



Πανεπιστήμιο Θεσσαλίας

Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Η/Υ

Διπλωματική Εργασία

**Χρήση του Αισθητήρα Kinect για
Αλληλεπίδραση Βασισμένη σε Χειρονομίες
Gesture Based Interaction Using the Kinect
Sensor**

Συγγραφέας:

Παναγιώτου Δημήτριος

Επιβλέποντες:

Ποταμιάνος Γεράσιμος

Μπέλλας Νικόλαος

Περίληψη

Η χρήση πολυτροπικών συστημάτων ολοένα αυξάνει, καθώς η πρόσβαση σε συστήματα που υποστηρίζουν φωνητική αναγνώριση αλλά και αναγνώριση χειρονομιών έχει γίνει ελεύθερη για την πλειοψηφία του κοινού. Η παρούσα διπλωματική αποτελεί την υλοποίηση ενός συστήματος απομακρυσμένου χειρισμού οχήματος με φωνητικές εντολές, με χειρονομίες, ή την ταυτόχρονη χρήση και των δύο.

Χρησιμοποιήθηκε η πλατφόρμα IrisTK για την ανάπτυξη της λειτουργίας με είσοδο φωνητικών εντολών, ενώ ο αισθητήρας Kinect v2 sensor για τη λειτουργία με είσοδο χειρονομιών. Το όχημα ελέγχεται από Raspberry Pi 2 model B και η επικοινωνία μεταξύ του χειριστή και του ρομπότ γίνεται με δύο network sockets.

Αρχικά συνδέσαμε τα μηχανικά μέρη του ρομπότ και το ρομπότ με τον μικροελεγκτή. Έπειτα δοκιμάσαμε την επικοινωνία μεταξύ front end και back end πάνω από το τοπικό δίκτυο. Αργότερα υλοποιήσαμε τη λειτουργία front end φωνητικών εντολών με το εργαλείο IrisTK που απλοποιεί τη δημιουργία πολυτροπικών συστημάτων διαλόγου. Ο προγραμματισμός έγινε σε γλώσσα Java. Ταυτόχρονα έγινε και η λειτουργία front end χειρονομιών σε γλώσσα C++, χρησιμοποιώντας μόνο τις προκαθορισμένες χειρονομίες. Έπειτα προγραμματίσαμε το back end της εφαρμογής που ήταν το server side τμήμα που τρέχει στο ρομπότ. Ο προγραμματισμός έγινε σε γλώσσα Python με τη βοήθεια της βιβλιοθήκης OpenCV για την επεξεργασία εικόνων. Επίσης δοκιμάσαμε τους χρονισμούς και των δύο λειτουργιών. Τέλος, προσθέσαμε τις νέες χειρονομίες στο ήδη υπάρχων σύνολο του Kinect και δοκιμάσαμε τη λειτουργία συνδυαστικής εισόδου.

Abstract

The use of multimodal systems is increasing, as more people gain access to systems which support speech and gesture recognition. The present thesis is an implementation of a system which controls a remote robot with voice input, gesture input, or both.

The IrisTK toolkit was used for the former mode and the Kinect v2 sensor for the latter. The robot is controlled by a Raspberry Pi 2 model B, and the connection between the user and the robot consists of two network sockets.

Initially, we built the robot and connected it to the Raspberry Pi. Then, we tested the connection between user and robot over the local network. Afterwards, we programmed the front end of voice input using IrisTK, which helps building multimodal dialog systems and is implemented in Java, as well as the front end of gesture input in C++, using only the default gestures. Next, we programmed the back end of the application, namely the server side code that runs on the Pi. We used Python for the main programming and the OpenCV library for image processing. We also tested the timings of the two modes. Finally, we added new gestures to the pre-existed Kinect sensor system and attempted to give both type of inputs simultaneously.

Ευχαριστίες

Καταρχήν θα ήθελα να ευχαριστήσω τον επικεφαλής επόπτη της διπλωματικής εργασίας κ.Γεράσιμο Ποταμιάνο, αναπληρωτή καθηγητή του τμήματος Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών, για την πολύτιμη βοήθειά και υποστήριξή του σε όλη τη διάρκεια της εκπόνησης της εργασίας. Επιπρόσθετα, θα ήθελα να ευχαριστήσω τον συνάδελφο και φίλο Δημήτρη Νικολάου, που ήταν ο άμεσος συνεργάτης μου και συντέλεσε εξίσου στη δημιουργία της υλοποίησης. Χωρίς την εργασία του δε θα ήταν δυνατόν να είχε αποτέλεσμα αυτή μας η προσπάθεια. Επίσης, είμαι ευγνώμων και στο δεύτερο μέλος της επιτροπής κ.Νικόλαο Μπέλλα που βοήθησε στο να ξεκινήσει η προσπάθειά μας αυτή με τις χρήσιμες συμβουλές του. Τέλος, για την πνευματική τους στήριξη και την αγάπη τους θα ήθελα να ευχαριστήσω την οικογένειά μου, τους φίλους μου, την κοπέλα μου και το σκυλάκι μου, που μου στάθηκαν σε κάθε δύσκολη στιγμή. Χωρίς όλους τους παραπάνω, τίποτα δε θα ήταν εφικτό.

ΠΕΡΙΕΧΟΜΕΝΑ

Περίληψη	I
Abstract	II
Ευχαριστίες	III
Περιεχόμενα	IV
Κατάλογος Σχημάτων	V
Κατάλογος Πινάκων	VI
Βιβλιογραφία	VII
1 Εισαγωγή.....	9
1.1 Αναγνώριση φωνητικών εντολών από υπολογιστές.....	9
1.1.1 Σύντομη ιστορική αναδρομή.....	9
1.1.2 Τομείς χρήσης σήμερα.....	11
1.2 Παρούσα υλοποίηση.....	12
1.3 Συνεισφορά της διπλωματικής.....	13
1.4 Περιεχόμενα κεφαλαίων.....	14
2 Αρχιτεκτονική του συστήματος.....	15
2.1 Αυτοκίνητο ελεγχόμενο από ενσωματωμένο Raspberry Pi.....	15
2.1.1 Μηχανικά μέρη του οχήματος.....	16
2.1.2 Raspberry Pi 2 Model B.....	17
2.2 Kinect v2 sensor για ανίχνευση κίνησης.....	20
2.3 Network Socket.....	23
3 Λειτουργία με είσοδο φωνητικές εντολές.....	24
3.1 Η πλατφόρμα IrisTK.....	24
3.1.1 Γραμματική SRGS.....	24
3.1.2 Εισαγωγή streaming στην πλατφόρμα.....	25
3.2 Περιγραφή λειτουργίας του συστήματος.....	26
3.3 Χαρτογράφηση εμποδίων στο δισδιάστατο επίπεδο.....	30
3.3.1 Συνθήκες λειτουργίας και δεδομένα.....	30
3.3.2 Επεξεργασία εικόνας – Εφαρμογή φίλτρων.....	31
3.3.3 2D απεικόνιση των εμποδίων.....	34
3.4 Πειράματα.....	35
3.4.1 Πειράματα κίνησης.....	35
3.4.2 Πειράματα mapping εμποδίων.....	37
4 Λειτουργία με είσοδο χειρονομίες.....	41
4.1 Σύνολο αποδεκτών χειρονομιών.....	41
4.2 Λειτουργία με συνδυαστική είσοδο φωνής και χειρονομιών.....	42
4.3 Πειράματα.....	43
5 Συμπεράσματα.....	46
5.1 Συνεισφορά της διπλωματικής εργασίας.....	46
5.2 Μελλοντική έρευνα.....	46

Κατάλογος Σχημάτων

Σχήμα 1.1: Πάνω: το 3d μοντέλο του προσώπου, η μάσκα που αυτό προβάλλεται, και ένα παράδειγμα προβολής. Κάτω: Το ρομπότ furhat στην RobotVille το Δεκέμβριο του 2011 (από)

Σχήμα 1.2: Block Diagram της αρχιτεκτονικής της υλοποίησης

Σχήμα 2.1: Αρχιτεκτονική του συστήματος

Σχήμα 2.2: Χαρακτηριστικά του Raspberry Pi 2 model B (απο).

Σχήμα 2.3: Συνδεσμολογία που χρησιμοποιήθηκε.

Σχήμα 2.4: Πάνω: Εξωτερική όψη του Kinect v2 sensor. Κάτω: Μετά την αφαίρεση καλύματος, φαίνονται οι αισθητήρες και οι εκπομποί των lasers (από).

Σχήμα 2.5: Διαδικασία εξαγωγής 3D point cloud μοντέλου. (α) Αρχική φωτογραφία (β) Depth map σε infrared (γ) 3d point cloud σε infrared, χρησιμοποιώντας το β (δ) Depthmap σε grayscale (ε) 3D point cloud σε grayscale

Σχήμα 2.6: Όλα τα σημεία – αρθρώσεις του ανθρώπινου σώματος που μπορεί να αναγνωρίσει με ακρίβεια το Kinect v2 sensor (από).

Σχήμα 2.7 Τα δύο network sockets που χρησιμοποιήθηκαν, ένα για το streaming της εικόνας πίσω στο IrisTK και ένα για τις εντολές προς το ρομπότ.

Σχήμα 3.1: Στιγμιότυπο τρεξίματος.

Σχήμα 3.2: Γλώσσα της υλοποίησης.

Σχήμα 3.3 FSM της λειτουργίας του συστήματος.

Σχήμα 3.4 : Θέση της κάμερας πάνω στο robot.

Σχήμα 3.5: (α) Αρχική εικόνα σε grayscale (β) Εφαρμογή threshold και contour finding/drawing για edge detection (γ) Αποκοπή της εικόνας για ελαχιστοποίηση error

Σχήμα 3.6: (α) Αρχική εικόνα εμποδίων με μπλε να είναι ένα εμπόδιο. (β) Η ζητούμενη απόσταση d είναι η πραγματική απόσταση του αντικειμένου από τη βάση της κάμερας. (γ) Σχηματισμός δύο τριγώνων. (δ) Υπολογισμός του c, h και θ. (ε) Υπολογισμός του d', που οδηγεί και στον υπολογισμό του ζητούμενου d.

Σχήμα 3.7: Αποτέλεσμα κάτοψης εμποδίων με εισοδο τη φωτογραφία του σχήματος 3.5.

Σχήμα 3.8: Πάνω: Φωτογραφία του τοπίου. Κάτω: Τελικό αποτέλεσμα του αλγορίθμου.

Σχήμα 3.9: Πάνω: Το εμπόδιο γύρω από το οποίο κινήθηκε το ρομπότ. Κάτω: Το αποτέλεσμα του αλγορίθμου.

Σχήμα 3.10: Πάνω: Η σκηνή του πειράματος. Κάτω: Η κάτοψη των εμποδίων σε δύο στάδια.

Σχήμα 4.1: Kinect v2 sensor default gestures.

Σχήμα 4.2: Πλήρες σύνολο αναγνωρίσιμων χειρονομιών.

Σχήμα 4.3: Πάνω: Αρχική εικόνα με δύο χειρονομίες ταυτόχρονα. Κάτω: Αναγνώριση χειρονομιών και εξαγωγή σκελετού.

Σχήμα 4.4: Confidence και μέσος όρος confidence πέντε διαδοχικών δοκιμών κάθε χειρονομίας.

Σχήμα 4.5: Real time μεταβολή του confidence ενώ δοκιμάζεται κάθε χειρονομία.

Κατάλογος Πινάκων

Πίνακας 2.1: Κυριότερα τεχνικά χαρακτηριστικά της webcam Microsoft lifecam VX-7000.

Πίνακας 2.2: Όλοι οι δυνατοί συνδυασμοί των pins του motor controller και πως μεταφράζονται για την κίνηση του ρομπότ. DF σημαίνει disfunction, αντίστοιχη κίνηση δεν υπάρχει και δεν προβλέπεται να έρθουν ποτέ σε αυτή την κατάσταση τα motors. Αν ωστόσο γίνει κάτι τέτοιο, η κίνηση θα είναι ελαττωματική.

Πίνακας 3.1: UPS phone set (από).

Πίνακας 3.2: Όλες οι αποδεκτές προτάσεις που αναγνωρίζει η γλώσσα της υλοποίησης.

Πίνακας 3.3: Ποσοστά αναγνώρισης για την εντολή Forward.

Πίνακας 3.4: Ποσοστά αναγνώρισης για την εντολή Backwards.

Πίνακας 3.5: Ποσοστά αναγνώρισης για την εντολή Left.

Πίνακας 3.6: Ποσοστά αναγνώρισης για την εντολή Right.

Πίνακας 4.1: Ποσοστό επι τοις εκατό χειρονομιών με confidence > 0.8.

ΚΕΦΑΛΑΙΟ 1^ο

Εισαγωγή

Η επικοινωνία ανθρώπου μηχανής αποτελεί καθημερινή πρακτική του σύγχρονου ανθρώπου, αφού κάθε τομέας της ζωής του περιλαμβάνει αλληλεπίδραση με μηχανές. Ακόμα και αν υποστηρίζουμε ότι πλέον η εξοικείωση με τέτοια συστήματα είναι μεγάλη, η αλήθεια είναι ότι ακόμα βρισκόμαστε πολύ μακριά από την αποτελεσματική επικοινωνία με τις μηχανές [6]. Η αποτελεσματικότητα αυτή έγκειται στην αλληλεπίδραση με τέτοια μηχανήματα, χωρίς να απομακρυνόμαστε από τις συνηθισμένες στον άνθρωπο τεχνικές επικοινωνίας, δηλαδή την ομιλία και την κίνηση.

1.1 Αναγνώριση φωνητικών εντολών από υπολογιστές

Η αναγνώριση φωνητικών εντολών από υπολογιστές αποτελεί ένα συνδυασμό διαφόρων προϋπάρχουσων τεχνολογιών όπως ψηφιακή επεξεργασία σημάτων, αναγνώριση προτύπων, φυσικής γλώσσας αλλά και της επιστήμης της γλωσσολογίας [7]. Βασικός της σκοπός της είναι η μετατροπή της φυσικής γλώσσας σε κείμενο (Speech To Text-STT).

1.1.1 Σύντομη ιστορική αναδρομή

Προσπάθειες για αλληλεπίδραση ανθρώπου μηχανής χρησιμοποιώντας αναγνώριση φωνής έχουν γίνει αναρίθμητες. Από τη δεκαετία του 50 κιόλας έχουμε ρομπότ που σκέφτονται μόνα τους και λαμβάνουν αποφάσεις. Η τεχνολογία αυτή αναπτύχθηκε ακόμα περισσότερο τις προηγούμενες δεκαετίες τόσο σε ρομπότ – οχήματα, σε ρομπότ –βραχίονες όσο και σε τεχνητής νοημοσύνης. Το ίδιο ισχύει και για επαφή με χειρονομίες, κάτι που τα τελευταία χρόνια έχει εξελιχθεί αρκετά λόγω της κυκλοφορίας ισχυρών depth cameras και sensors όπως είναι και το [Kinect v2](#) που χρησιμοποιείται και στην παρούσα εργασία [8]. Ο συνδυασμός τους όμως, δηλαδή συστήματα που δέχονται είσοδο ή παράγουν έξοδο με περισσότερους από έναν τρόπο (multimodal

systems) αποτέλεσαν αντικείμενο μελέτης τις τελευταίες δύο δεκαετίες. Συγκεκριμένα, το εργαλείο IrisTK το οποίο ανήκει στην κατηγορία των multimodal systems, μιας και επιτρέπει στο χρήστη τη δημιουργία multimodal συστημάτων με είσοδο/εξοδο φωνής και χειρονομιών και χρησιμοποιείται στην παρούσα υλοποίηση δημιουργήθηκε το 2011 από την [furhat robotics](http://www.furhat.com). Πρώτο τους σύστημα ήταν το ρομπότ furhat, φτιαγμένο για κοινωνική αλληλεπίδραση, προγραμματισμένο και αυτό με την πλατφόρμα IrisTK, μπορεί να συζητάει ταυτόχρονα με πολλαπλούς χρήστες (έχει δοκιμαστεί μέχρι πέντε), να απαντάει στις ερωτήσεις τους σύμφωνα με τη λογική που έχει προγραμματιστεί, αλλά και να τους κοιτάζει, αφού διαθέτει ανθρώπινα χαρακτηριστικά και έχει σχεδιαστεί ώστε να κοιτάζει τον συνομιλητή του. Στο σχήμα 1.1 φαίνεται ο τρόπος δημιουργίας του προσώπου του, καθώς και η συνομιλία με ακροατήριο στο κοινό του RobotVille, το Δεκέμβριο του 2011 [1].



Σχήμα 1.1: Πανω: το 3d μοντέλο του προσώπου, η μάσκα που αυτό προβάλλεται, και ένα παράδειγμα προβολής. Κάτω: Το ρομπότ furhat στην RobotVille το Δεκέμβριο του 2011 (από [1])

1.1.2 Τομείς χρήσης σήμερα

- **Ιατρική:** Χρησιμοποιείται ευρέως από γιατρούς σε νοσοκομεία για ιατρικές αναφορές ήδη από το 1994 [9]. Επίσης μεγάλη προσπάθεια έχει γίνει για να βοηθηθούν τα άτομα με ειδικές ικανότητες μέσω ASR λογισμικού (automatic speech recognition). Με τη χρήση ειδικών γραμματικών για κάθε περίπτωση ατόμου, έτσι ώστε να μειώνεται η πιθανότητα λάθους στις αναγνωρίσεις, είναι δυνατή η αλληλεπίδραση τους με το περιβάλλον τους, πολύ ευκολότερα απ'ότι νωρίτερα [10]. Τέλος, σε φάση ελέγχου βρίσκονται εφαρμογές που βοηθούν στην αποθεραπεία της εγκεφαλικής αρτηριοφλεβικής δυσπλασίας μετά την εγχείριση ασθενών, βοηθώντας τους να επανέλθει το λεκτικό και φρασσεολογικό επίπεδο των ασθενών εκεί που βρισκόταν.
- **Στρατιωτική:** Αναγνώριση φωνής χρησιμοποιείται ως κύριο μέσο για τον έλεγχο μη επανδρωμένων αεροσκαφών και ελικοπτέρων. Χρήση speech recognition γίνεται και σε πολιτικά αεροπλάνα όμως, για να αναγνωρίζουν μεταδόσεις που οι πιλότοι δεν είναι δυνατόν να ακούσουν λόγω θορύβου ή κάποιου άλλου συμβάντος, και να δράσουν αντίστοιχα [11].
- **Τηλεπικοινωνίες:** Ήδη εδώ και αρκετά χρόνια οι περισσότερες εταιρίες χρησιμοποιούν αυτόματους τηλεφωνητές για υπηρεσίες εξυπηρέτησης, τεχνικής υποστήριξης και επιστροφή τηλεφωνικών αριθμών σε ερώτημα των χρηστών [12].
- **Εκπαίδευση:** Όσον αφορά την εκπαίδευση, η αναγνώριση φωνής αποτελεί ένα ισχυρό εργαλείο για την εκμάθηση ξένης γλώσσας με χρήση αυτόματης μετάφρασης. Επίσης χρησιμοποιείται για την εκπαίδευση παιδιών με τύφλωση, χρησιμοποιώντας υπολογιστή ελεγχόμενο από φωνή. Επιπρόσθετα, πολλά παιδιά έχουν χρόνιες παθήσεις των μυών ή τενόντων, η κατάσταση των οποίων μπορεί να βελτιωθεί σε σημαντικό βαθμό από τη χρήση φωνητικού υποβοήθεια.
- **Καθημερινότητα:** Τα κινητά τηλέφωνα άρχισαν να υποστηρίζουν φωνητικές κλήσεις πριν από δεκαπέντε χρόνια, αλλά με τις σημερινές εφαρμογές για smartphones, φωνητικές εντολές χρησιμοποιούνται εκτός των κλήσεων και για υπηρεσίες αυτόματου σπιτιού, πληροφορίες καταλόγου, εύρεση τοποθεσίας GPS, και άλλες.

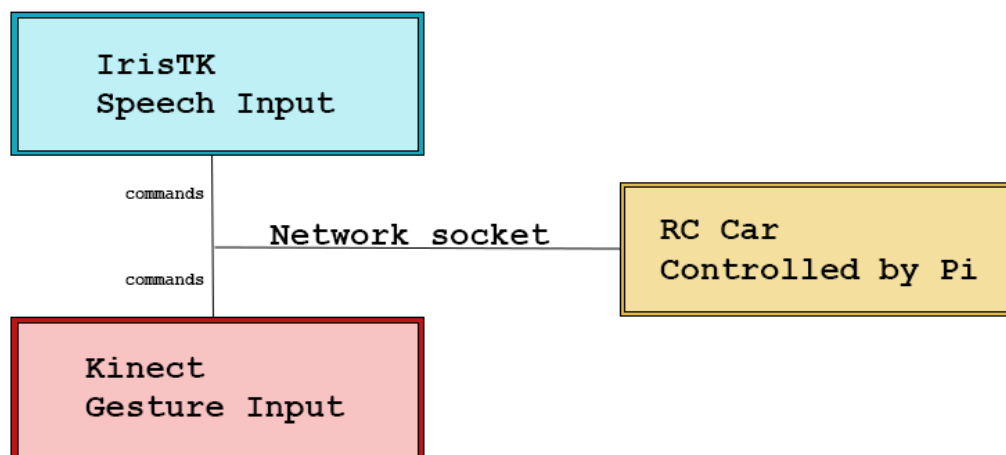
1.2 Η παρούσα υλοποίηση

Η παρούσα διπλωματική εργασία αποτελεί την υλοποίηση ενός συστήματος απομακρυσμένου χειρισμού οχήματος-ρομπότ μέσω φωνητικών εντολών και χειρονομιών.

Χρησιμοποιήθηκε η πλατφόρμα IrisTK για τον προγραμματισμό της λειτουργίας φωνητικών εντολών και ο αισθητήρας Kinect v2 για τον προγραμματισμό κίνησης μέσω χειρονομιών.

Οι ενέργειες που υποστηρίζονται από τη λειτουργία φωνητικών εντολών είναι οι εντολές κίνησης (εμπρόσθια κίνηση, οπισθοδρόμηση, αριστερή και δεξιά στροφή, σταμάτημα), εντολή φωτογραφίας η οποία αποτυπώνει μια φωτογραφία από την κάμερα που βρίσκεται πάνω στο όχημα, και τέλος εντολή για δημιουργία κάτοψης που παρουσιάζει όλα τα εμπόδια που εμφανίστηκαν στις φωτογραφίες που τραβήχθηκαν από το ρομπότ. Οι ενέργειες που υποστηρίζονται από τη λειτουργία χειρονομιών είναι μόνο εντολές κίνησης.

Το όχημα ελέγχεται από ένα Raspberry Pi 2 το οποίο κινεί το όχημα ανάλογα με την εντολή που λαμβάνει κάθε φορά. Οι εντολές αποστέλονται μέσω network sockets. Η αρχιτεκτονική του συστήματος περιγράφεται σύντομα στο σχήμα 1.2.



Σχήμα 1.2: Block Diagram της αρχιτεκτονικής της υλοποίησης

1.3 Συνεισφορά της διπλωματικής

Η παρούσα διπλωματική ασχολείται με τον προγραμματισμό ενός real-time συστήματος με είσοδο φωνητικές εντολές, χειρονομίες ή το συνδυασμό τους και έξοδο την κίνηση ενός ρομπότ-οχήματος σε γνωστό ή άγνωστο χώρο. Στη δεύτερη περίπτωση, οδηγούμενο από το χειριστή του, το όχημα μπορεί να αποτυπώσει μια κάτοψη του χώρου, με όλα τα εμπόδια που μπορεί να αντικρύσει.

Ερευνάται ακόμα η χρονική απόκριση του κάθε τύπου εισόδου, το ποσοστό επιτυχίας αναγνώρισης τόσο των φωνητικών εντολών όσο και των χειρονομιών.

Ένα άλλο αντικείμενο μελέτης της διπλωματικής ήταν η εισαγωγή νέων χειρονομιών στο σύνολο αναγνώρισης του Kinect v2 sensor, το οποίο πραγματοποιήθηκε στη γλώσσα προγραμματισμού C++.

1.4 Περιεχόμενα Κεφαλαίων

Κεφάλαιο 2^ο : Στο κεφάλαιο αυτό αναλύεται η αρχιτεκτονική του συστήματός μας, τα μηχανικά μέρη του ρομπότ καθώς και όλα τα περιφερειακά, εμπεριεχομένου και του Kinect v2 sensor.

Κεφάλαιο 3^ο : Στο κεφάλαιο αυτό αναλύεται η συμπεριφορά του συστήματος με είσοδο αποκλειστικά φωνητικά δεδομένα, παρουσιάζονται κάποια παραδείγματα πειραμάτων βήμα προς βήμα και η ακρίβεια που αυτά είχαν, τόσο για απλή κίνηση, όσο και για τη δισδιάστατη αναπαράσταση αγνώστου χώρου. Επίσης παρουσιάζονται οι γραμματικές που χρησιμοποιήσαμε, ο τρόπος που δημιουργούνται μέσω της πλατφόρμας IrisTK και η λειτουργία κάποιων κρίσιμων για την κατανόηση κομματιών κώδικα, που επεξηγούν βήμα προς βήμα τη διαδικασία.

Κεφάλαιο 4^ο : Στο κεφάλαιο αυτό γίνεται αρχικά μια αναφορά για τη λειτουργία του συστήματος με είσοδο χειρονομίες, παρουσιάζονται συνοπτικά οι νέες χειρονομίες που εισαγάγαμε και κάποια πειράματα χρήσης χειρονομιών για την κίνηση του ρομπότ μαζί με τα αποτελέσματά τους. Επιπρόσθετα, παρουσιάζεται η υλοποίηση της κοινής εισόδου φωνής και χειρονομιών, με τα αντίστοιχα πειράματα και αποτελέσματά τους.

Κεφάλαιο 5^ο : Στο κεφάλαιο αυτό παρουσιάζονται τα πειράματα και τα αποτελέσματα που εξήχθησαν από την κίνηση του οχήματος με διαφορετικές εισόδους, καθώς και της δισδιάστατης χαρτογράφησης, με μορφή γραφημάτων και πινάκων.

Κεφάλαιο 6^ο : Στο κεφάλαιο αυτό παρουσιάζεται η συζήτηση επάνω στη διπλωματική, τα αποτελέσματά της, τη συνεισφορά της, καθώς και τη μελλοντική δουλειά που μπορεί να γίνει με βάση την παρούσα εργασία.

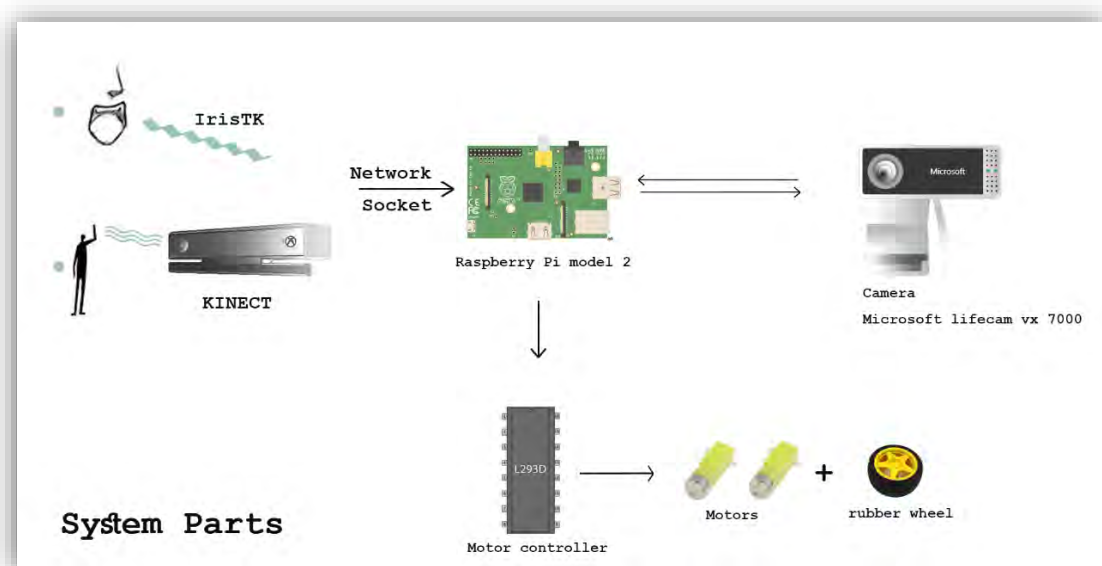
ΚΕΦΑΛΑΙΟ 2^ο

Αρχιτεκτονική του συστήματος

2.1 Αυτοκίνητο ελεγχόμενο από ενσωματωμένο Raspberry Pi

Για την υλοποίηση του συστήματός μας κατασκευάσαμε το παρακάτω ρομπότ-όχημα. Το όχημα κινείται με τη βοήθεια δύο ροδών οι οποίες ελέγχονται από έναν motor controller.

Ο motor controller τροφοδοτείται από τέσσερις μπαταρίες τύπου AA και ελέγχεται μέσω των pins του Raspberry Pi. Τέλος, το raspberry pi ελέγχεται μέσω δικτύου χρησιμοποιώντας network socket, μέσω του οποίου αποστέλονται οι εντολές ελέγχου του ρομπότ, είτε μέσω του IrisTK για φωνητική είσοδο, είτε μέσω κώδικα C++ για είσοδο από το Kinect. Η αρχιτεκτονική του συστήματος φαίνεται στο σχήμα 2.1 και μια φωτογραφία του ρομπότ στο σχήμα 2.2.



Σχήμα 2.1: Αρχιτεκτονική του συστήματος

2.1.1 Μηχανικά μέρη του οχήματος

Το ρομπότ αποτελείται από τα εξής μηχανικά μέρη:

- Shadow chassis: εξωτερικό περίβλημα.
- 2x Rubber wheels: ροδάκια που δένουν πάνω στο εξωτερικό περίβλημα.
- Breadboard Mini: πάνω στο οποίο γίνεται η διασύνδεση.
- Jumper Wires: για την καλωδίωση.
- 2x DC Gear Motors με χαρακτηριστικά:
 - RPM: 125r / minute
 - Current: 80 - 100mA
 - Reduction: 48:1
 - Output Torque: 0.8kg / cm
 - Dimensions: 70x22x18mm
 - Voltage: 3V
- [L293D Motor Controller](#) για να ελέγχει την κίνηση των motors.
- 4 Battery Pack για την τροφοδοσία του Motor Controller.


Για τη λήψη φωτογραφιών χρησιμοποιήθηκε μια web camera Microsoft VX7000. Στον πίνακα 2.1 φαίνονται τα κυριότερα χαρακτηριστικά της.

Version Information	
Product Name	Microsoft® LifeCam VX-7000
Product Version	Microsoft LifeCam VX-7000
Webcam Version	Microsoft LifeCam VX-7000
Product Dimensions	
Webcam Length	4.78 inches (121 millimeters)
Webcam Width	2.72 inches (69.0 millimeters)
Webcam Depth/Height	1.03 inches (26.0 millimeters)
Webcam Weight	4.66 ounces (132 grams)
Webcam Cable Length	72.0 inches (1829 millimeters)
Imaging Features	
Sensor	CMOS 2.0 MP sensor technology
Resolution	• Motion Video: 2 megapixel (1600 x 1190 pixels)* • Still Image: 7.6 megapixel (3200 x 2380 pixels) interpolated*
Field of View	71° diagonal field of view
Imaging Features	• Digital pan, digital tilt, and 5x digital zoom** • Fixed focus • Automatic image adjustment with manual override


Πίνακας 2.1: Κυριότερα τεχνικά χαρακτηριστικά της webcam Microsoft lifecam VX-7000.

2.1.2 Raspberry Pi 2 Model B

Το κυριότερο κομμάτι του συστήματος είναι φυσικά ο ελεγκτής του. Χρησιμοποιήθηκε το Raspberry Pi 2 Model B, τα χαρακτηριστικά του οποίου παρουσιάζονται στο σχήμα 2.2.



Raspberry Pi



Raspberry Pi 2, Model B

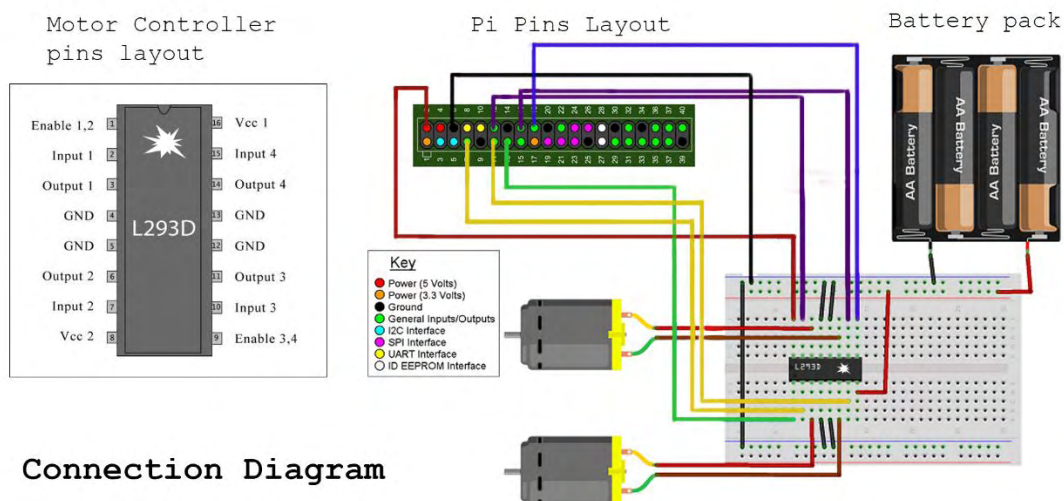
Product Name	Raspberry Pi 2, Model B
Product Description	The Raspberry Pi 2 delivers 6 times the processing capacity of previous models. This second generation Raspberry Pi has an upgraded Broadcom BCM2836 processor, which is a powerful ARM Cortex-A7 based quad-core processor that runs at 900MHz. The board also features an increase in memory capacity to 1Gbyte.
Specifications	
Chip	Broadcom BCM2836 SoC
Core architecture	Quad-core ARM Cortex-A7
CPU	900 Mhz
GPU	Dual Core VideoCore IV® Multimedia Co-Processor Provides Open GL ES 2.0, hardware-accelerated OpenVG, and 1080p30 H.264 high-profile decode Capable of 1Gpixel/s, 1.5Gtexel/s or 24GFLOPs with texture filtering and DMA infrastructure
Memory	1GB LPDDR2
Operating System	Boots from Micro SD card, running a version of the Linux operating system
Dimensions	85 x 56 x 17mm
Power	Micro USB socket 5V, 2A
Connectors:	
Ethernet	10/100 BaseT Ethernet socket
Video Output	HDMI (rev 1.3 & 1.4)
Audio Output	3.5mm jack, HDMI
USB	4 x USB 2.0 Connector
GPIO Connector	40-pin 2.54 mm (100 mil) expansion header: 2x20 strip Providing 27 GPIO pins as well as +3.3 V, +5 V and GND supply lines
Camera Connector	15-pin MIPI Camera Serial Interface (CSI-2)
JTAG	Not populated
Display Connector	Display Serial Interface (DSI) 15 way flat flex cable connector with two data lanes and a clock lane
Memory Card Slot	Micro SDIO

Σχήμα 2.2: Χαρακτηριστικά του Raspberry Pi 2 model B (απο [2]).

Αν και το Raspberry Pi γενικά δεν ανήκει στην κατηγορία μικροελεγκτών κλασικού τύπου (single board microcontrollers) που χρησιμοποιούνται σε παρόμοια projects, (όπως το Arduino) λόγω της μεγάλης σχετικά κατανάλωσης που έχει που δεν το κάνει τόσο ανεξάρτητο όσο τα άλλα, η υπολογιστική του δύναμη είναι πολλές φορές μεγαλύτερη και η δυνατότητά του να υποστηρίξει πλήρη λειτουργικά όπως τα windows 10 ([Internet of Things Edition](#)) ή κάποιες διανομές των linux (Ubuntu [Raspbian Jessie](#)-βασισμένο στο Debian, [Ubuntu Mate](#), [UbuntuSnappy](#)). Προτιμήθηκε η διανομή Raspbian Jessie διότι η συγκεκριμένη διανομή έχει προεγκατεστημένα αρκετά εργαλεία που χρειάστηκαν για τον προγραμματισμό που έγινε.

Το Pi τέλος, τροφοδοτείται από ένα Power Bank ZALMAN ZM-PB84IW (8400MAH).

Στο σχήμα 2.3 φαίνονται τα pin layouts του pi και του motor controller, καθώς και η συγκεκριμένη συνδεσμολογία που χρησιμοποιήσαμε.



Σχήμα 2.3: Συνδεσμολογία που χρησιμοποιήθηκε.

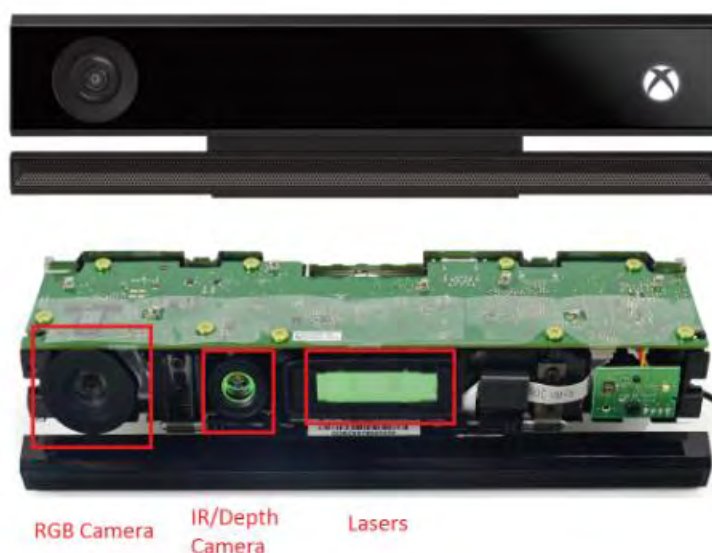
Παρατηρείται ότι κάθε motor απαιτεί τρία σήματα. Το ένα είναι σήμα ελέγχου, αν είναι Low το motor δεν μπορεί να λειτουργήσει, ενώ αν είναι High, το motor περιστρέφεται προς τη φορά που του προσδίδουν τα άλλα δύο σήματα, το ένα για τη φορά του ρολογιού και το άλλο για την αντίστροφη φορά του ρολογιού. Στον πίνακα 2.2 φαίνεται η κίνηση του οχήματος ανάλογα με τις τιμές των σημάτων αυτών.

Motor1E	Motor2E	Motor1A	Motor1B	Motor2A	Motor2B	KΙΝΗΣΗ
0	0	X	X	X	X	STOP
0	1	X	X	X	X	DF
1	0	X	X	X	X	DF
1	1	1	0	1	0	FORWARD
		0	1	0	1	BACKWARDS
		1	0	0	1	TURN RIGHT
		0	1	1	0	TURN LEFT
		0	0	0	0	DF
		0	0	0	1	
		0	0	1	0	
		0	0	1	1	
		0	1	0	0	
		0	1	1	1	
		1	0	0	0	
		1	0	1	1	
		1	1	0	0	
		1	1	0	1	
		1	1	1	0	
		1	1	1	1	

Πίνακας 2.2: Όλοι οι δυνατοί συνδυασμοί των pins του motor controller και πως μεταφράζονται για την κίνηση του ρομπότ. DF σημαίνει disfunction, αντίστοιχη κίνηση δεν υπάρχει και δεν προβλέπεται να έρθουν ποτέ σε αυτή την κατάσταση τα motors. Αν ωστόσο γίνει κάτι τέτοιο, η κίνηση θα είναι ελαττωματική.

2.2 Kinect v2 sensor για ανίχνευση κίνησης

Το Kinect είναι ένας αισθητήρας που πρωτοκυκλοφόρησε από τη Microsoft για την κονσόλα παιχνιδιών Xbox 360 τον Νοεμβριο του 2010, ενώ η αναβαθμισμένη του έκδοση Kinect v2 το 2014 (Σχήμα 2.4). Δύο χρόνια μετά τη λειτουργία του πρώτου, κυκλοφόρησε ο adaptor που το κάνει συμβατό και με υπολογιστή.

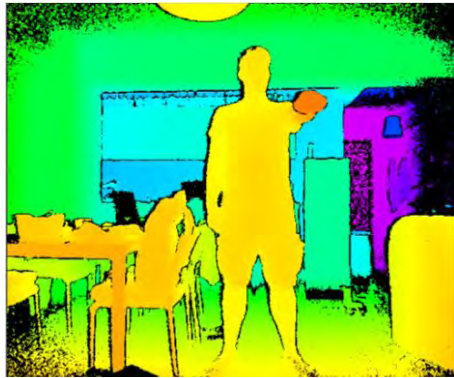


Σχήμα 2.4: Πάνω: Εξωτερική όψη του Kinect v2 sensor. Κάτω: Μετά την αφαίρεση καλύμματος, φαίνονται οι αισθητήρες και οι εκπομποί των lasers (από [3]).

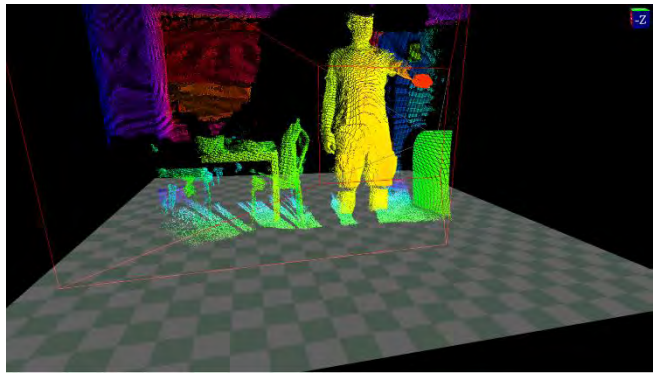
Η δεύτερη έκδοσή του, που χρησιμοποιείται και στην παρούσα υλοποίηση διαθέτει δύο κάμερες, μία RGB που λαμβάνει φωτογραφίες με χρώμα σε ανάλυση 1920x1080 pixels και μία IR (infrared) η οποία χρησιμοποιείται για να εξαχθούν χάρτες αποστάσεως σε πραγματικό χρόνο (real-time depthmaps) [13]. Έχει τη δυνατότητα να δημιουργεί τρισδιάστατα μοντέλα στο χώρο, παράγοντας αρχικά μια χαρτογράφηση βάθους (depth mapping), είτε σε infrared, είτε σε ασπρόμαυρες αποχρώσεις, και έπειτα το τελικό τρισδιάστατο μοντέλο (3d point cloud) εξάγοντας χαρακτηριστικά από το depth map. Στο σχήμα 2.5 παρουσιάζεται η διαδικασία κατασκευής ενός τρισδιάστατου μοντέλου σε infrared και σε ασπρόμαυρες αποχρώσεις.



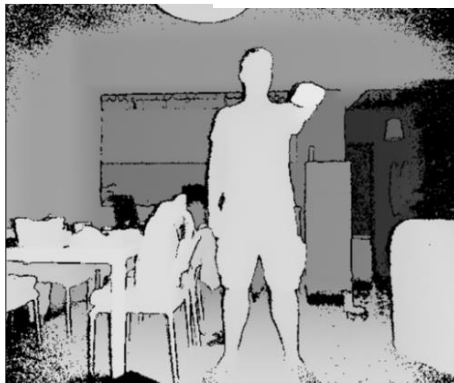
α



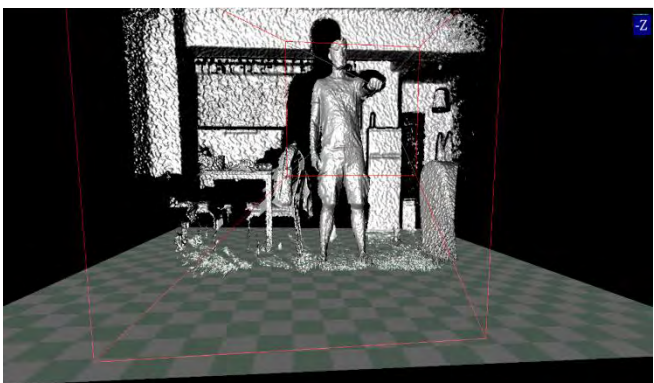
β



γ



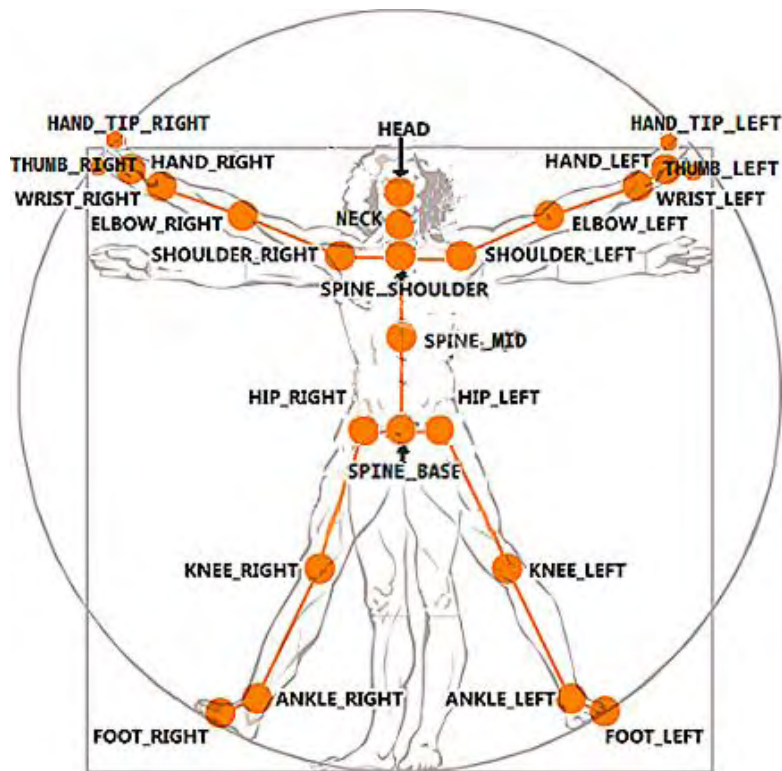
δ



ε

Σχήμα 2.5: Διαδικασία εξαγωγής 3D point cloud μοντέλου. (α) Αρχική φωτογραφία (β) Depth map σε infrared (γ) 3d point cloud σε infrared, χρησιμοποιώντας το β (δ) Depthmap σε grayscale (ε) 3D point cloud σε grayscale

Το Kinect λοιπόν μπορεί να αναγνωρίσει και να ανακατασκευάσει μοντέλα στον τρισδιάστατο χώρο με μεγάλη ευκολία, δίνοντας ταυτόχρονα τη δυνατότητα μέσω του SDK που παρέχεται να προγραμματιστεί κατάλληλα για οποιαδήποτε χρήση σχετίζεται με αναγνώριση κίνησης. Υποστηρίζει αναγνώριση μέχρι και έξι ατόμων ταυτόχρονα, με το σώμα κάθε ατόμου να χωρίζεται σε είκοσι έξι βασικές αρθρώσεις, έτσι ώστε οποιαδήποτε κίνηση κάνει το ανθρώπινο σώμα να μπορεί να αναγνωριστεί με ακρίβεια. Στο σχήμα 2.6 φαίνονται όλα αυτά τα σημεία.



Σχήμα 2.6: Όλα τα σημεία – αρθρώσεις του ανθρώπινου σώματος που μπορεί να αναγνωρίσει με ακρίβεια το Kinect v2 sensor (από [4]).

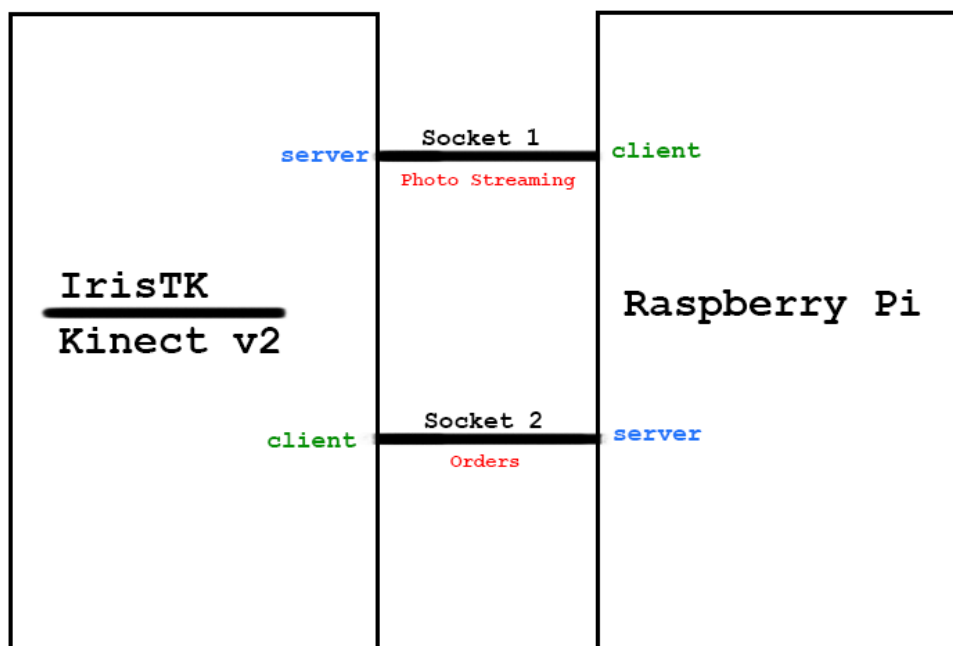
2.3 Network socket

Για να αποσταλεί η εντολή προς το Raspberry Pi που ελέγχει το ρομπότ, είτε είναι ηχητική εντολή, είτε μέσω χειρονομιών του χρήστη του Kinect v2 sensor, χρησιμοποιείται ένα network socket.

Ως network socket, ορίζεται ένα κανάλι επικοινωνίας μεταξύ μηχανών που συνυπάρχουν στο ίδιο δίκτυο. Επομένως, ένα socket μπορεί να υλοποιηθεί πάνω από το διαδίκτυο (internet socket) ή όπως στα πειράματά μας, πάνω από το τοπικό δίκτυο.

Στην παρούσα υλοποίηση, χρησιμοποιήθηκαν δύο διαφορετικά network sockets. Το πρώτο αφορά στην παρουσίαση του Live streaming απο την κάμερα του ρομπότ, που θεωρείται client side και την αποστολή των φωτογραφιών πίσω στον υπολογιστή που λογίζεται ως server side. Στο δεύτερο, οι ρόλοι του server και του client αντιστρέφονται, καθώς το ρομπότ λειτουργεί ως server και περιμένει να λάβει εντολές από πιθανούς clients, που μπορεί να είναι φωνητικές εντολές μέσω IrisTK ή χειρονομίες μέσω του Kinect v2 sensor.

Στο σχήμα 2.7 φαίνεται η λειτουργία των δύο αυτών sockets.



Σχήμα 2.7 Τα δύο network sockets που χρησιμοποιήθηκαν, ένα για το streaming της εικόνας πίσω στο IrisTK και ένα για τις εντολές προς το ρομπότ.

ΚΕΦΑΛΑΙΟ 3^ο

Λειτουργία με είσοδο φωνητικές εντολές

3.1 Η πλατφόρμα IrisTK

Η πλατφόρμα [IrisTK](#) δημιουργήθηκε από τους Gabriel Skantze και Samer Al Moubayed το 2013. Σκοπός της ήταν να παρέχει ένα εργαλείο για προγραμματισμό real-time συστημάτων που περιέχουν την πρόσωπο με πρόσωπο αλληλεπίδραση πολλών χρηστών μεταξύ τους [1]. Ωστόσο, πέραν της κύριας χρήσης τους, αποτελεί ένα πολύ γρήγορο εργαλείο προγραμματισμού συστημάτων που απαιτούν χρήση αναγνώρισης φωνής, εξαγωγή χαρακτηριστικών προσώπου, χειρονομιών ή κίνησης σε συνδυασμό με επικοινωνία ανθρώπου μηχανής. Το framework είναι προγραμματισμένο σε γλώσσα Java και χρησιμοποιεί ένα XML αρχείο για τη γρήγορη δημιουργία συστημάτων αλληλεπίδρασης βασισμένα σε γεγονότα, μέσω της μετατροπής του XML αρχείου σε κώδικα Java. Επίσης δίνει τη δυνατότητα στο χρήστη να χρησιμοποιεί διάφορα πρωτόκολλα γραμματικών για να δημιουργήσει τις γλώσσες που θα αναγνωρίζονται.

3.1.2 Γραμματική SRGS

Το IrisTK χρησιμοποιεί μόνο γραμματικές χωρίς συμφραζόμενα για να δημιουργήσει τη γραμματική που μπορεί να αναγνωρίσει από φωνητική είσοδο. Γραμματικές δηλαδή για τις οποίες ισχύει: $G = (V, \Sigma, R, S)$ όπου:

- V είναι το σύνολο μη τερματικών συμβόλων
- Σ είναι το σύνολο των τερματικών συμβόλων
- R είναι το σύνολο κανόνων
- S είναι η αρχική κατάσταση

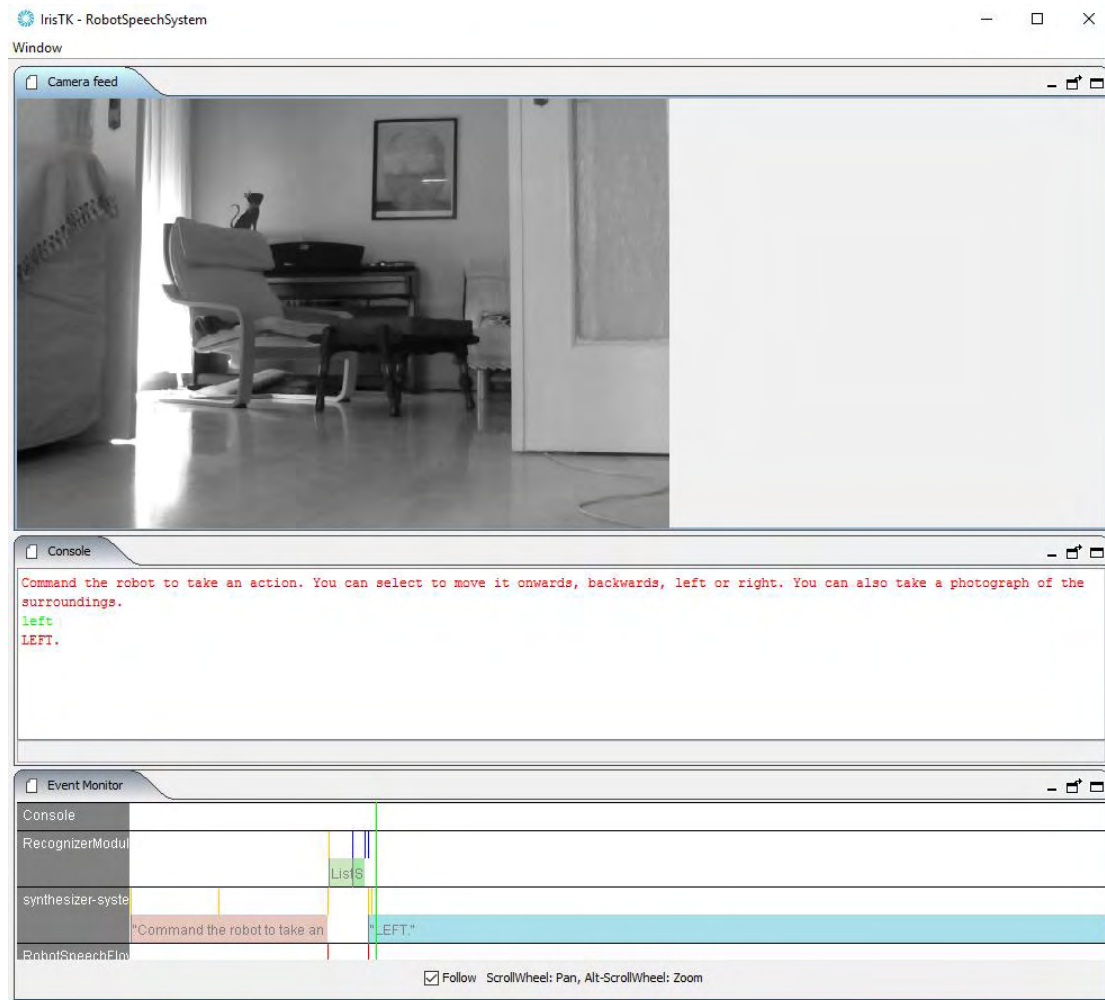
Η προεπιλεγμένη μορφή του IrisTK χρησιμοποιεί το πρωτόκολλο SRGS του οργανισμού [W3C](#) (world wide web consortium) το οποίο είναι σε μορφή XML και προτιμήθηκε στην παρούσα υλοποίηση. Εκτός του SRGS, το IrisTK υποστηρίζει και άλλα formats όπως το ABNF format που δεν αποτυπώνεται σε XML, και άρα είναι πιο ευανάγνωστο. Για να συνθέσει ομιλία, η πλατφόρμα χρησιμοποιεί τον κατάλογο UPS (universal phone set). Ο UPS περιέχει σύμβολα κατανοητά από υπολογιστή, τα οποία αντιστοιχίζονται σε ένα φώνημα σύμφωνα με αντίστοιχη χρήση του φωνήματος στη συγκεκριμένη γλώσσα. Στο σχήμα 3.1 παρουσιάζεται ο κατάλογος UPS.

UPS	SAMPA	IPA	English								
P	p	p	put	G	g	g	gut	OI	oi	ɔ.i	toy
B	b	b	big	NG	N	ŋ	sing	AI	ai	ɑ.i	bite
M	m	m	mat	H	h	h	help	IX	i@	i.ə	fear
F	f	f	fork	CH	tʃ	tʃ / tʃ	chin	EHX	e@	ɛ.ə	stairs
V	v	v	vat	JH	dʒ	d.ʒ / dʒ	joy	UWX	u@	u.ə	lure
TH	T	θ	thin	I	i:	i	feel	OWX	@U	o.ə	boa
DH	D	ð	then	U	u:	u	too	AOX	U@	ɔ.ə	four
T	t	t	talk	IH	ɪ	ɪ	fill	_S			SILENCE
D	d	d	dig	UH	U	ʊ	book				
N	n	n	no	O	@U	o	go				
DX	t	r	butter (US)	AX	@	ə	ago				
S	s	s	sit	EH	e	ɛ	pet				
Z	z	z	zap	ER	ɜ:	ɜ	bird (UK)				
L	l	l	lid	AH	{	ʌ	cut				
SH	S	ʃ	she	AO	Q	ɔ	dog				
ZH	Z	ʒ	pleasure	AE	{	æ	cat				
R	r	r	red	AA	A:	ɑ	father				
J	j	j	yard	Q	Q	ɒ	hot				
W	w	w	with	EI	ei	e.i	late				
K	k	k	cut	AU	aU	ɑ.ʊ	foul				

Πίνακας 3.1: UPS phone set (από [5]).

3.1.3 Εισαγωγή streaming στην πλατφόρμα

Το IrisTK δίνει τη δυνατότητα στο χρήστη να τοποθετήσει ή να απομακρύνει με απλό τρόπο οποιοδήποτε στοιχείο από το κύριο JFrame του, δηλαδή το κύριο παράθυρο εκτέλεσης. Με τη χρήση μιας απλής συνάρτησης getGUI μπορεί ο χρήστης να λάβει την τρέχουσα διάταξη του παραθύρου εκτέλεσης και με τη χρήση μια δεύτερης συνάρτησης addDockPanel μπορεί να προσθέσει κάποιο συστατικό στο παράθυρο αυτό. Στο συγκεκριμένο project, εκτός των default panels συνομιλίας και event motitor που χρησιμοποιήθηκαν, προστέθηκε ένα JPanel για Live Streaming της κάμερας που βρίσκεται πάνω στο ρομπότ, έτσι ώστε να παρέχει τη λειτουργικότητά του ακόμα και από απομακρυσμένη περιοχή. Λόγω μικρής ταχύτητας αποστολής των frames ωστόσο, το feed παρουσιάζεται σε grayscale για να μειωθεί η ταχύτητα αποστολής και εκτέλεσης. Ένα στιγμιότυπο τρεξίματος παρουσιάζεται στο σχήμα 3.1.



Σχήμα 3.1: Στιγμιότυπο τρεξίματος.

3.2 Περιγραφή λειτουργίας του συστήματος

Αρχικά το σύστημα περιμένει κάποια εντολή προς το ρομπότ. Αν η είσοδος που λάβει ανήκει στη γλώσσα που αναγνωρίζεται, τότε γίνεται αποδεκτή, και επιστρέφει στην αντίστοιχη μεταβλητή την προβλεπόμενη τιμή. Διαφορετικά, ζητάει εκ νέου είσοδο. Στο σχήμα 3.2 φαίνεται η γλώσσα της υλοποίησης σε μορφή SRGS XML και στον πίνακα 3.2 όλες οι αποδεκτές προτάσεις.

```

<?xml version="1.0" encoding="utf-8"?>
<grammar xml:lang="en-US" version="1.0" root="root"
  xmlns="http://www.w3.org/2001/06/grammar"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/06/grammar
http://www.iristk.net/xml/srgs.xsd" tag-format="semantics/1.0">

  <rule id="root" scope="public">
    <one-of>
      <item>forward<tag>out.action=1</tag></item>
      <item>go forward<tag>out.action=1</tag></item>
      <item>go straight<tag>out.action=1</tag></item>
      <item>move forward<tag>out.action=1</tag></item>
      <item>move straight<tag>out.action=1</tag></item>

      <item>move backwards<tag>out.action=2</tag></item>
      <item>move back<tag>out.action=2</tag></item>
      <item>go backwards<tag>out.action=2</tag></item>
      <item>go back<tag>out.action=2</tag></item>
      <item>backwards<tag>out.action=2</tag></item>

      <item>left<tag>out.action=3</tag></item>
      <item>turn left<tag>out.action=3</tag></item>
      <item>go left<tag>out.action=3</tag></item>

      <item>stop<tag>out.action=5</tag></item>

      <item>right<tag>out.action=4</tag></item>
      <item>turn right<tag>out.action=4</tag></item>
      <item>go right<tag>out.action=4</tag></item>

      <item>exit<tag>out.action=0</tag></item>

      <item>take a photo<tag>out.action=10</tag></item>
      <item>take a picture<tag>out.action=10</tag></item>
      <item>take a photograph<tag>out.action=10</tag></item>
      <item>photo<tag>out.action=10</tag></item>
      <item>picture<tag>out.action=10</tag></item>
      <item>photograph<tag>out.action=10</tag></item>

      <item>yes<tag>out.yes=1</tag></item>
      <item>no<tag>out.no=1</tag></item>

    </one-of>
  </rule>

</grammar>

```

Σχήμα 3.2: Γλώσσα της υλοποίησης.

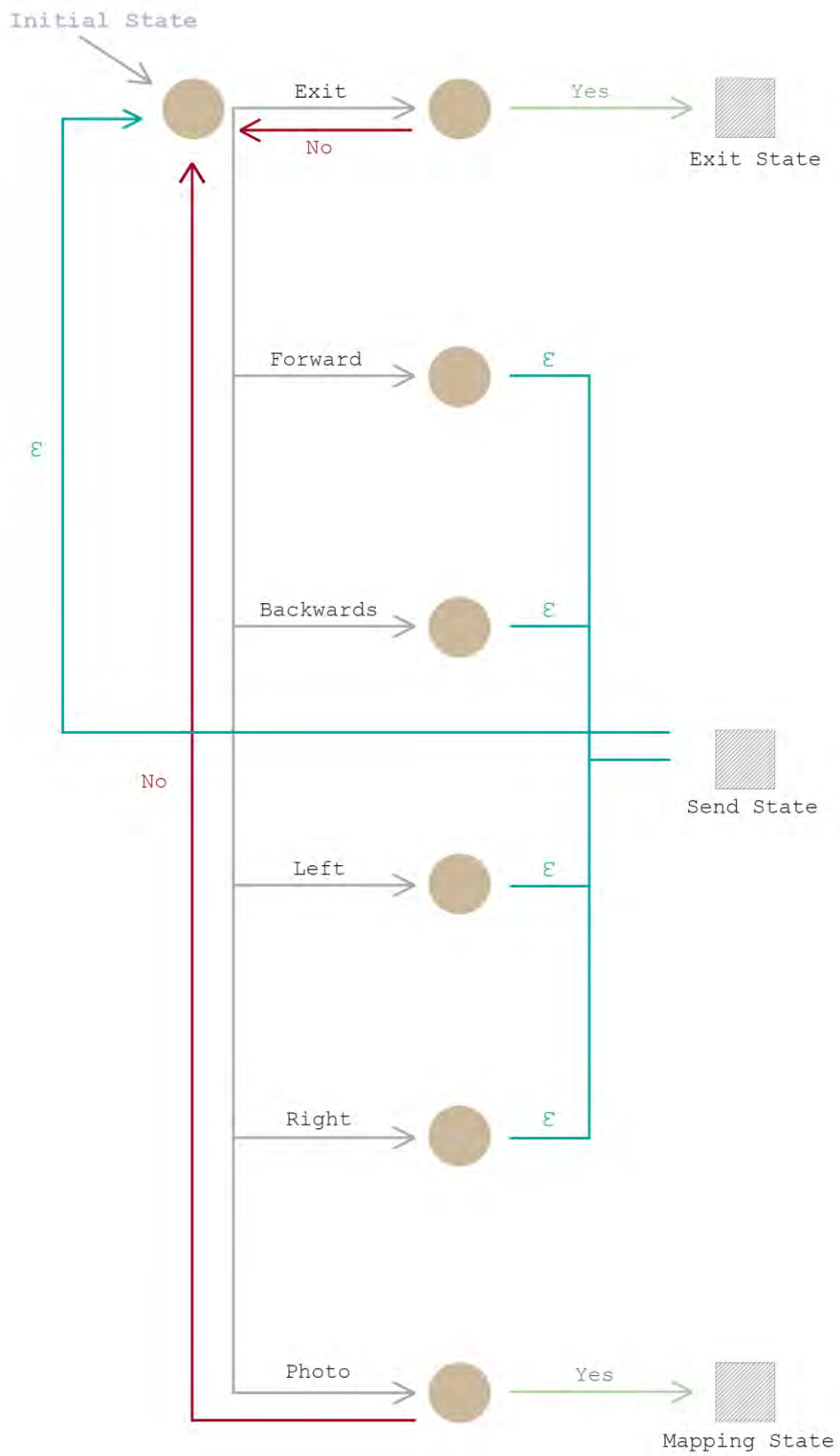
	Προτάσεις που ανήκουν στη γλώσσα					
Εμπρόσθια κίνηση	Forward	Go forward	Move forward	Go straight	Move straight	-
Οπίσθοδρομηση	Backwards	Go back	Go backwards	Move back	Move backwards	-
Αριστερή στροφή	Left	Turn left	Go left	-	-	-
Δεξιά στροφή	Right	Turn right	Go right	-	-	-
Σταμάτημα	Stop	-	-	-	-	-
Αποτύπωση φωτογραφίας	Photograph	Picture	Photo	Take a photo	Take a photograph	Take a picture
Όχι	No	-	-	-	-	-
Ναι	Yes	-	-	-	-	-
Έξοδος	Exit	-	-	-	-	-

Πίνακας 3.2: Όλες οι αποδεκτές προτάσεις που αναγνωρίζει η γλώσσα της υλοποίησης.

Για παράδειγμα, αν ο χρήστης δώσει ως είσοδο τη λέξη “Left”, η οποία ανήκει στη γλώσσα και άρα αναγνωρίζεται, η μεταβλητή action θα πάρει την τιμή 3 και θα την επιστρέψει στη ροή του προγράμματος, ώστε να αποστείλει στο ρομπότ την κίνηση που αντιστοιχεί σε αυτή.

Αφού αναγνωριστεί η κατάλληλη εντολή προς το ρομπότ, το σύστημα διαλόγου του IrisTK θα απαντήσει με τον προβλεπόμενο τρόπο, ή θα ζητήσει εκ νέου είσοδο αν δεν αναγνωρίστηκε η προηγούμενη εντολή, ή υπήρξε πολύς θόρυβος/απόλυτη ησυχία. Η εντολή αποστέλεται μέσω του socket στο ρομπότ, απο το οποίο χειρίζεται μέσω του Raspberry, έτσι ώστε να εκτελεστεί η κίνηση ή να βγει μια φωτογραφία ή και η κάτοψη εμποδίων.

Το FSM επικοινωνίας με όλες τις πιθανές μεταβάσεις καταστάσεων φαίνεται στο σχήμα 3.3, όπου το Send State αντιπροσωπεύει την αποστολή της εντολής μέσω του socket, το Exit State κλείνει ομαλά το πρόγραμμα και το Mapping State δημιουργεί και αποθηκεύει τη χαρτογράφηση των εμποδίων που το ρομπότ έχει εντοπίσει μέχρι στιγμής στις φωτογραφίες που έχει βγάλει. Τέλος, αφού ληφθεί η εντολή, το ρομπότ εκτελεί την προβλεπόμενη ενέργεια και περιμένει νέα εντολή.



Σχήμα 3.3 FSM της λειτουργίας του συστήματος.

3.3 Χαρτογράφηση εμποδίων στο δισδιάστατο επίπεδο

3.3.1 Συνθήκες λειτουργίας και δεδομένα

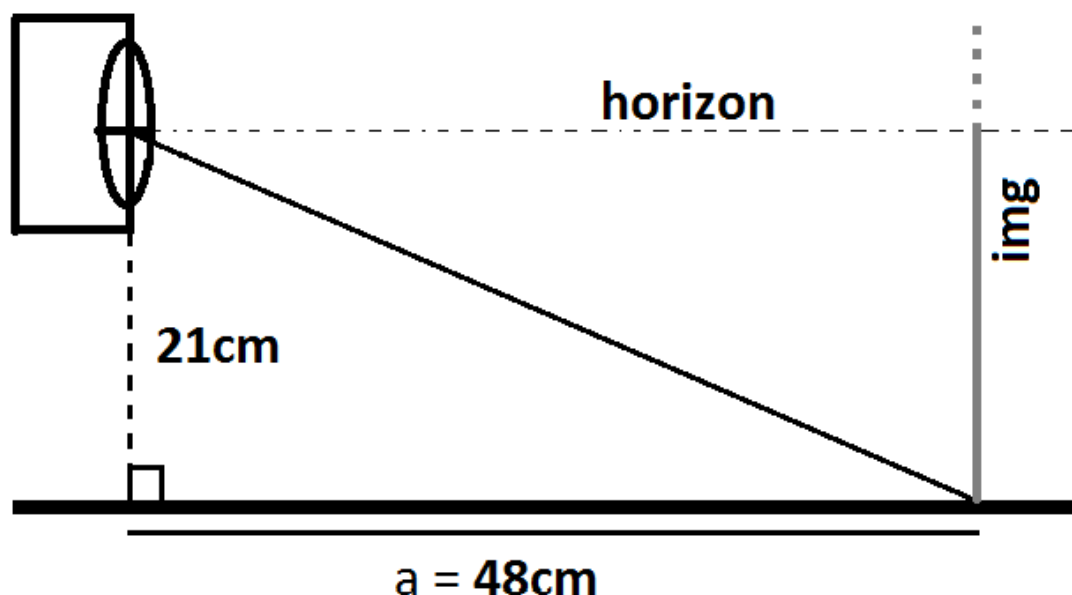
Η αναγνώριση βάθους σε μη ιδανικές και άγνωστες συνθήκες απαιτεί είτε δύο κάμερες ή ζεύγη εικόνων. Όμως σε εξιδανικευμένες συνθήκες μπορεί να επιτευχθεί μία βασική αναγνώριση βάθους, ακόμα και με μονές φωτογραφίες. Οι παραδοχές που γίνονται για να δουλέψει αυτή η μεθοδολογία είναι οι παρακάτω:

- Πρέπει να γνωρίζουμε την απόσταση της κάμερας από το έδαφος
- Πρέπει να γνωρίζουμε την κλίση της κάμερας
- Πρέπει το έδαφος να είναι επίπεδο και χωρίς κλίση
- Τέλος, πρέπει να γνωρίζουμε πόσο απέχει από την κάμερα το χαμηλότερο σημείο της φωτογραφίας

Στην δική μας περίπτωση ισχύει ότι:

- Η κάμερα βρίσκεται σε ύψος 21 cm
- Η κάμερα είναι παράλληλη με το έδαφος που σημαίνει ότι ο ορίζοντας βρίσκεται ακριβώς στη μέση της φωτογραφίας
- Η απόσταση του χαμηλότερου σημείου της φωτογραφίας είναι 48 cm

Στο σχήμα 3.4 φαίνεται η θέση της κάμερας στην παρούσα υλοποίηση.



Σχήμα 3.4 : Θέση της κάμερας πάνω στο robot.

Με δεδομένα τα παραπάνω, ο αλγόριθμος εφαρμόζει διάφορα φίλτρα στην φωτογραφία και χρησιμοποιεί βασικές τριγωνομετρικές ταυτότητες για να υπολογίσει τις αποστάσεις. Λειτουργεί ως εξής:

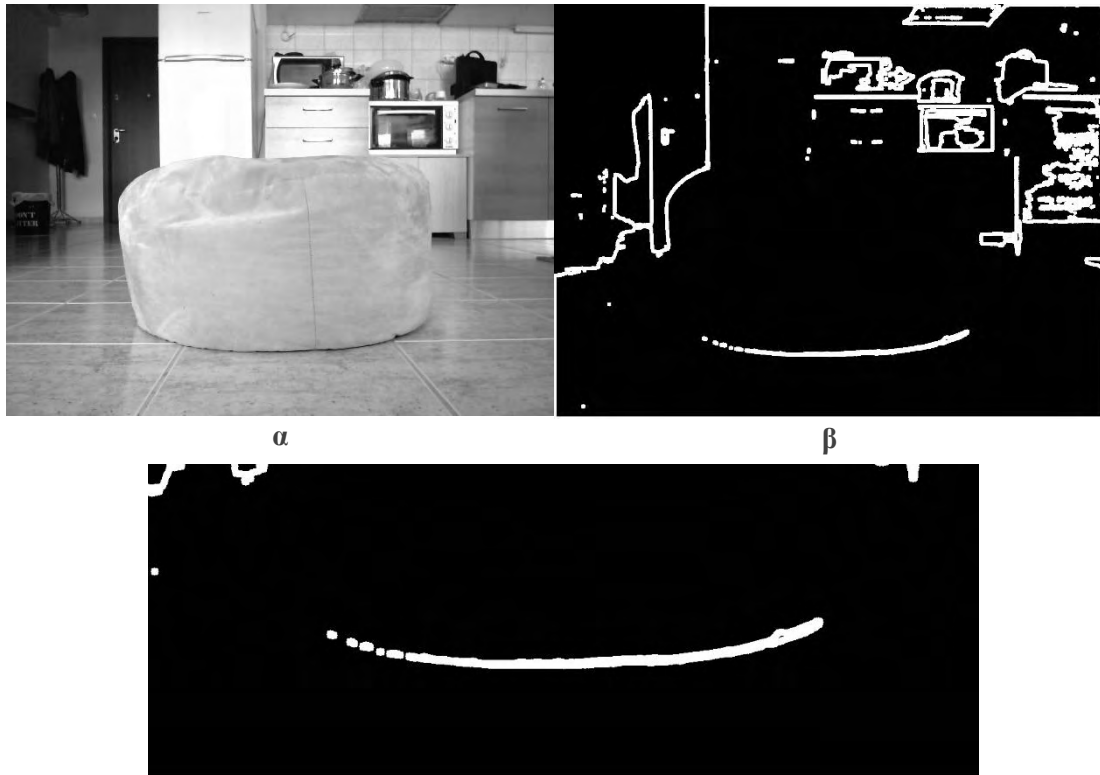
- 1) Βγάζει φωτογραφία και εφαρμόζει τα φίλτρα ώστε να ξεχωρίσουν τα εμπόδια
- 2) Σκανάρει την εικόνα από κάτω προς τα πάνω και από αριστερά προς τα δεξιά
- 3) Μόλις εντοπίσει ένα εμπόδιο τότε υπολογίζει την απόσταση και την γωνία του εμποδίου
- 4) Το σκανάρισμα συνεχίζεται μέχρι ο υπολογισμός να φτάσει στο δεξί όριο της εικόνας
- 5) Στο τέλος εξάγεται μια λίστα με τις αποστάσεις και τις γωνίες όλων των εμποδίων
- 6) Ο χειριστής μπορεί να επιλέξει την μετακίνηση του ρομπότ και την επανάληψη της παραπάνω διαδικασίας με νέα φωτογραφία από άλλο σημείο
- 7) Αλλιώς ο χειριστής μπορεί να επιλέξει τον τερματισμό του αλγορίθμου. Τότε οι λίστες με τα εμπόδια ζωγραφίζονται σε ένα τετράγωνο που αποτελεί και την κάτοψη των εμποδίων αυτών.

3.3.2 Επεξεργασία εικόνας – Εφαρμογή φίλτρων

Για να ξεχωρίσουμε τα εμπόδια σε μια φωτογραφία χρησιμοποιούμε τα παρακάτω φίλτρα.

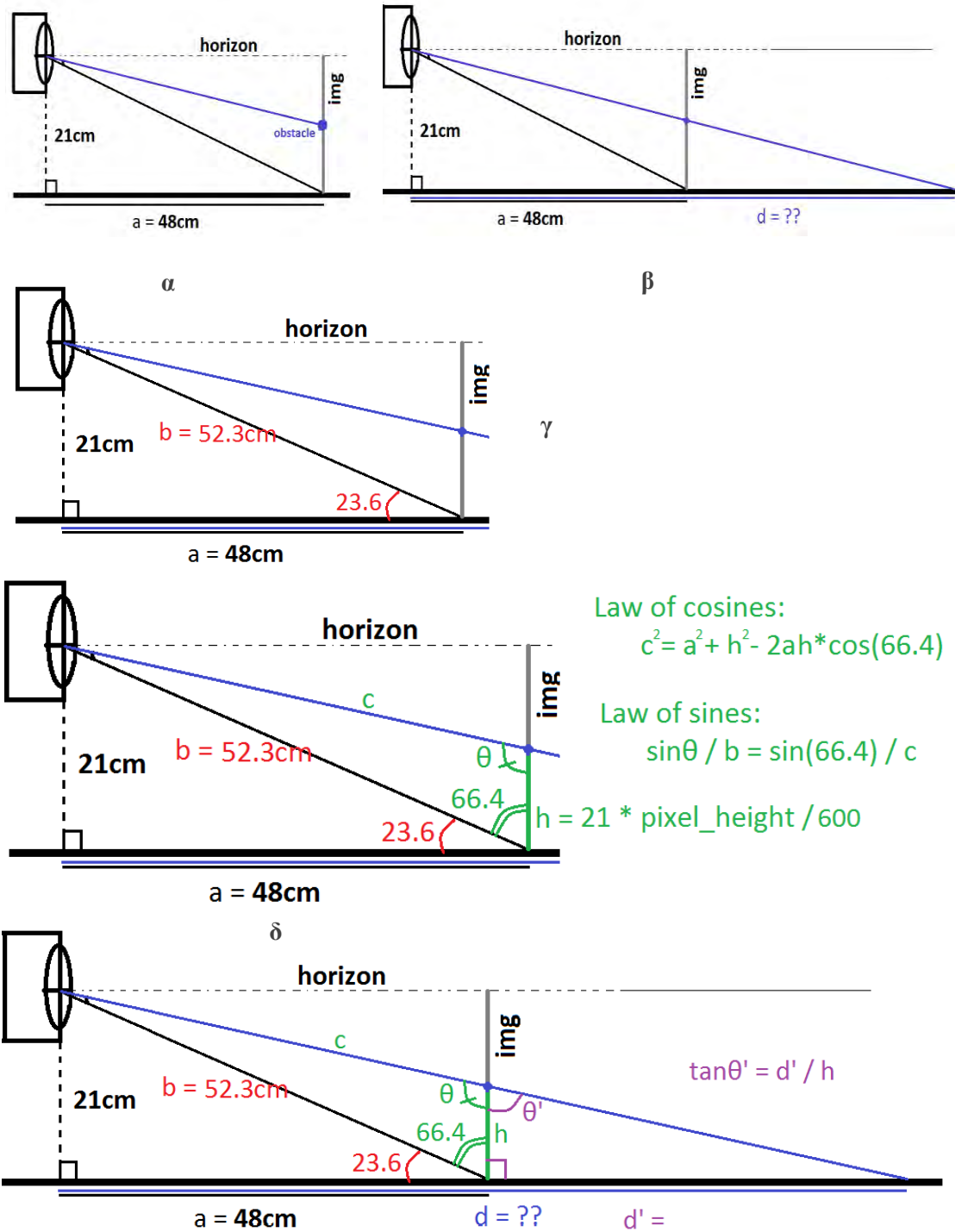
- Η εικόνα από έγχρωμη μετατρέπεται σε ασπρόμαρη.
- Εφαρμόζεται ένα απλό φίλτρο **threshold** με όριο το την τιμή 100, έτσι ώστε κάθε pixel με τιμή μικρότερη του 100 να γίνεται μηδέν (μαύρο).
- Βρίσκονται τα contours τα οποία σχεδιάζονται σε μια νέα εικόνα για να βρεθούν οι γωνίες και οι καμπύλες που σημαίνουν πιθανά εμπόδια.
- Αποκόπτεται ένα μέρος της εικόνας για πιο γρήγορους και ακριβείς υπολογισμούς.

Το σχήμα 3.5 απεικονίζει τη διαδικασία της επεξεργασίας που υποβάλλεται η εικόνα:



Σχήμα 3.5: (α) Αρχική εικόνα σε grayscale (β) Εφαρμογή threshold και contour finding/drawing για edge detection (γ) Αποκοπή της εικόνας για ελαχιστοποίηση error

Μετά το πέρας της επεξεργασίας της εικόνας, έρχεται η διαδικασία της αντιστοίχισης των εμποδίων της παραγόμενης εικόνας με απόσταση στον πραγματικό χώρο. Χάρη στις παραδοχές που κάναμε για τη θέση της κάμερας, μπορούμε να χρησιμοποιήσουμε απλές τριγωνομετρικές ταυτότητες για να υπολογίσουμε την απόσταση και την γωνία των εμποδίων με ικανοποιητική ακρίβεια. Το σχήμα 3.6 παρουσιάζει βηματικά όλα τα στάδια που ακολουθούνται για τον υπολογισμό της απόστασης των εμποδίων από τη βάση της κάμερας.

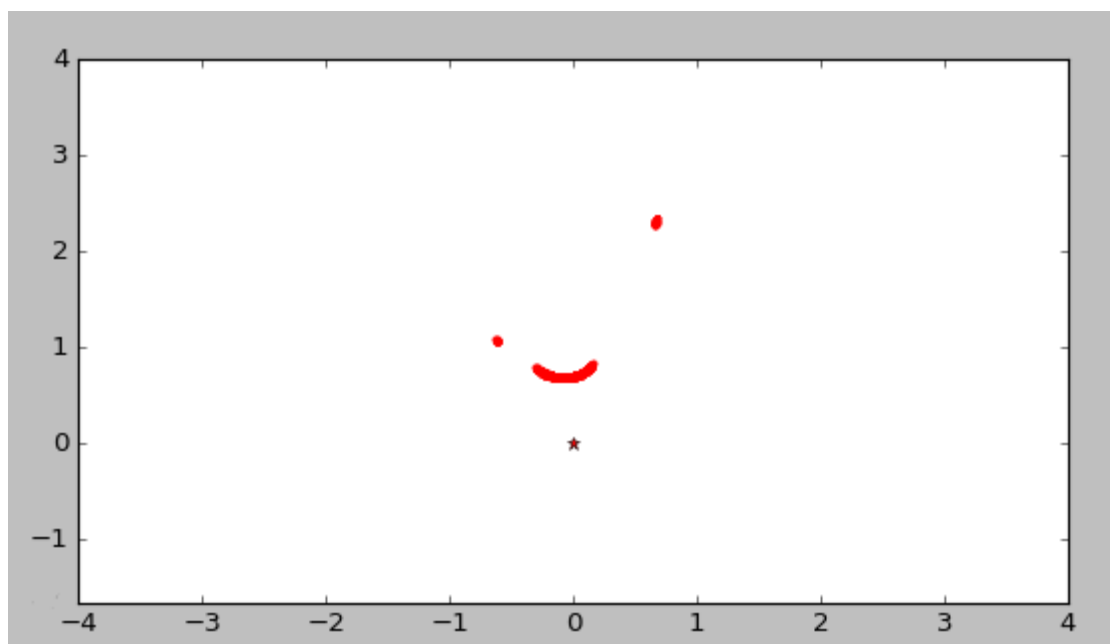


Σχήμα 3.6: (α) Αρχική εικόνα εμποδίων με μπλε να είναι ένα εμπόδιο. (β) Η ζητούμενη απόσταση d είναι η πραγματική απόσταση του αντικειμένου από τη βάση της κάμερας. (γ) Σχηματισμός δύο τριγώνων. (δ) Υπολογισμός του c , h και θ . (ε) Υπολογισμός του d' , που οδηγεί και στον υπολογισμό του ζητούμενου d .

Με παρόμοιο τρόπο αλλά χρησιμοποιώντας τα οριζόντια pixel υπολογίζουμε την γωνία των εμποδίων (δηλαδή το πόσο αριστερά ή δεξιά βρίσκεται το εμπόδιο σε σχέση με το ρομπότ).

3.3.3 2D απεικόνιση των εμποδίων

Με χρήση παρόμοιας τριγωνομετρίας όπως και προηγουμένως, γίνεται η αντιστοίχιση των αποστάσεων των εμποδίων στο δισδιάστατο επίπεδο. Λαμβάνοντας ως παραδοχή ότι η αρχική θέση του ρομπότ στο χώρο είναι η $(0,0)$ και πως η αρχική κατεύθυνση που «κοιτάζει» είναι 90° , τοποθετούνται τα εμπόδια κατάλληλα στο χώρο, ανάλογα με τις κινήσεις που έχουν προηγηθεί και τη σχετική τους απόσταση από την κάμερα. Το σχήμα 3.7 αποτυπώνει την κάτοψη εμποδίων από τη φωτογραφία του σχήματος 3.5.



Σχήμα 3.7: Αποτέλεσμα κάτοψης εμποδίων με εισοδο τη φωτογραφία του σχήματος 3.5.

3.4 Πειράματα

3.4.1 Πειράματα κίνησης

Η ποιότητα του μικροφώνου που χρησιμοποιείται για τις φωνητικές εντολές διαδραματίζει προφανώς μεγάλο ρόλο στη σωστή αναγνώριση της κάθε εντολής. Δοκιμάστηκαν τρία διαφορετικά μικρόφωνα, το ενσωματωμένο μικρόφωνο του Kinect v2 sensor, το μικρόφωνο της [webcam Logitech c270](#) και ένα μικρόφωνο χαμηλού κόστους [Trust Primo](#). Για την κίνηση του ρομπότ με είσοδο φωνητικές εντολές, δοκιμάστηκε η κάθε εντολή κίνησης από 100 φορές και από τρεις διαφορετικές αποστάσεις από το μικρόφωνο. Τα αποτελέσματα ανα εντολή παρουσιάζονται στους πίνακες 3.3, 3.4, 3.5, 3.6.

FORWARD COMMAND	C270 Microphone	Trust Primo	Kinect Microphone
10cm	83%	90%	93%
30cm	80%	88%	90%
1m	60%	70%	78%

Πίνακας 3.3: Ποσοστά αναγνώρισης για την εντολή Forward.

BACKWARDS COMMAND	C270 Microphone	Trust Primo	Kinect Microphone
10cm	84%	91%	95%
30cm	80%	88%	92%
1m	63%	72%	80%

Πίνακας 3.4: Ποσοστά αναγνώρισης για την εντολή Backwards.

LEFT COMMAND	C270 Microphone	Trust Primo	Kinect Microphone
10cm	90%	95%	98%
30cm	85%	92%	95%
1m	80%	85%	92%

Πίνακας 3.5: Ποσοστά αναγνώρισης για την εντολή Left.

RIGHT COMMAND	C270 Microphone	Trust Primo	Kinect Microphone
10cm	88%	90%	93%
30cm	82%	87%	90%
1m	75%	80%	85%

Πίνακας 3.6: Ποσοστά αναγνώρισης για την εντολή Right.

3.4.2 Πειράματα mapping εμποδίων

Η λειτουργία αυτή παρουσιάζει διαφορετικές συμπεριφορές ανάλογα με το περιβάλλον χρήσης, καθώς ο αλγόριθμος που δημιουργήθηκε δεν έχει μεγάλες ανοχές σε ακραίες τιμές φωτεινότητας ή σε ανώμαλο έδαφος. Στην πρώτη περίπτωση, η δημιουργία σκιάσεων των αντικειμένων οδηγεί στο να θεωρούνται οι ίδιες οι σκιές εμπόδια και έτσι να αποτυπώνονται, λανθασμένα, στην τελική δισδιάστατη αναπαράσταση.

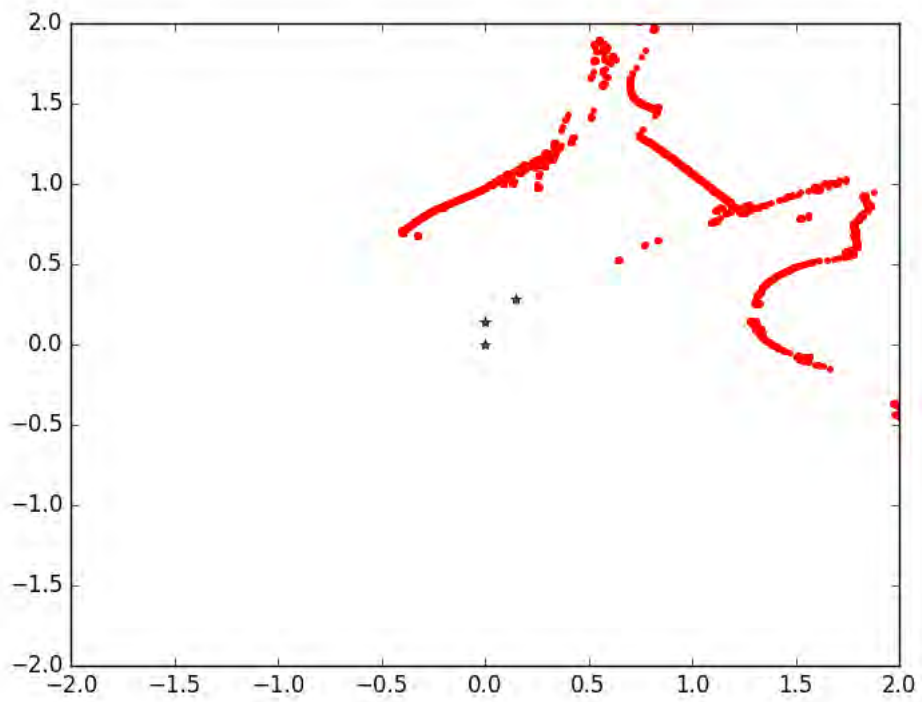
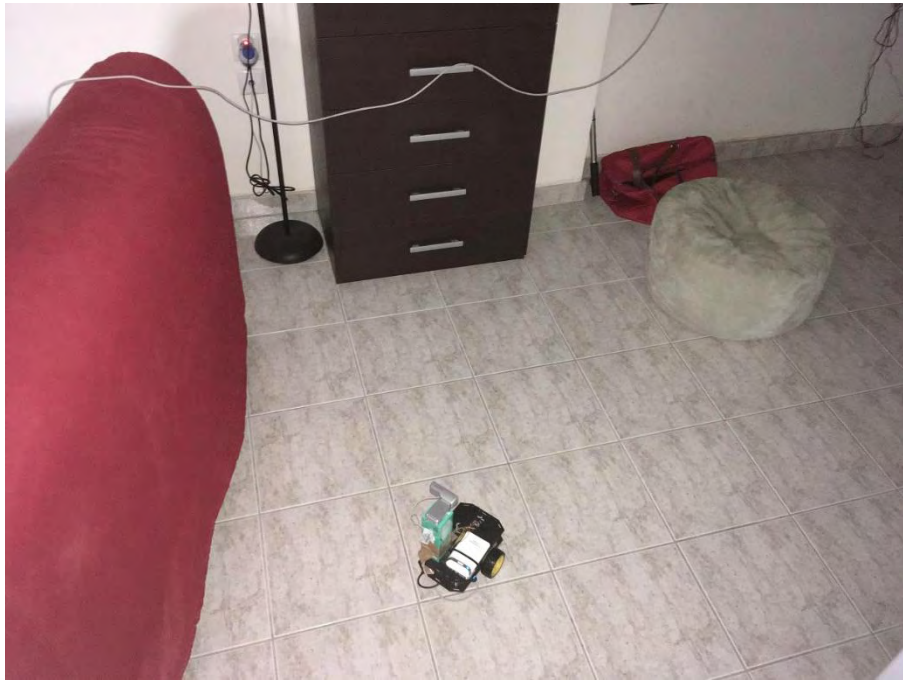
Στη δεύτερη περίπτωση, λόγω κίνησης του ρομπότ μόνο από δύο τροχούς και άρα εξισορρόπησής του από ένα πλαστικό μέρος, εμφανίζεται μεγάλη απόκλιση αν κατά την κίνησή του οχήματος σε κάποιο πιο ανώμαλο έδαφος το ρομπότ «κολλήσει» και δεν ολοκληρώσει την κίνησή του. Έτσι, ενώ οι υπολογισμοί αφορούν ολόκληρη την κίνηση στην πραγματικότητα υπάρχει απόκλιση, γι' αυτό και σε πειράματα με μεγάλο αριθμό κινήσεων και άρα συσσωρευμένων λαθών, η θέση ρομπότ και εμποδίων στην αναπαράσταση είναι λανθασμένη.

Για να δοκιμαστεί η λειτουργία αυτή, δημιουργήθηκαν τρία σενάρια που παρουσιάζονται παρακάτω, καθένα με διαφορετικές συνθήκες:

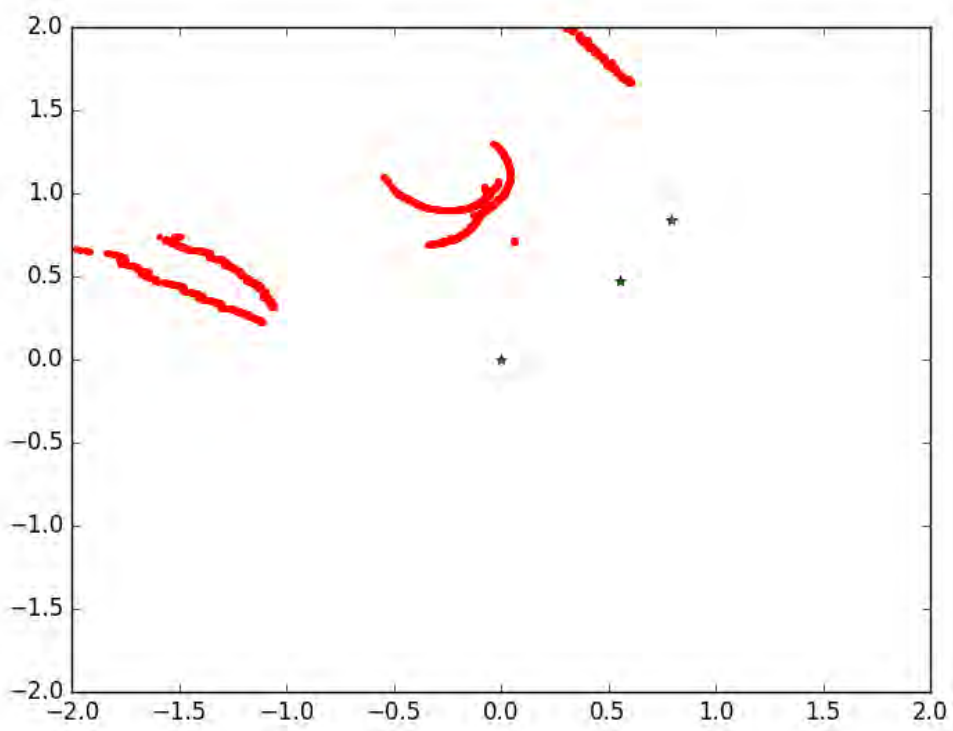
Πείραμα 1^ο: Το ρομπότ βγάζει τρεις φωτογραφίες με κατεύθυνση αριστερά, μπροστά και δεξιά κάνοντας μόνο στροφές και μικρές μετακινήσεις. Ο χώρος και ο χάρτης που παράγεται φαίνονται στο σχήμα 3.8. Παρατηρείται ότι τα βασικά σχήματα των εμποδίων καθώς και οι αποστάσεις τους υπολογίζονται σε ικανοποιητικό βαθμό.

Πείραμα 2^ο: Το ρομπότ βγάζει φωτογραφία ένα αντικείμενο περιστρεφόμενο γύρω του. Στο σχήμα 3.9 παρατηρείται πως εξαιτίας της παρατεταμένης κίνησης το ρομπότ δεν μπορεί υπολογίσει με μεγάλη ακρίβεια την θέση του, με αποτέλεσμα το εμπόδιο να φαίνεται κάπως παραμορφωμένο.

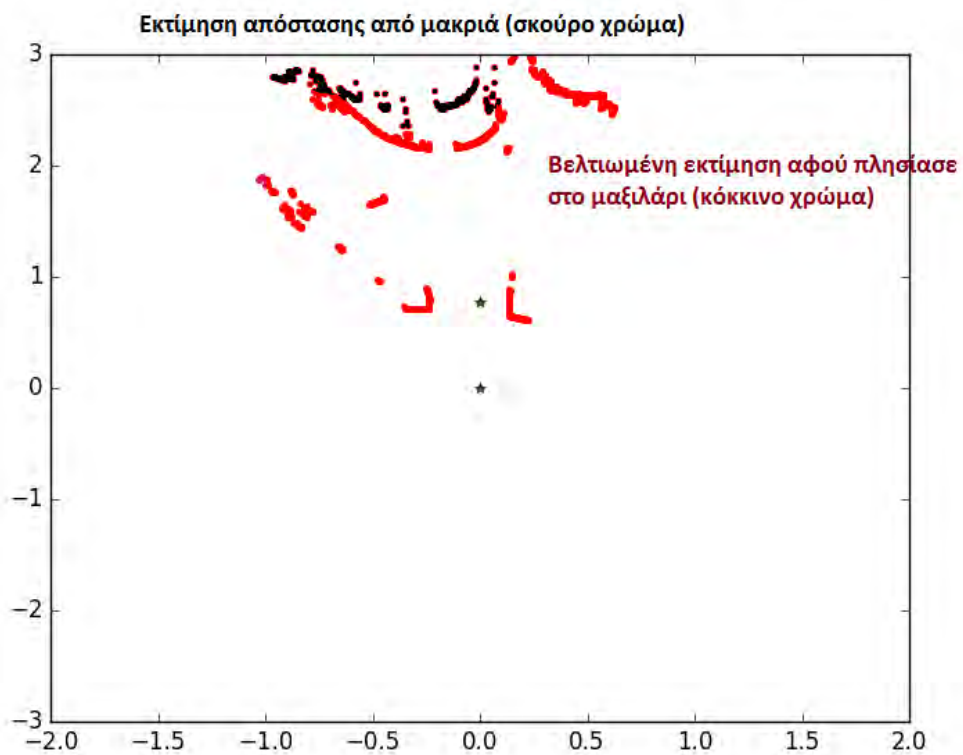
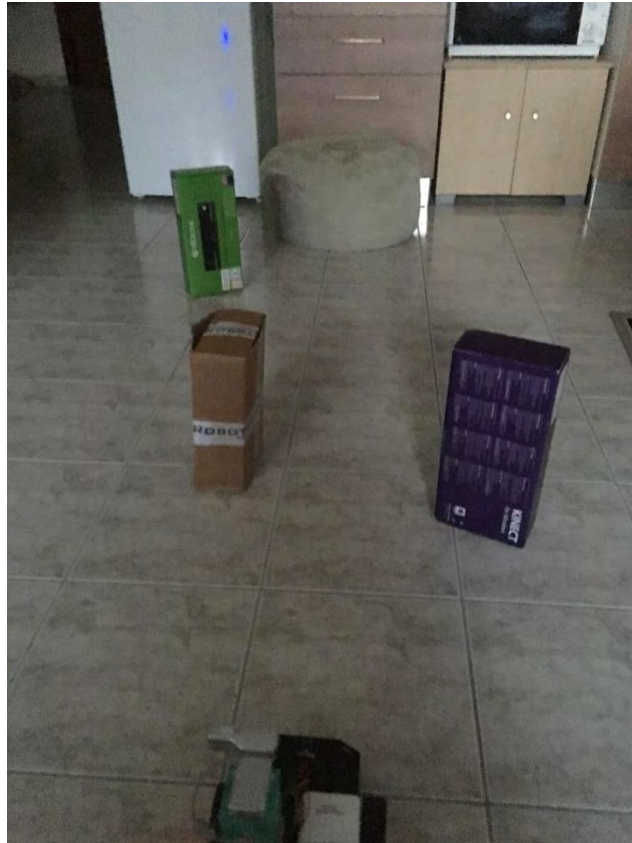
Πείραμα 3^ο: Στην τελευταία περίπτωση, που αντιστοιχεί στο σχήμα 3.10, το ρομπότ προχωράει ευθεία αντιμετωπίζοντας διάφορα εμπόδια. Εδώ φαίνεται ότι όσο το ρομπότ πλησιάζει προς ένα εμπόδιο, τόσο πιο ακριβείς υπολογισμούς κάνει για την απόστασή του. Συγκεκριμένα η θέση του μαξιλαριού αρχικά υπολογίζεται πιο μακριά απ' όσο πραγματικά είναι.



Σχήμα 3.8: Πάνω: Φωτογραφία του τοπίου. Κάτω: Τελικό αποτέλεσμα του αλγορίθμου.



Σχήμα 3.9: Πάνω: Το εμπόδιο γύρω από το οποίο κινήθηκε το ρομπότ. Κάτω: Το αποτέλεσμα του αλγορίθμου.



Σχήμα 3.10: Πάνω: Η σκηνή του πειράματος. Κάτω: Η κάτοψη των εμποδίων σε δύο στάδια.

ΚΕΦΑΛΑΙΟ 4^ο

Λειτουργία με είσοδο χειρονομίες

Η πλατφόρμα IrisTK δίνει τη δυνατότητα στο χρήστη να χρησιμοποιήσει το Kinect v2 sensor στην ανάπτυξη λογισμικού. Ωστόσο, επειδή υπήρξε η ανάγκη δημιουργίας νέων αναγνωρίσιμων χειρονομιών, κάτι που δεν υποστηρίζεται ακόμα από το IrisTK, η ανάπτυξη έγινε σε C++. Η λειτουργία αυτή δε θα αναλυθεί σε βάθος, διότι αποτελεί το κύριο αντικείμενο μελέτης συναδέλφου [14].

4.1 Σύνολο αποδεκτών χειρονομιών

Το Kinect v2 sensor είναι ικανό να ακολουθήσει πληθώρα από κινήσεις και χειρονομίες. Ωστόσο, όσον αφορά τον προγραμματισμό τους, υποστηρίζει εξ αρχής μόνο τρεις, οι οποίες παρουσιάζονται στο σχήμα 4.1.



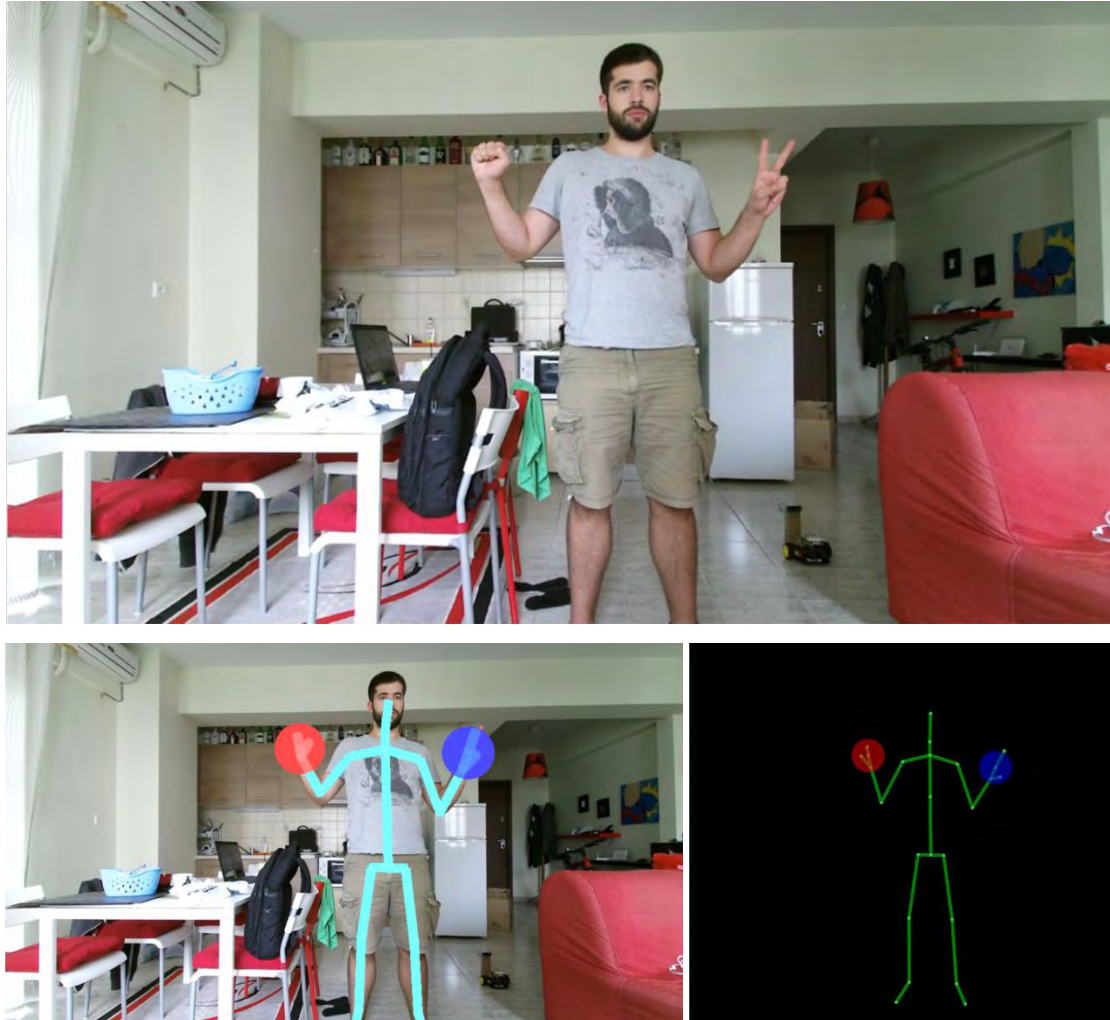
Σχήμα 4.1: Kinect v2 sensor default gestures.

Επομένως, για την κίνηση του ρομπότ, αλλά και τη σωστότερη εννοιολογική τους μετάφραση αναπτύχθηκαν περισσότερες. Το νέο σύνολο αποδεκτών χειρονομιών που μεταφράζονται σε κινήσεις του ρομπότ αποτυπώνεται στο σχήμα 4.2.



Σχήμα 4.2: Πλήρες σύνολο αναγνωρίσιμων χειρονομιών.

Κάθε φορά που αναγνωρίζεται μια χειρονομία, η αντίστοιχη εντολή αποστέλεται προς το Raspberry Pi μέσω network socket και αυτό υλοποιεί την κίνηση όπως και στις φωνητικές εντολές. Παραδείγματα αναγνώρισης χειρονομιών φαίνονται στο σχήμα 4.3.



Σχήμα 4.3: Πάνω: Αρχική εικόνα με δύο χειρονομίες ταυτόχρονα. Κάτω: Αναγνώριση χειρονομιών και εξαγωγή σκελετού.

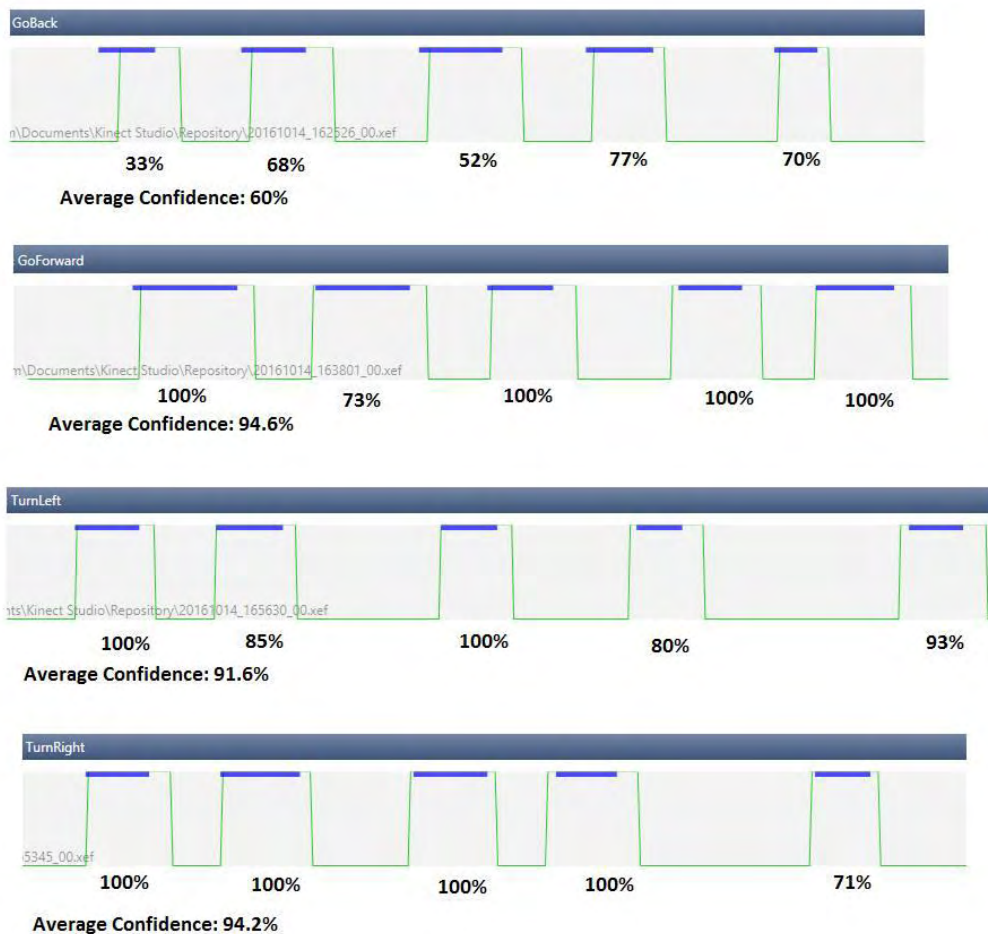
4.2 Λειτουργία με συνδυαστική είσοδο φωνής και χειρονομιών

Το σύστημα μπορεί τέλος να λειτουργήσει και με συνδυαστική είσοδο φωνητικών εντολών και χειρονομιών μέσω Kinect. Αφού το σύστημα δημιουργήθηκε για να δοκιμαστεί σε real time εφαρμογές, θεωρήθηκε σωστότερο να μην κρατώνται οι εντολές που δέχεται το ρομπότ σε ουρά αναμονής, αλλά να εκτελείται κατευθείαν η πιο πρόσφατη. Επομένως, ανεξαρτήτως εντολοδόχου, το ρομπότ θα εκτελέσει την κίνηση που έλαβε πιο πρόσφατα, αγχρηστεύοντας την κίνηση που έκανε μέχρι πρότινος,

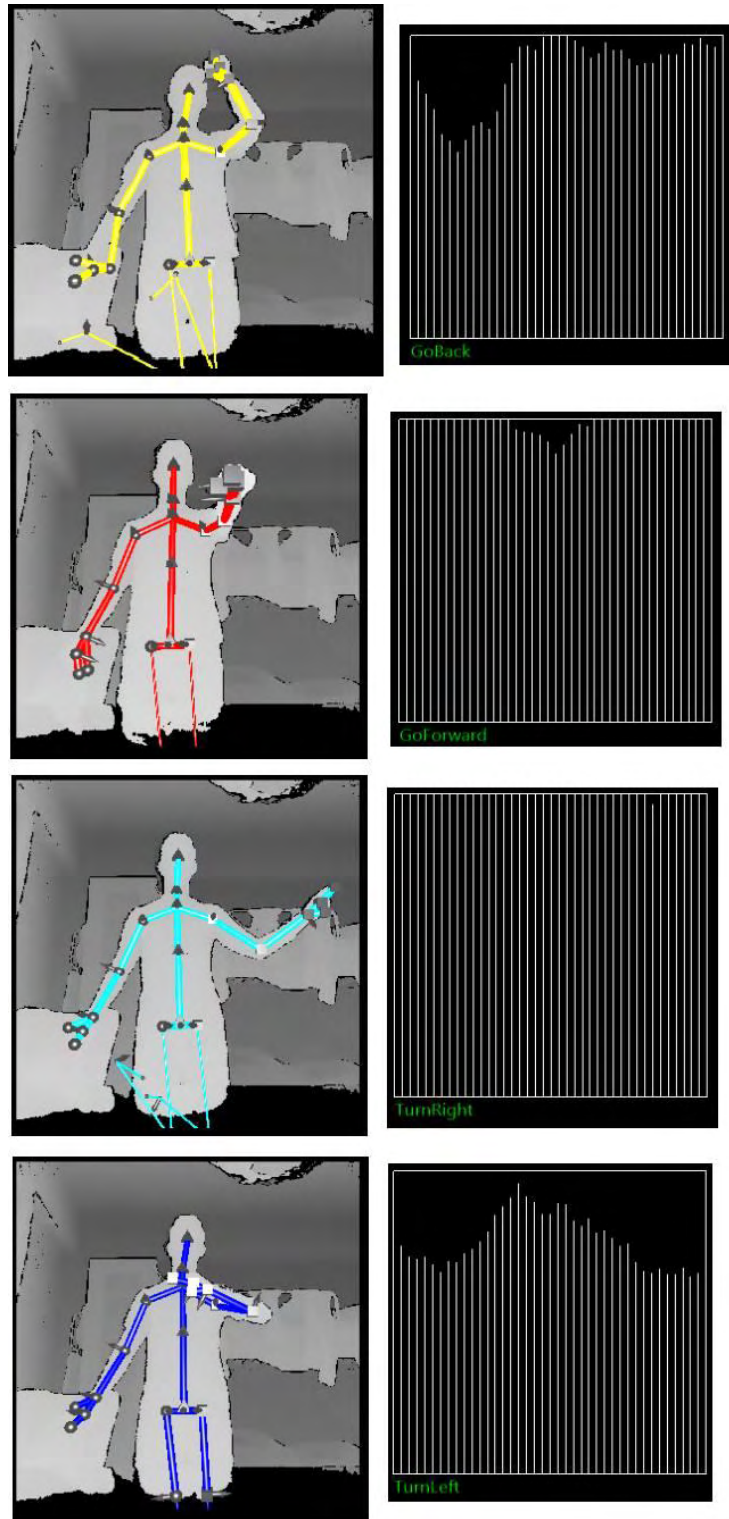
ακόμα και αν αυτή δεν έχει ολοκληρωθεί. Το ρομπότ έχει προγραμματιστεί να μπορεί να λαμβάνει εντολές ως και πέντε χρηστών ταυτόχρονα, αλλά αυτό μπορεί εύκολα να διευρυνθεί περαιτέρω, αν η χρήση του θεωρηθεί σωστή.

4.3 Πειράματα

Το SDK του Kinect v2 sensor δίνει τη δυνατότητα να δημιουργήσει ο χρήστης τις δικές του χειρονομίες. Επιπρόσθετα, παρέχει πληροφορίες σχετικά με τις καινούριες αυτές χειρονομίες. Πληροφορίες όπως το confidence της χειρονομίας, δηλαδή την εμπιστοσύνη που υπάρχει ότι η χειρονομία που δέχθηκε στην είσοδο είναι όντως η σωστή. Με αυτό τον τρόπο έγινε έλεγχος σχετικά με το ποσοστό λάθους που υπάρχει στην αναγνώριση χειρονομιών. Στο σχήμα 4.4 παρουσιάζονται οι χειρονομίες με τα αντίστοιχα confidence που έδωσαν σε πέντε συνεχόμενες προσπάθειες για αναγνώριση και το σχήμα 4.5 τη μεταβολή του confidence ενώ γίνεται η χειρονομία.



Σχήμα 4.4: Confidence και μέσος όρος confidence πέντε διαδοχικών δοκιμών κάθε χειρονομίας.



Σχήμα 4.5: Real time μεταβολή του confidence ενώ δοκιμάζεται κάθε χειρονομία.

Στην παρούσα υλοποίηση δεχόμαστε τις χειρονομίες με ποσοστό confidence μεγαλύτερο του 80%. Για το συγκεκριμένο νούμερο, στις εκατό χειρονομίες κάθε είδους, βρέθηκαν τα αποτελέσματα που παρουσιάζονται στον πίνακα 4.1.

	Forward	Backwards	Left	Right
Ποσοστό χειρονομιών με confidence > 0.8	75%	65%	93%	97%

Πίνακας 4.1: Ποσοστό επι τοις εκατό χειρονομιών με confidence > 0.8.

Συμπεράσματα

5.1 Συνεισφορά της διπλωματικής εργασίας

Στην παρούσα διπλωματική εργασία παρουσιάστηκε ένα σύστημα αλληλεπίδρασης μεταξύ ενός οχήματος-ρομπότ ελεγχόμενου από raspberry pi και ενός ή περισσότερων ανθρώπων. Χρησιμοποιήθηκε η πλατφόρμα IrisTK για τη γρήγορη και εύκολη ανάπτυξη λογισμικού για ανθρώπινη επικοινωνία και ο αισθητήρας Kinect v2 για την αναγνώριση χειρονομιών του χρήστη. Αναλύθηκε η λειτουργία των παραπάνω συστατικών μερών του συστήματος και παρουσιάστηκαν τα αποτελέσματα της προσπάθειας αυτής. Αποτελέσματα που θεωρούνται ικανοποιητικά, καθώς δημιουργήθηκε ένα πλήρες σύστημα απομακρυνσμένου χειρισμού του οχήματος μέσω φωνητικών εντολών και χειρονομιών σε πραγματικό χρόνο, με αμελητέα χρονική απόκλιση και ακρίβεια κίνησης πολύ ικανοποιητική για τα μηχανικά μέρη που χρησιμοποιήθηκαν. Αποδείχθηκε έτσι πως ένα σύστημα επικοινωνίας ανθρώπου μηχανής είναι αρκετά εύκολο να υλοποιηθεί υποστηρίζοντας και συνδυαστικές εισόδους που μπορούν, σε διαφορετικές περιπτώσεις ανθρώπων, να αποδειχθούν πολύτιμα εργαλεία καθημερινής χρήσης και επικοινωνίας.

5.2 Μελλοντική έρευνα

Η παρούσα μελέτη μπορεί να χρησιμοποιηθεί ως παράδειγμα χρήσης της πλατφόρμας IrisTK και του Kinect v2 sensor, καθώς και του συνδυασμού τους με κάποιο μικροελεγκτή. Αν και η λειτουργικότητα του ρομπότ ήταν πολύ περιορισμένη, εντούτις η πραγματοποίηση του project μπορεί να χρησιμοποιηθεί ως βάση πάνω στην οποία μπορεί να χτιστεί οποιοδήποτε σύστημα αλληλεπίδρασης ανθρώπινης φωνής ή κίνησης με μικροελεγκτή. Παραδείγματα μελλοντικής έρευνας αποτελούν εφαρμογές για άτομα με ειδικές ανάγκες, που μπορούν να εκμεταλευτούν κατάλληλα τη συνδυαστική είσοδο για να διευκολυνθούν. Επίσης όσον αφορά το εργαλείο IrisTK, μπορεί να χρησιμοποιηθεί και ως εργαλείο εκμάθησης της ανθρώπινης συμπεριφοράς και επικοινωνίας απο ανθρώπους με αυτισμό ή παρόμοιες παθήσεις. Τέλος, στοιχεία της παρούσας εργασίας μπορούν να ληφθούν και για κάποια ψυχαγωγική πλατφόρμα παιχνιδιών, καθώς οι περισσότερες ήδη υποστηρίζουν το Kinect.

Βιβλιογραφία

- [1] G. Skantze και S. A. Moubayed, «IrisTK: a Statechart-based Toolkit for Multi-Party Face-to-face Interaction,» σε *ICMI*, Santa Monica, CA, 2012.
- [2] «adafruit.com,» [Ηλεκτρονικό]. Available: <https://cdn-shop.adafruit.com/pdfs/raspberrypi2modelb.pdf>. [Πρόσβαση 25 08 2016].
- [3] L. Valgma, *3D reconstruction using Kinect v2 camera*, Bachelor Thesis, Faculty of Science and Technology, University of Tartu, 2016.
- [4] «Microsoft Developer Network,» Microsoft, [Ηλεκτρονικό]. Available: <https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx>. [Πρόσβαση 01 9 2016].
- [5] G. Skantze, «irisTK,» [Ηλεκτρονικό]. Available: http://www.irstk.net/speech_synthesis.html. [Πρόσβαση 25 09 2016].
- [6] G. D. Ryszard Tadeusiewicz, «Speech Man-Machine Communication,» σε *Man-Machine Interactions*, Springer Berlin Heidelberg, 2009, pp. 3-10.
- [7] L. R. Rabiner και B. Juang, *Fundamentals of speech recognition*, Prentice Hall, 1993.
- [8] J. Hernández-Aceituno, R. Arnay, J. Toledo και L. Acosta, «Using Kinect on an Autonomous Vehicle for Outdoors Obstacle Detection,» *IEEE Sensors Journal*, τόμ. 16, αρ. 10, pp. 3603 - 3610, 2016.
- [9] F. Eric, «Speech-Recognition Software Improves,» σε *Gains Recognition*, American Medical News, 1997, p. 19–21.
- [10] C. F. Jan Noyes, «Speech recognition technology for individuals with disabilities,» *Augmentative and Alternative Communication*, τόμ. 8, αρ. 4, pp. 297-303, 1992.
- [11] G. Bell, M. C.Schultz και J. T.Schultz, «Voice Recognition in Fighter Aircraft,» *Journal of Aviation/ Aerospace Education & Research*, τόμ. 10, αρ. 1, pp. 17-28, 2000.
- [12] L. R. Rabiner, «Applications of Speech Recognition in the Area of Telecommunications,» New Jersey, AT&T Labs, 1997, pp. 501-510.
- [13] E. Lachat, H. Macher, M.-A. Mittet, T. Landes και P. Grussenmeyer, «First Experiences with Kinect V2 Sensor for Close Range 3D Modelling,» σε *3D Virtual Reconstruction and Visualization of Complex Architectures*, Avila, Spain, 2015.

- [14] D. Nikolaou, «Gesture based Human Computer Interaction using Kinect v2 Sensor,»
Διπλωματική Εργασία, Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Η/Υ, Πανεπιστήμιο
Θεσσαλίας, 2017.