθεί ο εντοπισμός. Ως μέτρο ομοιότητας με τα βασικά καρέ χρησιμοποιείται το άθροισμα των τετραγωνικών διαφορών, με το καλύτερο αποτέλεσμα να δίνει το βασικό καρέ όπου εμφανίζεται η χαμηλότερη τιμή (Klein & Murray, 2008).

#### 4 Περιγραφή του αλγορίθμου ORB – SLAM

Όπως έχει ήδη αναφερθεί, ο αλγόριθμος που δοκιμάστηκε για την ολοκλήρωση της παρούσας εργασίας έχει βασιστεί στον αλγόριθμο ORB – SLAM (Mur-Artal et al., 2015), και αποτελεί ένα από τα ανοιχτά παραδείγματα εφαρμογής Monocular Visual SLAM του λογισμικού MATLAB. Σε αυτό το κεφάλαιο θα πραγματοποιηθεί μια σύντομη μελέτη της δημοσίευσης των Mur-Artal et al. (2015).

Ο αλγόριθμος ORB – SLAM είναι από τις πιο δημοφιλείς εφαρμογές Monocular Visual SLAM. Δεν αποτελεί άμεση μέθοδο, καθώς εντοπίζει χαρακτηριστικά σημεία στις εικόνες με βάση τον αλγόριθμο ORB των Rublee et al. (2011). Χρησιμοποιεί βασικά καρέ (keyframes) για τον υπολογισμό της θέσης και του προσανατολισμού της κάμερας, καθώς και για την χαρτογράφηση των τρισδιάστατων σημείων. Τα βασικά μέρη του αλγορίθμου παρουσιάζονται στο ακόλουθο διάγραμμα.



Εικόνα 27: Περιγραφή του αλγορίθμου ORB - SLAM (Πηγή: Mur-Artal et al. (2015)).

Σύμφωνα με τους Mur-Artal et al. (2015) ο εντοπισμός, η χαρτογράφηση και το κλείσιμο των βρόχων πραγματοποιούνται παράλληλα.

Η αρχικοποίηση του χάρτη πραγματοποιείται μεταξύ δυο καρέ, τα οποία είναι απαραίτητο να έχουν καλή γεωμετρία μεταξύ τους και πολλά ομόλογα σημεία. Για τον υπολογισμό του σχετικού προσανατολισμού αλλά και των τρισδιάστατων σημείων, θα πρέπει να ληφθεί υπόψιν η γεωμετρία του χώρου. Δηλαδή, αν το περιβάλλον που εμφανίζεται στις δυο εικόνες είναι σχετικά επίπεδο, τότε χρησιμοποιούνται οι σχέσεις της ομογραφίας με την μέθοδο DLT (Mur-Artal et al., 2015). Διαφορετικά, χρησιμοποιείται ο επιπολικός πίνακας με την λύση των οχτώ σημείων (Mur-Artal, et al., 2015). Και στις δυο λύσεις, φυσικά εφαρμόζεται RANSAC. Δεδομένου πως οι Mur-Artal et al., (2015) ήθελαν τον αλγόριθμο να είναι αυτόματος, υπολόγισαν το σκορ της κάθε μεθόδου και με βάση αυτό το σκορ δημιούργησαν μια ευρετική (heuristic) συνάρτηση, ώστε να αποφασίζει ο αλγόριθμος ποιο μοντέλο είναι κατάλληλο.

Αφού αρχικοποιηθεί ο χάρτη και υπάρχουν τα δυο βασικά καρέ, τα αρχικά τρισδιάστατα σημεία και η αρχική θέση και προσανατολισμός της κάμερας, ο αλγόριθμος προχωράει στον εντοπισμό.

Στην διαδικασία του εντοπισμού ξεκινάει ο αλγόριθμος να ελέγχει τα καρέ (δεδομένου πως υπάρχει ήδη μια αρχικοποίηση από προηγούμενως) για ομόλογα σημεία. Εντοπίζονται, λοιπόν χαρακτηριστικά σημεία και οι περιγραφείς τους μέσω του ORB, και έπειτα υπολογίζεται μια αρχική εκτίμηση της θέσης και του προσανατολισμού της κάμερας (Mur-Artal et al., 2015). Δεδομένου ότι ο εντοπισμός είναι πετυχημένος, πραγματοποιείται περεταίρω αναζήτηση ομόλογων σημείων προβάλλοντας τον χάρτη στο καρέ (Mur-Artal et al., 2015). Εφαρμόζοντας διάφορα κριτήρια για απαλοιφή σημείων ώστε να αποφευχθούν οι ακραίες αποκλίσεις (outliers), πραγματοποιείται η βελτιστοποίηση της θέσης και του προσανατολισμού της κάμερας (Mur-Artal et al., 2015). Στον εντοπισμό εντάσσεται και η διαδικασία επιλογής του νέου βασικού καρέ. Η διαδικασία εισαγωγής του νέου βασικού καρέ αναφέρθηκε στο προηγούμενο κεφάλαιο.

Παράλληλα με τον εντοπισμό, εφαρμόζεται και το τμήμα του αλγορίθμου που πραγματοποιεί χαρτογράφηση (τοπική χαρτογράφηση). Η χαρτογράφηση γίνεται σε κάθε βασικό καρέ που εντοπίζεται, το οποίο εισάγεται στον γράφο συνορατότητας (covisibility graph) και τον ενημερώνει για τις νέες σχέσεις σημείων – βασικών καρέ. Τα σημεία του χάρτη εισάγονται σε ένα αρκετά αυστηρό έλεγχο, όπου εάν πληρούν τις προϋποθέσεις μπορούν να παραμείνουν στον χάρτη. Επιπλέον, πραγματοποιείται αναζήτηση για περισσότερα ομόλογα σημεία στα συσχετισμένα βασικά καρέ, απορρίπτοντας τα σημεία εκείνα που δεν πληρούν τις επιπολικές δεσμεύσεις. Εφαρμόζοντας συνόρθωση δέσμης, βελτιστοποιείται το βασικό καρέ και εκείνα που συνδέονται με αυτό, καθώς και όλα τα σημεία. Εφαρμόζεται εξάλειψη και σε βασικά καρέ που επαναλαμβάνονται

Για τον κλείσιμο των βρόχων, όπως έχει αναφερθεί και παραπάνω, χρησιμοποιείται η τεχνική bag-of-words μέσω της οποίας δημιουργείται η βάση αναγνώρισης. Ελέγχεται η ομοιότητα του τρέχοντος καρέ με τα προηγούμενα βασικά καρέ ώστε να βρεθεί αν είναι υποψήφιο για κλείσιμο βρόγχου. Για τον κλείσιμο του βρόγχου υπολογίζεται ένας μετασχηματισμός ομοιότητας μεταξύ του τρέχοντος καρέ και εκείνου που εντοπίστηκε το κλείσιμο του βρόγχου. Έπειτα, διορθώνεται ο γράφος συνορατότητας με τα κοινά σημεία και τον μετασχηματισμό ομοιότητας ο οποίος εφαρμόζόμενος στην τρέχουσα θέση της κάμερας, συμβάλλει και στην συνόρθωση όλων των προηγούμενων ώστε να ικανοποιείται η γεωμετρία. Και τέλος πραγματοποιείται συνόρθωση του Essential Graph μέσω της συνόρθωσης γράφου για την ομοιόμορφη κατανομή του σφάλματος του κλεισίματος του βρόγχου αλλά και της απόκλισης της κλίμακας.

## 5 Περιγραφή του αλγορίθμου της εφαρμογής

Όπως αναφέρθηκε προηγουμένως, για την μελέτη του Monocular Visual SLAM εφαρμόστηκε η μέθοδος “Monocular Visual Simultaneous Localization and Mapping” του MATLAB. Η μέθοδος αυτή έχει βασιστεί στον αλγόριθμο ORB – SLAM, όπως περιεγράφηκε στο προηγούμενο κεφάλαιο. Παρόλα αυτά, για περαιτέρω δοκιμή του αλγορίθμου εφαρμόστηκαν επιπλέον οι αλγόριθμοι εντοπισμού και εξαγωγής χαρακτηριστικών σημείων SIFT και SURF.

### 5.1 Αρχικοποίηση του χάρτη

Όπως και στην περίπτωση του ORB – SLAM, ο αλγόριθμος ξεκινάει με την αρχικοποίηση του χάρτη θεωρώντας το πρώτο καρέ και αναζητώντας εκείνο το καρέ που θα προσφέρει την καλύτερη γεωμετρία. Αρχικά εντοπίζονται τα χαρακτηριστικά σημεία και εξάγονται οι περιγραφείς τους για το πρώτο καρέ. Έπειτα, εφαρμόζεται μια επαναληπτική διαδικασία στην οποία αναζητείται το καλύτερο δεύτερο καρέ, ώστε να υπολογιστούν τα τρισδιάστατα σημεία του χάρτη. Όπως στην δημοσίευση του ORB – SLAM έτσι και στον συγκεκριμένο αλγόριθμο, υπολογίζονται η ομογραφία και ο επιπολικός πίνακας μεταξύ των δύο καρέ και ορίζεται ένα σκορ ανά περίπτωση. Με βάση αυτό το σκορ υπολογίζεται η ευρετική συνάρτηση με το ίδιο κριτήριο όπως των Mur-Artal, Montiel, & Tardos (2015):

$$R_H = \frac{S_H}{S_H + S_F} \quad (30)$$



Όπου  $S_H$  και  $S_F$  είναι το σκορ που έχει προκύψει από την ομογραφία και τον επιπολικό πίνακα αντίστοιχα. Αν  $R_H > 0.45$  επιλέγεται η ομογραφία (Mur-Artal et al., 2015). Σκοπός είναι το μοντέλο που θα επιλεγεί να έχει το μικρότερο σφάλμα επαναπροβολής.

Τα ομόλογα σημεία που θα παραμείνουν μετά την εφαρμογή του γεωμετρικού περιορισμού (ομογραφία ή επιπολικός), χρησιμοποιούνται για τον υπολογισμό του σχετικού προσανατολισμού των δυο καρέ. Έχοντας θέση και προσανατολισμό της κάμερας, καθώς και τα ομόλογα σημεία, μπορούν να υπολογιστούν οι τρισδιάστατες συντεταγμένες των ομόλογων σημείων.

Αφού αρχικοποιηθεί ο χάρτης με τα 3D σημεία, ο αλγόριθμος προχωράει στην αποθήκευση των βασικών καρέ και των πληροφοριών που σχετίζονται με αυτά όπως οι περιγραφείς, οι θέσεις και οι προσανατολισμοί της κάμερας. Επίσης αποθηκεύονται, οι ακμές του γράφου (σχετικές και απόλυτες θέσεις) καθώς και οι σχετικές θέσεις που προκύπτουν μετά από κλείσιμο των βρόχων. Σε αυτό το σημείο δημιουργείται και ενημερώνεται ο γράφος.

Σε ξεχωριστό πίνακα αποθηκεύονται οι σημασιολογικές πληροφορίες που αφορούν τα τρισδιάστατα σημεία, δηλαδή σε ποια εικόνα προβάλλονται τα σημεία αυτά και αντίστοιχα ποια εικόνα «βλέπει» τα σημεία αυτά, και επιπρόσθετες ακόμα πληροφορίες σχετικά με τα σημεία αυτά.

Όπως έχει ήδη μελετηθεί, ο αλγόριθμος ORB – SLAM (Mur-Artal et al., 2015) χρησιμοποιεί bag-of-words για τον εντοπισμό κλεισίματος βρόχων αλλά και σε περιπτώσεις όπου έχει χαθεί ο εντοπισμός. Οι Mur-Artal et al. (2015) αναφέρουν πως η διαδικασία δημιουργίας του οπτικού λεξιλογίου πραγματοποιείται “offline”. Αντίστοιχα, και στον αλγόριθμο αυτόν δημιουργείται το οπτικό λεξιλόγιο σε ξεχωριστή διαδικασία από τον αλγόριθμο (offline). Το οπτικό λεξιλόγιο δημιουργείται από ένα μέρος του σετ δεδομένων και έχει δενδρική μορφή. Αφού δημιουργηθεί το οπτικό λεξιλόγιο εισάγεται στον αλγόριθμο, δημιουργείται η βάση μέσω του Inverted Index το οποίο επιφέρει ευκολότερο και γρηγορότερο εντοπισμό αντιστοιχιών. Στην βάση εισάγονται οι πληροφορίες των βασικών καρέ.

Δεδομένου ότι η αρχικοποίηση του χάρτη επηρεάζει την μετέπειτα διαδικασία, είναι αναγκαίο να πραγματοποιηθεί με την μεγαλύτερη δυνατή ακρίβεια. Γι’ αυτόν τον λόγο εφαρμόζεται μέθοδος δέσμης ώστε να βελτιστοποιηθούν τα σημεία και οι προσανατολισμοί μέσω μείωσης του σφάλματος επαναπροβολής.

## 5.2 Εντοπισμός θέσης και στροφών

Το κομμάτι του κώδικα που αναφέρεται στον εντοπισμό, ακολουθεί την ίδια διαδικασία με εκείνη που παρουσιάστηκε από τους (Mur-Artal, Montiel, & Tardos, 2015). Αρχικά εντοπίζονται και εξάγονται χαρακτηριστικά σημεία στο τρέχον καρέ και έπειτα αναζητούνται ομολογίες με το προηγούμενο βασικό καρέ και υπολογίζεται η θέση και ο προσανατολισμός της κάμερας χρησιμοποιώντας τον αλγόριθμο PnP (Perspective-n-Point). Σε αυτό το σημείο υπολογίζονται η θέση και ο προσανατολισμός της κάμερας με την Μέθοδο Δέσμης. Έτσι αναζητούνται περισσότερες ομολογίες μεταξύ του τρέχοντος καρέ και του τελευταίου βασικού καρέ μέσω της προβολής των σημείων του χάρτη που προέκυψαν από το προηγούμενο βασικό καρέ, στο τρέχον καρέ και εφαρμόζεται Μέθοδος Δέσμης για την συνόρθωση της θέσης και του

προσανατολισμού της κάμερας. Από όλα τα σημεία που εντοπίζονται, διατηρούνται μόνο όσα βρίσκονται σε καλή κλίμακα μπροστά στην κάμερα, ενώ τα υπόλοιπα απορρίπτονται. Πραγματοποιείται μια ακόμα φορά η διαδικασία της συνόρθωσης μόνο της θέσης και του προσανατολισμού της κάμερας με βάση τα σημεία που έχουν απομείνει.

Ο εντοπισμός συνεχίζεται με την προβολή των σημείων του τρισδιάστατου χάρτη στο τρέχον καρέ με 3D-to-2D ομολογίες ώστε να αναζητηθούν περισσότερα κοινά σημεία, όπως αναφέρεται και από τους Mur-Artal et al. (2015). Σε αυτό το σημείο, εφαρμόζονται τα αυστηρά κριτήρια που έχουν αναφέρει οι Mur-Artal et al. (2015) για την απόρριψη των σημείων που θεωρούνται outliers. Τα κριτήρια αυτά αφορούν κυρίως τον τρόπο με τον οποίο εμφανίζονται τα σημεία στο τρέχον καρέ και στον χάρτη.

Στη συνέχεια ελέγχεται αν το συγκεκριμένο καρέ μπορεί να αποτελέσει βασικό καρέ, χρησιμοποιώντας τα κριτήρια που περιγράφονται από τους Mur-Artal, Montiel, & Tardos (2015), τα οποία έχουν αναφερθεί σε προηγούμενο κεφάλαιο. Αν το τρέχον καρέ οριστεί ως βασικό, τότε συνορθώνεται ξανά η θέση και ο προσανατολισμός της κάμερας.

### 5.3 Χαρτογράφηση

Δεδομένου πως έχει οριστεί το τρέχον καρέ ως βασικό, ο αλγόριθμος προχωράει στην χαρτογράφηση. Αρχικά εισάγεται το νέο βασικό καρέ στον γράφο όπου και ενημερώνονται οι νέες σχέσεις μεταξύ των βασικών καρέ και αποθηκεύονται οι πληροφορίες που αφορούν την θέση και τον προσανατολισμό της κάμερας για το δεδομένο βασικό καρέ κ.α. Οι Mur-Artal, Montiel, & Tardos (2015) αφαιρούν τα σημεία του χάρτη που εμφανίζονται σε λιγότερα από 3 βασικά καρέ, θεωρώντας τα ως ακραίες τιμές (outliers). Το ίδιο πραγματοποιείται και στην συγκεκριμένη περίπτωση, δηλαδή αμέσως μετά την ενημέρωση του γράφου, αφαιρούνται οι ακραίες τιμές των σημείων του χάρτη.

Έπειτα, δημιουργούνται τα σημεία του χάρτη με τριγωνισμό των χαρακτηριστικών σημείων που έχουν υπολογιστεί από τις προηγούμενες διαδικασίες. Όπως στην δημοσίευση των Mur-Artal et al. (2015), υπολογίζονται οι επιπολικές δεσμεύσεις μεταξύ των δυο βασικών καρέ. Λαμβάνονται υπόψιν το σφάλμα επαναπροβολής, η κλίμακα, το βάθος στο οποίο εμφανίζονται και στις δυο εικόνες, η παράλλαξη και η συνέπεια της κλίμακας.

Αφού δημιουργηθούν τα καινούργια σημεία του χάρτη, εφαρμόζεται τοπικά η μέθοδος δέσμης, εφόσον δεν πραγματοποιείται για όλα τα καρέ αλλά για τα βασικά καρέ τα οποία συμβάλλουν στην δημιουργία των τρισδιάστατων σημείων. Η συνόρθωση δέσμης πραγματοποιείται για το τρέχον βασικό καρέ, για όσα συνδέονται με αυτό και για τα σημεία του χάρτη που εμφανίζονται σε αυτά τα βασικά καρέ (Mur-Artal, Montiel, & Tardos, 2015). Αφού υπολογιστούν οι βελτιωμένες τιμές από την μέθοδο δέσμης, ενημερώνεται και ο γράφος.

### 5.4 Κλείσιμο βρόχων

Η διαδικασία προχωράει στην εύρεση των καρέ που είναι πιθανά για κλείσιμο του βρόγχου. Πιο συγκεκριμένα, αν έχουν περάσει 20 βασικά καρέ τότε ξεκινάει ο έλεγχος

εάν το τρέχον βασικό καρέ είναι υποψήφιο για κλείσιμο του βρόγχου. Ορίζεται ένας ελάχιστος αριθμός ομόλογων σημείων και πραγματοποιείται αναζήτηση στην βάση για εκείνα τα βασικά καρέ που έχουν τις περισσότερες ομοιότητες, και κρατείται το ελάχιστο σκορ τους. Ο αριθμός των εικόνων μπορεί να αλλάξει ανάλογα με την εφαρμογή, για παράδειγμα μπορούν να αναζητηθούν τα 10 καλύτερα αποτελέσματα. Βρίσκονται τα σκορ ομοιότητας για τα βασικά καρέ που δεν σχετίζονται μεταξύ τους στον γράφο. Για να είναι έγκυρο θα πρέπει το σκορ του να είναι μεγαλύτερο είτε από το καλύτερο σκορ πολλαπλασιασμένο με ένα ποσοστό είτε από το ελάχιστο σκορ που υπολογίστηκε προηγουμένως.

Άλλη μια προϋπόθεση ώστε να αποτελεί κλείσιμο βρόγχου είναι να έχουν εντοπιστεί τουλάχιστον 3 διαδοχικά υποψήφια καρέ κλεισίματος βρόγχου (Mur-Artal et al., 2015). Όταν εντοπίζονται υποψήφια καρέ, υπολογίζεται η σχετική θέση και προσανατολισμός της κάμερας μεταξύ του τρέχοντος βασικού καρέ και του υποψήφιου. Αν το υποψήφιο καρέ δεν χρησιμοποιηθεί για το κλείσιμο του βρόγχου, εισάγεται στον γράφο η σύνδεση του με το τρέχον καρέ και η σχετική θέση και προσανατολισμός που υπολογίστηκαν.

Στην συγκεκριμένη εφαρμογή ο αλγόριθμος τερματίζεται όταν εντοπιστεί κλείσιμο του βρόγχου. Οπότε, αφού εντοπιστεί πραγματοποιείται συνόρθωση του γράφου ώστε να διορθωθεί η συνολική πορεία και ο προσανατολισμός της κάμερας.

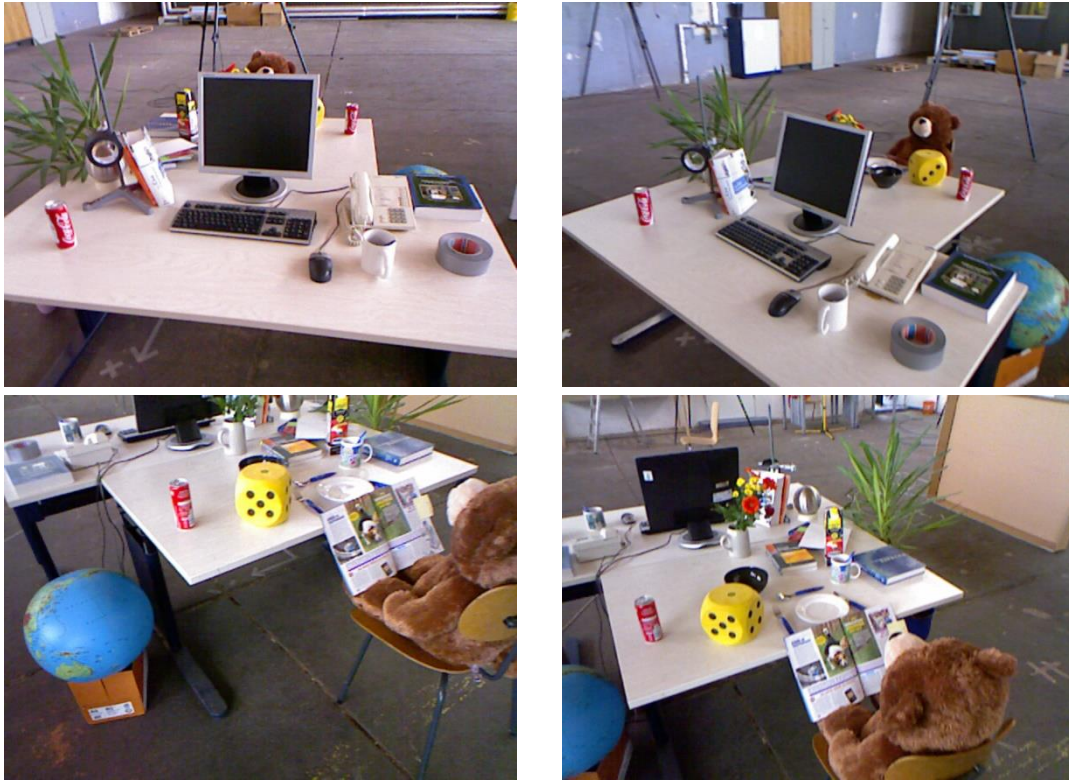
## 6 Αποτελέσματα

Ο αλγόριθμος που περιεγράφηκε παραπάνω δοκιμάστηκε με διαφορετικά σετ δεδομένων από το Τεχνικό Πανεπιστήμιο του Μονάχου (TUM). Τα συγκεκριμένα σετ δεδομένων παρέχουν εικόνες RGB αλλά και δεδομένα βάθους RGB – D. Τα δεδομένα αυτά λήφθηκαν μέσω του αισθητήρα Microsoft Kinect με ρυθμό ανανέωσης καρέ στα 30 Hz και η ανάλυση του αισθητήρα είναι 640×480. Το Kinect παρέχει και δεδομένα από επιταχυνσιόμετρο.

Στο σετ δεδομένων διατίθενται και τα πραγματικά δεδομένα (ground truth) της θέσης και του προσανατολισμού των λήψεων. Για τον υπολογισμό τους χρησιμοποιήθηκε ένα σύστημα εντοπισμού κίνησης με υψηλή ακρίβεια μέσω οχτώ κάμερες εντοπισμού μεγάλης ταχύτητας στα 100 Hz.

Στα δεδομένα παρέχονται και οι παράμετροι του εσωτερικού προσανατολισμού, όπου σε περιπτώσεις μη διόρθωσης των εικόνων από ακτινική διαστρόφη, δίνονται και οι παραμορφώσεις αυτές.

Χρησιμοποιήθηκαν δυο σετ δεδομένων: το Freiburg 2 desk validation (Εικόνα 28) και Freiburg 3 long office household (Εικόνα 29).

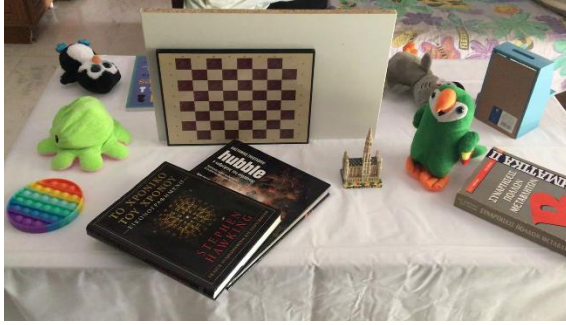


Εικόνα 28: Μέρος του σετ δεδομένων Freiburg 2 desk validation του TUM (<https://cvg.cit.tum.de/data/datasets/rgbd-dataset>)



Εικόνα 29: Μέρος του σετ δεδομένων Freiburg 3 long office household του TUM (<https://cvg.cit.tum.de/data/datasets/rgbd-dataset>)

Τέλος, αξιοποιήθηκε ένα σετ δεδομένων σε εσωτερικό χώρο όπου πάλι απεικονίζονται αντικείμενα σε ένα τραπέζι. Η τροχιά που ακολουθείται είναι κυκλική και τα δεδομένα λήφθηκαν με κάμερα κινητού τηλεφώνου με τα χαρακτηριστικά που έχουν αναφερθεί στο κεφάλαιο της βαθμονόμησης μηχανής.



Εικόνα 30: Σετ δεδομένων που δημιουργήθηκε όπου απεικονίζονται αντικείμενα πάνω σε τραπέζι.

Το σετ δεδομένων Freiburg 3 long office household περιέχει εικόνες διορθωμένες από ακτινική και εφαπτομενική διαστρόφη, ενώ δεν ισχύει το ίδιο για το σετ δεδομένων Freiburg 2 desk validation, το οποίο λήφθηκε υπόψιν για την διαδικασία του αλγορίθμου.

Για την αξιολόγηση των αποτελεσμάτων αλλά και την βέλτιστη σύγκρισή τους με τα πραγματικά δεδομένα (ground truth data), δημιουργήθηκε μια συνάρτηση όπου εφαρμόζομενη στο τέλος του αλγορίθμου συγκρίνει την υπολογισμένη στροφή και θέση της κάμερας με την πραγματική.

Η λογική που ακολουθήθηκε είναι αντίστοιχη με αυτή που αναφέρει ο Νάκης (2019) όπου με βάση την δημοσίευση των Geiger, Lenz & Urtasun (2012) το σφάλμα διακρίνεται σε γωνιακό και τροχιακό. Τα δεδομένα χωρίζονται σε υποπεριοχές ώστε να μελετηθεί η συμπεριφορά του σφάλματος όσο αυξάνεται η απόσταση. Ο χωρισμός που επιλέχθηκε είναι σε 10 υποπεριοχές των 2 μέτρων (2,4,...,20) βάση της πραγματικής τροχιάς που ακολουθήθηκε στα δυο σετ δεδομένων του TUM.

Η μεθοδολογία που υλοποιήθηκε από τον Νάκη (2019) είναι η εξής:

Αρχικά υπολογίζεται η αθροιστική απόσταση του κάθε βασικού καρέ από την αρχή της τροχιάς, έχοντας ως αναφορά τα πραγματικά δεδομένα. Έπειτα, ξεκινώντας από το πρώτο βασικό καρέ τόσο των αληθών όσο και των υπολογισμένων δεδομένων, βρίσκεται το βασικό καρέ με αθροιστική απόσταση πιο κοντά στην πρώτη υποπεριοχή. Δηλαδή, για την πρώτη υποπεριοχή, βρίσκεται εκείνο το καρέ που είναι πιο κοντά στα 2 μέτρα. Μεταξύ του πρώτου καρέ και του καρέ που βρέθηκε, υπολογίζεται ο σχετικός τους προσανατολισμός:

$$RelativeOr_{gTruth} = Pose^{-1}(gTruth)_i \times Pose(gTruth)_{nj} \quad (31)$$

$$RelativeOr_{VSLAM} = Pose^{-1}(VSLAM)_i \times Pose(VSLAM)_{nj} \quad (32)$$

Όπου ο δείκτης  $i$  αφορά το πρώτο καρέ, ενώ  $lj$  είναι το βασικό καρέ  $nj$  που βρίσκεται πιο κοντά στην υποπεριοχή  $n$ .

Έπειτα, υπολογίζεται ο “σχετικός προσανατολισμός” για τους σχετικούς προσανατολισμούς που υπολογίστηκαν προηγουμένως. Δηλαδή, υπολογίζεται πλέον ο σχετικός προσανατολισμός μεταξύ των αληθών και των υπολογισμένων δεδομένων, όπου αντιστοιχεί στην διαφορά μεταξύ των αληθών και των υπολογισμένων δεδομένων.

$$PoseError = RelativeOr_{VSLAM}^{-1} \times RelativeOr_{gTruth} \quad (33)$$

Οι σχέσεις που χρησιμοποιούνται για τον υπολογισμό των σχετικών προσανατολισμών, και στις δυο περιπτώσεις, αφορούν την μεταφορά των συστημάτων αναφοράς. Υπολογίζοντας τον αντίστροφο πίνακα μετασχηματίζονται τα υπολογισμένα από τον αλγόριθμο δεδομένα στο σύστημα των αληθινών δεδομένων, ώστε να είναι εφικτή η σύγκρισή τους. Αντίστοιχα και στις σχέσεις (31) και (32) του προηγούμενου βήματος, μετασχηματίζονται τα δεδομένα που βρίσκονται σε απόσταση 2 μέτρα μακριά από το αρχικό βασικό καρέ στο σύστημα αναφοράς του.

Από το αποτέλεσμα που προκύπτει από την σχέση (33) μπορούν να υπολογιστούν το γωνιακό και το τροχιακό σφάλμα. Το γωνιακό σφάλμα υπολογίζεται ως εξής:

$$RotationError = \arccos\left(\frac{TR(PoseError) - 1}{2}\right) \quad (34)$$

Το  $TR(PoseError) = PoseError_{11} + PoseError_{22} + PoseError_{33}$  είναι το ίχνος του πίνακα. Μέσω της L2 νόρμας μπορεί να βρεθεί το τροχιακό σφάλμα ως:

$$TrError = \sqrt{PoseError_{14}^2 + PoseError_{24}^2 + PoseError_{34}^2} \quad (35)$$

Δεδομένου πως τα σφάλματα είναι σε μοίρες ανά μέτρο και σε ποσοστό (%) αντίστοιχα, διαιρούνται με την απόσταση της κάθε υποπεριοχής.

Τέλος, επαναλαμβάνεται η συγκεκριμένη διαδικασία για όλα τα δεδομένα και για όλες τις υποπεριοχές. Δηλαδή, στην συγκεκριμένη εφαρμογή οι υποπεριοχές είναι της τάξης των 2 μέτρων (2, 4 έως 20), άρα αρχικά υπολογίζεται για το πρώτο βασικό καρέ με εκείνο που βρίσκεται πιο κοντά στα 2 μέτρα μακριά, για το δεύτερο αντίστοιχα εκείνο που βρίσκεται 2 μέτρα μακριά και ούτω καθεξής. Έπειτα, εφαρμόζεται ξανά για το πρώτο βασικό καρέ, αλλά αναζητείται εκείνο που απέχει 4 μέτρα. Με αυτόν τον τρόπο χωρίζονται τα δεδομένα σε υποπεριοχές και είναι εμφανής η συμπεριφορά και η αλλαγή των σφαλμάτων σε διάφορα διαστήματα δεδομένων.

### 6.1 Αποτελέσματα εντοπισμού και εξαγωγής χαρακτηριστικών σημείων ORB

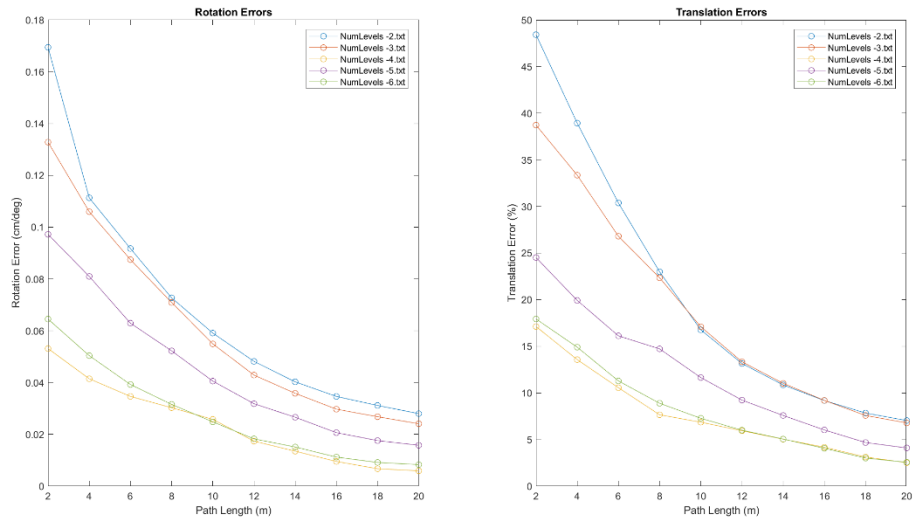
Αρχικά παρουσιάζονται τα αποτελέσματα του αλγορίθμου που προέκυψαν από τον εντοπισμό χαρακτηριστικών σημείων ORB σε κάθε καρέ. Πέρα από τον αλγόριθμο Visual SLAM, εντοπίστηκαν χαρακτηριστικά σημεία ORB και στην offline δημιουργία του bag-of-words.

Στο λογισμικό MATLAB ο αλγόριθμος ORB λαμβάνει δυο ορίσματα:

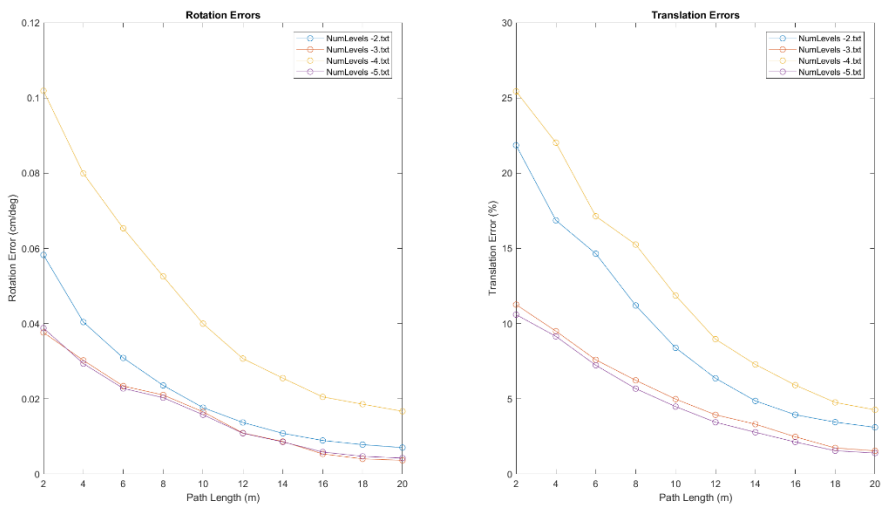
- Scale Factor: ορίζει το βήμα αναδειγματοληψίας της εικόνας στην πυραμίδα.
- Num Levels: ορίζει τον αριθμό των επιπέδων στην κλίμακα της πυραμίδας και εξαρτάται από το μέγεθος της εικόνας.

Πραγματοποιήθηκαν αρκετές δοκιμές μεταβολής των ορισμάτων αυτών με σκοπό να μελετηθεί κατά πόσο θα προκύψουν διαφορετικά αποτελέσματα. Στα ακόλουθα γραφήματα παρουσιάζονται τα αποτελέσματα τόσο για το γωνιακό όσο και για το τροχιακό σφάλμα.

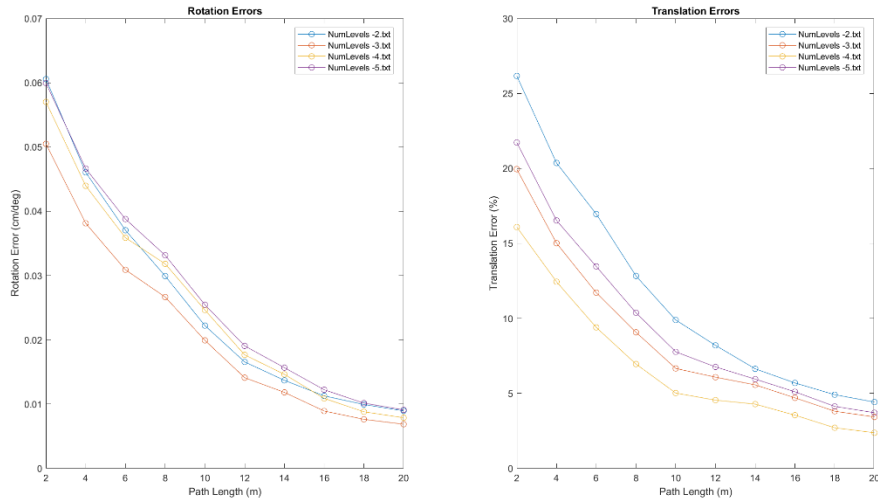
Στις ακόλουθες εικόνες παρουσιάζεται η διαφοροποίηση των τιμών στο γωνιακό σφάλμα (παρουσιάζεται στο αριστερό γράφημα) και στο τροχιακό σφάλμα (παρουσιάζεται στο διάγραμμα δεξιά) για διάφορες τιμές των μεταβλητών *Scale Factor* και *Num Levels*.



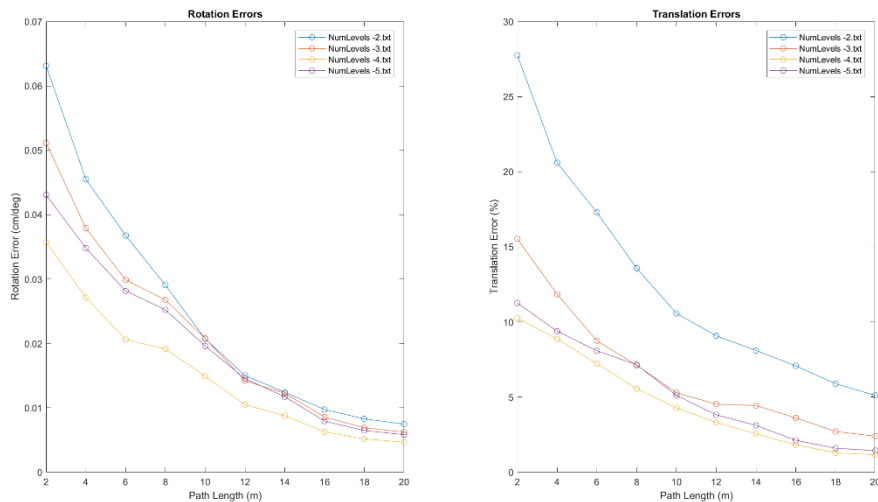
Εικόνα 31: Freiburg 2-Scale Factor παραμένει σταθερό και ίσο με 1.2 και αλλάζουν τα Num Levels.



Εικόνα 32: Freiburg 2-Scale Factor παραμένει σταθερό και ίσο με 1.3 και αλλάζουν τα Num Levels.



Εικόνα 33: Freiburg 2-Scale Factor παραμένει σταθερό και ίσο με 1.4 και αλλάζουν τα Num Levels.



Εικόνα 34: Freiburg 2-Scale Factor παραμένει σταθερό και ίσο με 1.5 και αλλάζουν τα Num Levels.

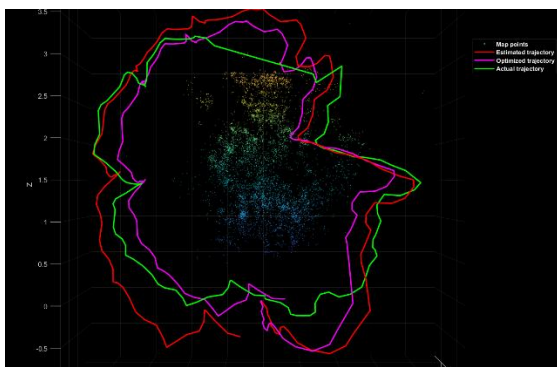
Μια αρχική παρατήρηση είναι πως οι τιμές του τροχιακού σφάλματος είναι αρκετά αυξημένες. Σε αρκετές δοκιμές εμφανίζεται αρκετά αυξημένο και το γωνιακό σφάλμα, όπως για παράδειγμα στην τιμή 1.2 του Scale Factor και Num Levels ίσο με 2 (Εικόνα 31), παρουσιάζεται σημαντικό γωνιακό σφάλμα, στα 2 μέτρα, που υπερβαίνει την τιμή 0.16 deg/cm και αντίστοιχα στο τροχιακό σφάλμα πλησιάζει την τιμή 50%. Στην Εικόνα 32 προκύπτει πως για την τιμή Scale Factor ίση με 1.3, τα καλύτερα αποτελέσματα επιφέρουν οι τιμές 3 και 5 της μεταβλητής Num Levels με το γωνιακό σφάλμα να είναι μικρότερο από 0.04 deg/cm στα 2 μέτρα, ενώ μειώνεται διαρκώς. Το τροχιακό σφάλμα αντιστοίχως για αυτές τις τιμές είναι κοντά στο 10% στα 2 μέτρα και μειώνεται όσο αυξάνεται η απόσταση.

Η τιμή Scale Factor 1.4 (Εικόνα 33) έχει καλύτερα αποτελέσματα ως προς το γωνιακό σφάλμα για την τιμή Num Levels ίση με 3 καθώς το σφάλμα είναι λίγο μεγαλύτερο από 0.05 deg/cm στα 2 μέτρα, ενώ στα 20 μέτρα πλησιάζει την τιμή 0.01 deg/cm. Το καλύτερο τροχιακό σφάλμα εμφανίζεται στην τιμή Num Levels ίση με 4, καθώς είναι

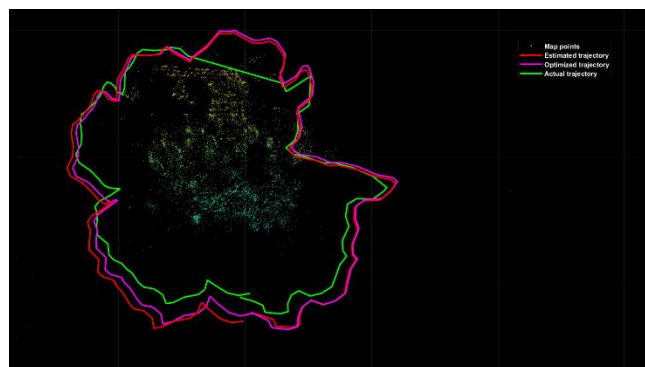


λίγο μεγαλύτερο από το 15%. Τέλος, για την μεταβλητή *Scale Factor* ίση με 1.5 (Εικόνα 34), το βέλτιστο τροχιακό και γωνιακό σφάλμα συναντώνται στην μεταβλητή *Num Levels* ίση με 4, όπου το γωνιακό σφάλμα είναι ανάμεσα 0.03 και 0.04 deg/cm ενώ το τροχιακό σφάλμα πλησιάζει το 10%. Στις περισσότερες περιπτώσεις, η τιμή της μεταβλητής *Num Levels* ίση με 2 έδωσε τα χειρότερα αποτελέσματα τόσο στο γωνιακό όσο και στο τροχιακό σφάλμα, με μόνη εξαίρεση την τιμή *Scale Factor* ίση με 1.3 (Εικόνα 32).

Στις παρακάτω εικόνες παρουσιάζεται ο τελικός χάρτης των σημείων αλλά και η τροχιά που ακολουθήθηκε αρχικά (κόκκινο), η βελτιστοποιημένη τροχιά μετά το κλείσιμο του βρόγχου και της συνόρθωσης του γράφου (μωβ) αλλά και η πραγματική τροχιά που ακολουθήθηκε (πράσινο), για τις τιμές *scale Factor* 1.3 (Εικόνα 35) και 1.5 (Εικόνα 36) και *num Levels* 4 και στις δυο περιπτώσεις.



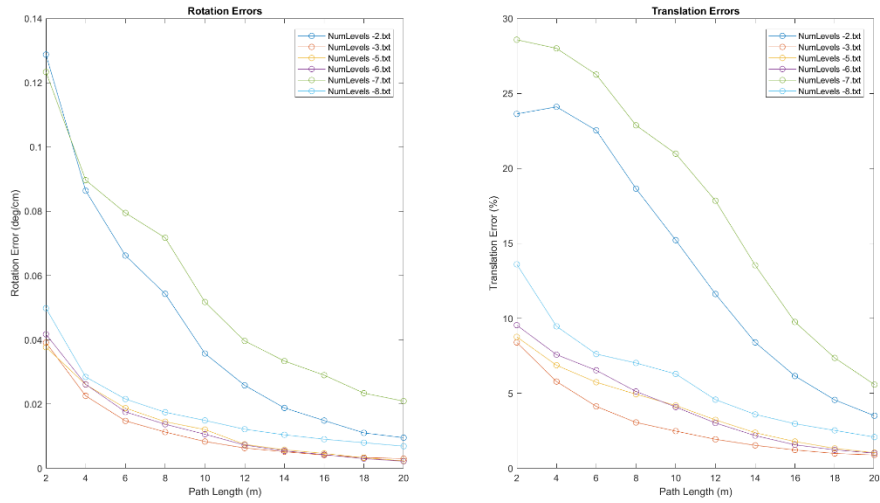
Εικόνα 35: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για *scale Factor* 1.3



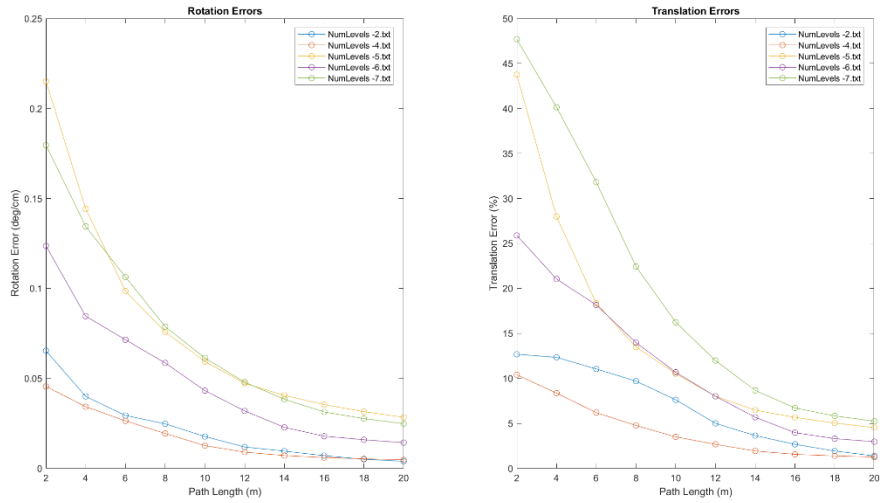
Εικόνα 36: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για *scale Factor* 1.5

Στην Εικόνα 35, παρατηρείται πως η βελτιστοποιημένη τροχιά αλλά και η υπολογισμένη τροχιά απέχουν αρκετά από την πραγματική. Αυτό είναι αναμενόμενο, δεδομένου πως στην Εικόνα 32 είναι εμφανές πως για την τιμή *Num Levels* ίση με 4, εμφανίζεται τροχιακό σφάλμα κοντά στο 25% στα δυο μέτρα. Στην Εικόνα 36, εμφανίζονται τα σημεία του χάρτη, και οι τροχιές για την τιμή *Scale Factor* 1.5 και *Num Levels* 4, όπου σε αυτή την περίπτωση το τροχιακό σφάλμα (Εικόνα 34) είναι ίσο με το 10% στα 2 μέτρα. Είναι οπτικά εμφανές πως η βελτιστοποιημένη τροχιά είναι πιο κοντά στην πραγματική στην Εικόνα 36 απ' ότι στην Εικόνα 35.

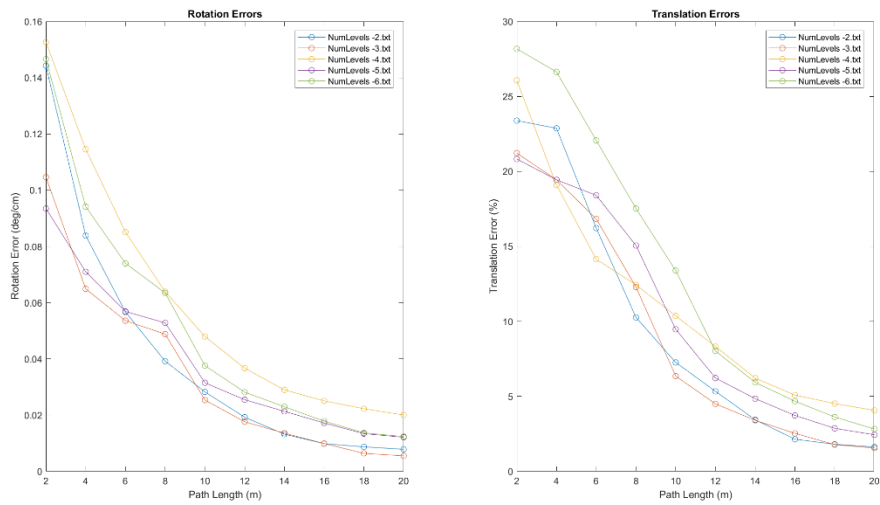
Αντίστοιχα για το σετ δεδομένων Freiburg 3 πραγματοποιήθηκαν επίσης δοκιμές με αλλαγές και στις δυο μεταβλητές (*Scale Factor* και *Num Levels*).



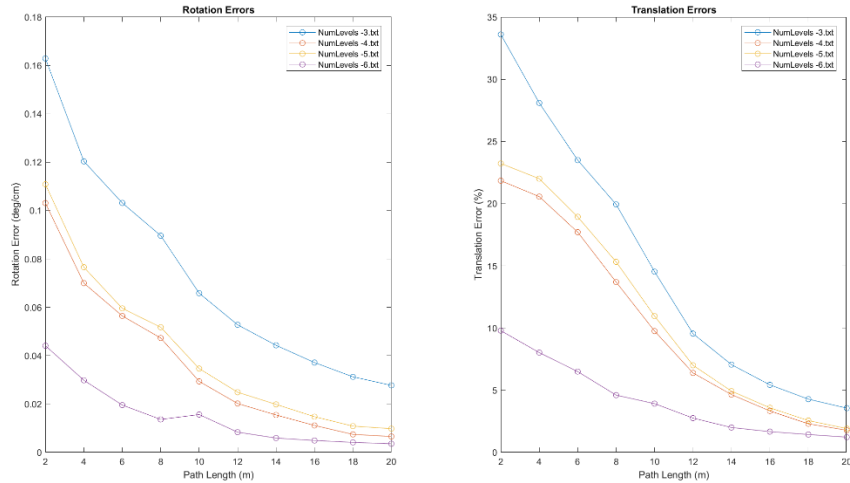
Εικόνα 37: Freiburg 3 - Scale Factor παραμένει σταθερό και ίσο με 1.2 και αλλάζουν τα Num Levels.



Εικόνα 38: Freiburg 3 - Scale Factor παραμένει σταθερό και ίσο με 1.3 και αλλάζουν τα Num Levels.



Εικόνα 39: Freiburg 3 - Scale Factor παραμένει σταθερό και ίσο με 1.4 και αλλάζουν τα Num Levels.



Εικόνα 40: Freiburg 3 - Scale Factor παραμένει σταθερό και ίσο με 1.5 και αλλάζουν τα Num Levels.

Και στην περίπτωση των δεδομένων Freiburg 3 παρατηρείται πως το τροχιακό σφάλμα σε αρκετές περιπτώσεις είναι σημαντικά αυξημένο, με τιμές στο 20 – 35%. Στην περίπτωση όπου οι τιμές Scale Factor και Num Levels είναι ίσες με 1.2 και 2 αντίστοιχα (Εικόνα 37), παρατηρείται πως τα σφάλματα είναι αυξημένα, δηλαδή το γωνιακό σφάλμα είναι λίγο μεγαλύτερο από την τιμή 1.2 deg/cm, ενώ το τροχιακό πλησιάζει το 25% στα 2 μέτρα. Η μεταβλητή Num Levels ίση με 3 (Εικόνα 37) έχει ως αποτέλεσμα στα 2 μέτρα το γωνιακό σφάλμα να είναι κοντά στην τιμή 0.04 deg/cm και το τροχιακό είναι μικρότερο από το 10%. Παρατηρείται πως για μεσαίες τιμές του Num Levels εμφανίζονται τα καλύτερα αποτελέσματα τόσο στο τροχιακό όσο και στο γωνιακό σφάλμα.

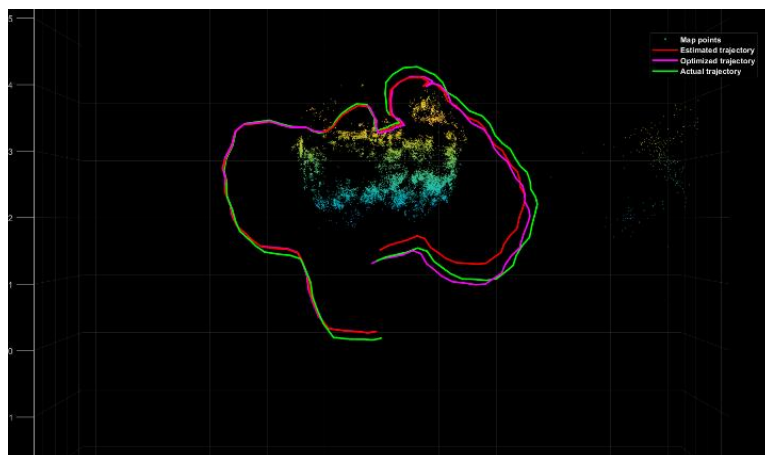
Για την μεταβλητή Scale Factor ίση με 1.3 (Εικόνα 38) και Num Levels ίση με 4 εμφανίζεται γωνιακό σφάλμα μικρότερο από 0.05 deg/cm ενώ το τροχιακό σφάλμα είναι στο 10%. Αν όμως η τιμή αυξηθεί κατά 1 (δηλαδή Num Levels ίσο με 5) εμφανίζεται αύξηση του σφάλματος, το γωνιακό είναι μεγαλύτερο από 0.2 deg/cm και το τροχιακό κοντά στο 45% στα 2 μέτρα.

Η μεταβλητή Scale Factor ίση με 1.4 (Εικόνα 39) εμφανίζει σχετικά αυξημένα σφάλματα για οποιαδήποτε τιμή της μεταβλητής Num Levels. Πιο συγκεκριμένα, στα 2 μέτρα το χαμηλότερο σφάλμα συναντάται για την μεταβλητή Num Levels ίση με 5, όπου το γωνιακό σφάλμα λαμβάνει είναι λίγο μικρότερο από 0.1 deg/cm, ενώ το τροχιακό σφάλμα είναι λίγο μεγαλύτερο από το 20%. Ακόμα, για την μεταβλητή Scale Factor ίση με 1.5 (Εικόνα 40), το καλύτερο γωνιακό και τροχιακό σφάλμα στα δύο μέτρα έχει η τιμή της μεταβλητής Num Levels ίση με 6. Το γωνιακό σφάλμα είναι λίγο μεγαλύτερο από 0.04 deg/cm, ενώ το τροχιακό είναι στο 10%. Αντίθετα, για την χαμηλότερη τιμή της Num Levels ίση με 3, το γωνιακό σφάλμα είναι μεγαλύτερο από 0.16 deg/cm ενώ το τροχιακό βρίσκεται κοντά στο 35%.

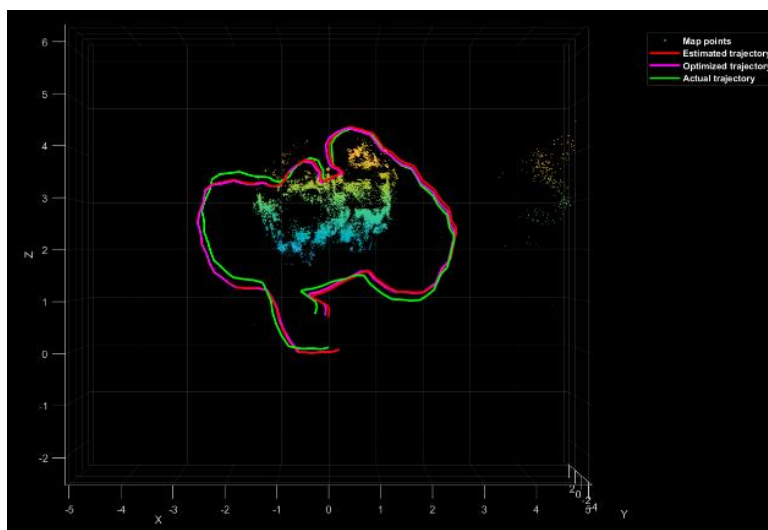
Εμφανίζονται ακολούθως τα τρισδιάστατα σημεία του χάρτη καθώς και η τροχιά της κάμερας που υπολογίστηκε αρχικά, η συνορθωμένη τροχιά και η πραγματική για τις τιμές:

- Scale Factor: 1.2 , Num Levels: 3 (Εικόνα 41).

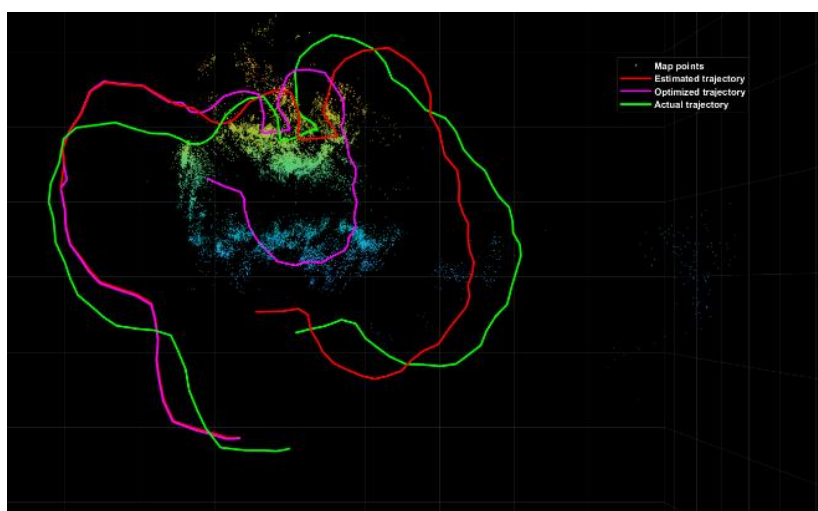
- *Scale Factor: 1.2 , Num Levels: 6* (Εικόνα 42).
- *Scale Factor: 1.3 , Num Levels: 7* (Εικόνα 43).



Εικόνα 41: Freiburg 3- Αποτελέσματα τροχιών και σημείων για scale Factor 1.2 num Levels 3



Εικόνα 42: Freiburg 3- Αποτελέσματα τροχιών και σημείων για scale Factor 1.2 num Levels 6



Εικόνα 43: Freiburg 3- Αποτελέσματα τροχιών και σημείων για scale Factor 1.3 num Levels 7

Στην [Εικόνα 43](#) είναι πράγματι εμφανές πως η τροχιά που υπολογίστηκε μετά την συνόρθωση του γράφου απέχει αρκετά από την πραγματική. Στην [Εικόνα 38](#) φάνηκε πως για τα 2 μέτρα το τροχιακό σφάλμα πλησίαζε το 50%. Η αρχική τροχιά ενώ εμφανίζει αρκετές αποκλίσεις, δεν παρουσιάζει τόσο σημαντικό πρόβλημα όσο εκείνη που υπολογίστηκε μετά την συνόρθωση του γράφου. Το πρόβλημα αυτό θα μπορούσε να υπάρξει λόγω λανθασμένης αντιστοίχισης των καρτέ που χρησιμοποιήθηκαν για το κλείσιμο του βρόγχου. Πιθανότατα οι ομολογίες να μην ήταν πράγματι οι σωστές και αυτό να επηρέασε την τροχιά των τελευταίων καρτέ. Αντιθέτως, στις άλλες δυο περιπτώσεις φαίνεται πως η υπολογισμένη τροχιά είναι αρκετά κοντά στην πραγματική.

## 6.2 Αποτελέσματα εντοπισμού και εξαγωγής χαρακτηριστικών σημείων SURF

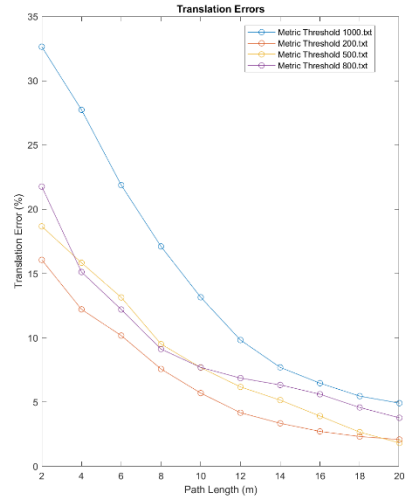
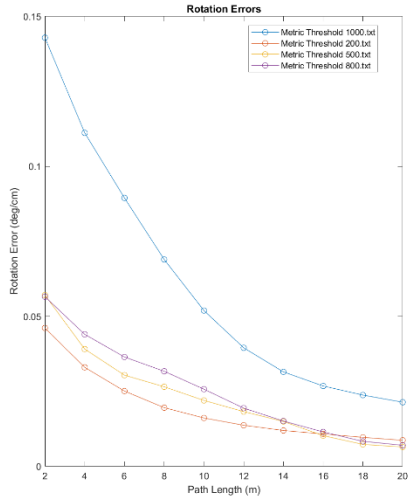
Στην συνέχεια, εφαρμόστηκε ο αλγόριθμος εντοπισμού και εξαγωγής χαρακτηριστικών σημείων SURF στα δεδομένα Freiburg 2 desk validation αλλά και στα δεδομένα Freiburg 3 long office. Στο MATLAB ο αλγόριθμος SURF περιλαμβάνει τις εξής μεταβλητές:

- **Metric Threshold:** Κατώφλι που χρησιμοποιείται για τον εντοπισμό σημείων. Τα σημεία που είναι κάτω από αυτό το κατώφλι απορρίπτονται.
- **Num Octaves:** Αριθμός των οκτάβων που θα χρησιμοποιηθούν στην πυραμίδα.
- **Num Scale Levels:** Ο αριθμός των επιπέδων της κλίμακας σε κάθε οκτάβα.

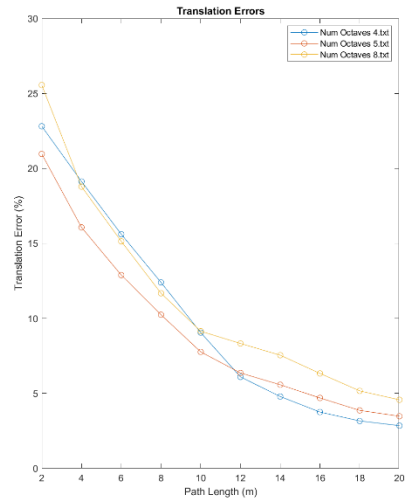
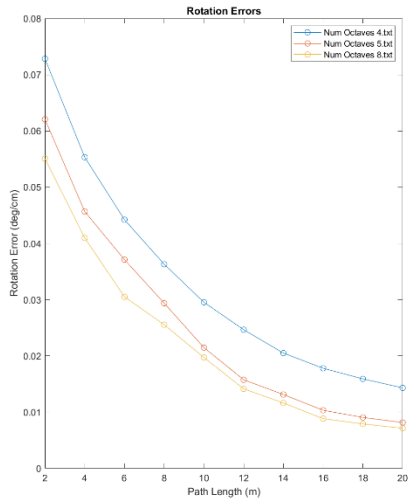
Πραγματοποιήθηκαν διαφορετικές δοκιμές για διαφορετικές τιμές των παραμέτρων αυτών. Οι μεταβολές των παραμέτρων πραγματοποιήθηκαν διατηρώντας σταθερές τις υπόλοιπες μεταβλητές. Δηλαδή, αλλάζοντας την τιμή της μεταβλητής *Metric Threshold* οι τιμές *Num Octaves* και *Num Scale Levels* παρέμειναν σταθερές. Αντίστοιχα, στην αλλαγή της παραμέτρου *Num Octaves*, επιλέχθηκαν σταθερές τιμές για τις υπόλοιπες παραμέτρους. Μόνο στην αλλαγή των *Num Scale Levels* οι παράμετροι δεν έμειναν σταθερές διότι ο αλγόριθμος δεν μπορούσε να πραγματοποιήσει τον εντοπισμό της θέσης με σωστό τρόπο.

. Η αλλαγή των παραμέτρων έγινε ως εξής:

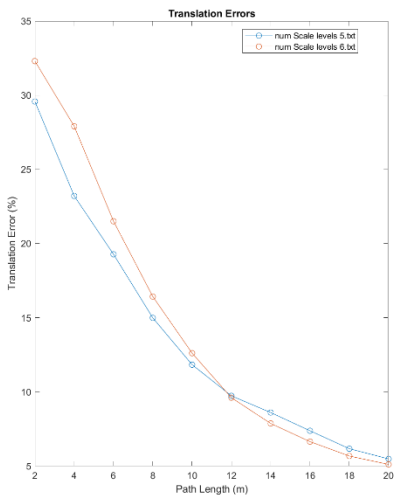
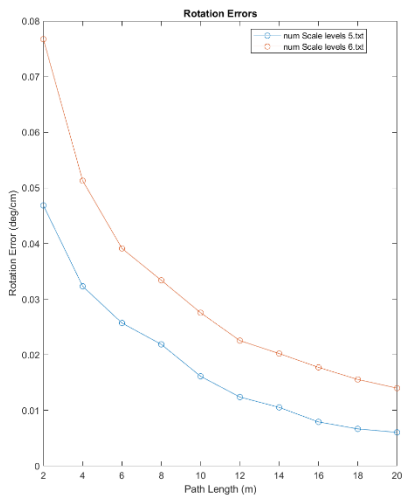
- **Metric Threshold:** Διατηρώντας παραμέτρους *Num Octaves*=3 και *Num Scale Levels* = 5 σταθερές.
- **Num Octaves:** Διατηρώντας παραμέτρους *Metric Threshold*=200.0 και *Num Scale Levels* = 5 σταθερές.
- **Num Scale Levels:** στην πρώτη περίπτωση *Metric Threshold*=800.0 και *Num Octaves* = 3 και στην δεύτερη *Metric Threshold*=200.0 και *Num Octaves* = 4.



Εικόνα 44: Freiburg 2-Αλλαγή στο Metric Threshold.



Εικόνα 45: Freiburg 2-Αλλαγή στο Num Octaves.

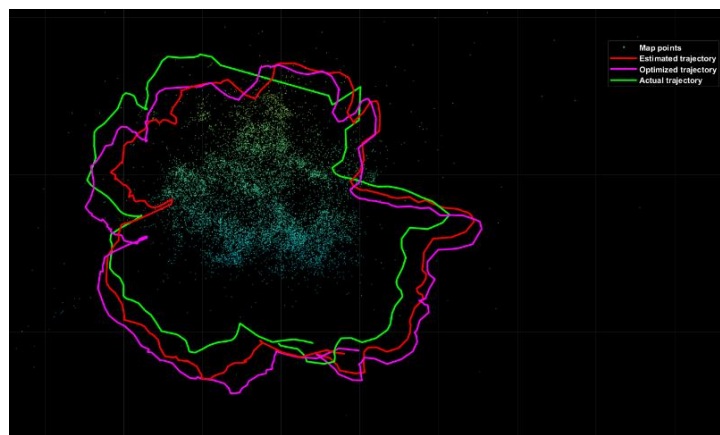


Εικόνα 46: Freiburg 2-Αλλαγή στο Num Scale Levels.

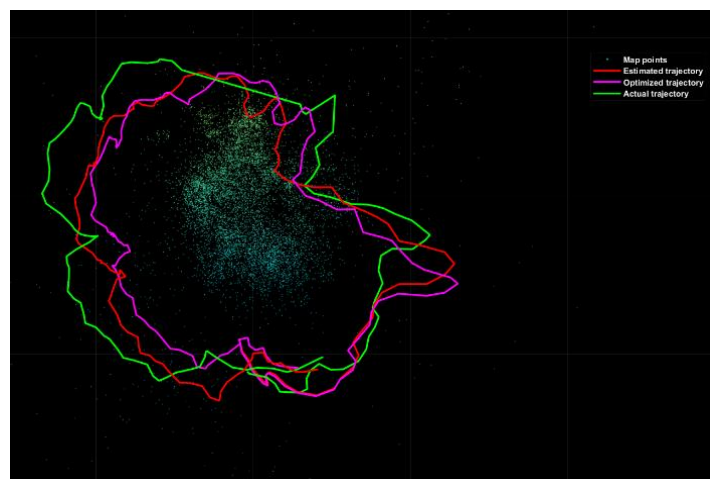
Παρατηρώντας την [Εικόνα 44](#), είναι εμφανές πως αυξάνοντας το *Metric Threshold* αυξάνονται αντίστοιχα τα γωνιακά και τροχιακά σφάλματα, ενώ τα καλύτερα αποτελέσματα λαμβάνονται από την μικρότερη τιμή που δοκιμάστηκε, δηλαδή ίσο με 200.0. Για εκείνη την τιμή της μεταβλητής *Metric Threshold* το γωνιακό σφάλμα έχει την τιμή 0.005 deg/cm, ενώ το τροχιακό είναι στο 15% στα 2 μέτρα. Για την τιμή του *Metric Threshold* ίση με 1000.0 το γωνιακό σφάλμα πλησιάζει τα 0.15 deg/cm και το τροχιακό βρίσκεται ανάμεσα στο 30-35% στα 2 μέτρα.

Αντίστοιχα η *Num Octaves* ([Εικόνα 45](#)) παρέχει καλύτερα αποτελέσματα για την τιμή 8 όπου το γωνιακό σφάλμα είναι ανάμεσα στα 0.05-0.06 deg/cm και το τροχιακό σφάλμα είναι ανάμεσα στο 20-25% στα 2 μέτρα. Η μεταβλητή *num Scale Levels* ([Εικόνα 46](#)) δεν ήταν εφικτό να δοκιμαστούν πολλές τιμές και παρατηρείται πως, στα 2 μέτρα, η τιμή 5 έχει καλύτερα αποτελέσματα από την τιμή 6.

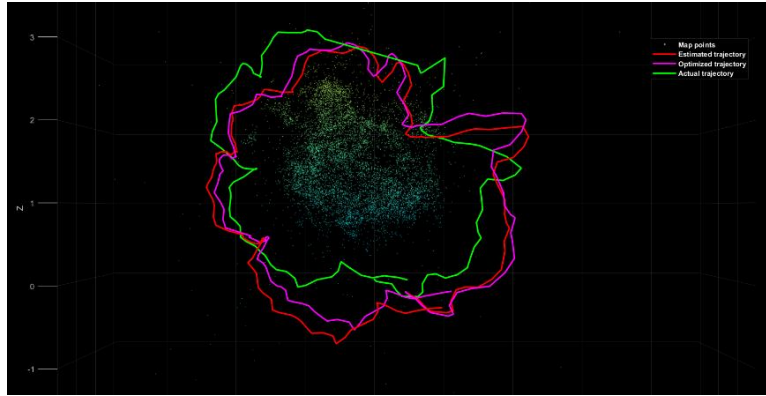
Παρακάτω παρουσιάζονται τα σημεία του χάρτη, η υπολογισμένη, η βελτιστοποιημένη και η πραγματική τροχιά για μερικές δοκιμές.



Εικόνα 47: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για *Metric Threshold* 500.



Εικόνα 48: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για *Num Octaves* 4.



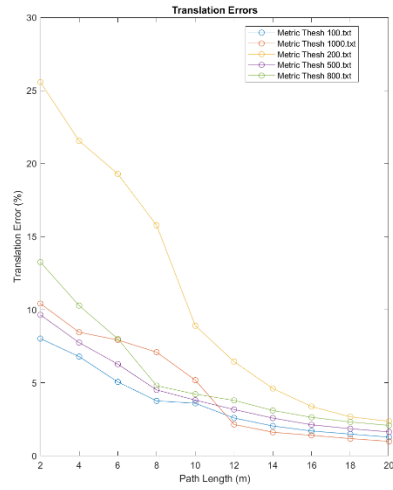
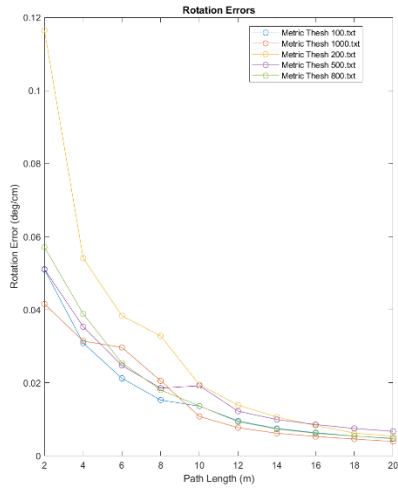
Εικόνα 49: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Num Scale Levels 5.

Και στις τρεις περιπτώσεις, το τροχιακό σφάλμα κυμαίνεται στις τιμές 20-35% στα 2 μέτρα. Πιο συγκεκριμένα, η Εικόνα 47 αφορά την τιμή της μεταβλητής *Metric Threshold* ίση με 500.0, όπου το τροχιακό σφάλμα (στα 2 μέτρα) είναι ανάμεσα στο 30-35%. Στην Εικόνα 48 παρουσιάζεται η τροχιά και τα σημεία του χάρτη για την τιμή 4 της μεταβλητής *Num Octaves* όπου το τροχιακό σφάλμα (στα 2 μέτρα) είναι ανάμεσα στο 20-25%. Τέλος, η Εικόνα 49 αφορά την τιμή 5 της μεταβλητής *Num Scale Levels* όπου το τροχιακό σφάλμα είναι 30%.

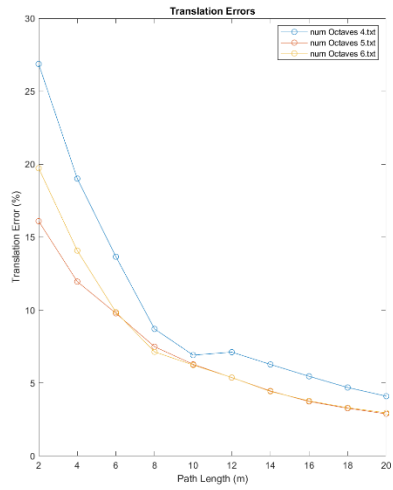
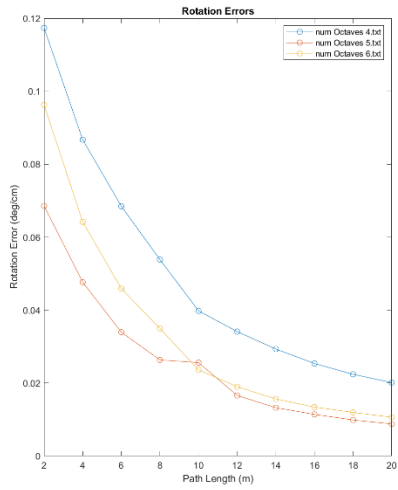
Όπως αναφέρθηκε προηγουμένως, πραγματοποιήθηκαν διαφορετικές δοκιμές για μια συγκεκριμένη μεταβλητή κάθε φορά, διατηρώντας σταθερές τις υπόλοιπες μεταβλητές. Και σε αυτή την περίπτωση η αλλαγή της μεταβλητής *Num Scale Levels* δεν πραγματοποιήθηκε για σταθερές παραμέτρους κάθε φορά. Πραγματοποιήθηκαν, λοιπόν, οι εξής αλλαγές:

- *Metric Threshold*: Διατηρώντας παραμέτρους *num Octaves*=3 και *num Scale Levels* = 4 σταθερές.
- *Num Octaves*: Διατηρώντας παραμέτρους *Metric Threshold*=1000.0 και *num Scale Levels* = 4 σταθερές.
- *Num Scale Levels*: στην πρώτη περίπτωση και στην τελευταία *Metric Threshold*=1000.0 ενώ στις δυο ενδιάμεσες *Metric Threshold*=800.0, ενώ το *num Octaves*=3

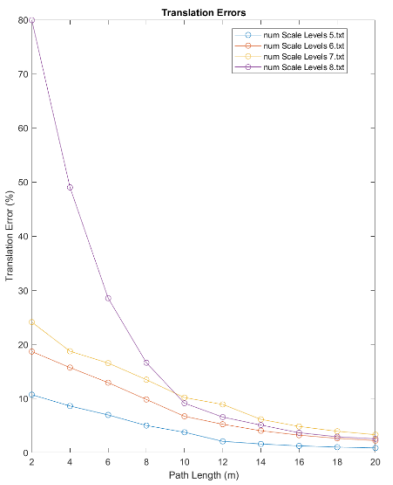
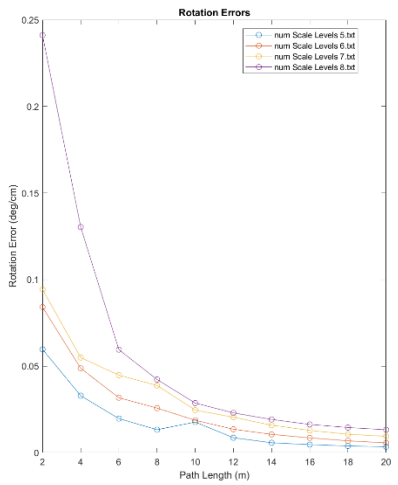




Εικόνα 50: Freiburg 3-Αλλαγή στο Metric Threshold..



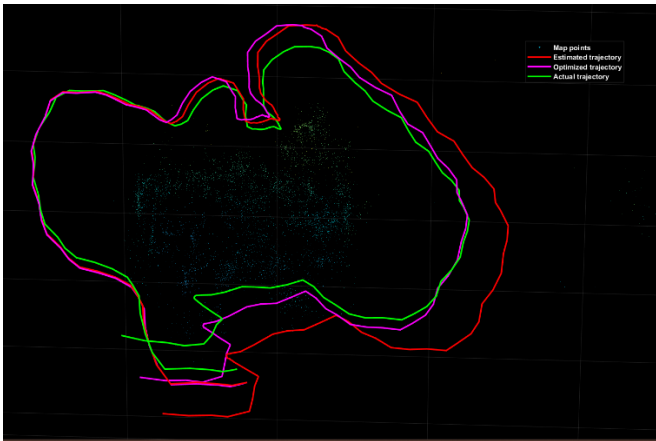
Εικόνα 51: Freiburg 3-Αλλαγή στο Num Octaves.



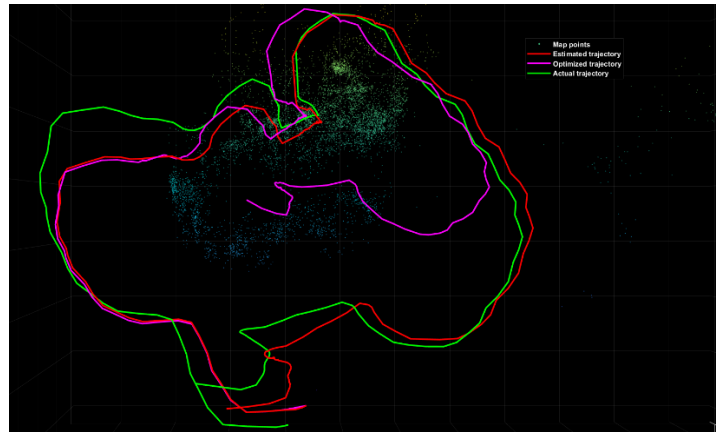
Εικόνα 52: Freiburg 3-Αλλαγή στο Num Scale Levels.

Παρατηρείται πως για μικρότερες τιμές της μεταβλητής *Metric Threshold* (Εικόνα 50), εμφανίζονται μικρότερα σφάλματα, με μόνη εξαίρεση την τιμή 200.0 όπου εμφανίζονται ακραίες τιμές σφάλματος τόσο στις γωνίες (μεγαλύτερο από 0.12 deg/cm) όσο και στην τροχιά (πάνω από 25%). Η τιμή 1000.0 έχει αρκετά καλό αποτέλεσμα ως προς το γωνιακό σφάλμα, κοντά στα 0.04 deg/cm στα 2 μέτρα, έπειτα εμφανίζει μια αύξηση κοντά στα 6 μέτρα η οποία μειώνεται αρκετά στα 10 μέτρα. Η συγκεκριμένη τιμή έχει αρκετά καλά αποτελέσματα και ως προς το τροχιακό σφάλμα όπου στα 2 μέτρα βρίσκεται κοντά στο 10%.

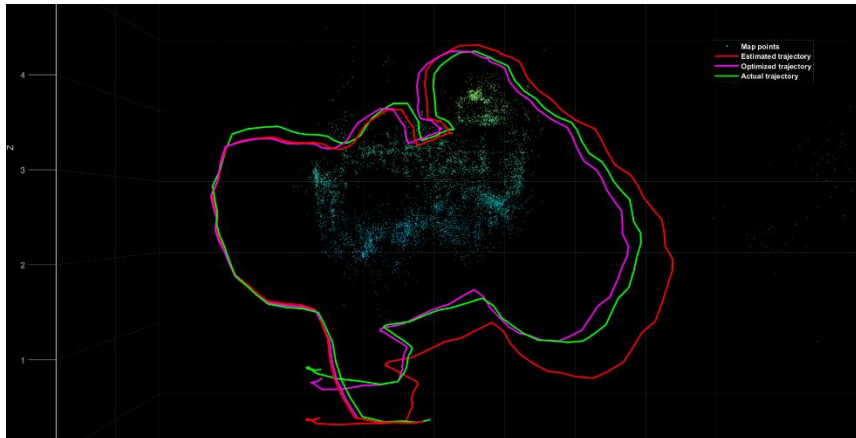
Ως προς την μεταβλητή *Num Octaves* (Εικόνα 51) παρατηρείται πως η καλύτερη τιμή είναι ίση με 5 όπου το γωνιακό σφάλμα είναι ανάμεσα στα 0.06-0.08 deg/cm και το τροχιακό σφάλμα, στα 2 μέτρα, βρίσκεται κοντά στο 15%. Αξίζει ακόμα να παρατηρηθεί πως το τροχιακό σφάλμα των τιμών 5 και 6 από τα 10 μέτρα και μετά συνταυτίζεται. Παρατηρώντας την Εικόνα 52 είναι εμφανές πως όσο αυξάνεται η τιμή της μεταβλητής *Num Scale Levels*, τόσο αυξάνεται και το σφάλμα. Πιο συγκεκριμένα, για την μεγαλύτερη τιμή της μεταβλητής, ίση με 8, συναντάται γωνιακό σφάλμα κοντά στο 0.25 deg/cm και τροχιακό σφάλμα κοντά στο 80%. Παρ' όλα αυτά η τιμή 5 εμφανίζει αρκετά καλά αποτελέσματα τόσο ως προς το γωνιακό (περίπου 0.05 deg/cm) όσο και ως προς το τροχιακό (10%).



Εικόνα 53: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για *Metric Threshold* 500.



Εικόνα 54: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για *num Octaves* 4



Εικόνα 55: Freiburg 2- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Num Scale Levels 5.

Και σε αυτή την περίπτωση παρατηρείται πως υπάρχει μια τιμή (*Num Octaves* ίσο με 4) που ενώ η αρχική τροχιά είναι σχετικά κοντά με την πραγματική, η τροχιά μετά την συνόρθωση έχει απομακρυνθεί αρκετά (Εικόνα 54). Αντιθέτως, η τροχιά του αποτελέσματος του σετ δεδομένων Freiburg 2 (Εικόνα 48) για την ίδια μεταβλητή num Octaves δεν έχει τόσο μεγάλη απόκλιση από την πραγματική. Για τις υπόλοιπες παραμέτρους φαίνεται πως η τροχιά δεν απέχει αρκετά από την πραγματική.

### 6.3 Αποτελέσματα εντοπισμού και εξαγωγής χαρακτηριστικών σημείων SIFT

Σε αυτή την δοκιμή, ο αλγόριθμος Visual SLAM πραγματοποιήθηκε με εξαγωγή χαρακτηριστικών σημείων μέσω του αλγορίθμου SIFT. Φυσικά, αυτό αφορά τόσο την εξαγωγή χαρακτηριστικών κατά την διάρκεια του αλγορίθμου αλλά και την δημιουργία του οπτικού λεξιλογίου με περιγραφείς SIFT.

Στο λογισμικό MATLAB ο αλγόριθμος SIFT λαμβάνει τις εξής παραμέτρους:

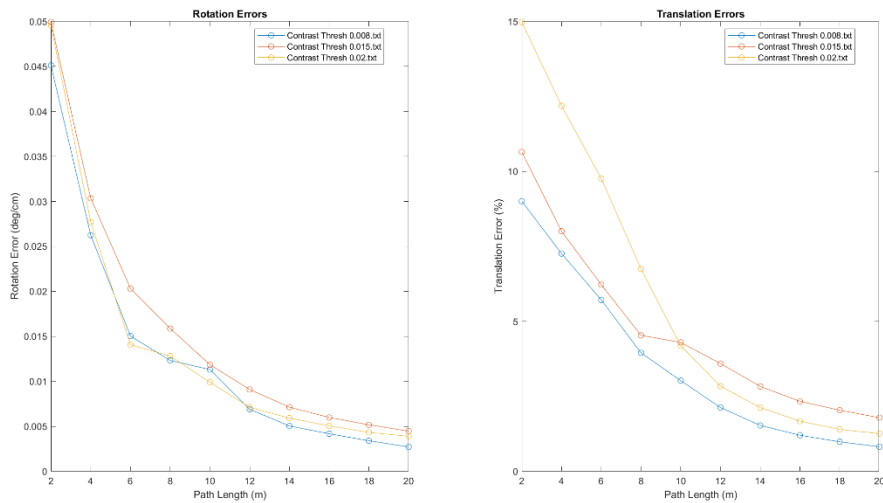
- Contrast threshold: ελάχιστη αντίθεση ώστε ένα σημείο να θεωρηθεί ως χαρακτηριστικό.
- Edge Threshold: μέτρο απόρριψης χαρακτηριστικών ανάλογα με την κλίση τους.
- Num Layers in Octave: ο αριθμός των επιπέδων που ανήκουν σε μια οκτάβα για την δημιουργία των επιπέδων κλίμακας.
- Sigma: Τυπική απόκλιση του φίλτρου Gauss που εφαρμόζεται στο επίπεδο των κλιμάκων.

Ένα πρόβλημα που δημιουργήθηκε κατά την εφαρμογή του Visual SLAM με χαρακτηριστικά σημεία SIFT ήταν πως για το σετ δεδομένων Freiburg 2 desk validation, ο εντοπισμός της θέσης και του προσανατολισμού της κάμερας τερματίζονταν στα πρώτα βασικά καρέ και δεν ήταν εφικτό να ολοκληρωθεί ο κώδικας. Παρ' όλα αυτά για το σετ δεδομένων Freiburg 3 long office household ο αλγόριθμος ήταν εφικτό να πραγματοποιηθεί.

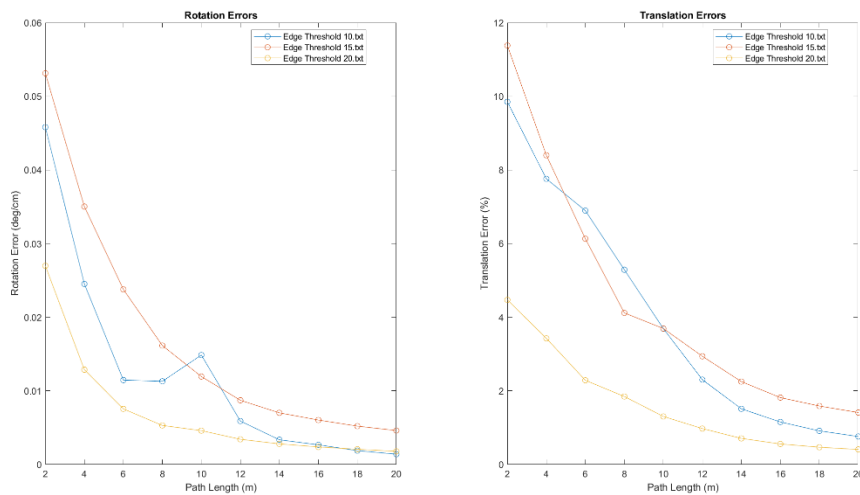
Όπως στον αλγόριθμο SURF, έτσι και για τον αλγόριθμο SIFT πραγματοποιήθηκαν διαφορετικές δοκιμές για διαφορετικές τιμές των παραμέτρων. Αλλάζοντας την μια παράμετρο οι υπόλοιπες διατηρήθηκαν σταθερές.

Οι αλλαγές που πραγματοποιήθηκαν και τα αποτελέσματα που προέκυψαν παρουσιάζονται ακολούθως.

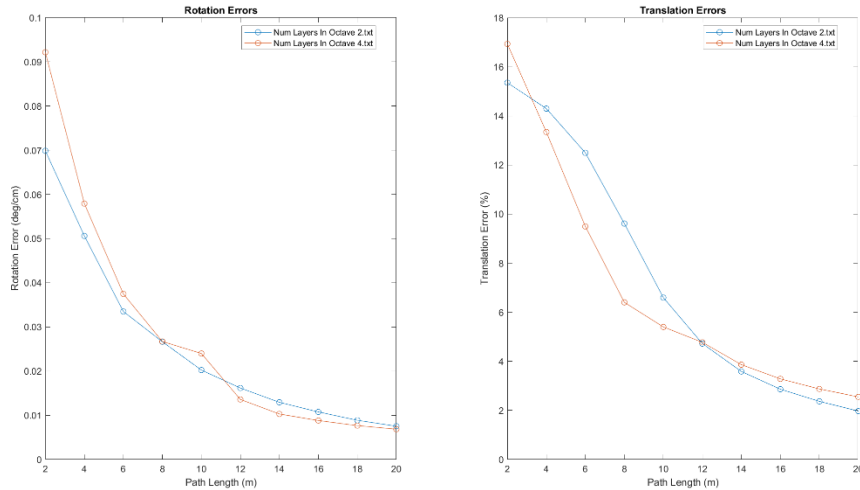
- Contrast threshold: Οι ακόλουθες παράμετροι διατηρήθηκαν σταθερές: Edge Threshold = 20.0, Num Layers In Octave = 3, Sigma = 1.6
- Edge Threshold: Οι ακόλουθες παράμετροι διατηρήθηκαν σταθερές: Contrast Threshold = 0.0133, Num Layers In Octave = 3, Sigma = 1.6
- Num Layers in Octave: Οι ακόλουθες παράμετροι διατηρήθηκαν σταθερές: Contrast Threshold = 0.0133, Edge Threshold = 20.0, Sigma = 1.6
- Sigma: Οι ακόλουθες παράμετροι διατηρήθηκαν σταθερές Contrast Threshold = 0.008, Edge Threshold = 20.0, Num Layers In Octave = 3



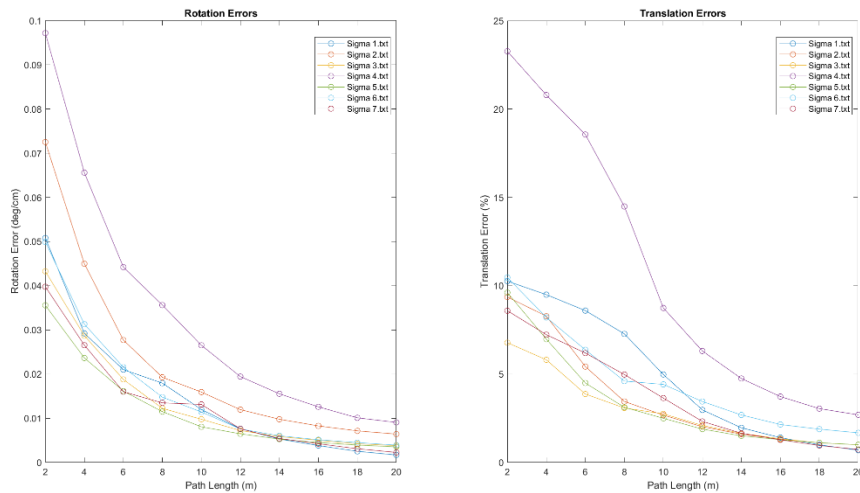
Εικόνα 56: Freiburg 3 – Αλλαγή στο Contrast Threshold.



Εικόνα 57: Freiburg 3 – Αλλαγή στο Edge Threshold.



Εικόνα 58: Freiburg 3 – Αλλαγή στο Num Layers In Octave.



Εικόνα 59: Freiburg 3 – Αλλαγή στο Sigma.

Από τα αποτελέσματα στην [Εικόνα 56](#) προκύπτει πως όσο μικρότερο είναι το *Contrast Threshold* τόσο βελτιώνονται τα αποτελέσματα στα σφάλματα στροφής και τροχιάς, ενώ το αντίθετο παρατηρείται για την μεταβλητή *Edge Threshold* ([Εικόνα 57](#)). Πιο συγκεκριμένα, για την τιμή *Contrast Threshold* ίση με 0.008 παρατηρείται, στα 2 μέτρα, γωνιακό σφάλμα κοντά στο 0.045 deg/cm και τροχιακό σφάλμα μικρότερο από το 10%. Για την μεταβλητή *Edge Threshold* η μεγαλύτερη τιμή (ίση με 20.0) στα 2 μέτρα παρουσιάζει γωνιακό σφάλμα μικρότερο του 0.03 deg/cm και αντίστοιχα το τροχιακό σφάλμα είναι κοντά στο 4%.

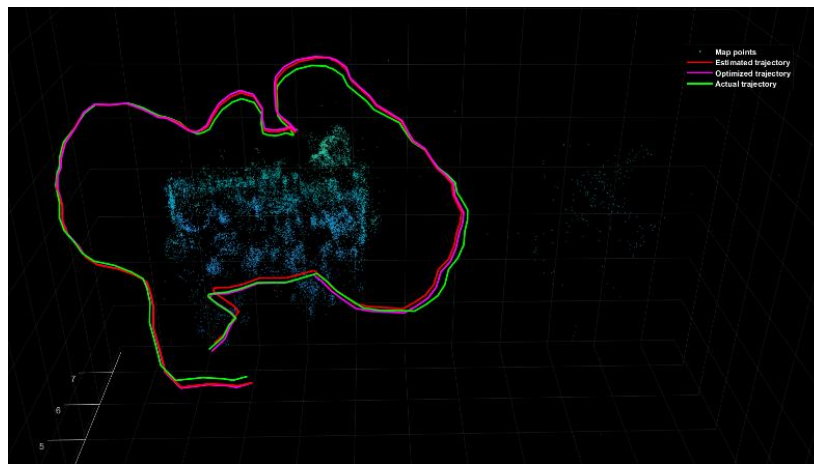
Σχετικά με την παράμετρο *Num Layers In Octave* ([Εικόνα 58](#)) παρατηρείται πως η τιμή 2 παρουσιάζει αρκετά ικανοποιητικά γωνιακά και τροχιακά σφάλματα, όπου στα 2 μέτρα το γωνιακό σφάλμα είναι κοντά στην τιμή 0.07 deg/cm και το τροχιακό είναι ανάμεσα στο 14 με 16%, ενώ στα επόμενα διαστήματα μειώνονται.

Ακόμα, η αλλαγή της παραμέτρου *Sigma* ([Εικόνα 59](#)) δεν ακολουθεί κάποια διαβαθμισμένη διαφοροποίηση, δηλαδή δεν εμφανίζονται καλύτερες τιμές με την

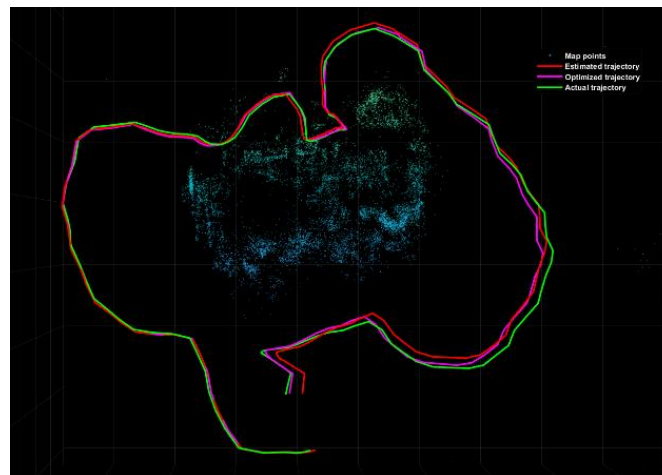
αύξηση ή την μείωση της. Τα καλύτερα αποτελέσματα για το γωνιακό σφάλμα φαίνεται να εμφανίζονται για την τιμή του Sigma ίση με 5, όπου το γωνιακό σφάλμα βρίσκεται ανάμεσα στα 0.03-0.04 deg/cm. Για την ίδια τιμή της παραμέτρου, το τροχιακό σφάλμα στα 2 μέτρα είναι λίγο χαμηλότερο από το 10%, ενώ μειώνεται στα επόμενα υποδιαστήματα. Τόσο ως προς το γωνιακό όσο ως προς το τροχιακό, το μεγαλύτερο σφάλμα εμφανίζεται στην τιμή Sigma ίση με 4, όπου στα δυο μέτρα το γωνιακό σφάλμα βρίσκεται κοντά στα 0.1 deg/cm και το τροχιακό κοντά στο 25%.

Παρόλα αυτά, είναι εμφανές πως οι περισσότερες τιμές της μεταβλητής *Sigma* επέφεραν αρκετά ικανοποιητικά αποτελέσματα ως προς τα σφάλματα.

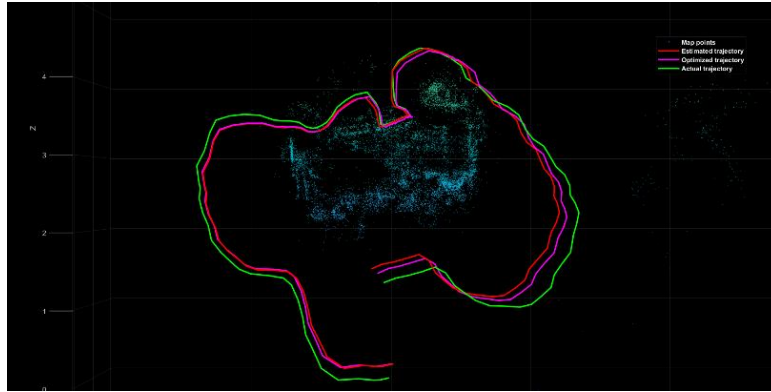
Στις ακόλουθες εικόνες παρουσιάζονται ο χάρτης των τρισδιάστατων σημείων, η υπολογισμένη, η αρχική και η πραγματική τροχιά.



Εικόνα 60: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Contrast Threshold 0.008



Εικόνα 61: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για Edge Threshold 20.0.



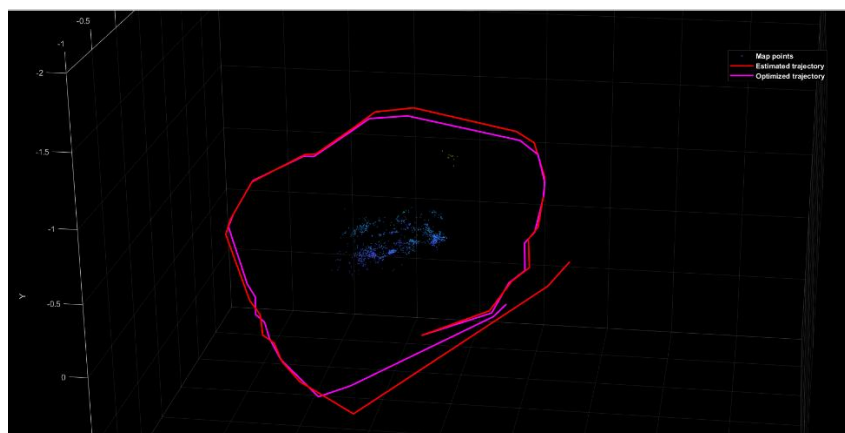
Εικόνα 62: Freiburg 3- Σύγκριση αποτελεσμάτων αρχικής, βελτιστοποιημένης και πραγματικής τροχιάς για  $\Sigma$  1.7

Στην [Εικόνα 60](#) εμφανίζονται τα σημεία του χάρτη και η τροχιάς για την μεταβλητή *Contrast Threshold* ίση με 0.08, όπου φαίνεται και στην [Εικόνα 56](#) εμφανίζει τροχιακό σφάλμα κοντά στο 10% στα 2 μέτρα, ενώ μειώνεται στα υπόλοιπα υποδιαστήματα. Στην [Εικόνα 61](#) φαίνονται τα σημεία του χάρτη και οι τροχιάς για την μεταβλητή *Edge Threshold* ίση με 20.0, όπου όπως έχει ήδη αναφερθεί έχει σφάλμα κοντά στο 4% στα 2 μέτρα. Η [Εικόνα 62](#) παρουσιάζει τα τρισδιάστατα σημεία που προέκυψαν, καθώς και τις τροχιάς της κάμερας για την τιμή 1.7 της μεταβλητής  $\Sigma$ . Είναι εμφανές από την [Εικόνα 59](#) πως το τροχιακό σφάλμα, στα δύο μέτρα, βρίσκεται κοντά στο 10%, ενώ στα υπόλοιπα υποδιαστήματα μειώνεται.

Γενικώς συμπεραίνεται πως ο αλγόριθμος SIFT αποδίδει αρκετά ικανοποιητικά αποτελέσματα τόσο ως προς το γωνιακό σφάλμα όσο και ως προς το τροχιακό σφάλμα. Ελάχιστες μεταβολές των παραμέτρων επέφεραν σημαντικά γωνιακά και τροχιακά σφάλματα.

#### 6.4 Αποτέλεσμα τροχιάς και τρισδιάστατου χάρτη στο σετ δεδομένων που δημιουργήθηκε

Ο αλγόριθμος Visual SLAM του MATLAB εφαρμόστηκε με χαρακτηριστικά σημεία ORB και για το σετ δεδομένων που απεικονίζει αντικείμενα πάνω σε τραπέζι και το αποτέλεσμα της τροχιάς αλλά και των σημείων στον τρισδιάστατο χάρτη παρουσιάζονται ακολούθως.

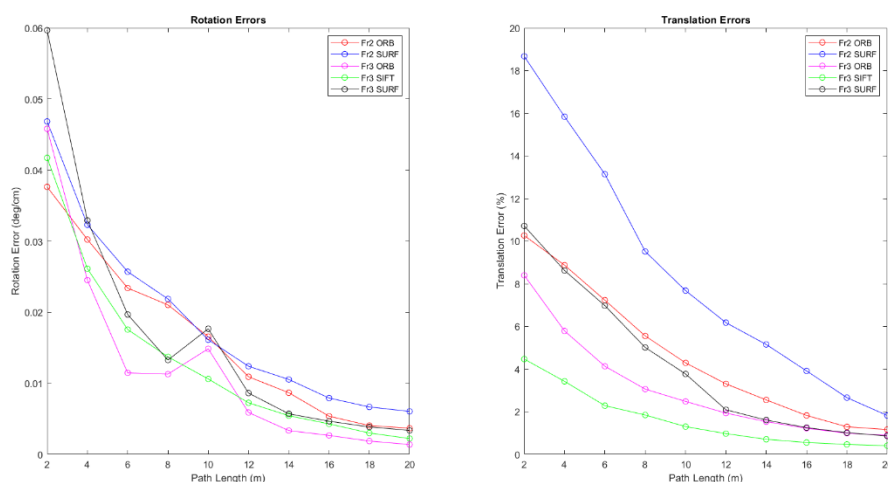


Εικόνα 63: Αποτέλεσμα αλγορίθμου Visual SLAM στο σετ δεδομένων που απεικονίζει τα αντικείμενα στο τραπέζι.

Παρατηρείται πως η σύγκριση της αρχικής τροχιάς με την βελτιστοποιημένη είναι αρκετά κοντά στο μεγαλύτερο μέρος της τροχιάς. Τα αντικείμενα που απεικονίζονται δεν είναι πολλά, όπως στα σετ δεδομένων Freiburg 2 και 3, γι' αυτόν τον λόγο ο χάρτης σημείων εμφανίζεται αρκετά αραιός. Η κάμερα ήταν προσανατολισμένη με τέτοιο τρόπο, ώστε το τραπέζι να βρίσκεται σχεδόν σε όλη την έκταση των καρτέ. Γι' αυτόν τον λόγο δεν εμφανίζονται και πολλά σημεία που να αποτελούν ακραίες αποκλίσεις (outliers) από το κυρίως απεικονιζόμενο αντικείμενο (το τραπέζι με τα αντικείμενα).

## 6.5 Γενικά συμπεράσματα για τις διαφορετικές μεθόδους εντοπισμού και εξαγωγής χαρακτηριστικών σημείων

Τα βέλτιστα αποτελέσματα για τα σετ δεδομένων Freiburg 2 και Freiburg 3 για όλες τις μεθόδους εξαγωγής χαρακτηριστικών σημείων από όλες τις δοκιμές εμφανίζονται στην [Εικόνα 64](#).



Εικόνα 64: Ελάχιστες τιμές σφαλμάτων από όλες τις δοκιμές και στα δυο σετ δεδομένων.

Είναι εμφανές πως στο γωνιακό και στο τροχιακό σφάλμα για τα δεδομένα Freiburg 2 ο αλγόριθμος ORB έχει καλύτερα αποτελέσματα, ενώ για το σετ δεδομένων Freiburg 3 ο αλγόριθμος SIFT επιφέρει, με διαφορά, τα βέλτιστα αποτελέσματα.

Παρατηρήθηκε ακόμα, πως ο εντοπισμός στο σετ δεδομένων Freiburg 2 δεν είναι τόσο ακριβής όσο στο σετ δεδομένων Freiburg 3, καθώς εμφανίζεται συστηματική απόκλιση τόσο της αρχικής όσο και της βελτιστοποιημένης τροχιάς από την πραγματική.

Άλλος ένας προβληματισμός ως προς τα δυο σετ δεδομένων είναι η μη δυνατότητα πραγματοποίησης του εντοπισμού της θέσης και του προσανατολισμού της κάμερας με τα χαρακτηριστικά σημεία SIFT για το σετ δεδομένων Freiburg 2. Αντιθέτως όταν εφαρμόστηκε ο αλγόριθμος SIFT στο σετ δεδομένων Freiburg 3, λήφθηκαν αποτελέσματα και μάλιστα με τα χαμηλότερα σφάλματα.

Τα αποτελέσματα, όπως προκύπτει από τις προηγούμενες δοκιμές, είναι αρκετά ακριβή και δεν αποκλίνουν σημαντικά από τις πραγματικές τροχιές. Απαιτεί όμως προσοχή κατά την συνόρθωση του γράφου, καθώς αν υπάρχουν προβληματικά δεδομένα, είτε στην βάση που χρησιμοποιείται για το κλείσιμο των βρόχων είτε στα βασικά καρτέ, μπορεί να δημιουργήσει πολύ σημαντικές αποκλίσεις της βελτιστοποιημένης τροχιάς τόσο από την αρχική όσο και από την πραγματική τροχιά.



## 7 Συμπεράσματα

Στην παρούσα διπλωματική εργασία πραγματοποιήθηκε μια εκτενής βιβλιογραφική ανασκόπηση των αλγορίθμων Visual SLAM, καθώς και των διαφορετικών τεχνικών που μπορούν να χρησιμοποιηθούν. Μέσω της βιβλιογραφικής αναζήτησης έγινε κατανοητή η θεωρητική πλευρά ενός προβλήματος Visual SLAM. Επιπλέον, παρουσιάστηκε μια εφαρμογή Visual SLAM βασισμένη σε έναν αρκετά γνωστό αλγόριθμο, τον ORB – SLAM των Mur-Artal et al. (2015). Κατά την εφαρμογή αυτήν, κατανοήθηκε η πρακτική πλευρά ενός προβλήματος Visual SLAM. Ο σκοπός, λοιπόν, της εργασίας ήταν η πλήρης κατανόηση του Visual SLAM τόσο θεωρητικά όσο και πρακτικά.

Για την επίτευξη του μελετήθηκαν αρκετές δημοσιεύσεις σχετικά με την εφαρμογή Visual SLAM, από τις οποίες φάνηκε πως το πρόβλημα του Visual SLAM δεν έχει μια λύση. Η λύση που θα επιλέξει ο καθένας εξαρτάται από τις ανάγκες της εφαρμογής του. Παρουσιάστηκαν αλγόριθμοι που χρησιμοποιούν φίλτρα αλλά και αλγόριθμοι που χρησιμοποιούν βασικά καρέ ενώ μελετήθηκε η διαφορά μεταξύ των δυο μεθόδων.

Έπειτα, επεξηγήθηκαν οι βασικές διαδικασίες που συναντώνται στους περισσότερους αλγορίθμους Visual SLAM, οι οποίες είναι: ο εντοπισμός, η χαρτογράφηση και το κλείσιμο των βρόχων. Ως προς τον εντοπισμό, παρουσιάστηκαν τόσο οι μέθοδοι αναγνώρισης σκηνης, όπως το Bag-of-words όσο και μερικοί αλγόριθμοι της βιβλιογραφίας όπου χρησιμοποιούνται τέτοιες μέθοδοι. Επιπλέον, αναφέρθηκε το πρόβλημα του επανεντοπισμού, το οποίο συναντάται σε αρκετούς αλγορίθμους, καθώς ο εντοπισμός μπορεί να χαθεί σε πολλές εφαρμογές πραγματικού χρόνου.

Δεδομένου πως ο αλγόριθμος Monocular Visual SLAM του MATLAB που εφαρμόστηκε βασίζεται στον αλγόριθμο ORB-SLAM, συντελέστηκε μια σύντομη αναφορά στον συγκεκριμένο αλγόριθμο, ο οποίος έχει τέσσερις βασικές επιμέρους διαδικασίες: την αρχικοποίηση του χάρτη, τον εντοπισμό, την χαρτογράφηση και το κλείσιμο των βρόχων.

Επεξηγήθηκε ο αλγόριθμος που εφαρμόστηκε, όπου χρησιμοποιεί τις ίδιες επιμέρους διαδικασίες με τον αλγόριθμο ORB-SLAM. Εφαρμόστηκε για τρία διαφορετικά σετ δεδομένων που αφορούσαν αντικείμενα σε γραφείο/τραπέζι. Αρχικά, εφαρμόστηκε ο αλγόριθμος με χαρακτηριστικά σημεία ORB μεταβάλλοντας τις παραμέτρους εισόδου, ώστε να ληφθούν διαφορετικά σημεία σε κάθε περίπτωση. έπειτα, εφαρμόστηκε εντοπισμός και συνταύτιση χαρακτηριστικών σημείων με τους αλγορίθμους SURF και SIFT, με σκοπό να μελετηθεί η συμπεριφορά του αλγορίθμου και το πως επηρεάζεται το αποτέλεσμα σε σχέση με τα χαρακτηριστικά σημεία.

Δημιουργήθηκε μια συνάρτηση όπου υπολογίζεται το γωνιακό και το τροχιακό σφάλμα, βασισμένη σε ήδη υπάρχουσα εφαρμογή. Με αυτόν τον τρόπο πραγματοποιήθηκε σύγκριση των διαφορετικών αποτελεσμάτων που προκύπτουν από τους αλγορίθμους εντοπισμού και συνταύτισης χαρακτηριστικών σημείων.

Με την ολοκλήρωση της παρούσας εργασίας συμπεραίνεται πως η διαδικασία του ταυτόχρονου εντοπισμού θέσης και χαρτογράφησης με μια κάμερα (Monocular Visual SLAM) μπορεί να παρέχει αρκετά ικανοποιητικά αποτελέσματα σε ότι αφορά το σφάλμα της τροχιάς αλλά και του προσανατολισμού της κάμερας. Δεδομένου ότι οι

θέσεις και οι προσανατολισμοί της κάμερας που υπολογίστηκαν είναι πολύ κοντά στα πραγματικά δεδομένα, μπορεί να θεωρηθεί με βεβαιότητα πως τα τρισδιάστατα σημεία που προκύπτουν μπορούν να περιγράψουν αρκετά καλά το μοντέλο του χώρου.

Όπως έγινε αντιληπτό στα προηγούμενα κεφάλαια, για να εφαρμοστεί η διαδικασία του Visual SLAM, είτε με μια κάμερα είτε με ζεύγος, είναι απαραίτητες οι γνώσεις των βασικών φωτογραμμετρικών διαδικασιών οι οποίες συμβάλλουν στην ανάκτηση τρισδιάστατης πληροφορίας από εικόνες έχοντας υπόψιν τις γεωμετρικές δεσμεύσεις του μοντέλου στον χώρο.

Μελετήθηκαν αρκετές δημοσιεύσεις σχετικά με την εφαρμογή Visual SLAM, από τις οποίες φάνηκε πως το πρόβλημα του Visual SLAM δεν έχει μια λύση. Η λύση που θα επιλέξει ο καθένας εξαρτάται από τις ανάγκες της εφαρμογής του.

Ο αλγόριθμος του MATLAB που χρησιμοποιήθηκε είναι ένα αρκετά καλό παράδειγμα εφαρμογής Monocular Visual SLAM όπου αλλάζοντας διάφορες παραμέτρους, μπορεί να παρατηρηθεί η διαφορά στα αποτελέσματα. Δεδομένου ότι ο αλγόριθμος χρησιμοποιεί μόνο την οπτική πληροφορία (εικόνες) είναι αναμενόμενο να βασίζεται αρκετά στα χαρακτηριστικά σημεία που εντοπίζονται. Βασίζεται τόσο ως προς την ποσότητα των σημείων αυτών όσο και ως προς τον αλγόριθμο εντοπισμού και εξαγωγής τους. Διαφορετικοί αλγόριθμοι σημείων παρέχουν και διαφορετικά αποτελέσματα του χάρτη των τρισδιάστατων σημείων αλλά και της πορείας της κάμερας (θέση και προσανατολισμός).

Τέλος, έγινε κατανοητό πως δεν επηρεάζουν μόνο οι αλγόριθμοι χαρακτηριστικών σημείων την πορεία του αλγορίθμου, αλλά εξαρτάται και τα σημεία της σκηνής στην πραγματικότητα. Ένας λειτουργικός αλγόριθμος για ένα σετ δεδομένων δεν εγγυάται την ίδια επιτυχία σε διαφορετικά δεδομένα.

## Βιβλιογραφία

---

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded up robust features. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part I (Vol. 9)*. Springer Berlin Heidelberg.

Brown, D.C., & Fryer, J.G. (1986). Lens distortion for close-range photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 52(1), pp. 51-58.

Castellanos, J. A., Neira, J., & Tardós, J. D. (2004). Limits to the consistency of EKF-based SLAM. *IFAC Proceedings Volumes*, 37(8), 716-721.

Cheng, J., Zhang, L., Chen, Q., Hu, X., & Cai, J. (2022). A review of visual SLAM methods for autonomous driving vehicles. *Engineering Applications of Artificial Intelligence*, 114, 104992,

Conrady, A. (1919). Decentered Lens Systems, *Monthly Notices of the Royal Astronomical Society*, Vol. 79, pp. 384- 390.

Cummins, M., & Newman, P. (2008). FAB-MAP: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6), pp. 647-665.

Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping: Part I. *IEEE Robotics & Automation Magazine*, 13(2), pp. 99-110.

Engel, J., Schöps, T., & Cremers, D. (2014, September). LSD-SLAM: Large-scale direct monocular SLAM. *European Conference on Computer Vision*. Cham: Springer International Publishing, pp. 834-849

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp. 381-395.

Forster, C., Pizzoli, M., & Scaramuzza, D. (2014). SVO: Fast semi-direct monocular visual odometry. *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 15-22

Galvez-Lopez, D., & Tardos, J. D. (2011). Real-time loop detection with bags of binary words. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. pp. 51-58.

Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*. IEEE. pp. 3354-3361

Grisetti, G., Kümmerle, R., Stachniss, C., & Burgard, W. (2010). A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2(4), pp. 31-43.

Grisetti, G., Stachniss, C., & Burgard, W. (2007). Fast and accurate SLAM with Rao-Blackwellized particle filters. *Robotics and Autonomous Systems*, 55(1), pp. 30-38.

- Harris, C., & Stephens, M. (1988). A combined corner and edge detector. In *Alvey Vision Conference* (Vol. 15, No. 50), pp. 10-5244.
- Hartley, R., & Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press.
- Heikkilä, J., & Silvén, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1106-1112
- Jégou, H., Douze, M., Schmid, C., & Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 3304-3311.
- Kazerouni, I. A., Fitzgerald, L., Dooly, G., & Toal, D. (2022). A survey of state-of-the-art on visual SLAM. *Expert Systems with Applications*, 205, 117734
- Klein, G., & Murray, D. (2007). Parallel tracking and mapping for small AR workspaces. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. IEEE, pp. 225-234
- Klein, G., & Murray, D. (2008). Improving the agility of keyframe-based SLAM. In *Computer Vision—ECCV 2008: 10th European Conference on Computer Vision*, Marseille, France, October 12-18, 2008, Proceedings, Part II (Vol. 10). Springer Berlin Heidelberg.
- Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., & Burgard, W. (2011). g2o: A general framework for graph optimization. In *2011 IEEE International Conference on Robotics and Automation*. IEEE. pp. 3607-3613
- Leonard, J. J., & Durrant-Whyte, H. F. (1991). Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(3), 376-382.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2), pp. 164-168.
- Li, S., & Yoon, H.S. (2023). Vehicle localization in 3D world coordinates using single camera at traffic intersection. *Sensors*, 23(7), 3661, pp. 1-15
- Lourakis, M. I. A., & Argyros, A. A. (2009). SBA: A software package for generic sparse bundle adjustment. *ACM Transactions on Mathematical Software (TOMS)*, 36(1), pp. 1-30.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE International Conference on Computer Vision*, Vol. 2. IEEE. pp. 1150-1157
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, pp. 91-110.
- Lu, F., & Milios, E. (1997). Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, 4, pp. 333-349.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2), pp. 431-441.

Mathworks, Monocular Visual Simultaneous Localization and Mapping, Example. Ανακτήθηκε από: <https://www.mathworks.com/help/vision/ug/monocular-visual-simultaneous-localization-and-mapping.html>

MathWorks. Using the Camera Calibrator App. Ανακτήθηκε από <https://www.mathworks.com/help/vision/ug/using-the-single-camera-calibrator-app.html>

Montemerlo, M., Thrun, S., Koller, D., & Wegbreit, B. (2002). FastSLAM: A factored solution to the simultaneous localization and mapping problem. *Proceedings of the AAAI/IAAI Conference*, pp. 593-598.

Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, 31(5), pp. 1147-1163.

Murphy, K. P. (1999). Bayesian map learning in dynamic environments. *Advances in Neural Information Processing Systems*, 12, pp. 1015-1021.

Neale, W. T., Hessel, D., & Terpstra, T. (2011). Photogrammetric measurement error associated with lens distortion (No. 2011-01-0286). SAE Technical Paper.

Newcombe, R. A., Lovegrove, S. J., & Davison, A. J. (2011). DTAM: Dense tracking and mapping in real-time. In *2011 International Conference on Computer Vision*. IEEE. pp. 2320-2327

Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6), pp. 756-770.

Nister, D., & Stewenius, H. (2006). Scalable recognition with a vocabulary tree. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2. IEEE, pp. 2161-2168

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*. IEEE. pp. 2564-2571.

Scaramuzza, D., & Fraundorfer, F. (2011). Visual odometry [tutorial]. *IEEE Robotics & Automation Magazine*, 18(4), pp. 80-92.

Scaramuzza, D., Martinelli, A., & Siegwart, R. (2006). A toolbox for easily calibrating omnidirectional cameras. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 5695-5701

Sivic, J., & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. *Proceedings of the Ninth IEEE International Conference on Computer Vision*. IEEE, pp. 1470-1477

Smith, R., Self, M., & Cheeseman, P. (1990). Estimating uncertain spatial relationships in robotics. In I.J. Cox & G.T. Wilfong (Eds.), *Autonomous Robot Vehicles*. New York: Springer-Verlag. pp. 167–193.

Stachniss, C. (2013/14). SLAM-Course - 01 - Introduction to Robot Mapping [Video]. YouTube. <https://www.youtube.com/watch?v=wVsfCnyt5jA>

Strasdat, H., Montiel, J. M. M., & Davison, A. J. (2012). Visual SLAM: Why filter? *Image and Vision Computing*, 30(2), pp. 65-77.

Technical University of Munich. (2011, September 30). Sequence Freiburg 2 desk. Last modified: 15:17. Retrieved from <https://cvg.cit.tum.de/data/datasets/rgbd-dataset/download>

Technical University of Munich. (2012, August 7). Sequence Freiburg 3 long office household. Last modified: 18:51. Retrieved from <https://cvg.cit.tum.de/data/datasets/rgbd-dataset/download>

Triggs, B., McLauchlan, P. F., Hartley, R. I., & Fitzgibbon, A. W. (2000). Bundle adjustment—a modern synthesis. *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms*, Corfu, Greece, September 21–22, 1999 Proceedings. Springer Berlin Heidelberg. pp. 298-372

Urban, S., Leitloff, J., & Hinz, S. (2015). Improved wide-angle, fisheye and omnidirectional camera calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, pp. 72-79.

Williams, B., Klein, G., Reid, I., & Sturm, J. (2009). A comparison of loop closing techniques in monocular SLAM. *Robotics and Autonomous Systems*, 57(12), pp. 1188-1197.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), pp. 1330-1334.

Γραμματικόπουλος, Λ., Καλησπεράκης, Η., Καρράς, Γ., Πέτσα, Ε., Τσιρώνης, Β. 2018, “Σημειώσεις Προβολικής Γεωμετρίας (για την Φωτογραμμετρία και την Όραση Υπολογιστών). [Πανεπιστημιακές Σημειώσεις], Αθήνα: Πανεπιστήμιο Δυτικής Αττικής (ΠΑΔΑ), Σχολή Μηχανικών, Τμήμα Τοπογραφίας και Γεωπληροφορικής.

Καλησπεράκης, Η. 2010 “Η Επιπολική Γεωμετρία στον Προβολικό και στον Ευκλείδειο χώρο”. [Διδακτορική Διατριβή], Εθνικό Μετσόβιο Πολυτεχνείο (ΕΜΠ). Σχολή Αγρονόμων και Τοπογράφων Μηχανικών (ΣΑΤΜ).

Νάκης, Π., 2019 “Ανάπτυξη και Αξιολόγηση Αλγορίθμου Οπτικής Οδομετρίας/SLAM”. [Μεταπτυχιακή Διπλωματική Εργασία], Πανεπιστήμιο Δυτικής Αττικής (ΠΑΔΑ). Σχολή Μηχανικών, Τμήμα Μηχανικών Τοπογραφίας και Γεωπληροφορικής.

Πέτσα, Ε. 2000 “Θεμελιώδεις Έννοιες και Θεμελιώδη Προβλήματα Της Φωτογραμμετρίας”, [Πανεπιστημιακές Σημειώσεις], Αθήνα: ΤΕΙ Αθήνας Τμήμα Τοπογραφίας.