



Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών  
και Μηχανικών Υπολογιστών  
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

**Σχεδίαση συστήματος απομακρυσμένου ελέγχου  
αντικειμένων μέσω στάσεων του σώματος και χειρονομιών  
με χρήση σκελετικών δεδομένων**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΓΕΩΡΓΙΟΣ ΠΑΡΑΣΚΕΥΟΠΟΥΛΟΣ**

**Επιβλέπων :** Στέφανος Κόλλιας  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2016





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών  
και Μηχανικών Υπολογιστών  
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

**Σχεδίαση συστήματος απομακρυσμένου ελέγχου  
αντικειμένων μέσω στάσεων του σώματος και χειρονομιών  
με χρήση σκελετικών δεδομένων**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**ΓΕΩΡΓΙΟΣ ΠΑΡΑΣΚΕΥΟΠΟΥΛΟΣ**

**Επιβλέπων :** Στέφανος Κόλλιας

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 13η Ιουλίου 2016.

.....  
Στέφανος Κόλλιας  
Καθηγητής Ε.Μ.Π.

.....  
Σταύρος Περαντώνης  
Διευθυντής Ερευνών, ΕΚΕΦΕ Δημόκριτος

.....  
Γιώργος Στάμου  
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2016

.....  
**Γεώργιος Παρασκευόπουλος**

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

Copyright © Γεώργιος Παρασκευόπουλος, 2016.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## Περίληψη

Η παρούσα διπλωματική εργασία μελετά τη σχεδίαση και υλοποίηση μιας Φυσικής Διεπαφής Χρήστη για τον έλεγχο των συσκευών σε ένα έξυπνο δωμάτιο με τη χρήση χειρονομιών και στάσεων του σώματος. Η άνοδος του Διαδικτύου των Αντικειμένων και η ανάγκη αλληλεπίδρασης με συσκευές στο φυσικό κόσμο εντείνει την ανάγκη για την υλοποίηση νέων διεπαφών που οδηγούνται από δράσεις του χρήστη. Συγκεκριμένα η εργασία μας επικεντρώνεται στα προβλήματα της ανίχνευσης πόζας και της αναγνώρισης χειρονομιών σε πραγματικό χρόνο και στο συνδυασμό αυτών σε μία αρχιτεκτονική μικροϋπηρεσιών η οποία θα εκτελείται σε ένα ετερογενές υπολογιστικό περιβάλλον.

Αρχικά εστιάζουμε στο πρόβλημα ανίχνευσης πόζας, για το οποίο εξάγουμε μια απλή και αποδοτική λύση με χρήση ταιριάσματος προτύπων πάνω σε γεωμετρικά χαρακτηριστικά που εξάγουμε από τον αισθητήρα Kinect. Επίσης παρέχουμε έναν εύκολο τρόπο για την επέκταση του αλγορίθμου σε νέες πόζες μέσω του ορισμού νέων προτύπων.

Στη συνέχεια προτείνουμε μια μέθοδο για την αναγνώριση χειρονομιών σε πραγματικό χρόνο. Συγκεκριμένα κατασκευάζουμε ένα νέο σύνολο δεδομένων και εξάγουμε ένα σύνολο γεωμετρικών χαρακτηριστικών χαμηλού επιπέδου και συγκρίνουμε την απόδοση πολλών, γνωστών αλγορίθμων μηχανικής πάνω σε αυτά τα χαρακτηριστικά. Η επίδοση της μεθόδου αξιολογείται με μία σειρά πειραμάτων που στοχεύουν να προσδιορίσουν τόσο την ακρίβεια της αναγνώρισης όσο και το χρόνο πρόβλεψης για κάθε ταξινομητή σε διάφορα σενάρια.

Τέλος συνδυάζουμε αυτούς τους αλγορίθμους σε μια αρχιτεκτονική την οποία ενσωματώνουμε στην πλατφόρμα SYNAISTHISI. Το σύστημα υλοποιείται με τη χρήση σύγχρονων τεχνολογιών και αρχών σχεδίασης και αποτελεί μια εφαρμογή στα πλαίσια του Διαδικτύου των Αντικειμένων. Επιπλέον παραθέτουμε μια ποσοτική και ποιοτική αξιολόγηση από τους χρήστες του συστήματος σε πραγματικά σενάρια.

## Λέξεις κλειδιά

Διαδίκτυο των Αντικειμένων, Μηχανική Μάθηση, Δεδομένα Σκελετού, Εξαγωγή χαρακτηριστικών, Kinect, Σύστημα Πραγματικού Χρόνου, Αλληλεπίδραση Ανθρώπου Μηχανής, Φυσική Διεπαφή Χρήστη, Αναγνώριση Χειρονομιών, Ανίχνευση Πόζας, Επικοινωνία Μηχανής με Μηχανή, M2M, MQTT



# Abstract

The objective of this thesis is the design and implementation of a Natural User Interface to enable the interaction with devices in a smart room using gestures and poses. The rise of Internet of Things brings up the need of designing novel, action based user interfaces to interact with devices in the physical world. Specifically our work is focused on the problems of pose detection, real time gesture recognition and composing them in a microservice architecture that will be executed in a heterogeneous computing environment.

Firstly we focus on the problem of pose detection and we devise a simple and efficient template matching based solution using geometric data extracted from the Kinect sensor. Additionally we provide a way to extend the algorithm to new poses by streamlining the definition of pose templates.

Next, we tackle the problem of real time gesture recognition. specifically we construct a real life dataset and extract a novel set of low level geometric features. Then we compare the performance of several well known machine learning algorithms on this set of features by evaluating the recognition accuracy and the prediction time of the classifiers in a set of different scenarios.

Finally we compose these techniques under an architecture which is integrated in the SYNAISTHISI platform. We make use of modern technologies and design principles to implement the final system as an Internet of Things application. Moreover we provide a quantitative and qualitative evaluation from the users in real world scenarios.

## **Key words**

Internet of Things, Machine Learning, Skeletal Data, Feature Extraction, Kinect, Real Time System, Human Computer Interaction, Natural User Interface, Gesture Recognition, Pose Detection, Machine to Machine Communication, M2M, MQTT





## Ευχαριστίες

Η εκπόνηση της παρούσας εργασίας πραγματοποιήθηκε στο Εργαστήριο Υπολογιστικής Νοημοσύνης στο Ε.Κ.Ε.Φ.Ε. Δημόκριτος υπό την καθοδήγηση και την καθοριστική συμβολή του Δρ. Ευάγγελου Σπύρου, τον οποίο θα ήθελα να ευχαριστήσω για τη συνεχή διάθεσή του να υποδείξει νέες ιδέες και προοπτικές οι οποίες ανέβασαν τον πήχη για αυτή την εργασία και για την αβίαστη συνεργασία μας. Επίσης θα ήθελα να ευχαριστήσω τους καθηγητές Στέφανο Κόλλια και Σταύρο Περαντώνη για την εμπιστοσύνη που μου έδειξαν και την προθυμία τους να προτείνουν νέες ιδέες και θεματικές.

Επίσης θα ήθελα να ευχαριστήσω τους γονείς μου Πέτρο και Ευγενία καθώς και την αδερφή μου Παναγιώτα για την αμείωτη υποστήριξη και ενθάρρυνση που μου προσφέρουν.

Γεώργιος Παρασκευόπουλος,

Αθήνα, 13η Ιουλίου 2016



# Περιεχόμενα

<b>Περίληψη</b> . . . . .	5
Abstract . . . . .	7
<b>Ευχαριστίες</b> . . . . .	9
<b>Περιεχόμενα</b> . . . . .	11
<b>Κατάλογος πινάκων</b> . . . . .	15
<b>Κατάλογος σχημάτων</b> . . . . .	17
<b>1. Εισαγωγή</b> . . . . .	19
1.1 Περίληψη της εργασίας . . . . .	19
1.2 Σχετικές Εργασίες . . . . .	20
<b>2. Το Διαδίκτυο των Αντικειμένων</b> . . . . .	23
2.1 Τι είναι το Διαδίκτυο των Αντικειμένων . . . . .	23
2.2 Η ιστορία του Διαδικτύου των Αντικειμένων . . . . .	24
2.2.1 1988: Ubiquitous computing . . . . .	25
2.2.2 Από τη δεκαετία του 90 στον 21ο αιώνα . . . . .	25
2.3 Η πλατφόρμα SYNAISTHISI . . . . .	26
2.3.1 Αρχιτεκτονική προσανατολισμένη σε υπηρεσίες . . . . .	26
2.3.2 Μεσισμικό προσανατολισμένο σε μηνύματα . . . . .	27

2.3.3	Η υλοποίηση . . . . .	28
<b>3.</b>	<b>M2M: Επικοινωνία Μηχανής με Μηχανή . . . . .</b>	<b>31</b>
3.1	Πρωτόκολλα M2M: Το μοντέλο εκδότη/συνδρομητή . . . . .	31
3.2	Η πλατφόρμα ROS . . . . .	33
3.3	Το πρωτόκολλο MQTT . . . . .	36
3.3.1	Περιγραφή του MQTT . . . . .	36
3.3.2	Λειτουργίες του MQTT . . . . .	37
<b>4.</b>	<b>Ο αισθητήρας Kinect . . . . .</b>	<b>43</b>
4.1	Το υλικό του Kinect . . . . .	43
4.2	Εξαγωγή της εικόνας βάθους . . . . .	44
4.2.1	Το υλικό του αισθητήρα . . . . .	45
4.2.2	Ο αλγόριθμος της 3D ανακατασκευής . . . . .	46
4.3	Εξαγωγή των αρθρώσεων του σκελετού . . . . .	49
4.3.1	Συλλογή δεδομένων και εξαγωγή χαρακτηριστικών . . . . .	49
4.4	Το περιβάλλον ανάπτυξης Kinect for Windows SDK . . . . .	51
4.4.1	Η ροή καρέ έγχρωμων εικόνων . . . . .	52
4.4.2	Η ροή εικόνων βάθους . . . . .	52
4.4.3	Η ροή σκελετικών δεδομένων . . . . .	53
4.4.4	Το Kinect SDK v2.0 . . . . .	55
<b>5.</b>	<b>Ανίχνευση Στάσεων του Σώματος . . . . .</b>	<b>57</b>
5.1	Προσδιορισμός του προβλήματος . . . . .	57
5.2	Εξαγωγή γεωμετρικών χαρακτηριστικών . . . . .	59
5.3	Αλγόριθμος Ανίχνευσης Πόζας . . . . .	60
<b>6.</b>	<b>Αναγνώριση χειρονομιών . . . . .</b>	<b>65</b>

6.1	Προσδιορισμός του προβλήματος . . . . .	65
6.2	Εξαγωγή Χαρακτηριστικών . . . . .	67
6.3	Αλγόριθμοι Μηχανικής Μάθησης . . . . .	69
6.4	Πειραματική αξιολόγηση της μεθόδου . . . . .	73
6.4.1	Κατασκευή του συνόλου δεδομένων . . . . .	73
6.4.2	Πειράματα . . . . .	75
<b>7.</b>	<b>Περίπτωση Χρήσης: Έλεγχος έξυπνου δωματίου . . . . .</b>	<b>81</b>
7.1	Το έξυπνο δωμάτιο . . . . .	81
7.2	Αρχιτεκτονική του Συστήματος . . . . .	82
7.2.1	Υπηρεσία Εντολών . . . . .	84
7.2.2	Λεπτομέρειες Υλοποίησης . . . . .	85
7.2.3	Αξιολόγηση του Aigaiο NUI . . . . .	87
<b>8.</b>	<b>Συμπεράσματα και Προεκτάσεις . . . . .</b>	<b>89</b>
	<b>Παράρτημα . . . . .</b>	<b>95</b>
<b>A.</b>	<b>Το XML Schema για τις πρότυπες πόζες . . . . .</b>	<b>95</b>



## Κατάλογος πινάκων

4.1	Τεχνικές προδιαγραφές του αισθητήρα βάθους στο Kinect 360 . . . . .	43
4.2	Τα μέρη του σώματος στην ενδιάμεση αναπαράσταση . . . . .	51
4.3	Σύγκριση των τεχνικών προδιαγραφών του Kinect 2.0 με το Kinect 360 [22]. . . . .	55
6.1	Χαρακτηριστικά που εξάγονται από τα σκελετικά δεδομένα . . . . .	68
6.2	Οι βέλτιστες παράμετροι για την εκπαίδευση κάθε ταξινομητή . . . . .	76
6.3	Το σκορ $F_1$ για κάθε χειρονομία χωριστά και το μέσο σκορ $F_1$ για όλες τις χειρονομίες	78
7.1	Αριθμός επιτυχημένων δράσεων ανά χρήστη κατά την αξιολόγηση του Aigaiο NUI .	88





## Κατάλογος σχημάτων

2.1	Η διαστρωματωμένη αρχιτεκτονική της πλατφόρμας SYNAISTHISI . . . . .	30
3.1	Αρχιτεκτονική pub/sub . . . . .	32
3.2	Παράδειγμα γράφου εφαρμογής στο ROS . . . . .	35
3.3	Flags για το μήνυμα CONNECT στο πρωτόκολλο MQTT . . . . .	37
3.4	Αποστολή μηνύματος MQTT με QoS 0 . . . . .	39
3.5	Αποστολή μηνύματος MQTT με QoS 1 . . . . .	39
3.6	Αποστολή μηνύματος MQTT με QoS 2 . . . . .	39
4.1	Το υλικό του Kinect 360 . . . . .	44
4.2	Εικόνα βάθους από τον αισθητήρα Kinect . . . . .	45
4.3	Μοτίβο κουκκίδων στην τεχνολογία PrimeSense . . . . .	46
4.4	Η ενδιάμεση αναπαράσταση των μερών του σώματος . . . . .	50
4.5	Οι αρθρώσεις που ανιχνεύονται από το Kinect SDK v1.8 . . . . .	53
4.6	Οι καταστάσεις λειτουργίας για την ανίχνευση των αρθρώσεων . . . . .	54
4.7	Η ιεραρχία των αρθρώσεων . . . . .	54
5.1	Παράδειγμα πόζας . . . . .	58
5.2	Το τρίγωνο που σχηματίζεται για τον υπολογισμό της γωνίας ανάμεσα σε δύο αρθρώσεις . . . . .	59
5.3	Απεικόνιση της γωνίας μεταξύ δύο αρθρώσεων στον τριγωνομετρικό κύκλο . . . . .	60
5.4	Η πόζα HANDSAPART . . . . .	62
5.5	Το πρότυπο XML για την πόζα HANDSAPART . . . . .	62

6.1	Μηχανή καταστάσεων για την αναγνώριση χειρονομιών με ενδιάμεσες πόζες . . . . .	67
6.2	Συγχρονισμένα έγχρωμα και σκελετικά καρτέ για δύο χειρονομίες . . . . .	74
6.3	Αποτελέσματα της K-πλής αντεπικύρωσης για διάφορους ταξινομητές . . . . .	77
6.4	Η μέση ακρίβεια των προβλέψεων του ταξινομητή ET με εκπαίδευση σε υποσύνολα χρηστών . . . . .	78
6.5	Ακρίβεια ανά χειρονομία αναγνώρισης για την εκπαίδευση σε υποσύνολα των χρηστών	79
6.6	Ακρίβεια ανά χειρονομία αναγνώρισης για την εκπαίδευση σε υποσύνολα των χρηστών με φιλτραρισμένο σύνολο δεδομένων . . . . .	80
6.7	Μέσος χρόνος αναγνώρισης σε διαφορετικές αρχιτεκτονικές . . . . .	80
7.1	Το MQTT API της έξυπνης αίθουσας . . . . .	82
7.2	Διάγραμμα των υπομονάδων του συστήματος Aigaiο NUI . . . . .	83
7.3	Οι υπομονάδες του Aigaiο NUI χωρισμένες σε υπηρεσίες . . . . .	84
7.4	Ενεργοποίηση εξαγωγής χαρακτηριστικών χειρονομίας . . . . .	86
7.5	Οι υπομονάδες του Aigaiο NUI και οι αντίστοιχες τεχνολογίες . . . . .	86

## Κεφάλαιο 1

# Εισαγωγή

Η μεταφορά πληροφορίας αποτελεί μια από τις πιο βασικές μορφές αλληλεπίδρασης με τον εξωτερικό κόσμο. Καθημερινά χρησιμοποιούμε μια πληθώρα τρόπων και διεπαφών για την επικοινωνία. Για την επικοινωνία με άλλους ανθρώπους μερικά από τα μέσα επικοινωνίας είναι ο προφορικός λόγος και ο τόνος της φωνής, οι εκφράσεις του προσώπου, ο γραπτός λόγος και η γλώσσα του σώματος, με αντίστοιχες διεπαφές το φωνητικό σύστημα, το πρόσωπο, τα χέρια και ολόκληρο το σώμα. Για την επικοινωνία με υπολογιστικά συστήματα χρησιμοποιούνται πιο έμμεσες διεπαφές, όπως πληκτρολόγια, ποντίκια, πάνελ πλήκτρων και οθόνες αφής. Στο παρελθόν έχουν γίνει προσπάθειες για τη χρήση πιο άμεσων διεπαφών με τους υπολογιστές, οι οποίες είναι ακόμα πιο αναγκαίες με την άνοδο του Διαδικτύου των Αντικειμένων και τον κατακλυσμό της αγοράς από μη συμβατικά υπολογιστικά συστήματα.

### 1.1 Περίληψη της εργασίας

Ο στόχος αυτής της εργασίας είναι η δημιουργία μιας φυσικής διεπαφής χρήστη για ένα έξυπνο δωμάτιο. Η διεπαφή αυτή επιτρέπει στο χρήστη να ελέγχει τις συσκευές που βρίσκονται στο δωμάτιο με τη χρήση χειρονομιών και στάσεων του σώματος, οι οποίες εξάγονται από τα σκελετικά δεδομένα που παρέχονται από τον αισθητήρα Kinect. Το υπό μελέτη δωμάτιο που θεωρούμε ως περίπτωση χρήσης αυτής της εργασίας αποτελεί μια αίθουσα συνεδριάσεων που εδράζεται στο Ε.Κ.Ε.Φ.Ε. Δημόκριτος. Στο δωμάτιο αυτό έχει εγκατασταθεί ένα σύνολο αυτοματισμών στα πλαίσια του έργου SYNAISTHISI<sup>1</sup>, τους οποίους θα αξιοποιήσουμε για την υλοποίηση της φυσικής διεπαφής χρήστη.

Ακολουθεί μια περιγραφή της δομής του κειμένου. Στο Κεφάλαιο 2 θα κάνουμε μια ιστορική αναδρομή στις καταβολές του Διαδικτύου των Αντικειμένων και θα εισάγουμε την πλατφόρμα SYNAISTHISI στην οποία έχουν ενσωματωθεί οι αλγόριθμοι που περιγράφονται σε αυτή την εργασία. Στο Κεφάλαιο 3 θα παρουσιάσουμε το μοντέλο εκδότη/συνδρομητή για την επικοινωνία μηχανής με μηχανή και θα παρουσιάσουμε τις δύο κυρίαρχες εναλλακτικές που εξετάσαμε για την υλοποίηση μιας υποδομής επικοινωνίας μεταξύ μηχανών, την πλατφόρμα ROS (Robot Operating System) και το πρωτόκολλο MQTT. Στο Κεφάλαιο 4 θα πραγματοποιήσουμε μια ανάλυση σε βάθος για το υλικό

---

<sup>1</sup> <http://iot.synaisthisi.iit.demokritos.gr/> (Last accessed 16/7/2016)

και το λογισμικό του αισθητήρα Microsoft Kinect. Η ανάλυση αυτή περιγράφει τις δυνατότητες του αισθητήρα, τις μεθόδους για την ανακατασκευή της εικόνας βάθους και εξαγωγής σκελετού και το προγραμματιστικό περιβάλλον για την ανάπτυξη εφαρμογών με το Kinect. Στο Κεφάλαιο 5 παρουσιάζουμε μια απλή και αποδοτική μέθοδο ανίχνευσης πόζας με χρήση σκελετικών δεδομένων. Στο Κεφάλαιο 6 παρουσιάζουμε αναλυτικά το σχεδιασμό, την κατασκευή του συνόλου δεδομένων και την πειραματική αξιολόγηση του αλγορίθμου αναγνώρισης χειρονομιών. Στο ίδιο Κεφάλαιο γίνεται μια σύντομη περιγραφή ενός συνόλου αλγορίθμων Μηχανικής Μάθησης, τους οποίους χρησιμοποιήσαμε. Τέλος στο Κεφάλαιο 7 παρουσιάζουμε την αρχιτεκτονική και τις λεπτομέρειες υλοποίησης του τελικού συστήματος μέσα από την περίπτωση χρήσης σε μία έξυπνη αίθουσα συνεδριάσεων.

## 1.2 Σχετικές Εργασίες

Η βασική ερευνητική πρόκληση για την υλοποίηση της διεπαφής είναι η εξαγωγή μιας μεθόδου για την αναγνώριση χειρονομιών με βάση τα σκελετικά δεδομένα η οποία θα παρέχει ακριβή αποτελέσματα σε πραγματικό χρόνο. Τα τελευταία χρόνια, πολλές ερευνητικές προσπάθειες έχουν εστιάσει στο πρόβλημα της αναγνώρισης των στάσεων του σώματος και χειρονομιών. Η πλειονότητα αυτών των προσεγγίσεων βασίζεται στη μοντέλοποίηση των σκελετικών αρθρώσεων του ανθρώπου και της κίνησής τους. Θεωρούμε αυτά τα δύο προβλήματα ανεξάρτητα, συνεπώς θα παρουσιάσουμε τις τελευταίες εξελίξεις και για τα δύο πεδία.

Οι Shotton et al. [1] παρουσιάζουν δύο προσεγγίσεις για την εκτίμηση πόζας από μια εικόνα βάθους. Η βασική προσφορά σε αυτές τις προσεγγίσεις είναι η χρήση ενός συνόλου συνθετικών εικόνων υψηλής μεταβλητότητας για την εκπαίδευση των αλγορίθμων. Η πρώτη προσέγγιση χρησιμοποιεί ταξινόμηση των μερών του σώματος σε κάθε pixel για την εντόπιση των σκελετικών αρθρώσεων, ενώ η δεύτερη πραγματοποιεί παλινδρόμηση απευθείας στις αρθρώσεις. Πρέπει να σημειωθεί ότι σε πρότερη εργασία των ίδιων βασίζεται και η μέθοδος εξαγωγής σκελετού που υιοθετεί το Kinect.

Οι Korppala και Saxena [2] προτείνουν μια μέθοδο για τη μοντελοποίηση χωροχρονικών συσχετίσεων ανάμεσα σε ανθρώπινες πόζες και αντικείμενα με τη χρήση Εξαρτημένων Τυχαίων Πεδίων (Conditional Random Fields). Επιλέγουν ανάμεσα σε πολλαπλές πιθανές δομές γράφων που συνιστούν μια δράση προσεγγίζοντας το γράφο με προσθετικά χαρακτηριστικά και χρησιμοποιούν τις προσεγγίσεις σαν βάση για τη σχεδίαση κινήσεων για την εξαγωγή άλλων πιθανών γραφικών δομών.

Οι Zanfir et al. [3] προτείνουν το framework *Moving Pose*. Χρησιμοποιούν έναν προσδιοριστή που βασίζεται σε πληροφορία πόζας και διαφορικές ποσότητες (ταχύτητα και επιτάχυνση) των αρθρώσεων του σκελετού. Αυτά τα χαρακτηριστικά συνδυάζονται με μια παραλλαγή του ταξινομητή KNN που λαμβάνει υπόψη την χρονική τοποθεσία ενός καρέ μέσα σε μια ακολουθία δράσεων.

Οι Yang and Ramanan [4] παρουσιάζουν μια μέθοδο για την ανίχνευση ανθρώπων και την εκτίμηση πόζας. Κάνουν χρήση ενός μοντέλου ανάμειξης που βασίζεται σε ένα σύνολο μη προσανατολισμένων τμημάτων. Αυτό το μοντέλο αναπαριστά τις χωρικές συσχετίσεις μεταξύ των τμημάτων και τις σχέσεις των αμοιβαίων εμφανίσεων ανάμεσα σε αναμίξεις των τμημάτων.

Οι Neverova et al. [5] προτείνουν μια μέθοδο ανίχνευσης και εντοπισμού χειρονομιών που βασίζεται σε τεχνικές Βαθιάς Μάθησης. Αρχικά καταγράφουν τη χωρική πληροφορία σε κάποια χωρική κλίμακα και έπειτα χρησιμοποιούν μια στρατηγική εκπαίδευσης που εκμεταλλεύεται την ακριβή αρχικοποίηση των επιμέρους πηγών δεδομένων και τη σταδιακή συγχώνευση των πηγών από την ισχυρότερη στην ασθενέστερη. Σε επόμενη εργασία τους [6] προτείνουν ένα σύστημα αναγνώρισης χειρονομιών με βάση ετερογενείς πηγές δεδομένων (καρέ βάθους, πόζες και ομιλία). Κάθε χειρονομία αποδομείται σε κινήσεις του σώματος μεγάλης κλίμακας και σε τοπικές κινήσεις. Αυτό εφαρμόζεται και στη χρονική κλίμακα, έτσι ώστε μια χειρονομία να μπορεί να θεωρηθεί ως ένα σύνολο δυναμικών ποζών. Έπειτα εξάγουν χαρακτηριστικά σε μικρά χωροχρονικά τμήματα και εκπαιδεύουν ένα Επαναληπτικό Νευρωνικό Δίκτυο (Recurrent Neural Network) για τη μοντελοποίηση χρονικών εξαρτήσεων μεγάλης κλίμακας, τη συγχώνευση δεδομένων και τελικά την αναγνώριση χειρονομιών.

Οι Bhattacharya et al. [7] παρουσιάζουν μια μέθοδο για την ταξινόμηση και την τμηματοποίηση χειρονομιών. Αυτό το επιτυγχάνουν με την εφαρμογή μεθόδων μηχανικής μάθησης στη ροή δεδομένων του Kinect. Αρχικά αναγνωρίζουν χειρονομίες από μια επεξεργασμένη ροή και κατασκευάζουν ένα σύνολο χειρονομιών. Έπειτα ανιχνεύουν τις χειρονομίες σε μία ροή δεδομένων που δεν έχει υποστεί επεξεργασία. Το τελευταίο βήμα είναι το φιλτράρισμα ακούσιων κινήσεων που δεν είναι τμήματα κάποιας χειρονομίας.

Οι Song et al. [8] προτείνουν μια μέθοδο παρακολούθησης του σώματος και των χεριών με τη χρήση μίας στερεοσκοπικής κάμερας. Ανασυγκροτούν τις πόζες του σώματος στον τρισδιάστατο χώρο και αναζητούν τα χέρια. Έπειτα, ταξινομούν τα χέρια με μια μέθοδο βασισμένη στην εμφάνιση με ένα SVM πολλών κλάσεων. Τα εξαγόμενα χαρακτηριστικά από το σώμα και τα χέρια παρέχονται σαν είσοδοι για την αναγνώριση χειρονομιών. Χρησιμοποιούν ένα κρυφό-δυναμικό Εξαρτημένο Τυχαίο Πεδίο με ένα χρονικό κυλιόμενο παράθυρο για τη συνεχή εκτέλεση της διεργασίας και ενισχύουν τα αποτελέσματα με τη χρήση πολυεπίπεδου φιλτραρίσματος στο επίπεδο της εισόδου και των προβλέψεων.

Οι Benilacqua et al. [9] προτείνουν ένα σύστημα που βασίζεται στα Κρυφά Μαρκοβιανά Μοντέλα για την ανάλυση χειρονομιών σε πραγματικό χρόνο. Το σύστημά τους παρέχει στην έξοδό του συνεχώς παραμέτρους που σχετίζονται με την εξέλιξη μιας χειρονομίας στο χρόνο και την αντίστοιχη πιθανοφάνεια. Αυτές οι παράμετροι υπολογίζονται συγκρίνοντας την εκτελούμενη χειρονομία με ένα σύνολο αποθηκευμένων χειρονομιών “αναφοράς”. Η μέθοδος βασίζεται στην ακριβή μοντελοποίηση πολυδιάστατων χρονικών καμπυλών. Η διαδικασία εκμάθησης απλοποιείται με τη χρήση πρότερης γνώσης, επιτρέποντας στο σύστημα να χρησιμοποιεί μόνο ένα παράδειγμα για κάθε κλάση.

Τέλος, η εργασία των Lai et al. [10] εστιάζει στην αναγνώριση χειρονομιών σε μικρή απόσταση με τη χρήση του αισθητήρα Kinect και προτείνουν δύο μεθόδους, οι οποίες βασίζονται στη χρήση χαρακτηριστικών που εξάγονται σε πραγματικό χρόνο από το σκελετικό μοντέλο του Kinect SDK. Με βάση αυτά τα χαρακτηριστικά εκπαιδεύονται δύο ταξινομητές KNN με διαφορετικές μετρικές, σε διαφορετικούς χώρους. Με αυτό το μοντέλο αναγνωρίζονται 8 διαφορετικές χειρονομίες.



## Κεφάλαιο 2

# Το Διαδίκτυο των Αντικειμένων

Ο όρος “Διαδίκτυο των Αντικειμένων”, ο οποίος στη συνέχεια θα αναφέρεται ως ΔτΑ, αφορά ένα ευρύ φάσμα από έννοιες και τεχνολογίες οι οποίες αποσκοπούν στη γεφύρωση του χάσματος μεταξύ του φυσικού κόσμου και του εικονικού υπολογιστικού περιβάλλοντος. Σε αυτό το κεφάλαιο θα προσπαθήσουμε να δώσουμε έναν περιεκτικό ορισμό για το ΔτΑ με μερικά παραδείγματα χαρακτηριστικών εφαρμογών. Έπειτα θα κάνουμε μια ιστορική αναδρομή από την οποία θα αναδειχθούν τεχνολογίες που διευκολύνουν την υλοποίηση λύσεων ΔτΑ σήμερα. Θα καταλήξουμε με την παρουσίαση της πλατφόρμας SYNAISTHISI, η οποία στοχεύει στην παροχή βασικών λειτουργιών για την ανάπτυξη ΔτΑ εφαρμογών και στην οποία ενσωματώθηκαν οι αλγόριθμοι που αναπτύχθηκαν στο πλαίσιο της παρούσας εργασίας.

## 2.1 Τι είναι το Διαδίκτυο των Αντικειμένων

Ως ΔτΑ περιγράφεται ένα περιβάλλον υπολογισμού στο οποίο υπάρχουν “αντικείμενα” τα οποία διαθέτουν μονάδες υπολογισμού και μοναδικούς προσδιοριστές. Τα αντικείμενα αυτά μπορούν να πραγματοποιούν τοπικά ενέργειες και να μεταδίδουν δεδομένα μέσω δικτύου μεταξύ τους και σε κεντρικούς εξυπηρετητές χωρίς ανθρώπινη παρέμβαση. Τα αντικείμενα μπορεί να είναι ηλεκτρικές συσκευές ευρείας κατανάλωσης ή συσκευές ειδικού σκοπού, των οποίων η λειτουργικότητα και η αυτονομία αυξάνονται με την εισαγωγή προγραμματιζόμενων ευφυϊών στοιχείων και τη σύνδεση στο δίκτυο. Η αυξημένη λειτουργικότητα μπορεί στοχεύει στη συλλογή δεδομένων, τη λήψη αυτόματων αποφάσεων, τον έλεγχο άλλων αντικειμένων στο δίκτυο και την παροχή διεπαφών για απομακρυσμένο έλεγχο αυτών των αντικειμένων. Παραθέτουμε μερικά παραδείγματα τέτοιων συσκευών:

- Ένας θερμοστάτης ο οποίος ρυθμίζει τη θερμοκρασία δωματίου είτε με είσοδο που λαμβάνεται από το χρήστη μέσω δικτύου, είτε συμπηφίζοντας δεδομένα πρόγνωσης καιρού και το ιστορικό προτιμήσεων των χρηστών που βρίσκονται στο δωμάτιο.
- Ένας έξυπνος βηματοδότης ο οποίος συλλέγει στατιστικά και ειδοποιεί το θεράποντα ιατρό του ασθενούς όταν εντοπίζει ανωμαλίες
- Ένας ανιχνευτής μονοξειδίου του άνθρακα σε ένα περιβάλλον εργασίας ο οποίος ανοίγει τον

εξαερισμό και ειδοποιεί τους εργαζόμενους όταν γίνει υπέρβαση των ανεκτών τιμών και συλλέγει στατιστικά για την αξιολόγηση του περιβάλλοντος εργασίας.

- Αισθητήρες και ενεργοποιητές (actuators) τοποθετημένοι σε ένα έξυπνο δίκτυο ενέργειας, οι οποίοι θα επιτρέπουν την αποτελεσματικότερη διαχείριση της παραγόμενης ενέργειας παρέχοντας δεδομένα για τη χρήση του δικτύου στους παρόχους και προτάσεις εξοικονόμησης ενέργειας στους κατ'αναλωτές.

Σε αυτά παραδείγματα βλέπουμε ότι αισθητήρες, υπολογιστικές μονάδες και ενεργοποιητές μπορούν να συνδυαστούν για την κατασκευή εφαρμογών στο διαδίκτυο των αντικειμένων. Οι επιμέρους μονάδες μπορούν να κατηγοριοποιηθούν σε υπηρεσίες, οι οποίες θα αναλαμβάνουν την ανάκτηση των δεδομένων (Sensing υπηρεσίες), την επεξεργασία τους (Processing υπηρεσίες) και την αλληλεπίδραση με το φυσικό κόσμο με βάση τα αποτελέσματα των υπολογισμών (Actuating υπηρεσίες). Οι υπηρεσίες που χωρίζονται στις παραπάνω κατηγορίες θα αναφέρονται συνολικά ως SPA. Οι SPA υπηρεσίες αποτελούν αφηρημένες, επαναχρησιμοποιήσιμες αναπαραστάσεις των φυσικών αντικειμένων που αλληλεπιδρούν με την πλατφόρμα. Η κατηγοριοποίηση αυτή των αντικειμένων σε υπηρεσίες μπορεί να συνδέσει το ΔτΑ με την έννοια Everything as a Service (EaaS) ?? από τον ανερχόμενο τομέα του υπολογιστικού νέφους (cloud computing). Με αυτόν τον τρόπο αντικείμενα του φυσικού κόσμου μπορούν να καταγράφουν δεδομένα ή να εκτελούν δράσεις κατ' απαίτηση σε ένα εννιαίο οικοσύστημα διασυνδεδεμένων αντικειμένων και υπηρεσιών.

Τα παραπάνω παραδείγματα εφαρμογών είναι ενδεικτικά για την πρόσθετη αξία και τις δυνατότητες ανάπτυξης που φέρνει η διασύνδεση έξυπνων αντικειμένων σε υπάρχοντα και νέα προϊόντα. Οι νέες λύσεις θα στοχεύουν στη βελτίωση του βιοτικού επιπέδου, την αυτοματοποίηση διαδικασιών και την αποδοτική διαχείριση πόρων.

## 2.2 Η ιστορία του Διαδικτύου των Αντικειμένων

Σε αυτή την ενότητα θα κάνουμε μια σύντομη ιστορική ανασκόπηση του ΔτΑ. Η διασύνδεση συσκευών στο φυσικό κόσμο δεν είναι μια νέα ιδέα, όμως η σημερινή ανάπτυξη της τεχνολογίας και η ελάττωση του κόστους των υπολογιστικών συστημάτων και της παροχής συνδεσιμότητας στο διαδίκτυο επιτρέπουν την υλοποίηση του σε μεγάλη κλίμακα. Θα ξεκινήσουμε αυτή την ενότητα παραθέτοντας τα λόγια του Nikola Tesla από μια συνέντευξη στο περιοδικό Colliers το 1926.

“When wireless is perfectly applied the whole earth will be converted into a huge brain, which in fact it is, all things being particles of a real and rhythmic whole.....and the instruments through which we shall be able to do this will be amazingly simple compared with our present telephone. A man will be able to carry one in his vest pocket.”



### 2.2.1 1988: Ubiquitous computing

Μια από τις πρώτες ιδέες που περιγράφει με ρεαλιστικό τρόπο την ενσωμάτωση των υπολογιστών στον φυσικό κόσμο είναι το όραμα του πανταχού παρόντος υπολογιστή (Ubiquitous Computing, ubicomp) που εξέφρασε ο Mark Weiser, επικεφαλής επικεφαλής επιστήμονας στο Xerox PARC. Ο όρος ubicomp αναφέρεται στην ενσωμάτωση μικρών, φθηνών και διασυνδεδεμένων υπολογιστικών μονάδων σε όλες τις εκφάνσεις της καθημερινής ζωής με σκοπό την αυτοματοποίηση τετριμμένων διαδικασιών και την μετάδοση πληροφορίας από και προς αυτές τις μονάδες. Αυτές οι μονάδες διαχωρίζονται σύμφωνα με το μέγεθός τους σε tabs, pads και boards

- tabs: Υπολογιστές μονάδες που μπορούν να ενσωματωθούν σε είδη ρουχισμού με μέγεθός μερικά εκατοστά
- pads: Φορητές συσκευές με μέγεθός μερικές δεκάδες εκατοστά
- boards: Διαδραστικές οθόνες μεγάλου μεγέθους (μερικά μέτρα)

Ένα από τα βασικά χαρακτηριστικά του ubicomp είναι ότι οι υπολογιστές πρέπει να λειτουργούν στο παρασκήνιο χωρίς το συνειδητό έλεγχο από ανθρώπους. Ο Mark Weiser κάνει παραλληλισμούς με άλλες τεχνολογίες οι οποίες αρχικά είχαν εξειδικευμένες διεπαφές και υψηλό κόστος αλλά στη συνέχεια ενσωματώθηκαν σε συσκευές και η χρήση τους έγινε απρόσκοπτη. Ένα παράδειγμα μιας τέτοιας τεχνολογίας είναι οι ηλεκτρικοί κινητήρες. Στην αρχή του 20ου αιώνα κάθε εργαστήριο είχε ένα μοναδικό κινητήρα, ο οποίος οδηγούσε τις υπόλοιπες συσκευές που απαιτούσαν μηχανική κίνηση με τη χρήση πολύπλοκων συστημάτων μετασχηματισμού κινητικής ενέργειας. Στη συνέχεια οι ηλεκτρικοί κινητήρες ενσωματώθηκαν σε συσκευές ευρείας κατανάλωσης και πλέον αυτές οι συσκευές περιέχουν έναν ή περισσότερους κινητήρες και δεν απαιτούν από το χρήστη τον έλεγχο, τη συντήρηση ή ακόμα και τη γνώση ύπαρξης του κινητήρα.

Αυτή την πορεία θα έπρεπε να ακολουθήσουν και οι υπολογιστές σύμφωνα με το ubicomp. Ο υπολογιστής αντί για ένα έξυπνο “κουτί” θα πρέπει να είναι ένα υποβοηθήμα για τις διεργασίες που μπορούν να αυτοματοποιηθούν παρέχοντας απλές, φυσικές διεπαφές στους χρήστες αυτών των συστημάτων. Μπορούμε να δούμε ότι πολλές ιδέες από το ubicomp έχουν γίνει ήδη μέρος της καθημερινότητας και ότι το ΔτΑ αποτελεί τη φυσική εξέλιξη και επέκταση αυτών των ιδεών.

### 2.2.2 Από τη δεκαετία του 90 στον 21ο αιώνα

Η πρώτη μη συμβατική συσκευή που συνδέθηκε στο διαδίκτυο αναπτύχθηκε για τους σκοπούς του συνεδρίου Interop και παρουσιάστηκε το 1990. Η συσκευή αυτή ήταν μία τοστιέρα που οι χρήστες μπορούσαν να ενεργοποιήσουν και να απενεργοποιήσουν με τη χρήση του πρωτοκόλλου TCP/IP. Τη συσκευή αυτή ακολούθησαν μια μηχανή αναψυκτικών που αναπτύχθηκε στο Carnegie Mellon, η οποία μετέδιδε μέσω διαδικτύου τη διαθεσιμότητα και τη θερμοκρασία των ροφημάτων

και μια καφετιέρα στο πανεπιστήμιο του Cambridge η οποία χρησιμοποιούσε μια κάμερα για να μεταδίδει τη στάθμη του καφέ. Εκτός από αυτά τα πιλοτικά μοντέλα τη δεκαετία του 90 υπήρξαν και σημαντικές τεχνολογικές εξελίξεις στο χώρο των τηλεπικοινωνιών, συγκεκριμένα αναπτύχθηκαν οι πρώτες βιομηχανικές εφαρμογές για την επικοινωνία μηχανής με μηχανή (M2M) και τυποποιήθηκε το πρωτόκολλο IPv6. Οι λύσεις M2M που αναπτύχθηκαν εκείνη τη δεκαετία ήταν κλειστού λογισμικού και δεν ακολουθούσαν κάποια τυποποίηση, οπότε η χρήση του περιοριζόταν σε ειδικευμένες εφαρμογές.

Το 1999 ήταν ένα σημαντικό έτος για την εξέλιξη του ΔτΑ. Ο πρώτος λόγος είναι ότι εκείνη τη χρονιά καθιερώθηκε για πρώτη φορά ο όρος “Internet of Things”. Ο όρος αυτός χρησιμοποιήθηκε για την παρουσίαση μιας βιομηχανικής εφαρμογής η οποία αφορούσε στη σύνδεση των RFID προσδιοριστών μιας γραμμής παραγωγής στο διαδίκτυο με σκοπό την αυτόματη καταμέτρηση και παρακολούθηση των αγαθών. Επίσης το 1999 αναπτύχθηκε η πρώτη έκδοση του πρωτοκόλλου MQTT, ενός πρωτοκόλλου δημοσίευσης - συνδρομής (publish - subscribe). Το 2003 ακολούθησε η τυποποίηση του ZigBee (IEEE 802.15.4), το οποίο αποσκοπεί στη δημιουργία Ασύρματων Δικτύων Προσωπικού Χώρου (WPAN) με συσκευές χαμηλού ενεργειακού αποτυπώματος.

Η ανάπτυξη και τυποποίηση ανοιχτών πρωτοκόλλων για M2M σε συνδυασμό με τη μείωση του κόστους και του μεγέθους των υπολογιστών προώθησαν την εξάπλωση του ΔτΑ και στον προσανατολισμό της έρευνας και της βιομηχανίας σε αυτή την κατεύθυνση. Η διαφορά αυτής της ενσάρκωσης της ιδέας του ΔτΑ με τις αρχικές ιδέες και των σημερινών λύσεων με τα αρχικά πρωτότυπα είναι ότι η προσέγγιση σήμερα βασίζεται σε πιο ώριμες υποδομές που μπορούν να υποστηρίξουν τεράστιο όγκο συσκευών, σε ανοιχτά πρωτόκολλα που επιτρέπουν την καινοτομία και τη διασύνδεση συστημάτων που έχουν αναπτυχθεί ανεξάρτητα και σε δομικές μονάδες των οποίων το χαμηλό κόστος επιτρέπει την ανάπτυξη προϊόντων ευρείας κατανάλωσης.

## **2.3 Η πλατφόρμα SYNAISTHISI**

Η πλατφόρμα SYNAISTHISI είναι μια ολοκληρωμένη πλατφόρμα ανάπτυξης, η οποία παρέχει τη δυνατότητα σε ένα ετερογενές σύνολο από συσκευές αισθητήρων και ενεργοποιητών να επικοινωνούν πάνω σε ετερογενή δίκτυα. Ο βασικός στόχος είναι η σύνθεση πληροφοριών με αυτόματο τρόπο που θα επιτρέψει την εξόρυξη γνώσης για την ανάπτυξη έξυπνων συστημάτων που θα βελτιώνουν τους χώρους διαβίωσης και εργασίας. Εκτός από την παροχή της υποδομής για την επικοινωνία, η πλατφόρμα SYNAISTHISI παρέχει δυνατότητες για την αυτοματοποιημένη ανακάλυψη, σύνθεση και διαχείριση των μονάδων (υπηρεσιών) που απαρτίζουν μια εφαρμογή. Στη συνέχεια θα περιγράψουμε συνοπτικά τα βασικά αρχιτεκτονικά μοτίβα της πλατφόρμας SYNAISTHISI.

### **2.3.1 Αρχιτεκτονική προσανατολισμένη σε υπηρεσίες**

Σύμφωνα με το σώμα προτυποποίησης Open Group, παραθέτουμε τον ακόλουθο ορισμό

**Ορισμός 1:** Η Αρχιτεκτονική Προσανατολισμένη σε Υπηρεσίες (SOA) είναι ένα αρχιτεκτονικό μοτίβο που επιτρέπει την ανάπτυξη συστημάτων στη βάση υπηρεσιών και των αποτελεσμάτων τους. Μια υπηρεσία:

- Είναι μια λογική αναπαράσταση μιας επαναλαμβανόμενης διαδικασίας με σαφώς ορισμένο αποτέλεσμα
- Είναι αυτοτελής
- Μπορεί να παράγεται ως σύνθεση άλλων υπηρεσιών
- Λειτουργεί σαν “μαύρο κουτί” για τις οντότητες που την καταναλώνουν

Όπως φαίνεται στον Ορισμό 1, οι υπηρεσίες είναι αυτοτελείς, χαλαρά συνδεδεμένες (loosely coupled) λειτουργικές μονάδες οι οποίες εκτελούν μια εργασία. Σε μία SOA οι διεπαφές των υπηρεσιών περιγράφονται με χρήση μεταδεδομένων σε κάποια γλώσσα περιγραφής. Τα μεταδεδομένα αυτά μπορούν να χρησιμοποιηθούν από συστήματα λογισμικού, ώστε η διαχείριση, ανακάλυψη και ένταξη των υπηρεσιών να γίνεται με αυτόματο τρόπο, ενώ διατηρείται η συνοχή του συστήματος. Οι υπηρεσίες συνήθως επικοινωνούν μεταξύ τους μέσω δικτύου.

Στα πλαίσια της πλατφόρμας SYNAISTHISI η υπηρεσία είναι η κεντρική έννοια. Μια υπηρεσία μπορεί να είναι κάθε αντικείμενο, συσκευή, άνθρωπος ή δικτυακή υπηρεσία και κατηγοριοποιείται σύμφωνα με το μοντέλο SPA. Εκτός από το λογικό διαχωρισμό των υπηρεσιών, η SYNAISTHISI επιτρέπει και την αυτόματη σύνθεση των υπηρεσιών με τη χρήση προδιαγραφών που παρέχονται από το χρήστη σε μορφή JSON. Σε αυτήν την προδιαγραφή ορίζεται ο τύπος ενός ή περισσότερων υπηρεσιών, η κατηγορία του φυσικού αντικειμένου στο οποίο αντιστοιχούν και ο γράφος διασύνδεσης μεταξύ τους. Δύο υπηρεσίες διασυνδέονται με την προώθηση της εξόδου της πρώτης στην είσοδο της δεύτερης. Ακολουθεί στον Κώδικα 2.1 ένα παράδειγμα εφαρμογής για τη χρήση χειρονομιών για τον έλεγχο ενός κλιματιστικού. Σε αυτή την εφαρμογή μια υπηρεσία τύπου S εξάγει τη σκελετική πληροφορία. Η πληροφορία εισάγεται σε μια υπηρεσία τύπου P, που αναγνωρίζει τη χειρονομία. Έπειτα η αναγνωρισμένη χειρονομία εισάγεται σε μια δεύτερη υπηρεσία τύπου P, που μεταφράζει τη χειρονομία στην επιθυμητή δράση η οποία μεταβιβάζεται σε μια υπηρεσία τύπου A που εκτελεί αυτή τη δράση. Ο γράφος της εφαρμογής είναι ο  $S1 \rightarrow P1 \rightarrow P2 \rightarrow A1$ .

### 2.3.2 Μεσισμικό προσανατολισμένο σε μηνύματα

Το δεύτερο αρχιτεκτονικό μοτίβο στο οποίο βασίζεται η πλατφόρμα SYNAISTHISI είναι το Μεσισμικό Προσανατολισμένο σε Μηνύματα (message oriented middleware, MOM). Το MOM είναι μια υποδομή η οποία επιτρέπει την ανταλλαγή μηνυμάτων σε ετερογενή κατανεμημένα συστήματα. Η βασική λειτουργία του MOM είναι η απλοποίηση της επικοινωνίας εφαρμογών που χρησιμοποιούν διαφορετικά δικτυακά πρωτόκολλα, λειτουργικά συστήματα με την εισαγωγή ενός επιπλέον επιπέδου ανάμεσα στη δικτυακή επικοινωνία και στην εφαρμογή. Αυτό το επίπεδο αφαίρεσης επιτρέπει στον

Κώδικας 2.1: Παράδειγμα εφαρμογής με σύνθεση υπηρεσιών στην πλατφόρμα SYNAISTHISI

```
{
  "S1": {
    "SERVICE TYPE": "S",
    "DEVICE CATEGORY": "FEATURE_EXTRACTION"
  },
  "P1": {
    "SERVICE TYPE": "P",
    "DEVICE CATEGORY": "GESTURE_RECOGNITION"
  },
  "P2": {
    "SERVICE TYPE": "P",
    "DEVICE CATEGORY": "GESTURE_TO_ACTION"
  },
  "A1": {
    "SERVICE TYPE": "A",
    "DEVICE CATEGORY": "AC_CONTROLLER"
  },
  "PIPING": { "S1":["P1"], "P1":["P2"], "P2":["A1"]}
}
```

προγραμματιστή τη δυνατότητα να χειρίζεται την επικοινωνία των μονάδων του συστήματος με διαφανή και ομοιόμορφο τρόπο. Η επικοινωνία σε ένα MOM είναι ασύγχρονη και βασίζεται συνήθως σε ένα μοντέλο εκδότη/συνδρομητή ή/και σε ουρές μηνυμάτων. Η πλατφόρμα SYNAISTHISI κάνει χρήση του πρωτοκόλλου MQTT για την υλοποίηση αυτού του αρχιτεκτονικού μοτίβου.

### 2.3.3 Η υλοποίηση

Η πλατφόρμα SYNAISTHISI δομείται σε μια διαστρωματωμένη αρχιτεκτονική 5 επιπέδων όπως φαίνεται στο Σχήμα 2.1. Τα επίπεδα αυτά είναι:

**Το επίπεδο των φυσικών αντικειμένων:** Αποτελείται από αντικείμενα που αλληλεπιδρούν και λαμβάνουν μετρήσεις από το φυσικό κόσμο. Αυτά τα αντικείμενα μπορεί να είναι αισθητήρες ή έξυπνες συσκευές στις οποίες παρέχεται η δυνατότητα σύνδεσης στο διαδίκτυο. Το σύστημα μπορεί να λαμβάνει δεδομένα από αυτά τα συστήματα και να τα ελέγχει μέσω προγραμματιστικών διεπαφών.

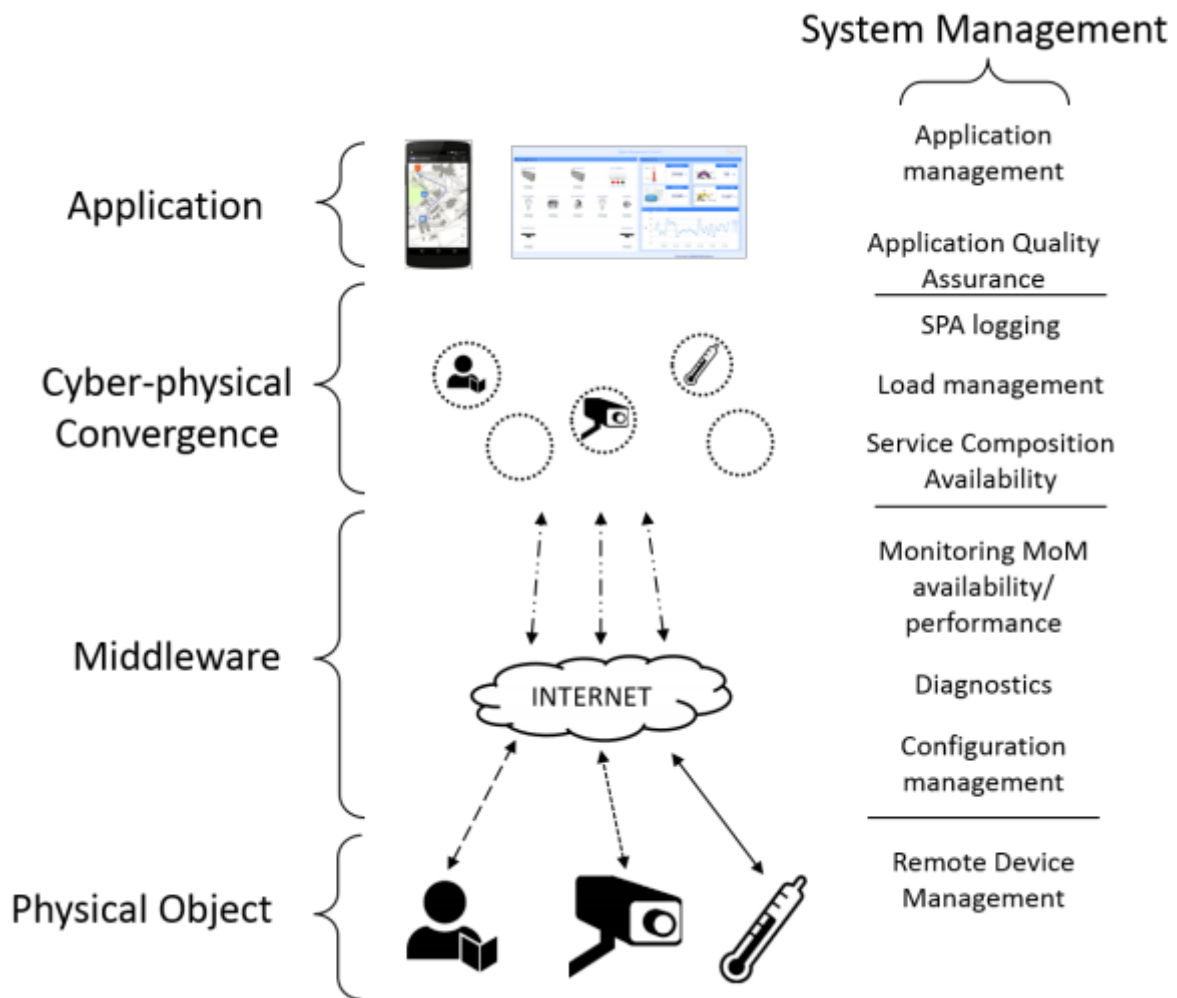
**Το μεσισμικό:** Το μεσισμικό (middleware) του συστήματος υλοποιείται με βάση μια MOM αρχιτεκτονική πάνω σε μία δικτυακή στοιβή TCP/IP. Ακολουθεί το μοντέλο εκδότη/συνδρομητή και υποστηρίζει τόσο σύγχρονη όσο και ασύγχρονη επικοινωνία πάνω σε ανοικτά πρωτόκολλα επικοινωνίας.

**Σύγκλιση ψηφιακών και φυσικών αντικειμένων:** Σε αυτό το επίπεδο τα φυσικά αντικείμενα εικονοποιούνται σε μία δεξαμενή υπηρεσιών. Μια υπηρεσία κατηγοριοποιείται σε έναν από τους τρεις τύπους (SPA), είναι επαναχρησιμοποιήσιμη και επιτρέπει στον τελικό χρήστη να αλληλεπιδρά με τα φυσικά αντικείμενα με έναν ευέλικτο τρόπο. Σε αυτό το επίπεδο υλοποιείται και η μονάδα Διαχείρισης Υπηρεσιών, η οποία προσφέρει δυνατότητες για την αυτόματη διαχείριση του κύκλου ζωής των υπηρεσιών αλλά και για την απευθείας διαχείριση των υπηρεσιών από τον χρήστη μέσω γραφικών και προγραμματιστικών διεπαφών.

**Το επίπεδο εφαρμογών:** Οι εφαρμογές υλοποιούνται στο υψηλότερο επίπεδο της πλατφόρμας SYN-AISTHISI και πραγματοποιούν σύνθετες διεργασίες που έχουν αξία για τον τελικό χρήστη. Οι εφαρμογές μπορούν να υλοποιηθούν σαν σύνθεση υπηρεσιών και αλληλεπιδρούν με τα αντικείμενα του φυσικού κόσμου μέσω αυτών.

**Επίπεδο διαχείρισης:** Αυτό το επίπεδο παρέχει στους διαχειριστές του συστήματος δυνατότητες ρύθμισης και διαχείρισης των εφαρμογών, υπηρεσιών και συσκευών. Επίσης καταγράφονται διαγνωστικές πληροφορίες και παρέχεται η δυνατότητα βελτιστοποίησης της επίδοσης και εξισορρόπησης του συνολικού φορτίου.

Εκτός από τα παραπάνω, σε κάθε επίπεδο της αρχιτεκτονικής παρέχονται δυνατότητες μόνιμης αποθήκευσης. Τα δεδομένα μπορούν να αποθηκευτούν σε βάσεις δεδομένων και κατανεμημένους αποθηκευτικούς χώρους υψηλής διαθεσιμότητας (high availability) με δυνατότητα δημιουργίας αντιγράφων (data replication). Τέλος μέσω της μονάδας ασφαλείας είναι δυνατή η πιστοποίηση και η διαχείριση των δικαιωμάτων των χρηστών και η επικύρωση, ο καθαρισμός και η ανωνυμοποίηση των δεδομένων.



**Σχήμα 2.1:** Η διαστρωματώμενη αρχιτεκτονική της πλατφόρμας SYNAISTHISI

## Κεφάλαιο 3

# M2M: Επικοινωνία Μηχανής με Μηχανή

Στο προηγούμενο κεφάλαιο πραγματοποιήσαμε μια ανάλυση των βασικών τεχνολογιών και αρχιτεκτονικών για την ανάπτυξη του ΔτΑ. Σε αυτό το κεφάλαιο θα εστιάσουμε σε μια από τις βασικές τεχνολογίες που υποστηρίζουν το ΔτΑ, της επικοινωνίας Μηχανής με Μηχανή (M2M). Σε αυτή την εργασία γίνεται εκτεταμένη χρήση αυτού του μοντέλου επικοινωνίας. Αρχικά θα παρουσιάσουμε το μοντέλο επικοινωνίας εκδότη/συνδρομητή το οποίο έχει κεντρικό ρόλο στην M2M επικοινωνία και έπειτα θα εξετάσουμε δυο βασικές τεχνολογίες που παρέχουν δυνατότητες M2M οι οποίες εξετάστηκαν για χρήση ως υποδομή επικοινωνίας (communication infrastructure), το Ρομποτικό Λειτουργικό Σύστημα (ROS) και το πρωτόκολλο MQTT.

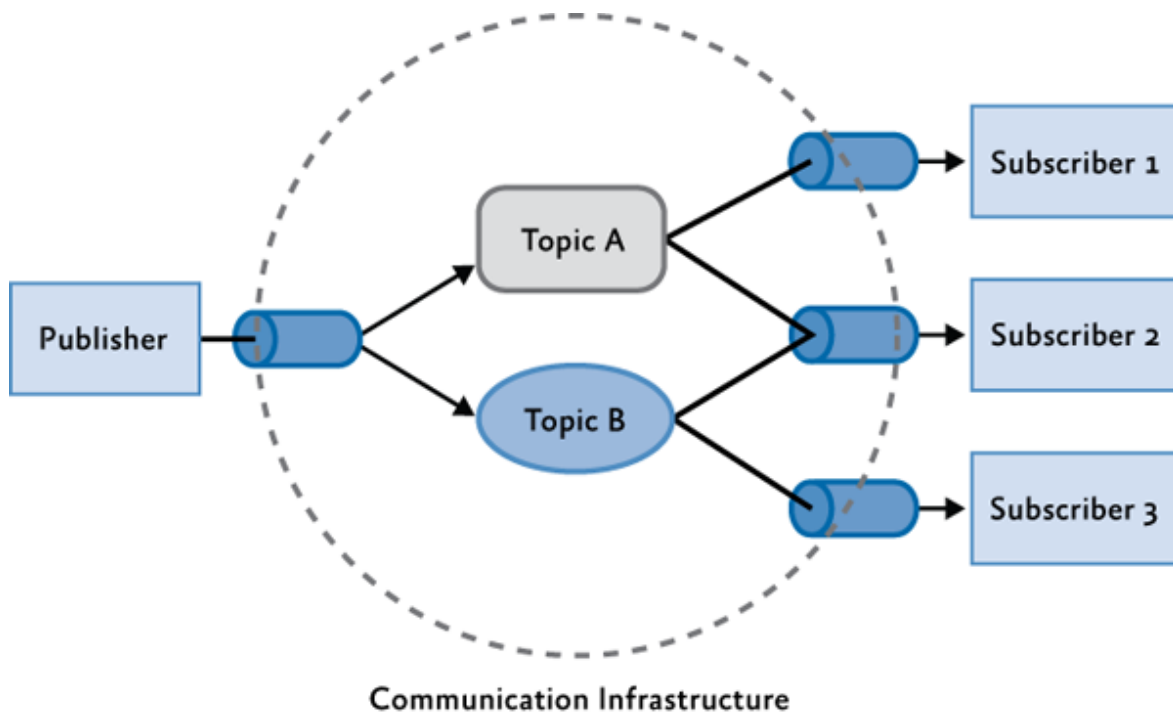
### 3.1 Πρωτόκολλα M2M: Το μοντέλο εκδότη/συνδρομητή

Όπως έχει αναφερθεί ήδη στα βασικά χαρακτηριστικά των εφαρμογών ΔτΑ συμπεριλαμβάνεται και η δυνατότητα των αντικειμένων να επικοινωνούν με άλλα αντικείμενα για την αποστολή εντολών και δεδομένων. Για αυτό το σκοπό έχουν παρουσιαστεί στο παρελθόν ένα πλήθος πρωτοκόλλων και έχει προταθεί μια πληθώρα αρχιτεκτονικών [11]. Σε αυτό το κεφάλαιο θα εστιάσουμε σε μια από τις κυριότερες αρχιτεκτονικές που χρησιμοποιούνται για την επίλυση του πρόβληματος της M2M επικοινωνίας, το μοντέλο εκδότη/συνδρομητή (publisher/subscriber, ή pub/sub).

Το μοντέλο pub/sub είναι μια εναλλακτική για το κλασσικό μοντέλο επικοινωνίας πελάτη/εξυπηρετητή και στοχεύει στο διαχωρισμό της οντότητας εκδότη που στέλνει ένα μήνυμα από την οντότητα συνδρομητή που λαμβάνει το μήνυμα. Για τα πρωτόκολλα pub/sub τόσο οι εκδότες όσο και οι συνδρομητές είναι πελάτες οι οποίοι επικοινωνούν με ένα κεντρικό διακομιστή (broker).

Η βασική λειτουργία του broker είναι η προώθηση των μηνυμάτων από τους εκδότες στους συνδρομητές. Πριν από την προώθηση των μηνυμάτων όμως είναι απαραίτητη μια διαδικασία φιλτραρίσματος, ώστε κάθε συνδρομητής να λάβει μόνο τα μηνύματα για τα οποία έχει εκδηλώσει ενδιαφέρον. Το φιλτράρισμα των μηνυμάτων επιτυγχάνεται είτε με έναν από τους παρακάτω τρόπους είτε με συνδυασμούς αυτών.

**Φιλτράρισμα με βάση το θέμα** Ο broker μπορεί να διαθέτει κάποια θέματα, τα οποία λειτουργούν



**Σχήμα 3.1:** Η αρχιτεκτονική ενός pub/sub συστήματος με θέματα. Ο Publisher δημοσιεύει μηνύματα στα Topics A, B. Τα μηνύματα που δημοσιεύονται στο Topic A λαμβάνονται από τους Subscribers 1, 2 ενώ τα μηνύματα του Topic B λαμβάνονται από τους Subscribers 2, 3.

ως endpoints για τους εκδότες και τους συνδρομητές. Ο όρος endpoint χρησιμοποιείται εδώ με τον τρόπο που ορίζονται τα endpoints στην επικοινωνία μέσω REST, δηλαδή μια συμβολοσειρά που αντιστοιχεί σε μια ιεραρχική οργάνωση των εκτεθειμένων πόρων και λειτουργιών του API. Εάν ένας συνδρομητής εγγραφεί σε ένα θέμα μπορεί να λάβει όλα τα μηνύματα που δημοσιεύονται σε αυτό το θέμα από έναν ή περισσότερους εκδότες. Στο σχήμα 3.1 φαίνεται εποπτικά πώς επικοινωνεί ένας Publisher με τρεις Subscribers μέσω θεμάτων.

**Φιλτράρισμα με βάση το περιεχόμενο** Οι συνδρομητές μπορούν υποβάλλουν στον broker περιορισμούς που πρέπει να ικανοποιούνται ή μοτίβα που πρέπει να περιέχονται στα πεδία ενός μηνύματος. Ένας συνδρομητής θα λάβει ένα μήνυμα, αν αυτό ικανοποιεί όλους τους περιορισμούς που έχει θέσει.

**Φιλτράρισμα με βάση τον τύπο** Στην αντικειμενοστραφή λογική είναι συνήθης πρακτική το φιλτράρισμα με βάση τον τύπο/κλάση ενός γεγονότος (μηνύματος). Αντίστοιχα ένας συνδρομητής μπορεί να επιλέξει να λαμβάνει μόνο τα μηνύματα που έχουν συγκεκριμένου τύπου (ή και υποτύπους αυτού).

Πλεονεκτήματα του pub/sub μοντέλου αποτελούν:

**Ο διαχωρισμός (decoupling) των οντοτήτων που επικοινωνούν** Η χρήση του broker απορρέει στο διαχωρισμό του εκδότη από τον συνδρομητή, καθώς κανένα από τα δύο μέρη της επικοινωνίας δε



χρειάζεται να έχει γνώση για την ύπαρξη του άλλου (πχ γνώση για τη διεύθυνση IP και την πόρτα) και όλη η επικοινωνία εκτελείται ασύγχρονα.

**Η κλιμάκωση (scalability)** Τα πρωτοκόλλα pub/sub μπορούν να έχουν μεγάλη κλιμάκωση (scalability), υποστηρίζοντας πολλές συνδέσεις μέσω

- κατανεμημένων και παράλληλων υλοποιήσεων των κεντρικών διακομιστών
- τη χρήση μηχανισμών ενδιάμεσης αποθήκευσης
- ισοσταθμιστών φορτίου

Όμως το pub/sub δεν αποτελεί καθολική λύση επικοινωνίας, καθώς έχει και μειονεκτήματα όπως:

**Σημασιολογική σύζευξη** Ο τύπος και η δομή των μηνυμάτων σε ένα σύστημα pub/sub πρέπει να είναι γνωστά εξ αρχής. Αυτό μπορεί να οδηγήσει σε ένα δύσκαμπτο σύστημα όπου είναι δύσκολο το refactoring, καθώς για την αλλαγή των πεδίων ενός μηνύματος πρέπει να γίνουν αλλαγές σε όλους τους συνδρομητές που λαμβάνουν αυτό το μήνυμα. Πρέπει να σημειωθεί όμως ότι αυτό το πρόβλημα δεν εντοπίζεται μόνο στην pub/sub λογική, αλλά σε μοντέλα πελάτη/εξυπηρετητή, όπως για παράδειγμα την επικοινωνία μέσω REST και JSON. Μια καλή πρακτική για την αντιμετώπιση αυτού του περιορισμού είναι η χρήση εκδόσεων (versioning) για τους τύπους μηνυμάτων, ώστε να είναι δυνατή η σταδιακή αναβάθμιση όλων των συνιστωσών του συστήματος και η διατήρηση της συμβατότητας με πρότερες εκδόσεις.

**Παράδοση μηνυμάτων** Λόγω της ασύγχρονης φύσης της pub/sub επικοινωνίας είναι δύσκολη η παροχή ισχυρών εγγυήσεων στους εκδότες για την παράδοση των μηνυμάτων και στους συνδρομητές για τη λήψη των μηνυμάτων από ένα συγκεκριμένο εκδότη. Αυτό το πρόβλημα δεν εντοπίζεται στη λογική πελάτη/εξυπηρετητή καθώς εκεί η επικοινωνία είναι σύγχρονη. Για την άμβλυνση αυτού του προβλήματος σε κάποιες υλοποιήσεις υπάρχουν διαβαθμίσεις ποιότητας εξυπηρέτησης (QoS) σε κάποιους τύπους μηνυμάτων και μηνύματα διαθήκης (last will) όπου ο broker ενημερώνει τους υπόλοιπους κόμβους του συστήματος όταν ένας κόμβος φαίνεται να έχει αποσυνδεθεί.

## 3.2 Η πλατφόρμα ROS

Η πλατφόρμα ROS (Robot Operating System) [12] είναι μια συλλογή από βιβλιοθήκες και frameworks για την ανάπτυξη λογισμικού για ρομποτικά συστήματα. Το ROS λειτουργεί ως ένα επίπεδο αφαίρεσης σε ένα περιβάλλον όπου το υλικό συνίσταται από ετερογενείς υπολογιστικές πλατφόρμες. Αυτό το επίπεδο αφαίρεσης έχει το χαρακτήρα λειτουργικού συστήματος, καθώς παρέχει αφαίρεση του υλικού, έλεγχο συσκευών χαμηλού επιπέδου, μηχανισμό ανταλλαγής μηνυμάτων μεταξύ διεργασιών, διαχείριση πακέτων και υλοποιήσεις συχνά χρησιμοποιούμενων λειτουργιών. Ακολούθως παραθέτουμε τις βασικές αρχές σχεδιασμού του ROS:

**Peer to peer (P2P) επικοινωνία** Τα συστήματα που βασίζονται στο ROS αποτελούνται από διεργασίες που τρέχουν σε ένα σύνολο υπολογιστικών μονάδων, οι οποίες συνδέονται μεταξύ τους σε ένα P2P δίκτυο. Αυτό οδηγεί σε μείωση των πακέτων που στέλνονται πάνω στο δίκτυο. Σε ρομποτικές εφαρμογές μεγάλης κλίμακας η μειωμένη κίνηση αποτελεί πλεονέκτημα και ενισχύει την απόδοση των συστημάτων, καθώς το δίκτυο στην γενική περίπτωση είναι ασύρματο και αργό.

**Πολυγλωσσία** Το ROS έχει σχεδιαστεί ώστε να είναι ανεξάρτητο από τη γλώσσα προγραμματισμού. Αυτό επιτυγχάνεται με τη χρήση μιας γλώσσας ορισμού διεπαφών (Interface Definition Language, IDL) στην οποία προσδιορίζονται οι τύποι και το περιεχόμενο των μηνυμάτων που ανταλλάσσονται ανάμεσα στις διεργασίες. Με βάση τα μηνύματα που ορίζονται στην IDL παράγονται αυτόματα από το ROS αντικείμενα στη γλώσσα προγραμματισμού που επιλέγει ο χρήστης τα οποία μπορούν να χρησιμοποιηθούν για την ανάπτυξη της εφαρμογής. Η P2P σύνδεση και η ρύθμιση των κόμβων πραγματοποιείται με χρήση XML-RPC, για το οποίο υπάρχουν υλοποιήσεις στις περισσότερες γλώσσες. Το ROS είναι συμβατό με τις C++, Lisp, Octave και Python.

**Δομοστοιχειωτή Αρχιτεκτονική Μικρού Πυρήνα (modular microkernel architecture)** Η αρχιτεκτονική αυτή στοχεύει στην υλοποίηση των ελάχιστων απαιτούμενων λειτουργιών στον πυρήνα του ROS και τη χρήση εξωτερικών υπηρεσιών και εργαλείων της λειτουργικότητας που απαιτούν οι συνιστώσες του ROS. Για τον περαιτέρω διαχωρισμό της λειτουργικότητας οι αλγόριθμοι και οδηγοί που μπορούν να επαναχρησιμοποιηθούν σε διαφορετικές εφαρμογές υλοποιούνται ως αυτοτελείς βιβλιοθήκες χωρίς εξαρτήσεις από το ROS μεταγλωττίζονται και συνδέονται ως ανεξάρτητες μονάδες από το ROS build system. Με βάση τις παραπάνω αρχές η λειτουργικότητα του ROS εντοπίζεται σε μικρές μονάδες, οι οποίες είναι εύκολα διαχειρίσιμες και ενημερώσιμες.

Στο ROS όλες οι εφαρμογές μπορούν να αναπαρασταθούν σαν ένας γράφος που αποτελείται από μονάδες λογισμικού που επικοινωνούν μεταξύ τους βασικές οντότητες είναι οι *κόμβοι (nodes)*, τα *μηνύματα (messages)*, τα *θέματα (topics)* και οι *υπηρεσίες (services)*.

Οι *κόμβοι* είναι οι βασικές μονάδες υπολογισμού στο ROS και αποτελούν τυπικές διεργασίες σε συστήματα βασισμένα στο UNIX που εκτελούνται σε ένα ετερογενές υπολογιστικό περιβάλλον. Ένας κόμβος επικοινωνεί με τους υπόλοιπους κόμβους ανταλλάσσοντας *μηνύματα*. Τα μηνύματα είναι δομές που κατασκευάζονται σε ένα σύστημα αυστηρών τύπων. Ένα μήνυμα κατασκευάζεται από πρωταρχικούς τύπους (integer, float, boolean κ.ο.κ.), από άλλα μηνύματα (σύνθεση) και από πίνακες πρωταρχικών τύπων και μηνυμάτων χρησιμοποιώντας την IDL. Ακολουθεί στον Κώδικα 3.1 ο ορισμός ενός μηνύματος που περιγράφει ένα PointCloud<sup>1</sup> από το πακέτο *sensor\_msgs* σε IDL. Το μήνυμα αυτό περιέχει πρωταρχικούς τύπους, σύνθεση μηνυμάτων και πίνακες.

Για την υλοποίηση ενός συστήματος περάσματος μηνυμάτων (message passing) από πολλούς κόμβους σε πολλούς κόμβους (many to many communication) γίνεται χρήση μιας αρχιτεκτονικής εκδότη/συνδρομητή (pub/sub). Ένας ή περισσότεροι κόμβοι εκδότες μπορούν να δημοσιεύσουν ένα μήνυμα σε ένα *θέμα*. Έπειτα οι εγγεγραμμένοι κόμβοι συνδρομητές σε αυτό το θέμα θα ειδοποιη-

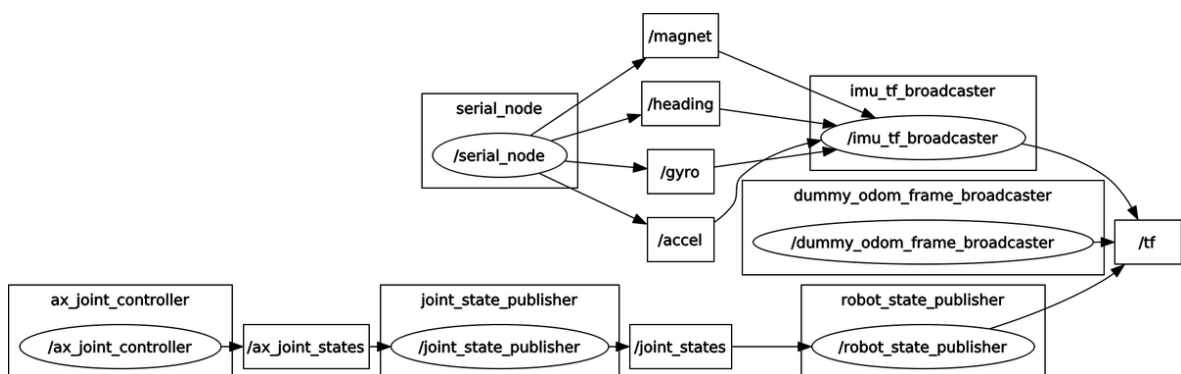
---

<sup>1</sup> Τα PointClouds είναι σύνολα από σημεία σε κάποιο σύστημα συντεταγμένων. Χρησιμοποιούνται από τη βιβλιοθήκη PCL για την αναπαράσταση 3D εικόνων

Κώδικας 3.1: Παράδειγμα μηνύματος στο ROS: Ένα μήνυμα που αναπαριστά ένα PointCloud

```
std_msgs/Header header
uint32 height
uint32 width
sensor_msgs/PointCloud[] fields
bool is_bigendian
uint32 point_step
uint32 row_step
uint8[] data
bool is_dense
```

θούν για τη δημοσίευση και θα λάβουν το περιεχόμενο του μηνύματος. Το θέμα λειτουργεί ως διεπαφή ανάμεσα στους κόμβους εκδότες και συνδρομητές, ώστε κάθε κόμβος να αγνοεί την ύπαρξη των υπολοίπων. Στο Σχήμα 3.2 απεικονίζεται ο γράφος μιας εφαρμογής στο ROS για τη λήψη δεδομένων για ένα robot από ένα σύνολο αισθητήρων. Στο γράφο απεικονίζονται τόσο οι κόμβοι όσο και τα σχετικά θέματα.



Σχήμα 3.2: Παράδειγμα γράφου μιας εφαρμογής συλλογής δεδομένων για την κατάσταση ενός ρομπότ στο ROS

Εκτός από το pub/sub μοντέλο επικοινωνίας, στο ROS υποστηρίζεται και ένα μοντέλο επικοινωνίας που προσεγγίζει αυτό του πελάτη-εξυπηρετητή μέσω των υπηρεσιών. Μια υπηρεσία ορίζεται από μια μοναδική συμβολοσειρά (URI) και ένα ζεύγος μηνυμάτων, την αίτηση (request) και την απάντηση (response). Αντίθετα με το μοντέλο pub/sub μόνο ένας κόμβος μπορεί να δέχεται αιτήσεις σε ένα URI.

### 3.3 Το πρωτόκολλο MQTT

Το MQTT (MQ Telemetry Transport) [13] είναι ένα ελαφρύ, ανοικτό, απλό και εύκολα υλοποιήσιμο δυαδικό pub/sub πρωτόκολλο μεταφοράς μηνυμάτων. Τα χαρακτηριστικά αυτά το κάνουν κατάλληλο για χρήση σε περιορισμένα περιβάλλοντα όπως επικοινωνία M2M και εφαρμογές ΔΤΑ όταν το εύρος ζώνης του δικτύου και το μικρό αποτύπωμα του κώδικα αποτελούν σημαντικούς παράγοντες βελτιστοποίησης. Λόγω του μικρών κεφαλίδων (headers) των πακέτων, ο όγκος των δεδομένων που μεταφέρονται μέσω του MQTT είναι μικρότερος από το πρωτόκολλο HTTP για την ίδια ποσότητα δεδομένων. Επίσης οι πελάτες του MQTT πρωτοκόλλου είναι εξαιρετικά απλοί στην υλοποίηση, οπότε μπορούν να εκτελούνται σε υπολογιστικές συσκευές με περιορισμένους πόρους. Το MQTT απαιτεί όπως και το HTTP την ύπαρξη μιας δικτυακής στοίβας (network stack) όπως η TCP/IP. Παρόλα αυτά υπάρχει και η προδιαγραφή MQTT-SN η οποία μπορεί να υλοποιηθεί πάνω στο Data Link με χρήση ενός δικτύου που βασίζεται στο IEEE 802.15.4, πχ το ZigBee [14].

#### 3.3.1 Περιγραφή του MQTT

Το MQTT έχει όλα τα πλεονεκτήματα και τους περιορισμούς της pub/sub λογικής. Λόγω της αποσύζευξης των κόμβων επικοινωνίας, το μόνο που χρειάζεται να γνωρίζει ένας πελάτης για να επικοινωνήσει με έναν άλλο είναι η διεύθυνση IP και η πόρτα στην οποία ακούει ο broker.

Τα αρχικά MQTT προέρχονται από το όνομα MQ Telemetry Transport. Το πρόθεμα MQ προέρχεται από τη σχέση του με την MQ σειρά προϊόντων της IBM, και έχει κυρίως ιστορική σημασία, καθώς το MQTT δεν υλοποιείται ως μια ουρά μηνυμάτων. Οι κύριες διαφορές ανάμεσα στο MQTT και σε μια παραδοσιακή ουρά μηνυμάτων είναι:

1. Σε μια ουρά μηνυμάτων ένα μήνυμα αποθηκεύεται μέχρι να καταναλωθεί. Το MQTT αντίθετα χρησιμοποιείται κυρίως σε εφαρμογές επικοινωνίας πραγματικού χρόνου, οπότε είναι πιθανό κάποια μηνύματα να μην “καταναλωθούν” από κάποιον πελάτη. Είναι δυνατό όμως να επιτευχθεί παρόμοια συμπεριφορά με τις ουρές μηνυμάτων για συγκεκριμένα μηνύματα θέτοντας το επίπεδο QoS, χρησιμοποιώντας μόνιμες συνεδρίες (persistent sessions) και αποθηκευμένα μηνύματα (retained messages).
2. Στις ουρές μηνυμάτων ένα μήνυμα θα διαβαστεί μόνο από τον πελάτη που θα το τραβήξει από την ουρά. Από την άλλη στο MQTT ένα μήνυμα θα αποσταλλεί σε κάθε πελάτη που είναι εγγεγραμμένος στο θέμα όπου εστάλει το μήνυμα.
3. Οι ουρές πρέπει δημιουργούνται στατικά πριν τη χρήση της. Η ανάλογη δομή στο MQTT, τα θέματα, μπορούν να δημιουργούνται δυναμικά.

Έγινε ήδη ο υπαινιγμός ότι το MQTT χρησιμοποιεί τη λογική του φιλτραρίσματος μηνυμάτων με βάση το θέμα. Κάθε μήνυμα περιέχει ένα θέμα στο οποίο αποστέλλεται και κάθε συνδρομητής εγγράφεται στα θέματα από τα οποία επιθυμεί να λαμβάνει μηνύματα.

### 3.3.2 Λειτουργίες του MQTT

Έχοντας κατηγοριοποιήσει το MQTT σε υψηλό επίπεδο, μπορούμε τώρα να περιγράψουμε τις βασικές λειτουργίες που παρέχει και τον τρόπο χρήσης του. Σε αυτή την ενότητα θα περιγράψουμε τη ροή εργασιών που πρέπει να εκτελέσει ένας πελάτης και τις επιλογές που παρέχονται για τη διασφάλιση της ποιότητας και της βελτιστοποίησης της επικοινωνίας.

#### CONNECT

Ένας πελάτης μπορεί να επιχειρήσει να συνδεθεί στο broker στέλνοντας ένα μήνυμα CONNECT. Σε αυτό το μήνυμα πρέπει να περιέχονται το μοναδικό αναγνωριστικό (id) του πελάτη και μια τιμή για το χρονικό διάστημα keepalive. Με το τελευταίο ο πελάτης δεσμεύεται να ανταλλάσσει με το broker τακτά PING requests κάθε φορά που περνάει το συγκεκριμένο διάστημα, ώστε να διασφαλίζεται ότι είναι δυνατή η επικοινωνία. Εκτός αυτών ένα μήνυμα CONNECT μπορεί να περιέχει και ένα σύνολο από flags που περιέχουν πληροφορίες για τη σύνδεση και τον τρόπο επικοινωνίας, όπως φαίνεται και στο Σχήμα 3.3

Bit	7	6	5	4	3	2	1	0
	User Name Flag	Password Flag	Will Retain	Will QoS		Will Flag	Clean Session	Reserved
byte 8	X	X	X	X	X	X	X	0

**Σχήμα 3.3:** Τα flags που περιέχονται σε ένα μήνυμα CONNECT που ανταλλάσσεται μεταξύ ενός πελάτη και του broker στο πρωτόκολλο MQTT

Ακολουθεί μια σύντομη περιγραφή των πεδίων που απεικονίζονται στο Σχήμα 3.3. Τα επιπλέον δεδομένα που προσδιορίζονται με βάση αυτά τα flags ορίζονται στο περιεχόμενο (payload) του μηνύματος.

**username** Προαιρετικό όνομα χρήστη για την πιστοποίηση του πελάτη

**password** Προαιρετικός κωδικός χρήστη για την πιστοποίηση του πελάτη

**will retain** Προσδιορίζει αν ο broker θα αποθήκευσει ένα last will μήνυμα για την ειδοποίηση των υπόλοιπων κόμβων σε περίπτωση απρόοπτης αποσύνδεσης του πελάτη.

**will QoS** Το QoS επίπεδο του μηνύματος last will

**will Flag** Προσδιορίζει αν η προδιαγραφή για το μήνυμα last will θα είναι παρούσα στο μήνυμα CONNECT

**clean session** Προσδιορίζει αν η συνεδρία θα είναι μόνιμη. Αν το πεδίο cleanSession είναι 0 τότε ο broker θα αποθηκεύει τις εγγραφές και τα μη παραδοθέντα μηνύματα του πελάτη όταν το QoS του

είναι 1 ή 2. Αν το πεδίο cleanSession είναι 1 τότε ο broker δεν αποθηκεύει τίποτα για αυτόν τον πελάτη και διαγράφει κάθε προηγούμενη μόνιμη συνεδρία.

Ο broker απαντά σε ένα μήνυμα CONNECT με ένα μήνυμα CONNACK, το οποίο περιέχει δύο πεδία

**session present** Προσδιορίζει την ύπαρξη μόνιμης συνεδρίας

0	Δεν υπάρχει συνεδρία
1	Υπάρχει συνεδρία

**return code** Προσδιορίζει αν η σύνδεση ήταν επιτυχής. Σε περίπτωση αποτυχημένης σύνδεσης επιστρέφει τον κωδικό σφάλματος

0	Σύνδεση επιτυχής
1	Μη συμβατή έκδοση πρωτοκόλλου
2	Απόρριψη id
3	Μη διαθέσιμος εξυπηρετητής
4	Λάθος πιστοποιητικά
5	Μη εξουσιοδοτημένη σύνδεση

Για τη δημοσίευση ενός μηνύματος χρησιμοποιούνται συνδυασμοί από τα μηνύματα

**PUBLISH** Δημοσίευση ενός νέου μηνύματος

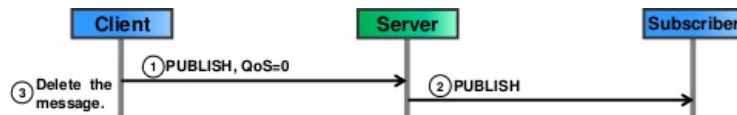
**PUBACK** Ο αποστολέας ενός πακέτου επιπέδου QoS 1 αποθηκεύει το πακέτο μέχρι να λάβει το μήνυμα PUBACK από το δέκτη για την επιβεβαίωση της αποστολής.

**PUBREC** Ο δέκτης ενός πακέτου επιπέδου QoS 2 ενημερώνει ότι αναγνωρίζει το πακέτο με ένα PUBREC μήνυμα

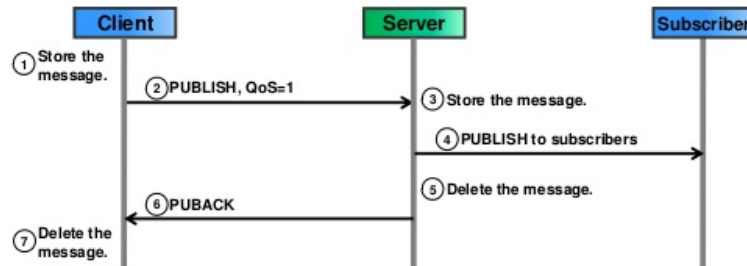
**PUBREL** Ο δέκτης ενός πακέτου επιπέδου QoS 2 αποθηκεύει το πακέτο μέχρις ότου λάβει ένα μήνυμα PUBREL από τον αποστολέα

**PUBCOMP** Ο δέκτης ενός πακέτου επιπέδου QoS 2 διαγράφει το state για αυτό το πακέτο και ολοκληρώνει το handshake με ένα PUBCOMP μήνυμα

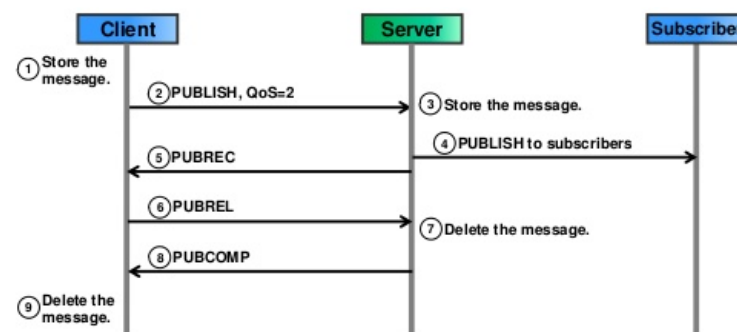
Παραθέτουμε τα ακολουθιακά διαγράμματα UML που περιγράφουν την αλληλεπίδραση εκδότη-broker-συνδρομητή κατά τη δημοσίευση ενός μηνύματος στα Σχήματα 3.4, 3.5 και 3.6 για QoS 0, 1 και 2 αντίστοιχα και την περιγραφή κάθε μηνύματος.



**Σχήμα 3.4:** Ακολουθιακό διάγραμμα UML που περιγράφει τη διαδικασία δημοσίευσης ενός μηνύματος με QoS 0. Όπως φαίνεται ο πελάτης εκδότης δεν έχει κάποια εγγύηση ότι κάποιος συνδρομητής θα λάβει το μήνυμά που έστειλε (fire and forget communication).



**Σχήμα 3.5:** Ακολουθιακό διάγραμμα UML που περιγράφει τη διαδικασία δημοσίευσης ενός μηνύματος με QoS 1. Μέσω του πρωτοκόλλου εγγυάται στον εκδότη ότι το μήνυμά του θα παραδοθεί τουλάχιστον μία φορά στον συνδρομητή με χρήση της χειραψίας 2 βημάτων.



**Σχήμα 3.6:** Ακολουθιακό διάγραμμα UML που περιγράφει τη διαδικασία δημοσίευσης ενός μηνύματος με QoS 2. Μέσω του πρωτοκόλλου εγγυάται στον εκδότη ότι το μήνυμά του θα παραδοθεί ακριβώς μία φορά στον συνδρομητή με χρήση της χειραψίας 4 βημάτων.

## PUBLISH

**packet id** Ένας μοναδικός προσδιοριστής για το μήνυμά. Χρησιμοποιείται όταν το QoS είναι μεγαλύτερο του 0

**topic name** Το όνομα του θέματος όπου θα δημοσιευτεί το μήνυμά

**qos** Το επίπεδο QoS για την παράδοση του συγκεκριμένου μηνύματος

0	Fire and Forget
1	Το μήνυμα θα παραδοθεί τουλάχιστον μια φορά
2	Το μήνυμα θα παραδοθεί ακριβώς μια φορά

**retain flag** Προσδιορίζει αν το μήνυμα θα σωθεί στο broker για να παραδοθεί σε συνδρομητές που θα εγγραφούν μετά την αποστολή του

**payload** Το περιεχόμενο του μηνύματος

**dup flag** Προσδιορίζει αν αυτό το μήνυμα είναι αντίγραφο κάποιου προηγούμενου που αποστέλλεται ξανά λόγω σφάλματος στην αρχική αποστολή

## PUBACK

**packet id** Ο μοναδικός προσδιοριστής του μηνύματος.

## PUBREC

**packet id** Ο μοναδικός προσδιοριστής του μηνύματος.

## PUBREL

**packet id** Ο μοναδικός προσδιοριστής του μηνύματος.

## PUBCOMP

**packet id** Ο μοναδικός προσδιοριστής του μηνύματος.

## SUBSCRIBE

Η εγγραφή σε ένα θέμα επιτυγχάνεται με την ανταλλαγή ενός ζεύγους μηνυμάτων SUBSCRIBE / SUBACK.

**packet id** Ο μοναδικός προσδιοριστής του μηνύματος.

**(topic names, qos)** Μια ακολουθία από ζεύγη που περιέχουν το όνομα θέματος και το επιθυμητό επίπεδο QoS για το συγκεκριμένο θέμα. Ένας συνδρομητής μπορεί με αυτό τον τρόπο να εγγραφεί σε πολλά θέματα στέλλοντας μόνο ένα μήνυμα



## SUBACK

**packet id** Ο μοναδικός προσδιοριστής του μηνύματος.

**return codes** Μια ακολουθία από κωδικούς επιστροφής που δείχνουν αν η εγγραφή ήταν επιτυχής για κάθε θέμα ξεχωριστά. Οι επιτρεπτές τιμές είναι

0	Επιτυχής εγγραφή. Μέγιστο QoS 0
1	Επιτυχής εγγραφή. Μέγιστο QoS 1
2	Επιτυχής εγγραφή. Μέγιστο QoS 2
128	Αποτυχία εγγραφής

Η διαγραφή από ένα θέμα επιτυγχάνεται με την ανταλλαγή ενός ζεύγους μηνυμάτων UNSUBSCRIBE / UNSUBACK.

## UNSUBSCRIBE

**packet id** Ο μοναδικός προσδιοριστής του μηνύματος.

**topic names** Μια ακολουθία από τα ονόματα των θεμάτων από τα οποία επιθυμεί να διαγραφεί ο συνδρομητής. Ένας συνδρομητής μπορεί με αυτό τον τρόπο να διαγραφεί από πολλά θέματα στέλνοντας μόνο ένα μήνυμα

## UNSUBACK

**packet id** Ο μοναδικός προσδιοριστής του μηνύματος.

Ο τελευταίος τύπος μηνυμάτων που θα αναφέρουμε, είναι τα PINGREQ - PINGRESP. Αυτά τα μηνύματα χρησιμοποιούνται ώστε οι κόμβοι να μπορούν να ελέγξουν αν η σύνδεση με το broker είναι δυνατή και ο broker να ενημερώνεται αν ένας κόμβος έχει χαθεί.



## Κεφάλαιο 4

### Ο αισθητήρας Kinect

Σε αυτό το κεφάλαιο θα γίνει μια παρουσίαση των αρχών λειτουργίας και των δυνατοτήτων του αισθητήρα Kinect [15]. Το Kinect αναπτύχθηκε από τη Microsoft ως απάντηση στο Nintendo Wii. Ο βασικός στόχος του είναι να παρέχει μια ανέπαφη φυσική διεπαφή χρήστη. Ο αισθητήρας αρχικά βρήκε εφαρμογή σε διεπαφές παιχνιδιών, ενώ μετέπειτα ακολούθησαν και εφαρμογές όπως το 3D scanning, η εικονική πραγματικότητα και η αυτόματη μετάφραση της νοηματικής γλώσσας. Μέχρι σήμερα έχουν κυκλοφορήσει δύο εκδόσεις του αισθητήρα, το Kinect 360 και το Kinect 2.0. Η εργασία αυτή υλοποιήθηκε με χρήση του Kinect 360, συνεπώς η ανάλυση μας θα εστιάσει σε αυτή την έκδοση. Η ανάλυση θα χωριστεί σε τρία επίπεδα, ξεκινώντας από το υλικό του αισθητήρα, προχωρώντας στην ανάλυση των ανεπεξέργαστων δεδομένων και καταλήγοντας στα δεδομένα που εκτίθενται στον προγραμματιστή από το Kinect SDK.

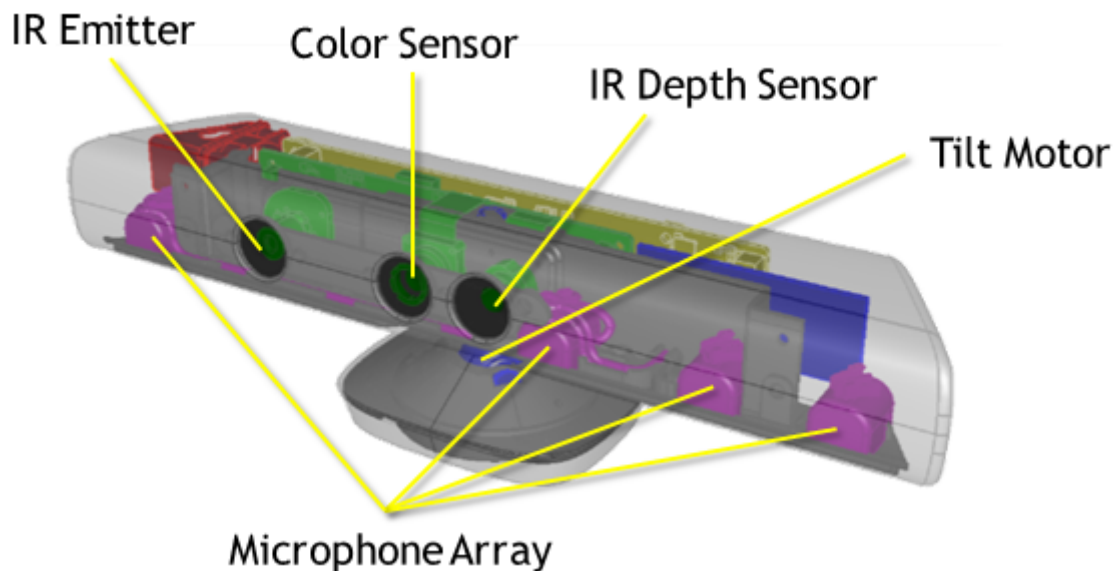
#### 4.1 Το υλικό του Kinect

Η βασική λειτουργία του Kinect είναι η εξαγωγή μιας τρισδιάστατης αναπαράστασης του χώρου η οποία γίνεται δυνατή με τη χρήση ενός εκπόμπου και μιας κάμερας υπερύθρων τοποθετημένων σε μια προκαθορισμένη απόσταση. Τα δύο αυτά εξαρτήματα αποτελούν τον αισθητήρα βάθους. Στον Πίνακα 4.1 παραθέτουμε τις τεχνικές προδιαγραφές του αισθητήρα βάθους όπως αναγράφονται στο εγχειρίδιο χρήσης του αισθητήρα [16]. Πρέπει να σημειωθεί επίσης ότι όπως αποδεικνύεται και από τους Khoshelham et al [17] η διακριτική ικανότητα και το σφάλμα του αισθητήρα αυξάνουν τετραγωνικά με την απόσταση από το χρήστη. Οι ελάχιστες και μέγιστες τιμές που δίνονται αναλογούν στην ελάχιστη και μέγιστη απόσταση λειτουργίας

Χαρακτηριστικά	Τιμές
Απόσταση λειτουργίας	0.5 m – 5 m
Ανάλυση εικόνας βάθους	320 × 240px
Τυχαίο σφάλμα	0.2 cm – 4 cm
Διακριτική ικανότητα	0.2 cm – 7 cm

**Πίνακας 4.1:** Τεχνικές προδιαγραφές του αισθητήρα βάθους στο Kinect 360

Το Kinect περιλαμβάνει δύο ακόμα βασικούς αισθητήρες, μια RGB κάμερα με ανάλυση εικόνας  $640 \times 480px$  σαν εναλλακτική πηγή οπτικών δεδομένων και μια συστοιχία 4 μικροφώνων για χρήση σε αναγνώριση φωνητικών εντολών και προσδιορισμού της θέσης του ομιλητή. Επιπρόσθετα παρέχεται ένας κινητήρας κλίσης με με εύρος από  $-27^\circ$  έως  $27^\circ$ , ο οποίος επιτρέπει την κάθετη σάρωση του χώρου, αυξάνοντας την κατακόρυφη γωνία θέασης και ένα επιταχυνσιόμετρο τριών αξόνων, το οποίο επιτρέπει τον προσδιορισμό του προσανατολισμού του Kinect. Στο Σχήμα 4.1 απεικονίζεται το Kinect 360 και τα βασικά εξαρτήματά από τα οποία αποτελείται.

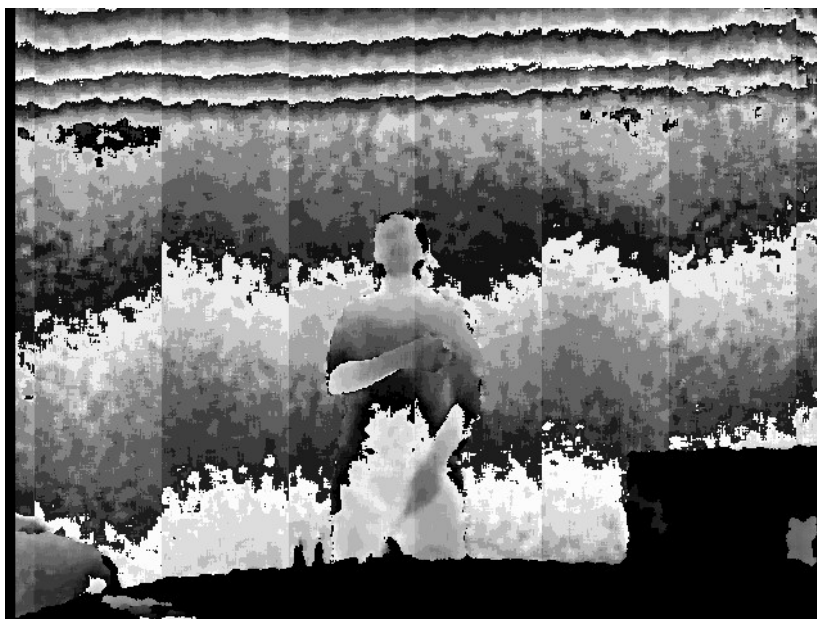


**Σχήμα 4.1:** Το υλικό του Kinect 360: Εκπομπός Υπερύθρων (IR Emitter), Αισθητήρας Χρώματος (Color Sensor), Αισθητήρας Υπερύθρων/Βάθους (IR Depth Sensor), Κινητήρας Κλίσης (Tilt Motor), Πίνακας Μικροφώνων (Microphone Array)

## 4.2 Εξαγωγή της εικόνας βάθους

Σε αυτή την ενότητα θα παρουσιάσουμε συνοπτικά τις μεθόδους και τους αλγορίθμους που χρησιμοποιεί το Kinect 360 για την ανακατασκευή της εικόνας βάθους. Η ανακατασκευή επιτυγχάνεται με τη χρήση της τεχνολογίας PrimeSense [18]. Λόγω της αγοράς των πατέντων που συνοδεύουν την τεχνολογία αυτή από την Apple inc., στη νεότερη έκδοση του Kinect (Kinect v2) το σύστημα της 3D ανακατασκευής έχει επανασχεδιαστεί και περιέχει αλλαγές στη σχεδίαση του υλικού και του λογισμικού. Η ανάλυσή μας εστιάζει στην τεχνολογία PrimeSense καθώς η εργασία αυτή δημιουργήθηκε με χρήση του Kinect 360 και είναι ενδεικτική για τη φιλοσοφία των μεθόδων 3D ανακατασκευής.

Η τεχνολογία PrimeSense αποτελείται από ένα συνδυασμό υλικού και λογισμικού, του οποίου τα βασικά μέρη θα περιγραφούν σε αυτή την ενότητα. Πρέπει να σημειωθεί ότι η παρακάτω ανάλυση βασίζεται στις αιτήσεις χορήγησης ευρεσιτεχνιών της PrimeSense και μπορεί να περιέχει κάποιες ανακρίβειες καθώς δεν είναι διαθέσιμες όλες οι πληροφορίες για το σύστημα.



**Σχήμα 4.2:** Εικόνα βάθους από τον αισθητήρα Kinect σε κλίμακα του γκρι

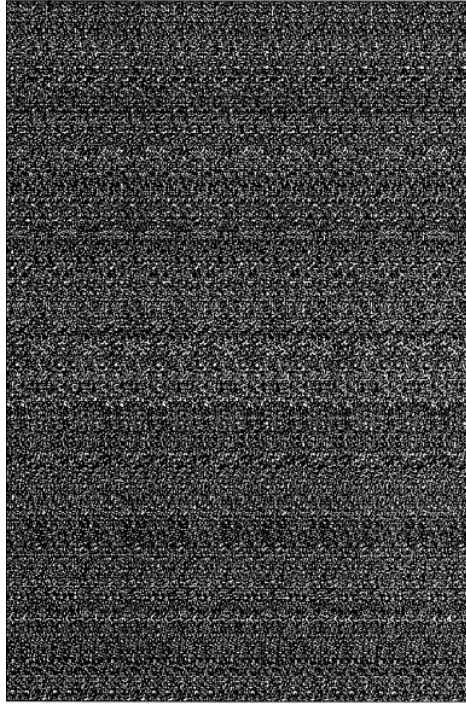
#### 4.2.1 Το υλικό του αισθητήρα

Ο αισθητήρας βάθους υλοποιεί μια παραλλαγή της μεθόδου δομημένου φωτός, η οποία συνίσταται από την προβολή στο χώρο ακτινοβολίας με ένα προκαθορισμένο μοτίβο και την ανίχνευση των παραμορφώσεων που προκαλούνται στο μοτίβο από τα αντικείμενα που βρίσκονται στο χώρο. Αυτό επιτρέπει την εξαγωγή της πληροφορίας του βάθους και άρα την τρισδιάστατη ανακατασκευή/ύψωση του χώρου. Ένα παράδειγμα της ανακατασκευασμένης εικόνας βάθους δίνεται στην Εικόνα 4.2.

Ο εκπομπός του Kinect αποτελείται από μια διαφάνεια στην οποία είναι τυπωμένο ένα μοτίβο κουκκίδων που τοποθετείται μπροστά από μια πηγή υπέρυθρου φωτός και προβάλλεται στο χώρο. Το μοτίβο αυτό πρέπει να είναι ασυσχέτιστο στα εγκάρσια επίπεδα στον άξονα της προβολής όπως περιγράφεται στον Ορισμό 2. Η χρήση ασυσχέτιστων μοτίβων λύνει το πρόβλημα της αναδίπλωσης της πληροφορίας. Αν για παράδειγμα χρησιμοποιούνταν ένα περιοδικό μοτίβο, τότε δεν θα υπήρχε διάκριση ανάμεσα σε σχετικές μετατοπίσεις των κουκκίδων μεγαλύτερων της περιόδου του προβαλλόμενου σήματος, καθώς σε αυτές τις μετατοπίσεις θα μπορούσε να προστεθεί κάποιο ακέραιο πολλαπλάσιο της περιόδου.

**Ορισμός 2:** Σε ένα προβαλλόμενο μοτίβο κουκκίδων οι σχετικές θέσεις των κουκκίδων θεωρούνται ασυσχέτιστες στα επίπεδα κάθετα στον άξονα προβολής όταν η αυτοσυσχέτιση των θέσεων των κουκκίδων ως συνάρτηση της εγκάρσιας μετατόπισης τους είναι:

- αμελητέα για κάθε τιμή της εγκάρσιας μετατόπισης μεγαλύτερη από το μέγεθος μιας κουκκίδας.
- μικρότερη από τη μέγιστη τιμή της εγκάρσιας μετατόπισης η οποία μπορεί να παρατηρηθεί σε όλο το εύρος των τιμών του βάθους που μπορούν να ανιχνευθούν από το σύστημα.



**Σχήμα 4.3:** Παράδειγμα προβαλλόμενου μοτίβου στην τεχνολογία PrimeSense [18]

Τα τυχαία μοτίβα και τα τεχνητά παραγόμενα ψευδοτυχαία μοτίβα είναι ασυσχέτιστα. Επίσης τμηματικά περιοδικά μοτίβα με  $n$ -πλή συμμετρία, όπου  $n = 5$  ή  $n \geq 7$  μπορούν να είναι ασυσχέτιστα. Η χρήση τμηματικά περιοδικών μοτίβων μπορεί να είναι πλεονεκτική καθώς το φάσμα συχνότητας του μοτίβου είναι γνωστό και έχει συγκεκριμένες κορυφές, συνεπώς η φασματική πληροφορία μπορεί να χρησιμοποιηθεί για την αποθρομβοποίηση των συλλεγόμενων εικόνων και την αντιστάθμιση της επίδρασης του περιβάλλοντος φωτός.

Η Εικόνα 4.3 περιέχει ένα ρεαλιστικό μοτίβο που προβάλλεται από το Kinect στο χώρο. Το μοτίβο αυτό απαρτίζεται από παράλληλες ζώνες κουκκίδων που δημιουργούνται από την ίδια ψευδοτυχαία κατανομή. Οι ζώνες εκτείνονται στον κατακόρυφο άξονα και επαναλαμβάνονται στον οριζόντιο. Αυτό αποτελεί ένα υβριδικό μοτίβο, το οποίο περιέχει και τυχαιότητα ώστε να εξασφαλίζεται η έλλειψη συσχέτισης, αλλά και η περιοδικότητα που μειώνει τις απαιτήσεις σε μνήμη και υπολογιστική πολυπλοκότητα.

#### 4.2.2 Ο αλγόριθμος της 3D ανακατασκευής

Ο αλγόριθμος αυτός παρέχεται υλοποιημένος σε ένα ASIC (Application Specific Integrated Circuit) και αποτελεί μια λύση που συμβιβάζει την ακρίβεια του τελικού χάρτη βάθους με τις απαιτήσεις ενός συστήματος πραγματικού χρόνου για χαμηλή αλγοριθμική πολυπλοκότητα και μικρό αποτύπωμα στη μνήμη.

Η μέθοδος εξαγωγής του 3D χάρτη του χώρου βασίζεται στη μέθοδο της τριγωνοποίησης

πάνω στο μοτίβο κουκκίδων που εκπέμπεται από τον PrimeSensor στο χώρο το οποίο στο εξής θα αναφέρεται ως το “μοτίβο αναφοράς”. Οι συντεταγμένες των κουκκίδων του μοτίβου αναφοράς για ένα προκαθορισμένο βάθος βρίσκονται αποθηκευμένες στη μνήμη του PrimeSensor. Το μοτίβο αυτό προβάλλεται στα αντικείμενα του χώρου όπως περιγράψαμε στην προηγούμενη ενότητα και ο δέκτης υπερύθρων αιχμαλωτίζει μια εικόνα από το ανακλώμενο μοτίβο.

Η μέθοδος της τριγωνοποίησης βασίζεται στη σύγκριση της εικόνας αναφοράς με την εικόνα ανάκλασης και στον συμπερασμό της μετατόπισης κάθε σημείου του μοτίβου από την αρχική του θέση. Αυτό οδηγεί στον υπολογισμό της εγκάρσιας μετατόπισης και άρα στον υπολογισμό του βάθους.

Ο υπολογισμός της διαφοράς μεταξύ των δύο εικόνων ανάγεται σε ένα πρόβλημα ταιριάσματος μεταξύ των δύο μοτίβων, δηλαδή την εύρεση για κάθε σημείο στο ανακλώμενο μοτίβο το αντίστοιχο σημείο στο μοτίβο αναφοράς. Οι προσεγγίσεις για την επίλυση αυτού του προβλήματος εκτείνονται από απλή εξαντλητική αναζήτηση, σε επέκταση περιοχών με πρόβλεψη (predictive region growing) μέχρι εξεζητημένους αλγόριθμους μέγιστης πιθανοφάνειας. Η επέκταση περιοχών αποτελεί έναν καλό συμβιβασμό ανάμεσα στην υπολογιστική πολυπλοκότητα και στην ακρίβεια της ανακατασκευής.

Η επέκταση περιοχών βασίζεται στην υπόθεση συνέχειας. Σύμφωνα με την υπόθεση συνέχειας δύο γειτονικά σημεία που ανήκουν στο ίδιο αντικείμενο θα έχουν μικρές διαφορές στην τιμή του βάθους τους, δηλαδή την τιμή της  $Z$  συντεταγμένης τους. Συνεπώς το βάθος ενός σημείου μπορεί να προβλεφθεί από τις τιμές του βάθους των γειτονικών του σημείων. Αν αυτή η πρόβλεψη είναι αρκετά καλή συμπεραίνεται ότι το σημείο ανήκει στην ίδια περιοχή με τα σημεία που χρησιμοποιήθηκαν για την πρόβλεψη. Παραθέτουμε σε ψευδοκώδικα τον Αλγόριθμο 2, όπου περιγράφεται η μέθοδος ταιριάσματος με επέκταση περιοχών, όπως περιγράφεται στις αιτήσεις ευρεσιτεχνίας του PrimeSensor [19].

---

**Algorithm 2** Αλγόριθμος ταιριάσματος σημείων με προβλεπτική επέκταση περιοχών

---

```
1: procedure RegionGrowing(reference, reflected)
2:   for all pixels in reflected do
3:     if pixel.isPartOf(speckle) then
4:       pixel.label  $\leftarrow$  UNKNOWN
5:     else
6:       pixel.label  $\leftarrow$  SHADOW
7:   while percentage(UNKNOWN pixels) > thresholds.UnknownPixels do
8:     P  $\leftarrow$  ChooseRandom(UNKNOWN pixel)
9:     window  $\leftarrow$  RegionAround(P, NxM)  $\triangleright$  Το παράθυρο είναι μια NxM περιοχή γύρω από το
    pixel P
10:    correlation  $\leftarrow$  CalculateCorrelation(window, reference)  $\triangleright$  υπολογισμός της συσχέτισης
    με το μοτίβο αναφοράς
11:    D  $\leftarrow$  CalculateDisparity(P, correlation)  $\triangleright$  ανομοιότητα (disparity) D=(dx,dy)
12:    if correlation > thresholds.correlation then
13:      anchor  $\leftarrow$  P
14:       $\triangleright$  Επέκταση περιοχής γύρω από το pixel υποστήριξης (anchor)
15:      queue  $\leftarrow$  NewFIFOQueue()
16:      anchor.label  $\leftarrow$  ACTIVE
17:      queue.push(anchor)
18:      while not queue.isEmpty() do
19:        P  $\leftarrow$  queue.pop()
20:        neighbors  $\leftarrow$  P + ( $\pm 1, \pm 1$ )  $\triangleright$  Οι γείτονες του P με συνεκτικότητα 4
    (4-connectivity) (πάνω, κάτω, δεξιά, αριστερά)
21:        for neighbor in neighbors do
22:          if neighbor.label == UNKNOWN then
23:            correlation  $\leftarrow$  CalculateCorrelation(neighbor, D + ( $\pm 1, \pm 1$ ))  $\triangleright$  η περιοχή
    αναζήτησης για τη συσχέτιση είναι η D + ( $\pm 1, \pm 1$ )
24:            if correlation > thresholds.correlation then
25:              pixel.label  $\leftarrow$  ACTIVE
26:              anchor  $\leftarrow$  P
27:              queue.push(P)
28:            else
29:              P.label  $\leftarrow$  EDGE
30:              depth  $\leftarrow$  P.depth
```

---



## 4.3 Εξαγωγή των αρθρώσεων του σκελετού

Η ανάκτηση μιας 3D αναπαράστασης του χώρου αποτελεί μια εξαιρετικά χρήσιμη δυνατότητα. Με τη χρήση της εικόνας βάθους οι λύσεις για πολλά δύσκολα προβλήματα της υπολογιστικής όρασης, όπως η αφαίρεση του φόντου της εικόνας γίνονται τετριμμένες. Σε αυτή την ενότητα θα δούμε ότι με τη χρήση της εικόνας βάθους μπορεί να δοθεί μια αποδοτική λύση πραγματικού χρόνου στο δύσκολο πρόβλημα της αυτόματης αναζήτησης των μερών του ανθρώπινου σώματος σε μια εικόνα μέσω της τμηματοποίησης της εικόνας. Η λύση αυτή βασίζεται σε ένα συνδυασμό αλγορίθμων επιβλεπόμενης μηχανικής μάθησης και υπολογιστικής όρασης, οπότε για την ακριβή περιγραφή της χρειάζεται να παρουσιάσουμε την προεπεξεργασία των δεδομένων, τη διαδικασία εκπαίδευσης και τα αποτελέσματα του αλγορίθμου

### 4.3.1 Συλλογή δεδομένων και εξαγωγή χαρακτηριστικών

Το σύνολο δεδομένων το οποίο χρησιμοποιήθηκε από τους Shotton et al. [20] αποτελείται από καρέ εικόνων βάθους από βίντεο στα οποία άνθρωποι εκτελούν διάφορες δραστηριότητες, όπως χορό, οδήγηση, χειρισμό μενού κ.ο.κ. Ο στόχος αυτού του συνόλου δεδομένων είναι να κωδικοποιηθεί μια ευρεία γκάμα από πόζες που λαμβάνει το ανθρώπινο σώμα κατά την εκτέλεση κινήσεων ώστε να παρέχει στο μοντέλο που θα εκπαιδευτεί με βάση αυτό μια βάση ώστε να είναι δυνατή η γενίκευσή του σε πόζες τις οποίες έχει δεν εντοπίσει στο παρελθόν. Το αρχικό σύνολο δεδομένων αποτελείται συνολικά από  $\sim 500K$  καρέ motion capture δεδομένων, τα οποία υποβάλλονται σε μια διαδικασία προεπεξεργασίας πριν την εξαγωγή των χαρακτηριστικών για τον καθαρισμό και την ενίσχυση των δεδομένων.

#### Προεπεξεργασία του συνόλου δεδομένων εκπαίδευσης

Το πρώτο βήμα κατά την προεπεξεργασία των δεδομένων είναι η υποδειγματοληψία των δεδομένων με τη χρήση συσταδοποίησης απομακρυσμένων γειτόνων. Αυτό φιλτράρει τα διαδοχικά καρέ στα οποία οι πόζες απέχουν μεταξύ τους λιγότερο από  $5cm$  και στοχεύει στην αφαίρεση των πλεονάζοντων εικόνων που προσθέτουν μικρή ποσότητα επιπλέον πληροφορίας στο σύνολο δεδομένων. Το αποτέλεσμα είναι ένα σύνολο δεδομένων που αποτελείται από  $100K$  καρέ τα οποία περιέχουν διακριτές πόζες.

Το επόμενο στάδιο της προεπεξεργασίας είναι ο εμπλουτισμός του συνόλου δεδομένων με τη χρήση συνθετικών δεδομένων. Η χρήση εικόνων βάθους κάνει δυνατή την παραγωγή ρεαλιστικών συνθετικών δεδομένων, καθώς οι εικόνες βάθους είναι εικόνες στην κλίμακα του γκρι, και άρα με τη χρήση αυτών αγνοούνται άσχετες πληροφορίες για το χρώμα ή την υφή του δέρματος και των ρούχων και το φόντο της εικόνας. Η σύνθεση των εικόνων πραγματοποιείται με τη χρήση τεχνικών της γραφικής υπολογιστών και τα συνθετικά δεδομένα εξαρτώνται από ένα σετ τυχαίων παραμέτρων. Οι

παράμετροι αυτές αφορούν το σχήμα και το μέγεθος του σώματος, τη θέση της κάμερας, τον προσανατολισμό του σώματος ως προς την κάμερα και το θόρυβο των παραγόμενων καρτέ. Με βάση τις τυχαίες παραμέτρους παράγονται παραλλαγές σε ένα σύνολο 3D πλεγμάτων που έχουν δημιουργηθεί με τη χρήση του λογισμικού Motioncapture της Autodesk. Οι τελικές συνθετικές εικόνες είναι πλήρως επισημασμένες και έτοιμες να συμπεριληφθούν στο σύνολο δεδομένων.

### Ενδιάμεση αναπαράσταση των μερών του σώματος

Κατά την περιγραφή της διαδικασίας κατασκευής του συνόλου δεδομένων αναφέραμε ότι τα δεδομένα είναι επισημασμένα. Ο Shotton et al προτείνουν τη χρήση μιας ενδιάμεσης αναπαράστασης των μερών του σώματος, όπου το σώμα χωρίζεται σε 31 διαφορετικές περιοχές, όπως περιγράφεται στον Πίνακα 4.2. Μια απεικόνιση της ενδιάμεσης αναπαράστασης δίνεται στην Εικόνα 4.4.



**Σχήμα 4.4:** Η ενδιάμεση αναπαράσταση των μερών του σώματος για την ανίχνευση των αρθρώσεων

### Εξαγωγή χαρακτηριστικών και εκπαίδευση των μοντέλων

Για την εκπαίδευση του ταξινομητή εξάγονται απλά χαρακτηριστικά χαμηλού επιπέδου από τις εικόνες βάθους. Τα χαρακτηριστικά αυτά εξάγονται απευθείας από τα pixels της εικόνας χωρίς να χρειάζεται κάποια προεπεξεργασία ή μετασχηματισμός της εικόνας και περιγράφονται τη Σχέση 4.1, όπου  $d_I(\mathbf{x})$  το βάθος στο pixel  $\mathbf{x}$  της εικόνας  $I$ , και  $\mathbf{u}, \mathbf{v}$  παράμετροι που περιγράφουν μετατοπίσεις σε κάποιο άλλο pixel της εικόνας.

$$f_{\theta}(I, \mathbf{x}) = d_I\left(\mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})}\right) - d_I\left(\mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})}\right) \quad (4.1)$$

Σε περίπτωση που κάποιο pixel  $\mathbf{y}$  βρίσκεται στο φόντο ή έξω από τα όρια της εικόνας το βάθος του  $d_I(\mathbf{y})$  λαμβάνει μια μεγάλη θετική τιμή. Τα χαρακτηριστικά αυτά δίνουν μια ασθενή περιγραφή για τα μέρη του σώματος, όμως το υπολογιστικό κόστος για τον υπολογισμό τους είναι αμελητέο και αν εκπαιδευτούν πολλοί διαφορετικοί συνδυασμοί αυτών των χαρακτηριστικών σε ένα τυχαίο δάσος αποφάσεων οι προβλέψεις που προκύπτουν έχουν ικανοποιητική ακρίβεια.

‘L: Left	R: Right
U: Upper	W: Lower
S: single Area	

Το x επισημαίνει ότι η περιοχή αποτελεί διαφορετική κλάση.

Body Part	<i>L</i>	<i>LU</i>	<i>LW</i>	<i>R</i>	<i>RU</i>	<i>RW</i>	<i>S</i>
Head		x	x		x	x	
Neck							x
Shoulder	x			x			
Arm		x	x		x	x	
Elbow	x			x			
Wrist	x			x			
Hand	x			x			
Torso		x	x		x	x	
Leg		x	x		x	x	
Knee	x			x			
Ankle	x			x			
Foot	x			x			

**Πίνακας 4.2:** Τα μέρη του σώματος στην ενδιάμεση αναπαράσταση

Τα χαρακτηριστικά αυτά χρησιμοποιούνται για την εκπαίδευση ενός μοντέλου Random Forests. Ο αλγόριθμος αυτός είναι ένα ensemble μοντέλο από Δέντρα Αποφάσεων και περιγράφεται αναλυτικά σε επόμενο κεφάλαιο. Η εκπαίδευση πραγματοποιείται σε ένα cluster 1000 πυρήνων και διαρκεί περίπου μία μέρα για ένα ensemble τριών δέντρων με ένα σετ δεδομένων εκπαίδευσης 900K καρέ. Στη συνέχεια εξάγονται οι τοποθεσίες των αρθρώσεων με τη μέθοδο της μέσης μετατόπισης (mean shift), η οποία συγκλίνει στην επικρατούσα τιμή ενός εκτιμητή πυκνότητας που κατασκευάζεται με βάση τις προβλέψεις. Οι αρθρώσεις αποφασίζονται ότι εντοπίζονται στα επικρατή pixels που ευρίσκονται από τον αλγόριθμο. Η μέση ακρίβεια (mean average precision) της πρόβλεψης των αρθρώσεων για ένα σετ από πραγματικά δεδομένα ελέγχου με 9000 καρέ είναι 91.4%.

## 4.4 Το περιβάλλον ανάπτυξης Kinect for Windows SDK

Το Kinect for Windows SDK (Kinect SDK) [21] είναι το επίσημο περιβάλλον προγραμματισμού της Microsoft για τον αισθητήρα Kinect. Στόχος του SDK είναι να παρέχει μια προγραμματιστική διεπαφή υψηλού επιπέδου για την ανάκτηση της σκελετικής πληροφορίας και της πληροφορίας βάθους από το Kinect. Το SDK υποστηρίζει τον προγραμματισμό σε C++, C# και VB.NET και η τελευταία έκδοσή του διατίθεται στην επίσημη ιστοσελίδα του. Μαζί με το Kinect SDK παρέχονται και οι οδηγοί συσκευής για το Kinect. Τα εκτελέσιμα είναι συμβατά με τις σύγχρονες εκδόσεις των Windows (7, 8 και 10).

Το επίσημο SDK είναι λογισμικό κλειστού κώδικα και όπως φαίνεται και από τις συμβατές πλατφόρμες και γλώσσες προγραμματισμού παρέχει περιορισμένη ευελιξία στους χρήστες του σε σχέση με άλλες εναλλακτικές ανοιχτού κώδικα όπως το OpenNI. Παρόλα αυτά παρέχει πιο εύρωστη και υψηλής ποιότητας σκελετική πληροφορία χωρίς την ανάγκη βαθμονόμησης του αισθητήρα και την απαίτηση από τον τελικό χρήστη να αρχικοποιήσει την ανίχνευση σκελετού με κάποια πόζα (T-rose). Σε αυτή την εργασία κάνουμε αποκλειστική χρήση των σκελετικών δεδομένων, συνεπώς χρησιμοποιήσαμε το επίσημο SDK για να εξασφαλίσουμε την ποιότητα της εμπειρίας του τελικού χρήστη συμβιβάζοντας την ευελιξία ως προς τον προγραμματισμό και την τελική πλατφόρμα εκτέλεσης. Η έκδοση του Kinect SDK που χρησιμοποιήθηκε και στην οποία θα εστιάσουμε είναι η έκδοση v1.8. Ο βασικός τρόπος ανάκτησης πληροφορίας από τον αισθητήρα με το Kinect SDK είναι μέσω τριών ροών δεδομένων (data streams) που παρέχουν τα καρτέ του βάθους, χρώματος και σκελετού αντίστοιχα. Στη συνέχεια θα αναλύσουμε κάθε μια από τις ροές χωριστά.

#### 4.4.1 Η ροή καρτέ έγχρωμων εικόνων

Τα καρτέ έγχρωμων εικόνων παρέχονται από τη ροή εικόνων χρώματος, αλλιώς Color Stream (CS), σε διαφορετικές αναλύσεις και κωδικοποιήσεις. Οι διαθέσιμες κωδικοποιήσεις είναι οι RGB, YUV και Bayer.

Στην κωδικοποίηση RGB το χρώμα αναπαρίσταται σαν ένα γραμμικό 32-bit bitmap της μορφής X8R8G8B8. Οι διαθέσιμες αναλύσεις είναι  $640 \times 480$  σε 30 καρτέ το δευτερόλεπτο και σε εικόνες υψηλής ανάλυσης  $1280 \times 960$  στα 12 καρτέ το δευτερόλεπτο.

Στην κωδικοποίηση YUV το χρώμα αναπαρίσταται σαν ένα γραμμικό 16-bit bitmap με διόρθωση gamma της μορφής UYVY. Λόγω του μικρότερου αριθμού bit που απαιτούνται, αυτή η κωδικοποίηση χρησιμοποιεί λιγότερη μνήμη. Τα καρτέ είναι διαθέσιμα σε ανάλυση  $640 \times 480$  στα 15 καρτέ το δευτερόλεπτο.

Η κωδικοποίηση Bayer έχει τα ίδια χαρακτηριστικά με την κωδικοποίηση RGB, με τη διαφορά ότι χρησιμοποιεί περισσότερα pixel για την αναπαράσταση του πράσινου χρώματος για να προσομοιώσει καλύτερα την φυσιολογία του ανθρώπινου ματιού.

#### 4.4.2 Η ροή εικόνων βάθους

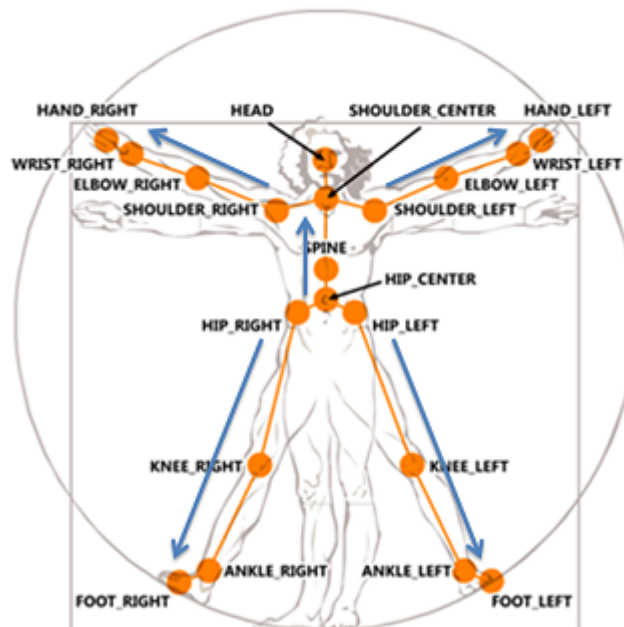
Τα καρτέ εικόνων βάθους παρέχονται από τη ροή εικόνων βάθους, αλλιώς Depth Stream (DS). Τα δεδομένα βάθους είναι εικόνες στην κλίμακα του γκρι όπου σε κάθε pixel αποθηκεύεται ένας απρόσημος αριθμός 16-bit. Τα πρώτα 3 bit αναπαριστούν τα δεδομένα διαχωρισμού των παιχτών και αποθηκεύουν την πληροφορία για τη συσχέτιση του ανιχνευόμενου σκελετού με τα δεδομένα βάθους. Τα υπόλοιπα 13 bit αποθηκεύουν την απόσταση του παίχτη από την κάμερα σε χιλιοστά. Αυτό θέτει τον περιορισμό ότι η μέγιστη απόσταση που μπορεί να κωδικοποιηθεί είναι 8 μέτρα.

Τα καρτέ βάθους είναι διαθέσιμα σε αναλύσεις  $640 \times 480$ ,  $320 \times 240$  και  $80 \times 60$  στα 30

καρέ το δευτερόλεπτο. Υπάρχουν δύο καταστάσεις λειτουργίας για την κάμερα βάθους, η προεπιλεγμένη που ανιχνεύει δεδομένα βάθους σε αποστάσεις από 0.8 έως 4 μέτρα και η κοντινή που παρέχει δεδομένα βάθους σε αποστάσεις από 0.4 έως 3 μέτρα.

#### 4.4.3 Η ροή σκελετικών δεδομένων

Σε προηγούμενη ενότητα περιγράψαμε αναλυτικά τον αλγόριθμο για την εξαγωγή των σκελετικών δεδομένων. Εδώ θα περιγράψουμε πώς η πληροφορία αυτή γίνεται διαθέσιμη στους χρήστες του Kinect SDK, συγκεκριμένα μέσω της ροής σκελετικών δεδομένων, Skeleton Stream (SS). Μέσω της SS παρέχεται η πληροφορία για τη θέση και τον προσανατολισμό των αρθρώσεων του χρήστη. Οι ανιχνευόμενες αρθρώσεις παρουσιάζονται στην Εικόνα 4.5. Σε κάθε αρθρωση ανατίθεται μια επισήμανση ως Ανιχνευμένη (Tracked), Συναγόμενη (Inferred) και Μη Ανιχνευμένη (Untracked), με βάση την εμπιστοσύνη της πρόβλεψης που προκύπτει από τον αλγόριθμο ανίχνευσης.

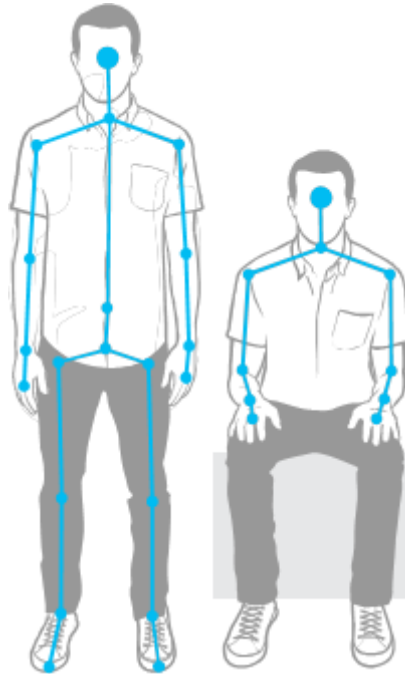


Σχήμα 4.5: Οι αρθρώσεις που ανιχνεύονται από το Kinect SDK v1.8

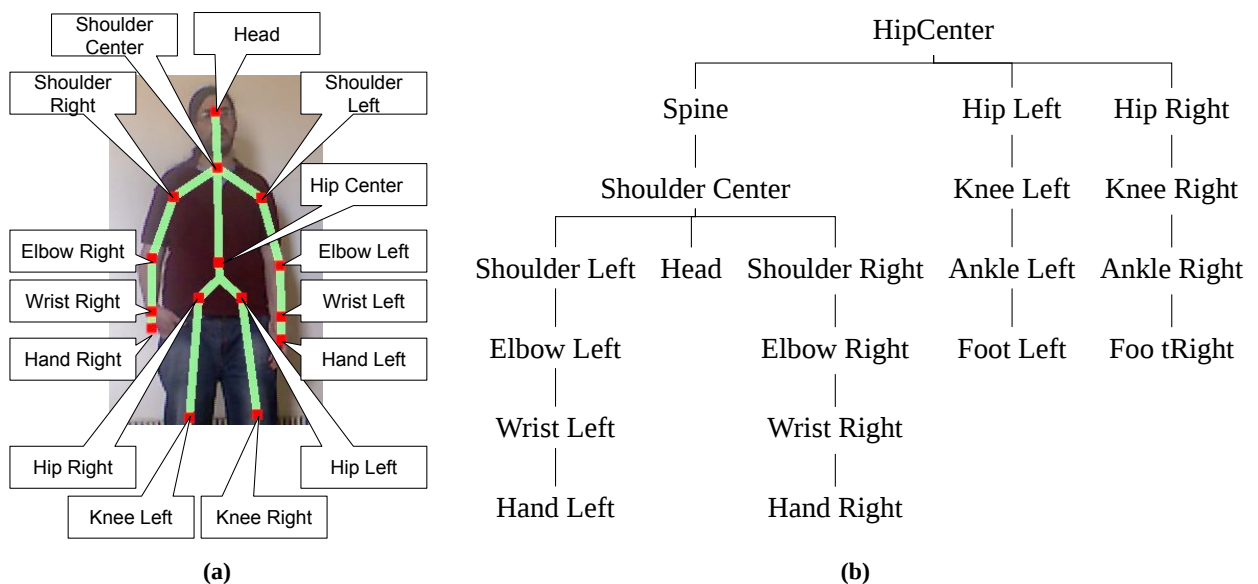
Η ανίχνευση έχει δύο καταστάσεις λειτουργίας, την προεπιλεγμένη όπου ανιχνεύονται όλες οι αρθρώσεις, και την κατάσταση καθιστών χρηστών, αλλιώς Seated Mode (SM), όπου ανιχνεύονται μόνο οι 10 αρθρώσεις στο πάνω μέρος του σώματος όπως φαίνεται στην Εικόνα 4.6. Στην SM η ανίχνευση γίνεται με βάση την κίνηση του χρήστη και όχι με βάση την απόσταση από το background όπως στην προεπιλεγμένη. Η SM καταναλώνει περισσότερους υπολογιστικούς πόρους και παράγει λιγότερα καρέ ανά δευτερόλεπτο, όμως παράγει καλύτερα αποτελέσματα όταν ο αισθητήρας είναι σε Near Mode.

Εκτός από τη θέση, σημαντική είναι και η πληροφορία για τον προσανατολισμό των αρ-

θρώσεων. Η αρθρώσεις είναι οργανωμένες σε μια δεντρική ιεραρχία όπως περιγράφεται στην Εικόνα 4.7. Με βάση αυτή την ιεραρχία είναι δυνατός ο υπολογισμός του μετασχηματισμού περιστροφής που απαιτείται για τη μετάβαση από μια άρθρωση πατέρα σε μία άρθρωση παιδί, καθώς και γωνίες μεταξύ των αρθρώσεων πατέρα - παιδί στον 3Δ χώρο.



**Σχήμα 4.6:** Οι καταστάσεις λειτουργίας κατά την ανίχνευση των αρθρώσεων. Από αριστερά προς τα δεξιά: Η προεπιλεγμένη κατάσταση και η κατάσταση καθιστών χρηστών.



**Σχήμα 4.7:** Η ιεραρχία των αρθρώσεων (a) πάνω στον ανθρώπινο σκελετό. (b) σε δεντρική δομή.

#### 4.4.4 Το Kinect SDK v2.0

Σε αυτή την ενότητα θα συγκρίνουμε τις τεχνικές προδιαγραφές και τα χαρακτηριστικά της νέας έκδοσης του αισθητήρα Kinect 2.0 με το Kinect 360 πάνω στο οποίο υλοποιήθηκε αυτή η εργασία. Το νέο SDK δεν είναι συμβατό με τις παλιότερες εκδόσεις καθώς το API έχει επανασχεδιαστεί και οι νέες δυνατότητες που προσφέρει βασίζονται στις υψηλότερες τεχνικές προδιαγραφές της νέας έκδοσης. Μια σύγκριση των τεχνικών προδιαγραφών παρατίθεται στον Πίνακα 4.3.

Χαρακτηριστικά	Kinect 360	Kinect 2.0
Ανάλυση εικόνας βάθους	320 × 240px	512 × 424px
Ανάλυση έγχρωμης εικόνας	640 × 480px	1920 × 1080px
Αριθμός ανιχνευόμενων σκελετών	2	6
Αριθμός αρθρώσεων ανά σκελετό	20	25
Οριζόντιο Οπτικό πεδίο	57°	70°
Κατακόρυφο Οπτικό πεδίο	43°	60°
Τύπος USB	2.0	3.0

**Πίνακας 4.3:** Σύγκριση των τεχνικών προδιαγραφών του Kinect 2.0 με το Kinect 360 [22].

Οι βελτιωμένες προδιαγραφές, και ειδικά η αυξημένη ανάλυση της κάμερας βάθους, έχουν οδηγήσει σε σημαντική βελτίωση της μονάδας ανίχνευσης σκελετού και σε ένα σύνολο από νέα χαρακτηριστικά που προσφέρονται από το Kinect SDK 2.0 [23].

- Ανίχνευση κλειστής/ανοιχτής παλάμης
- Ανίχνευση/αναγνώριση προσώπου
- Ανίχνευση κλίσης του σώματος
- Αναγνώριση χειρονομιών

Η αναγνώριση χειρονομιών που παρέχεται από το Kinect 2.0 επιτρέπει στο χρήστη να ορίσει νέες χειρονομίες. Οι χειρονομίες καταχωρούνται μέσω γραφικού περιβάλλοντος και χρησιμοποιούνται για την εκπαίδευση μοντέλων μηχανικής μάθησης (Adaboost και Random Forest Regression). Στο μέλλον θα εκτελέσουμε μια σειρά πειραμάτων για να συγκρίνουμε την ακρίβεια και την επίδοση της μεθόδου που περιγράφεται σε αυτή την εργασία με τη μέθοδο που παρέχεται από το Kinect 2.0, ενώ θα εργαστούμε πάνω στην επέκταση και τη βελτίωση της μεθόδου που καταχωρούνται νέες χειρονομίες προς αναγνώριση.





## Κεφάλαιο 5

# Ανίχνευση Στάσεων του Σώματος

Καθημερινά χρησιμοποιούμε το σώμα μας σαν διεπαφή για τη μεταφορά νοήματος. Η γλώσσα του σώματος αποτελεί έναν από τους βασικούς και ταυτόχρονα φυσικούς τρόπους μεταφοράς πληροφορίας σε άλλους ανθρώπους. Για τη χρήση της γλώσσας του σώματος για την επικοινωνία με υπολογιστικά συστήματα ή αντικείμενα είναι αναγκαίο να την αναλύσουμε και στα δομικά της συστατικά και να εξετάσουμε τα επιμέρους κομμάτια ξεχωριστά. Μια απλοϊκή ίσως αλλά αποδοτική αποδόμηση είναι ο διαχωρισμός των δομικών αυτών κομματιών σε στάσεις του σώματος, ή αλλιώς σε “πόζες” και σε “χειρονομίες”. Σε αυτό το κεφάλαιο θα ασχοληθούμε με την ανίχνευση στάσεων του σώματος και θα ασχοληθούμε με την αναγνώριση χειρονομιών στο επόμενο κεφάλαιο.

### 5.1 Προσδιορισμός του προβλήματος

Για να ξεκινήσουμε τη μελέτη μας πρέπει πρώτα να ορίσουμε το πρόβλημα και τις συνιστώσες του. Θα ξεκινήσουμε από τον ορισμό της πόζας. Ένα παράδειγμα πόζας για την απτή παρουσίαση του ορισμού φαίνεται στο Σχήμα 5.1.

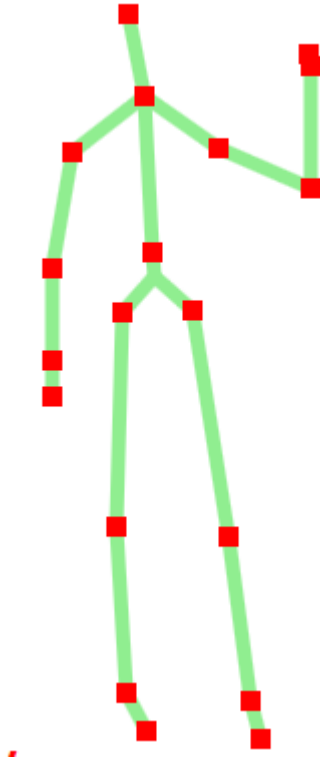
**Ορισμός 3:** Πόζα ονομάζουμε την προσωρινή παύση της κίνησης των αρθρώσεων του σώματος σε μία διακριτή διάταξη.

Για την ανάπτυξη της μεθόδου ανίχνευσης πόζας μπορούν να γίνουν κάποιες παραδοχές και να τεθούν περιορισμοί με βάση την περίπτωση χρήσης που μελετάμε, συγκεκριμένα την κωδικοποίηση μιας αναγνωρισμένης πόζας σαν σήμα ελέγχου για συσκευές σε ένα έξυπνο δωμάτιο.

#### Παραδοχές

**Ο φωτισμός του δωματίου** Το δωμάτιο έχει επαρκή φωτισμό και το Kinect έχει ρυθμιστεί για να λειτουργεί σε αυτές τις συνθήκες

**Η θέση και ο προσανατολισμός του χρήστη** Ο χρήστης τοποθετείται σε μια απόσταση που βρίσκεται μέσα στο διάστημα λειτουργίας του αισθητήρα και το μπροστινό μέρος του σώματός του είναι



**Σχήμα 5.1:** Μια πόζα που μπορεί να ανιχνευθεί με την προτεινόμενη μέθοδο (hello-pose)

προσανατολισμένο προς τον αισθητήρα για να είναι δυνατή η ανίχνευση του σκελετού και να αποφευχθεί η ανάγκη αντιστάθμισης της περιστροφής του χρήστη.

**Προκαθορισμένα σήματα ελέγχου** Τα σήματα ελέγχου, άρα και οι πόζες που πρέπει να ανιχνευτούν, ορίζονται στατικά και όχι κατά τη λειτουργία.

### Περιορισμοί

**Η ανίχνευση πραγματοποιείται σε πραγματικό χρόνο** Ο έλεγχος συσκευών γίνεται ζωντανά, οπότε η εξαγωγή χαρακτηριστικών και η ανίχνευση πρέπει να πραγματοποιούνται σε πραγματικό χρόνο

**Αναλλοίωτη βάθους και θέσης** Ο χρήστης μπορεί να μετακινηθεί ελεύθερα όσο βρίσκεται στο πεδίο λειτουργίας του αισθητήρα. Η ανίχνευση πρέπει να λειτουργεί ανεξάρτητα με τη θέση του χρήστη σε αυτό το πεδίο.

**Υποστήριξη δύο χρηστών** Είναι δυνατή η παράλληλη παρακολούθηση (tracking) του σκελετού δύο χρηστών και κατ' επέκταση η ανεξάρτητη αναγνώριση για τις πόζες τους,

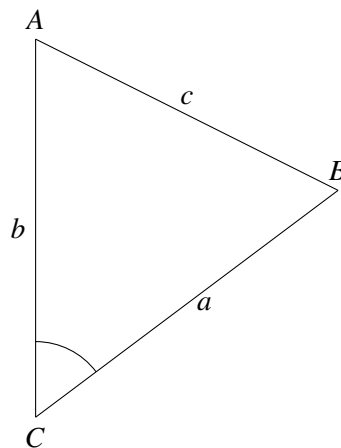
Λαμβάνοντας υπόψη τα παραπάνω μπορούμε να προχωρήσουμε στην παρουσίαση του αλγορίθμου. Ο αλγόριθμος ανίχνευσης πόζας βασίζεται στη μέθοδο που περιγράφεται από τους Jarret Webb και Ashley James στο [24]. Υλοποιήσαμε μια αποδοτική, ασύγχρονη παραλλαγή της περιγρα-

φόμενης μεθόδου σε node.js και την επεκτείνουμε με ένα XML schema για τον εύκολο ορισμό νέων ποζών.

## 5.2 Εξαγωγή γεωμετρικών χαρακτηριστικών

Επιλέξαμε να αντιμετωπίσουμε το πρόβλημα της ανίχνευσης πόζας με τη χρήση μίας γεωμετρικής προσέγγισης. Για την κατασκευή ενός γεωμετρικού μοντέλου για τις πόζες εκμεταλλευόμαστε την ιεραρχική διάταξη των αρθρώσεων και υπολογίζουμε τη γωνία ανάμεσα σε κάθε άρθρωση παιδί και τον πατέρα της σε ένα παγκόσμιο (global) σύστημα συντεταγμένων. Η χρήση των γωνιών μεταξύ των αρθρώσεων διατηρεί την αναλλοίωτη βάθους και θέσης, αφού οι γωνίες μεταξύ των αρθρώσεων δεν εξαρτώνται από τη θέση του χρήστη στο οπτικό πεδίο του αισθητήρα.

Ο υπολογισμός των γωνιών πραγματοποιείται με τη χρήση μιας μεθόδου τριγωνοποίησης. Θεωρούμε δύο σημεία  $A$  και  $B$  με γνωστές θέσεις, τα οποία αντιστοιχούν στις αρθρώσεις πατέρα  $J_A$  και παιδί  $J_B$  αντίστοιχα. Επιλέγουμε ένα τρίτο σημείο  $C$  στη θέση  $(x_p, 0)$  όπου  $x_p$  η  $x$  συντεταγμένη της θέσης της άρθρωσης πατέρα  $J_A$ . Τα σημεία  $A, B, C$  σχηματίζουν ένα τρίγωνο με πλευρές  $a, b, c$  όπως φαίνεται στο Σχήμα 5.2

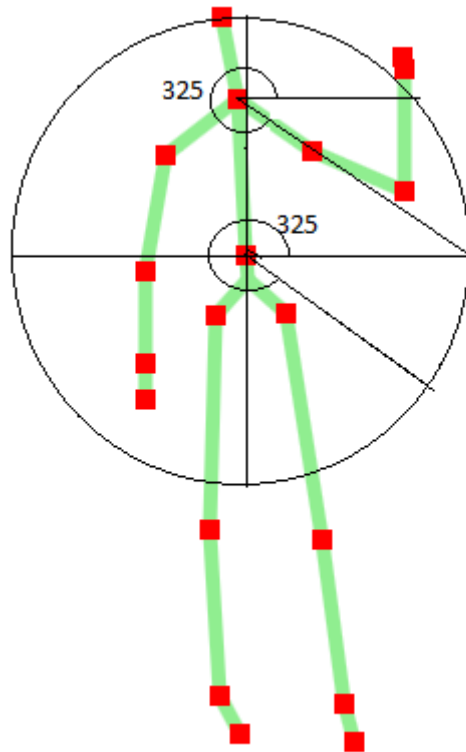


**Σχήμα 5.2:** Το τρίγωνο που σχηματίζεται για τον υπολογισμό της γωνίας ανάμεσα στις αρθρώσεις  $A$  και  $B$

Με βάση το νόμο των συνημιτόνων η γωνία  $\angle ACB$  μπορεί να υπολογιστεί από τον τύπο

$$C = \cos^{-1} \frac{a^2 + b^2 - c^2}{2ab} . \quad (5.1)$$

Στην πράξη ή μέθοδος αυτή οδηγεί στην απεικόνιση των γωνιών μεταξύ κάθε άρθρωσης παιδιού και πατέρα σε ένα τριγωνομετρικό κύκλο, αν λάβουμε τις θετικές φορές για τις γωνίες όπως φαίνεται στο Σχήμα 5.3 όπου λαμβάνουμε τις αρθρώσεις `ShoulderLeft` και `ShoulderCenter` ως παράδειγμα.



**Σχήμα 5.3:** Απεικόνιση της γωνίας μεταξύ δύο αρθρώσεων στον τριγωνομετρικό κύκλο.

Πρέπει να τονιστεί ότι αυτή η μέτρηση είναι η μοναδική που χρησιμοποιούμε για την αναγνώριση των στάσεων του σώματος καθώς δεν επηρεάζεται από τα φυσικά χαρακτηριστικά του χρήστη, όπως το ύψος και το βάρος του, τη θέση του στο πεδίο λήψης του αισθητήρα και την απόστασή του από την κάμερα. Αυτό οδηγεί στην κατασκευή ενός εξαιρετικά απλού μοντέλου, το οποίο επιτρέπει την ακριβή και γρήγορη αναγνώριση μιας πληθώρας από πόζες.

### 5.3 Αλγόριθμος Ανίχνευσης Πόζας

Έχοντας παράγει μια γεωμετρική μοντελοποίηση για τις πόζες μπορούμε να προχωρήσουμε στην ανάπτυξη της μεθόδου ανίχνευσης. Η απλότητα του μοντέλου μας επιτρέπει να προσανατολιστούμε σε μια λύση ταιριάσματος προτύπων (template matching), αντί για μια μέθοδο μηχανικής μάθησης που θα απαιτεί εκπαίδευση ταξινομητών ή μια σύνθετη μέθοδο αναγνώρισης προτύπων. Στην προτεινόμενη λύση ο χρήστης μπορεί να παρέχει ως είσοδο ένα αρχείο XML το οποίο περιγράφει τις γωνίες που λαμβάνουν ένα υποσύνολο των αρθρώσεων (ή και όλες οι αρθρώσεις) κατά την υιοθέτηση μιας πόζας. Τα XML templates για τις πόζες υπακούουν σε ένα XML schema, το οποίο κατασκευάσαμε για τον αυστηρό καθορισμό της μορφής των προτύπων. Σε αυτά τα πρότυπα καθορίζουμε την γωνία που πρέπει να σχηματίζεται ανάμεσα σε μία άρθρωση  $J$  και τον πατέρα της, καθώς και ένα περιθώριο απόκλισης από αυτή τη γωνία. Αν για μία άρθρωση δεν προσδιορίζονται αυτές οι

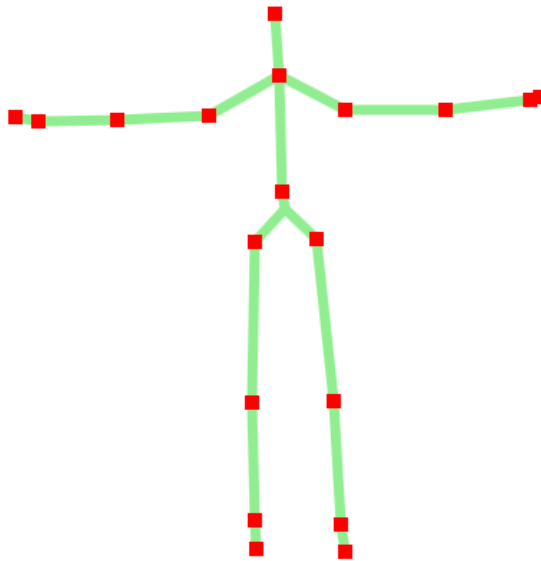
παράμετροι, τότε η απόκλιση θεωρείται ίση με  $\infty$  και η άρθρωση αυτή αγνοείται. Ακολουθεί τμήμα του κώδικα για το XML schema για μία μόνο άρθρωση, όπου φαίνεται η δομή του XML μιας πρότυπης πόζας. Με βάση αυτό το μερικό schema, στο χρήστη επιτρέπεται να ορίσει την επιθυμητή γωνία ανάμεσα στην άρθρωση *ElbowLeft* και τον πατέρα της (*ShoulderLeft*). Παρατηρούμε επίσης ότι δεν επιτρέπεται να ορίσουμε κάποια γωνία για την άρθρωση *HipCenter*, η οποία είναι η ρίζα του δέντρου των αρθρώσεων. Ο πλήρης κώδικας παρατίθεται στο Παράρτημα Α.

#### Κώδικας 5.0: Η βασική δομή του XML schema μιας πρότυπης πόζας μόνο για την άρθρωση *ElbowLeft*

```
<?xml version="1.0" encoding="utf-8"?>

<xs:schema id="PoseRules"
  attributeFormDefault="unqualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:PoseRules="urn:PoseRules"
  targetNamespace="urn:PoseRules"
  elementFormDefault="qualified">
  <xs:element name="Pose">
    <xs:complexType>
      <xs:sequence>
        <xs:element type="xs:string" name="DisplayName" minOccurs="1" maxOccurs="1" />
        <xs:element name="AngleConstraints" minOccurs="1" maxOccurs="1">
          <xs:complexType>
            <xs:all>
              <xs:element name="ElbowLeft" minOccurs="0" maxOccurs="1">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element type="xs:decimal" name="Desired" />
                    <xs:element type="xs:decimal" name="Deviation" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
              ...
            </xs:all>
            <xs:attribute type="xs:string" name="units" default="degrees" />
          </xs:complexType>
        </xs:element>
      </xs:sequence>
      <xs:attribute type="xs:string" name="id" use="required" />
    </xs:complexType>
  </xs:element>
</xs:schema>
```

Για την υλοποίηση του template matching, σε κάθε καρέ εξάγουμε τη γεωμετρική αναπαράσταση της πόζας και τη συγκρίνουμε με το πρότυπο μοντέλο της πόζας. Αν η σύγκριση είναι επιτυχής η πόζα θεωρείται αναγνωρισμένη. Η παραπάνω μέθοδος λειτουργεί ικανοποιητικά μόνο στην περίπτωση όπου ο χρήστης εκτελεί μια πόζα με απόλυτη ακρίβεια. Όμως σε ρεαλιστικές συνθήκες αυτό δεν είναι μια ρεαλιστική απαίτηση από τη μεριά του αλγορίθμου. Αυτός ο περιορισμός αίρεται αν παρέχουμε ένα περιθώριο σφάλματος για κάθε σύγκριση. Αυτό το “τέχνασμα” παρέχει στον αλγόριθμο ικανοποιητική δυνατότητα “γενίκευσης” χωρίς να έχει μεγάλο αντίκτυπο στην ακρίβεια και την ταχύτητα του αλγορίθμου. Συνεπώς το τελικό σύστημα ανίχνευσης πόζας αποτελείται από ένα σύνολο ανιχνευτών (detectors), κάθε ένας από τους οποίους ταιριάζει ένα πρότυπο. Οι ανιχνευτές δεν μοιράζονται δεδομένα, οπότε μπορούν να υλοποιηθούν ώστε να τρέχουν σε διαφορετικά threads, για να βελτιωθεί η απόδοση του συστήματος. Με τη χρήση αυτής της μεθόδου επιτύχαμε την αναγνώριση μιας πληθώρας από πόζες για ένα σχετικά μεγάλο σύνολο χρηστών. Στο Σχήμα 5.4 παρέχουμε ένα παράδειγμα πόζας και στον Κώδικα 5.1 το XML template που αντιστοιχεί σε αυτή την πόζα.



Σχήμα 5.4: Η πόζα HANDSAPART

Κώδικας 5.1: Το πρότυπο XML για την πόζα HandsApart

```

<?xml version="1.0" encoding="utf-8" ?>
<Pose id="HANDSAPART"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema
  -instance"
  xsi:schemaLocation="urn:PoseRules PoseRules.xsd">
  <DisplayName>Jack I'm Flying</DisplayName>
  <AngleConstraints units="degrees">
    <ElbowLeft>
      <Desired>180</Desired>
      <Deviation>15</Deviation>
    </ElbowLeft>
    <HandLeft>
      <Desired>180</Desired>
      <Deviation>15</Deviation>
    </HandLeft>
    <ElbowRight>
      <Desired>0</Desired>
      <Deviation>15</Deviation>
    </ElbowRight>
    <HandRight>
      <Desired>0</Desired>
      <Deviation>15</Deviation>
    </HandRight>
  </AngleConstraints>
</Pose>

```

Σχήμα 5.5: Το πρότυπο XML για την πόζα HANDSAPART

Παραθέτουμε τα βασικά τμήματα του αλγορίθμου ανίχνευσης πόζας σε ψευδοκώδικα. Αρχικά θα δείξουμε κάποιες βοηθητικές συναρτήσεις. Η Συνάρτηση 3 μετασχηματίζει μια γωνία ώστε να βρίσκεται στο διάστημα  $0^\circ - 360^\circ$ .

---

**Algorithm 3** Συνάρτηση ομαλοποίησης πόζας στο διάστημα  $0^\circ - 360^\circ$

---

```
1: function NormalizeAngle(angle)
2:   return mod(angle + 360, 360)
```

---

Η Συνάρτηση 4 υπολογίζει τη γωνία μεταξύ μιας άρθρωσης πατέρα και μιας άρθρωσης παιδιού σε μοίρες με βάση το νόμο των συνημιτόνων όπως περιγράψαμε στην Ενότητα 5.2.

---

**Algorithm 4** Συνάρτηση για τον υπολογισμό της γωνίας ανάμεσα σε ένα ζεύγος αρθρώσεων πατέρα-παιδιού

---

```
1: ▷ Οι παράμετροι parentPos και childPos είναι η θέση των αρθρώσεων πατέρα και παιδιού χωρίς
   τη συντεταγμένη βάθους
2: function CalculateAngle(parentPos, childPos)
3:   a ← EuclideanDistance(parentPos, childPos)
4:   b ← childPos.x
5:   childProjection ← createPos(0, childPos.y)
6:   c ← EuclideanDistance(parentPos, childProjection)
7:   angle ← acos((a * a + b * b - c * c) / (2 * a * b))
8:   if childPos.y > parentPos.y then
9:     angle ← 360 - angle
10:  return NormalizeAngle(angle)
```

---

Η Συνάρτηση 5 επιστρέφει *true* αν και μόνο αν μια γωνία βρίσκεται στο διάστημα  $[d - e, d + e]$ , όπου  $d$  η επιθυμητή τιμή της γωνίας και  $e$  ένα περιθώριο απόκλισης από αυτήν την τιμή, όπως αυτές οι τιμές καθορίζονται από το πρότυπο XML.

---

**Algorithm 5** Συνάρτηση για τη σύγκριση μιας γωνίας από το πρότυπο XML μιας πόζας με τη μετρούμενη γωνία

---

```
1: function CompareAngles(angle, desired, deviation)
2:   high ← NormalizeAngle(desired + deviation)
3:   low ← NormalizeAngle(desired - deviation)
4:   if low > high then
5:     low ← low - 360
6:     if angle > 270 then
7:       angle ← angle - 360
8:   return (angle < high) and (angle > low)
```

---

Τέλος παραθέτουμε την κύρια ρουτίνα ανίχνευσης πόζας στον Αλγόριθμο 6. Η ρουτίνα αυτή λαμβά-

νει ένα πρότυπο πόζας και τα χαρακτηριστικά, δηλαδή τις γωνίες των αρθρώσεων στο τρέχον καρέ, και υπολογίζει αν τα χαρακτηριστικά πληρούν τους περιορισμούς που επιβάλλονται από το πρότυπο με αναδρομή ουράς. Η συνάρτηση *IsPose* τρέχει για κάθε πόζα προς αναγνώριση με χρήση ταυτοχρονισμού (concurrency).

---

**Algorithm 6** Αλγόριθμος ανίχνευσης πόζας με ταίριασμα προτύπων

---

```

1: procedure JointAnglesInRange(tree, parent, accumulator)
2:   if isLeaf(parent) then return accumulator
3:   children ← getChildren(tree, parent)
4:   procedure checkAngles(inRange, currentChild) ▷ Αυτή η εμφωλευμένη συνάρτηση δίνεται
   ως όρισμα στη foldLeft
5:     angle ← NormalizeAngle(features[currentChild])
6:     desired ← constraints[currentChild].desired
7:     deviation ← constraints[currentChild].deviation
8:     newAccumulator ← inRange and CompareAngles(angle, desired, deviation)
9:     return JointAnglesInRange(getChildren(tree, parent), currentChild, newAccumulator) ▷
   Αναδρομή ουράς για όλα τα παιδιά
10:  return foldLeft(children, checkAngles, accumulator)
11: procedure IsPose(template, features)
12:  constraints ← template.angleConstraints
13:  jointHierarchy ← constructJointTree()
14:  return JointAnglesInRange(jointHierarchy, "HipCenter", true) ▷ Η αναδρομή αρχίζει από
   την άρθρωση ρίζα HipCenter με αρχική τιμή true στο συσσωρευτή

```

---



## Κεφάλαιο 6

### Αναγνώριση χειρονομιών

Στο προηγούμενο κεφάλαιο προτείναμε μια απλή και αποδοτική μέθοδο για την ανίχνευση στάσεων του σώματος και τη χρήση τους ως σήματα ελέγχου για τον έλεγχο ενός έξυπνου δωματίου. Σε αυτή την εργασία δεν περιοριζόμαστε όμως μόνο στο πεδίο των στάσεων του σώματος αλλά επεκτείνουμε τη μελέτη μας και στο πεδίο των κινούμενων χειρονομιών. Σε αυτό το κεφάλαιο θα αναλύσουμε το θεωρητικό υπόβαθρο, το σχεδιασμό και την υλοποίηση μιας μεθόδου για τη μετάφραση των χειρονομιών σε σήματα ελέγχου. Η μέθοδος αυτή είναι προσαρμοσμένη στην περίπτωση χρήσης μιας εφαρμογής ΔτΑ και έχει βελτιστοποιηθεί ώστε να εκτελείται σε ένα περιβάλλον περιορισμένων υπολογιστικών και δικτυακών πόρων.

#### 6.1 Προσδιορισμός του προβλήματος

Και σε αυτή την ανάλυση θα ξεκινήσουμε τη μελέτη μας με την παράθεση των σχετικών ορισμών, παραδοχών και περιορισμών. Αυτή η διαδικασία είναι απαραίτητη για την αξιολόγηση των διαφορετικών εναλλακτικών επιλύσεων και τη λήψη των σχεδιαστικών αποφάσεων. Πρέπει να σημειωθεί ότι εξακολουθούν να ισχύουν οι ίδιοι περιορισμοί και παραδοχές που αναφέρθηκαν στην Ενότητα 5.1.

**Ορισμός 4:** Ως χειρονομία ορίζεται μια συνεχής ακολουθία από πόζες.

Εναλλακτικά

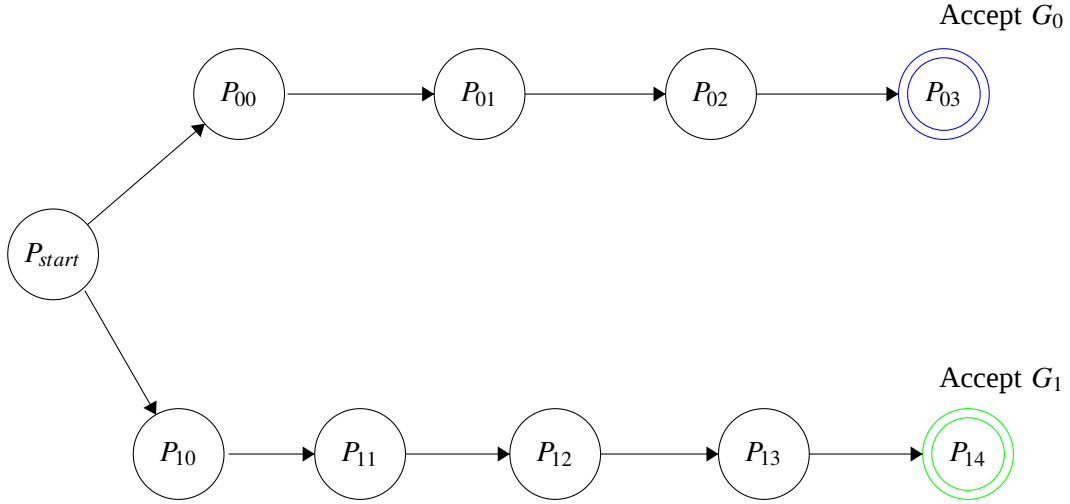
**Ορισμός 5:** Ως χειρονομία ορίζεται η μεταφορά ενός υποσυνόλου των αρθρώσεων του σκελετού από το σημείο  $A$  στο σημείο  $B$  πάνω σε μία προκαθορισμένη τροχιά.

Κάθε ένας από τους παραπάνω ορισμούς μπορεί να οδηγήσει σε ένα διαφορετικό σχεδιασμό της μονάδας αναγνώρισης. Στη συνέχεια θα παραθέσουμε μερικές από αυτές τις υποψήφιες εναλλακτικές για να εισάγουμε την προτεινόμενη λύση. Αν σύμφωνα με τον Ορισμό 4 θεωρήσουμε μια χειρονομία σαν μια συνεχή ακολουθία από πόζες, τότε μπορεί να προκύψει μια πολύ απλή μέθοδος

αναγνώρισης. Το πρώτο βήμα σε αυτή τη μέθοδο είναι η διακριτοποίηση της συνεχούς ακολουθίας, όπου λαμβάνουμε υπόψιν μόνο την αρχική και τελική πόζα στη χειρονομία και έναν πεπερασμένο αριθμό από ενδιάμεσες πόζες. Αυτή η διακριτή ακολουθία από πόζες μπορεί να είναι η είσοδος μιας πεπερασμένης μηχανής καταστάσεων (FSM) της οποίας κάθε τελική κατάσταση αντιστοιχεί σε μία αναγνωρισμένη χειρονομία. Ένα παράδειγμα μιας μηχανής καταστάσεων που αναγνωρίζει δύο χειρονομίες  $G_0, G_1$  με τις ενδιάμεσες πόζες  $P_{0i}$  και  $P_{1j}$  αντίστοιχα απεικονίζεται στο Σχήμα 6.1. Θεωρούμε επίσης και μια κοινή πόζα αρχής  $P_{start}$  που ενεργοποιεί την ανίχνευση.

Η παραπάνω μέθοδος βασίζεται στη χρήση ντετερμινιστικών μηχανών καταστάσεων. Για την κατασκευή ενός πιο σθεναρού συστήματος αναγνώρισης θα μπορούσαμε να προσανατολιστούμε σε μια λύση που βασίζεται στην εκπαίδευση ενός πιθανοτικού μοντέλου που βασίζεται σε μηχανές καταστάσεων όπως τα κρυφά μαρκοβιανά μοντέλα (HMM). Σε ένα τέτοιο μοντέλο οι ενδιάμεσες πόζες θεωρούνται ως παρατηρήσεις που καθορίζουν τις μεταβάσεις ανάμεσα στις κρυφές καταστάσεις του FSM. Το μοντέλο εκπαιδεύεται με βάση μια παραλλαγή του αλγορίθμου Expectation-Maximization, συνεπώς αναγνωρίζεται η χειρονομία η οποία μεγιστοποιεί την πιθανοφάνεια του μοντέλου, δηλαδή η χειρονομία με τη μέγιστη πιθανότητα δεδομένης της ακολουθίας από πόζες που έχει παρατηρηθεί. Αυτό παρέχει δύο πλεονεκτήματα. α) Η σθεναρότητα του μοντέλου σε περιπτώσεις σφάλματος στην ανίχνευση κάποιας ενδιάμεσης πόζας ή στην περίπτωση είσοδου ελλιπούς ακολουθίας. β) Μπορούμε να εφαρμόσουμε κατωφλίωση στη δεσμευμένη πιθανότητα της ανιχνευμένης χειρονομίας για τη διασφάλιση της ποιότητας των αποτελεσμάτων, δηλαδή να απορρίψουμε όλες τις αναγνωρισμένες χειρονομίες που η πιθανοφάνειά είναι μικρότερη από μία προκαθορισμένη τιμή.

Παρόλο που οι δύο παραπάνω μέθοδοι αποτελούν δελεαστικές λύσεις, καθώς έχουμε ήδη υλοποιήσει τη μονάδα ανίχνευσης πόζας, υποφέρουν από ένα βασικό μειονέκτημα: τον υπολογιστικό φόρτο στο σύστημα. Ο υπολογιστικός αυτός φόρτος οφείλεται στην παρουσία ενός μεγάλου αριθμού ανιχνευτών για τις ενδιάμεσες πόζες οι οποίοι εκτελούνται σειριακά ή παράλληλα και πραγματοποιούν ανίχνευση σε κάθε καρτέ. Επίσης αν η μονάδα ανίχνευσης πόζας μεταφερθεί στο cloud πρέπει σε κάθε καρτέ να μεταφέρονται τα εξαγόμενα γεωμετρικά χαρακτηριστικά της πόζας. Συνεπώς η ανίχνευση χειρονομιών με αυτή τη μέθοδο θα είναι εξαιρετικά ευαίσθητη στην ποιότητα του δικτύου και η καθυστέρηση του δικτύου θα επηρεάζει άμεσα την ακρίβεια των προβλέψεων και συνεπώς την εμπειρία του τελικού χρήστη. Για αυτούς τους λόγους αποφεύγουμε τη χρήση του υψηλού επιπέδου μοντέλου αναπαράστασης της χειρονομίας και επιλέγουμε τη χρήση χαμηλότερου επιπέδου γεωμετρικών χαρακτηριστικών που συνοψίζουν την πληροφορία από όλα τα καρτέ της κίνησης. Αυτά τα χαρακτηριστικά υποστηρίζονται από τον Ορισμό 5 και συνδυάζονται με ένα αποδοτικό αλγόριθμο μηχανικής μάθησης που εκτελείται στο cloud για την ταξινόμηση των χειρονομιών. Όπως θα περιγράψουμε και θα αποδείξουμε πειραματικά στην επόμενη ενότητα η εξαγωγή αυτών των χαρακτηριστικών είναι υπολογιστικά συμφέρουσα ενώ το διάνυσμα χαρακτηριστικών έχει πολύ μικρό μέγεθος και μεταφέρεται μέσω του δικτύου μόνο αφού ολοκληρωθεί η εκτέλεση της χειρονομίας και όχι σε κάθε καρτέ.



**Σχήμα 6.1:** Μηχανή καταστάσεων για την αναγνώριση των χειρονομιών  $G_0$  και  $G_1$ . Οι ενδιάμεσες καταστάσεις αποτελούν ενδιάμεσες πόζες σε κάθε χειρονομία.

## 6.2 Εξαγωγή Χαρακτηριστικών

Το προτεινόμενο σύνολο χαρακτηριστικών έχει εμπνευστεί εν μέρει από τα χαρακτηριστικά που προτείνονται από τον Sheng [25], το έργο του οποίου επέκτεινε αυτό του Rubine [26] από το χώρο των 2Δ στις 3Δ. Αρχικά θα παραθέσουμε το μαθηματικό φορμαλισμό που θα χρησιμοποιηθεί σε αυτή την ενότητα. Ορίζουμε μια δεδομένη άρθρωση ως  $J$ . Τότε οι αρθρώσεις  $J_c$  και  $J_p$  είναι αντίστοιχα οι αρθρώσεις παιδί και πατέρας της  $J$ . Επίσης ορίζουμε ως  $F_i$ ,  $i = 1, 2, \dots, N$  το  $i$ -οστό καρέ του βίντεο και ως  $\mathbf{v}_i^{(J)} = (v_x^{(J)}, v_y^{(J)}, v_z^{(J)})$  ορίζεται το διάνυσμα που αντιστοιχεί στις 3Δ συντεταγμένες της άρθρωσης  $J$  στο καρέ  $F_i$ . Επίσης ως  $\mathcal{V}^{(J)}$  ορίζεται το σύνολο όλων των συντεταγμένων των αρθρώσεων  $\mathbf{v}_i^{(J)}$ . Ως  $B(\mathcal{V}^{(J)})$  συμβολίζεται το πλαίσιο (bounding box) του  $\mathcal{V}^{(J)}$ , ενώ ως  $a_{B(\mathcal{V}^{(J)})}$  και  $b_{B(\mathcal{V}^{(J)})}$  ορίζονται τα δύο διαφορετικά μήκη των πλευρών του. Τα εξαγόμενα χαρακτηριστικά συνοψίζονται στον Πίνακα 6.1. Για την εξαγωγή αυτών των χαρακτηριστικών χρησιμοποιείται μόνο ένα υποσύνολο των αρθρώσεων (οι 3Δ συντεταγμένες τους). Συγκεκριμένα θεωρούμε ότι το  $J \in \{ElbowLeft, ElbowRight, HandLeft, HandRight\}$  καθώς οι χειρονομίες εκτελούνται με χέρια, οπότε η κίνηση αυτών των αρθρώσεων αποτελεί τον καθοριστικό παράγοντα για μια χειρονομία. Τα χαρακτηριστικά που επισημαίνονται με το σύμβολο \*, υπολογίζονται με χρήση μόνο για τις αρθρώσεις  $HandLeft$  και  $HandRight$  ενώ τα υπόλοιπα υπολογίζονται για όλες τις αρθρώσεις.

Εκτός των προαναφερθέντων υπολογίζονται η μέση και μέγιστη τιμή της γωνίας μεταξύ δύο αρθρώσεων πατέρα-παιδιού,  $\{J_p, J\}$ , (ή ισοδύναμα  $\{J, J_c\}$ ) για  $F_1, \dots, F_N$ . Από την ιεραρχία των αρθρώσεων προκύπτει ότι  $\{J_p, J\} \in \{\{ElbowLeft, HandLeft\}, \{ElbowRight, HandRight\}\}$ . Η γωνίες αυτές  $\theta_{pc}$  υπολογίζονται ανάμεσα στα ζεύγη των προαναφερθέντων αρθρώσεων με βάση τη Σχέση 6.1.

$$\theta_{pc} = \cos^{-1} \left( \frac{a_{pc}^2 + b_{pc}^2 - c_{pc}^2}{2a_{pc}b_{pc}} \right), \quad (6.1)$$

Feature name	Frames involved	Equation
Spatial angle	$F_2, F_1$	$\frac{\mathbf{v}_2^{(J)} - \mathbf{v}_1^{(J)}}{\ \mathbf{v}_2^{(J)} - \mathbf{v}_1^{(J)}\ }$
Spatial angle	$F_n, F_{n-1}$	$\frac{\mathbf{v}_n^{(J)} - \mathbf{v}_{n-1}^{(J)}}{\ \mathbf{v}_n^{(J)} - \mathbf{v}_{n-1}^{(J)}\ }$
Spatial angle	$F_n, F_1$	$\frac{\mathbf{v}_n^{(J)} - \mathbf{v}_1^{(J)}}{\ \mathbf{v}_n^{(J)} - \mathbf{v}_1^{(J)}\ }$
Total vector angle	$F_1, \dots, F_n$	$\sum_{i=1}^n \arccos \left( \frac{\mathbf{v}_i^{(J)} \cdot \mathbf{v}_{i-1}^{(J)}}{\ \mathbf{v}_i^{(J)}\  \ \mathbf{v}_{i-1}^{(J)}\ } \right)$
Squared total vector angle	$F_1, \dots, F_n$	$\sum_{i=1}^n \arccos \left( \frac{\mathbf{v}_i^{(J)} \cdot \mathbf{v}_{i-1}^{(J)}}{\ \mathbf{v}_i^{(J)}\  \ \mathbf{v}_{i-1}^{(J)}\ } \right)^2$
Total vector displacement	$F_n, F_1$	$\mathbf{v}_n^{(J)} - \mathbf{v}_1^{(J)}$
Total displacement	$F_1, \dots, F_n$	$\sum_{i=1}^n \mathbf{v}_i^{(J)} - \mathbf{v}_{i-1}^{(J)}$
Maximum displacement	$F_1, \dots, F_n$	$\max_n (\mathbf{v}_i^{(J)} - \mathbf{v}_{i-1}^{(J)})$
Bounding box diagonal length*	$F_1, \dots, F_n$	$\sqrt{a_{B(\mathcal{V}^{(J)})}^2 + b_{B(\mathcal{V}^{(J)})}^2}$
Bounding box angle*	$F_1, \dots, F_n$	$\arctan \frac{b_{B(\mathcal{V}^{(J)})}}{a_{B(\mathcal{V}^{(J)})}}$

**Πίνακας 6.1:** Χαρακτηριστικά που εξάγονται από τα σκελετικά δεδομένα

όπου

$$a_{pc} = \left( v_x^{(J)} - v_x^{(J_c)} \right)^2 + \left( v_y^{(J)} - v_y^{(J_c)} \right)^2, \quad (6.2)$$

$$b_{pc} = v_x^{(J)} \text{ και} \quad (6.3)$$

$$c_{pc} = \left( v_x^{(J_p)} \right)^2 + \left( v_y^{(J)} - v_y^{(J_p)} \right)^2. \quad (6.4)$$

Τέλος, μεταξύ των *HandLeft* (HL) και *HandRight* (HR) εξάγονται η μέγιστη

$$d_{\max} = \max_{i,j} \{ d(\mathbf{v}_i^{\text{HR}}, \mathbf{v}_j^{\text{HL}}) \}, \quad (6.5)$$

και η μέση απόσταση,

$$d_{\text{mean}} = \frac{1}{F^{(J)}} \sum_{i,j} d(\mathbf{v}_i^{\text{HR}}, \mathbf{v}_j^{\text{HL}}). \quad (6.6)$$

κατά την εκτέλεση της χειρονομίας. Η απόσταση  $d$  είναι η Ευκλείδεια απόσταση μεταξύ των σημείων και το  $F^{(J)}$  είναι ο αριθμός των καρέ που διαρκεί η χειρονομία. Η διάρκεια της χειρονομίας  $F^{(J)}$  περιλαμβάνεται επίσης στο διάλυσμα των χαρακτηριστικών.

Σύμφωνα με το [25] τα εξαγόμενα χαρακτηριστικά πρέπει να ορίζονται έτσι ώστε να ικανοποιούν τα παρακάτω κριτήρια: (α) πρέπει να υπολογίζονται σε σταθερό χρόνο  $O(1)$ ; (β) μικρές

αλλαγές στην εκτέλεση μιας χειρονομίας πρέπει να αντικατοπτρίζονται σε μικρές αλλαγές στο διάγραμμα των χαρακτηριστικών ( $\gamma$ ) τα χαρακτηριστικά πρέπει να υπάγονται σε κάποια σημασιολογία (γεωμετρική αναπαράσταση).

Όσον αφορά το ( $\alpha$ ), τα χαρακτηριστικά που προτείνουμε υπολογίζονται με χρήση αυξητικών ενημερώσεων, οι οποίες απαιτούν σταθερό χρόνο  $O(1)$  σε κάθε καρτέ. Τα κριτήρια ( $\beta$ ) και ( $\gamma$ ) ικανοποιούνται, καθώς τα χαρακτηριστικά βασίζονται κυρίως σε υπολογισμούς γωνιών και μετατοπίσεων, οπότε μικρές αλλαγές στην κίνηση των αρθρώσεων παράγουν μικρές μετατοπίσεις στο χώρο των χαρακτηριστικών. Επιπρόσθετα αυτά τα γεωμετρικά χαρακτηριστικά αποτελούν μια καλή σημασιολογική αναπαράσταση της κίνησης των αρθρώσεων ενός ανθρώπου όταν αυτός/η εκτελεί μια χειρονομία.

Το μέγεθος σε bytes των εξαγόμενων διανυσμάτων είναι εξαιρετικά μικρό, το οποίο οδηγεί σε αποδοτική χρήση του εύρους ζώνης. Ενδεικτικά ολόκληρο το dataset που κατασκευάσαμε περιέχει 820 σημεία και το μέγεθος του είναι 2.4MB σε μορφή συμπιεσμένου JSON (minified JSON). Οι παραπάνω ιδιότητες είναι καθοριστικές για την υλοποίηση μιας ρεαλιστικής μονάδας αναγνώρισης χειρονομιών σε πραγματικό χρόνο.

### 6.3 Αλγόριθμοι Μηχανικής Μάθησης

Τα χαρακτηριστικά που περιγράψαμε στην προηγούμενη ενότητα μπορούν να αποτελέσουν τη βάση για την εκπαίδευση ενός αλγορίθμου μηχανικής μάθησης. Η επιλογή του κατάλληλου αλγορίθμου για κάθε σύνολο δεδομένων μπορεί να υποβοηθηθεί από τη θεωρία αλλά τελικά κατά βάση μια πειραματική διαδικασία. Για το λόγο αυτό εκπαιδεύσαμε και αξιολογήσαμε ένα σύνολο ταξινομητών επιβλεπόμενης μάθησης που καλύπτουν τις βασικές κατηγορίες αλγορίθμων. Σε αυτή την ενότητα παρουσιάζουμε το απαραίτητο θεωρητικό υπόβαθρο για την κατανόηση της εσωτερικής λειτουργίας αυτών των αλγορίθμων, ενώ στην επόμενη θα προχωρήσουμε στην παρουσίαση των πειραματικών αποτελεσμάτων.

**K Κοντινότεροι Γείτονες (K Nearest Neighbors, KNN)** Ο αλγόριθμος KNN [27] είναι μία μη παραμετρική τεχνική ταξινόμησης. ο χαρακτηρισμός “μη παραμετρική τεχνική” υποδεικνύει ότι η εν λόγω τεχνική δε βασίζεται σε κάποιες *a priori* υποθέσεις για την κατανομή των δεδομένων. Ο KNN είναι ένας αλγόριθμος “οκνηρής μάθησης”. Αυτό σημαίνει ότι ο ταξινομητής δε δημιουργεί ένα γενικευμένο μοντέλο αναπαράστασης των δεδομένων στη φάση της εκπαίδευσης, αλλά χρησιμοποιεί απευθείας τα δεδομένα εκπαίδευσης για να εξάγει την απόφαση όταν παράγει προβλέψεις για νέα δεδομένα. Όταν ο ταξινομητής καλείται να προβλέψει την κλάση ενός νέου σημείου, εξετάζει τις κλάσεις των  $K$  κοντινότερων γειτόνων του και η τελική πρόβλεψη βασίζεται σε μια πλειοψηφική ψηφοφορία των γειτόνων. Λόγω της τοπικότητας του μοντέλου που παράγεται κατά τη διαδικασία της πρόβλεψης, ο KNN τείνει να έχει αρκετά καλή απόδοση σε μια μεγάλη γκάμα προβλημάτων, καθώς δεν παράγει ένα καθολικό στατικό μοντέλο στο οποίο βασίζονται όλες οι μελλοντικές προβλέψεις. Αυτή η τοπικότητα όμως κάνει τον KNN ευάλωτο στο θόρυβο που μπορεί να περιέχεται στο σύνολο

δεδομένων λόγω της εγγενούς διακύμανσης ή λόγω της μεθόδου που ακολουθήθηκε κατά τη μέτρηση των χαρακτηριστικών.

**Αφελής Μπεϋζιανός Ταξινομητής (Naive Bayes, NB)** Ο Naive Bayes [28] είναι ένας ακόμα απλός ταξινομητής. Σε αντίθεση με τον KNN, ο Naive Bayes είναι ένας αλγόριθμος “πρόθυμης μάθησης” και βασίζεται σε παραδοχές που αφορούν στην κατανομή των δεδομένων. Συγκεκριμένα οι προβλέψεις του αλγορίθμου βασίζονται στην ύστερη πιθανότητα των κλάσεων με βάση τα δεδομένα εκπαίδευσης, όπως περιγράφεται στην παρακάτω σχέση

$$NB(x, c) = \max_{c_i \in c} (P(c_i|x)) \quad (6.7)$$

Η ύστερη πιθανότητα  $P(c_i|x)$  υπολογίζεται από το γινόμενο της πιθανοφάνειας  $P(x|c_i)$  με την πρότερη πιθανότητα  $P(c_i)$  της κλάσης  $c_i$  όπως υποδεικνύεται από τον κανόνα του Bayes. Ο χαρακτηρισμός “αφελής”, υπάρχει λόγω της αφελούς υπόθεσης ότι τα χαρακτηριστικά είναι στατιστικώς ανεξάρτητα, και άρα η πιθανοφάνεια  $P$  για την κλάση  $c$  μπορεί να υπολογιστεί ως το γινόμενο των πιθανοφάνειών των επιμέρους χαρακτηριστικών

$$P(x|c) = \prod_{x_i \in x} (P(x_i|c)) \quad (6.8)$$

Οι επιμέρους πιθανοφάνειες υπολογίζονται για κάθε χαρακτηριστικό ανεξάρτητα με βάση μια προκαθορισμένη κατανομή, π.χ. πολυωνυμική κατανομή όταν τα χαρακτηριστικά παίρνουν διακριτές τιμές ή γκαουσιανή κατανομή όταν παίρνουν συνεχείς. Παρά την απλότητα του πιθανοτικού μοντέλου, ο αλγόριθμος Naive Bayes έχει ικανοποιητική απόδοση σε πολλά πρακτικά προβλήματα. Αυτό σε συνδυασμό με το γεγονός ότι ο αλγόριθμος δεν πάσχει από την “κατάρρα των μεγάλων διαστάσεων” (curse of dimensionality) και τους μικρούς χρόνους εκπαίδευσης καθώς και αποτίμησης των προβλέψεων κάνει τον αλγόριθμο κατάλληλη επιλογή σε αρκετές περιπτώσεις.

**Δέντρα Αποφάσεων (Decision Trees, DT)** Τα δέντρα αποφάσεων [29] είναι μη παραμετρικά μοντέλα πρόβλεψης τα οποία ταξινομούν τα παρατηρούμενα μοτίβα σε κλάσεις μέσω της υποβολής αλληλάλληλων ερωτήσεων (queries), όπου η απάντηση σε μια ερώτηση στην ακολουθία εξαρτάται από την απάντηση που δόθηκε στην προηγούμενη ερώτηση. Με αυτόν τον τρόπο κάθε ερώτηση παρέχει κάποια επιπλέον bits πληροφορίας που βοηθούν στην προβλεψη της κατηγορίας του προτύπου εισόδου. Τα δέντρα αποφάσεων είναι μοντέλα υψηλής απόκλισης (bias) - μικρής διακύμανσης (variance), το οποίο σημαίνει ότι η εκπαίδευση τους μπορεί να οδηγήσει σε υπερβολικά πολύπλοκα μοντέλα με περιορισμένη ικανότητα γενίκευσης. Στη συνέχεια θα παρουσιάσουμε τη διαδικασία εκπαίδευσης δυαδικών δέντρων αποφάσεων για την ταξινόμηση προτύπων σε δύο κλάσεις με τον αλγόριθμο CART. Ο CART κατασκευάζει τους κόμβους του τελικού δέντρου προσδιορίζοντας μια πλειάδα  $(\theta, t)$ , όπου  $\theta$  το χαρακτηριστικό στο οποίο θα γίνει η διχοτόμηση (split) και  $t$  το κατώφλι της διχοτόμησης. Πρότυπα στα οποία το χαρακτηριστικό  $\theta$  λαμβάνει τιμές μεγαλύτερες του  $t$  ταξινομούνται στο δεξί υποδέντρο, ενώ τα υπόλοιπα πρότυπα ταξινομούνται στο αριστερό υποδέντρο. Ο αλγόριθμος υιοθετεί μια “άπληστη” ευριστική μέθοδο για την εύρεση της πλειάδας  $(\theta, t)$ , προτιμώντας πλειάδες οι οποίες μεγιστοποιούν το κέρδος πληροφορίας σε κάθε κόμβο. Το κέρδος πληροφορίας διασφαλίζεται ελαχιστοποιώντας μια μετρική πρόσμειξης (impurity). Συνήθη παραδείγματα μετρικών πρόσμειξης αποτελούν η σχετική εντροπία των πιθανοτήτων κατανομών των κόμβων παιδιών και ο συντελεστής

Gini. Η επίδοση των εκπαιδευμένων δέντρων μπορεί να βελτιωθεί σημαντικά αν παρέχουμε στον αλγόριθμο εκπαίδευσης και μια συνθήκη πρόωρου τερματισμού, όπως το μέγιστο βάθος του δέντρου ή τον ελάχιστο αριθμό δειγμάτων που αντιστοιχούν σε κάθε φύλλο του δέντρου, με σκοπό τη μείωση της πολυπλοκότητας του εκπαιδευμένου μοντέλου και την αποφυγή της υπερεκπαίδευσης.

**Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines, SVM)** Οι μηχανές διανυσμάτων υποστήριξης [30] είναι μια κλάση ταξινομητών οι οποίοι στοχεύουν στην εύρεση του βέλτιστου υπερπιπέδου στον αλγεβρικό χώρο που σχηματίζεται από τα χαρακτηριστικά για το διαχωρισμό των δειγμάτων σε κλάσεις. Ως βέλτιστη επιφάνεια διαχωρισμού επιλέγεται η επιφάνεια που παρέχει το μέγιστο περιθώριο ανάμεσα στα δείγματα διαφορετικών κλάσεων. Παρόλο που τα SVM είναι γραμμικά μοντέλα, μπορούν να διαχειριστούν μη γραμμικά διαχωρίσιμα δεδομένα με τη χρήση συναρτήσεων που απεικονίζουν μη γραμμικά διαχωρίσιμα δεδομένα σε αλγεβρικούς χώρους σημαντικά μεγαλύτερων διαστάσεων όπου τα δείγματα θα είναι γραμμικά διαχωρίσιμα. Η απεικόνιση των δειγμάτων σε χώρους υψηλών διαστάσεων μπορεί να γίνεται είτε σαν τμήμα της προεπεξεργασίας είτε με τη χρήση του “κόλπου” των πυρήνων (kernel trick). Η πρώτη μέθοδος αποτελεί μια υπολογιστικά ακριβή, και στην περίπτωση που ο χώρος απεικόνισης είναι απείρων διαστάσεων μια υπολογιστικά αδύνατη διαδικασία. Η δεύτερη μέθοδος μπορεί να ενσωματωθεί σαν μέρος του αλγορίθμου, αντικαθιστώντας τον υπολογισμό κάθε εσωτερικού γινομένου με τον υπολογισμό μιας συνάρτησης πυρήνα  $k(x_i, x_j)$ . Αυτό οδηγεί στην έμμεση απεικόνιση σε περισσότερες διαστάσεις, ενώ η επιβάρυνση στην υπολογιστική πολυπλοκότητα εξαρτάται μόνο από τον υπολογισμό του πυρήνα. Οι πυρήνες που χρησιμοποιήθηκαν στην παρούσα εργασία είναι οι εξής

i. Γραμμικός Πυρήνας (Linear kernel)

$$k(x_i, x_j) = \langle x_i, x_j \rangle \quad (6.9)$$

ii. Γκαουσιανός Πυρήνας (RBF kernel)

$$k(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (6.10)$$

Στο εξής θα αναφέρουμε τις μηχανές διανυσμάτων υποστήριξης με γραμμικό πυρήνα ως LSVM ενώ αυτές με Γκαουσιανό πυρήνα ως RBFSVM.

**Τετραγωνική Διακριτική Ανάλυση (Quadratic Discriminant Analysis, QDA)** Η τετραγωνική διακριτική ανάλυση [31] είναι ένας Μπεϋζιανός ταξινομητής ο οποίος σχετίζεται στενά με τη γραμμική διακριτική ανάλυση. Το κριτήριο απόφασης στην QDA είναι ο λόγος των εκ των υστέρων πιθανοτήτων κατανομών, δηλαδή αποφασίζεται ότι το δείγμα  $X$  ανήκει στην κλάση  $i$  αν

$$\log\left(\frac{P(y=i|X)}{P(y=j|X)}\right) > 0 \xrightarrow{\text{Bayes rule}} \log\left(\frac{P(X|y=i)P(y=i)}{P(X|y=j)P(y=j)}\right) > 0. \quad (6.11)$$

Επίσης στην QDA γίνεται η υπόθεση ότι οι πρότερες κατανομές πυκνότητας πιθανότητας είναι κανονικές, δηλαδή

$$P(X|y=i) = \frac{1}{(2\pi)^n |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu_i)^T \Sigma_i^{-1} (X-\mu_i)}. \quad (6.12)$$

Ο υπολογισμός των πινάκων συνδιασποράς  $\Sigma_i$  μπορεί να γίνει με τη χρήση ενός συντελεστή κανονικοποίησης  $r$  ως

$$\Sigma_i = (1 - r)\Sigma_i + rI, 0 \leq r \leq 1, \quad (6.13)$$

όπου  $I$  ο μοναδιαίος πίνακας. Στην ειδική περίπτωση που οι πίνακες  $\Sigma_i$  είναι διαγώνιοι, τα χαρακτηριστικά θεωρούνται ανεξάρτητα και η QDA είναι ισοδύναμη με το Naive Bayes. Βλέπουμε επίσης ότι για αρκετά μεγάλες τιμές του συντελεστή  $r$  τόσο περισσότερο κυριαρχεί η διαγώνιος στον πίνακα συνδιασποράς, και άρα τόσο η QDA προσεγγίζει το Naive Bayes.

**Γραμμική Διακριτική Ανάλυση (Linear Discriminant Analysis, LDA)** Η γραμμική διακριτική ανάλυση [31] είναι μια Μπεϋζιανή στατιστική μέθοδος που στοχεύει στην εύρεση γραμμικών συνδυασμών από χαρακτηριστικά τα οποία διαχωρίζουν βέλτιστα τα δείγματα σε κλάσεις. Αυτό γεωμετρικά μεταφράζεται στην εύρεση της κατεύθυνσης ενός διανύσματος  $\vec{w}$  πάνω στο οποίο θα γίνει η προβολή των δειγμάτων  $\vec{x}$ , ώστε τα δείγματα να είναι διαχωρίσιμα σε κλάσεις από υπερεπίπεδα κάθετα στο διάνυσμα  $\vec{w}$ . Η LDA είναι μια ειδική περίπτωση της QDA, όπου

$$\forall(i, j) \Rightarrow \Sigma_i = \Sigma_j = \Sigma. \quad (6.14)$$

Με αυτήν την παραδοχή το κριτήριο απόφασης γίνεται ισοδύναμο με ένα εσωτερικό γινόμενο

$$\langle \vec{w}, \vec{x} \rangle > c, \quad (6.15)$$

όπου  $c$  μια σταθερά κατωφλίσωσης και

$$\vec{w} = \Sigma^{-1}(\vec{\mu}_i - \vec{\mu}_j). \quad (6.16)$$

Από τις παραπάνω εξισώσεις προκύπτει και η γεωμετρική ερμηνεία της LDA

**Προσαρμοστική Ενίσχυση (Adaboost)** Ο αλγόριθμος Adaboost [32] στοχεύει στον συνδυασμό πολλών αδύναμων ταξινομητών σε ένα σύνολο (ensemble) το οποίο αποτελεί έναν ισχυρό μέτα-ταξινομητή. Ως αδύναμος ορίζεται ένας ταξινομητής του οποίου η ακρίβεια είναι λίγο καλύτερη από την τυχαία επιλογή. Η ιδέα πίσω από την Adaboost είναι ότι ο μέτα-ταξινομητής θα έχει καλύτερη επίδοση από τους επιμέρους αδύναμους ταξινομητές, καθώς κάθε ένας από αυτούς θα παρέχει σωστές προβλέψεις (rule of thumb predictions) για υποσύνολα των δειγμάτων, ενώ ο συμψηφισμός τους θα παράγει ένα μοντέλο που μπορεί να κάνει προβλέψεις για ολόκληρο το σύνολο των δειγμάτων. Ο συμψηφισμός των ασθενών ταξινομητών γίνεται με την απόδοση βαρών σε κάθε έναν από αυτούς και η τελική πρόβλεψη γίνεται με ψηφοφορία, στην οποία η ψήφος κάθε αλγορίθμου έχει διαφορετικό βάρος. Ο αλγόριθμος κληρονομεί αυτή την ιδέα από τον Hedge( $\beta$ ), μια προγενέστερη μέθοδο boosting, με την διαφορά ότι οι παράμετροι που χρειάζονται για την εκπαίδευση του Hedge( $\beta$ ) αντί να δίνονται ως είσοδοι προσαρμόζονται με την σειριακή εκπαίδευση των ταξινομητών. Κατά τη σειριακή εκπαίδευση επιβραβεύονται περισσότερο οι ταξινομητές που δίνουν σωστές προβλέψεις για τα δείγματα τα οποία οι προηγούμενοι ταξινομητές απέτυχαν να ταξινομήσουν σωστά, διασφαλίζοντας ότι το σύνολο των ταξινομητών θα δίνει ακριβείς προβλέψεις για όλα, ή την πλειοψηφία των δειγμάτων. Αυτή η μέθοδος έχει το πλεονέκτημα ότι παράγει ισχυρούς ταξινομητές με πολύ καλή ακρίβεια, αλλά έχει το μειονέκτημα ότι είναι ευπαθής σε θόρυβο που πιθανώς υπάρχει στα δεδομένα εκπαίδευσης. Στην παρούσα εργασία χρησιμοποιήθηκε η Adaboost με Δέντρα Αποφάσεων (ABDT) και με Εξαιρετικά Τυχαία Δέντρα (ABET).



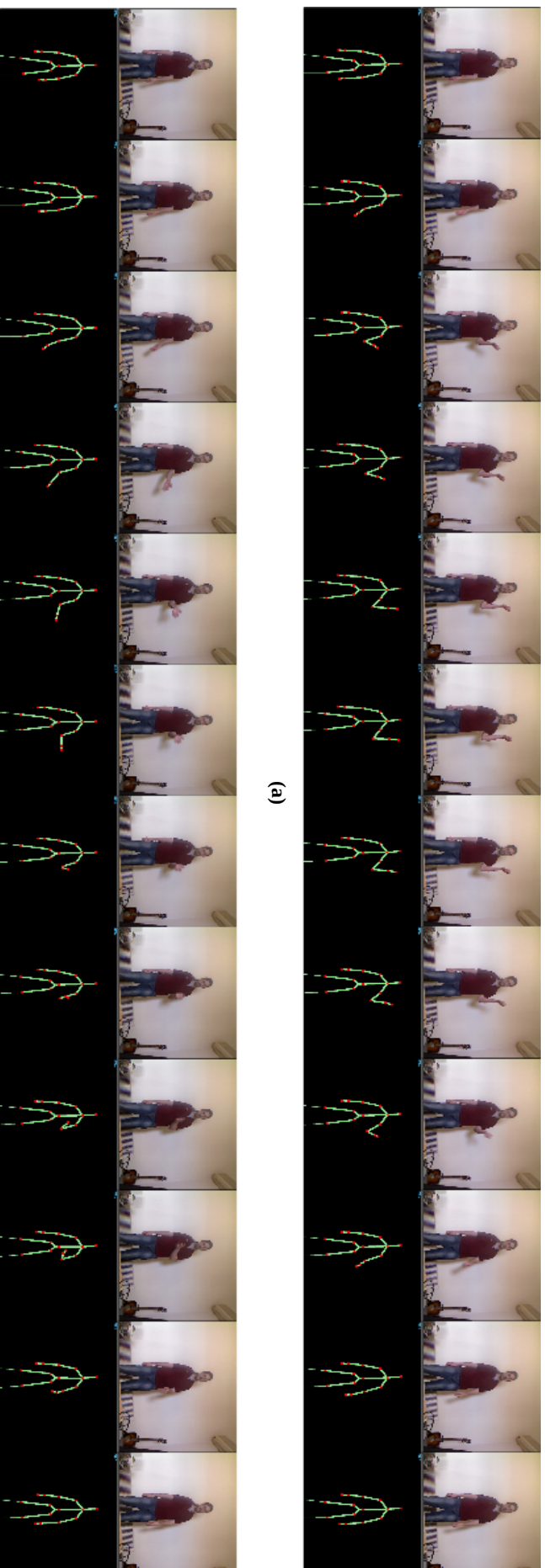
**Τυχαία Δάση Αποφάσεων (Random Forests, RF)** Τα τυχαία δάση αποφάσεων [33] είναι ένα σύνολο (ensemble) δέντρων αποφάσεων το οποίο δημιουργείται με την τεχνική “bagging” (Bootstrap Aggregating). Στην τεχνική του bagging παράγονται  $n$  διαφορετικά, τυχαία υποσύνολα δεδομένων εκπαίδευσης  $D_1, \dots, D_n$  από το σύνολο των δεδομένων εκπαίδευσης  $D$  χρησιμοποιώντας δειγματοληψία με επανατοποθέτηση. Έπειτα πάνω στα διαφορετικά υποσύνολα εκπαιδεύονται  $n$  ασθενείς ταξινομητές και η τελική πρόβλεψη λαμβάνεται με ισοβαρή ψηφοφορία των προβλέψεων των ταξινομητών. Στα Δάση Αποφάσεων οι ασθενείς ταξινομητές περιορίζονται σε Δέντρα Αποφάσεων και εισάγεται ένα επιπλέον επίπεδο τυχαιότητας, η “μέθοδος του Τυχαίου Υποχώρου”. Με αυτή τη μέθοδο, που εναλλακτικά ονομάζεται και bagging χαρακτηριστικών, τα δέντρα αποφάσεων εκπαιδεύονται σε διαφορετικά, τυχαία υποσύνολα του χώρου των χαρακτηριστικών με σκοπό τη μείωση της συσχέτισης ανάμεσα στα επιμέρους δέντρα. Η επιλογή τυχαίων υποσυνόλων δεδομένων εκπαίδευσης και χαρακτηριστικών για την εκπαίδευση των δέντρων στο σύνολο στοχεύει στην ελαχιστοποίηση της απόκλισης (bias) του συνολικού μοντέλου, ενώ η ομαλοποίηση που πραγματοποιείται με τον συμψηφισμό των δέντρων κάνει τον τελικό ταξινομητή μη ευπαθή στο θόρυβο στα δεδομένα εκπαίδευσης και στην υπερεκπαίδευση.

**Εξαιρετικά Τυχαία Δέντρα Αποφάσεων (Extra Trees, ET)** Τα εξαιρετικά τυχαία δέντρα αποφάσεων [34] είναι ακόμα σύνολο δέντρων αποφάσεων. Όμοια με τα Τυχαία Δάση χρησιμοποιούν τη μέθοδο του Τυχαίου Υποχώρου πραγματοποιώντας bagging χαρακτηριστικών. Σε αντίθεση όμως με τα Τυχαία Δάση τα δέντρα εκπαιδεύονται πάνω σε ολόκληρο το σύνολο δεδομένων εκπαίδευσης, χωρίς bagging, ενώ τα επιμέρους δέντρα εκπαιδεύονται με εντελώς τυχαίες πλειάδες διχοτόμησης  $(\theta, t)$ , αντί για τις τοπικά βέλτιστες πλειάδες που επιλέγονται στην εκπαίδευση των δέντρων για τα Τυχαία Δάση (πχ με τον αλγόριθμο CART). Τα εξαιρετικά τυχαία δέντρα οδηγούν σε ensemble μοντέλα με μικρότερη διακύμανση και μεγαλύτερη απόκλιση από τις υπόλοιπες tree ensemble τεχνικές, καθώς τα επιμέρους δέντρα τείνουν να είναι πιο πολύπλοκα, με μεγαλύτερο βάθος και περισσότερους κόμβους, όμως παρέχουν καλύτερους χρόνους εκπαίδευσης και σε πολλές περιπτώσεις βρίσκουν καλύτερη ισορροπία στο bias - variance tradeoff, συνεπώς εμπειρικά σε πολλές περιπτώσεις έχουν παρόμοια ή καλύτερη επίδοση από τις υπόλοιπες tree ensemble τεχνικές.

## 6.4 Πειραματική αξιολόγηση της μεθόδου

### 6.4.1 Κατασκευή του συνόλου δεδομένων

Για την αξιολόγηση της προεινόμενης μεθόδου κατασκευάσαμε ένα σύνολο δεδομένων. Στην κατασκευή του συμμετείχαν 10 χρήστες (7 άντρες και 3 γυναίκες) μεταξύ 22 και 36 χρονών. Επιλέξαμε ένα σύνολο από 8 χειρονομίες, συγκεκριμένα “swipes” (πάνω (up), κάτω (down), μέσα (in) και έξω (out)) για τα δύο χέρια. Στο Σχήμα 6.2 παρουσιάζονται οι ακολουθίες των έγχρωμων και των σκελετικών καρτέ για τις χειρονομίες “Swipe-Up” και “Swipe-In”, όπως αυτές εκτελούνται από ένα χρήστη με το δεξί χέρι. Οι χειρονομίες “Swipe-Down” και “Swipe-Out”, καθώς και οι χειρονομίες που εκτελούνται με το αριστερό χέρι είναι εύκολα αντιληπτές καθώς είναι συμμετρικές ως τις προαναφερθείσες. Επίσης αναφέρουμε ότι αυτά τα καρτέ έχουν εξαχθεί από παραδειγματικές λήψεις



**Σχήμα 6.2:** Συγχρονισμένα RGB και σκελετικά καρέ για τις χειρονομίες (α) *Swipe-Up* και (β) *Swipe-In*. Οι χειρονομίες εκτελούνται από τον ίδιο χρήστη

που παρουσιάστηκαν στους συμμετέχοντες για την κατανόηση της ορθής εκτέλεσης μιας χειρονομίας.

Ζητήθηκε από κάθε χρήστη να εκτελέσει κάθε χειρονομία τουλάχιστον 10 φορές και έπειτα ακολούθησε ο χειρωνακτικός καθαρισμός του συνόλου δεδομένων. Η διαδικασία αυτή παρήγαγε 820 αξιοποιήσιμες χειρονομίες. Για να αποφύγουμε τη μόλυνση του dataset με “θορυβώδη πρότυπα” πριν και έπειτα από κάθε χειρονομία και προκειμένου να αποφύγουμε την αυτόματη χρονική κατάτμηση του βίντεο, οι χρήστες εφοδιάστηκαν με ένα κουμπί εναλλαγής (toggle button), του οποίου διαδοχικά πατήματα σηματοδοτούν την αρχή και το τέλος κάθε χειρονομίας. Τα πατήματα του κουμπιού μεταδίδονται με τη χρήση ενός Arduino Yun μέσω MQTT στη μονάδα εξαγωγής χαρακτηριστικών και σηματοδοτούν την αρχή και το τέλος κάθε χειρονομίας. Στους χρήστες δόθηκε ένα περιθώριο 5 – 10 δοκιμών για να εξοικειωθούν με τη διαδικασία. Με αυτόν τον τρόπο απομονώσαμε τις χειρονομίες από τις υπόλοιπες κινήσεις. Για κάθε χειρονομία καταγράφηκαν τα καρέ RGB, βάθους και σκελετού και εξάγαμε τα σύνολα των χαρακτηριστικών που περιγράφονται στην Ενότητα 6.2 για τις κύριες αρθρώσεις που συμμετέχουν στις κινήσεις (αγκώνες (Elbows) και χέρια (Hands)).

#### 6.4.2 Πειράματα

Η πειραματική αξιολόγηση του συνόλου χαρακτηριστικών και των αλγορίθμων υπό εξέταση χωρίστηκε σε δύο φάσεις. Η πρώτη φάση πειραμάτων στοχεύει στη διερεύνηση της ακρίβειας αναγνώρισης για ένα γνωστό σύνολο από χρήστες. Συγκεκριμένα, υποθέτουμε ότι η αναγνώριση θα διεξαχθεί σε ένα σύνολο από χρήστες, οι οποίοι είναι γνωστοί εκ των προτέρων. Με αυτόν τον τρόπο μπορούμε να χρησιμοποιήσουμε δείγματα από όλους τους χρήστες για την κατασκευή του συνόλου εκπαίδευσης (training set) και του συνόλου αξιολόγησης (test set). Φυσικά ένα δείγμα δεν μπορεί να εμφανίζεται και στα δύο σύνολα. Η δεύτερη φάση πειραμάτων εστιάζει στη δοκιμή της μεθόδου σε ένα σύνολο από άγνωστους χρήστες, όπου απομονώνουμε κάποιους χρήστες εντελώς από το σύνολο εκπαίδευσης. Ο στόχος αυτού του πειράματος είναι να δείξει πόσο καλά γενικεύει ο αλγόριθμος αναγνώρισης σε χρήστες που δεν έχει ξανασυναντήσει.

Σε κάθε πείραμα, οι ταξινομητές εκπαιδεύτηκαν με τη χρήση ενός βέλτιστου σετ από παραμέτρους, το οποίο υπολογίστηκε με τη χρήση εξαντλητικής πλεγματικής αναζήτησης (exhaustive grid search). Στο πρώτο πείραμα οι αλγόριθμοι δοκιμάστηκαν με τη χρήση  $K$ -πλής αντεπικύρωσης ( $K$ -Fold cross validation,  $K$ -fold CV),  $K \in \{3, 5, 7, 10, 13, 15, 17, 20\}$ . Αυτή είναι μια καθιερωμένη στατιστική μέθοδος για την αξιολόγηση αλγορίθμων μηχανικής μάθησης, όπου το σύνολο δεδομένων χωρίζεται σε  $K$  ισομεγέθη υποσύνολα και το ένα από αυτά χρησιμοποιείται ως σύνολο επαλήθευσης, ενώ τα υπόλοιπα συγχωνεύονται και απαρτίζουν το σύνολο εκπαίδευσης. Ο αλγόριθμος εκπαιδεύεται με το σύνολο εκπαίδευσης και δοκιμάζεται στο σύνολο επαλήθευσης. Η διαδικασία αυτή επαναλαμβάνεται  $K$  φορές, ώστε κάθε υποσύνολο να χρησιμοποιηθεί μία φορά ως σύνολο επαλήθευσης. Η τελική ακρίβεια που παραθέτουμε είναι ο μέσος όρος των μετρήσεων για όλες τις εκτελέσεις. Για το πείραμά μας χρησιμοποιούμε μια παραλλαγή αυτής της μεθόδου, τη διαστρωματωμένη  $K$ -πλή αντεπικύρωση (stratified  $K$ -fold CV), όπου τα υποσύνολα δε χωρίζονται τυχαία, αλλά χρησιμοποιείται ως ευριστική για το χωρισμό των δειγμάτων σε υποσύνολα η διατήρηση περίπου ίσης της μέσης τιμής απόκρισης

(mean response value) σε κάθε υποσύνολο. Με αυτόν τον τρόπο διασφαλίζεται ότι κάθε υποσύνολο θα είναι αντιπροσωπευτικό για ολόκληρο το σύνολο δεδομένων. Παραθέτουμε τα αποτελέσματα στο Σχήμα 6.3, και τις βέλτιστες παραμέτρους για κάθε αλγόριθμό στον Πίνακα 6.2. Παρατηρούμε ότι ο αλγόριθμος ET αποδίδει καλύτερα από τους υπόλοιπους σε όλες τις περιπτώσεις ακολουθούμενος από τον RF. Μπορούμε να συμπεράνουμε λοιπόν ότι τα μοντέλα Tree Bagging μπορούν να διαχωρίσουν με μεγαλύτερη επιτυχία τα προτεινόμενα χαρακτηριστικά.

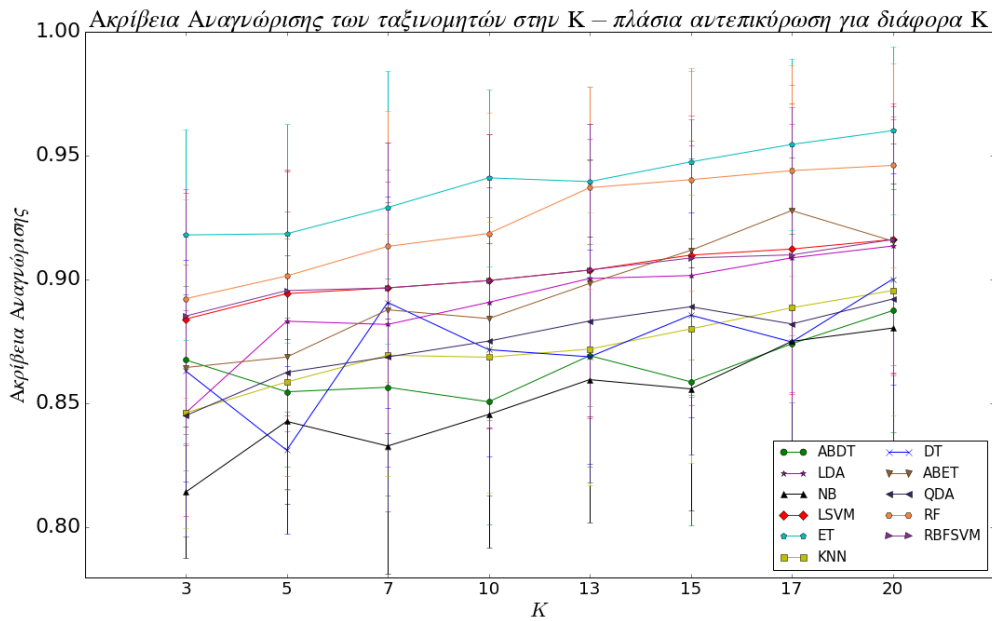
$\alpha$	learning rate	$n$	number of neighbors
$e$	number of estimators	$s$	search algorithm
$d$	maximum depth	$m$	metric between point $p$ and $q$
$f$	maximum number of features	$r$	regularization parameter

Classifier	Parameters
ABDT	$e = 103, \alpha = 621.6$
ABET	$e = 82, \alpha = 241.6$
DT	$d = 48, f = 49$
ET	$d = 17, f = 70, e = 96$
KNN	$n = 22, s = kd\_tree, m = \sum_{i=1}^n ( p_i - q_i )$
LSVM	$C = 0.0091$
QDA	$r = 0.88889$
RBFSVM	$C = 44.445, \gamma = 0.0001$
RF	$d = 27, f = 20, e = 75$

**Πίνακας 6.2:** Οι βέλτιστες παράμετροι για την εκπαίδευση κάθε ταξινομητή

Για τη δεύτερη σειρά πειραμάτων, επιλέγουμε να εστιάσουμε μόνο στον αλγόριθμο ET, καθώς διαφάνηκε ότι υπερέχει έναντι των υπόλοιπων από το πρώτο πείραμα. Στα πειράματα αυτά χωρίζουμε τα δείγματα με βάση τον χρήστη και κατασκευάζουμε σύνολα δεδομένων που αποτελούνται από 1 έως 9 χρήστες, ενώ οι υπόλοιποι χρήστες χρησιμοποιούνται για την κατασκευή του συνόλου επαλήθευσης. Καθώς δεν μπορούμε να χρησιμοποιήσουμε κάποια ευριστική για να επιλέγουμε τους χρήστες που θα συμμετάσχουν στο σύνολο εκπαίδευσης, κατασκευάζουμε διαφορετικά σύνολα εκπαίδευσης για όλους τους πιθανούς συνδυασμούς χρηστών και λαμβάνουμε το μέσο όρο και την τυπική απόκλιση από όλες τις μετρήσεις για την παρουσίαση του τελικού αποτελέσματος. Αυτό το πείραμα έχει δύο στόχους. Ο πρώτος είναι να δείξουμε τη δυνατότητα γενίκευσης του αλγορίθμου σε άγνωστους χρήστες. Ο δεύτερος είναι να βρούμε τον ελάχιστο αριθμό χρηστών που χρειάζεται να παρέχουν δείγματα για την εκπαίδευση ενός συστήματος που παράγει ικανοποιητικές προβλέψεις. Στον Πίνακα 6.3 παρέχουμε αναλυτικά την ακρίβεια για την οριακή περίπτωση όπου μόνο ένας χρήστης χρησιμοποιείται για το σύνολο επαλήθευσης (leave one out). Σε αυτό το πείραμα η μετρική μας είναι το  $F_1$  score, δηλαδή ο αρμονικός μέσος της ακρίβειας (precision)  $P$  και της ανάκλησης (recall)  $R$  που ορίζεται ως

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \quad (6.17)$$



**Σχήμα 6.3:** Αποτελέσματα της  $K$ -πλής αντεπικύρωσης ( $K$ -Fold CV) για όλους τους αλγορίθμους και για διάφορα  $K$

Η ακρίβεια ορίζεται ως

$$P = \frac{tp}{tp + fp}, \quad (6.18)$$

ενώ η ανάκληση ως

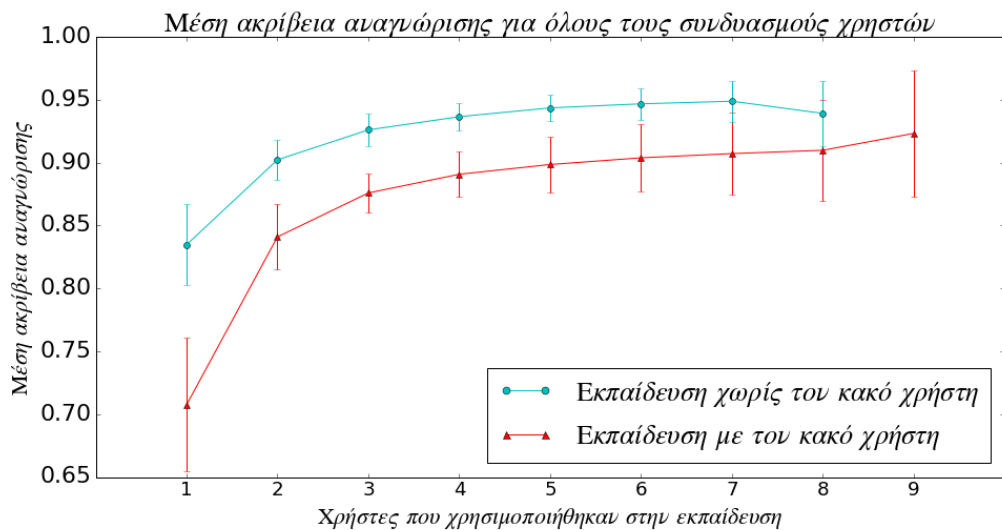
$$R = \frac{tp}{tp + fn}, \quad (6.19)$$

όπου  $tp$  είναι ο αριθμός των πραγματικά θετικών προβλέψεων,  $fp$  ο αριθμός των εσφαλμένων θετικών προβλέψεων και  $fn$  ο αριθμός των εσφαλμένων αρνητικών προβλέψεων.

Στο Σχήμα 6.4 βλέπουμε τη μέση ακρίβεια αναγνώρισης σαν συνάρτηση του αριθμού των χρηστών που περιέχονται στο σύνολο εκπαίδευσης. Από τα αποτελέσματα του Πίνακα 6.3 παρατηρήσαμε ότι ένας χρήστης είχε σημαντικά χειρότερα ποσοστά αναγνώρισης, οπότε επαναλάβαμε το πείραμα αποκλείοντας αυτό το χρήστη από το σύνολο δεδομένων. Από τα αποτελέσματα του πειράματος φαίνεται ότι για την εκπαίδευση μιας λειτουργικής μονάδας αναγνώρισης μπορούν να χρησιμοποιηθούν δείγματα ακόμα και από 3 – 4 χρήστες, χωρίς αυτό να έχει μεγάλη επίπτωση στην ακρίβεια των προβλέψεων. Τέλος στα Σχήματα 6.5 και 6.6 φαίνεται η ακρίβεια αναγνώρισης αυτού του πειράματος για κάθε χειρονομία. Αυτή η πληροφορία μπορεί να αξιοποιηθεί για τη βελτιστοποίηση της εμπειρίας χρήστη κατά το σχεδιασμό της Φυσικής Διεπαφής Χρήστη αναθέτοντας τις χειρονομίες με τη μεγαλύτερη ακρίβεια αναγνώρισης ως σήματα ελέγχου για τις πιο συχνά εκτελούμενες λειτουργίες.

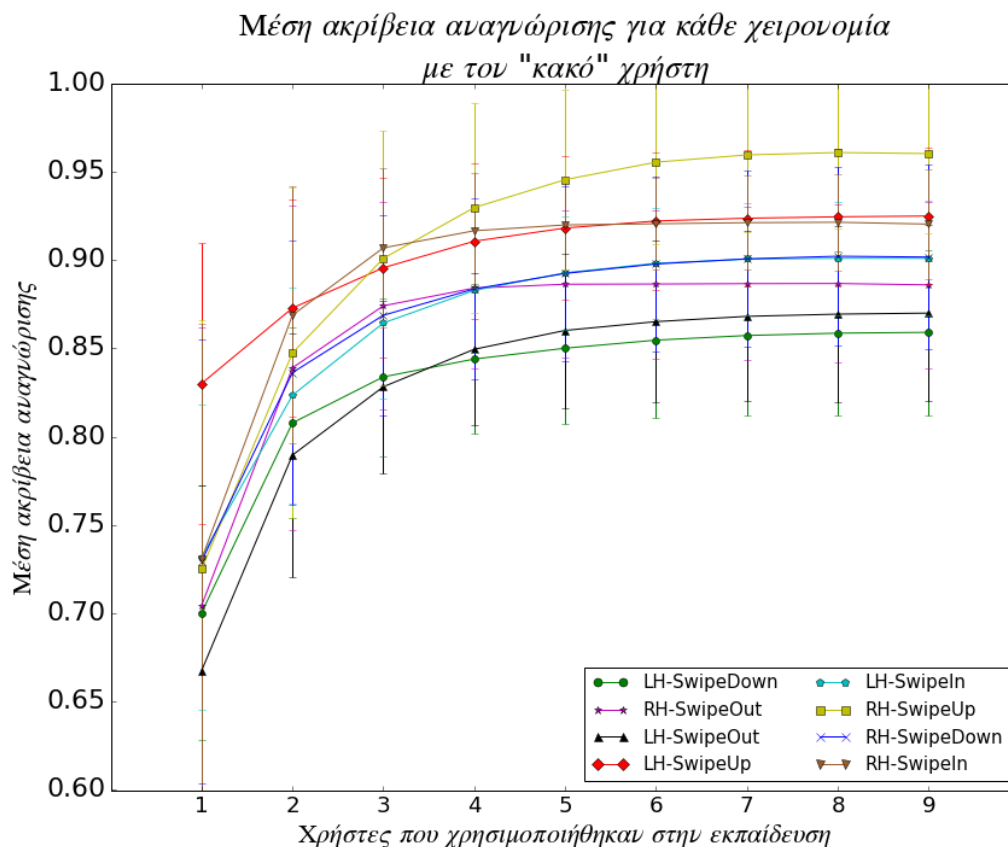
	User 1	User 2	User 3	User 4	User 5	User 6	User 7	User 8	User 9	User 10
LH-SwipeDown	0.76	0.83	1.00	0.82	1.00	0.80	1.00	1.00	1.00	0.96
LH-SwipeIn	0.38	0.92	0.84	1.00	1.00	0.92	1.00	1.00	1.00	1.00
LH-SwipeOut	0.61	0.93	0.86	1.00	1.00	0.89	1.00	1.00	0.97	1.00
LH-SwipeUp	0.69	0.90	1.00	0.84	1.00	0.83	1.00	1.00	0.97	0.96
RH-SwipeDown	0.78	1.00	0.95	-	1.00	1.00	0.92	1.00	0.87	1.00
RH-SwipeIn	0.64	1.00	0.67	-	1.00	1.00	1.00	1.00	0.89	0.96
RH-SwipeOut	0.61	1.00	0.80	-	1.00	1.00	0.95	1.00	1.00	0.95
RH-SwipeUp	0.40	1.00	0.95	-	1.00	1.00	1.00	1.00	0.96	1.00
Average	<b>0.62</b>	<b>0.94</b>	<b>0.88</b>	<b>0.92</b>	<b>1.00</b>	<b>0.92</b>	<b>0.99</b>	<b>1.00</b>	<b>0.96</b>	<b>0.97</b>

**Πίνακας 6.3:** Το σκορ  $F_1$  για κάθε χειρονομία χωριστά και το μέσο σκορ  $F_1$  για όλες τις χειρονομίες



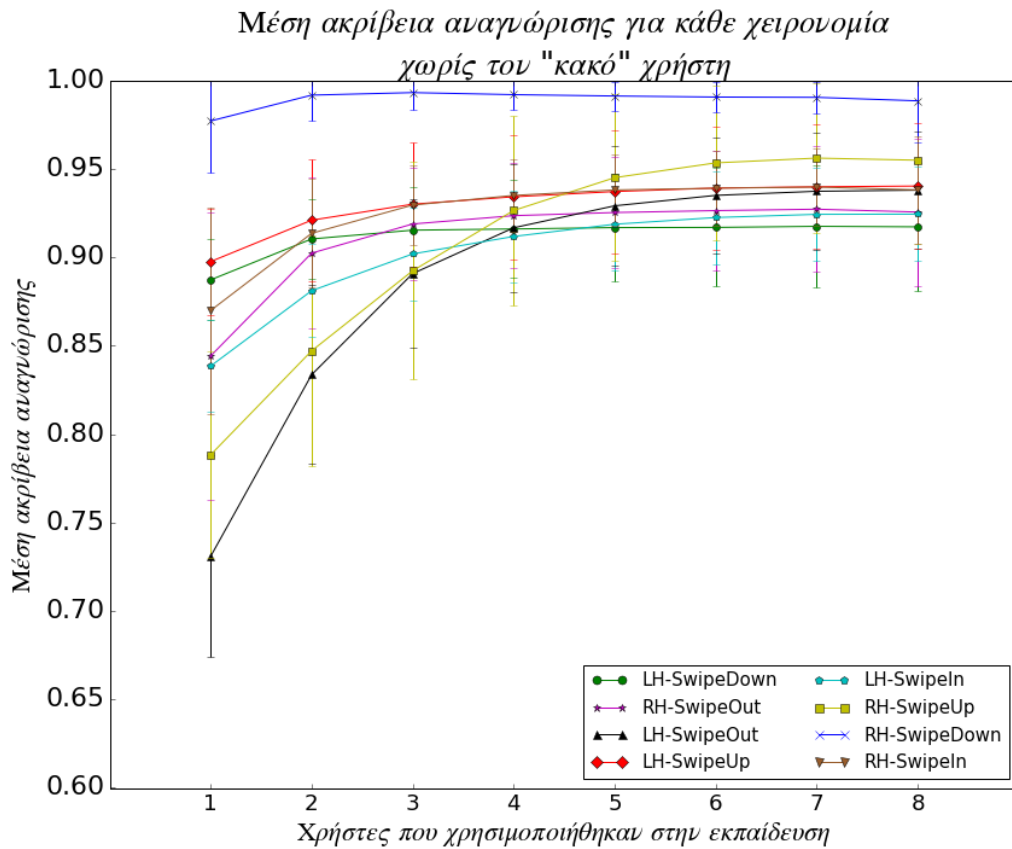
**Σχήμα 6.4:** Η μέση ακρίβεια των προβλέψεων του ταξινομητή ET με εκπαίδευση σε υποσύνολα χρηστών

Το τελικό πείραμα στοχεύει να αποδείξει τη βιωσιμότητα της χρήσης αυτής της προσέγγισης για την αναγνώριση χειρονομιών σε πραγματικό χρόνο και σε συστήματα με περιορισμένους πόρους. Για αυτό το σκοπό κατασκευάσαμε ένα benchmark, το οποίο μετράει το μέσο χρόνο που απαιτείται για την αναγνώριση ενός δείγματος από κάθε ταξινομητή. Φροντίσαμε ώστε στο σύνολο των δειγμάτων να περιλαμβάνονται δείγματα από κάθε χειρονομία και από διαφορετικούς χρήστες. Το μέγεθος του συνόλου εκπαίδευσης που χρησιμοποιείται για τους ταξινομητές είναι περίπου το 80% του συνολικού. Η πρόβλεψεις εκτελούνται 1000 φορές και ο τελικός χρόνος είναι ο μέσος χρόνος για κάθε εκτέλεση και κάθε χειρονομία. Το benchmark εκτελέστηκε σε δύο αρχιτεκτονικές, έναν ισχυρό επεξεργαστή Intel i7 3770 στα 3.4GHz με 8 πυρήνες και έναν ARMv7 rev 5 (v7l) με 4 πυρήνες, τον επεξεργαστή του Raspberry Pi 2. Τέλος επαναλάβαμε την τελευταία εκτέλεση στο Raspberry Pi 2, αλλά περιορίσαμε τον αριθμό των πυρήνων που μπορούν να χρησιμοποιηθούν από τους ταξινομητές σε 2 με χρήση

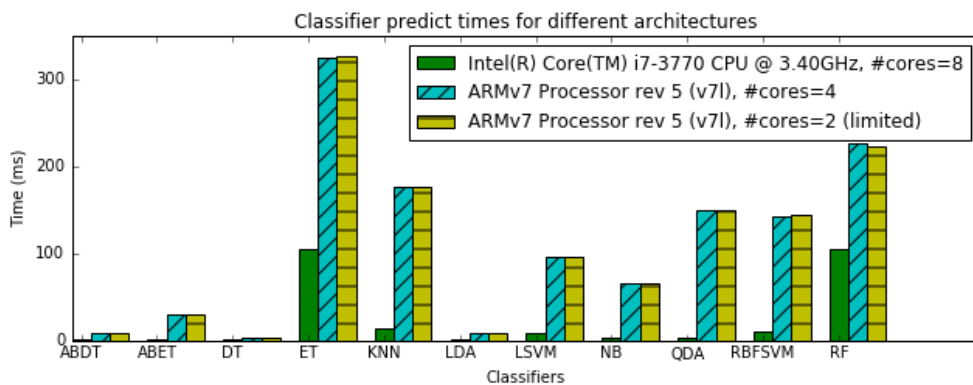


**Σχήμα 6.5:** Ακρίβεια ανά χειρονομία αναγνώρισης για την εκπαίδευση σε υποσύνολα των χρηστών

της εντολής *taskset* για να μελετήσουμε το χρόνο εκτέλεσης σε ένα ακόμα πιο περιορισμένο περιβάλλον. Στο Σχήμα 6.7 φαίνονται τα αποτελέσματα αυτών των πειραμάτων. Αρχικά παρατηρούμε ότι ο ταξινομητής ET που παράγει τις καλύτερες προβλέψεις είναι και ο πιο αργός, με μέσο χρόνο πρόβλεψης γύρω στα  $100ms$  στο ισχυρό σύστημα και στα  $300ms$  στο σύστημα περιορισμένων πόρων. Θεωρούμε ότι αυτή η καθυστέρηση είναι αποδεκτή σε ένα σύστημα πραγματικού χρόνου, όμως αν η εμπειρία χρήστη υποφέρει από αυτή την καθυστέρηση μπορούμε να συμβιβαστούμε με ελαφρώς χειρότερη ποιότητα προβλέψεων και να χρησιμοποιήσουμε τον LSVM ή ακόμα και την LDA όπου οι προβλέψεις πραγματοποιούνται πρακτικά στιγμιαία.



**Σχήμα 6.6:** Αποτελέσματα της εκπαίδευσης πάνω σε υποσύνολα των χρηστών χωρισμένα ανά χειρονομία χωρίς τον “κακό” χρήστη



**Σχήμα 6.7:** Μέσος χρόνος αναγνώρισης στον επεξεργαστή Intel i7 3770 ενός ισχυρού υπολογιστή και στο Raspberry Pi 2 για όλους τους ταξινομητές



## Κεφάλαιο 7

# Περίπτωση Χρήσης: Έλεγχος έξυπνου δωματίου

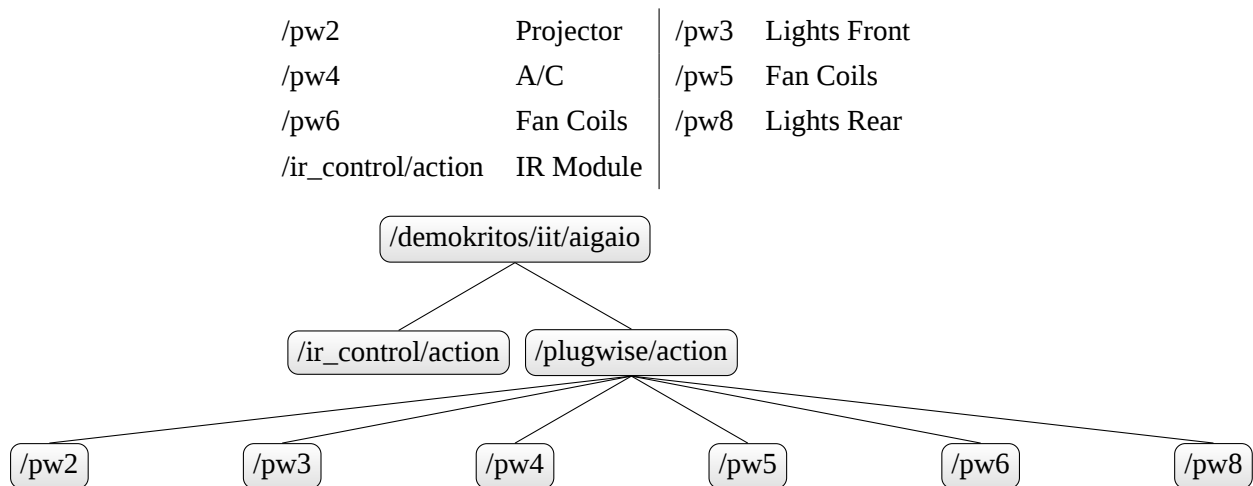
Σε αυτό το κεφάλαιο θα περιγράψουμε την αρχιτεκτονική του τελικού συστήματος και τη διαδικασία ενσωμάτωσης των προτεινόμενων μονάδων αναγνώρισης στάσεων του σώματος και χειρονομιών σε ένα έξυπνο δωμάτιο, συγκεκριμένα μια τυπική αίθουσα συνεδριάσεων που βρίσκεται στο Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών του Ε.Κ.Ε.Φ.Ε. Δημόκριτος.

### 7.1 Το έξυπνο δωμάτιο

Το δωμάτιο υπό μελέτη είναι η αίθουσα Αιγαίο στο Ε.Κ.Ε.Φ.Ε. Δημόκριτος. Στην αίθουσα αυτή έχει ήδη εγκατασταθεί μια πιλοτική εφαρμογή που παρέχει μηχανισμούς μέτρησης των περιβαλλοντικών συνθηκών και αυτοματισμούς για τη μετατροπή της αίθουσας σε ένα έξυπνο δωμάτιο. Η εφαρμογή αυτή εντάσσεται στο οικοσύστημα του ΔτΑ μέσα από την πλατφόρμα SYNAISTHISI και αποτελείται από ένα σύνολο λειτουργικών μονάδων τόσο στο επίπεδο του υλικού όσο και στο επίπεδο του λογισμικού. Όλες οι λειτουργικές μονάδες αντιμετωπίζονται σύμφωνα με τη λογική της Αρχιτεκτονικής Προσανατολισμένης σε Υπηρεσίες (SOA) ως υπηρεσίες, οι οποίες χωρίζονται σε αισθητήρες (υπηρεσίες τύπου S), ενεργοποιητές (υπηρεσίες τύπου A) και υπολογιστικές μονάδες (υπηρεσίες τύπου P). Οι υπηρεσίες συνδέονται και επικοινωνούν μέσω μιας υποδομής ανταλλαγής μηνυμάτων, η οποία βασίζεται στο πρωτόκολλο MQTT, η οποία παρέχεται από την πλατφόρμα SYNAISTHISI. Οι υπηρεσίες SPA συντονίζονται από μια κεντρική υπηρεσία λήψης αποφάσεων (DM) η οποία δέχεται μετρήσεις από τις υπηρεσίες τύπου S και επεξεργασμένα δεδομένα από τις υπηρεσίες τύπου P και με βάση αυτά ορίζει δράσεις για τις υπηρεσίες τύπου A. Οι δράσεις που ορίζονται καθορίζονται από ένα σύνολο αλγορίθμων τεχνητής νοημοσύνης, οι οποίοι στοχεύουν στην ελαχιστοποίηση της καταναλισκόμενης ενέργειας και την ταυτόχρονη βελτιστοποίηση της εμπειρίας του τελικού χρήστη.

Στο Σχήμα 7.1 παραθέτουμε την προγραμματιστική διεπαφή των υπηρεσιών τύπου A που έχουν εγκατασταθεί στο δωμάτιο. Η προγραμματιστική διεπαφή αποτελείται από ένα σύνολο από θέματα MQTT τα οποία αναπαρίστανται σε μια ιεραρχική δομή. Για την κατασκευή του πλήρους URI του θέματος ξεκινάμε από τη ρίζα του δέντρου και επισυνάπτουμε τη συμβολοσειρά του κόμβου μέχρι να φτάσουμε σε κάποιο φύλλο. Λόγω της πλήρους αποσύζευξης των επιμέρους υπηρεσιών που επιφέρει η χρήση του MQTT, η γνώση αυτής της απλής διεπαφής είναι αρκετή για να εντάξουμε τις

νέες υπηρεσίες στο δωμάτιο.



Σχήμα 7.1: Η ιεραρχία των θεμάτων στο MQTT API της έξυπνης αίθουσας

## 7.2 Αρχιτεκτονική του Συστήματος

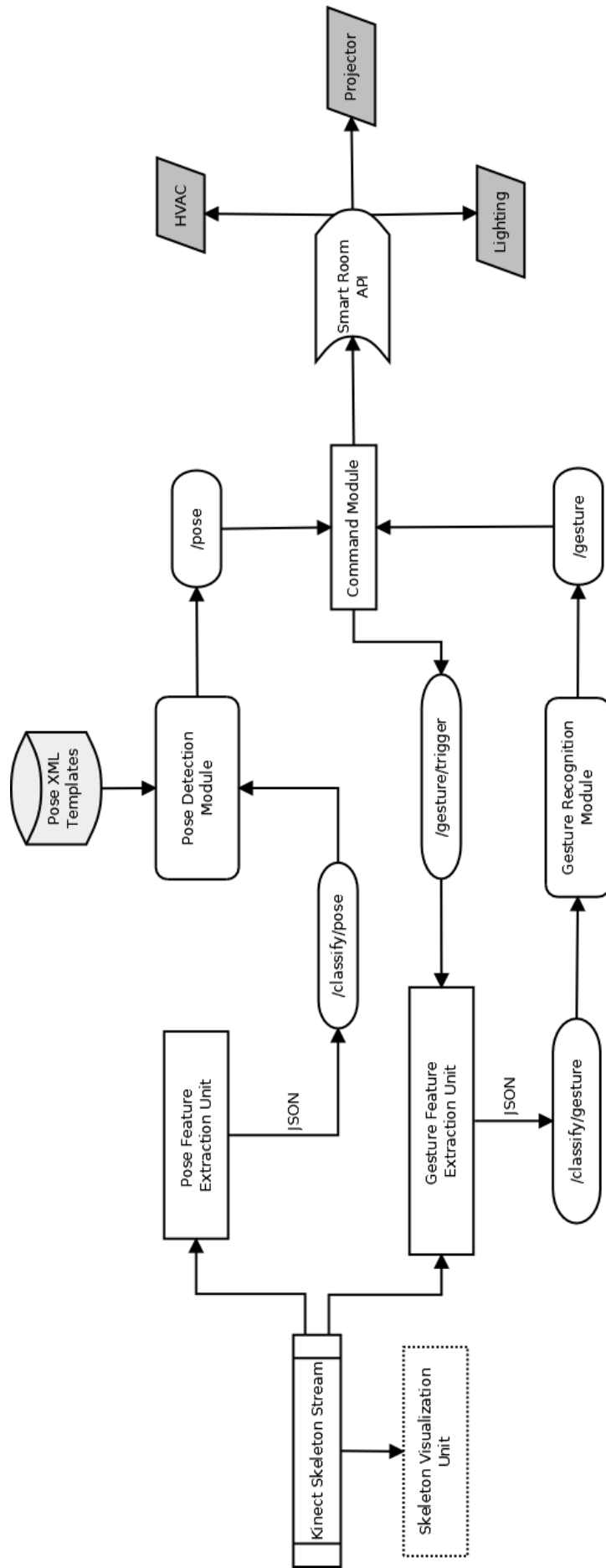
Η βασικές λειτουργικές μονάδες και διεπαφές των συστημάτων ανίχνευσης πόζας και αναγνώρισης χειρονομιών φαίνονται στο 7.2. Το σύστημα αυτό αποτελεί μια Φυσική Διεπαφή Χρήστη για την αίθουσα Αιγαίο, το Aigaio NUI. Η βασική δυνατότητα που προσφέρει το Aigaio NUI στον τελικό χρήστη είναι η δυνατότητα του απευθείας ελέγχου των συσκευών του δωματίου, παρακάμπτοντας την υπηρεσία λήψης αποφάσεων του δωματίου. Θα μπορούσε κάποιος να υποθέσει ότι ένα σύστημα αυτόματης αναγνώρισης φωνητικών εντολών θα είχε μικρότερη καμπύλη εκμάθησης και θα ήταν πιο φιλικό για τον τελικό χρήστη. Όμως πρέπει να σημειωθεί ότι η αίθουσα συνεδριάσεων έχει μεγάλα επίπεδα θορύβου και το σφάλμα αναγνώρισης φωνητικών εντολών θα είναι σημαντικό. Στη συνέχεια θα εξετάσουμε πιο αναλυτικά τα δομικά συστατικά του Aigaio NUI και θα τα διαχωρίσουμε σε υπηρεσίες.

**Μονάδα Εξαγωγής Σκελετού:** Αναλαμβάνει τη διαχείριση της διεπαφής με τον αισθητήρα Kinect και την εξόρυξη της σκελετικής πληροφορίας από τις ροές δεδομένων.

**Μονάδα Σχεδίασης Σκελετού:** Βοηθητική μονάδα για την οπτικοποίηση των αρθρώσεων σε πραγματικό χρόνο. Αναπτύχθηκε κυρίως για λόγους αποσφαλμάτωσης, αλλά μπορεί να ενσωματωθεί σε ένα γραφικό περιβάλλον το οποίο θα είναι διαθέσιμο στον τελικό χρήστη.

**Μονάδα Εξαγωγής Χαρακτηριστικών Πόζας:** Πραγματοποιεί την εξαγωγή των χαρακτηριστικών για την ανίχνευση πόζας, όπως περιγράφεται στην Ενότητα 5.2. Τα χαρακτηριστικά μετατρέπονται σε μορφή JSON και δημοσιεύονται σε ένα θέμα MQTT.

**Μονάδα Εξαγωγής Χαρακτηριστικών Χειρονομίας:** Κατασκευάζει το διάνυσμα χαρακτηριστικών της Ενότητας 6.2. Τα χαρακτηριστικά μετατρέπονται σε μορφή JSON και δημοσιεύονται σε ένα θέμα MQTT.



Σχήμα 7.2: Οι βασικές μονάδες λειτουργίας και διεπαφές για το σύστημα Aigaio NUI

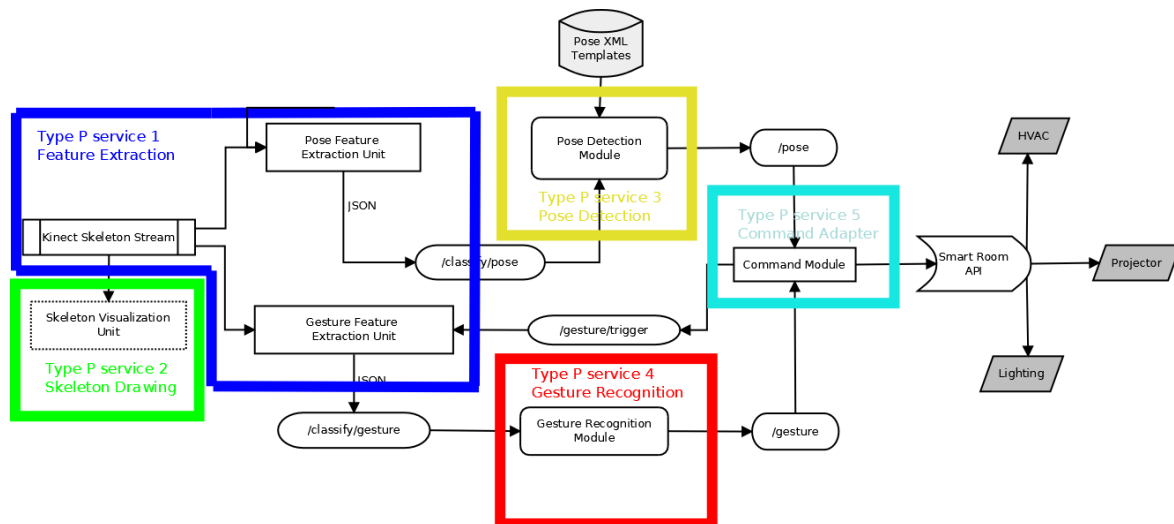
**Βάση Δεδομένων Πρότυπων XML Πόζας:** Αποθηκεύει τα πρότυπα XML για τις πόζες.

**Μονάδα Ανίχνευσης Πόζας:** Υλοποιεί τον αλγόριθμο ταιριάσματος προτύπων της Ενότητας 5.3. Δέχεται τα γεωμετρικά χαρακτηριστικά πόζας, δηλαδή τις γωνίες ανάμεσα στις αρθρώσεις σε JSON μορφοποίηση και δημοσιεύει την αναγνωρισμένη πόζα.

**Μονάδα Αναγνώρισης Χειρονομιών:** Πραγματοποιεί την αναγνώριση χειρονομιών με χρήση του ταξινομητή ET. Τα διανύσματα χαρακτηριστικών λαμβάνονται μέσω ενός θέματος MQTT σε JSON μορφοποίηση και δημοσιεύει την αναγνωρισμένη χειρονομία.

**Μονάδα Εντολών:** Η διεπαφή του Aigaio NUI με το έξυπνο δωμάτιο.

Το Aigaio NUI αποτελείται από 4 υπηρεσίες τύπου P, όπως αυτές περιγράφονται παρακάτω. Ο διαχωρισμός των μονάδων σε υπηρεσίες φαίνεται στο Σχήμα 7.3. Παρατηρούμε ότι οι περισσότερες υλοποιούνται ως αυτοτελείς υπηρεσίες, εκτός από τις μονάδες εξαγωγής σκελετού και χαρακτηριστικών, οι οποίες ομαδοποιούνται σε μια υπηρεσία τύπου S για να αποφευχθεί η μετάδοση ολόκληρης της σκελετικής πληροφορίας στις δύο μονάδες εξαγωγής χαρακτηριστικών μέσω του δικτύου.



**Σχήμα 7.3:** Οι υπομονάδες του Aigaio NUI χωρισμένες σε υπηρεσίες

## 7.2.1 Υπηρεσία Εντολών

Η υπηρεσία εντολών αποτελεί τη μοναδική διεπαφή των υπόλοιπων υπηρεσιών με το έξυπνο δωμάτιο και λειτουργεί ως adapter ανάμεσα στις εντολές (μοναδικοί αναγνωριστές των ανιχνευμένων πόζων και χειρονομιών) και στις δράσεις που εκτίθενται στο API του έξυπνου δωματίου.

Εκτός από αυτή τη λειτουργικότητα στην υπηρεσία εντολών υλοποιείται ακόμα μία βασική λειτουργία του κύκλου ζωής του Aigaio NUI: η πυροδότηση της καταγραφής των χαρακτηριστικών για την αναγνώριση χειρονομίας. Η αυτόματη χρονική κατάτμηση της ροής σκελετικών δεδομένων για την εύρεση των τμημάτων που περιέχουν κάποια χειρονομία είναι ένα δύσκολο ερευνητικό πρόβλημα, το οποίο τέθηκε εκτός της εμβέλειας αυτής της εργασίας. Αντί αυτού παρέχουμε μια απλή

εναλλακτική όπου κάποιες πόζες λειτουργούν ως ενεργοποιητές για τη μονάδα ανίχνευσης χαρακτηριστικών χειρονομίας. Με αυτόν τον τρόπο υλοποιούμε μία απλή και εύκολη στην εκμάθηση διεπαφή χρήστη, όπου οι πόζες ενεργοποίησης είναι οι πόζες στις οποίες ο χρήστης δείχνει προς την κατεύθυνση κάποιας συσκευής. Αυτή η συσκευή επιλέγεται ως η ενεργή συσκευή και ενεργοποιείται η εξαγωγή χαρακτηριστικών χειρονομίας. Στο χρήστη δίνεται ένα χρονικό περιθώριο να εκτελέσει τη χειρονομία, μετά από το οποίο η εξαγωγή χαρακτηριστικών σταματά. Η αναγνωρισμένη χειρονομία αποτελεί τη δράση την οποία θα εκτελέσει το έξυπνο δωμάτιο. Στο Σχήμα 7.4 φαίνεται το ακολουθιακό διάγραμμα που περιγράφει τη διαδικασία ενεργοποίησης της εξαγωγής χαρακτηριστικών χειρονομίας.

Αυτός ο σχεδιασμός της υπηρεσίας εντολών έχει πολλαπλά οφέλη:

- Για την προσαρμογή του Aigaiο NUI σε κάποιο άλλο δωμάτιο ή για την αλλαγή έκδοσης στο API του δωματίου αρκεί να αντικαταστήσουμε μόνο αυτή τη μονάδα, ενώ το υπόλοιπο σύστημα παραμένει αυτούσιο.
- Η αντικατάσταση της μεθόδου ενεργοποίησης της εξαγωγής γεωμετρικών χαρακτηριστικών με κάποια άλλη μέθοδο μπορεί να γίνει με την εισαγωγή νέων υπηρεσιών ή/και την τροποποίηση της μονάδας εντολών, χωρίς την τροποποίηση κάποιας άλλης υπηρεσίας, καθώς η υπηρεσία εξαγωγής χαρακτηριστικών χειρονομίας δέχεται τις εντολές εκκίνησης και παύσης μέσω MQTT.

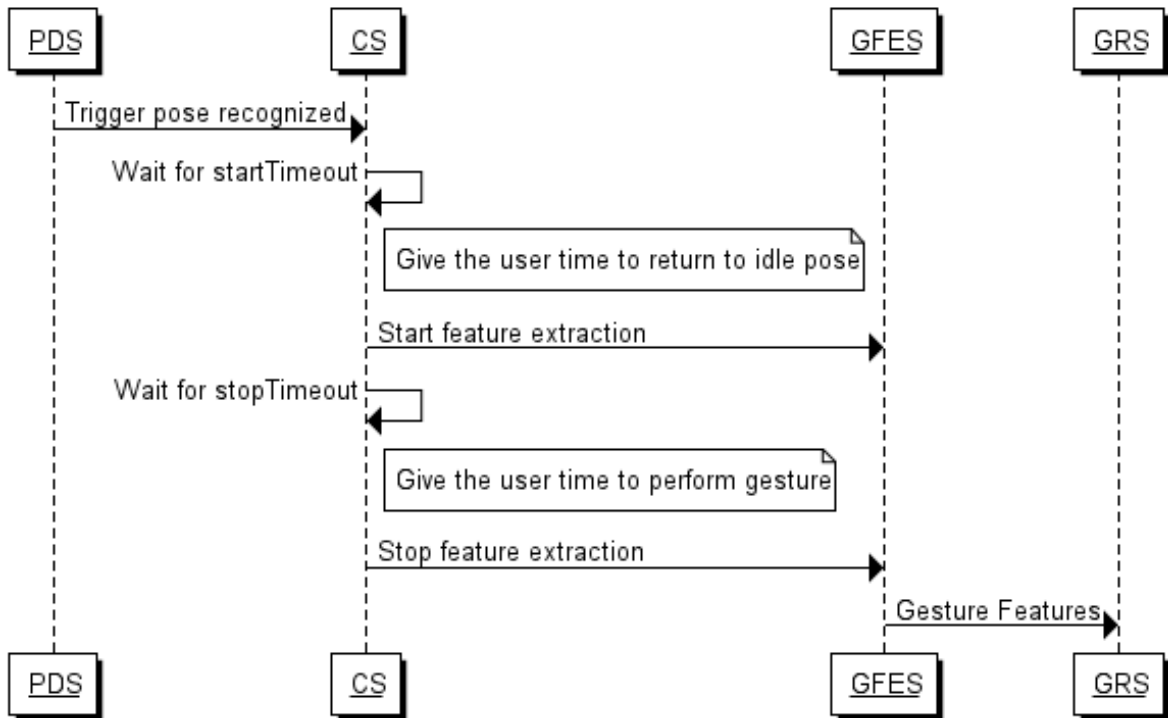
## 7.2.2 Λεπτομέρειες Υλοποίησης

Για την υλοποίηση των υπηρεσιών που περιγράψαμε παραπάνω έγινε χρήση μιας πληθώρας τεχνολογιών. Η αποσύζευξη των υπηρεσιών που εξασφαλίζεται από το πρωτόκολλο MQTT μας προσφέρει την ευελιξία να υλοποιήσουμε και να αποσφαλματώσουμε κάθε υπηρεσία σε απομόνωση και να εκμεταλλευτούμε τα χαρακτηριστικά και τη συμβατότητα βιβλιοθήκης διαφορετικών γλώσσων προγραμματισμού.

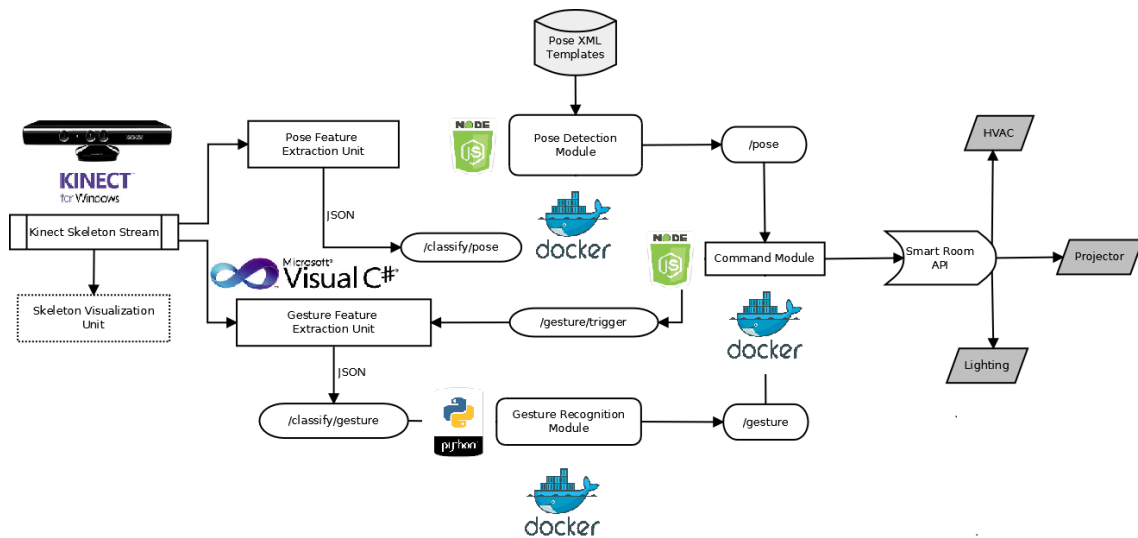
- Η *Υπηρεσία Εξαγωγής Χαρακτηριστικών* όπως και η *Υπηρεσία Σχεδίασης Σκελετού* υλοποιούνται σε C# σε μία native Windows εφαρμογή. Η χρήση της C# μας παρέχει τη δυνατότητα να χρησιμοποιήσουμε το Kinect SDK για την εξαγωγή των σκελετικών δεδομένων. Οι βασικές τεχνολογίες που χρησιμοποιήθηκαν φαίνονται στο Σχήμα 7.5.
- Η *Υπηρεσία Αναγνώρισης Χειρονομίας* υλοποιείται σε Python 2.7. Με αυτόν τον τρόπο είναι δυνατή η χρήση της βιβλιοθήκης scikit-learn [35], στην οποία παρέχονται αποδοτικές υλοποιήσεις μιας μεγάλης γκάμας αλγορίθμων μηχανικής μάθησης μέσω ενός απλού και αναλυτικά τεκμηριωμένου API. Ένα από τα κυριότερα χαρακτηριστικά αυτής της βιβλιοθήκης, είναι ότι επιτρέπει την εύκολη δοκιμή διαφορετικών αλγορίθμων και τεχνικών. Αυτό σε συνδυασμό με την εύκολη οπτικοποίηση των δεδομένων και των αποτελεσμάτων που παρέχει η βιβλιοθήκη matplotlib και τη χρήση του προγραμματιστικού περιβάλλοντος Ipython Notebook επιτρέπει τη διεκπεραίωση

- PDS Υπηρεσία Ανίχνευσης Πόζας
- CS Υπηρεσία Εντολών
- GFES Υπηρεσία Εξαγωγής Χαρακτηριστικών Χειρονομίας
- GRS Υπηρεσία Αναγνώρισης Χειρονομίας

### Gesture Recognition Trigger



Σχήμα 7.4: Ακολουθιακό διάγραμμα για την ενεργοποίηση της εξαγωγής χαρακτηριστικών χειρονομίας



Σχήμα 7.5: Οι υπομονάδες του Aigaiο NUI και οι αντίστοιχες τεχνολογίες

γρήγορων κύκλων ανάπτυξης (iterations) και εναλλακτικών πειραμάτων, το οποίο είναι καίριας σημασίας για την ανάπτυξη μιας εφαρμογής μηχανικής μάθησης. Η εγκατάσταση και εκτέλεση αυτής της υπηρεσίας γίνεται μέσα σε ένα Docker container, το οποίο παρέχει ένα ελεγχόμενο και σταθερό περιβάλλον εκτέλεσης και την αυτοματοποιημένη τοπική εγκατάσταση των απαιτούμενων πακέτων και βιβλιοθηκών.

- Η Υπηρεσία Ανίχνευσης Πόζας υλοποιείται σε node.js 6.2.2. Η εφαρμογή υλοποιείται με βάση το συναρτησιακό παράδειγμα προγραμματισμού, το οποίο επιβάλλει την αποφυγή μεταβολής της κατάστασης (state) του προγράμματος. Αυτό μας δίνει τη δυνατότητα να κάνουμε χρήση ταυτοχρονισμού (concurrency) με τη βιβλιοθήκη `async` σε λίγες γραμμές ευανάγνωστου και εύκολα συντηρήσιμου κώδικα. Και αυτή η υπηρεσία εκτελείται σε ένα Docker container.
- Η Υπηρεσία Εντολών υλοποιείται επίσης σε node.js 6.2.2. Η δυναμικότητα της γλώσσας επιτρέπει τη χρήση ενός δηλωτικού τρόπου για την αντιστοίχιση των εντολών με τα θέματα MQTT στο API του δωματίου. Και αυτή η υπηρεσία εκτελείται σε ένα Docker container.

### 7.2.3 Αξιολόγηση του Aigaiο NUI

Σε αυτή την ενότητα θα παρουσιάσουμε τα αποτελέσματα της αξιολόγησης των υπηρεσιών του Aigaiο NUI ως σύνολο με βάση την τελική εμπειρία χρήστη. Για αυτό το σκοπό έγιναν κάποια προκαταρκτικά πειράματα με ένα μικρό αριθμό χρηστών όπου ζητήθηκε από τους χρήστες να ελέγξουν τις “έξυπνες” συσκευές του δωματίου και μετρήσαμε τον αριθμό των επιτυχημένων προσπαθειών κατά τη διεκπεραίωση αυτών των εργασιών. Στο πείραμα συμμετείχαν 5 χρήστες και τους ζητήθηκε να εκτελέσουν τις παρακάτω δράσεις:

1. Άνοιγμα / Κλείσιμο μπροστινού φωτισμού
2. Άνοιγμα / Κλείσιμο πίσω φωτισμού
3. Άνοιγμα / Κλείσιμο κλιματιστικού
4. Άνοιγμα / Κλείσιμο ανεμιστήρα
5. Άνοιγμα / Κλείσιμο προβολέα

Σε αυτά τα πειράματα υλοποιήσαμε μία απλοϊκή διεπαφή χρήστη όπου χρησιμοποιούνται οι πόζες για την επιλογή συσκευών και την πυροδότηση της ανίχνευσης χειρονομίας και οι χειρονομίες για την επιλογή των δράσεων που θα εκτελέσουν οι συσκευές. Συγκεκριμένα για να στείλει ένας χρήστης μια εντολή σε μια συσκευή εκτελούνται τα παρακάτω βήματα:

1. Ο χρήστης δείχνει προς την κατεύθυνση της συσκευής.
2. Αυτό αναγνωρίζεται ως πόζα και πυροδοτεί την ανίχνευση χειρονομίας. Έχουμε υπολογίσει εκ’ των προτέρων τις πρότυπες πόζες για τις πόζες επιλογής συσκευών στο πεδίο λήψης του αισθητήρα Kinect που βρίσκεται στο δωμάτιο.

3. Ο χρήστης εκτελεί τη χειρονομία για την επιλογή της εκτελούμενης δράσης. Οι χειρονομίες α) LH-SwipeUp, β) RH-SwipeUp, γ) LH-SwipeIn και δ) RH-SwipeIn χρησιμοποιούνται για το άνοιγμα των συσκευών ενώ οι χειρονομίες ε) LH-SwipeDown, στ) RH-SwipeDown, ζ) LH-SwipeOut και η) RH-SwipeOut για το κλείσιμο των συσκευών.

Στον Πίνακα 7.1 φαίνονται τα αποτελέσματα της αξιολόγησης του Aigaiο NUI. Οι χρήστες εκτέλεσαν κάθε δράση 10 φορές και δείχνουμε το ποσοστό των επιτυχών δράσεων. Πρέπει να τονίσουμε ότι ως επιτυχής κρίνεται μια δράση όπου εκτελείται η επιθυμητή ενέργεια και ανιχνεύεται η σωστή χειρονομία. Θα δείξουμε τη σημασία αυτού του περιορισμού με ένα σενάριο. Ο χρήστης επιλέγει μια συσκευή και εκτελεί τη χειρονομία LH-SwipeOut. Το σύστημα έχει αναγνωρίσει λανθασμένα τη χειρονομία LH-SwipeDown. Σε αυτή την περίπτωση εκτελείται η σωστή λειτουργία (η συσκευή κλείνει) αλλά η δράση θεωρείται ανεπιτυχής.

	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5
Άνοιγμα μπροστινού φωτισμού	0.8	0.9	0.8	0.6	1.00
Κλείσιμο μπροστινού φωτισμού	0.7	0.8	0.9	0.7	0.8
Άνοιγμα πίσω φωτισμού	0.8	0.9	0.9	0.9	0.8
Κλείσιμο πίσω φωτισμού	1.00	0.6	0.9	0.5	0.7
Άνοιγμα κλιματιστικού	0.9	0.9	0.8	0.8	0.8
Κλείσιμο κλιματιστικού	0.8	0.7	0.9	0.7	0.9
Άνοιγμα ανεμιστήρα	0.7	0.8	0.8	0.8	0.9
Κλείσιμο ανεμιστήρα	0.7	0.9	1.00	0.8	0.7
Άνοιγμα προβολέα	0.7	1.00	0.8	0.7	0.8
Κλείσιμο προβολέα	0.8	0.9	0.9	0.8	0.8

**Πίνακας 7.1:** Αριθμός επιτυχημένων δράσεων ανά χρήστη κατά την αξιολόγηση του Aigaiο NUI

Σκοπεύουμε να επαναλάβουμε το παραπάνω πείραμα με μεγαλύτερο αριθμό και καλύτερη εκπαίδευση των χρηστών και με μία πιο σύνθετη διεπαφή χρήστη με περισσότερες επιλογές.



## Κεφάλαιο 8

### Συμπεράσματα και Προεκτάσεις

Σε αυτή την εργασία εξάγαμε δύο αποδοτικές μεθόδους για την ανίχνευση στάσεων του σώματος και χειρονομιών. Η ανίχνευση των στάσεων του σώματος βαζίζεται στη σύγκριση γεωμετρικών χαρακτηριστικών σε πρότυπες στάσεις με την παρατηρούμενη γεωμετρία των αρθρώσεων σε κάθε καρτέ. Για την αναγνώριση χειρονομιών εισάγαμε μια μέθοδο που στοχεύει στην αναγνώριση απλών χειρονομιών σε πραγματικό χρόνο. Προτείναμε ένα σύνολο χαρακτηριστικών που εξάγονται από τις 3Δ συντεταγμένες των αρθρώσεων και τα αξιολογήσαμε με τη χρήση ενός συνόλου αλγορίθμων μηχανικής μάθησης. Επίσης κατασκευάσαμε ένα νέο σύνολο δεδομένων που αποτελείται από δεδομένα χρώματος, βάθους και σκελετού για τις χειρονομίες. Η αξιολόγηση της προσέγγισής μας έγινε με βάση τις παραδοχές ότι ένα σύστημα οδηγούμενο από χειρονομίες μπορεί να χρησιμοποιηθεί από α) ένα περιορισμένο σύνολο από γνωστούς χρήστες και β) από ένα ανοικτό σύνολο από γνωστούς και άγνωστους χρήστες. τα πειραματικά αποτελέσματα δείχνουν ότι η προσέγγιση μας δίνει ικανοποιητικά αποτελέσματα και στις δύο περιπτώσεις. Επιπλέον διερευνήσαμε την επίδραση των “κακών χρηστών” και δείξαμε ότι απομονώνοντας αυτούς τους χρήστες από το σύνολο εκπαίδευσης μπορούμε να βελτιώσουμε σημαντικά την ακρίβεια αναγνώρισης. Τέλος, για να εξασφαλίσουμε ότι η αναγνώριση εκτελείται σε πραγματικό χρόνο, συγκρίναμε τους χρόνους πρόβλεψης εκπαιδευμένων ταξινομητών σε διαφορετικές αρχιτεκτονικές.

Οι παραπάνω αλγόριθμοι συνδυάστηκαν για την υλοποίηση μιας Φυσικής Διεπαφής Χρήστη για ένα έξυπνο δωμάτιο στα πλαίσια μια εφαρμογής ΔτΑ με στόχο την ενσωμάτωση αυτής της εφαρμογής στην πλατφόρμα SYNAISTHISI. Προτείναμε μια αρχιτεκτονική του συστήματος βασισμένη σε αυτοτελείς μικροϋπηρεσίες που επικοινωνούν μεταξύ τους μέσω μιας επικοινωνιακής υποδομής που βασίζεται στο πρωτόκολλο MQTT και εκτελούνται σε ένα ετερογενές υπολογιστικό περιβάλλον. Το τελικό σύστημα αξιολογήθηκε ποσοτικά και ποιοτικά από ένα μικρό σύνολο χρηστών οι οποίοι έκαναν χρήση της διεπαφής μας για να χειριστούν τις συσκευές του δωματίου. Τα αποτελέσματα της αξιολόγησης δείχνουν ότι το σύστημά μας μπορεί να χρησιμοποιηθεί σε πραγματικές συνθήκες.

Στο μέλλον θα εστιάσουμε στην επέκταση του συνόλου δεδομένων με την προσθήκη πιο σύνθετων χειρονομιών και περισσότερων δειγμάτων από τις υπάρχουσες εκτελούμενες από περισσότερους χρήστες. Επίσης στοχεύουμε να εφαρμόσουμε αυτούς τους αλγορίθμους για την αντιμετώπιση νέων, πιο σύνθετων προβλημάτων. Μια ακόμα πιθανή επέκταση είναι η υλοποίηση μιας δυναμικής μεθόδου για το διαχωρισμό των καρτέ του βίντεο που περιέχουν κάποια χειρονομία από τα υπόλοιπα.

Μια ακόμα πιθανή προέκταση είναι η μελέτη για τη μείωση του μεγέθους και του κόστους του απαιτούμενου εξοπλισμού. Τέλος σκοπεύουμε να διεξάγουμε μια εξαντλητική αξιολόγηση της διεπαφής που υλοποιήσαμε με περισσότερους χρήστες σε πιο σύνθετα σενάρια.

## Βιβλιογραφία

- [1] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, *et al.*, «Efficient human pose estimation from single depth images,» *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 12, pp. 2821–2840, 2013.
- [2] H. Koppula and A. Saxena, «Learning spatio-temporal structure from rgb-d videos for human activity detection and anticipation,» in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013, pp. 792–800.
- [3] M. Zanfir, M. Leordeanu, and C. Sminchisescu, «The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection,» in *Computer Vision (ICCV), 2013 IEEE International Conference on*, IEEE, 2013, pp. 2752–2759.
- [4] Y. Yang and D. Ramanan, «Articulated human detection with flexible mixtures of parts,» *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 12, pp. 2878–2890, 2013.
- [5] N. Neverova, C. Wolf, G. Paci, G. Somnavilla, G. W. Taylor, and F. Nebout, «A multi-scale approach to gesture detection and recognition,» in *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, IEEE, 2013, pp. 484–491.
- [6] —, «A multi-scale approach to gesture detection and recognition,» in *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, IEEE, 2013, pp. 484–491.
- [7] S. Bhattacharya, B. Czejdo, and N. Perez, «Gesture classification with machine learning using kinect sensor data,» in *Emerging Applications of Information Technology (EAIT), 2012 Third International Conference on*, IEEE, 2012, pp. 348–351.
- [8] Y. Song, D. Demirdjian, and R. Davis, «Continuous body and hand gesture recognition for natural human-computer interaction,» *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 1, p. 5, 2012.
- [9] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, «Continuous realtime gesture following and recognition,» in *Gesture in embodied communication and human-computer interaction*, Springer, 2010, pp. 73–84.
- [10] K. Lai, J. Konrad, and P. Ishwar, «A gesture-driven computer interface using kinect,» in *Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on*, IEEE, 2012, pp. 185–188.

- [11] C. Pereira and A. Aguiar, «Towards efficient mobile m2m communications: Survey and open challenges,» *Sensors (Basel)*, vol. 14, no. 10, pp. 19 582–19 608, 2014, 25333291[pmid], issn: 1424-8220. doi: 10.3390/s141019582. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4239907/>.
- [12] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, «Ros: An open-source robot operating system,» in *ICRA Workshop on Open Source Software*, 2009.
- [13] *Mqtt version 3.1.1*, OASIS, 2015. [Online]. Available: <http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html>.
- [14] A. S.-C. Urs Hunkeler Hong Linh Truong, «Mqtt-s — a publish/subscribe protocol for wireless sensor networks,» in *Communication Systems Software and Middleware and Workshops*, IEEE, 2008, pp. 791 –798.
- [15] Wikipedia. (2016). Kinect, [Online]. Available: <https://en.wikipedia.org/wiki/Kinect>.
- [16] Microsoft. (2016). Kinect for windows sensor components and specifications, [Online]. Available: <https://msdn.microsoft.com/en-us/library/jj131033.aspx>.
- [17] K. Khoshelham and S. O. Elberink, «Accuracy and resolution of kinect depth data for indoor mapping applications,» *Sensors (Basel)*, vol. 12, no. 2, pp. 1437–1454, 2012, sensors-12-01437[PII], issn: 1424-8220. doi: 10.3390/s120201437. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3304120/>.
- [18] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli, *Depth mapping using projected patterns*, US Patent 8,150,142, 2012. [Online]. Available: <https://www.google.gr/patents/US8150142>.
- [19] Z. Zalevsky, A. Shpunt, A. Maizels, and J. Garcia, *Method and system for object reconstruction*, WO Patent App. PCT/IL2006/000,335, 2007. [Online]. Available: <https://www.google.gr/patents/WO2007043036A1?cl=en>.
- [20] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, «Real-time human pose recognition in parts from a single depth image,» in *CVPR*, IEEE, 2011. [Online]. Available: <http://research.microsoft.com/apps/pubs/default.aspx?id=145347>.
- [21] Microsoft. (2016). Kinect for windows sdk, [Online]. Available: <https://msdn.microsoft.com/en-us/library/hh855347.aspx>.
- [22] —, (2016). Kinect hardware, [Online]. Available: <https://developer.microsoft.com/en-us/windows/kinect/hardware>.
- [23] —, (2016). Kinect for windows sdk 2, [Online]. Available: <https://msdn.microsoft.com/en-us/library/dn799271.aspx>.
- [24] J. Webb and J. Ashley, *Beginning Kinect Programming with the Microsoft Kinect SDK*, 1st. Berkely, CA, USA: Apress, 2012, isbn: 1430241047, 9781430241041.
- [25] J. Sheng, «A study of adaboost in 3d gesture recognition,» *Department of Computer Science, University of Toronto*, 2003.
- [26] D. Rubine, «Specifying gestures by example,» *ACM*, vol. 25, no. 4, 1991.

- [27] T. M. Cover and P. E. Hart, «Nearest neighbor pattern classification,» *Information Theory, IEEE Transactions on*, vol. 13, no. 1, pp. 21–27, 1967.
- [28] P. Domingos and M. Pazzani, «On the optimality of the simple bayesian classifier under zero-one loss,» *Machine learning*, vol. 29, no. 2-3, pp. 103–130, 1997.
- [29] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. CRC press, 1984.
- [30] V. N. Vapnik and V. Vapnik, *Statistical learning theory*. Wiley New York, 1998, vol. 1.
- [31] G. McLachlan, *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons, 2004, vol. 544.
- [32] Y. Freund and R. E. Schapire, «A decision-theoretic generalization of on-line learning and an application to boosting,» in *Computational learning theory*, Springer, 1995, pp. 23–37.
- [33] L. Breiman, «Random forests,» *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [34] P. Geurts, D. Ernst, and L. Wehenkel, «Extremely randomized trees,» *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.
- [35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, «Scikit-learn: Machine learning in Python,» *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.



## Παράρτημα Α

# Το XML Schema για τις πρότυπες πόζες

Κώδικας Α.1: Το XML schema για το πρότυπο XML μίας πόζας

```
<?xml version="1.0" encoding="utf-8"?>

<xs:schema id="PoseRules"
  attributeFormDefault="unqualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:PoseRules="urn:PoseRules"
  targetNamespace="urn:PoseRules"
  elementFormDefault="qualified">
  <xs:element name="Pose">
    <xs:complexType>
      <xs:sequence>
        <xs:element type="xs:string" name="DisplayName" minOccurs="1" maxOccurs="1" />
        <xs:element name="AngleConstraints" minOccurs="1" maxOccurs="1">
          <xs:complexType>
            <xs:all>
              <xs:element name="HipCenter" minOccurs="0" maxOccurs="0">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element type="xs:decimal" name="Desired" />
                    <xs:element type="xs:decimal" name="Deviation" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
              <xs:element name="Spine" minOccurs="0" maxOccurs="1">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element type="xs:decimal" name="Desired" />
                    <xs:element type="xs:decimal" name="Deviation" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
              <xs:element name="ShoulderCenter" minOccurs="0" maxOccurs="1">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element type="xs:decimal" name="Desired" />
                    <xs:element type="xs:decimal" name="Deviation" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
              <xs:element name="ShoulderLeft" minOccurs="0" maxOccurs="1">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element type="xs:decimal" name="Desired" />
                    <xs:element type="xs:decimal" name="Deviation" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
              <xs:element name="ElbowLeftAggregates" minOccurs="0" maxOccurs="1">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element type="xs:decimal" name="Desired" />
                    <xs:element type="xs:decimal" name="Deviation" />
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
              <xs:element name="WristLeft" minOccurs="0" maxOccurs="1">
                <xs:complexType>
```





```

    <xs:sequence>
      <xs:element type="xs:decimal" name="Desired" />
      <xs:element type="xs:decimal" name="Deviation" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="HipRight" minOccurs="0" maxOccurs="1">
  <xs:complexType>
    <xs:sequence>
      <xs:element type="xs:decimal" name="Desired" />
      <xs:element type="xs:decimal" name="Deviation" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="KneeRight" minOccurs="0" maxOccurs="1">
  <xs:complexType>
    <xs:sequence>
      <xs:element type="xs:decimal" name="Desired" />
      <xs:element type="xs:decimal" name="Deviation" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="AnkleRight" minOccurs="0" maxOccurs="1">
  <xs:complexType>
    <xs:sequence>
      <xs:element type="xs:decimal" name="Desired" />
      <xs:element type="xs:decimal" name="Deviation" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="FootRight" minOccurs="0" maxOccurs="1">
  <xs:complexType>
    <xs:sequence>
      <xs:element type="xs:decimal" name="Desired" />
      <xs:element type="xs:decimal" name="Deviation" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
</xs:all>
<xs:attribute type="xs:string" name="units" default="degrees" />
</xs:complexType>
</xs:element>
</xs:sequence>
<xs:attribute type="xs:string" name="id" use="required" />
</xs:complexType>
</xs:element>
</xs:schema>

```