

Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης



Σχολή Τεχνολογικών Εφαρμογών
Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων



Πτυχιακή Εργασία

Δημιουργία Φυσικής Διεπαφής Χρήστη και ενός

Context Information Manager (CIM), με τη χρήση του Microsoft Kinect

Diploma Thesis

Development of a Natural User Interface (NUI) and creation of a Context

Information Manager, using Microsoft Kinect

Βαρχαλαμάς Πέτρος (ΑΜ: 1978)

Επιβλέπων καθηγητής: Βιδάκης Νίκος

Επιτροπή Αξιολόγησης:

Ημερομηνία παρουσίασης: __/__/2013

Ευχαριστίες

Ένα μεγάλο ευχαριστώ στους γονείς μου που με έχουν στηρίξει με κάθε τρόπο, στον επιβλέποντα καθηγητή μου κ. Νίκο Βιδάκη για την σημαντική καθοδήγησή του, στον κ. Τριανταφυλλίδη, στον Ανέστη Βλασόπουλο και στον Μάκη Συντυχάκη για τη συνεργασία μας πάνω στο project, καθώς και στους υπόλοιπους συνεργάτες - συμφοιτητές και φίλους από το εργαστήριο ISTLab για οποιαδήποτε βοήθεια χρειάστηκα.

Abstract

This thesis presents a multimodal Natural User Interface (NUI), meaning a system for human-computer interaction that the user operates through intuitive actions related to natural, everyday human behavior.

The main purpose of this project focuses on facilitating this interaction by keeping the human actor in mind.

In addition, this work contributes to the improvement of the project MIDAS (Multimodal Interface Directed by Action Sentences) using Microsoft Kinect multi-sensor which provides RGB, depth and sound input, through all the advantages of the three basic Kinect's SDK functions, Speech-Recognition – Gesture Recognition – Skeleton Tracking.

Keywords:

Multimodal Interaction, Kinect, Multi-application, Action Sentences, Action Context Information Manager, MIDAS

Σύνοψη

Σε αυτή την πτυχιακή εργασία παρουσιάζεται μια πολυτροπική φυσική διεπαφή χρήστη, δηλαδή ένα σύστημα αλληλεπίδρασης ανθρώπου-υπολογιστή, όπου ο χρήστης λειτουργεί μέσω φυσικών κινήσεων που σχετίζονται με την καθημερινή ανθρώπινη συμπεριφορά.

Ο κύριος σκοπός αυτού του έργου εστιάζει στη διευκόλυνση αυτής της αλληλεπίδρασης, διατηρώντας τον ανθρώπινο παράγοντα κατά νου.

Επιπροσθέτως, σε αυτή την εργασία γίνεται η βελτίωση του ήδη υπάρχοντος προγράμματος MIDAS (Multimodal Interface Directed by Action Sentences)**Error! Reference source not found.**, χρησιμοποιώντας το Microsoft Kinect πολυαισθητήρα που παρέχει σαν είσοδο δεδομένων εικόνα από έγχρωμη κάμερα, κάμερα βάθους και ήχο, λαμβάνοντας όλα τα πλεονεκτήματα των τριών βασικών λειτουργιών του Kinect SDK, που είναι : αναγνώριση ομιλίας, εντοπισμός σκελετού και αναγνώριση χειρονομίας.

Περιεχόμενα

Περιεχόμενα	5
1 Εισαγωγή.....	11
1.1 Περίληψη	11
1.2 Δομή Εργασίας	12
2 Συσκευές Εισόδου.....	13
2.1 Εισαγωγή	13
2.2 Κλασικές συσκευές εισόδου.....	13
2.3 Συσκευές Πολυτροπικής Εισόδου – Multimodal Devices	13
2.4 Microsoft – Kinect.....	14
2.4.1 Σύγκριση Kinect for Windows και Kinect for XBOX 360	15
2.4.2 Αισθητήρας Βάθους (Depth Sensor) και Πομπός Υπέρυθρων (Infrared light) .	16
2.4.3 Κάμερα χρώματος (Color Camera)	17
2.4.4 Μικρόφωνα	18
2.4.5 Επιταχυνσιόμετρο (Accelerometer)	18
2.4.6 Led Light	19
2.4.7 Μηχανική Βάση (Motorized Tilt)	22
2.4.8 Τα διαθέσιμα APIs του Kinect	23
2.5 Asus - Xtion Pro και Xtion Pro LIVE	27
2.5.1 Ιστορικό εξέλιξης	27
2.5.2 Τα χαρακτηριστικά του Xtion Pro Live	28
2.6 Leap Motion – The Leap	29
2.6.1 Τα χαρακτηριστικά του Leap	30
2.6.2 Χρήση από το κοινό	31
2.6.3 Συνεργασίες της Leap Motion και μελλοντικά σχέδια.....	31

3	Αλληλεπίδραση Ανθρώπου με Υπολογιστή – Human Computer Interaction (HCI)	32
3.1	Εισαγωγή	32
3.2	Διεπαφές Χρήστη – User Interfaces	33
3.2.1	Είδη διεπαφής χρήστη	34
3.3	Φυσική Διεπαφή Χρήστη – Natural User Interface.....	39
3.3.1	Η φυσική διεπαφή στην επιστήμη των υπολογιστών.....	39
3.4	Πολυτροπικές Διεπαφές Χρήστη - Multimodal User Interfaces	40
4	Τεχνολογία και Εργαλεία Υλοποίησης.....	41
4.1	.NET Framework	41
4.2	Microsoft Visual C #	42
4.3	Microsoft Visual Studio 2010.....	42
4.4	Extensible Markup Language - XML.....	43
4.5	MySQL.....	44
4.6	Microsoft SDK for Kinect	45
5	Αρχιτεκτονική Συστήματος	49
5.1	Συσκευές και Δεδομένα Εισόδου	49
5.1.1	Αναγνώριση Ομιλίας - Speech Recognition	50
5.1.2	Αναγνώριση Χειρονομιών - Gesture Recognition	50
5.2	Αναγνώριση Προσώπου και Αντικειμένου - Face and Object Recognition	51
5.3	Sentence Compiler.....	51
5.4	Action Sentence Manipulator	52
5.5	Context Information Manager	52
6	Υλοποίηση	54
6.1	Εισαγωγή	54
6.2	Η Βάση δεδομένων του συστήματος.....	55
6.3	Gesture Recognition	57

6.4	Speech recognition	62
6.5	Περιγραφή Σεναρίου Χρήσης.....	63
6.5.1	Περιγραφή User Interface	67
6.5.2	Διάδραση με τις εφαρμογές.....	69
6.5.3	Customize Sentence	72
7	Συμπεράσματα και Μελλοντική Εργασία.....	75
	Βιβλιογραφία.....	76
	Αναφορά στις πηγές των εικόνων	78

Εικόνες

Εικόνα 1 – Classic peripherals	13
Εικόνα 2 - Microsoft Kinect για το Xbox 360	14
Εικόνα 3 - Microsoft Kinect για τα Windows.....	14
Εικόνα 4 – Το Kinect Fusion σε δράση	15
Εικόνα 5 – Το εσωτερικό της συσκευής Kinect	16
Εικόνα 6 - Η δομή υπέρυθρων από τον πομπό του Kinect	17
Εικόνα 7 - Εικόνα βάθους από το Kinect.....	17
Εικόνα 8 - RGB εικόνα από το Kinect.....	17
Εικόνα 9 - Η σειρά μικροφώνων μέσα στη συσκευή.....	18
Εικόνα 10 – The Kinect Accelerometer coordinate system	18
Εικόνα 11 - Το Επιταχυνσιόμετρο του Kinect.....	19
Εικόνα 12 - Κόκκινο, πράσινο και μπλε LED των 5mm	20
Εικόνα 13 – Ενυδρείο με κοράλλια υπό φωτισμό LED	21
Εικόνα 14 – Kinect’s LED light.....	22
Εικόνα 15 - Η μηχανοκίνητη βάση του Kinect	23
Εικόνα 16 - Calibration pose for OpenNI	25
Εικόνα 17 - Χάρτης βάθους και RGB κάμερα με τη χρήση του Libfreenect	25
Εικόνα 18 - Παράδειγμα χρήσης επιταχυνσιόμετρου με χρήση του CL NUI.	26
Εικόνα 19 - Finger gesture recognition με τη χρήση του Evoluce SDK.....	26
Εικόνα 20 – Asus Wavi Xtion Pro	27
Εικόνα 21 - Asus Xtion Pro LIVE	27
Εικόνα 22 - Το εσωτερικό της συσκευής Xtion Pro LIVE	28
Εικόνα 23 - The Leap.....	29
Εικόνα 24 – Leap’s smaller observation area and high resolution.....	30

Εικόνα 25 – Ένας από τους λόγους που προκλήθηκε το ατύχημα του Three Mile Island ήταν η ανεπαρκής εκπαίδευση του προσωπικού	34
Εικόνα 26 – Command Line Interface example.....	34
Εικόνα 27 – A Web-based User Interface example	35
Εικόνα 28 – Touch screen interface example	36
Εικόνα 29 – Graphical User Interface using WIMP style example	36
Εικόνα 30 – Zooming interface example	37
Εικόνα 31 – Brain – Computer Interface example.....	38
Εικόνα 32 – Η εξέλιξη των User Interfaces	39
Εικόνα 33 – Απεικόνιση της μετατροπής του Source code σε Bytecode και εν συνεχεία με το CLR σε Native code	42
Εικόνα 34 – Joins detected from Microsoft Skeleton Tracking.....	45
Εικόνα 35 – Face Detection	46
Εικόνα 36 – Joints tracked in seated mode	46
Εικόνα 37 – Απεικόνιση αρχιτεκτονικής MIDA.....	49
Εικόνα 38 – mydb_EER_Diagram.....	55
Εικόνα 39 – No sensor attached	64
Εικόνα 40 – Κύριο παράθυρο εφαρμογής.....	64
Εικόνα 41 - Τυχαίες περιπτώσεις αναγνώρισης χρήστη και αντικειμένου	65
Εικόνα 42 - Κεντρική σελίδα της εφαρμογής	67
Εικόνα 43 - Δημιουργία ολοκληρωμένης πρότασης με συνδυασμό speech & gesture recognition	69
Εικόνα 44 - Add Sentence procedure	72
Εικόνα 45 - Add your own command	73
Εικόνα 46 - Delete Sentence procedure	74
Εικόνα 47 - Update Sentence procedure	74

Πίνακες και Κώδικας

Πίνακας 1 - Χαρακτηριστικά Asus Xtion Pro Live vs. Microsoft Kinect	29
Πίνακας 2 -Μη στατικές και στατικές χειρονομίες που αναγνωρίζονται από το σύστημα	58
Πίνακας 3 - Συνθήκη μέτρησης των frames και πράξεις για την απόσταση	58
Πίνακας 4 - Αναγνώριση του Swing Gesture και της κατεύθυνσης του	59
Πίνακας 5 - Αλγόριθμος υπολογισμού της γωνίας του αγκώνα κάθε χεριού	60
Πίνακας 6 - Υπολογισμός στατικής χειρονομίας Hands Up	61
Πίνακας 7 - – Ενημέρωση Event Handler κάθε φορά που αναγνωρίζεται μια λέξη	62
Πίνακας 8 - Παράδειγμα γραμματικής του Speech Recognition	63
Πίνακας 9 - Έλεγχος βάσης για τον τύπο της πρότασης.....	70

1 Εισαγωγή

1.1 Περίληψη

Με την πάροδο των χρόνων και την εξέλιξη της τεχνολογίας, η σχέση μεταξύ ανθρώπου και υπολογιστή γίνεται ολοένα και πιο περίπλοκη. Η μεταξύ τους αλληλεπίδραση μελετάται από το επιστημονικό πεδίο του **Human Computer Interaction (HCI)** βλ. κεφάλαιο 3 και γίνεται με γνώμονα και τις δυο αυτές παραμέτρους.

Σ' ό,τι αφορά τη μηχανή, η αλληλεπίδραση σχετίζεται με την επιστήμη στα γραφικά, τα λειτουργικά συστήματα, τις γλώσσες προγραμματισμού, το εσωτερικό υλικό των συσκευών. Από την άλλη πλευρά, αυτά που χαρακτηρίζουν τον άνθρωπο - χρήστη, είναι η επικοινωνία, η γλωσσολογία, οι κοινωνικές επιστήμες η ψυχολογία και ο ανθρώπινος παράγοντας. Έτσι λοιπόν, λόγω του διεπιστημονικού χαρακτήρα του HCI, άνθρωποι με διαφορετικό αντικείμενο γνώσεων μπορούν να συμβάλουν στην επιτυχία του.

Η αλληλεπίδραση μεταξύ χρηστών και υπολογιστών γίνεται στο επίπεδο της φυσικής διεπαφής χρήστη (**Natural User Interface - NUI**) (βλ. κεφάλαιο 3.3), όπου παρουσιάζεται συνήθως σαν μια διεπαφή που είναι εύκολη στη χρήση και με το σκεπτικό να γίνει αόρατη με την επιτυχή μάθηση της.

Αυτό επιτυγχάνεται με τη δυνατότητα του χρήστη να χρησιμοποιεί φυσικές κινήσεις που σχετίζονται με την καθημερινότητά του για την αλληλεπίδρασή του με τον υπολογιστή. Για να γίνει όμως αυτό εφικτό, χρειάζεται την κατάλληλη τεχνολογία η οποία θα πρέπει παράλληλα να είναι και οικονομικά προσιτή για το μέσο χρήστη. Έτσι λοιπόν για τον εντοπισμό της κίνησης του ανθρώπινου σώματος χρειαζόμαστε εξοπλισμό, πιο εξειδικευμένο και κατά συνέπεια πιο ακριβό, όπως τη χρήση ψηφιακών αισθητήρων που να παρέχουν μεγάλη ακρίβεια και αποτελεσματικότητα στην αντίληψη των μηχανών – οπότε και μη εύκολα εμπορικά προσιτό προς το χρήστη.

Πρόσφατα όμως κυκλοφόρησαν προσιτές σε τιμή συσκευές, από πολύ μεγάλες εταιρείες στο χώρο της πληροφορικής όπως το *Kinect* της Microsoft, το *Xtion Pro LIVE* της Assus, και το *The Leap* της Leap Motion, οι οποίες με την ιδιαίτερη αρχιτεκτονική τους, τον συνδυασμό αισθητήρων και συσκευών εισόδου όπως RGB κάμερας, μικροφώνων και του πιο σημαντικού, την κάμερα βάθους παρέχουν ένα αξιόλογο εντοπισμό κίνησης (**Motion Tracking**), δίνοντας τη

δυνατότητα και την ελευθερία στους προγραμματιστές να αναπτύξουν περισσότερο ανθρωποκεντρικές εφαρμογές, και παράλληλα μια νέα πνοή στα όσα μέχρι τώρα γνωρίζουμε για την αλληλεπίδραση ανθρώπου μηχανής.

1.2 Δομή Εργασίας

Η ανάλυση της πτυχιακής γίνεται στα παρακάτω κεφάλαια στα οποία αναπτύσσονται :

1. Περιγραφή και παρουσίαση των συσκευών εισόδου/εξόδου σε αντιπαράθεση με τις εξελεγμένες συσκευές πολυτροπικής εισόδου όπως το Kinect, το XtionPro και το Leap.
2. Η σχέση μεταξύ ανθρώπου και υπολογιστή η οποία περιγράφεται από το επιστημονικό πεδίο του Human Computer Interaction (HCI).
3. Η τεχνολογία και τα εργαλεία που χρησιμοποιήθηκαν για την υλοποίηση της εφαρμογής.
4. Η αρχιτεκτονική του συστήματος με περιγραφή για κάθε κομμάτι της.
5. Η περιγραφή της εφαρμογής με λεπτομέρεια για κάθε πιθανή επιλογή του χρήστη και των δυνατοτήτων που του δίνονται.

2 Συσκευές Εισόδου

2.1 Εισαγωγή

Στην επιστήμη των υπολογιστών, μια συσκευή εισόδου είναι οποιαδήποτε περιφερειακό (κομμάτι του εξοπλισμού ηλεκτρονικών υπολογιστών) που χρησιμοποιείται για να παρέχει δεδομένα και σήματα ελέγχου σε ένα σύστημα επεξεργασίας πληροφοριών, όπως είναι ο υπολογιστής.

2.2 Κλασικές συσκευές εισόδου

Υπάρχουν πάρα πολλές συσκευές εισόδου, αλλά αυτά που έχουν ευρεία χρήση στην καθημερινότητα είναι το πληκτρολόγιο (keyboard), το ποντίκι (mouse), ο σαρωτής (scanner), το μικρόφωνο, η RGB κάμερα.

Κάθε συσκευή δίνει το δικό της σήμα εισόδου στο σύστημα που έχει σαν αποτέλεσμα να εκτελούνται κάποιες εντολές όπως για παράδειγμα :

- με το keyboard πληκτρολόγηση κειμένου
- με το mouse να γίνεται περιήγηση του κέρσορα σε κάποιο γραφικό περιβάλλον, drag and drop κάποιου αρχείου
- με το scanner σάρωση κάποιου εγγράφου
- με την κάμερα λήψη εικόνας
- με το μικρόφωνο λήψη ήχου



Εικόνα 1 – Classic peripherals

2.3 Συσκευές Πολυτροπικής Εισόδου – Multimodal Devices

Συσκευές εισόδου, που έχουν κουμπιά όπως χειριστήρια, ποντίκια και πληκτρολόγια, μπορούν να συνδυαστούν σε μια ενιαία φυσική συσκευή που θα μπορούσε να θεωρηθεί ως μια σύνθετη συσκευή. Πολλές συσκευές παιχνιδιών έχουν τέτοια.

Τεχνικά, ενώ τα ποντίκια είναι σύνθετες συσκευές, καθώς παρακολουθείται η κίνηση του δείκτη στο γραφικό περιβάλλον και παρέχουν τα κουμπιά για να κάνουν κλικ, τα σύνθετα

συστήματα, θεωρείται γενικά, ότι έχουν περισσότερες από δύο διαφορετικές μορφές της εισόδου. Συγκεκριμένα οι συσκευές που θα μελετηθούν στις παρακάτω ενότητες συνδυάζουν γνωρίσματα συσκευών εισόδου όπως RGB κάμερας, μικροφώνου, και του πιο σημαντικού, της κάμερας βάθους.

2.4 Microsoft – Kinect



Εικόνα 2 - Microsoft Kinect για το Xbox 360



Εικόνα 3 - Microsoft Kinect για τα Windows

Την 1 Ιουνίου του 2009, η Microsoft ανακοίνωσε ένα project με όνομα Project Natal. Το τελικό προϊόν του project, το οποίο έκανε την εμφάνισή του στην αγορά τον Νοέμβριο του 2010, ονομάστηκε Kinect **Error! Reference source not found.** από τις λέξεις kinetic (κινητικός) και connect (συνδέω), το οποίο είναι μια συσκευή εισόδου ανίχνευσης κίνησης από τη Microsoft για την βίντεο κονσόλα Xbox 360 και αλλά και για ανάπτυξη εφαρμογών σε Windows PC. Ο αισθητήρας βάθους, που σχεδιάστηκε και δημιουργήθηκε από την ισραηλινή εταιρία PrimeSense, Ltd [3].

Από εκείνη την περίοδο και μετά έφερε την επανάσταση όχι μόνο στον τομέα των βιντεοπαιχνιδιών, αλλά και στην ρομποτική, την ιατρική και σε πολλούς άλλους κλάδους που επωφελούνται από τις δυνατότητες αυτής της συσκευής.

Επίσης γεγονός είναι ότι η αναλογία απόδοσης – τιμής είναι αρκετά μεγάλη. Ένα παράδειγμα ενός ακριβού αισθητήρα είναι ο SwissRamger 4000, ο οποίος είχε κόστος περίπου \$10000 ενώ το Kinect αυτή τη στιγμή κοστολογείται λιγότερο από \$100, παρά το γεγονός ότι η απόδοσή του ως αισθητήρα είναι κατά πολύ ανώτερη σε σχέση με αυτή του SwissRamger 4000.

Είναι λοιπόν λογικό δύο μήνες μετά από την κυκλοφορία του, η Microsoft, να είχε πουλήσει 8 εκατομμύρια Kinect δίνοντάς της τον τίτλο της ηλεκτρονικής συσκευής με την ταχύτερη πώληση στο βιβλίο Guinness.

2.4.1 Σύγκριση Kinect for Windows και Kinect for XBOX 360

Οι δύο συσκευές μπορεί να μοιάζουν εξωτερικά, όμως έχουν κάποιες λίγες αλλά σημαντικές διαφορές, οι οποίες είναι οι εξής :

1. *Near - mode tracking*

Ενεργοποιεί την κάμερα ώστε να βλέπει αντικείμενα σε απόσταση 40 εκατοστών από τη συσκευή χωρίς να χάνεται η ακρίβεια της πληροφορίας.

2. *USB καλώδιο*

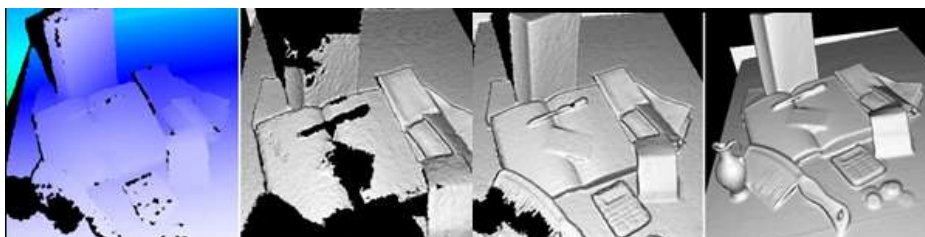
Διασφαλίζει την αξιοπιστία σε ένα ευρύ φάσμα των ηλεκτρονικών υπολογιστών και βελτιώνει τη συνύπαρξη με άλλα USB περιφερειακά.

3. *Εκτεταμένες ρυθμίσεις της κάμερας*

Παρέχει επιπλέον ρυθμίσεις όπως για τη φωτεινότητα, ώστε να μπορεί να γίνει καλύτερος συντονισμός.

4. *Kinect Fusion*

Με το Kinect Fusion, ο χρήστης παίρνοντας την εικόνα βάθους από την κάμερα του Kinect με πολλά ελλειπή στοιχεία, μέσα σε λίγα δευτερόλεπτα μπορεί να παράξει μια ρεαλιστική 3D ανακατασκευή μιας στατικής σκηνής, κουνώντας τον αισθητήρα τριγύρω. Αυτός είναι ένας τρόπος δημιουργίας point cloud ή 3D πλέγματος.



Εικόνα 4 – Το Kinect Fusion σε δράση

5. Τιμή προϊόντος

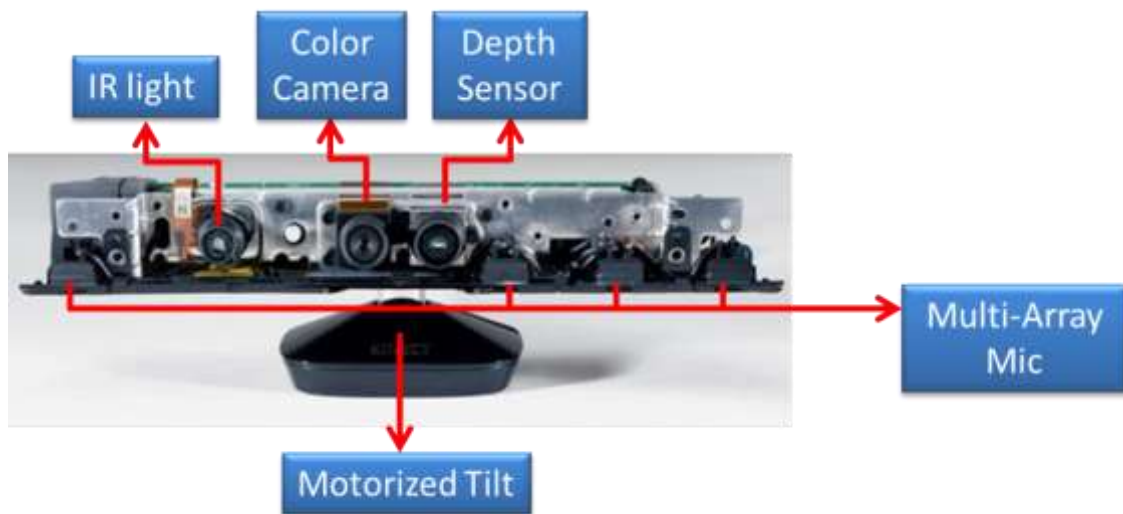
- Το Kinect for Windows κοστίζει κοντά στα 190 ευρώ.

- Το Kinect for XBOX 360 είναι πολύ πιο φτηνό, δηλαδή περίπου 50 ευρώ, για τη χρήση του όμως σε PC η αγορά ενός αντάπτορα ρεύματος είναι αναγκαία.

6. Licensing

Σύμφωνα με την ορθότητα της χρήσης της συσκευής που ορίζει η Microsoft, όταν γίνεται δημόσια παρουσίαση κάποιας εφαρμογής, θα πρέπει να χρησιμοποιείται το Kinect for Windows και όχι το Kinect for Xbox 360, διότι θα θεωρείται παράνομο.

Αυτός είναι και ο λόγος της μεγάλης διαφοράς κοστολόγησης των δύο συσκευών, δηλαδή η απόκτηση δικαιωμάτων.



Εικόνα 5 – Το εσωτερικό της συσκευής Kinect

2.4.2 Αισθητήρας Βάθους (Depth Sensor) και Πομπός Υπέρυθρων (Infrared light)

Η τεχνολογία Light Coding™ (ευρεσιτεχνία των Zalevsky, Z, et al.[4]) επιτρέπει στο Kinect να δημιουργήσει 3D χάρτες βάθους μιας σκηνής ή διαδικασία τριγωνισμού [5], σε πραγματικό χρόνο. Η πηγή λέιζερ εκπέμπει μια απλή δέσμη η οποία χωρίζεται σε πολλαπλές δέσμες από ένα φράγμα περίθλασης, για τη δημιουργία ενός σταθερού μοτίβου στιγμάτων τα οποία προβάλλονται πάνω στις επιφάνειες (βλ.Εικόνα 6) και ένας αισθητήρας εικόνας CMOS (Complementary Metal Oxide Semiconductor) δέχεται τις ανακλώμενες ακτίνες. Το PS1080 SoC (System on a Chip) – τσιπ φτιαγμένο από την PrimeSense που περιέχει το σύστημα του Kinect, ελέγχει την δομή από σημεία φωτός και να επεξεργάζεται τα δεδομένα από τον

αισθητήρα CMOS παράγοντας δεδομένα βάθους σε πραγματικό χρόνο [6]. Η μέγιστη ανάλυση της εικόνας βάθους που παράγει το (βλ. Εικόνα 7) PS1080 είναι 640x480, με συχνότητα 30 frames ανά δευτερόλεπτο. Στα 2 μέτρα απόστασης από τον αισθητήρα, έχει τη ακρίβεια 3 χιλιοστών σε ύψος και πλάτος και 1 εκατοστό σε βάθος. Η εμβέλεια ορθής λειτουργίας είναι από 0.8 μέχρι 3.5 μέτρα.



Εικόνα 6 - Η δομή υπέρυθρων από τον πομπό του Kinect



Εικόνα 7 - Εικόνα βάθους από το Kinect

2.4.3 Κάμερα χρώματος (Color Camera)

Το Kinect έχει επίσης μια ενσωματωμένη κάμερα χρώματος (Color CMOS - VNA38209015) με μέγιστη ανάλυση 1280x1024 για να έχουμε την πραγματική εικόνα πέρα από το χάρτη βάθους. Η συχνότητα λήψης της κάμερας είναι 30 fps και η εικόνα που παράγεται είναι αρκετά καλή, ώστε να χρησιμοποιηθεί σε αλγορίθμους αναγνώρισης προσώπου, δακτύλων ή οτιδήποτε άλλο χρειαζόμαστε στην εφαρμογή μας. Παράδειγμα εικόνας από την RGB κάμερα της συσκευής φαίνεται στην **Error! Reference source not found.**



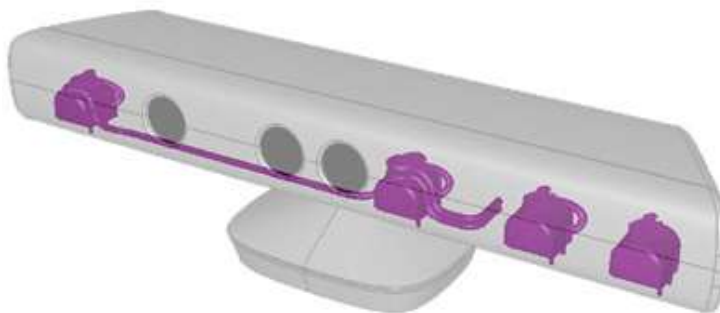
Εικόνα 8 - RGB εικόνα από το Kinect

2.4.4 Μικρόφωνα

Μια σειρά από 4 μικρόφωνα δίνει στο Kinect τη δυνατότητα όχι απλά να δέχεται ήχο, αλλά και να εντοπίζει την γωνία της πηγής του στη σκηνή, μετρώντας το ηχητικό πεδίο.

Η περιγραφή του ηχητικού πεδίου μπορεί να γίνει χρησιμοποιώντας φυσικές ποσότητες, όπως της ηχητικής πίεσης και της ταχύτητας των σωματιδίων. Με τη μέτρηση αυτών των ιδιοτήτων, είναι (έμμεσα) δυνατόν να ληφθεί μια κατεύθυνση πηγής του ήχου.

Στην Εικόνα 9 βλέπουμε τη θέση των μικροφώνων μέσα στη συσκευή η οποία επεξεργάζεται ξεχωριστά το καθένα από τα τέσσερα κανάλια τα οποία δέχονται 16-bit ήχο με συχνότητα δειγματοληψίας ίση με 16 kHz.

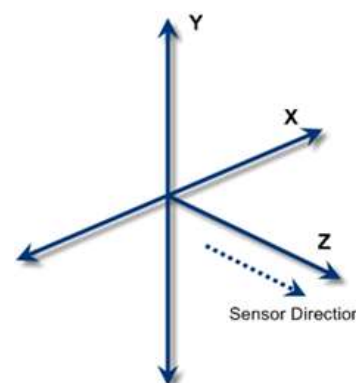


Εικόνα 9 - Η σειρά μικροφώνων μέσα στη συσκευή.

2.4.5 Επιταχυνσιόμετρο (Accelerometer)

Το επιταχυνσιόμετρο είναι μια συσκευή, η οποία όπως γίνεται αντιληπτό και από την ίδια τη λέξη, μετρά την επιτάχυνση. Η επιτάχυνση δεν είναι απαραίτητως επιτάχυνση συντεταγμένων (ρυθμός μεταβολής της ταχύτητας). Αντ' αυτού, το επιταχυνσιόμετρο βλέπει την επιτάχυνση που συνδέεται με το βάρος που έχει οποιοδήποτε αντικείμενο.

Τα επιταχυνσιόμετρα έχουν πολλαπλές εφαρμογές στη βιομηχανία και την επιστήμη. Πολύ ευαίσθητα επιταχυνσιόμετρα είναι συστατικά των αδρανειακών συστημάτων πλοήγησης για αεροσκάφη και πυραύλους. Επιταχυνσιόμετρα χρησιμοποιούνται



Εικόνα 10 – The Kinect Accelerometer coordinate system

επίσης σε υπολογιστές tablet, smartphones και ψηφιακές φωτογραφικές μηχανές, έτσι ώστε οι εικόνες στις οθόνες να εμφανίζονται πάντα σε όρθια θέση.

Το Kinect παρέχει ένα επιταχυνσιόμετρο τριών αξόνων (KXSD9-1026) το οποίο παρέχει την πληροφορία της θέσης της συσκευής σε σχέση με τη βαρύτητα, της οποίας το εύρος είναι μέχρι $2g$ – όπου g είναι η επιτάχυνση λόγω της βαρύτητας.

Τα δεδομένα που επιστρέφει το επιταχυνσιόμετρο είναι ένα 3D διάνυσμα που δείχνει στην κατεύθυνση της βαρύτητας, το οποίο είναι της μορφής Vector4 (x, y, z, w) με την τιμή του w να είναι 0.0. Επίσης το επιταχυνσιόμετρο είναι κεντραρισμένο στον αισθητήρα και είναι ένα σύστημα συντεταγμένων δεξιού χεριού με τη θετική τιμή του z άξονα να δείχνει την κατεύθυνση που κοιτάει ο αισθητήρας (βλ. Εικόνα 10). Η προκαθορισμένη τιμή του διανύσματος έχει την τιμή $(0, -1.0, 0, 0)$.



Εικόνα 11 - Το Επιταχυνσιόμετρο του Kinect.

Έτσι λοιπόν χάρη στα δεδομένα του αισθητήρα μπορεί να εντοπιστεί ένας “ασυνήθιστος” προσανατολισμός της συσκευής αλλά και με τη βοήθεια του SDK να πραγματοποιηθούν σενάρια επαυξημένης πραγματικότητας. Αυτό το χαρακτηριστικό του Kinect έχει χρησιμοποιηθεί ευρέως στον τομέα της ρομποτικής.

2.4.6 Led Light

Μια δίοδος εκπομπής φωτός (Light-Emitting Diode - LED) αποκαλείται ένας ημιαγωγός ο οποίος εκπέμπει φωτεινή ακτινοβολία στενού φάσματος όταν του παρέχεται μία ηλεκτρική τάση κατά τη φορά ορθής πόλωσης (forward-based). Εμφανίζονται ως πρακτικό ηλεκτρονικό στοιχείο, το 1962.



Το χρώμα του φωτός που εκπέμπεται εξαρτάται από την χημική σύσταση του ημιαγωγικού υλικού που χρησιμοποιείται, και μπορεί να είναι υπεριώδες, ορατό ή υπέρυθρο. Το μήκος κύματος του φωτός που εκπέμπεται, και, κατά συνέπεια, το χρώμα του, εξαρτάται από το χάσμα

ενέργειας των υλικών, τα οποία χρησιμοποιούνται για την δημιουργία του περάσματος p-n, όπου:

- Εικόνα 12 - Κόκκινο, πράσινο και μπλε LED των 5mm**
- p = Υλικό νοθευμένο με αποδέκτες
 - n = Υλικό νοθευμένο με δότες

Η βασική αρχή των LED είναι μια επαφή p - n η οποία πολώνεται ορθά για να εγγείη ηλεκτρόνια και οπές μέσα στις p- και n- πλευρές αντίστοιχα. Το εγγεόμενο φορτίο μειονότητας επανασυνδέεται με το φορτίο πλειονότητας στην περιοχή απογύμνωσης ή στην ουδέτερη περιοχή. Σε ημιαγωγούς άμεσου διάκενου η επανασύνδεση οδηγεί σε εκπομπή φωτός αφού η ακτινοβολία επανασύνδεσης, κυριαρχεί σε υλικά υψηλής ποιότητας. Σε υλικά έμμεσου χάσματος, η απόδοση εκπομπής φωτός είναι αρκετά φτωχή και οι περισσότερες από τις διαδρομές επανασύνδεσης, είναι μη ακτινοβόλες με παραγωγή θερμότητας παρά φωτός.

Οι εφαρμογές των LED μπορούν να χωριστούν σε τρεις κατηγορίες:

1. *Ορατή απεικόνιση, όπου το φως κατευθύνεται περισσότερο ή λιγότερο στο ανθρώπινο μάτι για να μεταφέρει ένα μήνυμα ή μια έννοια.*
 - ο Η χαμηλή κατανάλωση ενέργειας, το χαμηλό κόστος συντήρησης και το μικρό μέγεθος των σημερινών LED έχει οδηγήσει στην εφαρμογή τους σε μια πληθώρα εγκαταστάσεων και εξοπλισμού για την ορατή απεικόνιση. Τοποθετούνται σε μεγάλες επιφάνειες για την αποτύπωση δεδομένων και μηνυμάτων όπως στα στάδια, στους σταθμούς των τρένων και των λεωφορείων, στα αεροδρόμια, στα λιμάνια, στους δρόμους ως πινακίδες ή ως σηματοδότες, στις συναυλίες. Εξαιτίας της μεγάλης διάρκειας ζωής τους και των μικρών χρόνων απόκρισης, τα LED χρησιμοποιούνται στην αυτοκινητοβιομηχανία ως σηματοδότες φρένων και στροφής (φλας). Έτσι αυξάνεται και ο παράγοντας ασφάλειας που μπορεί να παρέχει ένα όχημα αφού ο οδηγός έχει στην διάθεσή του περισσότερο χρόνο αντίδρασης.

2. *Φωταγωγήση, όπου το φως από τα LED ανακλάται από αντικείμενα για να μπορούν αυτά να είναι ορατά.*
 - Με την εξέλιξη των LED υψηλής απόδοσης και ισχύος έγινε δυνατή η χρήση τους για φωτισμό και φωταγωγήση. Τα LED χρησιμοποιούνται στα φώτα των δρόμων ή σε αρχιτεκτονικές κατασκευές που απαιτείται φωτισμός με εναλλαγή χρωμάτων. Επίσης χρησιμοποιούνται και ως κύρια φώτα στα αυτοκίνητα, στις μοτοσυκλέτες και στα ποδήλατα.
 - Επίσης τα LED χρησιμοποιούνται όλο και περισσότερο στον φωτισμό ενυδρείων. Μπορούν να παρέχουν το απαιτούμενο φως με μικρότερη εκπομπή θερμότητας και έτσι βοηθούν στην συντήρηση της βέλτιστης θερμοκρασίας του ενυδρείου. Επειδή μπορούν να παράγουν φως με συγκεκριμένο μήκος κύματος είναι ιδανικά για να παρέχουν ένα συγκεκριμένο χρώμα-φάσμα για τον χρωματισμό των κοραλλιών, των ψαριών, των ανεμώνων κ.α.



Εικόνα 13 – Ενυδρείο με κοράλλια υπό φωτισμό LED

3. *Παράγωγή φωτός για μέτρηση και αλληλεπίδραση με διαδικασίες που δεν γίνονται αντιληπτές από το ανθρώπινο μάτι.*

Το φως έχει και άλλες χρήσεις έκτος από την όραση. Έτσι λοιπόν η χρήση τους επεκτείνεται και σε εφαρμογές επικοινωνίας, σε αισθητήρες αλλά και για περιπτώσεις όπου χρειάζεται αλληλεπίδραση φωτός με την ύλη.

- Το φως των LED μπορεί να τροποποιηθεί πολύ γρήγορα και έτσι χρησιμοποιείται στις οπτικές ίνες και στις οπτικές επικοινωνίες ελεύθερου χώρου. Αυτές περιλαμβάνουν τα τηλεχειριστήρια των τηλεοράσεων και των βίντεο, όπου χρησιμοποιούνται τα LED υπέρυθρου.
- Πολλά συστήματα αισθητήρων χρησιμοποιούν το φως ως πηγή σήματος. Τα LED είναι ιδανική πηγή για να ανταπεξέλθει στις απαιτήσεις των αισθητήρων. Χρησιμοποιούνται στους αισθητήρες κίνησης ή στις οθόνες αφής εντοπίζοντας το φως που ανακλάται από το δάχτυλο ή την ακίδα.
- Μια άλλη χρήση των LED είναι στις καλλιέργειες φυτών κυρίως γιατί είναι ενεργειακά αποδοτικά, παράγουν θερμότητα που δεν είναι ικανή να τα καταστρέψει γιατί είναι μικρή, και μπορούν να παρέχουν την βέλτιστη οπτική συχνότητα για την ανάπτυξή τους.

Στο Kinect, υπάρχει ένα LED το οποίο δεν έχει κάποια ιδιαίτερη χρησιμότητα πέρα από το να υποδεικνύει την κατάσταση της συσκευής, άρα θα λέγαμε ότι ανήκει στην κατηγορία ορατής απεικόνισης για τη μεταφορά μηνύματος.



Εικόνα 14 – Kinect’s LED light

2.4.7 Μηχανική Βάση (Motorized Tilt)

Η συσκευή διαθέτει επίσης μια μηχανικής βάση, η οποία ρυθμίζεται δυναμικά μέσω κώδικα. Η κίνηση της βάσης προέρχεται από ένα μικρό μοτέρ στο μέγεθος νομίσματος και τρία εύθραυστα πλαστικά γρανάζια, τα οποία είναι ευαίσθητα στη θερμότητα και ίσως αποτελούν

πιο αδύναμο σημείο της συσκευής. Η βάση δίνει στο Kinect τη δυνατότητα να στραφεί προς τα πάνω ή κάτω κατά 27°.

Λόγω όμως αυτού του μηχανοκίνητου μηχανισμού ανάκλισης απαιτείται περισσότερη ενέργεια από αυτή που μπορεί να του παρέχει μια USB θύρα. Για το λόγο αυτό η συσκευή κάνει χρήση ενός ειδικού καλωδίου τροφοδοσίας (περιλαμβάνεται στο πακέτο μαζί με τον αισθητήρα), το οποίο χωρίζει τη σύνδεση σε ξεχωριστές συνδέσεις USB και ισχύος. Η απαιτούμενη ενέργεια(12 Watt) παρέχεται από το ηλεκτρικό δίκτυο μέσω ενός μετασχηματιστή.



Εικόνα 15 - Η μηχανοκίνητη βάση του Kinect

2.4.8 Τα διαθέσιμα APIs του Kinect

- ✚ Το API είναι μια συντομογραφία του όρου **Application Programming Interface** (Διεπαφή προγραμματισμού εφαρμογών), το οποίο είναι μια προδιαγραφή για το πώς τα διάφορα προγράμματα θα αλληλεπιδρούν με το λειτουργικό σύστημα.
- ✚ Πρακτικά, στις περισσότερες των περιπτώσεων, ένα API είναι μια βιβλιοθήκη που συνήθως περιλαμβάνει προδιαγραφές για μεθόδους, δομές δεδομένων, κλάσεις αντικειμένων, και μεταβλητές.
- ✚ Άρα είναι διεπαφή ανάμεσα σε εφαρμογές, κι όχι διεπαφή χρήστη. Με τα APIs οι εφαρμογές επικοινωνούν χωρίς να γίνεται αντιληπτό από το χρήστη.

- ✚ Η χρήση ενός API σημαίνει ότι ένα πρόγραμμα οδήγησης ή άλλη εφαρμογή, είναι διαθέσιμο στον υπολογιστή για να κάνει κάποιες διαδικασίες .

Μέχρι στιγμής υπάρχουν αρκετά δημοφιλή APIs για την αποκωδικοποίηση των δεδομένων που παράγει το Kinect τα οποία είναι ελεύθερα προς χρήση όπως :

- το official SDK της Microsoft [7]
- OpenNI και NITE [8]
- Libfreenect [9]
- CLNUI [10]
- Evoluce SDK [11]

2.4.8.1 Microsoft SDK for Kinect

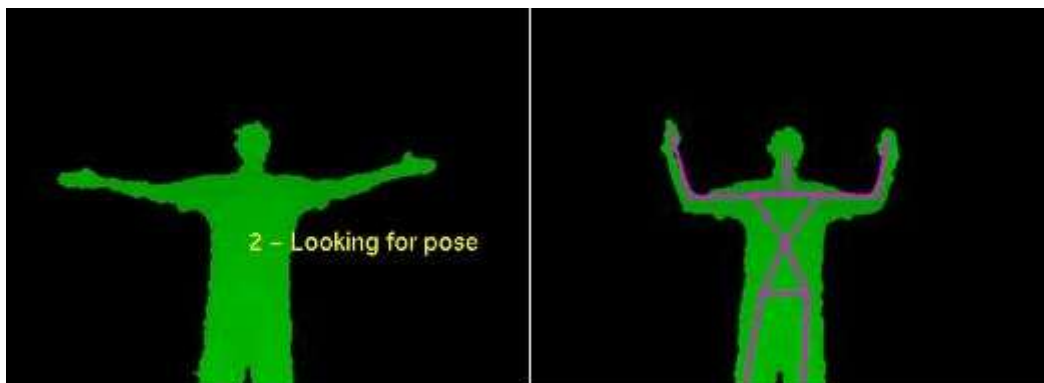
Αναλυτική περιγραφή ενότητας γίνεται παρακάτω (βλ. κεφάλαιο 4.6).

2.4.8.2 OpenNI και NITE

Δημιουργήθηκε από έναν μη κερδοσκοπικό οργανισμό ο οποίος απαρτίζεται από διάφορες εταιρίες, συμπεριλαμβανομένης και της PrimeSense Ltd. οι οποίες θέλησαν να θέσουν ένα βιομηχανικό πρότυπο λειτουργικότητας για τις συσκευές φυσική διεπαφής χρήστη – υπολογιστή (Natural User Interaction Devices). Κυκλοφόρησε το Δεκέμβριο του 2010. Το OpenNI αναπτύχθηκε σε C/C++ έτσι ώστε να μπορεί να χρησιμοποιηθεί από διάφορα λειτουργικά συστήματα, όπως Mac OSX, Ubuntu, Windows. Είναι το επίσημο λογισμικό των Xtion συσκευών της Asus, αλλά μπορεί να λειτουργήσει και με το Kinect.

Το OpenNI, όμως, συνοδεύεται και από μια ενδιάμεση βιβλιοθήκη η οποία λέγεται NITE τα αρχικά της οποίας σημαίνουν Natural Interaction Technology for End-user και είναι εξοπλισμένο με τεχνολογίες αναγνώρισης φωνής (Voice Recognition), αναγνώρισης χειρονομιών χειρών (Hand Gesture Recognition), και ανίχνευση σκελετού (Skeleton Tracking).

Η εγκατάσταση των drivers του OpenNi και του PrimeSense NITE Middleware είναι λίγο περίπλοκη καθώς πρέπει να γίνει εγκατάσταση σε κοινούς φακέλους του συστήματος, και αυτό γίνεται μόνο χειροκίνητα από το χρήστη.



Εικόνα 16 - Calibration pose for OpenNI

2.4.8.3 *Open Kinect – Libfreenect*

Το Libfreenect ήταν η πρώτη βιβλιοθήκη που δημιουργήθηκε για την αποκωδικοποίηση των δεδομένων του Kinect. Κυκλοφόρησε στις αρχές του Νοεμβρίου του 2010, λίγο μετά την κυκλοφορία του Kinect στην αγορά. Δημοσιεύθηκε στην ιστοσελίδα του Open Kinect στις 10 Νοεμβρίου όπου αποτέλεσε την αρχή της σύνταξης μιας μεγάλης κοινωνίας προγραμματιστών για τη χρήση του και την εξέλιξή του. Είναι αναπτυγμένο σε C και Python. Χαρακτηρίζεται API χαμηλού επιπέδου γιατί πέρα από την επικοινωνία με τη συσκευή δεν παρέχει μεθόδους όπως ανίχνευση σκελετού (Skeleton Tracking).

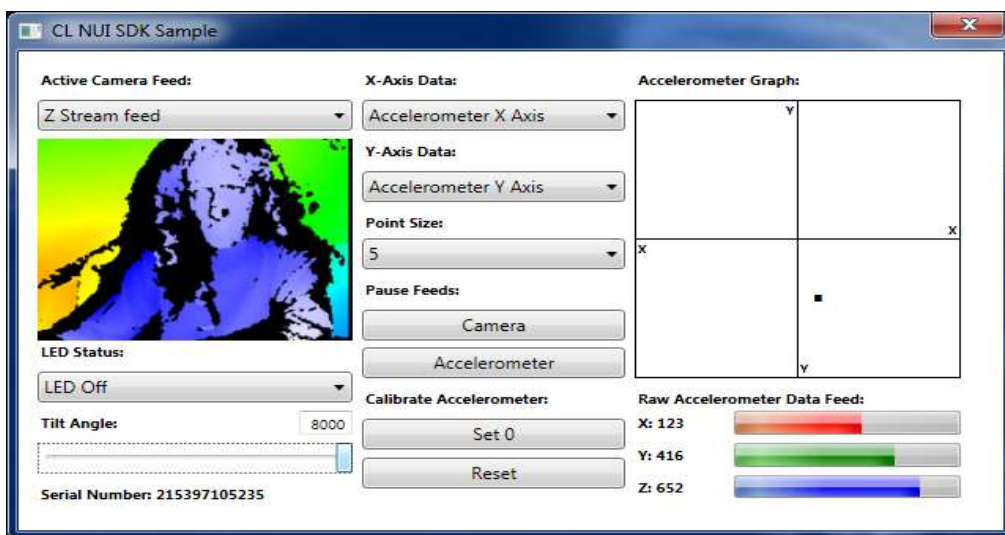


Εικόνα 17 - Χάρτης βάθους και RGB κάμερα με τη χρήση του Libfreenect

2.4.8.4 *CL NUI*

Η πρώτη έκδοση της πλατφόρμας CL NUI εκδόθηκε στις 8 Δεκεμβρίου 2010, και είναι ευρέως γνωστή στον κλάδο της ρομποτικής, γιατί είναι η μόνη βιβλιοθήκη που δίνει τη

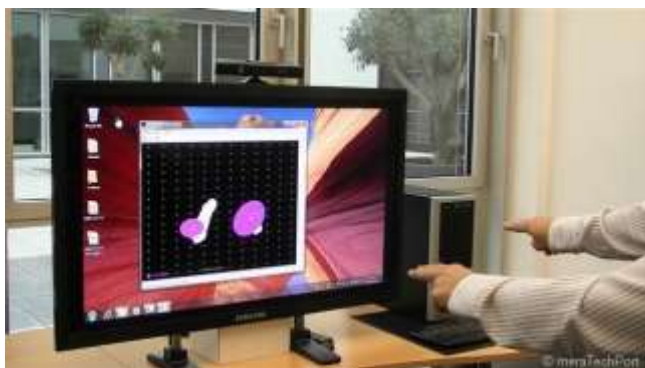
δυνατότητα πρόσβασης στο επιταχυνσιόμετρο του Kinect. Το CL NUI λειτουργεί μόνο σε Windows XP/Vista και 7 στις 32 αλλά και 64 bit εκδόσεις τους και παρέχει πρόσβαση στην κάμερα χρώματος, στον αισθητήρα βάθους, στο επιταχυνσιόμετρο, στο LED αλλά και στην μηχανοκίνητη βάση του Kinect. Χαρακτηρίζεται και αυτό ως API χαμηλού επιπέδου και απευθύνεται μονάχα στη συσκευή Microsoft Kinect. Το API συνοδεύεται από δύο παραδείγματα, εκ των οποίων το ένα κάνει χρήση του επιταχυνσιόμετρου.



Εικόνα 18 - Παράδειγμα χρήσης επιταχυνσιόμετρου με χρήση του CL NUI.

2.4.8.5 Evoluce SDK

Η εταιρία Evoluce ανέπτυξε ένα SDK βασισμένο στο OpenNI. Με τη χρήση της βιβλιοθήκης Emgu CV [12] πρόσθεσαν τη δυνατότητα ανίχνευσης των δακτύλων του χρήστη. Το Evoluce SDK δημοσιεύθηκε το Νοέμβριο του 2011 και πέρα από τις ιδιότητες του OpenNI παρέχει αναγνώριση φωνής (Speech Recognition), ανίχνευση χειρονομιών με τα δάκτυλα (Finger Gesture Recognition), υποστήριξη του Surface 2.0. Η χρήση του περιορίζεται μονάχα στα Windows 7.



Εικόνα 19 - Finger gesture recognition με τη χρήση του Evoluce SDK

2.5 Asus - Xtion Pro και Xtion Pro LIVE

2.5.1 Ιστορικό εξέλιξης

Το τέλος της άνοιξης του 2011 κυκλοφόρησε η πρώτη depth sensing συσκευή της Asus: το Asus Wavi Xtion Pro[13]. Απευθυνόταν καθαρά σε developers και περιείχε μονάχα τον αισθητήρα βάθους της Primesense Ltd.



Εικόνα 20 – Asus Wavi Xtion Pro

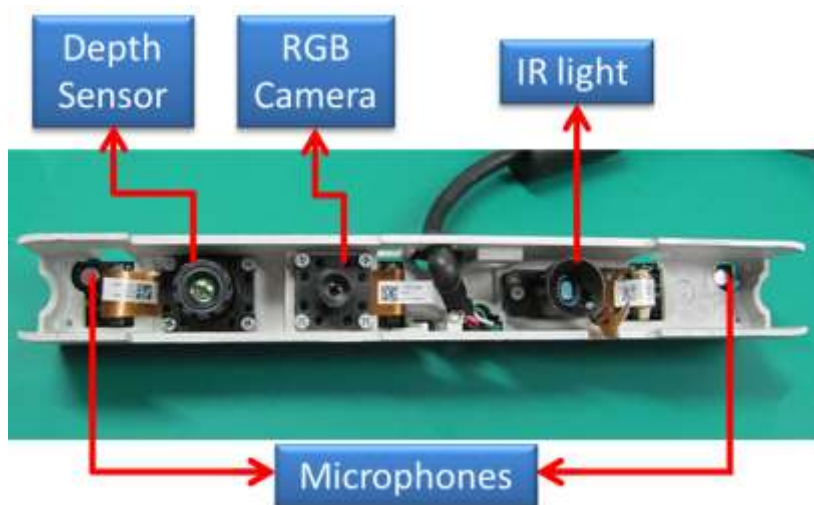
Η σύνδεσή του με τον υπολογιστή γινόταν μέσω USB θύρας και ένα μεγάλο πλεονέκτημά του ήταν ότι δεν χρειαζόταν εξωτερική τροφοδοσία. Αυτό το καθιστούσε πιο φορητό από το Kinect και σε συνδυασμό με το χαμηλό του βάρος βοηθούσε πάρα πολύ στο πεδίο της ρομποτικής. Όμως το Xtion Pro δεν είχε τόση απήχηση στην αγορά επειδή είχε αποκλειστικά και μόνο τον αισθητήρα βάθους. Έτσι, μετά από μερικούς μήνες η Asus σταμάτησε την παραγωγή του και κυκλοφόρησε μία νέα συσκευή με το όνομα Asus Xtion Pro Live.



Εικόνα 21 - Asus Xtion Pro LIVE

2.5.2 Τα χαρακτηριστικά του Xtion Pro Live

Το Xtion Pro Live [14] σχεδιάστηκε έτσι ώστε να μπορεί να συναγωνιστεί το Microsoft Kinect, εξοπλισμένη με πολλά όμοια χαρακτηριστικά (βλ. Πίνακας 1). Παρέχει μια κάμερα χρώματος, ένα αισθητήρα βάθους και 2 μικρόφωνα (βλ. Εικόνα 22).



Εικόνα 22 - Το εσωτερικό της συσκευής Xtion Pro LIVE

Κάποια από τα πλεονεκτήματα που παρέχει η νέα συσκευή της Asus είναι:

- Δεν χρειάζεται εξωτερική τροφοδοσία – Όπως και το προηγούμενο μοντέλο, το Xtion Pro LIVE, δεν χρειάζεται εξωτερική τροφοδοσία, δουλεύει με την ενέργεια που του παρέχει η USB θύρα.
- Μικρότερο σε μέγεθος – Το μέγεθός του (7" x 2" x 1.5") είναι αρκετά πιο εργονομικό από του Kinect.
- Χαμηλότερο βάρος – Το βάρος του (0.5 lb.), βοηθά πολύ στο πεδίο της ρομποτικής καθώς το βάρος το οποίο μπορούν να υποστηρίξουν οι ρομποτικές συσκευές είναι περιορισμένο.

	Asus Xtion Pro Live	Microsoft Kinect
Διαστάσεις	7" x 2" x 1.5"	12" x 3" x 2.5"
Βάρος	0.5 λίβρες (226~ γρ.)	3.0 λίβρες (1360~ γρ.)
Κάμερα χρώματος	1280x1024	1280x1024
Αισθητήρας Βάθους	PrimeSense	PrimeSense
Μικρόφωνα	2	4
Επιταχυνσιόμετρο	Όχι	Ναι
Μηχανοκίνητη Βάση	Όχι	Ναι
Εξωτερική τροφοδοσία	Όχι	Ναι

Πίνακας 1 - Χαρακτηριστικά Asus Xtion Pro Live vs. Microsoft Kinect

2.6 Leap Motion – The Leap

Η Leap Motion είναι μια εταιρεία που αναπτύσσει προηγμένη τεχνολογία ανίχνευσης κίνησης για την αλληλεπίδραση ανθρώπου-υπολογιστή.

Αρχικά εμπνευσμένη από την απογοήτευση γύρω από το 3D modeling, χρησιμοποιώντας το ποντίκι και το πληκτρολόγιο, θέλει να ισχυρίζεται ότι πραγματοποιώντας κάτι στην εικονική πραγματικότητα, πρέπει να είναι τόσο εύκολο όσο και στην πραγματική ζωή.

Μη δίνοντας στη δημοσιότητα τις κινήσεις της από το 2010, η εταιρεία ανακοινώνει το πρώτο της προϊόν το The Leap [15], στις 21 Μαΐου 2012.



Εικόνα 23 - The Leap

2.6.1 Τα χαρακτηριστικά του Leap

Η τρέχουσα διαθέσιμη έκδοση του Leap, είναι μια μικρή ενσύρματη USB περιφερειακή συσκευή που έχει σχεδιαστεί για να τοποθετηθεί σε μια επιφάνεια εργασίας, κοιτώντας προς τα πάνω. Χρησιμοποιώντας δύο κάμερες και τρία υπέρυθρα LED, το πεδίο αναγνώρισης της συσκευής είναι ημισφαιρικό, σε απόσταση περίπου 1 μέτρο.

Έχει σχεδιαστεί για να παρακολουθεί τα δάχτυλα (ή παρόμοια αντικείμενα, όπως ένα στυλό) που διασχίζουν μέσα στην προαναφερόμενη περιοχή, σε μια χωρική ακρίβεια περίπου 0,01 mm (βλ. Εικόνα 24). Επειδή η εικόνα είναι λίγο παραπλανητική το Leap βρίσκεται κάτω από το μολύβι ή στυλό που φαίνεται, με το οποίο ο χρήστης “γράφει” και βρίσκεται σε απόσταση από την οθόνη χωρίς να την ακουμπάει.



Εικόνα 24 – Leap’s smaller observation area and high resolution

Η μικρότερη περιοχή παρατήρησης και υψηλότερη ανάλυση της συσκευής διαφοροποιεί το προϊόν από το Kinect, το οποίο είναι πιο κατάλληλο για την παρακολούθηση ολόκληρου του σώματος σε ένα χώρο, στο μέγεθος ενός καθιστικού. Σε μια επίδειξη, το Leap δείχθηκε να εκτελεί καθήκοντα όπως η πλοήγηση σε μια ιστοσελίδα, χρησιμοποιώντας pinch-to-zoom χειρονομίες σε χάρτες, υψηλής ακρίβειας σχεδίασης, και το χειρισμό πολύπλοκων 3D data visualizations.

2.6.2 Χρήση από το κοινό

Η Leap Motion έχει διανείμει χιλιάδες μονάδες δωρεάν για τους προγραμματιστές που ενδιαφέρονται να δημιουργήσουν εφαρμογές για τη συσκευή. Ο ανιχνευτής κίνησης Leap είχε προγραμματισθεί να κυκλοφορήσει στο ευρύτερο κοινό, σύμφωνα με ανακοίνωση της εταιρείας, στις 22 Ιούλη 2013, με τιμή στις ΗΠΑ τα 79.99\$.

Ένα από τα θετικά στοιχεία που συνοδεύουν τη συσκευή είναι το Airspace [16], ηλεκτρονικό κατάστημα εφαρμογών. Αυτό είναι ένα πολύ μεγάλο πλεονέκτημα για τους προγραμματιστές από την άποψη της γρήγορης εύρεσης εφαρμογών, αλλά και για τους πελάτες, έτσι ώστε η διασφάλιση της απόκτησης των κατάλληλων και συμβατών με το Leap εφαρμογών, να είναι σίγουρη. Αυτή η κίνηση από την Motion Leap, αποτελεί αναπόσπαστο μέρος της στρατηγικής της παρασκευάστριας εταιρείας, δεδομένου ότι χωρίς το λογισμικό για τη χρήση αυτής μοναδικά ιδιαίτερης συσκευής, είναι μάλλον απίθανο ότι κάποιος θα «κολλήσει» μαζί της μακροπρόθεσμα.

2.6.3 Συνεργασίες της Leap Motion και μελλοντικά σχέδια

Έχει συνεργαστεί με την ASUS, από την οποία αναμένεται να κυκλοφορήσει στην αγορά υψηλής τεχνολογίας notebook και AiO (All in One) PCs, με την τεχνολογία του Leap, το 2013, όπως επίσης και με τη Hewlett Packard που ανακοίνωσε την ενσωμάτωση της τεχνολογίας στους HP υπολογιστές.

Ένα από τα ελάχιστα μελλοντικά σχέδια που η Leap Motion έχει γνωστοποιήσει είναι μια ασύρματη έκδοση της τρέχουσας συσκευής.

3 Αλληλεπίδραση Ανθρώπου με Υπολογιστή – Human Computer Interaction (HCI)

3.1 Εισαγωγή

Η αλληλεπίδραση μεταξύ ανθρώπου – υπολογιστή, μελετάται από το επιστημονικό πεδίο του Human Computer Interaction (HCI) με γνώμονα και τις δυο αυτές πλευρές και κατατάσσεται μεταξύ των πλέον αναπτυσσόμενων επιστημονικών και τεχνολογικών πεδίων. Αναπτύχθηκε κυρίως τη δεκαετία του 1980 και λογικό είναι, από τότε μέχρι σήμερα να έχουν αλλάξει πολλά πράγματα σ' αυτόν τον τομέα, λόγω του ότι για παράδειγμα η αρχική χρήση των υπολογιστών ήταν κυρίως για υπολογισμούς από επιστήμονες σε ερευνητικά εργαστήρια και τώρα επεκτείνεται σε σημείο να μπορεί να καλύψει ανάγκες επικοινωνίας και συνεργασίας στο πλαίσιο εκτέλεσης κοινωνικών δραστηριοτήτων της καθημερινής πρακτικής.

Από την πλευρά της μηχανής (υπολογιστή), η αλληλεπίδραση σχετίζεται με την επιστήμη στα γραφικά, τα λειτουργικά συστήματα, τις γλώσσες προγραμματισμού, το εσωτερικό υλικό των συσκευών, ενώ από την ανθρώπινη πλευρά σχετίζεται με πολλές επιστήμες όπως τη γλωσσολογία, την εργονομία, την κοινωνιολογία, τη γνωστική ψυχολογία, τις καλές τέχνες αλλά και τον ίδιο τον ανθρώπινο παράγοντα. Έτσι λοιπόν, λόγω του διεπιστημονικού χαρακτήρα του HCI, άνθρωποι με διαφορετικό αντικείμενο γνώσεων μπορούν να συμβάλουν στην επιτυχημένη εξέλιξη του.

Τα παραπάνω οδηγούν στο συμπέρασμα ότι η αλληλεπίδραση μεταξύ ανθρώπων και μηχανών δεν είναι τίποτα άλλο από στοιχεία που κατανοεί ο άνθρωπος, σε στοιχεία που αντιλαμβάνεται και ερμηνεύει ο υπολογιστής. Για να υπάρξει όμως αλληλεπίδραση θα πρέπει να γεφυρωθεί το χάσμα μεταξύ των δύο αυτών παραγόντων. Τη γεφύρωση αυτή αναλαμβάνει να πραγματοποιήσει η διεπαφή χρήστη που ορίζεται ως :

- ✚ το τμήμα της επικοινωνίας ανθρώπου – υπολογιστή που αφορά τους τρόπους, τις μεθόδους και τα εργαλεία που υλοποιούν την επικοινωνιακή διάδραση ανάμεσα στις οντότητες του χρήστη και του υπολογιστικού συστήματος.

Ένας βασικός στόχος του HCI είναι η βελτίωση αυτής της αλληλεπίδρασης, κάνοντας τους υπολογιστές πιο εύχρηστους και δεκτικούς στις ανάγκες των χρηστών. Συγκεκριμένα οι κύριοι τομείς ενασχόλησης του πεδίου είναι :

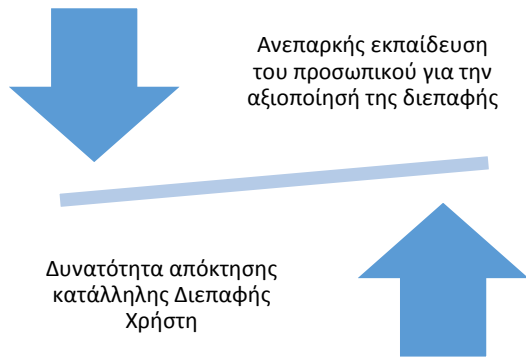
- *Μεθοδολογίες και διαδικασίες για το σχεδιασμό διεπαφών*
(δηλαδή, δίνεται ένα έργο και μια κατηγορία χρηστών, με σκοπό τον σχεδιασμό διεπαφής με τους συγκεκριμένους περιορισμούς, βελτιστοποιώντας τους παράγοντες όπως τη μάθηση ή την αποτελεσματικότητα της χρήσης)
- *Μέθοδοι για την ενσωμάτωση εργαλείων λογισμικού, βιβλιοθηκών και αποδοτικών αλγορίθμων στις διεπαφές*
- *Τεχνικές για την αξιολόγηση και τη σύγκριση των διεπαφών*
- *Ανάπτυξη νέων διεπαφών και τεχνικών αλληλεπίδρασης*

3.2 Διεπαφές Χρήστη – User Interfaces

Ο όρος διεπαφή χρήστη (user interface) όπως αναφέρθηκε και προηγουμένως είναι το σύνολο των συστατικών ενός συστήματος το οποίο επιτρέπει αμφίδρομη επικοινωνία μεταξύ συστήματος και χρήστη. Η διεπαφή χρήστη ενός συστήματος έχει σχέση με το ίδιο το σύστημα, το χρήστη του συστήματος και τον τρόπο που αλληλεπιδρούν μεταξύ τους. Ο όρος θέλει να δείξει το σημείο επαφής χρήστη και υπολογιστή, την γραμμή επαφής πίσω από την μεριά της οποίας βρίσκεται η μηχανή και από την άλλη μεριά ο άνθρωπος.

Έτσι λοιπόν η διεπαφή περιέχει στοιχεία που είναι τμήματα τόσο του υλικού του συστήματος - hardware (φυσικά), όσο και του λογισμικού - software (λογικά) που τρέχουν σε αυτό. Διεπαφές χρήστη υπάρχουν για διάφορα συστήματα, παρέχοντας :

- Είσοδο (Input), επιτρέποντας στους χρήστες να χειριστούν το σύστημα
- Έξοδο (Output), επιτρέποντας στο σύστημα να αναφέρει τα αποτελέσματα της χειραγώγησης των χρηστών



Εικόνα 25 – Ένας από τους λόγους που προκλήθηκε το ατύχημα του Three Mile Island ήταν η ανεπαρκής εκπαίδευση του προσωπικού

Η αλληλεπίδραση ανθρώπου-μηχανής σε ορισμένες περιπτώσεις είναι πολύ σημαντική, διότι κακοσχεδιασμένες διεπαφές μπορεί να οδηγήσουν σε πολλά απρόβλεπτα προβλήματα. Ένα κλασικό παράδειγμα αυτού, είναι το ατύχημα στο Three Mile Island, με την κατάρρευση του πυρήνα ενός αντιδραστήρα, με τις έρευνες να καταλήγουν στο συμπέρασμα ότι ο σχεδιασμός της διεπαφής ανθρώπου-μηχανής ήταν τουλάχιστον εν μέρει υπεύθυνος για την καταστροφή.[17][18]

3.2.1 Είδη διεπαφής χρήστη

Υπάρχουν πολλών ειδών διεπαφές χρήστη που αναλόγως τον σκοπό που θέλουν να επιτελέσουν και τις συσκευές εισόδου που έχει το εκάστοτε σύστημα προκύπτουν και οι κατηγορίες τους. Ορισμένες από αυτές είναι :

- Command Line Interface (CLI) → ο χρήστης παρέχει είσοδο πληκτρολογώντας μια σειρά εντολών με το πληκτρολόγιο του υπολογιστή

```
root@fsc2400:/home/larhzu# pkgtool --help
Usage: pkgtool [options] [command [command-arguments]]

Options:  -R dir,          Use a different root directory. This option
          --root dir      overrides the $ROOT environment variable.
          -q, --quiet     Suppress messages except wget progress bar.

Commands: check          Check for basic integrity of package database
                        and report problems found.
          purge          Empty the cache directory of downloaded packages.
          tools          Report which helper applications are installed.
          setup <source> <root_device>
                        Start the Slackware or Slackles installation to
                        the directory specified by --root or $ROOT.
                        Root device means the device on which the target
                        root filesystem is. E.g. /dev/hda1

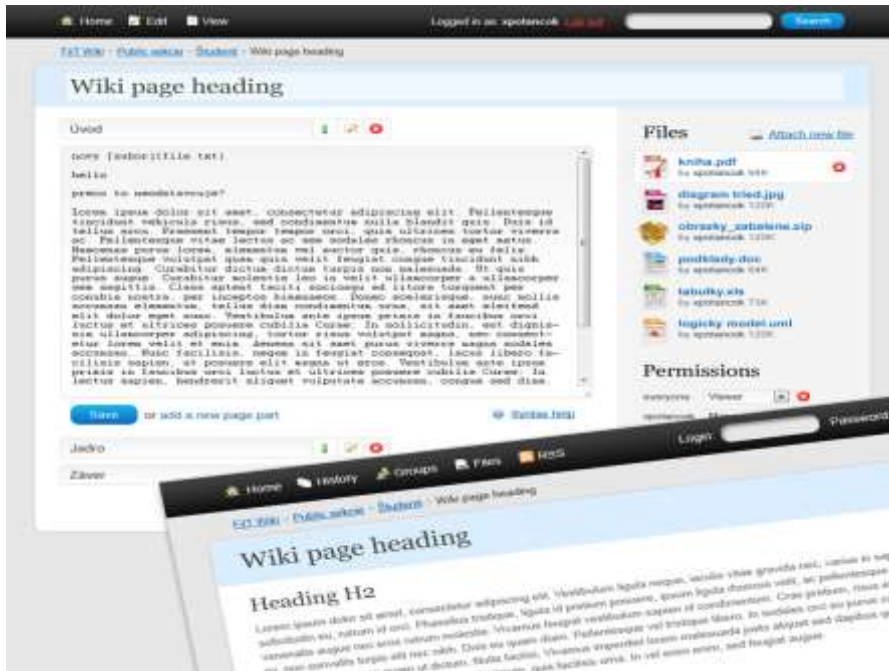
Interactive repository commands:
  menu <reponame> <search_regex>
  menu-desc <reponame> <search_regex>
  menu-updates <reponame>
  menu-dir <directory>

Non-interactive repository commands i.e. they ask nothing from the user:
  list
  add <reponame> <address>
  delete <reponame>
  rename <reponame> <newname>
  address <reponame> <newaddress>
  update <reponame>
  search <reponame> <search_regex>
  search-desc <reponame> <search_regex>
  install <reponame> <full_package_name>
  list-updates <reponame>
  install-updates <reponame>

If no command is specified pkgtool starts with the interactive dialog mode.
root@fsc2400:/home/larhzu#
```

Εικόνα 26 – Command Line Interface example

- Web-based User Interface → δέχονται είσοδο και παρέχουν έξοδο με τη δημιουργία ιστοσελίδων, οι οποίες μεταδίδονται μέσω του διαδικτύου και προβάλλονται από το χρήστη χρησιμοποιώντας ένα πρόγραμμα περιήγησης (web browser). Για τη δημιουργία τους χρησιμοποιούνται γλώσσες προγραμματισμού όπως Java, JavaScript, AJAX, Adobe Flex, PHP, Microsoft .NET, και πολλές άλλες.



Εικόνα 27 – A Web-based User Interface example

- Touch User Interface → χρησιμοποιούνται από συσκευές που έχουν οθόνες αφής. Δέχονται είσοδο από το άγγιγμα των δακτύλων ή μιας γραφίδας.



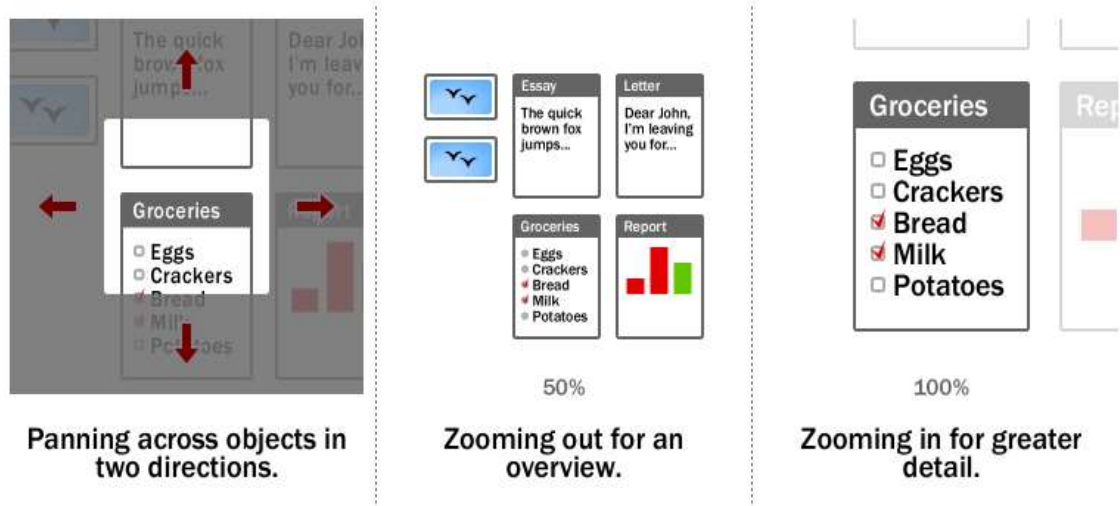
Εικόνα 28 – Touch screen interface example

- Graphical User Interface → δέχονται είσοδο μέσω συσκευών όπως το πληκτρολόγιο και το ποντίκι του υπολογιστή και να παρέχουν γραφική έξοδο στην οθόνη του υπολογιστή
 - WIMP (windows, icons, menus, pointers) → είναι ένα στυλ αλληλεπίδρασης που χρησιμοποιεί μια συσκευή εισόδου για τον έλεγχο της θέσης του δείκτη, τις περισσότερες φορές ένα ποντίκι, και παρουσιάζει πληροφορίες που οργανώνονται σε παράθυρα και εκπροσωπούνται με εικονίδια. Οι διαθέσιμες εντολές, συνήθως συγκεντρώνονται μαζί σε μενού, και οι ενέργειες εκτελούνται πραγματοποιώντας χρησιμοποιώντας το ποντίκι - κουνώντας το και κάνοντας κλικ.



Εικόνα 29 – Graphical User Interface using WIMP style example

- Object-oriented User Interface → είναι βασισμένο στη **μεταφορά**¹ του αντικειμενοστραφή προγραμματισμού, επιτρέποντας στους χρήστες να χειριστούν αντικείμενα και τις ιδιότητές τους. Παράδειγμα από τις πρώτες εφαρμογές που χρησιμοποιεί τέτοιου είδους interface είναι το Smalltalk-80.[19]
- Zooming User Interface → είναι γραφικά περιβάλλοντα στα οποία τα αντικείμενα που έχουν τις πληροφορίες εμφανίζονται σε διάφορα μεγέθη, και όπου ο χρήστης μπορεί να αλλάξει την κλίμακα της εμφανιζόμενης περιοχής προκειμένου να μπορεί να δει περισσότερες λεπτομέρειες, καθώς περιηγείται σε διάφορα έγγραφα.



Εικόνα 30 – Zooming interface example

¹ Με την έννοια της **μεταφοράς** – στη συγκεκριμένη περίπτωση της διεπαφής - εννοούμε ένα σύνολο από γραφικά της διεπαφής χρήστη, ενέργειες και διαδικασίες που εκμεταλλεύονται ειδικές γνώσεις που οι χρήστες έχουν ήδη από άλλους τομείς. Ο σκοπός της μεταφοράς διεπαφής, βασισμένος σε μια δραστηριότητα, ένα αντικείμενο, ή ένα συνδυασμό και των δύο, χρησιμοποιώντας αυτές τις ήδη υπάρχουσες οικίες γνώσεις, είναι να βοηθήσει τον χρήστη να κατανοήσει κάτι «άγνωστο» καλύτερα και γρηγορότερα.

- Natural User Interface (NUI) → είναι τύπος διεπαφής χρήστη που είναι βασισμένη η εφαρμογή που παρουσιάζεται σε αυτή την πτυχιακή εργασία – αναφέρεται και παρουσιάζεται παρακάτω.
 - Gesture User Interface → είναι ένας τύπος διεπαφής, που επιτρέπει στους χρήστες να αλληλεπιδρούν με τις συσκευές υπολογιστών μέσω της κίνησης των αντικειμένων και του ανθρώπινου σώματος.
 - Voice User Interface → καθιστά δυνατή την ανθρώπινη αλληλεπίδραση με τους υπολογιστές μέσω μιας πλατφόρμας αναγνώρισης φωνής / ομιλίας, προκειμένου να ξεκινήσει μια αυτοματοποιημένη υπηρεσία ή μια διαδικασία.
 - Brain - Computer Interfaces → δέχονται σαν είσοδο σήματα των νευρώνων του εγκεφάλου που μεταφράζονται από κατάλληλα προγράμματα σε ενέργειες. Με αυτού του είδους τις διεπαφές είναι δυνατό για κάποιον που έχει παραλύσει να χρησιμοποιήσει έναν υπολογιστή, μια μηχανοκίνητη αναπηρική καρέκλα ή ένα προσθετικό τεχνητό άκρο μέσω της σκέψης και μόνο.



Εικόνα 31 – Brain – Computer Interface example

3.3 Φυσική Διεπαφή Χρήστη – Natural User Interface

Στις δεκαετίες του 1970, 1980 και του 1990 ο Steve Mann αναπτύσσει μια σειρά από στρατηγικές user interfaces χρησιμοποιώντας τη φυσική αλληλεπίδραση με τον πραγματικό κόσμο, ως μια εναλλακτική λύση σε ένα περιβάλλον γραμμής εντολών (CLI) και γραφικής διεπαφής (GUI). Ο Mann αναφερόταν σε αυτό το έργο, ως Φυσικές Διεπαφές Χρήστη (Natural User Interfaces), Άμεσες Διεπαφές Χρήστη (Direct User Interfaces), και Μεταφορά – Ελεύθερος Υπολογισμός (Metaphor – Free Computing)[20]. Η αναφορά του Mann στη λέξη φυσικό αναφέρεται τόσο στις φυσικές ενέργειες που κάνουν οι ανθρώπινοι χρήστες, καθώς και στη χρήση της ίδιας της φύσης, δηλαδή τη φυσική (Φιλοσοφία της Φύσης), όπως και το φυσικό περιβάλλον.

Το 2006 ο Christian Moore ίδρυσε μια ανοικτή ερευνητική κοινότητα με στόχο να επεκτείνει τη συζήτηση και την ανάπτυξη που σχετίζονται με NUI τεχνολογίες [21]. Σε μια παρουσίαση του 2008 συνέδριο με θέμα «Predicting the Past», ο August de los Reyes, ως εκπρόσωπος της Microsoft χαρακτήρισε τα NUI ως το επόμενο εξελικτικό στάδιο που ακολουθεί την μετάβαση από το CLI στο GUI [22]. Μια πιο ακριβής περιγραφή της έννοιας αυτής θα μπορούσε να περιγράψει ως μια μετάβαση από το WIMP στο NUI.



Εικόνα 32 – Η εξέλιξη των User Interfaces

3.3.1 Η φυσική διεπαφή στην επιστήμη των υπολογιστών

Στην επιστήμη υπολογιστών μιλώντας για μια Φυσική Διεπαφή Χρήστη εννοούμε ένα σύστημα αλληλεπίδρασης ανθρώπου - υπολογιστή όπου ο χρήστης λειτουργεί μέσω φυσικών κινήσεων που σχετίζονται με την καθημερινή ανθρώπινη συμπεριφορά.

Επίσης αναφέρεται σε μια διεπαφή που είναι ουσιαστικά «αόρατη», ή γίνεται αόρατη με τις διαδοχικές αλληλεπιδράσεις με τους χρήστες της και στηριζόμενη στη γρήγορη εκμάθησή της, μετατρέπει έναν αρχάριο χρήστη σε έμπειρο. Παρόλο που η διεπαφή απαιτεί εκμάθηση, ο χρήστης διευκολύνεται από τον σχεδιασμό, συνεχίζοντας ευχάριστα την επιτυχή μάθηση.

3.4 Πολυτροπικές Διεπαφές Χρήστη - Multimodal User Interfaces

Πολύ σημαντικό κομμάτι των NUI είναι οι πολύ-τροπικές διεπαφές που συνδυάζουν διάφορους τρόπους εισόδου για τον χρήστη, εκτός των παραδοσιακών (πληκτρολόγιο, ποντίκι), όπως η ομιλία, η χρήση γραφίδας, η αφή, οι χειρονομίες. Αντιπροσωπεύουν μια καινούρια κατεύθυνση, παρεκκλίνοντας από τις τυπικές γραφικές διεπαφές, επειδή περιλαμβάνουν τεχνολογίες αναγνώρισης σχεδιασμένες να χειρίζονται συνεχόμενα και ταυτόχρονα εισόδους από παράλληλα εισερχόμενα κανάλια. Οι πολυτροπικές διεπαφές χρήστη αποτελούν αναπόσπαστο αντικείμενο έρευνας του πεδίου του HCI. Αυτό το νέο είδος διεπαφών στοχεύει στην αναγνώριση της ανθρώπινης επικοινωνίας έτσι ώστε η αλληλεπίδραση με τον υπολογιστή να είναι πιο φυσική για τον χρήστη, με όλο και λιγότερους περιορισμούς.

Μεγάλο πλεονέκτημα αυτών των διεπαφών είναι ότι έχουν καλύτερη προσβασιμότητα. Μια πολυτροπική εφαρμογή σωστά σχεδιασμένη μπορεί να χρησιμοποιηθεί από ανθρώπους με διάφορες αναπηρίες, όπως για παράδειγμα χρήστες με προβλήματα όρασης μπορούν να επωφεληθούν περισσότερο χρησιμοποιώντας την τροπικότητα της φωνής σε συνδυασμό με κάποια άλλη. Χρήστες που έχουν προβλήματα ακοής χρησιμοποιούν ευκολότερα την οπτική τροπικότητα και ούτω καθεξής.

4 Τεχνολογία και Εργαλεία Υλοποίησης

4.1 .NET Framework

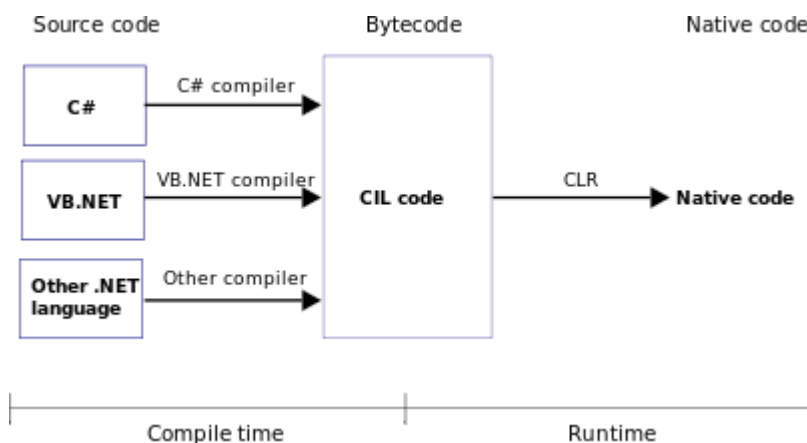
Το .NET Framework (τωρινή έκδοση 4) είναι μια επαναστατική πλατφόρμα που δημιουργήθηκε από τη Microsoft για την ανάπτυξη εφαρμογών. Αν και το .NET Framework της Microsoft τρέχει στο λειτουργικό σύστημα των Windows, είναι δυνατόν να βρεθούν εναλλακτικές εκδόσεις που θα δουλεύουν και σε άλλα λειτουργικά συστήματα. Επιπλέον, το Microsoft .NET Compact Framework (ουσιαστικά ένα υποσύνολο της πλήρους .NET Framework) χρησιμοποιείται σε προσωπικούς ψηφιακούς βοηθούς (Personal Digital Assistant – PDA) μέχρι και σε μερικά smartphones. Ένα από τα βασικά κίνητρα πίσω από το .NET Framework είναι η προβλεπόμενη χρήση του ως μέσο για την ενσωμάτωση διαφορετικών λειτουργικών συστημάτων.

Το .NET Framework έχει σχεδιαστεί έτσι ώστε να μπορεί να χρησιμοποιηθεί από πολλές γλώσσες προγραμματισμού, όπως της C #, C + +, Visual Basic, JScript, και από ακόμη παλιότερες γλώσσες, όπως η COBOL. Για να είναι αυτό εφικτό, οι .NET-εκδόσεις αυτών των γλωσσών δημιουργήθηκαν, και όλο και περισσότερες ανανεώνονται και κυκλοφορούν όλη την ώρα. Εκτός από το ότι όλες αυτές οι γλώσσες μπορούν να χρησιμοποιηθούν στο .NET Framework, μπορούν επίσης να επικοινωνούν μεταξύ τους. Για παράδειγμα είναι απολύτως εφικτό για προγραμματιστές C # για να κάνουν χρήση του κώδικα που γράφτηκε από προγραμματιστές της Visual Basic, και το αντίστροφο.

Στο εσωτερικό του αποτελείται κυρίως από μια γιγαντιαία βιβλιοθήκη κώδικα που χρησιμοποιείται από τις γλώσσες (όπως η C #) με τη βοήθεια των τεχνικών του αντικειμενοστραφή προγραμματισμού (OOP). Αυτή η βιβλιοθήκη κατηγοριοποιείται σε διαφορετικές ενότητες, οι οποίες χρησιμοποιούνται ανάλογα για τα θεμιτά αποτελέσματα. Για παράδειγμα, μία ενότητα περιέχει τα δομικά στοιχεία για τις εφαρμογές των Windows, ένα άλλο για τον προγραμματισμό του δικτύου, και ένα άλλο για την ανάπτυξη Ιστού. Ορισμένες ενότητες χωρίζονται σε πιο συγκεκριμένα υποστοιχεία, όπως μια ενότητα για την οικοδόμηση των υπηρεσιών Ιστού εντός της ενότητας για την ανάπτυξη Ιστού.

Εδώ χρειάζεται να γίνει αναφορά στο Common Language Runtime - CLR. Το CLR είναι εικονική μηχανή (virtual machine) του .NET Framework της Microsoft. Ο ρόλος του είναι η διαχείριση και η εκτέλεση των προγραμμάτων που έχουν δημιουργηθεί με το συγκεκριμένο

framework. Ο κώδικας που γίνεται compiled μετατρέπεται μέσω του CLR σε Common Intermediate Language - CIL κώδικα. Ο CIL κώδικας είναι μία αντικειμενοστραφής assembly γλώσσα και είναι η χαμηλότερη επιπέδου γλώσσα της Microsoft. Στην συνέχεια αυτός ο κώδικας μετατρέπεται μέσω του CLR σε εντολές μηχανής (machine instructions), οι οποίες τελικά εκτελούνται από τον επεξεργαστή.



Εικόνα 33 – Απεικόνιση της μετατροπής του Source code σε Bytecode και εν συνεχεία με το CLR σε Native code

4.2 Microsoft Visual C

Η Microsoft C # (προφέρεται C sharp) είναι μια νέα γλώσσα προγραμματισμού, δημιουργήθηκε το 2000, σχεδιασμένη για την οικοδόμηση ενός μεγάλου φάσματος επιχειρηματικών εφαρμογών που τρέχουν στο .NET Framework. Σαν μια εξέλιξη των C και C ++, η C # είναι απλή, σύγχρονη, έχει ασφάλεια τύπων, είναι αντικειμενοστραφής και ο σχεδιασμός της έγινε με σκοπό να ενσωματώσει πολλά από τα καλύτερα χαρακτηριστικά άλλων γλωσσών και την εξάλειψη των προβλημάτων τους. Προγραμματιστές που ξέρουν γλώσσες όπως C, C ++, Java και ορισμένες άλλες αντικειμενοστραφείς γλώσσες, είναι τυπικά έτοιμοι να δουλέψουν παραγωγικά στη C# μέσα σε μικρό χρονικό διάστημα.

4.3 Microsoft Visual Studio 2010

Για την υλοποίηση του προγράμματος ο κώδικας γράφτηκε στο Visual Studio 2010, που είναι ένα εργαλείο ανάπτυξης για το σύνολο του προγραμματισμού C #. Ένα εργαλείο ανάπτυξης, ή ολοκληρωμένο περιβάλλον ανάπτυξης (Integrated Development Environment – IDE) [23], όπως το VS δεν είναι απαραίτητο για την ανάπτυξη εφαρμογών C #, αλλά αυτό

κάνει τα πράγματα πολύ πιο εύκολα. Παρ' όλα αυτά ο κώδικας θα μπορούσε να γραφτεί σε μια εφαρμογή όπως το Σημειωματάριο, κάνοντας debug χρησιμοποιώντας τον compiler που αποτελεί μέρος του .NET Framework. Ωστόσο, γιατί να γίνει κάτι τέτοιο, όταν υπάρχει η δυνατότητα ενός IDE για να βοηθήσει.

Τα παρακάτω είναι μια μικρή λίστα με **μερικά χαρακτηριστικά** που καθιστούν το Visual Studio μια ελκυστική επιλογή για την ανάπτυξη εφαρμογών .NET :

- Αυτοματοποιεί τα βήματα που απαιτούνται για την κατάρτιση κώδικα και την ίδια στιγμή δίνει πλήρη έλεγχο σε οποιοσδήποτε επιλογές που χρησιμοποιούνται και πρέπει να παρακαμφθούν.
- Ο επεξεργαστής κειμένου VS είναι προσαρμοσμένος σε διάφορες γλώσσες (συμπεριλαμβανομένης της C #), έτσι ώστε να μπορεί να ανιχνεύει με έξυπνο τρόπο τα λάθη και να προτείνει κώδικα όπου ενδείκνυται, καθώς γίνεται πληκτρολόγηση. Αυτή η λειτουργία ονομάζεται IntelliSense.
- Το VS περιλαμβάνει σχεδιαστές για τα Windows Forms, φόρμες, και άλλες εφαρμογές, επιτρέποντας απλό drag-and-drop για το σχεδιασμό των στοιχείων UI.
- Πολλά project σε C # μπορεί να δημιουργηθούν με το "στερεότυπο" κώδικα που ήδη υπάρχουν, αντί να ξεκινήσει κάποιος από το μηδέν. Μερικά starter kits περιλαμβάνονται με την εγκατάσταση του VS, και μπορούν να βρεθούν πολλά περισσότερα online.
- Το VS δίνει τη δυνατότητα χρήσης προηγμένων τεχνικών εντοπισμού σφαλμάτων κατά την ανάπτυξη κάποιου project, έχοντας την ικανότητα να εκτελεί «γραμμή-γραμμή» τον κώδικα, γνωρίζοντας πάντα την κατάσταση της εφαρμογής.

4.4 Extensible Markup Language - XML

Η XML είναι μία γλώσσα σήμανσης, που περιέχει ένα σύνολο κανόνων για την ηλεκτρονική κωδικοποίηση κειμένων. Ορίζεται, κυρίως, στην προδιαγραφή XML 1.0 (XML 1.0 Specification), που δημιούργησε ο διεθνής οργανισμός προτύπων W3C (World Wide Web Consortium) [24] , αλλά και σε διάφορες άλλες σχετικές προδιαγραφές ανοιχτών προτύπων.

Στην παρούσα πτυχιακή, η γραμματική του speech recognition, είναι αποθηκευμένη σε ένα αρχείο grammar.xml .

4.5 MySQL

Η καθολική βάση δεδομένων του συστήματος είναι υλοποιημένη με τη τεχνολογία της MySQL. Η MySQL δεν είναι μόνο ένα σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων, είναι επίσης ένα ισχυρό εργαλείο οπτικού σχεδιασμού βάσεων δεδομένων (visual database design), το οποίο εμπλέκει την ανάπτυξη σε SQL, την διαχείριση, και το σχεδιασμό βάσης (database design). Ελέγχει την πρόσβαση στα δεδομένα, για να μπορούν να δουλεύουν πολλοί χρήστες ταυτόχρονα, για να παρέχει γρήγορη πρόσβαση και να διασφαλίζει ότι μόνο πιστοποιημένοι χρήστες μπορούν να έχουν πρόσβαση. Επίσης χρησιμοποιεί την SQL (Structured Query Language) την τυπική γλώσσα ερωτημάτων για βάσεις δεδομένων και υποστηρίζει τα τελευταία πρότυπα της αυτής της γλώσσας.

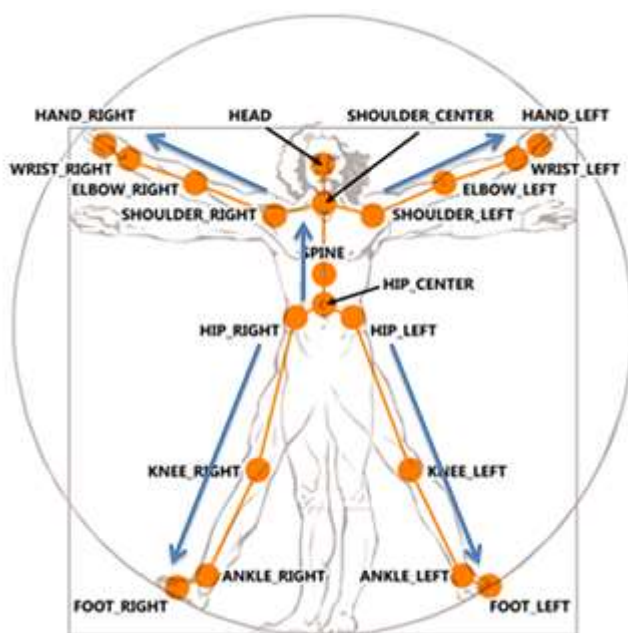
Τα κυριότερα πλεονεκτήματά της είναι η απόδοση, το ελάχιστο έως μηδενικό κόστος, η ευκολία στη χρήση της και η μεταφερσιμότητα. Επίσης έχει μια αρκετά καλή διεπαφή για την διαχείριση βάσεων (administration) μέσω του οποίου μπορεί ο διαχειριστής να δώσει δικαιώματα σε άλλους χρήστες, να κάνει τη διαχείριση της ασφάλειας και των συνδέσεων και πολλά άλλα.

Παράλληλα, η MySQL δίνει τη δυνατότητα του σχεδιασμού του μοντέλου μιας σχεσιακής βάσης δεδομένων στα πρότυπα της UML και στη συνέχεια την υλοποίησή της μέσω της λειτουργίας forward engineering. Με αυτό τον τρόπο γίνεται η αυτόματη δημιουργία των πινάκων που έχουν σχεδιαστεί χωρίς να χρειάζεται ο χρήστης να γράψει SQL κώδικα. Αυτή η διαδικασία προϋποθέτει πολύ καλή και προσεκτική σχεδίαση της βάσης χωρίς λάθη στην υλοποίηση. Επίσης υποστηρίζεται και η αντίστροφη διαδικασία που ονομάζεται reverse engineering, η οποία εξάγει το μοντέλο οντοτήτων συσχετίσεων από μία είδη υπάρχουσα υλοποιημένη βάση δεδομένων.

4.6 Microsoft SDK for Kinect

Στις 16 Ιουνίου του 2011 κυκλοφόρησε η beta έκδοση της βιβλιοθήκης της Microsoft για τη χρήση του Kinect στα Windows 7. Το SDK έδινε τη δυνατότητα στους προγραμματιστές να αναπτύξουν εφαρμογές σε C++, C# ή Visual Basic με τη χρήση του Microsoft Visual Studio 2010. Μερικές από τις αποκλειστικές δυνατότητες που προσέφερε είναι :

- Επεξεργασία ήχου με δυνατότητα υπολογισμού της πηγής (sound source estimation).
- Αναγνώριση ομιλίας (speech recognition).
- Υπολογισμός του σκελετού του χρήστη χωρίς να χρειάζεται αρχική στάση βαθμονόμησης (skeleton tracking without calibration pose) ανιχνεύοντας περισσότερες αρθρώσεις ανά χρήστη (αστραγάλους και καρπούς) (βλ. Εικόνα 34).
- Παραδείγματα με χρήση κώδικα και τεκμηρίωση (documentation).



Εικόνα 34 – Joints detected from Microsoft Skeleton Tracking

Στις αρχές του 2012 κυκλοφόρησε η πρώτη επίσημη έκδοση του SDK και μαζί της και η καινούργια έκδοση της συσκευής Kinect η οποία προοριζόταν αποκλειστικά για τη χρήση στον υπολογιστή. Το SDK είχε πολλές διαφορές στη σύνταξη του API της, παρείχε όμως το κατάλληλο documentation για να μπορούν οι προγραμματιστές να προσαρμόσουν τον κώδικά τους στα καινούρια δεδομένα. Έτσι πολλές εταιρείες άρχισαν την ανάπτυξη εφαρμογών με τη χρήση του SDK της Microsoft. Το Μάρτιο του 2012 ο Craig Eisler, γενικός διευθυντής του Kinect for Windows ανακοίνωσε ότι σχεδόν 350 εταιρίες συνεργάζονται με τη Microsoft για την ανάπτυξη διάφορων εφαρμογών με το Kinect για τα Windows.

Τον Μάιο του 2013 η Microsoft δημοσίευσε την Version 1.5 του Kinect for Windows SDK. Η έκδοση αυτή περιέχει μια καινούργια βιβλιοθήκη για ανίχνευση προσώπου (Face Detection),



Εικόνα 35 – Face Detection

αναγνώριση φωνής σε τέσσερις καινούργιες γλώσσες – Γαλλικά, Ισπανικά, Ιταλικά και Ιαπωνικά.

Όσον αφορά την ανίχνευση κίνησης του χρήστη, η Microsoft προσέθεσε την παράμετρο προσανατολισμού των αρθρώσεων του σκελετού (παράμετρο που παρείχε μέχρι τώρα μόνο το OpenNI) και ένα καινούργιο χαρακτηριστικό στο skeleton tracking :

- ο near ή seated mode με τη δυνατότητα να χρησιμοποιεί ο χρήστης μόνο το πάνω μέρος του σώματος του, με το Kinect να εντοπίζει και να χρησιμοποιεί μονάχα τις 10 αρθρώσεις του κεφαλιού και των χεριών (βλ. Εικόνα 36)



Εικόνα 36 – Joints tracked in seated mode

Η τελευταία έκδοση που κυκλοφόρησε τον Μάρτιο του 2013 είναι η 1.7. Τα καινούργια χαρακτηριστικά που έχει είναι :

- **Kinect Interactions :**

Ο συνδυασμός νέων ελέγχων καθιστά ευκολότερο για τους χρήστες να αλληλεπιδρούν με πιο φυσικό με τους υπολογιστές. Αυτοί περιλαμβάνουν εντολές όπως

 - "push" για επιλογή εικονικών αντικειμένων
 - "grip" για τη μετακίνηση και κύλιση
 - Αναγνώριση τεσσάρων χεριών ταυτόχρονα
- **Kinect Fusion :**

Καθίσταται πλέον δυνατό για τους προγραμματιστές να σχεδιάσουν για το Kinect for Windows εφαρμογές ανακατασκευής υψηλής ποιότητας, τριών διαστάσεων (3-D) απεικόνισης ανθρώπων και αντικειμένων σε πραγματικό χρόνο.
- **Καινούργια παραδείγματα και documentation σε**
 - OpenCV
 - MATLAB
 - Fusion
 - Interaction

4.6.1.1 *Hardware Requirements*

Για τη χρήση του SDK χρειαζόμαστε τις παρακάτω ελάχιστες απαιτήσεις:

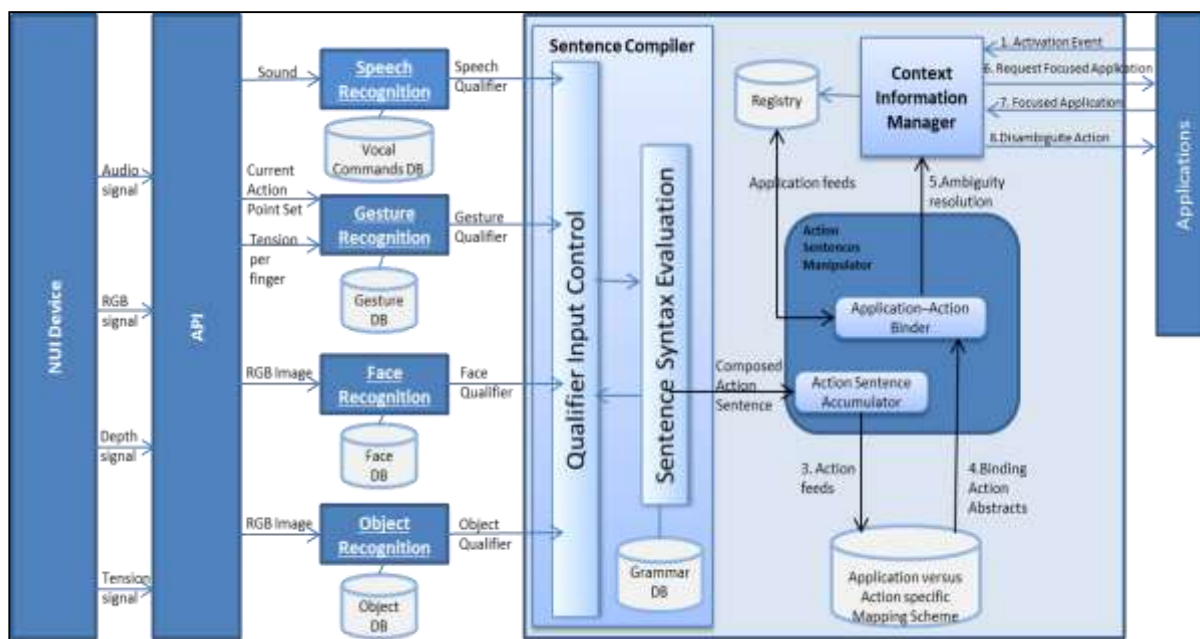
- 32-bit (x86) or 64-bit (x64) processors
- Dual-core, 2.66-GHz or faster processor
- USB 2.0 bus dedicated to the Kinect
- 2 GB of RAM
- Graphics card that supports DirectX 9.0c
- A Microsoft Kinect for Windows Sensor

4.6.1.2 *Software Requirements*

- .NET Framework 4
- Microsoft Visual Studio 2010 Express or other Visual Studio 2010 edition
- Windows 7 OS

5 Αρχιτεκτονική Συστήματος

Το σύστημα που παρουσιάζουμε είναι ένα πολυτροπικό σύστημα που βασίζεται στη φυσική αλληλεπίδραση του χρήστη, ενεργώντας σαν ένας ενδιάμεσος μεταξύ των συσκευών εισόδου και του λειτουργικού συστήματος. Η αρχιτεκτονική του αποτελείται από τρία βασικά κομμάτια τα οποία συνεργάζονται για (α) να δεχθούν δεδομένα από τις συσκευές εισόδου (input modalities), (β) να συντάσσουν τις εντολές που εντοπίστηκαν και (γ) να προσδιορίσουν ποιες εφαρμογές είναι ενεργοποιημένες και μπορούν να χρησιμοποιήσουν τις συνταγμένες εντολές. Τα κομμάτια αυτά της αρχιτεκτονικής είναι: ο “Sentence Compiler”, ο “Action Sentence Manipulator” και ο “Context Information Manager”.



Εικόνα 37 – Απεικόνιση αρχιτεκτονικής MIDA

5.1 Συσκευές και Δεδομένα Εισόδου

Ενσωματώνοντας τις αρχές της πολυτροπικότητας το σύστημά μας, υποστηρίζει διαφορετικές συσκευές εισόδου όπως : RGB φωτογραφικές μηχανές και αισθητήρες βάθους και ήχου, που καταγράφουν ένα ή περισσότερα χαρακτηριστικά της ανθρώπινης επικοινωνίας. Παράλληλα μέσω των API τους παίρνουμε τις κατάλληλες πληροφορίες ως εισόδους στους αλγόριθμους αναγνώρισης του κάθε χαρακτηριστικού. Κάθε συσκευή αντιπροσωπεύει μια

ξεχωριστή ενότητα της αρχιτεκτονικής του συστήματος μας, η οποία μπορεί να ενσωματωθεί ανεξάρτητα, χωρίς να επηρεάσει το υπόλοιπο σύστημα.

5.1.1 Αναγνώριση Ομιλίας - Speech Recognition

Στην επιστήμη υπολογιστών, η αναγνώριση ομιλίας είναι η μετάφραση των λέξεων από προφορικό σε γραπτό λόγο. Η απόδοση των συστημάτων αναγνώρισης ομιλίας συνήθως αξιολογείται όσον αφορά την ακρίβεια και την ταχύτητα. Η ακρίβεια συνήθως αξιολογείται με κάποιο ποσοστό σφάλματος λέξης (Word Error Rate - WER), ενώ η ταχύτητα μετριέται με παράγοντα τον πραγματικό χρόνο. Άλλα μέτρα της ακρίβειας περιλαμβάνουν το Single Word Error Rate – SWER και το Command Success Rate – CSR. Ωστόσο, η αναγνώριση ομιλίας (από μια μηχανή) είναι ένα πολύ πολύπλοκο πρόβλημα. Κάθε φωνή διαφέρει από την άποψη της προφοράς, της άρθρωσης, της τραχύτητας και πολλών άλλων.

Στο σύστημά μας χρησιμοποιείται το Speech Recognition API του Kinect. Τα Windows 7 και 8 χρησιμοποιούν το API των Windows. Οι διαφορές μεταξύ των δύο είναι πολύ λίγες, με σημαντικότερη ότι για το hardware του Kinect, η απόδοση του πρώτου API είναι καλύτερη.

5.1.2 Αναγνώριση Χειρονομιών - Gesture Recognition

Μια χειρονομία (Gesture) είναι ένα είδος μη-λεκτικής επικοινωνίας στην οποία κινήσεις του σώματος αποδίδουν συγκεκριμένα μηνύματα. Στην επιστήμη των υπολογιστών η αναγνώριση χειρονομιών είναι ένα θέμα με στόχο την ερμηνεία των ανθρώπινων χειρονομιών μέσω μαθηματικών αλγορίθμων. Χειρονομίες μπορεί να προέρχονται από οποιαδήποτε σωματική κίνηση ή κατάσταση, αλλά συνήθως προέρχονται από το πρόσωπο ή τα χέρια. Ο εντοπισμός και η αναγνώριση της στάσης του σώματος, το βάδισμα και οι ανθρώπινες συμπεριφορές είναι επίσης ένα θέμα των τεχνικών αναγνώρισης χειρονομιών.

Αναγνώριση χειρονομιών επιτρέπει στους ανθρώπους να επικοινωνούν με τους υπολογιστές (HCI) και να αλληλεπιδρούν φυσικά χωρίς μηχανικές συσκευές. Χρησιμοποιώντας αυτή την αναγνώριση, είναι δυνατόν κάνοντας μια χειρονομία να γίνεται κάποια απλή ή σύνθετη λειτουργία. Αυτό καθιστά περιττά συμβατικές συσκευές εισόδου, όπως ποντίκια, πληκτρολόγια, ακόμα και οθόνες αφής.

5.2 Αναγνώριση Προσώπου και Αντικειμένου - Face and Object Recognition

Το σύστημα αναγνώρισης προσώπου χρησιμοποιείται για την αυτόματη αναγνώριση ή την επαλήθευση ενός ατόμου από μια ψηφιακή εικόνα ή βίντεο. Ένας από τους τρόπους για να γίνει αυτό είναι με τη σύγκριση των επιλεγμένων χαρακτηριστικών του προσώπου από την εικόνα που είναι αποθηκευμένη στη βάση δεδομένων. Συνήθως χρησιμοποιείται σε συστήματα ασφαλείας και μπορεί να συγκριθεί με άλλες βιομετρικές μεθόδους όπως τα συστήματα αναγνώρισης ίριδας ματιού και δακτυλικών αποτυπωμάτων.

Η αναγνώριση αντικειμένων – στο computer vision [26], αυτό είναι το έργο της εξεύρεσης ενός συγκεκριμένου αντικειμένου σε εικόνες ή βίντεο. Εμείς οι άνθρωποι αναγνωρίζουμε ένα πλήθος αντικειμένων με λίγη προσπάθεια, παρά το γεγονός ότι η εικόνα των αντικειμένων μπορούν να ποικίλει αναλόγως την οπτική γωνία που βλέπετε, το μέγεθος και πολλά άλλα. Τα αντικείμενα μπορούν ακόμη και να αναγνωρίζονται ακόμη και όταν εν μέρει εμποδίζεται η όψη τους. Το έργο αυτό εξακολουθεί να αποτελεί πρόκληση για τα συστήματα όρασης υπολογιστών γενικότερα.

5.3 Sentence Compiler

Ο Sentence Compiler – SC, λαμβάνει τις εντολές που απορρέουν από τις διάφορες μορφές εισόδου. Κάθε εντολή είναι ένα δυναμικό μέρος μιας πρότασης. Ο σκοπός του SC είναι να συγκεντρώσει προτάσεις που αποτελούνται από αυτές τις εντολές.

Περιέχει δύο τμήματα, το Qualifier Input Control και το Sentence Syntax Evaluation. Το Qualifier Input Control είναι υπεύθυνο για τη συλλογή όλων των τύπων εισόδου (τον λόγο, την κίνηση, το αναγνωρισμένο πρόσωπο και το αναγνωρισμένο αντικείμενο). Η υποδοχή εισόδου είναι παράλληλη και διαδοχική. Το Sentence Syntax Evaluation αναλαμβάνει τον έλεγχο των λέξεων-εντολών που φτάνουν στον SC σαν είσοδο. Αν η λέξη που ελέγχθηκε είναι μέρος μιας προκαθορισμένης πρότασης, **που είναι αποθηκευμένη στη βάση δεδομένων**, τότε η πρόταση αποθηκεύεται σαν μη ολοκληρωμένη. Έτσι λοιπόν η ροή του προγράμματος, επιστρέφει στο Qualifier Input Control περιμένοντας την επόμενη λέξη. Αυτή η διαδικασία επαναλαμβάνεται έως ότου βρεθεί μια ολοκληρωμένη πρόταση από το Sentence Syntax Evaluation. Αν μια λέξη-εντολή δεν είναι μέρος μιας προκαθορισμένης πρότασης, σε συνδυασμό με τις προηγούμενες λέξη-εντολών που έχουν υποθεί, τότε αυτή απορρίπτεται από το σύστημα. Όταν συμπληρωθεί

και αναγνωρισθεί μια ολοκληρωμένη πρόταση, από τον SC, στέλνεται στον Action Sentence Manipulator.

5.4 Action Sentence Manipulator

Στον Action Sentence Manipulator – ASM, καταλήγουν οι ολοκληρωμένες προτάσεις που προέκυψαν από την διαδικασία που αναφέρθηκε παραπάνω. Ο ρόλος του στην ουσία είναι να χειριστεί τις ολοκληρωμένες προτάσεις, αντιμετωπίζοντάς τις πλέον σαν “actions” πάνω σε μια εφαρμογή. Συγκεκριμένα ο ASM αποτελείται από τον Action Sentence Accumulator.

Ο Action Sentence Accumulator ελέγχει την ολοκληρωμένη πρόταση που δέχθηκε από τον SC από την βάση του συστήματος. Ο έλεγχος έχει να κάνει με το αν η συγκεκριμένη πρόταση αντιστοιχεί σε μια συγκεκριμένη δράση (action), αποθηκεύοντας το όνομα της κατηγορίας που ανήκει (θα γίνει πιο κατανοητό όταν παρουσιαστεί παρακάτω η βάση του συστήματος).

Εν τέλει σε συνδυασμό με τις πληροφορίες που θα δεχτεί από τον Context Information Manager, θα πραγματοποιήσει το action σε κάποια εφαρμογή.

5.5 Context Information Manager

Ο ρόλος του Context Information Manager – CIM είναι να επικοινωνεί με το λειτουργικό σύστημα. Λειτουργεί δηλαδή σαν ενδιάμεσος μεταξύ του λειτουργικού συστήματος και της διεπαφής. Υπάρχουν τρεις μέθοδοι στην κλάση του CIM και οι έλεγχοι που γίνονται είναι οι εξής :

- Εύρεση της εφαρμογής που είναι focused.
- Τρεις διαδοχικοί έλεγχοι :
 - Έλεγχος όλων των process που είναι ενεργά και έχουν ανοιχτό παράθυρο (άρα θα βρίσκονται και στη γραμμή έναρξης).
 - Έλεγχος της βάσης δεδομένων για την εύρεση των εφαρμογών που έχουν σχέση με την ολοκληρωμένη πρόταση που έχει βγει από τον SC.
 - Έλεγχος των δικαιωμάτων του χρήστη που έχει για κάθε εφαρμογή.
 - Αποθήκευση των εφαρμογών που προκύπτουν από τους προηγούμενους ελέγχους.

- Απλός έλεγχος όλων των process που είναι ενεργά και αποθήκευση μόνο όσων έχουν ανοιχτό παράθυρο (άρα θα βρίσκονται και στη γραμμή έναρξης).

Όλες αυτές οι πληροφορίες αποθηκεύονται δυναμικά στη μνήμη. Όταν χρειαστεί μια τέτοιου είδους πληροφορία ο ASM, τότε επικοινωνεί με τον CIM.

6 Υλοποίηση

6.1 Εισαγωγή

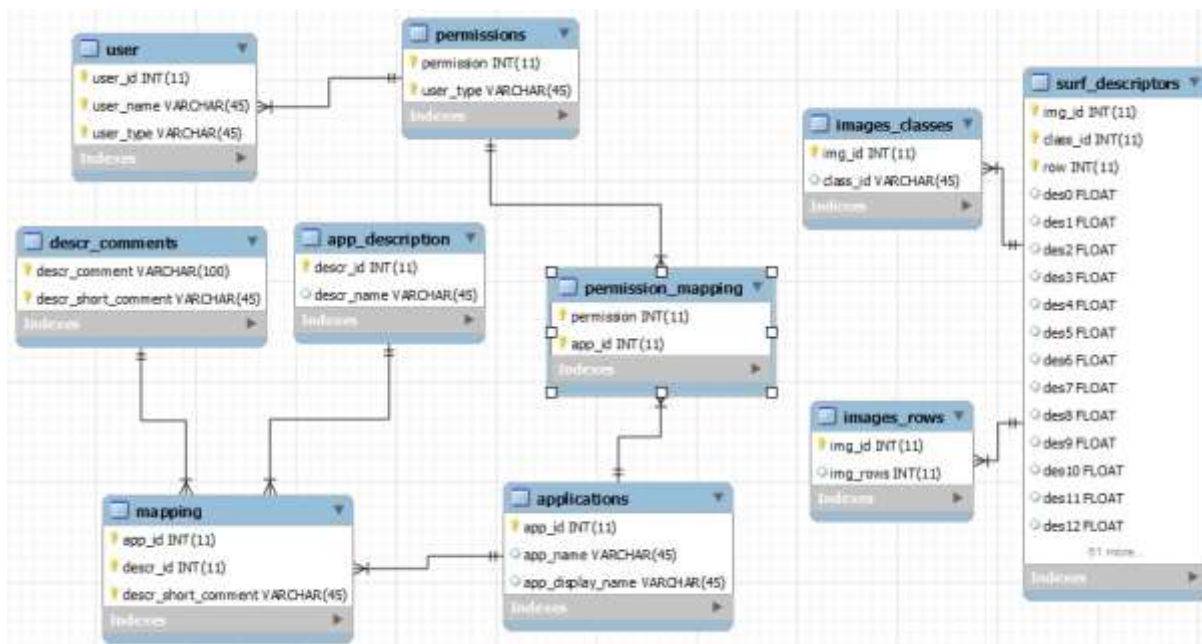
Στην παρούσα εργασία, το πρόγραμμα που αναπτύσσεται σε μια πολυτροπική φυσική διεπαφή χρήστη, βασίζεται σε δεδομένα εισόδου ήχου, βίντεο και επεξεργασία βάθους σε πραγματικό χρόνο. Η συσκευή εισόδου, μεταξύ του χρήστη και του συστήματος είναι το Microsoft Kinect και η επεξεργασία σημάτων πραγματοποιείται από το API της συσκευής. Για την υποστήριξη πολλαπλών εφαρμογών, το πρόγραμμα έχει σχεδιαστεί έτσι ώστε να τρέχει και στο παρασκήνιο χρησιμοποιώντας ως ενδιάμεσος μεταξύ των πολυτροπικών εισόδων και των ενεργών εφαρμογών που τρέχουν σε έναν υπολογιστή.

Για την επίδειξη της ιδέας παρουσιάζεται ένα σενάριο χρήσης που αποτελείται από τα εξής στάδια, δηλαδή:

1. Σύνδεση μέσω Face detection – recognition,
2. Επιλογή εφαρμογής με βάση το Object detection – recognition,
3. Έλεγχος των δικαιωμάτων του χρήστη,
4. Χρήση εφαρμογών με βάση τα βήματα 1 μέχρι 3, μέσω χειρονομιών (Gesture recognition) και φωνητικών εντολών (Speech recognition),
5. Ανανέωση, αφαίρεση και πρόσθεση ολοκληρωμένης πρότασης, στη βάση δεδομένων και στο xml της γραμματικής.

Σε αυτό το κεφάλαιο θα αναπτυχθούν αναλυτικά η βάση δεδομένων , το speech recognition και η γραμματική του και η ανάλυση των χειρονομιών που χρησιμοποιούνται στο gesture recognition .

6.2 Η Βάση δεδομένων του συστήματος



Εικόνα 38 – mydb_EER_Diagram

Οι οντότητες της βάσης είναι :

- **User** : Περιέχει το user_id, το user_name που είναι το όνομα του κάθε χρήστη και το user_type που είναι ο τύπος του χρήστη και ξένο κλειδί του πίνακα permissions.

Παράδειγμα

user_id	user_name	user_type
1	Tsampikos	User
2	Anestis	User
3	Petros	Admin

- **Permissions** : περιέχει το permission και το user_type

Παράδειγμα :

permission	user_type
0	admin
1	user

- **Applications** : περιέχει το app_id, το app_name που είναι το όνομα του process και το app_display_name που είναι το όνομα της εφαρμογής.

Παράδειγμα :

app_id	app_name	app_display_name
1	firefox	Mozilla Firefox
2	chrome	Google Chrome
3	mspaint	Microsoft Paint

- **Permission_mapping** : περιέχει το app_id που είναι ξένο κλειδί από τον πίνακα applications και το permission που είναι ξένο κλειδί από τον πίνακα permissions

Παράδειγμα :

app_id	permission
1	0
2	1
3	0

- **App_description** : περιέχει το descr_id και το descr_name που είναι η ολοκληρωμένη πρόταση από τις λέξεις που περνάνε από τον SC.

Παράδειγμα :

descr_id	descr_name
1	open up
2	close up
3	next right

- **Description_comments** : περιέχει το descr_comment που είναι η περιγραφή ενός action που στη συνέχεια περιγράφει μια ολοκληρωμένη πρόταση και το descr_short_comment που είναι το όνομα του action.

Παράδειγμα :

descr_comment	descr_short_comment
Close the application that will be choosen throw Speech Recognition.	Close
Next action is in accordance to the application choosen.	Next
Previous action is in accordance to the application choosen.	Previous

- **Mapping** : περιέχει το `app_id` που είναι ξένο κλειδί από τον πίνακα `applications`, το `descr_id` που είναι ξένο κλειδί από τον πίνακα `app_description` και το `descr_short_comment` που είναι ξένο κλειδί από τον πίνακα `description_comments`.

Παράδειγμα :

app_id	descr_id	descr_short_comment
1	2	Close
2	3	Next

Οι `Image_rows`, `image_classes` και `surf_descriptors` είναι βοηθητικοί πίνακες για το `face` και το `object recognition`.

6.3 Gesture Recognition

Οι χειρονομίες χωρίζονται σε στατικές και μη στατικές. Στην εφαρμογή αυτή χρησιμοποιήθηκαν και τα δύο είδη. Κατά τη διάρκεια της αναγνώρισης οι κινήσεις που ανιχνεύονται, μετατρέπονται σε κείμενο **και οι μη στατικές περνάνε** από τον `Sentence Compiler` για τη δημιουργία μιας πρότασης.

GESTURES		
<u>NON - STATIC</u>	<u>STATIC</u>	
Swing Right	Calibrate pose	_0_
Swing Left	Hands Up	\ 0 /
Swing Up	Hands Down	0
		/ \
Swing Down	Flying Mode	---0---
Push (in)		
Pull (out)		

Πίνακας 2 -Μη στατικές και στατικές χειρονομίες που αναγνωρίζονται από το σύστημα

Η Swing, η Push και η Pull χειρονομία είναι ουσιαστικά μία, αυτό που αλλάζει είναι η κατεύθυνση. Συγκεκριμένα ανά 10 frames ελέγχεται τη θέση του point of reference (π.χ. το joint του δεξιού καρπού του χρήστη) και αποθηκεύεται σε μια μεταβλητή Vector3D. Αρχικά ελέγχουμε την απόσταση μεταξύ των δύο σημείων χρησιμοποιώντας την μέθοδο Distance3D της βιβλιοθήκης System.Windows.Media.Media3D.

```

if (accCounter > 10)
{
    dist = Distance3D(curPointPos.X, curPointPos.Y, curPointPos.Z,
                    prevPointPos.X, prevPointPos.Y, prevPointPos.Z);

    Vector3D cur = prevPointPos - curPointPos;
}
    
```

Πίνακας 3 - Συνθήκη μέτρησης των frames και πράξεις για την απόσταση

Έπειτα αφαιρούμε το ένα σημείο από το άλλο και χρησιμοποιώντας την απόλυτη τιμή των αποτελεσμάτων ($|x|, |y|, |z|$) βρίσκουμε την κατεύθυνση της κίνησης. Ελέγχοντας την απόσταση και την κατεύθυνση της κίνησης αναγνωρίζουμε το gesture και αποστέλλουμε την κατάλληλη λέξη στον Sentence Compiler.

```

if (accCounter > 10)
    {
        dist = Distance3D(curPointPos.X, curPointPos.Y, curPointPos.Z,
            prevPointPos.X, prevPointPos.Y,
prevPointPos.Z);

        Vector3D cur = prevPointPos - curPointPos;

        if (dist > .3 && Math.Abs(cur.X) > Math.Abs(cur.Y) &&
            Math.Abs(cur.X) > Math.Abs(cur.Z))
        {
            if (cur.X < 0)
            {
                Console.WriteLine("Moved on X-Axis {0} to the Left",dist);
                NotifyPropertyChanged("left");
            }
            else
            {
                Console.WriteLine("moved on X-Axis {0} to the
Right",dist);
                NotifyPropertyChanged("right");
            }
            accCounter = 0;
            prevRightHand =
Kinect_Managers.Kinect_Device.skelMan.getScldSkeleton().ElementAt(11).Value;
            gestureCooldown();
            return;
        }
    }

```

Πίνακας 4 - Αναγνώριση του Swing Gesture και της κατεύθυνσης του

Η Calibrate pose ελέγχει πότε ο χρήστης έχει τα χέρια του όπως στην Εικόνα 24, δηλαδή να δημιουργεί ορθή γωνία με τους αγκώνες του. Αυτό επιτυγχάνεται καλώντας τη συνάρτηση GetBodySegmentAngle με τρία ορίσματα για το κάθε χέρι (τον καρπό, τον αγκώνα και τον ώμο) τα οποία είναι μεταβλητές τύπου Vector3D.

```
public double GetBodySegmentAngle(Vector3D joint1, Vector3D joint2, Vector3D
joint3)
{
    Vector3D vectorWristToElbow = (joint1 - joint2);
    Vector3D vectorElbowToShoulder = (joint2 - joint3);
    vectorWristToElbow.Normalize();
    vectorElbowToShoulder.Normalize();

    Vector3D crossProd = Vector3D.CrossProduct(vectorWristToElbow,
vectorElbowToShoulder);
    double crossProdLength = crossProd.Z;
    double dotProd = Vector3D.DotProduct(vectorWristToElbow,
vectorElbowToShoulder);
    double segmentAngle = Math.Atan2(crossProdLength, dotProd);

    // Convert the result to degrees.
    double degrees = segmentAngle * (180 / Math.PI);

    // Add the angular offset. Use modulo 360 to convert the value
calculated above to a range
    // from 0 to 360.
    degrees = (degrees % 360);

    degrees = CalculateReverseCoordinates(degrees);
    return degrees;
}
```

Πίνακας 5 - Αλγόριθμος υπολογισμού της γωνίας του αγκώνα κάθε χεριού

Στην αρχή υπολογίζονται τα διανύσματα μεταξύ καρπού – αγκώνα και αγκώνα – ώμου, στη συνέχεια το γινόμενο των διανυσμάτων [(vectorWristToElbow) x (vectorElbowToShoulder)] με τη μέθοδο CrossProduct που επιστρέφει ένα διάνυσμα Vector3D. Επειδή θα χρησιμοποιηθεί η μέθοδος Atan2 θα χρειαστούμε να για ορίσματά της δυο αριθμούς τύπου double. Έτσι λοιπόν το ένα όρισμα είναι η τιμή του διανύσματος crossProd

στον άξονα Z και η άλλη τιμή είναι πάλι το γινόμενο των διανυσμάτων που υπολογίστηκε προηγουμένως, αυτή τη φορά όμως σε μορφή αριθμού τύπου double. Με τη μέθοδο Atan2 θα επιστραφεί η γωνία της οποίας η εφαπτομένη είναι το πηλίκο των δύο συγκεκριμένων αριθμών. Τέλος με τις υπόλοιπες μετατροπές της γωνίας που υπολογίστηκε, παίρνουμε τη σωστή τιμή της γωνίας που ψάχναμε.

Οι υπόλοιπες στατικές χειρονομίες αναγνωρίζονται υπολογίζοντας την απόσταση του καρπού των χεριών από κάποιο σημείο του σώματος (skeleton joint) που είναι σταθερό. Για παράδειγμα η Hands up αναγνωρίζεται όταν η τιμή της απόστασης του διανύσματος του καρπού στον άξονα Y από το κέντρο των ώμων (shoulder center joint) στον άξονα Y, είναι μεγαλύτερη από κάποιον βολικό αριθμό (βολικό γιατί θα μπορούσε να είναι από έστω ένα cm μέχρι αρκετά cm χωρίς να υπάρχει ουσιαστική διαφορά), που καλό θα ήταν να είναι σχετικά μικρός, για να καλυφθούν περιπτώσεις χρήστη και με μεγάλα χέρια αλλά και με μικρότερα.

```
//          \0/
//   pose2   |
//          / \
// ( left && right elbow above ShoulderCenter )
//          if (elbowL.Y >
Kinect_Managers.Kinect_Device.skelMan.getScldSkeleton().ElementAt(2).Value.Y
//          &&
//          elbowR.Y >
Kinect_Managers.Kinect_Device.skelMan.getScldSkeleton().ElementAt(2).Value.Y
//          {
//          if (updistR > .5 && updistL > .5)
//          {
//          Console.WriteLine("both hands ABOVE level
position");
//          NotifyPropertyChanged("hands up");
//          accCounter1 = 0;
//          //poseCooldown();
//          return;
//          }
//          }
```

Πίνακας 6 - Υπολογισμός στατικής χειρονομίας Hands Up

6.4 Speech recognition

Για να αρχίσει η αναγνώριση της φωνής πρέπει να δημιουργηθεί μια γραμματική, η οποία θα κρατάει τις λέξεις που θέλουμε να αναγνωρίζονται, η μορφή της οποίας είναι σε ένα .xml αρχείο.

Με την κατάλληλη συγγραφή της γραμματικής οι δυνατότητες του χρήστη αυξάνονται κατά πολύ. Για παράδειγμα μια λέξη (Speech qualifier) μπορεί να αντιστοιχεί σε πολλές φωνητικές εντολές, όπως το «Yes» μπορεί να ανιχνευθεί λέγοντας «Yeah», «Yer», «Yo», «Yes». Στη δικιά μας γραμματική οι λέξεις δεν έχουν τέτοιες παραμέτρους, οπότε για να αναγνωριστούν πρέπει να υποθούν όπως είναι.

Κάθε λέξη που αναγνωρίζεται από τον αλγόριθμο της αναγνώρισης του συστήματος, συγκρίνεται με αυτές που έχουν προστεθεί στη γραμματική. Αν το αποτέλεσμα της σύγκρισης είναι θετικό ενημερώνεται ένας Event Handler και περνάνε στον Sentence Compiler για τη δημιουργία μιας πρότασης.

```
private void SreSpeechRecognized(object sender, SpeechRecognizedEventArgs e)
{
    recostring = e.Result.Text;

    if (e.Result.Confidence >= 0.30)

        NotifyPropertyChanged("recostring");
}
```

Πίνακας 7 -- Ενημέρωση Event Handler κάθε φορά που αναγνωρίζεται μια λέξη

```
<?xml version="1.0" encoding="utf-8"?>
<grammar
mode="voice"
version="1.0"
xml:lang="en-US"
root="commands"
xmlns="http://www.w3.org/2001/06/grammar">
  <rule id="commands" scope="public">
    <one-of>
      <item>
        <ruleref uri="#actions" />
      </item>
      <item>
        <ruleref uri="#numbers" />
      </item>
    </one-of>
  </rule>
  <rule id="actions">
    <one-of>
      <item> open </item>
      <item> close </item>
    </one-of>
  </rule>
  <rule id="numbers">
    <one-of>
      <item>one</item>
      <item>two</item>
      <item>three</item>
    </one-of>
  </rule>
</grammar>
```

Πίνακας 8 - Παράδειγμα γραμματικής του Speech Recognition

6.5 Περιγραφή Σεναρίου Χρήσης

Η εφαρμογή είναι υλοποιημένη σε Windows Presentation Foundation - WPF. Αποτελεί ένα μοντέλο προγραμματισμού το οποίο έχει υλοποιηθεί εξαρχής με τη λογική του πλήρους

διαχωρισμού της διεπαφής από τη λογική του προγράμματος που ονομάζεται αρχιτεκτονική Model – View – Viewmodel (MVVM). Με αυτόν τον τρόπο προγραμματιστές και σχεδιαστές συνεργάζονται με λιγότερες τεχνικές δυσκολίες.

Ξεκινώντας με το άνοιγμα της εφαρμογής γίνεται ένας έλεγχος για την εύρεση της συσκευής του Kinect. Σε περίπτωση που δεν βρεθεί συνδεδεμένο, τότε ένα pop-up παράθυρο ενημερώνει για την τρέχουσα κατάσταση τον χρήστη.



Εικόνα 39 – No sensor attached

Αφού ο έλεγχος αποβεί θετικός, με το άνοιγμα της εφαρμογής εμφανίζεται το κύριο παράθυρο της.



Εικόνα 40 – Κύριο παράθυρο εφαρμογής

Στο επάνω μέρος του παραθύρου υπάρχει ένα menu bar που δίνει τη δυνατότητα πλοήγησης αναλόγως την επιλογή που θα γίνει. Οι επιλογές θα εμφανιστούν παρακάτω, γιατί ούτως ή άλλως ο χρήστης δεν μπορεί να τις χρησιμοποιήσει χωρίς να ολοκληρώσει τα επόμενα βήματα. Το menu bar ανήκει στο κεντρικό παράθυρο ενώ από κάτω υπάρχει ένα frame που φιλοξενεί τις διάφορες σελίδες πλοήγησης.

Η πρώτη σελίδα που φορτώνεται με το άνοιγμα της εφαρμογής και του παραθύρου της είναι αυτή του Sign in και περιέχει στο πάνω μέρος της ένα κουμπί που πατώντας το αρχίζει το face και το object recognition. Σε αυτό το σημείο υπάρχουν δύο περιπτώσεις :

- Αναγνώριση κάποιου χρήστη και αναγνώριση κάποιου αντικειμένου.
- Αναγνώριση μόνο κάποιου χρήστη.

Με την αναγνώριση μόνο κάποιου αντικειμένου δεν μπορεί να γίνει είσοδος στο σύστημα και αξιοποίησή του.

Στην κάτω δεξιά γωνία βρίσκεται ένα πεδίο στο οποίο μπορεί να γραφτεί ο κωδικός για την σύνδεση στη βάση δεδομένων. (MySQL Workbench)

Με το πάτημα του κουμπιού λοιπόν, μέσα από ένα εύρος αποτελεσμάτων που φαίνεται παρακάτω, αναμένεται να επιστραφεί ένα από αυτά.



Εικόνα 41 - Τυχαίες περιπτώσεις αναγνώρισης χρήστη και αντικειμένου

Μετά την επιστροφή του πρώτου αποτελέσματος των αναγνώρισεων, σε περίπτωση που τα αποτελέσματα δεν είναι ικανοποιητικά δίνεται η ευκαιρία να επαναληφθεί η όλη διαδικασία πατώντας το κουμπί **Try Again**.

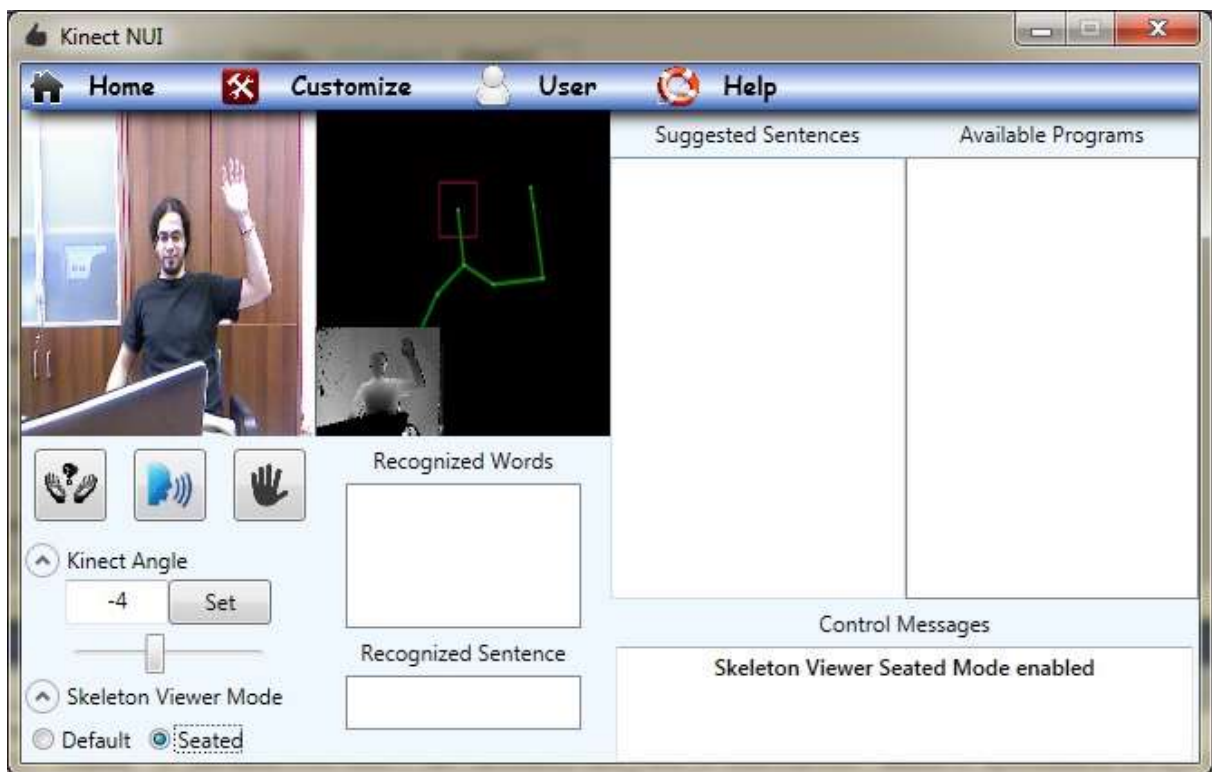
Στο κέντρο του παραθύρου δίνονται πληροφορίες όπως :

- Το όνομα του αναγνωρισμένου χρήστη και το user type του (user ή admin).
- Το είδος του αναγνωρισμένου αντικειμένου.
- Τα ονόματα των εφαρμογών που έχει δικαίωμα να χρησιμοποιήσει ο χρήστης.

Όταν τα αποτελέσματα είναι τα επιθυμητά, τότε πατώντας το κουμπί **Continue** επιτυγχάνεται είσοδος στο σύστημα και γίνεται πλοήγηση από την σελίδα του Sing in στην κεντρική σελίδα της εφαρμογής.

Στην πρώτη περίπτωση αναγνώρισης (χρήστη μαζί με αντικείμενο), αν το αντικείμενο είναι βιβλίο τότε παράλληλα με τη μετάβαση στην κεντρική σελίδα γίνεται και ενεργοποίηση του Google Chrome browser. Αν είναι παλέτα τότε ανοίγει η Ζωγραφική των Windows. Όλα όμως, αφού ελεγχθούν τα δικαιώματα του χρήστη από τη βάση δεδομένων. Στη δεύτερη περίπτωση αναγνώρισης απλά γίνεται πλοήγηση στην κεντρική σελίδα.

6.5.1 Περιγραφή User Interface



Εικόνα 42 - Κεντρική σελίδα της εφαρμογής

Το menu bar πλέον είναι προσβάσιμο από τον χρήστη.

Ανάλυση των επιλογών του menu :

- Home → επαναφέρει την κεντρική σελίδα στο προσκήνιο.
- Customize
 - Manage Sentence
 - Add Sentence → πρόσθεση πρότασης στη βάση και στη γραμματική
 - Remove Sentence → διαγραφή πρότασης από τη βάση
 - Update Sentence → ανανέωση υπάρχουσας πρότασης στη βάση και στη γραμματική
- User
 - User & Object information → Εμφάνιση στοιχείων αναγνωρισμένου χρήστη και αντικειμένου.
 - Sign in as different user → Επιστροφή στη σελίδα του Sign in και επανεκκίνηση διαδικασίας αναγνώρισης χρήστη και αντικειμένου για την είσοδο στο σύστημα.

- Help
 - Available Sentences → Εμφάνιση παραθύρου με τις υπάρχουσες ολοκληρωμένες προτάσεις και την λεπτομερή περιγραφή του action που εκτελούν.

Χωρίζοντας την σελίδα κάθετα στη μέση...

Στο αριστερό μέρος από πάνω αριστερά εμφανίζονται τα εξής :

- Δύο viewboxes όπου εμφανίζεται η εικόνα που παίρνουμε από το Kinect
 - Στο αριστερό είναι η εικόνα από την RGB κάμερα
 - Στο δεξί με το μαύρο φόντο απεικονίζεται ο Skeleton Viewer και κάτω αριστερά είναι η εικόνα από την depth κάμερα.
- Κάτω αριστερά υπάρχουν τρία κουμπιά
 - Πατώντας το αριστερό ενεργοποιείται το object recognition και απενεργοποιείται όταν επιστραφεί κάποιο αποτέλεσμα. Τη στιγμή που θα βρεθεί κάποιο αντικείμενο θα γίνει το action όπως στο Sign in, δηλαδή (user recognized + object recognized) ενεργοποιώντας την ανάλογη εφαρμογή.
 - Με το κεντρικό γίνεται εναλλαγή της κατάστασης (on/off) του Speech recognition.
 - Με το δεξί γίνεται εναλλαγή της κατάστασης (on/off) του Gesture recognition.

Ακριβώς από κάτω βρίσκονται δύο expander :

- Από το μενού του πρώτου δίνεται η δυνατότητα ρύθμισης της γωνίας κλίσης του Kinect, οι τιμές του οποίου προέρχονται από το επιταχυνσιόμετρο της συσκευής.
- Από το δεύτερο επιλέγουμε τον τρόπο εύρεσης των συνδέσμων (joints) του σώματος (βλ. σελίδα 31).

Δεξιά από τους expander υπάρχουν δύο listbox

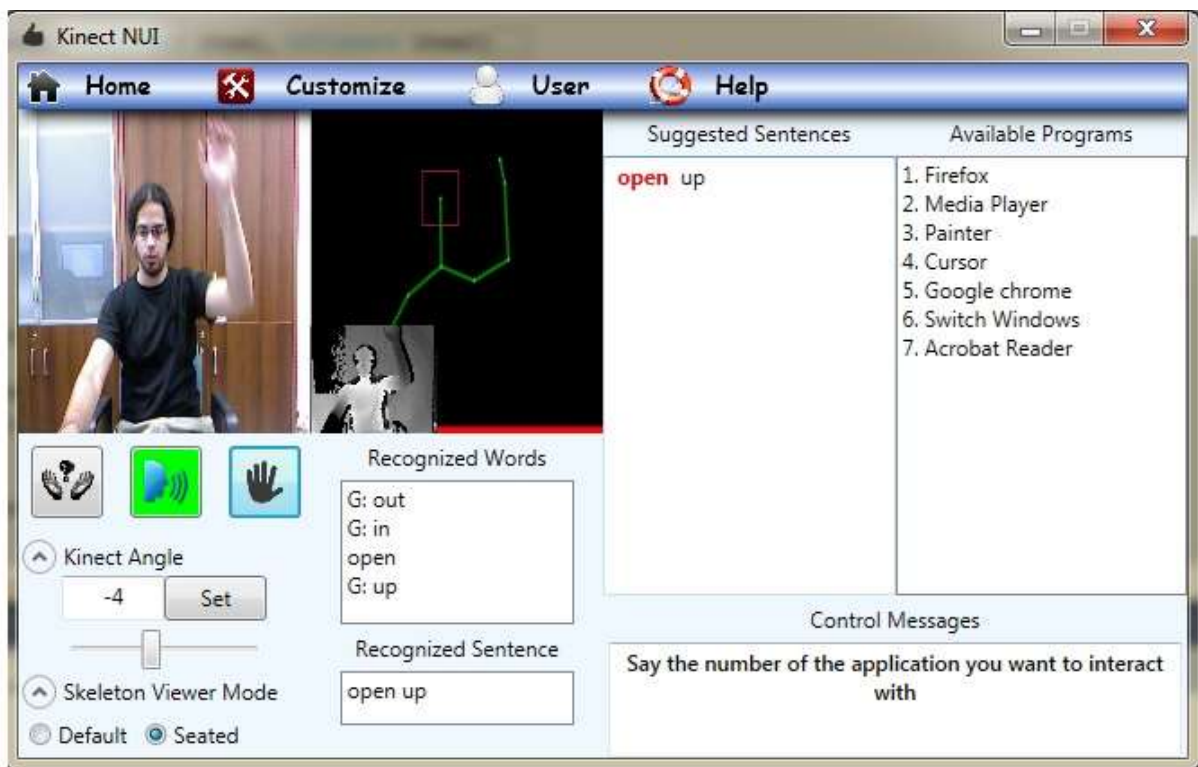
- Στο πάνω φαίνονται οι αναγνωρισμένες λέξεις που λέει ο χρήστης
- Στο κάτω εμφανίζεται η ολοκληρωμένη πρόταση, κάθε φορά που θα βρεθεί από το σύστημα.

Στο δεξί μισό της σελίδας υπάρχουν τρία listbox :

- Στο πάνω αριστερά εμφανίζονται – αφού υποθεί μια λέξη – οι ολοκληρωμένες προτάσεις που αποτελούνται από αυτή, έτσι ώστε ο χρήστης να είναι πάντα σε εγρήγορση για το ποια λέξη θα πει στη συνέχεια.
- Στο δεξί εμφανίζεται μια λίστα με τις διαθέσιμες εφαρμογές που μπορεί να κάνει κάποιο action ο χρήστης, όταν δεν υπάρχει focused εφαρμογή εκτός της υπάρχουσας ή όταν το action δεν προορίζεται για την εκάστοτε focused εφαρμογή.
- Τέλος στο κάτω μέρος εμφανίζονται βοηθητικά μηνύματα που ενημερώνουν τον χρήστη, κατά τη διάρκεια χρήσης της εφαρμογής.

6.5.2 Διάδραση με τις εφαρμογές

Ενεργοποιώντας το Speech και το Gesture recognition, ο χρήστης μπορεί να αρχίσει να λέει τις κατάλληλες λέξεις και κάνοντας τα διάφορα gesture, να δημιουργήσει μια πρόταση που θα ενεργοποιήσει ένα action.



Εικόνα 43 - Δημιουργία ολοκληρωμένης πρότασης με συνδυασμό speech & gesture recognition

Από τη στιγμή που θα δημιουργηθεί η ολοκληρωμένη πρόταση, σημαίνει ότι –όπως βλέπουμε στο σχήμα της αρχιτεκτονικής – θα περάσει από τον ASM και τον CIM. Αναλόγως τον τύπο της πρότασης θα γίνει και το κατάλληλο action.

Παραδείγματα προτάσεων :

- « open up » → περνώντας από τον ASM με τον έλεγχο στη βάση βλέπουμε ότι ανήκει στην κατηγορία Open (η αντιστοίχιση πρότασης με τον τύπο της φαίνεται στη βάση στον πίνακα mapping).

```
private void GetShortCommand(string completeSent)
{
    con = DB_Handle.OpenConnectinon(Pages.LoginPage.DB_pass);
    listScom = new List<string>();
    listdescrId = new List<int>();

    listdescrId.AddRange(DB_Handle.SelectIntCmd("SELECT descr_id FROM
mydb.app_description WHERE descr_name =' " + completeSent + "';", con));
    listScom.AddRange(DB_Handle.SelectStrCmd("SELECT descr_short_comment
FROM mydb.mapping WHERE descr_id = " + listdescrId[0] + "';", con));

    DB_Handle.CloseConnection(con);
}
```

Πίνακας 9 - Έλεγχος βάσης για τον τύπο της πρότασης

Η συγκεκριμένη πρόταση είναι η μόνη που δεν περνάει από τον CIM για το λόγο ότι οι πληροφορίες που παρέχονται από αυτόν δεν χρειάζονται ακόμα.

Στη συνέχεια ενημερώνεται ένας Event handler και έτσι στο listbox με τα **Available programs** εμφανίζονται οι εφαρμογές στις οποίες έχει δικαιώματα ο χρήστης να χειριστεί και στην προκειμένη περίπτωση να ανοίξει.

Έτσι λοιπόν πρέπει να διαλέξει μια και από τη στιγμή που θα εμφανιστεί η λίστα, η επιλογή αυτή, που γίνεται με voice command, πρέπει να είναι ένας αριθμός από αυτούς που υπάρχουν εκεί (one, two, three...) βλ. Εικόνα 43. Μετά την επιλογή, το action γίνεται στη συγκεκριμένη εφαρμογή ***.

- « close up » → περνώντας από τον ASM με τον έλεγχο στη βάση βλέπουμε ότι ανήκει στην κατηγορία Close. Στη συνέχεια καλείται ο CIM ενημερώνοντας τον ASM για όλες

τις ενεργές εφαρμογές με ανοιχτό παράθυρο (βρίσκονται στη γραμμή της έναρξης) αλλά και αυτή που είναι focused.

Αν την στιγμή που δημιουργήθηκε η ολοκληρωμένη πρόταση, υπήρχε κάποια εφαρμογή και ήταν focused τότε το action του close up θα γίνει σε αυτή.

Στην περίπτωση που δεν υπήρχε focused εφαρμογή τότε ενημερώνεται ο Event handler και στα **Available programs** εμφανίζεται η λίστα με τις εφαρμογές που επέστρεψε ο CIM. Ο χρήστης διαλέγει με voice command μια από αυτές.

- « next right » → περνώντας από τον ASM με τον έλεγχο στη βάση βλέπουμε ότι ανήκει στην κατηγορία Next. Στη συνέχεια καλείται ο CIM ενημερώνοντας τον ASM, επιστρέφοντας μια λίστα με εφαρμογές η οποία έχει προκύψει από τους εξής ελέγχους :

 - Έλεγχος και επιστροφή εφαρμογών μόνο με ανοιχτό παράθυρο.
 - Έλεγχος στη βάση των προηγούμενων εφαρμογών που έχουν σχέση με την πρόταση.
 - Έλεγχος των προηγούμενων για τα δικαιώματα του χρήστη.

Το action λοιπόν αν υπάρχει focused εφαρμογή θα γίνει εκεί, αλλιώς στα **Available programs** θα εμφανιστεί η λίστα που επέστρεψε ο CIM και ο χρήστης θα πρέπει να διαλέξει.

Για τα υπόλοιπα action, οι ενέργειες που γίνονται είναι ακριβώς οι ίδιες με αυτές του «next right» οπότε δεν υπάρχει λόγος ανάλυσής τους.

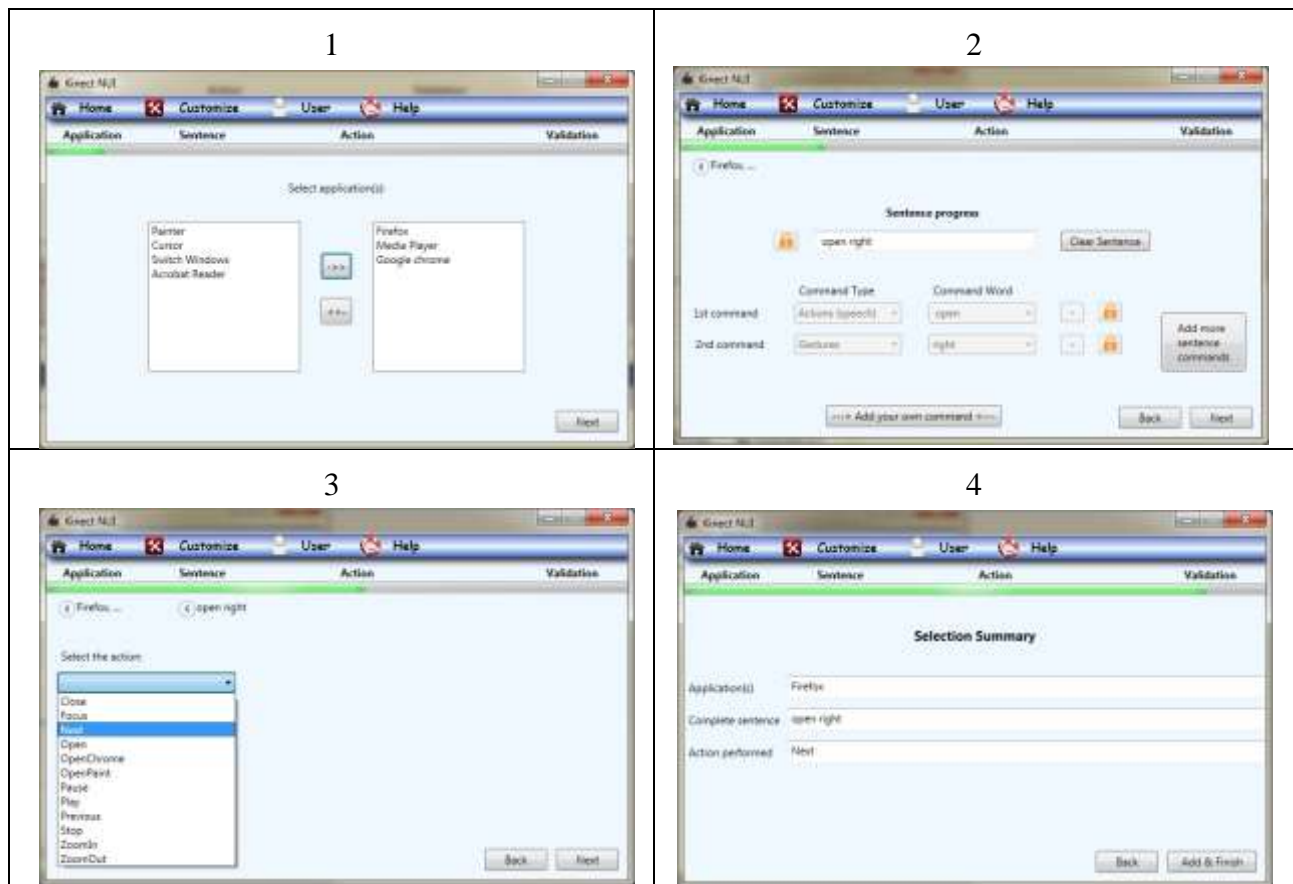
-----***-----

Αυτό που ισχύει σε κάθε περίπτωση είναι ότι, από τη στιγμή που θα εμφανιστεί η λίστα με τις διαθέσιμες εφαρμογές, αν για κάποιο λόγο η επόμενη λέξη από την ολοκληρωμένη πρόταση δεν είναι αριθμός τότε το action δεν θα γίνει, επειδή δεν θα έχει επιλεγθεί κάποια εφαρμογή και θα συνεχίσουν να αναγνωρίζονται λέξεις μέχρι να ξαναβρεθεί μια ολοκληρωμένη πρόταση.

-----***-----

6.5.3 Customize Sentence

6.5.3.1 Add sentence



Εικόνα 44 - Add Sentence procedure

- Επιλογή ενός ή πολλών εφαρμογών
- Σύνθεση νέας πρότασης
 - Επιλογή Command type (Actions ή Gestures)
 - Αναλόγως την επιλογή που έγινε, το Command word έχει διαφορετικές επιλογές
 - Πρόσθεση του διαλεγμένου Command word στην πρόταση πατώντας το κουμπί με το +. Η εξέλιξη της πρότασης φαίνεται στο textbox **Sentence progress**.
 - Πρόσθεση μέχρι και δύο ακόμα Command word στην πρόταση, πατώντας το κουμπί **Add more sentence commands**.

Παραπάνω από τρεις λέξεις, το ποσοστό σφάλματος της ολοκλήρωσης της πρότασης μεγαλώνει πολύ.

- Επιλογή τύπου προκαθορισμένου action.
- Στην τελευταία σελίδα γίνεται μια σύμπτυξη των επιλογών του χρήστη.

Πατώντας το **Add & Finish** κουμπί ενημερώνεται η βάση δεδομένων και το xml αρχείο της γραμματικής.

Κατά τη διάρκεια όλων των βημάτων γίνονται έλεγχοι για την αποφυγή σφαλμάτων όπως ίδιες εγγραφές στη βάση, η πρόταση να αποτελείται από το λιγότερο δυο λέξεις και άλλα.

Στο δεύτερο βήμα ο χρήστης, εκτός από το να συνθέσει μια πρόταση με τις λέξεις που ήδη υπάρχουν στο σύστημα, έχει τη δυνατότητα να προσθέσει δικές του λέξεις πατώντας το κουμπί **Add your own command**. Ένα pop-up παράθυρο με το μενού τις Εικόνας 44 επιτρέπει στον χρήστη να προσθέσει στο command type Actions τη λέξη.

Αφού προστεθεί ξαναφορτώνεται η σελίδα για τη σύνθεση της πρότασης και η καινούργια λέξη βρίσκεται στις επιλογές του.



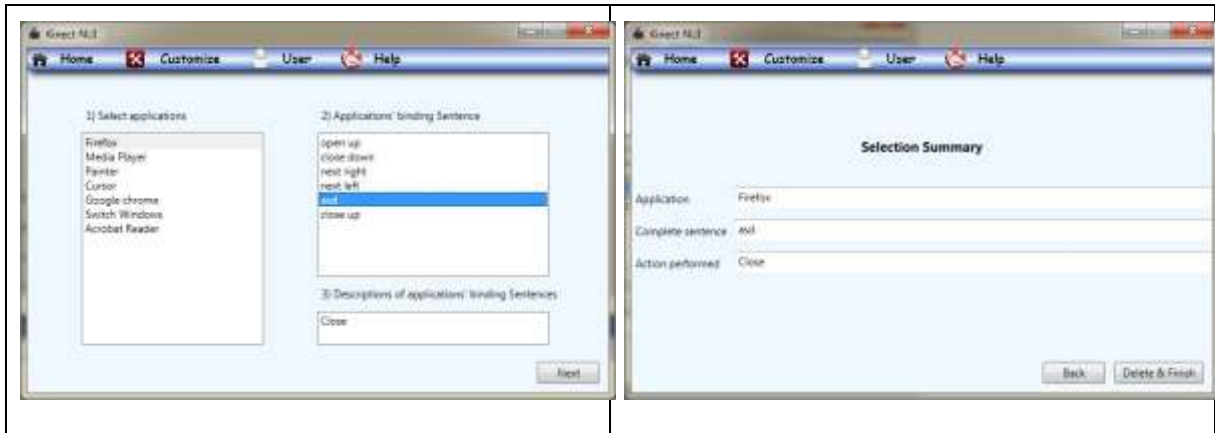
Εικόνα 45 - Add your own command

6.5.3.2 Delete Sentence

Η διαδικασία διαγραφής της πρότασης είναι αρκετά απλή.

- Πρώτα διαλέγουμε την εφαρμογή για την οποία ενδιαφερόμαστε .
- Κάθε εφαρμογή έχει προτάσεις με τις οποίες σχετίζεται και όταν είναι διαλεγμένη τότε εμφανίζονται οι προτάσεις αυτές στο διπλανό listbox. Παρομοίως κάθε πρόταση έχει έναν τύπο που την περιγράφει και όταν είναι διαλεγμένη στο επόμενο listbox εμφανίζεται ο τύπος της.

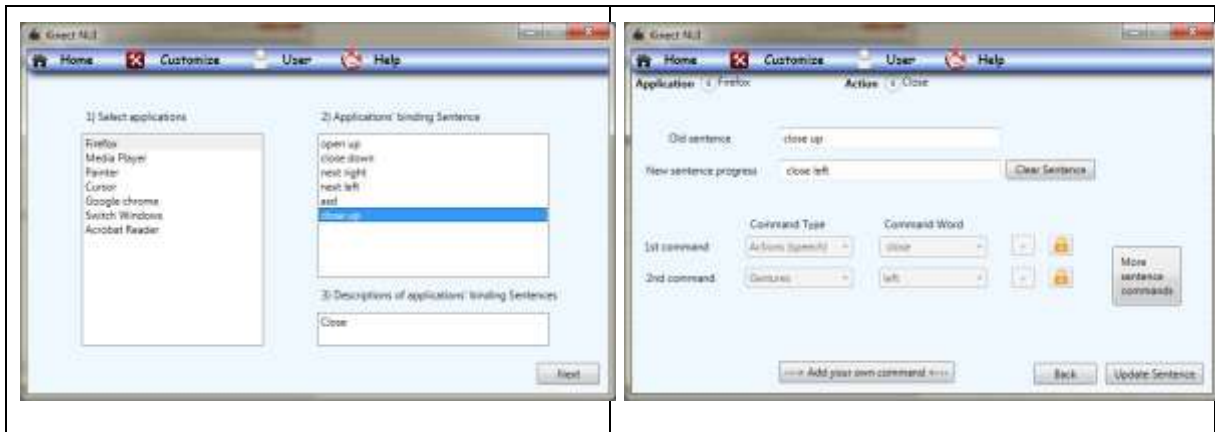
Συνεχίζοντας συνοψίζονται οι επιλογές και εν τέλει γίνεται η διαγραφή πατώντας **Delete & Finish**.



Εικόνα 46 - Delete Sentence procedure

6.5.3.3 Update Sentence

Με ακριβώς τον ίδιο τρόπο ξεκινάει η διαδικασία του update. Γίνεται επιλογή της εφαρμογής και μετά τις ολοκληρωμένες πρότασης. Στη συνέχεια πρέπει να γίνει σύνθεση της καινούργιας πρότασης διαλέγοντας πάλι από τις επιλογές των command type και command word. Σε περίπτωση που ο χρήστης θέλει να προσθέσει δική του λέξη, όπως και στο Add sentence, μπορεί να πατήσει το κουμπί «Add your own command» και να ακολουθήσει τη γνωστή διαδικασία.



Εικόνα 47 - Update Sentence procedure

7 Συμπεράσματα και Μελλοντική Εργασία

Δημιουργώντας πολυτροπικές εφαρμογές όπως αυτή που παρουσιάζεται στην παρούσα πτυχιακή, η χρήση συμβατικών συσκευών εισόδου/εξόδου τείνει να καταργηθεί, αφού η επικοινωνία μεταξύ ανθρώπου – υπολογιστή γίνεται όλο και με πιο φυσικό τρόπο, όπως για παράδειγμα με ομιλία, χειρονομίες, εκφράσεις προσώπου και συνδυασμό όλων αυτών, διευκολύνοντας σε μεγάλο βαθμό τον χρήστη ειδικά από τη στιγμή που θα εξοικειωθεί με τον τρόπο λειτουργίας της.

Στόχος ήταν ο επανασχεδιασμός και η βελτίωση του ήδη υπάρχοντος προγράμματος MIDAS (Multimodal Interface Directed by Action Sentences), χρησιμοποιώντας τον πολυαισθητήρα της Microsoft, Kinect, δημιουργώντας μια νέα διεπαφή χρήστη, με τη βοήθεια της οποίας οι δυνατότητες της εφαρμογής γίνονται πιο εμφανείς και κατανοητές από το χρήστη. Παράλληλα με τη χρήση του Context Information Manager στην αρχιτεκτονική του συστήματος και τον κατάλληλο σχεδιασμό της βάσης δεδομένων, η αλληλεπίδραση του χρήστη με τις εφαρμογές γενικεύεται έτσι ώστε ολοκληρώνοντας μια πρόταση (action), να μην απευθύνεται πάντα σε κάποια συγκεκριμένη εφαρμογή, αλλά σε ένα εύρος εφαρμογών που είναι ενεργές, πάντα σε συνδυασμό με τα δικαιώματα που του παρέχονται από τον διαχειριστή του συστήματος. Στην περίπτωση βέβαια που η εφαρμογή είναι «focused» το action πραγματοποιείται σε αυτή. Επίσης λόγω αυτής της γενίκευσης η μελλοντική επέκταση (πρόσθεση και χειρισμός καινούργιων εφαρμογών), μπορεί να γίνει χωρίς μεγάλες αλλαγές στο κομμάτι του κώδικα που είναι υπεύθυνο για αυτές τις ενέργειες.

Όσον αφορά τα σχέδια για μελλοντική εργασία, επιπλέον χαρακτηριστικά όπως η πρόσθεση καινούργιου χρήστη στο σύστημα, η βελτίωση του Sentence Compiler και της γραμματικής του Speech recognition για την επιλογή μονολεκτικής πρότασης αλλά και πολύ πιο σύνθετης, η χρήση καλύτερων αλγορίθμων αναγνώρισης προσώπου και αντικειμένου αλλά και η βελτίωση της διεπαφής χρήστη ώστε να μπορεί να φιλοξενήσει οποιαδήποτε καινούργια ιδέα, είναι μερικά από ένα μεγάλο σύνολο βελτιώσεων που θα μπορούσαν να υλοποιηθούν.

Βιβλιογραφία

- [1] N. Vidakis; M. Syntychakis; G. Triantafyllidis; D. Akoumianakis;. (2012). Multimodal Natural User Interaction for Multiple Applications: The Gesture – Voice Example
- [2] Microsoft Kinect Official WebPage <http://www.microsoft.com/en-us/kinectforwindows/>
- [3] PrimeSense Ltd, «The PrimeSensor(TM) Reference Design 1.08.»
- [4] Ευρεσιτεχνία αισθητήρα βάθους, ZALEVSKY, Z. et al.,
<http://patentscope.wipo.int/search/en/detail.jsf?docId=WO2007043036&recNum=1&maxRec=&office=&prevFilter=&sortOption=&queryString=&tab=PCT+Biblio>
- [5] K. Khoshelham. (2011). ACCURACY ANALYSIS OF KINECT DEPTH DATA
- [6] Z. S. A. M. A. e. a. Zalevsky, «Method and System for Object Reconstruction». USA Ευρεσιτεχνία 991,994, 14 Μάρτιος 2006
- [7] MICROSOFT KINECT SDK,
<http://www.microsoft.com/en-us/kinectforwindows/develop/developer-downloads.aspx>
- [8] OpenNI Nite , <http://openni.org/>
- [9] EVOLUCE SDK, <http://www.evolute.com/en/software/sdk-for-kinect.php>
- [10] OPENKINECT *Main Page*, http://openkinect.org/wiki/Main_Page
- [11] LABORATORIES, C. *About: CL NUI Platform*. Code Laboratories,
<http://codelaboratories.com/kb/nui>
- [12] Emgu CV, OpenCV wrapper http://www.emgu.com/wiki/index.php/Main_Page
- [13] Asus - Xtion Pro Official WebPage http://www.asus.com/Multimedia/Xtion_PRO/
- [14] Asus - Xtion Pro Live Official WebPage
http://www.asus.com/Multimedia/Xtion_PRO_LIVE/
- [15] Leap Motion - The Leap <https://www.leapmotion.com/>
- [16] Airspace download center <https://www.leapmotion.com/apps>
- [17] Ergoweb. “What is Cognitive Ergonomics?” Ergoweb.com Retrieved August 29, 2011.
- [18] "NRC: Background on the Three Mile Island Accident". Nrc.gov. Retrieved August 29, 2011. <http://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.html>
- [19] A few information about Smalltalk-80, <http://en.wikipedia.org/wiki/Smalltalk>
- [20] Intelligent Image Processing, John Wiley and Sons, 2001

- [21] Moore Christian (2006-07-15). "New Community Open". NUI Group Community.
- [22] de los Reyes, August (2008-09-25). "Predicting the Past". Web Directions South 2008. Sydney Convention Centre: Web Direct
- [23] Integrated Development Environment (IDE) , http://en.wikipedia.org/wiki/Integrated_development_environment
- [24] World Wide Web Consortium (W3C) , <http://en.wikipedia.org/wiki/W3C>
- [25] Karli Watson , Beginning Visual C# 2010
- [26] Computer Vision, http://en.wikipedia.org/wiki/Computer_vision
- [27] Vlasopoulos Anastasios, Ανάπτυξη ενός Qualifier Input Control με τη χρήση του Microsoft Kinect, 2013
- [28] Akoumianakis Dimosthenis, Διεπαφή Χρήστη – Υπολογιστή μια σύγχρονη προσέγγιση, 2006
- [29] Nikolas Vidakis, Anastasios Vlasopoulos, Tsampikos Kounalakis, Petros Varchalamas, Michalis Dimitriou, Gregory Kalliatakis, Efthimios Syntychakis, John Christofakis and Georgios Triantafyllidis , Multimodal Desktop Interaction: The Face – Object – Gesture – Voice Example (2013)
- [30] Wikipedia - http://en.wikipedia.org/wiki/Main_Page

Αναφορά στις πηγές των εικόνων

- I. pccure.eu
- II. msdn.microsoft.com/en-us/library/dn188670.aspx
- III. social.msdn.microsoft.com/Forums/en-US/456146ff-7415-470c-9408-9f83a932152b/about-kinect-inner-structure-microphone-array
- IV. msdn.microsoft.com/en-us/library/jj663790.aspx
- V. en.wikipedia.org/wiki/Light-emitting_diode
- VI. el.orphek.com/understanding-calcium-and-alkalinity
- VII. en.wikipedia.org/wiki/Common_Language_Runtime
- VIII.