



Σχολή Τεχνολογικών Εφαρμογών Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων



Πτυχιακή Εργασία

Ανάπτυξη ενός Qualifier Input Control με τη χρήση του Microsoft Kinect Sensor

Βλασόπουλος Αναστάσιος (ΑΜ: 1281)

Επιβλέπων καθηγητής: Νικόλαος Βιδάκης

Επιτροπή Αξιολόγησης:

Ημερομηνία παρουσίασης:

Ευχαριστίες.

Ευχαριστώ τους καθηγητές κ. Βιδάκη και κ. Τριανταφυλλίδη για την στήριξη και καθοδήγησή τους στα πλαίσια αυτής της πτυχιακής εργασίας. Επίσης ένα μεγάλο ευχαριστώ στους Πέτρο Βαρχαλαμά και Γιάννη Χριστοφάκη για την συνεργασία που είχαμε και γενικά σε όλα τα μέλη του εργαστηρίου.

Abstract

This thesis addresses the issue of human - computer interaction and presents the design, development and implementation of a multimodal natural user interface system. The development of new interaction techniques in recent years call for a change in the development and design methodology of interfaces to new and innovative techniques. The main purpose of this project was to familiarize the user in these new types of interactions which are evolving rapidly.

Beside that a second purpose of this project was the improvement of a multimodal natural user interface which name is MIDAS. This system refers to the natural interaction between a user - person and a computer machine. Microsoft Kinect, the Microsoft's multi-sensor, was used as input device of this system which has been developed in programming language C#. Also, Microsoft SDK for Kinect was studied to take all the advantages of the three basic Kinect functions , Speech-Recognition – Gesture Recognition –Skeleton Tracking.

Keywords:

Kinect, Multimodal Interaction, Natural User Interface, Gesture, Speech Recognition, Skeleton Tracking.

Περίληψη

Η παρούσα εργασία ασχολείται με το θέμα της αλληλεπίδρασης ανθρώπου - υπολογιστή και παρουσιάζει το σχεδιασμό, την ανάπτυξη και την υλοποίηση μιας φυσικής διεπαφής πολυτροπικού τύπου. Η ανάπτυξη των νέων τεχνικών αλληλεπίδρασης τα τελευταία χρόνια καλούν για μια αλλαγή στην ανάπτυξη και τη μεθοδολογία σχεδιασμού των διεπαφών με νέες και καινοτόμες τεχνικές. Ο κύριος σκοπός του λογισμικού που αναπτύχθηκε ήταν η εξοικείωση του χρήστη με αυτά τα νέα είδη των αλληλεπιδράσεων που εξελίσσονται με ταχείς ρυθμούς.

Παράλληλα ένας δεύτερος στόχος της παρούσας εργασίας ήταν η βελτίωση και η υλοποίηση ενός πολυτροπικού συστήματος που το όνομά του είναι MIDAS. Αυτό το σύστημα αναφέρεται στη φυσική αλληλεπίδραση μεταξύ ενός χρήστη - πρόσωπο και μιας μηχανής υπολογιστή. Το Microsoft Kinect, ένας πολύ-αισθητήρας της Microsoft, χρησιμοποιήθηκε ως συσκευή εισόδου του εν λόγω συστήματος το οποίο έχει αναπτυχθεί σε γλώσσα προγραμματισμού C#. Επίσης, μελετήθηκε το Microsoft Kinect SDK ούτως ώστε να αποκομιστούν όλα τα πλεονεκτήματα των τριών βασικών λειτουργιών του Kinect, Speech Recognition- Gesture Recognition- Skeleton Tracking.

Περιεχόμενα

1	Κεφάλαιο Εισαγωγή.....	10
1.1	Εισαγωγή	11
1.2	Δομή εργασίας.....	11
2	Κεφάλαιο Φυσική διεπαφή χρήστη υπολογιστή (Natural User Interface). 12	
2.1	Εισαγωγή κεφαλαίου.	13
2.2	Η φυσική διεπαφή στην επιστήμη των υπολογιστών.....	13
2.3	Τι σημαίνει ο όρος φυσική διεπαφή:	16
2.4	Επεξήγηση των έμφυτων ικανοτήτων	16
2.5	Η ικανότητα στη μάθηση:.....	16
3	Κεφάλαιο Φυσική διεπαφή χρήστη υπολογιστή (Natural User Interface). 18	
3.1	Εισαγωγή κεφαλαίου.	19
3.2	Πολυτροπική είσοδος και έξοδος	19
3.3	Πολυτροπική έξοδος.....	19
4	Κεφάλαιο Συσκευές εισόδου φυσικών διεπαφών.....	22
4.1	Εισαγωγή κεφαλαίου.	23
4.2	Στερεοσκοπικές κάμερες	23
4.3	Κάμερες βάθους.....	23
4.4	Αισθητήρες Αφής	24
4.5	Asus WA Vi Xtion (Αισθητήρες βάθους).....	26
4.6	PrimeSense (Πολύ-αισθητήρας).....	27
4.7	Kinect (Πολύ-αισθητήρας).....	28
4.7.1	Πιο αναλυτικά για το Kinect	28
4.7.2	Ανάλυση μέτρησης του βάθους.....	29
4.8	Τεχνολογία που επιλέχθηκε.....	31
5	Κεφάλαιο Χειρονομίες (Gestures).....	32
5.1	Χειρονομίες (Gestures).....	33
5.2	Προτυποποίηση των χειρονομιών	33

5.3	Αρχές σχεδίασης διεπαφών με χειρονομίες.....	34
5.3.1	Το άγγιγμα του Μίδα.....	34
5.3.2	Πολιτισμικά θέματα	34
5.3.3	Σειρά εκτέλεσης επιμέρους ενεργειών	34
5.3.4	Σχεδίαση με βάση το πλαίσιο χρήσης.....	35
5.3.5	Ανάδραση.....	35
5.3.6	Ολοκλήρωση και Αποσαφήνιση	35
5.4	Ταξινόμηση χειρονομιών.....	35
5.5	Χειρονομίες που επιλέχθηκαν	36
6	Κεφάλαιο Τεχνολογίες και εργαλεία υλοποίησης.....	38
6.1	Εισαγωγή κεφαλαίου.....	39
6.2	Microsoft VisualC#	39
6.3	MySQL	40
6.4	Microsoft SDK.....	40
6.4.1	Χαρακτηριστικά του Kinect SDK.....	41
6.4.2	Ανάλυση της λειτουργίας Skeleton Tracking.....	42
7	Κεφάλαιο Προσδιορισμό Συστήματος	44
7.1	Εισαγωγή Κεφαλαίου	45
7.2	Ανάλυση της αρχιτεκτονικής:	45
7.2.1	NUI Devices-APIs	46
7.2.2	Speech Recognition.....	46
7.2.3	Gesture Recognition (Αναγνώριση Χειρονομιών).....	46
7.2.4	Facial Expressions Recognition	47
7.2.5	Sentence Compiler	47
7.2.6	Action Sentence Manipulator	48
7.2.7	Context Information Manager (CIM).....	48
8	Κεφάλαιο Σχεδιασμός Συστήματος.....	50
8.1	Εισαγωγή κεφαλαίου	51
8.2	Σχεσιακή Βάση Δεδομένων.....	51
8.2.1	Μοντέλο οντοτήτων συσχετίσεων.....	52
8.3	Διάγραμμα κλάσεων.....	52
9	Κεφάλαιο Υλοποίηση	54

9.1	Εισαγωγή	55
9.2	Το κυρίως παράθυρο της εφαρμογής:	55
9.2.1	Listboxes	57
9.2.2	Οι ρυθμίσεις της εφαρμογής.....	58
10	Κεφάλαιο Περιπτώσεις Χρήσης (Use Cases).....	60
10.1	Εισαγωγή κεφαλαίου	61
10.2	Σενάριο 1:Χειρισμός της εφαρμογής Skype	61
10.2.1	Ανάλυση σεναρίου Skype	61
10.3	Σενάριο 2:Χειρισμός της εφαρμογής Media Player	66
10.3.1	Εργαλεία υλοποίησης σεναρίου Media Player.....	66
10.4	Ανάλυση σεναρίου Media Player	67
11	Επίλογος.....	70
11.1	Εισαγωγή Κεφαλαίου	71
11.2	Σύνοψη και συμπεράσματα	71
11.3	Μελλοντικές Επεκτάσεις.....	71
12	Λεξικό.....	72
13	Αναφορές.....	73

Πίνακας περιεχομένων εικόνων

Εικόνα 1 Επιτραπέζια οθόνη αφής.....	24
Εικόνα 2 Η συσκευή SUR40.....	25
Εικόνα 3 Ο αισθητήρας βάθους Xtion της Asus.....	26
Εικόνα 4 Απεικόνιση των επιμέρους τμημάτων του αισθητήρα του PrimeSense.....	27
Εικόνα 5. Απεικόνιση του Kinect και των αισθητήρων του.....	28
Εικόνα 6. Απεικόνιση ενός μοτίβου στιγμάτων (dot pattern) [10].....	30
Εικόνα 7. Αριστερά η την εικόνα με το μοτίβο στιγμάτων υπέρυθρης.....	31
Εικόνα 8. Εδώ απεικονίζεται μία χειρονομία ελέγχου για τον έλεγχο ενός avatar.....	36
Εικόνα 9. Απεικόνιση της λειτουργίας του CLR.....	40
Εικόνα 10. Απεικόνιση των είκοσι αρθρώσεων που αναγνωρίζει το Kinect.....	42
Εικόνα 11. Κάθετο πεδίο αναγνώρισης του Kinect.....	43
Εικόνα 12. Απεικόνιση αρχιτεκτονικής MIDA.....	45
Εικόνα 13. Απεικόνιση του Sentence Compiler.....	47
Εικόνα 14. Απεικόνιση του μοντέλου οντοτήτων συσχετίσεων της βάσης.....	51
Εικόνα 15. Οι κλάσεις που προέκυψαν από την αρχιτεκτονική του MIDAS.....	52
Εικόνα 16. Απεικόνιση του Main Window της εφαρμογής.....	55
Εικόνα 17. Use case diagram για το σενάριο Skype.....	62
Εικόνα 18. Καταγραφή των απαιτούμενων εισόδων για κάθε ενέργεια του σεναρίου.....	62
Εικόνα 19. Εκκίνηση της εφαρμογής Skype μέσω της φωνητικής εντολής openSkype.....	63
Εικόνα 20. Απεικόνιση του βοηθητικού παραθύρου με τις διαθέσιμες επαφές για κλήση.....	64
Εικόνα 21. Η εντολή Skype left και το αποτέλεσμα της.....	64
Εικόνα 22. Πραγματοποίηση κλήσης με φωνητική εντολή.....	65
Εικόνα 23. Άνοιγμα του φακέλου My Documents με φωνητική εντολή.....	65
Εικόνα 24. Το εργαλείο Spy++ καθώς ανιχνεύει το πάτημα ενός κουμπιού του πληκτρολογίου.....	66
Εικόνα 25. Καταγραφή των απαιτούμενων εισόδων για κάθε ενέργεια του σεναρίου.....	68
Εικόνα 26. Use case diagram για το σενάριο Media Player.....	68
Εικόνα 27. Η εντολή open player και το αποτέλεσμα της.....	69
Εικόνα 28. Απεικόνιση της εντολής player play και το αποτέλεσμα της.....	69

1 Κεφάλαιο

Εισαγωγή

Ενότητες κεφαλαίου.

1.1	Εισαγωγή	11
1.2	Δομή εργασίας	11

1.1 Εισαγωγή

Η αλληλεπίδραση ανθρώπου-μηχανής είναι το επιστημονικό πεδίο που μελετά την αλληλεπίδραση μεταξύ ανθρώπων και υπολογιστικών μηχανών. Θεωρείται ως το σημείο τομής μεταξύ της τεχνολογίας, της πληροφορικής και της ανθρώπινης συμπεριφοράς. Η αλληλεπίδραση μεταξύ χρηστών και υπολογιστών γίνεται στο επίπεδο της διαπαφής χρήστη, μέσω κατάλληλου λογισμικού και υλικού.

Από την κατασκευή των πρώτων συσκευών εισόδου (input devices) για υπολογιστές, όπως το ποντίκι και το πληκτρολόγιο, δυστυχώς δεν έχει εξελιχθεί πολύ ο τομέας εισόδου δεδομένων σε υπολογιστικά συστήματα, αφού μόνο τα τελευταία χρόνια έχουν κατασκευαστεί κάποιες καινοτόμες συσκευές, που βέβαια δεν έχουν αντικαταστήσει τις παλαιές κλασσικές μεθόδους.

Όπως είναι φυσικό, ο άνθρωπος έχει την τάση να προσπαθεί να προσαρμόσει τις μηχανές στη δική του φύση και κοινωνική συμπεριφορά με αποτέλεσμα να είναι όλο και πιο απαιτητικός από πλευράς εξέλιξης των μεθόδων επικοινωνίας ώστε να νοιώθει την απόλυτη εξοικείωση με αυτές. Γίνονται λοιπόν συνεχείς έρευνες και μελέτες πάνω στην τελειοποίηση ψηφιακών αισθητήρων που να παρέχουν μεγαλύτερη ακρίβεια και αποτελεσματικότητα στην αντίληψη των μηχανών είτε με Ψηφιακή Επεξεργασία Εικόνας, Βίντεο και άλλων Πολυμέσων, είτε με εκπομπή ιδανικών ακτινών για την κατανόηση του φυσικού τρισδιάστατου χώρου, είτε αισθητήρων που μετρούν άλλα φυσικά μεγέθη με αποτέλεσμα τη συνεχή βελτίωση σφαλμάτων αλλά και μεθόδων συλλογής και εισαγωγής δεδομένων.

Ένας συνδυασμός λοιπόν πρωτοποριακών αισθητήρων σε συνεργασία με την ανάπτυξη των απαραίτητων λογισμικών με υψηλή τεχνητή νοημοσύνη θα έδινε μια νέα πνοή στα όσα μέχρι τώρα γνωρίζουμε για την αλληλεπίδραση ανθρώπου μηχανής.

1.2 Δομή εργασίας

Η παρούσα πτυχιακή εργασία περιλαμβάνει ανασκόπηση των φυσικών διαπαφών χρήστη, τη εξέλιξή τους στην διάρκεια του χρόνου και το πώς αντιμετωπίζονται σήμερα. Αναλύονται επίσης οι πολυτροπικές διαπαφές, οι οποίες στο μέλλον θα παίξουν καθοριστικό ρόλο στην επικοινωνία μας με τους υπολογιστές.

Παράλληλα, ερευνήθηκαν οι συσκευές εισόδου σε φυσικά συστήματα διαπαφών όπως είναι οι στερεοσκοπικές κάμερες ή οι αισθητήρες βάθους και δόθηκε έμφαση στον πολύ-τροπικό αισθητήρα της Microsoft, το Kinect.

Στη συνέχεια αναλύθηκε διεξοδικά ο όρος χειρονομία από την σκοπιά της φυσικής αλληλεπίδρασης του χρήστη με τον υπολογιστή. Γίνεται μια ταξινόμηση των χειρονομιών και μελέτη του τρόπου σχεδιασμού αυτών.

Στις τελευταίες ενότητες αναλύονται οι τεχνικές υλοποίησης ενός πολυτροπικού συστήματος φυσικής διαπαφής με τις τεχνολογίες που είναι σήμερα διαθέσιμες. Αυτές είναι το SDK for Kinect, η γλώσσα προγραμματισμού C# και το σύστημα διαχείρισης βάσεων δεδομένων MySQL. Γίνεται επίσης παράθεση και ανάλυση του συστήματος και η παρουσίαση δύο σεναρίων χρήσης αυτού.

2 Κεφάλαιο

Φυσική διεπαφή χρήστη υπολογιστή (Natural User Interface)

Ενότητες Κεφαλαίου

2.1	Εισαγωγή κεφαλαίου.....	13
2.2	Η φυσική διεπαφή στην επιστήμη των υπολογιστών.....	13
2.3	Τι σημαίνει ο όρος φυσική διεπαφή:.....	16
2.4	Επεξήγηση των έμφυτων ικανοτήτων.....	16
2.5	Η ικανότητα στη μάθηση:.....	16

2.1 Εισαγωγή κεφαλαίου.

Στις δεκαετίες του 1970, '80 και του '90 ο Steve Mann αναπτύξει μια σειρά από στρατηγικές user-interface 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), "Άμεση Διεπαφή Χρήστη" (Direct User Interfaces), και Μεταφορικά –Ελεύθερος Υπολογισμός (Metaphor –Free Computing) [1]. 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.

Έπειτα ακολούθησαν πολλά χρόνια εξέλιξης, πάνω από 25, μέχρι να έρθουν οι φυσικές διεπαφές στο επίπεδο των της καθημερινών χρηστών. Ο λόγος που υπήρξε αυτή η καθυστέρηση εστιάζεται κυρίως στη δυσκολία κατασκευής κατάλληλων και αξιόπιστων συσκευών εισόδου για να υποστηρίξουν τέτοιες διεπαφές. Επίσης, στις προηγούμενες δεκαετίες υπήρχε μία έλλειψη οικειότητας του χρήστη με τέτοιου είδους διεπαφές με τον υπολογιστή.

2.2 Η φυσική διεπαφή στην επιστήμη των υπολογιστών

Ο όρος διεπαφή χρήστη (user interface) 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 πίσω από την άλλη μεριά ο άνθρωπος. Έτσι λοιπόν η διεπαφή περιέχει στοιχεία που είναι τμήματα τόσο του υλικού του συστήματος, όσο και του λογισμικού που τρέχει σε αυτό.

Ως στοιχεία του υλικού του συστήματος που περιλαμβάνονται στη διεπαφή χρήστη μπορούν να είναι μια οθόνη επαφής, μια φωτογραφίδα ή ένα ποντίκι. Μέρη του λογισμικού της διεπαφής χρήστη είναι, για παράδειγμα, τα μηνύματα λάθους, τα ηχητικά μηνύματα, τα εργαλεία πλοήγησης, εικόνες σύμβολα και αντικείμενα πάνω στην οθόνη, κάθε τι που διαθέτει το λογισμικό σαν στοιχείο αλληλεπίδρασης του συστήματος με το χρήστη. Με άλλα λόγια, ο όρος σε ότι αφορά τη λογισμική του υπόστασης, σημαίνει ένα σύνολο από οπτικές και ακουστικές παραμέτρους, που παρέχει ο υπολογιστής προς το χρήστη, μέσω του εκάστοτε εκτελούμενου προγράμματος, με σκοπό την καλύτερη επικοινωνία και συνεργασία μεταξύ ανθρώπου και μηχανής. Μπορούμε να παρομοιάσουμε τη διεπαφή χρήστη σαν κανάλι επικοινωνίας μεταξύ χρήστη και υπολογιστή.

Μέσα στον όρο επικοινωνία με το χρήστη περιλαμβάνεται η έννοια της αλληλεπίδρασης αυτού με το αντικείμενο. Τι μπορεί να κάνει ο χρήστης με το αντικείμενο, και τι μπορεί να κάνει το αντικείμενο για το χρήστη, είναι μέρη της επικοινωνίας αυτής. Η χρησιμότητα ή μη ενός αντικειμένου εξαρτάται πολλές φορές από την ικανότητα του καταναλωτή να το χειριστεί με επιτυχία. Εάν ένας χρήστης δεν μπορεί να εξακριβώσει πως λειτουργεί ένα αντικείμενο, τότε η επικοινωνία του με αυτό είναι ελλιπής και η χρηστικότητα του αντικειμένου θεωρείται αυτόματα αποτυχημένη.

Εδώ και πολλά χρόνια η αλληλεπίδραση ανθρώπου-υπολογιστή γίνεται μέσω εργαλείων όπως το ποντίκι, το πληκτρολόγιο ή μιας γραφίδας. Η φυσική διεπαφή χρήστη (NUI), τα τελευταία χρόνια, έρχεται να καταργήσει αυτές τις συσκευές ως ενδιάμεσες και βάζει τις φυσικές λειτουργίες του ανθρώπου, όπως είναι η ομιλία η κίνηση ακόμα και η σκέψη, στο προσκήνιο σαν άμεσο τρόπο αλληλεπίδρασης του χρήστη με τον υπολογιστή. Η νέα γενιά διεπαφών είναι αναμφισβήτητη η φυσική διεπαφή.

Στην επιστήμη των υπολογιστών, μια φυσική διεπαφή χρήστη, ή NUI είναι ο όρος που χρησιμοποιείται από τους σχεδιαστές και προγραμματιστές των διεπαφών ανθρώπου-μηχανής, για να αναφερθούν σε μια διεπαφή χρήστη που είναι καταρχάς ουσιαστικά αόρατο, ή που γίνεται αόρατο με τις διαδοχικές αλληλεπιδράσεις με τους χρήστες του, και κατά δεύτερον βασίζεται σε φυσικά στοιχεία. Η λέξη natural χρησιμοποιείται επειδή οι περισσότερες διεπαφές υπολογιστών χρησιμοποιούν συσκευές τεχνητού ελέγχου (artificial control devices) των οποίων η λειτουργίες θα πρέπει να εκπαιδευτούν. Η διασύνδεση απαιτεί μάθηση, η οποία μάθηση διευκολύνεται μέσω του σχεδιασμού που δίνει στον χρήστη την αίσθηση ότι είναι άμεση και συνεχώς επιτυχής η επικοινωνία. Έτσι, το «φυσικό» έχει σαν στόχο την εμπειρία του χρήστη - ότι δηλαδή η αλληλεπίδραση έρχεται φυσικά, όσο διαρκεί η αυτή και ότι η ίδια η διεπαφή είναι φυσική.

Τα NUIs αντιπροσωπεύουν μια ριζική αλλαγή στον τρόπο που οι άνθρωποι μπορούν να αλληλεπιδρούν με την τεχνολογία. Ο όρος φυσικές διεπαφές χρήστη περιγράφει στην πραγματικότητα μια ευρεία κατηγορία από τεχνολογίες που είναι ίσως πιο εύκολα αναγνωρίσιμες από τις παραδοσιακές μεθόδους εισαγωγής, συμπεριλαμβανομένων ποντίκια, πληκτρολόγια, και διάφορων ελεγκτών.

Ως στόχο έχουν να αφαιρέσουν τα ψυχικά και σωματικά εμπόδια για την τεχνολογία, ούτως ώστε να κάνουν τη χρήση των υπολογιστών πιο διαισθητική, και να επεκτείνουν τους τρόπους με τους οποίους οι χρήστες μπορούν να αλληλεπιδράσουν με την τεχνολογία.

Ένας σύντομος ορισμός της φυσικής διεπαφής χρήστη είναι ο εξής σύμφωνα με τον Joshua Blake [2]:

"Μια φυσική διεπαφή χρήστη, είναι μια διεπαφή η οποία έχει σχεδιαστεί για την επαναχρησιμοποίηση των υπάρχουσών δεξιοτήτων του χρήστη και έχει σαν σκοπό την κατάλληλη αλληλεπίδραση με το περιεχόμενο. "

Πρώτον, ο ορισμός αυτός μας λέει ότι τα φυσικά περιβάλλοντα χρήστη χρειάζονται σχεδιασμό, πράγμα που σημαίνει ότι απαιτούν σύνεση και συγκεκριμένες προσπάθειες σχεδιασμού εκ των προτέρων. Απαιτείται ιδιαίτερη προσοχή ούτως ώστε να εξασφαλιστεί η καταλληλότητα των αλληλεπιδράσεων για το χρήστη, την εργασία που θέλουμε να εκτελέσουμε, τη συσκευή, και το περιβαλλοντικό πλαίσιο στο οποίο αυτές εμφανίζονται.

Τίποτα δεν πρέπει να γίνεται τυχαία σε ένα τέτοιο σχεδιασμό. Θα πρέπει να αναγνωριστεί ο ρόλος που οι σχεδιαστές χρειάζεται να παίξουν και να βεβαιωθεί ότι η διαδικασία σχεδιασμού της διεπαφής έχει την ίδια προτεραιότητα με αυτή της ανάπτυξης.

Δεύτερον, η φράση "επαναχρησιμοποίηση των υπάρχουσών δεξιοτήτων" εστιάζει κυρίως στο πώς οι διεπαφές που δημιουργούνται θα είναι φυσικές και πλαισιώνει το σχεδιασμό γύρω από τις ανθρώπινες ικανότητες. Εδώ είναι όπου το "φυσικό" έρχεται στο προσκήνιο για να παίξει το ρόλο του. Οι χρήστες είναι έμπειροι σε πολλές δεξιότητες που έχουν αποκτήσει μόνο και μόνο επειδή είναι άνθρωποι. Έχουν εξασκήσει για χρόνια τις δεξιότητές τους, τόσο τις λεκτικές όσο και μη λεκτικές, για την επικοινωνία μεταξύ ανθρώπου με άνθρωπο καθώς και τις δεξιότητές τους για την αλληλεπίδραση ανθρώπου-περιβάλλοντος.

Η υπολογιστική ισχύς και η τεχνολογία εισόδου δεδομένων έχει εξελιχθεί σε ένα τέτοιο σημείο όπου θα μπορούσε να υπάρχει κάποιο όφελος από αυτές τις υπάρχουσες μη προγραμματιστικές δεξιότητες και φυσικές συμπεριφορές των χρηστών. Οι φυσικές διεπαφές επωφελούνται από αυτό, επιτρέποντας στους χρήστες να αλληλεπιδρούν με τους υπολογιστές χρησιμοποιώντας τις διαισθητικές τους λειτουργίες όπως το άγγιγμα, οι χειρονομίες, η ομιλία, ή ακόμα και η σκέψη. Έτσι με την παρουσία αυτών των διεπαφών οι χρήστες μπορούν να κατανοήσουν καλύτερα την επικοινωνία τους με τις μηχανές, μέσα από μεταφορές διεπαφών

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

(interface metaphors), οι οποίες αντλούνται από την εμπειρία που αποκτά ο χρήστης μέσω της αλληλεπίδρασής του με τον πραγματικό κόσμο.

Οι μεταφορές διεπαφών που χρησιμοποιούνται στις φυσικές διεπαφές βρίσκονται σε αντίθεση με αυτές που χρησιμοποιούνται στις γραφικές διεπαφές χρήστη. Οι τελευταίες περιλαμβάνουν πολλούς διαφορετικούς τύπους διεπαφών, αλλά στην πράξη, σχεδόν όλες ιστορικά έχουν χρησιμοποιήσει ένα συγκεκριμένο σύνολο μεταφορών αλληλεπίδρασης, όπως Παράθυρα, Εικονίδια, μενού, και συσκευές κατάδειξης. Όλα αυτά τα εργαλεία χρησιμοποιούνται ακόμα διότι η χρήση τους κρατάει εδώ και δεκαετίες, έτσι οι χρήστες είναι εξοικειωμένοι με αυτά λόγω της πολυετής χρήσης τους. Όμως αυτές τελικά οι τεχνικές αλληλεπίδρασης είναι σχεδιασμένες για μια εποχή, όπου η αιχμή της τεχνολογίας της αλληλεπίδρασης ήταν το ποντίκι και το πληκτρολόγιο.

Τέλος, οι φυσικές διεπαφές χρησιμοποιούν τις υπάρχουσες δεξιότητες και φυσικές συμπεριφορές για να επιτρέπουν στους χρήστες να αλληλεπιδρούν με τον πιο κατάλληλο τρόπο με το περιεχόμενο. Η φράση "αλληλεπιδρούν σωστά με το περιεχόμενο" εννοεί ότι το επίκεντρο των αλληλεπιδράσεων είναι σχετικό με το περιεχόμενο και υπάρχει η κατάλληλη αλληλεπίδραση με αυτό. Αυτό δεν σημαίνει ότι το περιβάλλον δεν μπορεί να έχει τα κατάλληλα εργαλεία ελέγχου, όπως κουμπιά ή πλαίσια ελέγχου, όταν αυτό είναι απαραίτητο. Το μόνο που σημαίνει είναι ότι οι έλεγχοι θα πρέπει να είναι δευτερεύουσας σημασίας ως προς το περιεχόμενο, καθώς και ότι η διασύνδεση θα πρέπει να επιτρέπει την πιο κατάλληλη αλληλεπίδραση με το περιεχόμενο σε μια συγκεκριμένη κατάσταση.

Καθώς οι υπολογιστές γίνονται όλο και πιο διάχυτοι στη ζωή και το περιβάλλον μας, οι χρήστες θα τους θεωρούν λιγότερο ως υπολογιστές και περισσότερο ως κοινά αντικείμενα ή χώρους από όπου μπορούν να έχουν πρόσβαση σε διάφορες πληροφορίες και περιεχόμενα. Τα ποντίκια και οι δρομείς είναι η καλύτερη επιλογή όταν χρειάζεται να γίνουν ακριβείς δραστηριότητες επιλογής και είναι πιο κατάλληλοι για εργασίες όπως η επεξεργασία εικόνας, αλλά αν σκεφτεί κανείς το ευρύτερο κόσμο όλων των πιθανών αλληλεπιδράσεων τώρα και στο μέλλον, οι αλληλεπιδράσεις με ποντίκια και δρομείς αποτελούν τη μειοψηφία. Χρησιμοποιώντας αυτές τις συσκευές εκεί που δεν είναι κατάλληλες, το μόνο που θα επιτευχθεί είναι η οπισθοδρόμηση ή η μη πρόοδος.

Το ερώτημα είναι αν η φυσική διεπαφή χρήστη θα αντικαταστήσει τη γραφική διεπαφή. Όμως αν δούμε την ιστορική διαδρομή των διεπαφών, θα διαπιστώσουμε ότι στην ουσία οι γραφικές διεπαφές αντικατέστησαν σε μεγάλο βαθμό τις διεπαφές γραμμών εντολών. Βέβαια δεν υπήρξε ποτέ πλήρης αντικατάσταση των διεπαφών αυτών, αφού ακόμα και σήμερα χρησιμοποιούνται σε συστήματα διαχείρισης, προγραμματιστικές εργασίες και γενικότερα όπου αυτά απαιτούνται. Φυσικά στις καθημερινές αλληλεπιδράσεις των χρηστών με τους υπολογιστές έχουν επικρατήσει οι γραφικές διεπαφές.

Σήμερα εμφανίζεται το ίδιο μοτίβο μεταξύ φυσικών και γραφικών διεπαφών. Οι φυσικές διεπαφές χρήστη θα αναλάβουν την σκυτάλη από τα γραφικά περιβάλλοντα χρήστη, αλλά οι γραφικές θα εξακολουθούν να υπάρχουν και να χρησιμοποιούνται, όταν αυτές κρίνονται ως πιο αποτελεσματικός τρόπος για να επιτευχθεί μια εξειδικευμένη εργασία. Οι εργασίες αυτές θα είναι πιθανώς δουλειές που απαιτούν μεγάλη ακρίβεια, όπως η γραφιστική. Όμως ακόμη και αυτές οι εφαρμογές θα επηρεαστούν από τα νέα πρότυπα αλληλεπίδρασης των φυσικών διεπαφών, οπότε μελλοντικά κάποιες από αυτές μπορεί να είναι υβριδικές. Η επανάσταση των φυσικών διεπαφών είναι αναπόφευκτη, αφού προσφέρουν στο χρήστη μεγαλύτερη ικανότητα κατανόησης της τεχνολογίας και περισσότερη ευχρηστία.

2.3 Τι σημαίνει ο όρος φυσική διεπαφή:

Εδώ χρειάζεται να εξετάσουμε τον όρο "φυσική διεπαφή". Πολύ συχνά χρησιμοποιείται η λέξη "διαισθητική" αντί για "φυσική". Για να κατανοηθεί καλύτερα ο όρος "φυσική διεπαφή", μπορούμε να αναφέρουμε την περιγραφή που δίνει ο Bill Buxton, ένας από τους πρωτοπόρους στην τεχνολογία των πολύ-επαφών και γενικότερα στο χώρο των φυσικών διεπαφών:

Μια διεπαφή θεωρείται ως φυσική όταν αυτή εκμεταλλεύεται τις δεξιότητες που έχουν αποκτήσει οι άνθρωποι μέσα από την εμπειρία της ζωής τους στο φυσικό κόσμο. ("An interface is natural if it exploits skills that we have acquired through a lifetime of living in the world").

Η περιγραφή αυτή είναι ενδιαφέρουσα για δύο λόγους. Πρώτον διότι συνδέει την έννοια της φυσικής με την ιδέα της επαναχρησιμοποίησης υφιστάμενων δεξιοτήτων. Δεύτερον, καθιστά σαφές ότι αυτές οι δεξιότητες δεν είναι μόνο οι έμφυτες ικανότητες με τις οποίες γεννιούνται οι άνθρωποι. Ο όρος "φυσικό" έχει να κάνει με τη χρησιμοποίηση των έμφυτων ικανοτήτων μαζί με τις δεξιότητες που έχουν αναπτυχθεί μέσα από την αλληλεπίδραση με το φυσικό περιβάλλον στην καθημερινή ζωή.

2.4 Επεξήγηση των έμφυτων ικανοτήτων και των αποκτηθέντων δεξιοτήτων

Όλοι οι άνθρωποι γεννιούνται με ορισμένες ικανότητες, και καθώς μεγαλώνουν ορισμένες άλλες ικανότητες ωριμάζουν από μόνες τους. Μερικά παραδείγματα από αυτές τις ικανότητες είναι το φαγητό, το περπάτημα, και η ομιλία. Επίσης οι άνθρωποι έχουν ορισμένες χαμηλού επιπέδου ικανότητες, οι οποίες χρησιμοποιούνται για τις βασικές λειτουργίες του σώματος, καθώς και για την αντίληψη του περιβάλλοντος. Μερικά παραδείγματα αυτών των χαμηλού επιπέδου ικανοτήτων είναι η αντίληψη των αλλαγών στο πεδίο της όρασης, η αντίληψη των διαφορών της υφής και του βάθους ακόμα και η απομόνωση ενός ήχου μεταξύ άλλων. Θα μπορούσαμε να πούμε ότι αυτές οι έμφυτες ικανότητες είναι σαν ένα πρόγραμμα οδηγού για τον εγκέφαλο και το σώμα μας.

2.5 Η ικανότητα στη μάθηση:

Οι άνθρωποι έχουν επίσης μια άλλη πολύ σημαντική ικανότητα. Έχουν την έμφυτη ικανότητα να μαθαίνουν, η οποία επιτρέπει την προσθήκη επιπλέον δεξιοτήτων στις φυσικές ικανότητες. Η μάθηση είναι ο πυρήνας της ανθρώπινης εμπειρίας. Χρειάζονται νέες δεξιότητες, προκειμένου ο άνθρωπος να αντιμετωπίσει και να προσαρμοστεί στο περιβάλλον. Οι δεξιότητες που αποκτούνται είναι διαφορετικές από τις έμφυτες ικανότητες, διότι χρειάζεται να επιλέξουμε για να μάθουμε μια δεξιότητα, ενώ οι έμφυτες ικανότητες ωριμάζουν αυτόματα. Μόλις ο άνθρωπος αποκτήσει μια δεξιότητα, γίνεται φυσικό και εύκολο να την επαναλάβει, όσο αυτός διατηρεί αυτή την ικανότητα.

Οι δεξιότητες και οι ικανότητες χρησιμοποιούνται για την ολοκλήρωση των διαφόρων εργασιών. Μια εργασία είναι μια μονάδα εργασίας που απαιτεί την ενέργεια του χρήστη και έχει ως σκοπό ένα συγκεκριμένο αποτέλεσμα. Οι εργασίες μπορεί να αποτελούνται από επιμέρους εργασίες.

Ένα παράδειγμα μιας εργασίας σε μια διεπαφή χρήστη είναι η αποστολή ενός e-mail. Για να σταλεί ένα e-mail, θα πρέπει να εκτελεστούν μια σειρά από επιμέρους διεργασίες, όπως η πληκτρολόγηση του σώματος του μηνύματος ή το πάτημα του κουμπιού "αποστολή". Χρησιμοποιούνται ειδικές ικανότητες για να ολοκληρωθεί κάθε τέτοια υπό-εργασία, καθώς πλησιάζουμε προς το συνολικό στόχο του που είναι η αποστολή του μηνύματος. Οι εργασίες και οι δεξιότητες είναι αλληλένδετες μεταξύ τους, αφού μια εργασία είναι κάτι που πρέπει να γίνει για να επιτευχθεί ένα αποτέλεσμα ενώ η δεξιότητα είναι η ικανότητα να επιτευχθεί αυτό το αποτέλεσμα.

3 Κεφάλαιο

Φυσική διεπαφή χρήστη υπολογιστή (Natural User Interface)

Ενότητες Κεφαλαίου

3.1	Εισαγωγή κεφαλαίου	19
3.2	Πολυτροπική είσοδος και έξοδος	19
3.3	Πολυτροπική έξοδος	19

3.1 Εισαγωγή κεφαλαίου.

Εδώ χρειάζεται να δούμε την έννοια της πολυτροπικότητας όσον αφορά την αλληλεπιδράσεις και τις διεπαφές. Η πολυτροπική αλληλεπίδραση (multimodal interaction) παρέχει στον χρήστη ενός συστήματος διαφορετικούς τρόπους να αλληλεπιδρά με αυτό. Μια πολυτροπική διεπαφή (multimodal interface) παρέχει αρκετά εργαλεία για την είσοδο και έξοδο των δεδομένων.

3.2 Πολυτροπική είσοδος και έξοδος

Υπάρχουν δύο διαφορετικά είδη πολυτροπικών διεπαφών που έχουν συγκλίνει. Μία είναι αυτή των εναλλακτικών μεθόδων εισόδου και αυτή της συνδυασμένης εισόδου/εξόδου δεδομένων. Το πρώτο είδος συνδυάζει διάφορες μεθόδους εισόδου δεδομένων από τον χρήστη, εκτός των παραδοσιακών (πληκτρολόγιο, ποντίκι), όπως η ομιλία, η χρήση γραφίδας, η αφή, οι χειρονομίες [2], το βλέμμα και οι κινήσεις του κεφαλιού ή του σώματος [3]. Μια κοινή διεπαφή αυτού του είδους είναι ο συνδυασμός μιας οπτικής τροπικότητας (όπως μια οθόνη, ένα πληκτρολόγιο και ένα ποντίκι) με κάποια φωνητική τροπικότητα (αναγνώριση ομιλίας για είσοδο δεδομένων, σύνθεση ομιλίας και προηχογραφημένους ήχους για έξοδο). Μπορούν όμως να χρησιμοποιηθούν και άλλες μέθοδοι όπως η είσοδος με βάση τη γραφίδα ή η απτική είσοδος/έξοδος. Οι πολυτροπικές διεπαφές χρήστη αποτελούν αντικείμενο έρευνας του πεδίου της αλληλεπίδρασης ανθρώπου-υπολογιστή (human-computer interaction, HCI).

Το πλεονέκτημα των πολλαπλών τρόπων εισόδου/εξόδου είναι η ευχρηστία: οι αδυναμίες μιας τροπικότητας συμπληρώνονται από τα πλεονεκτήματα κάποιας άλλης. Για παράδειγμα, σε μια φορητή συσκευή με μικρή οθόνη για γραφικό περιβάλλον και μικρό πληκτρολόγιο, μια λέξη μπορεί να είναι πιο δύσκολο να γραφτεί παρά να ειπωθεί. Αυτό είναι ιδιαίτερα σημαντικό όταν αυτές οι μικρές συσκευές χρησιμοποιούνται για την πρόσβαση και αναζήτηση σε καταλόγους πολυμέσων. Ένα άλλο παράδειγμα είναι η παροχή πληροφοριών ασθενή σε μια αίθουσα χειρουργείου, όπου τα μέλη της χειρουργικής ομάδας μπορούν να τις λαμβάνουν με διάφορους οπτικούς ή ακουστικούς τρόπους ώστε να λαμβάνονται όσο το δυνατόν πιο γρήγορα και σωστά.

Οι πολυτροπικές διεπαφές χρήστη μπορούν επίσης να είναι καλύτερα προσβάσιμες. Μια πολυτροπική εφαρμογή που έχει σχεδιαστεί σωστά μπορεί να χρησιμοποιηθεί από ανθρώπους με διάφορες αναπηρίες. Οι χρήστες με προβλήματα όρασης μπορούν να χρησιμοποιήσουν την τροπικότητα φωνής σε συνδυασμό με κάποιο πληκτρολόγιο. Οι χρήστες με προβλήματα ακοής μπορούν να χρησιμοποιήσουν την οπτική τροπικότητα και την φωνητική είσοδο. Χρήστες που έχουν προβλήματα λόγω της κατάστασης στην οποία βρίσκονται (για παράδειγμα όσοι χρησιμοποιούν γάντια σε περιβάλλον με θόρυβο, οδηγούν ή πρέπει να δώσουν τον κωδικό της πιστωτικής τους κάρτας ενώ είναι σε δημόσιο χώρο) μπορούν επίσης να χρησιμοποιήσουν την κατάλληλη τροπικότητα. Αν όμως μια πολυτροπική εφαρμογή απαιτεί από τον χρήστη της να χρησιμοποιεί όλους αυτούς τους τρόπους, τότε μάλλον δεν είναι σωστά σχεδιασμένη.

3.3 Πολυτροπική έξοδος

Το δεύτερο είδος πολυτροπικών συστημάτων δίνει στον χρήστη ερεθίσματα πολυμέσων και πολυτροπική έξοδο, κυρίως με τη μορφή εικόνας ή ήχου. Έχει αρχίσει να γίνεται χρήση και άλλων τροπικοτήτων στη σχεδίαση διεπαφών, όπως η αφή και η όσφρηση. Τα πολυτροπικά συστήματα εξόδου έχουν τα εξής προτερήματα: συνέργεια και πλεονασμός. Η πληροφορία που παρουσιάζεται με διαφορετικούς τρόπους λειτουργεί ενιαία και αφορά διαφορετικές πλευρές της ίδιας εργασίας. Η χρήση διαφορετικών τροπικοτήτων για την επεξεργασία της ίδιας πληροφορίας έχει σαν αποτέλεσμα υψηλότερο εύρος ζώνης μεταφοράς πληροφορίας [4][5][6].

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

Η πολυτροπική έξοδος χρησιμοποιείται κυρίως για να βελτιώνεται η αντιστοίχιση μεταξύ του μέσου επικοινωνίας και του περιεχομένου, και για τη διαχείριση της προσοχής σε περιβάλλοντα με μεγάλο όγκο δεδομένων όπου οι χρήστες πρέπει να έχουν συγκεντρωμένη την προσοχή τους .

Ένα σημαντικό βήμα στη σχεδίαση πολυτροπικών διεπαφών είναι η δημιουργία φυσικών αντιστοιχίσεων μεταξύ τροπικότητων και πληροφοριών ή εργασιών. Η ακοή διαφέρει από την όραση. Δεν έχει κατεύθυνση, είναι παροδική και λειτουργεί πάντα. Ειδικά η έξοδος ομιλίας, που είναι μορφή ηχητικής πληροφορίας, αποτελεί αντικείμενο μελέτης και έχουν αναπτυχθεί διάφορες τεχνικές για τη χρήση της ομιλίας. Οι Michaelis και Wiggins (1982) υποστήριξαν ότι η έξοδος ομιλίας θα έπρεπε να χρησιμοποιηθεί για απλά και σύντομα μηνύματα που δεν πρόκειται να χρειαστούν στο μέλλον. Επίσης θεώρησαν ότι θα ήταν καλύτερο η ομιλία να παράγεται όσο πιο σύντομα γίνεται και να πρέπει να απαντηθεί από τον χρήστη.

Η αίσθηση της αφής χρησιμοποιήθηκε για πρώτη φορά σαν μέσο επικοινωνίας κατά τα τέλη της δεκαετίας του 1950. Εκτός από πολλά υποσχόμενη, είναι και μοναδική. Σε αντίθεση με την ακοή και την όραση, τις δύο βασικές αισθήσεις της HCI, η αίσθηση της αφής είναι εγγύτατη, ανιχνεύει αντικείμενα που ακουμπούν στο σώμα και είναι δύο κατευθύνσεων γιατί υποστηρίζει τόσο την πρόσληψη ερεθισμάτων όσο και την παρέμβαση στο περιβάλλον.

Παραδείγματα ηχητικής αλληλεπίδρασης είναι τα ηχητικά εικονίδια των λειτουργικών συστημάτων στους ηλεκτρονικούς υπολογιστές, τα οποία σημαίνουν ενέργειες του χρήστη (όπως η διαγραφή ενός αρχείου, το άνοιγμα ενός φακέλου, ή κάποιο σφάλμα), και η έξοδος ομιλίας για τη βοήθεια κατά την πλοήγηση οχημάτων ή την προειδοποίηση του πιλότου στα σύγχρονα αεροσκάφη. Παραδείγματα σημάτων αφής είναι οι δονήσεις κατά το άναμμα του φωτός αλλαγής λωρίδας όταν υπάρχει άλλο αυτοκίνητο στη νεκρή γωνία του οδηγού, η δόνηση μιας θέσης αυτοκινήτου σαν σημάδι προειδοποίησης, και η δόνηση του μοχλού ελέγχου στα σύγχρονα αεροσκάφη σαν προειδοποίηση για επερχόμενες αναταράξεις.

Υπάρχουν αόρατοι χώροι διεπαφών που μπορούν να χρησιμοποιηθούν με την κατάλληλη τεχνολογία αισθητήρων, όπως οι κάμερες και οι αισθητήρες υπέρυθρου φωτός ή υπερήχων [7].

4 Κεφάλαιο

Συσκευές εισόδου φυσικών διεπαφών

Ενότητες κεφαλαίου.

4.1	Εισαγωγή κεφαλαίου	23
4.2	Στερεοσκοπικές κάμερες	23
4.3	Κάμερες βάθους	23
4.4	Αισθητήρες Αφής	24
4.5	Asus WA Vi Xtion (Αισθητήρες βάθους)	26
4.6	PrimeSense (Πολύ-αισθητήρας)	27
4.7	Kinect (Πολύ-αισθητήρας)	28
4.7.1	Πιο αναλυτικά για το Kinect	28
4.7.2	Ανάλυση μέτρησης του βάθους	29
4.8	Τεχνολογία που επιλέχθηκε	31

4.1 Εισαγωγή κεφαλαίου.

Εδώ θα πρέπει να δούμε τον σημαντικό ρόλο που παίζουν οι συσκευές εισόδου και για την υλοποίηση των φυσικών διεπαφών χρήστη. Στις επόμενες ενότητες θα αναλυθούν οι διάφορες συσκευές εισόδου, οι οποίες ανιχνεύουν με φυσικό τρόπο τις εντολές του χρήστη.

4.2 Στερεοσκοπικές κάμερες

Οι στερεοσκοπικές κάμερες είναι εμπνευσμένες από τον τρόπο που λειτουργεί η ανθρώπινη όραση. Διαθέτουν δύο τουλάχιστον αισθητήρες RGB και οι οπτικοί φακοί τους παρατηρούν τον χώρο από δύο διαφορετικές γωνίες. Ειδικά για την περίπτωση των δύο οπτικών φακών, εφαρμόζοντας την μέθοδο του τριγωνισμού στα ζεύγη εικόνων που παράγονται κατά σταθερά χρονικά διαστήματα από τον αισθητήρα, είναι δυνατό να εξαχθεί η πληροφορία βάθους.

Ο υπολογισμός του βάθους μέσω του τριγωνισμού είναι βέβαια γνωστός, όμως ακόμα και με αυτό το θεωρητικό υπόβαθρο υπάρχει δυσκολία στο να έχουμε την ακριβή θέση κάθε σημείου στο χώρο λόγω των πολλαπλών τύπων θορύβου που πολλές φορές είναι απροσδιόριστη η συμβολή τους. Η μέθοδος του τριγωνισμού εμφανίζεται με πολλές διαφορετικές εκδοχές που εστιάζουν στην εκτίμηση της θέσης του κάθε σημείου στο χώρο υπολογίζοντας ένα σφάλμα μέτρησης και προσπαθώντας να το ελαχιστοποιήσουν.

4.3 Κάμερες βάθους

Σε αυτή την κατηγορία ανήκουν οι κάμερες που παρέχουν ροή εικόνων με κάθε pixel να αντιστοιχεί στην απόσταση του αισθητήρα από το σημείο που καταγράφηκε και βασίζονται στις εξής τεχνολογίες :

- Time-Of-Life (TOF) που χρησιμοποιεί ένα μοναδικό ή κινούμενο Laser. Η τεχνική αυτή εκμεταλλεύεται τους εκπεμπόμενους οπτικούς παλμούς και υπολογίζει την καθυστέρηση του φωτός κατά τη διάδοσή του, την ανάκλασή του και τελικά την επιστροφή του στον αισθητήρα. Οι παλμοί που εκπέμπονται είναι συνήθως διαμορφωμένα σήματα υπέρυθρης ακτινοβολίας ή παλμοί φωτός και λαμβάνονται από κατάλληλους αισθητήρες CMOS ή φωτοδιόδους που είναι ενσωματωμένοι στο συνολικό σύστημα.
- Structured Light Technique, που είναι η τεχνική του δομημένου φωτισμού. Η τεχνολογία αυτή βασίζεται στην εκπομπή φωτός με ένα συγκεκριμένο πρότυπο. Ο φωτισμός μπορεί να είναι στο ορατό φως ή ακόμα και στο υπέρυθρο, ενώ το πρότυπο προβολής είναι συνήθως ένα πλέγμα που για το ορατό φως χρησιμοποιείται δίχρωμο τύπου σκακιέρας.

Κάμερες οι οποίες βασίζονται στην τεχνολογία TOF εμφανίζονται πρόσφατα στα μέσα της δεκαετίας του 2000 όταν δηλαδή η τεχνολογία των ημιαγωγών ήταν επαρκής ώστε να υποστηρίξει τη δημιουργία πολύ στενών παλμών. Το κόστος προμήθειας είναι εξαιρετικά μεγάλο της τάξης των \$10000.

Από τους αισθητήρες δομημένου φωτισμού παρουσιάζουν ιδιαίτερο ενδιαφέρον εκείνοι που χρησιμοποιούν υπέρυθρη ακτινοβολία. Χαρακτηριστικά παραδείγματα είναι οι αισθητήρες PrimeSense και Kinect οι οποίοι θα αναλυθούν περαιτέρω στις ενότητες που ακολουθούν.

4.4 Αισθητήρες Αφής

Ένας αισθητήρας αφής μετράει διαρκώς μεταβολές ασκούμενης δύναμης σε μια περιοχή. Αυτό επιτυγχάνεται με τη χρήση ενός πίνακα $m \times n$ ο οποίος αποτελείται από ξεχωριστά όμοια στοιχεία τα οποία αποκαλούνται δυναμοστοιχεία. Η πληροφορία που λαμβάνεται από το κάθε στοιχείο μπορεί να συνδυαστεί έτσι ώστε να μας δώσει μια εικόνα αφής του αντικειμένου. Επιπλέον και εκτός από το σχήμα και το μέγεθος, μπορούν να εξαχθούν πληροφορίες όπως η υφή και η θερμική αγωγιμότητα του αντικειμένου. Αυτού του είδους οι αισθητήρες είναι απαραίτητοι για ρομποτικές εφαρμογές αλλά και για εφαρμογές της καθημερινής μας ζωής, όπως τα κινητά τηλέφωνα ή τα GPS

Η επιφάνεια ενός αισθητήρα αφής πρέπει να είναι κατασκευασμένη από ανθεκτικό και υποχωρητικό υλικό, ενώ η απόσταση των δυναμοστοιχείων θα πρέπει να είναι σταθερή και να μην παρουσιάζει φαινόμενα υστέρησης. Επιπλέον χρειάζεται να υπάρχει μονότονη σχέση μεταξύ δύναμης και απόκρισης αν και αυτή δεν είναι απαραίτητο να είναι γραμμική. Οι διαστάσεις του αισθητήρα καθώς και τα ιδιαίτερα χαρακτηριστικά των δυναμοστοιχείων εξαρτώνται από την εφαρμογή.

Η μετάδοση της πληροφορίας γίνεται συνήθως με χρήση της χωρητικής ή αντιστατικής μετατροπής. Στην αντιστατική μετατροπή χρησιμοποιείται μια αντίσταση η οποία μεταβάλλεται ανάλογα με τη δύναμη η οποία ασκείται στην επιφάνεια. Αντίστοιχα στην χωρητική μετατροπή μετριέται η μεταβολή στην χωρητικότητα ενός πυκνωτή που προκαλείται από την άσκηση δύναμης πάνω στην επιφάνεια.

Από τη σκοπιά των συσκευών εισόδου σε φυσικές διεπαφές, μια σημαντική εφαρμογή της τεχνολογίας των αισθητήρων αφής είναι οι οθόνες αφής. Μια οθόνη αφής είναι μια συσκευή εισόδου που μπορεί να ανιχνεύσει την παρουσία και τη θέση από ένα άγγιγμα μέσα στην περιοχή της οθόνης. Οι οθόνες αυτές έχουν τη δυνατότητα να ανιχνεύσουν χειρονομίες των δακτύλων πάνω στην επιφάνεια της οθόνης. Πλέον με τη τεχνολογία της πολύ-επαφής (multi touch), υπάρχει η δυνατότητα να ανιχνευτούν χειρονομίες που προέρχονται από το συνδυασμό αφής ακόμα και τεσσάρων δακτύλων.



Εικόνα 1 Επιτραπέζια οθόνη αφής

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

Οι οθόνες αφής είναι αρκετά διαδεδομένες σε διάφορα εμπορικά προϊόντα όπως τα κινητά τηλέφωνα, τις δορυφορικές συσκευές πλοήγησης ή τους προσωπικούς υπολογιστές. Επίσης χρησιμοποιούνται στο τομέα της βιομηχανίας αλλά και στο τομέα των υπηρεσιών όπως είναι η έκδοση εισιτηρίων.

Πρόσφατα η Samsung σε συνεργασία με τη Microsoft κατασκεύασαν το SUR40, μια συσκευή που παρέχει αρκετή υπολογιστική ισχύ και η οποία ελέγχεται από το χρήστη αποκλειστικά μέσω της αφής. Στην ουσία είναι μια οθόνη αφής σαράντα ιντσών με αναγνώριση πολλαπλών επαφών και πολλαπλών χρηστών.



Εικόνα 2 Η συσκευή SUR40.

Η ραγδαία εξέλιξη στην τεχνολογία των οθονών αφής φέρνει συνεχώς νέα δεδομένα στο τομέα των διεπαφών, οι οποίες γίνονται όλο και πιο φυσικές. Η αφή σαν μέθοδο εισόδου σε ένα σύστημα τείνει να καθιερωθεί και μαζί με τις υπόλοιπες φυσικές λειτουργίες του ανθρώπου (ομιλία, χειρονομίες) θα αποτελέσει στο μέλλον τον κύριο τρόπο με τον οποίο θα επιτυγχάνεται η επικοινωνία ανθρώπου – μηχανής.

4.5 Asus WA Vi Xtion (Αισθητήρες βάθους)

Μια εναλλακτική πρόταση στον τομέα των συσκευών εισόδου φυσικού τύπου θα διατεθεί στην αγορά από την Asus σε συνεργασία με την PrimeSense στο τρίμηνο του 2011, όπως παρουσιάστηκε στην διεθνή έκθεση ψηφιακής βιομηχανίας CeBit. Το Xtion διαθέτει έναν αισθητήρα βάθους αλλά και δυνατότητες ασύρματης μετάδοσης δεδομένων ήχου και εικόνας από έναν υπολογιστή. Είναι δυνατό ένας ασύρματος αναμεταδότης να μεταφέρει πληροφορίες από τον υπολογιστή σε ένα δεύτερο αναμεταδότη σε απόσταση 25 μέτρων.



Εικόνα 3 Ο αισθητήρας βάθους Xtion της Asus

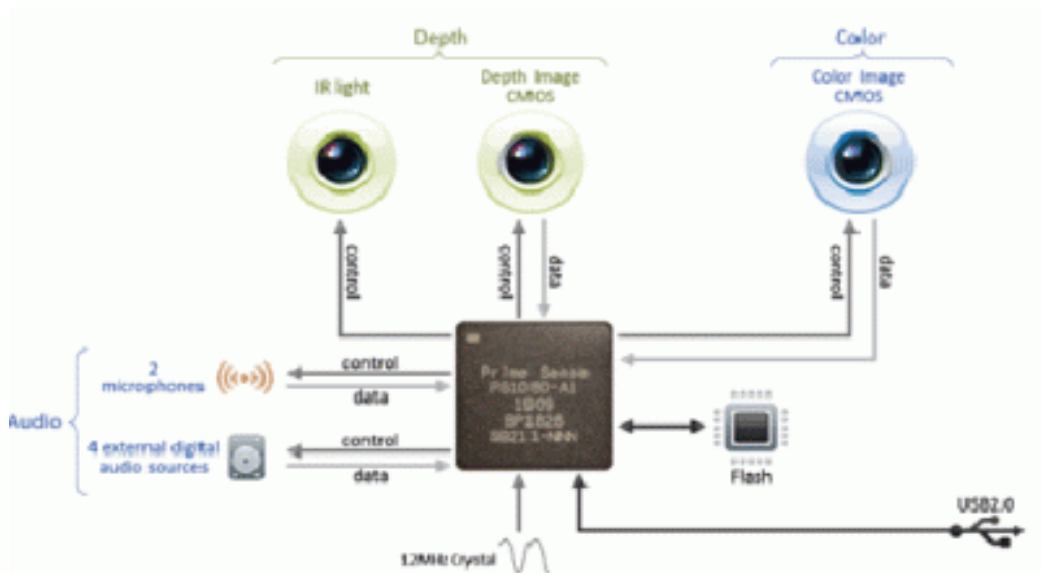
Τα τεχνικά χαρακτηριστικά του Asus Xtion Pro Live είναι τα εξής:

- Εύρος λειτουργίας 0,8 – 3,5 μέτρα.
- Εύρος θέασης: οριζόντια 58°, κάθετα 45° και διαγώνια 70°.
- Σύνδεση με USB. 2.0
- Μέγιστη ανάλυση εικόνας: 640 X 480
- Ρυθμός μετάδοσης εικόνας: 30 fps.
- Ανάλυση video : SXGA (1280*1024)
- Δύο μικρόφωνα.

Αν και με σαφώς πιο περιορισμένες δυνατότητες σε σχέση με τον πολύ αισθητήρα Kinect είναι μια πολύ καλή εναλλακτική λύση που αξιοποιεί τα ίδια τα εργαλεία λογισμικού τα οποία είναι διαθέσιμα και στον PrimeSense και στο Kinect. Αυτή τη στιγμή κοστολογείται στα \$160 και προορίζεται για ένα κοινό φάσμα εφαρμογών που δεν μπορεί να ικανοποιήσει το Kinect λόγω της αδυναμίας ασύρματης μετάδοσης δεδομένων.

4.6 PrimeSense (Πολύ-αισθητήρας)

Ο αισθητήρας PrimeSense είναι προϊόν της ομώνυμης ιδιωτικής Ισραηλινής εταιρίας που ιδρύθηκε το 2005 και έμελλε να ανατρέψει με την τεχνολογία του το κατεστημένο στην τρισδιάστατη απεικόνιση και ανακατασκευή του χώρου. Αν και ο ίδιος ο αισθητήρας δεν έτυχε ιδιαίτερης εμπορικότητας, ο απερχόμενος και πιο ισχυρός που η ίδια εταιρία κατασκεύασε για λογαριασμό της Microsoft ήταν από πολλές απόψεις επαναστατικός.



Εικόνα 4 Απεικόνιση των επιμέρους τμημάτων του αισθητήρα του PrimeSense

Ο αισθητήρας είναι στην πραγματικότητα ένας πολύ-αισθητήρας καθώς διαθέτει μία RGBκάμερα, μία κάμερα βάθους, δύο μικρόφωνα και τέσσερις εισόδους για ψηφιακές ηχητικές πηγές, ενώ συνδέεται με εξωτερικά συστήματα μέσω της ευρύτατα διαδεδομένης θύρας USB2. Ο αισθητήρας βάθους δίνει ανάλυση εικόνας VGA 640x480 pixels με χωρική ανάλυση x/y 3 mmσε απόσταση 2 μέτρα και ανάλυση βάθους 1 cmομοίως σε απόσταση 2 μέτρων από τον αισθητήρα. Επιπλέον, έχει ρυθμό μετάδοσης εικόνων 60 fps, και εύρος λειτουργίας 0.8-3.5 μέτρα. Είναι ιδιαίτερα μικρό σε διαστάσεις με ελάχιστες απαιτήσεις σε ενέργεια 2.25 Watt, ενώ ο RGB αισθητήρας διαθέτει ανάλυση UXGA 1600x1200 pixels.

Η PrimeSense εισήγαγε την τεχνολογία Light Coding με την οποία μπορεί να εξαχθεί το βάθος της σκηνής σε μικρή απόσταση από τον αισθητήρα. Ο αισθητήρας βάθους αποτελείται από δύο ανταγωνιστικούς αισθητήρες, έναν υπέρυθρης ακτινοβολίας που προβάλλει κωδικοποιημένο υπέρυθρο φως που είναι αόρατο στο ανθρώπινο μάτι, και έναν αλληλοεξαρτώμενο CMOS αισθητήρα εικόνας που λαμβάνει το κωδικοποιημένο υπέρυθρο σήμα. Ο CMOS αισθητήρας είναι συνδεδεμένος με ένα εσωτερικό επεξεργαστή που υποβάλλει τις εικόνες στη βασική επεξεργασία εφαρμόζοντας παράλληλες τεχνικές. Ο επεξεργαστής αποκωδικοποιεί την υπέρυθρη εικόνα και πραγματοποιεί ταχύτατη εξαγωγή του βάθους για κάθε pixel και επιπλέον εκτελεί αλγόριθμο ταύτισης των pixelστης εικόνας βάθους με την αντίστοιχη εικόνα που προέρχεται από τον RGB αισθητήρα.

4.7 Kinect (Πολύ-αισθητήρας)

Το Kinect είναι μια συσκευή εισόδου ανίχνευσης κίνησης και ήχου (πολύ-αισθητήρας) που κατασκευάστηκε από την PrimeSense για λογαριασμό της Microsoft, με σκοπό να πλαισιώσει το Xbox 360, την κονσόλα παιχνιδιών, αλλά και τα Windows. Παρουσιάστηκε για πρώτη φορά σε μια διεθνή έκθεση ηλεκτρονικών βιντεοπαιχνιδιών τον Ιούνιο του 2009. Για μαζική παραγωγή στην παγκόσμια αγορά διατέθηκε σταδιακά το Νοέμβριο του 2010. Από εκείνη την περίοδο και μετά έχει φέρει την επανάσταση στην ρομποτική αλλά και στον τομέα των βιντεοπαιχνιδιών, αφού αυτοί οι κλάδοι προσπάθησαν να επωφεληθούν από τη χρήση της συγκεκριμένης συσκευής. Επίσης και στον ιατρικό τομέα έχουν αναπτυχθεί πολλές εργασίες με επίκεντρο το Kinect. Ξέφυγε δηλαδή αρκετά γρήγορα από τον αρχικό στόχο της Microsoft, που ήταν η χρήση του σαν ένα ελεγκτής σε βιντεοπαιχνίδια, και επεκτάθηκε και στους υπολογιστές. Αυτό έγινε με την ανάπτυξη προγραμματιστικών εργαλείων και βιβλιοθηκών όπως είναι το OpenNI ή το Microsoft SDK for Kinect, με τα οποία μπορούν να αναπτυχθούν εφαρμογές για υπολογιστές και οι οποίες θα αναπτυχθούν αναλυτικότερα στις επόμενες ενότητες.

Ο λόγος για τον οποίο το Kinect σάρωσε την παγκόσμια αγορά έγκειται κυρίως στο γεγονός 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 στα \$120, και η απόδοσή του σαν αισθητήρα είναι κατά πολύ κατώτερη σε σχέση με αυτή του Kinect.



Εικόνα 5. Απεικόνιση του Kinect και των αισθητήρων του.

4.7.1 Πιο αναλυτικά για το Kinect

Το Kinect βασίζεται σε τεχνολογίες λογισμικού που αναπτύσσονται από τη Rare, θυγατρική εταιρεία της Microsoft Game Studios, και την τεχνολογία της ισραηλινής PrimeSense πάνω στις κάμερες και τις συσκευές λήψης βίντεο. Η PrimeSense ανέπτυξε ένα σύστημα που μπορεί να ανιχνεύσει χειρονομίες και καθιστά δυνατό τον έλεγχο της συσκευής χρησιμοποιώντας μια υπέρυθρη κάμερα, δύο φακούς και ένα ειδικό μικροτσίπ για την παρακολούθηση της κίνησης των αντικειμένων και των ατόμων σε τρεις διαστάσεις. Αυτό το σύστημα τρισδιάστατης σάρωσης, που ονομάζεται Light Coding, χρησιμοποιεί μια παραλλαγή τρισδιάστατης εικόνας η οποία μπορεί να αναπαρασταθεί σε τρισδιάστατο χώρο.

Ο αισθητήρας Kinect είναι μια συσκευή που αποτελείται από μια οριζόντια διάταξη που συνδέεται με μια μικρή βάση με μηχανοκίνητο άξονα και έχει σχεδιαστεί για να τοποθετείται κατά μήκος πάνω ή κάτω από την οθόνη που προβάλλει το λογισμικό που το χρησιμοποιεί. Η συσκευή διαθέτει έναν RGB φακό, έναν αισθητήρα βάθους που είναι συνδυασμός δύο φακών και ένα πολλαπλό μικρόφωνο τα οποία λειτουργούν με το συγκεκριμένο λογισμικό που περιλαμβάνει το Microsoft Kinect SDK και παρέχουν πλήρη τρισδιάστατη καταγραφή της κίνησης, αναγνώριση προσώπου και δυνατότητες αναγνώρισης φωνής.

Ο αισθητήρας βάθους αποτελείται από έναν υπέρυθρο προβολέα λέιζερ σε συνδυασμό με ένα αισθητήρα CMOS, ο οποίος καταγράφει δεδομένα τρισδιάστατου βίντεο κάτω από οποιοδήποτε συνθήκες φωτισμού. Η απόσταση ανίχνευσης του αισθητήρα βάθους είναι ρυθμιζόμενη, και το λογισμικό του Kinect είναι σε θέση να βαθμονομήσει αυτόματα τον αισθητήρα με βάση το περιβάλλον παιχνιδιού και το φυσικό περιβάλλον του παίκτη, ακόμα και αν υπάρχουν στο χώρο και άλλα αντικείμενα.

Ο αισθητήρας Kinect εξάγει βίντεο με ρυθμό καρτέ 30 Hz. Η ροή βίντεο χρησιμοποιεί το σύστημα χρωμάτων RGB 8-bit και ανάλυση 640×480 pixel με χρωματικό φίλτρο Bayer, ενώ η μονόχρωμη αισθητήρια ροή βίντεο χρησιμοποιεί ανάλυση 640×480 pixel με βάθος 11-bit, το οποίο παρέχει 2.048 επίπεδα ευαισθησίας. Ο αισθητήρας Kinect έχει ένα πρακτικό όριο που κυμαίνεται από 1,2 έως 3,5 μέτρα απόσταση. Η περιοχή που καλύπτει το οπτικό πεδίο του Kinect είναι περίπου 6 τ.μ., αν και ο αισθητήρας μπορεί να διατηρήσει την εστίαση παρακολούθησης σε ένα διευρυμένο φάσμα περίπου 0,7 έως 6 μέτρων. Ο αισθητήρας έχει οπτικό πεδίο 57° οριζόντια και 43° κατακόρυφα, ενώ ο κινητήριος κεντρικός άξονας είναι σε θέση να κινηθεί σε εύρος γωνίας έως και 27° προς τα πάνω ή προς τα κάτω. Το οριζόντιο πεδίο του αισθητήρα Kinect στην ελάχιστη απόσταση θέασης του (περίπου 0,8 μέτρα) είναι περίπου 87 εκατοστά, και το κατακόρυφο περίπου 63 εκατοστά, δηλαδή αντιστοιχούν περίπου 1,3 χιλιοστά ανά εικονοστοιχείο. Το πολλαπλό μικρόφωνο διαθέτει τέσσερις μικροφωνικές συσκευές και καθεμία λειτουργεί με κανάλι των 16-bit ήχου με ρυθμό δειγματοληψίας 16 kHz.

Επειδή ο αισθητήρας Kinect έχει μηχανοκίνητο μηχανισμό ανάκλησης απαιτεί 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 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.

Πιο συνοπτικά τα μέρη από τα οποία αποτελείται το Kinect είναι τα εξής:

- Μία RGB κάμερα με ανάλυση VGA 640×480 pixels, βάθους χρώματος 8-bit και ρυθμό μετάδοσης εικόνων 30fps.
- Τέσσερα μικρόφωνα 16-bit κβάντισης με συχνότητα δειγματοληψίας 16kHz το καθένα.
- Μία κάμερα βάθους ανάλυσης VGA 640×480 pixels
- Led τριών αποχρώσεων (πράσινο κόκκινο πορτοκαλί) ανάλογα με την κατάσταση που βρίσκεται ο αισθητήρας.
- Μία βάση με κινητήρα αλλαγής κλίσης κατά την οριζόντια κλίμακα

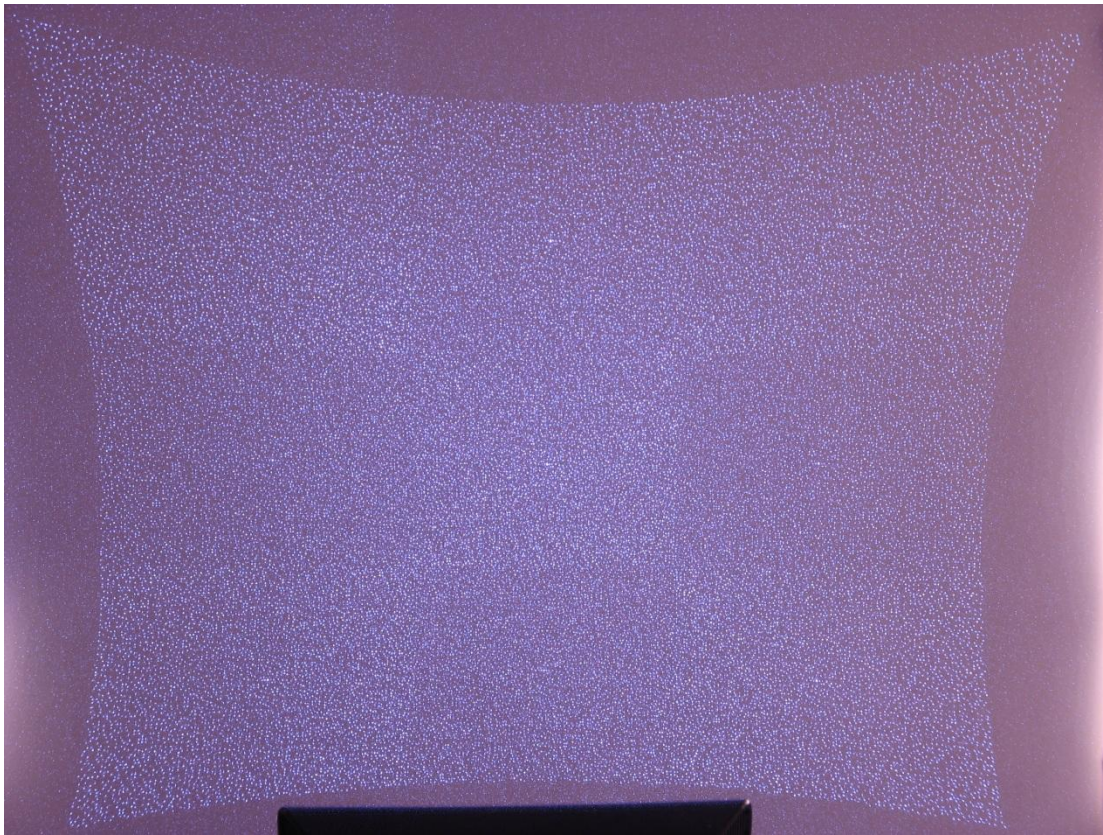
4.7.2 Ανάλυση μέτρησης του βάθους

Εδώ έχει μεγάλο ενδιαφέρον η ανάλυση του τρόπου που το Kinect αντιλαμβάνεται και αναγνωρίζει το βάθος, αφού αυτό είναι που στην ουσία καθιέρωσε το Kinect ως μία αξιόπιστη και φτηνή λύση για τη χρήση του ως συσκευή πολυτροπικής εισόδου.

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

Η λειτουργία του αισθητήρα βάθους είναι παρόμοια με αυτή του PrimeSense αφού στην ουσία το Kinect αποτελεί τη μετεξέλιξή του. Όμως στην πραγματικότητα ο ακριβής και πραγματικός τρόπος με τον οποίο λειτουργεί το Kinect παραμένει ένας γρίφος. Ούτε η κατασκευάστρια εταιρία (PrimeSense) αλλά ούτε και η πωλήτρια εταιρία (Microsoft) δεν έχουν ακόμα δημοσιεύσει datasheets για τη λειτουργία της αναγνώρισης βάθους. Το μόνο που έχουμε διαθέσιμο είναι κάποιες γενικές αναφορές των εταιριών και η όσο το δυνατόν καλύτερη παρατήρηση και κατανόηση των διάφορων ερευνητών που έχουν ασχοληθεί με την αποκρυπτογράφηση της λειτουργίας του.

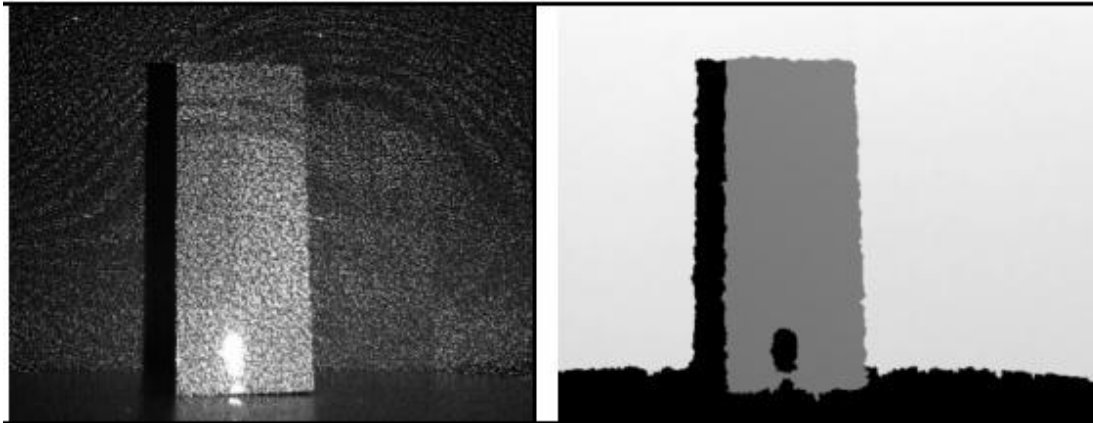
Οι κατασκευαστές του Kinect περιγράφουν τη μέτρηση του βάθους ως διαδικασία τριγωνισμού [8]. Η πηγή λέιζερ εκπέμπει μια απλή δέσμη η οποία χωρίζεται σε πολλαπλές δέσμες από ένα φράγμα περίθλασης για τη δημιουργία ενός σταθερού μοτίβου στιγμάτων τα οποία προβάλλονται πάνω στις επιφάνειες. Αυτό το μοτίβο στιγμάτων συλλαμβάνεται από την υπέρυθρη κάμερα και συσχετίζεται με ένα πρότυπο αναφοράς.



Εικόνα 6. Απεικόνιση ενός μοτίβου στιγμάτων (dot pattern) [10].

Το πρότυπο αναφοράς αποκτιέται με τη λήψη ενός επιπέδου σε μία γνωστή απόσταση από τον αισθητήρα, το οποίο στη συνέχεια αποθηκεύεται στην μνήμη του αισθητήρα. Όταν ένα στίγμα προβάλλεται πάνω σε ένα αντικείμενο, του οποίου η απόσταση με τον αισθητήρα είναι μικρότερη ή μεγαλύτερη από αυτή του επίπεδο αναφοράς, η θέση του στίγματος στην υπέρυθρη εικόνα θα μετατοπιστεί στην κατεύθυνση της γραμμής βάσης που υπάρχει μεταξύ του πομπού λέιζερ και του κέντρου της υπέρυθρης κάμερας. Αυτές οι μετατοπίσεις μετρήθηκαν για όλες στίγματα με μια απλή διαδικασία συσχέτισης, η οποία αποφέρει μια διαφορετική-ανόμοια εικόνα. Για καθένα εικονοστοιχείο η απόσταση από τον αισθητήρα μπορεί στη συνέχεια να ανακτηθεί από τη σύγκριση των δύο αυτών διαφορετικών εικόνων.

Παρακάτω φαίνεται ένα παράδειγμα της μέτρησης του βάθους, το πώς δηλαδή από την εικόνα με το μοτίβο στιγμάτων υπέρυθρης ακτινοβολίας αποδίδεται η εικόνα του βάθους.



Εικόνα 7. Αριστερά η την εικόνα με το μοτίβο στιγμάτων υπέρυθρης ακτινοβολίας και δεξιά η αντίστοιχη της εικόνα βάθους [8].

4.8 Τεχνολογία που επιλέχθηκε

Η τεχνολογία που επιλέχθηκε σαν συσκευή εισόδου στο σύστημα που αναπτύχθηκε στην παρούσα εργασία είναι το Microsoft Kinect και το SDK for Kinect της Microsoft. Η επιλογή αυτής της τεχνολογίας έγινε με κριτήριο τις δυνατότητες που προσφέρει αλλά και το κόστος σε σχέση με τις υπόλοιπες συσκευές εισόδου.

Οι δυνατότητες που προσφέρει το Kinect είναι πάρα πολλές όσον αφορά τη χρήση του ως πολυτροπική συσκευή εισόδου. Όπως αναλύθηκε στην παραπάνω ενότητα, παρέχεται η δυνατότητα αναγνώρισης φωνής, ομιλίας και προσώπου. Με τη χρήση της βιβλιοθήκης Microsoft SDK for Kinect, το οποίο ανανεώνεται και βελτιώνεται αρκετά συχνά, όλες αυτές οι λειτουργίες είναι εύκολα διαχειρίσιμες. Δίνεται η δυνατότητα να χρησιμοποιηθούν αντικειμενοστραφείς γλώσσες όπως η C# ή η C++ και έτσι η διαχείριση γίνεται σε υψηλό επίπεδο προγραμματισμού. Φυσικά υπάρχει και το OpenNI σαν βιβλιοθήκη για το Kinect η οποία χρησιμοποιεί τη Java σαν γλώσσα προγραμματισμού, όμως σε σχέση με το SDK της Microsoft έχει κάποια μειονεκτήματα. Ένα από αυτά είναι η συγκεκριμένη στάση που πρέπει να έχει το σώμα του χρήστη ούτως ώστε να αναγνωρίσει το σκελετό του.

Φυσικά, όσον αφορά το κόστος της συσκευής, η σύγκριση με τις υπόλοιπες παλαιότερες συσκευές πολυτροπικής εισόδου είναι συντριπτικά υπέρ του Kinect. Με το κόστος του σήμερα να μην ξεπερνάει τα 150 ευρώ, αποτελεί την πιο αξιόπιστη και οικονομική λύση. Εδώ χρειάζεται να τονιστεί ότι παρόλο το χαμηλό κόστος, η δυνατότητες και η αξιοπιστία του Kinect είναι ανάλογες και σε πολλά σημεία υπερτερούν σε σχέση με τις υπόλοιπες πιο ακριβές τεχνολογίες.

Όπως γίνεται κατανοητό το Kinect σε συνδυασμό με το Microsoft SDK είναι μια αξιόπιστη συσκευή πολυτροπικής εισόδου με πολλές δυνατότητες και χαμηλό κόστος.

5 Κεφάλαιο

Χειρονομίες (Gestures)

Ενότητες κεφαλαίου

5.1	Χειρονομίες (Gestures)	33
5.2	Προτυποποίηση των χειρονομιών	33
5.3	Αρχές σχεδίασης διεπαφών με χειρονομίες	34
5.3.1	Το άγγιγμα του Μίδα	34
5.3.2	Πολιτισμικά θέματα	34
5.3.3	Σειρά εκτέλεσης επιμέρους ενεργειών	34
5.3.4	Σχεδίαση με βάση το πλαίσιο χρήσης	35
5.3.5	Ανάδραση	35
5.3.6	Ολοκλήρωση και Αποσαφήνιση	35
5.4	Ταξινόμηση χειρονομιών	35
5.5	Χειρονομίες που επιλέχθηκαν	36

5.1 Χειρονομίες (Gestures)

Οι χειρονομίες είναι προϊόν της αλληλεπίδρασης μεταξύ των ανθρώπων και αποτελούνται από κινήσεις του σώματος και του προσώπου, οι οποίες αντικαθιστούν την προφορική επικοινωνία. Οι χειρονομίες είναι απαραίτητες για την φυσική και παραστατική ανθρώπινη επικοινωνία. Η φυσικότητα είναι το ζητούμενο, αλλά κάποιες διεπαφές μπορεί να χρησιμοποιήσουν έντονες χειρονομίες γεγονός που αποτελεί βασικό σχεδιαστικό ζήτημα.

Η κύρια ταξινόμηση των χειρονομιών γίνεται με βάση την σημασία, την λειτουργία και την περιγραφή τους. Η ταξινόμιά τους βέβαια προχωρά σε μεγάλο βάθος διακρίνοντας πολλές κατηγορίες χειρονομιών που μπορούν να χρησιμοποιηθούν σε αντίστοιχες διεπαφές.

Οι διεπαφές με χειρονομίες διακρίνονται στις μηχανικές και απτικές διεπαφές, στις διεπαφές «υπολογιστικής όρασης» και στις διεπαφές που βασίζονται σε εκφράσεις του προσώπου. Ο τρόπος που επικοινωνούμε με αυτές τις διεπαφές χαρακτηρίζεται ως τεχνική αλληλεπίδρασης και αφορά την χειρονομία που κάνουμε και την απόκριση του συστήματος σε αυτή. Οι τεχνικές αλληλεπίδρασης που χρησιμοποιούνται στις διεπαφές με χειρονομίες σχετίζονται με την εγγύτητα, την κίνηση του σώματος, το δείξιμο, τον κυματισμό του χεριού, την εισαγωγή των χεριών, την περιστροφή αντικειμένου, το πάτημα με το πόδι, το ταρακούνημα και την κλίση του σώματος.

Η αλληλεπίδραση με διεπαφές είναι μια καινούργια πρόκληση στην βιομηχανία. Η πρόοδος που έχει γίνει στο μέγεθος, την κατανάλωση ενέργειας και το κόστος των επεξεργαστών, των μνημών, των καμερών και άλλων αισθητηριακών συσκευών έχει ανοίξει τον δρόμο στον έλεγχο τους με χειρονομίες και κινήσεις. Ένας καινούργιος κόσμος αλληλεπίδρασης δημιουργείται ή έστω αυτό ισχυρίζονται όλοι. Οι καινούργιες αλληλεπιδράσεις έχουν και νέο εμπορικό όνομα, χαρακτηρίζονται ως «φυσικές» και έτσι έχουμε τις «Φυσικές Διεπαφές Χρήστη».

Οι διεπαφές με χειρονομίες δεν είναι κάτι καινούργιο. Οι χειρονομίες χρησιμοποιούνται σε διεπαφές εδώ και πολύ καιρό. Η εργασία του Brad Myers (1998) κάνει μια ανασκόπηση σε συστήματα που κάνουν χρήση χειρονομιών τα οποία ήταν εμπορικά από το 1960. Στην πρωτοποριακή εργασία του Myron Krueger (1983) σχετικά με την εικονική πραγματικότητα, γίνεται μια εισαγωγή στην αλληλεπίδραση με την χρήση χειρονομιών πάνω σε μεγάλες προβαλλόμενες εικόνες. Τα συστήματα πολλαπλών σημείων επιλογής άρχισαν να εμφανίζονται την δεκαετία του 1980 και όπως αναφέρεται στην εργασία του Bill Buxton το πρώτο τέτοιο σύστημα δημιουργήθηκε το 1982. Εξειδικευμένοι αισθητήρες για τον εντοπισμό την ανθρώπινης θέσης και κίνησης εξετάζονται εδώ και πολύ καιρό από την βιομηχανία των παιχνιδιών. Επίσης μουσικά όργανα που χειρίζονται με χειρονομίες και πολύ-απτικές συσκευές εισόδου μουσικής αποτελούν πεδίο έρευνας και εμπορικής εφαρμογής εδώ και μισό αιώνα καθώς το The remin, ένα ηλεκτρονικό πιάνο που κάνει χρήση χειρονομιών προτυποποιήθηκε από τον δημιουργό του το 1928.

5.2 Προτυποποίηση των χειρονομιών

Οι περισσότερες χειρονομίες δεν είναι ούτε φυσικές ούτε είναι εύκολο να τις μάθεις και να τις θυμάσαι. Ακόμα και το απλό κούνημα του κεφαλιού είναι πολύπλοκο όταν περιπλέκονται πολλοί πολιτισμοί. Οι Δυτικοί που ταξιδεύουν στην Ινδία δυσκολεύονται να κατανοήσουν τον ινδικό τρόπο κουνήματος του κεφαλιού ο οποίος είναι μια διαγώνια μείξη του δυτικού κάθετου κουνήματος που σημαίνει «ναι» και το οριζόντιου κουνήματος που σημαίνει «όχι». Παρομοίως, οι χειρονομίες κουνήματος του χεριού για να πούμε «γεια», «αντίο», «έλα εδώ» κ.α. πραγματοποιούνται διαφορετικά σε διαφορετικές κουλτούρες.

5.3 Αρχές σχεδίασης διεπαφών με χειρονομίες

Ο σχεδιασμός των διεπαφών με βάση τις χειρονομίες είναι αρκετά δύσκολος, αφού η δημιουργία προτύπων χειρονομιών πηγάζει από τις φυσικές κινήσεις του ανθρώπου, οι οποίες είναι σε αρκετές περιπτώσεις αρκετά πολύπλοκες. Έτσι έχουν αναπτυχθεί πολλές μέθοδοι για τη σχεδίαση των διεπαφών που χρησιμοποιούν τις χειρονομίες σαν μέθοδο εισόδου, οι οποίες αναλύονται στις επόμενες υποενότητες.

5.3.1 Το άγγιγμα του Μίδα

1. Η διεπαφή πρέπει να καταλαβαίνει πότε να αρχίζει και πότε να σταματά την ερμηνεία μιας χειρονομίας.
2. Μπορούμε να καθορίσουμε κάποιες χειρονομίες ως εναρκτήριες και καταληκτικές για μια σύνοδο όπου το σύστημα θα αναγνωρίζει τις κινήσεις μας.
3. Όπως ο Μίδας δεν ήθελε ότι άγγιζε να γίνεται χρυσός, έτσι κι εμείς δεν θέλουμε όποτε κινούμαστε να ενεργοποιείται το σύστημα

5.3.2 Πολιτισμικά θέματα

1. Είναι γνωστό πως οι μη προφορική επικοινωνία εξαρτάται από πολιτισμικά στοιχεία.
2. Ο πολιτισμός μπορεί να επηρεάζει την σημασία της χειρονομίας, τον ρυθμό εκτέλεσης της και την ταχύτητά της.
3. Όταν μια χειρονομία δεν ταιριάζει σε κάποιον πολιτισμό τότε μπορούμε να προσαρμόσουμε την διεπαφή μας ανάλογα με τον χρήστη και να επιλέγει ο ίδιος τις χειρονομίες που του ταιριάζουν για να χρησιμοποιεί.

5.3.3 Σειρά εκτέλεσης επιμέρους ενεργειών

1. Καθορίζεται το κάθε βήμα που πρέπει να εκτελέσει ο χρήστης στην διεπαφή μας για να φέρει σε πέρας μια συγκεκριμένη διεργασία.

Παράδειγμα:

- Ο χρήστης κάνει μια χειρονομία για εισαγωγή και η χειρονομία εκτελείται στο σημείο τοποθέτησης του αντικειμένου
- Η διεπαφή επιστρέφει ένα σύνολο από αντικείμενα για εισαγωγή
- Ο χρήστης κάνει μια χειρονομία επιλογής και διαλέγει το αντικείμενο που θέλει
- Η διεπαφή αφαιρεί την λίστα αντικειμένων και εισάγει το αντικείμενο που επιλέχθηκε

5.3.4 Σχεδίαση με βάση το πλαίσιο χρήσης

1. Η εφαρμογή θα πρέπει να ερμηνεύει τις χειρονομίες σε σχέση με το πλαίσιο χρήσης.
 - Χωρικές ζώνες – Ο χώρος γύρω από τον χρήστη μπορεί να χωριστεί σε διάφορες ζώνες διάδρασης, όπως χώρος μενού, χώρος εργασίας, χώρος εντολών κ.α. Η θέση στην οποία γίνεται μια χειρονομία καθορίζει και το αποτέλεσμα της.
 - Τρόπος χρήσης – Μπορούμε να επαναχρησιμοποιούμε χειρονομίες με διαφορετικούς τρόπους. Όπως για παράδειγμα το ποντίκι στο Photoshop όπου ανάλογα με τον τρόπο που επιλέγουμε να χρησιμοποιηθεί μπορεί να ζωγραφίσει, να αποκόψει εικόνα κ.α.

5.3.5 Ανάδραση

1. Η χρήση χειρονομιών για την επικοινωνία με το σύστημα μπορεί να κάνει τον χρήστη να νιώσει αποκομμένος. Πρέπει επομένως να τροφοδοτούμε τον χρήστη με ανάδραση η οποία θα τον πληροφορεί ότι το σύστημα καταλαβαίνει την παρουσία και τις κινήσεις του.
2. Η ανάδραση του συστήματος προς τον χρήστη δεν πρέπει να είναι μόνο οπτική αλλά και ακουστική και απτική.

5.3.6 Ολοκλήρωση και Αποσαφήνιση

Η σχεδίαση μιας αποκλειστικά και μόνο διεπαφής χειρονομιών είναι δύσκολη εργασία. Η ταξινόμηση συνεργασίας όταν έχουμε πολλές τροπικότητες φαίνεται παρακάτω:

- Μεταφορά: Πρώτα εκτελείται η μια τροπικότητα και μετά η άλλη.
- Ισοδυναμία: Χρήση οποιασδήποτε τροπικότητας θα φέρει το ίδιο αποτέλεσμα.
- Συμπληρωματικότητα: Βάλε αυτό εκεί (Λες «βάλε» δείχνοντας το αντικείμενο και μετακινείς το χέρι σου στην τοποθεσία και λες «εκεί» για να ολοκληρωθεί η διαδικασία).
- Πλεονασμός: Πολλές ταυτόχρονες τροπικότητες.
- Ειδικότητα: Μόνο μια τροπικότητα έχει πρόσβαση σε κάποια λειτουργία.

5.4 Ταξινόμηση χειρονομιών

Οι χειρονομίες που χρησιμοποιούνται για την αλληλεπίδραση με μια φυσική διεπαφή είναι αρκετές. Κάποιες από αυτές είναι οι κινήσεις των δακτύλων, των χεριών, των ποδιών και γενικότερα του σώματος : Επίσης η κατεύθυνση του βλέμματος μπορεί να θεωρηθεί σαν χειρονομία καθώς και οι εκφράσεις του προσώπου.

Οι χειρονομίες μπορούν να ταξινομηθούν ως προς τη λειτουργία τη σημασία και την περιγραφή τους. Η σημασία έχει να κάνει με την επικοινωνία και το σκοπό της χειρονομίας. Η λειτουργία περιγράφει την επίδραση μια χειρονομίας πάνω σε μια διεπαφή και η περιγραφή έχει να κάνει με το πώς εκτελείται η χειρονομία, ποιες δηλαδή είναι οι απαραίτητες κινήσεις.

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

Με βάση τη σημασία των χειρονομιών, μπορούμε να τις κατατάξουμε στις συνειδητές και στις αυθόρμητες. Συνειδητές είναι οι χειρονομίες οι οποίες έχουν νόημα ακόμα και αν δεν ακολουθούνται από ομιλία. Αντίθετα οι αυθόρμητες πλαισιώνονται και από την ομιλία ούτως ώστε να έχουν νόημα. Σε μια φυσική διεπαφή οι συνειδητές χειρονομίες παίζουν καθοριστικό ρόλο, ενώ για τις αυθόρμητες χρειάζονται πολυτροπικές διεπαφές για να υποστηρίξουν και την ομιλία σαν είσοδο.

Με βάση τη λειτουργία των χειρονομιών μπορούμε να τις ταξινομήσουμε σε χειρονομίες εντολών, δεικτικές, χειρισμού και ελέγχου. Οι χειρονομίες εντολών δίνουν πρόσβαση στις λειτουργίες συστήματος. Οι δεικτικές χειρονομίες χρησιμεύουν στην επιλογή των αντικειμένων πάνω σε μια διεπαφή. Με τη σειρά τους οι χειρονομίες χειρισμού είναι απαραίτητες για τη διαχείριση και επεξεργασία δισδιάστατων ή τρισδιάστατων αντικειμένων, όπως η περιστροφή μιας 3Dεικόνας. Τέλος, οι χειρονομίες ελέγχου έχουν να κάνουν με τον έλεγχο ενός αντικειμένου, όπως μια κάμερα ή ένα Avatar.



Εικόνα 8. Εδώ απεικονίζεται μία χειρονομία ελέγχου για τον έλεγχο ενός avatar.

Τέλος, με βάση την περιγραφή των χειρονομιών υπάρχει η ταξινόμησή σε τρεις κατηγορίες.

- Στατικές, οι οποίες προϋποθέτουν συγκεκριμένη θέση των χεριών και των δακτύλων χωρίς να λαμβάνεται υπόψη κάποια κίνηση.
- Δυναμικές, οι οποίες βασίζονται στην κίνηση χωρίς να παίζει ρόλο ο χρόνος.
- Χώρο-χρονικές, οι οποίες στην ουσία είναι υποκατηγορία των δυναμικών και σχετίζονται με την συνεχή κίνηση μέσα σε ένα συγκεκριμένο χρονικό διάστημα.

5.5 Χειρονομίες που επιλέχθηκαν

Στην πολυτροπική εφαρμογή που αναπτύχθηκε για την παρούσα πτυχιακή εργασία χρησιμοποιήθηκαν μη στατικές χειρονομίες. Το σημείο ενδιαφέροντος είναι η αλλαγή κατεύθυνσης των χεριών, η οποία παρακολουθείται και καταγράφεται. Με αυτόν τον τρόπο, όταν για παράδειγμα

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

το δεξί χέρι του χρήστη αλλάζει κατεύθυνση από αριστερά προς τα δεξιά τότε αυτή η αλλαγή θεωρείται από το σύστημα ως χειρονομία – εντολή.

Πιο συγκεκριμένα, ανά δέκα frames αποθηκεύεται η θέση του καρπού του δεξιού χεριού. Αυτή η θέση θεωρείται ότι είναι η προηγούμενη η οποία ελέγχεται πάντα με την εκάστοτε τωρινή θέση του καρπού. Με έναν μαθηματικό αλγόριθμο υπολογίζεται η διαφορά μεταξύ προηγούμενης και τωρινής θέσης και έτσι υπολογίζεται η αλλαγή της κατεύθυνσης. Στη συνέχεια αφού έχει αναγνωριστεί η χειρονομία, το σύστημα μπαίνει σε κατάσταση αγνόησης χειρονομιών για ένα δευτερόλεπτο, ούτως ώστε η επαναφορά του χεριού στην αρχική κατάσταση να μην θεωρηθεί ως μία ακολουθούμενη χειρονομία του χρήστη.

Με αυτό τον τρόπο γίνεται η αναγνώριση των μη στατικών χειρονομιών από το σύστημα, οι οποίες έχουν σχεδιαστεί με τέτοιο τρόπο ούτως ώστε να είναι εύκολες στην υλοποίησή του από το χρήστη.

6 Κεφάλαιο

Τεχνολογίες και εργαλεία υλοποίησης

Ενότητες κεφαλαίου

6.1	Εισαγωγή κεφαλαίου	39
6.2	Microsoft VisualC#	39
6.3	MySQL	40
6.4	Microsoft SDK	40
6.4.1	Χαρακτηριστικά του Kinect SDK	41
6.4.2	Ανάλυση της λειτουργίας Skeleton Tracking	42

6.1 Εισαγωγή κεφαλαίου.

Στο κεφάλαιο αυτό θα αναλυθούν διεξοδικά όλες οι τεχνολογίες που χρησιμοποιήθηκαν για την τελική υλοποίηση της διεπαφής. Ο κύριος κορμός της εφαρμογής υλοποιήθηκε με τη C# σαν γλώσσα προγραμματισμού και συγκεκριμένα το .NET framework. Επίσης χρησιμοποιήθηκε το Microsoft SDK for Kinectσαν εργαλειοθήκη για την αξιοποίηση των λειτουργιών που προσφέρει το Kinect. Η βάση δεδομένων που επικοινωνεί με την εφαρμογή υλοποιήθηκε με τη χρήση της Mysql.

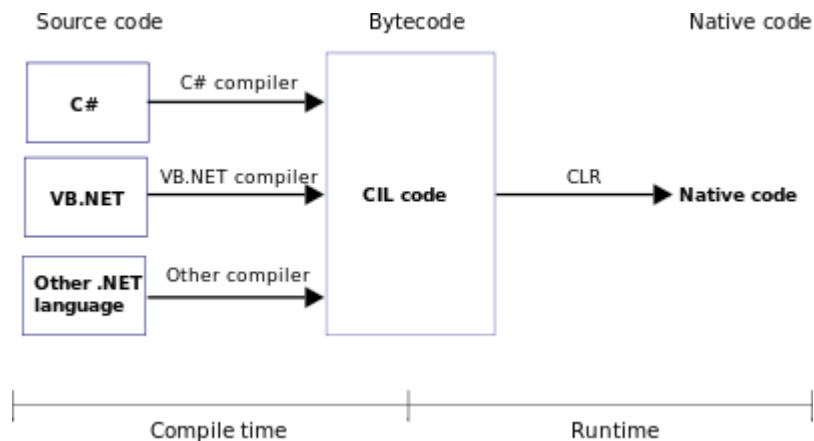
6.2 Microsoft VisualC#

Τον Ιανουάριο του 1999, ο Anders Heijlberg διαμόρφωσε μια ομάδα για την δημιουργία μιας νέας γλώσσας προγραμματισμού, με στόχο η γλώσσα αυτή να είναι μία αντικειμενοστραφή γλώσσα που να μοιάζει με τη C. Η αρχική ονομασία της νέας αυτής γλώσσας ήταν Cool, όμως η Microsoftπέλεξε να μην κρατήσει αυτό το όνομα για εμπορικούς λόγους. Με την δημοσιοποίηση του .NET Framework τον Ιούλιο του 2000 στο Professional Developers Conference η γλώσσα μετονομάστηκε σε C Sharp.

Τα κυριότερα χαρακτηριστικά της C#είναι τα εξής [11]:

- Strong Typing. Θέτει κανόνες για το πώς διαχειρίζονται δεδομένα διαφορετικού τύπου.
- Imperative. Είναι επιτακτική γλώσσα.
- Declarative. Είναι δηλωτική γλώσσα, εκφράζει δηλαδή τη λογική ενός υπολογισμού χωρίς να περιγράφει τη ροή ελέγχου.
- Functional. Είναι λειτουργική, αντιμετωπίζει δηλαδή τους υπολογισμούς σαν μια εξέλιξη μαθηματικών συναρτήσεων.
- Generic.
- Object-oriented programming. Είναι αντικειμενοστραφής γλώσσα, κάτι που είναι και η βασικότερη ιδιότητα της γλώσσας.
- Component-oriented.

Εδώ χρειάζεται να γίνει αναφορά στο CLR (Common Language Runtime). ΤοCLR εικονική μηχανή (virtual machine) του NET frameworkτης Microsoft.Ο ρόλος του είναι η διαχείριση και η εκτέλεση των προγραμμάτων που έχουν δημιουργηθεί με το συγκεκριμένο framework. Ο κώδικας που γίνεται compiledμετατρέπεται μέσω του CLRσε CIL(Common Intermediate Language) κώδικα. Ο CILκώδικας είναι μία αντικειμενοστραφής assemblyγλώσσα και είναι η χαμηλότερη επιπέδου γλώσσα της Microsoft. Στην συνέχεια αυτός ο κώδικας μετατρέπεται μέσω του CLRσε εντολές μηχανής (machine instructions), οι οποίες τελικά εκτελούνται από τον επεξεργαστή.



Εικόνα 9. Απεικόνιση της λειτουργίας του CLR και το πώς μετατρέπεται Byte code σε Native code

6.3 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. Συνεπώς η MySQL είναι ένας πολυνηματικός διακομιστής πολλαπλών χρηστών. Χρησιμοποιεί την SQL (Structured Query Language) την τυπική γλώσσα ερωτημάτων για βάσεις δεδομένων και υποστηρίζει τα τελευταία πρότυπα της αυτής της γλώσσας. Η MySQL είναι διαθέσιμη από το 1996 αλλά η ιστορία της ξεκινά από το 1979 .

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 για να αποφευχθούν τα 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.

6.4 Microsoft SDK

Το Kinect SDK είναι ένα σύνολο εργαλείων για την ανάπτυξη εφαρμογών. Αυτό το SDK παρέχει το κατάλληλο περιβάλλον για τον χειρισμό του Kinect μέσω του προγραμματισμού. Το Kinect for Windows SDK περιλαμβάνει προγράμματα οδήγησης υπολογιστή.

Η πρώτη έκδοση του Kinect SDK δημοσιεύτηκε τον Ιούνιο του 2011 [12] από τη Microsoft για μη εμπορικούς σκοπούς. Αυτή η έκδοση περιελάμβανε τους κατάλληλους οδηγούς του Kinect για το λειτουργικό σύστημα Windows 7 (παλαιότερες εκδόσεις λειτουργικών συστημάτων της Microsoft δεν υποστηρίζονται). Το Microsoft SDK επιτρέπει στους προγραμματιστές να αναπτύξουν εφαρμογές με τη βοήθεια του Visual Studio, το οποίο είναι και το επίσημο περιβάλλον ανάπτυξης (IDE) της Microsoft, χρησιμοποιώντας τις εξής γλώσσες προγραμματισμού :

- C++
- C#
- Visual Basic

6.4.1 Χαρακτηριστικά του Kinect SDK

Τα χαρακτηριστικά του SDK έχουν να κάνουν με την επεξεργασία του ήχου που δέχεται σαν είσοδο η συσκευή αλλά και με τη διαδικασία Skeleton Tracking και εγγραφής video. Παρακάτω γίνεται μια συνοπτική αναφορά και επεξήγηση των χαρακτηριστικών αυτών.

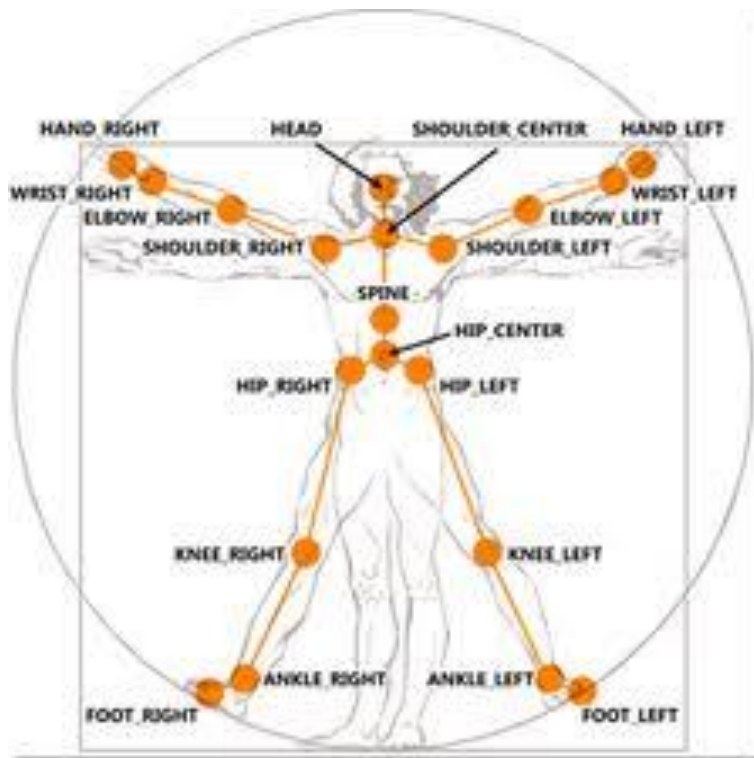
- Επεξεργασία δεδομένων από τους αισθητήρες του Kinect. Οι προγραμματιστές μπορούν να πάρουν και να επεξεργαστούν κατάλληλα τα δεδομένα από τους διάφορους αισθητήρες του Kinect (κάμερα βάθους- RGB κάμερα και μικρόφωνα).
- Skeleton Tracking. Αυτή είναι ίσως η σημαντικότερη δυνατότητα που δίνει το Kinect SDK. Με αυτή τη δυνατότητα μπορούν οι προγραμματιστές να ανιχνεύσουν και να παρακολουθήσουν την κίνηση (move tracking) έως δύο ανθρώπων. Αυτό θα αναλυθεί περαιτέρω στην επόμενη ενότητα.
- Ανίχνευση και επεξεργασία ήχου. Κάποιες από τις λειτουργίες του Kinect που προσφέρει το SDK όσον αφορά τον ήχο είναι το noise suppression (αποθρομβοποίηση) και το echo cancelation (καταστολή της ηχούς). Αυτές οι λειτουργίες χρησιμοποιούνται για την προσδιορισμό της προέλευσης του ήχου, υπάρχει δηλαδή η δυνατότητα να αναγνωριστεί κατά προσέγγιση από ποιο σημείο της σκηνής προέρχεται ο ήχος. Επίσης υπάρχει η δυνατότητα για αναγνώριση φωνής (voice recognition) αναγνωρίζοντας μέχρι τώρα τις εξής γλώσσες : Αγγλικά, Γαλλικά, Ισπανικά, Γιαπωνέζικα, Ιταλικά, και τα Ιταλικά. Μάλιστα στην περίπτωση των Αγγλικών Γαλλικών και Ισπανικών υποστηρίζονται και κάποιες από τις διαλέκτους των γλωσσών αυτών (όπως τα Αγγλικά που χρησιμοποιούνται στην Αγγλία και τα Αγγλικά που μιλιούνται στις Ηνωμένες Πολιτείες Αμερικής).
- Επίσης, ένα σημαντικό χαρακτηριστικό του SDK for Kinect είναι η τεκμηρίωσή του, η οποία είναι αρκετά λεπτομερής και επεξηγηματική με πολλά παραδείγματα σε επίπεδο προγραμματισμού αλλά και λειτουργιών του Kinect.

Το Microsoft SDK for Kinect μπορεί κάποιος να το βρει και να το κατεβάσει δωρεάν από τη σελίδα της Microsoft. Όμως οι προγραμματιστές μπορούν να το χρησιμοποιήσουν με τον όρο που η ίδια η εταιρεία θέτει και ο οποίος είναι ότι οι εφαρμογές που θα αναπτυχθούν με αυτό το εργαλείο θα πρέπει να μην είναι εμπορικές. Η Microsoft το διαθέτει μόνο για ερευνητικούς σκοπούς.

Πλέον μέχρι και την περίοδο που γράφεται αυτή η πτυχιακή, το SDK υποστηρίζεται από τα Windows 7 και τα Windows 8. Οι ελάχιστες απαιτήσεις συστήματος σε επίπεδο hardware είναι 32-bit επεξεργαστής διπύρηνος με πάνω από τα 2.66 Ghz χρονισμό και πάνω από 2 GB σε μνήμη RAM. Επίσης η κάρτα γραφικών πρέπει να υποστηρίζει DirectX 9 και φυσικά να υπάρχει είσοδος USB 2.

6.4.2 Ανάλυση της λειτουργίας Skeleton Tracking

Το Skeleton Tracking είναι μία από τις κυριότερες λειτουργίες που μας δίνει το SDK for Kinect. Επιτρέπει στο Kinect να αναγνωρίσει ανθρώπινα σώματα και να ακολουθήσει τις δράσεις τους και την κίνησή τους. Χρησιμοποιώντας την κάμερα υπέρυθρης ακτινοβολίας (IR), το Kinect μπορεί να αναγνωρίσει έως έξι χρήστες στο οπτικό πεδίο του αισθητήρα. Από αυτούς τους έξι χρήστες, μέχρι δύο μπορεί το Kinect να παρακολουθεί τις κινήσεις τους με λεπτομέρεια. Με αυτόν τον τρόπο μπορεί να εντοπίσει τις αρθρώσεις των εντοπισμένων χρηστών στο χώρο και να παρακολουθεί τις κινήσεις τους κατά την διάρκεια του χρόνου. Τα σημεία των αρθρώσεων που αναγνωρίζονται φαίνονται στην παρακάτω εικόνα.



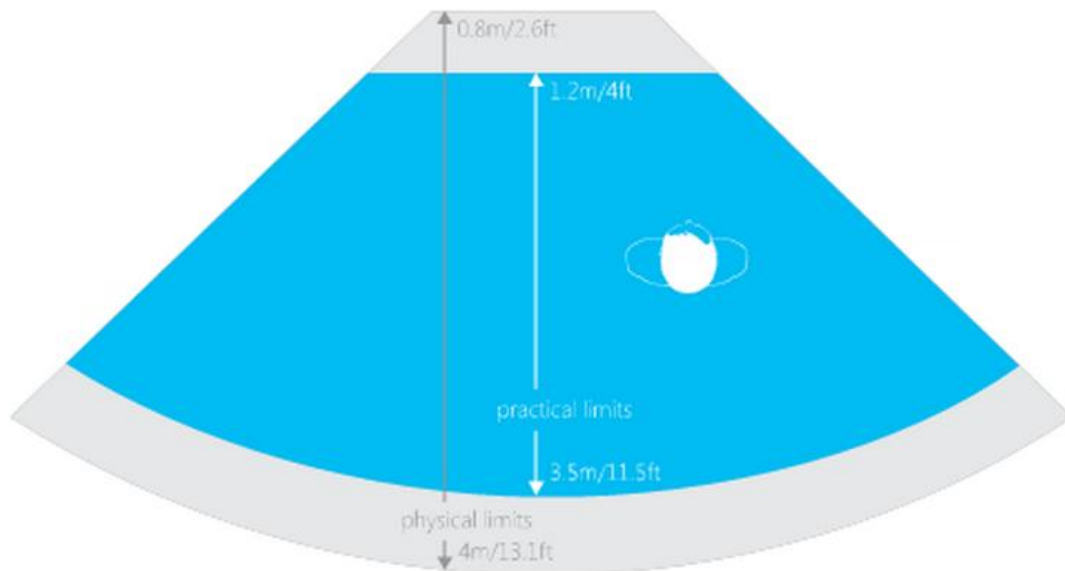
Εικόνα 10. Απεικόνιση των είκοσι αρθρώσεων που αναγνωρίζει το Kinect

Η ανίχνευση σκελετού σταδιακά έχει βελτιστοποιηθεί για να αναγνωρίζει χρήστες σε όρθια αλλά και καθιστή στάση. Σε όρθια στάση η συσκευή αναγνωρίζει είκοσι αρθρώσεις ενώ σε καθιστή στάση μόνο δέκα, αφού όσες από αυτές τις αρθρώσεις βρίσκονται κάτω από το Shoulder Center δεν είναι ορατές στη στάση αυτή.

Για να αναγνωριστεί ένας χρήστης, χρειάζεται απλώς να είναι μπροστά από τον αισθητήρα και απλά ο αισθητήρας να μπορεί να δει το κεφάλι του και το πάνω μέρος του σώματός του. Με το Kinect SDK δεν χρειάζεται καμία συγκεκριμένη πόζα ή δράση βαθμονόμησης για να αναγνωριστεί κάποιος χρήστης.

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

Στη λειτουργία default range mode, το Kinect μπορεί να αναγνωρίσει ανθρώπους που στέκονται μεταξύ 0,8 μέτρα και 4,0 μέτρα μακριά από τον αισθητήρα. Βέβαια η πρακτική απόσταση για την ομαλή λειτουργία του tracking είναι από 1,2 έως 3,5 μέτρα περίπου.



Εικόνα 11. Κάθετο πεδίο αναγνώρισης του Kinect

Στη λειτουργία near range mode, η οποία υποστηρίζεται από την έκδοση 1.5 του SDK, οι σκελετοί των χρηστών μπορούν να αναγνωριστούν από πιο κοντινή απόσταση, μεταξύ 0,4 και 3 μέτρα. Εδώ χρειάζεται να ειπωθεί ότι την λειτουργία αυτή δεν την υποστηρίζει το Kinect for Xbox αλλά μόνο το Kinect for PC, το οποίο κοστολογείται περίπου στην διπλάσια τιμή του πρώτου (\$249 την παρούσα περίοδο).

Σε συνδυασμό λοιπόν του Skeleton Tracking και των αρθρώσεων που προκύπτουν από αυτή τη λειτουργία γίνεται δυνατή η χρήση των χειρονομιών σαν βασικό εργαλείο για τον καθορισμό εντολών φυσικού τύπου από τους χρήστες. Για παράδειγμα, αφού είναι δυνατή η παρακολούθηση του καρπού ενός χρήστη στη περίοδο του χρόνου, μπορεί να καθοριστεί μια εντολή που θα αντιστοιχεί στην μετακίνηση του χεριού του χρήστη από δεξιά προς τα αριστερά μέσα σε ένα συγκεκριμένο χρονικό διάστημα.

7 Κεφάλαιο

Προσδιορισμό Συστήματος

Ενότητες κεφαλαίου

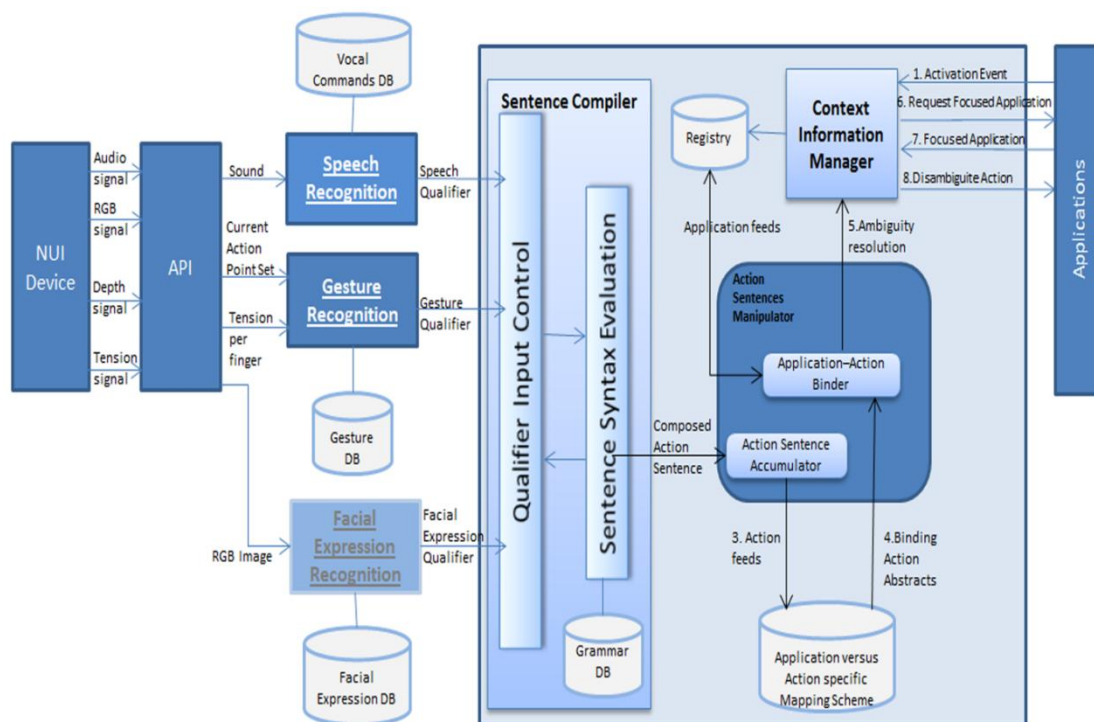
7.1	Εισαγωγή Κεφαλαίου	45
7.2	Ανάλυση της αρχιτεκτονικής:	45
7.2.1	NUI Devices-APIs	46
7.2.2	Speech Recognition	46
7.2.3	Gesture Recognition (Αναγνώριση Χειρονομιών)	46
7.2.4	Facial Expressions Recognition	47
7.2.5	Sentence Compiler	47
7.2.6	Action Sentence Manipulator	48
7.2.7	Context Information Manager (CIM)	48

7.1 Εισαγωγή Κεφαλαίου

Στις ενότητες 4, 5 και 6 αναλύθηκαν οι διάφορες πολυτροπικές συσκευές εισόδου, οι χειρονομίες σαν τρόπο επικοινωνίας μεταξύ ανθρώπου – υπολογιστή και οι διαθέσιμες τεχνολογίες που μπορούν να χρησιμοποιηθούν για την υλοποίηση φυσικών διεπαφών χρήστη υπολογιστή. Σε αυτό το κεφάλαιο θα γίνει ανάλυση της υλοποίησης μιας τέτοιας διεπαφής με τη χρήση του Microsoft Kinect σαν συσκευή εισόδου και της τεχνολογίας Microsoft SDK for Kinect σαν βιβλιοθήκη χειρισμού αυτής της συσκευής.

7.2 Ανάλυση της αρχιτεκτονικής:

Η εφαρμογή που αναπτύχθηκε στα πλαίσια της παρούσας πτυχιακής εργασίας είναι μια πολυτροπική διεπαφή χρήστη υπολογιστή. Βασίστηκε σε προηγούμενη πτυχιακή του Ευθύμιου Συντυχάκη και σκοπός της είναι η εξέλιξη και βελτίωση του MIDAS (Multimodal Interface Directed by Action Sentences) [13]. Το MIDAS είναι ένα σύστημα πολυτροπικής διεπαφής χρήστη, το οποίο χρησιμοποιεί τις κινήσεις του σώματος όπως οι χειρονομίες αλλά και την ομιλία σαν εισόδους για την αλληλεπίδραση του χρήστη με το σύστημα. Κάθε τέτοια είσοδος αναγνωρίζεται σαν μια λέξη (word). Η αρχιτεκτονική του φαίνεται στο παρακάτω σχήμα.



Εικόνα 12. Απεικόνιση αρχιτεκτονικής MIDAS.

Το σύστημα αποτελείται από επιμέρους λειτουργικές μονάδες (modules). Αυτές οι μονάδες σύμφωνα με την φιλοσοφία που σχεδιάστηκε το σύστημα είναι ανεξάρτητες μεταξύ τους. Είναι η λογική του Black Box, σύμφωνα με την οποία κάθε αυτόνομη μονάδα επικοινωνεί με κάποιες από τις υπόλοιπες μέσω μηνυμάτων εισόδου. Μια τυχόν αλλαγή που θα γίνει σε μία μονάδα δεν θα πρέπει να επηρεάζει τη δομή μιας άλλης μονάδας. Παρακάτω γίνεται μια πιο λεπτομερής ανάλυση των επιμέρους μονάδων του συστήματος.

7.2.1 NUI Devices–APIs

Οι συσκευές εισόδου (NUI Devices) είναι οι συσκευές που μπορούν να χρησιμοποιηθούν από το σύστημα για γίνει η καταγραφή των φυσικών τρόπων επικοινωνίας των χρηστών μιας και μιλάμε για μια φυσική διεπαφή. Έτσι αυτές οι συσκευές μπορεί να είναι διάφοροι αισθητήρες όπως αισθητήρας πίεσης, μια κάμερα ή ένας αισθητήρας βάθους.

Στη συνέχεια υπάρχει το module API. Το συγκεκριμένο module της αρχιτεκτονικής αντιπροσωπεύει τις διάφορες βιβλιοθήκες που υποστηρίζουν την κάθε συσκευή εισόδου, όπως είναι για παράδειγμα το Microsoft SDK for Kinect που αναλύθηκε παραπάνω. Τα APIs δίνουν τις κατάλληλες εισόδους στα επόμενα modules του συστήματος, Speech Recognition, Gesture Recognition, Facial Expression Recognition.

7.2.2 Speech Recognition

Στο module Speech Recognition γίνεται η αναγνώριση και αποτύπωση του προφορικού λόγου σε κείμενο. Αυτό που στην ουσία γίνεται είναι η αναγνώριση κάποιων λέξεων που προφέρονται από το χρήστη και η αποτύπωση αυτών των λέξεων σε κείμενο. Με αυτό τον τρόπο μπορεί το σύστημα να αναγνωρίσει τις φωνητικές εντολές των χρηστών. Οι φωνητικές εντολές αυτές είναι αποθηκευμένες σε μια βάση δεδομένων. Αν μια λέξη δεν περιέχεται μέσα σε αυτή τη βάση τότε το σύστημα δεν την θεωρεί γνωστή και την απορρίπτει σε περίπτωση που ειπωθεί από το χρήστη. Η λέξη που αναγνωρίζεται και στη συνέχεια στέλνεται σαν είσοδος στον Sentence Compiler, επιλέγεται από τον προγραμματιστή. Αυτό σημαίνει ότι για κάθε λέξη που προφέρει ο χρήστης μπορεί να αντιστοιχεί σε παραπάνω από μία εντολή (Speech Qualifier).

Αυτό γίνεται διότι η λειτουργία αναγνώρισης ομιλίας δεν εκπαιδεύεται από το σύστημα. Όλες λοιπόν οι λέξεις που βγαίνουν σαν έξοδο από το Speech Recognition είναι προκαθορισμένες. Βέβαια υπάρχει η δυνατότητα, όπως θα αναλυθεί σε επόμενη ενότητα, να γίνει αλλαγή στο περιεχόμενο της βάσης των προφορικών εντολών τη στιγμή που τρέχει η διεπαφή ανάλογα με την κατάσταση που βρίσκεται εκείνη τη στιγμή το σύστημα. Αυτή η αλλαγή έχει σκοπό να αποκλείσει κάποιες λέξεις τις οποίες ο χρήστης δεν τις χρειάζεται σε μια δεδομένη στιγμή αλληλεπίδρασης με τη διεπαφή και έτσι να αποφεύγεται η σύγχυση μεταξύ κάποιων λέξεων που μοιάζουν μεταξύ τους, π.χ. το round με το ground έχουν παρόμοια προφορά.

7.2.3 Gesture Recognition (Αναγνώριση Χειρονομιών)

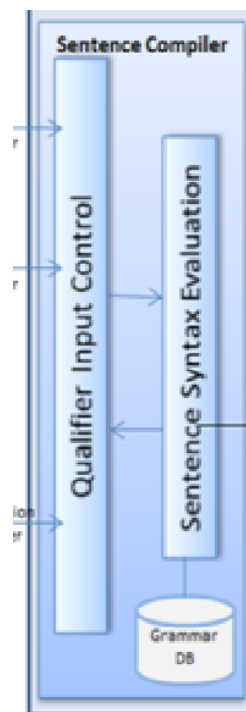
Η αναγνώριση χειρονομιών καταγράφει τις χειρονομίες και γενικότερα τις κινήσεις του σώματος και τις μετατρέπει κι αυτές σε κείμενο. Οι χειρονομίες αυτές μπορεί να είναι οι κινήσεις των χεριών ή και άλλων μερών του σώματος όπως τα πόδια ή τα δάκτυλα των χεριών. Κάθε χειρονομία που θέλουμε να αναγνωρίζεται από το σύστημα είναι καταγεγραμμένη από πριν και έχει αντιστοιχηθεί σε ένα gesturequalifier όπως αντίστοιχα έχει γίνει με τη λειτουργία του Speech Recognition. Η καταγραφή, η αντιστοίχιση και η αποθήκευση των χειρονομιών και των αντίστοιχων gesturequalifier γίνονται σε μία βάση δεδομένων, την GestureDB όπως φαίνεται και από το σχήμα. Εδώ έχουμε την απόδοση κάθε χειρονομίας σε μία λέξη (gesture qualifier). Αυτή η λέξη αποτελεί την έξοδο του Gesture Recognition και την είσοδο του Sentence Compiler.

7.2.4 Facial Expressions Recognition

Οι εκφράσεις του προσώπου αποτελούν κι αυτές ένα είδος μη-λεκτικής επικοινωνίας του ανθρώπου όπως είναι και οι χειρονομίες του σώματος. Στο σύστημα γίνεται η αναγνώριση αυτών των εκφράσεων του χρήστη, όπως είναι οι κινήσεις των φρυδιών ή των χειλιών και στη συνέχεια αυτές αποτυπώνονται σε μία λέξη (facial expression qualifier) με τον ίδιο τρόπο που γίνεται και για τα υπόλοιπα δύο είδη αναγνώρισης όπως αναλύθηκαν παραπάνω. Κάθε έκφραση προσώπου αντιστοιχίζεται πάλι σε μια λέξη. Η αντιστοίχιση αυτή είναι αποθηκευμένη σε μία βάση δεδομένων, την FacialExpressionDB, η οποία είναι και αυτή προκαθορισμένη από το σύστημα. Έπειτα το facial expression qualifier περνάει σαν είσοδος στον Sentence Compiler.

7.2.5 Sentence Compiler

Στον Sentence Compiler καταλήγουν όλες οι εντολές που προέρχονται από το τις συσκευές εισόδου του πολυτροπικού συστήματος. Ο ρόλος του είναι η σύνταξη προτάσεων δράσεις (action sentence) οι οποίες στη συνέχεια στέλνονται στον Action Sentence Manipulator.



Εικόνα 13. Απεικόνιση του Sentence Compiler.

Το module Sentence Compiler αποτελείται από δύο κύρια μέρη. Το πρώτο μέρος είναι το Qualifier Input Control. Εκεί καταλήγουν καταρχάς οι τα διάφορα είδη εισόδων. Ο τρόπος με τον οποίο γίνεται η υποδοχή των εισόδων είναι παράλληλος και συνεχόμενος. Όταν, για παράδειγμα, εντοπιστεί μια χειρονομία ή μια φωνητική εντολή, αρχίζει μια διαδικασία σύνταξης μιας πρότασης δράσης. Η εντολή που εντοπίστηκε περνά στο Sentence Syntax Evaluation, το δεύτερο μέρος, ως εν δυνάμει πρόταση για να ελεγχθεί εάν υπάρχει στη Grammar database. Η συγκεκριμένη βάση είναι υλοποιημένη με το σχεσιακό σύστημα διαχείρισης βάσεων δεδομένων MySQL και περιέχει όλες τις δομημένες προτάσεις. Οι προτάσεις αυτές αποτελούνται από τις λέξεις-εντολές που προκύπτουν από τους διαφορετικούς Qualifiers που αναλύθηκαν παραπάνω.

Στην συνέχεια γίνεται ο έλεγχος από τον Sentence Syntax Evaluation. Αν η πρόταση που ελέγχεται αποτελεί μέρος σε τουλάχιστον μια από τις δομημένες προτάσεις που βρίσκεται στην βάση, τότε η εντολή αυτή κρατιέται. Η ροή του συστήματος γυρίζει πίσω στον Qualifier Input Control, όπου αναμένεται μια καινούρια εντολή. Όταν έρθει η επόμενη εντολή γίνεται εκ νέου η παραπάνω διαδικασία, μέχρι να καταλήξει το σύστημα στην τελική ολοκληρωμένη πρόταση. Αν μια εντολή-λέξη δεν είναι αυτή που αναμενόταν από τον Sentence Syntax Evaluation, δεν αποτελεί δηλαδή μέρος μιας ολοκληρωμένης πρότασης γενικά ή δεν ταιριάζει με την προηγούμενη εντολή για την δόμηση μιας ολοκληρωμένης πρότασης, τότε η εντολή απορρίπτεται από το σύστημα και περιμένει την επόμενη. Αφού λοιπόν βρεθεί μια ολοκληρωμένη πρόταση, τότε αυτή στέλνεται στον Action Sentence Manipulator.

7.2.6 Action Sentence Manipulator

Στον Action Sentence Manipulator καταλήγουν οι ολοκληρωμένες προτάσεις που προέκυψαν από την διαδικασία που αναφέρθηκε παραπάνω. Ο ρόλος του Manipulator στην ουσία είναι να χειριστεί τις ολοκληρωμένες προτάσεις, αντιμετωπίζοντάς τις πλέον σαν Actions πάνω σε μια εφαρμογή.

Πιο συγκεκριμένα, ελέγχει την ολοκληρωμένη πρόταση που δέχθηκε από τον Sentence Compiler μέσα από την βάση του συστήματος. Η σύγκριση έχει να κάνει με το αν η συγκεκριμένη πρόταση αντιστοιχεί σε μια συγκεκριμένη δράση (action) πάνω σε μια συγκεκριμένη εφαρμογή (application). Αν συμβαίνει αυτό, τότε γίνεται έλεγχος για τον αν η συγκεκριμένη εφαρμογή είναι focused στο λειτουργικό σύστημα και μόνο τότε εφαρμόζεται το συγκεκριμένο action στην εφαρμογή. Αυτό γίνεται μέσω του Context Information Manager, του οποίου η λειτουργία θα αναλυθεί παρακάτω.

Βέβαια στον έλεγχο που αναφέρθηκε μέσω του CIM, υπάρχουν και εξαιρέσεις, μιας και μερικές δράσεις –actions δεν προϋποθέτουν τη focused κατάσταση της εφαρμογής ή ακόμα και τη running. Τέτοιες εντολές θα μπορούσε να είναι για παράδειγμα η open chrome, η οποία αντιστοιχεί στο άνοιγμα του περιηγητή google chrome. Όπως γίνεται κατανοητό, τέτοιου είδους εντολές δεν εξαρτώνται από την κατάσταση (state) της εφαρμογής και έτσι στην ουσία παρακάμπτεται ο Context Information Manager.

Αν μια εντολή δράσης πάνω σε μια εφαρμογή δεν μπορεί να εκτελεστεί λόγω του ότι αυτή δεν είναι focused από το εκάστοτε λειτουργικό σύστημα τότε αυτή δεν εκτελείται, προβάλλοντας το κατάλληλο μήνυμα στο χρήστη. Αν η εφαρμογή είναι σε κατάσταση focused τότε η εντολή εκτελείται κανονικά.

7.2.7 Context Information Manager (CIM)

Ο ρόλος του Context Information Manager είναι να επικοινωνεί με το λειτουργικό σύστημα. Λειτουργεί δηλαδή σαν ενδιάμεσος μεταξύ λειτουργικού συστήματος και διεπαφής. Ελέγχει επικοινωνώντας με το λειτουργικό σύστημα το ποιες εφαρμογές είναι ανοιχτές, αλλά και το ποια είναι η κατάσταση (state) των εφαρμογών αυτών.

Όλες αυτές οι πληροφορίες αποθηκεύονται δυναμικά στη μνήμη. Όταν χρειαστεί μια τέτοιου είδους πληροφορία ο Action Sentence Manipulator, τότε επικοινωνεί με τον Context Information Manager και ο τελευταίος παρέχει τη δεδομένη κατάσταση όλων ή μέρος των εφαρμογών αλλά και το ποιες εφαρμογές είναι διαθέσιμες εκείνη τη δεδομένη χρονική στιγμή. Αυτό γίνεται επικοινωνώντας ο CIM με το λειτουργικό σύστημα και τη Registry για να πάρει όλες τις πληροφορίες για όλες τις εφαρμογές και τις καταστάσεις τους. Αν η δράση σε μια εφαρμογή πληροί τις προϋποθέσεις προς εκτέλεση με βάση τις πληροφορίες που παρέχει ο CIM, τότε αυτή αποδοσμεύεται και εκτελείται, αλλιώς παρακάμπτεται.

Ένα παράδειγμα για την κατανόηση της λειτουργίας του CIM είναι για παράδειγμα αν δοθεί η εντολή reload firefox ο Action Sentence Manipulator θα καταλάβει ότι πρόκειται για το action reload πάνω στην εφαρμογή firefox. Προτού όμως δοθεί η εντολή προς εκτέλεση, ο ASM θα επικοινωνήσει με τον CIM. Αν ο CIM δεν βρει τη συγκεκριμένη εφαρμογή ανοιχτή στο λειτουργικό σύστημα, τότε το action δεν θα εκτελεστεί. Επίσης αν η εφαρμογή firefox είναι τρέχει αλλά παράλα αυτά δεν είναι focused σύμφωνα με τον CIM, τότε πάλι η εφαρμογή δεν θα εκτελεστεί και συνεχιστεί η ροή του προγράμματος κανονικά. Η μόνη περίπτωση για να γίνει το reload θα είναι ο περιηγητής firefox να είναι ανοικτός και focused, κάτι που ο CIM έχει την ευθύνη να ελέγξει. Έτσι γίνεται κατανοητό ότι ο Context Information Manager παίζει πολύ σημαντικό ρόλο στο σύστημα, αφού είναι ο μόνος διάυλος επικοινωνίας με το λειτουργικό σύστημα.

8 Κεφάλαιο

Σχεδιασμός Συστήματος

Ενότητες Κεφαλαίου

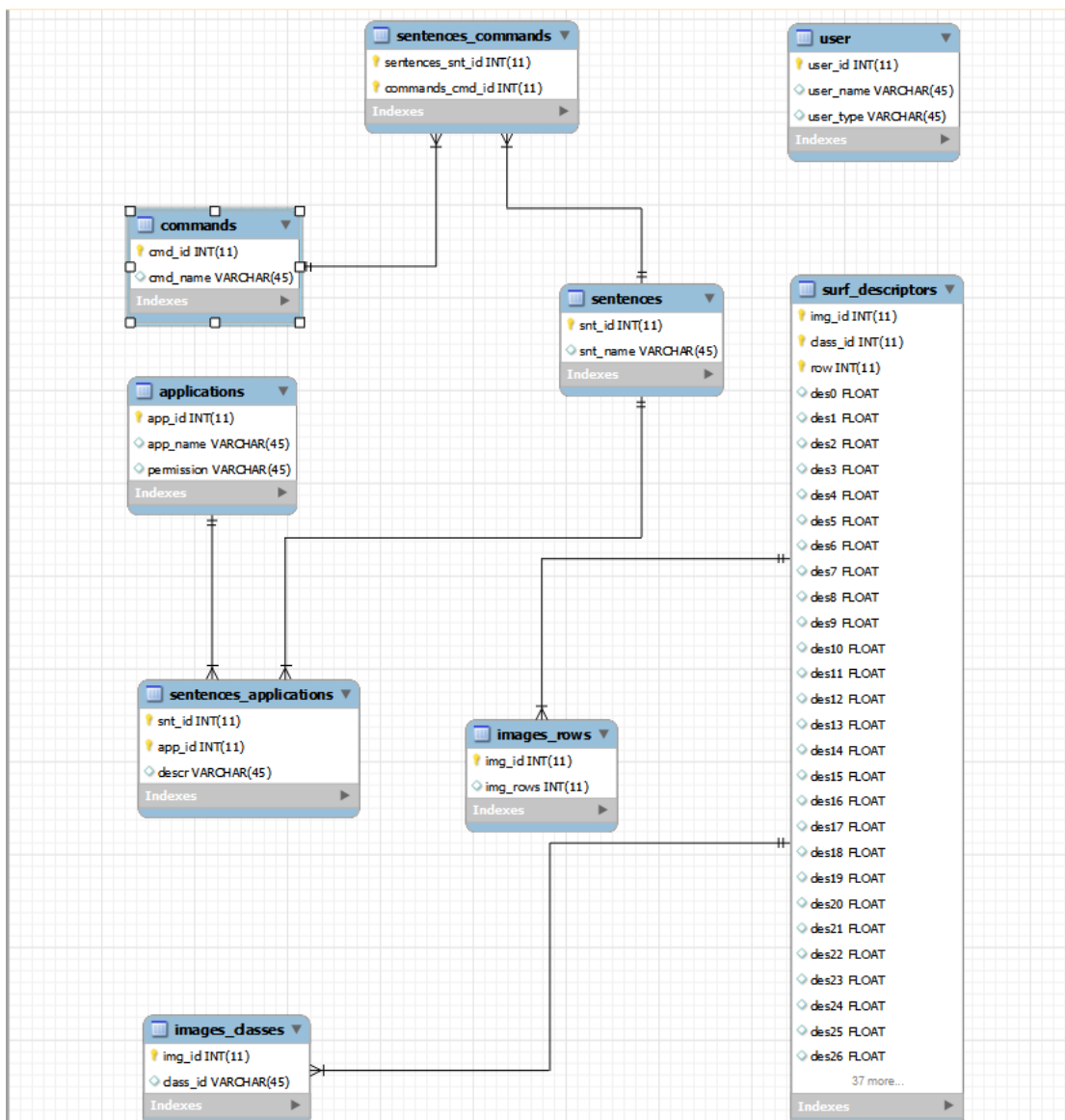
8.1	Εισαγωγή κεφαλαίου	51
8.2	Σχεσιακή Βάση Δεδομένων	51
8.2.1	Μοντέλο οντοτήτων συσχετίσεων	52
8.3	Διάγραμμα κλάσεων	52

8.1 Εισαγωγή κεφαλαίου

Σε αυτή την ενότητα θα αναλυθούν οι πίνακες σεναρίων, τα διαγράμματα σεναρίων χρήσης καθώς και τα διαγράμματα κλάσεων της εφαρμογής. Επίσης θα αποτυπωθεί η σχεδίαση της σχεσιακής βάσης δεδομένων με τη βοήθεια των μοντέλων οντοτήτων συσχετίσεων.

8.2 Σχεσιακή Βάση Δεδομένων

Πριν γίνει η τελική υλοποίηση της βάσης χρειάστηκε να γίνει το μοντέλο οντοτήτων συσχετίσεων με τη βοήθεια που μας παρέχει το εργαλείο Data Modeling του MySQL Workbench. Το μοντέλο που προέκυψε από την ανάλυση των απαιτήσεων για τη βάση της εφαρμογής φαίνεται στην παρακάτω εικόνα.



Εικόνα 14. Απεικόνιση του μοντέλου οντοτήτων συσχετίσεων της βάσης

8.2.1 Μοντέλο οντοτήτων συσχετίσεων.

Συνοπτικά οι οντότητες του μοντέλου είναι :

- Sentences. Είναι ο πίνακας όπου θα αποθηκεύονται όλες οι προκαθορισμένες ολοκληρωμένες προτάσεις του συστήματος.
- Applications. Είναι ο πίνακας που θα περιέχει τις διάφορες εφαρμογές του λειτουργικού συστήματος, τις οποίες ο χρήστης μπορεί να διαχειρίζεται μέσω των εντολών του.
- Commands. Εδώ θα είναι στην ουσία οι λέξεις (words) που θα περιέχονται σε κάθε ολοκληρωμένη πρόταση.
- Users. Είναι ο πίνακας που περιέχει τα στοιχεία κάθε χρήστη του συστήματος.
- Surf_descriptors. Είναι ο πίνακας που περιέχει δεδομένα για τα χαρακτηριστικά διάφορων εικόνων. Αυτές οι εικόνες θα χρησιμοποιηθούν σαν βάση σύγκρισης στη λειτουργία Object-Detection.
- Sentences_commands, sentences_applications, images_classes, images_rows. Αυτές οι επιμέρους οντότητες είναι κάποιιοι βοηθητικοί πίνακες, οι οποίοι προκύπτουν από τις συσχετίσεις μεταξύ των υπόλοιπων οντοτήτων μεταξύ τους. Χρειάζονται κυρίως για την εύκολη ανάκτηση κάποιων δεδομένων.

Από το παραπάνω μοντέλο προκύπτει το καθολικό σχήμα της βάσης του συστήματος MIDAS.

8.3 Διάγραμμα κλάσεων

Βάση της αρχιτεκτονικής του συστήματος που αναλύθηκε σε προηγούμενη ενότητα, για την υλοποίηση της διεπαφής σε προγραμματιστικό επίπεδο προέκυψε το διάγραμμα κλάσεων που φαίνεται στην παρακάτω εικόνα. Οι περισσότερες κλάσεις απεικονίζουν τις οντότητες της αρχιτεκτονικής, ενώ άλλες προέκυψαν κατά την διάρκεια της υλοποίησης της διεπαφής.



Εικόνα 15. Οι κλάσεις που προέκυψαν από την αρχιτεκτονική του MIDAS.

9 Κεφάλαιο

Υλοποίηση

Ενότητες κεφαλαίου

9.1	Εισαγωγή	55
9.2	Το κυρίως παράθυρο της εφαρμογής:	55
9.2.1	Listboxes	57
9.2.2	Οι ρυθμίσεις της εφαρμογής	58

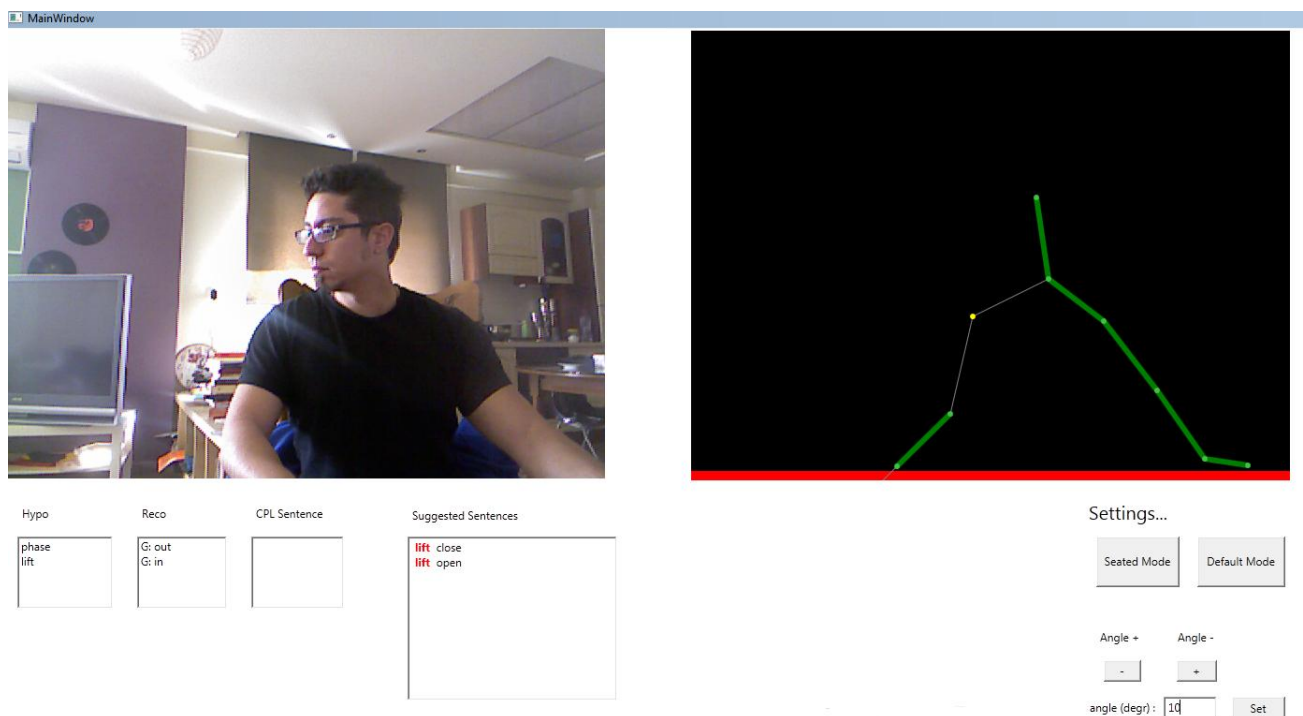
9.1 Εισαγωγή

Το παρόν σύστημα χρησιμοποιεί το Kinect για συσκευή εισόδου, αφού όπως προαναφέρθηκε είναι συσκευή πολυτροπικής εισόδου (multimodal input device). Το Kinect προσφέρει αναγνώριση ομιλίας και υποστηρίζει την αναγνώριση χειρονομιών (Gesture Recognition). Επίσης μέσω της ανάλυσης του βάθους και της RGBεικόνας που προσφέρει το Kinect μπορεί να γίνει αναγνώριση προσώπου (Face Recognition) και αναγνώρισης αντικειμένων (Object Detection).

Στις παρακάτω ενότητες αναλύεται ο τρόπος υλοποίησης των δύο σεναρίων με τη χρήση της C# σαν κύρια γλώσσα προγραμματισμού.

9.2 Το κυρίως παράθυρο της εφαρμογής:

Το κυρίως παράθυρο της εφαρμογής είναι υλοποιημένο με το client-side framework της Microsoft που ονομάζεται WPF (Windows Presentation Foundation). Αποτελεί ένα μοντέλο προγραμματισμού το οποίο έχει υλοποιηθεί εξ αρχής με τη λογική του πλήρη διαχωρισμού της διεπαφής από τη λογική του προγράμματος. Είναι ένα τεράστιο API το οποίο δίνει υποστήριξη σε υπηρεσίες όπως animation, έλεγχος media, 2D και 3D graphics, data-binding και typography. Παράλληλα, επιτρέπει τον πλήρη καθορισμό της μορφής που έχουν τα διάφορα controls. Τα πάντα περιγράφονται μέσω της XAML κι έτσι τα projects μπορούν να μοιραστούν σε ρόλους όπου για παράδειγμα ο developer ασχολείται με τη business λογική και ο designer με το οπτικό κομμάτι. Απαιτεί .NET Framework 3.0 ή νεότερης έκδοσης.



Εικόνα 16. Απεικόνιση του Main Window της εφαρμογής.

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

Όπως φαίνεται το main window της εφαρμογής περιέχει δύο viewbox, τέσσερα listbox, ένα text box και πέντε buttons. Το πρώτο viewbox απεικονίζει την RGB εικόνα που προέρχεται από την κάμερα του Kinect. Στο δεύτερο viewbox έχουμε την απεικόνιση του σκελετού, ο οποίος ανήκει στον παίχτη που έχει κάνει tracking το Kinect.

Η απεικόνιση του σκελετού έγινε με την ένωση των Joints που βρίσκει το Kinect. Δημιουργήθηκε η μέθοδος DrawBones And Joints μέσα στη κλάση Kinect_Skeleton_Manager, η οποία ενώνει τα joints που χρειάζονται ούτως ώστε να φτιαχτεί ο σκελετός του tracked player. Παρακάτω παραθέτουμε ένα μέρος κώδικα για την υλοποίηση αυτής της μεθόδου.

```
Private void DrawBonesAndJoints(Skeleton skeleton, DrawingContext drawingContext)
```

```
{
```

```
// Render Torso
```

```
this.DrawBone(skeleton, drawingContext, JointType.Head,
```

```
JointType.ShoulderCenter);
```

```
// Left Leg
```

```
this.DrawBone(skeleton, drawingContext, JointType.HipLeft,
```

```
JointType.KneeLeft);
```

```
this.DrawBone(skeleton, drawingContext, JointType.KneeLeft,
```

```
JointType.AnkleLeft);
```

```
this.DrawBone(skeleton, drawingContext, JointType.AnkleLeft,
```

```
JointType.FootLeft);
```

```
// Right Leg
```

```
this.DrawBone(skeleton, drawingContext, JointType.HipRight,  
JointType.KneeRight);
```

```
this.DrawBone(skeleton, drawingContext, JointType.KneeRight,  
JointType.AnkleRight);
```

```
this.DrawBone(skeleton, drawingContext, JointType.AnkleRight,  
JointType.FootRight);
```

Η κλάση Skeleton_Manager λοιπόν είναι υπεύθυνη για την απεικόνιση του σκελετού ενός player. Είναι subscribed στο event Skeleton FrameReady με έναν event handler. Το συγκεκριμένο event γίνεται fired για κάθε skeleton frame που είναι διαθέσιμο. Με αυτόν τον τρόπο ανανεώνεται η εικόνα με το καινούριο position των Joints και του σκελετού.

9.2.1 Listboxes

Αν ο χρήστης πει μια λέξη, τότε το Speech Recognition του Kinect μπορεί να μην την αναγνωρίσει κατευθείαν, αλλά να την υποθέσει. Η λέξη που έχει υποθέσει το Speech Recognition απεικονίζεται στο πρώτο ListBox με τίτλο Hypo.

Τώρα, αν ο χρήστης πει καθαρά τη λέξη και το Speech Recognition την αναγνωρίσει κατευθείαν, τότε η λέξη αυτή εμφανίζεται στο ListBox με όνομα Reco. Επίσης το ListBox αυτό καταγράφει και τις όποιες χειρονομίες (Gestures) αναγνωριστούν από το σύστημα έχοντας το γράμμα G στην αρχή.

Την ακρίβεια που θέλουμε κάθε φορά για την αναγνώριση την ομιλίας μπορεί να ρυθμιστεί προγραμματιστικά μέσα από την κλάση Speech_Recognition, από μία ιδιότητα που προσφέρει το Απίτου Kinect και λέγεται Confidence όπως φαίνεται στο παρακάτω block από κώδικα.

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

    if(e.Result.Confidence >= 0.30)
        NotifyPropertyChanged("recostring");
}
```

Στο παραπάνω παράδειγμα το Confidence έχει χαμηλό threshold. Αν χρειάζεται μεγαλύτερη αξιοπιστία στο αποτέλεσμα τότε το κατώφλι μπορεί να αυξηθεί ανάλογα.

Στο τρίτο ListBox με όνομα CPL Sentence, καταγράφεται η αναγνωρισμένη ολοκληρωμένη πρόταση που συντέθηκε από τις διάφορες είδους εντολές του χρήστη, είτε μέσω της ομιλίας είτε μέσω χειρονομιών είτε μέσω του συνδυασμού αυτών. Εδώ έχουμε την εμπλοκή του Sentence Compiler, ο οποίος στην ουσία ελέγχει το αν οι προτάσεις του συντίθενται είναι έγκυρες και μόνο τότε αυτές εμφανίζονται.

Το τελευταίο ListBox εμφανίζονται οι προτεινόμενες προτάσεις τις οποίες προτείνει το σύστημα στο χρήστη ανάλογα με την τελευταία αναγνωρισμένη λέξη (word) που έχει αναγνωρισθεί από το σύστημα, είτε μέσω του Gesture είτε μέσω του Speech Recognition. Με κόκκινο χρώμα εμφανίζεται η προηγούμενη λέξη ενώ η υπόλοιπη προτεινόμενη- ολοκληρωμένη πρόταση που μπορεί να υποστηρίξει το σύστημα είναι με μαύρο χρώμα. Για παράδειγμα αν ο χρήστης πει την λέξη open και αυτή αναγνωριστεί από το σύστημα, τότε αμέσως θα εμφανιστούν όλες οι προτάσεις που περιέχουν τη λέξη αυτή, όπως open window – open Firefox – open Skype και άλλες. Αυτή η λειτουργία υλοποιήθηκε στα πλαίσια της βοήθειας του χρήστη έτσι ώστε να μπορεί να χρησιμοποιεί με μεγαλύτερη ευκολία τη διεπαφή

9.2.2 Οι ρυθμίσεις της εφαρμογής

Στο δεξιό μέρος του παραθύρου παρουσιάζονται οι ρυθμίσεις της εφαρμογής. Στα πρώτα δύο buttons ο χρήστης μπορεί να επιλέξει αν θέλει να θέλει το Kinect να λειτουργεί σε seated ή default mode. Το default mode το Kinect αναγνωρίζει και τα είκοσι joints του σώματος ενώ στο seated mode αναγνωρίζει μόνο τα δέκα. Αυτό προγραμματιστικά γίνεται αλλάζοντας τη μεταβλητή Tracking Mode από το interface του Kinect σε Seated και Default αντίστοιχα.

```
Private void button4_Click(object sender, RoutedEventArgs e)
{
    Kinect_Device.ksensor.SkeletonStream.TrackingMode =
        SkeletonTrackingMode.Seated;
}
```

Τα επόμενα δύο Buttons ρυθμίζουν τη γωνία της βάσης του Kinect κατά δύο μοίρες πάνω ή κάτω κατά τον οριζόντιο άξονα. Για κάθε αλλαγή της γωνίας πρέπει να μεσολαβήσει ένα δευτερόλεπτο από την προηγούμενη αλλαγή. Σε προγραμματιστικό επίπεδο αυτό ρυθμίζεται από τη μεταβλητή ElevationAngle, όπως φαίνεται παρακάτω.

```
int new_Angle = Convert.ToInt32(this.textBox1.Text);
Kinect_Device.ksensor.ElevationAngle = new_Angle;
```

Η τελευταία ρύθμιση έχει να κάνει κι αυτή με τη γωνία της βάσης του Kinect, η οποία παίρνει την τιμή που θα βάλει ο χρήστης στο textbox. Η τιμή αυτή είναι σε μοίρες και θα πρέπει να είναι μεταξύ -27 και 27 μοιρών.

10 Κεφάλαιο

Περιπτώσεις Χρήσης (Use Cases)

Ενότητες κεφαλαίου

10.1	Εισαγωγή κεφαλαίου	61
10.2	Σενάριο 1:Χειρισμός της εφαρμογής Skype	61
10.2.1	Ανάλυση σεναρίου Skype	61
10.3	Σενάριο 2:Χειρισμός της εφαρμογής Media Player	66
10.3.1	Εργαλεία υλοποίησης σεναρίου Media Player	66
10.4	Ανάλυση σεναρίου Media Player	67

10.1 Εισαγωγή κεφαλαίου

Σε αυτή την ενότητα θα αναλυθούν δύο σενάρια χρήσης του πολυτροπικού συστήματος MIDAS. Τα πρώτο έχει να κάνει με την εφαρμογή Skype και τον χειρισμό αυτής. Το δεύτερο σενάριο αφορά την εφαρμογή Media Player. Και στα δύο σενάρια θα χρησιμοποιηθούν οι χειρονομίες και η ομιλία για την δημιουργία εντολών προς το σύστημα.

10.2 Σενάριο 1: Χειρισμός της εφαρμογής Skype

Το πρώτο σενάριο έχει να κάνει με το Skype που είναι μια desktop εφαρμογή πραγματοποίησης κλήσεων και ανταλλαγής μηνυμάτων μέσω του διαδικτύου. Εδώ χρειάζεται να ειπωθεί ότι για το χειρισμό της εφαρμογής Skype χρησιμοποιήθηκε μία βοηθητική βιβλιοθήκη που έχει γραφτεί ειδικά για αυτήν την εφαρμογή και είναι γραμμένη σε C++. Αυτή η βιβλιοθήκη είναι σε μορφή dll αρχείου και στην ουσία δίνει στους προγραμματιστές την δυνατότητα να χειριστούν το Skype και να πάρουν δεδομένα από αυτό χωρίς να χρειαστούν να πάνε σε low level προγραμματισμό. Το αρχείο που περιέχει τον unmanaged code λέγεται Interop.SKYPE4COMLib.dll και εισάγεται σαν reference στην εφαρμογή με τον παρακάτω τρόπο.

```
Using System.Runtime.InteropServices;
```

Με αυτόν τον τρόπο προγραμματιστικά για παράδειγμα μπορεί να διαχειριστεί κάποιος τους χρήστες ανάλογα με την κατάσταση σύνδεσής τους (online-offline-away). Επίσης μπορεί να γίνει τερματισμός της εφαρμογής ή εκκίνηση της με μία απλή προγραμματιστική εντολή. Φυσικά υποστηρίζονται σχεδόν όλες οι λειτουργίες που προσφέρει το Skype, έτσι είναι δυνατόν να υλοποιηθούν αρκετά εύκολα σε προγραμματιστικό επίπεδο.

10.2.1 Ανάλυση σεναρίου Skype

Οι συνολικές λειτουργίες του σεναρίου αυτού και ο τύπος της εντολής εισόδου που υποστηρίζει η κάθε μία από αυτές φαίνεται στον παρακάτω πίνακα. Αμέσως μετά απεικονίζεται το use case diagram για το συγκεκριμένο σενάριο.

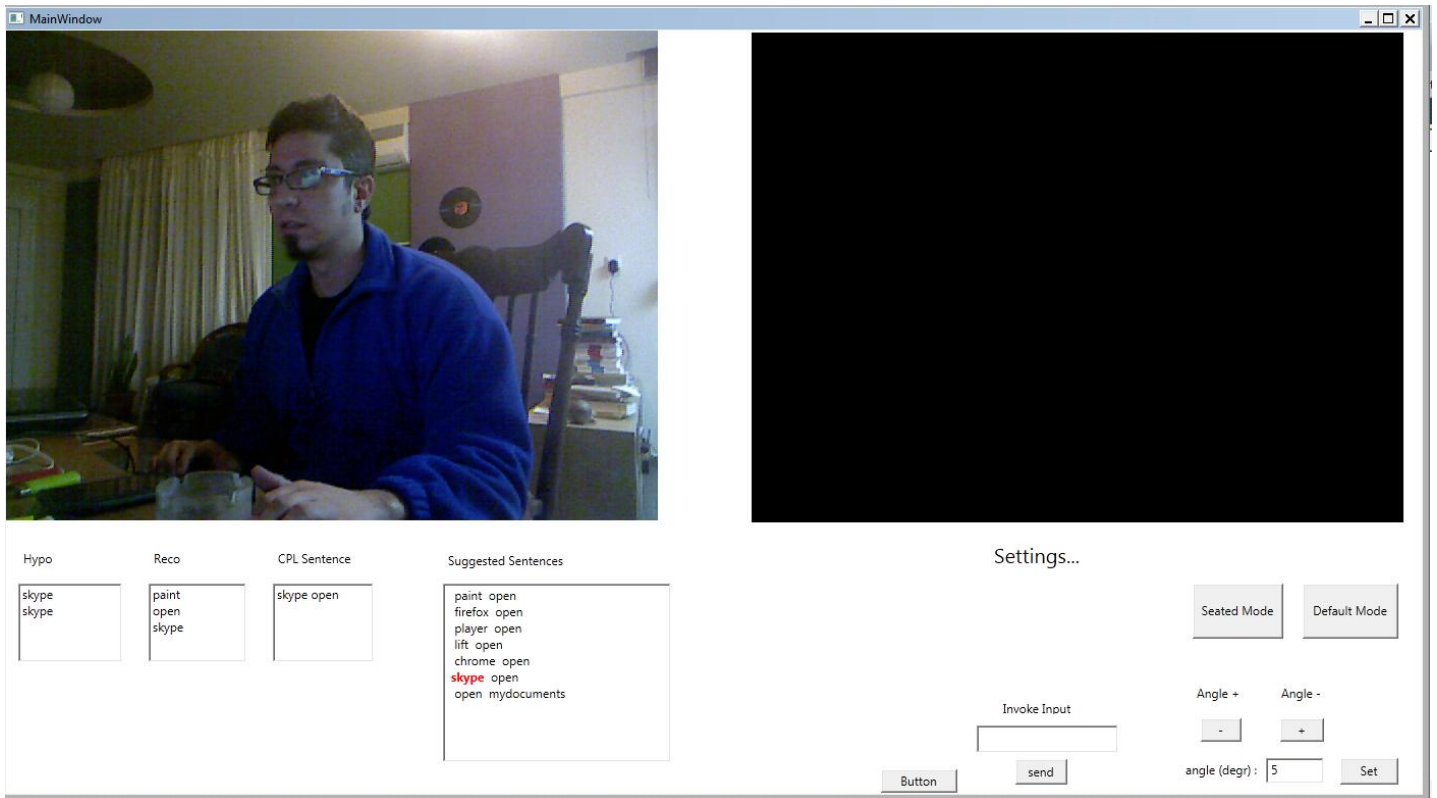
Scenario 1 Skype	
Action	Required Inputs
Activate Skype	Voice
Login	Voice, Face Recognition
Drag Window	Voice, Gesture
Call Contact	Voice
Open Mydocuments Folder	Voice
End Call	Gesture

Εικόνα 18. Καταγραφή των απαιτούμενων εισόδων για κάθε ενέργεια του σεναρίου



Εικόνα 17. Use case diagram για το σενάριο Skype.

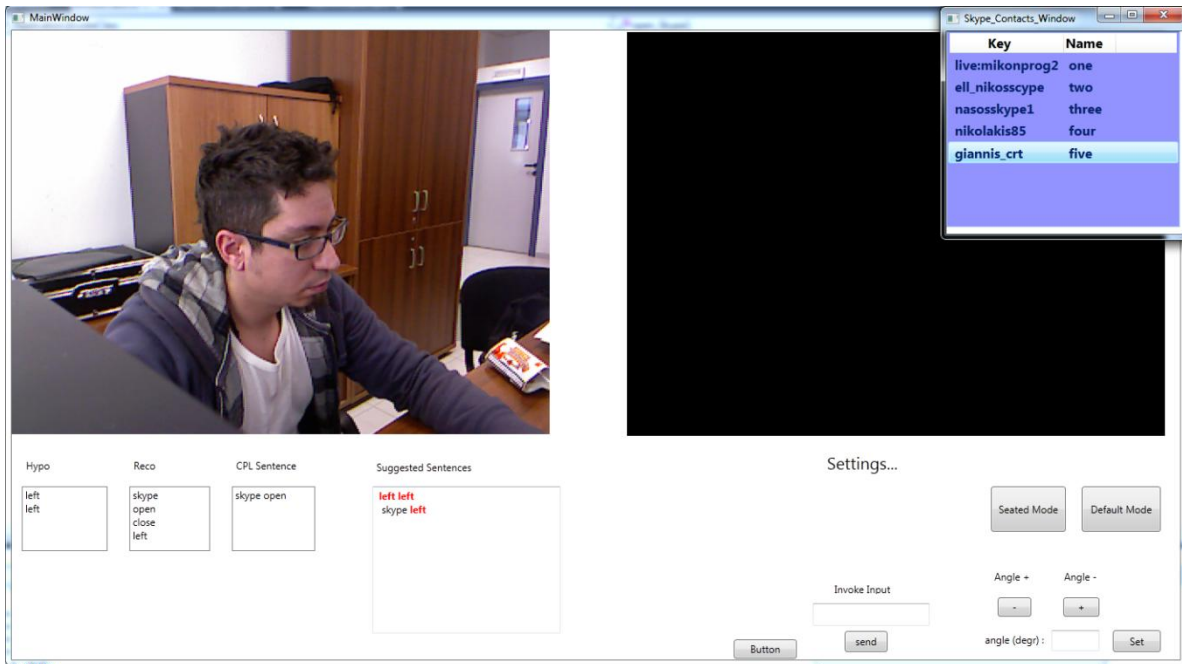
Το σενάριο ξεκινάει με την εκκίνηση της εφαρμογής και την είσοδο του χρήστη (login) με το όνομα χρήστη και τον κωδικό του. Αυτό γίνεται μέσω της φωνητικής εντολής open Skype.



Εικόνα 19. Εκκίνηση της εφαρμογής Skype μέσω της φωνητικής εντολής open Skype.

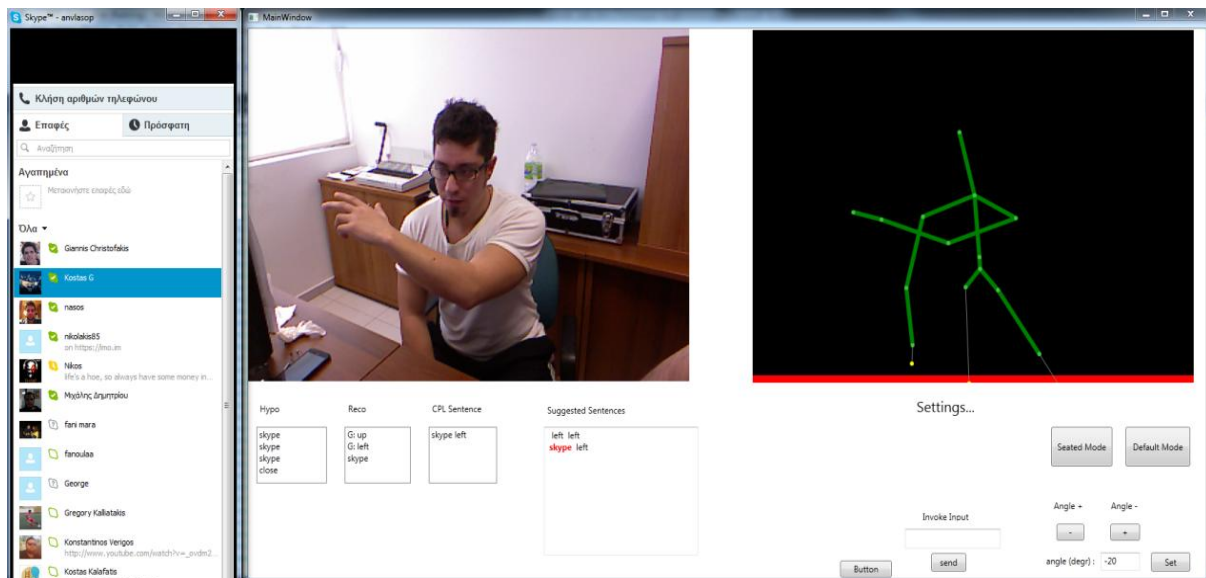
Αφού γίνει εκκίνηση του Skype, γίνεται η είσοδος του χρήστη στο λογαριασμό του. Στη συνέχεια εμφανίζεται ένα παράθυρο με τις διαθέσιμες online επαφές, τις οποίες ο χρήστης του συστήματος μπορεί να καλέσει. Το συγκεκριμένο παράθυρο έχει στη μία στήλη τα ονόματα των χρηστών και στην άλλη τις αντίστοιχες φωνητικές εντολές που αντιστοιχούν στην εντολή κλήσης κάθε ενός από αυτούς. Κάθε τέτοια εντολή απαρτίζεται από τη λέξη dial και στη συνέχεια από έναν αριθμό, αυξανόμενος κατά ένα ανά κάθε εντολή (dial one, dial two, dial three ...). Αυτό γίνεται για να είναι ξεκάθαρες οι εντολές, αφού οι αγγλικοί αριθμοί προφέρονται εύκολα από τον κάθε χρήστη.

Η εμφάνιση του βοηθητικού παραθύρου με τους διαθέσιμους χρήστες και τις φωνητικές εντολές που αντιστοιχούν στην κλήση τους χρειάστηκε, σε επίπεδο υλοποίησης την δυναμική εγγραφή στη βάση του συστήματος των διαθέσιμων ονομάτων – φίλων του χρήστη αλλά και οι εντολές με τις κατάλληλες παραμέτρους για την κλήση αυτών. Επίσης χρειάζεται να γίνει δυναμική αλλαγή στο αρχείο XML που χρησιμοποιεί η λειτουργία Speech Recognition του Kinect. Στο XML αυτό είναι καταγεγραμμένες όλες οι λέξεις που θέλουμε να αναγνωρίζονται από το Speech Recognition. Έτσι γίνεται κατανοητό ότι αυτό το αρχείο θα πρέπει να αλλάζει on runtime, αφού ο προγραμματιστής δεν μπορεί να ξέρει από πριν ποιες επαφές θα είναι διαθέσιμες σε μια συγκεκριμένη στιγμή και για ένα συγκεκριμένο χρήστη.



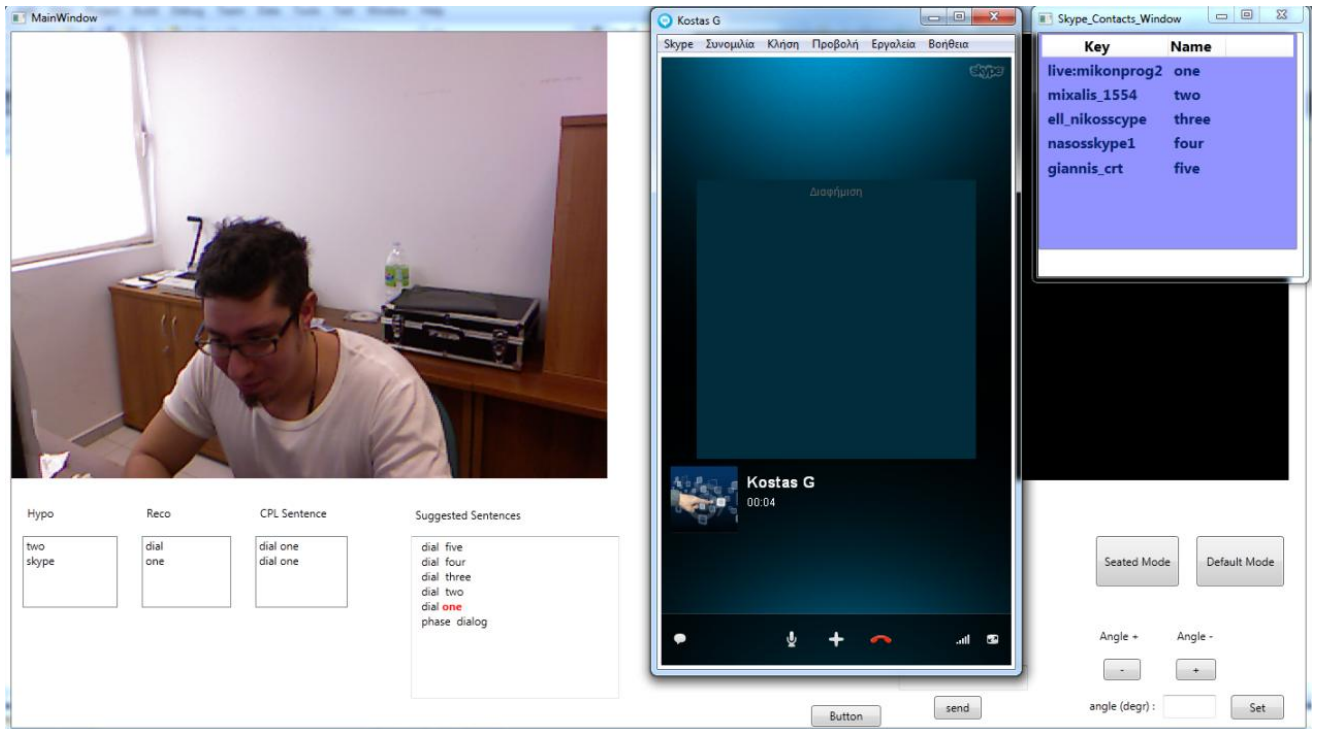
Εικόνα 20. Απεικόνιση του βοηθητικού παραθύρου με τις διαθέσιμες επαφές για κλήση.

Στη συνέχεια έχει ο χρήστης με μια προκαθορισμένη χειρονομία σε συνδυασμό με μια φωνητική εντολή θέτει το παράθυρο της εφαρμογής Skype στο πάνω αριστερό μέρος της οθόνης του υπολογιστή του. Η συγκριμένη εντολή – πρόταση είναι η Skype left. Η λέξη Skype απορρέει από την φωνητική εντολή Skype και η λέξη left αντιστοιχεί στη χειρονομία του δεξιού χεριού από δεξιά προς τα αριστερά όπως φαίνεται στην παρακάτω εικόνα.



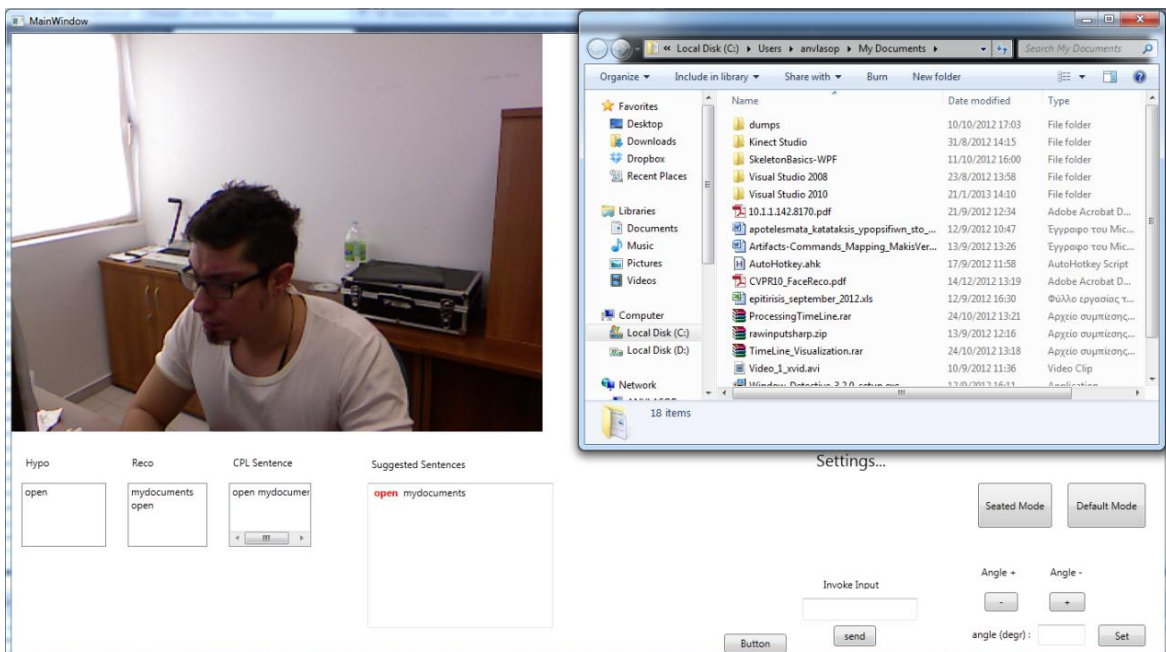
Εικόνα 21. Η εντολή Skype left και το αποτέλεσμα της.

Έπειτα με μία φωνητική εντολή όπως αυτές που προαναφέρθηκαν παραπάνω μπορεί ο χρήστης να πραγματοποιήσει μία κλήση σε κάποιον από τους διαθέσιμους φίλους του. Όπως φαίνεται στην παρακάτω εικόνα, με τη φωνητική εντολή dial one πραγματοποιείται κλήση της πρώτης στη λίστα επαφής.



Εικόνα 22. Πραγματοποίηση κλήσης με φωνητική εντολή.

Επίσης με μια χειρονομία θα μπορεί ο χρήστης όποτε θέλει να τελειώσει την κλήση που βρίσκεται σε εξέλιξη. Παράλληλα υπάρχει η δυνατότητα του χρήστη να ανοίξει την τον φάκελο My documents που βρίσκεται στον υπολογιστή του για να μπορέσει να κάνει αποστολή ένα αρχείο του σε κάποιον άλλο χρήστη όπως φαίνεται στην παρακάτω εικόνα.



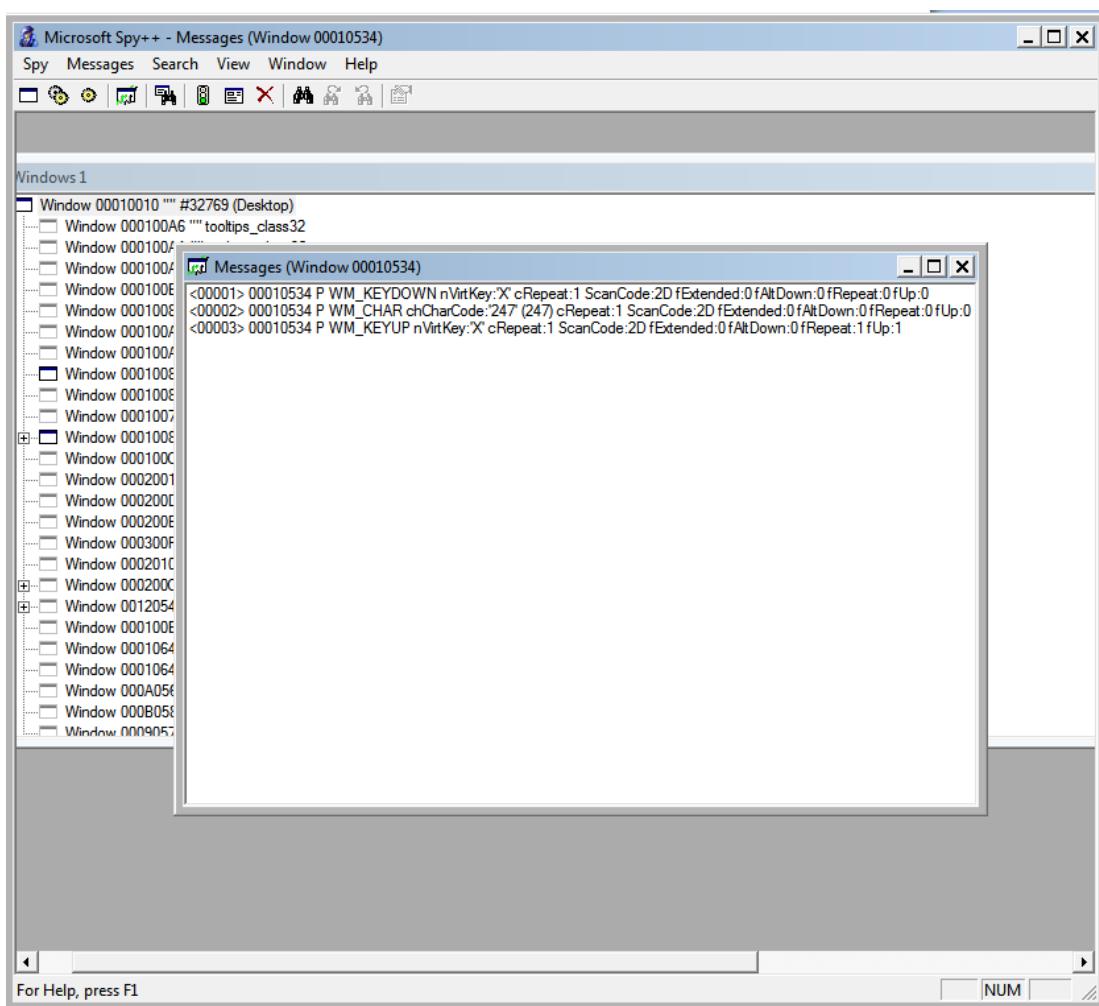
Εικόνα 23. Ανοίγμα του φακέλου My Documents με φωνητική εντολή.

10.3 Σενάριο 2: Χειρισμός της εφαρμογής Media Player

Το δεύτερο σενάριο έχει να κάνει με την εφαρμογή Windows Media Player, μία εφαρμογή των Windows για την αναπαραγωγή μουσικής και video.

10.3.1 Εργαλεία υλοποίησης σεναρίου Media Player.

Για την υλοποίηση αυτής της λειτουργίας χρησιμοποιήθηκε το εργαλείο Spy++. Το εργαλείο αυτό δίνει τη δυνατότητα να εντοπιστούν τα μηνύματα που στέλνονται στο λειτουργικό σύστημα όταν ο χρήστης δώσει κάποια εντολή σε αυτό μέσω των διαφόρων συσκευών εισόδου. Για παράδειγμα όταν ο χρήστης πατάει ένα κουμπί του πληκτρολογίου του, τότε στέλνεται ένα μήνυμα στο λειτουργικό σύστημα π.χ. WM_KeyDown(1522, 0x100, 0x00000002, 0x00000010).



Εικόνα 24. Το εργαλείο Spy++ καθώς ανιχνεύει το πάτημα ενός κουμπιού του πληκτρολογίου.

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων

Το μήνυμα αυτό στην ουσία είναι μία μέθοδος, όπως η WM_KeyDown, η οποία καλείται με διάφορες παραμέτρους. Η μέθοδος – συνάρτηση αυτή είναι υλοποιημένη σε C++ και βρίσκεται πακεταρισμένη σε μια βιβλιοθήκη – APIτων Windows. Η βιβλιοθήκη αυτή είναι ένα αρχείο dll, τοUser32.dll και για να χρησιμοποιηθεί χρειάζεται να γίνει importστον κώδικα και στη συνέχεια να δηλωθεί ποια μέθοδο θα χρησιμοποιήσει από τον προγραμματιστή για την περαιτέρω υλοποίηση, όπως φαίνεται στο παρακάτω παράδειγμα.

```
[DllImport("User32.dll")]
PublicstaticexternintFindWindow(
    stringstrClassName,
    stringstrWindowName);
```

Το παραπάνω παράδειγμα δείχνει την εισαγωγή της βιβλιοθήκης User32.dllκαι στη συνέχεια δηλώνεται η μέθοδος FindWindowμαζί με τις παραμέτρους που θα πάρει όταν αυτή κληθεί. Αφού κληθεί, η συνάρτηση επιστρέφει έναν αριθμό τύπου Integer, ο οποίος αντιστοιχεί σε ένα παράθυρο που είναι ανοιχτό στο λειτουργικό σύστημα. Με αυτό τον αριθμό δίνεται η δυνατότητα να αναφερόμαστε στο συγκεκριμένο παράθυρο που μας ενδιαφέρει και να δίνονται σε αυτό διάφορες εντολές προγραμματιστικά.

Αυτός είναι ο τρόπος με τον οποίο μπορεί να χρησιμοποιηθεί unmanaged code, όπως είναι ένα αρχείο dllμε κώδικα C++, σε managed code όπως είναι ο κώδικας υλοποίησης αυτής της εφαρμογής. Με αυτή τη μέθοδο μπορεί να συνδυαστούν blocksαπό κώδικα υλοποιημένα με διαφορετική γλώσσα προγραμματισμού (C++, C#). Γενικότερα το .NET framework δίνει αυτή την δυνατότητα με όλες τις γλώσσες προγραμματισμού που υποστηρίζουν αυτό το framework.

Οι συναρτήσεις που χρησιμοποιήθηκαν για την υλοποίηση του σεναρίου Media Player είναι οι εξής:

- FindWindow(string strClassName, string strWindowName)
- SendMessage(int hWnd, intMsg, intwParam, intlParam);

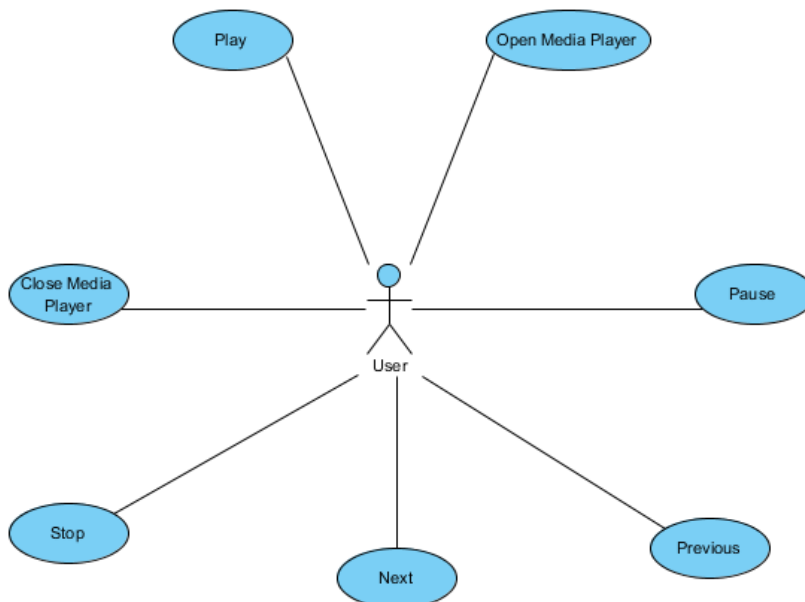
Η FindWindow χρησιμοποιείται για να βρεθεί ο αριθμός τύπου Integer που αντιστοιχεί στο παράθυρο του Media Player που είναι ήδη ανοιχτό στο λειτουργικό σύστημα. Η SendMessageμε τη σειρά της είναι αναγκαία στο να αποσταλεί ένα μήνυμα σε ένα συγκεκριμένο παράθυρο. Για παράδειγμα αν θέλουμε η εφαρμογή Media Player να αρχίσει την αναπαραγωγή ενός τραγουδιού, τότε χρειάζεται να σταλεί ένα μήνυμα στο παράθυρο της εφαρμογής με τις κατάλληλες παραμέτρους.

10.4 Ανάλυση σεναρίου Media Player

Οι συνολικές λειτουργίες του σεναρίου αυτού και ο τύπος της εντολής εισόδου που υποστηρίζει η κάθε μία από αυτές φαίνεται στον παρακάτω πίνακα. Αμέσως μετά απεικονίζεται το use case diagram για το συγκεκριμένο σενάριο.

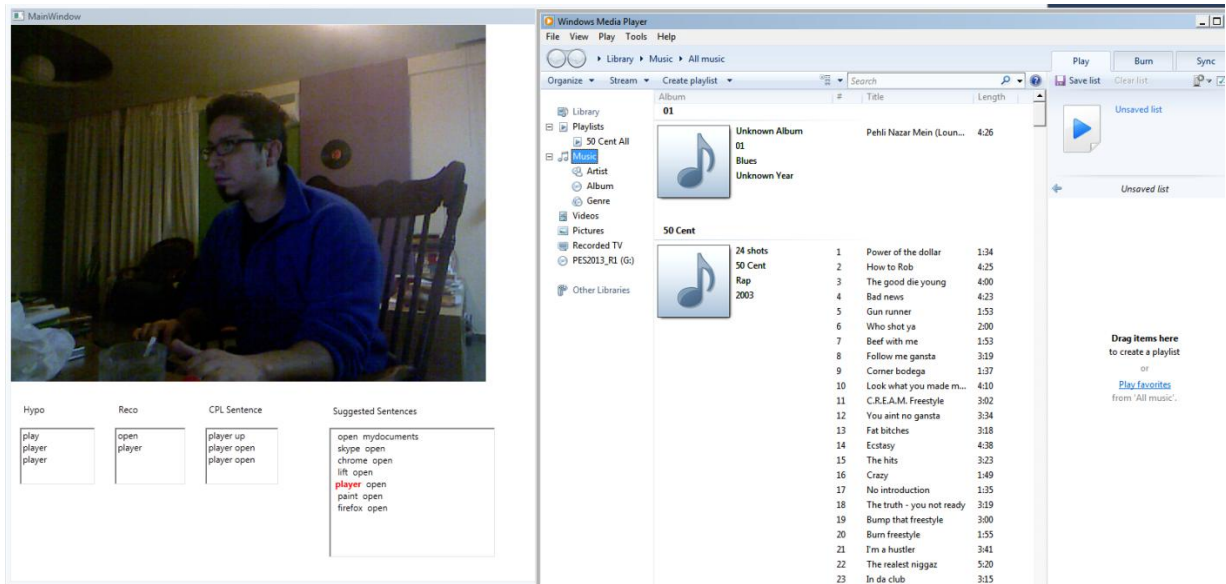
Scenario 2 Media Player	
Action	Required Inputs
Open Media Player	Voice
Play a song	Voice
Pause	Voice
Stop Playing	Voice
Close Media Player	Voice
Next	Voice
Previous	Voice

Εικόνα 25. Καταγραφή των απαιτούμενων εισόδων για κάθε ενέργεια του σεναρίου.



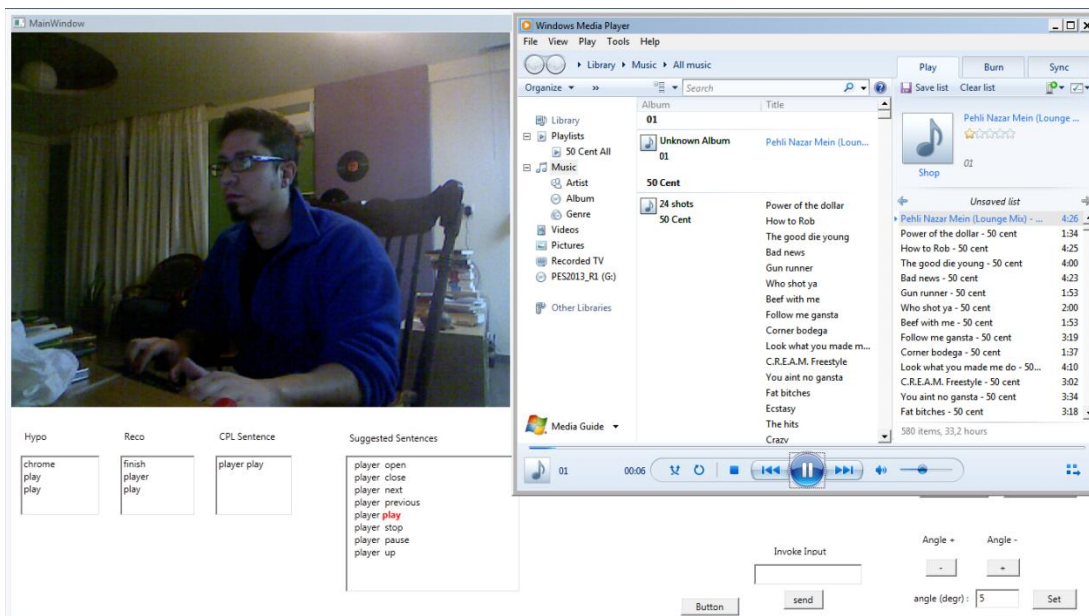
Εικόνα26. Use case diagram για το σενάριο Media Player.

Το σενάριο ξεκινάει με την εκκίνηση της εφαρμογής Media Player. Η εντολή για αυτή την ενέργεια η φωνητική εντολή open player. Η εντολή - πρόταση αυτή είναι καταγεγραμμένη στη βάση του συστήματος και επειδή είναι φωνητική κάθε λέξη από την οποία αποτελείται είναι αποθηκευμένη και μέσα στο αρχείο Grammar.xml. Το αρχείο αυτό είναι απαραίτητο για τη λειτουργία του Speech Recognition, αφού αυτή θα είναι η βάση δεδομένων για την αναγνώριση ομιλίας. Αν μία λέξη δεν βρίσκεται στο Grammar.xml τότε η λέξη δεν θα αναγνωρίζεται από το Speech Recognition του Kinect.



Εικόνα 27. Η εντολή open player και το αποτέλεσμα της.

Αφού ανοίξει η εφαρμογή τότε ο χρήστης μπορεί να δώσει την φωνητική εντολή player play και ο media player θα παίξει το πρώτο τραγούδι της λίστας.



Εικόνα 28. Απεικόνιση της εντολής player play και το αποτέλεσμα της.

Με τον ίδιο τρόπο γίνεται και το σταμάτημα της αναπαραγωγής του Media Player. Πιο συγκεκριμένα με την εντολή player stop η εφαρμογή σταματάει την αναπαραγωγή. Με την εντολή player pause γίνεται η παύση της αναπαραγωγής. Επίσης με τις εντολές player next και player previous ο Media Player προχωράει στην αναπαραγωγή του επόμενου ή του προηγούμενου τραγουδιού αντίστοιχα. Τέλος με την φωνητική εντολή Player close η εφαρμογή media player σταματάει τη λειτουργία της.

11 Επίλογος

Ενότητες επιλόγου

11.1 Εισαγωγή Κεφαλαίου	71
11.2 Σύνοψη και συμπεράσματα	71
11.3 Μελλοντικές Επεκτάσεις	71

11.1 Εισαγωγή Κεφαλαίου

Σαν επίλογο για την παρούσα πτυχιακή εργασία θα αναπτυχθούν τα συμπεράσματα που προέκυψαν από αυτή. Παράλληλα γίνεται μια αναφορά για τις ενδεχόμενες μελλοντικές επεκτάσεις και βελτιώσεις της εργασίας.

11.2 Σύνοψη και συμπεράσματα

Στόχος αυτής της πτυχιακής εργασίας ήταν ο επανασχεδιασμός και η υλοποίηση μιας εφαρμογής φυσικής αλληλεπίδρασης μεταξύ ανθρώπου και υπολογιστή. Για το σκοπό αυτό δημιουργήθηκε ένα λογισμικό συμβατό με το λειτουργικό σύστημα των Windows 7 το οποίο προσφέρει στο χρήστη την φυσική αλληλεπίδρασή του με δύο συγκεκριμένες εφαρμογές που υποστηρίζονται από αυτό το λειτουργικό σύστημα.

Πιο συγκεκριμένα το λογισμικό που αναπτύχθηκε δίνει την δυνατότητα στο χρήστη να εξοικειωθεί τις αλληλεπιδράσεις φυσικού τύπου με τον υπολογιστή, καθώς το μέλλον των αλληλεπιδράσεων ανθρώπου – υπολογιστή τείνει όλο και περισσότερο να γίνεται με φυσικό τρόπο. Έτσι καθίσταται πολύ σημαντική η σταδιακή απεμπλοκή των χρηστών από τις παραδοσιακές μεθόδους αλληλεπίδρασης και η σταδιακή αφομοίωση των νέων και καινοτόμων μορφών φυσικών διεπαφών.

Μια συνοπτική αναφορά των διαφόρων βημάτων για την υλοποίηση αυτού του λογισμικού γίνεται παρακάτω:

- Κατασκευάστηκε ένα γραφικό διαδιάστατο περιβάλλον το οποίο αποτέλεσε και το βασικό κορμό πάνω στον οποίο στηρίχτηκε η διεπαφή.
- Δημιουργήθηκε το καθολικό σχήμα της βάσης δεδομένων από το οποίο η εφαρμογή θα μπορεί να ανακτά τα κατάλληλα δεδομένα για τη λειτουργία της.
- Χρησιμοποιήθηκε το Microsoft Kinect σαν βασική συσκευή πολύ-τροπικής εισόδου του συστήματος.
- Αναπτύχθηκαν αλγόριθμοι για τον έλεγχο εφαρμογών από τον άνθρωπο με τη βοήθεια των χειρονομιών και της ομιλίας του.

Τέλος, η δημιουργία αυτού του λογισμικού αποδεικνύει ότι οι μέθοδοι ανάπτυξης λογισμικού και οι σημερινοί τεχνολογία των αισθητήρων μπορούν να αποδώσουν νέους τρόπους της ανθρώπινης επικοινωνίας με τον υπολογιστή. Οι κλασικές συσκευές εισόδου όπως είναι το ποντίκι ή το πληκτρολόγιο μπορούν να αντικατασταθούν από τους νέους πολύ-αισθητήρες και με τις κατάλληλες αλλαγές στο σχεδιασμό και την ανάπτυξη των λογισμικών δίνεται η δυνατότητα να προχωρήσει η επικοινωνία ανθρώπου - υπολογιστή σε ένα πιο σύγχρονο και εναλλακτικό επίπεδο ακόμα και σε άλλους τομείς της επιστήμης, όπως η ιατρική.

11.3 Μελλοντικές Επεκτάσεις.

Στο ζήτημα της μελλοντικής επέκτασης, παρουσιάζεται το θέμα των εναλλακτικών συσκευών εισόδου με μεγαλύτερη ακρίβεια από αυτή που προσφέρει το Kinect. Επίσης η ανάπτυξη ειδικών αλγορίθμων για το Skeleton Tracking και την αναγνώριση φωνής, έτσι ώστε να περιοριστούν τα περιθώρια σφαλμάτων που εκ των πραγμάτων παρουσιάζονται.

Παράλληλα υπάρχει η δυνατότητα σε μελλοντική εργασία να ενσωματωθούν και άλλου είδους συσκευές όπως συσκευές αναγνώρισης πίεσης, που με τη σειρά τους μπορούν να συνδυαστούν με τον τομέα της ρομποτικής.

12 Δεξικό

- Natural User Interface (NUI) : Φυσική Διεπαφή Χρήστη
- Software Development Kit (SDK) : Εργαλείο Ανάπτυξης Λογισμικού
- Application Program Interface (API) : Διεπαφή Προγραμματισμού Εφαρμογής
- Entity Relationship Model (ER Model) : Μοντέλο οντοτήτων συσχετίσεων
- Context Information Manager (CIM) : Πλαίσιο Διαχείρισης Πληροφοριών
- Complementary Metal Oxide Semiconductor (CMOS) : Συμπληρωματικού οξειδίου μετάλλου-ημιαγωγού
- Joints : Αρθρώσεις
- Data Base (DB) : Βάση Δεδομένων
- Infrared Ray (IR) : Υπέρυθρη Ακτινοβολία

13 Αναφορές

- [1] Intelligent Image Processing, John Wiley and Sons, 2001.
- [2] Kettebekov, Sanshzar, and Rajeev Sharma (2001). "Toward Natural Gesture/Speech Control of a Large Display." Proceedings EHCI '01 Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction Pages 221-234
- [3] Marius Vassiliou, V. Sundareswaran, S. Chen, R. Behringer, C. Tam, M. Chan, P. Bangayan, and J. McGee (2000), "Integrated Multimodal Human-Computer Interface and Augmented Reality for Interactive Display Applications," in Darrel G. Hopper (ed.) Cockpit Displays VII: Displays for Defense Applications (Proc. SPIE . 4022), 106-115. [ISBN 0-8194-3648-8]
- [4] Oviatt, S. (2002), Multimodal interfaces, Jacko, J.; Sears, A, The Human-Computer Interaction Handbook, Lawrence Erlbaum
- [5] Bauckhage, C.; Fritsch, J.; Rohlfing, K.J.; Wachsmuth, S.; Sagerer, G. (2002). Evaluating integrated speech-and image understanding. Int. Conf. on Multimodal Interfaces.
- [6] Ismail, N.A.; O'Brien, E.A. (2008). Enabling Multimodal Interaction in Web-Based Personal Digital Photo Browsing. Int. Conf. on Computer and Communication Engineering
- [7] Brooks, A.; Petersson, E. (2007). SoundScapes: non-formal learning potentials from interactive VEs. SIGGRAPH
- [8] K. Khoshelham. (2011). ACCURACY ANALYSIS OF KINECT DEPTH DATA
- [9] <http://www.futurepicture.org/?p=129>. (2010). Kinect Hacking 105: Full Resolution, Public Domain Images of the Speckle Pattern.
- [10] C# Language Specification (4th ed.). Ecma International. June 2006. Retrieved January 26, 2012.
- [11] Released: Windows SDK for Windows 7 and .NET Framework 4
- [12] N. Vidakis; M. Syntychakis; G. Triantafyllidis; D. Akoumianakis;. (2012). Multimodal Natural User Interaction for Multiple Applications: The Gesture – Voice Example