



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

Υλοποίηση αλγορίθμου εντοπισμού θέσης
Ρομπότ μέσω οπτικής αναγνώρισης στατικών
αντικειμένων.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΑΡΙΣΤΕΙΔΗ ΣΜΑΡΝΑΚΗ

Επιβλέπων: Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

ΕΡΓΑΣΤΗΡΙΟ ΔΙΑΧΕΙΡΙΣΗΣ ΚΑΙ ΒΕΛΤΙΣΤΟΥ ΣΧΕΔΙΑΣΜΟΥ ΔΙΚΤΥΩΝ ΤΗΛΕΜΑΤΙΚΗΣ
Αθήνα, Ιούλιος 2020



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής
Εργαστήριο Διαχείρισης και Βέλτιστου Σχεδιασμού Δικτύων Τηλεματικής

Υλοποίηση αλγορίθμου εντοπισμού θέσης
Ρομπότ μέσω οπτικής αναγνώρισης στατικών
αντικειμένων.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΑΡΙΣΤΕΙΔΗ ΣΜΑΡΝΑΚΗ

Επιβλέπων: Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 24η Ιουλίου 2020.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Συμεών Παπαβασιλείου
Καθηγητής Ε.Μ.Π.

.....
Θεοδώρα Βαρβαρίγου
Καθηγήτρια Ε.Μ.Π.

.....
Ιωάννα Ρουσσάκη
Επίκ. Καθηγήτρια Ε.Μ.Π.

Αθήνα, Ιούλιος 2020

(Υπογραφή)

.....
ΑΡΙΣΤΕΙΔΗΣ ΣΜΑΡΝΑΚΗΣ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

© 2020 – All rights reserved



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Επικοινωνιών, Ηλεκτρονικής και Συστημάτων Πληροφορικής
Εργαστήριο Διαχείρισης και Βέλτιστου Σχεδιασμού Δικτύων Τηλεματικής

Copyright ©–All rights reserved Αριστείδης Σμαρνάκης, 2020.

Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Περίληψη

Τα τελευταία χρόνια έχει πραγματοποιηθεί ραγδαία πρόοδος στον τομέα της ρομποτικής και των ρομποτικών εφαρμογών τόσο σε βιομηχανικό όσο και οικιακό επίπεδο. Μια κατηγορία ρομποτικών μηχανισμών που εμφανίζει όλο και περισσότερη δυναμική στους κύκλους των ερευνητών είναι τα κινούμενα ρομπότ. Μέσω αυτών πραγματοποιείται μια πληθώρα εφαρμογών που έχουν ως κορμό την αυτόνομη πλοήγηση της πλατφόρμας ώστε να εν συνεχεία να δυνατή η εκτέλεση πληθώρας εφαρμογών όπως η επιτήρηση χώρων ή η μετακίνηση βαρέων φορτίων. Προκειμένου να επιτευχθούν τέτοιου είδους διεργασίες είναι απαραίτητο να είναι γνωστή η θέση του ρομπότ ως προς το περιβάλλον του κατά την διάρκεια της κίνησης αυτού. Σκοπός αυτής της διπλωματικής εργασίας είναι η υλοποίηση και αξιολόγηση μιας υβριδικής μεθόδου εντοπισμού θέσης του ρομπότ για εσωτερικούς χώρους η οποία βασίζεται στην οπτική αναγνώριση στατικών αντικειμένων. Η μέθοδος κάνει χρήση αλγορίθμων αναγνώρισης αντικειμένων μέσω βαθιάς μάθησης και όρασης υπολογιστών προκειμένου να αναγνωρίσει τα αντικείμενα και τις συντεταγμένες τους ώστε τελικά μέσω αυτών να υπολογίσει την απόλυτη θέση του ρομπότ εντός του χώρου εργασίας του. Ακολούθως της ανάπτυξης της μεθόδου ακολουθεί πειραματική αξιολόγηση τα αποτελέσματα της οποία καταδεικνύουν τις σημαντικές δυνατότητες της μεθόδου και καθώς και πιθανές περιοχές που επιδέχονται βελτίωση.

Λέξεις Κλειδιά

Ρομποτική, Εντοπισμός Θέσης, Βαθιά Μάθηση, Συνελικτικό Νευρωνικό Δίκτυο, Όραση Υπολογιστών, Επεξεργασία Εικόνων, Αναγνώριση Αντικειμένων, Εξαγωγή Χαρακτηριστικών, Εξισώσεις Συγγραμμικότητας, Φωτογραμμετρική Οπισθοτομία, Μεθοδος Ελαχίστων Τετραγώνων.

Abstract

For the past decades there is an increasing interest in the field of robotics which has lead to the development of innovative solutions to several everyday or industrial problems using robots. One of the most common categories used are the mobile robots. The main task performed by these robots is the one called autonomous navigation. Having this as their core task, mobile robot are used for a wide range of applications varying from surveillance to transportation of heavy cargo. However, in order to achieve any of the above, the robot must be aware of its absolute position with respond to its environment at any given moment. This problem is called robot localisation. The purpose of this diploma thesis is to implement and then test a hybrid indoor robot localisation method which utilises static object detection via camera. This method consist of three main parts. Firstly, it uses a deep learning algorithm for object detection in order to identify static objects in a picture of the robots surroundings, secondly it extract the coordinates of said objects and finally it calculates the absolute position of the robot. After the development phase, an experiment was conducted in order to test and evaluate the method. The results show the method to be quit accurate but they also exploit several of its weaknesses.

Keywords

Robotics, Mobile Robots, Robot Localisation, Deep Learning, Convolutional Neural Network - CNN, Faster R-CNN algorithm, ResNet101, Computer Vision, Image Processing, Object Detection, Feature Extraction, SIFT, RANSAC, Collinearity Equations, Space Resection, Least Squares Approximation.

Ευχαριστίες

Με την εκπόνηση της παρούσας εργασίας κλείνει ένα κεφάλαιο της ζωής μου το οποίο ήταν γεμάτο όμορφες στιγμές αλλά και δύσκολες προκλήσεις. Μέσα απο τα χρόνια μου εντός του πολυτεχνείου έκανα γνωριμίες οι οποίες με εξέλιξαν ως άτομο και ειδικά χάρη στην διπλωματική μου πέρασα από την σφαίρα του φοιτητή σε αυτή του μηχανικού.

Η παρούσα διπλωματική εργασία εκπονήθηκε το ακαδημαϊκό έτος 2019-2020 υπό την επίβλεψη του κυρίου Συμεών Παπαβασιλείου, καθηγητή της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσοβίου Πολυτεχνείου, τον οποίο και ευχαριστώ που μου έδωσε την δυνατότητα να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα. Επιπλέον θα ήθελα να ευχαριστήσω ιδιαίτερος τους Δημήτρη Δεχουνιώτη, Δημήτρη Σπαθαράκη και Μάριο Αυγέρη για την διαρκή βοήθεια και την εξαιρετική συνεργασία που είχαμε κατα την διάρκεια εκπόνησης της εργασίας μου παρά το γεγονός ότι συνέπεσε στην διάρκεια της πρωτοφανούς πανδημίας του COVID-19.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου που βρισκόταν πάντα δίπλα μου καθώς και τους φίλους μου τόσο εντός όσο και εκτός της σχολής, και ιδιαίτερα τους κολλητούς μου, Γιάννη και Αλέξη.

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	5
Περιεχόμενα	8
Κατάλογος Σχημάτων	11
Κατάλογος Πινάκων	13
1 Εισαγωγή	15
1.1 Εντοπισμός Θέσης	16
1.1.1 Πλαίσια Αναφοράς	16
1.1.2 Τυπικά Προβλήματα	17
1.2 Αντικείμενο διπλωματικής	18
1.2.1 Το πρόβλημα	18
1.2.2 Συνεισφορά	19
1.3 Οργάνωση Τόμου	19
2 ΒΙΒΛΙΟΓΡΑΦΙΑ	21
2.1 Ο εντοπισμός θέσης σε εσωτερικούς χώρους.	21
2.2 Εντοπισμός Θέσης και Όραση Υπολογιστών	21
2.2.1 Μέθοδοι συσχέτισης με μοντέλα κτιρίων.	22
2.2.2 Μέθοδοι βασισμένες σε εικόνες.	22
2.2.3 Μέθοδοι βασισμένες σε ορόσημα.	22
3 Θεωρητική ανάλυση	27
3.1 Αναγνώριση αντικειμένων μέσω Βαθιάς Μάθησης.	28
3.1.1 Αναγνώριση αντικειμένων (object detection).	28
3.1.2 Βαθιά Μάθηση - Ο αλγόριθμος Faster R-CNN.	28
3.2 Τεχνικές Όρασης Υπολογιστών - Αλγόριθμος SIFT	38

3.3	Φωτογραμμετρική Οπισθοτομία (Space resection).	40
3.3.1	Συνθήκη Συγγραμμικότητας (Collinearity Condition)	40
3.3.2	Γραμμικοποίηση εξισώσεων συγγραμμικότητας.	41
3.3.3	Λύση Ελαχίστων Τετραγώνων.	42
4	Υλοποίηση Αλγορίθμου Εντοπισμού Θέσης μέσω Ανίχνευσης Στα- τικών Αντικειμένων.	45
4.1	Το πρόβλημα	45
4.2	Περιγραφή της υλοποίησης.	46
4.2.1	Ο Αλγόριθμος	46
4.2.2	Το Στάδιο αναγνώρισης Στατικών Αντικειμένων	47
4.2.3	Το στάδιο Υπολογισμού Θέσης.	49
5	Πείραμα και Αξιολόγηση των Αποτελεσμάτων	53
5.1	Λεπτομέρειες υλοποίησης	53
5.1.1	Ο χώρος εργασίας	53
5.1.2	Προετοιμασία Πειράματος - Αποσυνδεδεμένη Φάση.	53
5.1.3	Αποτελέσματα Υποσταδίων - Συνδεδεμένη Φάση.	55
5.2	Αξιολόγηση της Μεθόδου.	61
5.2.1	Καθορισμός Πειράματος.	61
5.2.2	Αποτελέσματα της Μεθόδου	64
5.2.3	Σχολιασμός Αποτελεσμάτων	73
6	Συμπεράσματα και Μελλοντική Εργασία.	75
6.1	Σύνοψη Συμπερασμάτων Αξιολόγησης.	75
6.2	Μελλοντικές Προεκτάσεις	75
	Βιβλιογραφία	77

Κατάλογος Σχημάτων

3.1	Τα τρία βασικά υποστάδια της μεθόδου	27
3.2	Το συνολικό δίκτυο του αλγορίθμου Faster R-CNN.	29
3.3	Η αρχιτεκτονική κάθε δομικού συνόλου του δικτύου ResNet101	30
3.4	Η αρχιτεκτονική του δικτύου ResNet101	32
3.5	Η αρχιτεκτονική του Region Proposal Network	33
3.6	Αναπαράσταση του τρόπου λειτουργίας του RPN μέσω αγκυρών	34
3.7	Οπτικοποίηση των πλεονάζοντων πλαισίων περιγράμματος που προκύπτουν ως έξοδος του RPN	36
3.8	Η αρχιτεκτονική του επιπέδου Κατηγοριοποίησης.	37
3.9	Η τεχνική των οκτάβων.	39
4.1	Ο συνολικός αλγόριθμος εντοπισμού της θέσης του ρομπότ χωρίζεται σε 2 στάδια, (i) την αναγνώριση στατικών αντικειμένων στις εικόνες δοκιμής και (ii) τον υπολογισμό της θέσης του ρομπότ.	46
4.2	Το μοτίβο 6×9 που χρησιμοποιήθηκε για την βαθμονόμηση της κάμερας.	51
5.1	Κάτοψη του χώρου εργασίας εκτός κλίμακας για σκοπούς οπτικοποίησης.	54
5.2	Οι πόρτες 1,2,5,6,9 από τις συνολικές 9 που χρησιμοποιούνται ως στατικά αντικείμενα για τον αλγόριθμο.	55
5.3	Δύο εικόνες αναφοράς του στατικού αντικειμένου DOOR1. Τα τονισμένα σημεία αντιστοιχούν στα σημεία ελέγχου το εν λόγω στατικού αντικειμένου.	56
5.4	Μία εκ των εικόνων δοκιμής που περιέχει τα στατικά αντικείμενα DOOR4, DOOR5	56
5.5	Αποτέλεσμα του σταδίου αναγνώρισης. Ο αλγόριθμος faster R-CNN εντόπισε τα στατικά αντικείμενα DOOR4,DOOR5.	57
5.6	Το αποτέλεσμα του αλγορίθμου ταιριάσματος. Τα σημεία που είναι τονισμένα από τους κύκλους αναπαριστούν τα σημεία ελέγχου. Τα ενωμένα σημεία μεταξύ των εικόνων είναι τα έγκυρα ζεύγη χαρακτηριστικών μέσω των οποίων φτιάχνεται ο πίνακας H_{homo}	59
5.7	Το τελικό αποτέλεσμα του Αλγορίθμου 1. Με διαφορετικά χρώματα φαίνονται τα σημεία ελέγχου που προέκυψαν από διαφορετικά στατικά αντικείμενα.	60

5.8	Η μείωση των σφαλμάτων της κάθε μεταβλητής κατάστασης $\omega, \phi, \kappa, x, y, z$ καθώς αυξάνεται ο αριθμός επαναλήψεων.	61
5.9	Η γωνία Θ αντιπροσωπεύει την γωνία σκόπευσης η οποία θα εναλλάσσεται σε θέσεις στο εύρος $[0,180]$ μοιρών για τις διαφορετικές περιπτώσεις δοκιμής. . .	62
5.10	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 33 πρωινές λήψεις.	64
5.11	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 14 λήψεις στις οποίες εναλλάσσεται η απόσταση. Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη. . .	65
5.13	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 6 λήψεις που πραγματοποιήθηκαν απο τυχαίες θέσεις. Στο οριζόντιο άξονα φαίνεται το ID της κάθε εικόνας.	65
5.12	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 13 λήψεις στις οποίες εναλλάσσεται η γωνία θέασης Θ . Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη. . .	66
5.14	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 33 μεσημεριανές λήψεις.	66
5.15	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 14 λήψεις στις οποίες εναλλάσσεται η απόσταση. Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη. . .	67
5.16	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 13 λήψεις στις οποίες εναλλάσσεται η γωνία θέασης Θ . Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη. . .	68
5.17	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 6 λήψεις που πραγματοποιήθηκαν απο τυχαίες θέσεις. Στο οριζόντιο άξονα φαίνεται το ID της κάθε εικόνας.	68
5.18	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 33 απογευματινές λήψεις.	69
5.19	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 14 λήψεις στις οποίες εναλλάσσεται η απόσταση. Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη. . .	69
5.20	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 13 λήψεις στις οποίες εναλλάσσεται η γωνία θέασης Θ . Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη. . .	70
5.21	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 6 λήψεις που πραγματοποιήθηκαν απο τυχαίες θέσεις. Στο οριζόντιο άξονα φαίνεται το ID της κάθε εικόνας.	70
5.22	Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για όλες τις λήψεις. Στο υπόμνημα φαίνεται η ώρα εκτέλεσης της κάθε λήψης.	71
5.23	Οι χρόνοι εκτέλεσης της μεθόδου για κάθε Πρωινή Λήψη. Στους οριζόντιους άξονες φαίνεται το ID κάθε εικόνας	72

5.25 Οι χρόνοι εκτέλεσης της μεθόδου για κάθε Πρωινή Λήψη. Στους οριζόντιους άξονες φαίνεται το ID κάθε εικόνας	72
5.24 Οι χρόνοι εκτέλεσης της μεθόδου για κάθε Πρωινή Λήψη. Στους οριζόντιους άξονες φαίνεται το ID κάθε εικόνας	73

Κατάλογος Πινάκων

3.1	Η αρχιτεκτονική του νευρωνικού δικτύου ResNet101 χωρισμένη στα 5 βασικά της στάδια. Κάθε στάδιο αποτελείται από τα δομικά σύνολα που φαίνονται εντός των παρενθέσεων.	31
5.1	Τα ground truth των σημείων όπου εναλλάσσεται η απόσταση.	63
5.2	Τα ground truth των σημείων όπου εναλλάσσεται η γωνία θέασης της κάμερας.	63
5.3	Τα ground truth των σημείων που βρίσκονται σε τυχαίες θέσεις.	63
5.4	Το μέσο σφάλμα της μεθόδου για κάθε φάση λήψης.	73

Κεφάλαιο 1

Εισαγωγή

Τα τελευταία χρόνια η πρόοδος στον τομέα της Ρομποτικής σε συνδυασμό με την διαρκώς μειούμενη τάση στο κόστος προμήθειας Ρομποτικών μηχανισμών έχει οδηγήσει στην υιοθέτηση Ρομποτικών λύσεων για μια πληθώρα προβλημάτων τόσο στην βιομηχανία όσο και στην καθημερινή ζωή. Μια μεγάλη κατηγορία ρομποτικών συστημάτων που αποκτά ολοένα και μεγαλύτερη δυναμική είναι τα κινητά Ρομπότ (mobile robots), δηλαδή ρομποτ που δεν είναι καθηλωμένα σε ένα σταθερό σημείο του χώρου εργασίας τους (workspace). Χαρακτηριστικά, τέτοιες εφαρμογές εντοπίζονται σε μεγάλα νοσοκομειακά συγκροτήματα όπου τα ρομπότ χρησιμοποιούνται για την αποδοτικότερη μεταφορά αντικειμένων και εγγράφων, σε μεγάλες αποθήκες όπου συνήθως αλληλεπιδρά μεγάλος αριθμός ρομποτ ώστε να επιτευχθεί η μέγιστη δυνατή επίδοση κατά την μεταφορά και τοποθέτηση αντικειμένων από και προς τα ράφια, ή ακόμα και για ατομική χρήση όπως παραδείγματος χάριν αυτόνομες ηλεκτρικές σκούπες.

Για να εκτελεστεί επιτυχώς η οποιαδήποτε ενέργεια, είναι αναγκαίο το ρομπότ να αλληλεπιδράσει με τον περιβάλλοντα χώρο. Προκειμένου να μπορεί να αντιλαμβάνεται τον χώρο εργασίας του, καθώς και τυχόν αλλαγές που μπορεί να λάβουν χώρα με την πάροδο του χρόνου, ενσωματώνονται στο ρομπότ αισθητήρες (sensors) όπως κάμερες, επιταχυνσιόμετρα (accelerometers), αισθητήρες ακτίνων λέιζερ (lidars), GPS και άλλοι. Αντίθετα, προκειμένου το ρομπότ να καταφέρει να δράσει εντός του χώρου εργασίας του και να εκτελέσει εργασίες, ενσωματώνονται ενεργοποιητές (actuators) και χειριστές (manipulators) όπως πνευματικά στοιχεία, κινητήρες (rotors) ή βραχίονες [6].

Μια από τις μεγαλύτερες προκλήσεις σε εφαρμογές ρομποτικής με κινητά ρομπότ είναι η εκτέλεση αυτόνομης πλοήγησης (autonomous navigation). Προκειμένου το ρομποτικό σύστημα να εκτελέσει επιτυχημένη πλοήγηση βασίζεται σε τέσσερις θεμελιώδεις διαδικασίες:

- **Ικανότητα Αντίληψης (perception)**

Η εξαγωγή χρήσιμης πληροφορίας μέσω επεξεργασίας των δεδομένων που προέρχονται από τους ενσωματωμένους αισθητήρες.

- **Εντοπισμός Θέσης (Localisation)**

Η αναγνώριση της θέσης του ρομπότ στο περιβάλλον που δρα.

- **Λήψης αποφάσεων (Cognition)**

Η ικανότητα του ρομπότ να αποφασίζει πως θα δράσει προκειμένου να πετύχει τους καθορισμένους στόχους.

- **Έλεγχος της Κίνησης (motion control)**

Οι απαιτούμενοι νόμοι ελέγχου προκειμένου να δοθεί σωστό έναυσμα στο μέσο κίνησης του ρομπότ (κατα βάση στους τροχούς) ώστε να επιτευχθεί ο στόχος.

Εκ των τεσσάρων ακρογωνιαίων λίθων για την επιτυχημένη πλοήγηση, ο εντοπισμός θέσης παρουσιάζει το μεγαλύτερο ερευνητικό ενδιαφέρον τα τελευταία χρόνια εντός της επιστημονικής κοινότητας με αποτέλεσμα να έχουν αναπτυχθεί καινοτόμες λύσεις για το εν λόγω πρόβλημα τόσο σε εφαρμογές εσωτερικού όσο και εξωτερικού χώρου.

1.1 Εντοπισμός Θέσης

Το πρόβλημα του εντοπισμού θέσης είναι εκείνο που απαντά στο **θεμελιώδες ερώτημα: Πού βρίσκεται το Ρομπότ την παρούσα χρονική στιγμή;**

Μία αξιόπιστη απάντηση σε αυτή την ερώτηση είναι απαραίτητη προκειμένου να παρθεί η σωστή απόφαση για την ενέργεια που θα εκτελεστεί στην επόμενη χρονική στιγμή. Για τον Άνθρωπο, το πρόβλημα εντοπισμού της θέσης αποτελεί μια αυθόρμητη διαδικασία κατά την οποία αφού πατρατηρήσει τον περιβάλλοντα χώρο μέσω των αισθητήριων οργάνων του (όραση, ακοή) δημιουργεί συνειδητά ή υποσυνείδητα ένα χάρτη του χώρου, αναγνωρίζει τη θέση του σε αυτόν και σχεδιάζει ένα μονοπάτι της διαδρομής που επιθυμεί να ακολουθήσει [1,2,3]. Κατά αντιστοιχία, η πληροφορία που χρησιμοποιείται για τον εντοπισμό της θέσης συλλέγεται από τους ενσωματωμένους αισθητήρες μέσω των οποίων το ρομπότ παρακολουθεί τόσο την δική του κίνηση όσο και το περιβάλλον γύρω του.

1.1.1 Πλαίσια Αναφοράς

Απο την ίδια τη φύση της θεμελιώδους ερώτησης *Σε ποιά θέση βρίσκεται το ρομπότ;*, προκύπτει η ανάγκη ορισμού πλαισίων αναφοράς (reference frames) ως προς τα οποία θα εκφράζεται η θέση που εντοπίζεται κάθε χρονική στιγμή. Ο λόγος που γίνεται χρήση πολλαπλών πλαισίων αναφοράς έχει να κάνει με την ανάγκη υιοθέτησης διαφορετικών στρατηγικών κατά το στάδιο Λήψης αποφάσεων (cognition) ανάλογα με την εφαρμογή και τον στόχο που καλείται να διεκπεραιώσει το κινητό ρομπότ. Συγκεκριμένα, προκειμένου ο εντοπισμός θέσης να είναι αξιόπιστος, μπορεί να απαιτείται πληροφορία για την απόλυτη θέση του ρομπότ ως προς την Γη, την απόλυτη θέση του στα πλαίσια του χώρου εργασίας του, καθώς και την σχετική του θέση ως προς την έναρξη της κίνησης του ή ως προς τα αντικείμενα που το περιβάλλον. Χρησιμοποιούνται λοιπόν τα εξής **Πλαίσια Αναφοράς**:

1. **Γέω-σχετικό, Γέω-σταθερο Πλαίσιο Αναφοράς (ECEF) - Earth Centered, Earth Fixed Reference Frame**

Πρόκειται για ένα *Απόλυτο πλαίσιο αναφοράς* το οποίο ταυτίζεται με το κέντρο μάζας της γης. Το είδος αισθητήρων που χρησιμοποιεί αυτό το πλαίσιο αναφοράς είναι αισθητήρες

του Παγκόσμιου Συστήματος Δορυφορικής Πλοήγησης (Global Navigation Satellite System - GNSS) με πιο διάσημο το Παγκόσμιο Σύστημα Στιγματοθέτησης (Global Positioning System - GPS).

2. Πλαίσιο Αναφοράς Χάρτη (Map reference frame)

Συνήθως το πρόβλημα εντοπισμού θέσης είναι άρρηκτα συνδεδεμένο με την χρήση ή και δημιουργία ενός χάρτη του χώρου εργασίας. Ο χάρτης αυτός είτε φορτώνεται στον εγκέφαλο του ρομπότ επειτα από προηγούμενη χαρτογράφηση του χώρου, είτε δημιουργείται και ενημερώνεται ζωντανά μέσω μιας διαδικασίας που ονομάζεται Ταυτόχρονος Εντοπισμός Θέσης και Χαρτογράφηση (Simultaneous Localisation and Mapping - SLAM). Με βάση αυτόν το χάρτη, ορίζεται το σημείο μηδέν, το οποίο αντιστοιχεί και στο Πλαίσιο Αναφοράς Χάρτη, το οποίο και αποτελεί ακόμα ένα Απόλυτο Πλαίσιο Αναφοράς. Το Πλαίσιο Αναφοράς Χάρτη χρησιμοποιείται κατα κόρον για τον Σχεδιασμό Μονοπατιού (Path Planning) ώστε το ρομπότ να φθάσει σε μία συγκεκριμένη θέση του χάρτη.

3. Πλαίσιο Αναφοράς Οδομετρίας (Odometry Reference Frame)

Οδομετρία ορίζεται η μέθοδος κατα την οποία για να εντοπιστεί η θέση σε κάθε χρονική στιγμή γίνεται χρήση αισθητήρων που εποπτεύουν την κίνηση του ίδιου του οχήματος όπως κωδικοποιητές τροχών (wheel encoders) και Αδρανειακές Μονάδες Μέτρησης (Inercial Measurement Unit - IMU) σε συνδυασμό με το μαθηματικό μοντέλο για την κίνηση του συγκεκριμένου ρομπότ (παραδείγματος χάρη μοντέλο διαφορικής οδήγησης ή ολονομικό μοντέλο). Ως Πλαίσιο Αναφοράς Οδομετρίας ορίζεται εκείνο που είναι καρφωμένο στο σημείο εκκίνησης της κίνησης του ρομπότ, δηλαδή πρόκειται για μία σχετικιστική προσέγγιση της θέσης εντός του δοσμένου χάρτη του χώρου εργασίας.

1.1.2 Τυπικά Προβλήματα

Ο Εντοπισμός Θέσης βασίζεται σε πολύ μεγάλο βαθμό στους αισθητήρες που είναι υπεύθυνοι για την Αντίληψη του χώρου εργασίας του ρομπότ και την καταγραφή της κατάστασης της κίνησης του. Αυτό οδηγεί στην δημιουργία προβλημάτων που οφείλονται στους εξής παράγοντες:

- **Ακρίβεια Μετρήσεων.**

Παρότι υπάρχουν κάποια είδη αισθητήρων προηγμένης τεχνολογίας που δύνανται να προσφέρουν μετρήσεις μεγάλης ακρίβειας όπως αισθητήρες δέσμης Λείζερ (LiDAR), συχνά δεν επιλέγονται λόγω του υψηλού κόστους. Έτσι σε πολλές εφαρμογές χρησιμοποιούνται αισθητήρες που προσφέρουν μικρότερη ακρίβεια στις μετρήσεις του και άρα ο Εντοπισμός Θέσης καθίσταται δυσκολότερος. Χαρακτηριστικό παράδειγμα αυτού του προβλήματος είναι οι αισθητήρες GPS που η παρούσα τεχνολογία μπορεί να παράσχει ακρίβεια μερικών μέτρων πράγμα που την καθιστά μη αποδεκτή για μια σειρά ρομποτικών εφαρμογών όπως ανθρωποκεντρικές εφαρμογές ή νανο-ρομπότ.

- **Θόρυβος Αισθητήρων.**

Οι ενσωματωμένοι αισθητήρες παρουσιάζουν θόρυβο στις μετρήσεις τους πράγμα που εισάγει μεγάλο βαθμό αναξιοπιστίας στην πληροφορία που καλείται να επεξεργαστεί ο εγκέφαλος του ρομποτ ώστε να εκτελέσει τον Εντοπισμό της Θέσης. Παραδείγματος χάριν, οι αισθητήρες LiDAR προσφέρουν μετρήσεις υψηλής ακρίβειας αλλά εισάγουν θόρυβο αν ο χώρος εργασίας παρουσιάζει συνθήκες ομίχλης που αλλοιώνει την ποιότητα της ανακλώμενης δέσμης και άρα την αξιοπιστία των μετρήσεων.

- **Συμβατότητα Αισθητήρα - Εφαρμογής - Χώρου εργασίας.**

Είναι συχνό φαινόμενο να υπάρχουν αισθητήρες που δύνανται να λύσουν προβλήματα στο στάδιο του Εντοπισμού Θέσης για κάποια εφαρμογή, αλλά η χρήση τους είναι αδύνατη επειδή δεν είναι συμβατοί με τον χώρο εργασίας για της εφαρμογής. Τρανταχτό παράδειγμα αποτελούν ξανά οι αισθητήρες GPS που παρότι προσφέρουν την απόλυτη θέση του ρομποτ για εξωτερικές εφαρμογές, η χρήση τους σε εσωτερικούς ή κεκαλυμμένους χώρους είναι αδύνατη καθώς δεν μπορούν να εντοπίσουν το σήμα του δορυφόρου.

- **Αναδίπλωση Πληροφορίας - Sensor Aliasing**

Το πρόβλημα του sensor Aliasing έχει να κάνει με το γεγονός ότι η μέτρηση απο ένα μόνο είδος αισθητήρα δεν δύναται να ενθυλακώσει όλη απαραίτητη πληροφορία για τον υπολογισμό της κατάστασης του ρομποτ (state) σε μια δεδομένη χρονική στιγμή και άρα τον υπολογισμό της θέσης του. Χαρακτηριστικά, ένας αισθητήρας δέσμης Υπερήχων (ultrasonic) παρέχει πληροφορία μόνο για την απόσταση προς ένα εμπόδιο αγνοώντας πληροφορία για ανάγλυφο, χρώμα και σκληρότητα. Συνεπώς, όταν γίνεται χρήση μόνο ενός είδους αισθητήρα, ακόμα και αν δεν παρουσίαζε καθόλου θόρυβο, ο Εντοπισμός θέσης καθίσταται ανακριβής και άρα απαιτούνται εξεζητημένες τεχνικές επεξεργασίας των δεδομένων.

1.2 Αντικείμενο διπλωματικής

1.2.1 Το πρόβλημα

Παρά την διαρκώς αυξανόμενη ζήτηση για ρομποτικές εφαρμογές πλοήγησης σε εσωτερικούς χώρους, δεν έχει αναπτυχθεί ακόμα κάποια τεχνολογία για Εσωτερικό Εντοπισμό Θέσης τόσο αξιόπιστη όσο είναι οι αισθητήρες GNSS για εφαρμογές εξωτερικού χώρου, διότι το σήμα του δορυφόρου αδυνατεί να διαπεράσει συμπαγή αντικείμενα όπως τοίχους. Επιπλέον, συχνά είναι αδύνατη και η χρήση αισθητήρων μετάδοσης Ραδιοκυμάτων (RF transmission) εξαιτίας του δαιδαλώδους ανάγλυφου που παρουσιάζουν συχνά οι χώροι εργασίας σε εσωτερικά περιβάλλοντα. Ταυτόχρονα, μέθοδοι που βασίζονται στην Οδομετρία παρουσιάζουν το πολύ σημαντικό μειονέκτημα του σωρευτικού σφάλματος στον Εντοπισμό Θέσης, ενώ πιθανοί αισθητήρες που είναι συμβατοί με εσωτερικούς χώρους, παρουσιάζουν υψηλά επίπεδα θορύβου και aliasing. Δημιουργείται, λοιπόν, η ανάγκη για την ανάπτυξη τεχνικών Απόλυτου Εντοπισμού Θέσης (Absolute Localisation) σε εσωτερικά περιβάλλοντα οι οποίες πρέπει να

εξασφαλίζουν:

- Υψηλή ακρίβεια πρόβλεψης θέσης.
- Αξιόπιστες μετρήσεις ανεξάρτητα από το χρόνο λειτουργίας του συστήματος.
- Δυνατότητα κλιμακωσιμότητας όσον αφορά το εύρος του χώρου εργασίας.
- Μειωμένα επίπεδα θορύβου.

Στα πλαίσια της διπλωματικής εργασίας υλοποιήθηκε η αλγοριθμική λύση που προτείνεται στην επιστημονική δημοσίευση [7]. Υλοποιήθηκε ένας υβριδικός αλγόριθμος Εντοπισμού Απόλυτης Θέσης για εσωτερικό περιβάλλον που παρουσιάζει προβλέψεις ακριβείας εκατοστών. Ο αλγόριθμος βασίζεται στην οπτική αναγνώριση αντικειμένων που παραμένουν σταθερά σε μια σκηνή εσωτερικού χώρου, όπως πόρτες ή παράθυρα, ώστε να υπολογίσει την θέση της κάμερας εντός του χώρου και κατ' επέκταση την θέση του Ρομπότ. Ακολούθησε πειραματική αξιολόγηση των παραμέτρων του αλγορίθμου και της απόδοσης του σε διάφορες καταστάσεις.

1.2.2 Συνεισφορά

Τα πειραματικά αποτελέσματα δείχνουν ότι ο αλγόριθμος είναι μια καλή επιλογή στο πρόβλημα του Εντοπισμού Απόλυτης Θέσης σε χώρο εργασίας εσωτερικού περιβάλλοντος. Ο αλγόριθμος καταφέρνει να δώσει μία εύρωστη και αξιόπιστη λύση με ακρίβεια που συναγωνίζεται τις συμβατικές μεθόδους εντοπισμού θέσης ως προς την ακρίβεια των εκτιμήσεων θέσης. Ταυτόχρονα, τα πειραματικά αποτελέσματα καταδεικνύουν πιθανές αδυναμίες του αλγορίθμου στους χρόνους εκτέλεσης και απόκρισης.

1.3 Οργάνωση Τόμου

Η παρούσα διπλωματική εργασία είναι οργανωμένη σε έξι κεφάλαια. Εργασίες σχετικές με το αντικείμενο της διπλωματικής παρουσιάζονται στο Κεφάλαιο 2. Στο Κεφάλαιο 3 παρουσιάζεται το θεωρητικό υπόβαθρο, καθώς και οι βασικές έννοιες και τεχνικές που ήτο χρήσιμες στη διπλωματική εργασία. Στο Κεφάλαιο 4 παρουσιάζεται εκτενώς το πρόβλημα, και οι αλγόριθμοι που χρησιμοποιήθηκαν στην διπλωματική για την επίλυση του. Στο Κεφάλαιο 5 παρουσιάζουμε τα πειραματικά αποτελέσματα των μετρήσεων και ακολουθεί ανάλυση αυτών και παρουσίαση των συμπερασμάτων της διπλωματικής. Το Κεφάλαιο 6 παρέχει ιδέες και κατευθύνσεις για την προοπτική μελλοντικής εξέλιξής της.

Κεφάλαιο 2

ΒΙΒΛΙΟΓΡΑΦΙΑ

2.1 Ο εντοπισμός θέσης σε εσωτερικούς χώρους.

Με εξαίρεση τους αισθητήρες λήψης σημάτων GNSS, το μεγαλύτερο εύρος αισθητήρων και ραδιοσημάτων (RF) δεν έχουν σχεδιαστεί με πρωταρχικό στόχο τον εντοπισμό θέσης αλλά μπορούν να χρησιμοποιηθούν για αυτόν τον σκοπό μέσω χρήσης διαφορετικών αρχών και αλγορίθμων. Μεταξύ αυτών συγκαταλέγονται τεχνικές όπως η αναγνώριση αποτυπώματος Wi-Fi [9], γεωμαγνητισμού [10] ή Bluetooth [11] οι οποίες είναι διάσημες για την απόδοση τους και την ανεξαρτησία τους όσον αφορά τον χώρο εργασίας στον οποίο εφαρμόζονται. Τέτοιες μέθοδοι δύνανται να πετύχουν την θέση του ρομπότ με ακρίβεια δύο έως πέντε μέτρων αλλά συνήθως εμφανίζουν μεγάλα επίπεδα παρεμβολών όταν χρησιμοποιούνται σε δυναμικά περιβάλλοντα.

2.2 Εντοπισμός Θέσης και Όραση Υπολογιστών

Τις τελευταίες δεκαετίες υπάρχει μεγάλο ερευνητικό ενδιαφέρον πάνω σε μεθόδους εντοπισμού θέσης που βασίζονται στην επεξεργασία εικόνων και την όραση των υπολογιστών. Οι οπτικές μέθοδοι σε συνδυασμό με συμβατικές μεθόδους που αναφέρθηκαν παραπάνω δύνανται να επιφέρουν πιο αξιόπιστα αποτελέσματα σε δυναμικά περιβάλλοντα.

Στην ερευνητική εργασία [16] παρουσιάζεται μια απλή αλλά ισχυρή μέθοδος αναγνώρισης θέσης σε εσωτερικούς χώρους που επιτρέπει τον λεπτομερή εντοπισμό θέσης και προσανατολισμού του χρήστη. Η συνεισφορά αυτής της μεθόδου έγκειται στον συνδυασμό ενός συστήματος αναγνώρισης εικόνων με έναν αλγόριθμο υπολογισμού απόστασης χρησιμοποιώντας μόνο την κάμερα της κινητής πλατφόρμας χωρίς την ανάγκη οποιασδήποτε προϋπάρχουσας υποδομής.

Ο διαχωρισμός των οπτικών μεθόδων για τον εντοπισμό θέσης και κατ'επέκταση την υλοποίηση πλοήγησης μπορεί να χωριστεί σε τρεις κατηγορίες όπως φαίνεται παρακάτω.

2.2.1 Μέθοδοι συσχέτισης με μοντέλα κτιρίων.

Τέτοιες μέθοδοι συμπεριλαμβάνουν αλγορίθμους που εντοπίζουν αντικείμενα σε εικόνες και έπειτα τα αντιστοιχίζουν με εκείνα που βρίσκονται στην βάση δεδομένων του κτιρίου.

Η μέθοδος αυτή χρησιμοποιήθηκε στην ερευνητική εργασία [12] και βασίστηκε σε ένα ψηφιακό χωρο-σημασιολογικό μοντέλο του εσωτερικού ενός κτιρίου σε συνδυασμό με έναν αισθητήρα εικόνας-απόστασης (Range Imaging Sensor). Σε αντίθεση με συμβατικές μεθόδους εντοπισμού θέσης για εσωτερικούς χώρους, η συγκεκριμένη πρακτική δεν προαπαιτεί την ύπαρξη τοπικής φυσικής υποδομής όπως θερμά σημεία ασύρματων τοπικών δικτύων (WLAN hot spots) ή σηματοδότες αναφοράς.

2.2.2 Μέθοδοι βασισμένες σε εικόνες.

Οι μέθοδοι αυτές λειτουργούν εν γένει ως εξής, μέσω της κάμερας του ρομπότ λαμβάνεται μια φωτογραφία η οποία δημιουργεί την εικόνα δοκιμής (testing image). Στην συνέχεια, η εικόνα αυτή συγκρίνεται με ένα σύνολο εικόνων που ονομάζονται εικόνες συσχέτισης (reference images) και επιλέγεται το ζευγάρι με το μεγαλύτερο σκορ. Οι εικόνες συσχέτισης έχουν ληφθεί σε φάση εκτός σύνδεσης.

Οι Kim και Jun υλοποίησαν στην ερευνητική εργασία [13] μια μέθοδο που βασίζεται στην παραπάνω αρχή. Προκειμένου να αναγνωρίζει το ρομπότ την θέση του, προ-κατασκεύασαν μια βάση δεδομένων με εικόνες και μοντέλα θέσης που αποτελούνται από σημεία και μονοπάτια μεταξύ διαφορετικών θέσεων σε ένα εσωτερικό περιβάλλον. Στην συνέχεια, η θέση το ρομπότ αναγνωρίζεται κάνοντας χρήση αυτή της *a priori* γνώσης για την διαρρύθμιση του εσωτερικού περιβάλλοντος σε συνδυασμό με την χρήση διαφόρων αλγορίθμων Όρασης Υπολογιστών. Η μέθοδος σχεδιάστηκε για εφαρμογές Επαυξημένης Πραγματικότητας (Augmented Reality - AR) για την παροχή περαιτέρω πληροφορίας θέσης.

2.2.3 Μέθοδοι βασισμένες σε ορόσημα.

Πρόκειται για μεθόδους στις οποίες γίνεται χρήση σημαδιών αναφοράς ή ορόσημων (landmarks) τα οποία έχουν εγκατασταθεί σε προκαθορισμένες θέσης εντός του χώρου εργασίας του ρομπότ. Τα σημάδια αυτά συνήθως είναι είτε κωδικοποιημένοι στόχοι όπως ομόκεντροι δακτύλιοι και barcodes ή μοτίβα αποτελούμενα από χρωματιστές ρίζες ή βούλες.

Η κεντρική ιδέα αυτών των μεθόδων παρουσιάζεται στην έρευνα [41]. Σε μεθόδους βασισμένες σε ορόσημα, παρέχεται στο ρομπότ ο χάρτης του χώρου εργασίας καθώς και μία ακολουθία από ορόσημα τα οποία αναμένεται να εντοπιστούν κατά την διάρκεια της πλοήγησης. Κατά την διάρκεια της πλοήγησης, ανατίθεται σε ένα σύστημα όρασης να αναζητήσει και να αναγνωρίσει ορόσημα σε κάθε εικόνα που λαμβάνει. Μόλις τα ορόσημα αναγνωριστούν εντός της εικόνας, το ρομπότ κάνει χρήση του χάρτη ώστε να εκτιμήσει την θέση, συγκρίνοντας τις παρατηρήσεις (εικόνα) ως προς τις προσδοκώμενες θέσεις που είναι αποθηκευμένες σαν περιγραφές των οροσήμων σε κάποια βάση δεδομένων. Συγκεντρωτικά, οι υπολογισμοί της διαδικασίας αυτής χωρίζονται στα εξής τέσσερα βήματα:

1. Συγκέντρωση πληροφορίας από τους αισθητήρες.

Για πλοήγηση βασισμένη στην όραση υπολογιστών το στάδιο αυτό περιλαμβάνει την συγκέντρωση και ψηφιοποίηση εικόνων από τις κάμερες.

2. Εντοπισμός Οροσήμεων.

Συνήθειες πρακτικές για αυτό το στάδιο είναι εξαγωγή χαρακτηριστικών στις εικόνες (ακμές, γωνίες), η εξομάλυνση, φιλτράρισμα και ο χωρισμός σε υποπεριοχές με βάση διαφορές στην φωτεινότητα (gray levels), στους χρωματισμούς (RGB colour) και το βάθος (depth).

3. Υλοποίηση ταιριάσματος ανάμεσα στις παρατηρήσεις και τις αναμενόμενες τιμές.

Σε αυτό το βήμα, το σύστημα προσπαθεί να αναγνωρίσει τα ορόσημα που παρατηρήθηκαν αναζητώντας στην βάση δεδομένων πιθανές αντιστοιχίσεις με τα πατηρημένα ορόσημα με έχοντας ως γνώμονα κάποιο κριτήριο μέτρησης.

4. Υπολογισμός Θέσης.

Μόλις πραγματοποιηθεί κάποια αντιστοίχιση ή σύνολο αντιστοιχίσεων, το σύστημα πρέπει να υπολογίσει την θέση ως συναρτηση των ορόσημων που παρατηρήθηκαν και των θέσεων τους στην βάση δεδομένων.

Έχοντας ως κύριο σκελετό την παραπάνω σειρά διεργασιών, τα τελευταία χρόνια έχει πραγματοποιηθεί μια μεγάλη ποικιλία δημοσιεύσεων προς αυτή την κατεύθυνση. Συγκεκριμένα, οι κατηγορίες εντοπισμού θέσης με την χρήση οροσήμεων μπορούν να χωριστούν σε μεθόδους που κάνουν χρήση φυσικών οροσήμεων και μεθόδους με τεχνητά ορόσημα.

Μέθοδοι Φυσικών Οροσήμεων

Στο [39] αναπτύχθηκε η μέθοδος "Αντιστοίχισης Οροσήμεων, Τριγωνισμού, Ανακατασκευής και Σύγκρισης" - "Landmark Matching, Triangulation, Reconstruction, and Comparison" (**LTRC**). Η μέθοδος αυτή αναπτύχθηκε με αφορμή τα παρακάτω προβλήματα:

- Πολλοί γενικευμένοι αλγόριθμοι εκτίμησης θέσης είναι τρωτοί ως προς την αβεβαιότητα που εισάγεται από ορόσημα τα οποία δεν είναι μοναδικώς ορισμένα.
- Σε τεχνικές εντοπισμού θέσης που έχουν ως κορμό την όραση υπολογιστών, είναι σύνηθες φαινόμενο να μην αξιοποιείται πλήρως η πληροφορία των εικόνων υψηλών διαστάσεων.

Ο αλγόριθμος απόλυτου εντοπισμού θέσης **LTRC** υπερνικά τα προβλήματα αυτά καθώς εξαγάγει φυσικά ορόσημα για την εκτέλεση μιας φάσης αδρού ταιριάσματος χαρακτηριστικών πριν την δημιουργία της λίστας εκτιμήσεων θέσης μέσω τριγωνισμού. Στην συνέχεια ακολουθούν οι φάσεις ανακατασκευής και σύγκρισης προκειμένου να βαθμονομηθούν οι πιθανές

εκτιμήσεις. Προκειμένου να δοκιμαστεί η μέθοδος χρησιμοποιήθηκε μια γλώσσα προγραμματισμού βασισμένη σε διερμηνέα καθώς και ένα ρομπότ με ενσωματωμένο σύστημα όρασης πανοραμικής κάλυψης του χώρου. Η μέθοδος παρουσίασε σημαντική βελτίωση όσον αφορά την ακρίβεια των προβλέψεων σε σύγκριση με την συμβατική μέθοδο RANSAC [35]. Επιπλέον, η μέθοδος είναι εύρωστη ακόμα και όταν παρέχονται χάρτες με ανακριβή δεδομένα.

Στην ερευνητική εργασία [40] παρουσιάζεται μια στρατηγική καθολικού αυτό-εντοπισμού θέσης για μικρά αυτόνομα ρομπότ που δρουν σε γνωστό μεν, δυναμικό δε περιβάλλον κάνοντας χρήση στερεοσκοπικής κάμερας. Το εν λόγω ρομπότ εκινείτο πάνω σε προκαθορισμένο μονοπάτι, οπότε ο απόλυτος εντοπισμός θέσης κατέστη απαραίτητος κατά την εκκίνηση της πλοήγησης ή όταν το ρομπότ ξέφευγε της πορείας του λόγω της δυναμικότητας του περιβάλλοντος. Σε αντίθεση με συμβατικές μεθόδους που λύνουν αυτό το πρόβλημα μέσω πυκνών σαρωτών λέιζερ, πολυκατευθυντικών καμερών ή τεχνητών οροσήμεων, εδώ ο υπολογισμός της απόστασης γίνεται με βάση την εξαγωγή χαρακτηριστικών από τις στερεοσκοπικές εικόνες. Συγκεκριμένα, η θέση του ρομπότ υπολογίζεται μετρώντας την απόσταση ως προς ευδιάκριτα φυσικά χαρακτηριστικά όπως χρωματικές εναλλαγές, γωνίες, και διασταυρώσεις γραμμών. Σε αντίθεση με μεθόδους που βασίζονται σε μετρήσεις γωνιών, εδώ απαιτούνται μόνο 2 ευδιάκριτα ορόσημα.

Μέθοδοι Τεχνητών Οροσήμεων

Στην ερευνητική εργασία [15] χρησιμοποιήθηκε ένα μοτίβο barcode το οποίο κολλήθηκε σε διαφορετικά σημεία εντός του χώρου εργασίας του ρομπότ. Στην συνέχεια, η κάμερα εντοπίζει αυτά τα σημάδια ώστε να εξαγάγει κάθε φορά την θέση του ρομπότ καθώς και περαιτέρω πληροφορία.

Στην ερευνητική εργασία [17] παρουσιάζεται μια μέθοδος αυτό-εντοπισμού θέσης για πλοήγηση ρομπότ σε εσωτερικούς χώρους που βασίζεται σε αξιόπιστο σχεδιασμό και αναγνώριση τεχνητών ορόσημων. Κάθε ένα από αυτά τα ορόσημα αποτελείται από μοτίβα ομόκεντρων δακτυλίων σε λευκό και μαύρο χρώμα ώστε να κωδικοποιείται αξιόπιστα ή ταυτότητα του κάθε ορόσημου υπό τον φωτισμό του περιβάλλοντος. Ένα κινητό ρομπότ καθώς πλοηγείται στον χώρο καταγράφει με την ενσωματωμένη κάμερα του ορόσημα στο περιβάλλον του. Εντός της εικόνας, τώρα, τα ορόσημα εντοπίζονται και αναγνωρίζονται χρησιμοποιώντας έναν δι-επίπεδο αλγόριθμο αναγνώρισης:

1. Μία καθολική διαδικασία αναγνώρισης αρχικά εξαγάγει υποψήφιες περιοχές οροσήμεων από όλο το εύρος της εικόνας και γίνεται προσπάθεια να εντοπιστεί επαρκής αριθμός οροσήμεων.
2. Αν καταστεί αναγκαίο, μια τοπική διαδικασία αναγνώρισης ενισχύει τις μη αναγνωρισμένες περιοχές ενδιαφέροντος που επηρεάζονται από την φωτεινότητα ή έλλειψη πληροφορίας και τις επανεξετάζει.

Τέλος, τα αναγνωρισμένα ορόσημα χρησιμοποιούνται προκειμένου να εντοπιστεί η θέση

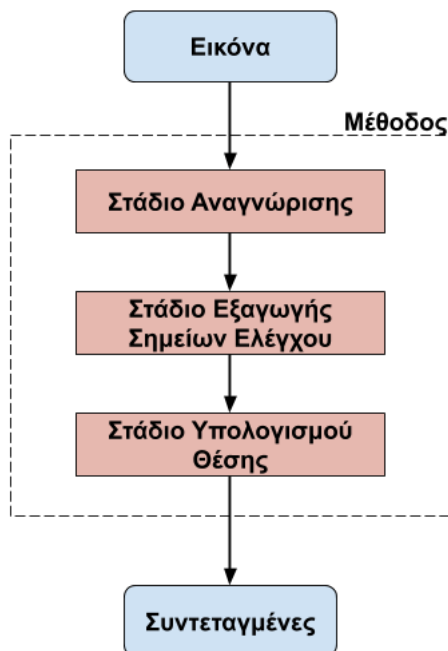
και ο προσανατολισμός την ενσωματωμένης κάμερας εντός του χώρου εργασίας μέσω της γεωμετρικής σχέσης ανάμεσα στα πλαίσια της εικόνας και του περιβάλλοντος.

Στην ερευνητική εργασία [38] παρουσιάζεται μία υπολογιστικά απαιτητική διαδικασία αυτο-εντοπισμού θέσης για χρήση σε πλοήγηση αυτόνομων ρομπότ εσωτερικού χώρου. Η τεχνική που ακολουθούν βασίζεται σε μια μορφή τριγωνισμού με δύο ευθείες (bilateration), εκμεταλλευόμενοι αρχές της προβολικής γεωμετρίας (projective geometry). Γίνεται χρήση δύο στρατηγικά τοποθετημένων ορόσημων στο περιβάλλον του ρομπότ, η απόσταση του από τα οποία υπολογίζεται με εξαγωγή χαρακτηριστικών από φωτογραφίες τους. Για την επιτυχή εξαγωγή των χαρακτηριστικών, απαιτείται η πραγματοποίηση επαρκών λήψεων των ορόσημων. Η προϋπόθεση αυτή ικανοποιείται με τη σάρωση του περιβάλλοντος του ρομπότ, από την κάμερα η οποία βρίσκεται συνδεδεμένη σε αυτό, έως ότου αναγνωριστούν δύο επαρκείς λήψεις. Ως εκ τούτου, κρίνεται απαραίτητη η χρησιμοποίηση λογισμικού επεξεργασίας εικόνας σε πραγματικό χρόνο.

Κεφάλαιο 3

Θεωρητική ανάλυση

Στην ενότητα αυτή περιγράφονται κάποιες βασικές και χρήσιμες έννοιες και γίνεται επεξήγηση της απαιτούμενης θεωρίας που απαιτείται για την υλοποίηση της μεθόδου. Τα τρία βασικά στάδια της μεθόδου τα οποία και αναλύονται στα επόμενα υποκεφάλαια φαίνονται στο σχήμα 3.1



Σχήμα 3.1: Τα τρία βασικά υποστάδια της μεθόδου

3.1 Αναγνώριση αντικειμένων μέσω Βαθιάς Μάθησης.

3.1.1 Αναγνώριση αντικειμένων (object detection).

Το πρόβλημα της αναγνώρισης αντικειμένων ορίζεται ως η διαδικασία κατα την οποία, δοθείσης μίας εικόνας, πρώτα αναγνωρίζεται η θέση αντικειμένων εντός αυτής (object localisation) και στην συνέχεια αναγνωρίζεται η κατηγορία στην οποία ανήκει το κάθε αντικείμενο (object classification). Συνεπώς, η παραδοσιακή διαδικασία αναγνώρισης αντικειμένων μπορεί να χωριστεί στις εξής τρεις ενέργειες [18]:

1. Πρόταση περιοχών χρησιμής πληροφορίας (Region Proposal):

Η πιθανότητα εμφάνισης διαφόρων αντικειμένων σε τυχαίες θέσεις και μεγέθη εντός της εικόνας οδηγούν στην υιοθέτηση μιας πολυ-κλιμακωτής τεχνικής συρόμενου παραθύρου προκειμένου να σαρωθεί ολόκληρη η εικόνα. Παρότι αυτή η εξαντλητική τεχνική δύναται να εντοπίσει όλες τις περιοχές ενδιαφέροντος παρουσιάζει μεγάλο αριθμό πλεονάζοντων παραθύρων οξύνοντας την πολυπλοκότητα ενώ σε περίπτωση που επιλεχθεί σταθερός αριθμός παραθύρων υπάρχει κίνδυνος μη ικανοποιητικών αποτελεσμάτων.

2. Εξαγωγή χαρακτηριστικών (feature extraction).

Για την αναγνώριση των διαφορετικών αντικειμένων απαιτείται η εξαγωγή οπτικών χαρακτηριστικών τα οποία θα παράσχουν σημασιολογική και ευσταθή αναπαράσταση αυτών.

3. Κατηγοριοποίηση (Classification).

Τέλος, απαιτείται η ύπαρξη τεχνικής κατηγοριοποίησης των αναγνωρισμένων αντικειμένων ώστε να ξεχωρίζεται το εκάστοτε αντικείμενο από όλες τις υπόλοιπες κατηγορίες.

3.1.2 Βαθιά Μάθηση - Ο αλγόριθμος Faster R-CNN.

Τα τελευταία χρόνια έχει γίνει μια στροφή στην προς την χρήση τεχνικών Βαθιάς Μάθησης για την αντιμετώπιση του προβλήματος αναγνώρισης αντικειμένων. Συγκεκριμένα, έχουν αναπτυχθεί νέες μέθοδοι πρότασης περιοχών [21] βασισμένες πάνω στα λεγόμενα συνελικτικά νευρωνικά δίκτυα επικέντρωσης περιοχής (region-based convolutional neural network R-CNN) τα οποία είναι ασύμφορα όσον αφορά την υπολογιστική πολυπλοκότητα. Αλγόριθμοι όπως ο Fast R-CNN και ο Faster R-CNN καταφέρνουν να μειώσουν δραστικά τους χρόνους εκτέλεσης και να προσφέρουν αποτελέσματα σχεδόν σε πραγματικό χρόνο.

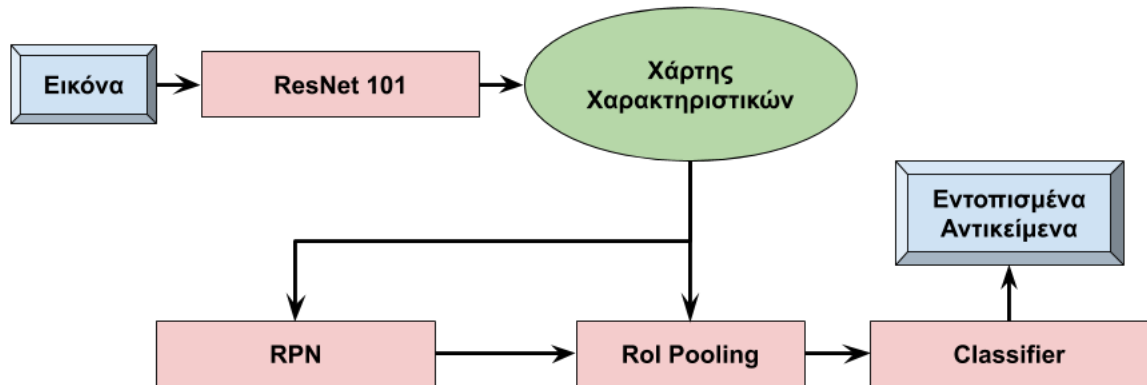
Στα πλαίσια της παρούσας εργασίας υλοποιήθηκε ο αλγόριθμος Faster R-CNN προκειμένου να επιλυθεί το πρόβλημα της αναγνώρισης αντικειμένων. Ο αλγόριθμος ενσωματώνει τα τρία βασικά βήματα της αναγνώρισης αντικειμένων, region proposal, feature extraction, classification σε ένα πλήρες και συνεχόμενο δικτύωμα ενώ ταυτόχρονα μειώνει δραστικά τους υπολογισμούς επιταχύνοντας έτσι την διαδικασία αναγνώρισης. Ο τρόπος με τον οποίο καταφέρνει να επιτύχει αυτή την βελτίωση στους χρόνους οφείλεται στην υιοθέτηση ενός νέου δικτύου στο τελικό στάδιο του προγόνου του Fast R-CNN το οποίο ονομάζεται Δίκτυο Πρότασης Περιοχών (Region Proposal Network -RPN) το οποίο και θα εξηγηθεί παρακάτω.

Η Δομή του δικτύου.

Ο ακριβής τίτλος του δικτύου που χρησιμοποιήθηκε στην παρούσα εργασία ονομάζεται:

Faster R-CNN ResNet 101

και η δομή του φαίνεται στην εικόνα 3.2: Πιο λεπτομερώς, το Δικτύωμα αποτελείται από



Σχήμα 3.2: Το συνολικό δίκτυο του αλγορίθμου Faster R-CNN.

τα εξής διαδοχικά στάδια:

1. Υπολειπόμενο Νευρωνικό Δίκτυο (Residual Neural Network - ResNet).

Η αρχιτεκτονική ενός ResNet [22] βασίζεται στην δομή ενός απλού πολυεπίδεδου συνελικτικού νευρωνικού δικτύου όπως είναι το VGG [24]. Τα συνελικτικά επίπεδα αποτελούνται κυρίως από 3×3 φίλτρα και ακολουθούν δύο βασικές αρχές σχεδιασμού: (i) Αν το προηγούμενο επίπεδο παράξει χάρτη χαρακτηριστικών αμετάβλητου μεγέθους τότε το τρέχον επίπεδο έχει τον ίδιο αριθμό φίλτρων με το προηγούμενο επίπεδο, (ii) αν το μέγεθος του χάρτη μειωθεί στο μισό, τότε το τρέχον επίπεδο έχει διπλάσιο αριθμό φίλτρων από το προηγούμενο επίπεδο ώστε να διατηρηθεί η ίδια πολυπλοκότητα ανά επίπεδο. Τέλος το δικτύωμα ολοκληρώνεται με ένα καθολικό επίπεδο μέσης υποδειγματοληψίας (average pooling) και ένα επίπεδο πλήρους διασύνδεσης των νευρώνων (fully connected layer).

Υπολειπόμενη Μάθηση

Η καινοτομία που εισάγει το ResNet στην παραπάνω απλή αρχιτεκτονική είναι η ιδέα της υπολειπόμενης μάθησης (residual learning). Αντί να αναμένεται από κάθε επίπεδο να ταιριάζει την είσοδο του σε κάποια υποκείμενη συνάρτηση χαρτογράφησης, αφήνεται συνειδητά σε κάθε επίπεδο να πραγματοποιήσει ένα ταίριασμα σε μια υπολειπόμενη συνάρτηση χαρτογράφησης. Έτσι, για την $H(x)$ υποκείμενη συνάρτηση, τα στοιβαγμένα

μη γραμμικά επίπεδα αφήνονται να κάνουν χαρτογράφηση της μορφής $F(x) := H(x) - x$ και άρα, η αρχική συνάρτηση ανάγεται στην $F(x) + x$. Παρότι και οι δύο αναπαραστάσεις δύνανται να πλησιάσουν ασυμπτωτικά την επιθυμητή συνάρτηση, η δεύτερη περίπτωση απλοποιεί πολύ διαδικασία εκπαίδευσης του δικτύου. Η υπολειπόμενη μάθηση υιοθετείται με μερικά από τα στοιβαγμένα επίπεδα.

Προκειμένου να υλοποιηθούν οι παραπάνω φορμαλισμοί, χρησιμοποιείται η μέθοδος των παρακάμφσεων (Shortcuts) [22] ανά κάποιο αριθμό δομικών συνόλων (building block). Υπό αυτή την έννοια, ένα δομικό σύνολο ορίζεται εν γένει ως:

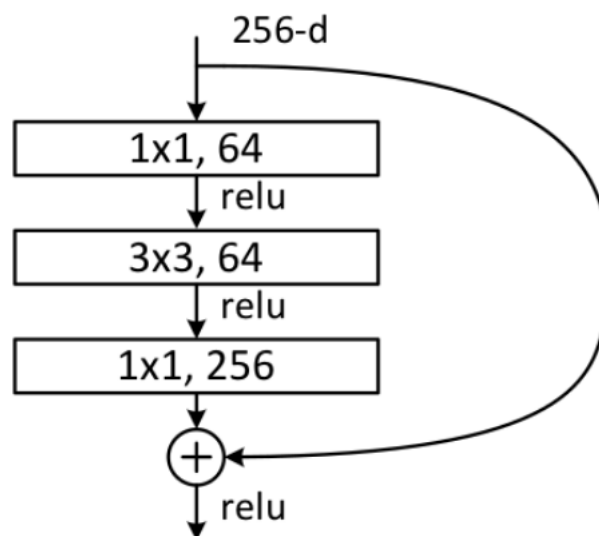
$$y = f(x, \{W_i\}) + x \quad (3.1)$$

όπου x και y είναι τα διανύσματα εισόδου και εξόδου αντίστοιχα σε εξεταζόμενα δομικά σύνολα. Η συνάρτηση $f(x, \{W_i\})$ αναπαριστά την υπολειπόμενη χαρτογράφηση που αποσκοπείται να γνωστοποιηθεί και ο δείκτης i υποδηλώνει τον αριθμό των επιπέδων από τα οποία αποτελείται το κάθε δομικό σύνολο.

Στην παρούσα εργασία χρησιμοποιήθηκε μια από τις πιο βαθιές Αρχιτεκτονικές Συμφόρισης (Deeper Bottleneck Architectures) και συγκεκριμένα το 101-layer ResNet.

Η Αρχιτεκτονική του ResNet101.

Σε αυτή την αρχιτεκτονική, το δομικό σύνολο σχεδιάζεται με την αρχή της συμφόρισης (bottleneck) ως εξής. Για κάθε συνάρτηση σφάλματος F χρησιμοποιείται μια στοίβα τριών επιπέδων. Τα επίπεδα αυτά είναι 1×1 , 3×3 και 1×1 συνελκτικά, με τα 1×1 επίπεδα να είναι υπεύθυνα για την μείωση και αποκατάσταση των διαστάσεων ώστε τα διανύσματα εισόδου-εξόδου του 3×3 επιπέδου να είναι χαμηλότερων διαστάσεων, μειώνοντας έτσι την πολυπλοκότητα. Παρακάτω, στην εικόνα 3.3 φαίνεται μία αναπαράσταση ενός δομικού συνόλου με μία σύνδεση παράκαμψης.

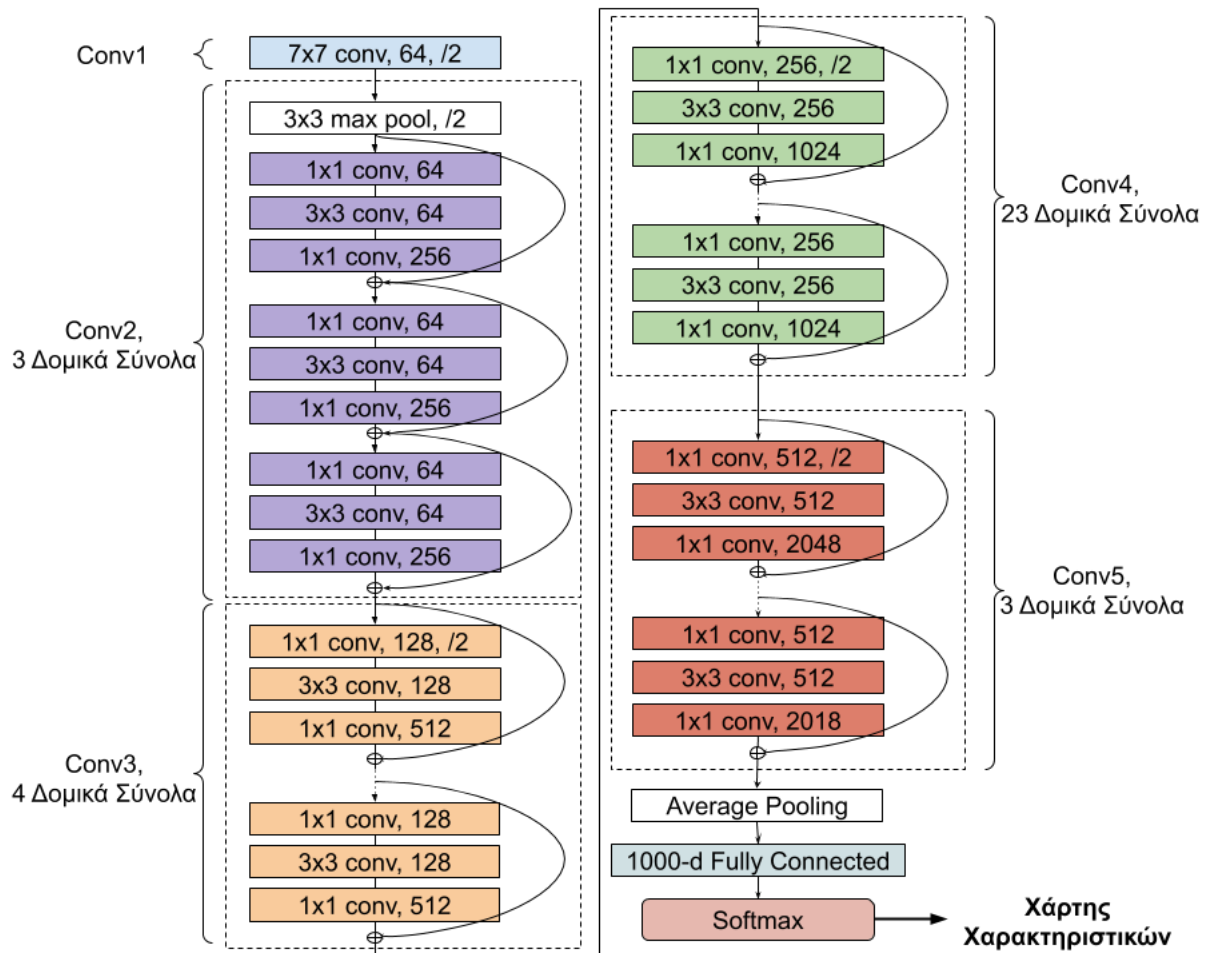


Σχήμα 3.3: Η αρχιτεκτονική κάθε δομικού συνόλου του δικτύου ResNet101

Τίτλος Επιπέδου	Μέγεθος χάρτη χαρακτηριστικών	Αρχιτεκτονική του ResNet101
conv1	112×112	$7 \times 7, 64, \text{stride } 2$
		$3 \times 3, \text{max pool, stride } 2$
conv2	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$
conv4	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$
conv5	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-dc fc, softmax

Πίνακας 3.1: Η αρχιτεκτονική του νευρωνικού δικτύου ResNet101 χωρισμένη στα 5 βασικά της στάδια. Κάθε στάδιο αποτελείται από τα δομικά σύνολα που φαίνονται εντός των παρενθέσεων.

Συνολικά το Δίκτυο έχει την μορφή που φαίνεται στον πίνακα 3.1. Ενώ εποπτικά, η αρχιτεκτονική που περιγράφεται από τον πίνακα 3.1 φαίνεται στην εικόνα 3.4.



Σχήμα 3.4: Η αρχιτεκτονική του δικτύου ResNet101

Το αποτέλεσμα της επεξεργασίας που γίνεται στις εικόνες από το ResNet 101 είναι η εξαγωγή του χάρτη χαρακτηριστικών, ο οποίο θα χρησιμοποιηθεί στο επόμενο ακριβώς βήμα για την δημιουργία περιοχών ενδιαφέροντος στην εικόνα.

2. Δίκτυο Πρότασης Περιοχών (Region Proposal Network - RPN).

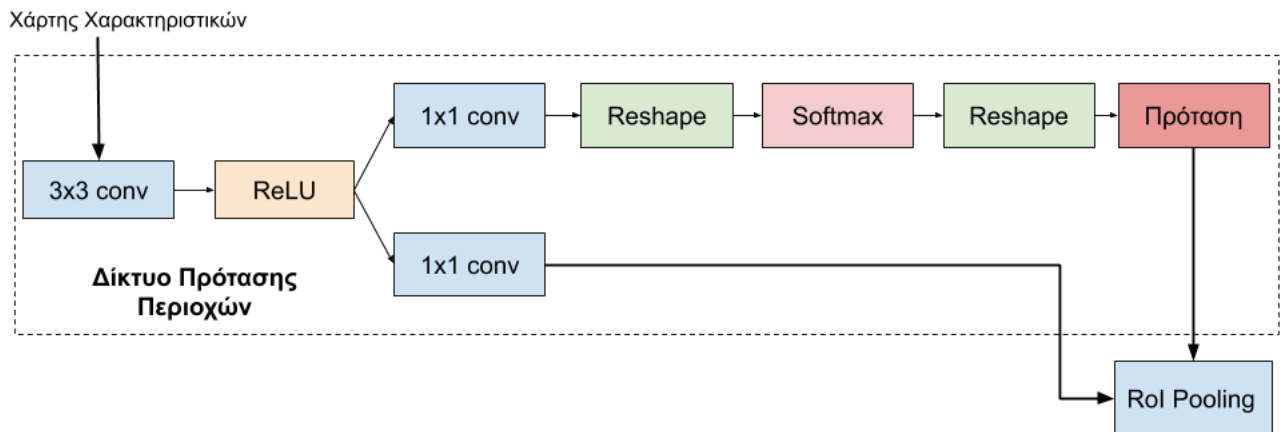
Σε περίπτωση που ζητούνταν απλώς η κατηγοριοποίηση των εικόνων, τότε η επεξεργασία της εικόνας χάρη στο ResNet101 θα αρκούσε. Εντούτοις, ο στόχος της εργασίας σε αυτό το στάδιο είναι η αναγνώριση και κατηγοριοποίηση αντικειμένων εντός της ίδιας εικόνας. Προκειμένου να επιτευχθεί αυτός ο στόχος, το μοντέλο Faster R-CNN εισάγει την έννοια του Δικτύου Πρότασης Περιοχών - RPN [20].

Εν γένει, η βασική αρχή λειτουργίας ενός RPN είναι η εξής, το δίκτυο δέχεται μια εικόνα οποιουδήποτε μεγέθους ως είσοδο και παράγει ως έξοδο ένα σύνολο από προτάσεις ορθογωνικών περιοχών αναθέτοντας παράλληλα στην κάθε μία βαθμολογία που υποδηλώνει το κατά πόσο τοις εκατό αυτή η περιοχή ανήκει σε κάποια κλάση αντικειμένου αντί για το φόντο (objectness score). Με απλά λόγια δηλαδή, το RPN δέχεται ως είσοδο τον

χάρτη χαρακτηριστικών και παράγει ως έξοδο την εικόνα με πλαισιωμένα τα αντικείμενα που έχουν εντοπιστεί εντός της εικόνας καθώς και ένα το ποσο βεβαιότητας εύστοχου εντοπισμού.

Η Αρχιτεκτονική του RPN.

Για να δημιουργηθούν οι προτάσεις περιοχών, ένα μικρό δικτύωμα παρεμβάλεται ανάμεσα στον συνελκτικό χάρτη χαρακτηριστικών που προέκυψε από το στάδιο ResNet101 και το στάδιο Υποδειματοληψίας Περιοχών Ενδιαφέροντος (Region of Interest Pooling - ROI pooling). Αυτό το δίκτυο είναι πλήρως συνδεδεμένο (fully connected) σε ένα 3×3 συρόμενο χωρικό παράθυρο του χάρτη χαρακτηριστικών που χρησιμοποιείται ως είσοδος. Κάθε συρόμενο παράθυρο χαρτογραφείται με την σειρά του σε ένα χαμηλότερης διάστασης διάνυσμα. Το διάνυσμα αυτό διοχετεύεται σε δύο πανομοιότυπα πλήρως συνδεδεμένα επίπεδα, το επίπεδο παλινδρόμησης περιγράμματος (box regression layer - reg) και το επίπεδο ταξινόμησης περιγράμματος (box classification layer - cls). Επιπλέον, χρησιμοποιείται μια συνάρτηση ενεργοποίησης ReLU [25] στην έξοδο του εμβόλιμου 3×3 συνελκτικού δικτύου. Στο Σχήμα 3.5 φαίνεται η γραφική αναπαράσταση του RPN.

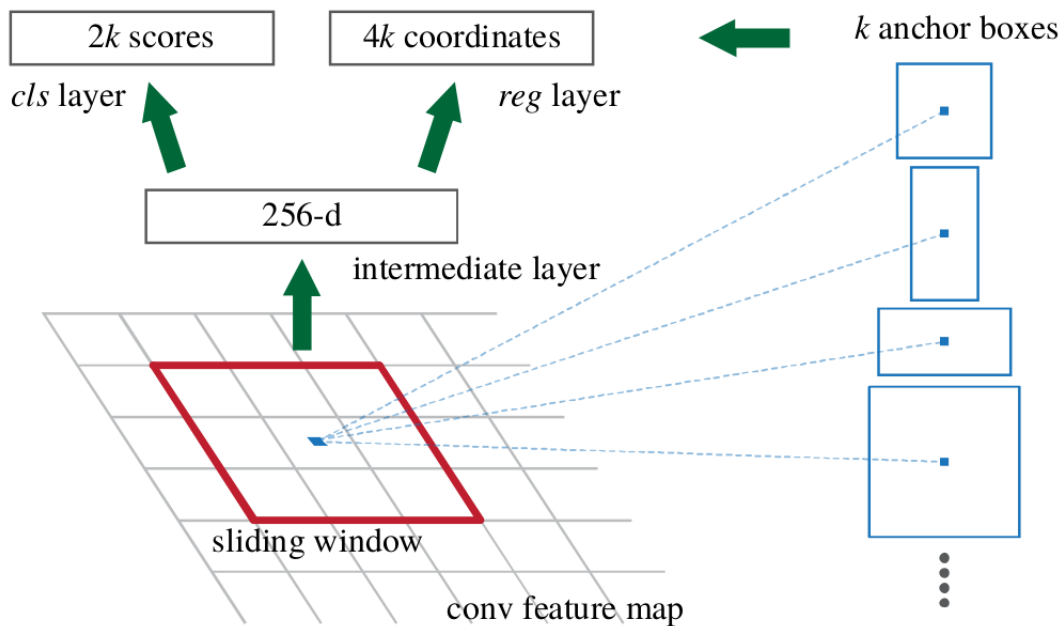


Σχήμα 3.5: Η αρχιτεκτονική του Region Proposal Network

Ο τρόπος λειτουργίας του RPN.

Για κάθε τοποθεσία του συρόμενου παραθύρου, προβλέπονται ταυτόχρονα κ πλαίσια (περιοχές) πρόβλεψης. Το reg επίπεδο όπως ορίστηκε παραπάνω, δίνει ως έξοδο τις συντεταγμένες των κ περιγραμμάτων και άρα παράγει $4κ$ εξόδους (μία για κάθε γωνία των περιγραμμάτων). Το cls επίπεδο που παράγει την εκτίμηση “είναι αντικείμενο / δεν είναι αντικείμενο” για κάθε πρόταση, παράγει $2κ$ εξόδους. Οι κ προβλέψεις παραμετροποιούνται ως προς κ πλαίσια συσχέτισης τα οποία ονομάζονται άγκυρες (anchors). Κάθε άγκυρα ευθυγραμμίζεται με το τρέχον συρόμενο παράθυρο και συσχετίζεται με κάποια

κλίμακα (scale) και κάποιο λόγο μήκους-πλάτους (aspect ratio). Εδώ χρησιμοποιούνται 3 κλίμακες και 3 aspect ratios και άρα παράγονται $K = 9$ άγκυρες για κάθε θέση του συρόμενου παραθύρου. Θεωρώντας ταυτόχρονα ότι ο χάρτης χαρακτηριστικών που προέκυψε από το ResNet101 έχει μέγεθος $W \times H$, παράγονται συνολικά $N = W \times H \times K$ άγκυρες, δηλαδή N συνολικά περιοχές εντός της εικόνας στις οποίες το δίκτυο έχει εντοπίσει ένα αντικείμενο. Ο λόγος που το RPN γεννά τόσο μεγάλο αριθμό προτάσεων είναι σκόπιμος. Αν κάποιο αντικείμενο δεν εντοπιζόταν σε αυτή την πρώτη φάση, θα ήταν αδύνατο να κατηγοριοποιηθεί σε μία από τις κλάσεις των αντικειμένων που ζητείται να εντοπιστούν από το συνολικό δίκτυο. Παρακάτω φαίνεται γραφικά πως λειτουργεί το RPN μέσω των αγκυρών.



Σχήμα 3.6: Αναπαράσταση του τρόπου λειτουργίας του RPN μέσω αγκυρών

Η εκπαίδευση του RPN.

Ανατίθεται θετική ετικέτα σε δύο είδη αγκυρών: (i) η άγκυρα ή άγκυρες με την υψηλότερη τιμή αλληλοεπικάλυψης Intersection-over-Union (IoU) με ένα πλαίσιο ground-truth, ή (ii) η άγκυρα που έχει αλληλοεπικάλυψη IoU υψηλότερη του 0.7 με οποιοδήποτε πλαίσιο ground-truth. Αξιοσημείωτο είναι ότι ένα μοναδικό πλαίσιο ground-truth μπορεί να αναθέσει θετικές ετικέτες σε πολλαπλές άγκυρες. Ανατίθεται αρνητική ετικέτα σε μη-θετικές άγκυρες αν ο λόγος IoU είναι χαμηλότερος του 0.3 ως προς όλα τα πλαίσια ground-truth. Τέλος, άγκυρες στις οποίες δεν έχει ανατεθεί ούτε αρνητική ούτε θετική ετικέτα δεν συνεισφέρουν στην διαδικασία εκπαίδευσης.

Έχοντας ορίσει τις παραπάνω έννοιες, το δίκτυο RPN και κατ'επέκταση το συνολικό

δίκτυο εκπαιδεύεται καθώς συγκλίνει η παρακάτω συνάρτηση κόστους:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (3.2)$$

Εδώ, i είναι ο δείκτης μια άγκυρας σε ένα μικρο-σύνολο και p_i είναι η προβλεπόμενη πιθανότητα η άγκυρα i να είναι αντικείμενο. Η ετικέτα του ground-truth p_i^* είναι 1 αν η άγκυρα είναι θετική και 0 αν η άγκυρα είναι αρνητική. Το t_i είναι το διάνυσμα που αναπαριστά τις 4 παραμετροποιημένες συντεταγμένες του προβλεπόμενου περιγράμματος, και t_i^* είναι το πλαίσιο ground-truth σχετιζόμενο με θετική άγκυρα. Το σφάλμα κατηγοριοποίησης (classification loss) L_{cls} είναι λογαριθμικό σφάλμα πάνω στις δύο κλάσεις (“αντικείμενο” vs “οχι αντικείμενο”). Για το σφάλμα παλινδρόμησης (regression loss) χρησιμοποιείται η ποσότητα $L_{reg}(t_i, t_i^*) = R(t_i - t_i^*)$ όπου το R είναι η εύρωστη συνάρτηση κόστους που περιγράφεται στο [26]. Ο όρος $p_i * L_{reg}$ υποδηλώνει ότι το σφάλμα παλινδρόμησης ενεργοποιείται μόνο για θετικές άγκυρες (δηλαδή με $p_i^* = 1$). Οι έξοδοι των επιπέδων cls και reg όπως αυτά περιγράφονται παραπάνω αποτελούνται από τα μεγέθη $\{p_i\}$ και $\{t_i\}$ αντίστοιχα. Οι δύο αυτοί όροι κανονικοποιούνται μέσω των N_{cls} και N_{reg} αντίστοιχα, και εξισορροπούνται χάρη στο βάρος λ .

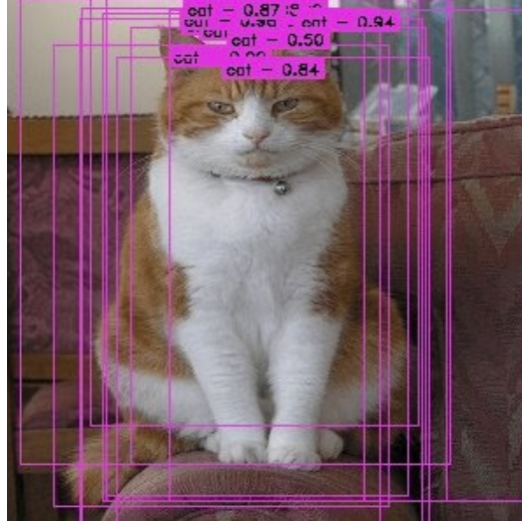
3. Επίπεδο Υποδειγματοληψίας περιοχών ενδιαφέροντος (ROI Pooling).

Ο μεγάλος αριθμός προτάσεων που παράγει το RPN έρχεται με κάποιο κόστος όπως προβλήματα στην απόδοση του συνολικού δικτύου που δυσκολεύουν την πραγματοποίηση αναγνώρισης αντικειμένων σε πραγματικό χρόνο ή το γεγονός ότι καθίσταται αδύνατο να εκπαιδευτεί το συνολικό δίκτυο με ένα πέρασμα.

Για να υπερκεραστούν τέτοιου είδους προβλήματα, γίνεται χρήση του επιπέδου Υποδειγματοληψίας Περιοχών Ενδιαφέροντος (ROI Pooling) [27]. Το δίκτυο αυτό δέχεται δύο εισόδους:

- Έναν σταθερό χάρτη χαρακτηριστικών προερχόμενο από ένα βαθύ συνελικτικό νευρωνικό δίκτυο. Στην παρούσα εργασία ο χάρτης αυτός παρέχεται στο ROI Pooling από το ResNet101.
- Έναν πίνακα $N \times 5$ ο οποίος αναπαριστά τις N περιοχές ενδιαφέροντος. Η πρώτη στήλη περιέχει τον δείκτη της εικόνας και οι υπόλοιπες 4 τις συντεταγμένες των τεσσάρων γωνιών των περιγραμμάτων που αναπαριστούν τις περιοχές ενδιαφέροντος.

Στην συνέχεια το δίκτυο επεξεργάζεται για κάθε περιοχή ενδιαφέροντος τον χάρτη χαρακτηριστικών εφαρμόζοντας συντελεστές κλίμακας προκαθορισμένου μεγέθους. Με αυτόν τον τρόπο δημιουργούνται υπο-χάρτες χαρακτηριστικών καθορισμένου μεγέθους. Κάθε ένας από αυτούς του υπο-χάρτες είναι το κομμάτι εκείνο του συνολικού χάρτη χαρακτηριστικών που περικλείει ένα αντικείμενο χωρίς να συμπεριλαμβάνει την περίσσεια πληροφορίας που παράγει το RPN γύρω από ένα αντικείμενο, όπως φαίνεται στην εικόνα 3.7



Σχήμα 3.7: Οπτικοποίηση των πλεονάζοντων πλαισίων περιγράμματος που προκύπτουν ως έξοδος του RPN

4. Επίπεδο Κατηγοριοποίησης

Στο τελικό στάδιο του δικτύου εκτελείται η κατηγοριοποίηση των αναγνωρισμένων αντικειμένων σε μία από τις προκαθορισμένες κλάσεις που έχει ορίσει ο μηχανικός μέσω των ετικετών. Πρόκειται για ένα στάδιο αποτελούμενο από διαδοχικά πλήρως συνδεδεμένα επίπεδα (fully connected), ένα τμήμα σιγμοειδούς συνάρτησης ενεργοποίησης (softmax) και ένα τμήμα παλινδρόμησης των πλαισίων περιγράμματος. Το τμήμα της softmax κατηγοριοποίησης παράγει τις πιθανότητες κάθε Περιοχής Ενδιαφέροντος (RoI) να ανήκει σε μία από τις k κατηγορίες ή την κατηγορία του φόντου. Το τμήμα παλινδρόμησης πλαισίων περιγράμματος χρησιμοποιείται για ένα τελευταίο καλιμπράρισμα των πλαισίων που προέκυψαν από τον στάδιο RPN.

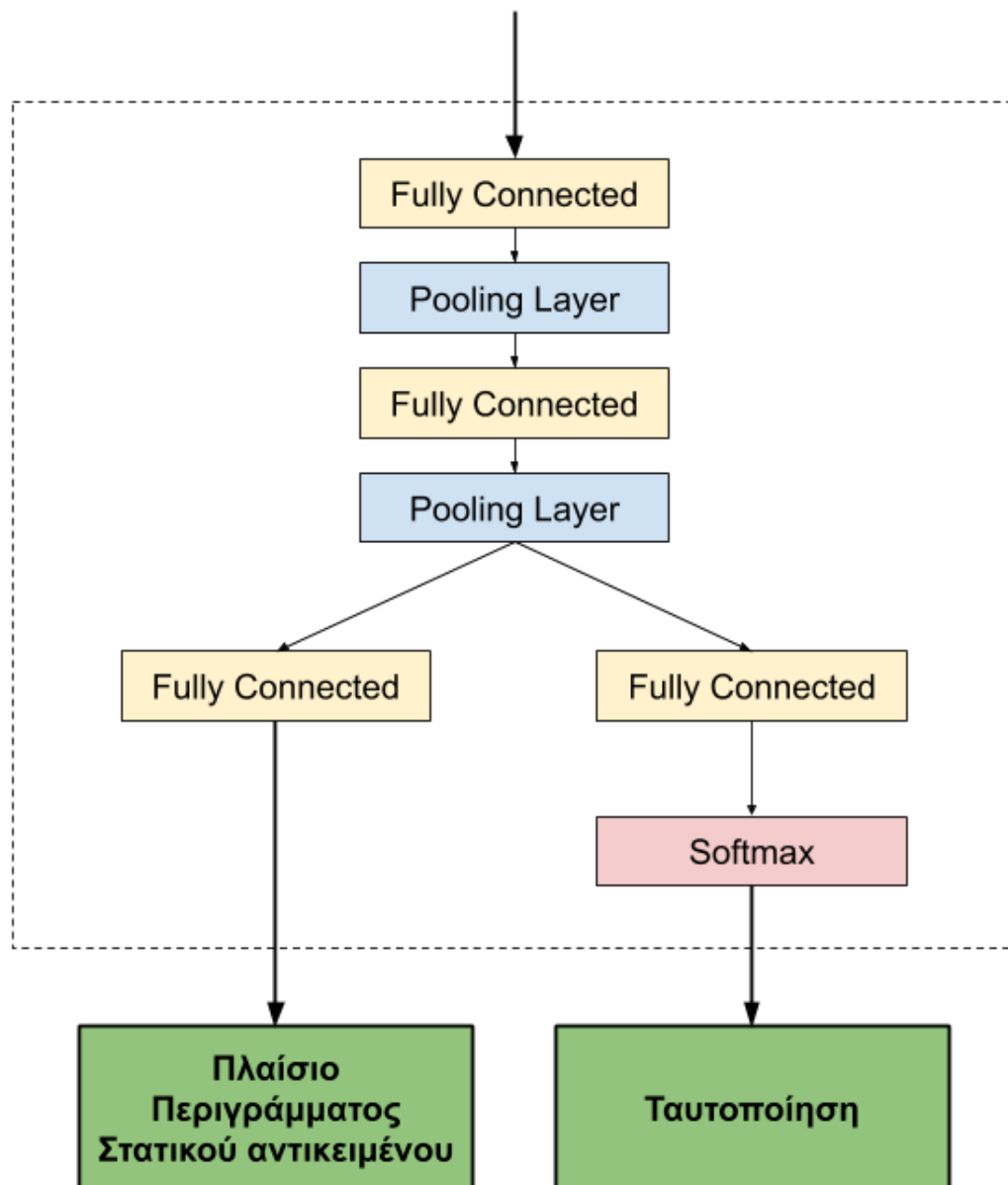
Η αρχιτεκτονική αυτού του σταδίου φαίνεται στο σχήμα 3.8.

Ο τρόπος λειτουργίας.

Η εκπαίδευση του τελικού σταδίου γίνεται μέσω ελαχιστοποίησης της συνάρτησης κόστους L που απεικονίζεται στην σχέση 3.3. Πρόκειται για έναν συνδυασμό της ελαχιστοποίησης σφάλματος του τμήματος κατηγοριοποίησης (L_{cls}) και του τμήματος παλινδρόμησης πλαισίων περιγράμματος (L_{loc}).

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v) \quad (3.3)$$

Το διάνυσμα p είναι αυτό των πιθανοτήτων κάθε Περιοχής ενδιαφέροντος να ανήκει σε κάθε μία από τις $k + 1$ κλάσεις με $k = 0, 1, 2, \dots, n$. Το σφάλμα κατηγοριοποίησης $L_{cls}(p, u)$ δίνεται από το μέγεθος $\log(p)$ που είναι το λογαριθμικό σφάλμα ως προς την πραγματική κλάση u .



Σχήμα 3.8: Η αρχιτεκτονική του επιπέδου Κατηγοριοποίησης.

Για την συνάρτηση σφάλματος L_{loc} του σταδίου παλινδρόμησης ισχύει η σχέση 3.4.

$$L_{loc}(t^u, v) = \sum_{i \in \{x, y, z, h\}} \text{smooth}_{L_1}(t_i^u - v_i) \quad (3.4)$$

Το στάδιο παλινδρόμησης πλαισίων παράγει 4 μετατοπίσεις πλαισίων παλινδρόμησης (bounding box regression offsets) t_i^k με $i = x, y, w, h$ όπου (x, y) αντιστοιχεί στις συντεταγμένες της πάνω-αριστερά γωνίας του πλαισίου και w, h υποδηλώνουν το πλάτος και το ύψος του πλαισίου. Οι πραγματικοί στόχοι της παλινδρόμησης πλαισίου για μια

κλάση u αντιπροσωπεύονται από την ποσότητα v_i με $i = x, y, w$, και h όταν $u \geq 1$, καθώς το φόντο με $u = 0$ δεν έχει πλαίσιο ground-truth.

3.2 Τεχνικές Όρασης Υπολογιστών - Αλγόριθμος SIFT

Μια από τις δημοφιλέστερες τεχνικές ανίχνευσης σημείων ενδιαφέροντος προτάθηκε από τον Lowe το 1999 με την ονομασία Scale Invariant Feature Transform (SIFT) [28]. Στην ουσία πρόκειται για μια μέθοδο εντοπισμού σημείων ενδιαφέροντος και εξαγωγής χαρακτηριστικών γύρω από κάθε σημείο. Έγινε ιδιαίτερος δημοφιλής τόσο για την απόδοση της όσο και για την ταχύτητα της.

Το στάδιο εντοπισμού (key localization) της μεθόδου SIFT, βασίζεται σε κλιμακώ-χωρικό (scale-space) φιλτράρισμα κάνοντας χρήση του Λαπλασιανού τελεστή σε Γκαουσιανή Κατανομή - **Laplacian of Gaussian (LoG)** για διάφορες τιμές της μεταβλητής σ που αντιπροσωπεύει την κλίμακα. Συγκεκριμένα, η LoG ανιχνεύει συστάδες - blobs σε διάφορα μεγέθη λόγω της αλλαγής του σ . Τα υποψήφια σημεία ενδιαφέροντος προκύπτουν ως τα τοπικά ακρότατα της μορφής (x, y, σ) στο τρισδιάστατο scale-space πεδίο που προκύπτει από το φιλτράρισμα με μια κανονικοποιημένη LoG. Βέβαια, μια τέτοια διαδικασία έχει μεγάλο υπολογιστικό κόστος. Για τον λόγο αυτό, προτάθηκε από τον Lowe μια προσέγγιση της κανονικοποιημένης LoG, που συμβολίζεται ως LoGn, από διαφορές Γκαουσιανών (Difference of Gaussian - DoG). Η προσέγγιση αυτή βασίζεται στην παρατήρηση:

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \simeq \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \quad (3.5)$$

από την οποία προκύπτει ότι φιλτράρισμα με διαφορά Γκαουσιανών (γειτονικής κλίμακας) προσεγγίζει το φιλτράρισμα με LoGn:

$$G(x, y, k\sigma) - G(x, y, \sigma) \simeq (k - 1)\sigma^2 \nabla^2 G \quad (3.6)$$

Συνεπώς τα σημεία ενδιαφέροντος προκύπτουν ως τα τοπικά ακρότατα της απόκρισης DoG φιλτραρίσματος. Η απόκριση αυτή ορίζεται ως:

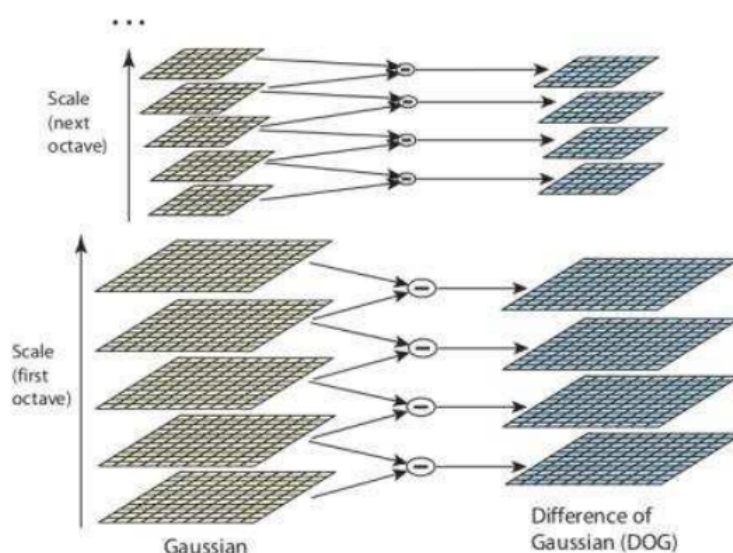
$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (3.7)$$

Η χρήση της διαφοράς Γκαουσιανών, συστηματοποιείται με την χρήση Γκαουσιανών πυραμίδων για πολυκλιμακωτή ανίχνευση σημείων ενδιαφέροντος. Η Γκαουσιανή πυραμίδα μιας εικόνας αποτελεί μια μορφή πολυκλιμακωτής αναπαράστασης. Το βασικό χαρακτηριστικό της προσέγγισης με Γκαουσιανή πυραμίδα είναι η διαδοχική υποδειγματοληψία της εικόνας για κλίμακες που αντιστοιχούν σε δύναμη του 2 ($\sigma = \sigma_0 2^k$,

$k = 1, 2, \dots$). Η υποδειγματοληψία έχει ως αποτέλεσμα τον υποδιπλασιασμό της εικόνας σε κάθε διάσταση. Μια Γκαουσιανή πυραμίδα οργανώνεται σε οκτάβες, που αποτελούν ένα σύνολο από επιμέρους Γκαουσιανά φίλτρα της εικόνας ανάμεσα σε διαδοχικές υποδειγματοληψίες. Για την εξαγωγή των σημείων ενδιαφέροντος με την μέθοδο SIFT προτείνεται η χρήση 4 οκτάβων με 5 επίπεδα κλίμακας σε κάθε οκτάβα. Η κλίμακα για μια οκτάβα k και το i -στο επίπεδο εντός της οκτάβας προκύπτει σύμφωνα με την σχέση $\sigma_{k_i} = (\sigma_0 2^k) \sqrt{2^i}$.

Η εξαγωγή υποψήφιων σημείων ενδιαφέροντος συνοψίζεται ως εξής:

- Δημιουργία Γκαουσιανής πυραμίδας για ένα σύνολο από κλίμακες.
- Υπολογισμός της διαφοράς Γκαουσιανών (DOG) μεταξύ διαδοχικών κλιμάκων εντός κάθε οκτάβας. Η διαδικασία αυτή φαίνεται στο Σχήμα 3.9.
- Εξαγωγή των υποψήφιων σημείων ενδιαφέροντος ως τοπικά ακρότατα σε γειτονίες $3 \times 3 \times 3$ της προκύπτουσας πολυκλιμακωτής αναπαράστασης DOG.



Σχήμα 3.9: Η τεχνική των οκτάβων.

Τα υποψήφια σημεία ενδιαφέροντος που προκύπτουν από την παραπάνω διαδικασία είναι κατά κανόνα πολλά σε πλήθος. Η τελική επιλογή γίνεται ως εξής:

- Εφαρμογή αναπτύγματος Taylor γύρω από τα σημεία ενδιαφέροντος για επανεκτίμηση τους με μεγαλύτερη ακρίβεια (subpixel estimation-interpolation). Αυτή η διαδικασία οδηγεί σε πιο ευσταθή σημεία ενδιαφέροντος που αντιστοιχούν σε πραγματικά κλιμακό-χωρικά (scale-space) ακρότατα. Συγκεκριμένα, γίνεται χρήση του αναπτύγματος Taylor (προσέγγιση με το πολύ τετραγωνικούς όρους σφάλματος) της $D(x, y, \sigma)$ θεωρώντας ως σημείο αναφοράς το εκάστοτε σημείο ενδιαφέροντος $x_0 = (x_0, y_0, \sigma_0)$. Αν υπολογίσουμε την D και τις παραγώγους της στο

σημείο x_0 και ορίσουμε ως $x = (x, y, \sigma)$ την μετατόπιση από το σημείο αναφοράς x_0 , η D μπορεί να γραφτεί ως:

$$D(x) = D(x_0) + \left(\frac{\partial D}{\partial x} \right)^\top x + \frac{1}{2} x^\top \frac{\partial^2 D}{\partial x^2} x \quad (3.8)$$

- Για να υπολογιστεί το νέο ακρότατο, με μεγαλύτερη ακρίβεια, αρκεί να βρούμε τα σημεία μηδενισμού της παραγώγου της παραπάνω συνάρτησης, από όπου προκύπτει:

$$\hat{x} = \left(\frac{\partial^2 D}{\partial x^2} \right)^{-1} \frac{\partial D}{\partial x} \quad (3.9)$$

- Απόρριψη σημείων χαμηλής αντίθεσης (low contrast). Το στάδιο αυτό γίνεται με χρήση ενός προκαθορισμένου κατωφλίου για την φωτεινότητα της προκύπτουσας $D(x, y, \sigma)$ για το κάθε σημείο ενδιαφέροντος. Επειδή έχει γίνει λεπτομερής εκτίμηση των σημείων σύμφωνα με το παραπάνω βήμα, για τον υπολογισμό της φωτεινότητας στα τελικά σημεία ενδιαφέροντος θα χρησιμοποιήσουμε επίσης ανάπτυγμα Taylor:

$$D(\hat{x}) = D + \frac{1}{2} \left(\frac{\partial D}{\partial x} \right)^\top \hat{x} \quad (3.10)$$

- Τέλος, λόγω της φύσης του LoG τελεστή και της προσέγγισής του, DoG, είναι πιθανό να προκύψουν πολλά σημεία που αντιστοιχούν σε ακμές. Τέτοια σημεία θα απορριφθούν κάνοντας χρήση κάποιου κριτηρίου γωνιότητας.

3.3 Φωτογραμμετρική Οπισθοτομία (Space resection).

Φωτογραμμετρική Οπισθοτομία είναι εκείνη η διαδικασία με την οποία προσδιορίζονται τα έξι στοιχεία του εξωτερικού προσανατολισμού μιας λήψης χρησιμοποιώντας τις φωτογραφικές μετρήσεις σημείων ελέγχου ενός αντικειμένου του οποίου οι συντεταγμένες των χωρικών σημείων ελέγχου XYZ είναι γνωστές. Στις παραμέτρους αυτές συμπεριλαμβάνονται οι τρεις συντεταγμένες θέσης του μέσου λήψης (X_L, Y_L, Z_L) καθώς και τρία στοιχεία γωνιακής περιστροφής (ω, ϕ, κ) . Το μέσο με το οποίο επιτυγχάνεται αυτό είναι οι εξισώσεις συγγραμμικότητας.

3.3.1 Συνθήκη Συγγραμμικότητας (Collinearity Condition)

Η συνθήκη συγγραμμικότητας ορίζει ότι για οποιαδήποτε φωτογραφία, ο σταθμός έκθεσης (το κέντρο λήψης της κάμερας - exposure station) κείται στην ίδια ευθεία με κάθε σημείο ελέγχου του αντικειμένου και το αντίστοιχο σημείο εντός της εικόνας [29].

Στην Φωτογραμμετρική Οπισθοτομία μέσω συγγραμμικότητας απαιτούνται οι συντεταγμένες τουλάχιστον τριών σημείων του αντικειμένου και οι αντίστοιχες συντεταγμένες

σημείων εικόνας. Οι βασικές εξισώσεις συγγραμμικότητας που πρέπει να λυθούν για την εύρεση των εξωτερικών παραμέτρων (extrinsic parameters) ορίζονται παρακάτω:

$$x_a = x_0 - f \frac{m_{11}(X_A - X_L) + m_{21}(Y_A - Y_L) + m_{31}(Z_A - Z_L)}{m_{13}(X_A - X_L) + m_{23}(Y_A - Y_L) + m_{33}(Z_A - Z_L)} \quad (3.11)$$

$$y_a = y_0 - f \frac{m_{12}(X_A - X_L) + m_{22}(Y_A - Y_L) + m_{32}(Z_A - Z_L)}{m_{13}(X_A - X_L) + m_{23}(Y_A - Y_L) + m_{33}(Z_A - Z_L)} \quad (3.12)$$

Σε αυτό το σύστημα εξισώσεων τα επιμέρους μεγέθη είναι οι εξής:

- (x_0, y_0, f) είναι οι εσωτερικές παράμετροι της κάμερας (intrinsic parameters) οι οποίες γίνεται να ανακτηθούν μέσω τεχνικών βαθμονόμησης της κάμερας (camera calibration) [30].
- (X_L, Y_L, Z_L) είναι οι χωρικές συντεταγμένες την κέντρου λήξης, δηλαδή της κάμερας.
- m_{ij} είναι τα εννέα συνημίτονα κατεύθυνσης που σχετίζονται με τον εξωτερικό προσανατολισμό της κάμερας και έχουν τιμές:

$$m_{11} = \cos \phi \cos \kappa$$

$$m_{12} = \sin \omega \sin \phi \cos \kappa + \cos \omega \sin \phi$$

$$m_{13} = -\cos \omega \sin \phi \cos \kappa + \sin \omega \sin \kappa$$

$$m_{21} = -\cos \phi$$

$$m_{22} = -\sin \omega \sin \phi \sin \kappa + \cos \omega \cos \kappa$$

$$m_{23} = \cos \omega \sin \phi \sin \kappa + \sin \omega \cos \kappa$$

$$m_{31} = \sin \phi$$

$$m_{32} = -\sin \omega \cos \phi$$

$$m_{33} = \cos \omega \cos \phi$$

- (X, Y, Z) είναι οι χωρικές συντεταγμένες (space coordinates) ενός σημείου ελέγχου και (x, y) είναι οι αντίστοιχες συντεταγμένες εικονοστοιχείων (pixel coordinates) εντός της εικόνας.

3.3.2 Γραμμικοποίηση εξισώσεων συγγραμμικότητας.

Οι εξισώσεις 3.11, 3.12 παρουσιάζουν μεγάλο βαθμό μη γραμμικότητας καθιστώντας απο δύσκολη έως αδύνατη την επίλυσή τους δίχως την γραμμικοποίησή τους. Για την γραμμικοποίησή τους γίνεται χρήση του αναπτύγματος Taylor. Οι μετρήσεις των συντεταγμένων εικόνας x_a, y_a είναι σταθεροί όροι όπως επίσης και οι intrinsic παράμετροι της κάμερας x_0, y_0, f .

Οι εξισώσεις 3.11, 3.12 ξαναγράφονται ως:

$$F = x_0 - f \frac{r}{q} = x_a \quad (3.13)$$

$$G = y_0 - f \frac{s}{q} = y_a \quad (3.14)$$

με:

$$q = m_{13}(X_A - X_L) + m_{23}(Y_A - Y_L) + m_{33}(Z_A - Z_L)$$

$$r = m_{11}(X_A - X_L) + m_{21}(Y_A - Y_L) + m_{31}(Z_A - Z_L)$$

$$s = m_{12}(X_A - X_L) + m_{22}(Y_A - Y_L) + m_{32}(Z_A - Z_L)$$

Στην συνέχεια, σύμφωνα με το Θεώρημα Taylor οι εξισώσεις 3.13,3.14 μπορούν να εκφραστούν σε γραμμικοποιημένη μορφή με χρήση μερικών παραγώγων ως προς τους αγνώστους:

$$F_0 + \left(\frac{\partial F}{\partial \omega} \right)_0 d\omega + \left(\frac{\partial F}{\partial \phi} \right)_0 d\phi + \left(\frac{\partial F}{\partial \kappa} \right)_0 d\kappa + \left(\frac{\partial F}{\partial X_L} \right)_0 dX_L + \left(\frac{\partial F}{\partial Y_L} \right)_0 dY_L + \left(\frac{\partial F}{\partial Z_L} \right)_0 dZ_L = x_a \quad (3.15)$$

$$G_0 + \left(\frac{\partial G}{\partial \omega} \right)_0 d\omega + \left(\frac{\partial G}{\partial \phi} \right)_0 d\phi + \left(\frac{\partial G}{\partial \kappa} \right)_0 d\kappa + \left(\frac{\partial G}{\partial X_L} \right)_0 dX_L + \left(\frac{\partial G}{\partial Y_L} \right)_0 dY_L + \left(\frac{\partial G}{\partial Z_L} \right)_0 dZ_L = y_a \quad (3.16)$$

Εδώ, F_0 και G_0 είναι οι συναρτήσεις F και G αποτιμημένες στις αρχικές συνθήκες των 9 αγνώστων. Οι όροι εντός των παρενθέσεων είναι οι μερικές παράγωγοι των συναρτήσεων F και G ως προς τις υποφαινόμενους αγνώστους, αποτιμημένες στις αρχικές συνθήκες. Τα διαφορικά d στους αγνώστους $\omega, \phi, \kappa, X_L, Y_L, Z_L$ είναι οι άγνωστες βελτιώσεις που θα εφαρμοστούν στις αρχικές συνθήκες.

3.3.3 Λύση Ελαχίστων Τετραγώνων.

Προκειμένου να χρησιμοποιηθούν οι εξισώσεις 3.15,3.16 σε μία λύση με την μέθοδο Ελαχίστων Τετραγώνων, είναι αναγκαίο να συμπεριληφθούν όροι σφαλμάτων προκειμένου οι εξισώσεις να είναι συνεπείς. Οι παρακάτω εξισώσεις ενσωματώνουν τους όρους σφαλμάτων:

$$b_{11}d\omega + b_{12}d\phi + b_{13}d\kappa - b_{14}dX_L - b_{15}dY_L - b_{16}dZ_L = J + \nu_{x_a} \quad (3.17)$$

$$b_{21}d\omega + b_{22}d\phi + b_{23}d\kappa - b_{24}dX_L - b_{25}dY_L - b_{26}dZ_L = K + \nu_{y_a} \quad (3.18)$$

Με τους όρους J και K να ισούνται με $x_a - F_0$ και $y_a - G_0$ αντίστοιχα. Οι όροι b είναι συντελεστές που ισούνται με τις μερικές παραγώγους των εξισώσεων 3.15,3.16 οι οποίοι αναλύονται παρακάτω με $\Delta X, \Delta Y, \Delta Z$ να ισούνται με $X_A - X_L, Y_A - Y_L, Z_A - Z_L$ αντίστοιχα. Οι αριθμητικές τιμές αυτών των συντελεστών αποκτώνται χρησιμοποιώντας τις αρχικές προσεγγίσεις των αγνώστων.

$$\begin{aligned} b_{11} &= \frac{f}{q^2} [r(-m_{33}\Delta Y + m_{32}\Delta Z) - q(-m_{13}\Delta Y + m_{12}\Delta Z)] \\ b_{12} &= \frac{f}{q^2} [r(\cos \phi \Delta X + \sin \omega \sin \phi \Delta Y - \cos \omega \sin \phi \Delta Z) \\ &\quad - q(-\sin \phi \cos \kappa \Delta X + \sin \omega \cos \phi \cos \kappa \Delta Y - \cos \omega \cos \phi \cos \kappa \Delta Z)] \\ b_{13} &= -\frac{f}{q} (m_{21}\Delta X + m_{22}\Delta Y + m_{23}\Delta Z) \\ b_{14} &= \frac{f}{q^2} (rm_{31} - qm_{11}) \\ b_{15} &= \frac{f}{q^2} (rm_{32} - qm_{12}) \\ b_{16} &= \frac{f}{q^2} (rm_{33} - qm_{13}) \\ J &= x_a - x_0 + f \frac{r}{q} \\ \\ b_{21} &= \frac{f}{q^2} [s(-m_{33}\Delta Y + m_{32}\Delta Z) - q(-m_{23}\Delta Y + m_{22}\Delta Z)] \\ b_{22} &= \frac{f}{q^2} [s(\cos \phi \Delta X + \sin \omega \sin \phi \Delta Y - \cos \omega \sin \phi \Delta Z) \\ &\quad - q(-\sin \phi \sin \kappa \Delta X + \sin \omega \cos \phi \sin \kappa \Delta Y - \cos \omega \cos \phi \sin \kappa \Delta Z)] \\ b_{23} &= -\frac{f}{q} (m_{11}\Delta X + m_{12}\Delta Y + m_{13}\Delta Z) \\ b_{24} &= \frac{f}{q^2} (sm_{31} - qm_{21}) \\ b_{25} &= \frac{f}{q^2} (sm_{32} - qm_{22}) \\ b_{26} &= \frac{f}{q^2} (sm_{33} - qm_{23}) \\ K &= y_a - y_0 + f \frac{s}{q} \end{aligned}$$

Κεφάλαιο 4

Υλοποίηση Αλγορίθμου Εντοπισμού Θέσης μέσω Ανίχνευσης Στατικών Αντικειμένων.

Στην ενότητα αυτή παρουσιάζεται η υλοποίηση του αλγορίθμου εντοπισμού θέσης μέσω ανίχνευσης στατικών αντικειμένων όπως περιγράφεται στην δημοσίευση [7]. Επεξηγούνται λεπτομερώς όλοι οι επιμέρους αλγόριθμοι καθώς και τα εργαλεία που χρησιμοποιήθηκαν ώστε να υλοποιηθούν.

4.1 Το πρόβλημα

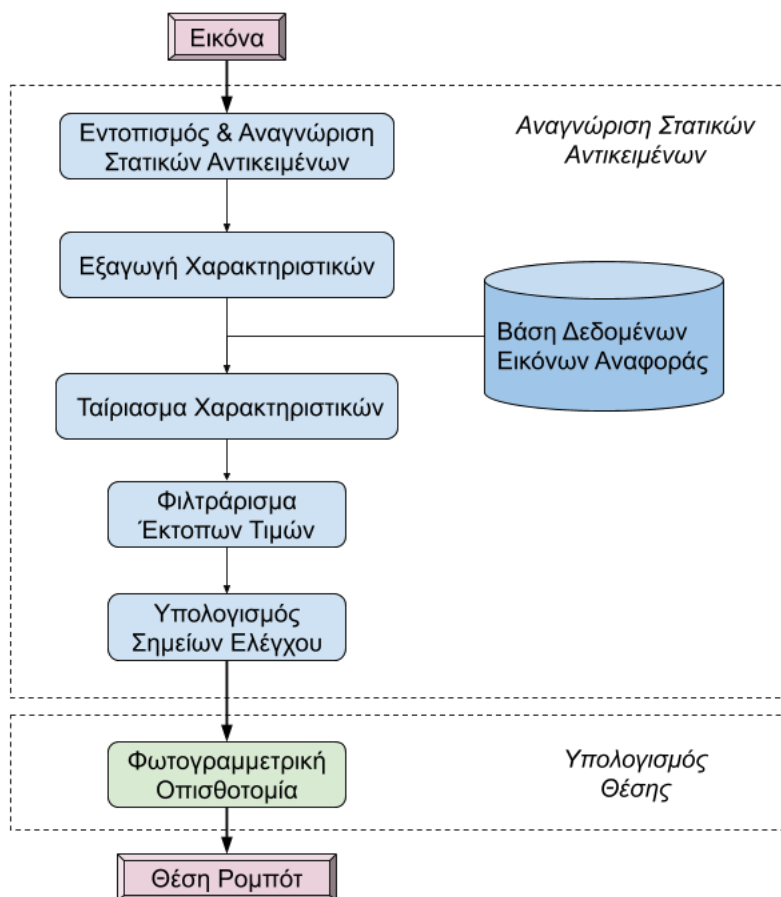
Το πρόβλημα εντοπισμού θέσης μέσω ανίχνευσης στατικών αντικειμένων ορίζεται ως εξής. Αρχικά, ως Στατικά Αντικείμενά ορίζονται αντικείμενα τα οποία είναι κοινά και στάσιμα εντός του εσωτερικού χώρου, είναι ευρέως διάσπαρτα σε αυτόν και περαμένουν σε αυτές τις θέσεις καλοδιατηρημένα για μεγάλο διάστημα. Χαρακτηριστικό παράδειγμα τέτοιων αντικειμένων για ένα εσωτερικό περιβάλλον είναι οι Πόρτες και τα Παράθυρα. Προκειμένου να εντοπιστεί η θέση του ρομπότ εντός του εσωτερικού χώρου, ενσωματώνεται σε αυτό ένας οπτικός αισθητήρας (στην παρούσα περίπτωση η κάμερα ενός έξυπνου κινητού τηλεφώνου - smartphone) ώστε να γίνουν λήψεις του περιβάλλοντος χώρου. Σκοπός είναι το σύστημα να εντοπίσει τα στατικά αντικείμενα που βρίσκονται εντός της κάθε λήψης και στην συνέχεια, συνδυάζονται αλγορίθμοι βαθιάς μάθησης και όρασης υπολογιστών να εντοπίσει την θέση του σημείου σκόπευσης (δηλαδή της κάμερας) και κατ' επέκταση του ρομπότ.

4.2 Περιγραφή της υλοποίησης.

4.2.1 Ο Αλγόριθμος

Η βασική ιδέα του αλγορίθμου είναι να χρησιμοποιηθεί η εικόνα που λαμβάνεται από την ενσωματωμένη κάμερα του ρομπότ προκειμένου να εντοπίζεται η θέση του σε εσωτερικούς χώρους. Η διαδικασία αυτή χωρίζεται σε 2 υπο-ενότητες όπως φαίνεται και στο σχήμα 1.

1. Την αναγνώριση των στατικών αντικειμένων που αποσκοπεί στον εντοπισμό και την αναγνώριση στατικών αντικειμένων εντός της εικόνας και στην συνέχεια στην εξαγωγή σημείων ελέγχου - control points τα οποία συμβάλλουν στον υπολογισμό της θέσης.
2. Τον υπολογισμό της θέσης του ρομπότ χρησιμοποιώντας τα σημεία ελέγχου του προηγούμενου βήματος και την μέθοδο φωτογραμμετρικής οπισθοτομίας.



Σχήμα 4.1: Ο συνολικός αλγόριθμος εντοπισμού της θέσης του ρομπότ χωρίζεται σε 2 στάδια, (i) την αναγνώριση στατικών αντικειμένων στις εικόνες δοκιμής και (ii) τον υπολογισμό της θέσης του ρομπότ.

Ένα ακόμα πολύ βασικό χαρακτηριστικό της μεθόδου είναι ότι αποτελείται από 2 φάσεις, την αποσυνδεδεμένη (offline) και την συνδεδεμένη (online). Στην πρώτη, συμπεριλαμβά-

νται διαδικασίες οι οποίες είναι απαραίτητες για την προετοιμασία της μεθόδου (εκπαίδευση νευρωνικού, ενημέρωση βάσης δεδομένων κλπ) και η δεύτερη πρόκειται για την εν λειτουργία κατάσταση της μεθόδου μέσω της οποία εντοπίζεται η θέση του ρομπότ.

Η υλοποίηση του παραπάνω αλγορίθμου πραγματοποιήθηκε μέσω ανάπτυξης μίας σειράς από python scripts τα οποία υλοποιούν κάθε επιμέρους κομμάτι του αλγορίθμου και στο τέλος συνθέτουν όλα τα κομμάτια σε μία από-άκρη-σε-άκρη δομή (end-to-end pipeline). Όλοι οι κώδικες είναι αναρτημένοι και διαθέσιμοι στο διαδίκτυο σαν repository του github [31].

4.2.2 Το Στάδιο αναγνώρισης Στατικών Αντικειμένων

1. Εντοπισμος και Αναγνωση στατικών αντικειμένων.

Σε αυτό το στάδιο γίνεται χρήση του αλγορίθμου Faster R-CNN όπως αυτό περιγράφηκε στην ενότητα 3.2 προκειμένου να εντοπιστούν εντός της εικόνας που ελήφθη από την κάμερα όλα τα στατικά αντικείμενα. Για την υλοποίηση του αλγορίθμου αυτό χρησιμοποιήθηκε η δομή - framework Tensorflow.

Το tensorflow είναι στην ουσία μια από άκρη-σε-άκρη πλατφόρμα ανοιχτή στο ευρύ κοινό και αφοσιωμένη στην μηχανική μάθηση. Συμπεριλαμβάνει ένα κατανοητό και ευέλικτο οικοσύστημα από εργαλεία, βιβλιοθήκες και κοινοτικούς πόρους που δίνουν την δυνατότητα σε ερευνητές να υλοποιούν εφαρμογές αισιμής στον τομέα της μηχανικής μάθησης καθώς και τους προγραμματιστές να δημιουργήσουν εφαρμογές υποστηριζόμενες από μηχανική μάθηση. Παρέχει σταθερές Διασυνδέσεις Προγραμματισμού Εφαρμογών - Application Programming Interface (API) σε Python και C++ συμβατά για χρήση τόσο σε Κέντρικες Μονάδες Επεξεργασίας (CPU) όσο και με την υποστήριξη Μονάδων Επεξεργασίας Γραφικών (GPU).

Στην παρούσα εργασία χρησιμοποιήθηκε η API TensorFlow Object Detection API [23] η οποία είναι χτισμένη σε ένα επίπεδο πάνω από τον βασικό πυρήνα του TensorFlow. Η API αυτή δίνει την δυνατότητα στον χρήστη να εκτελέσει αναγνώριση πολλαπλών αντικειμένων εντός μιας εικόνας επιλέγοντας μία από τις αρχιτεκτονικές - μοντέλα νευρικών δικτύων τα οποία υλοποιεί. Όλες οι ενέργειες αυτού του βήματος πραγματοποιήθηκαν κατά την διάρκεια της αποσυνδεδεμένης φάσης της μεθόδου.

Προετοιμασία Δεδομένων - Εκπαίδευση Νευρωνικού Δικτύου.

Πρώτη ενέργεια πριν την χρήση του νευρωνικού δικτύου είναι η προ-επεξεργασία των δεδομένων εκπαίδευσης (Data Pre-Processing). Σε αυτή την φάση, αφού έχουν συλλεχτεί όλες οι εικόνες που συντελούν το σύνολο δεδομένων εκπαίδευσης (training dataset), προσαρμόζεται το μέγεθος τους σε 1920x1080 pixel. Στην συνέχεια, είναι αναγκαία η διαδικασία απονομής ετικετών (annotation) στα αντικείμενα που βρίσκονται εντός της κάθε εικόνας αφού η αναγνώριση αντικειμένων αποτελεί υποκατηγορία της επιβλεπόμενης μηχανικής μάθησης.

Για την διαδικασία του annotation χρησιμοποιήθηκε το εργαλείο labelImg [32]. Μέσω αυτού του εργαλείου δημιουργείται χειροκίνητα ένα πλαίσιο περιγράμματος (bounding box) γύρω από κάθε στατικό αντικείμενο που εντοπίζει εντός της εικόνας και του αναθέτει την ανάλογη ταμπέλα (label). Στην συνέχεια προκειμένου να αυξηθεί το σύνολο δεδομένων (dataset augmentation) έγινε περιστροφή των εικόνων του training dataset και των αντίστοιχων πλαισίων γύρω από τα αντικείμενα με την βοήθεια python και bash scripts. Αφού ολοκληρωθεί το Preprocessing των δεδομένων εκπαίδευσης ακολουθεί η διαδικασία της ίδιας της εκπαίδευσης του δικτύου. Όταν η διαδικασία αυτή γίνεται εκ του μηδενός, προκειμένου η εκπαίδευση να είναι επαρκής ώστε το νευρωνικό να καταφέρνει να εντοπίζει τα αντικείμενα και ταυτόχρονα να γενικεύει (αποφυγή overfitting) απαιτούνται datasets με χιλιάδες ή ακόμα και εκατομμύρια εικόνες. Αυτό είναι ένα πολύ σημαντικό πρόβλημα καθώς η διαδικασία συλλογής δεδομένων είναι χρονοβόρα και απαιτεί μεγάλο ανθρώπινο δυναμικό ενώ ταυτόχρονα απαιτούνται πολύ ισχυροί υπολογιστικοί πόροι προκειμένου να εκπαιδευτεί το νευρωνικό σε αποδεκτά χρονικά πλαίσια. Ο τρόπος με τον οποίο το TensorFlow βοηθάει τους χρήστες να υπερκεράσουν αυτό το πρόβλημα είναι παρέχοντάς τους μια σειρά από προ-εκπαιδευμένα μοντέλα όλων των διαθέσιμων αρχιτεκτονικών σε μια μεγάλη ποικιλία από αντικείμενα. Τα μοντέλα αυτά έχουν προ-εκπαιδευτεί στο dataset COCO [33] της Microsoft. Για τα ζητούμενα της παρούσας εργασίας χρησιμοποιήθηκε το προ-εκπαιδευμένο μοντέλο faster_rcnn_resnet101_coco [34]. Για να γίνει το πέρασμα από το γενικευμένο αυτό μοντέλο σε ένα μοντέλο που θα αναγνωρίζει τα ζητούμενα στατικά αντικείμενα, χρησιμοποιήθηκε το Παγωμένο Γράφημα Συμπεράσματος (Frozen Inference Graph) του προ-εκπαιδευμένου μοντέλου προκειμένου να συνεχιστεί η εκπαίδευσή του λαμβάνοντας πλέον και τις κλάσεις των στατικών αντικειμένων. Χάρη σε αυτή την διαδικασία, χρειάστηκαν μόλις 200 εικόνες για το training dataset το οποίο έπειτα από την διαδικασία του augmentation έγινε μεγέθους 800 εικόνων.

Τελικά, μετά το πέρας αυτής της διαδικασίας, προκύπτει το Frozen Inference Graph που χρησιμοποιείται στην συνέχεια για την αναγνώριση των στατικών αντικειμένων κατά εκτέλεση του συνολικού αλγορίθμου με ένασμα εισόδου τις εικόνες ελέγχου (test images).

2. Εξαγωγή συντεταγμένων σημείων Ελέγχου (control points coordinates).

Μετά το πέρας του εντοπισμού των στατικών αντικειμένων εντός της ληφθείσας εικόνας είναι αναγκαίο να εξαχθούν σημεία ελέγχου. Ως σημεία ελέγχου ορίζονται εκείνα τα φυσικά σημεία χαρακτηριστικών των στατικών αντικειμένων στα οποία μπορούμε να μετρήσουμε με ακρίβεια και σχετική τις συντεταγμένες. Κατασκευάζοντας μια σχέση ανάμεσα συντεταγμένες εικονοστοιχείων στην εικόνα και τις χωρικές συντεταγμένες των σημείων ελέγχου είναι δυνατόν να ανακτηθεί η θέση του σταθμού λήψης της φωτογραφίας και κατ' επέκταση του ίδιου του ρομπότ. Συνεπώς, είναι επιτακτική ανάγκη να βρεθούν οι συντεταγμένες εικονοστοιχείων των σημείων ελέγχου στις εικόνες ελέγχου.

Ακολουθείται η εξής στρατηγική.

- Στην αποσυνδεδεμένη φάση, λαμβάνονται εικόνες των στατικών αντικειμένων από διαφορετικές γωνίες λήψης και αποθηκεύονται ως ένα επιπλέον dataset, επονομαζόμενο “εικόνες αναφοράς” - reference images. Σε κάθε μία από αυτές τις εικόνες, μετρούνται ακριβώς οι συντεταγμένες εικονοστοιχείων των σημείων ελέγχου και αποθηκεύονται στην βάση δεδομένων.
- Στην συνέχεια, αφού εντοπιστούν τα υπάρχοντα στατικά αντικείμενα εντός των εικόνων δοκιμής κατά την διάρκεια της συνδεδεμένης φάσης, πραγματοποιείται τόσο σε αυτές, όσο και στις αντίστοιχες εικόνες αναφοράς εξαγωγή χαρακτηριστικών. Για την εξαγωγή χαρακτηριστικών εφαρμόζεται στις εικόνες ο αλγόριθμος SIFT ό τρόπος λειτουργίας του οποίου περιγράφηκε στην ενότητα 3. Η υλοποίηση του πραγματοποιήθηκε με την συνεισφορά της βιβλιοθήκης της Python, OpenCV, η οποία παρέχει μια μεγάλη ποικιλία αλγορίθμων επεξεργασίας εικόνων και όρασης υπολογιστών.
- Έπειτα εφαρμόζεται ένας αλγόριθμος ταιριάσματος χαρακτηριστικών για να εξαχθούν αρκετά ζεύγη χαρακτηριστικών ομωνυμίας. Κατά κανόνα, όταν συνδυάζονται σημεία χαρακτηριστικών που έχουν προκύψει από τον αλγόριθμο SIFT παρατηρούνται λανθασμένα ταιριάσματα τα οποία θα προκαλέσουν προβλήματα στα επόμενα στάδια. Για τον λόγο αυτό είναι απαραίτητο το επόμενο βήμα.
- Φιλτράρισμα των άκυρων ζευγών. Πραγματοποιούνται 2 κύκλοι φιλτραρίσματος. Ο πρώτος είναι σύμφωνα με τον λόγο του Lowe [28] και ο δεύτερος είναι μέσω εφαρμογής του αλγορίθμου Random Sample Consensus - RANSAC [35]. Μέσω αυτών των τεχνικών φιλτραρίσματος προκύπτουν μόνο έγκυρα ζεύγη ομωνυμίας.
- Υπολογισμός του Ομογραφικού Πίνακα (Homography Matrix). Ο πίνακας αυτός αναπαριστά μία χαρτογράφηση που συσχετίζει τα εικονοστοιχεία της εικόνας δοκιμής με την αντίστοιχη εικόνα αναφοράς. Τελικά, οι συντεταγμένες εικονοστοιχείων των σημείων ελέγχου στις εικόνες δοκιμής μπορούν να υπολογιστούν χρησιμοποιώντας τον ομογραφικό πίνακα και τις γνωστές συντεταγμένες εικονοστοιχείων των σημείων ελέγχου στις εικόνες αναφοράς.

Ο αλγόριθμος αυτός φαίνεται παρακάτω ως Αλγόριθμος 1.

4.2.3 Το στάδιο Υπολογισμού Θέσης.

Η γεωμετρική σχέση μεταξύ των σημείων ελέγχου στον χώρο και την εικόνα μπορεί να αναπαρασταθεί με την βοήθεια των εξισώσεων συγγραμμικότητας όπως αυτές περιγράφηκαν στην ενότητα 3.3. Οι εξισώσεις αυτές χρησιμοποιούνται μέσω της μεθόδου Φωτογραμμετρικής Οπισθοτομίας προκειμένου να αξιοποιηθεί η πληροφορία που προέκυψε από το προηγούμενο στάδιο ώστε να υπολογιστεί η θέση του σταθμού λήψης και άρα του ρομπότ.

Για την υλοποίηση της μεθόδου καθώς και την ανάπτυξη της μεθόδου ελαχίστων τετραγώνων δημιουργήθηκε ένα python script και χρησιμοποιήθηκε η βιβλιοθήκη NumPy [42] που

Αλγόριθμος 1. Εξαγωγή των Συντεταγμένων Εικονοστοιχείων των Σημείων Ελέγχου των Εικόνων Δοκιμής

Είσοδος: block της εικόνας δοκιμής που περιέχει Στατικό Αντικείμενο.

Διαδικασία:

- 1: Φέρε από την βάση δεδομένων τις εικόνες αναφοράς που αντιστοιχούν στα Στατικά Αντικείμενα που βρίσκονται στο block εικόνας Δοκιμής;
- 2: Εξήγαγε χαρακτηριστικά στις εικόνες αναφοράς και στο block εικόνας Δοκιμής μέσω του τελεστή SIFT;
- 3: Πραγματοποίησε ταιρίασμα χαρακτηριστικών προκειμένου να δημιουργηθούν ζευγάρια χαρακτηριστικών ομωνυμίας.
- 4: Εκτέλεσε φιλτράρισμα μέσω του νόμου του Lowe και φιλτράρισμα RANSAC ώστε να απορριφθούν τα εσφαλμένα ταιριάσματα; Τα εναπομείναντα ζεύγη βαφτίζονται P_{test} για την εικόνα Δοκιμής και P_{ref} για την εικόνα Αναφοράς;
- 5: Υπολόγισε τον πίνακα Ομογραφίας H_{homo} μέσω επίλυσης της σχέσης:

$$P_{test} = H_{homo} \times P_{ref}$$

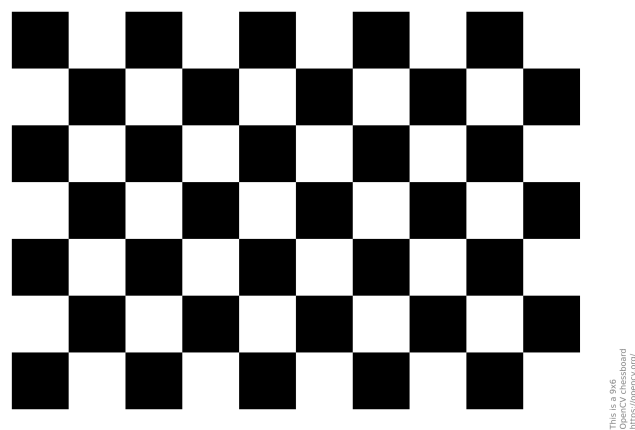
- 6: Υπολόγισε τις συντεταγμένες εικονοστοιχείων των σημείων ελέγχου στην εικόνα Δοκιμής - (CPT) μέσω της παρακάτω σχέσης, όπου CPT_{ref} είναι το σύνολο συντεταγμένων εικονοστοιχείων των σημείων ελέγχου στις εικόνες Αναφοράς:

$$CPT = H_{homo} \times CPT_{ref}$$

Έξοδος: CPT

παρέχει μια ευρεία ποικιλία από μαθηματικά εργαλεία. Σημαντικές παρατηρήσεις για αυτή την ενότητα είναι οι εξής:

- Προκειμένου να υλοποιηθεί η μέθοδος Φωτογραμμετρικής Οπισθοτομίας ήταν απαραίτητο να ανακτηθούν οι εσωτερικές παράμετροι της κάμερας που χρησιμοποιήθηκε (intrinsic parameters). Για να επιτευχθεί αυτό, εκτελέστηκε βαθμονόμηση της κάμερας με την βοήθεια της βιβλιοθήκης OpenCV [36]. Κατά την διαδικασία αυτή, χρησιμοποιήθηκε ένα μοτίβο σκακιέρας όπως φαίνεται στην εικόνα 4.2 το οποίο ελήφθη από την κάμερα από διαφορετικές αποστάσεις και γωνίες ώστε να εξαχθούν τελικά οι εσωτερικές παράμετροι x_0, y_0, f .



Σχήμα 4.2: Το μοτίβο 6×9 που χρησιμοποιήθηκε για την βαθμονόμηση της κάμερας.

- Στην υλοποίηση της μεθόδου ελαχίστων τετραγώνων, χρησιμοποιήθηκαν 2 διακοπτικές συνθήκες, η πρώτη ήταν η συνθήκη σύγκλισης της μεθόδου η οποία ενεργοποιείτο όταν η μεγιστική νόρμα του διανύσματος σφαλμάτων σύγκλισης δεν υπερέβαινε μια ελάχιστη τιμή, η δεύτερη ήταν συνθήκη ελέγχου του αριθμού επαναλήψεων και ενεργοποιείτο έπειτα από 60 επαναλήψεις.

Κεφάλαιο 5

Πείραμα και Αξιολόγηση των Αποτελεσμάτων

Στο κεφάλαιο αυτό θα παρουσιαστούν οι τεχνικές λεπτομέρειες και η πειραματική αξιολόγηση του αλγορίθμου που περιγράφηκε.

5.1 Λεπτομέρειες υλοποίησης

5.1.1 Ο χώρος εργασίας

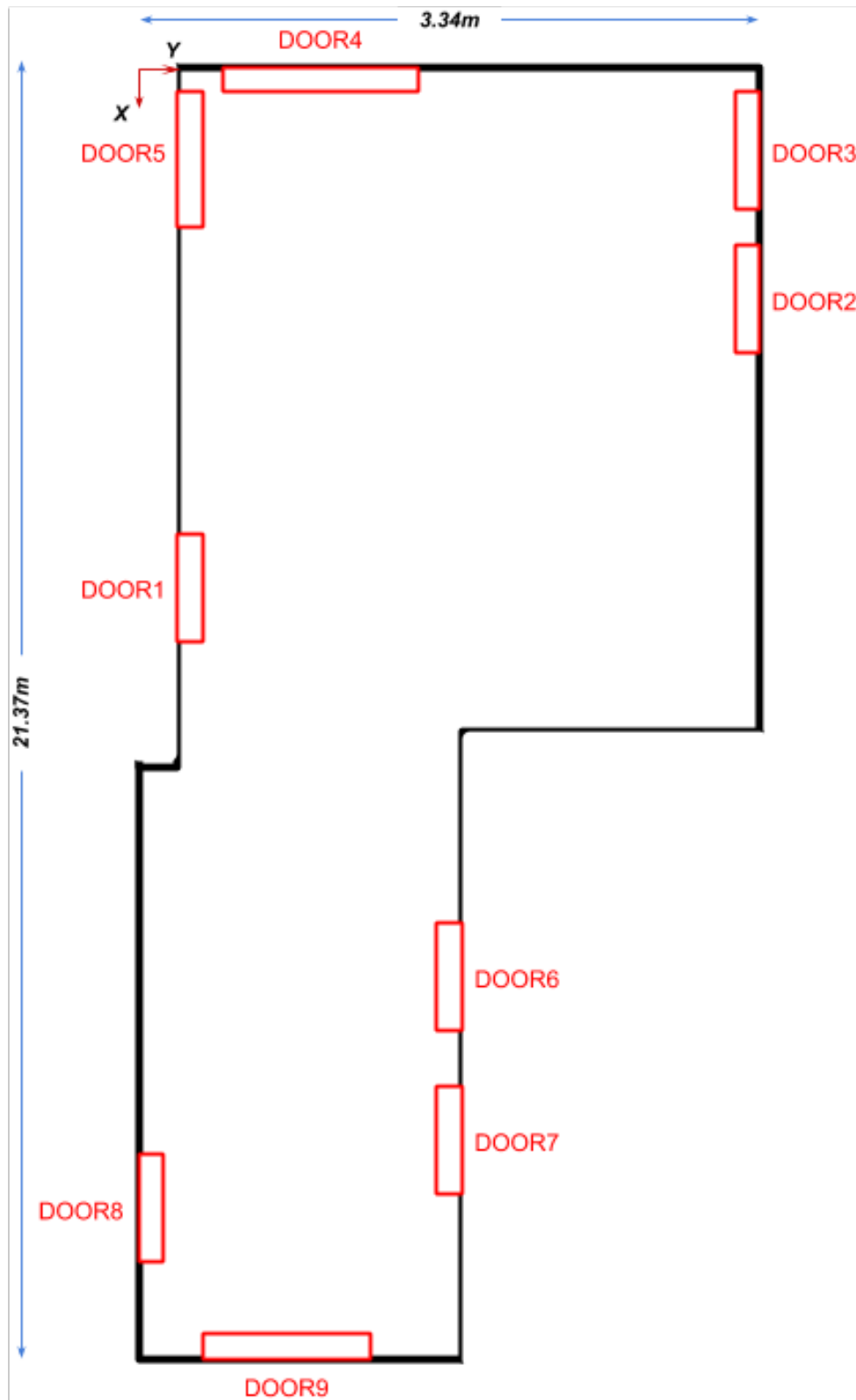
Για την δοκιμή της μεθόδου επιλέχθηκε ως χώρος εργασίας μια εκ των αιθουσών διαδρόμου στα νέα κτίρια της σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσοβίου Πολυτεχνείου. Η κάτοψη του χώρου εργασίας φαίνεται στο σχήμα 5.1 όπως επίσης και το παγκόσμιο πλαίσιο αναφορά (global reference frame) που θεσπίστηκε για την εκτέλεση του εντοπισμού θέσης.

Στον χώρο αυτό υπάρχουν 9 πόρτες που χρησιμοποιούνται ως στατικά αντικείμενα. Σε κάθε μία από τις πόρτες δόθηκε μία ετικέτα της μορφή $DOOR_i$ με $i = 1, 2, \dots, 9$. Ενδεικτικά παρακάτω στο Σχήμα 5.2 φαίνονται οι 5 από τις 9 πόρτες.

5.1.2 Προετοιμασία Πειράματος - Αποσυνδεδεμένη Φάση.

Στην αποσυνδεδεμένη φάση έγιναν οι εξής ενέργειες.

- Λήψη 200 φωτογραφιών των στατικών αντικειμένων από τυχαίες γωνίες λήψης και αποστάσεις ώστε να εκπαιδευτεί το νευρωνικό.
- Μέτρηση όλου του χώρου εργασίας του ρομπότ. Στην φάση αυτή μετρήθηκαν οι συντεταγμένες (X, Y, Z) των χωρικών σημείων ελέγχου των στατικών αντικειμένων και αποθηκεύτηκαν στην βάση δεδομένων.
- Λήψη των εικόνων αναφοράς των στατικών αντικειμένων και καταγραφή των συντεταγμένων εικονοστοιχείων των σημείων ελέγχων των στατικών αντικειμένων. Στο σχήμα



Σχήμα 5.1: Κάτοψη του χώρου εργασίας εκτός κλίμακας για σκοπούς οπτικοποίησης.



(α') Door 1



(β') Door 2



(γ') Door 6



(δ') Door 5



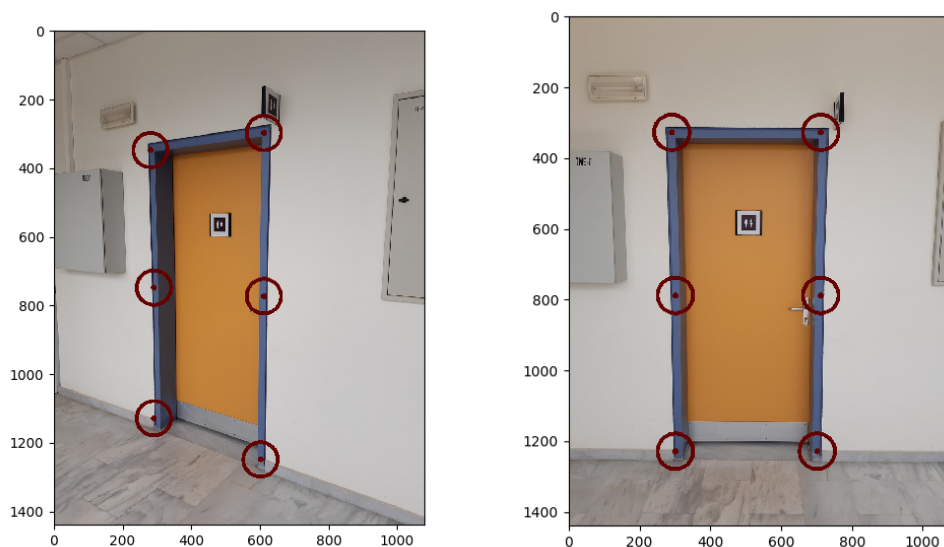
(ε') Door 9

Σχήμα 5.2: Οι πόρτες 1,2,5,6,9 από τις συνολικές 9 που χρησιμοποιούνται ως στατικά αντικείμενα για τον αλγόριθμο.

5.3 φαίνεται 2 εκ των εικόνων αναφοράς του στατικού αντικειμένου DOOR1 μαζί με τα 6 σημεία ελέγχου του.

5.1.3 Αποτελέσματα Υποσταδίων - Συνδεδεμένη Φάση.

Εδώ παρατίθενται ενδεικτικά τα αποτελέσματα των υποσταδίων της μεθόδου κατά την διάρκεια επεξεργασίας μια εικόνας Δοκιμής στα πλαίσια της Συνδεδεμένης φάσης. Σαν είσοδος



(α') Πρόσθια όψη της **Door 1 reference**.

(β') Πλάγια όψη της **Door 1 reference**.

Σχήμα 5.3: Δύο εικόνες αναφοράς του στατικού αντικειμένου DOOR1. Τα τονισμένα σημεία αντιστοιχούν στα σημεία ελέγχου το εν λόγω στατικού αντικειμένου.

στο σύστημα χρησιμοποιήθηκε η παρακάτω εικόνα του σχήματος 5.4.



Σχήμα 5.4: Μία εκ των εικόνων δοκιμής που περιέχει τα στατικά αντικείμενα DOOR4, DOOR5

Στάδιο Αναγνώρισης.

Το πρώτο στάδιο επεξεργασίας είναι ο αλγόριθμος faster R-CNN ο οποίος θα χωρίσει την εικόνα δοκιμής σε image blocks που περιέχουν ένα στατικό αντικείμενο και θα τα κατηγοριοποιήσει σε μία από τις 9 κλάσεις. Στο Σχήμα 5.5 φαίνεται η έξοδος αυτού του σταδίου. Τα δύο αυτά block εικόνας θα χρησιμοποιηθούν ως είσοδος για τον Αλγόριθμο 1 όπως αυτός περιγράφηκε στην ενότητα 4.



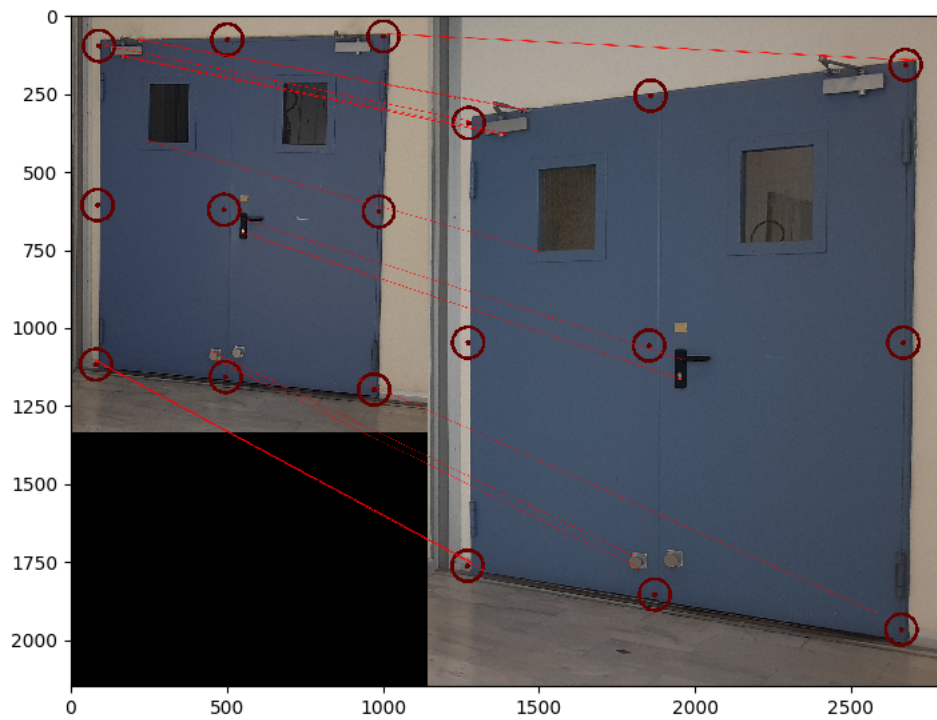
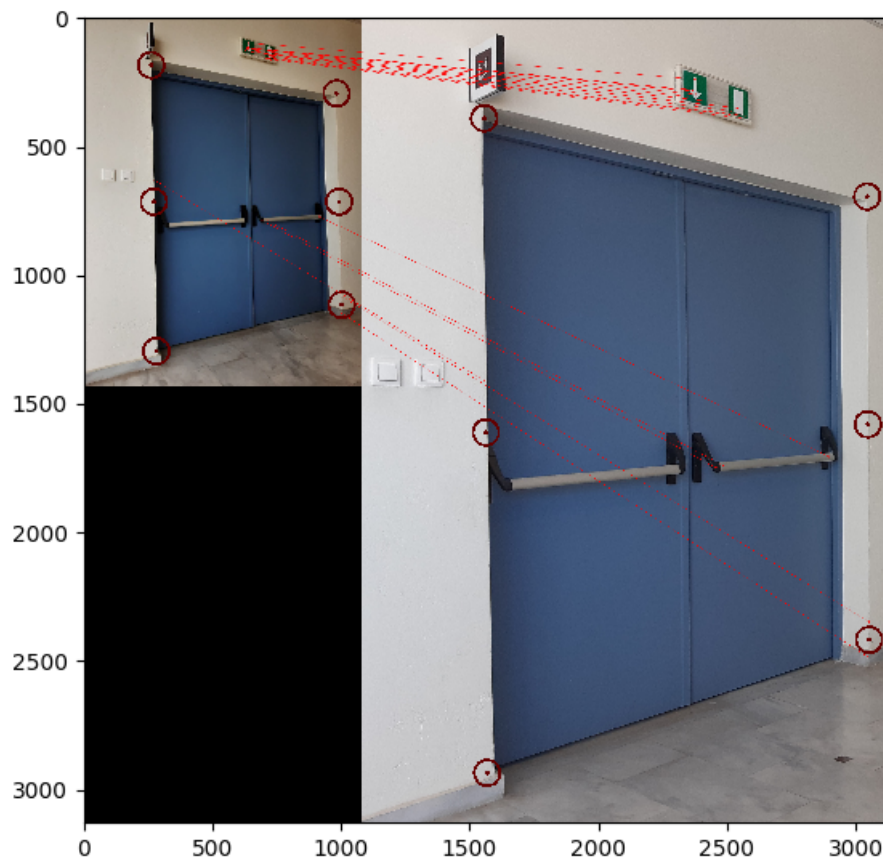
Σχήμα 5.5: Αποτέλεσμα του σταδίου αναγνώρισης. Ο αλγόριθμος faster R-CNN εντόπισε τα στατικά αντικείμενα DOOR4,DOOR5.

Αλγόριθμος 1.

Ο αλγόριθμος 1 ευθύνεται για την εξαγωγή των συντεταγμένων εικονοστοιχείων των στατικών αντικειμένων εντός της εικόνας Δοκιμής.

- Αρχικά, για κάθε block εικόνας που ανιχνεύτηκε απο το νευρωνικό ανακτώνται από την βάση οι αντίστοιχες εικόνες αναφοράς
- Στην συνέχεια πραγματοποιείται ο αλγόριθμος SIFT και ο αλγόριθμος ταιριάσματος ανάμεσα στο block εικόνας και κάθε μία απο τις αντίστοιχες εικόνες αναφοράς.
- Έπειτα εκτελούνται οι δύο κύκλοι φιλτραρισματος και επιλέγεται η εικόνα αναφοράς που δίνει τα περισσότερα έγκυρα ζεύγη χαρακτηριστικών με το block εικόνας ώστε να εξαχθεί ο ομογραφικός πίνακας. Μέσω του ομογραφικού πίνακα αποκτώνται οι συντεταγμένες εικονοστοιχείων των σημείων ελέγχου του block εικόνας δοκιμής. Στο

παρακάτω σχήμα 5.6 παρουσιάζεται το αποτέλεσμα της εν λόγω διαδικασίας για την εικόνα δοκιμής που χρησιμοποιήθηκε.

(α) Στατικό αντικείμενο **Door4**.(β') Στατικό αντικείμενο **Door5**.

Σχήμα 5.6: Το αποτέλεσμα του αλγορίθμου ταιριάσματος. Τα σημεία που είναι τονισμένα από τους κύκλους αναπαριστούν τα σημεία ελέγχου. Τα ενωμένα σημεία μεταξύ των εικόνων είναι τα έγκυρα ζεύγη χαρακτηριστικών μέσω των οποίων φτιάχνεται ο πίνακας H_{homo} .

- Τέλος, υπολογίζονται οι συντεταγμένες εικονοστοιχείων των στατικών αντικειμένων ως μέρος της συνολικής εικόνας Δοκιμής και αποθηκεύονται σε μία λίστα προκειμένου να χρησιμοποιηθούν στο επόμενο στάδιο, την Φωτογραμμετρική Οπισθοτομία. Στο σχήμα 5.7 φαίνεται το τελικό αποτέλεσμα του αλγορίθμου 1.

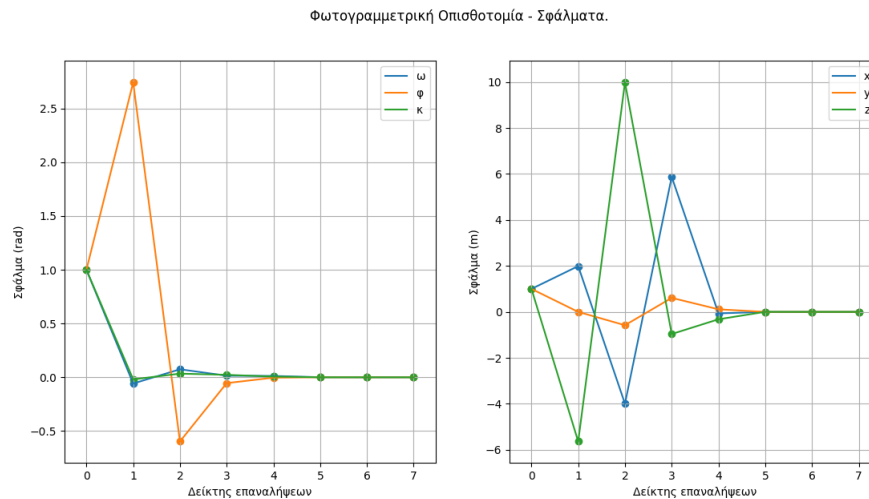


Σχήμα 5.7: Το τελικό αποτέλεσμα του Αλγορίθμου 1. Με διαφορετικά χρώματα φαίνονται τα σημεία ελέγχου που προέκυψαν από διαφορετικά στατικά αντικείμενα.

Φωτογραμμετρική Οπισθοτομία - Τελικό αποτέλεσμα.

Το τελικό στάδιο του συστήματος είναι ο συνδυασμός των συντεταγμένων εικονοστοιχείων που μόλις υπολογίστηκαν, των χωρικών συντεταγμένων και των εσωτερικών παραμέτρων της κάμερας ώστε να υπολογισθεί η θέση του σταθμού λήψης μέσω Φωτογραμμετρικής Οπισθοτομίας.

Οι αρχικές συνθήκες που δίνονται στην μέθοδο είναι μεν τυχαίες, αλλά φροντίζεται να είναι συντεταγμένες εντός του χώρου εργασίας. Στην συντριπτική πλειονότητα των περιπτώσεων, ο αλγόριθμος συγκλίνει σε λιγότερες από 10 επαναλήψεις. Στο σχήμα 5.8 φαίνεται η πορεία των σφαλμάτων των μεταβλητών κατάστασης καθώς εξελίσσεται ο αλγόριθμος για την δεδομένη εικόνα δοκιμής.



Σχήμα 5.8: Η μείωση των σφαλμάτων της κάθε μεταβλητής κατάστασης ω , ϕ , κ , x , y , z καθώς αυξάνεται ο αριθμός επαναλήψεων.

Όταν τελικά συγκλίνει ο αλγόριθμος φωτογραμμετρικής Οπισθοτομίας, δίνει σαν έξοδο την τελική θέση του σταθμού λήψης ως προς το απόλυτο πλαίσιο αναφοράς του χώρου εργασίας όπως αυτό φαίνεται στο Σχήμα 5.1. Για την παρούσα εικόνα δοκιμής, η λήψη έγινε από το σημείο:

$$X_{gt} = 4.80m, Y_{gt} = 3.60m, Z_{gt} = 1.26m$$

και ο αλγόριθμος έδωσε έξοδο:

$$X_L = 4.802m, Y_L = 3.161m, Z_L = 1.188m$$

5.2 Αξιολόγηση της Μεθόδου.

5.2.1 Καθορισμός Πειράματος.

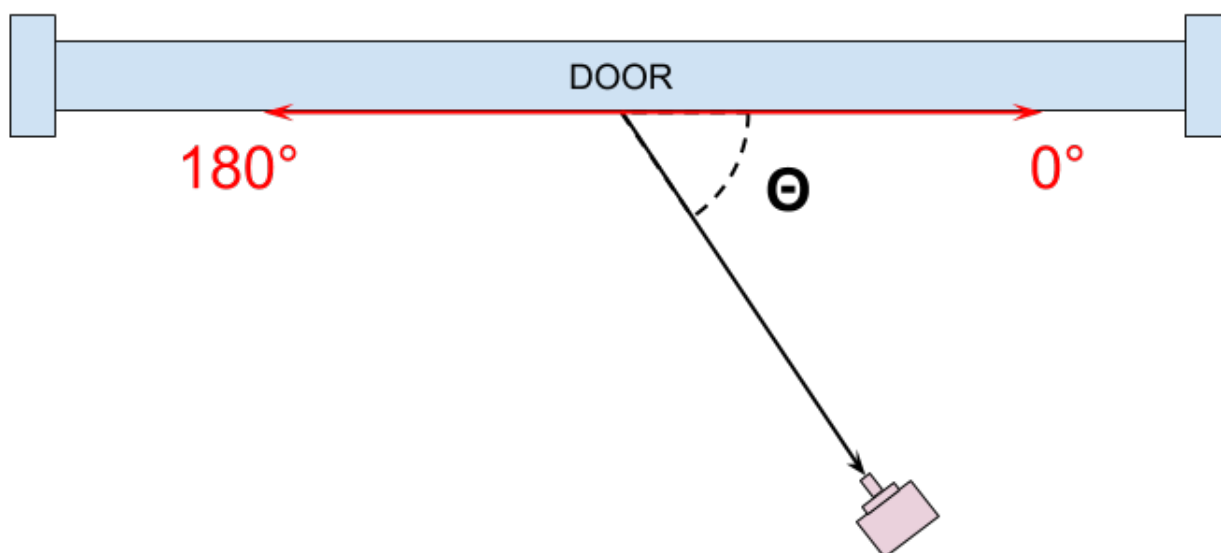
Το πρώτο βήμα για να αξιολογηθεί η μέθοδος είναι να οριστεί το είδος του πειράματος που θα πραγματοποιηθεί. Σκοπός του πειράματος είναι να αξιολογηθεί η δυνατότητα της μεθόδου να εκτελεί επιτυχώς τις παρακάτω ενέργειες:

1. **Έγκυρος εντοπισμός θέσης:** Το ποσοστό στο οποίο καταφέρνει η μέθοδος να παράξει έγκυρο αποτέλεσμα, δηλαδή εντός του χώρου εργασίας του ρομπότ.
2. **Ακρίβεια εντοπισμού θέσης:** Ο Βαθμός ακρίβειας στον εντοπισμό της θέσης, δηλαδή το πόσο κοντά στον ground truth βρίσκεται η πρόβλεψη της μεθόδου. Ως μετρική χρησιμοποιήθηκε η ευκλείδεια απόσταση μεταξύ πρόβλεψης και ground truth.
3. **Χρόνος εκτέλεσης:** Η ταχύτητα στην οποία η μέθοδος παράγει μία πρόβλεψη.

Καθορισμός του ground truth.

Για να ελεγχθεί η μέθοδος έγιναν συνολικά 99 λήψεις. Όλες οι λήψεις πραγματοποιήθηκαν από σταθερό ύψος χάρη στην χρήση ενός τρίποδου και με την κάμερα πάντα κάθετη ως προς δάπεδο. Προκειμένου να ελεγχθεί η επίδραση του φωτός, πραγματοποιήθηκαν 3 φάσεις, η πρωινή, η μεσημεριανή και η απογευματινή, με 33 λήψεις η κάθε μία. Για τις 33 λήψεις της κάθε φάσης, μετρήθηκε ακριβώς το ground truth με την χρήση συμβατού μέτρου και διαβήτη. Τα 33 αυτά σημεία επιλέχθηκαν ως εξής:

- Δεκατέσσερα σημεία χρησιμοποιήθηκαν για να αξιολογηθεί η μέθοδος καθώς η κάμερα εστιάζει σε ένα στατικό αντικείμενο και με σταθερή την γωνία λήψης αλλάζει η απόσταση του σταθμού λήψης από αυτό.
- Δεκατρία σημεία χρησιμοποιήθηκαν για να αξιολογηθεί η μέθοδος καθώς η κάμερα εστιάζει σε ένα στατικό αντικείμενο και με σταθερή την απόσταση του σταθμού λήψης από αυτό, αλλάζει η γωνία σκόπευσης προς αυτό. Ως γωνία λήψης θεωρούμε την γωνία που σχηματίζεται μεταξύ της κάμερας, τους κέντρου της πόρτας και του δεξιού κουφώματος με γνώμονα την φορά σκόπευσης της κάμερας προς την πόρτα. Η γωνία αυτή φαίνεται στο Σχήμα 5.9



Σχήμα 5.9: Η γωνία Θ αντιπροσωπεύει την γωνία σκόπευσης η οποία θα εναλλάσσεται σε θέσεις στο εύρος $[0,180]$ μοιρών για τις διαφορετικές περιπτώσεις δοκιμής.

- Έξι σημεία χρησιμοποιήθηκαν για να αξιολογηθεί η μέθοδος από τυχαία σημεία λήψης.

Στους παρακάτω πίνακες φαίνονται τα ground truths των τριαντατριών λήψεων.

ID	Πόρτες Παρούσες στην Λήψη	X_{gr}	Y_{gr}	Y_{gr}	Απόσταση
1	DOOR4	2.5	1.40	1.26	2.5
2	DOOR4,DOOR5	4.0	1.40	1.26	4.0
3	DOOR4,DOOR5	5.5	1.40	1.26	5.5
4	DOOR3,DOOR4,DOOR5	7.0	1.40	1.26	7.0
5	DOOR2,DOOR3,DOOR4,DOOR5	8.5	1.40	1.26	8.5
6	DOOR2,DOOR4,DOOR5	10.0	1.40	1.26	10
7	DOOR1,DOOR4	11.5	1.40	1.26	11.5
8	DOOR9	18.87	1.40	1.26	2.5
9	DOOR9	17.37	1.40	1.26	4.0
10	DOOR8,DOOR9	15.87	1.40	1.26	5.5
11	DOOR7,DOOR8,DOOR9	14.37	1.40	1.26	7.0
12	DOOR6,DOOR7,DOOR8,DOOR9	12.87	1.40	1.26	8.5
13	DOOR6,DOOR7,DOOR9	11.37	1.40	1.26	10
14	DOOR6,DOOR9	9.87	1.40	1.26	11.5

Πίνακας 5.1: Τα ground truth των σημείων όπου εναλλάσσεται η απόσταση.

ID	Πόρτες Παρούσες στην Λήψη	X_{gr}	Y_{gr}	Y_{gr}	Απόσταση	Γωνία Θ
15	DOOR1,DOOR9	6.48	2.23	1.26	3.0	30
16	DOOR1	7.39	2.94	1.26	3.0	60
17	DOOR1	8.79	3.30	1.26	3.0	90
18	DOOR1	9.99	3.03	1.26	3.0	120
19	DOOR1,DOOR5	10.96	2.27	1.26	3.0	150
20	DOOR2,DOOR3	5.0	2.7	1.26	2.8	30
21	DOOR2,DOOR3	4.03	1.47	1.26	2.8	60
22	DOOR2,DOOR3	2.42	0.93	1.26	2.8	90
23	DOOR2,DOOR3	0.75	1.15	1.26	2.8	120
24	DOOR5	0.86	2.88	1.26	3.0	60
25	DOOR5	2.55	3.3	1.26	3.0	90
26	DOOR5	3.78	3	1.26	3.0	120
27	DOOR4,DOOR5	4.75	2.27	1.26	3.0	150

Πίνακας 5.2: Τα ground truth των σημείων όπου εναλλάσσεται η γωνία θέασης της κάμερας.

ID	Πόρτες Παρούσες στην Λήψη	X_{gr}	Y_{gr}	Y_{gr}
28	DOOR4,DOOR5	4.80	3.60	1.26
29	DOOR2,DOOR3,DOOR4	6.41	0.75	1.26
30	DOOR1,DOOR4,DOOR5	10.54	2.70	1.26
31	DOOR8,DOOR9	15.32	1.76	1.26
32	DOOR6,DOOR7	12.94	0.75	1.26
33	DOOR6,DOOR7,DOOR9	12.94	0.75	1.26

Πίνακας 5.3: Τα ground truth των σημείων που βρίσκονται σε τυχαίες θέσεις.

5.2.2 Αποτελέσματα της Μεθόδου

Σε αυτή την υποενότητα παρουσιάζονται τα αποτελέσματα για τις 99 λήψεις όσον αφορά τις μετρικές που αναφέρθηκαν στο προηγούμενο υποκεφάλαιο.

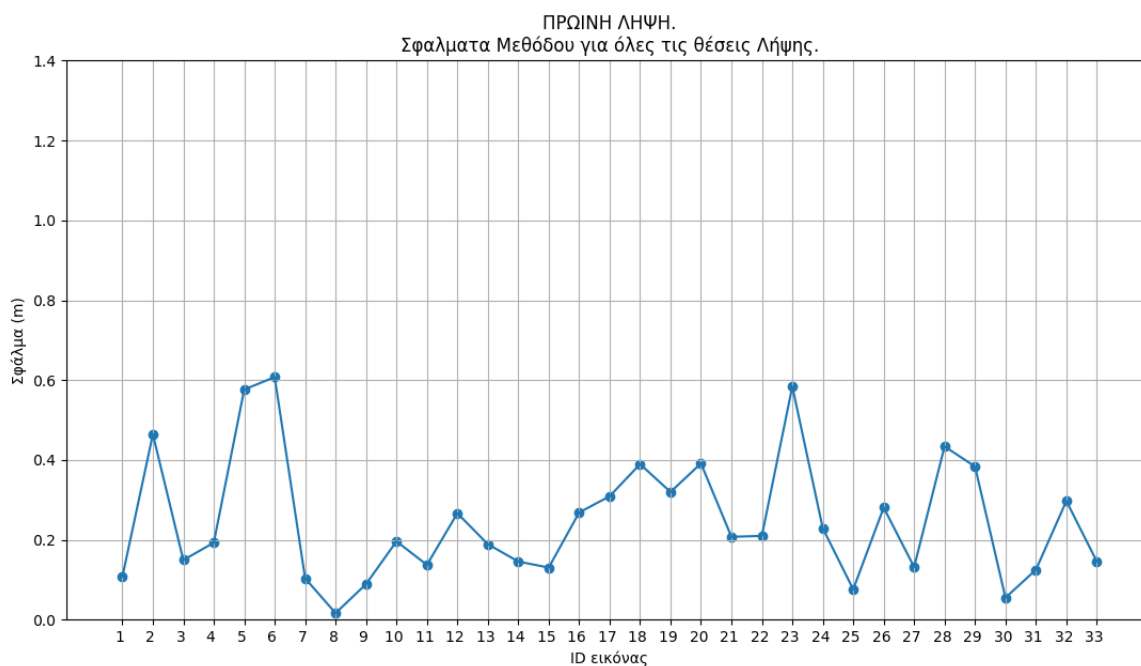
1. Ποσοστό επιτυχίας Μεθόδου

Η μέθοδος καταφέρνει να παράξει έγκυρη πρόβλεψη, δηλαδή πρόβλεψη εντός του χώρου εργασίας του ρομπότ για όλες τις περιπτώσεις δοκιμής που του τέθηκαν.

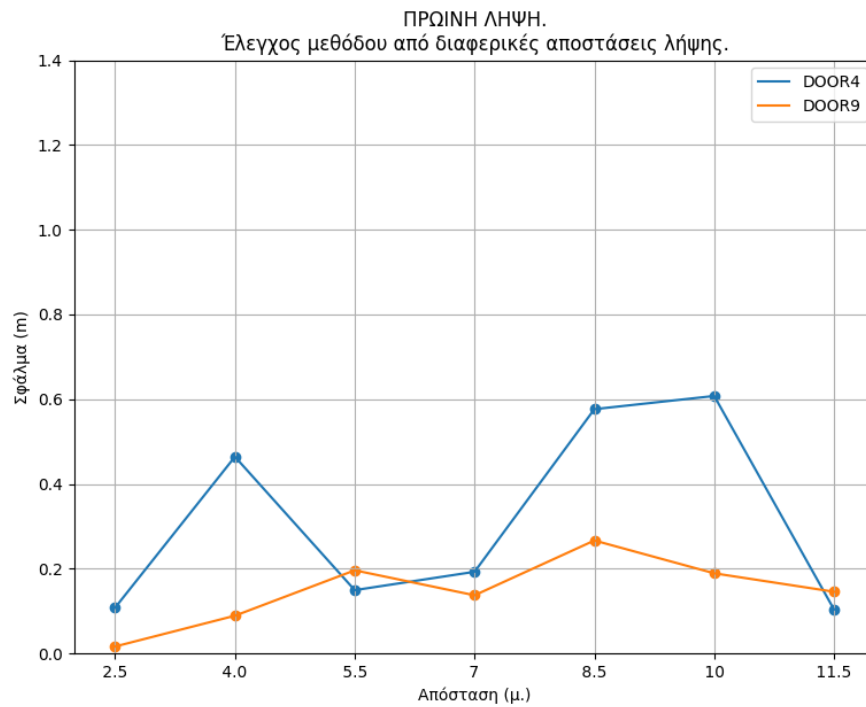
2. Ακρίβεια Μεθόδου

Στα παρακάτω διαγράμματα φαίνονται τα σφάλματα της μεθόδου για κάθε μια από τις περιπτώσεις που προαναφέρθηκαν. Για κάθε χρονική περίοδο λήψης, αρχικά παρουσιάζεται το σύνολο των λήψεων και στην συνέχεια παρουσιάζεται ένα διάγραμμα για τιμές που αντιστοιχούν στον καθένα από τους παραπάνω πίνακες 5.1, 5.2 και 5.3.

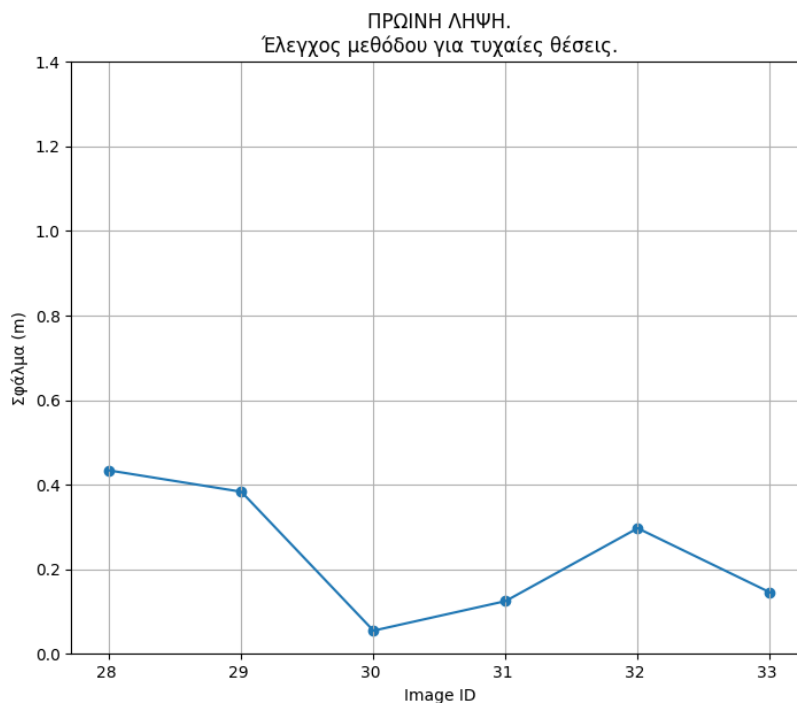
Πρωινές Λήψεις.



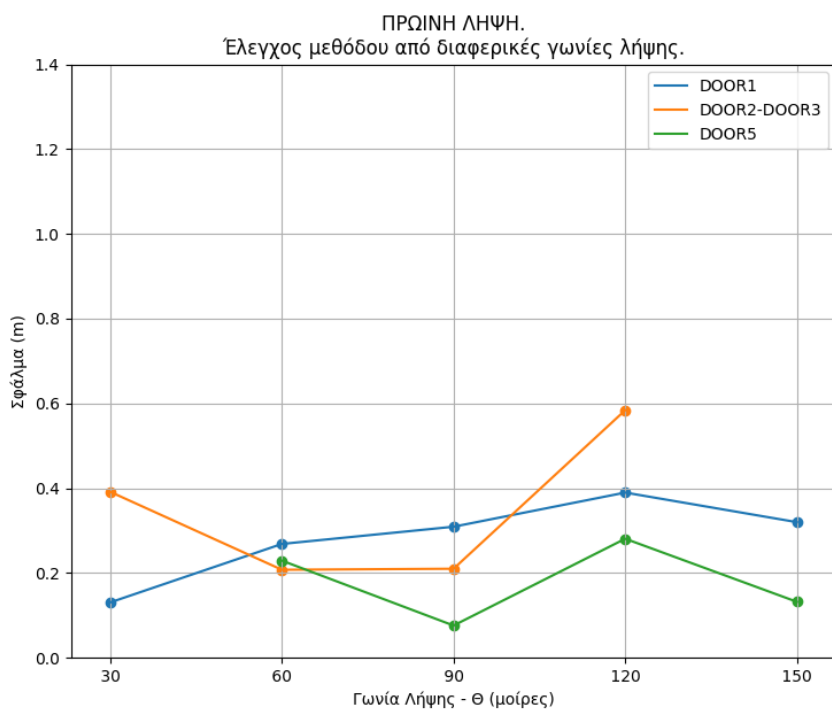
Σχήμα 5.10: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 33 πρωινές λήψεις.



Σχήμα 5.11: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 14 λήψεις στις οποίες εναλλάσσεται η απόσταση. Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη.



Σχήμα 5.13: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 6 λήψεις που πραγματοποιήθηκαν από τυχαίες θέσεις. Στο οριζόντιο άξονα φαίνεται το ID της κάθε εικόνας.

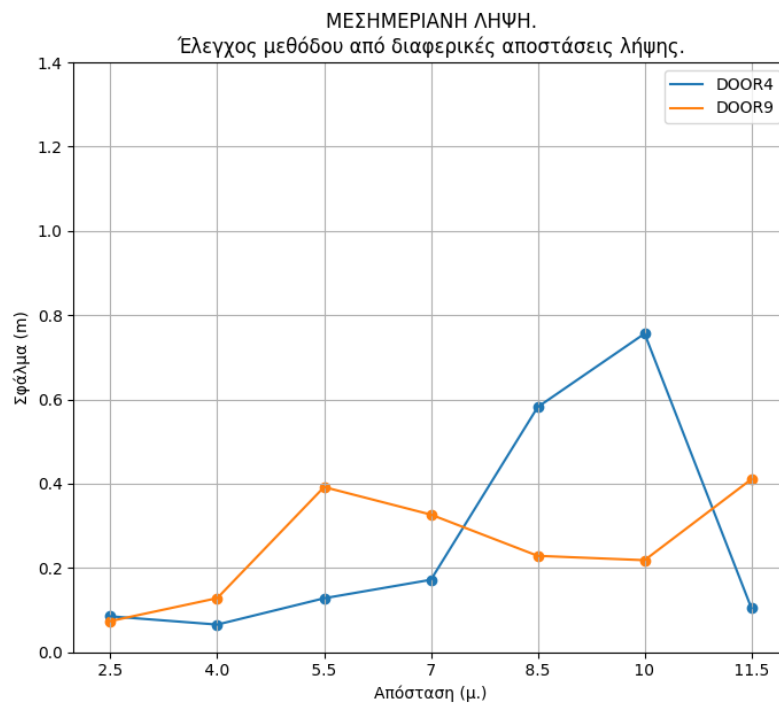


Σχήμα 5.12: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 13 λήψεις στις οποίες εναλλάσσεται η γωνία θέασης Θ . Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη.

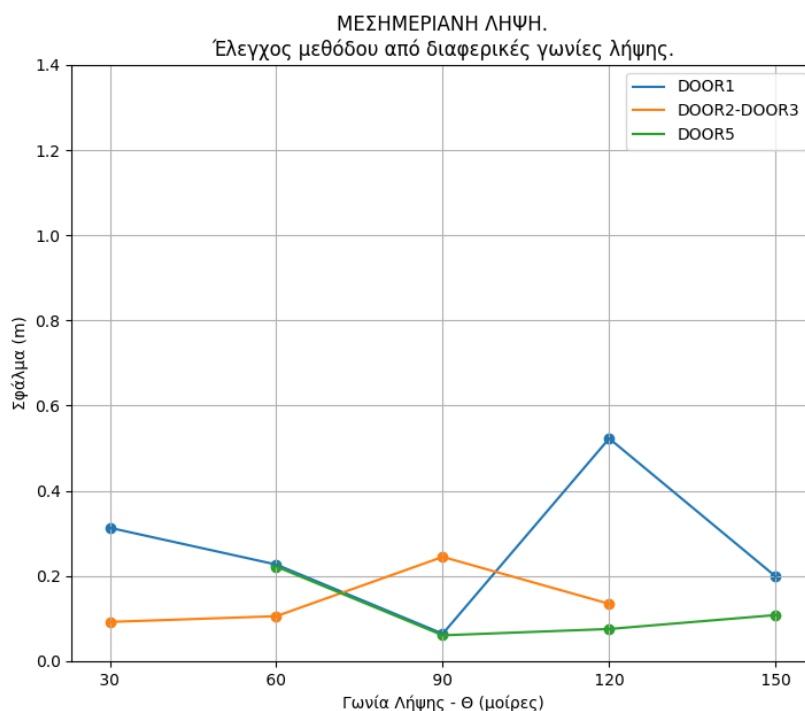
Μεσημεριανές Λήψεις.



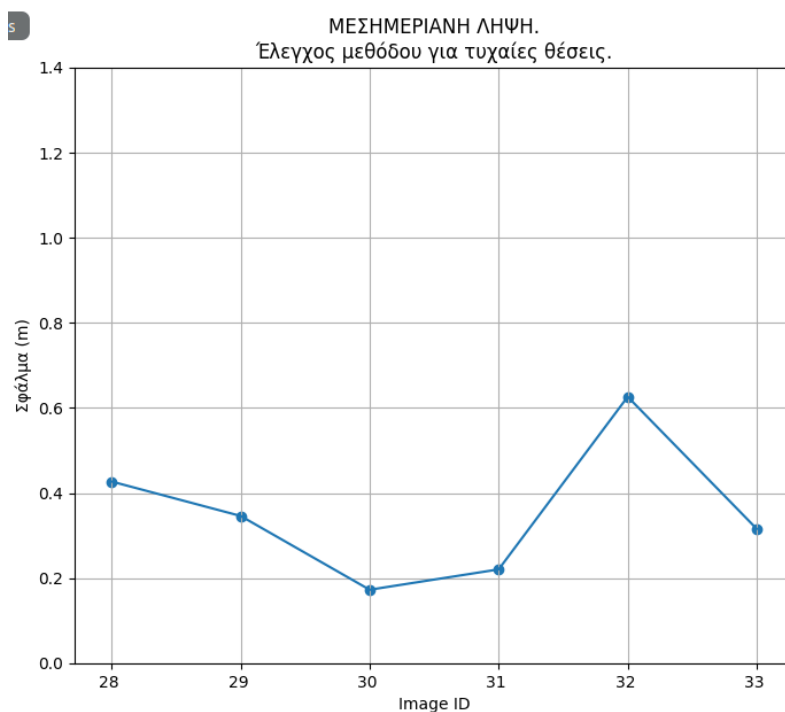
Σχήμα 5.14: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 33 μεσημεριανές λήψεις.



Σχήμα 5.15: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 14 λήψεις στις οποίες εναλλάσσεται η απόσταση. Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη.

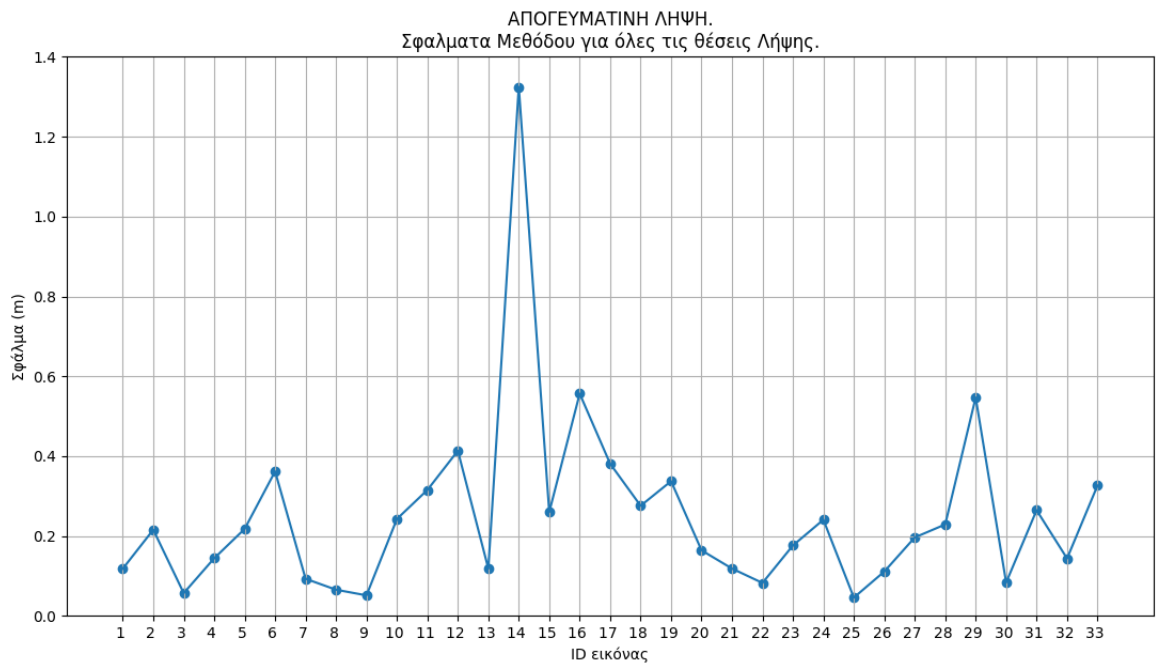


Σχήμα 5.16: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 13 λήψεις στις οποίες εναλλάσσεται η γωνία θέασης θ . Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη.

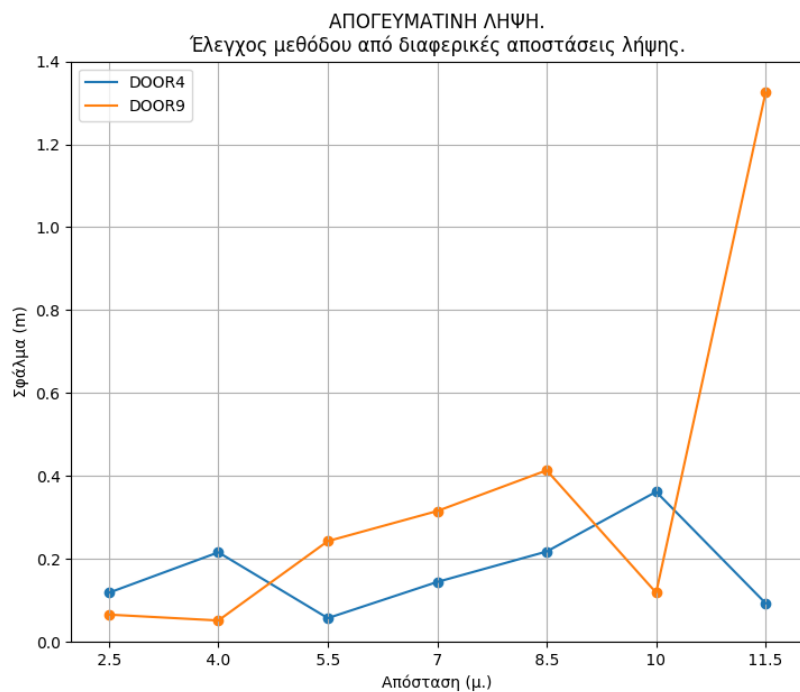


Σχήμα 5.17: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 6 λήψεις που πραγματοποιήθηκαν από τυχαίες θέσεις. Στο οριζόντιο άξονα φαίνεται το ID της κάθε εικόνας.

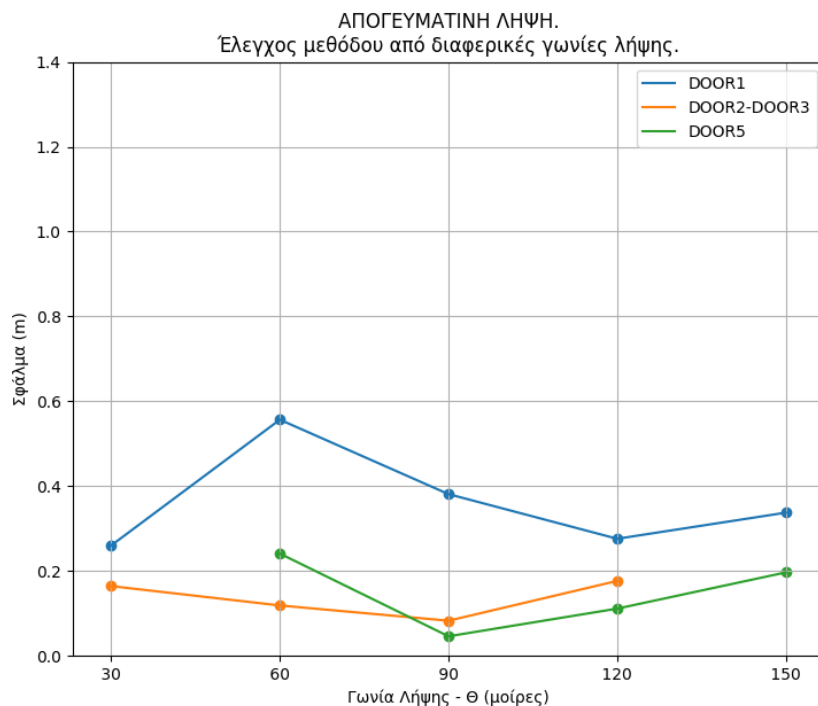
Απογευματινές Λήψεις.



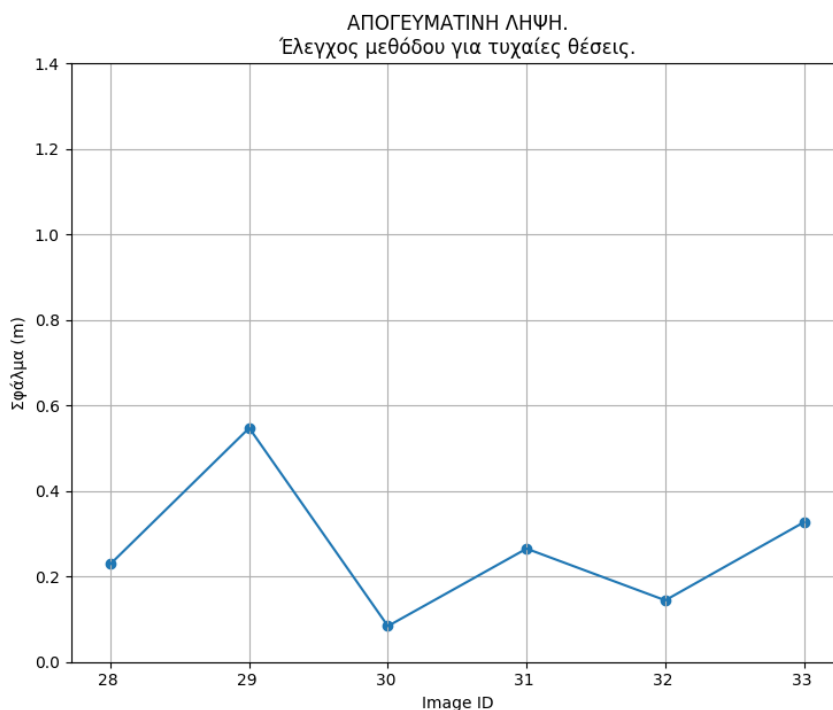
Σχήμα 5.18: Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 33 απογευματινές λήψεις.



Σχήμα 5.19: Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για τις 14 λήψεις στις οποίες εναλλάσσεται η απόσταση. Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη.

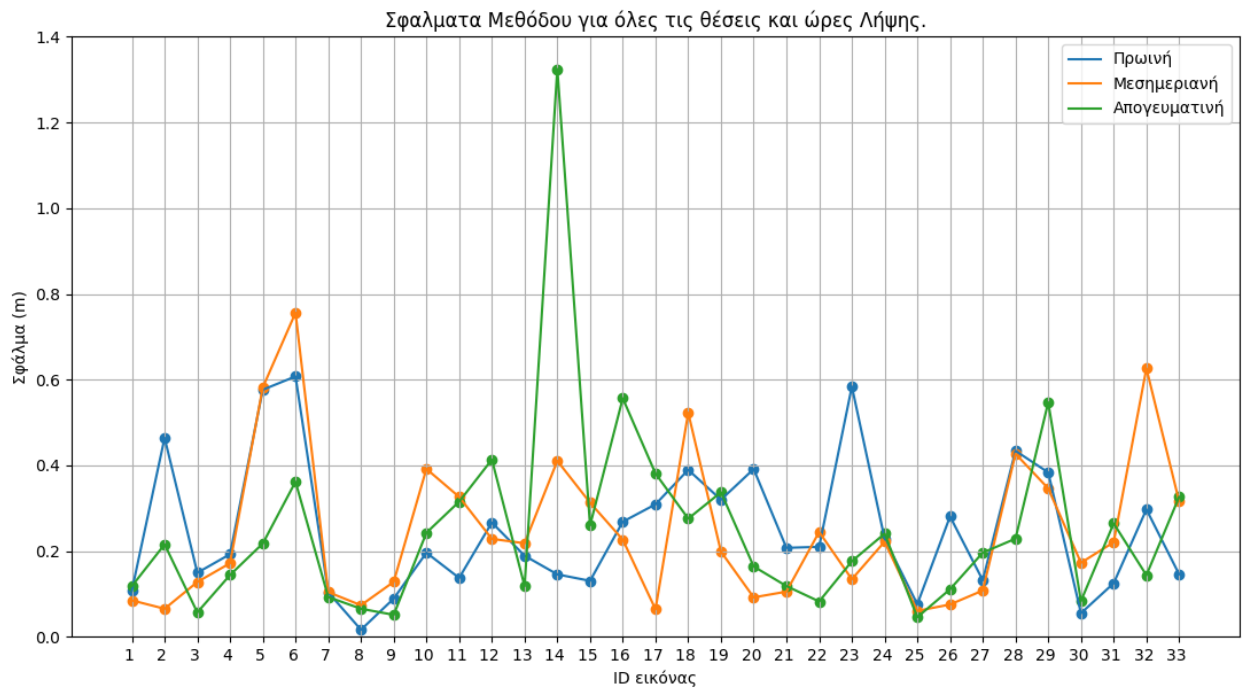


Σχήμα 5.20: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 13 λήψεις στις οποίες εναλλάσσεται η γωνία θέασης Θ . Εντός του υπομνήματος δηλώνεται το στατικό αντικείμενο στο οποίο εστιάζει η κάμερα σε κάθε λήψη.



Σχήμα 5.21: Οι αποκλίσεις την πρόβλεψης της μεθόδου από το ground truth για τις 6 λήψεις που πραγματοποιήθηκαν από τυχαίες θέσεις. Στο οριζόντιο άξονα φαίνεται το ID της κάθε εικόνας.

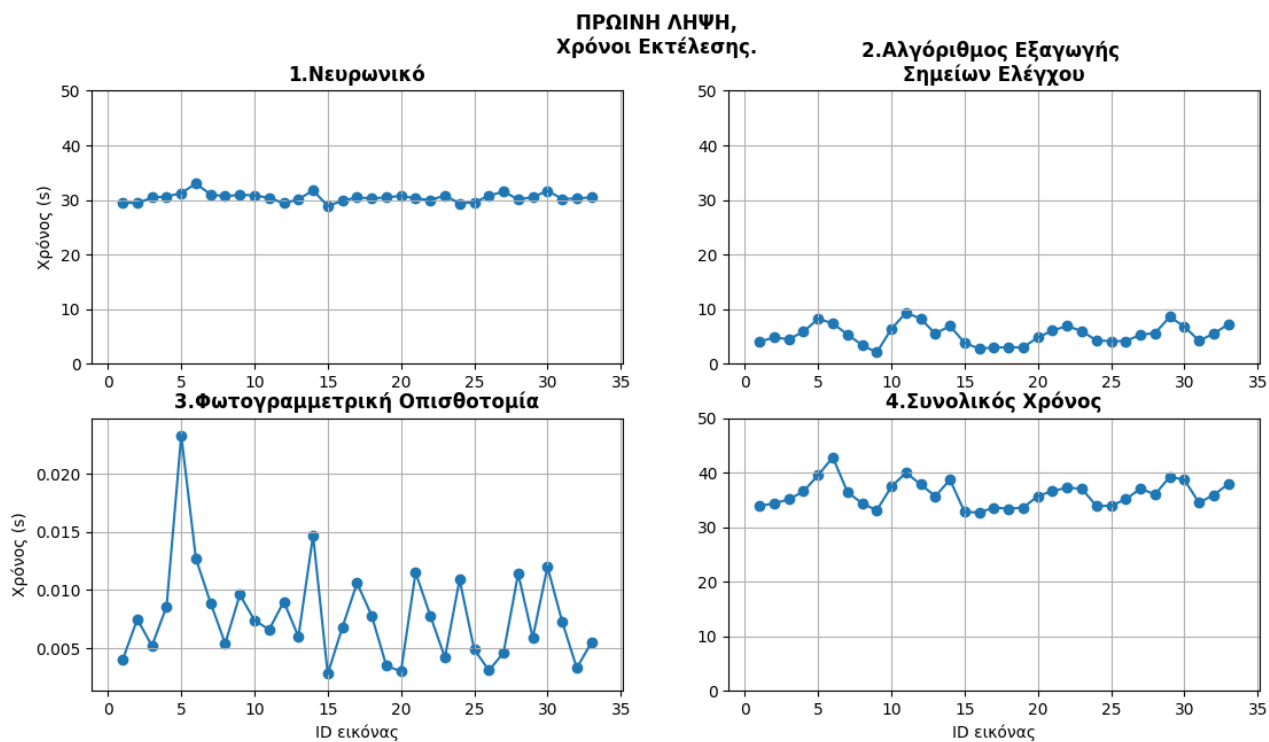
Τέλος προκειμένου να εξεταστεί η επίδραση της φωτεινότητας στην ακρίβεια της μεθόδου, παρουσιάζεται το σχήμα 5.22 που περιέχει όλες τις λήψεις κατηγοριοποιημένες ανά ώρα λήψης.



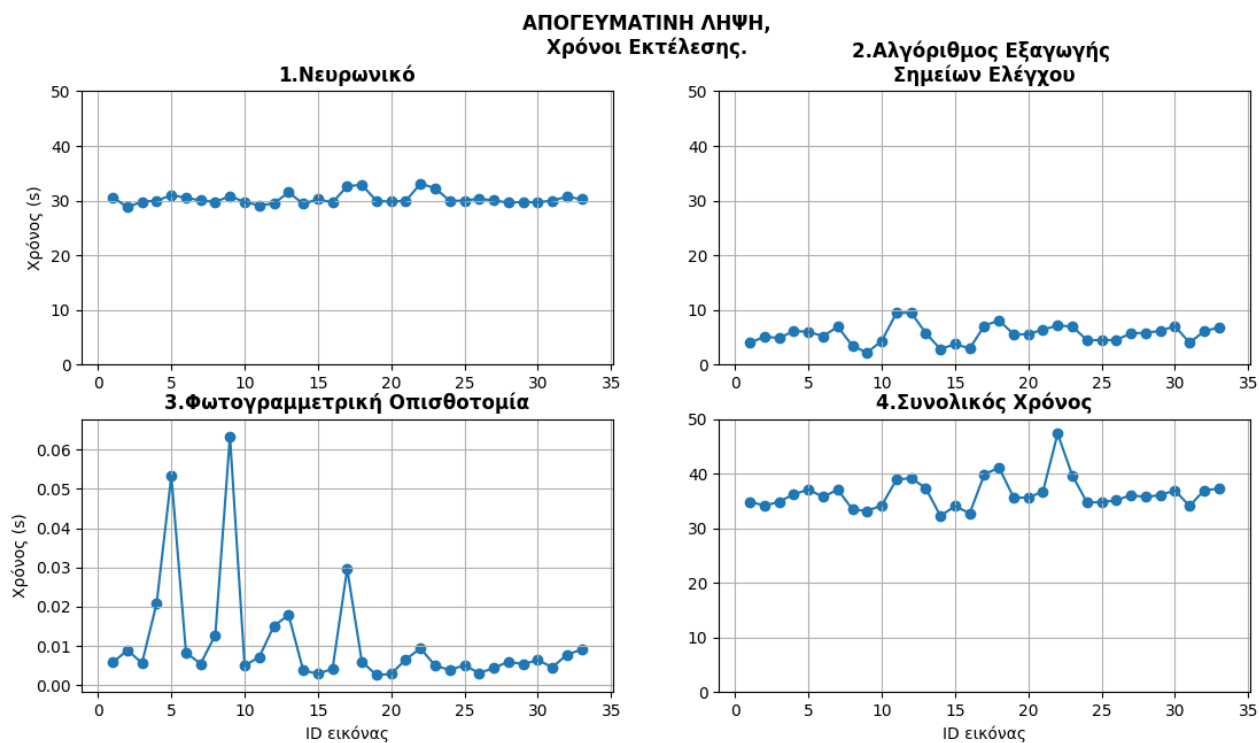
Σχήμα 5.22: Οι αποκλίσεις την πρόβλεψης της μεθόδου απο το ground truth για όλες τις λήψεις. Στο υπόμνημα φαίνεται η ώρα εκτέλεσης της κάθε λήψης.

3. Χρόνοι Εκτέλεσης

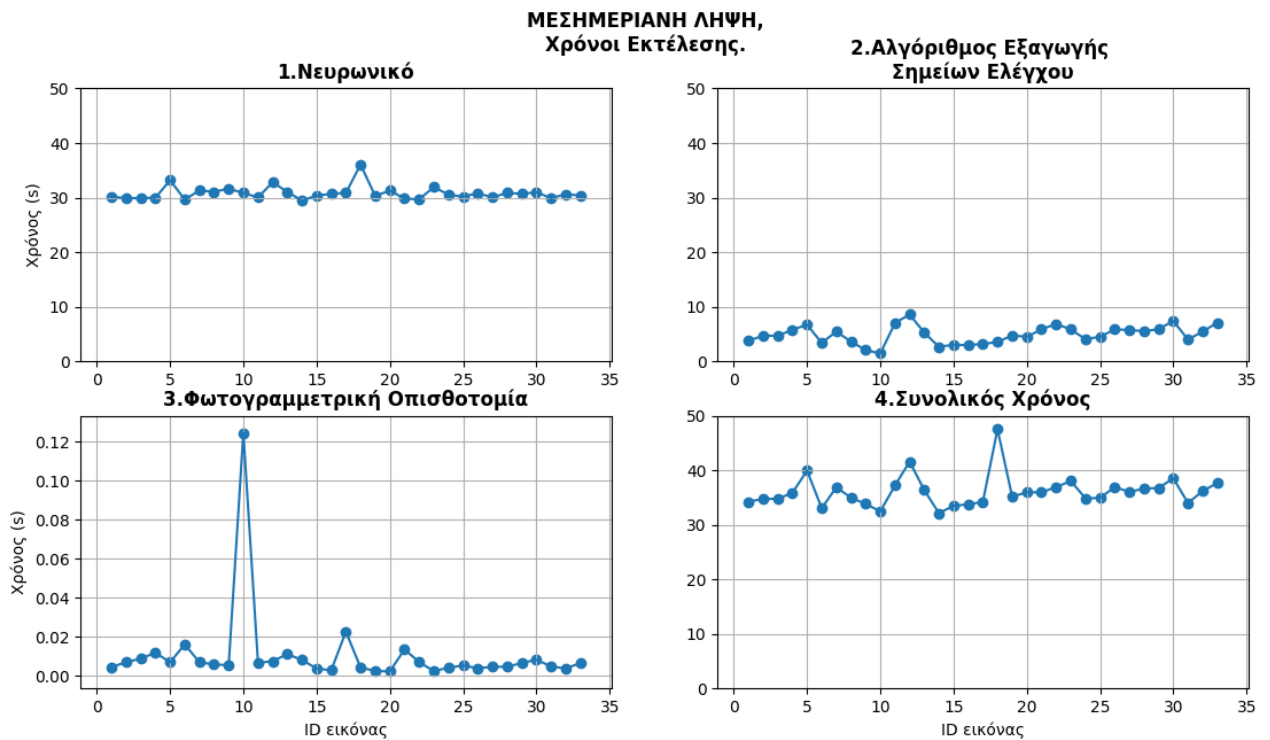
Στα παρακάτω διαγράμματα 5.23, 5.24, 5.25 παρουσιάζονται οι χρόνοι εκτέλεσης των τριών υποσταδίων της μεθόδου καθώς και ο συνολικός χρόνος εκτέλεσης της κάθε λήψης.



Σχήμα 5.23: Οι χρόνοι εκτέλεσης της μεθόδου για κάθε Πρωινή Λήψη. Στους οριζόντιους άξονες φαίνεται το ID κάθε εικόνας



Σχήμα 5.25: Οι χρόνοι εκτέλεσης της μεθόδου για κάθε Πρωινή Λήψη. Στους οριζόντιους άξονες φαίνεται το ID κάθε εικόνας



Σχήμα 5.24: Οι χρόνοι εκτέλεσης της μεθόδου για κάθε Πρωινή Λήψη. Στους οριζόντιους άξονες φαίνεται το ID κάθε εικόνας

5.2.3 Σχολιασμός Αποτελεσμάτων

Αφορμόμενοι από τα αποτελέσματα του πειράματος εφαρμογής της μεθόδου εντοπισμού θέσης μέσω οπτικής αναγνώρισης στατικών αντικειμένων είναι απαραίτητο να τονιστούν τα εξής:

1. **Σχετικά με την ικανότητα της μεθόδου να παραγάγει έγκυρα αποτελέσματα:**
Η μέθοδος δύναται να παράξει αξιόπιστα αποτελέσματα εντός του χώρου εργασίας του ρομπότ σε όλες τις περιπτώσεις.
2. **Σχετικά με την Ακρίβεια της μεθόδου:**
Η μέθοδος παρουσιάζει ακρίβεια καλύτερη του 0.6 μέτρου στο 96% των περιπτώσεων. Συγκεκριμένα οι μέσες τιμές των σφαλμάτων σε κάθε μία απο τις φάσεις λήψης ήταν:

Φάση Λήψης	Μέσο Σφάλμα (m)
Πρωινή	0.249
Μεσημεριανή	0.247
Απογευματινή	0.251

Πίνακας 5.4: Το μέσο σφάλμα της μεθόδου για κάθε φάση λήψης.

Δηλαδή η μέθοδος παρουσιάζει ακρίβεια $residual \simeq 0.25m$ χωρίς να επηρεάζεται από την φωτεινότητα του περιβάλλοντος χώρου. Ενδεικτικά, η ακρίβεια του πιο διάσημου μέσου απόλυτου εντοπισμού θέσης, του GPS κυμαίνεται από ένα έως τρία μέτρα, πράγμα που καθιστά την παρούσα μέθοδο μεγάλης ακριβείας. Επιπλέον, η ακρίβεια της πρόβλεψης φαίνεται να μην επηρεάζεται από την απόσταση του σταθμού λήψης από το στατικό αντικείμενο (σχήματα 5.11, 5.15, 5.19) ενώ φαίνεται να υπάρχει μία μικρή βελτίωση στο αποτέλεσμα όταν κάμερα σκεπεύει κάθετα ($\Theta = 90 \text{ deg}$ το στατικό αντικείμενο (σχήματα 5.12, 5.16, 5.20)). Τέλος αξίζει να σχολιαστεί το εξής, στις περιπτώσεις στις οποίες παρατηρείται πολύ υψηλότερο σφάλμα σε σχέση με τους μέσους όρους, συνήθως η κάμερα ελάμβανε μόλις δύο ή ένα στατικά αντικείμενα στο οπτικό της πεδίο ή ο αλγόριθμος εξαγωγής σημείων ελέγχου δεν κατάφερε να εντοπίσει αρκετά έγκυρα ζεύγη χαρακτηριστικών μεταξύ block εικόνας δοκιμής και των εικόνων αναφοράς με αποτέλεσμα το εν λόγω block να αγνοείται, και άρα το στάδιο φωτογραμμετρικής οπισθοτομίας να λαμβάνει ανεπαρκή πληροφορία ώστε να εξάγει την τελική θέση σωστά.

3. Σχετικά με τους χρόνους εκτέλεσης:

Ο μέσος χρόνος εκτέλεσης της μεθόδου είναι

$$t_{all} = 36,1 \text{ sec}$$

Σε αυτό το χρόνο εκτέλεσης, καθένα από τα επιμέρους τμήματα της μεθόδου συνεισφέρει ως εξής:

- Το κομμάτι που συνεισφέρει περισσότερο είναι αυτό της αναγνώρισης των στατικών αντικειμένων μέσω του αλγορίθμου faster R-CNN που κατα μέσο όρο διαρκεί περίπου τριάντα δευτερόλεπτα. Αυτή η καθυστέρηση είναι αναμενόμενη καθώς επιλέχθηκε ο αλγόριθμος Faster R-CNN με συνελικτικό δίκτυο το ResNet101 το οποίο στον συμβιβασμό ταχύτητας-ακρίβειας εστιάζει στην ακρίβεια με αποτέλεσμα να έχει υψηλότερους χρόνους εκτέλεσης σε σχέση με άλλες αρχιτεκτονικές που θα μπορούσαν να επιλεγθούν για τον πυρήνα του Faster R-CNN.
- Η επόμενη μεγαλύτερη συνεισφορά στον χρόνο εκτέλεσης είναι ο αλγόριθμος εξαγωγής σημείων ελέγχου ο οποίος κατα μέσο όρο διαρκεί πέντε δευτερόλεπτα.
- Τέλος, ο αλγόριθμος φωτογραμμετρική Οπισθοτομίας, διαρκεί μόλις κλάσματα του δευτερολέπτου καθώς κατα μέσο όρο συγκλίνει στις έξι επαναλήψεις.

Κεφάλαιο 6

Συμπεράσματα και Μελλοντική Εργασία.

6.1 Σύνοψη Συμπερασμάτων Αξιολόγησης.

Στα πλαίσια αυτής της διπλωματικής εργασίας, υλοποιήθηκε και έγινε αξιολόγηση στον αλγόριθμο εντοπισμού θέσης ρομπότ σε εσωτερικούς χώρους μέσω οπτικής αναγνώρισης στατικών αντικειμένων. Ο αλγόριθμος χωρίζεται σε τρία στάδια, τον εντοπισμό και αναγνώριση στατικών αντικειμένων (πόρτες) σε μία λήψη του χώρου εργασίας, την εξαγωγή σημείων ελέγχου εντός της εικόνας δοκιμής και τέλος τον μαθηματικό υπολογισμό της θέσης του σταθμού λήψης και κατ' επέκταση του ρομπότ μέσω της μεθόδου φωτογραμμετρικής οπισθοτομίας. Συνοψίζοντας τα πειραματικά αποτελέσματα, προκύπτει το συμπέρασμα ότι αυτός ο αλγόριθμος εντοπισμού θέσης για εσωτερικούς χώρους είναι αρκετά αποδοτικός αφού δύναται να παράξει αξιόπιστες προβλέψεις για την απόλυτη θέση του ρομπότ εντός του χώρου εργασίας του και μάλιστα με ακρίβεια μόλις μερικών εκατοστών, πράγμα που τον καθιστά ικανό να συναγωνιστεί τις συμβατικές μεθόδους εντοπισμού θέσης.

Βασικά πλεονεκτήματα του αλγορίθμου είναι η ακρίβεια που προσφέρει σε συνδυασμό με την αποδέσμευση που προσφέρει όσον αφορά την χρήση πολύπλοκων και ακριβών υποδομών προκειμένου να υλοποιηθεί.

Κύριο μειονέκτημα είναι ο χρόνος καθυστέρησης που υπάρχει μεταξύ της κάθε πρόβλεψης θέσης πράγμα που τον καθιστά μη συμβατό με εφαρμογές που απαιτούνται πολύ χαμηλοί χρόνοι απόκρισης για την εκτέλεση των εργασιών του ρομπότ.

6.2 Μελλοντικές Προεκτάσεις

Το έργο της παρούσας διπλωματικής μπορεί να επεκταθεί και στις εξής ακόλουθες ενδεικτικές, αλλά συγχεκριμένες κατευθύνσεις:

- Προτείνεται η μελέτη εναλλακτικών αλγορίθμων εξαγωγής χαρακτηριστικών ως μέρος του Αλγορίθμου Εξαγωγής Σημείων Ελέγχου προκειμένου να καταστεί πιο επιτυχημένη

η διαδικασία εξαγωγής έγκυρων ζευγών χαρακτηριστικών και άρα να δημιουργούνται καλύτεροι ομογραφικοί πίνακες. Συγκεκριμένα, προτείνεται ο αλγόριθμος BOLD [37] ο οποίος έχει τις ίδιες αρχές λειτουργίας με τον αλγόριθμο SIFT αλλά εξειδικεύεται στην εξαγωγή χαρακτηριστικών σε αντικείμενα με πολύ ομαλό ανάγλυφο - Textureless Objects όπως είναι οι πόρτες που επιλέχθηκαν στην παρούσα εργασία ως στατικά αντικείμενα.

- Προτείνεται η μελέτη διαφορετικών αρχιτεκτονικών συνελικτικού δικτύου στον πυρήνα του Faster R-CNN όπως παραδείγματος χάριν το ResNet50 προκειμένου να ελεγχθεί με ποια επιτυγχάνεται ο βέλτιστος συμβιβασμός μεταξύ των μετρικών ακρίβεια-χρόνος απόκρισης.

Βιβλιογραφία

- [1] Fyhn, Marianne, Sturla Molden, Menno P. Witter, Edvard I. Moser, and May-Britt Moser. “Spatial Representation in the Entorhinal Cortex.” *Science* (New York, N.Y.) 305, no. 5688 (August 27, 2004): 1258–64. <https://doi.org/10.1126/science.1099901>.
- [2] Sargolini, Francesca, Marianne Fyhn, Torkel Hafting, Bruce L. McNaughton, Menno P. Witter, May-Britt Moser, and Edvard I. Moser. “Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex.” *Science* (New York, N.Y.) 312, no. 5774 (May 5, 2006): 758–62. <https://doi.org/10.1126/science.1125572>.
- [3] O’Keefe, J., and J. Dostrovsky. “The Hippocampus as a Spatial Map. Preliminary Evidence from Unit Activity in the Freely-Moving Rat.” *Brain Research* 34, no. 1 (November 12, 1971): 171–75. [https://doi.org/10.1016/0006-8993\(71\)90358-1](https://doi.org/10.1016/0006-8993(71)90358-1).
- [4] “Robot Localization: An Introduction - Huang - - Major Reference Works - Wiley Online Library.” Accessed July 9, 2020. <https://onlinelibrary.wiley.com/doi/full/10.1002/047134608X.W8318>.
- [5] Press, The MIT. “Probabilistic Robotics — The MIT Press.” The MIT Press. Accessed July 9, 2020. <https://mitpress.mit.edu/books/probabilistic-robotics>.
- [6] Rigelsford, Jon. “Introduction to Autonomous Mobile Robots.” *Industrial Robot: An International Journal* 31, no. 6 (January 1, 2004): 534–35. <https://doi.org/10.1108/ir.2004.31.6.534.3>.
- [7] Xiao, Aoran, Ruizhi Chen, Deren Li, Yujin Chen, and Dewen Wu. “An Indoor Positioning System Based on Static Objects in Large Indoor Scenes by Using Smartphone Cameras.” *Sensors* (Basel, Switzerland) 18, no. 7 (July 11, 2018). <https://doi.org/10.3390/s18072229>.
- [8] Nīkitenko, Agris, Aleksis Liekna, Mārtiņš Ekmanis, Guntis Kuļikovskis, and Ilze Andersone. “Single Robot Localisation Approach for Indoor Robotic Systems through Integration of Odometry and Artificial Landmarks” 14 (2013): 50–58.
- [9] Youssef, Moustafa, and Ashok Agrawala. “The Horus WLAN Location Determination System.” In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services - MobiSys '05*, 205. Seattle, Washington: ACM Press, 2005. <https://doi.org/10.1145/1067170.1067193>.

- [10] Haverinen, Janne, and Anssi Kemppainen. “Global Indoor Self-Localization Based on the Ambient Magnetic Field.” *Robotics Auton. Syst.*, 2009. <https://doi.org/10.1016/j.robot.2009.07.018>.
- [11] Chen, Liang, Heidi Kuusniemi, Yuwei Chen, Jingbin Liu, Ling Pei, Laura Ruotsalainen, and Ruizhi Chen. “CONSTRAINT KALMAN FILTER FOR INDOOR BLUETOOTH LOCALIZATION.” In *2015 23RD EUROPEAN SIGNAL PROCESSING CONFERENCE (EUSIPCO)*, 1915–19. IEEE, 2015. <https://researchportal.helsinki.fi/en/publications/constraint-kalman-filter-for-indoor-bluetooth-localization>.
- [12] Kohoutek, Tobias K., Rainer Mautz, and Andreas Donaubaauer. “Real-Time Indoor Positioning Using Range Imaging Sensors.” edited by Nasser Kehtarnavaz and Matthias F. Carlsohn, 77240K. Brussels, Belgium, 2010. <https://doi.org/10.1117/12.853688>.
- [13] Kim, Jongbae, and Heesung Jun. “Vision-Based Location Positioning Using Augmented Reality for Indoor Navigation.” *IEEE Transactions on Consumer Electronics* 54, no. 3 (August 2008): 954–62. <https://doi.org/10.1109/TCE.2008.4637573>.
- [14] Möller, Andreas, Matthias Kranz, Robert Huitl, Stefan Diewald, and Luis Roalter. “A Mobile Indoor Navigation System Interface Adapted to Vision-Based Localization,” 4:1-4:10, 2012. <https://doi.org/10.1145/2406367.2406372>.
- [15] Mulloni, Alessandro, Daniel Wagner, István Barakonyi, and Dieter Schmalstieg. “Indoor Positioning and Navigation with Camera Phones.” *IEEE Pervasive Computing* 8 (April 1, 2009): 22–31. <https://doi.org/10.1109/MPRV.2009.30>.
- [16] M. Werner, M. Kessel and C. Marouane, ”Indoor positioning using smartphone camera,” 2011 International Conference on Indoor Positioning and Indoor Navigation, Guimaraes, 2011, pp. 1-6. doi: 10.1109/IPIN.2011.6071954
- [17] Zhong, Xu, Yu Zhou, and Hanyu Liu. “Design and Recognition of Artificial Landmarks for Reliable Indoor Self-Localization of Mobile Robots.” *International Journal of Advanced Robotic Systems* 14, no. 1 (January 1, 2017): 1729881417693489. <https://doi.org/10.1177/1729881417693489>.
- [18] Zhao, Zhong-Qiu, Peng Zheng, Shou-tao Xu, and Xindong Wu. “Object Detection with Deep Learning: A Review.” *ArXiv:1807.05511 [Cs]*, April 16, 2019. <http://arxiv.org/abs/1807.05511>.
- [19] “Elements of Photogrammetry with Application in GIS, Fourth Edition: Wolf, Paul, DeWitt, Bon, Wilkinson, Benjamin: 9780071761123: Amazon.Com: Books.” Accessed July 9, 2020. <https://www.amazon.com/Elements-Photogrammetry-Application-GIS-Fourth/dp/0071761128>.

- [20] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” In *Advances in Neural Information Processing Systems 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, 91–99. Curran Associates, Inc., 2015. <http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf>.
- [21] Uijlings, J. R. R., K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. “Selective Search for Object Recognition.” *International Journal of Computer Vision* 104, no. 2 (September 2013): 154–71. <https://doi.org/10.1007/s11263-013-0620-5>.
- [22] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition.” In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–78, 2016. <https://doi.org/10.1109/CVPR.2016.90>.
- [23] “TensorFlow.” Accessed July 12, 2020. <https://www.tensorflow.org/>.
- [24] Simonyan, Karen, and Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” *ArXiv:1409.1556 [Cs]*, April 10, 2015. <http://arxiv.org/abs/1409.1556>.
- [25] Nair, Vinod, and Geoffrey E Hinton. “Rectified Linear Units Improve Restricted Boltzmann Machines,” n.d., 8.
- [26] GitHub. “Tensorflow/Docs.” Accessed July 12, 2020. <https://github.com/tensorflow/docs>.
- [27] Girshick, Ross. “Fast R-CNN.” In *2015 IEEE International Conference on Computer Vision (ICCV)*, 1440–48, 2015. <https://doi.org/10.1109/ICCV.2015.169>.
- [28] Lowe, David G. “Distinctive Image Features from Scale-Invariant Keypoints.” *International Journal of Computer Vision* 60, no. 2 (November 2004): 91–110. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>.
- [29] “Space Resection in Photogrammetry Using Collinearity Condition without Linearisation: Survey Review: Vol 42, No 315.” Accessed July 9, 2020. <https://www.tandfonline.com/doi/abs/10.1179/003962609X451681>.
- [30] Zhang, Z. “A Flexible New Technique for Camera Calibration.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, no. 11 (November 2000): 1330–34. <https://doi.org/10.1109/34.888718>.
- [31] Smarnakis, Aris. *Smarnakis/Robot_localisation*. Python, 2020. https://github.com/smarnakis/robot_localisation.
- [32] <https://github.com/tzutalin/labelImg>

- [33] <https://cocodataset.org>
- [34] https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tfl_detection_zoo.m
- [35] Fischler, Martin A., and Robert C. Bolles. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." *Communications of the ACM* 24, no. 6 (June 1, 1981): 381–395. <https://doi.org/10.1145/358669.358692>.
- [36] <https://opencv.org/>
- [37] Tombari, Federico, Alessandro Franchi, and Luigi Di. "BOLD Features to Detect Texture-Less Objects." In *2013 IEEE International Conference on Computer Vision*, 1265–72, 2013. <https://doi.org/10.1109/ICCV.2013.160>.
- [38] M. Avgeris, D. Spatharakis, N. Athanasopoulos, D. Dechouniotis and S. Papavassiliou, "Single Vision-Based Self-Localization for Autonomous Robotic Agents," 2019 7th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Istanbul, Turkey, 2019, pp. 123-129, doi: 10.1109/FiCloudW.2019.00035.
- [39] D. C. K. Yuen and B. A. MacDonald, "Vision-based localization algorithm based on landmark matching, triangulation, reconstruction, and comparison," *IEEE Trans. Robot.*, vol. 21, no. 2, pp. 217–226, Apr. 2005.
- [40] A. Bais and R. Sablatnig, "Landmark Based Global Self-localization of Mobile Soccer Robots," in *Computer Vision – ACCV 2006*, vol. 3852, P. J. Narayanan, S. K. Nayar, and H.-Y. Shum, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 842–851.
- [41] G. N. Desouza and A. C. Kak, "Vision for mobile robot navigation: a survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237-267, Feb. 2002, doi: 10.1109/34.982903.
- [42] <https://numpy.org/>

