



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

## Ποιοτική Ανάλυση και Αναγνώριση Χειρονομιών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μαυρουδέας Γεώργιος

**Επιβλέπων:** Ανδρέας Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π

Αθήνα, Ιούνιος 2017

---

---



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Ποιοτική Ανάλυση και Αναγνώριση Χειρονομιών**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μαυρουδέας Γεώργιος

**Επιβλέπων:** Ανδρέας Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή επιτροπή την .....

.....  
Ανδρέας Σταφυλοπάτης  
Καθηγητής Ε.Μ.Π  
Ε.Μ.Π

.....  
Κωνσταντίνα Νικήτα  
Καθηγητής Ε.Μ.Π

.....  
Γεώργιος Στάμου  
Επίκουρος Καθηγητής

Αθήνα, Ιούλιος 2017

---

.....

Μαυρουδέας Γεώργιος

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

Copyright © Μαυρουδέας Γεώργιος, 2017.

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 αποτελεσματική επικοινωνία ανθρώπου μηχανής μέσω φυσικών κινήσεων οδήγησε στη μελέτη των ανθρώπινων χειρονομιών στο πλαίσιο της χρησιμοποίησης τους για το σκοπό αυτό. Το επόμενο βήμα ήταν η εξαγωγή της γνώσης από τη μελέτη, στη δημιουργία συστημάτων ικανά να αναγνωρίζουν αυτόματα τις κινήσεις του ανθρώπου και να εκτελούν λειτουργίες με βάση αυτές. Σκοπός της παρούσας διπλωματικής εργασίας είναι, η δημιουργία ενός τέτοιου συστήματος, ικανό να προβαίνει στην αυτόματη αναγνώριση μιας συγκεκριμένης κλάσης χειρονομιών που βασίζονται σε κινήσεις των χεριών

Για τη διεκπαιραίωση αυτού του εγχειρήματος χρησιμοποιούνται τεχνικές που βασίζονται σε αυτό-οργανούμενους χάρτες (SOM), στις Μαρκοβιανές Αλυσίδες (Markov Chains) και στα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models). Τέλος η διπλωματική αγγίζει και το κομμάτι της εκφραστικότητας αναλύοντας τη θεωρία της και εξάγοντας στατιστικές παραμέτρους που σχετίζονται με αυτήν

### Λέξεις-Κλειδιά

Νευρωνικά δίκτυα, αυτο-οργανούμενοι χάρτες, επικοινωνία ανθρώπου μηχανής, αναγνώριση χειρονομιών, μαρκοβιανά μοντέλα, μαρκοβιανές αλυσίδες, μηχανική μάθηση, KMM, EAM,

---

## **Abstract**

The need for efficient human computer communication through physical movement lead to the study of human gestures for the purpose mention above. The next step as an aftereffect of that study was the utilization of that knowledge in order to built systems capable of identyfing human movements in an automatic manner and to operate based on them. The purpose of this diploma thesis is, the creation of a system with the ability of recognizing automatically a specific class of human gestures based on hand and arm movements.

To be able to accomplish the aformentioned task we utilize technics based on self organizing maps (SOM), Markov chains and Hidden Markov Models (HMM). Finally the following thesis covers and a part of expressivity, analyzing its theory and extracting several statistical measures rerated to it.

## **Key words**

Neural networks, self organizing maps, human computer interaction, gesture recognition, expressivity, markov models, markov chains, HMM, HCI, levenstein, machine learning

---



## Ευχαριστίες

Καταρχάς, θα ήθελα να ευχαριστήσω θερμά τον καθ.Σταφυλοπάτη Ανδρέα ο οποίος με εμπιστεύτηκε αναθέτοντάς μου την εκπόνηση της παρούσας διπλωματικής εργασίας και έτσι μου έδωσε τη δυνατότητα να ασχοληθώ με ένα πολύ ενδιαφέρον, για εμένα προσωπικά, επιστημονικό αντικείμενο, όπως είναι αυτό της Επιστήμης των Υπολογιστών και συγκεκριμένα το κομμάτι της Μηχανικής Μάθησης.

Ζωτικής σημασίας ήταν η βοήθεια των Γεωργίου Σιόλα και Γεωργίου Καρυδάκη τους οποίους και ευχαριστώ. Καθ' όλη τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας, με τις γνώσεις και την καθοδήγησή τους με βοήθησαν να την φέρω εις πέρας.

Δεν θα μπορούσα, επίσης, να μην ευχαριστήσω την οικογένειά μου και τους φίλους μου για την αγάπη και στήριξή τους κατά τη διάρκεια της φοίτησης μου στο ΕΜΠ

---



---

# Πίνακας Περιεχομένων

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Εισαγωγή</b>   | <b>6</b>  |
| <b>2</b> | <b>Το Πρόβλημα της Αναγνώρισης Χειρονομιών-Εκφραστικότητα</b>                           | <b>8</b>  |
| 2.1      | Γενικά . . . . .  | 8         |
| 2.2      | Χειρονομίες . . . . .   | 9         |
| 2.2.1    | Ύψος . . . . .  | 10        |
| 2.2.2    | Τεχνολογία εισόδου στο σύστημα . . . . .  | 12        |
| 2.2.3    | Χώρος Εφαρμογής . . . . .   | 15        |
| 2.2.4    | Έξοδος ή Απόκριση Συστήματος . . . . .  | 17        |
| 2.3      | Ανάλυση προβλήματος . . . . .   | 17        |
| 2.3.1    | Ανίχνευση . . . . .   | 18        |
| 2.3.2    | Παρακολούθηση . . . . .   | 20        |
| 2.3.3    | Αναγνώριση . . . . .  | 22        |
| 2.4      | Κορυφαίες Μέθοδοι(State of the Art) . . . . .   | 25        |
| 2.4.1    | Κρυφό Μαρκοβιανό Μοντέλο Εισόδου/Εξόδου<br>(Input/Output Hidden Markov Model) . . . . . | 26        |
| 2.4.2    | Εκπαίδευση . . . . .  | 27        |
| 2.4.3    | Αναγνώριση . . . . .  | 27        |
| 2.5      | Εκφραστικότητα . . . . .  | 28        |
| <b>3</b> | <b>Τεχνικές μηχανικής μάθησης</b>   | <b>32</b> |
| 3.1      | Γενικά . . . . .  | 32        |
| 3.2      | Αυτο Οργανούμενοι Χάρτες . . . . .  | 32        |
| 3.2.1    | Εισαγωγή . . . . .  | 32        |
| 3.2.2    | Ο Χάρτης . . . . .  | 34        |
| 3.2.3    | Διαδικασία Ανταγωνισμού . . . . .   | 35        |
| 3.2.4    | Διαδικασία Συνεργασίας . . . . .  | 36        |
| 3.2.5    | Διαδικασία Προσαρμογής . . . . .  | 40        |
| 3.2.6    | Οι Δύο φάσεις της Διαδικασίας Προσαρμογής:<br>Διάταξης και Σύγκλισης . . . . .          | 41        |
| 3.3      | Κρυφά Μαρκοβιανά Μοντέλα (KMM) . . . . .  | 43        |

|          |  |            |
|----------|--|------------|
| 3.3.1    | Διακριτές Διαδικασίες Μάρκοφ . . . . .   | 43         |
| 3.3.2    | Επέκταση της Θεωρίας στα Κρυφά Μαρκοβιανά Μοντέλα                                    | 44         |
| 3.3.3    | Στοιχεία του KMM . . . . .   | 45         |
| 3.3.4    | Τα τρία βασικά προβλήματα που χαρακτηρίζουν τα Κρυφά<br>Μαρκοβιανά μοντέλα . . . . . | 46         |
| 3.3.5    | Λύσεις των τριών προβλημάτων του KMM . . . . .                                       | 47         |
| <b>4</b> | <b>Το Σύστημα μας</b>  | <b>57</b>  |
| 4.1      | Το Σύνολο δεδομένων (Dataset) . . . . .  | 57         |
| 4.2      | Το Σύστημα Αναγνώρισης . . . . .   | 58         |
| 4.2.1    | Γενικά . . . . .   | 58         |
| 4.2.2    | Μοντελοποίηση των Χειρονομιών . . . . .  | 61         |
| 4.2.3    | Μοντέλο Θέσης . . . . .  | 62         |
| 4.2.4    | Μοντέλο κατεύθυνσης . . . . .  | 64         |
| 4.2.5    | Levenstein . . . . .   | 66         |
| 4.2.6    | Ταξινόμηση . . . . .   | 67         |
| <b>5</b> | <b>Αριθμητικά Αποτελέσματα</b>   | <b>73</b>  |
| 5.1      | Εισαγωγή . . . . .   | 73         |
| 5.2      | Γραφήματα Αναγνώρισης(Accuracy)<br>Ανά Κλάση . . . . .                               | 76         |
| 5.2.1    | Δεδομένα Συντεταγμένων . . . . .   | 76         |
| 5.2.2    | Δεδομένα αξελερόμετρων . . . . .   | 88         |
| 5.3      | Πίνακες Ταξινόμησης(Confusion Matrices) . . . . .                                    | 94         |
| 5.4      | Συμπεράσματα Αποτελεσμάτων . . . . .   | 98         |
| 5.5      | Αποτελέσματα Εκφραστικότητας . . . . .   | 98         |
| <b>6</b> | <b>Επίλογος</b>  | <b>100</b> |
| 6.1      | Συμπεράσματα . . . . .   | 100        |
| 6.2      | Μελλοντική Δουλειά . . . . .   | 102        |

# Κατάλογος σχημάτων

|      |  |    |
|------|--|----|
| 2.1  | Διαχωρισμός Χειρονομιών . . . . .  | 10 |
| 2.2  | Η κατανομή των χειρονομιών με βάση το στυλ τους . . . . .  | 12 |
| 2.3  | Ένα παιδί χρησιμοποιεί χειρονομίες μπροστά από γραφικό περιβάλλον αυτόματης αναγνώρισης χειρονομιών . . . . .  | 13 |
| 2.4  | Γάντια με αισθητήρες για μέτρηση παραμέτρων της κινήσεως του χεριού (αριστερά ) και κάμερα καταγραφής του χρήστη (δεξιά) .   | 14 |
| 2.5  | Διαχωρισμός Χειρονομιών ως προς την είσοδο τους . . . . .  | 15 |
| 2.6  | Το Ποσοστό των Χειρονομιών ανά μέρος του σώματος . . . . .   | 18 |
| 2.7  | Αναγνώριση Χειρονομιών στην Ιατρική . . . . .  | 20 |
| 2.8  | Μέρη της Αναγνώρισης Χειρονομιών . . . . .   | 22 |
| 2.9  | Κάμερα Kinect για εντοπισμό χειρονομιών σε εφαρμογές αναγνώρισης . . . . .   | 25 |
| 2.10 | Οπτική Αναπαράσταση του KMM Εισόδου/Εξόδου . . . . .   | 26 |
| 3.1  | Το προτότυπο μοντέλο του Konohen . . . . .   | 34 |
| 3.2  | Ένα διδιάστατο πλέγμα νευρώνων . . . . .   | 35 |
| 3.3  | Τοπογραφική Συνάρτηση Γειτονιάς Γκάους . . . . .   | 38 |
| 3.4  | Αλυσίδα Μαρκόφ 5 καταστάσεων . . . . .   | 43 |
| 3.5  | Απεικόνιση του Εμπρόσθιου Αλγορίθμου που βασίζεται στον δυναμικό προγραμματισμό . . . . .  | 50 |
| 3.6  | Παρουσίαση της ακολουθίας καταστάσεων που απαιτούνται για τον υπολογισμό του γεγονότος το σύστημα να είναι στην κατάσταση $S_i$ τον χρόνο $t$ και στην κατάσταση $S_j$ τον χρόνο $t + 1$ . . | 53 |
| 3.7  | Μοντέλο bakis . . . . .  | 56 |
| 4.1  | Οπτικοποίηση των Χειρονομιών . . . . .   | 60 |
| 4.2  | Μία αναπαράσταση της διαδικασίας μάθησης βασισμένη στις συντεταγμένες του χεριού . . . . .   | 65 |
| 4.3  | Η διαδικασία αξιολόγησης του συστήματος μας με δεδομένα τις συντεταγμένες του χεριού . . . . .   | 68 |
| 4.4  | Παρουσίαση του συστήματος μάθησης βασισμένο στα δεδομένα εισόδου από τα αξελερόμετρα . . . . .   | 70 |

|      |   |    |
|------|---|----|
| 5.1  | Πείραμα 1: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 76 |
| 5.2  | Πείραμα 2: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 77 |
| 5.3  | Πείραμα 3: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 78 |
| 5.4  | Πείραμα 4: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 79 |
| 5.5  | Πείραμα 5: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 80 |
| 5.6  | Πείραμα 6: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 81 |
| 5.7  | Πείραμα 7: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 82 |
| 5.8  | Πείραμα 1: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 83 |
| 5.9  | Πείραμα 2: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 84 |
| 5.10 | Πείραμα 3: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 85 |
| 5.11 | Πείραμα 4: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 86 |
| 5.12 | Μέση Τιμή Ποσοστού Αναγνώρισης των Μοντέλων . . . . .     | 87 |
| 5.13 | Πείραμα 1: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 88 |
| 5.14 | Πείραμα 2: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 89 |
| 5.15 | Πείραμα 1: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 90 |
| 5.16 | Πείραμα 2: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 91 |
| 5.17 | Πείραμα 3: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 92 |
| 5.18 | Πείραμα 4: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση . . . . | 93 |
| 5.19 | Μέση Τιμή Ποσοστού Αναγνώρισης των Μοντέλων . . . . .     | 94 |

# Κατάλογος Πινάκων

|     |  |    |
|-----|--|----|
| 4.1 | Πίνακας των Χειρονομιών που περιέχει το σύνολο δεδομένων . . | 59 |
| 5.1 | Πίνακας Παρουσίασης Πιθανοτήτων . . . . .                    | 74 |
| 5.2 | Πίνακας Συνδυασμών Πιθανοτήτων . . . . .                     | 75 |
| 5.3 | Πίνακας Ταξινόμησης 1 . . . . .                              | 95 |
| 5.4 | Πίνακας Ταξινόμησης 2 . . . . .                              | 97 |

# Κεφάλαιο 1

## Εισαγωγή

Το θέμα της παρούσας διπλωματικής είναι η αναγνώριση χειρονομιών και η εξαγωγή εκφραστικών παραμέτρων. Πρόκειται για δύο διαφορετικές κλάσεις προβλημάτων οι οποίες ωστόσο, συνδέονται μεταξύ τους σε μεγάλο βαθμό.

Οι χειρονομίες αποτελούνε μέσω έκφρασης και ως τέτοιο συνιστούν αναπόσπαστο κομμάτι στην επικοινωνία μεταξύ των ανθρώπων, είτε αυτόνομα, είτε συνοδεύοντας την ομιλία. Στόχος του επιστημονικού πεδίου της αναγνώρισης χειρονομιών είναι, σε πρώτο βήμα να καταφέρει να μοντελοποιήσει με αποτελεσματικό τρόπο τις χειρονομίες και στην συνέχεια να καταφέρει να βρει έναν τρόπο για την αυτόματη αναγνώριση τους από κάποιο υπολογιστικό σύστημα. Η μοντελοποίηση αυτή μπορεί να γίνει με πολλούς τρόπους και αποτελεί διεπιστημονικό ζήτημα. Για παράδειγμα ένας τρόπος, αν μιλάμε για χειρονομίες που περιλαμβάνουν κινήσεις ολόκληρου του χεριού, θα ήταν να τις μοντελοποιήσουμε ως ακολουθίες συντεταγμένων, δηλαδή με βάση την τροχιά τους (η μέθοδος αυτή μάλιστα χρησιμοποιείται στη συγκεκριμένη εργασία). Εφόσον πραγματοποιηθεί με επιτυχία το κομμάτι αυτό, στη συνέχεια καλείται να δημιουργήσει κάποια μέθοδο η οποία θα είναι ικανή, μέσω κάποιας διαδικασίας να προβεί στην αυτόματη αναγνώριση. Εδώ εμπλέκεται ο τομέας της Μηχανικής Μάθησης (Machine Learning), ο οποίος προσφέρει ένα πλήθος από αλγορίθμους για τη λύση του συγκεκριμένου ζητήματος.

Η αποδοτική λύση του προβλήματος της αυτόματης αναγνώρισης χειρονομιών αποτελεί τον προθάλαμο για την ανάπτυξη εφαρμογών που θα εκμεταλλεύονται την φυσική αλληλεπίδραση ανθρώπου μηχανής (Human Computer Interaction). Τέτοιες εφαρμογές έχουν αλλάξει εντελώς τον τρόπο που αντιλαμβάνομαστε την διεπαφή χρήστη, υπολογιστικού συστήματος και έχουν επιφέρει μεγάλες αλλαγές στην καθημερινότητα με την δημιουργία πληθώρας εφαρμογών.

Τέτοιες εφαρμογές αποτελούνε τα έξυπνα δωμάτια και οι έξυπνες συσκευές, τα οποία έχουν τη δυνατότητα ρύθμισης των παραμέτρων τους με την εκτέλεση χειρονομιών από το χρήστη, ο απομακρυσμένος έλεγχος μηχανών, η πλοηγηση σε εικονικά περιβάλλοντα όπου ο χρήστης έχει την δυνατότητα να εκτελεί χειρονομίες οι οποίες αντιγράφονται στο εικονικό περιβάλλον που βρίσκεται προσφέροντας του μία ρεαλιστική εμπειρία κτλ.

Η εξαγωγή των εκφραστικών χαρακτηριστικών μίας χειρονομίας σύμφωνα με κάποιο μοντέλο εκφραστικότητας, έχει ως σκοπό την αποδόμηση της στα καθαρά στατιστικά χαρακτηριστικά της με απώτερο στόχο την αυτόματη εύρεση του συναισθηματικού περιεχομένου της. Το αποτέλεσμα αυτό στη συνέχεια θα μπορούσε να χρησιμοποιηθεί για την αυτόματη πρόβλεψη της συναισθηματικής κατάστασης του εκτελεστή. Τα δεδομένα αυτά θα μπορούσαν να χρησιμοποιηθούν για την δημιουργία ρεαλιστικών χαρακτήρων (avatars) σε βινετοπαιχνίδια ή και πάλι σε έξυπνα δωμάτια για την αυτόματη ρύθμιση των παραμέτρων τους ανάλογα τη διάθεση του χρήστη.

Στη συνέχεια παρουσιάζεται ο τρόπος με τον οποίο διαρθρώνεται το κείμενο στην παρούσα εργασία:

1. **Κεφάλαιο 2:** Στο κεφάλαιο αυτό θα αναλύσουμε το πρόβλημα της αναγνώρισης χειρονομιών και της εκφραστικότητας, λαμβάνοντας υπ'όψιν όλες τις πτυχές τους. Αρχίζουμε περιγράφοντας αναλυτικά την έννοια των χειρονομιών, τις διάφορες τεχνικές που χρησιμοποιούνται ή έχουν χρησιμοποιηθεί για την αναγνώριση τους, τα στάδια της αναγνώρισης καθώς και μία από τις βέλτιστες μεθόδους για την επίλυση του προβλήματος. Τέλος περιγράφουμε την έννοια της εκφραστικότητας και το πως έχει μοντελοποιηθεί και αναλυθεί από διάφορους ερευνητές.
2. **Κεφάλαιο 3:** Γίνεται εκτενής θεωρητική ανάλυση των τεχνικών μηχανικής μάθησης που χρησιμοποιησάμε για την επίλυση του προβλήματος προς μελέτη.
3. **Κεφάλαιο 4:** Στο κομμάτι αυτό γίνεται αναλυτική περιγραφή του συστήματος μας. Αναλύοντας κάθε κομμάτι του χωριστά.
4. **Κεφάλαιο 5:** Παρουσιάζεται ένα κομμάτι των αριθμητικών αποτελεσμάτων που εξάγαμε και θεωρήσαμε ότι αποτελούν αντιπροσωπευτικό κομμάτι της μελέτης μας.
5. **Κεφάλαιο 6:** Παρουσιάζεται ένας σύντομος επίλογος της δουλειάς μας καθώς και πιθανή χρησιμοποίηση της μεθόδου για μελλοντικές εργασίες.



## Κεφάλαιο 2

# Το Πρόβλημα της Αναγνώρισης Χειρονομιών-Εκφραστικότητα

### 2.1 Γενικά

Η ραγδαία εξέλιξη της τεχνολογίας και η συνεχώς αυξανόμενη χρήση ευφυών συσκευών και συστημάτων από ανεξαρτήτου ηλικιάς, φύλου και υποβάθρου μονάδων σε ατομικό πλαίσιο αλλά και σε επίπεδο εταιριών, οργανισμών και οργανώσεων, αποτέλεσε υπομόχλιο για την διεξαγωγή έρευνας και μελέτης τεχνικών για την επικοινωνία ανθρώπου μηχανής. Συγκεκριμένα η σκέψη της επικοινωνίας με τα υπολογιστικά συστήματα με έναν πιο φυσικό αλλά και πιο προσητό τρόπο, είχε ως αποτέλεσμα εκτενείς ερευνητικές προσπάθειες στην ενσωμάτωση του προφορικού λόγου (φωνής) αλλά και των χειρονομιών σε αυτήν την αλληλεπίδραση. Συσκευές εισόδου όπως το συμβατικό πληκτρολόγιο (keyboard) και το ποντίκι (mouse), αν και αναπόσπαστα κομμάτια αυτής της διμερούς επικοινωνίας δεν είναι ικανά να προσφέρουν το κάτι παραπάνω που θα αλλάξει για πάντα τον τρόπο με τον οποίο αντιλαμβανόμαστε την τεχνολογία γύρω μας.

Λαμβάνοντας τα παραπάνω υπόψιν καταλαβαίνουμε πόσο σημαντική είναι η επίλυση του προβλήματος της αναγνώρισης χειρονομιών με τρόπο που να επιτρέπει τη χρησιμοποίησή της σε μία πληθώρα εφαρμογών με σκοπό πάνω από όλα την χάραξη ενός νέου τρόπου σκέψης, αντίληψης και έκφρασης στον τομέα της τεχνολογίας. Στη συνέχεια θα γίνει μια εκτεταμένη ανάλυση του προβλήματος και των θεωρητικών του μερών αρχίζοντας από τα βασικά μέρη του και καταλήγοντας στα πιο ειδικά κομμάτια του.

## 2.2 Χειρονομίες

Όπως γίνεται κατανοητό και από το όνομα του προβλήματος (**αναγνώριση χειρονομιών**), βασικό του κομμάτι είναι η χειρονομίες (gestures) τις οποίες θέλουμε να αναγνωρίσουμε. Για το λόγο αυτό η μελέτη τους και η κωδικοποίηση τους με αποτελεσματικό τρόπο είναι ζωτικής σημασίας για την κατανόηση του προβλήματος, αλλά ταυτόχρονα και για τη μοντελοποίηση του με αποδοτικό τρόπο. Σύμφωνα με την έρευνα που παρουσιάζεται στο [1] οι χειρονομίες είναι εκφραστικές με νόημα κινήσεις του σώματος συμπεριλαμβανομένων φυσικών κινήσεων, των δακτύλων, καρπών, χεριών, κεφαλιού ή σώματος με στόχο:

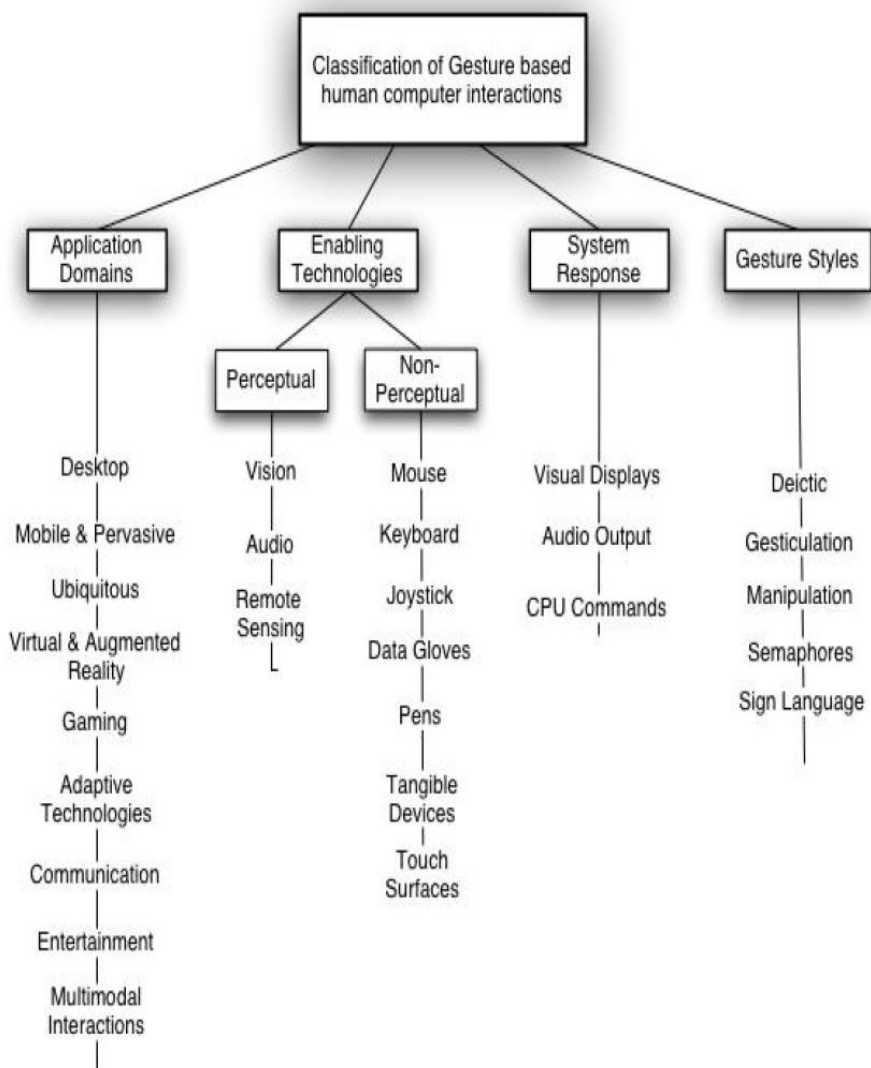
1. *Να επικοινωνήσουν σημαντική πληροφορία.*
2. *Να επιδράσουν στο περιβάλλον ή με το περιβάλλον.*

Μία χειρονομία μπορεί να ειδωθεί επίσης ως ένα μέσο συμπίεσης πληροφορίας με στόχο την μετάδοση της από τον πομπό με ελάχιστη προσπάθεια και την αποκωδικοποίηση της από τον δέκτη.

Στα πλαίσια του υπολογιστικού περιβάλλοντος χρήστη (computer interfaces) σύμφωνα και με το [2], οι χειρονομίες μπορούν να διαχωριστούν σε δύο κατηγορίες:

- **Ασύγχρονες (Offline) Χειρονομίες:** Οι χειρονομίες που επεξεργάζονται αφού ο χρήστης (user) έχει αλληλεπιδράσει με το αντικείμενο, όπως για παράδειγμα χειρονομίες για την ενεργοποίηση κάποιου καταλόγου.
- **Σύγχρονες (online) Χειρονομίες:** Πρόκειται για χειρονομίες που ανιχνεύονται κατά τη διάρκεια της κίνησης του χρήστη συνοδεύοντας τον σε πραγματικό χρόνο. Για παράδειγμα χειρονομίες για μεγέθυνση κάποιου γραφικού περιβάλλοντος, περιστροφή κάποιας εικόνας σε συγκεκριμένες μοίρες σύμφωνα με την προτίμηση του χρήστη κτλ.

Επιπροσθέτως αφού έχουμε διαχωρίσει τις χειρονομίες σε σύγχρονες και ασύγχρονες στο [3] προτείνεται ο διαχωρισμός τους σε τέσσερις κατηγορίες με βάση το ύφος τους (style) τον τομέα στον οποίο εφαρμόζονται (application domain), την τεχνολογία εισόδου στο σύστημα (input) και την απόκριση του συστήματος (output).



Σχήμα 2.1: Διαχωρισμός Χειρονομιών [3]

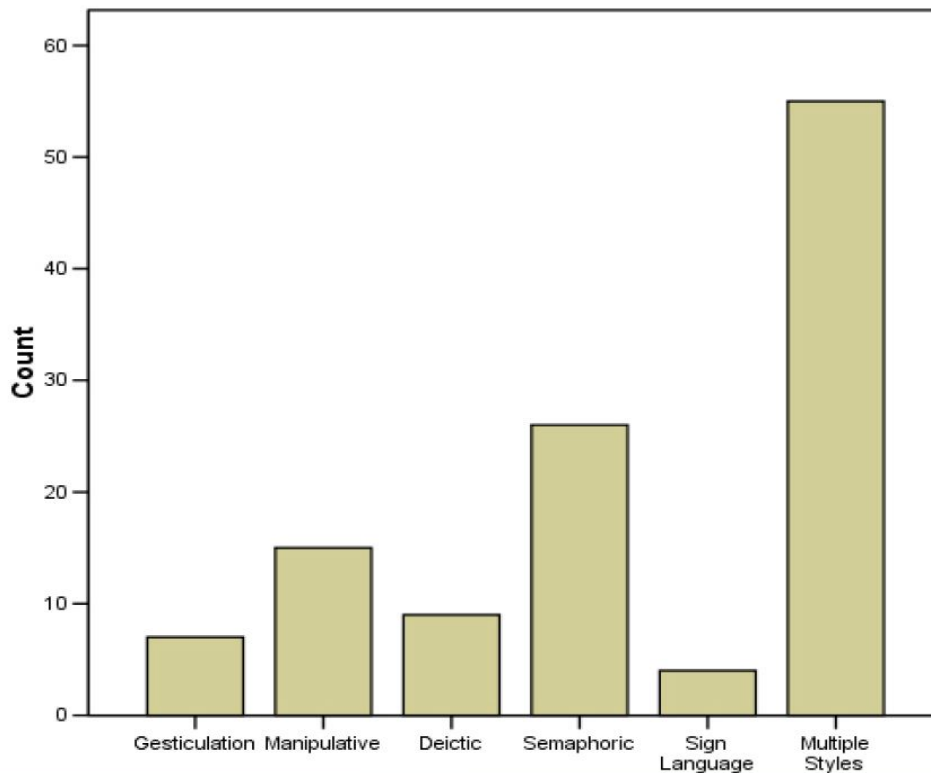
### 2.2.1 Ύφος

Με βάση το ύφος της χειρονομίας σύμφωνα με τους ορισμούς που έχουν δοθεί στα [4], [5],[6], μπορούμε να τις διαχωρίσουμε σε πέντε κατηγορίες:

- **Δεικτικές (Deictic Gestures):** Οι δεικτικές χειρονομίες, όπως φανερώνει και το όνομα τους, περιλαμβάνουν την υπόδειξη, με στόχο την ανάδειξη της ταυτότητας ενός αντικειμένου ή την χωρική του τοποθεσία μέσα στο περιβάλλον στο οποίο βρίσκεται. Για παράδειγμα όταν θέλουμε να φανερώσουμε την τοποθεσία ενός αντικειμένου ή να υποδείξουμε σε

ποιο άτομο αναφερόμαστε, συνήθως προτάσουμε ένα από τα δύο χέρια μας κάθετα στον κορμό και διατείνουμε τον δείκτη της παλάμης προς το σημείο ενδιαφέροντος.

- **Χειρονομίες χειρισμού (Manipulative Gestures):** Πρόκειται για χειρονομίες με σκοπό τον έλεγχο κάποιας οντότητας (στον πραγματικό ή στον εικονικό κόσμο), εγκαθιδρύοντας μία στενή σχέση μεταξύ των πραγματικών κινήσεων του μέλους που πραγματοποιεί την χειρονομία (παλάμη /χέρι) με την οντότητα εκείνη η οποία είναι αντικείμενο του χειρισμού. Για παράδειγμα η κίνηση του κέρσορα στην οθόνη του υπολογιστή μέσω των κινήσεων του χεριού στο ποντίκι.
- **Επικοινωνιακές χειρονομίες (Communicative or Semaphoric Gestures):** Εμπνευσμένες από συστήματα σημάτων που χρησιμοποιούν σημαίες, φώτα ή και μέρη του σώματος, οι επικοινωνιακές χειρονομίες αποτελούν οποιοδήποτε σύστημα χειρονομιών που χρησιμοποιεί κάποιο λεξιλόγιο, στατικών ή δυναμικών χειρονομιών συνήθως της παλάμης και του χεριού. Τέτοιες χειρονομίες αν και δεν χρησιμοποιούνται σε μεγάλο βαθμό στις διαπροσωπικές σχέσεις μεταξύ ανθρώπων, αποτελούν πολύ σημαντικό κομμάτι στον τομέα της αναγνώρισης χειρονομιών και χρησιμοποιούνται σε έξυπνα δωμάτια (smart rooms) και σε εφυή περιβάλλοντα (intelligent environments).
- **Συνδιαστικές χειρονομίες (Gesticulation):** Πρόκειται για μία κατηγορία χειρονομιών η οποία διέπεται από μία φυσικότητα. Είναι πολύ διαδομένες στην επικοινωνία μεταξύ των ανθρώπων και συνδυάζονται συνήθως και με λόγο, ή όπως θα μπορούσαμε να πούμε συνοδεύουν το λόγο. Η διαφορά τους με τις επικοινωνιακές χειρονομίες είναι ότι σε αντίθεση με την πρώτη αυτές δεν μαθαίνονται αλλά είναι το επακόλουθο του αυθορμητισμού του εκτελεστή τους και έρχονται φυσικά με το λόγο. Πολλές φορές αναφέρονται και ως περιγραφικές ή εικονικές χειρονομίες με σκοπό τη διασαφήνιση μίας προφορικής περιγραφής, ενός φυσικού σχήματος, ή μορφής. Για παράδειγμα ενώ περιγράφω με λόγια έναν κύλινδρο, ταυτόχρονα προσπαθώ με κινήσεις του χεριού να τον απεικονίσω στο συνομιλητή μου.
- **Χειρονομίες που περιγράφουν μία γλώσσα (Language Gestures):** Χειρονομίες που χρησιμοποιούνται για να περιγράψουν μία δομημένη γλώσσα. Διακατέχονται από εσωτερική δομή, όπως σειρά, με σκοπό να επικοινωνήσουν επακριβώς κάποιο νόημα, όπως θα γινόταν και στον γραπτό ή προφορικό λόγο. Ένα παράδειγμα είναι η τεχνητή γλώσσα χειρονομιών για άτομα με προβλήματα στην ακοή.



Σχήμα 2.2: Η κατανομή των χειρονομιών με βάση το στυλ τους [3]

### 2.2.2 Τεχνολογία εισόδου στο σύστημα

Με βάση την τεχνολογία εισόδου στο σύστημα οι χειρονομίες χωρίζονται σε αυτές που εισάγονται με κάποια συσκευή με την οποία ο χρήστης έχει φυσική επαφή (non-perceptual-input) και σε αυτές στις οποίες η τεχνολογία εισόδου περιλαμβάνει οπτικά ερεθίσματα, ακουστικά ερεθίσματα και ανίχνευση κίνησης του χρήστη, χωρίς την φυσική αλληλεπίδραση του με κάποια συσκευή εισόδου (perceptual-input).

Κάποιες από τις τεχνολογίες όπου εντάσσονται οι χειρονομίες της πρώτης κατηγορίας είναι:

1. **Ποντίκι και ψηφιακή πένα:** Χρησιμοποιούνται από χειρονομίες οι οποίες μεταφράζουν τις κινήσεις του χεριού διαμέσου των συσκευών αυτών σε κατά προσέγγιση μετακινήσεις του δείκτη στο γραφικό περιβάλλον χρήστη.
2. **Είσοδος αφής και πίεσης:** Χρησιμοποίηση αισθητήρων αφής και πίεσης για την εισαγωγή των χειρονομιών στην ηλεκτρονική συσκευ-

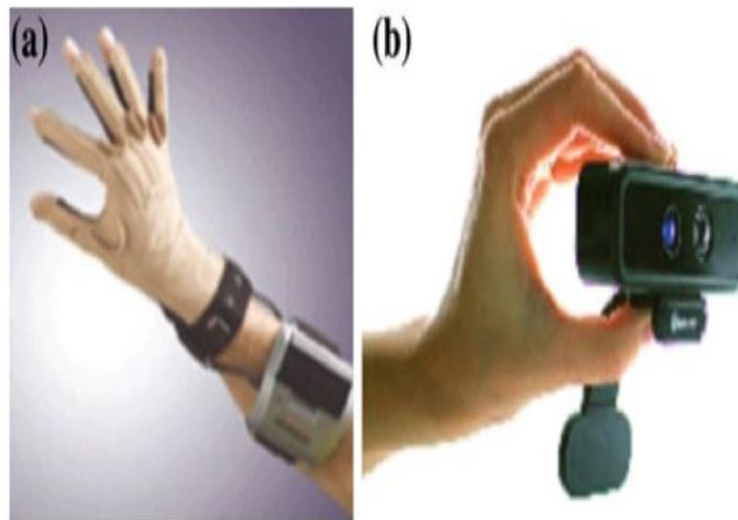


Σχήμα 2.3: Ένα παιδί χρησιμοποιεί χειρονομίες μπροστά από γραφικό περιβάλλον αυτόματης αναγνώρισης χειρονομιών [2]

ή.Επιτρέπει μία πιο φυσική αλληλεπίδραση συσκευής, χρήστη, καθιστώντας περιρτή τη χρήση ενδιαμέσων συσκευών όπως το ποντίκι.

3. **Ηλεκτρονική Ανίχνευση με Αισθητήρες σε κάποιο σημείο του σώματος:** Πρόκειται για μία από τις πρώτες μεθόδους ανίχνευσης χειρονομιών, όπως κινήσεις παλάμης ή χεριού.Το χαρακτηριστικό τους είναι η δυνατότητα να ανιχνεύσουν χώρο, θέση και προσανατολισμό δια μέσο ηλεκτρομαγνητικών αισθητήρων.
4. **Ηλεκτρονικά γάντια:** Συσκευή που όπως φανερώνει και το όνομα της φοριέται στο χέρι σαν γάντι.Περιέχει ενσωματωμένους αισθητήρες και έχει τη δυνατότητα να ανιχνεύσει πιο εκλεπτυσμένες χειρονομίες με μεγαλύτερη λεπτομέρεια.Χρησιμοποιούνται σε εφαρμογές εικονικής πραγματικότητας (virtual reality) και χειρισμού 3-Δ ψηφιακών αντικειμένων σε υπολογιστικά περιβάλλοντα.
5. **Συσκευές ανίχνευσης με την χρησιμοποίηση υπερύθρων:** Με τη βοήθεια συσκευών μετάδοσης και πρόσληψης υπέρυθρων ακτίνων γίνεται η ανίχνευση των χειρονομιών του χρήστη.Χρησιμοποιούνται κυρίως σε έξυπνα περιβάλλοντα για τον απομακρυσμένο ελέγχο συσκευών.

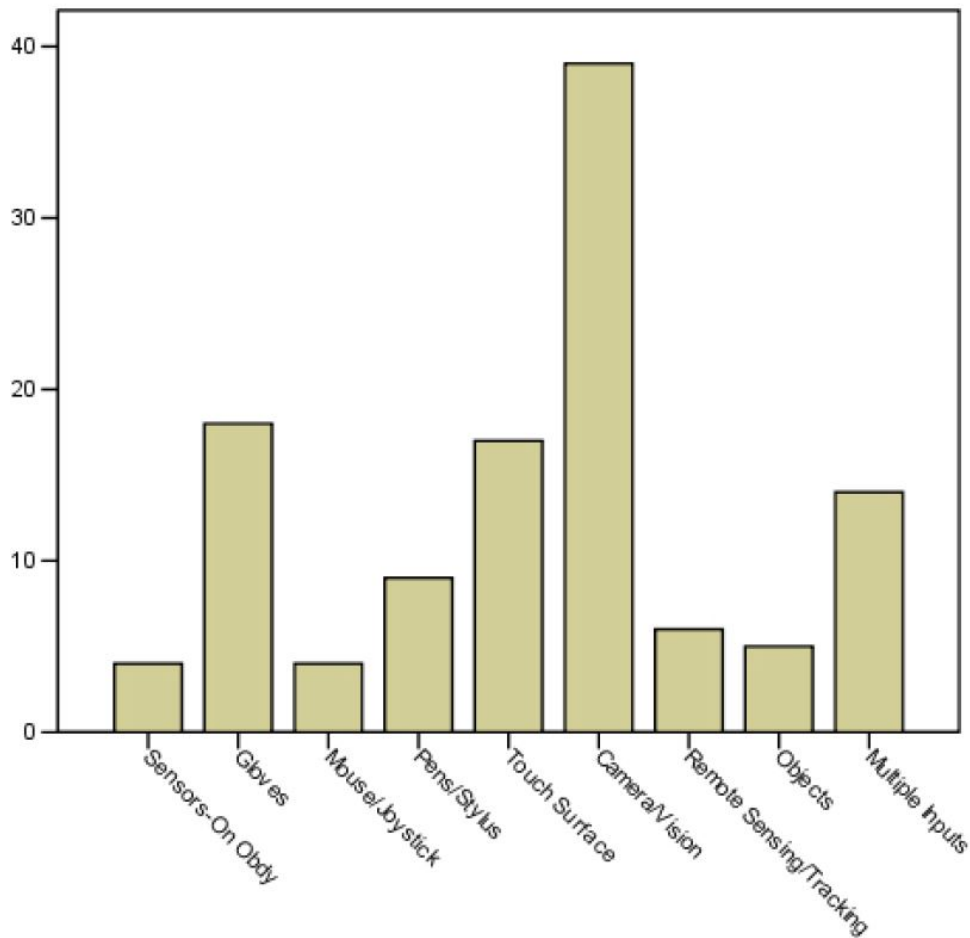
6. **Ηχητική είσοδος:** Συσκευές που συνδέουν διάφορους ήχους με συγκεκριμένες χειρονομίες στο πλαίσιο ενός υπολογιστικού περιβάλλοντος. Για παράδειγμα το χτύπημα τον δακτύλων πάνω σε μία επιφάνεια να συνδέεται με κάποια συγκεκριμένη δράση στην ψηφιακή συσκευή.
- 



Σχήμα 2.4: Γάντια με αισθητήρες για μέτρηση παραμέτρων της κινήσεως του χεριού (αριστερά) και κάμερα καταγραφής του χρήστη (δεξιά) [2]

Στην δεύτερη κατηγορία έχουμε τεχνολογίες που δεν χρειάζονται τη χρήση κάποιας συσκευής άμεσα από το χρήστη.

1. **Όραση υπολογιστών (Computer Vision):** Η όραση υπολογιστών ήταν ένας από τους κυρίαρχους τεχνολογικούς παράγοντες που επηρέασαν τον τύπο των χειρονομιών που τέθηκαν αντικείμενο έρευνας στις αλληλεπιδράσεις μεταξύ ανθρώπου μηχανής. Συνήθως χρησιμοποιείται συστοιχία καμερών για τον προσδιορισμό των κινήσεων του χρήστη σε συνδιασμό με δυνατούς αλγορίθμους ανίχνευσης βασισμένους στην φυσιολογία και στα χαρακτηριστικά που προσφέρει το σήμα της εικόνας. Ένα μειονέκτημα της χρησιμοποίησης αυτής της μεθόδου είναι η δυσκολία της να προσαρμοστεί σε απότομες αλλαγές φωτισμού.
2. **Απομακρυσμένοι αισθητήρες:** Συσκευές οι οποίες χρησιμοποιούν ηλεκτρικά πεδία για την αναγνώριση χειρονομιών. Χρησιμοποιούνται κυρίως για την ανίχνευση του ανθρώπου στο περιβάλλον, την ανίχνευση κίνησης, αλλά και πίεσης.



Σχήμα 2.5: Διαχωρισμός Χειρονομιών ως προς την είσοδο τους [3]

### 2.2.3 Χώρος Εφαρμογής

Στο σημείο αυτό θα προσπαθήσουμε να κάνουμε ένα διαχωρισμό των χειρονομιών με βάση το πεδίο εφαρμογής τους. Σε πολλές περιπτώσεις έχουμε τη δημιουργία ενός συστήματος αναγνώρισης με την ικανότητα της επαρκούς ανταπόκρισης στις απαιτήσεις πολλών τομέων και εφαρμογών.

1. **Εικονική και Επαυξημένη Πραγματικότητα**(Virtual and Augmented Reality): Είναι ένα από τα πεδία εφαρμογών που παρουσιάζουν το μεγαλύτερο ενδιαφέρον και έχουν λάβει ένα τεράστιο κομμάτι της επιστημονικής έρευνας στο συγκεκριμένο χώρο. Περιλαμβάνει διαδραστικό χαρακτήρα μεταξύ χρήστη και υπολογιστικού περιβάλλοντος δίνοντας τη δυνατότητα στο χρήστη να έχει μία ολοκληρωμένη εμπειρία εφάμιλλη της



πραγματικότητας. Υπάρχουν τέσσερις κύριες κατηγορίες αλληλεπιδράσεων που μπορούμε να εντάξουμε σε αυτήν την ταξινόμηση.

- **Διεισδυτικές Αλληλεπιδράσεις και Αβαταρς:** Αισθητήρες οι οποίοι τοποθετούνται σε σημεία του σώματος του χρήστη βοηθάνε στην αναγνώριση των κινήσεων του, και την αποτύπωση τους σε έναν εικονικό χαρακτήρα στα πλαίσια ίσως κάποιου διαδραστικού παιχνιδιού.
  - **Χειρισμός Αντικειμένων και Πλοήγηση σε Εικονικούς Κόσμους:** Χρησιμοποιούνται σε εφαρμογές τύπου CAD όπου ο χρήστης έχει την δυνατότητα να επεξεργαστεί εικονικά αντικείμενα στον χώρο με κινήσεις του χεριού και άλλων μερών του σώματος. Επίσης με τη βοήθεια ειδικών αισθητήρων στο χώρο αλλά και συσκευές τοποθετημένες στο χρήστη, μπορεί να γίνει η ανίχνευση του στο χώρο και η αίσθηση πλοήγησης σε διάφορους εικονικούς κόσμους.
  - **Επαυξημένη πραγματικότητα:** Δίνει τη δυνατότητα στους χρήστες με την εκτέλεση χειρονομιών να επιτέλεσουν λειτουργίες κάποιας εφαρμογής.
  - **Ρομποτική και Απομακρυσμένη Παρουσία:** Συνήθως εφαρμογές τέτοιου τύπου χρησιμοποιούνται σε στρατιωτικές επιχειρήσεις και διαστημικές εξερευνήσεις. Εκπαιδευμένοι χρήστες είναι υπεύθυνοι για τον χειρισμό ρομπότ από τα οποία έχουν πρώτου προσώπου εικόνα.
2. **Εφαρμογές Γραφικών και Ζωγραφικής:** Χρησιμοποίηση συσκευών όπως ποντίκι, ειδικών ψηφιακών πενών, αισθητήρων πίεσης σε οθόνες αφής, αλλά και πιο μοντέρνα συστήματα με χρησιμοποίηση καμερών για την τριών διαστάσεων διαχείριση αντικειμένων .
3. **Παιχνίδια:** Πρόκειται για επέκταση των εφαρμογών της εικονικής πραγματικότητας. Χρησιμοποίηση χειρονομιών για την ενσωμάτωση σε περιβάλλον παιχνιδιών με τη βοήθεια αισθητήρων κίνησης και ευαίσθητων καμερών με ικανότητα ανίχνευσης σε τρεις διαστάσεις. Δύο παραδείγματα είναι το Nintendo Wii controller όπου χρησιμοποιεί αισθητήρες όπως επιταχυνσιόμετρα για την προσέγγιση των κινήσεων του χεριού και η Microsoft Kinect RGBD κάμερα η οποία έχει τη δυνατότητα να ανιχνεύει κάποιο άτομο στο χώρο και στη συνέχεια με τη χρησιμοποίηση δυνατών αλγορίθμων να αναγνωρίζει διάφορες χειρονομίες αναλόγως του πλαισίου στο οποίο χρησιμοποιείται.

4. Επιγραμματικά κάποια άλλα πεδία εφαρμογής που αξιώνουν τη χρησιμοποίηση χειρονομιών είναι τα *επεμβατικά περιβάλλοντα, περιβάλλοντα κινητών συσκευών, περιβάλλοντα επικοινωνιών, ιατρικής και έξυπνων δωματίων.*

#### 2.2.4 Έξοδος ή Απόκριση Συστήματος

Ένα από τα χαρακτηριστικά που καθορίζει τις διαφορετικές αλληλεπιδράσεις χρησιμοποιώντας χειρονομίες είναι η απόκριση του συστήματος ή το αποτέλεσμα όπου μία χειρονομία οδηγεί. Με βάση αυτό το κριτήριο διαχωρίζονται τέσσερις κατηγορίες: Ακουστική, οπτική 2-Δ και 3-Δ και αποκρίσεις Επεξεργαστή

##### 1. Οπτική Απόκριση:

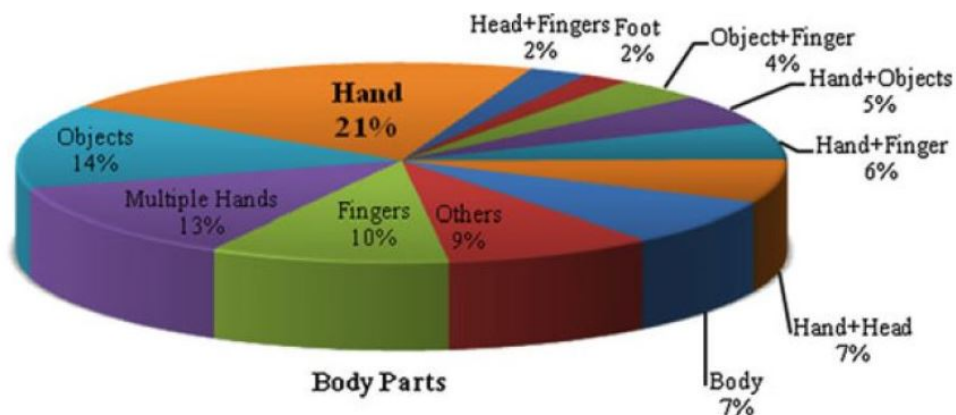
- **2 Διαστάσεων Οπτική Έξοδος:** Περιλαμβάνει οθόνες επιτραπέζιων υπολογιστών (Desktop Computers' Screens), μεγάλες οθόνες, προβολείς και φορητές ή κινητές συσκευές. Η απόφαση να συμπεριληφθεί ένα συγκεκριμένο μέσο προβολής στον σχεδιασμό αλληλεπιδράσεων βασισμένες στις χειρονομίες εξαρτάται από πολλούς παράγοντες, όμως σπάνια βασίζεται στον τύπο της συσκευής επικοινωνίας όπου ο σχεδιαστής θέλει να δημιουργήσει.
- **3 Διαστάσεων Οπτική Έξοδος:** Συνήθως στην αλληλεπίδραση με 3Δ μέσα προβολής όπως στερεοσκοπικές οθόνες υπάρχουν και 2Δ και 3Δ χειρονομίες που χρησιμοποιούνται για αυτήν την αλληλεπίδραση. Συνήθως όμως όταν μιλάμε για 3Δ περιβάλλοντα στρεφόμαστε συνήθως σε εφαρμογές εικονικής πραγματικότητας.

2. **Ακουστική έξοδος:** Χρησιμοποιείται συνήθως σε κινητά και επεμβατικά περιβάλλοντα ,πχ στην οδήγηση .

3. **Κεντρική Μονάδα Επεξεργασίας ή Έξοδος που μεταφράζεται σε κάποια εντολή:** Πολλές φορές το αποτέλεσμα της επικοινωνίας ανθρώπου υπολογιστή δια μέσου χειρονομιών δεν οδηγεί στην εμφάνιση του σε κάποια συγκεκριμένη συσκευή εξόδου, αλλά χρησιμοποιείται ως μία εντολή εισόδου για άλλες συσκευές ή εφαρμογές.

### 2.3 Ανάλυση προβλήματος

Όπως έχει ήδη καταστεί φανερό, υπάρχουν αρκετά είδη χειρονομιών. Στην παρούσα εργασία έχουμε επικεντρωθεί στο πρόβλημα της αναγνώρισης δυναμικών



Σχήμα 2.6: Το Ποσοστό των Χειρονομιών ανά μέρος του σώματος [7]

χειρονομιών και συγκεκριμένα χειρονομιών οι οποίες εκτελούνται με τη βοήθεια της κινήσεως ολόκληρου του χεριού του εκτελεστή τους. Το πρόβλημα όπως αναφέρεται και στο [7] μπορεί να διαχωριστεί σε τρία θεμελιώδη μέρη: **Ανίχνευση (Detection)**, **Παρακολούθηση (Tracking)**, **Αναγνώριση (Recognition)**

### 2.3.1 Ανίχνευση

Ένας μεγάλος αριθμός από μεθόδους έχει προταθεί στην βιβλιογραφία για την αναγνώριση χειρονομιών με βάση τις κινήσεις των χεριών. Αυτές συνήθως χρησιμοποιούν αρκετούς τύπους από οπτικά χαρακτηριστικά και σε πολλές εφαρμογές συνδυασμό τους. Τέτοια χαρακτηριστικά είναι το χρώμα του δέρματος, το σχήμα, η κίνηση αλλά και ανατομικές αναπαραστάσεις μοντέλων των χεριών.

Στη συνέχεια θα κάνουμε μια συνοπτική παρουσίαση ορισμένων από τις παραπάνω μεθόδους που αναφέρθηκαν.

- Χρώμα:** Το χρώμα του δέρματος έχει χρησιμοποιηθεί σε πολλές τεχνικές για ανίχνευση του χεριού. Μία σημαντική απόφαση που πρέπει να παρθεί για την παροχή ενός τέτοιου μοντέλου είναι και ο χώρος χρώματος που θα χρησιμοποιηθεί. Οι πιο συχνά χρησιμοποιούμενοι χώροι είναι οι RGB, HSV, YCrCb, YUV. Γενικά υπάρχει μια προτίμηση σε χώρους οι οποίοι διαχωρίζουν την χρωματικότητα από τη φωτεινότητα καθώς δίνουν μια ευελιξία στη μέθοδο να έχει κάποιο βαθμό ανεξαρτησίας από τις διακυμάνσεις του φωτισμού. Γενικά οι μέθοδοι που χρησιμοποιούν το χρώμα του δέρματος είναι επιρρεπείς σε λάθη που αφορούν τις γρήγορες μεταβολές φωτεινότητας αλλά και χρώματος του δέρματος.

- **Σχήμα:** Το χαρακτηριστικό σχήμα των χεριών έχει χρησιμοποιηθεί για την ανίχνευση τους στις εικόνες με πολλούς τρόπους. Ένας από αυτούς περιλαμβάνει την εξαγωγή των περιγραμμάτων των αντικειμένων της εικόνας.

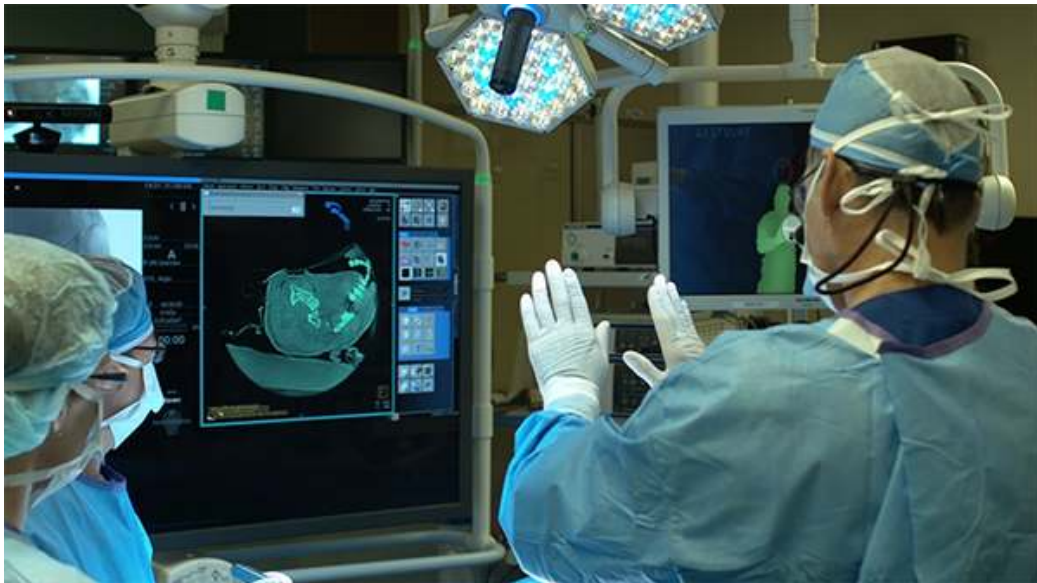
Αυτά αν αναγνωριστούν με επιτυχή τρόπο τότε είναι ικανά να οδηγήσουν σε μία αποτελεσματική τεχνική ανεξάρτητη της θέσης θέασης της εικόνας, του χρώματος του χεριού και του φωτισμού. Τεχνικές εξαγωγής περιγραμμάτων βασίζονται συνήθως στην ανίχνευση ακμών. Πολλές φορές τέτοιες τεχνικές μπορούν να ανιχνεύσουν και ακμές διαφορετικές από αυτές των χεριών, για το λόγο αυτό χρειάζεται προεπεξεργασία των δεδομένων ώστε να αυξηθεί η αξιοπιστία της ανίχνευσης.

Ακόμα μία μέθοδος αυτής της κατηγορίας περιλαμβάνει τοπικούς τοπογραφικούς περιγραφητές, οι οποίοι χρησιμοποιούνται για το ταίριασμα ενός μοντέλου με τις ακμές της εικόνας.
- **Τιμές των πίξελ (Pixel):** Μέθοδοι οι οποίες έχουν ως στόχο την εκπαίδευση ταξινομητών με τεχνικές μηχανικής μάθησης σε ένα προκαθορισμένο σύνολο δεδομένων, που αποτελείται από πρότυπες εικόνες που απεικονίζουν το αντικείμενο που πρέπει να ανιχνευτεί (τα χέρια στη δική μας περίπτωση). Όμως η αυτόματη ανίχνευση χαρακτηριστικών με τέτοιες τεχνικές δεν είναι πολύ αξιόπιστη ακόμα. Μία ακτίνα αισιοδοξίας αποτελούν οι τεχνικές ενίσχυσης (boosting) οι οποίες έχουν ως φιλοσοφία τον συνδυασμό πολλών αδύναμων ταξινομητών για τη δημιουργία ενός ενιαίου με συντριπτικά καλύτερα αποτελέσματα.
- **Μοντέλα 3Δ:** Μία κατηγορία προσεγγίσεων χρησιμοποιεί 3Δ μοντέλα χεριών για την ανίχνευση των χεριών στις εικόνες. Ένα από τα πλεονεκτήματα αυτών των μεθόδων είναι ότι μπορούν να καταφέρουν ανίχνευση ανεξαρτήτως γωνίας θέασης. Τέτοια μοντέλα χρησιμοποιούν συνήθως σημεία ή ακμές των εικόνων σε συνδυασμό με προκαθορισμένα ανατομικά χαρακτηριστικά των χεριών. Σκοπός τους είναι να χωρέσουν το μοντέλο με όσο μεγαλύτερη ακρίβεια γίνεται στο αντικείμενο, εδώ χέρι, που θέλουν να αναγνωρίσουν.
- **Κίνηση:** Χρησιμοποιείται σε λίγες μεθόδους για αναγνώριση χεριών. Ο λόγος που συμβαίνει αυτό είναι ότι χρειάζονται ελεγχόμενες συνθήκες εργαστηρίου και επιπροσθέτως υποθέτουν ότι η μόνη κίνηση που συμβαίνει στην εικόνα είναι εξαιτίας της κινήσεως των χεριών. Τέτοιες μέθοδοι περιλαμβάνουν την ανίχνευση των χεριών σε μία πρώτη εικόνα, στατική, συνήθως με τη βοήθεια ανιχνευτών δέρματος και στη συνέχεια με την μέθοδο της οπτικής ροής, την ανίχνευση της μετατόπισης των πίξελ από το προηγούμενο πλαίσιο στο επόμενο.

### 2.3.2 Παρακολούθηση

Δεδομένου ότι η λειτουργία της ανίχνευσης είναι αρκετά γρήγορη ώστε να μπορέσει να ακολουθήσει τον ρυθμό των εικόνων στο βίντεο, θα μπορούσε να χρησιμοποιηθεί και για την παρακολούθηση. Όμως η παρακολούθηση των χεριών είναι μία αρκετά δύσκολη διεργασία καθώς τα χέρια μπορεί να κινούνται πολύ γρήγορα μέσα στην ακολουθία εικόνων. Η σημασία της αποτελεσματικής παρακολούθησης είναι διττή. Κατ' αρχάς παρέχει την διαπλασιασική σύνδεση των εμφανίσεων των χεριών, με αποτέλεσμα να καταγράφει χαρακτηριστικά που σχετίζονται με την τροχιά τους. Η τροχιά παρέχει σημαντικές πληροφορίες σχετικά με το είδος και την ποιότητα της χειρονομίας και μπορεί να δωθεί ως είσοδος σε εφαρμογές που βασίζονται σε αυτή (πχ ψηφιακή ζωγραφική). Δεύτερον η παρακολούθηση παρέχει ένα τρόπο για τη διατήρηση εκτιμήσεων που σχετίζονται με τις παραμέτρους του μοντέλου, τις μεταβλητές του, και χαρακτηριστικών που δεν είναι άμεσα παρατηρήσιμα σε μία δεδομένη χρονική στιγμή.

Στη συνέχεια θα παρουσιάσουμε ορισμένες μεθόδους παρακολούθησης.



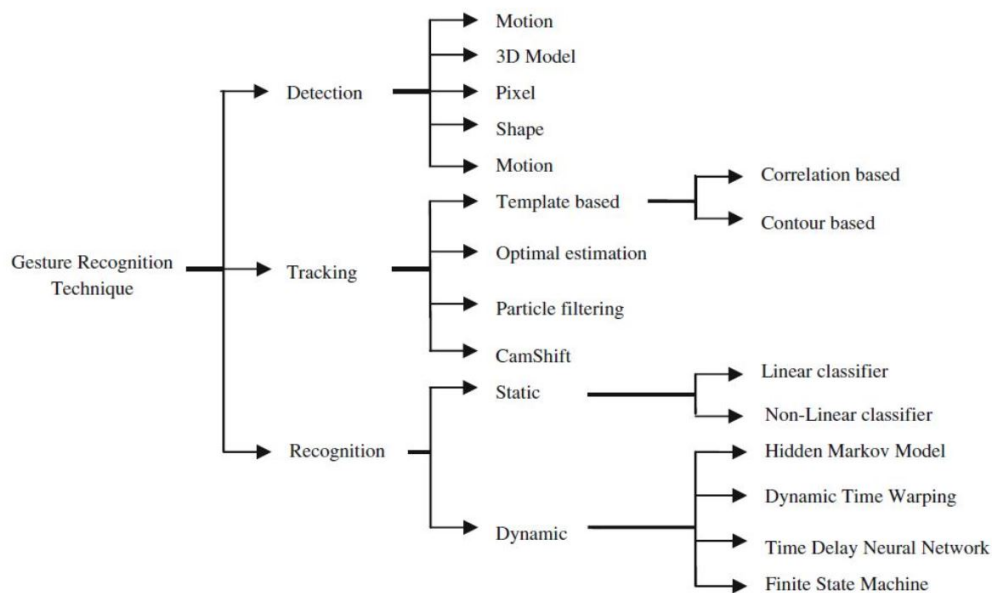
Σχήμα 2.7: Αναγνώριση Χειρονομιών στην Ιατρική : φωτογραφία του John Petis

- **Βασισμένες σε πρότυπα**(Template based): Μοιάζουν πολύ με τις μεθόδους που χρησιμοποιούνται για την ανίχνευση, προσδίδοντας όμως περαιτέρω περιορισμούς που έχουν ως στόχο την μείωση του χώρου αναζήτησης μέσα στην εικόνα. Μία μέθοδος αυτής της οικογενείας είναι η

παρακολούθηση χαρακτηριστικών βασισμένη στη συσχέτιση (Correlation-based feature tracking). Όταν ανιχνευθεί το χέρι μέσα σε ένα πλαίσιο, οι περιοχές της εικόνας όπου ανιχνεύθηκε χρησιμοποιούνται ως προτύπα για την ανίχνευση του χεριού στο επόμενο πλαίσιο. Η μέθοδος αυτή κάνει την υπόθεση ότι τα χέρια θα εμφανιστούν στην ίδια τοπολογική γειτονία από πλαίσιο σε πλαίσιο. Η παρακολούθηση με βάση τα περιγράμματα (Contour based tracking) είναι μία άλλη μέθοδος αυτής της οικογενείας. Χρησιμοποιούνται παραμορφώσιμες συναρτήσεις περιγραμμάτων, οι λεγόμενες συναρτήσεις φιδιών (snake functions). Σκοπός τους είναι η ανίχνευση περιοχών των χεριών σε συνεχόμενα πλαίσια εικόνων. Συνήθως τα όρια των περιοχών αυτών καθορίζονται από παραγώγους των εντάσεων των πίξελ. Η τεχνική αρχικοποιείται τοποθετώντας ένα περίγραμμα κοντά στην περιοχή ενδιαφέροντος. Στη συνέχεια το περίγραμμα αυτό παραμορφώνεται επαναληπτικά προς κοντινές ακμές ώστε να προσαρμοστεί καλύτερα στην πραγματική περιοχή του χεριού. Η προσαρμογή αυτή γίνεται δια μέσου της βελτιστοποίησης ενός συναρτησιακού ενέργειας. Όταν αυτού του είδους οι συναρτήσεις χρησιμοποιούνται για την παρακολούθηση, ένα μοντέλο ενεργού σχήματος (active shape model) εφαρμόζεται σε κάθε πλαίσιο και η σύγκλιση της συνάρτησης φιδιού στο συγκεκριμένο πλαίσιο χρησιμοποιείται ως αρχικό σημείο για το επόμενο.

- **Βέλτιστης Εκτίμησης (Optimal Estimation):** Η παρακολούθηση χαρακτηριστικών έχει μελετηθεί εκτενώς στα πλαίσια της όρασης υπολογιστών. Έχοντας αυτό κατά νου το περιβάλλον της βέλτιστης εκτίμησης από το φίλτρο Κάλμαν έχει χρησιμοποιηθεί ευρύτερα στο να μετατρέπει παρατηρήσεις σε εκτιμήσεις. Τα πλεονεκτήματά του είναι η απόδοση του σε πραγματικό χρόνο, η διαχείριση της αβεβαιότητας και η παροχή προβλέψεων για τα διαδοχικά πλαίσια του βίντεο.
- **Φίλτρα Σωματιδίων (Particle Filtering):** Η παρακολούθηση βασισμένη στα φίλτρα σωματιδίων και η εφαρμογή τους σε συστήματα αναγνώρισης χειρονομιών είναι μία μέθοδος που απέκτησε δημοσιότητα πρόσφατα. Τα φίλτρα σωματιδίων είναι πολύ αποτελεσματικά στην εκτίμηση της κατάστασης δυναμικών συστημάτων από πληροφορίες αισθητήρων. Η κεντρική ιδέα είναι η αναπαράσταση των πυκνοτήτων με ένα σύνολο δειγμάτων. Αυτό έχει ως αποτέλεσμα την ικανότητα τους να παριστούν ένα μεγάλο εύρος πυκνοτήτων πιθανότητας, επιτρέποντας την, πραγματικού χρόνου εκτίμηση μη γραμμικών, μη γκαουσιανών δυναμικών συστημάτων.
- **CamShift:** Ο αλγόριθμος βασίζεται στις αρχές του mean shift αλγορίθμου. Αυτός είναι μία μέθοδος παρακολούθησης που χρησιμοποιεί μοντέλα εμφάνισης βασισμένα στην πυκνότητα για την αναπαράσταση των

στόχων. Η μέθοδος παρακολουθεί τους στόχους βρίσκοντας το πρότυπο με την κοντινότερη κατανομή σε μία ακολουθία πλαισίων. Χρησιμοποιείται ευρέως λόγω της σχετικής απλότητάς του και το μικρό του υπολογιστικό κόστος. Όμως έχει ένα μειονέκτημα, αδυνατεί να προσαρμοστεί στην κλίμακα παρακολούθησης, με αποτέλεσμα να αποτυγχάνει όταν ο στόχος πλησιάζει ή απομακρύνεται από την κάμερα. Για να ξεπεραστεί το πρόβλημα αυτό δημιουργήθηκε ο CamShift (Continuous Adaptive mean-shift)



Σχήμα 2.8: Μέρη της Αναγνώρισης Χειρονομιών [7]

### 2.3.3 Αναγνώριση

Ο συνολικός στόχος της αναγνώρισης χειρονομιών που γίνονται από κινήσεις των χεριών είναι η κατανόηση της σημασιολογίας όπου η θέση και ο προσανατολισμός τους εκφράζει. Αυτό είναι και το αντικείμενο της διπλωματικής αυτής στο μεγαλύτερο της μέρος, δηλαδή να καταφέρουμε να αναγνωρίσουμε, με βάση συγκεκριμένα χαρακτηριστικά που εξάγουμε από μία ακολουθία πλαισίων, μία δομή όπου θα φανερωθεί το είδος της χειρονομίας.

Οι χειρονομίες χεριών από την πλευρά της αναγνώρισης, βασισμένης στην όραση υπολογιστών μπορούν να χωριστούν σε στατικές (πόζες χεριών) και σε δυναμικές (χειρονομίες που εμπεριέχουν σε μεγάλο βαθμό το κομμάτι της κίνησης). Για τις στατικές χειρονομίες, θα μπορούσε να χρησιμοποιηθεί ένας

γενικός ταξινομητής, όπως ένα νευρωνικό δίκτυο ταξινόμησης εικόνων ή μία μηχανή υποστήριξης διανυσμάτων(Support Vector Machine-SVM), ή ακόμα και ένας μηχανισμός ταιριάσματος προτύπων. Όμως οι δυναμικές χειρονομίες περιέχουν το χρονικό παράγοντα και απαιτούν τεχνικές όπου μπορούν να χειριστούν αυτό το χαρακτηριστικό, όπως είναι τα κρυφά μαρκοβιανά μοντέλα (Hidden Markov Models-HMMs). Στη συνέχεια θα δούμε μία επισκόπηση των αλγορίθμων μάθησης που χρησιμοποιούνται ή έχουν χρησιμοποιηθεί για την διακπεραίωση της εργασίας της αναγνώρισης χειρονομιών.

- **K-means:** Σκοπός του συγκεκριμένου αλγορίθμου είναι ο προσδιορισμός  $k$  σημείων, που ονομάζονται κέντρα με σκοπό την ελαχιστοποίηση του σφάλματος μεταξύ των διαφόρων κλάσεων (διαφορετικές ομάδες δεδομένων μέσα στο χώρο δειγμάτων του προβλήματος). Το σφάλμα μεταξύ των κλάσεων ορίζεται ως το άθροισμα των αποστάσεων όλων των σημείων από τα αντίστοιχα κέντρα των κλάσεων τους. Ο αλγόριθμος αρχίζει τοποθετώντας τυχαία  $k$  σημεία στο χώρο δειγμάτων. Κάθε πιξελ της εικόνας στη συνέχεια, ανατιθεται στο κοντινότερο κέντρο κλάσης (ένα από τα  $k$  σημεία που τοποθετήθηκαν τυχαία προηγουμένως). Αφού ανατεθούν όλα τα σημεία σε κάποιο κέντρο επαναλαμβάνεται η διαδικασία με νέα κέντρα τον μέσο όρο των σημείων που έχουν ανατεθεί σε κάθε κλάση. Όταν ο αλγόριθμος συγκλίνει, δηλαδή, είτε τα κέντρα σταματήσουν να μετακινούνται, είτε μετακινούνται με κάποια ανεκτή μικρή απόσταση, τότε λέμε ότι το μοντέλο μας έχει εκπαιδευτεί και τα  $k$  αυτά σημεία που έχουν οριστεί ως κέντρα αποτελούν τα προτότυπα μας και θεωρούνται αντιπροσωπευτικά σημεία του χώρου δειγμάτων.
- **K-Κοντινότερος Γείτονας (K-nearest neighbor):** Πρόκειται για μία μέθοδο για ταξινόμηση βασισμένη στο πόσο κοντά είναι το δείγμα προς ταξινόμηση σε σχέση με όλο τον χώρο δειγμάτων εκπαίδευσης. Ο αλγόριθμος αυτός ανήκει στην κατηγορία των αλγορίθμων αργής μάθησης (lazy learning) καθώς κάνει μόνο εκτενής τοπική αναζήτηση. Στο τέλος το δείγμα ταξινομείται με βάση την ψηφο των  $k$  κοντινότερων γειτόνων του, κατά πλειοψηφία. Αυτό σημαίνει ότι στην πιο απλή εκδοχή του, δηλαδή με  $k = 1$ , το δείγμα ταξινομείται στην κλάση του γείτονα με την ελάχιστη απόσταση. Γενικά όταν έχουμε ταξινόμηση ανάμεσα σε δύο κλάσεις, είναι θεμιτό το  $k$  να οριστεί κάποιος περιττός αριθμός ώστε να μην έχουμε ισοπαλίες στις ψήφους. Οι συνηθέστεροι τύποι απόστασης που χρησιμοποιούνται για την εύρεση των γειτόνων είναι η μανχάταν και η ευκλείδεια.
- **Mean shift clustering:** Είναι μία μη παραμετρική τεχνική που δεν χρειάζεται προηγούμενη γνώση του αριθμού των κλάσεων και δεν περι-



ορίζει το σχήμα των κλάσεων. Η κύρια ιδέα πίσω από τον συγκεκριμένο αλγόριθμο είναι η μεταχείριση των σημείων στον  $n$ -διάστατο χώρο χαρακτηριστικών ως μία εμπειρική πυκνότητα πιθανότητας όπου οι πυκνές περιοχές απεικονίζουν τα τοπικά μέγιστα της διέπουσας κατανομής. Για κάθε σημείο του χώρου των δεδομένων χρησιμοποιούνται τεχνικές κατάβασης στην τοπικά υπολογισμένη πυκνότητα μέχρι συγκλίσεως.

- **Διανύσματα Υποστήριξης (Support vector machines-SVMs):** Είναι ένας μη γραμμικός ταξινομητής, που συνήθως εμφανίζει καλύτερη απόδοση από τους υπόλοιπους ταξινομητές. Η βασική ιδέα τους είναι το μη γραμμικό πέρασμα του συνόλου δεδομένων σε έναν χώρο χαρακτηριστικών μεγαλύτερης διάστασης, όπου τα δεδομένα θα είναι γραμμικά διαχωρίσιμα. Ο αλγόριθμος διανυσμάτων υποστήριξης έχει την ιδιότητα να διαχωρίζει δύο κλάσεις παρέχοντας το μεγαλύτερο δυνατό εύρος διαχωρισμού τους.
- **Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models):** Πρόκειται για μία από τις πιο διαδεδομένες τεχνικές αναγνώρισης δεδομένων με δυναμικό και ακολουθιακό χαρακτήρα. Η θεωρία τους θα αναλυθεί εκτενώς αργότερα μέσα στο κείμενο, καθώς αποτελούν μία από τις βασικές τεχνικές που χρησιμοποιήθηκαν για την περάτωση της συγκεκριμένης διπλωματικής.
- **Dynamic time wrapping -DTW:** Χρησιμοποιείται κυρίως για το βέλτιστο ταίριασμα μεταξύ δύο σημάτων. Υπολογίζει την απόσταση μεταξύ οποιονδήποτε δύο σημείων τους. Χρησιμοποιεί αυτούς τους υπολογισμούς για τη δημιουργία ενός πίνακα απόστασης και στη συνέχεια βρίσκει το μονοπάτι με το λιγότερο κόστος μέσα σε αυτόν. Το μονοπάτι αυτό παριστάνει το ιδανικό ταίριασμα μεταξύ των δύο σημάτων.
- **Νευρωνικά Δίκτυα, Καθυστέρησης Χρόνου:** Πρόκειται για μια ειδική κατηγορία τεχνητών νευρωνικών δικτύων, με πλεονεκτήματα κυρίως σε εφαρμογές πραγματικού χρόνου. Είναι ουσιαστικά μία επέκταση των πολυεπίπεδων νευρωνικών δικτύων. Βασίζονται σε χρονικές καθυστερήσεις που δίνουν τη δυνατότητα στους νευρώνες να αποθηκεύσουν την ιστορία των σημάτων εισόδου.
- **Μηχανή Πεπερασμένης Κατάστασης (Finite State Machine):** Πρόκειται για μία δομή που περιέχει καταστάσεις. Χρησιμοποιείται πολύ συχνά στη δημιουργία εφαρμογών αυτόματης διόρθωσης ορθογραφίας και αυτόματης μετάφρασης. Περιέχει κάποια αρχική και κάποια τελική κατάσταση. Αν κατά το τέλος λήψης των δεδομένων εισόδου καταφέρει να φτά-

σει σε κάποια τελική κατάσταση, λέμε ότι η μηχανή αναγνώρισε μία οντότητα (πχ κάποια λέξη ή κάποια χειρονομία).



Σχήμα 2.9: Κάμερα Kinect για εντοπισμό χειρονομιών σε εφαρμογές αναγνώρισης

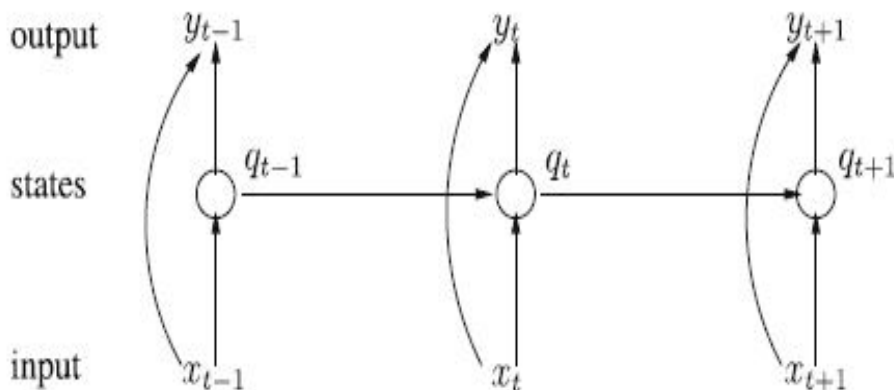
## 2.4 Κορυφαίες Μέθοδοι(State of the Art)

Γενικά ο τομέας της αναγνώρισης χειρονομιών στο κομμάτι της έρευνας πάσχει από ένα συγκεκριμένο πρόβλημα. Αυτό είναι ότι οι διάφορες μεθόδους που αναπτύσσονται για τη διακρίση του κομματιού της αναγνώρισης, δοκιμάζονται σε διαφορετικά δεδομένα εισόδου στην πλειοψηφία τους. Αυτό έχει ως αποτέλεσμα να μην υπάρχει ένα κοινό σημείο αναφοράς για σύγκριση που να καταδεικνύει κάποια μέθοδο σαφώς ανώτερη από κάποια άλλη. Παρ'όλα αυτά μπορούμε να πούμε ότι κάποιες συγκεκριμένες τεχνικές έχουν ξεχωρίσει και χρησιμοποιούνται κυρίως από ερευνητές αλλά και εταιρείες στη βιομηχανία. Στη συνέχεια θα σας παρουσιάσουμε μία από αυτές τις μεθόδους βασισμένη στα *Κρυφά Μαρκοβιανά Μοντέλα* όπως παρουσιάζεται στο [8].

### 2.4.1 Κρυφό Μαρκοβιανό Μοντέλο Εισόδου/Εξόδου (Input/Output Hidden Markov Model)

Πρόκειται για μία επέκταση των KMM. Τα Μαρκοβιανά Μοντέλα Εισόδου/Εξόδου έχουν την ικανότητα να διαχωρίσουν μία χρονική ακολουθία χρησιμοποιώντας έναν αλγόριθμο εκπαίδευσης με επίβλεψη ταυριάζοντας ένα σύνολο δεδομένων εισόδου σε ένα σύνολο δεδομένων εξόδου. Τα δεδομένα εισόδου αντιστοιχούν στις ακολουθίες παρατηρήσεων (χαρακτηριστικά από πλαίσιο σε πλαίσιο σε μία ακολουθία εικόνων ενός βίντεο) και η ακολουθία εξόδου αντιστοιχεί στην κλάση της χειρονομίας που θέλουμε να αναγνωρίσουμε.

Έστω  $x_{1..T} = \{x_1, \dots, x_t, \dots, x_T\}$  μία ακολουθία εισόδου, δηλαδή οι παρατηρήσεις (που αντιστοιχούν στην ακολουθία εξόδου του Κρυφού Μαρκοβιανού Μοντέλου), και  $y_{1..T} = \{y_1, \dots, y_t, \dots, y_T\}$  μία ακολουθία εξόδου, όπου  $y_t \in \{1 \dots C\}$  είναι ο δείκτης της κατηγορίας της χειρονομίας σε κάθε χρονικό βήμα. Το  $C$  είναι ο αριθμός των κατηγοριών των χειρονομιών και  $T$  είναι το μέγεθος των ακολουθιών εισόδου και εξόδου. Η αρχιτεκτονική του Μαρκοβιανού Μοντέλου Εισόδου/Εξόδου αποτελείται επίσης από ένα σύνολο  $N$  καταστάσεων, οπότε έστω  $q_t = \{1 \dots N\}$  να είναι η κατάσταση το χρόνο  $t$ . Με κάθε κατάσταση σχετίζονται δύο υπό συνθήκη κατανομές: Μία για τις πιθανότητες μετάβασης (transition probabilities) και μία για της πιθανότητες εκπομπής (emission probabilities). Η πιθανότητα εκπομπής  $\Pr(y_t | q_t = i, x_t) \forall i \in \{1, \dots, N\}$  τον χρόνο  $t$  εξαρτάται από την τρέχουσα κατάσταση  $q_t$ , αλλά επίσης εξαρτάται και από το  $x_t$ . Η πιθανότητα μετάβασης μεταξύ των καταστάσεων  $\Pr(q_t = j | q_{t-1} = i, x_t) \forall (i, j) \in \{1, \dots, N\}$  εξαρτάται από την προηγούμενη κατάσταση αλλά και από την παρατήρηση  $x_t$ .



Σχήμα 2.10: Οπτική Αναπαράσταση του KMM Εισόδου/Εξόδου [8]

Τα Μαρκοβιανά Κρυφά Μοντέλα εισόδου εξόδου εξαρτώνται από το χρόνο εφόσον και η πιθανότητα μετάβασης αλλά και εκπομπής εξαρτώνται από το  $x_t$ . Αυτό έχει ως αποτέλεσμα η δυναμική του συστήματος να μην είναι καθορισμένη όπως συμβαίνει στα απλά Κρυφά Μαρκοβιανά Μοντέλα αλλά να εξελίσσεται στο χρόνο και να είναι συνάρτηση της ακολουθίας εισόδου.

Η δόμηση ενός προβλήματος με ακολουθίες εισόδου με συνεχείς τιμές δια μέσο της μοντελοποίησης των υπό συνθήκη κατανομών με μοντέλα συνδυασμών Γκαουσιανών (Gaussian Mixtures Models -GMMS) είναι αρκετά δύσκολη. Για την αντιμετώπιση αυτού του προβλήματος μπορεί να χρησιμοποιηθεί μία κωδικοποίηση των δεδομένων με τη βοήθεια ενός καταλόγου κωδικολέξεων όπως πχ η μέθοδος Κβάντισης Διανυσμάτων (Vector Quantization (VQ)).

### 2.4.2 Εκπαίδευση

Έστω  $\mathbf{X}$  το σύνολο ακολουθιών εισόδου και  $\mathbf{y}$  το σύνολο ακολουθιών εξόδου που θέλουμε να μοντελοποιήσουμε. Η εκπαίδευση της μηχανής Εισόδου/Εξόδου προκύπτει από τη μεγιστοποίηση της πιθανοφάνειας

$$L = \Pr(\mathbf{y}|\mathbf{X}, \Lambda) = \prod_{k=1}^K \Pr(y_{1...T_k} | x_{1...T_k}, \Lambda) \quad (2.1)$$

Όπου το  $\Lambda$  παριστάνει το σύνολο των παραμέτρων του μοντέλου,  $K$  είναι το μήκος της ακολουθίας εισόδου εξόδου,  $x_{1...T_k} \in \mathbf{X}$  είναι η  $k_{\sigma\tau\eta}$  ακολουθία εισόδου μήκους  $T_k$  και  $y_{1...T_k} \in \mathbf{y}$  είναι η  $k_{\sigma\tau\eta}$  ακολουθία εξόδου μήκους  $T_k$ . Ο αλγόριθμος Μεγιστοποίησης Προσδοκίας (Expectation Maximization) χρησιμοποιείται για την εκπαίδευση των παραμέτρων του μοντέλου.

### 2.4.3 Αναγνώριση

Έστω  $x_{1...T}$  η ακολουθία εισόδου για αναγνώριση. Η ακολουθία ανατίθεται στην κλάση με τη μεγαλύτερη υπό συνθήκη πιθανότητα σε κάθε πλαίσιο έτσι ώστε

$$c_{max} = \operatorname{argmax}_{c \in \{1, \dots, C\}} \Pr(y_1 = c, \dots, y_T = c | x_{1...T}) \quad (2.2)$$

Η συγκεκριμένη πλειοψηφική μέθοδος ανάθεσης κατηγορίας της ακολουθίας για αναγνώριση, συμπλέει με δουλειές πάνω στην αναγνώριση χειρονομιών που χρησιμοποιούν τα κρυφά μαρκοβιανά μοντέλα εισόδου/εξόδου.

## 2.5 Εκφραστικότητα

Στη συνέχεια θα παρουσιάσουμε τη θεωρία σχετική με την εκφραστικότητα των χειρονομιών και τον ρόλο της στο πλαίσιο της αλληλεπίδρασης ανθρώπου μηχανής, όπως παρουσιάζεται και στα [9],[10],[11],[12],[13].

Το συναίσθημα και η φυσική του έκφραση είναι ένα αναπόσπαστο κομμάτι της κοινωνικής επικοινωνίας, καθώς ενημερώνει τους γύρω μας πώς νιώθουμε και επηρεάζει κοινωνικές τάσεις. Η μελέτη του συναισθήματος και της έκφρασης απασχόλησε από τις αρχές του 19<sup>ου</sup> αιώνα την επιστημονική κοινότητα.

Παρά το γεγονός ότι τα συναισθήματα μπορούν να εκφραστούν με διάφορες φυσικές εκφάνσεις και λειτουργίες, οι εκφράσεις προσώπου είναι η κατηγορία η οποία έχει ερευνηθεί εκτενώς από τους μελετητές του χώρου. Αυτό είναι φυσικό επακόλουθο αν σκεφτούμε τις πληροφορίες που μπορούμε να εξάγουμε για την συναισθηματική κατάσταση ενός ανθρώπου βασιζόμενοι στις εκφράσεις του προσώπου του. Για παράδειγμα ένα χαμόγελο, θα μπορούσε να φανερώσει την χαρά ενός ατόμου ή ένα σκυθρωπό πρόσωπο τον προβληματισμό ή τη λύπη του.

Λιγότερο εκτενείς είναι οι μελέτες που σχετίζονται με την εκφραστική πληροφορία που χειρονομίες του υπόλοιπου σώματος μεταδίδουν. Για παράδειγμα η στάση του σώματος ή οι κινήσεις που κάνουμε όταν είμαστε λυπημένοι ή χαρούμενοι αντίστοιχα. Το κομμάτι αυτής της διπλωματικής σχετίζεται με αυτές ακριβώς τις χειρονομίες και της εξαγωγής από αυτές κάποιων χαρακτηριστικών που θα μπορούσαν να χρησιμοποιηθούν ως βάση για την αυτόματη αναγνώριση εκφραστικότητας από υπολογιστικά συστήματα, με απώτερο στόχο την αυτόματη αποκωδικοποίηση του συναισθηματικού περιεχόμενου τους.

Ένα από τα βασικά ερωτήματα της μελέτης της εκφραστικότητας του σώματος έχει υπάρξει το τί πλευρές των συναισθημάτων εκφράζονται μέσα από διάφορες φυσικές συμπεριφορές και επιπροσθέτως τί οι παρατηρητές συμπαιρνούν μέσα από αυτές τις συμπεριφορές. Συγκεκριμένα υπάρχει μία άξια αναφοράς αντιπαράθεση ως προς το αν η τοποθέτηση του ανθρώπινου σώματος (πόζα) και η κίνηση μπορούν με αξιόπιστο τρόπο να μεταδώσουν τις κατηγορίες συναισθημάτων ή από την οπτική γωνία των διαστάσεων (μετρικές που δημιουργήθηκαν για να μετράνε κάποιες πτυχές των συναισθημάτων, οι πιο γνωστές και ευρέως χρησιμοποιούμενες είναι το *activation* που μετράει το πόσο δυναμική είναι μία χειρονομία και το *valence* που μετράει το πόσο ευχάριστο είναι το περιεχόμενο της) να επικοινωνήσουν την ένταση του συναισθήματος. Μία σχετική ερώτηση πάνω σε αυτό το θέμα είναι το ποια χαρακτηριστικά της κίνησης του σώματος είναι

σχετικά με την έκφραση του συναισθήματος. Αυτά τα χαρακτηριστικά μπορεί να σχετίζονται με την πόζα, τη μορφή της κίνησης, την ποσότητα της κίνησης ή τη δυναμική της.

Η μελέτη της σωματικής έκφρασης αντιμετωπίζει πληθώρα θεωρητικές και μεθοδολογικές προκλήσεις, που σχετίζονται με ερωτήματα όπως το πώς ορίζονται τα συναισθήματα και πώς αναπαρίσταται η συναισθηματική έκφραση.

Ένα από τα κύρια ζητήματα που επηρεάζει τη μελέτη της συναισθηματικής έκφρασης αφορά το διαχωρισμό μεταξύ της κωδικοποίησης του συναισθήματος από κάποιον άνθρωπο στη συμπεριφορά του και της αποκωδικοποίησης του από τον δέκτη μέσω της παρατήρησης του πομπού. Είναι προφανές ότι η κωδικοποίηση του συναισθήματος σε μία συμπεριφορά που δεν περιλαμβάνει λόγο μπορεί να έχει μεγάλη διαφορά από την αποκωδικοποίηση της (δηλαδή τι καταλαβαίνει ο δέκτης), αν λάβουμε υπ' όψιν ότι υπάρχουν πολλές φορές και λεπτές ισορροπίες μεταξύ των συναισθημάτων.

Η σκόπιμη εκτέλεση μία χειρονομιάς με σκοπό την μετάδοση κάποιου συναισθήματος, είναι πολύ δύσκολη στην καλύτερη και αδύνατη σε ορισμένες περιπτώσεις, με αποτέλεσμα να οδηγεί σε διαφορές στη δυναμική, στον χρονισμό και στη συμμετρία της από περίπτωση σε περίπτωση. Ακόμα υπάρχουν ερωτηματικά στο κατά πόσο μία εκφραστική συμπεριφορά σχετίζεται με κάποια συναισθηματική κατάσταση ή απλά χρησιμοποιείται για να επηρεάσει τον δέκτη.

Κύριο πρόβλημα στην εξαγωγή συναισθημάτων δια μέσου των χειρονομιών του σώματος είναι ότι δεν υπάρχει ένας καθολικός τρόπος εξαγωγής χαρακτηριστικών υπό μελέτη. Σε αντίθεση χειρονομίες που σχετίζονται με το πρόσωπο, όπως εκφράσεις προσώπου βασίζονται σε ένα καθορισμένο σύστημα κωδικοποίησης το **FACS-(Facial Action Coding System)**. Για το λόγο αυτό οι ερευνητές πρέπει από μόνοι τους να βρουν τον τρόπο για να αναπαραστήσουν τα σημεία του σώματος που παίρνουν μέρος στη χειρονομία, τη δυναμική τους και τη μορφή της κίνησης.

Στη συνέχεια θα περιγράψουμε ορισμένα από τα συστήματα κωδικοποίησης που έχουν χρησιμοποιηθεί για τη μελέτη της έκφρασης του συναισθήματος δια μέσου χειρονομιών που σχετίζονται με το σώμα.

- **Αναπαράσταση πόζας και θέσης των σημείων του σώματος:** Η πόζα συνήθως αναφέρεται στην στατική τοποθέτηση ολόκληρου

του σώματος.Ορίζεται ως η θέση και προσανατολισμός συγκεκριμένων μερών του σώματος.

- **Αναπαράσταση κίνησης:** Η κινητική ανάλυση που βασίζεται στον Laban περιέγραψε τις ενέργειες των μερών του σώματος στο χρόνο χρησιμοποιώντας ένα σύστημα σημειώσεων που αποτελούσαν από σύμβολα που παριστούσαν τα ποιοτικά και ποσοτικά χαρακτηριστικά της κίνησης.Τα σύμβολα οργανώνονται σε πέντε κατηγορίες:**Σώμα**(μερη του σώματος που εμπλέκονται στη χειρονομία),**Χώρο**(τοπικές κατευθύνσεις και μονοπάτι της κίνησης), **Προσπάθεια**(πώς το σώμα εκτελεί κινήσεις) και **Σχήμα**(μορφές που το σώμα κάνει στο χώρο)
- **Αναπαράσταση χειρονομίας:** Όλες οι χειρονομίες είναι κινήσεις, αλλά όχι όλες οι κινήσεις χειρονομίες.

Έχοντας τα παραπάνω κατά νου, μπορούμε να προχωρήσουμε στις τεχνικές. Μία μέθοδος κομματιάσματος και ταξινομήσεως προτάθηκε για την ενσωμάτωση της δυναμικής στην περιγραφή της χειρονομίας από τον D.McNeil.Το κομματάσιμα γίνεται χρησιμοποιώντας την λογική τριών φάσεων της κίνησης της χειρονομίας: προετοιμασία, κυρίως κίνηση και υποχώρηση στην αρχική θέση.Η κυρίως κίνηση είναι το μέρος της χειρονομίας που εμπεριέχει το νόημα της.Πριν από αυτό το κομμάτι το μέρος του σώματος που εκτελεί την χειρονομία παίρνει την κατάλληλη θέση μέσα στο χώρο για την εκτέλεση της (φάση προετοιμασίας). Στη συνέχεια μετά την χειρονομία επιστρέφει στην αρχική του θέση (φάση υποχώρησης).Η ταξινόμηση αποτελείται από το μέρος της περιγραφής ορισμένων χαρακτηριστικών της χειρονομίας, όπως το σχήμα του χεριού, τον προσανατολισμό του κτλ.

Κάποιες θεωρίες συναισθήματος ισχυρίζονται ότι υπάρχουν διακριτά συναισθήματα.Τα συνηθέστερα είναι ο φόβος, ο θυμός ,η λύπη, η χαρά, η έκπληξη, η αηδία αλλά και το ενδιαφέρον, η ντροπή και οι ενοχή.Θεωρίες διαστάσεων ορίζουν χώρους που υπολογίζουν ομοιότητες και διαφορές στις συναισθηματικές καταστάσεις.Για παράδειγμα το μοντέλο PAD(pleasure,arousal,dominance), περιγράφει την συναισθηματική κατάσταση με τιμές ευχαρίστησης, διέγερσης και επικράτησης.

Κάποιες θεωρίες δίνουν έμφαση στη διασύνδεση της νόησης και του συναισθήματος και σύμφωνα με αυτές ένα συναίσθημα δημιουργείται από την εκτίμηση της κατάστασης σε σχέση με τα πιστεύω μας, τις επιθυμίες μας, αλλά και τις προθέσεις μας.

Το πώς αναπαρίστανται τα συναισθήματα παίζει καθοριστικό ρόλο στην έρευνα.Μία προσέγγιση σε κατηγορικό μοντέλο συναισθήματος θα ψάξει για δι-

ασύνδεση μεταξύ των κατηγοριών των συναισθημάτων και την έκφραση τους μέσα στην συμπεριφορά. Από την άλλη μεριά μία προσέγγιση βασισμένη στο μοντέλο των διαστάσεων επικεντρώνεται στην ένταση και στην ευχαρίστηση και συνεπώς ψάχνει για αυτές τις διαστάσεις στην έκφραση του σώματος.

Η επιστήμη της κίνησης, η οποία μελετάει τις μετακινήσεις του σώματος που δεν περιέχουν ομιλία, υποστηρίζει ότι η δυναμική τους μπορεί να είναι επαρκής για την εξαγωγή μίας διέπουσας συναισθηματικής κατάστασης. Μελέτες έδειξαν ότι η ταχύτητα, η επιτάχυνση και η απότομη αλλαγή κατεύθυνσης είναι κινήσεις που επηρεάζονται άμεσα από την συναισθηματική κατάσταση.

Οι πιο πολλές μελέτες της κίνησης έλαβαν μία καθολική προσέγγιση, βασισμένη στη δυναμική ολόκληρου του σώματος. Το περπάτημα συγκεκριμένα έχει λάβει το μεγαλύτερο κομμάτι της μελέτης. Το γενικό επίπεδο της δραστηριότητας της κίνησης και ο χώρος στον οποίο αυτή εξαπλώνεται φαίνεται να είναι σημαντικά χαρακτηριστικά για την αναγνώριση του συναισθήματος από την κίνηση. Θυμωμένες κινήσεις τείνουν να είναι μεγάλες, γρήγορες με απότομες μεταβολές, ενώ κινήσεις που υποδηλώνουν φόβο ή λύπη είναι λιγότερο ενεργητικές, μικρότερες και πιο αργές.

Τέλος, βασισμένες στην επιστήμη της ψυχολογίας έχουν οριστεί ορισμένες παράμετροι εκφραστικότητας, οι οποίες, εμπεριέχουν αρκετή πληροφορία για τη συναισθηματική κατάσταση του ατόμου. Αυτές είναι:

1. **Χώρική εξάπλωση:** Δηλαδή πόσο μεγάλη είναι η χειρονομία στο χώρο
2. **Χρονική εξάπλωση:** Ταχύτητα χειρονομίας
3. **Ρευστότητα:** Δηλαδή το πώς δύο συνεχόμενες χειρονομίες εκφράζονται
4. **Επανάληψη:** Πόσο συχνά η χειρονομία επαναλαμβάνετε
5. **Δύναμη:** Πόσο δυνατή είναι η χειρονομία
6. **Ολική Ενεργοποίηση:** Η συνολική ποιότητα της κίνησης σε ένα συγκεκριμένο κανάλι μετάδοσης



## Κεφάλαιο 3

# Τεχνικές μηχανικής μάθησης

### 3.1 Γενικά

Σε αυτό το σημείο θα γίνει μία παρουσίαση των τεχνικών μηχανικής μάθησης που χρησιμοποιήσαμε για την αντιμετώπιση του προβλήματος της αναγνώρισης χειρονομιών σε αυτή τη διπλωματική εργασία αλλά και του συνολικού συστήματος που δημιουργήσαμε για να εξάγουμε το τελικό αποτέλεσμα.

Δύο ήταν οι κύριες τεχνικές που υλοποιήθηκαν συνδυαστικά για την διακρίβωση αυτής της εργασίας. Η πρώτη εμπίπτει σε μία κατηγορία νευρωνικών δικτύων που ονομάζεται αυτοοργανούμενοι χάρτες (SOM-Self Organizing Maps). Η δεύτερη βασίζεται στις διαδικασίες Μαρκόφ και συγκεκριμένα στις αλυσίδες Μαρκόφ και στα Κρυφά Μαρκοβιανά Μοντέλα. Στη συνέχεια θα παρουσιαστεί αναλυτικά η θεωρία πίσω από αυτές τις μεθόδους. Η παρουσίαση της θεωρίας βασίζεται στο [14] για τους αυτοοργανούμενους χάρτες, ενώ για τα μαρκοβιανά μοντέλα έγινε άντληση πληροφοριών από τα [15], [16], [17].

### 3.2 Αυτο Οργανούμενοι Χάρτες

#### 3.2.1 Εισαγωγή

Όπως αναφέρθηκε και προηγουμένως οι αυτο οργανούμενοι χάρτες ανήκουν στην κατηγορία των τεχνητών νευρωνικών δικτύων. Τα δίκτυα αυτά βασίζονται στην ανταγωνιστική μάθηση, με την έννοια, ότι οι νευρώνες εξόδου ανταγωνίζονται μεταξύ τους ώστε να είναι εκείνοι που θα ενεργοποιηθούν σε μία συγκεκριμένη είσοδο. Ο νευρώνας εξόδου που εν τέλει καταφέρνει να ενεργοποιηθεί νικώντας κατά μία έννοια τους αντιπάλους του ονομάζεται νευρώνας νικητής.

Σε έναν τέτοιο χάρτη, οι νευρώνες τοποθετούνται ως κόμβοι σε ένα πλέγμα μίας ή συνηθέστερα δύο διαστάσεων. Οι διαστάσεις του χάρτη μπορεί να είναι και περισσότερες από δύο όμως εκτός από το ότι δεν συνηθίζεται, ένας χάρτης σε παραπάνω από τρεις διαστάσεις χάνει το πλεονέκτημα της οπτικοποίησης. Οι νευρώνες ρυθμίζονται επιλεκτικώς μέσα από μία διαδικασία μάθησης ώστε να ανταποκρίνονται σε συγκεκριμένα ερεθίσματα εισόδων. Οι τοποθεσίες των νευρώνων πάνω στον χάρτη ρυθμίζονται με τέτοιο τρόπο ώστε να δημιουργήσουν μία συγκεκριμένη δομή πάνω του που θα έχει νόημα, με την έννοια ότι γειτονικοί νευρώνες συνηθίζουν να ανταποκρίνονται σε κοντινές εισόδους του N-διάστατου χώρου. Για το λόγο αυτό οι αυτο οργανούμενοι χάρτες χαρακτηρίζονται από τη δημιουργία ενός τοπογραφικού χάρτη των προτύπων εισόδου, στον οποίο οι χωρικές τοποθεσίες των νευρώνων στο πλέγμα συνδέονται με τα εσωτερικά στατιστικά χαρακτηριστικά που περιέχονται στα πρότυπα εισόδου.

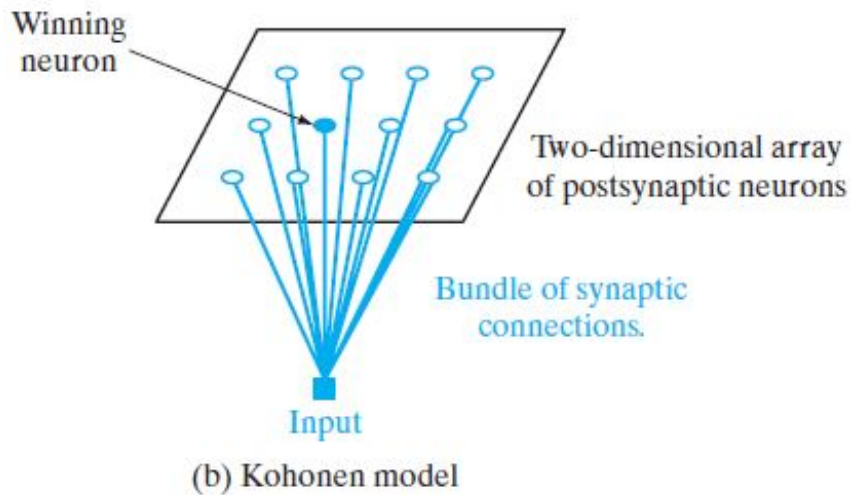
Ως νευρωνικό μοντέλο οι αυτο οργανούμενοι χάρτες παρέχουν μία σύδεση μεταξύ δύο επιπέδων προσαρμογής:

- Κανόνες προσαρμογής σε μικροσκοπικό επίπεδο ενός νευρώνα
- Προσαρμογή σε επίπεδο στρωμάτων νευρώνων

Οι αυτο οργανούμενοι χάρτες είναι μη γραμμικοί από τη φύση τους.

Γενικά υπάρχουν δύο είδη αυτο οργανούμενων χαρτών, το μοντέλο του Willshaw-von der Malsburg's και το μοντέλο του Kohonen. Το δεύτερο είναι και το μοντέλο που θα ασχοληθούμε. Οι υπολογιστικοί χάρτες τέτοιου είδους διέπονται από τέσσερις ιδιότητες

1. Σε κάθε χάρτη, οι νευρώνες δρουν παράλληλα και επεξεργάζονται κομμάτια πληροφορίας που είναι παρόμοια στη φύση τους, αλλά ίσως προέρχονται από διαφορετικά σημεία του χώρου εισόδου.
2. Σε κάθε επίπεδο αναπαράστασης κάθε κομμάτι πληροφορίας κρατιέται στο περιβάλλον στο οποίο έχει νόημα.
3. Νευρώνες που ανταποκρίνονται σε πληροφορίες κοντινές μεταξύ τους βρίσκονται κοντά ώστε να επικοινωνούν με συνάψεις μικρού μήκους



Σχήμα 3.1: Το προτότυπο μοντέλο του Kohonen [14]

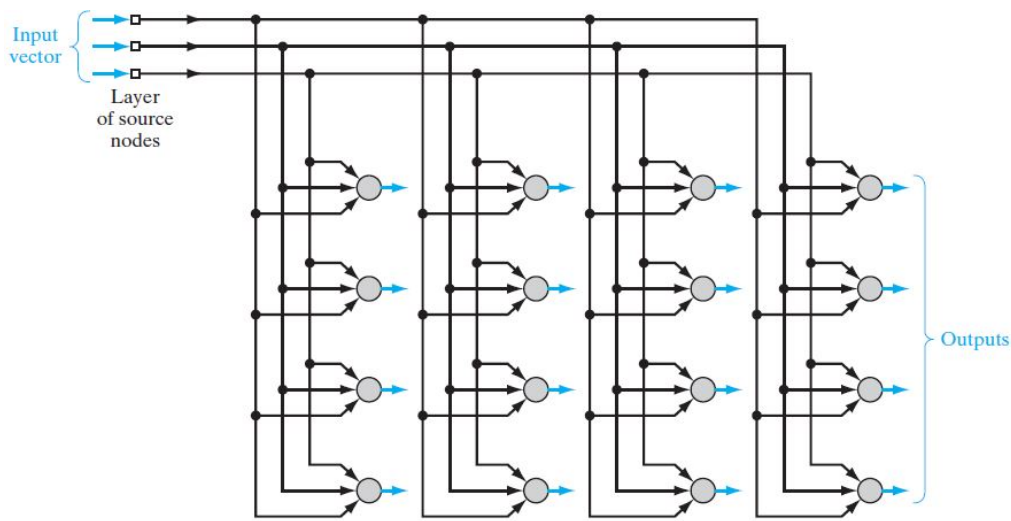
4. Οι χάρτες αυτοί βοηθούν στην οπτικοποίηση δεδομένων υψηλότερων διαστάσεων σε χαμηλότερες.

### 3.2.2 Ο Χάρτης

Ο κύριος στόχος του αυτο οργανούμενου χάρτη είναι η μετατροπή ενός αυθαίρετου σήματος εισόδου οποιασδήποτε διάστασης σε έναν μονοδιάστατο ή διδιάστατο χάρτη, με αυτή τη μετατροπή να λαμβάνει χώρο με έναν τοπολογικά διατεταγμένο τρόπο.

Όπως βλέπουμε και από το σχήμα 3.2 κάθε νευρώνας στο πλέγμα είναι πλήρως συνδεδεμένος σε όλους τους κόμβους πηγής. Αυτό το δίκτυο αναπαριστά ένα δόμημα εμπρόσθιας τροφοδότησης, με ένα υπολογιστικό στρώμα που αποτελείται από νευρώνες που διατάσσονται σε γραμμές και στήλες.

ο αλγόριθμος που είναι υπεύθυνος για τον σχηματισμό του χάρτη, αρχικοποιεί σε πρώτο μέρος τα βάρη των συνάψεων του δικτύου. Αυτό μπορεί να γίνει ορίζοντας μικρές τυχαίες τιμές από μία γεννήτρια τυχαίων αριθμών. Αυτό έχει ως αποτέλεσμα μία αρχική τυχαία τοποθέτηση των νευρώνων στο χάρτη καθώς τα συναπτικά τους βάρη είναι αυτά που στην ουσία καθορίζουν το τελικό οπ-



Σχήμα 3.2: Ένα διδιάστατο πλέγμα νευρώνων [14]

τικό αποτέλεσμα. Μετά την αρχικοποίηση των νευρώνων υπάρχουν κυρίως τρία στάδια για την τελική τοποθέτησή τους στο χώρο:

1. **Ανταγωνισμός:** Για κάθε πρότυπο εισόδου, οι νευρώνες στο δίκτυο υπολογίζουν τις τιμές τους ως προς την είσοδο με βάση κάποια ορισμένη συνάρτηση. Ο νευρώνας με τη μεγαλύτερη τιμή ανακηρύσσεται νικητής της διαδικασίας.
2. **Συνεργασία:** Ο νικητής ορίζει μία τοπολογική γειτονιά νευρώνων που θα διεγερθούν, με βάση τη δομή του αρχικού χάρτη, προσφέροντας μία βάση συνεργασίας.
3. **Προσαρμογή νευρώνων:** Αυτή η διαδικασία έχει ως στόχο να δώσει πλεονέκτημα στους νευρώνες που είχαν διεγερθεί στο προηγούμενο στάδιο αυξάνοντας τα συναπτικά τους βάρη σε σχέση με την συγκεκριμένη είσοδο. Αυτό έχει ως αποτέλεσμα οι νευρώνες να σχηματίζουν κάποιες υποομάδες οι οποίες αντιδρούν περισσότερο σε συγκεκριμένα πρότυπα εισόδου.

### 3.2.3 Διαδικασία Ανταγωνισμού

Έστω  $m$  η διάσταση του χώρου εισόδου των δεδομένων. Ας θεωρήσουμε ένα τυχαίο δείγμα του χώρου εισόδου που ορίζεται ως

$$\mathbf{x} = [x_1, x_2, \dots, x_m]^T \quad (3.1)$$

Το διάνυσμα βάρους κάθε νευρώνα στο δίκτυο έχει την ίδια διάσταση με το χώρο εισόδου. Έστω το συναπτικό βάρος του νευρώνα  $j$  να συμβολίζεται με

$$\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jm}]^T, \quad j = 1, 2, \dots, l \quad (3.2)$$

Όπου  $l$  είναι ο συνολικός αριθμός νευρώνων στο δίκτυο. Για να βρούμε το καλύτερο ταίριασμα του διανύσματος εισόδου  $\mathbf{x}$  με το διάνυσμα συναπτικού βάρους  $\mathbf{w}_j$ , συγκρίνουμε το εσωτερικό γινόμενο  $w_j^T x$  για  $j = 1, 2, \dots, l$  και επιλέγουμε το μεγαλύτερο. Αυτή η μέθοδος υποθέτει ότι το ίδιο κατώφλι εφαρμόζεται σε όλους τους νευρώνες. Το κατώφλι είναι η αρνητική πόλωση. Έτσι επιλέγοντας το νευρώνα με το μεγαλύτερο εσωτερικό γινόμενο  $w_j^T x$ , ουσιαστικά προσδιορίζουμε τον νευρώνα με βάση τον οποίο θα κεντραριστεί η τοπολογική γειτονιά των νευρώνων που θα διεγερθούν.

Αξίζει να αναφερθεί ότι το κριτήριο του εσωτερικού γινομένου είναι ισοδύναμο με το κριτήριο ελάχιστης απόστασης με μόνη προϋπόθεση τα βάρη να είναι κανονικοποιημένα με μοναδιαίο μέτρο. Αν ορίσουμε με  $i(x)$  τον δείκτη του νευρώνα ο οποίος ικανοποιεί το κριτήριο του εσωτερικού γινομένου για είσοδο  $\mathbf{x}$ , τότε μπορούμε να βρούμε αυτόν τον δείκτη εφαρμόζοντας την παρακάτω συνθήκη η οποία αποτελεί και την ουσία του ανταγωνισμού μεταξύ των νευρώνων.

$$i(x) = \underset{j}{\operatorname{argmin}} \|x - w_j\|, \quad j \in A \quad (3.3)$$

Όπου με  $A$  συμβολίζουμε το σύνολο των νευρώνων πάνω στο πλέγμα. Ουσιαστικά το  $i(x)$  είναι αυτό που μας αφορά καθώς θέλουμε να προσδιορίσουμε την ταυτότητα του νικητή νευρώνα. Σύμφωνα με τα προηγούμενα μπορούμε να φτάσουμε στο ακόλουθο συμπέρασμα

*Ο συνεχής χώρος προτύπων εισόδου απεικονίζεται σε έναν διακριτό χώρο εξόδου που αποτελείται από τους νευρώνες, μέσω της διαδικασίας του ανταγωνισμού τους μέσα στο δίκτυο.*

### 3.2.4 Διαδικασία Συνεργασίας

Ο νικητής νευρώνας εντοπίζει το κέντρο της τοπολογικής γειτονιάς των νευρώνων που θα συνεργαστούν. Ένα ερώτημα που γεννάται είναι το πώς μπορούμε να προσδιορίσουμε αυτή την γειτονιά με τρόπο που να έχει νόημα νευροβιολογικώς.

Η επιστήμη της νευροβιολογίας υποστηρίζει ότι υπάρχει απόδειξη πλευρικής επικοινωνίας μεταξύ ομάδων διεγερμένων νευρώνων. Συγκεκριμένα ένας νευρώνας που πυροδοτείται τείνει να διεγείρει τους γειτονικούς του νευρώνες περισσότερο από τους πιο μακρινούς του. Η παρατήρηση αυτή μας οδηγεί να εισάγουμε μία τοπολογική γειτονιά γύρω από τον νικητή νευρώνα  $i$  και να την κάνουμε να εξασθενεί ομαλώς με την απόσταση. Συγκεκριμένα, έστω ότι το  $h_{j,i}$  δηλώνει την τοπολογική γειτονιά κεντραρισμένη στον νικητή νευρώνα  $i$  η οποία περιλαμβάνει και ένα σύνολο διεγερμένων νευρώνες γείτονες, ένας εκ των οποίων συμβολίζεται με  $j$ . Έστω  $d_{j,i}$  η απόσταση μεταξύ του νευρώνα νικητή  $i$  και του διεγερμένου νευρώνα  $j$ . Τότε μπορούμε να υποθέσουμε ότι η συνάρτηση τοπολογικής γειτονιάς  $h_{j,i}$  εξαρτάται από την απόσταση  $d_{j,i}$  έτσι ώστε να ικανοποιούνται δύο κύριες απαιτήσεις:

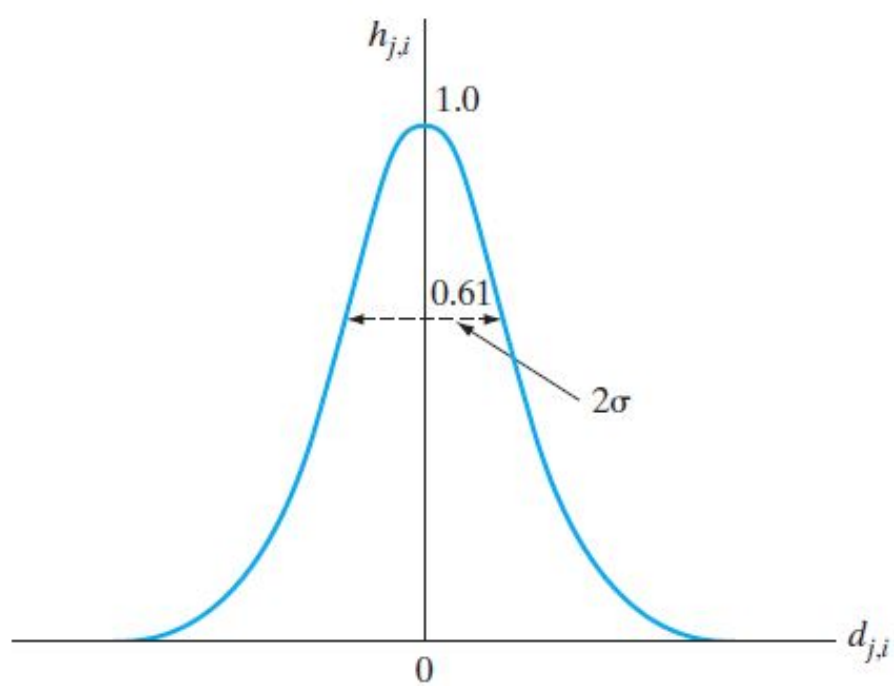
1. Η συνάρτηση της τοπολογικής γειτονιάς  $h_{j,i}$  να είναι συμμετρική γύρω από το μέγιστο σημείο που ορίζεται από το  $d_{j,i} = 0$ , δηλαδή να παίρνει τη μέγιστη τιμή της πάνω στον νικητή νευρώνα (που προφανώς απέχει από τον εαυτό του μηδενική απόσταση)
2. Το ύψος της συνάρτησης τοπολογικής γειτονιάς  $h_{j,i}$  να μειώνεται μονοτονικά με την αύξηση της διαμήκης απόστασης  $d_{j,i}$ , τείνοντας στο μηδέν για  $d \rightarrow \infty$ , η τελευταία είναι απαραίτητη συνθήκη για σύγκλιση.

Μία καλή επιλογή της  $h_{j,i}$  που ικανοποιεί τις παραπάνω προϋποθέσεις είναι αυτή της Γκαουσιανής συνάρτησης

$$h_{j,i}(x) = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right), \quad j \in A \quad (3.4)$$

η οποία είναι ανεξάρτητη της μετατόπισης (δηλαδή ανεξάρτητη από την τοποθεσία του νευρώνα νικητή). Η παράμετρος  $\sigma$  είναι το πλάτος δράσης της τοπολογικής γειτονιάς όπως φαίνεται και στην εικόνα 3.3. Ρυθμίζει το πλήθος των νευρώνων που θα θεωρηθούν ότι βρίσκονται στην τοπολογική γειτονιά του νικητή νευρώνα

Για να είναι δυνατόν να υπάρξει το κομμάτι της συνεργασίας είναι απαραίτητο η τοπολογική γειτονιά να εξαρτάται από την απόσταση  $d_{j,i}$  μεταξύ του νικητή νευρώνα  $i$  και του νευρώνα διέγερσης  $j$  στο χώρο εξόδου, παρά από κάποια άλλη μετρική απόστασης στο χώρο εισόδου. Η απόσταση αυτή όταν μιλάμε για μονοδιάστατο πλέγμα είναι απλά η διαφορά της θέσης των δύο νευρώνων  $d_{j,i} = |i - j|$ . Στη συνηθέστερη περίπτωση όμως, δηλαδή όταν έχουμε διδιάστατο πλέγμα η απόσταση αυτή ορίζεται ως η νόρμα μεταξύ των διανυσμάτων



Σχήμα 3.3: Τοπογραφική Συνάρτηση Γειτονιάς Γκάους [14]

θέσης των νευρώνων στο πλέγμα

$$d_{j,i}^2 = \|\mathbf{r}_j - \mathbf{r}_i\|^2 \quad (3.5)$$

και συγκεκριμένα το  $r_i$  είναι το διάνυσμα θέσης του νικητή νευρώνα και το  $r_j$  το διάνυσμα θέσης κάποιου άλλου νευρώνα του πλέγματος.

Ένα άλλο σημαντικό στοιχείο του αλγορίθμου των αυτο οργανούμενων χαρτών είναι ότι το μέγεθος της τοπολογικής γειτονιάς επιτρέπεται να συρρικνώνεται με το χρόνο. Αυτό επιταγχάνεται κάνοντας το πλάτος  $\sigma$  της τοπολογικής γειτονιάς  $h_{j,i}$  να μικραίνει με το χρόνο. Μία συχνή τακτική για την εξάρτηση του  $\sigma$  από το διακριτό χρόνο  $n$  είναι η εκθετική μείωσή του που περιγράφεται από

$$\sigma(n) = \sigma_0 \exp\left(-\frac{n}{\tau_1}\right), \quad n = 0, 1, 2, \dots \quad (3.6)$$

Όπου  $\sigma_0$  είναι η τιμή του  $\sigma$  στην εκκίνηση του αλγορίθμου του χάρτη και  $\tau_1$  είναι η σταθερά χρόνου η οποία επιλέγεται από τον σχεδιαστή. Άρα η συνάρτηση της τοπολογικής γειτονιάς σύμφωνα με τα νέα δεδομένα διαμορφώνεται ως

$$h_{j,i(x)}(n) = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2(n)}\right), \quad n = 0, 1, 2, \dots \quad (3.7)$$

Έτσι καθώς ο διακριτός χρόνος αυξάνεται (ο αριθμός επαναλήψεων του αλγορίθμου) το πλάτος  $\sigma(n)$  μειώνεται εκθετικά, και η τοπολογική γειτονιά μειώνεται με αντίστοιχο τρόπο. Εδώ είναι σημαντικό να παρατηρήσουμε ότι η συνάρτηση της τοπολογικής γειτονιάς θα έχει πάντα τιμή μονάδα για τον νικητή νευρώνα  $i$  εφόσον η απόσταση του από τον εαυτό του είναι μηδενική, που έχει ως αποτέλεσμα η Γκαουσιανού τύπου συνάρτηση γειτονιάς να παίρνει την τιμή ένα.

Υπάρχει και ένας ακόμα χρήσιμος τρόπος να δούμε την μεταβολή της συνάρτησης γειτονιάς  $h_{j,i(x)}(n)$  γύρω από το νικητή νευρώνα  $i(x)$  στο πέρασμα του χρόνου  $n$ . Ο σκοπός μίας συνάρτησης γειτονιάς  $h_{j,i(x)}(n)$  με μεγάλο εύρος είναι να συσχετίσει την κατεύθυνση των ανανεώσεων των βαρών από έναν μεγάλο αριθμό νευρώνων στο πλέγμα. Όσο το εύρος μειώνεται, τόσο λιγότερος είναι και ο αριθμός των νευρώνων που επηρεάζονται από μία συγκεκριμένη είσοδο



### 3.2.5 Διαδικασία Προσαρμογής

Πρόκειται για την τελευταία διαδικασία, την διαδικασία προσαρμογής των συνάψεων, στον αυτο οργανούμενο σχηματισμό του χάρτη. Για να είναι το δίκτυο αυτο-οργανούμενο το συναπτικό βάρος  $w_j$  του νευρώνα  $j$  του δικτύου, πρέπει να αλλάζει σε σχέση με το πρότυπο εισόδου  $x$ . Το ερώτημα είναι ποιός θα είναι ένας σωστός τρόπος ώστε να γίνεται αυτή η αλλαγή. Η λύση έρχεται χρησιμοποιώντας μία παραλλαγή της Χεμπιανής Μάθησης, παραλλάζοντας την Χεμπιανή Υπόθεση, εισάγοντας έναν όρο λήθης  $g(y_j)w_j$  με  $w_j$  να είναι το διάνυσμα βάρους της σύναψης του νευρώνα  $j$  και  $g(y_j)$  μία θετική βαθμωτή συνάρτηση της απόκρισης  $y_j$ . Η μόνη απαίτηση για την συνάρτηση  $g(y_j)$  είναι ότι ο σταθερός όρος στην επέκταση της σε σειρά *Taylor* να είναι μηδέν, ώστε να μπορούμε να γράψουμε

$$g(y_j) = 0 \quad \text{for } y_j = 0 \quad (3.8)$$

Δοθέντος μίας τέτοιας συνάρτησης μπορούμε να εκφράσουμε την αλλαγή του διανύσματος βάρους του νευρώνα  $j$  στο πλέγμα ως :

$$\Delta w_j = \eta y_j \mathbf{x} - g(y_j) \mathbf{w}_j \quad (3.9)$$

Όπου  $\eta$  είναι η παράμετρος ρυθμού μάθησης του αλγορίθμου. Ο πρώτος όρος του δεξιού μέρους της παραπάνω εξίσωσης είναι ο Χεμπιανός όρος και ο δεύτερος είναι ο όρος λήθης. Για τον όρο λήθης χρησιμοποιούμε μία γραμμική συνάρτηση για το  $g(y_j)$ , δηλαδή

$$g(y_j) = \eta y_j \quad (3.10)$$

Επίσης για έναν νικητή νευρώνα  $i(x)$  ορίζουμε την έξοδο ως

$$y_j = h_{j,i(x)} \quad (3.11)$$

Συνδυάζοντας τα παραπάνω παίρνουμε:

$$\Delta w_j = \eta h_{j,i(x)}(x - w_j), \quad \left\{ \begin{array}{l} i: \text{ νικητής νευρώνας} \\ j: \text{ διεγερμένος νευρώνας} \end{array} \right\} \quad (3.12)$$

Τέλος χρησιμοποιώντας την λογική του διακριτού χρόνου, δοθέντος του συναπτικού διανύσματος βάρους  $w_j(n)$  του νευρώνα  $j$  το χρόνο  $n$ , ορίζουμε το

ανανεωμένο διάνυσμα βάρους  $w_j(n+1)$  στο χρόνο  $n+1$  με την εξίσωση

$$w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x(n) - w_j(n)) \quad (3.13)$$

η οποία εφαρμόζεται σε όλους τους νευρώνες του πλέγματος που βρίσκονται μέσα στην τοπολογική γειτονιά του νευρώνα νικητή. Η εξίσωση ανανέωσης έχει την ιδιότητα να μετακινεί το συναπτικό διάνυσμα βάρους του νικητή νευρώνα και των γειτόνων του προς το διάνυσμα εισόδου  $x$ . Μετά από επαναλαμβανόμενη παρουσίαση των δεδομένων εκπαίδευσης στον αλγόριθμο, τα διανύσματα των συναπτικών βαρών τείνουν να ακολουθούν την κατανομή των διανυσμάτων εισόδου. Για αυτό το λόγο ο αλγόριθμος οδηγεί σε τοπολογική διάταξη του χάρτη χαρακτηριστικών στο χώρο εισόδου, με την έννοια ότι νευρώνες που βρίσκονται κοντά στο πλέγμα θα τείνουν να έχουν όμοια συναπτικά διανύσματα βαρών.

Στο σημείο αυτό να προσθέσουμε ότι και η παράμετρος μάθησης  $\eta$  θα πρέπει να αλλάζει με το χρόνο ώστε να έχουμε μία στοχαστική προσέγγιση του αλγορίθμου. Συγκεκριμένα θα πρέπει να αρχίζει σε μία αρχική τιμή  $\eta_0$  και μετά να μειώνεται σταδιακά καθώς αυξάνεται ο χρόνος. Αυτή η προϋπόθεση μπορεί να ικανοποιηθεί με τον παρακάτω τρόπο:

$$\eta(n) = \eta_0 \exp\left(-\frac{n}{\tau_2}\right), \quad n = 0, 1, 2, \dots, \quad (3.14)$$

Όπου το  $\tau_2$  είναι μία ακόμα σταθερά του αλγορίθμου του χάρτη.

### 3.2.6 Οι Δύο φάσεις της Διαδικασίας Προσαρμογής: Διάταξης και Σύγκλισης

Αρχίζοντας από μία αρχική κατάσταση πλήρους αποδιοργάνωσης, ο αλγόριθμος του αυτο οργανούμενου χάρτη (AOX) καταφέρνει να οδηγήσει σε μία οργανωμένη αναπαραστάση των προτύπων ενεργοποίησης του χώρου εισόδου, δεδομένου ότι οι παράμετροι του αλγορίθμου διαλέγονται με σωστό τρόπο. Μπορούμε να διαχωρίσουμε το στάδιο της εκπαίδευσης σε δύο φάσεις: μία πρώτη φάση στην οποία γίνεται η κυρίως ξετύλιξη του χάρτη, αποκτώντας σχεδόν την τελική μορφή που θα έχει, και στη συνέχεια μία φάση σύγκλισης όπου συμβαίνουν μικροαλλαγές στο χάρτη με σκοπό την καλύτερη αναπαράσταση του χώρου εισόδου:

1. **Φάση Διάταξης:** Η πρώτη φάση της διαδικασίας προσαρμογής όπου συμβαίνει η τοπολογική διάταξη. Η διαδικασία αυτή μπορεί να χρειαστεί

παραπάνω από 1000 επαναλήψεις του αλγορίθμου. Για το λόγο αυτό πρέπει να γίνει προσεκτικός χειρισμός στην επιλογή των μεταβλητών ρυθμού μάθησης  $\eta(n)$  και των παραμέτρων της συνάρτησης γειτονιάς  $h_{j,i(x)}(n)$ .

- Γενικά στη φάση αυτή ο ρυθμός μάθησης πρέπει αρχικά να παίρνει σχετικά μεγάλες τιμές και σταδιακά να ελαττώνεται, βρισκόμενος όμως πάντα πάνω από ένα κατώφλι. Ο λόγος που αρχικά θέλουμε να παίρνει μεγάλες τιμές είναι για να έχουμε μεγάλες μετακινήσεις στο χάρτη και να γίνεται γρήγορα η ζετούλιξή του.
- Η συνάρτηση γειτονιάς θέλουμε αρχικά να συμπεριλαμβάνει όλους τους νευρώνες του πλέγματος με κέντρο τον νικητή νευρώνα και στη συνέχεια να μικραίνει αυτή την περιοχή ώστε τελικά η περιοχή αυτή να περιλαμβάνει μόνο τον ίδιο νικητή νευρώνα.

2. **Φάση σύγκλισης:** Η δεύτερη αυτή φάση της διαδικασίας προσαρμογής χρειάζεται για να μικρο-ρυθμίσουμε ορισμένες λεπτομέρειες στον χάρτη και συνεπώς να προσφέρουμε μία ακριβής στατιστική ποσοτικοποίηση του χώρου εισόδου. Γενικά οι επαναλήψεις του αλγορίθμου σε αυτή τη φάση εξαρτώνται σε μεγάλο βαθμό από την διάσταση της εισόδου, αλλά σαν γενικός κανόνας οι επαναλήψεις είναι περίπου 500 φορές ο αριθμός των νευρώνων στο πλέγμα. Αυτό σημαίνει ότι η συγκεκριμένη φάση μπορεί να χρειαστεί και δεκάδες χιλιάδες επαναλήψεις του αλγορίθμου. Για το λόγο αυτό η παράμετρος μάθησης και η συνάρτηση γειτονιάς πρέπει να οριστούν με παραμέτρους έτσι ώστε να βοηθήσουν τη διαδικασία .

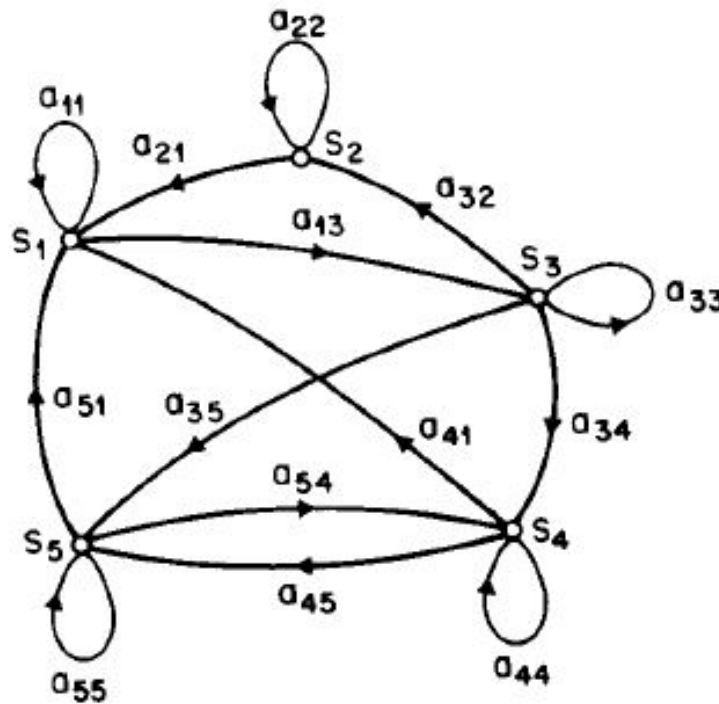
- Συγκεκριμένα για καλή στατιστική ακρίβεια ο ρυθμός μάθησης πρέπει να κρατηθεί σε χαμηλές τιμές στη φάση της σύγκλισης. Προσοχή πρέπει να δώθει ώστε η τιμή του να μην φτάσει ποτέ στο μηδέν καθώς μπορεί να δημιουργηθούν ανεπιθύμητες καταστάσεις στο χάρτη.
- Η τοπολογική γειτονιά  $h_{j,i(x)}$  πρέπει να περιέχει το πολύ τους κοντινότερους ή άμεσους γείτονες της στο πλέγμα (εδώ να σημειώσουμε ότι πολλές φορές υπάρχει μία σύγκληση στον διαχωρισμό του χάρτη από το πλέγμα, ο χάρτης είναι η αναπαράσταση των νευρώνων μέσα από τα συναπτικά τους βάρη τα οποία σχετίζονται με τη θέση των νευρώνων στο πλέγμα αλλά δεν ταυτίζονται, από την άλλη το πλέγμα είναι μία αρχική αναπαράσταση μίας, δύο, ή σπάνια τριών διαστάσεων δομή η οποία ορίζει γεωμετρικές ιδιότητες μεταξύ των νευρώνων, όπως την πραγματική τους θέση, τους γείτονές τους και τις απόστασεις μεταξύ τους).

Κάπου εδώ κλείνουμε με την παρουσίαση της θεωρίας για τους αυτο οργανούμενους χάρτες (Self Organizing Maps) και προχωράμε στην παρουσίαση της θεωρίας των Κρυφών Μαρκοβιανών Μοντέλων (KMM) από τα [15],[16],[17].

### 3.3 Κρυφά Μαρκοβιανά Μοντέλα (KMM)

#### 3.3.1 Διακριτές Διαδικασίες Μάρκοφ

Έστω σύστημα το οποίο μπορεί να περιγραφεί κάθε χρονική στιγμή ως μία από τις  $N$  διακριτές καταστάσεις  $S_1, S_2, \dots, S_N$  όπως φαίνεται και στο παρακάτω σχήμα. Σε τακτά χρονικά διαστήματα το σύστημα αλλάζει κατάσ-



Σχήμα 3.4: Αλυσίδα Μάρκοφ 5 καταστάσεων [15]

ταση (ίσως πίσω στην προηγούμενη), σύμφωνα με ένα σύνολο πιθανοτήτων που σχετίζονται με την κατάσταση. Συμβολίζουμε τις χρονικές στιγμές που σχετίζονται με την αλλαγή κατάστασης ως  $t = 1, 2, \dots$ , επίσης συμβολίζουμε την κατάσταση του συστήματος τον χρόνο  $t$  ως  $q_t$ . Για μία πλήρη πιθανοτική αναπαράσταση του συστήματος θα χρειαζόταν περιγραφή της τωρινής κατάστασης αλλά και όλων των προηγούμενων. Όμως στην ειδική περίπτωση των Αλυσίδων Μάρκοφ πρώτης τάξης που θα μας απασχολήσουν, η πιθανοτική αναπαράσταση

μειώνεται μόνο στην τωρινή και αμέσως προηγούμενη κατάσταση, δηλαδή

$$\begin{aligned} \Pr[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \dots] \\ = \Pr[q_t = S_j | q_{t-1} = S_i]. \end{aligned} \quad (3.15)$$

Επίσης ασχολούμαστε μόνο με τις διαδικασίες εκείνες όπου είναι ανεξάρτητες από το χρόνο, οδηγώντας σε ένα σύνολο πιθανοτήτων μετάβασης της μορφής

$$\Pr[q_t = S_j | q_{t-1} = S_i], \quad 1 \leq i, j \leq N \quad (3.16)$$

Με τους συντελεστές μετάβασης να ικανοποιούν τους στοχαστικούς περιορισμούς:

$$a_{ij} \geq 0 \quad (3.17)$$

$$\sum_{j=1}^N a_{ij} = 1 \quad (3.18)$$

Η παραπάνω στοχαστική διαδικασία ονομάζεται παρατηρήσιμο Μαρκοβιανό μοντέλο, καθώς η έξοδος του μοντέλου είναι το σύνολο των καταστάσεων σε κάθε χρονική στιγμή, όπου κάθε κατάσταση αντιστοιχεί σε κάποιο φυσικό γεγονός

### 3.3.2 Επέκταση της Θεωρίας στα Κρυφά Μαρκοβιανά Μοντέλα

Μέχρι στιγμής κάθε Μαρκοβιανό μοντέλο αποτελούταν από καταστάσεις οι οποίες αντιστοιχούσαν σε παρατηρήσιμα πραγματικά γεγονότα. Αυτό συντελεί έναν περιοριστικό παράγοντα για πολλές εφαρμογές. Στο κομμάτι αυτό θα θεωρήσουμε την περίπτωση όπου η παρατήρηση είναι μία πιθανοτική συνάρτηση της κατάστασης. Το κρυφό Μαρκοβιανό μοντέλο είναι μία διπλά ενσωματωμένη στοχαστική διαδικασία με μία διέπουσα στοχαστική διαδικασία η οποία δεν είναι παρατηρήσιμη (δηλαδή είναι κρυφή), και μπορεί να παρατηρηθεί από ένα άλλο σύνολο στοχαστικών διαδικασιών που παράγουν την ακολουθία των παρατηρήσεων.

### 3.3.3 Στοιχεία του KMM

Ένα Κρυφό Μαρκοβιανό μοντέλο χαρακτηρίζεται από τα παρακάτω στοιχεία:

1.  $N$ , ο αριθμός των καταστάσεων του μοντέλου. Αν και οι καταστάσεις είναι κρυφές, πολλές φορές υπάρχει φυσική σύνδεση τους με πραγματικά φαινόμενα. Γενικά οι καταστάσεις είναι συνδεδεμένες μεταξύ τους με τρόπο ώστε να υπάρχει πλήρης προσβασιμότητα από τη μία στην άλλη, φυσικά υπάρχουν και άλλοι τύποι μη πλήρων διασυνδέσεων που έχουν πρακτικό ενδιαφέρον. Συμβολίζουμε τις καταστάσεις με  $S = \{S_1, S_2, \dots, S_N\}$  και την κατάσταση το χρόνο  $t$  με  $q_t$
2.  $M$ , τον αριθμό των συμβόλων των παρατηρήσεων ανά κατάσταση, δηλαδή το μέγεθος του αλφαβήτου μας. Αυτό αντιστοιχεί με τη φυσική έξοδο του μοντέλου ή όπως είναι γνωστή ως η παρατήρηση. Συμβολίζουμε τα σύμβολα με  $V = \{v_1, v_2, \dots, v_m\}$
3. Την κατανομή πιθανότητας των μεταβάσεων κατάστασης  $A = \{a_{ij}\}$  όπου

$$a_{ij} = \Pr[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N \quad (3.19)$$

Για την ειδική περίπτωση όπου κάθε κατάσταση μπορεί να προσπελαστεί από οποιαδήποτε άλλη με ένα βήμα ισχύει ότι  $a_{ij} > 0 \forall i, j$ . Σε άλλες περιπτώσεις KMM θα μπορούσαμε να έχουμε πιθανότητες μετάβασης μηδενικές για ορισμένα ζευγάρια καταστάσεων.

4. Την κατανομή πιθανότητας των συμβόλων στην κατάσταση  $j$ ,  $B = \{b_j(k)\}$  όπου

$$b_j(k) = \Pr[v_k \text{ at } t | q_t = S_j], \quad \begin{matrix} 1 \leq j \leq N \\ 1 \leq k \leq M \end{matrix} \quad (3.20)$$

5. Την κατανομή πιθανότητας των αρχικών καταστάσεων (δηλαδή πόσο πιθανό είναι να ξεκινήσει το μοντέλο από μία συγκεκριμένη κατάσταση)  $\pi = \{\pi_i\}$  με

$$\pi_i = \Pr[q_1 = S_i], \quad 1 \leq i \leq N \quad (3.21)$$

Δοθέντος κατάλληλων τιμών για τα  $N, M, A, B, \pi$  το KMM μπορεί να χρησιμοποιηθεί ως γεννήτρια για να παράξει μία ακολουθία παρατηρήσεων,

$$O = O_1 O_2 \cdots O_T \quad (3.22)$$

(όπου κάθε παρατήρηση  $O_t$  είναι ένα από τα σύμβολα στο  $V$  και  $T$  είναι ο αριθμός των παρατηρήσεων στην ακολουθία), με τον ακόλουθο τρόπο :

1. Διάλεξε μία αρχική κατάσταση  $q_1 = S_i$  σύμφωνα με την αρχική κατανομή  $\pi$
2. Θέσε  $t = 1$
3. Διάλεξε  $O_t = v_k$  σύμφωνα με την κατανομή των συμβόλων στην κατάσταση  $S_i$  δηλαδή την  $b_i(k)$
4. Πήγαινε σε μία νέα κατάσταση  $q_{t+1} = S_j$  σύμφωνα με τις πιθανότητες μεταβάσεων της κατάστασης  $S_i$  δηλαδή την  $a_{ij}$
5. Θέσε  $t = t + 1$  και επανάλαβε την διαδικασία από το βήμα 3 αν  $t \leq T$ , αλλιώς σταμάτησε τη διαδικασία.

Από εδώ και πέρα θα λέμε ότι ένα Μαρκοβιανό μοντέλο χαρακτηρίζεται πλήρως από τις παραμέτρους του  $A, B, \pi$ . Θα το συμβολίζουμε συνεπώς:

$$\lambda = (A, B, \pi) \quad (3.23)$$

### 3.3.4 Τα τρία βασικά προβλήματα του χαρακτηρίζουν τα Κρυφά Μαρκοβιανά μοντέλα

Γενικά όταν έχουμε ένα τέτοιο μοντέλο υπάρχουν τρία βασικά προβλήματα που πρέπει να λυθούν ώστε να είναι χρήσιμο σε εφαρμογές. Αυτά είναι τα ακόλουθα:

1. Δοθέντος της ακολουθίας παρατηρήσεων  $O = O_1 O_2 \cdots O_T$  και του μοντέλου  $\lambda = (A, B, \pi)$  πώς υπολογίζουμε αποτελεσματικά την πιθανότητα της ακολουθίας παρατηρήσεων  $\Pr(O|\lambda)$  δοθέντος του μοντέλου ?

2. Δοθέντος της ακολουθίας παρατηρήσεων  $O = O_1 O_2 \cdots O_T$ , και του μοντέλου  $\lambda$ , πώς διαλέγουμε την βέλτιστη ακολουθία καταστάσεων  $Q = q_1 q_2 \cdots q_t$  ?
3. Πώς προσαρμόζουμε τις παραμέτρους του μοντέλου  $\lambda = (A, B, \pi)$  για να μεγιστοποιήσουμε την  $\Pr(O|\lambda)$  ?

Το πρώτο είναι το πρόβλημα της αξιολόγησης, δηλαδή, έχοντας ένα μοντέλο και μία ακολουθία παρατηρήσεων, πώς υπολογίζουμε την πιθανότητα η ακολουθία αυτή να έχει παραχθεί από το μοντέλο μας.

Το δεύτερο πρόβλημα έχει στόχο την ανακάλυψη του κρυμμένου κομματιού του μοντέλου μας, δηλαδή της σωστής ακολουθίας των καταστάσεων. Εδώ πρέπει να τονίσουμε ότι η λύση δεν είναι μοναδική και εξαρτάται από τον τρόπο που θα ορίσουμε εμείς το τί σημαίνει σωστή ακολουθία.

Το τρίτο πρόβλημα έχει στόχο την βελτιστοποίηση των παραμέτρων του μοντέλου έτσι ώστε να περιγράφει με τον καλύτερο τρόπο μία συγκεκριμένη ακολουθία παρατηρήσεων. Η ακολουθία αυτή ονομάζεται ακολουθία εκπαίδευσης, καθώς χρησιμοποιείται για να εκπαιδεύσει το μοντέλο μας.

### 3.3.5 Λύσεις των τριών προβλημάτων του KMM

#### Λύση στο πρόβλημα 1

Θέλουμε να υπολογίσουμε την πιθανότητα της ακολουθίας παρατηρήσεων  $O = O_1 O_2 \cdots O_T$  δοθέντος του μοντέλου  $\lambda$  δηλαδή την  $\Pr(O|\lambda)$ . Ο πιο απλός τρόπος για να γίνει αυτό είναι μετρώντας κάθε πιθανή ακολουθία καταστάσεων. Προφανώς αυτός ο τρόπος αν αναλογιστούμε ότι έχουμε  $N$  καταστάσεις και η ακολουθία έχει μήκος  $T$  οδηγεί σε υπολογισμούς της τάξης του  $N^T$ . Έστω ακολουθία

$$Q = q_1 q_2 \cdots q_T \quad (3.24)$$

με  $q_1$  την αρχική κατάσταση. Η πιθανότητα της ακολουθίας παρατήρησης  $O$  για την συγκεκριμένη ακολουθία καταστάσεων είναι :

$$\Pr(O|Q, \lambda) = \prod_{t=1}^T \Pr(O_t/q_t, \lambda) \quad (3.25)$$



υποθέτοντας στατιστική ανεξαρτησία των παρατηρήσεων. Άρα παίρνουμε

$$\Pr(O|Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \cdots b_{q_T}(O_T) \quad (3.26)$$

Η πιθανότητα μίας τέτοιας ακολουθίας καταστάσεων μπορεί να γραφεί ως

$$\Pr(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T} \quad (3.27)$$

Η από κοινού πιθανότητα των  $O$  και  $Q$ , δηλαδή η πιθανότητα αυτά τα δύο να συμβούν ταυτοχρόνως είναι απλώς το γινόμενο των δύο παραπάνω όρων, δηλαδή

$$\Pr(O, Q|\lambda) = \Pr(O|Q, \lambda) \Pr(Q, \lambda). \quad (3.28)$$

Η πιθανότητα του  $O$  (δεδομένου του μοντέλου) αποκτάται αθροίζοντας την από κοινού πιθανότητα πάνω σε όλες τις δυνατές ακολουθίες καταστάσεων  $q$  δίνοντας

$$\Pr(O|\lambda) = \sum_{all\ Q} \Pr(O|Q, \lambda) \Pr(Q|\lambda) = \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \cdots a_{q_{T-1} q_T} b_{q_T}(O_T) \quad (3.29)$$

Ο παραπάνω τύπος μεταφράζεται ως: Αρχικά είμαστε στην κατάσταση  $q_1$  με πιθανότητα  $\pi_{q_1}$  και παράγουμε το σύμβολο  $O_1$  με πιθανότητα  $b_{q_1}$ . Ο χρόνος περνάει στην επόμενη χρονική στιγμή και μεταβαίνουμε στην επόμενη κατάσταση παράγοντας το επόμενο σύμβολο. Αυτό συνεχίζει μέχρι να παράξουμε όλη την ακολουθία. Τέλος αθροίζουμε όλες τις πιθανές ακολουθίες καταστάσεων που μπορούν να παράξουν τη συγκεκριμένη ακολουθία.

Όμως όπως αναφέρθηκε και προηγουμένως ο απευθείας υπολογισμός της παραπάνω πιθανότητας είναι υπολογιστικά πολύ μεγάλος της τάξης του  $N^T$ , επιτακτική είναι η ανάγκη για έναν γρηγορότερο τρόπο υπολογισμού. Τέτοιος τρόπος υπάρχει, βασίζεται δε στην ισχύ του δυναμικού προγραμματισμού και ονομάζεται Μπρος-Πίσω (forward-backward) αλγόριθμος, του οποίου σε αυτό το σημείο χρειαζόμαστε μόνο το Προς τα Μπροστά (forward) κομμάτι του.

Στη συνέχεια θα γίνει παρουσίαση του Εμπρόσθιου αλγορίθμου. Έστω η μεταβλητή forward  $\alpha_t(i)$  που ορίζεται ως

$$\alpha_t(i) = \Pr(O_1 O_2 \cdots O_t, q_t = S_i | \lambda) \quad (3.30)$$

να είναι η πιθανότητα της ακολουθίας παρατήρησης μέχρι την στιγμή  $t$  και η κατάσταση εκείνη τη στιγμή να είναι η  $S_i$  δοθέντος του μοντέλου  $\lambda$ . Μπορούμε να λύσουμε ως προς  $\alpha_t(i)$  επαγωγικά με τον εξής τρόπο:

1. Αρχικοποίηση:

$$\alpha_1(i) = \pi b_i(O_1), \quad 1 \leq i \leq N, \quad (3.31)$$

2. Επαγωγή:

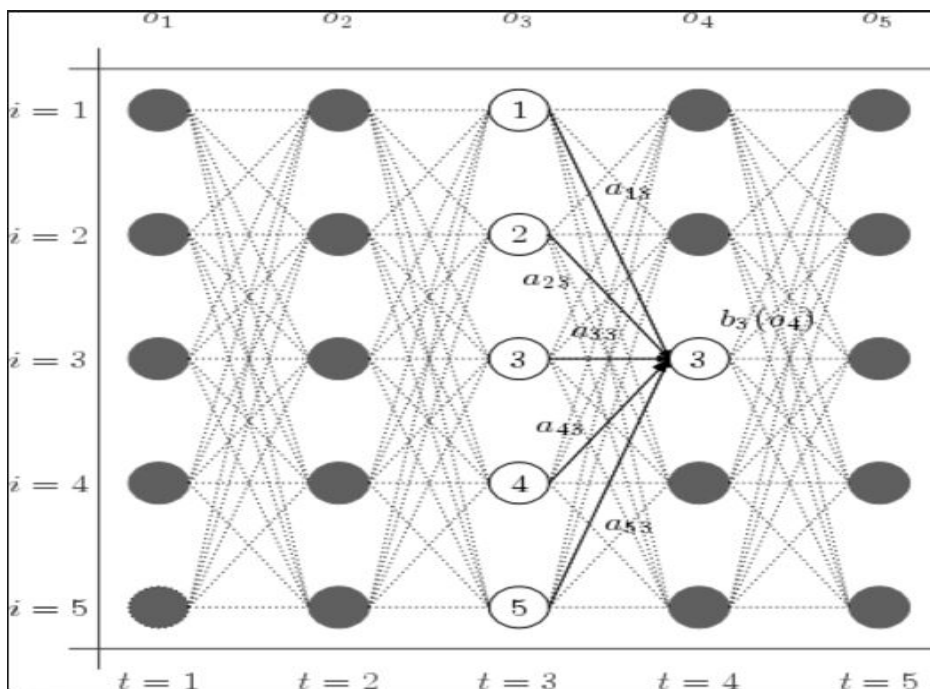
$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1 \quad (3.32)$$

$$1 \leq j \leq N \quad (3.33)$$

3. Τερματισμος:

$$\Pr(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (3.34)$$

Το πρώτο βήμα αρχικοποιεί την Εμπρόσθια πιθανότητα ως την από κοινού πιθανότητα της κατάστασης  $S_i$  και της αρχικής παρατήρησης  $O_1$ . Το δεύτερο βήμα, το βήμα της επαγωγής που είναι και η καρδιά του αλγορίθμου, αθροίζει τις πιθανότητες να είμαστε σε κάθε μια από τις  $N$  καταστάσεις στο χρόνο  $t$  και να είμαστε στην κατάσταση  $S_j$  το χρόνο  $t+1$ , πάλι για όλες τις  $N$  καταστάσεις, στο τέλος οι πιθανότητες που προκύπτουν είναι η συνολικές πιθανότητες να βρισκομαστε σε κάθε μία από τις  $N$  ξεχωριστές καταστάσεις έχοντας παρατηρήσει όλη την ακολουθία, έτσι στο τρίτο και τελευταίο βήμα απλά αθροίζουμε αυτές τις πιθανότητες.



Σχήμα 3.5: Απεικόνιση του Εμπρόσθιου Αλγορίθμου που βασίζεται στον δυναμικό προγραμματισμό (πηγή: Βικιπένδια)

Στο σημείο αυτό να σημειωθεί ότι με τη χρησιμοποίηση δυναμικού προγραμματισμού ουσιαστικά μειώσαμε την πολυπλοκότητα του αλγορίθμου σε  $N^2 \cdot T$  περίπου .

Με το ίδιο τρόπο που ορίσαμε την Εμπρόσθια μεταβλητή, μπορούμε να ορίσουμε και την Οπίσθια μεταβλητή  $\beta_t(i)$  ως:

$$\beta_t(i) = P(O_{t+1}O_{t+2} \dots O_T | q_t = S_i, \lambda) \quad (3.35)$$

η οποία περιγράφει την πιθανότητα να παρατηρήσουμε την ακολουθία από το χρόνο  $t+1$  μέχρι το τέλος και την χρονική στιγμή  $t$  να είμαστε στην κατάσταση  $S_i$  δεδομένου του μοντέλου  $\lambda$ . Με το ίδιο τρόπο χρησιμοποιώντας επαγωγή ακολουθούμε τα παρακάτω βήματα για να βρούμε το  $\beta$  σε κάθε στάδιο:

1. Αρχικοποίηση:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (3.36)$$

## 2. Επαγωγή

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j),$$

$$t = T - 1, T - 2, \dots, 1, 1 \leq i \leq N. \quad (3.37)$$

### Λύση στο Πρόβλημα 2

Αντιθέτως με το πρόβλημα 1 στο οποίο μπορεί να δοθεί ακριβής λύση, το πρόβλημα 2 έχει διάφορες πιθανές λύσεις αναλόγως τον τρόπο που θα ορίσουμε ποιά είναι η βέλτιστη ακολουθία καταστάσεων που σχετίζεται με την ακολουθία παρατηρήσεων. Το κριτήριο που θα χρησιμοποιήσουμε εμείς, το οποίο μάλιστα είναι και το πιο διαδεδομένο είναι να βρούμε τη μοναδική εκείνη ακολουθία(μονοπάτι) για να μεγιστοποιήσουμε το  $P(Q|O, \lambda)$ . Ο αλγόριθμος για την επίτευξη της παραπάνω εργασίας βασίζεται και αυτός στον δυναμικό προγραμματισμό και ονομάζεται αλγόριθμος Viterbi

Αλγόριθμος Viterbi: Για να βρούμε την μοναδική καλύτερη ακολουθία καταστάσεων  $Q = \{q_1, q_2, \dots, q_T\}$  για την δοσμένη ακολουθία  $O = \{O_1 O_2 \dots O_T\}$  πρέπει να ορίσουμε την ποσότητα

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} \Pr[q_1 q_2 \dots q_t = i, O_1 O_2 \dots O_t | \lambda] \quad (3.38)$$

όπου το  $\delta$  είναι το καλύτερο σκορ (υψηλότερη πιθανότητα) κατά μήκος ενός μονοπατιού το χρόνο  $t$  που μετράει τις πρώτες  $t$  παρατηρήσεις και τελειώνει στην κατάσταση  $S_i$ . Με επαγωγή έχουμε

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] \cdot b_j(O_{t+1}) \quad (3.39)$$

Για να ανακτήσουμε την ακολουθία καταστάσεων, πρέπει να κρατήσουμε σε κάθε βήμα τους δείκτες που μας δίνουν το μέγιστο για κάθε  $t, j$ . Αυτό το κά-  
νουμε δια μέσου του πίνακα  $\psi_t(j)$ . Η πλήρης διαδικασία περιγράφεται παρακάτω:

#### 1. Αρχικοποίηση:

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (3.40)$$

$$\psi_1(i) = 0 \quad (3.41)$$

2. Αναδρομή:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T$$

$$1 \leq j \leq N \quad (3.42)$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T$$

$$1 \leq j \leq N \quad (3.44)$$

3. Τερματισμός:

$$p^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (3.45)$$

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)] \quad (3.46)$$

4. Ανάκτηση μονοπατιού προς τα πίσω:

$$q_t^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1 \quad (3.47)$$

Μπόρει να καταστεί προφανές ότι ο αλγόριθμος viterbi είναι σχεδόν ίδιος με τον Εμπρόσθιο αλγόριθμο, με τη διαφορά ότι αντί σε κάθε βήμα να παίρνουμε το άθροισμα των μεταβλητών του προηγούμενου βήματος παίρνουμε το μέγιστο από αυτές.

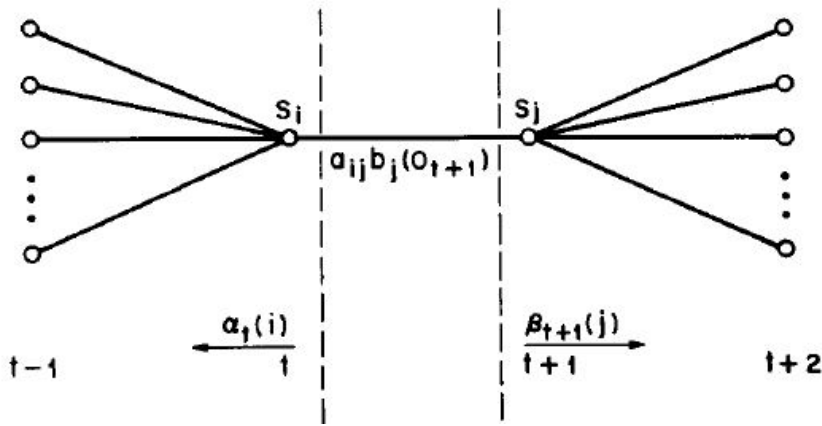
### **Λύση του τρίτου προβλήματος**

Το τρίτο πρόβλημα το οποίο είναι και το δυσκολότερο στα Κρυφά Μαρκοβιανά Μοντέλα είναι ο προσδιορισμός μίας μεθόδου για την προσαρμογή των μεταβλητών του μοντέλου  $(A, B, \pi)$  με στόχο τη μεγιστοποίηση της πιθανότητας της ακολουθίας παρατηρήσεων δοθέντος του μοντέλου. Καθώς πρόκειται για ένα πρόβλημα με μεγάλη πολυπλοκότητα ακόμα δεν υπάρχει αναλυτικός τρόπος για τη λύση του αλλά και ούτε βέλτιστος τρόπος. Ωστόσο υπάρχει η δυνατότητα

επιλογής του μοντέλου  $\lambda = (A, B, \pi)$  έτσι ώστε το  $P(O|\lambda)$  να μεγιστοποιείται τοπικώς χρησιμοποιώντας μία επαναληπτική διαδικασία όπως ο Baum-Welch αλγόριθμος, που είναι μία υπό κατηγορία του αλγορίθμου Μεγιστοποίησης Προσδοκίας.

Αρχικά ορίζουμε τη μεταβλητή  $\xi_t(i, j)$  όπου είναι η πιθανότητα να είμαστε στην κατάσταση  $S_i$  το χρόνο  $t$  και στην κατάσταση  $S_j$  τον χρόνο  $t+1$  δοθέντος του μοντέλου και των παρατηρήσεων, δηλαδή

$$\xi_t(i, j) = \Pr(q_t = S_i, q_{t+1} = S_j | O_t, \lambda) \quad (3.48)$$



Σχήμα 3.6: Παρουσίαση της ακολουθίας καταστάσεων που απαιτούνται για τον υπολογισμό του γεγονότος το σύστημα να είναι στην κατάσταση  $S_i$  τον χρόνο  $t$  και στην κατάσταση  $S_j$  τον χρόνο  $t+1$  [15]

Σύμφωνα με τις μεταβλητές του Εμπρόσθιου και του Οπίσθιου αλγορίθμου μπορούμε να γράψουμε το  $\xi_t(i, j)$  στη μορφή

$$\xi_t(i, j) = a_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (3.49)$$

$$= \frac{a_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N a_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \quad (3.50)$$

όπου ο αριθμητής είναι  $\Pr(q_t = S_i, q_{t+1} = S_j, O|\lambda)$  και ο παρονομαστής είναι ο όρος κανονικοποίησης, δηλαδή το  $P(O|\lambda)$

Ορίζουμε το  $\gamma_t(i)$  ως την πιθανότητα του να είσαι στην κατάσταση  $S_i$  τη στιγμή  $t$ , δοθέντος της ακολουθίας παρατήρησεων και του μοντέλου. Έτσι μπορούμε να συσχετίσουμε το  $\gamma_t(i)$  με το  $\xi_t(i, j)$  αθροίζοντας πάνω στα  $j$  παίρνοντας

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (3.51)$$

τώρα, αθροίζοντας τα  $\gamma_t(i)$  πάνω στο  $t$ , παίρνουμε μία ποσότητα που μπορεί να μεταφραστεί ως ο αριθμός των φορών που περιμένουμε να επισκεφθούμε την κατάσταση  $S_i$ , ή αντίστοιχα τον προσδοκώμενο αριθμό μεταβάσεων στην κατάσταση αυτή. Όμοια η άθροιση του  $\xi_t(i, j)$  πάνω στο  $t$  μας δίνει αντίστοιχα τον προσδοκώμενο αριθμό μεταβάσεων από την κατάσταση  $i$  στην  $j$ . Δηλαδή

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{Προσδοκώμενος αριθμός μεταβάσεων από την } S_i \quad (3.52)$$

$$(3.53)$$

$$\sum_{t=1}^{T-1} \xi_t(i, j) = \text{Προσδοκώμενος αριθμός μεταβάσεων από την } S_i \text{ στην } S_j \quad (3.54)$$

Χρησιμοποιώντας τις παραπάνω φόρμουλες μπορούμε να βρούμε μία μέθοδο επανεκτίμησης των παραμέτρων του KMM. Ένα σύνολο λογικών τέτοιων τύπων

είναι :

$$\tilde{\pi}_i = \frac{\text{προσδοκώμενη συχνότητα στην κατάσταση } S_i \text{ το χρόνο } t = 1}{\text{δηλαδή το } \gamma_1(i)} \quad (3.55)$$

$$\begin{aligned} \tilde{a}_{ij} &= \frac{\text{προσδοκώμενος αριθμός μεταβάσεων από την κατάσταση } S_i \text{ στην } S_j}{\text{προσδοκώμενος αριθμός μεταβάσεων από την κατάσταση } S_i} \\ &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \end{aligned} \quad (3.56)$$

$$\begin{aligned} \tilde{b}_j(k) &= \frac{\text{προσδοκώμενος αριθμός να είσαι στην κατάσταση } j \text{ και να παρατηρείς το σύμβολο } v_k}{\text{προσδοκώμενος αριθμός να είσαι στην κατάσταση } j} \\ &= \frac{\sum_{\substack{t=1 \\ s.t. O_t=v_k}} \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)} \end{aligned} \quad (3.57)$$

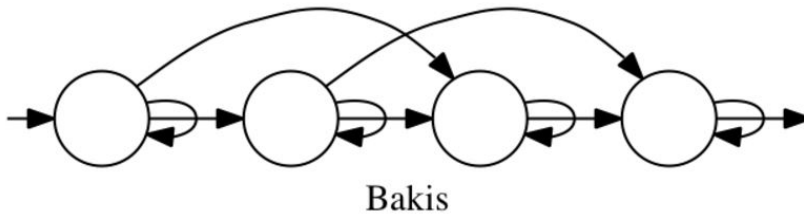
Αν ορίσουμε το τωρινό μοντέλο ως  $\lambda = (A, B, \pi)$  και χρησιμοποιήσουμε τους παραπάνω τύπους για να βρούμε το καινούργιο μοντέλο με τις νέες παραμέτρους  $\tilde{\lambda} = (\tilde{A}, \tilde{B}, \tilde{\pi})$ , τότε έχει αποδειχθεί από τον Baum και τους συνεργάτες του ότι είτε το αρχικό μοντέλο είναι ακριβώς ίδιο με το επαναυπολογισμένο, είτε ότι το επαναυπολογισμένο είναι πιο πιθανό, με την έννοια ότι περιγράφει καλύτερα την ακολουθία παρατηρήσεων  $P(O|\lambda) \leq P(O|\tilde{\lambda})$ .

Έτσι βασιζόμενοι στο προηγούμενο θεώρημα, ακολουθούμε επαναληπτικά την διαδικασία μέχρι να βρούμε το καλύτερο δυνατό μοντέλο που να περιγράφει την ακολουθία μας.

Στο σημείο αυτό θα κλείσουμε τη θεωρία των Μαρκοβιανών μοντέλων αναφέροντας ένα τύπο Κρυφού Μαρκοβιανού μοντέλου το bakis ή left-right model. Πρόκειται για έναν ιδιαίτερο τύπο KMM ο οποίος χρησιμοποιείται όταν το πρόβλημα που έχουμε να αντιμετωπίσουμε περιέχει ακολουθιακά δεδομένα, όπως πχ στην αναγνώριση φωνής, ή στην περίπτωση του θέματος που ασχολούμαστε σε αυτή



τη διπλωματική, στην αναγνώριση χειρονομιών. Ο συγκεκριμένος τύπος έχει το χαρακτηριστικό να μην έχει πλήρως συνδεδεμένες μεταξύ τους καταστάσεις (δηλαδή από οποιαδήποτε κατάσταση η μηχανή να μπορεί να μεταβεί σε οποιαδήποτε άλλη), όπως συνηθίζεται και συγκεκριμένα αρχίζοντας από μία αρχική κατάσταση και αν σκεφτούμε ότι όλες οι καταστάσεις είναι διατεταγμένες σε μία γραμμή, υπάρχει η ικανότητα μετάβασης μόνο σε μία μεταγενέστερη κατάσταση ή να παραμείνει στην ίδια.



Σχήμα 3.7: Μοντέλο bakis

# Κεφάλαιο 4

## Το Σύστημα μας

### 4.1 Το Σύνολο δεδομένων (Dataset)

Όπως έχει αναφερθεί ήδη, το κυριάρχο θέμα της παρούσας διπλωματικής εργασίας είναι η αναγνώριση χειρονομιών. Για το λόγο αυτό είναι απαραίτητο να αποκτήσουμε τα κατάλληλα δεδομένα για την μελέτη του προβλήματος αυτού. Αυτά τα δεδομένα μας παρέχονται από το πανεπιστήμιο του Texas, Dallas των Ηνωμένων Πολιτειών της Αμερικής [18]. Πρόκειται για ένα συγκεντρωτικό σύνολο δεδομένων που περιέχει συνδυασμό χειρονομιών που εκτελούνται με τα χέρια αλλά και με τα πόδια. Εμείς ασχολούμαστε με την αναγνώριση χειρονομιών που βασίζονται σε κινήσεις του χεριού. Συνεπώς απομονώσαμε το κομμάτι των χειρονομιών εκείνο που μας ενδιαφέρει.

Το συγκεκριμένο σύνολο δεδομένων αποτελείται από RGB βίντεο, εικόνες βάθους, δεδομένα θέσεων αρθρώσεων των ανθρώπων που εκτελούν τις χειρονομίες (σκελετούς) και σήματα όπως γραμμικές και γωνιακές επιταχύνσεις των περιοχών ενδιαφέροντος (χέρι και πόδι). Όλα αυτά τα δεδομένα έχουν αποκτηθεί από μία Kinect κάμερα και από έναν αισθητήρα μέτρησης επιταχύνσεων (αξελερόμετρο) ο οποίος φοριέται από τον εκτελεστή της χειρονομίας.

Η κάμερα που χρησιμοποιήθηκε έχει την δυνατότητα να τραβήξει εικόνα χρώματος με ανάλυση  $640 \times 480$  πίξελ και 16-bit εικόνα βάθους. Ο ρυθμός πλαισίων είναι περίπου 30 πλαίσια ανά δευτερόλεπτο. Με τη βοήθεια του Kinect SDK ανιχνεύθηκαν οι 20 αρθρώσεις του σκελετού και οι 3Δ χωρικές τους θέσεις.

Ο αισθητήρας ανίχνευσης κίνησης που χρησιμοποιήθηκε είναι κατασκευή του πανεπιστημίου του Τέξας ο οποίος έχει μικρό κόστος. Αποτελείται από (α) έναν αισθητήρα MEMS 9 αξόνων, με δυνατότητα μέτρησης επιτάχυνσης 3

αξόνων, 3 αξόνων γωνιακή ταχύτητα και 3 αξόνων μαγνητική δύναμη ( $\beta$ ) έναν μικροελεγκτή 16-bit χαμηλής ισχύος ( $\gamma$ ) ένα χαμηλής ισχύος δέκτη Bluetooth που στέλνει τα δεδομένα ασύρματα στον υπολογιστή και ( $\delta$ ) μία πλατφόρμα για την δυνατότητα εντολών ελέγχου και μετάδοσης δεδομένων. Γενικά στα δεδομένα δεν χρησιμοποιήθηκαν τα μαγνητικά δεδομένα. Ο ρυθμός δειγματισμού του αξελερόμετρου ήταν 50 κύκλοι και το εύρος μέτρησης του ήταν  $\pm 8g$  για την επιτάχυνση και  $\pm 1000$  μοίρες/δευτερόλεπτο για την περιστροφή.

Το σύνολο δεδομένων περιέχει 27 δράσεις που εκτελούνται από 8 άτομα (4 άντρες και 4 γυναίκες). Κάθε εκτελεστής επαναλαμβάνει 4 φορές κάθε χειρονομία. Το σύνολο δεδομένων περιέχει 861 ακολουθίες δεδομένων. Περιέχονται διάφορες χειρονομίες από πολυποίκιλες δραστηριότητες της καθημερινής ζωής, όπως για παράδειγμα αθλητικές δραστηριότητες, καθημερινές δράσεις και χειρονομίες χεριού. Παρακάτω δίνεται ο πίνακας που περιέχει αναλυτικά τις χειρονομίες. Από αυτές τις χειρονομίες κρατήσαμε μόνο όσες σχετίζονταν με το δικό μας πρόβλημα, δηλαδή αποτελούνταν από κινήσεις του ενός μόνο χεριού, δηλαδή κρατήσαμε τις χειρονομίες με αριθμούς 1, 2, 3, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 20, 21. Στη συνέχεια δείχνουμε στιγμιότυπα από τις χειρονομίες που χρησιμοποιήσαμε έτσι όπως εκτελέστηκαν από τους πρωταγωνιστές

Με αυτές τις εικόνες κλείνουμε λοιπόν την παρουσίαση του συνόλου δεδομένων που χρησιμοποιήσαμε και συνεχίζουμε με την παρουσίαση του συστήματός μας.

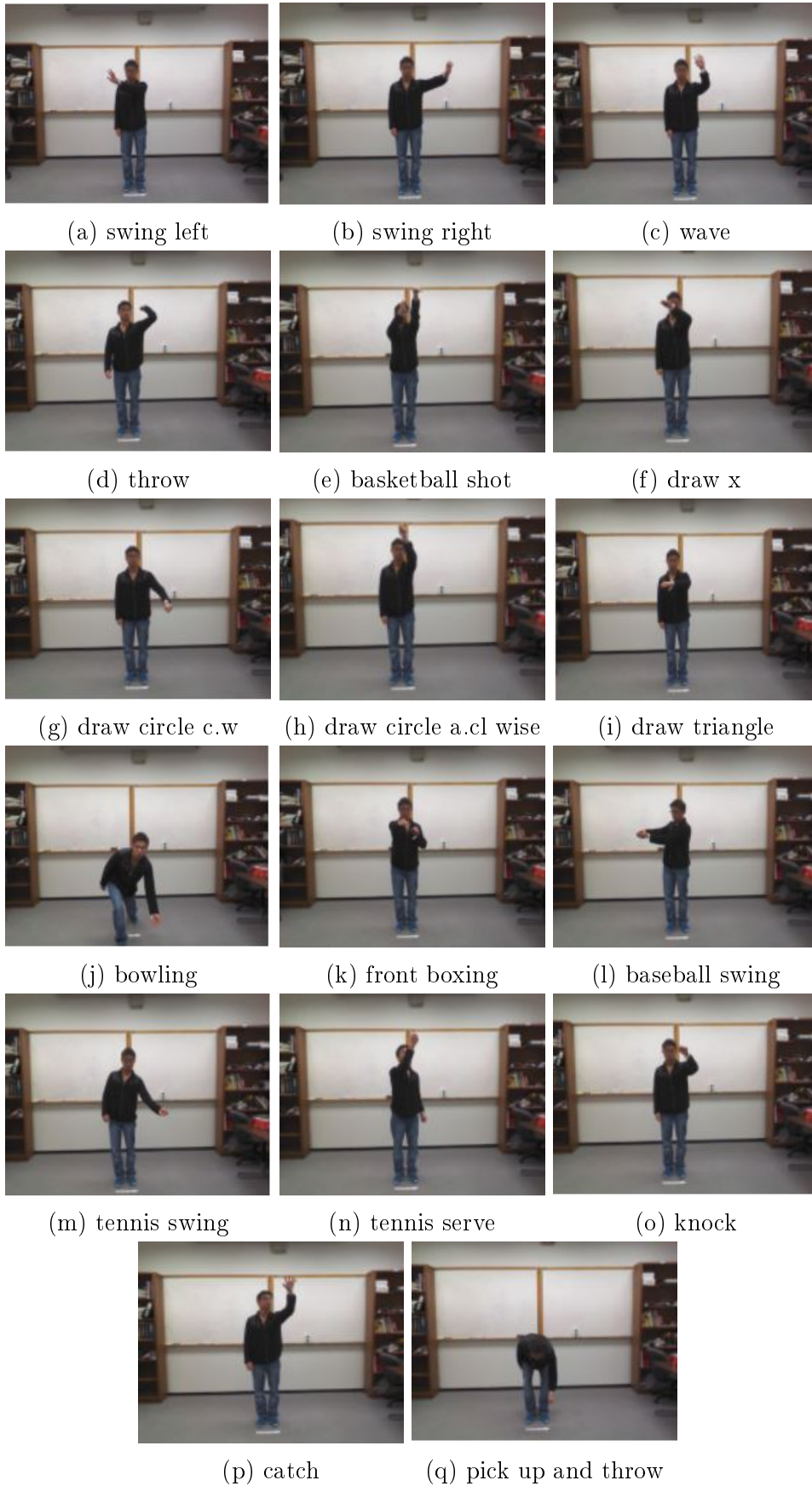
## 4.2 Το Σύστημα Αναγνώρισης

### 4.2.1 Γενικά

Το σύστημα που χρησιμοποιήσαμε για την αναγνώριση αποτελεί μία επέκταση της δουλειάς στο [19], κάνοντας χρήση ακολουθιών για αναγνώριση στις τρεις διαστάσεις. Πρόκειται για μία μέθοδο που συνδυάζει τους αυτοοργανούμενους χάρτες, τα Μαρκοβιανά μοντέλα και άλλες στατιστικές μετρικές για την αποτελεσματική αναγνώριση δυναμικών χειρονομιών οι οποίες εκτελούνται με τη χρήση του χεριού των υποκειμένων. Οι αυτοοργανούμενοι χάρτες προσφέρουν την ικανότητα μοντελοποίησης των χειρονομιών σε ακολουθίες από σύμβολα, στη συνέχεια αυτά τα σύμβολα χρησιμοποιούνται για ( $\alpha$ ) την εξαγωγή στατιστικών μετρικών ( $\beta$ ) για την εκπαίδευση διαφόρων μοντέλων Μαρκόβ όπως απλές Μαρκοβιανές Αλυσίδες (Markov Chains) και Κρυφά Μοντέλα Μαρκόβ (Hidden Markov Models), όπου θα χρησιμοποιηθούν στη συνέχεια για τη διεκπεραίωση της εργασίας της αυτόματης αναγνώρισης.

| Χειρονομίες όπου φοριέται το αξελερόμετρο στο δεξί χέρι |   |                    |
|---|---|--------------------|
| 1   | μετακίνηση δεξιού χεριού από δεξιά στα αριστερά                   | (swipe_left)       |
| 2   | μετακίνηση δεξιού χεριού δεξιά                                    | (swipe_right)      |
| 3   | χαρτισμός δεξιού χεριού   | (wave)             |
| 4   | χειροκρότημα δύο χεριών   | (clap)             |
| 5   | πέταγμα με δεξιά χέρι   | (throw)            |
| 6   | δίπλωμα χεριών στο στήθος   | (arm_cross)        |
| 7   | βολή μπάσχετ  | (basketball_shoot) |
| 8   | σχηματισμός του X με το δεξί χέρι                                 | (draw_x)           |
| 9   | σχηματισμός κύκλου με το δεξί χέρι(με φορά ρολογιού)              | (draw_circle_CW)   |
| 10  | σχηματισμός κύκλου με το δεξί χέρι(αντίθετα από τη φορά ρολογιού) | (draw_circle_CCW)  |
| 11  | σχηματισμός τριγώνου  | (draw_triangle)    |
| 12  | μποουλινγκ  | (bowling)          |
| 13  | μπροστινό μπόξ  | (boxing)           |
| 14  | χτύπημα του μπέιζμπολ από τα δεξιά                                | (baseball_swing)   |
| 15  | χτύπημα του τένις   | (tennis_swing)     |
| 16  | λύγισμα χεριού  | (arm_curl)         |
| 17  | σερβίς τένις  | (tennis_serve)     |
| 18  | σπρώξιμο με δύο χέρια   | (push)             |
| 19  | χτύπημα πόρτας  | (knock)            |
| 20  | πιάσιμο αντικειμένου με δεξιά χέρι                                | (catch)            |
| 21  | σηκώνω και ρίχνω με δεξιά χέρι                                    | (pickup_throw)     |
| Χειρονομίες όπου φοριέται το αξελερόμετρο στο δεξί πόδι |   |                    |
| 22  | επι τόπου τρέξιμο   | (jog)              |
| 23  | επι τόπου περπάτημα   | (walk)             |
| 24  | κάθονται και σηκώνονται   | (sit2stand)        |
| 25  | σηκώνονται και κάθονται   | (stand2sit)        |
| 26  | προβολή ποδιού  | (lunge)            |
| 27  | κάθισμα   | (squat)            |

Πίνακας 4.1: Πίνακας των Χειρονομιών που περιέχει το σύνολο δεδομένων [18]



Σχήμα 4.1: Οπτικοποίηση των Χειρονομιών [18]

## 4.2.2 Μοντελοποίηση των Χειρονομιών

Όπως έχει αναφερθεί και προηγουμένως εργαζόμαστε πάνω σε ένα σύνολο δεδομένων το οποίο αποτελείται από 17 κλάσεις χειρονομιών. Για μία πιο γενική περιγραφή δεν θα αναφερθούμε ειδικά σε αυτές αλλά θα επεκτείνουμε τη συζήτηση υποθέτοντας πως έχουμε  $c$  διαφορετικές κλάσεις. Άρα έστω

$$D = \{G_1, \dots, G_c\} \quad (4.1)$$

το σύνολο δεδομένων μας το οποίο περιέχει  $c$  κλάσεις χειρονομιών  $G_1, \dots, G_c$ . Κάθε κλάση με τη σειρά της περιέχει ένα σύνολο χειρονομιών που εμπίπτουν σε αυτή την κλάση, δηλαδή,

$$G_i = \{G_{i1}, G_{i2}, \dots, G_{in}\}. \quad (4.2)$$

Για ευκολία θα αναφερομαστε σε κάθε χειρονομία ως  $G$  και θα ξεκαθαρίζεται από το νόημα αν αναφερόμαστε σε κλάση ή σε συγκεκριμένο δείγμα χειρονομίας του συνόλου δεδομένων.

Οι χειρονομίες που χρησιμοποιηθήκαν τέθηκαν αντικείμενο μοντελοποίησης με δύο τρόπους. Ο πρώτος κάνει χρήση της ακολουθίας των 3Δ συντεταγμένων του χεριού, με αποτέλεσμα κάθε χειρονομία (gesture) να αποτελείται από μία ακολουθία 3Δ συντεταγμένων, ανάλογα με την διάρκεια της, δηλαδή  $G = \{(x_1, y_1, z_1), \dots, (x_n, y_n, z_n)\}$ , με το  $n$  να φανερώνει τον αριθμό των πλαισίων του βίντεο από το οποίο εξάχθηκαν οι συντεταγμένες. Ο δεύτερος τρόπος αναπαράστασης ήταν με την βοήθεια των δεδομένων που παρείχε ο αισθητήρας του αζελερόμετρου,

$$G = \{(a_{11}, a_{21}, a_{31}, v_{11}, v_{21}, v_{31}), \dots, (a_{1n}, a_{2n}, a_{3n}, v_{1n}, v_{2n}, v_{3n})\}$$

όπου τα  $a_i = (a_{1i}, a_{2i}, a_{3i})$ ,  $v_i = (v_{1i}, v_{2i}, v_{3i})$  συμβολίζουν τη 3Δ γραμμική επιτάχυνση και την 3Δ γωνιακή ταχύτητα αντίστοιχα στο πλαίσιο  $i$ .

Έχοντας προσδιορίσει μεμονωμένα τους τρόπους αναπαράστασης των χειρονομιών, είτε ως ακολουθία συντεταγμένων είτε ως ακολουθία επιταχύνσεων και ταχυτήτων, προχωρούμε στα πιο τεχνικά θέματα της αναπαράστασης. Συγκεκριμένα πρέπει

να σημειώσουμε ως προς τις 3Δ συντεταγμένες των χεριών ότι για να δημιουργήσουμε ένα μοντέλο ανεξάρτητο από τη σχετική θέση του ατόμου μέσα στην εικόνα (η οποία προφανώς και θα επηρεάζει την τιμή των συντεταγμένων του χεριού), χρησιμοποιήσαμε τις συντεταγμένες των χεριών με σημείο αναφοράς το κεφάλι του χρήστη. Επίσης για να καταφέρουμε να κάνουμε τις χειρονομίες ανεξάρτητες από τα γεωμετρικά χαρακτηριστικά των εκτελεστών (μήκος χεριού, εύρος κίνησης) κανονικοποιήσαμε τα δείγματα μας. Ο τρόπος κανονικοποίησης ήταν η αφαίρεση της μέσης τιμής, ώστε να έχουμε ένα σύνολο δεδομένων μη-δενικής μέσης τιμής και επιπροσθέτως μοναδιαίας διακύμανσης. Στη συνέχεια θα γίνει αναλυτική περιγραφή του συστήματος. Η μέθοδος που ακολουθείται μπορεί να εφαρμοστεί και στους δύο τρόπους αναπαράστασης που χρησιμοποιήσαμε.

### 4.2.3 Μοντέλο Θέσης

Το πρώτο στάδιο του συστήματος μας, έχει ως σκοπό την μετατροπή των ακολουθιών συντεταγμένων ή επιταχύνσεων και γωνιακών ταχυτήτων σε ακολουθίες από σύμβολα. Για το λόγο αυτό χρησιμοποιήθηκε η μέθοδος των αυτοοργανούμενων χαρτών. Οι χάρτες αυτοί όπως έχει ήδη αναφερθεί στο κομμάτι ανάλυσης της θεωρίας τους έχουν την δυνατότητα αναπαράστασης του χώρου εισόδου (εδώ τις ακολουθίες κάθε χειρονομίας), από ένα σύνολο διατεταγμένων νευρώνων, που χαρακτηρίζονται από τα συναπτικά τους βάρη. Στην περίπτωση των ακολουθιών συντεταγμένων πήραμε όλες τις ακολουθίες  $(x, y, z)$  όλων των χειρονομιών, όλων των κλάσεων και τις τροφοδοτήσαμε στον αλγόριθμο μάθησης του χάρτη. Το αποτέλεσμα ήταν να πάρουμε μία πλήρη διακριτή αναπαράσταση του 3Δ χώρου εισόδου (3Δ συντεταγμένες όλων των χειρονομιών στο χώρο) σε μορφή νευρώνων στο πλέγμα. Ο αλγόριθμος στο τέλος αποφασίζει τα κατάλληλα βάρη των νευρώνων όπου θα μοντελοποιούν το χώρο εισόδου με τον βέλτιστο τρόπο. Στην περίπτωση των γραμμικών επιταχύνσεων και των γωνιακών ταχυτήτων χρησιμοποιήσαμε δύο χάρτες έναν για την αναπαράσταση κάθε ομάδας, αλλά και έναν χάρτη για την συνολική αναπαράσταση των δεδομένων. Τελικά μετά από συγκρίσεις καταλήξαμε ότι η χρησιμοποίηση ενός χάρτη για τα 6Δ διανύσματα επιταχύνσεων και ταχυτήτων δεν ήταν αποτελεσματική λόγω της έλλειψης μεγάλου αριθμού δεδομένων εκπαίδευσης, οπότε τελικά κρατήσαμε δύο ξεχωριστούς χάρτες. Όπως και στην περίπτωση με τις ακολουθίες συντεταγμένων πήραμε όλες τις ακολουθίες επιταχύνσεων  $a_1, a_2, a_3$  και όλες τις ακολουθίες γωνιακών ταχυτήτων  $v_1, v_2, v_3$  για να εκπαιδεύσουμε κάθε έναν από τους δυο χάρτες αντίστοιχα.

Εδώ να σημειώσουμε ότι δεν χρειάζεται να εκπαιδεύσουμε έναν χάρτη για κάθε ξεχωριστή κλάση, καθώς το αποτέλεσμα που θα πάρουμε εκπαιδεύοντας το χάρτη με τα συνολικά δεδομένα του συνόλου δεδομένων, θα αποτελεί μια

ολική αναπαράσταση που θα περιέχει όλες τις χειρονομίες, αλλά θα μπορεί πιθανώς να αναπαραστήσει και χειρονομίες οι οποίες δεν εμπεριέχονται στο σύνολο δεδομένων.

Αφου εκπαιδεύσουμε τους χάρτες μας για κάθε ένα από τα διαφορετικά σύνολα δεδομένων που έχουμε (αξελερόμετρα ,3Δ συντεταγμένες ), ακολουθούμε τη διαδικασία κωδικοποίησης των ακολουθιών κάθε χειρονομίας σε ακολουθίες συμβόλων, που παριστούν τους νευρώνες στο πλέγμα του χάρτη. Η διαδικασία είναι η ίδια για κάθε ένα από τα σύνολα δεδομένων οπότε θα αναλυθεί με ενιαίο τρόπο.

Έστω η χειρονομία,

$$G_i = \{(\text{δεδομένα 1ου πλαισίου}), \dots, (\text{δεδομένα νιοστού πλαισίου})\} \quad (4.3)$$

όπου τα δεδομένα είναι είτε οι 3-Δ συντεταγμένες της χειρονομίας στο χρόνο, είτε οι γωνιακές ταχύτητες, είτε οι γραμμικές επιταχύνσεις, αναλόγως το σύνολο δεδομένων που χρησιμοποιούμε. Με βάση τον εκπαιδευμένο χάρτη και του πίνακα βαρών του (δηλαδή τον πίνακα που περιέχει τα συναπτικά βάρη κάθε νευρώνα του χάρτη, στο πλέγμα), βρίσκουμε τον νευρώνα που βρίσκεται πιο κοντά σε κάθε δεδομένο πλαισίου της χειρονομίας (Best Matching Unit-BMU). Έτσι αν πούμε  $T$  τον τελεστή που επιδρά πάνω στη χειρονομία και την μετατρέπει σε ακολουθία συμβόλων νευρώνων, τότε έχουμε :

$$T(G_i) = (u_1, u_2, \dots, u_n) : u_i = BMU(\text{δεδομένα πλαισίου } i \text{ χειρονομίας}), \\ i \in [1, n] \quad (4.4)$$

με την συνάρτηση BMU να επιστρέφει τον δείκτη του πιο κοντινού νευρώνα στο δεδομένο της χειρονομίας που του δίνεται .Δηλαδή για παράδειγμα αν έχουμε χρησιμοποιήσει  $k$  νευρώνες στο χάρτη μας, η συνάρτηση θα συγκρίνει είτε τις συντεταγμένες του δεδομένου, είτε τις ταχύτητες, είτε τις επιταχύνσεις του δεδομένου, ανάλογα με το σύνολο δεδομένων που χρησιμοποιήσαμε για την εκπαίδευση του χάρτη και θα επιστρέψει έναν αριθμό από το 1 μέχρι το  $k$  που θα είναι ο δείκτης του πιο κοντινού νευρώνα.

Στη συνέχεια αφού βρούμε για κάθε χειρονομία του συνόλου δεδομένων μας την αναπαράσταση της ως ακολουθία συμβόλων νευρώνων, αφαιρούμε τα



συνεχόμενα ίδια σύμβολα ώστε σε κάθε χρονικό βήμα να έχουμε διαφορετικό σύμβολο. Δηλαδή :

$$\tilde{G} = N(T(G)) = \{\tilde{u}_1, \dots, \tilde{u}_m\} : m \leq n, \quad \tilde{u}_t \neq \tilde{u}_{t-1} \forall t \in [2, m] \quad (4.5)$$

όπου  $N$  η συνάρτηση που αφαιρεί τα συνεχόμενα ίδια σύμβολα  $u_i$  από τις ακολουθίες κάθε χειρονομίας. Ο λόγος που το κάνουμε αυτό είναι για να μειώσουμε τη διάσταση του προβλήματος. Συνήθως κατά τη διάρκεια εκτέλεσης της χειρονομίας από πλαίσιο σε πλαίσιο θα υπάρχει μικρή απόκλιση των τιμών των χαρακτηριστικών η οποία δεν είναι σημαντική για τον διαχωρισμό των χειρονομιών μεταξύ τους, αυτή η μικρή διαφορά θα αποτυπωθεί στο πρόβλημα ως ταίριασμα των δεδομένων στον ίδιο νευρώνα, όποτε την αποβάλλουμε αφαιρώντας την πλεονάζουσα πληροφορία.

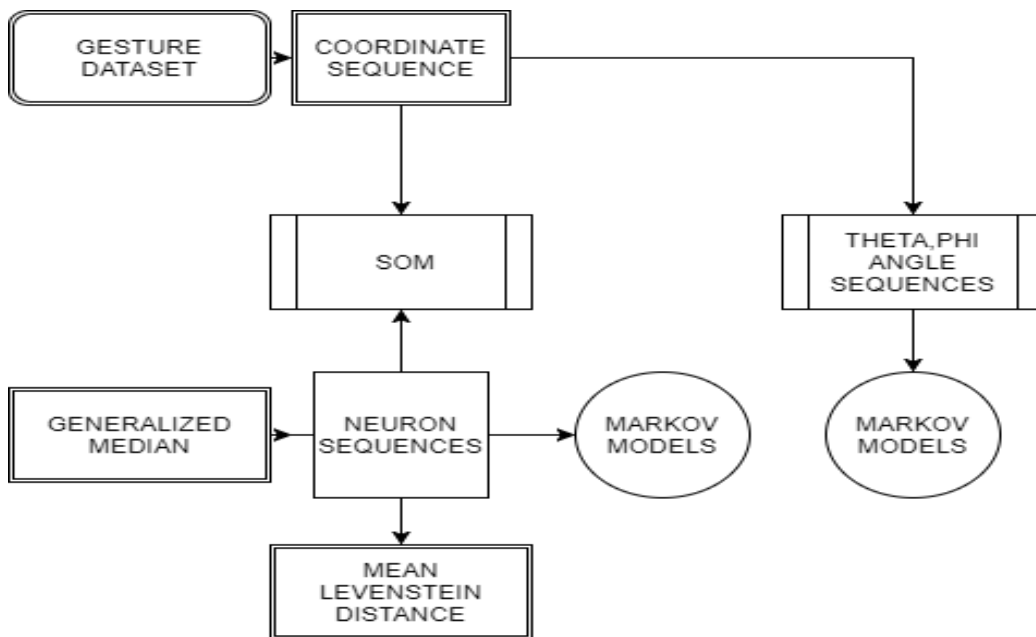
Έχοντας πλέον μετατρέψει την κάθε χειρονομία του συνόλου δεδομένων σε ακολουθία νευρώνων, συγκεντρώνουμε κάθε κλάση χειρονομιών ξεχωριστά και χρησιμοποιούμε τα δεδομένα για να εκπαιδεύσουμε για κάθε μία ένα Μαρκοβιανό μοντέλο. Αυτό μπορεί να είναι είτε μία απλή αλυσίδα Μαρκοβ,ε είτε ένα Κρυφό Μαρκοβιανό μοντέλο. Στην περίπτωση της αλυσίδας Μαρκοβ απλώς εκπαιδεύουμε έναν πίνακα διαστάσεων  $NN \times NN$ , με το  $NN$  να αντιπροσωπεύει τον αριθμό των νευρώνων που χρησιμοποιήσαμε στο χάρτη, για κάθε κλάση. Κάθε νευρώνας αναπαριστά μία κατάσταση στη μαρκοβιανή αλυσίδα και ο πίνακας έστω  $Mm_i$  περιέχει τις πιθανότητες μετάβασης από την μία κατάσταση στην άλλη. Στην περίπτωση εκπαίδευσης κρυφών μαρκοβιανών μοντέλων οι ακολουθίες των νευρώνων αναπαριστούν τα σύμβολα που παρατηρούμε στο μοντέλο κάθε χρονική στιγμή και οι καταστάσεις θα μπορούσαμε να πούμε ότι αναπαριστούν τη διάρκεια των χειρονομιών. Έτσι καταλήγουμε με  $c$  μοντέλα Μαρκόβ ένα για κάθε κλάση:

$$\begin{aligned} MM^{som} &= \{MM_1^{som}, \dots, MM_c^{som}\} \\ \tilde{D}_j &= \{\tilde{G}_1, \dots, \tilde{G}_{n_j}\} \rightarrow MM_j^{som} \end{aligned} \quad (4.6)$$

όπου το  $MM^{som}$  συμβολίζει το σύνολο των μοντέλων Μαρκόβ που φτιάξαμε και το  $D_j$  το σύνολο των  $n_j$  χειρονομιών που ανήκουν στην κλάση  $j$

#### 4.2.4 Μοντέλο κατεύθυνσης

Το μοντέλο αυτό χρησιμοποιήθηκε μόνο στην περίπτωση του συνόλου δεδομένων που αποτελούταν από τις 3Δ συντεταγμένες, καθώς δεν έχει κάποια



Σχήμα 4.2: Μία αναπαράσταση της διαδικασίας μάθησης βασισμένη στις συντεταγμένες του χεριού

φυσική σημασία για το άλλο σύνολο. Το μοντέλο αυτό βασίζεται στην οπτική ροή και έχει σκοπό να αναδείξει τις διαφορές των κατευθύνσεων των τροχιών των χειρονομιών, μέσω των γωνιών τους από πλαίσιο σε πλαίσιο. Συγκεκριμένα για κάθε χειρονομία, όπως έχουμε αναφέρει ήδη, έχουμε μία ακολουθία 3Δ συντεταγμένων της τροχιάς των χεριών του εκτελεστή σε σχέση με το κεφάλι. Από πλαίσιο σε πλαίσιο εξάγουμε δύο διαφορετικές γωνίες  $(\phi, \theta)$ , όπου  $\phi$  είναι η γωνία που σχηματίζει η ακολουθία συντεταγμένων δύο διαδοχικών πλαισίων στο επίπεδο  $x - y$ :

$$\phi = \arctan \left( \frac{|y_i - y_{i-1}|}{|x_i - x_{i-1}|} \right) \quad (4.7)$$

και  $\theta$  η γωνία που σχηματίζουν δύο συνεχόμενα πλαίσια με τον άξονα  $z$

$$\theta = \arctan \left( \frac{\sqrt{(y_i - y_{i-1})^2 + (x_i - x_{i-1})^2}}{|z_i - z_{i-1}|} \right). \quad (4.8)$$

Συνεπώς έχουμε άλλες δύο αναπαραστάσεις του συνόλου δεδομένων μας ως ακολουθίες γωνιών του  $(\theta, \phi)$ . Οπότε για κάθε χειρονομία εξάγουμε μία τέτοια αναπαράσταση. Στη συνέχεια για να περιορίσουμε τον δυνατό αριθμό των

διαφορετικών γωνιών, τις χβαντίζουμε σε 8 διαφορετικές δυνατές τιμές .Στη συνέχεια ομαλοποιούμε τις ακολουθίες έτσι ώστε να περιορίσουμε τυχόν αστάθειες στην κίνηση του χεριού και τέλος αφαιρούμε τις διαδοχικές εμφανίσεις ίδιων γωνιών.Άρα παίρνουμε δύο ακόμα μοντέλα αναπαράστασης του συνόλου δεδομένων που βασίζονται στις γωνίες :

$$OF_\phi(G) = \{v_1, \dots, v_m\} : v_i = W_T(Q(\phi)) \quad (4.9)$$

$$OF_\theta(G) = \{l_1, \dots, l_m\} : l_i = W_T(Q(\theta)) \quad (4.10)$$

όπου το  $Q$  είναι η συνάρτηση χβάντισης των γωνιών και  $W_T$  το median φίλτρο που χρησιμοποιούμε για την ομαλοποίηση.Στη συνέχεια εφαρμόζουμε και τον συντελεστή αφαίρεσης συνεχόμενων ίδιων συμβόλων  $N$  και παίρνουμε την αναπαράσταση κάθε χειρονομίας ως:

$$G_{i_\phi} = N(OF_\phi(G)) = \{v_1, \dots, v_m\} \quad (4.11)$$

$$G_{i_\theta} = N(OF_\theta(G)) = \{l_1, \dots, l_m\} \quad (4.12)$$

Έχοντας στην διάθεση μας αυτά τα μοντέλα δημιουργούμε αντίστοιχα όπως και προηγουμένως  $c$  διαφορετικά Μαρκοβιανά μοντέλα, ένα για κάθε κλάση, ένα για κάθε μοντέλο αναπαράστασης γωνιών:

$$MM^{of_\phi} = \{MM_1^{of_\phi}, \dots, MM_c^{of_\phi}\} : D_{\phi_i} = \{G_{1_{\phi_i}}, \dots, G_{n_{\phi_i}}\} \rightarrow MM_i^{of_\phi} \quad (4.13)$$

$$MM^{of_\theta} = \{MM_1^{of_\theta}, \dots, MM_c^{of_\theta}\} : D_{\theta_i} = \{G_{1_{\theta_i}}, \dots, G_{n_{\theta_i}}\} \rightarrow MM_i^{of_\theta} \quad (4.14)$$

#### 4.2.5 Levenstein

Ένα τελευταίο μοντέλο που δημιουργείται ανά κλάση χειρονομιών είναι η Γενικευμένη μέση τιμή (Generalized Median) του συνόλου  $D_j$ .Αυτή ορίζεται ως η ακολουθία σε ένα σύνολο ακολουθιών  $S$  που το άθροισμα των αποστάσεων όλων των ακολουθιών από αυτή είναι το μικρότερο δυνατό, σύμφωνα με την απόσταση Levenstein.Αν η ακολουθία με τη συγκεκριμένη ιδιότητα βρίσκεται μέσα στο σύνολο των ακολουθιών τότε ονομάζεται Γενικευμένη μέση τιμή του συνόλου (*Generalized Set Median*).Αν  $M_j$  συμβολίζει τη Γενικευμένη μέση

τιμή του  $\tilde{D}_j$  και  $L$  την απόσταση Levenstein τότε

$$M_j = \operatorname{argmin}_g \sum_{\tilde{G} \in \tilde{D}_j} L(g, \tilde{G}). \quad (4.15)$$

Η μέση απόσταση Levenstein μεταξύ των μελών του  $\tilde{D}_j$  και του  $M_j$ , υπολογίζεται επίσης και συμβολίζεται με  $ML_j$ . Αυτή μετράει τις αποκλίσεις μέσα στο σύνολο δεδομένων και χρησιμοποιείται στο επομένο στάδιο της ταξινόμησης των χειρονομιών:

$$ML_j = \frac{\sum_{i=1}^{n_j} L(\tilde{G}_i, M_j)}{n_j} \quad (4.16)$$

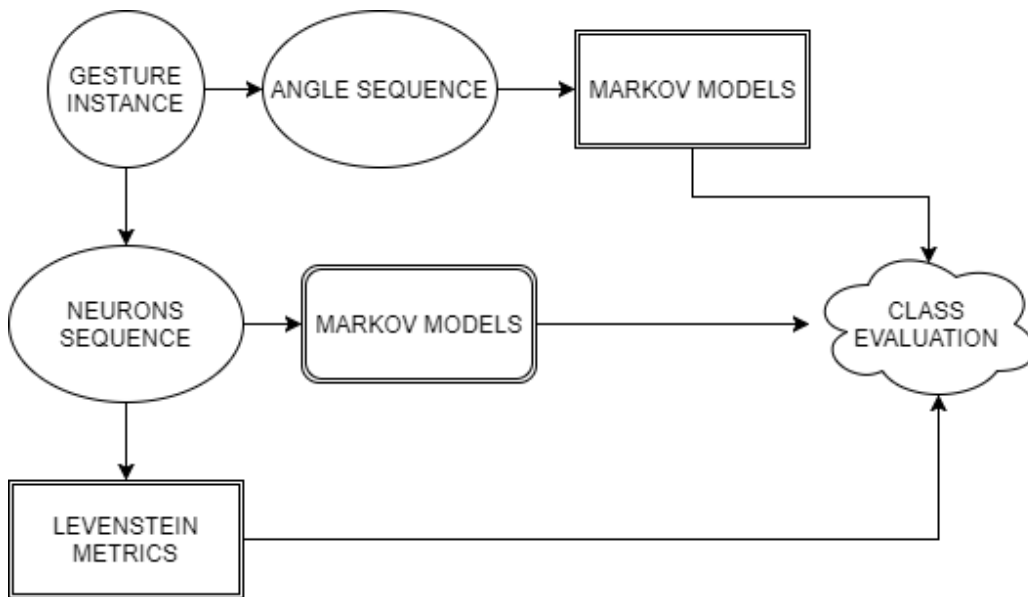
#### 4.2.6 Ταξινόμηση

Για την ταξινόμηση χρησιμοποιήθηκαν τρεις διαφορετικές μέθοδοι. Και οι τρεις βασίζονται στα Μαρκοβιανά Μοντέλα. Ο πρώτος αφορά απλές Μαρκοβιανές αλυσίδες, δηλαδή εκπαιδευμένα μοντέλα ένα για κάθε κλάση, όπου κάθε δεδομένο προς αναγνώριση αξιολογείται σε καθένα από αυτά και αυτό που παράγει τη μεγαλύτερη πιθανότητα, θεωρείται και το μοντέλο της κλάσης στην οποία θα ταξινομηθεί. Ο δεύτερος τρόπος βασίζεται στα Κρυφά Μαρκοβιανά Μοντέλα και ακολουθεί επίσης την ίδια τακτική, δηλαδή βρίσκουμε τις πιθανότητες που παράγει κάθε ένα από τα μοντέλα μας και στη συνέχεια αυτό με την μεγαλύτερη τιμή, είναι και αυτό στο οποίο θα ταξινομηθεί το δεδομένο για αναγνώριση. Η τρίτη μέθοδος βασίζεται πάλι στις Μαρκοβιανές Αλυσίδες, λαμβάνοντας όμως υπ' όψιν εκτός από τις τιμές των μοντέλων και την απόσταση των νευρώνων του χάρτη στο πλέγμα.

Συνοδευτικά με αυτά τα μοντέλα όπως έχει ήδη σημειωθεί χρησιμοποιούνται και το μοντέλο Levenstein καθώς και στην περίπτωση του συνόλου δεδομένων με τις συντεταγμένες και το μοντέλο κατεύθυνσης, για την ενίσχυση της σωστής ταξινόμησης.

#### **Αλυσίδες Μαρκοφ :**

- Δεδομένα Συντεταγμένων: Έστω  $G$  η ακολουθία χειρονομίας που θέλουμε να αναγνωρίσουμε, όπως ήδη έχουμε αναφέρει, την μετατρέπουμε σε τρεις ακολουθίες δεδομένων, μία από σύμβολα νευρώνων, μία από γωνίες του



Σχήμα 4.3: Η διαδικασία αξιολόγησης του συστήματος μας με δεδομένα τις συντεταγμένες του χεριού

x-y επιπέδου,  $\varphi$ , και μία από ακολουθίες γωνιών με τον άξονα z,  $\theta$ , άρα αρχικά βρίσκουμε τις πιθανότητες

$$P_{1i} = \Pr(G|MM_i^{som}) \quad \forall i \in [1, \dots, c] \quad (4.17)$$

$$P_{2i} = \Pr(G|MM_i^{\phi}) \quad \forall i \in [1, \dots, c] \quad (4.18)$$

$$P_{3i} = \Pr(G|MM_i^{\theta}) \quad \forall i \in [1, \dots, c] \quad (4.19)$$

$$P_{all_i} = P_{1i} \cdot P_{2i} \cdot P_{3i} \quad \forall i \in [1, \dots, c] \quad (4.20)$$

όπου είναι  $P_{1i}, P_{2i}, P_{3i}$  οι πιθανότητες η χειρονομία να ανήκει σε κάθε μία από τις  $c$  κλάσεις σύμφωνα με τα τρία μοντέλα, με το  $P_{all_i}$  να είναι ο συνδυασμός αυτών. Στη συνέχεια βρίσκουμε την:

$$\Pr(G|M_j) = \frac{ML_j}{L(G, M_j)} \quad (4.21)$$

η οποία δεν είναι πιθανότητα αλλά είναι ένα μέτρο ομοιότητας της ακολουθίας σε σχέση με τη μέση απόσταση Levenstein και την Γενικευμένη μέση τιμή του συνόλου, μετρικές που είχαμε εξάγει από τα δεδομένα εκπαίδευσης. Τελικά η κλάση  $i$  στην οποία ταξινομείται η εν λόγω χειρονομία

αποκτάται από

$$\text{class} = \underset{i}{\operatorname{argmax}}(P_{all_i} \cdot \Pr(G|M_j)), \quad i \in [1, \dots, c] \quad (4.22)$$

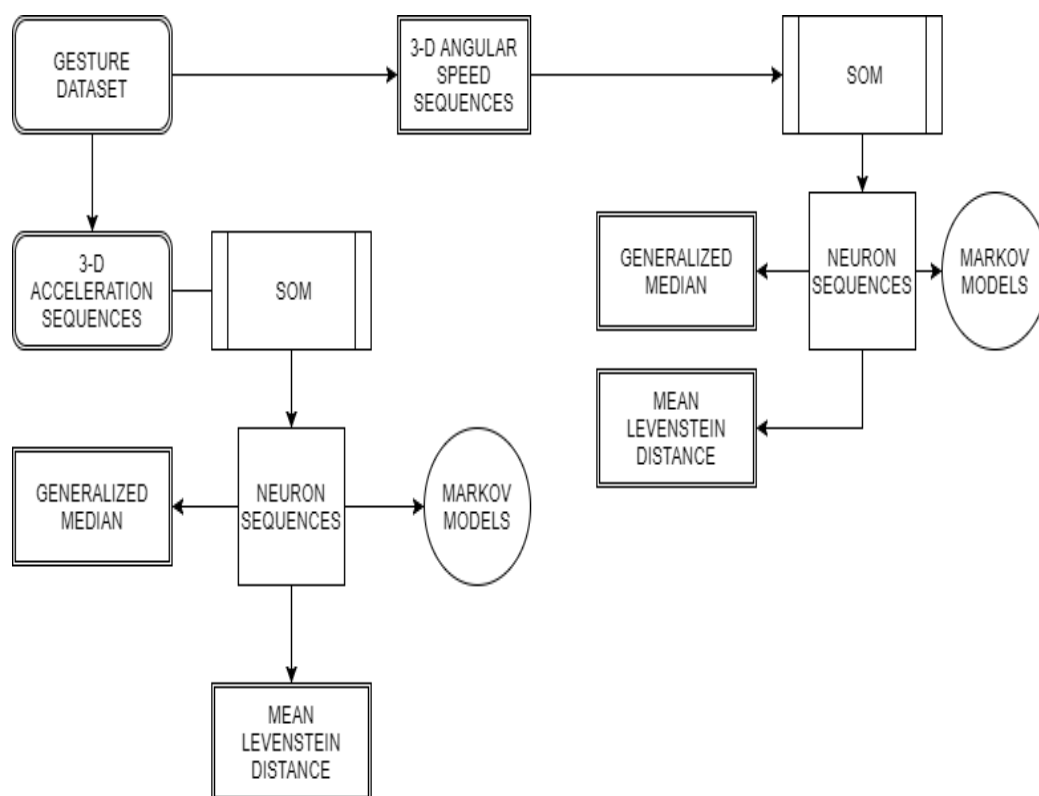
- Δεδομένα Αξελερόμετρων: Σε αυτήν την περίπτωση ακολουθούμε ακριβώς την ίδια τακτική όπως και προηγουμένως με τη διαφορά ότι πλέον δεν έχουμε τα μοντέλα γωνιών. Όπως έχουμε αναφέρει και προηγουμένως εκπαιδεύσαμε διαφορετικά μοντέλα για τα δεδομένα με τις γωνιακές ταχύτητες και για τα δεδομένα των γραμμικών επιταχύνσεων. Στην συνέχεια ενώσαμε τις πιθανότητες των δύο και βρήκαμε μία συνολική πιθανότητα η οποία χρησιμοποιήθηκε για την ταξινόμηση. Συγκεκριμένα έστω  $P_{acc_i}$  η συνολική πιθανότητα της χειρονομίας  $G$  να ανήκει στην κλάση  $i$  σύμφωνα με το μοντέλο των επιταχύνσεων και  $P_{sp_i}$  συνολική πιθανότητα της χειρονομίας  $G$  να ανήκει στην κλάση  $i$  σύμφωνα με το μοντέλο των ταχυτήτων τότε :

$$\text{class} = \underset{i}{\operatorname{argmax}}(P_{sp_i} \cdot P_{acc_i}), \quad i \in [1, \dots, c] \quad (4.23)$$

### **Κρυφά Μαρκοβιανά Μοντέλα:**

Η τακτική που ακολουθείται εδώ είναι ακριβώς η ίδια με αυτήν των μαρκοβιανών αλυσίδων. Το μόνο που αλλάζει είναι ο τρόπος εκπαίδευσης των μοντέλων που γίνεται με το τον αλγόριθμο Baum-Welch και ο τρόπος αξιολόγησης των δεδομένων αναγνώρισης, όπου χρησιμοποιείται ο αλγόριθμος Προς τα Εμπρός για την αποτίμηση της πιθανότητας του μοντέλου. Να σημειωθεί ότι θα μπορούσε να χρησιμοποιηθεί και ο Viterbi, αλλά προτιμήσαμε τον Προς τα Εμπρός καθώς μας δίνει την ολική πιθανότητα μία ακολουθία να έχει παραχθεί από το μοντέλο και όχι την μέγιστη .

Όπως και προηγουμένως μαζί με τις πιθανότητες Μαρκόβ από τα διάφορα μοντέλα χρησιμοποιήθηκε και η μετρική που βασίζεται στην απόσταση Levenshtein. Εδώ να σημειώσουμε ότι η μετρική αυτή εκτός από τις άλλες λειτουργίες που επιτελεί, οι οποίες έχουν αναφερθεί προηγουμένως, βοηθάει να υπάρχει μία ισάξια μεταχείριση των χειρονομιών. Δηλαδή καθώς οι ακολουθίες έχουν διαφορετικό μέγεθος είναι λογικό αυτές με το μικρότερο να ευνοούνται από τα ακολουθιακά μοντέλα που έχουμε φτιάξει, καθώς θα παράγουν μεγαλύτερες πιθανότητες. Έτσι η μετρική αυτή βοηθάει δίνοντας κάποια ποινή σε αυτές η οποία



Σχήμα 4.4: Παρουσίαση του συστήματος μάθησης βασισμένο στα δεδομένα εισόδου από τα αξελερόμετρα

αντισταθμίζει με κάποιον τρόπο το πλεονέκτημα τους έναντι των μεγαλύτερων χειρονομιών .

### **Αλυσίδες Μαρκοφ με Απόσταση Νευρώνων στο πλέγμα:**

Έστω  $G_k$  μία χειρονομία και  $\tilde{G}_k, G_{k_\phi}, G_{k_\theta}$ , οι μετασχηματισμένες ακολουθίες σε σύμβολα σύμφωνα με τα μοντέλα θέσης και γωνιών (αν πρόκειται για τα δεδομένα συντεταγμένων, αλλιώς μόνο τα μοντέλα θέσης για τα δεδομένα από τα αξελερόμετρα). Η πιθανότητα αυτή η χειρονομία να ανήκει στην κλάση  $j$  είναι:

$$P(\tilde{G}_k | MM_j^{som}) = \frac{\prod_{i=1}^q S_i^{som}}{q} : q = |\tilde{G}_k| \quad (4.24)$$

Το  $S_i$  συμβολίζει την πιθανότητα του συμβόλου  $u_i$  της ακολουθίας η οποία έχει μήκος  $q$ . Οι παραπάνω τιμές υπολογίζονται ως :

$$S_1^{som} = \max_z (NF_{u_1}^{som}(z) \pi_j^{som}) \quad (4.25)$$

$$S_i^{som} = \max_z (NF_{u_i}^{som}(z) P(z | u_{i-1}, MM_j^{som})) \quad (4.26)$$

Δηλαδή για την πρώτη κατάσταση το σύστημα βρίσκει τον κόμβο που έχει την μεγαλύτερη από κοινού πιθανότητα :

- να είναι κοντά στο  $u_1$  που είναι  $NF_{u_1}^{som}(z)$
- να είναι η πρώτη κατάσταση στο  $MM_j^{som}$  που είναι  $\pi_j^{som}$

Για τους κόμβους από  $[2, q]$ , εφαρμόζεται μία παρόμοια αναζήτηση με την δεύτερη πιθανότητα να είναι η πιθανότητα μετάβασης  $p(z | u_{i-1}, MM_j^{som})$ . Η  $NF_{u_i}^{som}(z)$  είναι η απόσταση του νευρώνα  $z$  από τον κόμβο  $u_i$  έτσι όπως ορίζεται από την Γκαουσιανή συνάρτηση γειτονιάς του Αυτο Οργανούμενου Χάρτη με το  $u_i$  ως κέντρο. Το μέγιστο  $z$  που βρίσκεται σε κάθε βήμα χρησιμοποιείται ως αρχή για την εύρεση της επόμενης κατάστασης, δηλαδή :

$$u_i = \operatorname{argmax}_z (S_i^{som}) : i \in [1, q] \quad (4.27)$$

Μία παρόμοια τεχνική εφαρμόζεται και για το μοντέλο των γωνιών, με τη διαφορά καθώς δεν έχουμε συνάρτηση γειτονιάς βάζουμε τιμή  $1/2$  στους άμεσους γείτονες κάποιας γωνίας και τιμή  $1/4$  στους αμέσως επόμενους (δύο γωνίες θεωρούνται γειτονικές αν απέχουν τη μικρότερη απόσταση αριθμητικά με βάση τον κβαντισμό ίσης απόστασης που έχουμε κάνει).



Τέλος συνδιάζουμε όλα τα μοντέλα και φτάνουμε στην τελική αποτίμησή τους, ταξινομώντας την χειρονομία στην κλάση με την μεγαλύτερη πιθανότητα.

## Κεφάλαιο 5

# Αριθμητικά Αποτελέσματα

### 5.1 Εισαγωγή

Όπως αναφέραμε και προηγουμένως χρησιμοποιήσαμε τρεις τεχνικές αξιολόγησης των δεδομένων μας, οι δύο έχουν σαν βάση τις αλυσίδες Μαρκόβ και η τρίτη τα Κρυφά Μαρκοβιανά Μοντέλα. Πραγματοποιήσαμε αρκετούς συνδυασμούς παραμέτρων για την εύρεση των μοντέλων που μας δίνουν τα υψηλότερα ποσοστά αναγνώρισης. Κάποιες από αυτές ήταν: (1) ο αριθμός νευρώνων στο πλέγμα του Αυτό Οργανούμενου χάρτη, (2) ο αριθμός καταστάσεων στα Κρυφα Μαρκοβιανά μοντέλα, (3) ο τρόπος κανονικοποίησης των δεδομένων και (4) η ομαλοποίηση ή όχι των γωνιών. Από αυτές τις δοκιμές σε αυτή την ενότητα παρουσιάζουμε αυτές με το μεγαλύτερο ενδιαφέρον.

Εκτός από τις παραπάνω παραμέτρους χρησιμοποιήσαμε και συνδυασμούς των πιθανοτήτων των μοντέλων που συνδυάσαμε για να δούμε, ποια από αυτά έχουν τη μεγαλύτερη βαρύτητα στο κομμάτι της αποτελεσματικής αναγνώρισης και ποια σε ορισμένες περιπτώσεις μπορεί η χρησιμοποίησή τους να μειώνει την απόδοση του ολικού συστήματος. Στον παρακάτω πίνακα φαίνονται αυτές οι μετρικές ανάλογα τον τύπο δεδομένων και το σύστημα που χρησιμοποιείται σε καθένα από αυτά.

| Σύστημα με Δεδομένα Συντεταγμένων |   |
|-----------------------------------|---|
| $P_{pos}$ :                       | Πιθανότητα αποτίμησης του Μοντέλου Θέσης                        |
| $P_{\phi}$ :                      | Πιθανότητα αποτίμησης του Μοντέλου γωνίας στο x-y επίπεδο       |
| $P_{\theta}$ :                    | Πιθανότητα αποτίμησης του Μοντέλου γωνίας με τον άξονα z        |
| $P_{leven}$ :                     | Μετρική απόστασης Levenstein                                    |
| Σύστημα με Δεδομένα Αξελερόμετρου |   |
| $P_{acc}$ :                       | Πιθανότητα αποτίμησης του Μοντέλου με 3Δ γραμμικές επιταχύνσεις |
| $P_{vel}$ :                       | Πιθανότητα αποτίμησης του Μοντέλου με 3Δ γωνιακές ταχύτητες     |
| $L_{acc}$ :                       | Μετρική απόστασης Levenstein με επιταχύνσεις                    |
| $L_{vel}$ :                       | Μετρική απόστασης Levenstein με δεδομένα ταχύτητας              |

Πίνακας 5.1: Πίνακας Παρουσίασης Πιθανοτήτων

Από τις μετρικές που παρουσιάζονται παραπάνω, πήραμε διάφορους συνδυασμούς, όπως προαναφέρθηκε ώστε να εξετάσουμε τη λειτουργία των μοντέλων χωριστά και συνάμα να αξιολογήσουμε τη συμβολή τους αλλά και τις περιπτώσεις όπου οι χρήση ορισμένων οδηγούν σε χειρότερα αποτελέσματα. Παρακάτω παρουσιάζεται ο πίνακας με τους συνδυασμούς αυτούς οι τιμές του οποίου θα χρησιμοποιηθούν παρακάτω στην παρουσίαση των αποτελεσμάτων.

| Πίνακας Συνδυασμών     |   |                        |   |
|------------------------|---|------------------------|---|
| Δεδομένα Συντεταγμένων |   | Δεδομένα Αξελερόμετρου |   |
| Συνδυασμός             | Τιμή  | Συνδυασμός             | Τιμή  |
| 1                      | $P_{pos} \cdot P_{\phi} \cdot P_{\theta} \cdot P_{leven}$ | 1                      | $P_{acc} \cdot L_{acc} \cdot P_{vel} \cdot L_{vel}$ |
| 2                      | $P_{pos} \cdot P_{\phi}$                                  | 2                      | $P_{acc} \cdot L_{acc}$                             |
| 3                      | $P_{pos} \cdot P_{\theta} \cdot P_{leven}$                | 3                      | $P_{vel} \cdot L_{vel}$                             |
| 4                      | $P_{\phi} \cdot P_{\theta} \cdot P_{leven}$               | 4                      | $P_{acc} \cdot P_{vel}$                             |
| 5                      | $P_{pos} \cdot P_{\phi} \cdot P_{leven}$                  | -                      | -   |

Πίνακας 5.2: Πίνακας Συνδυασμών Πιθανοτήτων

Να σημειώσουμε επίσης ότι ως προς τον αριθμό των νευρώνων στο πλέγμα του Αυτο Οργανούμενου χάρτη χρησιμοποιήθηκαν χάρτες με: 36, 64, 100, 144, 289 νευρώνες όλοι με εξαγωνική μορφή πλέγματος και Manhattan μετρική απόστασης. Από αυτά ενδιαφέρον στη συγκεκριμένη περίπτωση εμφάνισαν τα πλέγματα με αριθμούς νευρώνων 64, 100, αποτελέσματα των οποίων παρουσιάζονται στη συνέχεια.

Ως προς τον αριθμό καταστάσεων στα Κρυφά Μαρκοβιανά μοντέλα χρησιμοποιήθηκαν: 3, 5, 7, 9, 11, 13, 15, 17 καταστάσεις. Από τα συνολικά πειράματα στην κατηγορία αυτή παρουσιάζονται κάποια με το μεγαλύτερο ενδιαφέρον.

Τέλος από τους Συναδυασμούς πιθανοτήτων παρουσιάζονται και σε αυτήν την περίπτωση ορισμένες περιπτώσεις που θεωρήσαμε ότι έπρεπε να συμπεριληφθούν.

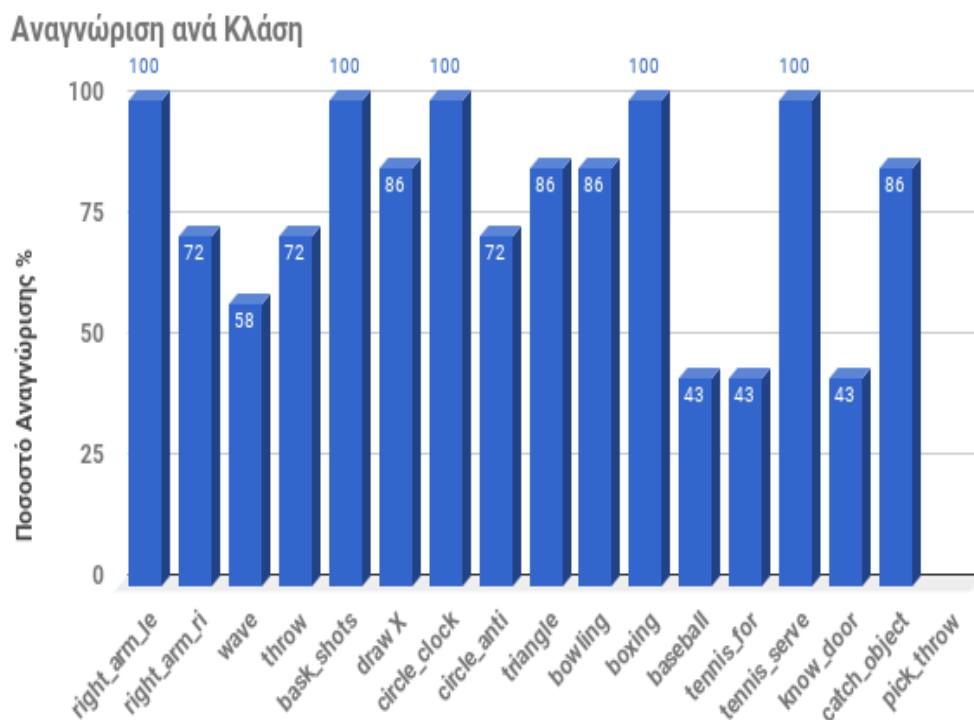
## 5.2 Γραφήματα Αναγνώρισης (Accuracy) Ανά Κλάση

### 5.2.1 Δεδομένα Συντεταγμένων

#### Αλυσίδες Μαρκόφ

1. Πείραμα (1):

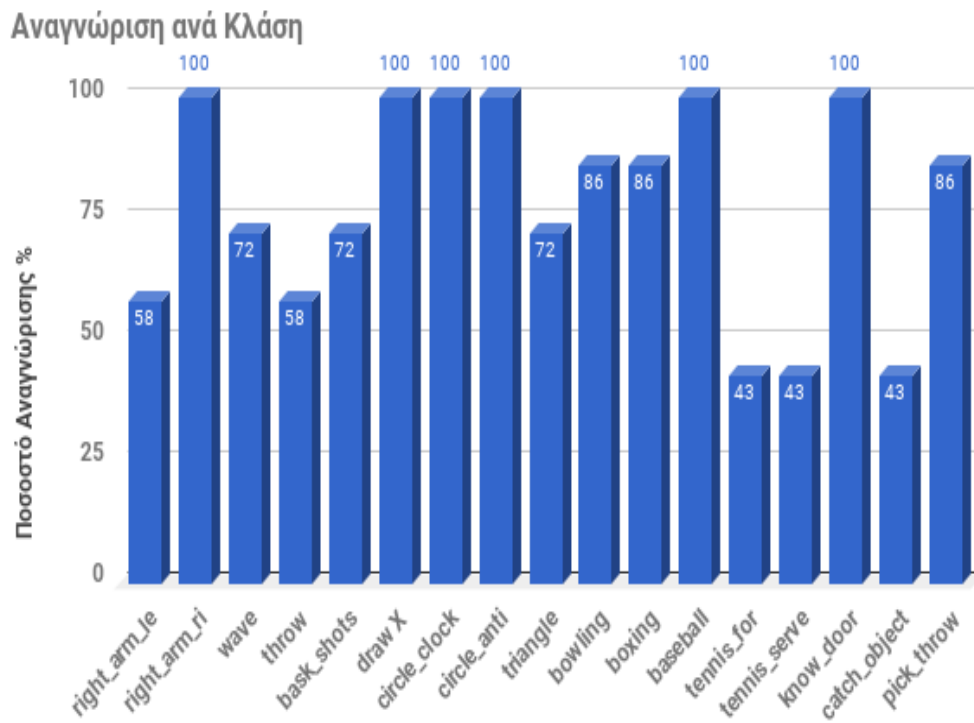
- Αριθμός Νευρώνων: 100
- Ποσοστό Αναγνώρισης: 77%
- Κανονικοποίηση: NAI
- Ομαλοποίηση: NAI
- Συνδυασμός: 1



Σχήμα 5.1: Πείραμα 1: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

2. Πείραμα (2):

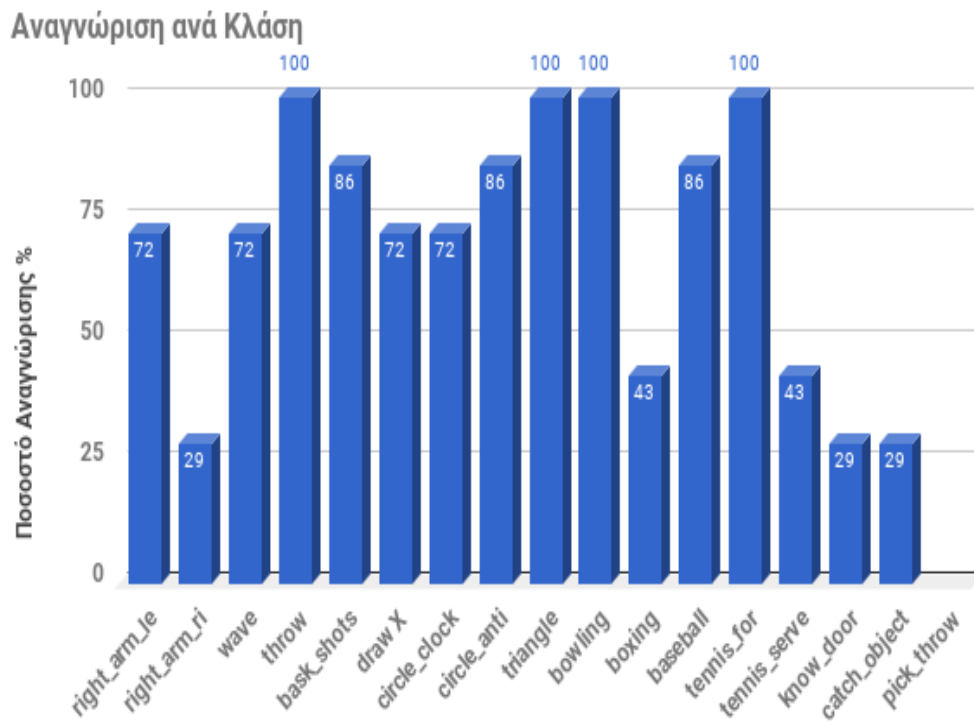
- Αριθμός Νευρώνων: 100
- Ποσοστό Αναγνώρισης: 78%
- Κανονικοποίηση: ΝΑΙ
- Ομαλοποίηση: ΝΑΙ
- Συνδυασμός: 3



Σχήμα 5.2: Πείραμα 2: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

### 3. Πείραμα (3)

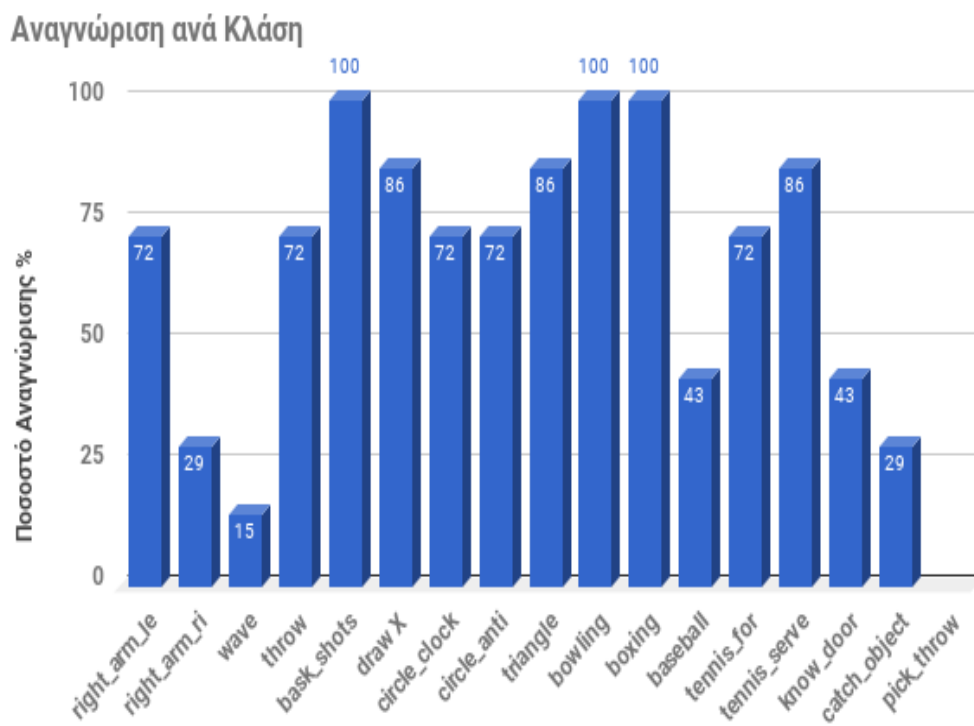
- Αριθμός Νευρώνων: 100
- Ποσοστό Αναγνώρισης: 69%
- Κανονικοποίηση: ΟΧΙ
- Ομαλοποίηση: ΟΧΙ
- Συνδυασμός: 1



Σχήμα 5.3: Πείραμα 3: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

#### 4. Πείραμα (4)

- Αριθμός Νευρώνων: 100
- Ποσοστό Αναγνώρισης: 67%
- Κανονικοποίηση: ΟΧΙ
- Ομαλοποίηση: ΟΧΙ
- Συνδυασμός: 5

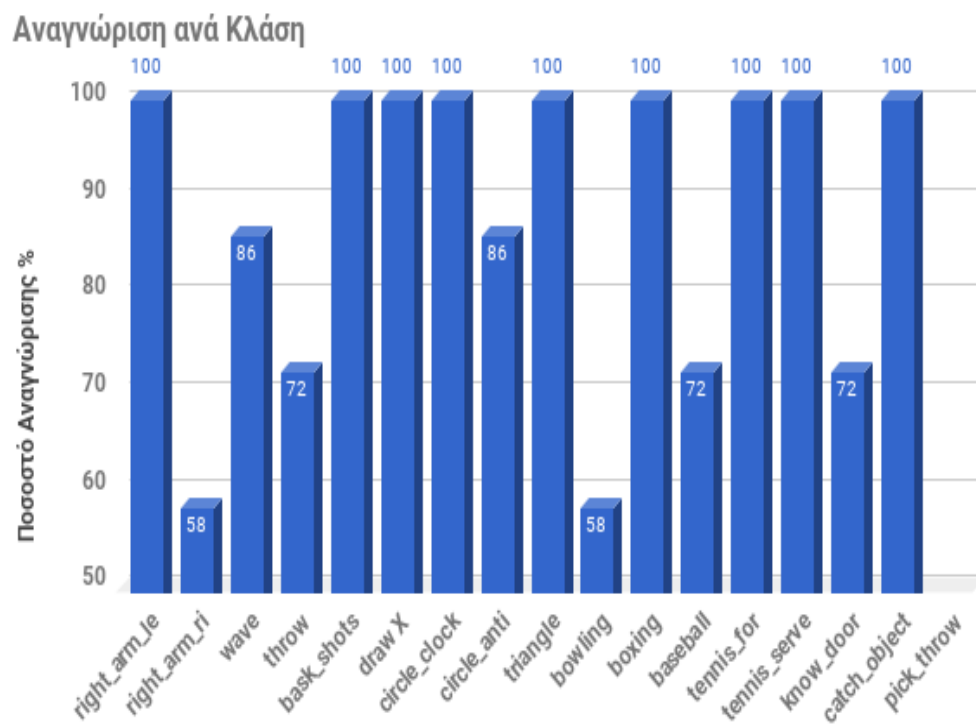


Σχήμα 5.4: Πείραμα 4: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση



## 5. Πείραμα (5)

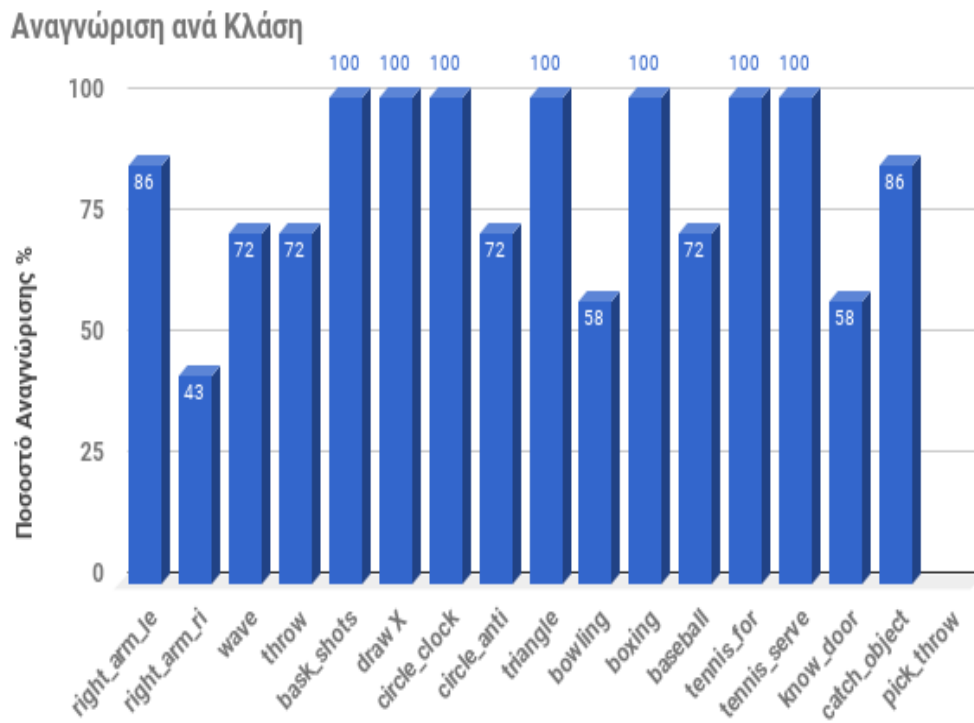
- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 89%
- Κανονικοποίηση: ΝΑΙ
- Ομαλοποίηση: ΟΧΙ
- Συνδυασμός: 1



Σχήμα 5.5: Πείραμα 5: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

## 6. Πείραμα (6)

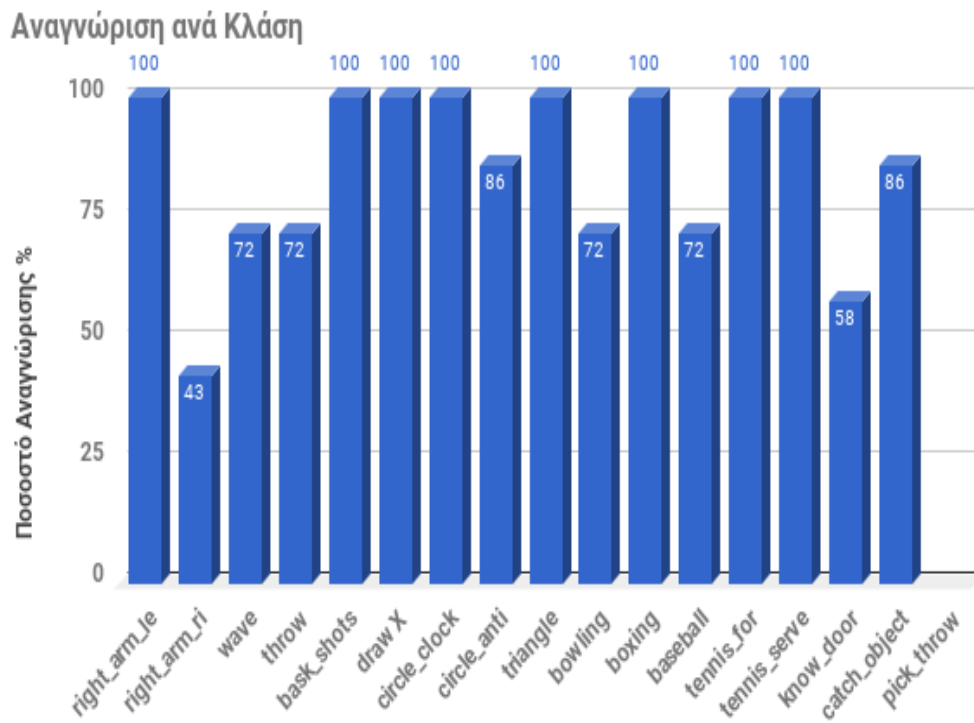
- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 84%
- Κανονικοποίηση: NAI
- Ομαλοποίηση: NAI
- Συνδυασμός: 5



Σχήμα 5.6: Πείραμα 6: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

## 7. Πείραμα (7)

- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 86%
- Κανονικοποίηση: NAI
- Ομαλοποίηση: NAI
- Συνδυασμός: 2

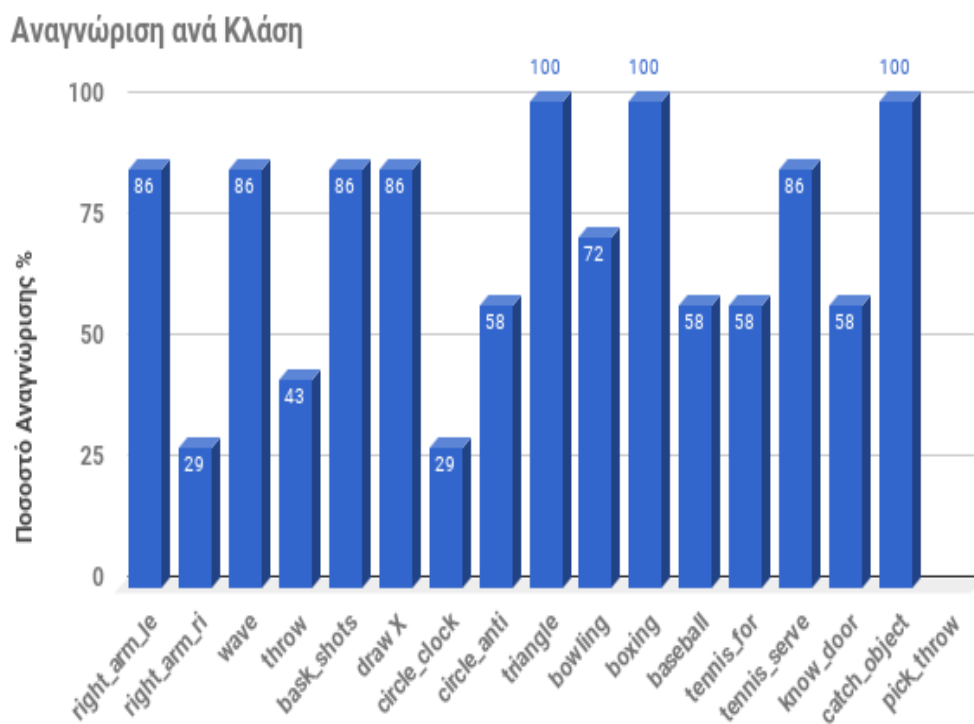


Σχήμα 5.7: Πείραμα 7: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

## Κρυφά Μαρκοβιανά Μοντέλα

### 1. Πείραμα (1)

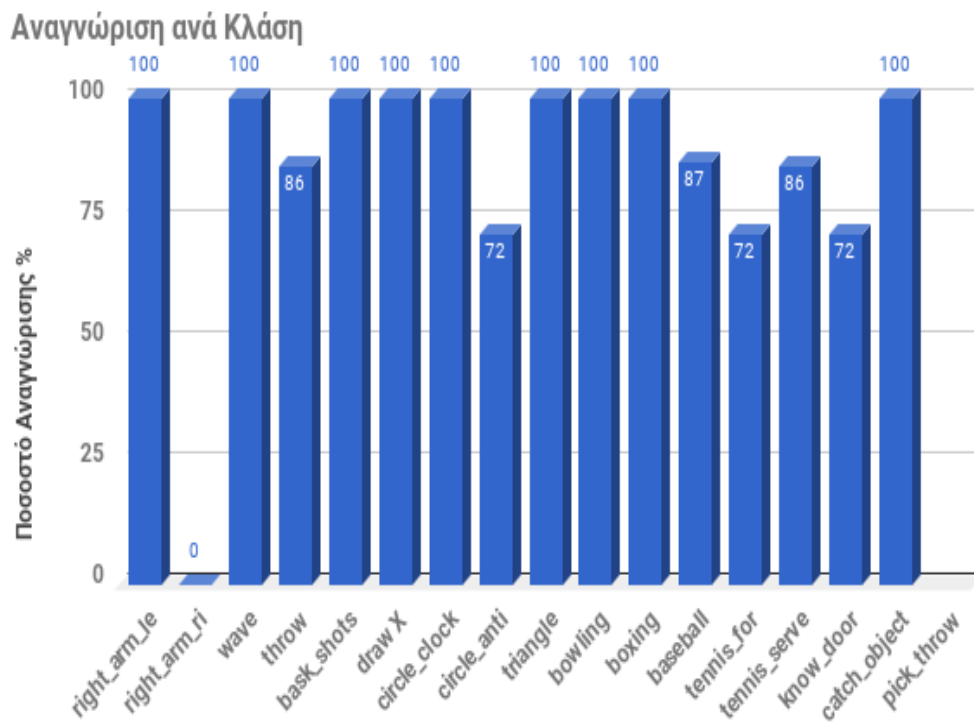
- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 71%
- Κανονικοποίηση: NAI
- Ομαλοποίηση: NAI
- Αριθμός Καταστάσεων: 5
- Συνδυασμός: 2



Σχήμα 5.8: Πείραμα 1: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

## 2. Πείραμα (2)

- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 87%
- Κανονικοποίηση: NAI
- Ομαλοποίηση: NAI
- Αριθμός Καταστάσεων: 17
- Συνδυασμός: 2

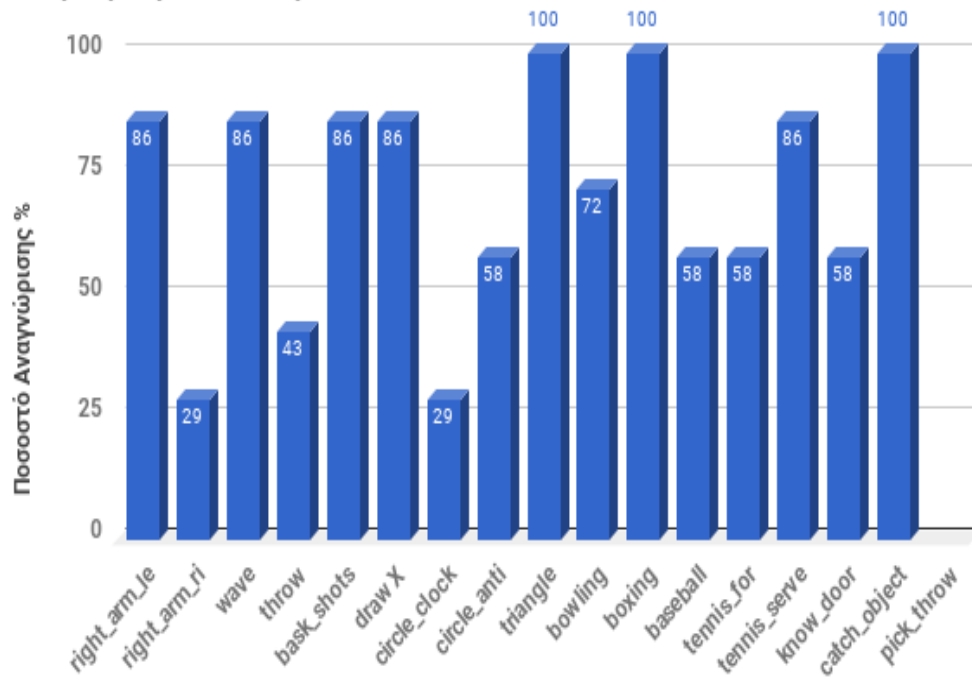


Σχήμα 5.9: Πείραμα 2: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

### 3. Πείραμα (3)

- Αριθμός Νευρώνων: 100
- Ποσοστό Αναγνώρισης: 64%
- Κανονικοποίηση: NAI
- Ομαλοποίηση: NAI
- Αριθμός Καταστάσεων: 3
- Συνδυασμός: 2

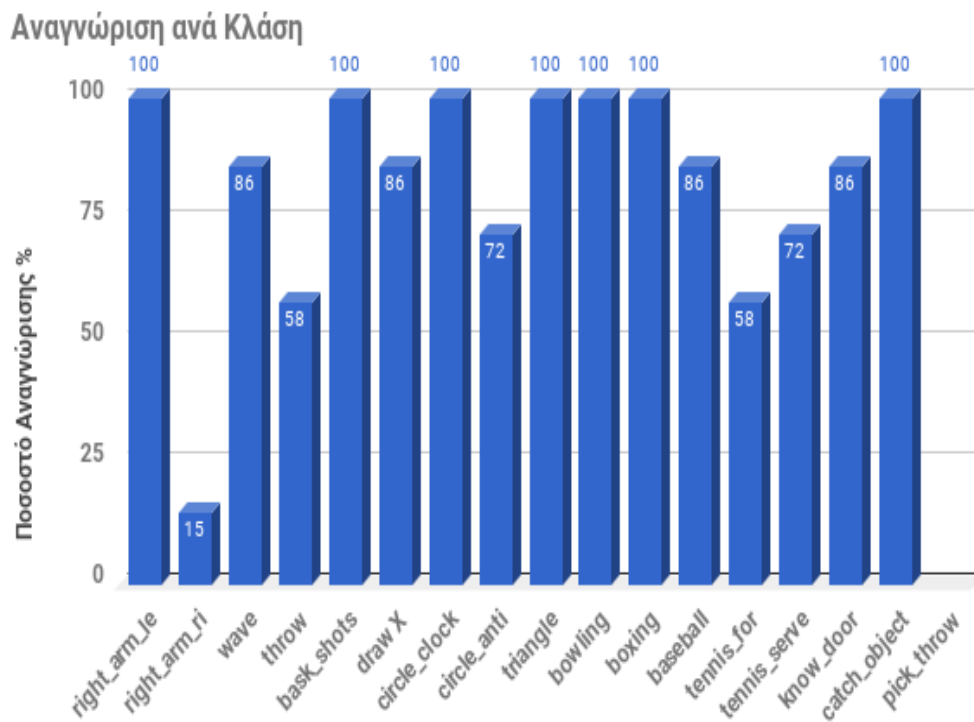
Αναγνώριση ανά Κλάση



Σχήμα 5.10: Πείραμα 3:Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

#### 4. Πείραμα (4)

- Αριθμός Νευρώνων: 100
- Ποσοστό Αναγνώρισης: 69%
- Κανονικοποίηση: NAI
- Ομαλοποίηση: NAI
- Αριθμός Καταστάσεων: 11
- Συνδυασμός: 2

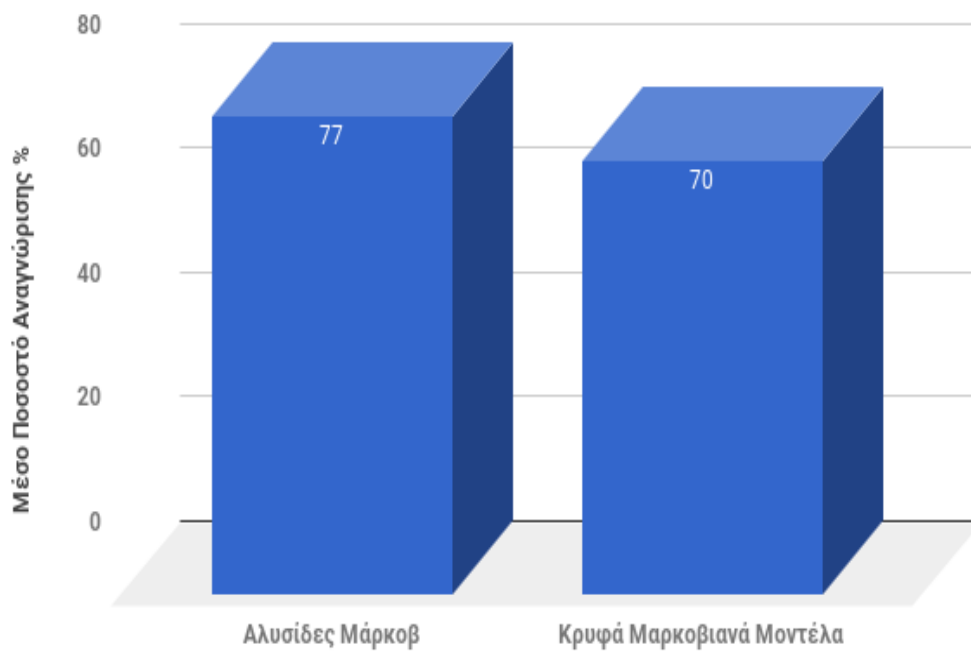


Σχήμα 5.11: Πείραμα 4: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

### 5. Μέση Τιμή Ποσοστού Αναγνώρισης των Μοντέλων :

Η μέση τιμή πάρθηκε χρησιμοποιώντας όλα τα αποτελέσματα, όλων των συνδυασμών των παραμέτρων των μοντέλων που δοκιμάσαμε στο στάδιο των πειραμάτων και προκύπτουν από πάνω από 50 δοκιμές .

#### Αλυσίδες Μαρκόβ εναντίον ΚΜΜ



Σχήμα 5.12: Μέση Τιμή Ποσοστού Αναγνώρισης των Μοντέλων

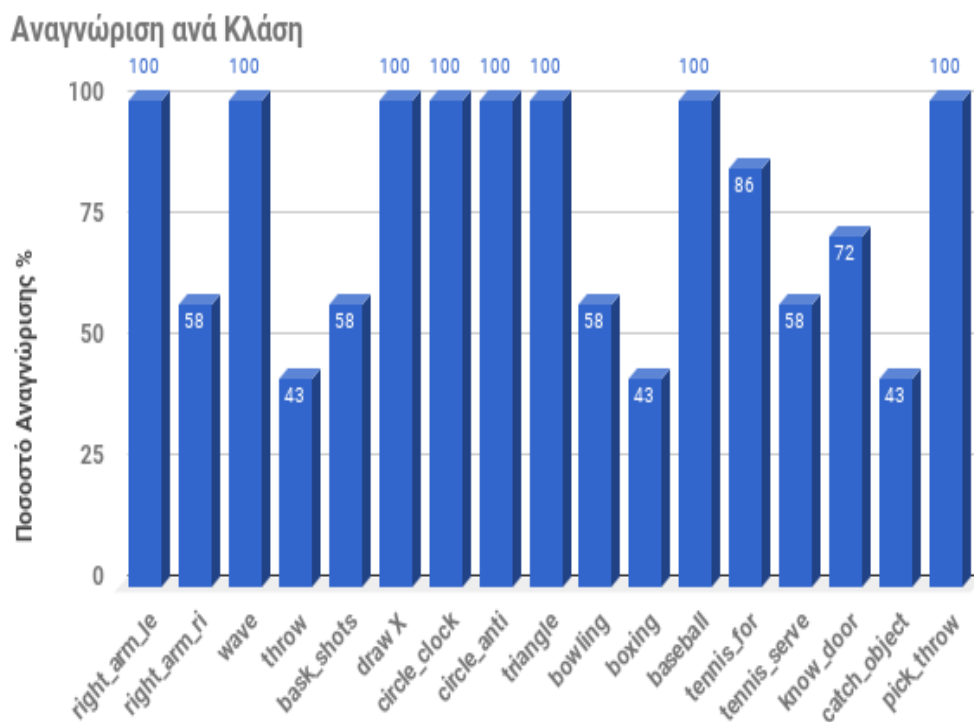


## 5.2.2 Δεδομένα αξιολογήσεων

### Αλυσίδες Μαρκόφ

#### 1. Πείραμα (1)

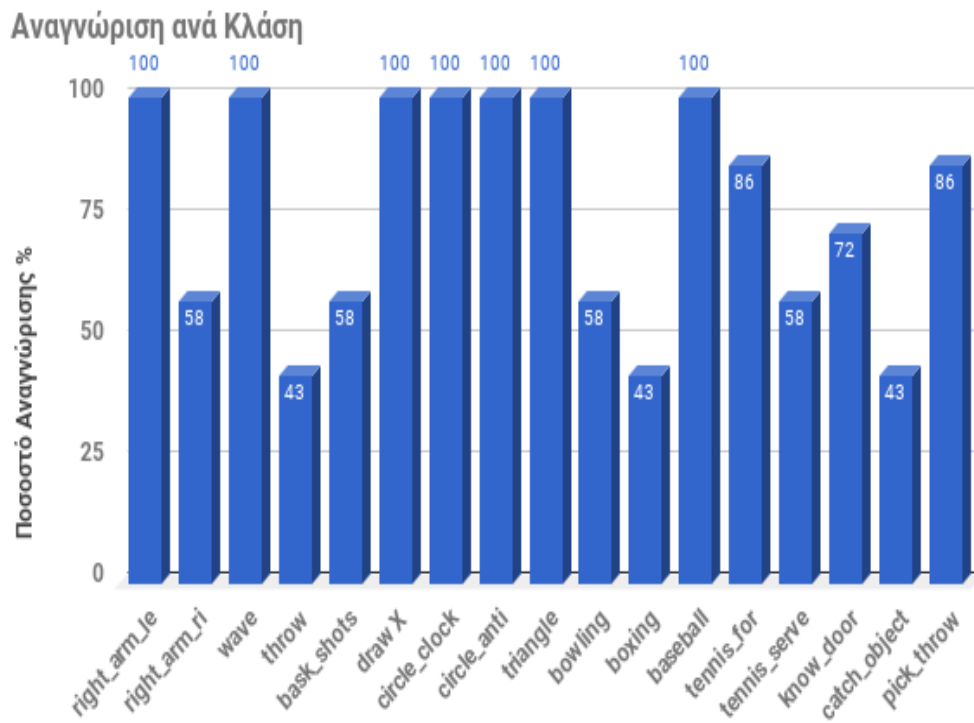
- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 78%
- Κανονικοποίηση: ΟΧΙ
- Ομαλοποίηση: ΟΧΙ
- Συνδυασμός: 1



Σχήμα 5.13: Πείραμα 1: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

## 2. Πείραμα (2)

- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 77%
- Κανονικοποίηση: ΟΧΙ
- Ομαλοποίηση: ΟΧΙ
- Συνδυασμός: 2

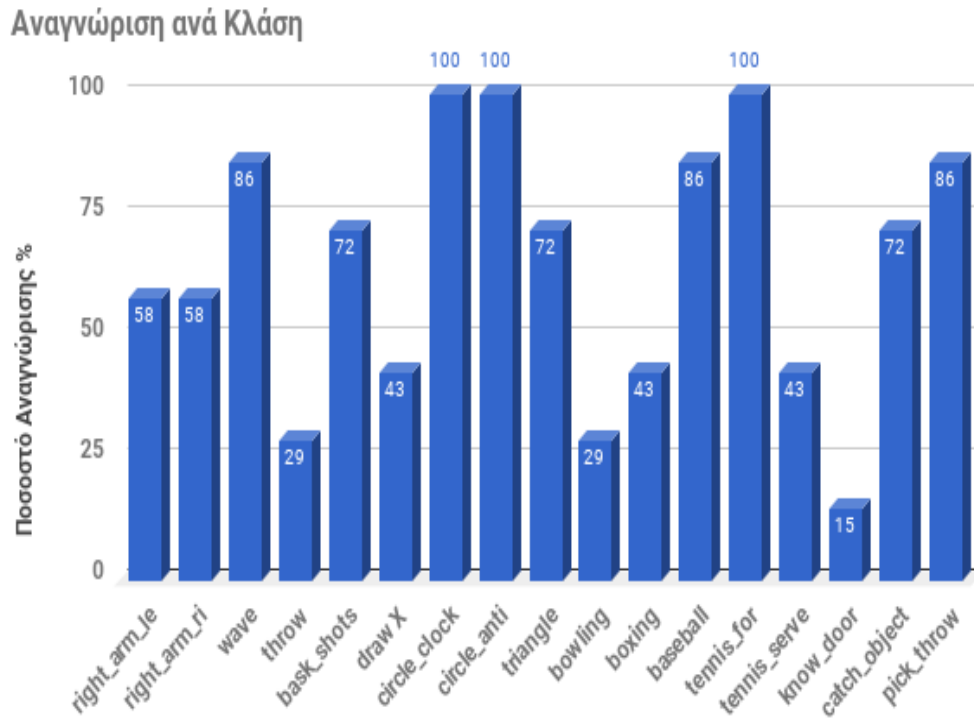


Σχήμα 5.14: Πείραμα 2: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

## Κρυφά Μαρκοβιανά Μοντέλα

### 1. Πείραμα (1)

- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 64%
- Κανονικοποίηση: ΟΧΙ
- Ομαλοποίηση: ΟΧΙ
- Αριθμός Καταστάσεων: 3
- Συνδυασμός: 1

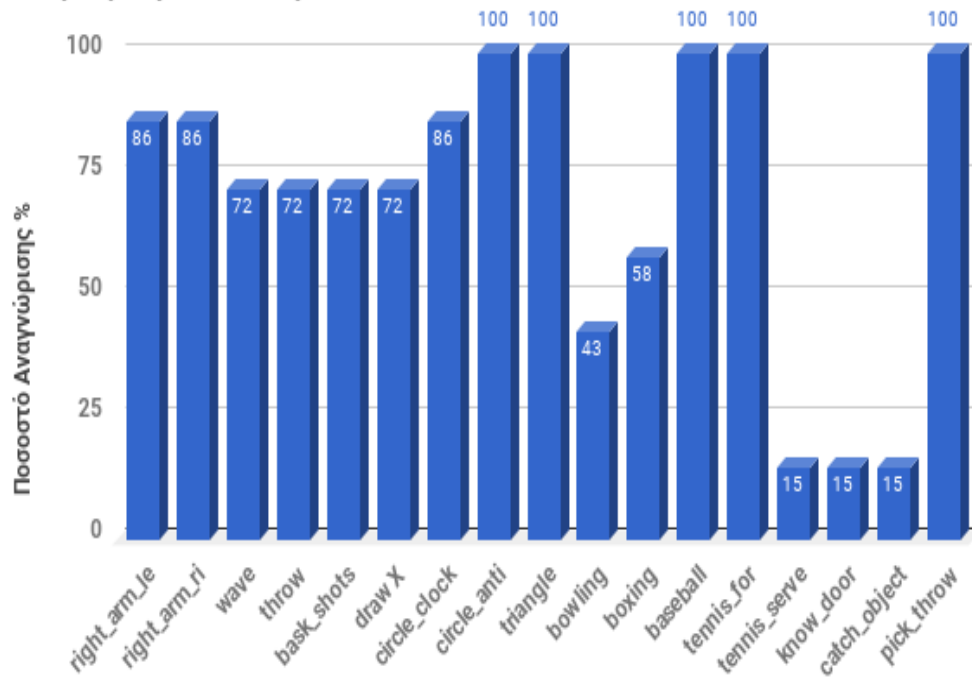


Σχήμα 5.15: Πείραμα 1: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

## 2. Πείραμα (2)

- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 77%
- Κανονικοποίηση: ΟΧΙ
- Ομαλοποίηση: ΟΧΙ
- Αριθμός Καταστάσεων: 11
- Συνδυασμός: 1

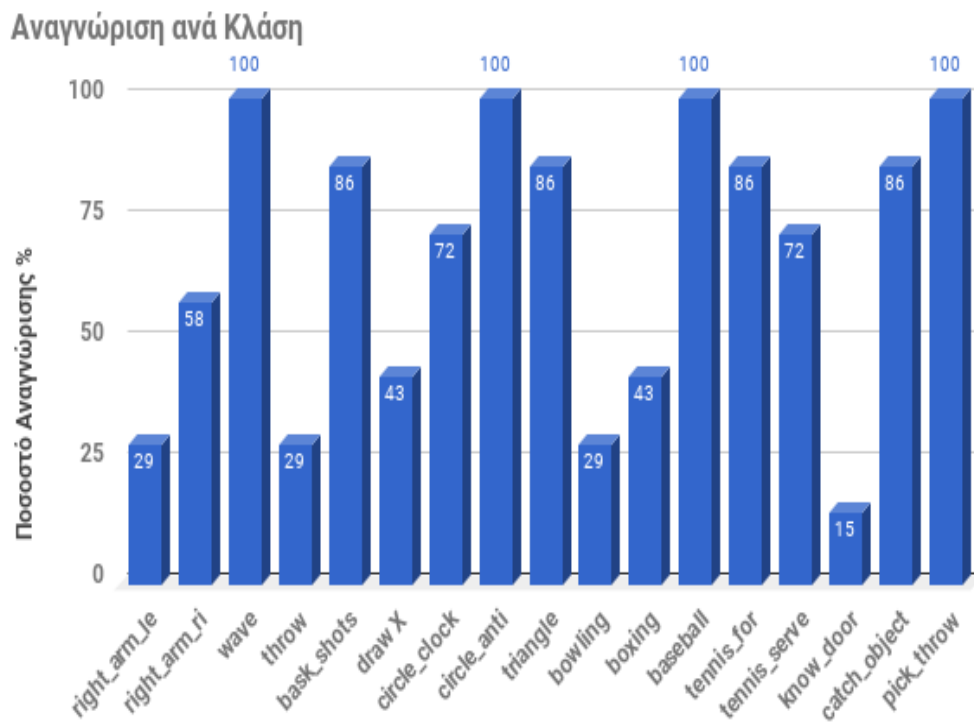
Αναγνώριση ανά Κλάση



Σχήμα 5.16: Πείραμα 2: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

### 3. Πείραμα (3)

- Αριθμός Νευρώνων: 100
- Ποσοστό Αναγνώρισης: 67%
- Κανονικοποίηση: ΟΧΙ
- Ομαλοποίηση: ΟΧΙ
- Αριθμός Καταστάσεων: 3
- Συνδυασμός: 1

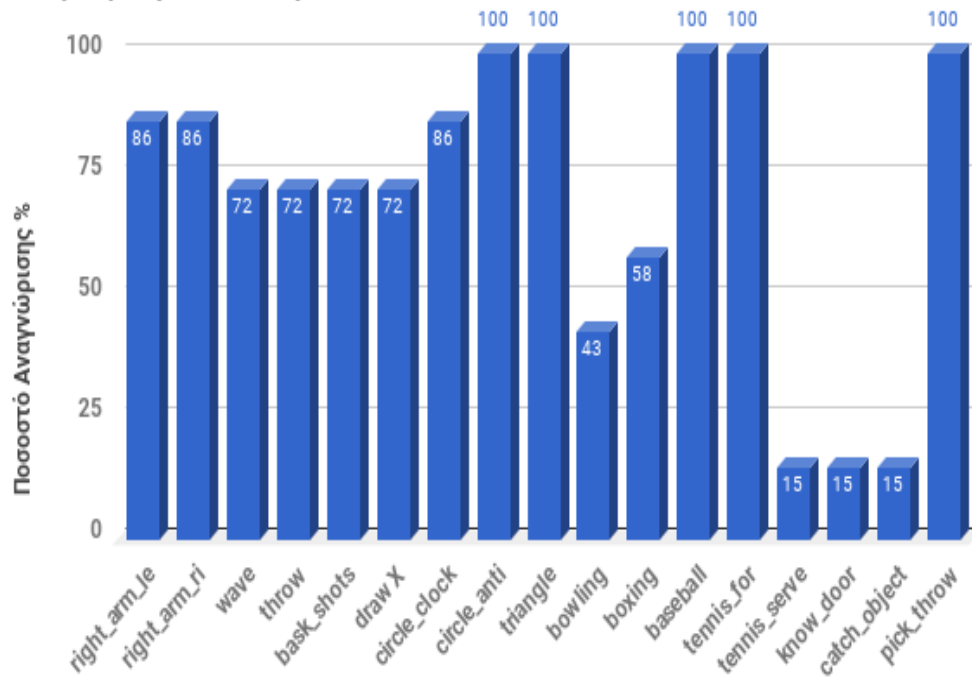


Σχήμα 5.17: Πείραμα 3: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

#### 4. Πείραμα (4)

- Αριθμός Νευρώνων: 100
- Ποσοστό Αναγνώρισης: 77%
- Κανονικοποίηση: ΟΧΙ
- Ομαλοποίηση: ΟΧΙ
- Αριθμός Καταστάσεων: 11
- Συνδυασμός: 1

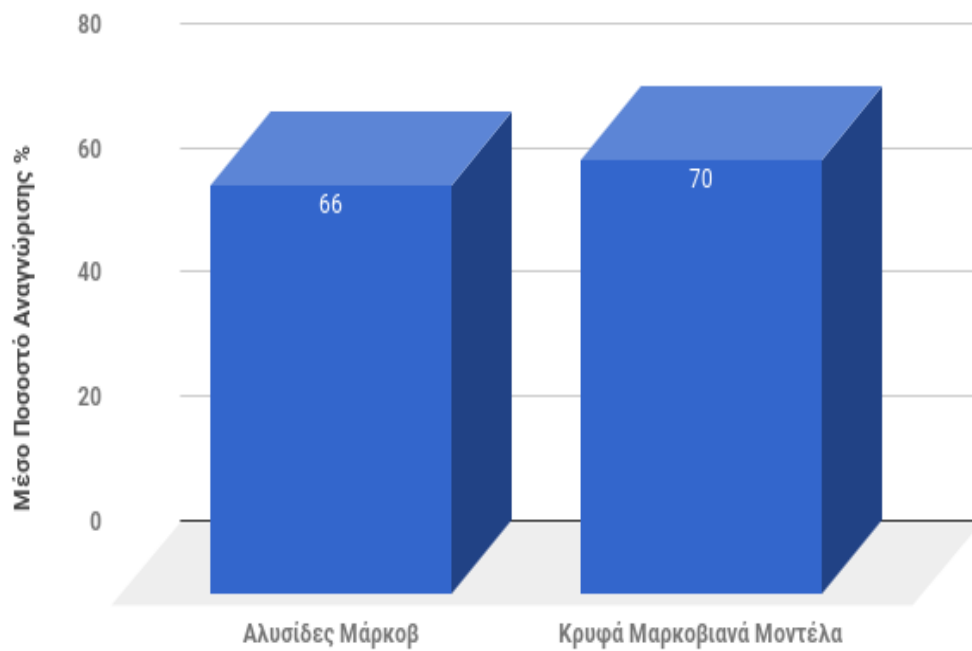
Αναγνώριση ανά Κλάση



Σχήμα 5.18: Πείραμα 4: Γράφημα Ποσοστού Αναγνώρισης ανα Κλάση

## 5. Μέση Τιμή Ποσοστού Αναγνώρισης των Μοντέλων

### Αλυσίδες Μάρκοβ εναντίον ΚΜΜ



Σχήμα 5.19: Μέση Τιμή Ποσοστού Αναγνώρισης των Μοντέλων

## 5.3 Πίνακες Ταξινόμησης (Confusion Matrices)

Στο σημείο αυτό θα απεικονίσουμε τους Πίνακες Ταξινόμησης για ορισμένα δείγματα συγκεκριμένων δοκιμών.

1. Δείγμα (1)

- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 89%
- Κανονικοποίηση: NAI
- Ομαλοποίηση: OXI
- Συνδυασμός: 1

| Πίνακας Ταξινόμησης |                |                |                |                |                |                |                |                |                |                 |                 |                 |                 |                 |                 |                 |                 |
|---------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Κλάσεις%            | c <sub>1</sub> | c <sub>2</sub> | c <sub>3</sub> | c <sub>4</sub> | c <sub>5</sub> | c <sub>6</sub> | c <sub>7</sub> | c <sub>8</sub> | c <sub>9</sub> | c <sub>10</sub> | c <sub>11</sub> | c <sub>12</sub> | c <sub>13</sub> | c <sub>14</sub> | c <sub>15</sub> | c <sub>16</sub> | c <sub>17</sub> |
| c <sub>1</sub>      | 100            | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>2</sub>      | 0              | 100            | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>3</sub>      | 0              | 0              | 57.14          | 0              | 0              | 14.28          | 0              | 28.57          | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>4</sub>      | 0              | 0              | 0              | 85.71          | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 14.28           |
| c <sub>5</sub>      | 0              | 0              | 0              | 0              | 71.24          | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 14.28           | 0               | 14.28           |
| c <sub>6</sub>      | 0              | 0              | 0              | 0              | 0              | 100            | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>7</sub>      | 0              | 0              | 0              | 0              | 0              | 0              | 100            | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>8</sub>      | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 100            | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>9</sub>      | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 14.28          | 85.71          | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>10</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 100             | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>11</sub>     | 0              | 0              | 0              | 0              | 14.28          | 0              | 0              | 0              | 0              | 14.28           | 57.14           | 0               | 0               | 0               | 14.28           | 0               | 0               |
| c <sub>12</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 100             | 0               | 0               | 0               | 0               | 0               |
| c <sub>13</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 28.57           | 71.42           | 0               | 0               | 0               | 0               |
| c <sub>14</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 100             | 0               | 0               | 0               |
| c <sub>15</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 100             | 0               | 0               |
| c <sub>16</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 28.57           | 71.42           | 0               |
| c <sub>17</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 100             |

Πίνακας 5.3: Πίνακας Ταξινόμησης 1



Στο σημείο αυτό θα ορίσουμε δύο μετρικές, οι οποίες μπορούν να εξαχθούν απευθείας από υπολογισμούς του Πίνακα Ταξινόμησης. Αυτές είναι η Μέση Γενικευμένη Ακρίβεια (Mean Generalized Precision) και η Μέση Γενικευμένη Ανάκληση (Mean Generalized Recall). Η Γενικευμένη Ακρίβεια μας δίνει το κλάσμα των δεδομένων που ταξινομήθηκαν σωστά στην κλάση τους σε σχέση με όλα τα δεδομένα που ταξινομήθηκαν σε αυτήν την κλάση. Η Γενικευμένη Ανάκληση μας δίνει το ποσοστό των δεδομένων που ταξινομήθηκαν στην κλάση τους σε σχέση με όλα τα δεδομένα που ανήκουν σε αυτήν την κλάση. Άρα η μέση τιμή είναι ο μέσος όρος αυτών των δεικτών πάνω σε όλες τις κλάσεις δηλαδή:

- Μέση Γενικευμένη Ακρίβεια:  $MGAκ = \left( \sum_i^c \frac{M_{ii}}{\sum_j M_{ji}} \right) / c$
- Μέση Γενικευμένη Ανάκληση:  $MGA = \left( \sum_i^c \frac{M_{ii}}{\sum_j M_{ij}} \right) / c$

Παρακάτω δίνεται η Μέση Γενικευμένη Ακρίβεια και η Μέση Γενικευμένη Ανάκληση για το συγκεκριμένο Δείγμα:

- MGAκ: 91.03%
- MGA: 88.23%

## 2. Δείγμα (2)

- Αριθμός Νευρώνων: 64
- Ποσοστό Αναγνώρισης: 64%
- Κανονικοποίηση: ΝΑΙ
- Ομαλοποίηση: ΟΧΙ
- Αριθμός καταστάσεων: 5
- Συνδυασμός: 2

| Πίνακας Ταξινόμησης |                |                |                |                |                |                |                |                |                |                 |                 |                 |                 |                 |                 |                 |                 |
|---------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Κλάσεις%            | c <sub>1</sub> | c <sub>2</sub> | c <sub>3</sub> | c <sub>4</sub> | c <sub>5</sub> | c <sub>6</sub> | c <sub>7</sub> | c <sub>8</sub> | c <sub>9</sub> | c <sub>10</sub> | c <sub>11</sub> | c <sub>12</sub> | c <sub>13</sub> | c <sub>14</sub> | c <sub>15</sub> | c <sub>16</sub> | c <sub>17</sub> |
| c <sub>1</sub>      | 57.14          | 0              | 0              | 0              | 0              | 28.57          | 0              | 0              | 14.28          | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>2</sub>      | 0              | 100            | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>3</sub>      | 0              | 0              | 14.28          | 0              | 0              | 0              | 0              | 14.28          | 0              | 0               | 0               | 42.85           | 0               | 0               | 0               | 0               | 28.57           |
| c <sub>4</sub>      | 0              | 0              | 0              | 85.71          | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 14.28           |
| c <sub>5</sub>      | 0              | 0              | 0              | 0              | 71.24          | 0              | 14.28          | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 14.28           |
| c <sub>6</sub>      | 0              | 0              | 0              | 0              | 0              | 100            | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>7</sub>      | 0              | 0              | 14.28          | 0              | 0              | 0              | 85.71          | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>8</sub>      | 0              | 0              | 57.14          | 0              | 0              | 0              | 14.28          | 28.57          | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>9</sub>      | 0              | 0              | 28.57          | 0              | 0              | 0              | 28.57          | 0              | 42.85          | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 0               |
| c <sub>10</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 71.42           | 0               | 0               | 0               | 0               | 0               | 0               | 28.57           |
| c <sub>11</sub>     | 0              | 0              | 0              | 0              | 28.57          | 0              | 0              | 0              | 14.28          | 0               | 42.85           | 0               | 0               | 0               | 0               | 0               | 14.28           |
| c <sub>12</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 100             | 0               | 0               | 0               | 0               | 0               |
| c <sub>13</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 57.14           | 42.85           | 0               | 0               | 0               | 0               |
| c <sub>14</sub>     | 0              | 0              | 0              | 0              | 28.57          | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 57.14           | 0               | 0               | 14.28           |
| c <sub>15</sub>     | 0              | 0              | 0              | 0              | 28.57          | 14.28          | 14.28          | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 14.58           | 28.57           | 0               |
| c <sub>16</sub>     | 0              | 0              | 0              | 0              | 14.28          | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 71.42           | 14.28           |
| c <sub>17</sub>     | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0              | 0               | 0               | 0               | 0               | 0               | 0               | 0               | 100             |

Πίνακας 5.4: Πίνακας Ταξινόμησης 2

Παρακάτω δίνεται η Μέση Γενικευμένη Ακρίβεια και η Μέση Γενικευμένη Ανάκληση για το συγκεκριμένο Δείγμα:

- ΜΓΑκ: 74.73%
- ΜΓΑ: 63.86%

## 5.4 Συμπεράσματα Αποτελεσμάτων

Από το μικρό δείγμα αποτελεσμάτων που παρατίθενται μπορούμε να δούμε ότι οι χειρονομίες που δεν ταξινομούνται σωστά στις περισσότερες περιπτώσεις, ταξινομούνται σε κοντινές κλάσεις ως προς το περιεχόμενο της χειρονομίας που εκτελείται.

Γενικά παρατηρήσαμε ότι οι αλυσίδες Μαρκόβ παρούσιαζαν σχεδόν τα ίδια αποτελέσματα με τα Κρυφά Μαρκοβιανά Μοντέλα. Για μεγάλα πλέγματα Αυτό Οργανούμενων Χαρτών είχαμε μείωση της αναγνώρισης. Αυτό οφείλεται στο μέγεθος των δεδομένων εκπαίδευσης καθώς, τα δεδομένα γινόντουσαν πολύ αραιά αφήνοντας πολλούς βαθμούς ελευθερίας στο σύστημα, κωδικοποιώντας τα δεδομένα με μεγάλο πλήθος συμβόλων. Τα καλύτερα αποτελέσματα και για τις δύο μεθόδους τα είχαμε με πλέγματα 64 νευρώνων. Εδώ να σημειωθεί ότι με πολύ μικρότερα πλέγματα είχαμε το αντίθετο αποτέλεσμα, δηλαδή πλέον τα δεδομένα ήταν πολύ πυκνά και δεν γινόταν σωστή αναπαράσταση του χώρου εισόδου. Για τα ΚΜΜ μπορούμε να πούμε ότι οι καταστάσεις παριστούν τη χρονική διάρκεια της κίνησης και στο πλαίσιο αυτό είδαμε ότι αριθμός καταστάσεων σχετικά μεγάλος έδινε τα καλύτερα αποτελέσματα. Τέλος η τρίτη μέθοδος των αλυσίδων Μαρκόβ που χρησιμοποιούσε και την απόσταση των νευρώνων στο πλέγμα δεν αναφέρθηκε στα αποτελέσματα καθώς τα ποσοστά επιτυχίας της άγγιζαν μόλις το 20% για το συγκεκριμένο σύνολο δεδομένων. Αυτό συνέβει γιατί για κάθε δεδομένο αξιολόγησης και για κάθε κλάση έδινε την μέγιστη πιθανότητα, οδηγώντας σε ταξινόμηση του, ανεξάρτητη της κλάσης.

Όσον αφορά τα δεδομένα με τα αξελερόμετρα, ήταν φυσικό να παρουσιάσουν χειρότερα αποτελέσματα καθώς η μέθοδος είχε σχεδιαστεί κυρίως για δεδομένα συντεταγμένων και η φιλοσοφία της ήταν η μοντελοποίηση της τροχιάς του χεριού για αναγνώριση.

## 5.5 Αποτελέσματα Εκφραστικότητας

Στην περίπτωση της εκφραστικότητας δεν δημιουργήσαμε κάποιο μοντέλο κατάταξης. Εξάγαμε απλώς τα στατιστικά χαρακτηριστικά **Spatial ext, Fluidity**

**,Power,Overall Activation,Temporal**, από το σύνολο δεδομένων συντεταγμένων. Τα χαρακτηριστικά αυτά μπορούν να χρησιμοποιηθούν στη συνέχεια για δημιουργία συστήματος αυτόματης κατάταξης των δεδομένων στις διαστάσεις Ενεργοποίησης/Ικανοποίησης (activation, valence), εφόσον έχουν ήδη αξιολογηθεί από χρήστες.

# Κεφάλαιο 6

## Επίλογος

### 6.1 Συμπεράσματα

Έχοντας πάντα κατά νου το σχετικό μικρό σύνολο δεδομένων, αλλά και την πολυπλοκότητα των κινήσεων που τέθηκαν αντικείμενο αναγνώρισης, τα αποτελέσματα των πειραμάτων παρουσιάστηκαν ενθαρρυντικά. Πετύχαμε υψηλά ποσοστά αναγνώρισης τα οποία πλησιάζουν αυτά των κορυφαίων μεθόδων και ξεπερνάνε κατά πολύ τα αποτελέσματα των εκδοτών του συγκεκριμένου συνόλου δεδομένων. Ο συνδυασμός των Αυτό Οργανούμενων χαρτών με τα Μαρκοβιανά μοντέλα, είναι μία τεχνική η οποία μπορεί να κωδικοποιήσει με αποδοτικό τρόπο τα δεδομένα εισόδου και στη συνέχεια να μοντελοποιήσει το ακολουθιακό στοιχείο που τα διέπει. Στην περίπτωση μάλιστα των απλών Μαρκοβιανών αλυσίδων η τεχνική μπορεί να χρησιμοποιηθεί και για σύγχρονη αναγνώριση. Μία αρνητική πτυχή είναι ο μεγάλος χρόνος που χρειάζεται για την εκπαίδευση του Αυτό Οργανούμενου χάρτη που έχει ως αποτέλεσμα την αδυναμία δοκιμής πολλών παραμέτρων, που σχετίζονται με αυτόν, για την ρύθμιση ενός συστήματος. Το θετικό όμως είναι ότι εφόσον ρυθμιστούν κατάλληλα οι παραμέτροι του, στη συνέχεια μπορεί να χρησιμοποιηθεί και αυτός με αρκετά μικρό κόστος.

Όσον αφορά τα πιο τεχνικά κομμάτια της μελέτης μας, κάναμε τις εξής παρατηρήσεις:

- Αριθμός Νευρώνων στο Πλέγμα: Είδαμε ότι όσο μεταβάλλουμε τον αριθμό των νευρώνων έχουμε αντίστοιχα αυξομειώσεις στην απόδοση των συστημάτων μας. Για το συγκεκριμένο πρόβλημα, όπου ο χώρος εκτέλεσης των χειρονομιών ήταν περιορισμένος και εκτεινόταν μέχρι το πολύ την έκταση που μπορεί να καλύψει ένα ανθρώπινο χέρι, ο αριθμός των νευρώνων έπρεπε να μείνει σχετικά μικρός δηλαδή της τάξης

των διαφορετικών τιμών των δεδομένων του προβλήματος. Αυτό είχε ως αποτέλεσμα να καλύπτεται όλο το εύρος των χειρονομιών χωρίς να υπάρχει πλεονάζουσα πληροφορία που θα επηρέαζε στη εκπαίδευση των Μαρκοβιανών Μοντέλων. Αντιθέτως με το πρόβλημα της αραιότητας που εμφανίζεται όταν έχουμε μεγάλο αριθμό νευρώνων στο πλέγμα, παρουσιάζεται μεγάλη πυκνότητα μη ικανή να καλύψει το χώρο των χειρονομιών όταν έχουμε υπερβολικά μικρό αριθμό νευρώνων, καθώς χάνεται πολύτιμη πληροφορία.

- Αριθμός Καταστάσεων KMM: Όπως ήδη έχουμε αναφέρει σε κομμάτια του κειμένου, θα μπορούσαμε να παρομοιάσουμε τον αριθμό καταστάσεων των KMM με τη φυσική έννοια της διάρκειας της χειρονομίας. Δηλαδή μεγάλος αριθμός καταστάσεων παριστά μία μεγάλη σε διάρκεια χειρονομία και μικρός αριθμός το αντίθετο. Καθώς στο σύνολο δεδομένων όλες οι χειρονομίες έχουν προσεγγιστικά την ίδια διάρκεια, ανάλογα, και σχετικός αριθμός καταστάσεων παρουσίασε τα καλύτερα αποτελέσματα.
- Σύγκριση Μεθόδων: Γενικά και οι δύο μέθοδοι μπορούμε να πούμε ότι εμφάνισαν κοντινά μεταξύ τους αποτελέσματα. Αν και τα Κρυφά Μαρκοβιανά Μοντέλα αποτελούν πιο γενική μέθοδο σε σχέση με τις Μαρκοβιανές Αλυσίδες, χάνουν σε απλότητα έχοντας πολλούς βαθμούς ελευθερίας, με αποτέλεσμα κατά γενικό κανόνα να χρειάζονται πολύ περισσότερα δεδομένα για αποτελεσματική εκπαίδευση. Άρα σύμφωνα και με τον κανόνα της απλότητας, μπορούμε να πούμε ότι θα προτιμούσαμε τις Μαρκοβιανές Αλυσίδες ως πιο απλό μοντέλο από τα δύο.
- Δεδομένα Συντεταγμένων: Η μέθοδος αυτή σχεδιάστηκε εξ'αρχής για την επίλυση του προβλήματος των χειρονομιών με δεδομένα εισόδου τις τροχίες των χεριών. Συνεπώς αποτελεί λογικό επακόλουθο να εξάγονται καλύτερα αποτελέσματα με το συγκεκριμένο σύνολο δεδομένων. Πιστεύουμε ότι με τη χρησιμοποίηση ενός αρκετά μεγαλύτερου συνόλου δεδομένων τα αποτελέσματα θα γινότουσαν αρκετά καλύτερα, καθώς θα έδιναν την ευκαιρία στα μοντέλα μας να λάβουν υπ'όψιν και άλλες περιπτώσεις.
- Δεδομένα Μετρήσεων Αξελερόμετρου: Τα δεδομένα αυτά καθώς αποτελούνταν από γραμμικές 3Δ επιταχύνσεις και γωνιακές 3Δ ταχύτητες ήταν φυσικό να περιέχουν μεγαλύτερο εύρος τιμών, το οποίο εξαρτάται σε μεγάλο βαθμό από τον τρόπο εκτέλεσης των χειρονομιών από τα υποκείμενα. Αυτό είχε ως αποτέλεσμα την ελλειπή αναπαράσταση του χώρου εισόδου στα μοντέλα μας και συνεπώς χειρότερα αποτελέσματα αναγνώρισης. Όπως ήταν αναμενόμενο τα KMM παρουσίασαν καλύτερες επιδόσεις από τις Μαρκοβιανές Αλυσίδες καθώς η κωδικοποίηση των δεδομένων

μέσω του Αυτό Οργανούμενου Χάρτη δεν έχει κάποια φυσική σημασία όπως στην περίπτωση των δεδομένων Συντεταγμένων.

- Πολυπλοκότητα: Όπως αναφέρθηκε και προηγουμένως οι Μαρκοβιανές Αλυσίδες έχουν την ικανότητα να παρουσιάσουν αποτέλεσμα σε γραμμικό χρόνο (ακολουθώντας απλά τον πίνακα μεταβάσεων). Από την άλλη τα Κρυφα Μαρκοβιανά Μοντέλα χρειάζονται χρόνο τετραγωνικό ως προς τις καταστάσεις τους και γραμμικό ως προς το μέγεθος της ακολουθίας που έχουν να αξιολογήσουν.

## 6.2 Μελλοντική Δουλειά

Η μέθοδος αυτή θα μπορούσε να χρησιμοποιηθεί σε πληθώρα εφαρμογών, όπου η αναγνώριση χειρονομιών αποτελεί κομβικό σημείο. Συγκεκριμένα τα περιβάλλοντα εικονικής και επαυξημένης πραγματικότητας στον χώρο των βιντεο παιχνιδιών ή και άλλων εφαρμογών, είναι τέτοιοι τομείς. Επίσης ο απομακρυσμένος έλεγχος μηχανημάτων που λαμβάνουν εντολές σε μορφή χειρονομιών θα μπορούσε να αποτελέσει ένα ακόμα χώρο αξιοποίησης της μεθόδου.

Όσον αφορά τις παραμέτρους εκαφραστικότητας αυτές θα μπορούσαν να χρησιμοποιηθούν ως χαρακτηριστικά σε συστήματα αξιολόγησης συναισθηματικών καταστάσεων 3Δ χειρονομιών.

# Βιβλιογραφία

- [1] S. Mitra and T. Acharya, “Gesture recognition: A survey,” *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART C:APPLICATIONS AND REVIEWS*, vol. 37, no. 3, 2007.
- [2] D. Latapie and P. Boswell, “Gesture recognition,” *Wikipedia*, 2011.
- [3] M. Karam and M. schraefel, *A taxonomy of Gestures in Human Computer Interaction*.
- [4] Quek, McNeil, and Bryl, “Multimodal human discourse:gesture and speech,” *ACM Trans. Computer Human Interaction*, vol. 9, no. 3, pp. 171–193, 2002.
- [5] S. Kettebekov, “Exploiting prosodic structuring of coverbal gesticulation,” *In Proceedings of the 6th international conference on Multimodal interfaces, IEEE Computer Society*, p. 151, 2004.
- [6] A. Wexelbat, “An approach to natural gesture in virtual environments,” *ACM Trans. Computer-Human Interaction*, vol. 2, no. 3, pp. 179–200, 1995.
- [7] S. S. Rautaray and A. Agrawal, “Vision based hand gesture recognition for human computer interaction::a survey,” *Springer Science +Business Media*, 2012.
- [8] A. Just and S. Marcel, “A comparative study of two-state-of-the-art sequence processing techniques for hand gesture recognition,” *Computer Vision and Image Understanding*, vol. 113, pp. 532–543, 2009.
- [9] M. et al, “Associating gesture expressivity with affective representations,” *Engineering Applications of Artificial Intelligence*, 2016.
- [10] G. Caridakis and K. Karpouzis, “Full body expressivity analysis in 3d natural interaction: a comparative study,” *Information Systems application*, 2011.



- [11] P. Elkman and w. v. Friesen, “The repertoire of nonverbal behavior:categories,origins,usage and coding,” *Semiotica*, vol. 1, pp. 49–98, 1969.
- [12] E. Paul, “Differential communication of affect by head and body cues,” *Journal of Personality and Social Psychology*, vol. 2, p. 726, 1965.
- [13] P. Elkman and W. V. Friesen, “Head and body cues in the judgment of emotion:a reformulation,” *Journal of Communication*, vol. 22, no. 353-374, 1972.
- [14] S. Haykin, *Neural Networks and Learning Machines*. Pearson, Prentice Hall, 2009.
- [15] L. R. Rabiner, *A tutorial in Hidden Markov Models and Selected Applications in Speech Recognition*.
- [16] M. Stamp, “A revealing introduction to hidden markov models,” Department of Computer Science San Jose State University, dec 2015.
- [17] A. B. Poritz, “Hidden markov models:a guided tour,” Institute for Defense Analyses,Communications research division ,Princeton,NJ 08540.
- [18] C. Chen, R. Jafari, and N. Kehtarnavaz, “Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor,” *Proceedings of IEEE International Conference on Image Processing*, 2015.
- [19] G. Caridakis, K. Karpouzis, A. Drosopoulos, and S. Kollias, “Somm: Self organizing markov map for gesture recognition,” *Pattern Recognition Letters*, vol. 31, pp. 52–59, 2010.