



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**«Αναγνώριση Χειρονομιών με Χρήση Κινούμενης Εικόνας Βάθους»**

**Αλέξανδρος Γ. Μήτσου**

**Επιβλέπων Καθηγητής: Γεώργιος Σταμούλης  
Συνεπιβλέπων Καθηγητής: Σπύρου Ευάγγελος**

**ΛΑΜΙΑ  
ΙΟΥΛΙΟΣ 2018**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**



**«Αναγνώριση Χειρονομιών με Χρήση Κινούμενης Εικόνας Βάθους»**

**Μήτσου Αλέξανδρος**

**AM: 2113182**

**Επιβλέπων Καθηγητής: Γεώργιος Σταμούλης**  
**Συνεπιβλέπων Καθηγητής: Σπύρου Ευάγγελος**

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις (1), που προβλέπονται από τις διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί χωρίς να τα περικλείω σε εισαγωγικά και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία και τη σελίδα. Η αυτολεξεί παράθεση χωρίς αναφορά στη πηγή είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από πίνακα ή σχέδιο, όπως στα παραθέματα.

2. Δέχομαι ότι η αυτολεξεί παράθεση χωρίς εισαγωγικά, ακόμα και αν συνοδεύεται από αναφορά στην πηγή σε κάποιο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.

3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια.

4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1896 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»

## Περίληψη

Η είσοδος της τεχνολογίας στην ανθρώπινη καθημερινότητα αυξάνεται ολοένα και περισσότερο μέρα με τη μέρα και έτσι, νέες εφαρμογές της Πληροφορικής εισέρχονται σε αυτήν. Ένας από τους βασικούς τομείς είναι η Τεχνητή Νοημοσύνη, η οποία στοχεύει να αλλάξει την κοινωνία προς το καλύτερο.

Σκοπός της παρούσας πτυχιακής εργασίας είναι η αναγνώριση ανθρώπινων χειρονομιών σε πραγματικό χρόνο με χρήση του αισθητήρα Kinect, διερευνώντας διάφορες τεχνικές μηχανικής μάθησης. Αρχικά, ορίζεται ένα σύνολο χειρονομιών και με χρήση της κάμερας βάθους συγκεντρώνονται μετρήσεις από τις σκελετικές αρθρώσεις. Οι μετρήσεις αυτές χρησιμοποιούνται για την κατασκευή ενός συνόλου δεδομένων. Στη συνέχεια εκπαιδεύονται διάφοροι ταξινομητές με σκοπό την αναγνώριση των χειρονομιών.

Αρχικά, στην παρούσα εργασία παρουσιάζονται τα τεχνικά χαρακτηριστικά του Kinect, και ο τρόπος με τον οποίο πραγματοποιείται η εξαγωγή χαρακτηριστικών από τις σκελετικές αρθρώσεις. Στη συνέχεια παρουσιάζονται οι αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν. Έπειτα παρουσιάζονται τα πειραματικά αποτελέσματα που προέκυψαν. Τέλος, η εργασία κλείνει με τα συμπεράσματα που προέκυψαν και πιθανές μελλοντικές επεκτάσεις.

**Θεματική περιοχή:** Αναγνώριση Κίνησης

**Λέξεις Κλειδιά:** Αναγνώριση Χειρονομιών, Kinect, Σκελετικές Αρθρώσεις, Αλληλεπίδραση Ανθρώπου-Μηχανής, Μηχανική Μάθηση

## **ΕΥΧΑΡΙΣΤΙΕΣ**

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω όλους εκείνους που συνέβαλλαν και βοήθησαν στην πραγματοποίηση αυτής της πτυχιακής εργασίας.

Ιδιαίτερα, θα ήθελα να ευχαριστήσω τον επιβλέποντα επίκουρο καθηγητή του ΤΕΙ Στερεάς Ελλάδας κ. Ευάγγελο Σπύρου, για την βοήθειά του στην περάτωση αυτής της εργασίας. Επίσης, θα ήθελα να ευχαριστήσω την Βοηθό Έρευνας του ΕΚΕΦΕ-«Δημόκριτος» κ. Ειρήνη Μαθέ για την υπομονή της και για τη συνολική υποστήριξη καθ' όλη τη διάρκεια εκπόνησης της παρούσας εργασίας.

## ΠΕΡΙΕΧΟΜΕΝΑ

1. Εισαγωγή και βασικές έννοιες .....	1
1.1 Εισαγωγή .....	1
1.2 Σκοπός της εργασίας .....	1
1.3 Βασικές έννοιες και ορισμοί .....	2
1.3.1 Ορισμός χειρονομίας και κίνησης .....	2
1.3.2 Εικόνα βάθους .....	3
1.3.3 Φυσική διεπαφή χρήστη (Natural User Interface) .....	4
1.3.4 Μηχανική Μάθηση .....	5
1.4 Δομή εργασίας .....	6
2. Ο αισθητήρας Kinect .....	7
2.1 Εισαγωγή και σύντομη ιστορία .....	7
2.2 Τεχνικές Προδιαγραφές .....	9
2.3 Εφαρμογές χρήσης του Kinect .....	10
2.4 Κάμερες Βάθους και APIs .....	11
3. Εξαγωγή Χαρακτηριστικών .....	12
3.1 Αναγνώριση Χειρονομιών .....	12
3.2 Επιλογή Χαρακτηριστικών .....	12
3.3 Αναπαράσταση Σκελετικών Αρθρώσεων .....	16
4. Σχετικές Εργασίες .....	17
5. Μέθοδοι Και Μοντέλα .....	19
5.1 Εκπαίδευση μοντέλων Μηχανικής Μάθησης .....	19
5.2 Η μέθοδος Ανάλυσης Κύριων Συνιστωσών (PCA) .....	20
5.3 Μηχανές Διανυσμάτων Υποστήριξης (SVMs) .....	21
5.4 Μέθοδοι Πυρήνων .....	22
5.5 Πολυεπίπεδο Perceptron (Multilayer Perceptron) .....	23
5.6 Extra Trees (Extremely Randomized Trees) .....	24
5.7 AdaBoost .....	24
6. Πειράματα–Αποτελέσματα .....	25
6.1 Σύνολο Δεδομένων .....	25
6.2 Πειραματικά αποτελέσματα .....	26
7. Μελλοντικές εργασίες και συμπεράσματα .....	29

Αναφορές .....	30
----------------	----



## ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικόνα 1: Η συνολική διαδικασία αναγνώρισης χειρονομιών .....	1
Εικόνα 2: Ευθυγραμμισμένες RGB και σκελετικές εικόνες ενδεικτικών χειρονομιών.....	2
Εικόνα 3: Εικόνα βάθους: Αριστερά απεικονίζεται η οθόνη ενός Η/Υ Δεξιά απεικονίζεται μία ανθρώπινη μορφή.....	3
Εικόνα 4: Αλληλεπίδραση χρήστη-μηχανής.....	4
Εικόνα 5: Από αριστερά προς τα δεξιά: α) Το Kinect β) RGB εικόνα γ) εικόνα βάθους δ) Σκελετός.....	7
Εικόνα 6: Η ιεραρχία των σκελετικών αρθρώσεων.....	8
Εικόνα 7: Αναπαράσταση του αισθητήρα Kinect.....	9
Εικόνα 8: Οι ανθρώπινες σκελετικές αρθρώσεις που έχουν εξαχθεί με χρήση του Kinect SDK.....	16
Εικόνα 9: Διαδικασία εκπαίδευσης ενός μοντέλου.....	19
Εικόνα 10: Στιγμιότυπο της μεθόδου ανάλυσης κύριων συνιστωσών για τρεις κλάσεις.....	20
Εικόνα 11: Ταξινόμηση κλάσεων με SVM.....	21
Εικόνα 12: Δομή Perceptron πολλαπλών επιπέδων.....	23
Εικόνα 13: Ποσοστό αξιολόγησης ακρίβειας των ταξινομητών για το πρώτο πείραμα.....	27
Εικόνα 14: Ποσοστό αξιολόγησης ακρίβειας των ταξινομητών για το δεύτερο πείραμα.....	28

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

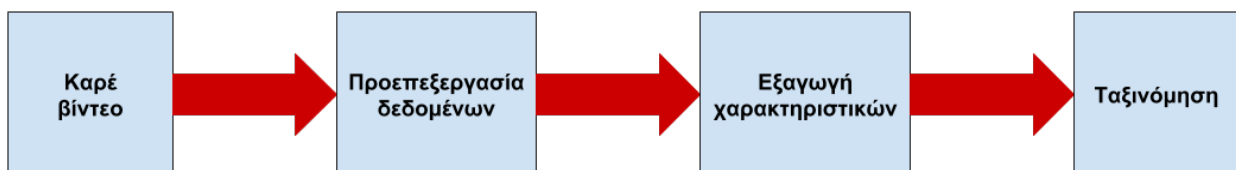
Πίνακας 1: Τα σύμβολα και η επεξήγηση αυτών που χρησιμοποιήθηκαν σε αυτήν την εργασία.....	12
Πίνακας 2: Τα προτεινόμενα χαρακτηριστικά, που εξήχθησαν από της σκελετικές αρθρώσεις.....	13
Πίνακας 3: Βέλτιστες παράμετροι ταξινομητών (α: ρυθμός μάθησης, e: αριθμός εκτιμητών, γ: Συντελεστής πυρήνα, C: ποινή του όρου σφάλματος).....	26

# 1.Εισαγωγή και βασικές έννοιες

## 1.1 Εισαγωγή

Η ανθρώπινη επικοινωνία βασίζεται στη χρήση ετερογενών μέσων έκφρασης, όπως η ομιλία, οι εκφράσεις του προσώπου, το βλέμμα και η στάση του σώματος. Συγκεκριμένα, τόσο οι στάσεις όσο και οι χειρονομίες μπορούν να χρησιμοποιηθούν για την επικοινωνία, από την στιγμή που μεταφέρουν κάποιο νόημα. Η αναγνώριση της στάσης και των χειρονομιών έχουν σκοπό την αναγνώριση τέτοιων νοηματικών εκφράσεων, οι οποίες εκτελούνται από τον άνθρωπο. Αυτές οι εκφράσεις τυπικά περιλαμβάνουν την κίνηση της παλάμης, των χεριών, της κεφαλής, εκφράσεων του προσώπου και σε κάποιες περιπτώσεις και όλο το σώμα. Η αναγνώρισή τους συχνά αποσκοπεί στη διευκόλυνση της διασύνδεσης μεταξύ ανθρώπων και υπολογιστών. Αυτή η διευκόλυνση αποσκοπεί στη χρήση εφαρμογών χωρίς τη χρήση συσκευών, όπως το πληκτρολόγιο και το ποντίκι. Οι υπάρχουσες εφαρμογές στοχεύουν στην έκφραση και την αποκωδικοποίηση της νοηματικής γλώσσας. Επίσης, η αναγνώριση χειρονομιών χρησιμοποιείται σε παιχνίδια, ιατρικές εφαρμογές, υπηρεσίες «Διαδικτύου των Πραγμάτων», εφαρμογές απομακρυσμένης πρόσβασης και εικονική πραγματικότητα.

## 1.2 Σκοπός της εργασίας



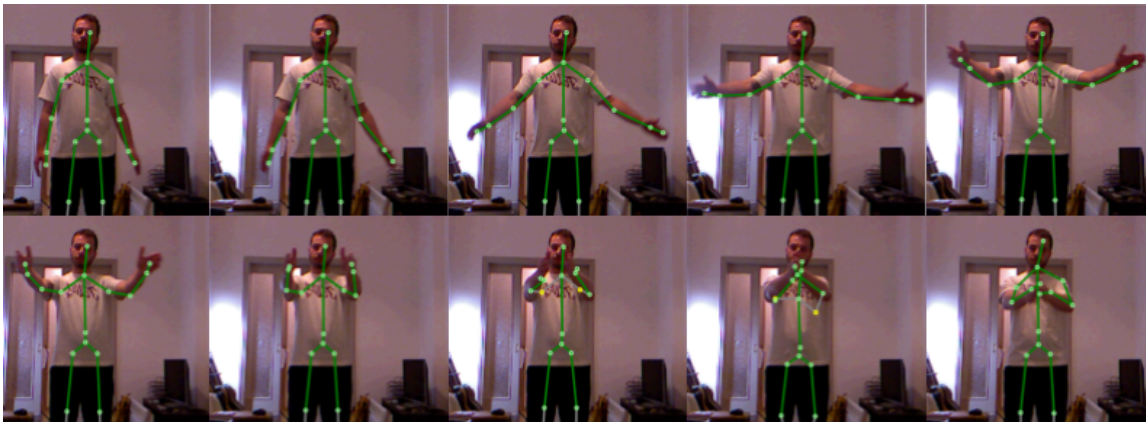
Εικόνα 1: Η συνολική διαδικασία αναγνώρισης χειρονομιών

Σκοπός της παρούσας εργασίας είναι η αναγνώριση χειρονομιών με τη χρήση κάμερας. Ο χρήστης στέκεται μπροστά από την κάμερα και κάποια μέρη του σώματός του/της ανιχνεύονται και παρακολουθούνται. Τα χαρακτηριστικά, εξάγονται και συγκρίνονται με εκπαιδευμένα μοντέλα ούτως ώστε να μεταφραστούν σε προκαθορισμένες χειρονομίες. Πιο συγκεκριμένα χρησιμοποιούμε τον αισθητήρα Kinect, καθώς και το Πακέτο Ανάπτυξης Λογισμικού (SDK) για την εξαγωγή και παρακολούθηση των σκελετικών αρθρώσεων. Από αυτές τις αρθρώσεις εξάγουμε τα διανυσματικά χαρακτηριστικά τους και έπειτα υπολογίζουμε ένα σύνολο στατιστικών με τη βοήθεια αλγορίθμων μηχανικής μάθησης. Αυτοί οι αλγόριθμοι χρησιμοποιούνται για να

ταξινομηθούν οι χειρονομίες με βάση το είδος της εκάστοτε χειρονομίας. Στόχος είναι, να μπορεί το κάθε μοντέλο που φτιάχνουμε να αναγνωρίζει την κάθε χειρονομία. Η Εικόνα 1 απεικονίζει όλη τη δομή της διαδικασίας προς αναγνώριση χειρονομιών που ακολουθείται σε αυτήν την εργασία.

## 1.3 Βασικές έννοιες και ορισμοί

### 1.3.1 Ορισμός χειρονομίας και κίνησης



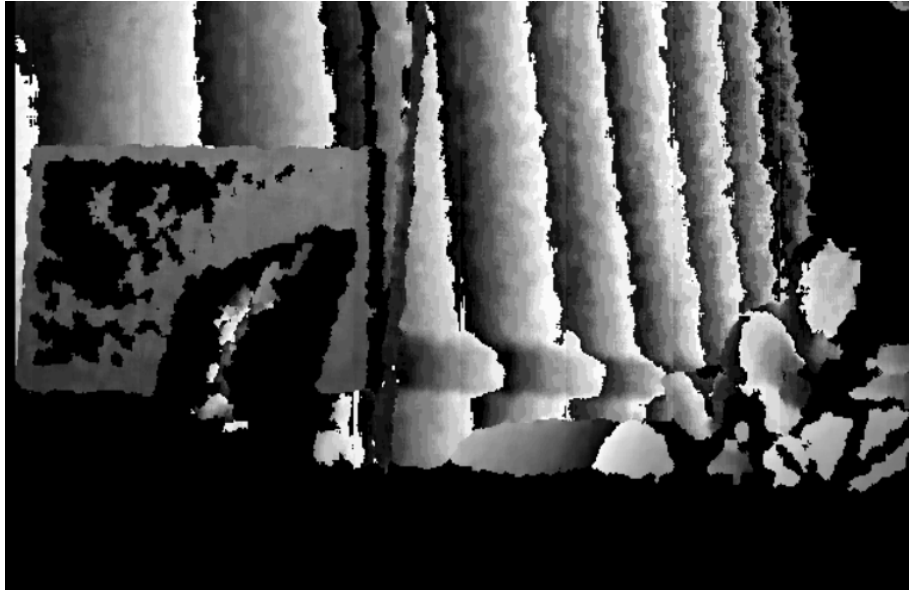
Εικόνα 2: RGB και σκελετικές εικόνες ενδεικτικών χειρονομιών

Δεν είναι απλό να δοθεί ένας ορισμός της χειρονομίας. Σύμφωνα με το λεξικό Merriam-Webster, μία χειρονομία μπορεί να ορισθεί ως «μία κίνηση συνήθως του σώματος ή των άκρων η οποία εκφράζει ή δίνει μία έμφαση σε μια ιδέα, ένα συναίσθημα, ή μία συμπεριφορά», ή «η χρήση της κίνησης των άκρων ή του σώματος ως ένα τρόπος έκφρασης». Στην επιστημονική βιβλιογραφία, υπάρχουν διάφοροι ορισμοί της. Στους τομείς της υπολογιστικής όρασης και της αναγνώρισης προτύπων, έχει διαμορφωθεί ένας γενικός ορισμός της από τους Mitra και Acharya [1] ως εξής: «Οι χειρονομίες είναι ακριβείς, νοηματικές κινήσεις του σώματος οι οποίες περιλαμβάνουν φυσικές κινήσεις των δακτύλων, των παλαμών, των χεριών, της κεφαλής, του προσώπου ή του σώματος με πρόθεση: 1) να μεταφέρουν νοηματική πληροφορία ή 2) να επιδρούν με το περιβάλλον.» Πιο συγκεκριμένα, μία χειρονομία (μία κίνηση του χεριού) μπορεί να οριστεί ως η μεταφορά ενός (υπο)συνόλου αρθρώσεων του χεριού (π.χ., παλάμες, καρποί, αγκώνες) από ένα σημείο A σε ένα σημείο B, με χρήση της ασαφούς προκαθορισμένης τροχιάς.

Ως κίνηση στη φυσική ορίζεται η μεταβολή της θέσης ενός σώματος ως προς ένα σύστημα αναφοράς. Καθώς το ανθρώπινο σώμα απαρτίζεται από μέλη (όπως οι σκελετικές αρθρώσεις), η κίνηση μπορεί να οριστεί ως η μεταβολή των θέσεων αυτών

στον τρισδιάστατο (3D) χώρο. Για παράδειγμα η Εικόνα 2 απεικονίζει ενδεικτικά μία από τις χειρονομίες που χρησιμοποιούνται σε αυτήν την εργασία.

### 1.3.2 Εικόνα βάθους



Εικόνα 3: Εικόνα βάθους: Αριστερά απεικονίζεται η οθόνη ενός Η/Υ. Δεξιά απεικονίζεται μία ανθρώπινη μορφή

Οι εικόνες βάθους (Εικόνα 3) είναι ένας διαφορετικός τρόπος να αναπαρασταθεί η πληροφορία η οποία, εξάγεται από την κάμερα βάθους. Ουσιαστικά, είναι ευκολότερο για το υπολογιστικό σύστημα να κατανοήσει που βρίσκεται ένα αντικείμενο μέσα στην εικόνα. Σε αυτού του είδους τις εικόνες, το χρώμα κάθε εικονοστοιχείου προσδιορίζει πόσο μακριά από την κάμερα βρίσκεται το κάθε τμήμα της εικόνας, δηλαδή τα φωτεινότερα σημεία αντιστοιχούν σε μικρή απόσταση, ενώ τα σκοτεινότερα σε μεγάλη. Ωστόσο, ανάλογα με τα χαρακτηριστικά της εκάστοτε κάμερας, μπορεί να ακολουθείται η αντίθετη αναπαράσταση.

Έτσι, μπορεί να βρεθεί εύκολα η αρχή και το τέλος ενός αντικειμένου μέσα στην εικόνα. Μία εικόνα βάθους περιέχει ακριβείς πληροφορίες στον 3D χώρο σχετικά με το τι βρίσκεται μπροστά της. Επομένως, με αυτό τον τρόπο είναι δυνατό να ανακατασκευαστούν 3D μοντέλα, τα οποία είναι εύκολα διαχειρίσιμα.

### 1.3.3 Φυσική διεπαφή χρήστη (Natural User Interface)



Εικόνα 4: Αλληλεπίδραση χρήστη-μηχανής

Οι πρόσφατες εξελίξεις στις συσκευές εισόδου/εξόδου αποκαλύπτουν ότι, ο τρόπος με τον οποίο επικοινωνούμε με τις ψηφιακές οθόνες, τείνει να αλλάζει. Περιφερειακές συσκευές όπως το ποντίκι και το πληκτρολόγιο, αντικαθίστανται από διεπαφές αφής και κίνησης οι οποίες είναι γνωστές ως «φυσικές διεπαφές χρήστη» (Natural User Interfaces - NUIs). Αυτός ο όρος, ο οποίος απεικονίζεται αφηρημένα στην Εικόνα 4, είναι μία σχετικά νέα έννοια στο πεδίο της ανάδρασης ανθρώπου-μηχανής. Η λέξη «φυσική (natural)», χρησιμοποιείται για το λόγο ότι, οι περισσότερες διεπαφές υπολογιστών είναι τεχνητές συσκευές ελέγχου, όπου οι χρήστες πρέπει να εκπαιδευτούν για τον τρόπο λειτουργίας τους. Δηλαδή, ένα NUI βασίζεται σε έναν χρήστη ο οποίος, μπορεί να πραγματοποιήσει φυσικές κινήσεις ή χειρονομίες κατά τις οποίες ανακαλύπτει και ανακτά γρήγορα τον έλεγχο των εφαρμογών ή του περιεχομένου της οθόνης. Έτσι, αν και έχουν γίνει διάφορες απόπειρες για να δοθεί ο ορισμός ενός NUI, δεν έχει βρεθεί κάποιος που να το περιγράφει επακριβώς. Μία απόπειρα ορισμού ενός NUI είναι η ακόλουθη:

«Ένα NUI δεν είναι αυστηρά ένα CLI (command line interface) ή ένα GUI (graphical user interface)».

Το NUI είναι ένας τύπος διεπαφής χρήστη, σχεδιασμένος με τέτοιο τρόπο, ώστε να επιτρέπει αλληλεπίδραση μεταξύ χρήστη και μηχανής με όσο το δυνατόν πιο φυσικό τρόπο. Τέτοιου είδους διεπαφές διέπονται από διαφάνεια, καθώς δεν είναι ορατές στους χρήστες. Ουσιαστικά το NUI είναι μία αλληλεπιδραστική μεθοδολογία η οποία, εστιάζει σε ανθρώπινες ικανότητες όπως η αντίληψη της έκφρασης. Αυτές οι διεπαφές, θα πρέπει να

αντιμετωπίσουν τη ροή ποικίλων τρόπων επικοινωνίας, οι οποίοι, βασίζονται σε δεξιότητες που έχουν μάθει οι άνθρωποι μέσω της αλληλεπίδρασής τους με τον πραγματικό κόσμο.

Οι εφαρμογές των NUIs περιλαμβάνουν συλλογή δεδομένων από διάφορους αισθητήρες που ανιχνεύουν κίνηση, χειρονομίες, ήχο, ερμηνεία των δεδομένων που προέρχονται από τον/τους αισθητήρα-ες και τέλος, στην απεικόνισή τους. Αυτές οι εφαρμογές ποικίλουν ανάλογα με το πεδίο έρευνας. Ενδεικτικά τέτοιες εφαρμογές, μπορεί να είναι η αναγνώριση χειρονομιών με χρήση κάμερας [2], ο έλεγχος ραδιολογικών εικόνων κατά τη διάρκεια χειρουργικής επέμβασης [3], ασύρματος έλεγχος νευρομυϊκής συσκευής ηλεκτρικού διεγέρτη με χρήση έξυπνου τηλεφώνου [4], κ.α.

### 1.3.4 Μηχανική Μάθηση

Ως μάθηση, θα μπορούσε να οριστεί η χρήση των δεδομένων που λαμβάνει κανείς από το περιβάλλον με στόχο, να βελτιώνεται σταδιακά μέσα από παράθεση εκπαιδευτικού υλικού. Ως μηχανική μάθηση στα υπολογιστικά συστήματα, καλείται η δημιουργία μοντέλων από ένα σύνολο δεδομένων. Πιο συγκεκριμένα, ένας αυστηρότερος ορισμός κατά τον Mitchell [5] είναι: «Ένα πρόγραμμα υπολογιστή θεωρείται ότι μαθαίνει από την εμπειρία  $E$  σε σχέση με μια κατηγορία εργασιών  $T$  και ένα μέτρο απόδοσης  $P$ , αν η απόδοσή του σε εργασίες της κατηγορίας  $T$ , όπως αποτιμούνται από την  $P$ , βελτιώνονται με την εμπειρία  $E$ ». Η μηχανική μάθηση, είναι ουσιαστικά ένα πεδίο της τεχνητής νοημοσύνης, το οποίο αφορά αλγόριθμους και μοντέλα τα οποία, επιτρέπουν στα υπολογιστικά συστήματα να μαθαίνουν χρησιμοποιώντας στατιστικές μεθόδους που εξυπηρετούν στην ανάλυση των δεδομένων.

Αν και υπάρχουν παραλλαγές για τον ορισμό των τύπων της μηχανικής μάθησης, συνήθως, είναι εφικτό αυτές να διαχωριστούν σε κατηγορίες αναλόγως με τη φύση του εκάστοτε προβλήματος, εμπίπτοντας σε μία εκ των δύο ακόλουθων κατηγοριών.

- **Επιβλεπόμενη μάθηση (supervised learning):**

Σε αυτό το είδος μάθησης το υπολογιστικό σύστημα καλείται να «μάθει» μία συνάρτηση η οποία αποκαλείται συνάρτηση στόχος (target function) και αποτελεί περιγραφή του παραγόμενου μοντέλου από τα δεδομένα. Επιπλέον, τα δεδομένα για εκπαίδευση που δίνονται στον εκάστοτε αλγόριθμο και περιλαμβάνουν την επιθυμητή λύση, ονομάζονται ετικέτες (labels). Μία τυπική διαδικασία η οποία πραγματοποιείται σε τέτοιου είδους μάθηση είναι η ταξινόμηση (classification).

- **Μη επιβλεπόμενη μάθηση (unsupervised learning):**  
Σε αυτό το είδος μάθησης, το υπολογιστικό σύστημα καλείται να «ανακαλύψει» από μόνο του συσχετίσεις σε ένα σύνολο δεδομένων (τα οποία δεν έχουν ετικέτες), δημιουργώντας έτσι πρότυπα αγνώστου αριθμού και ύπαρξης.

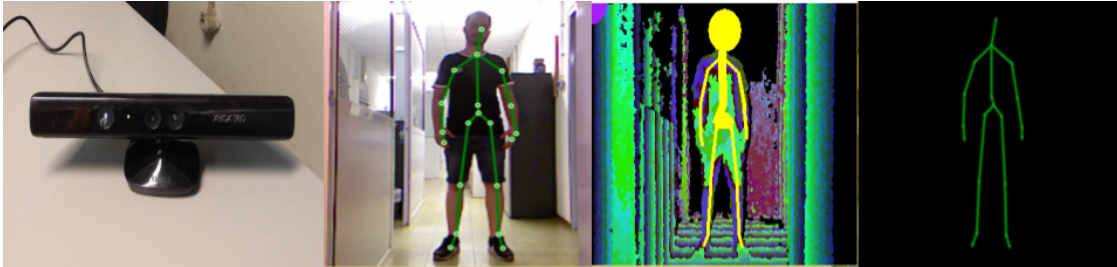
## 1.4 Δομή εργασίας

Το υπόλοιπο αυτής της εργασίας, έχει την εξής δομή: Στο 2<sup>ο</sup> κεφάλαιο αναφέρονται τα τεχνικά χαρακτηριστικά της κάμερας Kinect. Στο 3<sup>ο</sup> κεφάλαιο περιγράφονται τα χαρακτηριστικά που εξάγονται από τα καρέ βίντεο. Στο 4<sup>ο</sup> κεφάλαιο γίνεται αναφορά σε σχετικές εργασίες αναγνώρισης χειρονομιών. Στο 5<sup>ο</sup> και στο 6<sup>ο</sup> κεφάλαιο αναφέρονται οι μέθοδοι και τα μοντέλα που χρησιμοποιήθηκαν στο πλαίσιο αυτής της εργασίας και παρουσιάζονται τα πειραματικά αποτελέσματα. Στο 7<sup>ο</sup> κεφάλαιο γίνεται ανασκόπηση της εργασίας και καταγράφονται τα συμπεράσματα που προέκυψαν κατά την υλοποίηση αυτής. Επιπλέον, γίνεται αναφορά σε μελλοντική εργασία.



## 2. Ο αισθητήρας Kinect

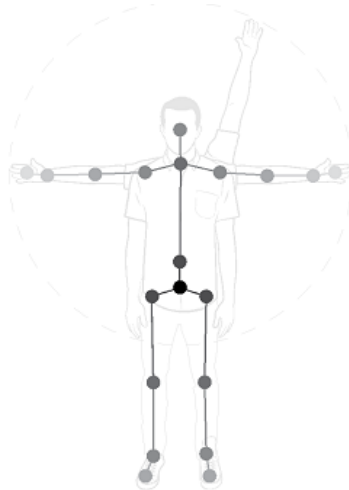
### 2.1 Εισαγωγή και σύντομη ιστορία



Εικόνα 5: Από αριστερά προς τα δεξιά: α) Το Kinect β) RGB εικόνα γ) εικόνα βάθους δ) Σκελετός

Το Kinect [6] είναι μία περιφερειακή συσκευή για το XBOX 360 (καθώς και για το XBOX One στη δεύτερη του έκδοση) της Microsoft. Η συγκεκριμένη συσκευή δεν κατασκευάστηκε ολοκληρωτικά από την Microsoft, αλλά είναι αποτέλεσμα χρόνιας και εκτεταμένης ακαδημαϊκής έρευνας στον τομέα της αναγνώρισης κίνησης και φωνής. Η Microsoft ξεκίνησε την ανάπτυξή του, υπό την κωδική ονομασία «Project Natal». Διέθεσε το προϊόν στην αγορά τον Νοέμβριο του 2010, θέλοντας έτσι να αλλάξει τον τρόπο με τον οποίο αλληλεπιδρούσαν οι χρήστες με τις παιχνιδομηχανές τους, μέσω της φυσικής διεπαφής χρήστη. Λόγω του χαμηλού του κόστους, το Kinect είναι ουσιαστικά, η πρώτη συσκευή η οποία εισήγαγε την έννοια και τη χρήση του όρου της φυσικής διεπαφής χρήστη στους μαζικούς χρήστες.

Προκειμένου να εξαχθούν τα σκελετικά δεδομένα, γίνεται χρήση του Kinect, το οποίο παρέχει σε πραγματικό χρόνο βίντεο RGB, εικόνα βάθους και παρακολούθηση (tracking) σκελετικής εμποπτείας, όπως απεικονίζεται στην Εικόνα 5. Βασίζεται στην εργασία των Shotton κ.α [2], παρέχοντας έναν εύρωστο αλγόριθμο για κάθε άτομο, σε οποιοδήποτε φόντο, χωρίς να χρειάζεται ένα βήμα βαθμονόμησης. Ένας άνθρωπος περιγράφεται με ένα δομημένο γράφο αρθρώσεων, ο οποίος παρουσιάζει τα βασικά σημεία του σώματός του (δηλ. παλάμες, πόδια, κεφάλι, ώμοι, κλπ.). Για κάθε άρθρωση, οι αντίστοιχες 3D συντεταγμένες εξάγονται σε πραγματικό χρόνο. Αυτά τα δεδομένα αποτελούνται από την ταυτότητα και τις 3D συντεταγμένες κάθε άρθρωσης. Πρέπει να σημειώσουμε ότι οι αρθρώσεις είναι οργανωμένες σε ιεραρχική δομή, όπου υπαινίσσεται μία σχέση πατρικού-θυγατρικού. Για παράδειγμα η ρίζα είναι το ισχίο και οι απόγονοί της είναι η σπονδυλική στήλη, το αριστερό και δεξιό τμήμα της λεκάνης κλπ.

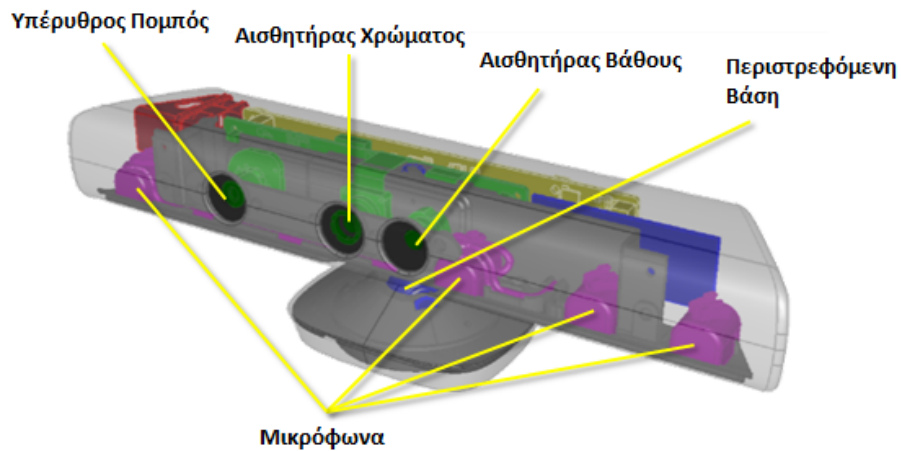


Hip Center					
Spine			Hip Left	Hip Right	
Shoulder Center			Knee Left	Knee Right	
Shoulder Left	Head	Shoulder Right	Ankle Left	Ankle Right	
Elbow Left		Elbow Right	Foot Left	Foot Right	
Wrist Left		Wrist Right			
Hand Left		Hand Right			

Εικόνα 6: Η ιεραρχία των σκελετικών αρθρώσεων

Το Kinect μπορεί να αναγνωρίσει έως έξι άτομα και να εποπτεύσει δύο (όρθια είτε καθιστά). Οι σκελετικές αρθρώσεις που μπορεί να εποπτεύσει απεικονίζονται ιεραρχικά δομημένες στην παραπάνω Εικόνα 6. Αν και το Kinect υποστηρίζεται έως και σήμερα από την Microsoft, η εταιρεία αποφάσισε να διακόψει την παραγωγή του το 2017.

## 2.2 Τεχνικές Προδιαγραφές



Εικόνα 7: Αναπαράσταση του αισθητήρα Kinect

Το Kinect απαρτίζεται από τις ακόλουθες συσκευές (Εικόνα 7):

- **RGB κάμερα:**  
Αυτή η κάμερα αποθηκεύει δεδομένα σε ανάλυση 640x960 , παράγοντας μιας καλής ποιότητας εικόνα με 30 καρέ ανά δευτερόλεπτο. Υπάρχει επίσης και η δυνατότητα να ανέβει η ανάλυση στα 1280x960 με λιγότερα καρέ ανά δευτερόλεπτο.
- **Έναν υπέρυθρο πομπό και έναν αισθητήρα βάθους:**  
Το Kinect έχει τη δυνατότητα να δημιουργήσει 3D χάρτες βάθους. Έτσι, ο πομπός εκπέμπει μία σειρά από υπέρυθρες δέσμες φωτός οι οποίες προβάλλονται στο χώρο ως σημεία. Ο αισθητήρας βάθους τα δέχεται ως ανακλώμενες δέσμες φωτός. Οι ανακλώμενες δέσμες μετατρέπονται σε πληροφορίες βάθους οι οποίες μετρούν την απόσταση μεταξύ ενός αντικειμένου και του αισθητήρα.
- **Μία συστοιχία μικροφώνων:**  
Το Kinect περιέχει τέσσερα μικρόφωνα, τα οποία δίνουν τη δυνατότητα καταγραφής του ήχου, καθώς και εύρεσης της τοποθεσίας προέλευσης της ηχητικής πηγής. Η συχνότητα δειγματοληψίας είναι ορισμένη στα 16 KHz και ο φορέας της παλμοκωδικής διαμόρφωσης στα 24bit .

- **Επιταχυνσιόμετρο:**  
Είναι μία ενσωματωμένη συσκευή η οποία μετράει την επιτάχυνση. Μπορεί να χρησιμοποιηθεί για να δώσει τον προσανατολισμό του Kinect. Είναι φτιαγμένο για 2G εύρος, με G να δηλώνει τη βαρυτική επιτάχυνση. Το επιταχυνσιόμετρο επιστρέφει τις 3Δ συνταγμένες υπό μορφή 4Δ, εισάγοντας με την τιμή της 4<sup>ης</sup> συντεταγμένης ορισμένη πάντα στο 0.00
- **Μηχανικά περιστρεφόμενη βάση:**  
Ρυθμίζει την κλίση του Kinect με εύρος  $\pm 27^\circ$ .

## 2.3 Εφαρμογές χρήσης του Kinect

Το Kinect, είναι ευρέως διαδεδομένο σε εφαρμογές διαφορετικών πεδίων, όπως, η ιατρική, η εικονική πραγματικότητα, η νοηματική γλώσσα, η εκπαίδευση, τα παιχνίδια, η εγκληματολογία, κλπ. Η συνεχής ανάπτυξη εφαρμογών φέρνει τη χρήση τέτοιων συσκευών πιο κοντά στην ανθρώπινη καθημερινότητα και δραστηριότητα. Αναφορικά τέτοιες εφαρμογές είναι:

- Επίβλεψη ηλικιωμένων σε έξυπνα σπίτια και βοήθεια αυτών σε περίπτωση πτώσης [7].
- Μυοσκελετικοί τραυματισμοί μουσικών [8]: Οι μουσικοί, λόγω της συχνής και πολύωρης καθημερινής μελέτης, αντιμετωπίζουν πολλές φορές τραυματισμούς λόγω κόπωσης ή λανθασμένης τοποθέτησης των χεριών τους στο εκάστοτε όργανο. Με τη χρήση του Kinect μπορούν να αποφευχθούν τέτοιου είδους τραυματισμοί, καθώς γίνεται δυνατή η επίβλεψη αυτών σε πραγματικό χρόνο.
- Νοηματική Γλώσσα [9]: Πολλές φορές, οι άνθρωποι οι οποίοι δεν έχουν τη δυνατότητα να μιλήσουν, περιθωριοποιούνται και αποκλείονται από κοινωνικές εργασίες και δραστηριότητες. Με τη βοήθεια του Kinect, τους δίνεται η δυνατότητα να μετατρέψουν τις κινήσεις της νοηματικής γλώσσας σε ομιλία, βοηθώντας στις καθημερινές τους εργασίες.
- Αξιολόγηση ασθενών με Parkinson μέσω χρήσης του Kinect [10].

## 2.4 Κάμερες Βάθους και APIs

Εκτός του Kinect υπάρχουν και άλλες αξίες αναφοράς κάμερες βάθους οι οποίες παρουσιάζονται ονομαστικά παρακάτω:

- ASUS Xtion Pro<sup>1</sup>
- Ercon Systems<sup>2</sup>
- Occipital Structure<sup>3</sup>
- Intel SR300<sup>4</sup>
- Orbbec Astra<sup>5</sup>
- Intel RealSense R200<sup>6</sup>
- Stereolabs ZED Stereo Camera<sup>7</sup>
- Intel RealSense F200<sup>8</sup>
- Orbbec Persee<sup>9</sup>
- DUO mini Ix<sup>10</sup>
- Leap Motion Controller<sup>11</sup>

Ακόμη εκτός του Microsoft SDK, υπάρχουν και τα ακόλουθα APIs για εποπτεία σκελετικών αρθρώσεων, τα οποία παρουσιάζονται ονομαστικά παρακάτω:

- Leap SDK<sup>12</sup>
- Bone API<sup>13</sup>
- NUITRACK SDK<sup>14</sup>
- OpenNI<sup>15</sup>
- Open Kinect Project<sup>16</sup>
- Jester API<sup>17</sup>
- UsdSkel API<sup>18</sup>

---

<sup>1</sup> [https://www.asus.com/gr/3D-Sensor/Xtion\\_PRO/](https://www.asus.com/gr/3D-Sensor/Xtion_PRO/)

<sup>2</sup> <http://ercon.com/>

<sup>3</sup> <https://structure.io/>

<sup>4</sup> <https://software.intel.com/en-us/realsense/sr300>

<sup>5</sup> <https://orbbec3d.com/product-astra/>

<sup>6</sup> <https://software.intel.com/en-us/realsense/previous>

<sup>7</sup> <https://www.stereolabs.com/zed/>

<sup>8</sup> <https://www.intel.com/>

<sup>9</sup> <https://orbbec3d.com/>

<sup>10</sup> <https://duo3d.com/product/duo-minilx-lv1>

<sup>11</sup> <https://www.leapmotion.com/>

<sup>12</sup> <https://developer.leapmotion.com/>

<sup>13</sup> <https://www.npmjs.com/package/bone-api>

<sup>14</sup> <https://nuitrack.com/>

<sup>15</sup> <http://www.openni.ru/>

<sup>16</sup> [https://openkinect.org/wiki/Main\\_Page](https://openkinect.org/wiki/Main_Page)

<sup>17</sup> <https://github.com/jolie/jester>

<sup>18</sup> [https://graphics.pixar.com/usd/docs/api/usd\\_skel\\_page\\_front.html](https://graphics.pixar.com/usd/docs/api/usd_skel_page_front.html)

## 3. Εξαγωγή Χαρακτηριστικών

### 3.1 Αναγνώριση Χειρονομιών

Η αναγνώριση των χειρονομιών είναι ένα αρκετά πολύπλοκο ζήτημα, διότι υπάρχουν πολλοί αστάθμητοι παράγοντες που πρέπει να ληφθούν υπόψη. Έτσι, το υπολογιστικό σύστημα, για να είναι ικανό να αντιλαμβάνεται την κάθε χειρονομία θα πρέπει να έχει προηγηθεί ένας διαχωρισμός των δειγμάτων των καρτέ βίντεο, προκειμένου να έχουν ταξινομηθεί οι χειρονομίες με βάση την κατηγορία (νόημα) της κάθε μιας. Με βάση τα παραπάνω μπορεί να γίνει ανακατασκευή 3Δ μοντέλων μέσω ανάλυσης διαφορετικών στιγμιότυπων από τα καρτέ βίντεο του κάθε ατόμου. Για να γίνει αυτό εφικτό πρέπει πρώτα να εξαχθούν κάποια χαρακτηριστικά (τα οποία είναι χωρικά εντοπισμένα, όπως περιοχές εικόνας) από όλα τα καρτέ βίντεο τα οποία παρουσιάζονται σε επόμενη ενότητα. Ουσιαστικά στόχος είναι η μάθηση της συνάρτησης κατά την οποία θα μπορούν να αντιστοιχίζονται οι χειρονομίες με βάση την κίνηση του κάθε αντικειμένου στον 3Δ χώρο.

### 3.2 Επιλογή Χαρακτηριστικών

Πίνακας 1. Τα σύμβολα και η επεξήγηση αυτών που χρησιμοποιήθηκαν σε αυτήν την εργασία

Σύμβολο	Ορισμός
$J$	μια δοθείσα άρθρωση
$J_c, J_p$	άρθρωση παιδί, γονέας, αντίστοιχα
$F_i$	ένα δοσμένο πλαίσιο βίντεο, $i = 1, \dots, N$
$v_i^J$	διάνυσμα 3Δ συντεταγμένων του $J$ στο $F_i$
$v_{x,i}^{(J)}, v_{y,i}^{(J)}, v_{z,i}^{(J)}$	οι 3Δ συντεταγμένες του $v_i^J$
$\mathcal{J}$	το σύνολο όλων των αρθρώσεων
$V^J$	το σύνολο όλων των διανυσμάτων $v_i^J$ , με $J \in \mathcal{J}$ , για $i = 1, \dots, N$
$B(\bullet)$	ένα τρισδιάστατο πλαίσιο οριοθέτησης της συνόλου διανυσμάτων
$a_{B(\bullet)}, b_{B(\bullet)}$	τα μήκη των πλευρών του $B(\bullet)$

**Πίνακας 2.** Τα προτεινόμενα χαρακτηριστικά, που εξήχθησαν από τις σκελετικές αρθρώσεις

<b>Όνομα Χαρακτηριστικού</b>	<b>Εμπλεκόμενα Καρέ</b>	<b>Εξίσωση</b>
Χωρική γωνία (Spatial Angle)	$F_2, F_1$	$\arccos \frac{v_2^{(J)} - v_1^{(J)}}{\ v_2^{(J)} - v_1^{(J)}\ }$
Χωρική γωνία (Spatial Angle)	$F_n, F_{n-1}$	$\arccos \frac{v_n^{(J)} - v_{n-1}^{(J)}}{\ v_n^{(J)} - v_{n-1}^{(J)}\ }$
Χωρική γωνία (Spatial Angle)	$F_n, F_1$	$\arccos \frac{v_n^{(J)} - v_1^{(J)}}{\ v_n^{(J)} - v_1^{(J)}\ }$
Συνολικό διάνυσμα γωνίας (Total Vector Angle)	$F_1 \dots F_n$	$\sum_{i=1}^n \arccos \left( \frac{v_i^{(J)} \cdot v_{i-1}^{(J)}}{\ v_i^{(J)}\  \cdot \ v_{i-1}^{(J)}\ } \right)$
Τετραγωνικό συνολικό διάνυσμα γωνίας (Squared Total Vector Angle)	$F_1 \dots F_n$	$\sum_{i=1}^n \arccos \left( \frac{v_i^{(J)} \cdot v_{i-1}^{(J)}}{\ v_i^{(J)}\  \cdot \ v_{i-1}^{(J)}\ } \right)^2$
Διάνυσμα Ευκλείδειας μετατόπισης (Total Vector Displacement)	$F_n, F_1$	$\ v_n^{(J)} - v_1^{(J)}\ $
Συνολική μετατόπιση (Total Displacement)	$F_1 \dots F_n$	$\sum_{i=1}^n \ v_i^{(J)} - v_{i-1}^{(J)}\ $
Μέγιστη μετατόπιση (Maximum Displacement)	$F_1 \dots F_n$	$\max_{i=2, \dots, n} (\ v_i^{(J)}\  - \ v_{i-1}^{(J)}\ )$
Μήκος διαγωνίου περιβάλλοντος κιβωτίου (Bounding Box Diagonal Length)	$F_1 \dots F_n$	$\sqrt{a_{B(v^{(J)})}^2 + b_{B(v^{(J)})}^2}$
Γωνία διαγωνίου περιβάλλοντος κιβωτίου (Bounding Box Angle)	$F_1 \dots F_n$	$\arctan \frac{b_{B(v^{(J)})}}{a_{B(v^{(J)})}}$

Για την έναρξη της διαδικασίας της αναγνώρισης των χειρονομιών θα πρέπει αρχικά, να οριστεί ένα σύνολο χαρακτηριστικών, τα οποία εξάγονται από τα εκάστοτε καρέ εικόνων του Kinect. Το προτεινόμενο σύνολο χαρακτηριστικών, είναι μερικής εμπνευσμένο από εκείνο που προτάθηκε από τον Sheng [11], του οποίου η εργασία επέκτεινε εκείνη του Rubine [12], από τον 2D στον 3D χώρο. Σύμφωνα με τον Sheng, τα εξαγόμενα χαρακτηριστικά θα πρέπει να είναι ορισμένα έτσι ώστε, να ικανοποιούν τα ακόλουθα κριτήρια: α) θα πρέπει να απαιτούν σταθερό χρόνο, β) οι μικρές αλλαγές στις χειρονομίες θα πρέπει να αντανakλούνται σε μικρές αλλαγές στα εξαγόμενα χαρακτηριστικά, και γ) τα χαρακτηριστικά θα πρέπει να έχουν εννοιολογικό νόημα.

Πιο συγκεκριμένα, έστω  $\mathbf{J}$  μία δοθείσα άρθρωση. Τότε  $\mathbf{J}_c$  και  $\mathbf{J}_p$  είναι οι συντεταγμένες απόγονοι και πρόγονοι αντίστοιχα. Επίσης, έστω  $\mathbf{F}_i, i = 1, 2, \dots, N$  ένα καρέ βίντεο και έστω  $\mathbf{v}_i^{(j)} = (v_{x,i}^{(j)}, v_{y,i}^{(j)}, v_{z,i}^{(j)})$  ένα διάνυσμα το οποίο αντιστοιχεί τις 3D συντεταγμένες του  $\mathbf{J}$  στο  $i$ -οστο πλαίσιο  $F$ . Επιπλέον, έστω  $V^{(j)}$  το σύνολο όλων των  $v_i^{(j)}$ . Με  $B(V^{(j)})$  προσδιορίζουμε την οριοθέτηση του  $V^{(j)}$  με,  $a_{B(V^{(j)})}$  και  $b_{B(V^{(j)})}$  ως τα δύο διαφορετικά μήκη των πλευρών του. Τα εξαγόμενα χαρακτηριστικά απεικονίζονται στον Πίνακα 2. Θα πρέπει να ληφθεί υπόψιν ότι κάθε χαρακτηριστικό υπολογίζεται για μία δοθείσα άρθρωση  $\mathbf{J}$  και μπορεί να περιλαμβάνει διαφορετικό υποσύνολο καρέ. Ο συμβολισμός αυτών συνοψίζεται στον Πίνακα 1.

Ειδικότερα, εξάγεται η χωρική γωνία μεταξύ των δύο πρώτων πλαισίων (δηλ. ,  $F_1$  και  $F_2$ ), των δύο τελευταίων (δηλ. ,  $F_{N-1}$  και  $F_N$ ) και του πρώτου με το τελευταίο πλαίσιο (δηλ. ,  $F_1$  και  $F_N$ ), χρησιμοποιώντας σε όλες τις περιπτώσεις το εσωτερικό γινόμενο των αντίστοιχων διανυσμάτων. Επιπλέον, εξάγεται η συνολική και η τετραγωνική συνολική γωνία των διανυσμάτων, μεταξύ των  $F_1$  και  $F_N$  χρησιμοποιώντας το εσωτερικό γινόμενο των αντίστοιχων διανυσμάτων, δηλ. ,  $v_1^{(j)}$  και  $v_N^{(j)}$ .

Η συνολική διανυσματική μετατόπιση είναι το γινόμενο της διαφοράς των διανυσμάτων που ανταποκρίνονται στις αρχικές και τελικές θέσεις του  $\mathbf{J}$ , όσο η συνολική μετατόπιση είναι το άθροισμα των γινομένων όλων των διαφορών μεταξύ των θέσεων του  $\mathbf{J}$ , εντός δύο οποιονδήποτε δοσμένων διαδοχικών πλαισίων. Όμοια, η μέγιστη μετατόπιση είναι το γινόμενο της μέγιστης διαφοράς μεταξύ των θέσεων του  $\mathbf{J}$  εντός οποιονδήποτε δύο πλαισίων.

Τέλος, κατασκευάζεται ένα τρισδιάστατο «κουτί» οριοθέτησης (bounding box) για το σύνολο όλων των θέσεων των αρθρώσεων εντός όλων των πλαισίων και σχηματίζεται σύμφωνα με το μήκος και τη γωνία της διαγωνίου του. Εδώ θα πρέπει να ληφθεί υπόψιν ότι, από τα σκελετικά δεδομένα χρησιμοποιείται μόνο ένα υποσύνολο αρθρώσεων (δηλ. τις 3D συντεταγμένες τους). Ακόμη, να σημειωθεί ότι,  $\mathbf{J} \in \{\text{Αριστερός/Δεξιός Αγκώνας}, \text{Αριστερή/Δεξιά Παλάμη}, \text{Αριστερός/Δεξιός Καρπός}, \text{Αριστερός/Δεξιός/Κεντρικός Ώμος}, \text{Κέντρο Γοφών}\}$ . ( βλέπε Εικόνα 2). Επίσης μεταξύ δύο αρθρώσεων προγόνων και αρθρώσεων απογόνων δηλ.  $\{\mathbf{J}_p, \mathbf{J}\}$ , (ή ισοδύναμα  $\{\mathbf{J}, \mathbf{J}_c\}$ )



) υπολογίζεται η αρχική και η τελική (δηλ. στο  $F_1$  και στο  $F_2$ , αντίστοιχα), μέση και μέγιστη γωνία (π.χ. για το  $F_1$  και  $F_2$ ). Η γωνία  $\theta_{pc}$  που χρησιμοποιήθηκε σε αυτούς τους υπολογισμούς ορίζεται ως:

$$\theta_{pc} = \cos^{-1} \left( \frac{a_{pc}^2 + b_{pc}^2 + c_{pc}^2}{2a_{pc}b_{pc}} \right), \quad (1)$$

όπου 
$$a_{pc}^2 = \left( v_x^{(J)} - v_x^{(Jc)} \right)^2 + \left( v_y^{(J)} - v_y^{(Jc)} \right)^2, \quad (2)$$

$$b_{pc} = v_x^{(J)} \quad \text{και} \quad (3)$$

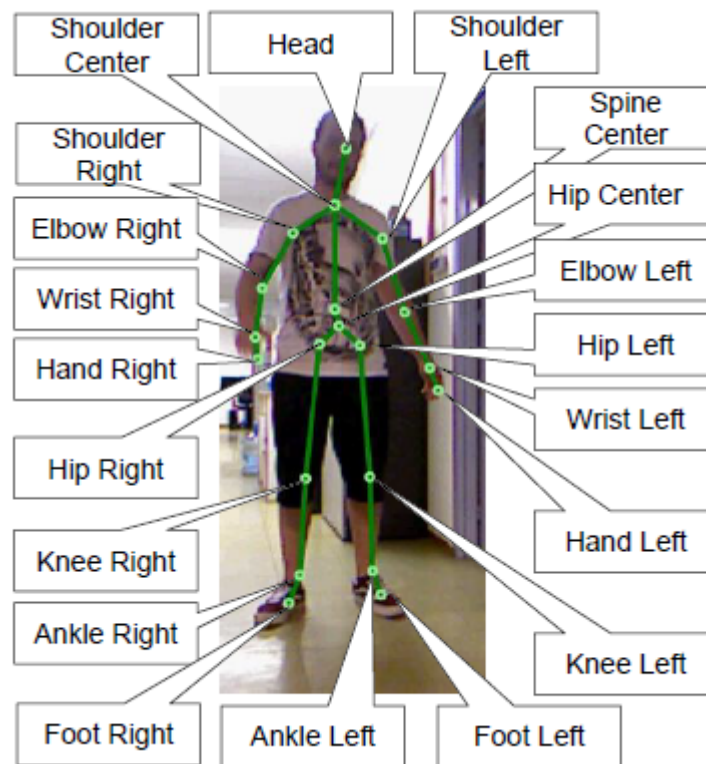
$$c_{pc}^2 = \left( v_x^{(Jp)} \right)^2 + \left( v_y^{(J)} - v_y^{(Jc)} \right)^2 \quad (4)$$

Τέλος, μεταξύ της αριστερής (HL) και δεξιάς παλάμης (HR) εξάγεται η μέγιστη και η ελάχιστη απόσταση (που περιλαμβάνονται στην χειρονομία), οι οποίες ορίζονται ως:

$$d_{max} = \max_{i,j} \{d(v_i^{HR}, v_j^{HL})\}, \quad (5)$$

και 
$$d_{mean} = \frac{1}{F(J)} \sum_{i,j} d(v_i^{HR}, v_j^{HL}), \text{ αντίστοιχα.} \quad (6)$$

### 3.3 Αναπαράσταση Σκελετικών Αρθρώσεων



Εικόνα 8: Οι σκελετικές αρθρώσεις του ανθρώπου που εξάγονται με χρήση του Kinect SDK

Τα σκελετικά δεδομένα, αναπαρίστανται γεωμετρικά σε τρεις διαστάσεις. Ουσιαστικά η απεικόνιση τους στο χώρο λαμβάνει υπόψη τις παραμέτρους του χρόνου, της μετατόπισης στο επίπεδο και του βάθους. Στην Εικόνα 8 παρουσιάζεται ο τρόπος αναπαράστασης των ανθρώπινων σκελετικών αρθρώσεων. Αυτές οι αρθρώσεις σχεδιάζονται στο προγραμματιστικό περιβάλλον της γλώσσας Python με χρήση της βιβλιοθήκης Pygame, έτσι ώστε να γίνονται εμφανείς κατά τη διαδικασία παρακολούθησης και να εξάγονται μετρήσεις για την κάθε άρθρωση ξεχωριστά.

## 4. Σχετικές Εργασίες

Κατά τη διάρκεια των τελευταίων ετών, πολλές ερευνητικές προσπάθειες έχουν εστιάσει στο πρόβλημα της αναγνώρισης χειρονομιών. Σε αυτό το κεφάλαιο παρουσιάζονται ερευνητικές εργασίες, οι οποίες επιχειρούν να αναγνωρίσουν παρόμοιες χειρονομίες, με αυτές που αναγνωρίζονται σε αυτήν την εργασία.

Οι παραδοσιακές προσεγγίσεις της μηχανικής μάθησης, όπως τα τεχνητά νευρωνικά δίκτυα (ANN), οι μηχανές διανυσμάτων στήριξης (SVM), τα δένδρα αποφάσεων (DT) ή οι ταξινομητές K-κοντινότερων γείτονων (KNN), έχουν χρησιμοποιηθεί ευρέως. Οι Bhattacharya κ.α. [13], χρησιμοποίησαν μια χαρακτηριστική χειρονομία για να υποδείξει την αρχή της διαδικασίας αναγνώρισης και να αποφύγουν χρονική κατάτμηση. Η προσέγγισή τους, χρησιμοποίησε SVM (Support Vector Machines) και DT (Decision Trees) πάνω σε 3D συντεταγμένες σκελετικών αρθρώσεων. Οι Lai κ.α. [14], επέλεξαν τη σπονδυλική στήλη ως σημείο αναφοράς, υπολογίζοντας τις αποστάσεις από τους αγκώνες και τις παλάμες και τις χρησιμοποίησαν ως χαρακτηριστικά. Για την αναγνώριση των χειρονομιών χρησιμοποίησαν KNN (K- Nearest Neighbour) ταξινομητή. Οι Mangera κ.α. [15], χρησιμοποίησαν 3D αρθρώσεις οι οποίες υπολογίστηκαν σε εξαγμένα καρέ-κλειδιά από ακολουθίες και αλληλουχίες ANNs (Artificial Neural Networks), προκειμένου, να ταξινομήσουν αρχικά την πλευρά της χειρονομίας (αριστερή/δεξιά) και στη συνέχεια να την αναγνωρίσουν. Οι Miranda κ.α. [16], χρησιμοποίησαν SVMs (Support Vector Machines) για την αναγνώριση χαρακτηριστικών στάσεων. Αυτοί μοντελοποίησαν χειρονομίες ως ακολουθίες από στάσεις κλειδιά και χρησιμοποίησαν δάση αποφάσεων (Decision Forests) για την αναγνώριση αυτών. Οι Ting κ.α. [17], χρησιμοποίησαν τετραδιάστατες συντεταγμένες για την περιγραφή των αρθρώσεων και SVMs για την αναγνώριση των χειρονομιών.

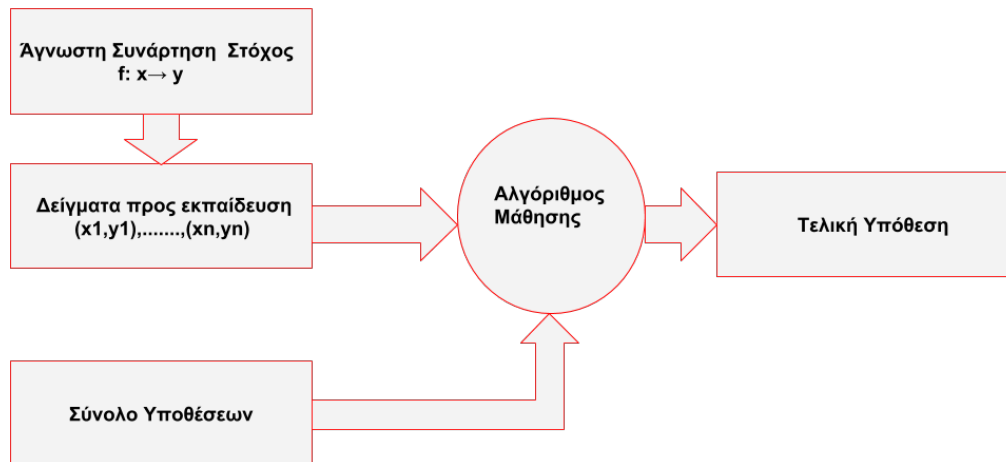
Αφού γενικά, οι χειρονομίες είναι χρονικές ακολουθίες, οι οποίες ποικίλλουν σε ταχύτητα (για παράδειγμα όταν εκτελούνται από διαφορετικούς χρήστες), ο δυναμικός αλγόριθμος χρονικής στρέβλωσης (Dynamic Time Wrapping algorithm) [18], έχει χρησιμοποιηθεί σε πολλές εργασίες για το λόγο ότι είναι ικανός να αντιμετωπίσει με ακρίβεια αυτό το πρόβλημα. Οι Celebi κ.α. [19], χρησιμοποίησαν δυναμικό αλγόριθμο χρονικής στρέβλωσης σε κανονικοποιημένα τρισδιάστατα σημεία για κάθε άτομο. Οι Reyes κ.α. [20], πρότειναν μία παραλλαγή με βάρη του δυναμικού αλγορίθμου χρονικής στρέβλωσης, χρησιμοποιώντας το λαιμό ως σημείο αναφοράς. Οι Ribó κ.α. [21], παρουσίασαν μια παρόμοια προσέγγιση, η οποία βασίζεται σε δυναμικό αλγόριθμο χρονικής στρέβλωσης, KNN ταξινομητές και κάποιους ευρετικούς αλγορίθμους για τη βελτίωση της απόδοσης. Οι Ibañez κ.α. [22], χρησιμοποίησαν τρισδιάστατες αρθρώσεις και σύγκριναν το δυναμικό αλγόριθμο στρέβλωσης με HMM (Hidden Markov Models), αποκτώντας την ίδια ακρίβεια.

Κρυφά Μαρκοβιανά Μοντέλα (HMM) και Ανάμεικτα Γκαουσιανά Μοντέλα (GMM), επίσης έχουν συχνά χρησιμοποιηθεί, λόγω της ικανότητάς τους να λύνουν χρονικά

προβλήματα αναγνώρισης προτύπων. Οι Anuj κ.α. [23], μείωσαν το πρόβλημα 3D παρακολούθησης σε 2D, προβάλλοντας τις παρακολουθούμενες αρθρώσεις σε ένα σύστημα συντεταγμένων προσκολλημένο στον χρήστη και ευθυγραμμισμένο προς αυτόν. Η βασισμένη σε κανόνες προσέγγισή τους εξαρτήθηκε από HMM και προσαρμοστικά όρια. Η τεχνική που προτάθηκε από τους Gonzalez-Sanchez και Ruig [24], βασίστηκε στην αφαίρεση του φόντου και στην ανίχνευση χεριών/προσώπου. Οι τρισδιάστατες θέσεις των χεριών και της κεφαλής χρησιμοποιήθηκαν σε GHMM (General Hidden Markov Models). Οι Gu κ.α. [25], εφάρμοσαν HMM σε συστοιχίες από τρισδιάστατο σύνολο αρθρώσεων. Οι Tran και Trivedi [26], ανίχνευσαν το κεφάλι και τα χέρια εκπαιδεύοντας στη συνέχεια GMMs (General Markov Models) και επιβάλλοντας κινηματικούς περιορισμούς στη διαδικασία. Τέλος οι Yin και Davis [27], επόπτευαν χέρια και άντλησαν χαρακτηριστικά διανύσματα τα οποία αντιπροσώπευαν το σχήμα του χεριού, βασισμένα σε HOGs (Histogram Of Oriented Gradients) και χαρακτηριστικά κίνησης. Έπειτα, χρησιμοποίησαν ιεραρχικό μοντέλο HMM για αναγνώριση σε πραγματικό χρόνο.

## 5. Μέθοδοι Και Μοντέλα

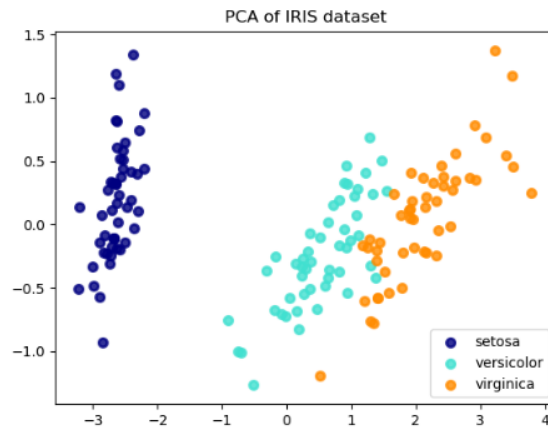
### 5.1 Εκπαίδευση μοντέλων Μηχανικής Μάθησης



Εικόνα 9: Διαδικασία εκπαίδευσης ενός μοντέλου

Σε αυτό το στάδιο της διαδικασίας, ο εκάστοτε αλγόριθμος μηχανικής μάθησης, προσπαθεί να αποκτήσει τις κατάλληλες πληροφορίες, οι οποίες, θα του επιτρέψουν κατά το στάδιο του ελέγχου να ερευνήσει και να εντοπίσει τα χαρακτηριστικά που διέπουν κάθε χειρονομία. Αυτές οι πληροφορίες είναι ουσιαστικά, τα δεδομένα προς εκπαίδευση τα οποία θα πρέπει να περιέχουν τη σωστή απάντηση, η οποία είναι γνωστή ως «στόχος» . Έπειτα, ο αλγόριθμος βρίσκει πρότυπα στα δεδομένα προς εκπαίδευση τα οποία αντιστοιχίζουν τα χαρακτηριστικά των δεδομένων εισόδου στο γνώρισμα-στόχο, δηλαδή στην απάντηση προς πρόβλεψη. Ο αλγόριθμος μάθησης έχει ως αποτέλεσμα ένα μοντέλο μηχανικής μάθησης το οποίο εντοπίζει αυτά τα πρότυπα. Επομένως, το κάθε μοντέλο, μπορεί να χρησιμοποιηθεί για προβλέψεις σε νέα δεδομένα, για το οποία ο στόχος είναι άγνωστος. Η διαδικασία εκπαίδευσης απεικονίζεται στην Εικόνα 9.

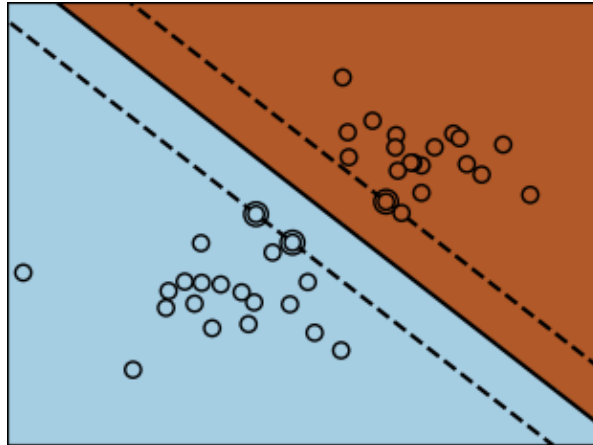
## 5.2 Η μέθοδος Ανάλυσης Κύριων Συνιστωσών (PCA)



Εικόνα 10: Στιγμιότυπο της μεθόδου ανάλυσης κύριων συνιστωσών για τρεις κλάσεις

Η ανάλυση κύριων συνιστωσών, είναι μία στατιστική μέθοδος η οποία μετατρέπει μία ομάδα παρατηρήσεων δυνητικά συσχετισμένων μεταβλητών σε μία ομάδα νέων, μη-γραμμικά συσχετισμένων μεταβλητών, οι οποίες καλούνται κύριες συνιστώσες. Ουσιαστικά αυτή η μέθοδος είναι ένας ορθογώνιος μετασχηματισμός, κατά τον οποίο αφαιρείται η συσχέτιση των μεταβλητών και μειώνονται οι διαστάσεις τους. Είναι ένα σημαντικό εργαλείο στην επεξεργασία σημάτων, το οποίο εφαρμόζεται συχνά στην επεξεργασία εικόνων, με σκοπό τη συμπίεση δεδομένων, τη μείωση των διαστάσεων σε ένα σύνολο δεδομένων ή την αποσυσχέτιση δεδομένων, όπως π.χ. απεικονίζει η Εικόνα 10 .

## 5.3 Μηχανές Διανυσμάτων Υποστήριξης (SVMs)



Εικόνα 11: Ταξινόμηση κλάσεων με SVM

Οι μηχανές διανυσμάτων υποστήριξης, είναι μοντέλα μάθησης με επίβλεψη και ανήκουν στην κατηγορία των γραμμικών ταξινομητών, καθώς επιτυγχάνουν το διαχωρισμό των δεδομένων με τη χρήση υπερεπιπέδων. Αυτά τα μοντέλα σχετίζονται με αλγορίθμους μάθησης, οι οποίοι αναλύουν δεδομένα, αναγνωρίζουν πρότυπα και χρησιμοποιούνται για ταξινόμηση και ανάλυση παλινδρόμησης. Δοθέντος ενός συνόλου παραδειγμάτων προς εκπαίδευση, όπου κάθε παράδειγμα είναι προκαθορισμένο να ανήκει σε κάποια κατηγορία, ένας αλγόριθμος μηχανής διανυσμάτων υποστήριξης, κατασκευάζει ένα μοντέλο στο οποίο αναθέτει νέα παραδείγματα σε μία κατηγορία ή σε άλλη. Επιπλέον, μία μηχανή διανυσμάτων υποστήριξης είναι μία αναπαράσταση των παραδειγμάτων ως σημεία στο χώρο, τα οποία αντιστοιχίζονται με τέτοιο τρόπο, ώστε παραδείγματα διαφορετικών κατηγοριών να έχουν διαιρεθεί έχοντας όσο το δυνατόν μέγιστο περιθώριο. Δηλαδή ο ταξινομητής, προσπαθεί να διαχωρίσει το σύνολο των παραδειγμάτων, ούτως ώστε εκείνα τα παραδείγματα που ανήκουν στην ίδια κατηγορία, να βρίσκονται στην ίδια πλευρά του υπερεπιπέδου. Παράδειγμα τέτοιου υπερεπιπέδου απεικονίζεται στην Εικόνα 11.

## 5.4 Μέθοδοι Πυρήνων

Οι συναρτήσεις πυρήνα, είναι μία τάξη αλγορίθμων που συχνά χρησιμοποιείται στο πεδίο της αναγνώρισης προτύπων. Ουσιαστικά, είναι απεικονίσεις των διανυσμάτων εισόδου  $\mathbf{x}$  στο σύνολο των πραγματικών αριθμών  $\mathbb{R}$ , οι οποίες έχουν συγκεκριμένη μορφή και ιδιότητες που γενικεύουν σε μεγάλο βαθμό τις εφαρμογές των αλγορίθμων ταξινόμησης. Σύμφωνα με τη μέθοδο των πυρήνων, τα διανύσματα εισόδου μετασχηματίζονται με κατάλληλο τρόπο, ούτως ώστε, να επιτευχθεί καλύτερη λύση στο εκάστοτε πρόβλημα ταξινόμησης. Για να πραγματοποιηθεί το προαναφερθέν, η μέθοδος πυρήνων πηγαίνει από τον χώρο εισόδου σε έναν μετασχηματισμένο χώρο χαρακτηριστικών ο οποίος, έχει υψηλότερη διάσταση με μία απεικόνιση  $\varphi(x)$ . Η συνάρτηση πυρήνα ορίζεται ως:

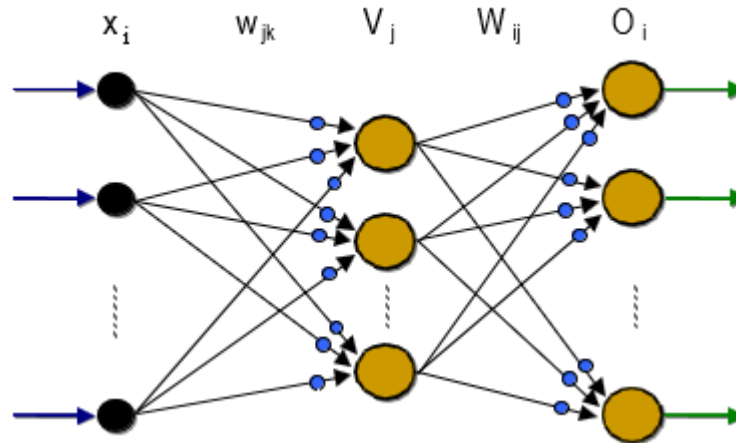
$$k(x, x') = \varphi^T(x)\varphi(x') = \langle \varphi(x)\varphi(x') \rangle \quad (7)$$

Είναι δηλαδή το εσωτερικό γινόμενο μεταξύ των διανυσμάτων  $\varphi(x)$  και  $\varphi(x')$ .

Αν και υπάρχουν διαφορετικά είδη πυρήνα όπως: RBF (Radial Basis Function), Linear (Γραμμικός), Polynomial (Πολυωνυμικός), Sigmoid (Σιγμοειδής), σε αυτήν την εργασία χρησιμοποιήθηκαν μόνο οι δύο πρώτοι (σε συνδυασμό με SVM). Τέλος, πολλοί αλγόριθμοι, είναι δυνατόν να λειτουργήσουν με χρήση μεθόδων πυρήνα, αναφορικά, μερικοί από αυτούς είναι: Μηχανές Διανυσμάτων Υποστήριξης (SVMs), Ανάλυση Κύριων Συνιστωσών (PCA), Γκαουσιανές Διεργασίες.



## 5.5 Πολυεπίπεδο Perceptron (Multilayer Perceptron)



Εικόνα 12: Δομή Perceptron πολλαπλών επιπέδων

Ένα MLP (Multilayer Perceptron), ανήκει στην κλάση δικτύων πρόσθιας τροφοδότησης και αποτελείται από παραπάνω από έναν νευρώνες. Τα δίκτυα πρόσθιας τροφοδότησης είναι η πιο απλή μορφή νευρωνικών δικτύων και χαρακτηριστικό τους είναι ότι η ροή πληροφορίας είναι μονής κατεύθυνσης. Τέτοια δίκτυα αποτελούνται από ένα επίπεδο εισόδου (το οποίο δέχεται σήματα εισόδου) και ένα ή περισσότερα ενδιάμεσα κρυφά επίπεδα, καθώς και ένα επίπεδο εξόδου (το οποίο λαμβάνει μία απόφαση ή κάνει μία πρόβλεψη). Αυτού του είδους τα δίκτυα περιλαμβάνουν ουσιαστικά, τουλάχιστον τρία επίπεδα κόμβων όπου, κάθε κόμβος είναι ένας νευρώνας ο οποίος, χρησιμοποιεί μία μη-γραμμική συνάρτηση ενεργοποίησης. Μεταξύ νευρώνων ίδιου επιπέδου δεν υπάρχουν συνδέσεις, όμως υπάρχει διασύνδεση μεταξύ δύο διαδοχικών επιπέδων.

Πιο συγκεκριμένα: το πρώτο επίπεδο που βρίσκεται αριστερά στην Εικόνα 12, καλείται επίπεδο εισόδου και αποτελείται από ένα σύνολο νευρώνων  $\{x_i | x_1, x_2, \dots, x_m\}$  το οποίο αναπαριστά τα χαρακτηριστικά εισόδου. Κάθε νευρώνας κρυφού επιπέδου, μετασχηματίζει τις τιμές προηγούμενου επιπέδου με ένα γραμμικό άθροισμα με βάρη  $w_1x_1 + w_2x_2 + \dots + w_mx_m$ , το οποίο ακολουθείται από μία μη-γραμμική συνάρτηση ενεργοποίησης  $g(\bullet): R \rightarrow R$ . Το επίπεδο εξόδου λαμβάνει τιμές από το τελευταίο κρυφό επίπεδο και τις μετασχηματίζει σε τιμές εξόδου. Τα MLPs, έχουν ως πλεονέκτημα ότι μπορούν να μάθουν μη-γραμμικά μοντέλα, καθώς και να μάθουν μοντέλα σε πραγματικό χρόνο (on-line learning)

Τα MLPs χρησιμοποιούνται συχνά σε προβλήματα μάθησης με επίβλεψη και καλύπτουν μεγάλο εύρος εφαρμογών του πραγματικού κόσμου όπως: Συμπίεση δεδομένων, οδήγηση, αναγνώριση προτύπων/υπολογιστική όραση, αναγνώριση ομιλίας, ανάλυση πρωτεϊνικών δομών, πρόβλεψη σοδειάς, κ.α.

## 5.6 Extra Trees (Extremely Randomized Trees)

Ο αλγόριθμος Extra Trees [28], δημιουργεί ένα σύνολο δένδρων αποφάσεων ή παλλινδρόμησης. Αυτό το σύνολο, αναπτύσσεται με βάση την κλασσική διαδικασία από πάνω προς τα κάτω. Οι δύο βασικές του διαφορές σε σχέση με άλλες μεθόδους ανάλυσης με βάση το δένδρο είναι, ότι, α) οι κόμβοι διασπώνται με εντελώς τυχαία σημεία αποκοπής και β) χρησιμοποιείται όλο το μαθησιακό δείγμα ως προς την ανάπτυξη των δένδρων. Τα Extra Trees, εκπαιδεύονται, χρησιμοποιώντας μέθοδο bagging ή μέθοδο τυχαίων υποχώρων όπως σε έναν κοινό αλγόριθμο Random Forests. Σε αυτόν τον αλγόριθμο, αντί να υπολογίζεται κάθε φορά το τοπικό βέλτιστο χαρακτηριστικό, επιλέγεται μία τυχαία τιμή για τον διαχωρισμό από τον σύνολο δεδομένων προς εκπαίδευση. Η διαδικασία επαναλαμβάνεται αρκετές φορές έως ότου, να δημιουργηθεί ένα μοντέλο από το αρχικό δείγμα μάθησης.

## 5.7 AdaBoost

Ο AdaBoost, είναι ένας αλγόριθμος μηχανικής μάθησης, ο οποίος μπορεί να χρησιμοποιηθεί σε συνδυασμό με πολλούς άλλους τύπους αλγορίθμων μάθησης, με στόχο τη βελτίωση της απόδοσης. Ο AdaBoost, είναι προσαρμοστικός αλγόριθμος υπό την έννοια ότι, εστιάζει σε προβλήματα ταξινόμησης με στόχο τη μετατροπή ενός συνόλου αδύναμων ταξινομητών σε δυνατότερους, βελτιώνοντάς τους. Είναι ένας αλγόριθμος ευαίσθητος σε θορυβώδη δεδομένα και υπερβολικές τιμές. Η βασική ιδέα αυτού του αλγορίθμου, είναι να θέσει βάρη τόσο στους ταξινομητές, όσο και στα δείγματα προς εκπαίδευση, αναγκάζοντας έτσι τους ταξινομητές να επικεντρώσουν το ενδιαφέρον τους σε δείγματα τα οποία είναι δύσκολο να ταξινομηθούν σωστά. Αυτή η διαδικασία εκτελείται σειριακά, με βάση ότι δύο βάρη ρυθμίζονται σε κάθε βήμα, καθώς οι επαναλήψεις του αλγορίθμου συνεχίζονται.

## 6. Πειράματα–Αποτελέσματα

### 6.1 Σύνολο Δεδομένων

Για το σκοπό της αξιολόγησης της μεθόδου που χρησιμοποιήθηκε, κατασκευάστηκε ένα πραγματικό σύνολο δεδομένων από 10 χρήστες (7 άνδρες και 3 γυναίκες), ηλικίας 22 έως 36 με διαφορετικό ύψος, το οποίο κυμαίνεται από 155 εκ. έως 190 εκ. Ορίστηκε ένα σύνολο χειρονομιών το οποίο περιλαμβάνει τις εξής 15 διαφορετικές κινήσεις/χειρονομίες που απαρτίζουν το σύνολο δεδομένων:

1. Χειροκρότημα (Clapping)
2. Ύψωση και των δύο χεριών (Two-hands Raise)
3. Σφίξιμο και των δύο χεριών (Two-hands Squeeze)
4. Κύκλοι με τα δύο χέρια (Two-hands Circle)
5. Διαγώνιο σύρσιμο με το αριστερό χέρι (Swipe-diagonal-left)
6. Σύρσιμο προς τα κάτω με το αριστερό χέρι (Swipe-down-left)
7. Σύρσιμο προς τα πάνω με το δεξί χέρι (Swipe-down right)
8. Σύρσιμο (προς τα δεξιά) με το αριστερό χέρι (swipe-left)
9. Σύρσιμο (προς τα αριστερά) με το δεξί χέρι (swipe-right)
10. Σύρσιμο προς τα πάνω με το αριστερό χέρι (swipe-up-left)
11. Σύρσιμο προς τα πάνω με το δεξί χέρι (swipe-up-right)
12. Κυματισμός με το δεξί χέρι (wave-right)
13. Περιστροφή των καρπών με τα δύο χέρια (Two-hands-wrist-rotation)
14. Μεγέθυνση (Zoom-in)
15. Σμίκρυνση (Zoom-out)

Για τη δημιουργία του συνόλου δεδομένων, οι χρήστες τοποθετήθηκαν σε απόσταση περίπου 2 μέτρων από το Kinect, ούτως ώστε να εξασφαλιστεί ότι ο σκελετός θα εξάγεται με επιτυχία. Από κάθε ακολουθία η οποία αντιστοιχούσε σε χειρονομία εξήχθησαν τα ακατέργαστα χρωματικά RGB-D (Red, Green, Blue -Depth) και τα σκελετικά δεδομένα. Τα RGB-D δεδομένα έχουν χρησιμοποιηθεί μόνο για σκοπούς οπτικής απεικόνισης, αφού η μεθοδολογία που χρησιμοποιήθηκε στη συγκεκριμένη εργασία βασίστηκε αποκλειστικά σε σκελετικά δεδομένα. Αρχικά παρουσιάστηκαν στους χρήστες οι 15 χειρονομίες, με στόχο να εξοικειωθούν με αυτές πριν τη διαδικασία συλλογής των χειρονομιών. Κατά τη διάρκεια της διαδικασίας, ζητήθηκε από τους χρήστες να εκτελέσουν κάθε χειρονομία τουλάχιστον 5 φορές. Έτσι για κάθε χειρονομία έγινε καταγραφή των σκελετικών δεδομένων και εξήχθη το

σύνολο των χαρακτηριστικών που περιγράφονται στο 3<sup>ο</sup> κεφάλαιο, για όλες τις εμπλεκόμενες αρθρώσεις σε αυτές τις χειρονομίες, δηλαδή παλάμες, καρπούς, αγκώνες, ώμους, κεφαλή και ισχίο.

## 6.2 Πειραματικά αποτελέσματα

Πραγματοποιήθηκαν πειράματα, με ένα σύνολο ευρέως χρησιμοποιηθέντων ταξινομητών, οι οποίοι, προέρχονται από την βιβλιοθήκη Scikit-Learn [29] και είναι οι ακόλουθοι: α) Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines), τόσο γραμμικές όσο και με πυρήνα ακτινικής συνάρτησης βάσης (rbf), β) Πολυεπίπεδους Νευρώνες (Multilayer Perceptron), και τέλος γ) AdaBoost με Extra Trees (ABET). Στον Πίνακα 3 απεικονίζονται οι ταξινομητές που χρησιμοποιήθηκαν στο πλαίσιο αυτής της εργασίας με τις βέλτιστες παραμέτρους τους. Σε κάθε πείραμα, οι ταξινομητές εκπαιδεύονταν με αξιοποίηση μίας βέλτιστης τιμής παραμέτρου, η οποία είχε καθοριστεί με χρήση διασταύρωσης (Cross Validation).

Πιο συγκεκριμένα όπως απεικονίζεται στον Πίνακα 3: στον RBFSVM χρησιμοποιήθηκαν ως παράμετροι οι: C (προαιρετική παράμετρος) με τιμή 44.445 η οποία αφορά την ποινή του όρου σφάλματος και η παράμετρος  $\gamma$  η οποία είναι ο συντελεστής πυρήνα με τιμή 0.0001. Στον LSVM χρησιμοποιήθηκε μόνο η παράμετρος C με τιμή 0.0091. Ακολούθως για τον ABET χρησιμοποιήθηκαν οι παράμετροι  $e$  και  $\alpha$ , με το  $e$  να προσδιορίζει τον αριθμό των εκτιμητών με τιμή 82 και το  $\alpha$  να προσδιορίζει τον ρυθμό μάθησης με τιμή 241.6. Όσον αφορά τον ταξινομητή MLP δεν έγινε χρήση κάποιας παραμέτρου, αφήνοντας έτσι τις προκαθορισμένες παραμέτρους.

**Πίνακας 3:** Βέλτιστες παράμετροι ταξινομητών ( $\alpha$ : ρυθμός μάθησης,  $e$ : αριθμός εκτιμητών,  $\gamma$ : Συντελεστής πυρήνα, C: ποινή του όρου σφάλματος)

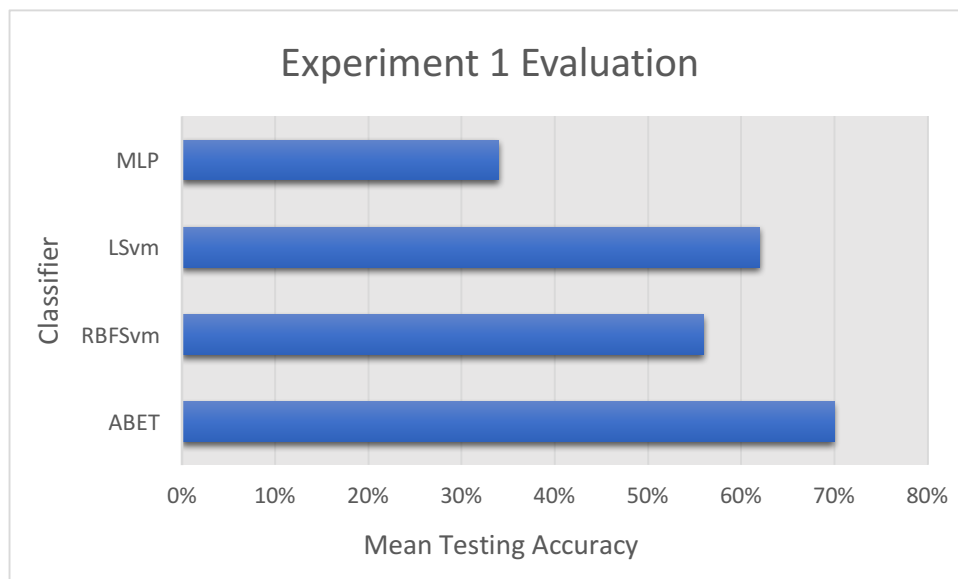
Ταξινομητής	Παράμετροι
RBFSVM	$C = 44.445, \gamma = 0.0001$
LSVM	$C = 0.0091$
ABET	$e = 82, \alpha = 241.6$
MLP	default παράμετροι

Για την πειραματική αξιολόγηση αυτής της εργασίας, εκτελέστηκαν δύο σειρές εκτενών πειραμάτων. Στόχος του πρώτου πειράματος ήταν να ερευνηθεί η απόδοση των μοντέλων που δημιουργήθηκαν, για ένα «γνωστό» σύνολο χρηστών από όλες τις χειρονομίες. Έτσι, κατέστη δυνατό και στα δύο πειράματα, να χρησιμοποιηθούν δείγματα χειρονομιών για εκπαίδευση από τους 9 χρήστες, αφήνοντας 1 χρήστη κάθε φορά για έλεγχο. Ο στόχος του δεύτερου πειράματος ήταν να ερευνηθεί η απόδοση για ένα σύνολο «γνωστών» χρηστών (όπως και στο πρώτο πείραμα), με τη διαφορά ότι δε

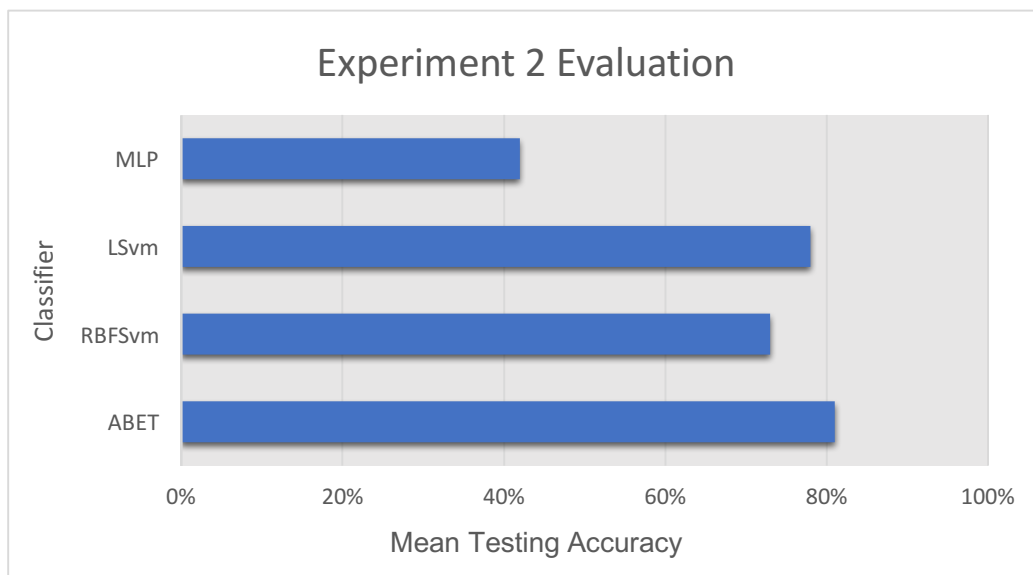
χρησιμοποιήθηκαν όλες οι χειρονομίες για ταξινόμηση. Έτσι, ο αριθμός των κλάσεων μειώθηκε κατά 5, αφού λάβαμε υπόψιν τα αποτελέσματα που έδωσαν οι ταξινομητές του Πίνακα 3 στο πρώτο πείραμα. Επομένως, το σύνολο των χειρονομιών περιλάμβανε τις ακόλουθες χειρονομίες:

1. Χειροκρότημα (Clapping)
2. Ύψωση και των δύο χεριών (Two-hands Raise)
3. Κύκλοι με τα δύο χέρια (Two-hands Circle)
4. Διαγώνιο σύρσιμο με το αριστερό χέρι (Swipe-diagonal-left)
5. Σύρσιμο προς τα κάτω με το αριστερό χέρι (Swipe-down-left)
6. Σύρσιμο προς τα κάτω με το δεξί χέρι (Swipe-down-right)
7. Σύρσιμο (προς τα δεξιά) με το αριστερό χέρι (swipe-left)
8. Σύρσιμο (προς τα αριστερά) με το δεξί χέρι (swipe-right)
9. Σύρσιμο προς τα πάνω με το αριστερό χέρι (swipe-up-left)
10. Σύρσιμο προς τα πάνω με το δεξί χέρι (swipe-up-right)

Στην συνέχεια παρουσιάζονται τα αποτελέσματα της αξιολόγησης των ταξινομητών α) για το πρώτο πείραμα (Εικόνα 13) β) για το δεύτερο πείραμα (Εικόνα 14)



Εικόνα 13: Ποσοστό αξιολόγησης ακρίβειας των ταξινομητών για το πρώτο πείραμα



Εικόνα 14: Ποσοστό αξιολόγησης ακρίβειας των ταξινομητών για το δεύτερο πείραμα

Όπως είναι εμφανές από την Εικόνα 13 , η ταξινόμηση για το πρώτο πείραμα δεν δίνει ορθά αποτελέσματα, για διάφορους λόγους, όπως: λανθασμένο ή ελλιπές σύνολο δεδομένων, λανθασμένες παράμετροι στους ταξινομητές, λάθος διαχωρισμός δειγμάτων προς εκπαίδευση και τεστ, κ.α. Παρ' όλα αυτά, ο ταξινομητής ABET παρουσιάζει την καλύτερη απόδοση με ποσοστό 70%. Στη συνέχεια ο επόμενος καλύτερος ταξινομητής ήταν ο LSVM με ποσοστό 62%, ξεπερνώντας τον RBFSVM, ο οποίος αξιολογήθηκε με ποσοστό 56%. Τέλος, για το πρώτο πείραμα, το MLP παρουσίασε τη χειρότερη απόδοση με ποσοστό 34%.

Στη συνέχεια, με βάση αυτά τα αποτελέσματα, για την επίτευξη καλύτερης απόδοσης, μειώθηκε ο αριθμός των κλάσεων κατά 5. Η Εικόνα 14 παρουσιάζει το ποσοστό της μέσης ακρίβειας προς αξιολόγηση που επετεύχθη από κάθε ταξινομητή για το δεύτερο πείραμα. Καλύτερη απόδοση είχε ο ταξινομητής ABET, ο οποίος πέτυχε ποσοστό 81%. Στη συνέχεια, ο επόμενος καλύτερος ταξινομητής ήταν ο LSVM με ποσοστό 78%, ξεπερνώντας τον τον RBFSVM, ο οποίος αξιολογήθηκε με ποσοστό 73%. Τέλος, για το δεύτερο πείραμα, χειρότερη απόδοση με ποσοστό 42% παρουσίασε ο MLP. Όπως ήταν αναμενόμενο, με τη μείωση του αριθμού των κλάσεων, αυξήθηκε η συνολική ακρίβεια του συστήματος, κάνοντας το να έχει καλύτερη απόδοση.

## 7. Μελλοντικές εργασίες και συμπεράσματα

Η αναγνώριση χειρονομιών και οι συναφείς εφαρμογές σε αυτήν, αποτελούν ένα αρκετά ενδιαφέρον πεδίο έρευνας που συγκεντρώνει ολοένα και αυξανόμενη προσοχή τα τελευταία χρόνια. Η παρούσα πτυχιακή, επικεντρώθηκε στο πρόβλημα της αναγνώρισης χειρονομιών μελετώντας με χρήση αλγορίθμων μηχανικής μάθησης χαρακτηριστικά που προσδιορίζουν το είδος της εκάστοτε χειρονομίας από ένα σύνολο απλών χειρονομιών που εκτελέστηκαν σε πραγματικό χρόνο. Αρχικά δημιουργήθηκε ένα σύνολο δεδομένων που περιλαμβάνει εικόνες RGB και βάθους, αλλά και σκελετικά δεδομένα για διάφορες χειρονομίες. Έπειτα, παρουσιάστηκε μία μέθοδος εξαγωγής διανυσματικών χαρακτηριστικών κατά την οποία, εξήχθησαν χαρακτηριστικά από τις 3Δ θέσεις σκελετικών αρθρώσεων και τέλος, παρουσιάστηκαν τα αποτελέσματα από κάθε αλγόριθμο μηχανικής μάθησης. Καλύτερα αποτελέσματα έδωσε ο ταξινομητής ABET και στα δύο πειράματα, όπου στο δεύτερο η απόδοσή του αυξήθηκε κατά 11% .

Ως σκέψεις για μελλοντική επέκταση αυτής της εργασίας, θα ήταν ο εμπλουτισμός του συνόλου χειρονομιών με περισσότερες χειρονομίες, η χρήση περισσότερων Kinect ή άλλων καμερών βάθους με σκοπό την αναγνώριση χειρονομιών περισσότερων ατόμων, έτσι ώστε να δημιουργηθεί ένα μεγαλύτερο σύνολο δεδομένων. Ακόμη, για την καταγραφή των δεδομένων από το Kinect, ο χρήστης θα μπορούσε να ορίζει μέσω φωνητικών εντολών την αρχή και το τέλος της καταγραφής της χειρονομίας, καθώς το Kinect έχει τη δυνατότητα καταγραφής του ήχου. Επιπρόσθετα θα ήταν ικανοποιητικό αν ο αριθμός των καρτέ βίντεο ήταν ίδιος για όλα τα δείγματα. Ως μελλοντική επέκταση θα ήταν η χρήση αυτής της εργασίας, σε εφαρμογές IoT (Internet Of Things) υπηρεσίες οι οποίες, περιλαμβάνουν έξυπνες συσκευές που θα μπορούσαν να λειτουργήσουν με αναγνώριση κινήσεων. Μία άλλη εφαρμογή θα μπορούσε να ήταν σε οχήματα τελευταίας τεχνολογίας όπου η οδήγηση γίνεται μέσω χειρονομιών. Τέλος, ως η σημαντικότερη επέκταση αυτής της εργασίας στο μέλλον, θα ήταν η εκπαίδευση και η δημιουργία μοντέλων με χρήση αλγορίθμων από το πεδίο της βαθιάς μάθησης ( Deep Learning).

## Αναφορές

1. Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3), 311-324.
2. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., ... & Blake, A. (2011, June). Real-time human pose recognition in parts from single depth images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (pp. 1297-1304). Ieee.
3. Nestorov, N., Hughes, P., Healy, N., Sheehy, N., & OHare, N. (2016, June). Application of natural user interface devices for touch-free control of radiological images during surgery. In *Computer-Based Medical Systems (CBMS), 2016 IEEE 29th International Symposium on* (pp. 229-234). IEEE.
4. Sweeney, D., Quinlan, L. R., & ÓLaighin, G. (2015, August). Smartphone app design for the wireless control of a neuromuscular electrical stimulator device with integrated randomization allocation process for RCT applications. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (pp. 4574-4577). IEEE.
5. Mitchell, T. M. (1997). Machine learning. WCB.
6. Zhang, Z. (2012). Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2), 4-10.
7. Galvão, Y. M., Albuquerque, V. A., Fernandes, B. J., & Valença, M. J. (2017, November). Anomaly detection in smart houses: Monitoring elderly daily behavior for fall detecting. In *Computational Intelligence (LA-CCI), 2017 IEEE Latin American Conference on* (pp. 1-6). IEEE.
8. Li, M., Savvidou, P., Willis, B., & Skubic, M. (2014, August). Using the Kinect to detect potentially harmful hand postures in pianists. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 762-765). IEEE.
9. Muttena, S., Sriram, S., & Shiva, R. (2014, November). Mapping gestures to speech using the kinect. In *Science Engineering and Management Research (ICSEMR), 2014 International Conference on* (pp. 1-5). IEEE.
10. Salimpour, Y., Chien, J. H., Liu, C. C., Lee, S., Guadix, S., Mills, K. A., & Anderson, W. S. (2017, July). Touch-free reaching task for Parkinson's disease patients: A motion sensing approach. In *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE* (pp. 3918-3921). IEEE.



11. Sheng, J. (2003). A study of adaboost in 3d gesture recognition. *Department of Computer Science, University of Toronto*.
12. Rubine, D. (1991). *Specifying gestures by example* (Vol. 25, No. 4, pp. 329-337). ACM.
13. Bhattacharya, S., Czejdo, B., & Perez, N. (2012, November). Gesture classification with machine learning using kinect sensor data. In *Emerging Applications of Information Technology (EAIT), 2012 Third International Conference on*(pp. 348-351). IEEE.
14. Lai, K., Konrad, J., & Ishwar, P. (2012, April). A gesture-driven computer interface using Kinect. In *Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on*(pp. 185-188). IEEE.
15. Mangera, R., Senekal, F., & Nicolls, F. (2014). Cascading neural networks for upper-body gesture recognition. Avestia Publishing.
16. Miranda, L., Vieira, T., Martinez, D., Lewiner, T., Vieira, A. W., & Campos, M. F. (2012, August). Real-time gesture recognition from depth data through key poses learning and decision forests. In *Graphics, patterns and images (SIBGRAPI), 2012 25th SIBGRAPI conference on* (pp. 268-275). IEEE.
17. Ting, H. Y., Sim, K. S., Abas, F. S., & Besar, R. (2014). Vision-based human gesture recognition using Kinect sensor. In *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications* (pp. 239-244). Springer, Singapore.
18. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees, The Wadsworth Statistics and Probability Series, Wadsworth International Group, Belmont California (pp. 356).
19. Celebi, S., Aydin, A. S., Temiz, T. T., & Arici, T. (2013, February). Gesture recognition using skeleton data with weighted dynamic time warping. In *VISAPP (1)* (pp. 620-625).
20. Reyes, M., Dominguez, G., & Escalera, S. (2011, November). Featureweighting in dynamic timewarping for gesture recognition in depth data. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*(pp. 1182-1188). IEEE.
21. Ribó, A., Warcho, D., & Oszust, M. (2016). An Approach to Gesture Recognition with Skeletal Data Using Dynamic Time Warping and Nearest Neighbour Classifier. *International Journal of Intelligent Systems and Applications*, 8(6), 1.
22. Ibanez, R., Soria, Á., Teyseyre, A., & Campo, M. (2014). Easy gesture recognition for Kinect. *Advances in Engineering Software*, 76, 171-180.

23. Anuj, A., Mallick, T., Das, P. P., & Majumdar, A. K. (2015, December). Robust control of applications by hand-gestures. In *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2015 Fifth National Conference on* (pp. 1-4). IEEE.
24. Gonzalez-Sanchez, T., & Puig, D. (2011). Real-time body gesture recognition using depth camera. *Electronics Letters*, 47(12), 697-698.
25. Gu, Y., Do, H., Ou, Y., & Sheng, W. (2012, December). Human gesture recognition through a kinect sensor. In *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*(pp. 1379-1384). IEEE.
26. Tran, C., & Trivedi, M. M. (2012). 3-D posture and gesture recognition for interactivity in smart spaces. *IEEE Transactions on Industrial Informatics*, 8(1), 178-187.
27. Yin, Y., & Davis, R. (2014, July). Real-time continuous gesture recognition for natural human-computer interaction. In *Visual Languages and Human-Centric Computing (VL/HCC), 2014 IEEE Symposium on* (pp. 113-120). IEEE.
28. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1), 3-42.
29. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825-2830.
30. Mitchell, T. M. (1997). *Machine learning*. WCB.
31. Steinberg, G. (2012). Natural user interfaces. In *ACM SIGCHI conference on human factors in computing systems*.
32. Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.