



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**«Μοντελοποίηση και Αναγνώριση ανθρώπινων
Χειρονομιών στην Επικοινωνία Ανθρώπου -
Υπολογιστή»**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωνσταντίνος Γ. Βολουδάκης
Κωνσταντίνος Α. Μαρμαροκόπος

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής

Αθήνα , Ιούλιος 2005



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

**«Μοντελοποίηση και Αναγνώριση ανθρώπινων
Χειρονομιών στην Επικοινωνία Ανθρώπου -
Υπολογιστή»**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωνσταντίνος Γ. Βολουδάκης
Κωνσταντίνος Α. Μαρμαροκόπος

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 31^η
Ιουνίου 2005

.....
Κόλλιας Στέφανος
Καθηγητής

.....
Τσανάκας Παναγιώτης
Καθηγητής

.....
Σταφυλοπάτης Ανδρέας
Γεώργιος
Καθηγητής

Αθήνα , Ιούλιος 2005

.....
Κωνσταντίνος Γ. Βολουδάκης ,2005
Κωνσταντίνος Α. Μαρμαροκόπος ,2005

Διπλωματούχοι Ηλεκτρολόγοι Μηχανικοί και Μηχανικοί
Υπολογιστών Ε.Μ.Π.

Copyright © Κωνσταντίνος Γ. Βολουδάκης
Copyright © Κωνσταντίνος Α. Μαρμαροκόπος

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.

ΕΥΧΑΡΙΣΤΙΕΣ

Καταρχήν, θα ήθελα να ευχαριστήσω τον Καθηγητή μας, κ. Στέφανο Κόλλια, που μας εμπιστεύθηκε με την παρούσα διπλωματική εργασία. Με την στάση του, μας βοήθησε να εργαστούμε απρόσκοπτα για την ολοκλήρωσή της. Οι γνώσεις και οι εμπειρίες που αποκομίσαμε μέσα από την διαδρομή αυτής της εργασίας είναι πολύ σημαντικές και θα σίγουρα θα μας φανούν χρήσιμες στο μέλλον.

Εν συνεχεία, θα ήθελα να ευχαριστήσω τον κ. Κωνσταντίνο Καρπούζη για την πολύτιμη καθοδήγηση, συνεργασία και βοήθεια που μας προσέφερε, στην επίλυση των προβλημάτων που ανέκυπταν κατά διάρκεια της διπλωματικής εργασίας.

Θα ήθελα, επίσης, να ευχαριστήσω τους γονείς μου για την αμέριστη συμπαράσταση τους –ηθική και υλική-, την υπομονή και την καθοδήγηση τους, καθ' όλη την διάρκεια των σπουδών μου. Το μεγαλύτερο μερίδιο της επιτυχούς ολοκλήρωσης της φοίτησης μου, το οφείλω σε αυτούς.

Ευχαριστώ το φίλο και συνεργάτη μου στην συγγραφή της παρούσης διπλωματικής εργασίας, Κωνσταντίνο Μαρμαροκόπο. Ειλικρινά, δεν γνωρίζω εάν θα τα είχαμε καταφέρει χωρίς τη συμβολή του. Οι δυσκολίες που περάσαμε και οι τρόποι που καταφέρναμε κάθε φορά να τις αντιμετωπίζουμε, μα πάνω απ' όλα οι συνεχείς εναλλαγές συναισθημάτων, από την πρώτη μέχρι και την τελευταία στιγμή, θα μου μείνουν αξέχαστες.

Τέλος, θέλω να ευχαριστήσω, τρεις φίλους, που πιστεύω ότι χωρίς τη συμπαράστασή τους τα πράγματα θα ήταν πολύ διαφορετικά. Ήσασταν πάντα δίπλα μου, όταν σας χρειάστηκα...Κώστα, Νίκο και Αγγέλα, σας ευχαριστώ.

Κωνσταντίνος Βολουδάκης

Θέλω να ευχαριστήσω το φίλο και συνάδελφό μου Κώστα Βολουδάκη για την άψογη συνεργασία μας. Επίσης, τους γονείς μου για τη συμπαράστασή τους, τον Κώστα Καρπούζη για τη πολύτιμη βοήθειά του και φυσικά, τον επιβλέποντα καθηγητή μου κ. Στέφανο Κόλλια, που μας εμπιστεύθηκε αυτή τη διπλωματική εργασία.

Κωνσταντίνος Μαρμαροκόπος

ΠΕΡΙΕΧΟΜΕΝΑ

1 Εισαγωγή	8
1.1 Κίνητρα	8
1.2 Οργάνωση της Διπλωματικής	9
2 Επικοινωνία ανθρώπου υπολογιστή	10
2.1 Εισαγωγή	10
2.2 Η Έννοια και η Εξέλιξη της ΕΑΥ	11
2.3 Στόχοι της ΕΑΥ	12
2.4 Λειτουργικότητα: Θεμελιώδης επιδίωξη της ΕΑΥ	12
2.5 Στοιχεία ενός συστήματος ΕΑΥ	13
2.6 Είσοδος Δεδομένων	14
2.7 ΕΑΥ και Όραση Υπολογιστών	14
2.8 ΕΑΥ και Αναγνώριση Χειρονομιών	15
3 Μοντελοποίηση και αναγνώριση των ανθρωπίνων δραστηριοτήτων και αλληλεπιδράσεων	17
3.1 Μοντελοποίηση του Ανθρωπίνου Σώματος	18
3.2 Επίπεδα Λεπτομερειών	20
3.3 Μέθοδοι Αναγνώρισης Ανθρωπίνων Ενεργειών	22
4 Ανθρώπινη κίνηση και χειρονομίες	25
4.1 Αναγνώριση της ανθρώπινης κίνησης	25
4.2 Οι Χειρονομίες	27
4.2.1 Βιολογικός και Κοινωνιολογικός Ορισμός και Ταξινόμηση των Χειρονομιών	27
4.2.2 Ταξινόμηση χειρονομιών	28
4.2.3 Τυπολογία Χειρονομιών	29
4.3 Αναγνώριση Φωνής και Κειμένου: Θέματα παράλληλα με την Αναγνώριση Χειρονομιών	30
4.4 Αναγνώριση Χειρονομιών με HMM	31
5 Τα Κρυφά Μαρκοβιαννά Μοντέλα	33
5.1 Μοντελοποίηση των φυσικών Διεργασιών	33
5.2 Τα HMM	34
5.3 Τρία Βασικά Προβλήματα	35
5.4 Τύποι HMM	42
6 Περιγραφή του συστήματος	46

6.1 Γενική Περιγραφή του Συστήματος	46
6.2 Ο Αλγόριθμος Εντοπισμού του Δέρματος και των κινούμενων Μελών	48
6.3 Επιλογή των Χαρακτηριστικών Γνωρισμάτων	54
6.4 Ταξινόμηση με Χρήση HMM	62
6.4.1 Τα Κρυφά Μαρκοβιανά Μοντέλα	62
6.4.2 Χρήση των HMM στο Σύστημα μας	63
7 Υλοποίηση με Διακριτά HMM	67
7.1 Κβαντισμός των Χαρακτηριστικών	67
7.2 Υλοποίηση	68
7.2.1 Εκπαίδευση των Μοντέλων	68
7.2.2 Αποκωδικοποίηση	73
7.3 Πειραματικά Αποτελέσματα	79
8 Υλοποίηση με Συνεχή HMM	81
8.1 HMM με συνεχείς Κατανομές Εξόδου και Γκαουσιανά Μείγματα	81
8.1.1 Ο Αλγόριθμος Baum – Welch για Γκαουσιανά Μείγματα	82
8.2 Υλοποίηση	85
8.3 Πειραματικά Αποτελέσματα	87
9 Σύνοψη	90
Βιβλιογραφία	92

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ

Στα πλαίσια της παρούσας διπλωματικής εργασίας μελετάται η χρήση των HMM στην αναγνώριση χειρονομιών. Προκειμένου να γίνουν πλήρως κατανοητές οι εφαρμογές των HMM στο αντικείμενο της αναγνώρισης χειρονομιών, η διπλωματική αυτή περιγράφει το σχεδιασμό, την υλοποίηση και τα πειράματα που διενεργήθηκαν πάνω σε δύο συστήματα εκπαίδευσης HMM και αναγνώρισης χειρονομιών με HMM. Το πρώτο σύστημα βασίζεται στη χρήση HMM με διακριτές κατανομές εξόδου, ενώ το δεύτερο με συνεχείς.

1.1 Κίνητρα

Από μικρά παιδιά μαθαίνουμε την ειδική σημασία που περιέχουν ορισμένες κινήσεις του ανθρώπινου σώματος. Για παράδειγμα, όταν ένας άνθρωπος υψώσει το χέρι του, προς το μέρος ενός άλλου ανθρώπου, με ανοιχτή και τεντωμένη την παλάμη και το κουνήσει κατά την οριζόντια διεύθυνση (χαιρετισμός), ο άνθρωπος προς τον οποίο απευθύνεται αυτή η χειρονομία, καταλαβαίνει πως η χειρονομία αυτή είναι ένδειξη φιλίας και ανταποκρίνεται αντιστοίχως.

Η κατανόηση των ανθρώπινων συναισθημάτων μπορεί να θεωρηθεί ως ένα πρόβλημα αναγνώρισης προτύπων. Προκειμένου ο άνθρωπος να μεταφέρει οπτικά μηνύματα σε κάποιον δέκτη, εκφράζει τα συναισθήματά του με κάποια πρότυπα. Τα πρότυπα αυτά, ονομάζονται ,γενικά, χειρονομίες και εμφανίζουν μεγάλη ποικιλία, παρόλα αυτά, τις περισσότερες φορές, διακρίνονται σαφώς μεταξύ τους και περικλείουν κάποιο συγκεκριμένο νόημα. Για παράδειγμα η χειρονομία “wave” (χαιρετισμός) εμφανίζει διαφοροποιήσεις, γιατί μπορεί η θέση του χεριού –ακόμα και του ίδιου ανθρώπου – να διαφέρει μερικά εκατοστά σε σχέση με κάποια προηγούμενη εκτέλεση της ίδιας χειρονομίας, είναι διακριτή γιατί οι άνθρωποι μπορούν εύκολα να καταλάβουν τότε κάποιος τους χαιρετάει και τέλος έχει συγκεκριμένο νόημα το οποίο έχει συμφωνηθεί να είναι το «γεια».

Η έρευνα πάνω στο αντικείμενο της αναγνώρισης χειρονομιών έχει ως κύριο κίνητρο και τελικό στόχο την βελτίωση του τρόπου αλληλεπίδρασης ανθρώπου και υπολογιστή. Αν για παράδειγμα, ένας υπολογιστής μπορεί να αναγνωρίσει και να κατανοήσει το νόημα της χειρονομίας που κάνει κάποιος χρήστης, θα μπορεί να ανταποκριθεί κατάλληλα. Ενδεικτικά αναφέρουμε δύο παραδείγματα, επισημαίνοντας πως οι εφαρμογές που έχουν τα συστήματα αναγνώρισης χειρονομιών είναι πάρα πολλές : ο χρήστης θα μπορεί να κάνει zoom in / out σε φωτογραφίες που υπάρχουν προς επεξεργασία στην οθόνη, ανάλογα με την χειρονομία που πραγματοποιεί, όπως επίσης, θα μπορεί να μετακινεί αντικείμενα της επιφάνειας εργασίας με χειρονομίες αντί του ποντικού.

Η διπλωματική αυτή εργασία, παρουσιάζει μια μέθοδο δημιουργίας ενός συστήματος επικοινωνίας με χειρονομίες, χρησιμοποιώντας, όπως είπαμε, Hidden Markov Models. Τα HMM είναι ένα διπλά στοχαστικά μοντέλα και είναι κατάλληλα για αναπαράσταση των στοχαστικών ιδιοτήτων των χειρονομιών. Τα HMM υιοθετούνται, λοιπόν, για την αναπαράσταση των χειρονομιών, και οι παράμετροι

τους εκτιμώνται από τα δεδομένα εκπαίδευσης. Οι χειρονομίες αναγνωρίζονται μετά την αξιολόγηση των εκπαιδευμένων HMM, με βάση το κριτήριο της πιο πιθανής απόδοσης. Η προτεινόμενη μέθοδος μπορεί να έχει πιθανές εφαρμογές σε ποικίλα προβλήματα επικοινωνίας ανθρώπου-μηχανής, όπως αναγνώριση συναισθημάτων ή ανάπτυξη συστημάτων καθορισμού εντολών με βάση χειρονομίες.

1.2 Οργάνωση της Διπλωματικής

Η παρούσα διπλωματική εργασία αποτελείται από εννέα κεφάλαια. Επιδίωξή μας ήταν η ομαλή εισαγωγή του αναγνώστη στις έννοιες της επικοινωνίας ανθρώπου – υπολογιστή και των μεθόδων μοντελοποίησης της ανθρώπινης κίνησης και η μαθηματική θεμελίωση των Κρυφών Μαρκοβιανών μοντέλων, προτού αναλυθεί το σύστημα αναγνώρισης χειρονομιών που υλοποιήσαμε.

Συγκεκριμένα στο Κεφάλαιο 2, γίνεται η εισαγωγή του αναγνώστη στην έννοια της επικοινωνίας ανθρώπου – υπολογιστή, κλάδος της οποίας είναι η αναγνώριση χειρονομιών. Περιγράφεται, συνοπτικά, η θεμελίωσή και εξέλιξη της, καθορίζονται οι στόχοι της και αναφέρονται οι βασικές αρχές και τα στοιχεία που πρέπει να διέπουν τα συστήματα αλληλεπίδρασης ανθρώπου – υπολογιστή.

Στο Κεφάλαιο 3, ο αναγνώστης εισάγεται στις έννοιες της μοντελοποίησης και αναγνώρισης των ανθρωπίνων κινήσεων και δραστηριοτήτων. Εξετάζονται ζητήματα που αφορούν τους τρόπους μοντελοποίησης του ανθρωπίνου σώματος και των ανθρωπίνων κινήσεων καθώς και του επίπεδο λεπτομερειών που πρέπει να έχουν τα μοντέλα, ανάλογα με την εφαρμογή για την οποία προορίζονται.

Στο Κεφάλαιο 4 εξετάζονται ειδικότερα οι ανθρώπινες χειρονομίες. Αναλύεται η φύση τους, πράγμα απαραίτητο για την σωστή σχεδίαση ενός συστήματος αναγνώρισης χειρονομιών. Δίνεται ένας «ορισμός» του όρου «ανθρώπινη χειρονομία» και παραθέτονται οι κατηγορίες τις οποίες χωρίζονται οι χειρονομίες, με βάση τα ιδιαίτερα χαρακτηριστικά που εμφανίζουν. Τέλος, παρατίθενται ορισμένες μέθοδοι αναγνώρισης των χειρονομιών.

Στο Κεφάλαιο 5 ο αναγνώστης εισάγεται στην έννοια των HMM και ακολουθεί η μαθηματική τους θεμελίωση. Παρουσιάζονται τα τρία βασικά προβλήματα που αφορούν τα HMM και παρατίθενται οι αλγόριθμοι που έχουν αναπτυχθεί για την επίλυσή τους. Αναφέρονται επίσης κάποιες εξειδικευμένες κατηγορίες HMM, που χρησιμοποιήθηκαν στην υλοποίηση του συστήματός μας.

Στο Κεφάλαιο 6 δίνεται η γενική περιγραφή και η δομή του συστήματός που υλοποιήσαμε. Εξετάζονται ζητήματα σχετικά με την αναγνώριση του δέρματος στις ακολουθίες εικόνων, την εξαγωγή των κατάλληλων χαρακτηριστικών γνωρισμάτων που χρησιμοποιήθηκαν για την εκπαίδευση των HMM και την αναγνώριση των χειρονομιών και παρουσιάζεται ο τρόπος με τον οποίον επιτυγχάνεται η αναγνώριση με χρήση των HMM.

Στα Κεφάλαια 7 και 8, αναλύεται, αντιστοίχως, η υλοποίηση του συστήματος με διακριτά και συνεχή HMM. Εξηγείται συνοπτικά ο κώδικας του Matlab και οι ιδιαιτερότητες και τα προβλήματα που παρουσιάστηκαν. Τέλος, παρατίθενται και σχολιάζονται τα πειραματικά αποτελέσματα.

Στο Κεφάλαιο 9, το οποίο είναι και το τελευταίο, αναφέρονται χρήσιμα συμπεράσματα καθώς και προτάσεις για μελλοντικές εργασίες.

Τέλος υπάρχει η **βιβλιογραφία**, όπου αναφέρονται τα επιστημονικά άρθρα και τα βιβλία τα οποία συμβουλευθήκαμε κατά την ανάπτυξη της εφαρμογής μας.

ΚΕΦΑΛΑΙΟ 2

ΕΠΙΚΟΙΝΩΝΙΑ ΑΝΘΡΩΠΟΥ ΥΠΟΛΟΓΙΣΤΗ

2.1 Εισαγωγή

Η ραγδαία ανάπτυξη της επιστήμης των υπολογιστών, το δεύτερο μισό του εικοστού αιώνα, είχε σαν αποτέλεσμα την σταδιακή διείσδυση τους σε ολοένα και περισσότερες πτυχές της ζωής μας. Από τα τεράστια σε όγκο, δύσχρηστα και φτωχά σε απόδοση υπολογιστικά συστήματα, που χρησιμοποιούνταν για συγκεκριμένες επιστημονικές ή στρατιωτικές κυρίως εφαρμογές, έχουμε πια περάσει σε μια εποχή που ο προσωπικός υπολογιστής, καθώς και οι υπόλοιπες ηλεκτρονικές συσκευές, αποτελούν πλέον σημαντικό εργαλείο και είναι αναπόσπαστο κομμάτι της καθημερινότητας μας. Σε αυτό συνετέλεσαν πολλοί παράγοντες μεταξύ των οποίων, η μείωση του μεγέθους των υπολογιστών, η αύξηση της υπολογιστικής ισχύος των νέων συστημάτων, η διεύρυνση του πεδίου εφαρμογών τους και η ραγδαία μείωση του κόστους τους. Η μεγάλη πρόοδος και οι νέες μέθοδοι σχεδίασης τόσο του υλικού, όσο και του λογισμικού των υπολογιστών, έχουν οδηγήσει σε αυτή την μεγάλη μείωση του κόστους. Ένα υπολογιστικό σύστημα, το ο οποίο πριν λίγες μόνο δεκαετίες, είχε μέγεθος ίσο με ένα ολόκληρο δωμάτιο κτιρίου και κόστιζε μερικά εκατομμύρια ευρώ, τώρα πλέον βρίσκεται τυπωμένο σε ένα chip πυριτίου διαστάσεων μερικών εκατοστών και κοστίζει μερικές δεκάδες ευρώ. Η χρήση του πυριτίου στην κατασκευή των ολοκληρωμένων κυκλωμάτων, έχει κάνει την απόκτηση ενός υπολογιστή τόσο οικονομική ώστε εκατοντάδες εκατομμύρια ανθρώπων παγκοσμίως χρησιμοποιούν πλέον τους υπολογιστές στην εργασία τους, στην βιομηχανία, στην διοίκηση, ακόμα και στην προσωπική τους ζωή (διασκέδαση κτλ.).

Ολοένα και περισσότερο, στην καθημερινή μας ζωή, καλούμαστε να ασχοληθούμε με τον προγραμματισμό συσκευών οι οποίες μας έχουν γίνει πλέον απαραίτητες. Όμως κοινή διαπίστωση είναι, πως η χρήση των συσκευών αυτών δεν είναι καθόλου απλή. Παρόλη την προσπάθεια που καταβάλλεται για την ανάπτυξη συσκευών και εφαρμογών φιλικών προς τον χρήστη, οι περισσότεροι άνθρωποι δεν μπορούν να εξοικειωθούν εύκολα με τις νέες συσκευές που καλούνται να χρησιμοποιήσουν.

Οι υπολογιστές και οι υπόλοιπες ηλεκτρονικές συσκευές πρέπει να σχεδιάζονται βασισόμενες στο ότι οι άνθρωποι που θα τις χρησιμοποιήσουν, έχουν στο μυαλό τους συγκεκριμένες εργασίες που θέλουν να επιτελέσουν με αυτές και, παράλληλα, θέλουν να έχουν την δυνατότητα να συνεχίζουν τις καθημερινές τους εργασίες απρόσκοπτα. Για να επιτευχθεί αυτό όμως πρέπει οι σχεδιαστές των συστημάτων να σκέφτονται με τους όρους εργασίας των τελικών χρηστών και να μεταφράζουν αυτή την γνώση σε ένα λειτουργικό, εκτελέσιμο σύστημα. Η ανάπτυξη όμως εύρωστων διεπιφανειών, ικανών να αντιμετωπίζουν όλα τα λάθη και τις αβλεψίες των χρηστών, είναι εν τέλει ένα δύσκολο και πολύπλοκο πρόβλημα, το οποίο θα πρέπει να λαμβάνεται υπ' όψιν καθ' όλη τη διαδικασία σχεδίασης ενός συστήματος και να μην αποτελεί απλά το τελευταίο στάδιο της σχεδίασης. Αυτό ακριβώς είναι και το αντικείμενο της Επικοινωνίας Ανθρώπου – Υπολογιστή (EAY).

2.2 Η Έννοια και η Εξέλιξη της ΕΑΥ

Η συστηματική μελέτη της απόδοσης του ανθρώπου ξεκίνησε στις αρχές του εικοστού αιώνα, στα εργοστάσια, δίνοντας έμφαση σε χειρονακτικές εργασίες. Ο Β΄ Παγκόσμιος Πόλεμος αποτέλεσε το έναυσμα για την μελέτη της επικοινωνίας ανθρώπου – μηχανής, καθώς οι αντιμαχόμενες πλευρές προσπαθούσαν να παράγουν όσον το δυνατόν πιο αποτελεσματικά οπλικά συστήματα. Το 1949 ιδρύθηκε η Ερευνητική Κοινότητα για την Εργονομία. Ο όρος εργονομία αναφέρεται στην απόδοση του χρήστη στο περιβάλλον οποιουδήποτε συστήματος, υπολογιστικού, μηχανικού ή χειροκίνητου. Με την ευρεία διάδοση των υπολογιστών, ένας διαρκώς αυξανόμενος αριθμός ερευνητών επικεντρώθηκαν στην μελέτη της επικοινωνίας της διάδρασης (interaction) μεταξύ ανθρώπων και υπολογιστών, εξετάζοντας τις φυσικές, θεωρητικές και ψυχολογικές εκφάνσεις αυτής της διαδικασίας. Η έρευνα αυτή ξεκίνησε με τον όρο «επικοινωνία ανθρώπου - μηχανής» αλλά λόγω του ιδιαίτερου ενδιαφέροντος και της ανάπτυξης των υπολογιστών μετονομάστηκε σε «επικοινωνία ανθρώπου – υπολογιστή».

Η ΕΑΥ άπτεται πολλών πεδίων αλλά ειδικά στην σχεδίαση υπολογιστικών συστημάτων αποτελεί ζήτημα κεφαλαιώδους σημασίας. Για όλα τα άλλα πεδία μπορεί να θεωρηθεί ως ένας εξειδικευμένος κλάδος της διαδικασίας σχεδίασης, παρά τις πληροφορίες ζωτικής σημασίας που μπορεί να παρέχει. Στην σχεδίαση όμως υπολογιστικών συστημάτων πρέπει να αποτελεί αναπόσπαστο κομμάτι της διαδικασίας αυτής. Από αυτή την άποψη λοιπόν, η ΕΑΥ περιλαμβάνει την σχεδίαση, την υλοποίηση και την αξιολόγηση διαδραστικών συστημάτων σε σχέση με τον χρήστη.

Όταν μιλάμε για «επικοινωνία ανθρώπου – υπολογιστή», δεν αναφερόμαστε απαραίτητα σε ένα μεμονωμένο χρήστη με ένα προσωπικό υπολογιστή. Με τον όρο χρήστη (user) μπορεί να εννοούμε ένα μεμονωμένο χρήστη, μια ομάδα χρηστών που δουλεύουν μαζί ή μια αλληλουχία χρηστών, καθένας εκ των οποίων ασχολείται με κάποιο επιμέρους θέμα μιας εργασίας. Με τον όρο υπολογιστής (computer) αναφερόμαστε σε οποιαδήποτε τεχνολογία υπολογιστών – από τους προσωπικούς υπολογιστές γενικής χρήσης, μέχρι τα μεγάλης κλίμακας υπολογιστικά συστήματα, τα συστήματα ελέγχου διεργασιών (process control systems), ή τα ενσωματωμένα συστήματα (embedded systems). Το σύστημα μπορεί να περιλαμβάνει και άλλα μέρη εκτός από τους υπολογιστές. Τέλος με τον όρο διάδραση (interaction) εννοούμε οποιαδήποτε επικοινωνία μεταξύ ενός χρήστη και ενός υπολογιστή, άμεση ή έμμεση. Η άμεση διάδραση συνεπάγεται ένα διάλογο με ανάδραση και έλεγχο καθ' όλη τη διάρκεια της εκτέλεσης της εργασίας. Η έμμεση διάδραση μπορεί να περιλαμβάνει μαζική επεξεργασία ή επεξεργασία στο παρασκήνιο. Το σημαντικό γεγονός όμως της όλης διαδικασίας είναι ότι ο χρήστης αλληλεπιδρά / επικοινωνεί με τον υπολογιστή προκειμένου να πετύχει κάποιο στόχο.

2.3 Στόχοι της ΕΑΥ

Οι στόχοι της ΕΑΥ είναι η ανάπτυξη χρήσιμων, λειτουργικών και ασφαλών συστημάτων. Προκειμένου να αναπτυχθούν τέτοια συστήματα ο σχεδιαστής πρέπει να προσπαθήσει :

- Να καταλάβει τους παράγοντες που καθορίζουν τον τρόπο με τον οποίο οι άνθρωποι χρησιμοποιούν την τεχνολογία.
- Να αναπτύξει εργαλεία και τεχνικές που θα του επιτρέψουν να δημιουργήσει κατάλληλα συστήματα.
- Να επιτύχει αποδοτική, λειτουργική και ασφαλή επικοινωνία.
- Να βάλει στην υψηλότερη θέση τον παράγοντα άνθρωπο.

Επιγραμματικά, το σημαντικότερο ζήτημα, όσον αφορά την σχεδίαση ενός συστήματος επικοινωνίας ανθρώπου μηχανής είναι η τοποθέτηση του χρήστη στη βάση της σχεδίασης του συστήματος. Οι ανάγκες, προτιμήσεις και ικανότητες των χρηστών κατά την λειτουργία ενός συστήματος, πρέπει να είναι οι παράγοντες που θα κατευθύνουν την σχεδίαση των συστημάτων αυτών. Δεν θα πρέπει οι χρήστες να αλλάζουν τον τρόπο με τον οποίο χρησιμοποιούν ένα σύστημα, προκειμένου να μπορέσουν να χρησιμοποιήσουν ένα τέτοιο σύστημα. Αντίθετα το σύστημα θα πρέπει να είναι σχεδιασμένο με τέτοιο τρόπο, ώστε να ικανοποιεί τις απαιτήσεις τους.

Για την σχεδίαση ενός διαδραστικού συστήματος απαιτούνται γνώσεις από πολλά επιστημονικά πεδία : την ψυχολογία και την γνωστική επιστήμη (γνώση για την αντίληψη του χρήστη, την γνωστική του ικανότητα και την ικανότητα επίλυσης προβλημάτων) , την εργονομία (γνώση για τις φυσικές ικανότητες του χρήστη), την κοινωνιολογία (κατανόηση του ευρύτερου πλαισίου μέσα στο οποίο λαμβάνει χώρα η διάδραση), την επιστήμη των υπολογιστών (ανάπτυξη της απαιτούμενης τεχνολογίας), τη σχεδίαση γραφικών (για την δημιουργία μιας αποτελεσματικής παρουσίασης της διεπιφάνειας), την συγγραφή τεχνικών κειμένων (για την σύνταξη των συνοδευτικών εγχειριδίων του συστήματος) κ.α. . Για το λόγο αυτό, παρόλο που ο όρος ΕΑΥ περιλαμβάνει όλα τα ανωτέρω μέρη, στην πράξη υπάρχουν εξειδικευμένοι επιστήμονες που επικεντρώνουν το ενδιαφέρον τους σε καθένα από τους παραπάνω τομείς. Έτσι, το τελικό σύστημα, αποτελεί το προϊόν της συνεργασίας εξειδικευμένων επιστημόνων όλων των παραπάνω κλάδων.

2.4 Λειτουργικότητα: Θεμελιώδης επιδίωξη της ΕΑΥ

Αν και δεν υπάρχει κάποια ενοποιημένη θεωρία πάνω στην ΕΑΥ , υπάρχει μια θεμελιώδης αρχή πάνω στην οποία πρέπει να βασίζεται η σχεδίαση οποιουδήποτε διαδραστικού συστήματος και αυτή είναι ότι οι άνθρωποι χρησιμοποιούν τους υπολογιστές για να εκτελέσουν κάποια εργασία. Η παραδοχή αυτής της αρχής μας οδηγεί σε τρεις βασικούς πόλους ενδιαφέροντος : τον άνθρωπο, τον υπολογιστή και την εργασία που εκτελείται. Το σύστημα πρέπει να υποστηρίζει την εργασία του χρήστη, πράγμα το οποίο μας οδηγεί σε ένα τέταρτο σημείο εστίασης , την ευχρηστία/λειτουργικότητα: εάν το σύστημα υποχρεώνει τον χρήστη να υιοθετήσει ένα πολύπλοκο και δύσκολο τρόπο εργασίας τότε δεν είναι εύχρηστο. Αν και η κάθε εργασία δεν μπορεί να αποσαφηνιστεί και να οριστεί αυστηρά, οι υπάρχουσες γνώσεις και τα συγκεκριμένα χαρακτηριστικά και οι ιδιαιτερότητες της κάθε εργασίας, αρκούν για να παράσχουν μια πρώτη κατεύθυνση στη σχεδίαση. Οι

τρέχουσες εργασίες του χρήστη μελετώνται και κατόπιν υποστηρίζονται από υπολογιστές, οι οποίοι με την σειρά τους, μπορούν να επηρεάσουν την φύση της αρχικής εργασίας και να προκαλέσουν την εξέλιξή της.

Η λειτουργικότητα, λοιπόν, είναι μία από τις βασικές ιδέες στην επικοινωνία ανθρώπου μηχανής. Σκοπός είναι η ανάπτυξη εύκολων στην εκμάθηση και λειτουργία συστημάτων. Ένα λειτουργικό σύστημα πρέπει να είναι:

- εύκολο στην εκμάθηση της λειτουργίας του
- εύκολο στην απομνημόνευση του τρόπου λειτουργίας του
- αποδοτικό
- ασφαλές
- ενδιαφέρον / διασκεδαστικό κατά την χρήση

Στο παρελθόν οι υπολογιστές χρησιμοποιούνταν κυρίως από ειδικούς και η επικέντρωση του ενδιαφέροντος στην διεπιφάνεια ήταν μία πολυτέλεια η οποία δεν ετύγγανε ιδιαίτερης προσοχής. Σήμερα όμως οι σχεδιαστές δεν μπορούν να παραμελούν την διεπιφάνεια για χάρη της λειτουργικότητας του συστήματός τους καθώς πλέον αυτά τα δύο διαπλέκονται σε βαθμό τέτοιο, που ο διαχωρισμός του ενός από το άλλο δεν είναι δυνατός. Εάν η διεπιφάνεια είναι φτωχή, η λειτουργικότητα του συστήματος επισκιάζεται. Εάν πάλι υπάρχει μια καλοσχεδιασμένη διεπιφάνεια, ο χρήστης μπορεί να χρησιμοποιήσει πιο αποδοτικά το σύστημα. Με τον τρόπο αυτό αυξάνεται η λειτουργικότητά του.

Πολλά από τα συστήματα και προϊόντα που χρησιμοποιούμε στην καθημερινή μας ζωή (όπως κινητά τηλέφωνα, ηλεκτρονικές ατζέντες, τηλεχειριστήρια, ΑΤΜ, εκτυπωτές, ηχητικά συστήματα, φωτοτυπικά μηχανήματα) είναι σχεδιασμένα χωρίς να έχει ληφθεί υπ' όψιν τους η λειτουργικότητα του συστήματος. Αυτό οδηγεί σε λάθη και χαμένο χρόνο κατά την λειτουργία των συστημάτων.

Οι βασικοί παράγοντες, οι οποίοι επηρεάζουν το κατά πόσο ένα σύστημα είναι λειτουργικό, είναι ο τρόπος με τον οποίο ο υπολογιστής δέχεται τα δεδομένα εισόδου, η ύπαρξη ανάδρασης, δηλαδή η δυνατότητα του υπολογιστή να «καθοδηγεί» το χρήστη στη σωστή χρησιμοποίηση του συστήματος και το κατά πόσο είναι προφανής και ευκολοκατανόητη η κάθε λειτουργία .

2.5 Στοιχεία ενός συστήματος ΕΑΥ

Η διάδραση ανθρώπου υπολογιστή απαιτεί την ύπαρξη κάποιων μέσων επικοινωνίας, καθώς και την δυνατότητα και των δύο πλευρών να επεξεργάζονται και να ανταποκρίνονται στα ερεθίσματα – πληροφορίες που δέχονται. Κάποια από τα βασικά στοιχεία κάθε συστήματος ΕΑΥ είναι τα εξής :

-Το υλικό: πρέπει να έχει την δυνατότητα να μετατρέπει τα σήματα που αποστέλλονται αμφίδρομα ανάμεσα στον «εύστροφο» αλλά «ανακριβή» και επιρρεπή σε λάθη άνθρωπο και τον «χαζό» αλλά «αξιόπιστο» υπολογιστή.

-Το πρόγραμμα : ο υπολογιστής πρέπει να είναι σωστά προγραμματισμένος ώστε να μπορεί να ανταποκρίνεται σε όλα τα δεδομένα που δέχεται ακόμα και αν αυτά δεν έχουν κάποιο ιδιαίτερο νόημα στην συγκεκριμένη εφαρμογή και σε καμία περίπτωση να μην διακόπτει την επικοινωνία.

- Την επικοινωνία μέσω φυσικής γλώσσας και χειρονομιών: τα τελευταία χρόνια γίνεται συνεχώς και μεγαλύτερη προσπάθεια από τους σχεδιαστές συστημάτων ΕΑΥ για την υιοθέτηση της φυσικής γλώσσας, ως μέσον επικοινωνίας μεταξύ ανθρώπου

και υπολογιστή. Στόχος είναι η συνεχής μείωση της πολυπλοκότητας της επικοινωνίας και η όσο το δυνατόν μικρότερη ανάγκη ύπαρξης αυστηρών κανόνων λειτουργίας και εκπαίδευσης του συστήματος από τον χρήστη.

2.6 Είσοδος Δεδομένων

Η είσοδος των δεδομένων σε ένα διαδραστικό σύστημα, όπως προαναφέρθη, είναι πολύ σημαντικό κεφάλαιο και έχει να κάνει τόσο με την ευχρηστία όσο και με την λειτουργικότητα και απόδοση του συστήματος. Είναι πολύ σημαντικό ο υπολογιστής να δέχεται εύκολα, γρήγορα και σωστά τα δεδομένα.

Η είσοδος των δεδομένων μπορεί να γίνει με δύο τρόπους: με μαζική (batch) καταχώρηση δεδομένων και με διαδραστική (interactive) χρήση του υπολογιστή. Η μαζική καταχώρηση δεδομένων χρησιμοποιείται όταν πρέπει να καταχωρηθούν μεγάλες ποσότητες δεδομένων στο σύστημα, σε σαφώς καθορισμένη μορφή, και ο χρήστης ενδιαφέρεται κυρίως για την, όσο το δυνατόν, πιο γρήγορη είσοδο τους. Η διαδραστική δράση συνεπάγεται την είσοδο των δεδομένων από ένα χρήστη ο οποίος κάθεται μπροστά από τον υπολογιστή και εισάγει τα δεδομένα όταν του το ζητάει ο υπολογιστής. Γενικά μπορούμε να πούμε πως ο υπολογιστής χρησιμοποιείται ως ένα εργαλείο για την διαχείριση και την ανάκτηση πληροφοριών. Δεν προστίθενται νέες πληροφορίες στην μνήμη της μηχανής αλλά εισάγονται πληροφορίες με την μορφή εντολών, οι οποίες λένε στον υπολογιστή να εκτελέσει μια συγκεκριμένη ενέργεια ή σύνολο ενεργειών. Όσον αφορά την μαζική εισαγωγή δεδομένων, η διάδραση του ανθρώπου με τον υπολογιστή είναι ελάχιστη (πχ. η εισαγωγή δεδομένων με διάτρητες κάρτες, που διαβάζονται από μια μονάδα ανάγνωσης). Οι διαδραστικές συσκευές εισόδου, μπορούν να διαχωριστούν σε δύο ευρείες κατηγορίες :αυτές που επιτρέπουν την εισαγωγή κειμένου και αυτές που επιτρέπουν την ένδειξη, μετακίνηση και επιλογή συγκεκριμένων στοιχείων πάνω στην οθόνη του υπολογιστή. Η πρώτη κατηγορία απαρτίζεται από συσκευές όπως τα πληκτρολόγια και τα συστήματα αναγνώρισης φωνής, ενώ η δεύτερη από τα ποντίκια, τα joysticks και τις οθόνες επαφής. Τέλος, σημειώνουμε πως τα συστήματα αναγνώρισης χειρονομιών μπορούν να χρησιμοποιηθούν και για την εισαγωγή κειμένου (αναγνώριση και τύπωση στην οθόνη μεμονωμένων χαρακτήρων) αλλά και για την αντικατάσταση άλλων εντολών από κάποια χειρονομία (πχ. το zoom in / out κατά την επεξεργασία μιας εικόνας ή η μετακίνηση κάποιου στοιχείου, που απεικονίζεται στην οθόνη, αφού προηγουμένως έχει αναγνωριστεί η αντίστοιχη χειρονομία).

Η σύνδεση μιας βιντεοκάμερας στον υπολογιστή μας δίνει την δυνατότητα να έχουμε ένα ακόμα κανάλι εισόδου δεδομένων (όραση υπολογιστών). Δεδομένης της ύπαρξης πολλών αλγορίθμων επεξεργασίας των παραγόμενων εικόνων, γίνεται κατανοητό πως οι δυνατότητες διάδρασης του υπολογιστή αυξάνονται σημαντικά.

2.7 ΕΑΥ και Όραση Υπολογιστών

Ένα σύστημα όρασης θα πρέπει να δίνει στον υπολογιστή την δυνατότητα να αναγνωρίζει τους χρήστες του και να προσαρμόζει το σύστημα σύμφωνα με τις απαιτήσεις τους. Ωστόσο, η όραση υπολογιστή μπορεί να έχει βαθύτερη επίδραση στην διάδραση, καθώς μπορεί να δίνει στον υπολογιστή την δυνατότητα να εκτελεί τις εργασίες του ανεξάρτητα από τον χρήστη, εξαλείφοντας την ανάγκη συνεχούς διάδρασης χρήστη – υπολογιστή.

Τα συστήματα όρασης υπολογιστή βασίζονται συνήθως σε μια προσέγγιση η οποία ξεκινάει από την επεξεργασία της εικόνας σε χαμηλό επίπεδο και σταδιακά την επεξεργάζεται σε ολοένα και πιο υψηλό επίπεδο. Η εικόνα αντιμετωπίζεται σαν ένα σύνολο από pixels, τα οποία αναλύονται σταδιακά για να εξαχθεί η σημασία τους. Τυπικά παραδείγματα λειτουργιών επεξεργασίας χαμηλού επιπέδου είναι η αφαίρεση του θορύβου από την εικόνα και η προσαρμογή των επιπέδων φωτεινότητας. Παράδειγμα επεξεργασίας του αμέσως υψηλότερου επιπέδου είναι η ανίχνευση και λέπτυνση των περιγραμμάτων των αντικειμένων που τυχόν περιέχει η εικόνα. Κατόπιν, σε ακόμα υψηλότερο επίπεδο επεξεργασίας, οι παραγόμενες γραμμές θα μπορούσαν να ενώνονται για την δημιουργία ολόκληρων σχημάτων και σ' ένα επόμενο στάδιο θα μπορούσαν να χρησιμοποιούνται υψηλότερου επιπέδου πληροφορίες, τριών διαστάσεων, για την εξαγωγή περιγραφών των σχημάτων από τα περιγράμματα που υπάρχουν στην εικόνα. Η υφή και το χρώμα, μπορούν επίσης να χρησιμοποιηθούν σαν βοηθητικά στοιχεία (όπως πχ. σε μια εικόνα στην οποία θέλουμε να απομονώσουμε το δέρμα των χεριών και του προσώπου ενός ανθρώπου) μέχρι να φτάσουμε στο υψηλότερο δυνατό επίπεδο περιγραφής της εικόνας, βάσει των αντικειμένων που περιλαμβάνει και των σχετικών θέσεών τους. Ωστόσο, τα προβλήματα που πρέπει να επιλυθούν κατά την σχεδίαση ενός συστήματος όρασης είναι πολύπλοκα και για το λόγο αυτό η πρόοδος σε αυτό τον τομέα γίνεται με σχετικά μικρούς ρυθμούς αν αποτελεί αντικείμενο με έντονο ερευνητικό ενδιαφέρον.

2.8 ΕΑΥ και Αναγνώριση Χειρονομιών

Οι χειρονομίες είναι ένα στοιχείο της διάδρασης ανθρώπου υπολογιστή το οποίο έχει ιδιαίτερη σημασία στα πολυτροπικά συστήματα. Η δυνατότητα ελέγχουν του υπολογιστή με συγκεκριμένες κινήσεις των χεριών οι οποίες αντιστοιχούν σε συγκεκριμένες εντολές, θα μας παρείχε σημαντικά πλεονεκτήματα σε πολλές περιπτώσεις που δεν είναι δυνατή η χρήση πληκτρολογίου ή σε περιπτώσεις που ο χρήστης χρησιμοποιεί τις υπόλοιπες αισθήσεις του για την εκτέλεση κάποιας άλλης εργασίας. Για ορισμένες κατηγορίες χρηστών με ειδικές ανάγκες, όπως πχ. προβλήματα ακοής, οι χειρονομίες παρέχουν ένα μέσο επικοινωνίας με το οποίο είναι ιδιαίτερα εξοικειωμένοι. Οι χειρονομίες όμως εκτός του ότι παρουσιάζουν μεγάλες διαφοροποιήσεις από χρήστη σε χρήστη, είναι και από τη φύση τους στοχαστικές διαδικασίες. Για το λόγο αυτό είναι δύσκολος ο σχεδιασμός συστημάτων τα οποία θα αναγνωρίζουν, με μεγάλη ακρίβεια, μεγάλο πλήθος χειρονομιών, οι οποίες θα πραγματοποιούνται από πολλούς διαφορετικούς χρήστες. Εκτός από αυτό, η τεχνολογία αναγνώρισης χειρονομιών έχει μεγάλο κόστος. Απαιτείται είτε ένα σύστημα όρασης μέσω υπολογιστή (οπότε ο χρήστης θα κάνει τις χειρονομίες σε κάποια κάμερα που είναι συνδεδεμένη με τον υπολογιστή και στη συνέχεια θα ακολουθεί η επεξεργασία των εικόνων αυτών και η αναγνώριση της εκάστοτε χειρονομίας) είτε ειδικός εξοπλισμός όπως το dataglove. Το dataglove, παρέχει ευκολότερη πρόσβαση σε εξαιρετικά ακριβείς πληροφορίες, αλλά είναι μια τεχνολογία άβολη για τους χρήστες καθώς αυτοί υποχρεώνονται να φορούν ειδικό γάντι από συνθετικό υλικό. Η ερμηνεία των δειγματοληπτούμενων δεδομένων είναι μια πολύ δύσκολη διαδικασία καθώς τα προβλήματα που ανακύπτουν κατά το διαχωρισμό μιας χειρονομίας σε επιμέρους τμήματα είναι πολλά. Μια ομάδα επιστημόνων από το Toronto ανέπτυξε ένα σύστημα αναγνώρισης χειρονομιών το οποίο μεταφράζει τις κινήσεις του χεριού σε συνθετική ομιλία, χρησιμοποιώντας πέντε νευρωνικά δίκτυα τα οποία δουλεύουν παράλληλα για την εκμάθηση και κατόπιν την ερμηνεία διαφορετικών τμημάτων της εισόδου.

Το Media Room στο MIT χρησιμοποιεί μια διαφορετική προσέγγιση για την ενσωμάτωση χειρονομιών στη διάδραση. Το Media Room έχει ένα τοίχο ο οποίος λειτουργεί σα μια μεγάλη οθόνη, με μικρότερες οθόνες επαφής εκατέρωθεν του χρήστη, ο οποίος κάθεται σε μια καρέκλα που είναι τοποθετημένη στο κέντρο του δωματίου. Ο χρήστης μπορεί να μετακινείται στις διάφορες πληροφορίες που εμφανίζονται, χρησιμοποιώντας τις οθόνες επαφής, ένα joystick ή μέσω φωνής. Η ενσωμάτωση των χειρονομιών στην διάδραση επιτυγχάνεται με τη χρήση ενός κύβου αναγνώρισης θέσης ο οποίος προσαρτάται σε ένα ειδικό μπρασελέ που φοράει ο χρήστης. Το σύστημα «Put that there» χρησιμοποιεί τις προερχόμενες από τις χειρονομίες πληροφορίες σε συνδυασμό με αναγνώριση ομιλίας για να δώσει στον χρήστη τη δυνατότητα να υποδείξει τι θα πρέπει να μετακινηθεί και που, με μια απλή ένδειξη. Αυτή είναι μια πολύ πιο φυσιολογική μορφή διάδρασης σε σύγκριση με το να καθορίζουμε προφορικά τι είναι αυτό που πρέπει να μετακινηθεί και να περιγράψουμε που πρέπει να πάει, ενώ έχει επίσης το πλεονέκτημα της συνοπτικότητας. Μια επέκταση του παραπάνω συστήματος θα ήταν η προσάρτηση μιας συσκευής παρακολούθησης της κίνησης των ματιών αντί της αναγνώρισης των χειρονομιών για τον έλεγχο της οθόνης.

ΚΕΦΑΛΑΙΟ 3

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΤΩΝ ΑΝΘΡΩΠΙΝΩΝ ΔΡΑΣΤΗΡΙΟΤΗΤΩΝ ΚΑΙ ΑΛΛΗΛΕΠΙΔΡΑΣΕΩΝ

ΕΙΣΑΓΩΓΗ

Η επεξεργασία ακολουθιών εικόνων έχει προοδεύσει από το επίπεδο της απλής αναγνώρισης της κίνησης στο να αναγνωρίζει τις πράξεις και αλληλεπιδράσεις ως ξεχωριστά γεγονότα. Η αναγνώριση της ανθρώπινης δραστηριότητας σε ένα video παρέχει δυνατότητα για ανάπτυξη πολλών εφαρμογών όπως η αυτόματη παρακολούθηση, ιατρικές διαγνώσεις, ανάλυση videos από διάφορα σπορ, και η επικοινωνία ανθρώπου – υπολογιστή. Επίσης λόγω της προόδου της τεχνολογίας υπάρχει πλέον η δυνατότητα για επεξεργασία εικόνων και videos σε πραγματικό χρόνο. Με βάση τα παραπάνω, γίνεται κατανοητό, πως η αναγνώριση των ανθρώπινων δραστηριοτήτων σε ακολουθίες video θα είναι μία από τις βασικότερες εφαρμογές του μέλλοντος.

Η αναγνώριση της ανθρώπινης δραστηριότητας από τον υπολογιστή περιλαμβάνει την κατανόηση της ανθρώπινης κίνησης. Η αναγνώριση όμως της ανθρώπινης κίνησης είναι ένα ιδιαίτερος περίπλοκο αντικείμενο. Η δομή και το σχήμα του ανθρώπινου σώματος δεν μπορεί να είναι σαφώς καθορισμένο, λόγω της ύπαρξης πολλών αρθρώσεων και λόγω της ύπαρξης των ενδυμάτων. Επίσης, οι αλλαγές στην φωτεινότητα της εικόνας καθώς και ο θόρυβος που προέρχεται από τις σκιές, δυσκολεύουν ακόμα περισσότερο τις προσπάθειες για αναγνώριση των ανθρώπινων κινήσεων. Για παράδειγμα, η αναγνώριση δραστηριοτήτων σε εξωτερικούς χώρους επηρεάζεται σημαντικά από τις αλλαγές του καιρού και του φωτισμού.

Η κατανόηση της ανθρώπινης κίνησης, μπορεί να προσεγγιστεί με διάφορα επίπεδα λεπτομερειών, ανάλογα με την πολυπλοκότητα της εκάστοτε κίνησης. Η μοντελοποίηση και η αναγνώριση της ανθρώπινης συμπεριφοράς προϋποθέτει τον χαρακτηρισμό και την ταξινόμηση των διαφόρων ειδών κίνησης. Μια ιδέα που εφαρμόστηκε αρχικά για την επίλυση αυτού του ζητήματος ήταν η ταξινόμηση της κίνησης σε « αλλαγή, γεγονός, επεισόδιο και ιστορία» ώστε να υπάρξει αποτύπωση των διαφορετικών διαστάσεων του προβλήματος. Η κάθε διάσταση σχετίζεται και με διαφορετικό όγκο πληροφορίας που απαιτείται για την επίτευξη αναγνώρισης. Μια διαφορετική προσέγγιση είναι ο διαχωρισμός της κίνησης σε «κινήσεις, δραστηριότητα, ενέργεια ». Σε αυτού του είδους την ταξινόμηση οι *κινήσεις* είναι εξατομικευμένες στοιχειώδεις κινήσεις οι οποίες δεν απαιτούν την συλλογή δεδομένων από κάποια ακολουθία για να αναγνωριστούν. Αντίθετα η *δραστηριότητα* αναφέρεται σε μια ακολουθία κινήσεων ή καταστάσεων όπου η μόνη πραγματική γνώση είναι τα στατιστικά χαρακτηριστικά της ακολουθίας. Μεγάλο μέρος της πρόσφατης έρευνας πάνω στην αναγνώριση χειρονομιών εμπίπτει σε αυτή την κατηγορία αναγνώρισης κινήσεων. Τέλος, οι *ενέργειες* είναι γεγονότα μεγαλύτερης διάρκειας τα οποία συνήθως περιλαμβάνουν αλληλεπιδράσεις με το περιβάλλον. Το

αντικείμενο της αναγνώρισης *ενεργειών* είναι συναφές και αλληλένδετο με το αντικείμενων της όρασης υπολογιστών και της τεχνητής νοημοσύνης.

Στην συνέχεια θα δώσουμε έμφαση στο υψηλό-επίπεδο της αναγνώρισης της ανθρώπινης κίνησης δηλαδή στις πράξεις και τις αλληλεπιδράσεις και πιο συγκεκριμένα στην μοντελοποίηση του ανθρώπινου σώματος, στο επίπεδο των λεπτομερειών που χρειάζονται για την αναγνώριση των ανθρωπίνων πράξεων, μεθόδους αναγνώρισης των πράξεων και υψηλού επιπέδου αναγνώριση σκηνών. Η υψηλού επιπέδου αναγνώριση των ανθρώπινων κινήσεων απαιτεί προηγουμένως πολλά βήματα επεξεργασίας χαμηλού – επιπέδου όπως τμηματοποίηση, εντοπισμό, ανάκτηση μορφής και εξαγωγή τροχιάς στα οποία όμως δεν θα αναφερθούμε εκτενώς.

3.1 Μοντελοποίηση του Ανθρώπινου Σώματος

Η ανάλυση της κίνησης του ανθρώπινου σώματος αποτελεί μέρος της αναγνώρισης κίνησης στα πλαίσια της όρασης υπολογιστών. Πολλές μελέτες στην ανάλυση κίνησης χρησιμοποιούν τον διαχωρισμό της κίνησης σε άκαμπτες και μη, βασιζόμενες στο κατά πόσο το αντικείμενο που κινείται είναι στερεό ή όχι. Μια άλλη προσέγγιση είναι η θεώρηση της ανθρώπινης κίνησης σαν αρθρωτή κίνηση, η οποία βεβαίως μπορεί να θεωρηθεί υποκατηγορία της μη άκαμπτης κίνησης. Η αρθρωτή κίνηση του ανθρώπινου σώματος αποτελείται από μικρότερες άκαμπτες κινήσεις των ξεχωριστών μελών του σώματος, η ολική όμως κίνηση είναι μη άκαμπτη.

Μπορούμε να κατατάξουμε τις μελέτες που χρησιμοποιούν την προσέγγιση της αρθρωτής κίνησης σε αυτές που χρησιμοποιούν μια *a priori* μορφή του μοντέλου και σε αυτές που δεν χρησιμοποιούν κάποιο μοντέλο αλλά βασίζονται στην εικόνα του αντικειμένου. Ο διαχωρισμός βασίζεται στο κατά πόσο χρησιμοποιείται στην ανάλυση της κίνησης *a priori* γνώση για τη μορφή του αντικειμένου. Και οι δύο προσεγγίσεις έχουν πλεονεκτήματα και μειονεκτήματα. Οι προσεγγίσεις που βασίζονται στην εικόνα του αντικειμένου μπορούν να έχουν εφαρμογή σε πιο ιδιαίτερες περιπτώσεις αφού δεν απαιτούν κάποιο συγκεκριμένο μοντέλο για το αντικείμενο. Παρόλα αυτά η προσέγγιση με βάση την εικόνα είναι πιο ευαίσθητη στο θόρυβο αφού γιατί δεν διαθέτουν κάποιο μηχανισμό διαχωρισμού του θορύβου από το οπτικό σήμα εισόδου. Από την άλλη η προσέγγιση που βασίζεται σε γνώση του μοντέλου του κινούμενου αντικειμένου μπορεί να συνδυάσει αποδοτικά την γνώση του σχήματος και την είσοδο του οπτικού σήματος, κάνοντας την καλύτερη να αναγνωρίζει υψηλού – επιπέδου πολύπλοκες κινήσεις. Το βασικό μειονέκτημα αυτού του είδους προσέγγισης είναι ότι προϋποθέτει επιπλέον βήματα επεξεργασίας και επιλογής του μοντέλου και εκτίμηση των παραμέτρων του μοντέλου ώστε να ταιριάζει στην δοσμένη οπτική είσοδο. Επίσης, για την προσθήκη κάποιας νέας ενέργειας ή κίνησης μπορεί να χρειαστούν σημαντικής δυσκολίας υπολογισμοί για την αναβάθμιση του μοντέλου.

Οι προσεγγίσεις που βασίζονται στην εικόνα δημιουργούν μια αναπαράσταση του σώματος βασιζόμενες στην ανίχνευση των κατάλληλων χαρακτηριστικών της εικόνας ενώ οι προσεγγίσεις με βάση το μοντέλο δημιουργούν την αναπαράσταση του σώματος με το να προσαρμόσουν στα δεδομένα της εικόνας, τις προκαθορισμένες παραμέτρους ενός παραμετρικού μοντέλου του ανθρώπινου σώματος. Η διαδικασία προσαρμογής των παραμέτρων γίνεται είτε με την βελτιστοποίηση κάποιου κριτηρίου όπως των ελαχίστων τετραγώνων είτε με κάποια στοχαστική διαδικασία δειγματοληψίας όπως η μέθοδος στοιχειώδους φιλτραρίσματος.

Σε κάθε προσέγγιση, το ανθρώπινο σώμα μπορεί να αναπαρασταθεί με διάφορα επίπεδα λεπτομερειών, τα οποία περιλαμβάνουν ευλύγιστα κουτιά, ραβδόμορφα σχήματα, δισδιάστατα χωρία ή τρισδιάστατους όγκους, ανάλογα με την πολυπλοκότητα του μοντέλου που απαιτεί η συγκεκριμένη εφαρμογή. Η αναπαράσταση με βάση τα ευλύγιστα κουτιά είναι ένα από τα πιο απλά μοντέλα του ανθρώπινου σώματος. Η ικανότητα αναπαράστασης που προσφέρει είναι περιορισμένη. Το μοντέλο των ευλύγιστων κουτιών είναι χρήσιμο κυρίως σε ακολουθίες εικόνων που το ανθρώπινο σώμα είναι τόσο μικρό ώστε να καταλαμβάνει μόνο μερικά pixels. Η αναπαράσταση με βάση ραβδόμορφα σχήματα θεωρεί το σώμα σαν μια σύνθεση από παραλληλόγραμμα (ραβδιά) και τις μεταξύ τους αρθρώσεις. Η θεώρηση αυτή βασίζεται στην παρατήρηση ότι η κίνηση του ανθρώπινου σώματος προέρχεται κυρίως από την κίνηση των οστών (τα οποία ουσιαστικά αναπαριστούνται με τα ραβδιά). Η αναπαράσταση με δισδιάστατα χωρία βασίζεται στην ιδέα ότι το ανθρώπινο σώμα που εμφανίζεται στην εικόνα είναι η προβολή του τρισδιάστατου ανθρώπινου σώματος σε ένα δισδιάστατο χώρο. Για το λόγο αυτό προσεγγίζει το σώμα με χωρία χωρίς καθορισμένα όρια. Τα τρισδιάστατα, ογκομετρικά μοντέλα προσπαθούν να περιγράψουν με λεπτομέρειες το τρισδιάστατο ανθρώπινο σώμα με χρήση πολύεδρων στερεών όπως ελλειπτικούς κυλίνδρους, γενικευμένους κώνους ή σφαίρες. Τα παραπάνω μοντέλα αναφέρθηκαν με σειρά αυξανόμενης πολυπλοκότητας αλλά και αυξανόμενου επιπέδου λεπτομερειών. Τα πιο λεπτομερή μοντέλα μπορούν να αναπαραστήσουν πιο πολύπλοκες πλευρές της ανθρώπινης δραστηριότητας, αλλά απαιτούν σημαντικά μεγαλύτερη υπολογιστική πολυπλοκότητα. Για παράδειγμα, τα τρισδιάστατα μοντέλα απαιτούν στερεομετρικά χαρακτηριστικά τα οποία για να αποκτηθούν χρειάζονται πολλές κάμερες.

Το επίπεδο των λεπτομερειών που χρειάζεται για την αναπαράσταση του ανθρώπινου σώματος, εξαρτάται από την εφαρμογή. Για παράδειγμα, σε μερικές εφαρμογές δεν υπάρχει ανάγκη για αναπαράσταση ολόκληρου του σώματος ή για αναπαράσταση πολλών λεπτομερειών σε ορισμένα σημεία αυτού. Σε τέτοιες περιπτώσεις μια πιο απλή προσέγγιση μπορεί να δώσει ικανοποιητικά αποτελέσματα. Ένα τέτοιο παράδειγμα υλοποίησης είναι η μοντελοποίηση κάθε παίκτη ενός αγώνα αμερικάνικου ποδοσφαίρου με χρήση ενός ευλύγιστου κουτιού. Ο εντοπισμός γινόταν με το να διατηρείται το κουτί στην εικόνα διαμέσου των frames.

Το επίπεδο των λεπτομερειών σχετίζεται με περιορισμούς στις φυσικές διαστάσεις των αισθητήρων καταγραφής (κάμερες). Όσο μικρότερη είναι η χωρική ανάλυση σε μια εικόνα τόσο μεγαλύτερη είναι η περιοχή η οποία καταγράφεται, και αντιστρόφως. Για το λόγο αυτό πρέπει να υπάρξει κάποιου είδους συμβιβασμός ανάμεσα στη χωρική ανάλυση και το εύρος καταγραφής. Για να αποφευχθεί το παραπάνω πρόβλημα μπορεί να υιοθετηθεί η χρήση πολλών διασκορπισμένων καμερών ώστε να καλυφθεί με υψηλή ανάλυση ολόκληρος ο χώρος. Ένας άλλος τρόπος επίλυσης του συγκεκριμένου προβλήματος είναι η χρήση κάμερας με περιστρεφόμενη κεφαλή και φακούς εστίασης.

Τα βασικά χαρακτηριστικά που χρησιμοποιούνται για την ανίχνευση του ανθρώπινου δέρματος είναι το χρώμα και η ένταση. Υπάρχουν όμως και πολλά άλλα. Ένα από αυτά είναι το «ιστορικό κίνησης της εικόνας» που αποτελείται από τα συσσωρευμένα κατά την εξέλιξη της ακολουθίας, δυαδικά μέρη του προσκηνίου της εικόνας. Με βάση το ιστορικό κίνησης μπορούν να αναγνωριστούν διαφορετικές ενέργειες ενός ανθρώπου μέσα σε μια ακολουθία. Ένα άλλο χαρακτηριστικό που χρησιμοποιείται για τον εντοπισμό των ανθρώπων και των ενεργειών τους είναι η ταχύτητα των κινούμενων pixel. Pixel τα οποία κινούνται με παρόμοια ταχύτητα θεωρείται πως ανήκουν στο ίδιο μέλος του ανθρώπινου σώματος.

Επίσης σε διάφορες μελέτες έχουν χρησιμοποιηθεί χαρακτηριστικά όπως ο όγκος, οι ακμές, ο ήχος, το χρώμα, η ταχύτητα της οπτικής ροής και η ένταση για την εξαγωγή της δραστηριότητας. Ένα τέτοιο παράδειγμα είναι ο συνδυασμός της έντασης των pixel με τον ήχο της μουσικής για την αναγνώριση των διάφορων ειδών χορού από ζευγάρια ανθρώπων σε κοινωνικές εκδηλώσεις.

Σε μια πρόσφατη μελέτη χρησιμοποιήθηκαν πολλαπλά επίπεδα επεξεργασίας της εικόνας (επίπεδο pixel, μικρών μερών του σώματος, μελών, ακολουθία εικόνων) για την δημιουργία ενός μοντέλου του ανθρωπίνου σώματος, βασισμένου στην «εικόνα του αντικειμένου». Τα pixel ομαδοποιούνται ανάλογα με την ομοιομορφία στο χρώμα τους και στη συνέχεια οι ομάδες που δημιουργούνται, ενώνονται για να αποτελέσουν κάποιο μέλος (πχ. τα κεφάλι, το στήθος, τη λεκάνη, το πρόσωπο, τα μαλλιά, τα χέρια και τα πόδια). Τέλος τα μέλη του σώματος εντοπίζονται κατά μήκος της ακολουθίας των εικόνων.

3.2 Επίπεδα Λεπτομερειών

Μεγάλο μέρος της έρευνας πάνω στην αναγνώριση των πράξεων έχει επικεντρωθεί στην ανάλυση των δραστηριοτήτων ενός μόνο ατόμου. Η αναγνώριση όμως των αλληλεπιδράσεων περιλαμβάνει αλληλεπιδράσεις μεταξύ δύο ατόμων, μεταξύ μιας ομάδας τριών ή περισσότερων ατόμων, μεταξύ ανθρώπου και υπολογιστή και μεταξύ ανθρώπου και κάποιου αντικειμένου. Γενικά, κάθε μία από τις παραπάνω περιπτώσεις απαιτεί διαφορετικό επίπεδο ανάλυσης της εικόνας και διαφορετικό τρόπο αναπαράστασης του προς αναγνώριση σκηνηκού. Όσο περισσότεροι άνθρωποι περιλαμβάνονται στην εικόνα τόσο λιγότερα pixel θα καταλαμβάνει ο καθένας από αυτούς, συντελώντας έτσι σε μια εικόνα χαμηλής ανάλυσης. Για το λόγο αυτό κάθε περίπτωση χρειάζεται και διαφορετικές μεθόδους επεξεργασίας. Η αναγνώριση των πράξεων και των αλληλεπιδράσεων μπορεί να επιτευχθεί με διαφορετικά επίπεδα λεπτομερειών κατά την ανάλυση. Αυτά είναι : το γενικό επίπεδο, το ενδιάμεσο και το λεπτομερές.

Στο γενικό επίπεδο, κάθε άτομο αναπαρίσταται με ένα ξεχωριστό κινούμενο, ευλύγιστο κουτί ή έλλειψη. Σε αυτό το επίπεδο, η αναγνώριση των αλληλεπιδράσεων του ανθρώπου, περιορίζεται στο γενικό επίπεδο κατανόησης των προτύπων κίνησης αυτών των κουτιών ή ελλείψεων. Οι εφαρμογές παρακολούθησης χρησιμοποιούν πολλές φορές αυτό, το γενικό, επίπεδο. Ένα τέτοιο σύστημα, είναι ένα σύστημα αναγνώρισης των αλληλεπιδράσεων (συνάντηση, αποχαιρετισμός κτλ) δύο πεζών σε ένα δρόμο. Το σύστημα αναγνωρίζει κάθε έναν από τους πεζούς σαν ένα ευλύγιστο κουτί και κατόπιν κατατάσσει την κίνηση των δύο κουτιών σε κάποιο από τα υπάρχοντα πρότυπα. Ένα άλλο σύστημα που κάνει χρήση αυτού του γενικού επιπέδου λεπτομερειών, είναι ένα σύστημα το οποίο αναλύει μια ακολουθία εικόνων ανθρώπων, που παίζουν αμερικάνικο ποδόσφαιρο. Κάθε παίκτης αναπαριστάται με ένα ευλύγιστο ορθογώνιο κουτί. Η αναγνώριση επιτυγχάνεται μέσω της μελέτης της αλληλεπίδρασης των κουτιών, έχοντας υπ' όψιν τους κανόνες του αμερικάνικου ποδοσφαίρου, που αφορούν τις επιτρεπτές αλληλεπιδράσεις μεταξύ των παικτών.

Στο ενδιάμεσο επίπεδο λεπτομερειών, τα άτομα αναπαρίστανται με διακριτά τα βασικά μέρη του σώματός τους, όπως το κεφάλι, τον κορμό, τα χέρια και τα πόδια. Πολλές μέθοδοι έχουν προταθεί για την τμηματοποίηση των βασικότερων μελών του ανθρωπίνου σώματος. Ένα σύστημα ενδιάμεσου επιπέδου λεπτομερειών είναι αυτό της αναγνώρισης πολλών ατόμων σε κάποιο χώρο. Εφαρμόζοντας διαδοχικές αφαιρέσεις του παρασκηνίου, επιτυγχάνεται η τμηματοποίηση των περιοχών του

προσκήνιου στις οποίες βρίσκεται μια ομάδα ατόμων. Κατόπιν προβάλλεται μια δυαδική εικόνα του προσκήνιου προκειμένου να εντοπιστούν τα κεντροειδή των κεφαλιών του κάθε ατόμου της ομάδας.

Τέλος στο λεπτομερές επίπεδο, το οποίο άλλωστε είναι αυτό που θα μας απασχολήσει και στη συνέχεια, έχουν γίνει αρκετές μελέτες που αφορούν την αναγνώριση ανθρωπίνων δραστηριοτήτων, οι οποίες προέρχονται από την κίνηση συγκεκριμένων μόνο μελών. Αυτές οι μελέτες στοχεύουν ουσιαστικά στην ανάπτυξη συστημάτων αλληλεπίδρασης ανθρώπου – υπολογιστή, τα οποία θα βασίζονται στη χρήση χειρονομιών. Η αναγνώριση χειρονομιών για διεπαφές ανθρώπου-υπολογιστή είναι αντικείμενο εκτεταμένης έρευνας. Οι χειρονομίες, όπως επίσης και οι κινήσεις του βραχίονα θεωρείται πως είναι δυνατόν να χρησιμοποιηθούν για την εισαγωγή οπτικών εντολών στον έλεγχο του υπολογιστή. Σε αυτού του είδους τις εφαρμογές το υπόλοιπο σώμα δεν μας ενδιαφέρει. Αντίθετα, πολύ σημαντική είναι η υψηλή χωρική ανάλυση της εικόνας του βραχίονα και του χεριού, που χρησιμοποιείται σαν είσοδος. Οι χειρονομίες του χεριού (παλάμης) χρησιμοποιούνται κυρίως για να αναπαραστήσουν αριθμητικά ψηφία ή γράμματα, ενώ οι κινήσεις του βραχίονα για να υποδείξουν λειτουργίες όπως η εστίαση ή η μετακίνηση του δρομέα. Η οπτική αναγνώριση των εντολών μέσω χειρονομιών είναι στενά συνδεδεμένη με την γενικότερη αναγνώριση των ανθρωπίνων ενεργειών, καθώς και στις δύο περιπτώσεις ο υπολογιστής χρειάζεται να κατατάξει και να «μεταφράσει» τις ανθρώπινες κινήσεις. Πολλά από τα συστήματα τα οποία έχουν υλοποιηθεί, χρησιμοποιούν στατιστικές μεθόδους για να συγκρίνουν τις ακολουθίες εισόδου με τα αποθηκευμένα μοντέλα χειρονομιών. Άλλα πάλι, χρησιμοποιούν τόσο στατιστικές μεθόδους για τις χειρονομίες όσο και δυναμικές για την αναγνώριση της κίνησης του βραχίονα.

Η αναγνώριση της αλληλεπίδρασης ανάμεσα στον άνθρωπο και τα αντικείμενα είναι ένα εξίσου σημαντικό κεφάλαιο της οπτικής παρακολούθησης, στο οποίο χρησιμοποιούνται πληροφορίες, οι οποίες αφορούν το περιβάλλον ή τα αντικείμενα ώστε να γίνει κατανοητό το « τι συμβαίνει στην σκηνή». Υπάρχουν συστήματα τα οποία συνδυάζουν πληροφορίες που αφορούν το περιβάλλον με αλγόριθμους, που βασίζονται στο οπτικό σήμα. Για παράδειγμα, για ένα στούντιο τηλεόρασης, όπου λαμβάνει χώρα μια εκπομπή μαγειρικής, έχει αναπτυχθεί ένα σύστημα με κάμερες, οι οποίες αυτόματα εστιάζουν και επιλέγουν την σημαντικότερη σκηνή που θα προβληθεί. Σε αυτό το σύστημα χρησιμοποιούνται τόσο οπτικές όσο και πληροφορίες σχετικές με το περιβάλλον, ώστε να επιτευχθεί η αναγνώριση της αλληλεπίδρασης του ανθρώπου με τα γύρω αντικείμενα.

Συνήθως η αλληλεπίδραση ανθρώπου και αντικειμένων απαιτεί ένα μόνο φορέα σε αντίθεση με την αλληλεπίδραση μεταξύ ανθρώπων, η οποία, προφανώς, απαιτεί τουλάχιστον δύο ανεξάρτητους φορείς, ώστε να πραγματοποιηθεί. Για τον λόγο αυτό, κατά την αναγνώριση της αλληλεπίδρασης δύο ανθρώπων, μελετώνται οι κινήσεις του κάθε ξεχωριστού φορέα σε σχέση με τις κινήσεις του άλλου φορέα. Για παράδειγμα, ας θεωρηθεί ένα σύστημα μοντελοποίησης και αναγνώρισης ανθρωπίνων αλληλεπιδράσεων σε μια πλατεία. Το σύστημα ταξινομεί τις αλληλεπιδράσεις μεταξύ δύο ανθρώπων σε κλάσεις όπως: ο ένας άνθρωπος να ακολουθεί τον άλλον, ο ένας άνθρωπος να αλλάζει την διαδρομή που ακολουθεί προκειμένου να συναντήσει τον άλλον, η προσέγγιση των δύο ανθρώπων κ.α.. Η ταξινόμηση γίνεται με σύγκριση κάθε ακολουθίας εισόδου με στατιστικά μοντέλα κινήσεων, τα οποία υπάρχουν ήδη αποθηκευμένα στην μνήμη του υπολογιστή. Άλλα συστήματα τα οποία έχουν αναπτυχθεί εστιάζουν είτε στην λεπτομερή αναγνώριση μεμονωμένων ατόμων σε μια ακολουθία εικόνων, παραμελώντας τις μεταξύ τους αλληλεπιδράσεις, είτε στη

λεπτομερή αναγνώριση των αλληλεπιδράσεων μεταξύ δύο ή περισσότερων ατόμων χρησιμοποιώντας πολλές φορές εκτός από την οπτική και ακουστική πληροφορία.

3.3 Μέθοδοι Αναγνώρισης Ανθρωπίνων Ενεργειών

Η αναγνώριση των ανθρώπινων ενεργειών επιτυγχάνεται με την ταξινόμηση των δεδομένων που προέρχονται από κάποιο video. Υπάρχουν δύο τρόποι για να επιτευχθεί η ταξινόμηση: της απευθείας αναγνώρισης και της αναγνώρισης μέσω της ανακατασκευής. Στην περίπτωση της απευθείας αναγνώρισης οι ανθρώπινες ενέργειες αναγνωρίζονται απευθείας από τα δεδομένα της εικόνας, χωρίς την ανακατασκευή στιγμιότυπων του ανθρώπινου σώματος. Για παράδειγμα, ένα σύστημα αναγνώρισης της συμπεριφοράς των πεζών (περπάτημα, τρέξιμο, συνάντηση δύο ανθρώπων). Το σύστημα αυτό δεν χρησιμοποιεί κάποιο μοντέλο για το ανθρώπινο σώμα αλλά την περιοδικότητα της κίνησης των ανθρώπων. Η περιοδικότητα αυτή είναι αποτέλεσμα της αιώρησης των χεριών και των ποδιών κατά το βάδισμα. Κάθε ακολουθία εικόνων θεωρείται πως είναι ένας χώρος με τρεις διαστάσεις, δύο χωρικές και μία χρονική. Οι επαναλαμβανόμενες κινήσεις των πεζών στην εικόνα, γεννούν κάποιες καμπύλες στον τρισδιάστατο χώρο. Η αναγνώριση των ενεργειών των ανθρώπων επιτυγχάνεται με σύγκριση των καμπυλών που προέρχονται από μια ακολουθία εισόδου με τις καμπύλες αναφοράς που υπάρχουν για την κάθε ανθρώπινη δραστηριότητα.. Από την άλλη, στην περίπτωση της αναγνώρισης μέσω της ανακατασκευής, πρώτα γίνεται η ανακατασκευή του αντικειμένου με βάση την εικόνα και μετά γίνεται η αναγνώριση των ανθρώπινων ενεργειών. Παράδειγμα χρησιμοποίησης της παραπάνω μεθόδου, είναι ένα σύστημα αναγνώρισης της σκηνης «ψηλά τα χέρια» κατά την διάρκεια μιας ληστείας. Η αναγνώριση επιτυγχάνεται αφού πρώτα γίνει η τμηματοποίηση του σώματος σε επιμέρους μέλη και κατόπιν εξεταστεί η σχετική θέση των χεριών.

Υπάρχουν όμως και κάποια υβριδικά συστήματα, τα οποία δεν υιοθετούν αυστηρά την μια ή την άλλη μέθοδο αλλά δανείζονται χαρακτηριστικά και των δύο. Ένας άλλος τρόπος διαχωρισμού των μεθόδων αναγνώρισης είναι σε στατικές και δυναμικές. Η αναγνώριση της ανθρώπινης κίνησης προϋποθέτει την ανάλυση μιας σειράς εικόνων χρονικά διασυνδεδεμένων : της ακολουθίας του video. Η ακολουθία του video μπορεί να αναλυθεί, είτε χρησιμοποιώντας κάποια στατική αναπαράσταση κάθε frame ξεχωριστά, είτε χρησιμοποιώντας κάποια δυναμική αναπαράσταση ολόκληρης της ακολουθίας. Η προσέγγιση η οποία χρησιμοποιεί στατική αναπαράσταση αναλύει πρώτα τα ξεχωριστά frame και κατόπιν συνδυάζει τα αποτελέσματα στην ακολουθία ενώ η προσέγγιση η οποία χρησιμοποιεί δυναμική αναπαράσταση αντιμετωπίζει ολόκληρη την ακολουθία (ή κάποιο προκαθορισμένο μέρος αυτής) σαν την βασική μονάδα ανάλυσης. Αρχικά, ένα σύστημα το οποίο βασίστηκε στην στατική αναπαράσταση, χρησιμοποιούσε «ραβδόμορφες φιγούρες» σε κάθε frame για να αναλύσει τα διαφορετικά στιγμιότυπα ενός ατόμου. Με βάση ένα συγκεκριμένο στιγμιότυπο κάθε φορά, γινόταν η εξαγωγή των συναισθημάτων και των ενεργειών του ατόμου. Η «ραβδόμορφη φιγούρα» κατασκευαζόταν από τον χρήστη, με σταδιακή τοποθέτηση των μελών του σώματος του ατόμου, και γινόταν ξεχωριστή ανάλυση για κάθε frame χωρίς να δύνεται προσοχή σε τυχόν διασυνδέσεις μεταξύ των frame της ακολουθίας. Οι περισσότερες στατικές προσεγγίσεις υιοθέτησαν την μέθοδο του ταιριάσματος του περιγράμματος για να πετύχουν αναγνώριση.

Οι περισσότερες μελέτες, που χρησιμοποιούν δυναμική αναπαράσταση εφαρμόζουν τις μεθόδους της «Δυναμικής Περιτύλιξης του Χρόνου» (Dynamic Time

Wrapping – DTW) ή των «Κρυφών Μαρκοβιανών Μοντέλων» (Hidden Markov Models – HMM). Η DTW είναι μια μέθοδος σύγκρισης ακολουθιών η οποία χρησιμοποιείται σε πολλές εφαρμογές, όπως η σύγκριση κομματιών DNA στη μικροβιολογία, η σύγκριση συμβολοσειρών στην μετάδοση σήματος, και στην ανάλυση των κελαηδισμάτων πουλιών όπως και στην αναγνώριση φωνής. Η DTW διαχειρίζεται τις διαφορές ανάμεσα στις ακολουθίες με χρήση μεθόδων εισαγωγής-διαγραφής, συμπίεσης – διαστολής και αφαίρεσης των υποακολουθιών. Με τον καθορισμό μιας ποσότητας η οποία θα καθορίζει το κατά πόσο διαφέρουν οι ακολουθίες πριν και μετά την εκτέλεση των παραπάνω λειτουργιών, γίνεται και η ταξινόμηση των ακολουθιών. Η DTW μπορεί επίσης να εφαρμοστεί και σε ακολουθίες εικόνων. Παρόλα αυτά, παρουσιάζει το μειονέκτημα ότι δεν λαμβάνει υπ’ όψιν της, τυχόν διασυνδέσεις οι οποίες υπάρχουν ανάμεσα σε γειτονικές – χρονικά-ακολουθίες. Σε πραγματικές καταστάσεις μια ακολουθία παρουσιάζει, συνήθως, μεγαλύτερη συσχέτιση με κοντινότερες ακολουθίες απ’ ότι με τις πιο απομακρυσμένες (χρονικά).

Τα HMM λαμβάνουν υπ’ όψιν τους αυτή τη χρονική συσχέτιση των ακολουθιών, με την δημιουργία μιας αλυσίδας Markov. Τα HMM υποθέτουν πως η παρατηρούμενη ακολουθία (ή ακολουθία εξόδου), είναι το αποτέλεσμα μιας στοχαστικής, «κρυφής» διαδικασίας η οποία αποτελείται από έναν προκαθορισμένο αριθμό «κρυφών» καταστάσεων. Ένα HMM αποτελείται από ένα πεπερασμένο σύνολο «κρυφών» καταστάσεων, ένα σύνολο καταστάσεων εξόδου, πιθανότητες μετάβασης μεταξύ των «κρυφών» καταστάσεων, πιθανότητες μετάβασης από κάποια κρυφή κατάσταση σε κάποια συγκεκριμένη έξοδο και πιθανότητες για την αρχική κατάσταση. Η επιτυχία των HMM σε εφαρμογές αναγνώρισης φωνής, οδήγησε στην χρησιμοποίησή τους και σε εφαρμογές αναγνώρισης κινήσεων σε εικόνες. Τα δεδομένα φωνής αναπαρίστανται από καλά ορισμένες δομικές μονάδες της φυσικής γλώσσας (τα φωνήματα). Αντίθετα, η όραση υπολογιστών δεν προσφέρει κάποια δομική μονάδα για τις εικόνες που να έχει γενική εφαρμογή. Δεν είναι δυνατή, λοιπόν, η μετατροπή των εικόνων σε σύμβολα με κάποιο γενικό τρόπο. Για το λόγο αυτό, για να υπάρξει αποτελεσματική αναγνώριση των πολύπλοκων ενεργειών και αλληλεπιδράσεων, δημιουργούνται πολύπλοκες δομές που αποτελούνται από διασυνδεδεμένα μονοδιάστατα HMM. Τα HMM είναι πιο δημοφιλή για την δυναμική αναπαράσταση ενεργειών από την DTW λόγω της ικανότητας που τους παρέχει η στοχαστική τους φύση, να συμπεριλάβουν και να διαχειριστούν την αβεβαιότητα η οποία υπάρχει σε όλες τις κινήσεις, ενέργειες και χειρονομίες μίας ακολουθίας εικόνων.

Ένας σημαντικός περιορισμός στην χρήση των HMM είναι το γεγονός ότι δεν μπορούν να διαχειριστούν αποδοτικά τρεις ή περισσότερες ξεχωριστές διαδικασίες. Για να αντιμετωπιστεί αυτό το πρόβλημα, αναπτύχθηκαν τα «Δυναμικά Δίκτυα του Bayes» (Dynamic Bayesian Networks – DBNs) σαν μια γενίκευση των HMM. Τα DBNs είναι κατευθυντικά γραφικά μοντέλα μιας στοχαστικής διαδικασίας και γενικεύουν τα HMM αναπαριστώντας τόσο τις κρυφές καταστάσεις όσο και τις καταστάσεις εξόδου με μεταβλητές κατάστασης, οι οποίες μπορεί να έχουν πολύπλοκες αλληλεξαρτήσεις. Οι αλληλεξαρτήσεις μεταξύ των μεταβλητών κατάστασης μπορούν να αναπαρασταθούν αποτελεσματικά με τη δομή των κατευθυντικών γραφικών μοντέλων. Ως παράδειγμα, αναφέρουμε ένα σύστημα αναγνώρισης αλληλεπιδράσεων μεταξύ δύο ανθρώπων με χρήση ενός Bayesian δικτύου με ιεραρχική μορφή (BN). Στο σύστημα αυτό τα ταυτόχρονα στιγμιότυπα των μελών των σωμάτων των ατόμων αναγνωρίζονται στα χαμηλά επίπεδα του BN ενώ το ολικό στιγμιότυπο του σώματος στα υψηλότερα του BN. Τέλος, ένα δυναμικό

Βαesian δίκτυο επεξεργάζεται την εξέλιξη των στιγμιοτύπων των διάφορων μελών του σώματος.

ΚΕΦΑΛΑΙΟ 4

ΑΝΘΡΩΠΙΝΗ ΚΙΝΗΣΗ ΚΑΙ ΧΕΙΡΟΝΟΜΙΕΣ

ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο αναλύεται η φύση των ανθρωπίνων κινήσεων και χειρονομιών, πράγμα απαραίτητο για την σωστή σχεδίαση ενός συστήματος αναγνώρισης χειρονομιών. Δίνεται ένας «ορισμός» του όρου «ανθρώπινη χειρονομία» και επιχειρείται ο διαχωρισμός τους και η ταξινόμησή της ανάλογα με τα ιδιαίτερα χαρακτηριστικά που εμφανίζουν ορισμένες ομάδες χειρονομιών. Τέλος, παρατίθενται κάποια πρώτα στοιχεία που αφορούν τους τρόπους αλλά και τα προβλήματα που παρουσιάζονται κατά την αναγνώριση των χειρονομιών.

4.1 Αναγνώριση της ανθρώπινης κίνησης

Η αναγνώριση και η ερμηνεία της κίνησης του ανθρώπινου σώματος είναι ένα δύσκολο αλλά συναρπαστικό πρόβλημα. Υπάρχουν πολλές μορφές κίνησης του ανθρώπινου σώματος, όπως κινήσεις που συνδέονται με τον τρόπο που περπατάμε, επικοινωνούμε, και εκτελούμε κάποια εργασία. Μέσα στις κινήσεις αυτές του ανθρώπινου σώματος, υπάρχουν κρυμμένες πληροφορίες για την πρόθεση, τη διάθεση, τις ιδέες και ακόμη και την προσωπικότητα των ανθρώπων. Για παράδειγμα, έχει αποδειχθεί, ότι είναι δυνατό να προβλεφθεί η εγκληματική δραστηριότητα κάποιου ανθρώπου, με την παρατήρηση του ανθρώπου αυτού, κατά την επικοινωνία του με άλλους ανθρώπους, μέσω της γλώσσας του σώματος του. Είναι επίσης δυνατό, να αναγνωριστεί ένα άτομο, ή το γένος του, από τον βηματισμό ή τη γενική στάση σώματός του, όταν περπατάει .

Οι κινήσεις του ανθρώπινου σώματος μπορεί να ταξινομηθούν γενικά σε βηματισμό ή στάση, δράση, χειρονομία και σε πιο συγκεκριμένες κινήσεις, όπως είναι κινήσεις για την παραγωγή κάποιας νοηματικής γλώσσας. Ο βηματισμός ή η στάση είναι συνήθως μια ασυνείδητη μορφή κίνησης του σώματος, η οποία μπορεί να παρατηρηθεί όταν ένα άτομο περπατάει. Οι ενέργειες είναι συνήθως κινήσεις του σώματος με τις οποίες το άτομο επιδρά συνειδητά πάνω σε κάποιο αντικείμενο. Η χειρονομία είναι μια υποσυνείδητη μορφή επικοινωνίας, η οποία συμπληρώνει τη δυνατότητα για επικοινωνία που έχει ένα άτομο. Η νοηματική γλώσσα είναι μια συνειδητή μορφή επικοινωνιακής γλώσσας μεταξύ των ανθρώπων. Όλες αυτές οι μορφές κίνησης του σώματος μπορούν να ερμηνευθούν ως εκδηλώσεις της ανθρώπινης συμπεριφοράς. Ως εκ τούτου, μπορούμε να δούμε ότι η συμπεριφορά μπορεί να είναι συνειδητή, υποσυνείδητη, επικοινωνιακή ή ενεργός. Με την αυξανόμενη χρήση των καμερών ασφαλείας , υπάρχει ανάγκη ανάπτυξης αυτόματων συστημάτων τα οποία θα βοηθηθούν τους εργαζόμενους ασφαλείας. Άλλες εφαρμογές περιλαμβάνουν την αυτόματη αναγνώριση νοηματικής γλώσσας, τις διαεπαφές άνθρωπου-υπολογιστή και τα εργαλεία επίβλεψης ή εκπαίδευσης.

Το πρόβλημα της ερμηνείας των ανθρώπινων κινήσεων με χρήση της όρασης υπολογιστών είναι σύνθετο αλλά και πολύ ενδιαφέρον. Η πολυπλοκότητα του προβλήματος έγκειται κυρίως στην τετρα-διάστατη φύση της ανθρώπινης κίνησης

(δηλαδή χωρική και χρονική). Επίσης, ένα ακόμα πρόβλημα είναι ότι η ίδια χειρονομία μπορεί να διαφοροποιείται σημαντικά από πρόσωπο σε πρόσωπο, από πολιτισμό σε πολιτισμό και μερικές φορές ακόμη και όταν πραγματοποιείται πολλές φορές από το ίδιο άτομο. Τρεις από τις προαναφερθείσες, τέσσερις, διαστάσεις βρίσκονται στο χώρο. Η τέταρτη, είναι ο χρόνος και συσχετίζεται επομένως με τη δυναμική της κάθε κίνησης. Οι πληροφορίες που αφορούν την χρονική εξέλιξη μιας χειρονομίας είναι ιδιαίτερες σημαντικές, δεδομένου ότι δείχνουν που αρχίζει και που τελειώνει κάθε χειρονομία. Χωρίς αυτές τις πληροφορίες, θα ήταν δύσκολο να διακριθεί η μία χειρονομία από την άλλη, δεδομένου ότι πολλές χειρονομίες περιλαμβάνουν σχεδόν όμοια κομμάτια ακολουθιών στις τροχιές τους. Παραδείγματος χάριν, οι χειρονομίες : «δείχνοντας προς τα δεξιά» και «χαιρετώντας με το δεξί χέρι», αρχίζουν με την ίδια κίνηση : την ανύψωση δηλαδή του χεριού πάνω από το ύψος της μέσης και κατόπιν την κίνηση του χεριού και του μπράτσου πίσω στην αρχικές θέσεις τους. Τα ιδιαίτερα χαρακτηριστικά γνωρίσματα αυτών των δύο χειρονομιών, που τις διαφοροποιούν, μπορούν να εντοπισθούν μόνο στη μέση της τροχιάς των χειρονομιών αυτών. Ως εκ τούτου μια μέθοδος τμηματοποίησης των χειρονομιών σύμφωνα με το που αρχίζουν και που τελειώνουν και ο διαχωρισμός τους στις επιμέρους φάσεις τις κινήσεώς τους, μπορεί να βοηθήσει σημαντικά τη διαδικασία αναγνώρισης.

Επεκτείνοντας το ίδιο παράδειγμα, εάν μια ομάδα φίλων βρίσκεται στον ίδιο χώρο και παρατηρήσουμε κάποιον από την ομάδα να εκτελεί μια χειρονομία χαιρετισμού, είναι απίθανο, οι υπόλοιποι, να εκτελέσουν πάλι την ίδια χειρονομία, χωρίς προηγουμένως να εκτελέσουν άλλες χειρονομίες. Για παράδειγμα είναι πιθανότερο το άτομο που έκανε την χειρονομία χαιρετισμού, να ανταλλάξει χειραψία με κάποιον άλλον. Καταλαβαίνουμε λοιπόν ότι το περιεχόμενο είναι επίσης πολύ σημαντικός παράγοντας στην αναγνώριση της ανθρώπινης κίνησης. Το περιεχόμενο μπορεί να καθορίζεται σε σχέση με τις προηγούμενες ή και τις επόμενες χειρονομίες, αλλά και από την αλληλεπίδραση του ατόμου με αντικείμενα ή άλλους ανθρώπους που βρίσκονται στο ίδιο περιβάλλον. Για να εξηγήσουμε καλύτερα τον παράγοντα «περιεχόμενο», μπορούμε να αναφερθούμε, εκ νέου, στο προηγούμενο παράδειγμά μας, της χειρονομίας υπόδειξης και της χειρονομίας χαιρετισμού και να ισχυριστούμε ότι είναι αδύνατο να αναγνωριστούν και οι δύο χειρονομίες προτού αυτές ολοκληρωθούν. Αυτό συμβαίνει επειδή κατά την διάρκεια της εξέλιξης της χειρονομίας παρεμβάλλουμε ενστικτωδώς μικρές κινήσεις. Η χειρονομία της υπόδειξης για παράδειγμα, μπορεί να περιλαμβάνει την προετοιμασία του χεριού για την θέση υπόδειξης, ενώ ακόμη το χέρι κινείται. Αντιστοίχως, για την χειρονομία του χαιρετισμού, υπάρχει η κίνηση του ανοίγματος της παλάμης, η οποία προετοιμάζεται για το χαιρετισμό, ενώ ακόμη το χέρι ανεβαίνει.

Ένα ακόμη χρήσιμο οπτικό στοιχείο θα μπορούσε να είναι το γύρισμα του κεφαλιού προς την κατεύθυνση που δείχνει η χειρονομία υπόδειξης. Η αναγνώριση χειρονομιών με βάση το περιεχόμενό τους έχει μελετηθεί από τους Sherrah και Gong, που χρησιμοποίησαν την θέση του κεφαλιού για να θέσουν περιορισμούς όσον αφορά την σημασία της χειρονομίας, σε ένα σύστημα αναγνώρισης χειρονομίας για τηλεδιάσκεψη. Τέτοιες λεπτομέρειες στις κινήσεις του ανθρώπινου σώματος μπορούν να εξαχθούν μόνο με τον εντοπισμό των συνδυασμών των κινήσεων των άκρων. Συχνά, οι μικρές διαφορές ανάμεσα στις χειρονομίες μπορούν να εντοπιστούν μόνο με βάση αυτούς τους συνδυασμούς κινήσεων των άκρων. Προφανώς, εάν εντοπιστούν όλες αυτές οι μικρές κινήσεις που λαμβάνουν χώρα κατά την εξέλιξη της χειρονομίας, είναι πολύ πιθανόν να έχουμε μια αποτελεσματικότερη αναγνώριση.

Το ιδανικό θα ήταν η δημιουργία ενός συστήματος που θα αναγνωρίζει τις χειρονομίες που γίνονται σε πραγματικό χρόνο αλλά κάτι τέτοιο περιορίζει σημαντικά το χρόνο υπολογισμού που έχει στην διάθεσή του ο υπολογιστής για να διεκπεραιώσει την διαδικασία αναγνώρισης. Επομένως, είναι προτιμότερο να βρεθεί η χρυσή τομή ανάμεσα στα αντιμαχόμενα χαρακτηριστικά της καλής ακρίβειας, του μικρού χρόνου επεξεργασίας καθώς επίσης και της δυνατότητας του συστήματος να προσαρμοστεί στις αλλαγές στο θέμα ή το περιεχόμενο. Η μείωση του χρόνου υπολογισμού επιτυγχάνεται συχνά με κάποια μορφή εξαγωγής ή αναπαράστασης χαρακτηριστικών γνωρισμάτων των χειρονομιών, η οποία έχει ως στόχο την συμπύκνωση των πληροφοριών των κινήσεων και την απαλλαγή από δεδομένα τα οποία δεν χρησιμεύουν στην διαδικασία αναγνώρισης.

Ένα ακόμα ζήτημα που έχει να κάνει με την αναγνώριση χειρονομίας είναι η επιλογή των τεχνικών βιντεοσκόπησης που θα χρησιμοποιηθούν για την καταγραφή της ανθρώπινη κίνησης. Η κίνηση λαμβάνει χώρα, ως γνωστόν, σε τρεις διαστάσεις. Το πρόβλημα είναι ότι οι πιο κοινές τεχνικές χρησιμοποιούν δισδιάστατες ακολουθίες εικόνων. Δημιουργούνται έτσι προβλήματα που έχουν να κάνουν με το κλείσιμο και την αντίληψη του βάθους της εικόνας. Η έρευνα αυτή την στιγμή κινείται προς την κατεύθυνση της διαμόρφωσης ενός τρισδιάστατου μοντέλου του ανθρώπινου σκελετού, αλλά κάτι τέτοιο αφενός θα απαιτούσε περισσότερες από μια κάμερες για την υλοποίησή του και αφετέρου δεν είναι απαραίτητως πιο χρήσιμο όταν πρόκειται να χρησιμοποιηθεί σε εφαρμογές που αφορούν τον πραγματικό κόσμο. Παραδείγματος χάριν, η χρήση δύο καμερών αντί για μία για τηλεοπτική επιτήρηση, θα σήμαινε αυτόματα διπλασιασμό του κόστους. Από τα παραπάνω, συμπεραίνουμε πως το πρόβλημα της ευφυούς ερμηνείας της ανθρώπινης συμπεριφοράς μπορεί να χωριστεί σε τρία στάδια : την εξαγωγή και την αναπαράσταση των χαρακτηριστικών των κινήσεων, τον εντοπισμό της τροχιάς και το τελικό στάδιο της αναγνώρισης.

4.2 Οι Χειρονομίες

Ο πρωταρχικός στόχος της έρευνας που διεξάγεται πάνω στην αναγνώριση χειρονομιών είναι η δημιουργία ενός συστήματος το οποίο θα μπορεί να αναγνωρίζει συγκεκριμένες ανθρώπινες χειρονομίες και να τις χρησιμοποιεί για να μεταφέρει πληροφορίες ή για να ελέγξει κάποια συσκευή. Για την κατανόηση του τι ακριβώς είναι μια χειρονομία παραθέτουμε στη συνέχεια τους ορισμούς που δίνουν οι βιολόγοι και οι κοινωνιολόγοι για τον όρο «χειρονομία». Πολύ σημαντικά ερωτήματα τα οποία πρέπει να απαντηθούν πριν ξεκινήσει η σχεδίαση ενός συστήματος αναγνώρισης χειρονομιών είναι το πώς κωδικοποιείται η πληροφορία σε μία χειρονομία, το πώς χρησιμοποιούν οι άνθρωποι τις χειρονομίες για να επικοινωνούν μεταξύ τους και το πώς ορίζεται και χρησιμοποιείται από τους μηχανικούς ο όρος «χειρονομία».

4.2.1 Βιολογικός και Κοινωνιολογικός Ορισμός και Ταξινόμηση των Χειρονομιών

Από βιολογική και κοινωνιολογική σκοπιά δεν μπορεί να υπάρξει κάποιος αυστηρός ορισμός για τις χειρονομίες.. Για το λόγο αυτό οι ερευνητές έχουν την δυνατότητα να απεικονίζουν και να ταξινομούν τις χειρονομίες σύμφωνα με τα

υποκειμενικά τους κριτήρια. Η έρευνα πάνω στην αναγνώριση ομιλίας και γραφής έχει αποδώσει χρήσιμες μεθόδους σχεδίασης συστημάτων αναγνώρισης και κριτηρίων για την ταξινόμηση τέτοιων συστημάτων. Για το λόγο αυτό μελετώνται επίσης και συστήματα αναγνώρισης χειρονομιών, τα οποία χρησιμοποιούνται για τον έλεγχο συσκευών αποθήκευσης και απεικόνισης. Οι συσκευές αυτές μπορεί να βρίσκονται είτε στον ίδιο χώρο με τον χρήστη είτε σε κάποιο απομακρυσμένο περιβάλλον.

Οι άνθρωποι χρησιμοποιούν συχνά τις χειρονομίες για να επικοινωνούν. Η χρήση μιας χειρονομίας μπορεί να είναι ιδιαιτέρως απλή (μια χειρονομία η οποία δείχνει προς μια κατεύθυνση) αλλά μπορεί να είναι και ιδιαιτέρως σύνθετη (μετάδοση μέσω χειρονομιών των χαρακτηριστικών κάποιου χώρου / αντικειμένου). Οι ενδείξεις που υπάρχουν οδηγούν στο συμπέρασμα πως οι χειρονομίες δεν συμπληρώνουν απλά την φυσική γλώσσα αλλά αποτελούν αναπόσπαστο κομμάτι της διαδικασίας γέννησης της γλώσσας.

Οι βιολόγοι δίνουν τον παρακάτω γενικό ορισμό για τις χειρονομίες « η έννοια της χειρονομίας πρέπει να συμπεριλάβει όλα τα είδη κινήσεων τις οποίες συνδυάζει ένα άτομο προκειμένου να επικοινωνήσει» . Υπάρχουν χειρονομίες οι οποίες σχετίζονται με την ομιλία (gesticulation) και υπάρχουν και χειρονομίες οι οποίες εμφανίζονται ανεξάρτητα από την ομιλία και ονομάζονται αυτόνομες. Οι αυτόνομες χειρονομίες είναι δυνατόν να οργανωθούν με τέτοιο τρόπο ώστε να δημιουργήσουν την δική τους «γλώσσα» επικοινωνίας, όπως είναι η Αμερικανική Νοηματική Γλώσσα (American Sign Language – ASL). Οι αυτόνομες χειρονομίες μπορούν επίσης να χρησιμοποιηθούν για να δώσουν εντολές κίνησης. Παρακάτω εξετάζονται διάφοροι ορισμοί τους οποίους δίνουν οι κοινωνιολόγοι και οι βιολόγοι για τις χειρονομίες ώστε να βρεθεί εάν υπάρχουν κάποιες χειρονομίες οι οποίες είναι ιδανικές να χρησιμοποιηθούν για επικοινωνία και έλεγχο διαφόρων συσκευών.

4.2.2 Ταξινόμηση χειρονομιών

Μια πρώτη μέθοδος ταξινόμησης είναι ο διαχωρισμός των χειρονομιών σε τέσσερις κατηγορίες : ενέργειας – συμβολικές, προφανείς και μη, αυτόνομες σημειολογικές (ο όρος ‘σημειολογικές’ αναφέρεται σε μια γενικότερη φιλοσοφική θεωρία των σημάτων και των συστημάτων που πραγματεύεται την χρήση τους τόσο στην τεχνητή όσο και στη φυσική γλώσσα) – πολυσημειολογικές, φυγόκεντρες – κεντρομόλες.

Ο πρώτος διαχωρισμός είναι σε χειρονομίες ενέργειας και συμβολικές χειρονομίες. Οι χειρονομίες ενέργειας, όπως υποδηλώνει και το όνομά τους , δεν έχουν κάποια μεταφορική / συμβολική σημασία . Παραδείγματα τέτοιων χειρονομιών είναι το κόψιμο ξύλων ή το μέτρημα χρημάτων. Παράδειγμα συμβολικών χειρονομιών είναι το σήμα του «OK» ή η το νόημα για οτοστόπ. Υπάρχουν όμως και χειρονομίες ενέργειας στις οποίες είναι δυνατόν να αποδοθεί κάποιο συμβολικό νόημα.(σημειογέννεση) όπως για παράδειγμα, σε μια κατασκοπική νουβέλα , όπου το αντικείμενο που κρατάει ένας πράκτορας στο χέρι του μπορεί να έχει πολύ σημαντικό νόημα. Ο διαχωρισμός αυτός δίνει την δυνατότητα στους ερευνητές να χρησιμοποιήσουν για τον έλεγχο συσκευών χειρονομίες οι οποίες ουσιαστικά αναπαριστούν κάποια πραγματική κίνηση.

Ο διαχωρισμός σε προφανείς και μη χειρονομίες αναφέρεται στην ευκολία με την οποία είναι δυνατόν οι υπόλοιποι να ερμηνεύσουν κάποια χειρονομία.. Ο όρος «προφανής χειρονομία» έχει συσχετιστεί με την έννοια της οικουμενικότητας, σύμφωνα με την οποία ορισμένες χειρονομίες έχουν καθορισμένο, διαπολιτισμικό

νόημα. Στην πραγματικότητα το νόημα κάθε χειρονομίας εξαρτάται σε πολύ μεγάλο βαθμό από τον πολιτισμό μέσα στον οποίο αναπτύσσεται. Μέσα σε μια κοινωνία οι χειρονομίες έχουν καθορισμένα νοήματα, όμως δεν υπάρχει καμία γνωστή χειρονομία ή κίνηση του σώματος η οποία να έχει το ίδιο νόημα σε όλες τις κοινωνίες. Ακόμα και στην ASL μερικά νοήματα έχουν τόσο ξεκάθαρη σημασία που μπορεί να τα αναγνωρίσει και κάποιος ο οποίος δεν γνωρίζει την ASL. Αυτό σημαίνει πως οι χειρονομίες οι οποίες θα καθορίσουν την λειτουργία μιας συσκευής μπορούν να επιλεγθούν σχεδόν αυθαίρετα.

Ο διαχωρισμός σε φυγόκεντρες και κεντρομόλες έχει να κάνει με το κατά πόσο μια χειρονομία έχει συγκεκριμένη κατεύθυνση ή όχι. Οι φυγόκεντρες χειρονομίες κατευθύνονται συνήθως προς κάποιο αντικείμενο σε αντίθεση με τις κεντρομόλες. Οι ερευνητές επικεντρώνουν το ενδιαφέρον τους σε χειρονομίες οι οποίες έχουν ως στόχο τον έλεγχο κάποιου αντικειμένου ή την επικοινωνία με κάποιο συγκεκριμένο άτομο, από κάποια ομάδα ατόμων.

Οι χειρονομίες οι οποίες ανήκουν σε ένα αυτόνομο σημειολογικό σύστημα είναι αυτές οι οποίες χρησιμοποιούνται σε κάποια νοηματική γλώσσα όπως η ASL. Από την άλλη οι χειρονομίες οι οποίες δημιουργούνται ως επί μέρους τμήματα κάποιας πολυσημειολογικής δραστηριότητας, είναι χειρονομίες οι οποίες «συνοδεύουν» κάποια άλλη γλώσσα όπως για παράδειγμα τον προφορικό λόγο. Οι ερευνητές επικεντρώνουν το ενδιαφέρον τους σε χειρονομίες οι οποίες παράγονται έχοντας το δικό τους, ξεχωριστό σημειολογικό νόημα αν και υπάρχουν και κάποιες εξαιρέσεις.

4.2.3 Τυπολογία Χειρονομιών

Ένας άλλος τρόπος κατηγοριοποίησης των χειρονομιών είναι σε : αυθαίρετες, μιμητικές και δεικτικές.

Στις μιμητικές χειρονομίες, οι κινήσεις που πραγματοποιούνται αναπαριστούν την μορφή κάποιου αντικειμένου ή κάποιο χαρακτηριστικό του. Για παράδειγμα, η χειρονομία κατά την οποία το χέρι κινείται κατά μήκος του πηγουνιού, μπορεί να χρησιμοποιηθεί για να αναπαραστήσει μια κατσίκια, κάνοντας αναφορά στο γέني της. Η σημασία τέτοιων χειρονομιών κατά κανόνα είναι προφανής. Οι μιμητικές χειρονομίες είναι ιδιαίτερες χρήσιμες στις νοηματικές γλώσσες.

Οι δεικτικές χειρονομίες, χρησιμοποιούνται για να δείξουν ένα αντικείμενο και κάθε τέτοια χειρονομία έχει κάποιο προφανές νόημα, διαφορετικό κάθε φορά, ανάλογα με το χώρο μέσα στον οποίο πραγματοποιείται. Οι δεικτικές χειρονομίες χωρίζονται σε εξειδικευμένες, γενικές και λειτουργικές. Οι εξειδικευμένες χειρονομίες αναφέρονται σε κάποιο συγκεκριμένο αντικείμενο. Οι γενικές χειρονομίες αναφέρονται σε μια ομάδα αντικειμένων. Οι λειτουργικές χειρονομίες αναπαριστούν κάποιες προθέσεις, όπως για παράδειγμα το να δείχνει κάποιος μια καρέκλα, ζητώντας με τον τρόπο αυτό άδεια για να κάτσει. Οι δεικτικές χειρονομίες είναι, επίσης, χρήσιμες στις νοηματικές γλώσσες.

Οι αυθαίρετες χειρονομίες είναι αυτές, των οποίων η ερμηνεία πρέπει να δοθεί από κάποιον, αφού δεν έχουν κάποιο προφανές νόημα. Αν και δεν συναντώνται συχνά μέσα σε κάποιο κοινωνικό περιβάλλον, άπαξ και το νόημά τους διασαφηνιστεί, είναι δυνατόν να χρησιμοποιηθούν και να γίνουν κατανοητές, χωρίς να υπάρχει ανάγκη για περαιτέρω διευκρινήσεις. Ένα τέτοιο παράδειγμα είναι το «σετ» των χειρονομιών που χρησιμοποιούνται για τις εγχειρίσεις στον εγκέφαλο. Οι αυθαίρετες χειρονομίες είναι χρήσιμες γιατί είναι δυνατόν να αναπτυχθούν εξειδικευμένες χειρονομίες κάθε φορά, οι οποίες θα έχουν εφαρμογή στον έλεγχο της συσκευής που μας ενδιαφέρει. Στις χειρονομίες αυτές, οι οποίες αναπτύσσονται για κάποιο

συγκεκριμένο σκοπό, δίνεται από πριν κάποιο αυθαίρετο νόημα και το οποίο δεν υπάρχει ανάγκη να διασαφηνίζεται κάθε φορά που γίνεται χρήση της χειρονομίας.

4.3 Αναγνώριση Φωνής και Κειμένου: Θέματα παράλληλα με την Αναγνώριση Χειρονομιών

Τα συστήματα αναγνώρισης φωνής και κειμένου έχουν πολλά κοινά στοιχεία με τα συστήματα αναγνώρισης φωνής. Όλα τα παραπάνω συστήματα αναγνωρίζουν «κάτι», το οποίο κινείται διαγράφοντας μια τροχιά στο χώρο και στον χρόνο. Κατανοώντας τη φύση της αναγνώρισης κειμένου και φωνής, και τους τρόπους ταξινόμησής τους, είναι δυνατόν αν προκύψουν χρήσιμα συμπεράσματα και όσον αφορά την ανάπτυξη κάποιου συστήματος αναγνώρισης χειρονομιών.

Τα κλασικά συστήματα αναγνώρισης φωνής «ταιριάζουν» την μετασχηματισμένη, με κάποιο τρόπο, ομιλία με κάποια αποθηκευμένη αναπαραστάση. Τα περισσότερα συστήματα χρησιμοποιούν κάποια μορφή φασματικής ανάλυσης, όπως τα φασματικά πρότυπα ή τα Hidden Markov Models. Τα συστήματα αναγνώρισης φωνής ταξινομούνται ανάλογα με το αν διαθέτουν ή όχι τα παρακάτω χαρακτηριστικά:

- **Ανεξαρτησία από τον ομιλητή:** Το επιθυμητό είναι ένα σύστημα να μπορεί να αναγνωρίσει με ακρίβεια την ομιλία οποιουδήποτε ομιλητή, χωρίς να απαιτείται προηγουμένως η εκπαίδευση του συστήματος με την φωνή του εκάστοτε ομιλητή. Πάντως τα συστήματα τα οποία εξαρτώνται από τον ομιλητή, είναι πιο ακριβή, δεδομένου ότι δεν χρειάζεται να προσαρμόζονται σε μεγάλες διαφορές στον τρόπο εκφοράς μιας λέξης.
- **Συνεχές ή διακριτό:** Το χαρακτηριστικό αυτό αφορά το αν το σύστημα μπορεί να αναγνωρίσει συνεχείς προτάσεις ή εάν απαιτείται ο ομιλητής να χωρίζει τις λέξεις που εκφωνεί με μικρά κενά. Τα συστήματα αναγνώρισης απομονωμένων λέξεων πάντως, εμφανίζουν μεγαλύτερο βαθμό αναγνωρισιμότητας, κάτι το οποίο εν μέρει οφείλεται στο γεγονός πως το σύστημα γνωρίζει πότε τελειώνει κάθε λέξη.
- **Μέγεθος του λεξιλογίου:** Όταν όλα τα υπόλοιπα χαρακτηριστικά ενός συστήματος είναι δεδομένα, έχει παρατηρηθεί ότι όσο μικρότερο είναι το λεξιλόγιο που μπορεί να αναγνωρίσει το σύστημα, τόσο μεγαλύτερος είναι ο βαθμός αναγνωρισιμότητας που επιτυγχάνεται.
- **Βαθμός αναγνώρισης:** Τα περισσότερα εμπορικά προϊόντα ισχυρίζονται πως επιτυγχάνουν βαθμούς αναγνώρισης τουλάχιστον 95%. Αν και το ποσοστό αυτό φαίνεται ιδιαίτερος υψηλό, είναι μετρημένο σε συνθήκες εργαστηρίου. Χαρακτηριστικό είναι πως ο ίδιος ο άνθρωπος, παρουσιάζει ποσοστό αναγνώρισης περίπου 99.2%

Τα συστήματα τα οποία έχουν την δυνατότητα να αναγνωρίζουν λέξεις από μεγάλο λεξιλόγιο χρησιμοποιούν τα Κρυφά Μαρκοβιανά Μοντέλα (HMM). Τα HMM χρησιμοποιούνται επίσης και από αρκετά συστήματα αναγνώρισης χειρονομιών. Σε μερικά συστήματα αναγνώρισης φωνής, οι καταστάσεις του HMM αντιστοιχούν στις φωνητικές μονάδες. Ο πίνακας με τις πιθανότητες μετάβασης

καθορίζει ποια θα είναι η επόμενη κατάσταση. Ο όρος «κρυφό» αναφέρεται στο είδος του Μαρκοβιανού μοντέλου που χρησιμοποιείται. Σε αυτού του είδους τα μοντέλα οι παρατηρήσεις εξόδου αποτελούν μια στοχαστική διαδικασία που εξαρτάται από την κάθε κατάσταση. Ο πλήρης καθορισμός ενός Μαρκοβιανού μοντέλου απαιτεί τις παρακάτω πληροφορίες : την κατανομή πιθανότητας ανάμεσα στις καταστάσεις, την κατανομή πιθανότητας εξόδου κάθε κατάστασης, την κατανομή πιθανότητας της αρχικής κατάστασης. Κατά την διαδικασία αναγνώρισης φωνής, δημιουργείται ένα HMM για κάθε λέξη η οποία περιέχεται στο λεξικό. Κατόπιν, για κάθε ακολουθία φωνητικών μονάδων η οποία αποτελεί την είσοδο, υπολογίζεται η πιθανότητα ,αυτή, να έχει προέλθει από το κάθε ένα από τα HMM του λεξικού.

Από μια οπτική γωνία, η αναγνώριση κειμένου, μπορεί να θεωρηθεί πως εντάσσεται στο γενικότερο πλαίσιο της αναγνώρισης χειρονομιών. Οι on-line (ή δυναμικές) συσκευές αναγνώρισης κειμένου, αναγνωρίζουν το γραπτό κείμενο καθώς ο χρήστης γράφει. Οι συσκευές αυτές, έχουν το πλεονέκτημα ότι συλλαμβάνουν την δυναμική πληροφορία του γραψίματος, συμπεριλαμβανομένου του αριθμού των φορών που το στυλό ακουμπάει την οθόνη καθώς και την κατεύθυνση και την ταχύτητα που έχει την κάθε φορά. Τα συστήματα αυτά παρέχουν στον χρήστη την δυνατότητα να διορθώνει τυχόν λάθη που προκύπτουν κατά την διαδικασία αναγνώρισης , την στιγμή που προκύπτουν.

Τα περισσότερα συστήματα αντιλαμβάνονται το κείμενο σαν μια ακολουθία συντεταγμένων των σημείων. Η διαδικασία της αναγνώρισης είναι ιδιαίτερος δύσκολη, αφού υπάρχουν αρκετοί τρόποι για να γραφτεί ο ίδιος χαρακτήρας. Ένα άλλο πρόβλημα είναι το γεγονός πως οι χαρακτήρες αλληλεπικαλύπτονται, πρόβλημα το οποίο είναι παρόμοιο με αυτό που εμφανίζεται στην αναγνώριση φωνής (όπου είναι δυνατόν να αλληλεπικαλύπτονται οι διάφορες λέξεις). Επίσης, είναι δυνατόν διαφορετικοί χαρακτήρες να έχουν παρόμοια εμφάνιση. Για την αντιμετώπιση των παραπάνω προβλημάτων , τα συστήματα αναγνώρισης κειμένου, αρχικά προεπεξεργάζονται το κείμενο και κατόπιν αναγνωρίζουν την μορφή των χαρακτήρων. Το στάδιο της προεπεξεργασίας περιλαμβάνει το διαχωρισμό των χαρακτήρων που επικαλύπτονται και την εξαγωγή του θορύβου.

4.4 Αναγνώριση Χειρονομιών με HMM

Η χρήση των HMM για την αναγνώριση χειρονομιών έγινε, όπως προαναφέρθη, μετά την επιτυχημένη εφαρμογή τους στο πρόβλημα της αναγνώρισης φωνής. Οι ομοιότητες μεταξύ ομιλίας και χειρονομίας υπαινίσσονται πως οι τεχνικές που είναι αποτελεσματικές για το ένα πρόβλημα είναι πολύ πιθανόν να είναι αποτελεσματικές και για το άλλο.

Κατ' αρχάς, οι χειρονομίες, όπως και ο προφορικός λόγος, ποικίλλουν ανάλογα με την τοποθεσία, το χρόνο, και τους κοινωνικούς παράγοντες. Δεύτερον, οι κινήσεις του σώματος, όπως και οι διάφοροι ήχοι, κωδικοποιούν κάποιο συγκεκριμένο νόημα. Τρίτον, η διαδοχή των κινήσεων που κάνει κάποιος καθώς μιλάει, μπορεί να παρομοιαστεί με τους συντακτικούς κανόνες. Επομένως, οι γλωσσικές μέθοδοι μπορούν να χρησιμοποιηθούν στην αναγνώριση χειρονομίας.

Από τη άλλη πλευρά, η αναγνώριση χειρονομιών έχει τα δικά της ιδιαίτερα χαρακτηριστικά και προβλήματα. Για να αξιολογηθεί μια διεπαφή βασισμένη σε χειρονομίες, απαιτούνται μερικά συγκεκριμένα κριτήρια όπως είναι η το ποσοστό αναγνώρισης των νοηματικών χειρονομιών, εάν γίνεται χρήση ειδικών αισθητήρων,

εάν χρησιμοποιούνται αποδοτικοί αλγόριθμοι εκπαίδευσης, και εάν η εφαρμογή είναι ακριβής, αποδοτική και εάν η αναγνώριση γίνεται σε πραγματικό χρόνο .

Οι νοηματικές χειρονομίες μπορούν να είναι πολύ σύνθετες, περιέχοντας ταυτόχρονες κινήσεις διάφορων σημείων. Εντούτοις, αυτές οι σύνθετες χειρονομίες πρέπει να περιγραφούν στον υπολογιστή με τρόπο όσο το δυνατόν πιο απλό και σαφή. Αρχικά, κάθε εφαρμογή, έχει ένα στάδιο εκπαίδευσης, στο οποίο συλλέγονται παραδείγματα διαφορετικών χειρονομιών και χρησιμοποιούνται για την εκπαίδευση των μοντέλων. Τα εκπαιδευμένα μοντέλα αναπαριστούν πλέον όλες τις χειρονομίες που το σύστημα είναι σε θέση να αναγνωρίζει.

Επίσης, καθώς η χειρονομία είναι μια εκφραστική κίνηση, είναι φυσικό να μπορεί να περιγραφεί από ένα ακολουθιακό μοντέλο. Με βάση αυτές τις εκτιμήσεις, το HMM είναι κατάλληλο για την αναγνώριση χειρονομίας. Ένα πολυδιάστατο HMM είναι σε θέση να αναγνωρίσει και χειρονομίες πολλαπλών διαδρομών που αποτελούν την γενικότερη περίπτωση χειρονομιών. Απ' την άλλη, μια χειρονομία μοναδικής διαδρομής, μπορεί, συνήθως, να μεταφραστεί σε δισδιάστατη ή τρισδιάστατη χρονική ακολουθία στον Καρτεσιανό χώρο. Δηλαδή μια χειρονομία μοναδικής διαδρομής $g(x, Y, z, t)$ μπορούν να αναλυθεί σε $X(t)$, $Y(t)$, και $Z(t)$. Τέλος, η χρήση πολυδιάστατων HMM παρέχει τη δυνατότητα χρήσης πολλαπλών χαρακτηριστικών γνωρισμάτων για κάθε χειρονομία πράγμα το οποίο βοηθάει στην αύξηση των ποσοστών αναγνώρισης.

ΚΕΦΑΛΑΙΟ 5

ΤΑ ΚΡΥΦΑ ΜΑΡΚΟΒΙΑΝΑ ΜΟΝΤΕΛΑ

ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό παρέχεται το κατάλληλο μαθηματικό υπόβαθρο που αφορά τα HMM. Διασαφηνίζεται η έννοια των Κρυφών Μαρκοβιανών Μοντέλων, δίνεται η μαθηματική τους θεμελίωση καθώς και παρουσιάζονται οι αλγόριθμοι που χρησιμοποιούνται για την λύση των ζητημάτων που αφορούν την χρήση των HMM. Τέλος, περιγράφονται διάφοροι τύποι HMM.

5.1 Μοντελοποίηση των Φυσικών Διεργασιών

Οι φυσικές διεργασίες έχουν συνήθως ως αποτέλεσμα παρατηρήσιμες εξόδους οι οποίες ονομάζονται σήματα. Τα σήματα μπορεί να είναι είτε διακριτά (όπως χαρακτήρες ενός αλφάβητου, κβαντισμένα διανύσματα κτλ.) είτε συνεχή (όπως δείγματα φωνής, μετρήσεις θερμοκρασίας, μουσική κτλ.). Η πηγή των σημάτων μπορεί να είναι είτε στάσιμη (οι στατιστικές της ιδιότητες να μην μεταβάλλονται με τον χρόνο) είτε μη στάσιμη (τα χαρακτηριστικά του σήματος μεταβάλλονται με το χρόνο).

Ένα θεμελιώδες πρόβλημα είναι η ανάπτυξη μοντέλων για την περιγραφή τέτοιων φυσικών σημάτων. Υπάρχουν πολλοί λόγοι που καθιστούν χρήσιμη την ανάπτυξη τέτοιων μοντέλων. Πρώτον, η ύπαρξη ενός μοντέλου για το σήμα μπορεί να αποτελέσει τη βάση για την θεωρητική περιγραφή ενός συστήματος επεξεργασίας σήματος το οποίο θα επεξεργαστεί το φυσικό σήμα και θα παράγει την επιθυμητή έξοδο (πχ. εξαγωγή του θορύβου από ένα μεταδιδόμενο σήμα φωνής). Δεύτερον, μέσω της μελέτης του μοντέλου μπορούμε να εξάγουμε χρήσιμα συμπεράσματα για την φυσική πηγή του σήματος, την οποία πολλές φορές δεν μπορούμε να την έχουμε διαθέσιμη. Τέλος, ο σημαντικότερος λόγος μοντελοποίησης των σημάτων, είναι ότι συνήθως τα μοντέλα δίνουν πολύ ικανοποιητικά αποτελέσματα όταν χρησιμοποιούνται για την ανάπτυξη πρακτικών συστημάτων (πχ. συστήματα πρόβλεψης, συστήματα αναγνώρισης κτλ.)

Υπάρχουν πολλές δυνατότητες επιλογής για τον τύπο του μοντέλου που θα χρησιμοποιηθεί για να περιγράψει τις ιδιότητες ενός φυσικού σήματος. Μια πρώτη διάκριση είναι ο διαχωρισμός τους σε ντετερμινιστικά και στατιστικά. Τα ντετερμινιστικά μοντέλα χρησιμοποιούν κάποιες γνωστές ιδιότητες του σήματος και η εξαγωγή του μοντέλου γίνεται απευθείας (με εκτίμηση μόνο μερικών παραμέτρων). Τα στατιστικά μοντέλα μπορούν να μας δώσουν πληροφορίες που αφορούν μόνο τις στατιστικές ιδιότητες του σήματος. Στα στατιστικά μοντέλα περιλαμβάνονται οι γκαουσιανές διαδικασίες, οι διαδικασίες Poisson όπως και τα Κρυφά Μαρκοβιανά Μοντέλα (HMM). Η βασική θεώρησης στα στατιστικά μοντέλα είναι ότι το σήμα μπορεί να χαρακτηριστεί σαν μια παραμετρική τυχαία διαδικασία, της οποίας οι παράμετροι μπορούν να καθοριστούν με σαφές τρόπο.

Στη συνέχεια μελετάται διεξοδικά ένα από τα προαναφερθέντα στατιστικά μοντέλα, το οποίο και χρησιμοποιήθηκε για την ανάπτυξη του συστήματος αναγνώρισης χειρονομιών και το οποίο είναι τα HMM.

5.2 Τα HMM

Ένα Κρυφό Μαρκοβιανό Μοντέλο είναι η αναπαράσταση μιας Μαρκοβιανής Διαδικασίας η οποία δεν μπορεί να είναι παρατηρήσιμη. Η ιδιαιτερότητα των Κρυφών Μαρκοβιανών Μοντέλων (HMM) είναι ότι οι καταστάσεις του μοντέλου δεν αντιστοιχούν σε κάποιο φυσικό γεγονός όπως συμβαίνει στα απλά Μαρκοβιανά Μοντέλα. Τα HMM είναι λοιπόν μία διπλά στοχαστική διαδικασία περιλαμβάνοντας μια διαδικασία η οποία δεν είναι παρατηρήσιμη (είναι «κρυφή») και μία η οποία είναι, και παράγει την ακολουθία παρατηρήσιμων εξόδων.

Κάθε κατάσταση του μοντέλου χαρακτηρίζεται από δύο σετ πιθανοτήτων : την πιθανότητα μετάβασης και είτε μια διακριτή κατανομή πιθανότητας εξόδου είτε μια συνεχή συνάρτηση πυκνότητας πιθανότητας εξόδου οι οποίες, με δεδομένη την κατάσταση καθορίζουν την δεσμευμένη πιθανότητα εκπομπής κάποιου από τα σύμβολα εξόδου (που υπάρχουν σε ένα πεπερασμένο αλφάβητο) ή κάποιου συνεχούς τυχαίου διανύσματος.

Ένα HMM έχει τα ακόλουθα χαρακτηριστικά :

1) N , είναι ο αριθμός καταστάσεων του μοντέλου. Παρόλο που οι καταστάσεις είναι κρυφές, για πολλές πρακτικές εφαρμογές υπάρχει κάποια φυσική σημασία για τις καταστάσεις ή ομάδες καταστάσεων του μοντέλου. Οι καταστάσεις συνήθως συνδέονται με τέτοιο τρόπο ώστε κάθε κατάσταση να μπορεί να έχει ως επόμενη οποιαδήποτε από τις άλλες (εργοδικό μοντέλο). Ωστόσο υπάρχουν και άλλοι τύποι HMM (όσον αφορά την διασύνδεση των καταστάσεων) οι οποίοι είναι ιδιαίτερος χρήσιμοι για συγκεκριμένες εφαρμογές και οι οποίοι θα παρουσιαστούν στη συνέχεια. Οι καταστάσεις θα συμβολίζονται στο εξής ως $S = \{S_1, S_2, \dots, S_N\}$ και η κατάσταση την χρονική στιγμή t ως q_t .

2) M , είναι ο αριθμός των διακριτών συμβόλων παρατήρησης ανά κατάσταση (πχ. το μέγεθος του διακριτού αλφάβητου). Τα σύμβολα εξόδου αντιστοιχούν στη φυσική έξοδο του συστήματος το οποίο μοντελοποιείται. Τα σύμβολα παρατήρησης θα αναφέρονται στο εξής ως $V = \{v_1, v_2, \dots, v_M\}$.

3) Την κατανομή πιθανότητας μετάβασης των καταστάσεων $A = \{a_{ij}\}$ όπου

$$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i], \quad 1 \leq i, j \leq N.$$

Για την ειδική περίπτωση που κάθε κατάσταση μπορεί να οδηγήσει σε οποιαδήποτε από τις υπόλοιπες σε ένα μόνο βήμα ισχύει $a_{ij} > 0$ για όλα τα i, j . Για άλλους τύπους HMM θα έχουμε $a_{ij} = 0$ για παραπάνω από ένα ζευγάρι (i, j)

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

$$b_j(k) = P[v_k \text{ όταν } t \mid q_t = S_j], \quad \begin{matrix} 1 \leq j \leq N \\ 1 \leq k \leq M \end{matrix}$$

5) Την κατανομή της αρχικής κατάστασης $\pi = \{\pi_i\}$ όπου

$$\pi_i = P[q_1 = S_i], \quad 1 \leq j \leq N$$

Όταν δοθούν κατάλληλες τιμές στα N, M, A, B και π το HMM μπορεί να χρησιμοποιηθεί σαν γεννήτρια ώστε να μας δώσει μια ακολουθία εξόδων

$$O = O_1 O_2 \dots O_T$$

όπου κάθε έξοδος O_i είναι ένα από τα σύμβολα από το V και T είναι το μήκος της ακολουθίας.

Είναι φανερό από τα παραπάνω ότι για τον πλήρη προσδιορισμό ενός HMM πρέπει να καθοριστούν δύο παράμετροι του μοντέλου (N και M), να καθοριστούν τα σύμβολα εξόδου καθώς επίσης και τρία στατιστικά μεγέθη (A, B και π). Για λόγους συντομίας, όταν γίνεται αναφορά στο σετ παραμέτρων του μοντέλου, χρησιμοποιείται ο ακόλουθος συμβολισμός $\lambda = (A, B, \pi)$.

5.3 Τρία Βασικά Προβλήματα

Με δοσμένη τη μορφή του HMM που περιγράφηκε προηγουμένως, υπάρχουν τρία βασικά προβλήματα στα οποία πρέπει να δοθεί ικανοποιητική λύση προτού το μοντέλο μπορέσει να χρησιμοποιηθεί σε εφαρμογές του πραγματικού κόσμου. Τα προβλήματα αυτά είναι τα ακόλουθα :

A) Με δεδομένη την ακολουθία εξόδων $O = O_1 O_2 \dots O_T$ και το μοντέλο $\lambda = (A, B, \pi)$ πώς μπορεί να υπολογιστεί αποδοτικά η πιθανότητα $P(O | \lambda)$, δηλαδή η πιθανότητα η συγκεκριμένη ακολουθία να έχει «γεννηθεί» από το συγκεκριμένο μοντέλο.

B) Με δεδομένη την ακολουθία εξόδων $O = O_1 O_2 \dots O_T$ και το μοντέλο $\lambda = (A, B, \pi)$ πώς μπορεί να προσδιοριστεί μία ακολουθία καταστάσεων $Q = q_1 q_2 \dots q_T$ η οποία είναι η βέλτιστη κατά τρόπο που να έχει κάποια φυσική σημασία. (πχ. να εξηγήει την ακολουθία των εξόδων).

Γ) Με δεδομένη την ακολουθία εξόδων $O = O_1 O_2 \dots O_T$ πώς μπορούν να προσδιοριστούν οι βέλτιστες παράμετροι $\lambda = (A, B, \pi)$ του μοντέλου ώστε να μεγιστοποιείται η πιθανότητα $P(O | \lambda)$.

ΠΡΟΒΛΗΜΑ Α

Το πρόβλημα αυτό ονομάζεται πρόβλημα εκτίμησης. Πρέπει να εκτιμηθεί η πιθανότητα, δεδομένης της ακολουθίας εξόδων και του μοντέλου, η συγκεκριμένη ακολουθία να προήλθε από το συγκεκριμένο μοντέλο. Μπορεί επίσης να θεωρηθεί και ως πρόβλημα «ταιριάσματος», του πόσο πολύ «ταιριάζει» η συγκεκριμένη ακολουθία με το μοντέλο. Η τελευταία αυτή προσέγγιση είναι ιδιαίτερα χρήσιμη όταν έχουμε μια δοσμένη ακολουθία και πολλά μοντέλα από τα οποία μπορεί η ακολουθία αυτή να έχει προέλθει. Η λύση στο ΠΡΟΒΛΗΜΑ Α θα μας δώσει και την απάντηση σε ποιο μοντέλο «ταιριάζει» περισσότερο η δοσμένη ακολουθία..

ΠΡΟΒΛΗΜΑ Β

Εδώ ουσιαστικά πρέπει να αποκαλυφθεί το «κρυφό» μέρος του μοντέλου (πχ. να βρεθεί η σωστή ακολουθία καταστάσεων). Είναι προφανές ότι για όλα τα μοντέλα,

πλην ορισμένων εκφυλισμένων καταστάσεων, δεν υπάρχει «σωστή» ακολουθία. Αντ' αυτού επιχειρείται η βελτιστοποίηση κάποιου κριτηρίου ώστε να δοθεί η καλύτερη δυνατή λύση. Δυστυχώς, υπάρχουν αρκετά κριτήρια βελτιστοποίησης τα οποία μπορούν να χρησιμοποιηθούν και έτσι σε κάθε περίπτωση, αναλόγως με την εφαρμογή και το σκοπό που έχει η «αποκάλυψη» της ακολουθίας των καταστάσεων, επιλέγεται το πιο κατάλληλο.

ΠΡΟΒΛΗΜΑ Γ

Το πρόβλημα αυτό αφορά την βελτιστοποίηση των παραμέτρων του μοντέλου ώστε να περιγράφουν όσο το δυνατόν καλύτερα πώς προέκυψε η δοσμένη ακολουθία.. Η ακολουθία που χρησιμοποιείται για την εκπαίδευση του μοντέλου ονομάζεται ακολουθία εκπαίδευσης καθώς με βάση αυτή γίνεται η εκμάθηση του μοντέλου (βελτιστοποίηση των παραμέτρων του). Το πρόβλημα της εκπαίδευσης του μοντέλου είναι ιδιαίτερος σημαντικό καθώς χρειάζεται οι παράμετροι του μοντέλου να προσαρμοστούν κατά βέλτιστο τρόπο στην ακολουθία εξόδων ώστε το μοντέλο να περιγράφει όσο το δυνατόν καλύτερα το φυσικό φαινόμενο το οποίο καλείται να μοντελοποιήσει.

Στη συνέχεια θα δοθεί συνοπτικά το μαθηματικό υπόβαθρο που χρησιμοποιείται για την λύση των τριών αυτών προβλημάτων.

ΛΥΣΗ ΣΤΟ ΠΡΟΒΛΗΜΑ Α

Θέλουμε να υπολογίσουμε την πιθανότητα η δεδομένη ακολουθία να έχει προέλθει από το συγκεκριμένο μοντέλο δηλ $P(O | \lambda)$. Ο πιο προφανής τρόπος είναι αριθμώντας κάθε πιθανή ακολουθία καταστάσεων, μήκους T (όσο είναι και το μήκος της ακολουθίας εξόδων). Ας υποθέσουμε ότι έχουμε μια τέτοια ακολουθία

$$Q = q_1 q_2 \dots q_T \quad (1)$$

όπου q_1 είναι η αρχική κατάσταση. Η πιθανότητα να προκύψει η δεδομένη ακολουθία εξόδων με ακολουθία καταστάσεων την (1), υποθέτοντας στατιστική ανεξαρτησία των παρατηρήσεων είναι

$$P(O | Q, \lambda) = \prod_{t=1}^T P(Q_t | q_t, \lambda) \quad (2)$$

η οποία γράφεται και

$$P(O | Q, \lambda) = b_{q_1}(O_1) \cdot b_{q_2}(O_2) \dots b_{q_T}(O_T) \quad (3)$$

Η πιθανότητα να προκύψει μια τέτοια ακολουθία Q είναι

$$P(Q | \lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \dots a_{q_{T-1} q_T} \quad (4)$$

Η δεσμευμένη πιθανότητα να έχουμε την ακολουθία O , δεδομένου ότι η ακολουθία καταστάσεων είναι η Q , είναι το γινόμενο των δύο παραπάνω όρων δηλαδή

$$P(O, Q | \lambda) = P(O | Q, \lambda) P(Q, \lambda) \quad (5)$$

Η πιθανότητα να έχουμε την ακολουθία εξόδων O (με δεδομένο το μοντέλο) μπορεί να υπολογιστεί ως άθροισμα των παραπάνω δεσμευμένων πιθανοτήτων για όλες τις πιθανές ακολουθίες q και είναι

$$P(O | \lambda) = \sum_{all Q} P(O | Q, \lambda) P(Q, \lambda) = \sum_{q^1, q^2, \dots, q^T} \pi_{q^1} b_{q^1}(O_1) a_{q^1 q^2} b_{q^2}(O_2) \dots a_{q^{T-1} q^T} b_{q^T}(O_T) \quad (6)$$

Εύκολα γίνεται αντιληπτό πως η παραπάνω λύση μπορεί να είναι η πιο προφανής αλλά το υπολογιστικό της κόστος είναι τεράστιο. Ο υπολογισμός της $P(O | \lambda)$ σύμφωνα με την σχέση (6) περιλαμβάνει $2T * N^T$ υπολογισμούς πράγμα το οποίο σημαίνει πως για ένα μικρό σχετικά μοντέλο με $N=5$ και μήκος ακολουθίας $T=100$ θα χρειαστεί ο αστρονομικός αριθμός των 10^{72} υπολογισμών. Η ανάγκη για μια πιο αποδοτική λύση είναι προφανής. Η λύση αυτή είναι η εμπρός – πίσω διαδικασία (forward-backward procedure) και περιγράφεται στη συνέχεια.

Η ΕΜΠΡΟΣ – ΠΙΣΩ ΔΙΑΔΙΚΑΣΙΑ

Θεωρούμε την προς τα εμπρός μεταβλητή $\alpha_t(i)$ η οποία ορίζεται ως εξής

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i | \lambda)$$

η οποία είναι η πιθανότητα να έχει προκύψει, με δεδομένο το μοντέλο, το μέρος της ακολουθίας $O_1 O_2 \dots O_t$ (μέχρι την χρονική στιγμή t) και κατάσταση την χρονική στιγμή t την S_i . Μπορούμε να υπολογίσουμε επαγωγικά το $\alpha_t(i)$ ως εξής:

1) Αρχικοποίηση

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

2) Επαγωγή

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

$$1 \leq j \leq N$$

3) Τερματισμός

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

Το βήμα 1 αρχικοποιεί την προς τα εμπρός πιθανότητα ως την δεσμευμένη πιθανότητα της κατάστασης S_i όταν η πρώτη παρατήρηση είναι η O_1 . Το βήμα της επαγωγής αποτελεί και την «καρδιά» της διαδικασίας. Η κατάσταση S_j , την χρονική στιγμή $t+1$, μπορεί να προέλθει από τις N δυνατές καταστάσεις S_i , $1 \leq i \leq N$ της

χρονικής στιγμής t . Αφού η μεταβλητή $a_t(i)$ είναι η δεσμευμένη πιθανότητα να έχουμε ακολουθία εξόδου $O_1 O_2 \dots O_t$ όταν η κατάσταση την χρονική στιγμή t είναι η S_i , το γινόμενο $a_t(i) * a_{ij}$ είναι η πιθανότητα να έχουμε ακολουθία εξόδου $O_1 O_2 \dots O_t$ και η κατάσταση S_j την χρονική στιγμή $t+1$ έχει προέλθει από την κατάσταση S_i της χρονικής στιγμής t . Αθροίζοντας αυτό το γινόμενο πάνω στις N πιθανές καταστάσεις S_i την χρονική στιγμή t , με $1 \leq i \leq N$, μας δίνει την πιθανότητα, την χρονική στιγμή $t+1$, να προκύψει η κατάσταση S_j με δεδομένη τη μέχρι εκείνη τη στιγμή ακολουθία εξόδου. Αφού υπολογιστεί το S_j , το $a_{t+1}(j)$ υπολογίζεται εύκολα με πολλαπλασιασμό του αθροίσματος που έχει προκύψει με την πιθανότητα στην κατάσταση j να έχουμε την έξοδο O_{t+1} δηλ. $b_j(O_{t+1})$. Ο παραπάνω υπολογισμός γίνεται για όλες τις καταστάσεις j , με $1 \leq j \leq N$, για δεδομένο t και εν συνεχεία επαναλαμβάνεται για $t=1,2,\dots,T-1$. Τέλος, το τρίτο βήμα μας δίνει την επιθυμητή τιμή της $P(O | \lambda)$ σαν το άθροισμα όλων των τελικών προς τα εμπρός μεταβλητών $a_T(i)$.

Η αποδοτικότητα του συγκεκριμένου αλγόριθμου είναι σαφώς καλύτερη από αυτή του προηγούμενου. Οι υπολογισμοί που χρειάζονται είναι της τάξεως του $N^2 * T$ πράγμα το οποίο σημαίνει ότι για μοντέλο ίδιο με το προηγούμενο ($N=5$ και $T=100$) χρειάζονται 3000 υπολογισμοί (αντί για 10^{72}). Το γεγονός που αυξάνει την αποδοτικότητα του αλγόριθμου είναι ότι αφού υπάρχουν μόνο N καταστάσεις, όλες οι δυνατές ακολουθίες θα προκύπτουν από τις ίδιες N καταστάσεις κάθε χρονική στιγμή. Έτσι την χρονική στιγμή $t=1$ υπολογίζουμε τις τιμές $a_t(i)$ με $1 \leq i \leq N$. Τις χρονικές στιγμές $t=2,3,\dots,T$ χρειάζεται να υπολογίσουμε μόνο τις τιμές $a_t(j)$ με $1 \leq j \leq N$, και ο κάθε υπολογισμός περιλαμβάνει μόνο N προηγούμενες τιμές του $a_{t-1}(i)$ γιατί καθεμιά από τις N καταστάσεις είναι η επόμενη κάποιας από τις N , ίδιες, καταστάσεις της προηγούμενης χρονικής στιγμής.

Με αντίστοιχο τρόπο ορίζεται και η προς τα πίσω διαδικασία και η προς τα πίσω μεταβλητή $\beta_t(i)$. Σημειώνουμε όμως ότι η προς τα πίσω διαδικασία χρησιμοποιείται για την επίλυση του προβλήματος της εκπαίδευσης του μοντέλου και δεν χρειάζεται για την επίλυση του προβλήματος A .

ΛΥΣΗ ΣΤΟ ΠΡΟΒΛΗΜΑ Β

Όπως προαναφέραμε, σε αυτό το πρόβλημα δεν μπορεί να υπάρξει συγκεκριμένη λύση όπως αυτή που δόθηκε στο ΠΡΟΒΛΗΜΑ Α. Υπάρχουν πολλοί τρόποι επίλυσης αυτού του προβλήματος, της εύρεσης δηλαδή, της βέλτιστης, κατά μια έννοια, ακολουθίας καταστάσεων, που σχετίζεται με τη δεδομένη ακολουθία εξόδων. Η μεγαλύτερη δυσκολία έγκειται στον ορισμό της βέλτιστης ακολουθίας, καθώς υπάρχουν αρκετά κριτήρια βελτιστοποίησης. Για παράδειγμα, ένα κριτήριο βελτιστοποίησης είναι η επιλογή των καταστάσεων q_t έτσι ώστε να είναι η καθεμία ξεχωριστά πιο πιθανή. Αυτό το κριτήριο μεγιστοποιεί τον αναμενόμενο αριθμό «σωστών» ξεχωριστών καταστάσεων. Για να εφαρμόσουμε την παραπάνω σκέψη στη λύση του ΠΡΟΒΛΗΜΑΤΟΣ Β ορίζουμε την ακόλουθη μεταβλητή

$$\gamma_t(i) = P(q_t = S_i | O, \lambda).$$

δηλαδή την πιθανότητα να βρισκόμαστε στην κατάσταση S_i την χρονική στιγμή t , δεδομένης της ακολουθίας εξόδου O καθώς και του μοντέλου λ . Κάνοντας χρήση των προς τα εμπρός και προς τα πίσω μεταβλητών η παραπάνω σχέση γράφεται

$$\gamma_t(i) = \frac{a_t(i) \beta_t(i)}{P(O | \lambda)} = \frac{a_t(i) \beta_t(i)}{\sum_{i=1}^N a_t(i) \beta_t(i)}$$

αφού το $a_t(i)$ περιλαμβάνει το μέρος της ακολουθίας μέχρι την χρονική στιγμή t ($O_1 O_2 \dots O_t$) ενώ το $b_t(i)$ περιλαμβάνει την εναπομένουσα ακολουθία μέχρι το T ($O_{t+1} O_{t+2} \dots O_T$), με δεδομένη την κατάσταση S_i την χρονική στιγμή t . Ο παράγοντας κανονικοποίησης $\sum_{i=1}^N a_t(i) b_t(i)$ κάνει την $\gamma_t(i)$ μετρήσιμη, έτσι ώστε

$\sum_{i=1}^N \gamma_t(i) = 1$. Χρησιμοποιώντας το $\gamma_t(i)$ μπορούμε να λύσουμε το πρόβλημα εύρεσης της ξεχωριστά πιο πιθανής κατάστασης q_t την χρονική στιγμή t ως εξής

$$q_t = \operatorname{argmax} [\gamma_t(i)], \quad 1 \leq i \leq N, \quad 1 \leq t \leq T \quad (7)$$

Αν και η παραπάνω λύση μεγιστοποιεί τον αριθμό των σωστών καταστάσεων (διαλέγοντας κάθε χρονική στιγμή την πιο πιθανή κατάσταση), είναι δυνατόν να παρατηρηθούν προβλήματα με την συνολική ακολουθία που θα προκύψει. Αν για παράδειγμα το συγκεκριμένο HMM έχει καταστάσεις μεταξύ των οποίων υπάρχει μηδενική πιθανότητα μετάβασης ($a_{ij} = 0$ για κάποια i και j) είναι πιθανό η ακολουθία καταστάσεων που θα προκύψει ως βέλτιστη, να μην είναι καν δυνατόν να προέλθει από το συγκεκριμένο μοντέλο. Αυτό συμβαίνει διότι η σχέση (7) καθορίζει απλά την πιο πιθανή κατάσταση κάθε χρονική στιγμή χωρίς να λαμβάνει υπ όψιν την την πιθανότητα εμφάνισης ακολουθιών καταστάσεων

Μια λύση στο παραπάνω πρόβλημα είναι η επιλογή ενός διαφορετικού κριτηρίου βελτιστοποίησης. Για παράδειγμα, θα μπορούσε να χρησιμοποιηθεί ως κριτήριο προς μεγιστοποίηση ο αριθμός των σωστών ζευγαριών καταστάσεων (q_t, q_{t+1}) ή των τριάδων καταστάσεων (q_t, q_{t+1}, q_{t+2}) κτλ.. Αν και αυτά τα κριτήρια μπορεί να είναι λογικά για μια σειρά εφαρμογών το πιο διαδεδομένο κριτήριο είναι της εύρεσης του βέλτιστου μονοπατιού (ακολουθίας καταστάσεων) δηλαδή η μεγιστοποίηση της πιθανότητας $P(Q | O, \lambda)$ που ισοδυναμεί με την μεγιστοποίηση της πιθανότητας $P(Q, O | \lambda)$. Η μέθοδος που έχει αναπτυχθεί για τον υπολογισμό του βέλτιστου μονοπατιού ονομάζεται αλγόριθμος του Viterbi και βασίζεται σε τεχνικές γραμμικού προγραμματισμού.

ΑΛΓΟΡΙΘΜΟΣ VITERBI

Ορίζουμε την ποσότητα

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

δηλαδή $\delta_{t+1}(i)$ είναι η μεγαλύτερη πιθανότητα κατά μήκος ενός απλού μονοπατιού την χρονική στιγμή t , που περιλαμβάνει τις πρώτες t παρατηρήσεις και τελειώνει στην κατάσταση S_i . Με επαγωγή έχουμε

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

Στην πραγματικότητα για να υπολογίσουμε το βέλτιστο μονοπάτι πρέπει να ακολουθήσουμε το όρισμα που μεγιστοποιεί την παραπάνω σχέση για κάθε t και j . Αυτό το κάνουμε με χρήση ενός ακόμα πίνακα, του $\psi_t(j)$. Ο αλγόριθμος είναι ο εξής

1) Αρχικοποίηση

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

$$\psi_1(i) = 0$$

2) Επανάληψη

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

$$\psi_t(j) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad \begin{matrix} 2 \leq t \leq T \\ 1 \leq j \leq T \end{matrix}$$

3) Τερματισμός

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

$$q^*_T = \operatorname{argmax}_{1 \leq j \leq N} [\delta_T(i)]$$

4) Εξαγωγή του μονοπατιού

$$q^*_t = \psi_{t+1}(q^*_{t+1}), \quad t = T-1, T-2, \dots, 1.$$

ΛΥΣΗ ΣΤΟ ΠΡΟΒΛΗΜΑ Γ

Το τρίτο και δυσκολότερο πρόβλημα είναι ο προσδιορισμός μιας μεθόδου που να προσαρμόζει τις παραμέτρους του μοντέλου (A, B, π) ώστε να μεγιστοποιείται η πιθανότητα η δοσμένη ακολουθία εξόδων να έχει προέλθει από το μοντέλο. Δεν υπάρχει κάποιος αναλυτικός τρόπος εύρεσης των παραμέτρων που θα μεγιστοποιήσουν αυτή την πιθανότητα. Στην πραγματικότητα, δοσμένης μιας πεπερασμένης ακολουθίας εξόδων σαν δεδομένα εκπαίδευσης, δεν υπάρχει βέλτιστος τρόπος εκτίμησης των παραμέτρων του μοντέλου. Μπορούμε παρόλα αυτά να διαλέξουμε $\lambda = (A, B, \pi)$ έτσι ώστε η $P(O | \lambda)$ να μεγιστοποιείται τοπικά με χρήση κάποιας επαναληπτικής μεθόδου όπως ο αλγόριθμος Baum – Welch ή με χρήση τεχνικών κλίσης. Παρακάτω θα περιγράψουμε μία επανάληψη του αλγόριθμου Baum – Welch για την επιλογή των παραμέτρων του μοντέλου.

Για να περιγράψουμε αυτή την μέθοδο επανεκτίμησης (επαναληπτική ενημέρωση και βελτίωση) των παραμέτρων του HMM ορίζουμε την πιθανότητα $\xi_t(i, j)$, να είμαστε στην κατάσταση s_i τον χρόνο t και στην s_j τον χρόνο $t+1$

$$\xi_t(i, j) = P[q_t = s_i, q_{t+1} = s_j | O, \lambda]$$

Η παραπάνω πιθανότητα μπορεί να οριστεί και συναρτήσει των παραμέτρων $a_t(i)$ και $b_t(j)$ του αλγόριθμου forward-backward.

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(O|\lambda)}$$

Παραπάνω έχει οριστεί η πιθανότητα $\gamma_t(i)$ ως η πιθανότητα να είμαστε στην κατάσταση s_i τον χρόνο t για δεδομένο μοντέλο και ακολουθία παρατηρήσεων O . Αν αθροίσουμε για όλα τα j , τις τιμές της $\xi_t(i, j)$ λαμβάνουμε την $\gamma_t(i)$

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

Αν αθροίσουμε την $\gamma_t(i)$ για όλες τις χρονικές στιγμές t παίρνουμε τον αναμενόμενο αριθμό των φορών κατά τις οποίες το σύστημα φτάνει στην κατάσταση s_i στην διάρκεια του χρόνου των παρατηρήσεων, ή ισοδύναμα, τον αναμενόμενο αριθμό των μεταφορών που γίνονται από την κατάσταση s_i εξαιρούμενης από το άθροισμα της χρονικής στιγμής κατά την οποία $t=T$. Από τα παραπάνω προκύπτει ότι το άθροισμα της χρονικής στιγμής $\xi_t(i, j)$ στο χρόνο είναι ο αναμενόμενος αριθμός των μεταφορών από την κατάσταση s_i στην κατάσταση s_j .

$$\sum_{t=1}^{T-1} \gamma_t(i) = \text{αναμενόμενος αριθμός μεταβάσεων από την } s_i$$

$$\sum_{\tau=1}^{T-1} \xi_{\tau}(i, j) = \text{αναμενόμενος αριθμός μεταβάσεων από την } s_i \text{ στην } s_j$$

Χρησιμοποιώντας τις παραπάνω σχέσεις έχουμε μια μέθοδο επανεκτίμησης των παραμέτρων HMM. Οι εξισώσεις είναι για τα A , B και π είναι

$$a'_{ij} = \frac{\text{αναμενόμενος αριθμός μεταβίσεων από την } s_i \text{ στην } s_j}{\text{αναμενόμενος αριθμός μεταβάσεων από την } s_i}$$

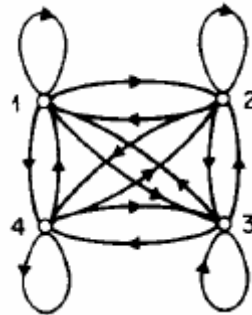
$$b'_j(k) = \frac{\text{αναμενόμενη πιθανότητα της κατάστασης } s_j \text{ και ταυτόχρονη παρατήρηση του συμβόλου } v_k}{\text{αναμενόμενη πιθανότητα της } s_j}$$

$$\pi'_i = \text{αναμενόμενη συχνότητα στην κατάσταση } s_i \text{ τη χρονική στιγμή } t=1$$

Αν υποθέσουμε ότι έχουμε το υπάρχον μοντέλο, $\lambda=(A,B,\pi)$ και το μοντέλο $\lambda'=(A',B',\pi')$, το δεύτερο είναι πιο πιθανό ότι παρήγαγε τις ακολουθίες παρατηρήσεων. Οι παράμετροι του μοντέλου επανεκτιμούνται έως ότου οι νέες τιμές να είναι ίδιες με τις προηγούμενες.

5.4 Τύποι HMM

Μέχρι στιγμής έχουμε θεωρήσει μόνο την ειδική περίπτωση του εργοδικού ή πλήρους διασυνδεδεμένου HMM, στο οποίο κάθε κατάσταση θα μπορούσε να προσεγγιστεί από οποιαδήποτε άλλη κατάσταση σε ένα μόνο βήμα. (σύμφωνα με τον αυστηρό ορισμό σε ένα εργοδικό μοντέλο κάθε κατάσταση μπορεί να προσεγγιστεί από οποιαδήποτε άλλη με ένα πεπερασμένο αριθμό βημάτων). Για το μοντέλο με $N=4$ που φαίνεται στο σχήμα 1 ισχύει ότι όλοι οι συντελεστές μετάβασης a_{ij} είναι θετικοί.



Σχήμα (1)

Για μερικές εφαρμογές, όπως είναι η αναγνώριση φωνής καθώς και η αναγνώριση χειρονομιών, έχει βρεθεί ότι άλλοι τύποι HMM μοντελοποιούν καλύτερα το φυσικό σήμα απ' ό,τι ένα εργοδικό μοντέλο. Ένα τέτοιο μοντέλο φαίνεται στο σχήμα 2. Αυτό το μοντέλο καλείται αριστερό-δεξί μοντέλο ή αλλιώς Bakis γιατί έχει



Σχήμα (2)

την ιδιότητα ότι καθώς αυξάνεται ο χρόνος αυξάνεται επίσης και ο αριθμός της κατάστασης στην οποία βρίσκεται ή το πολύ παραμένει η ίδια (δεν μπορεί να επιστρέψει σε κάποια προηγούμενη κατάσταση προχωρώντας έτσι από τα αριστερά προς τα δεξιά). Είναι προφανές ότι τα μοντέλα αυτού του τύπου έχουν την επιθυμητή ιδιότητα να μοντελοποιούν παραστατικά σήματα τα οποία αλλάζουν με τον χρόνο. Η θεμελιώδης ιδιότητα όλων των Bakis μοντέλων είναι ότι για τους συντελεστές μετάβασης ισχύει $a_{ij} = 0$ για $j < i$. Επίσης οι πιθανότητες των αρχικών καταστάσεων έχουν την ιδιότητα

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases}$$

καθώς η ακολουθία πρέπει να ξεκινάει από την κατάσταση 1 (και να τελειώνει στην κατάσταση N). Πολλές φορές στα Bakis μοντέλα τίθενται επιπρόσθετοι περιορισμοί

στους συντελεστές μετάβασης ώστε να εξασφαλιστεί ότι δεν θα συμβαίνουν μεγάλες μεταβάσεις. Ένας τέτοιος περιορισμός είναι πχ. $a_{ij} = 0$, $j > i + \Delta$. Για το μοντέλο του σχήματος 2 , είναι $\Delta=2$ και ο πίνακας μετάβασης είναι ο παρακάτω

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}$$

Είναι προφανές ότι για ένα τέτοιο μοντέλο οι συντελεστές της τελευταίας κατάστασης ορίζονται ως

$$a_{NN} = 1$$

$$a_{Ni} = 0, \text{ για } i < N$$

Παρόλο που διαχωρίσαμε τα HMM σε εργοδικά και δεξιά – αριστερά μοντέλα, υπάρχουν ακόμα πολλοί άλλοι τύποι και συνδυασμοί. Όπως το HMM του σχήματος 3 , που δείχνει ένα HMM το οποίο δημιουργείται με την παράλληλη σύνδεση δύο Bakis μοντέλων. Το μοντέλο αυτό υπακούει μεν στους περιορισμούς των Bakis μοντέλων για τους συντελεστές μετάβασης, αλλά έχει επιπλέον χαρακτηριστικά που δεν συναντώνται στα Bakis μοντέλα.

Τέλος, πρέπει να τονίσουμε πως οι περιορισμοί στα Bakis ή σε άλλα μοντέλα, δεν επηρεάζουν την διαδικασία εκτίμησης των παραμέτρων. Αυτό συμβαίνει γιατί όταν η αρχική τιμή για κάποιο στοιχείο του πίνακα μετάβασης είναι 0, το στοιχείο αυτό θα παραμείνει 0 και μετά την ολοκλήρωση της διαδικασίας της εκτίμησης των παραμέτρων.

5.5 Συνεχείς Ποσότητες Παρατήρησης στα HMM

Όλα όσα έχουμε αναφέρει μέχρις εδώ αφορούν HMM των οποίων οι έξοδοι είναι διακριτά σύμβολα τα οποία προέρχονται από ένα πεπερασμένο αλφάβητο και για το λόγο αυτό μπορούσαμε να χρησιμοποιήσουμε μια διακριτή κατανομή πιθανότητας για την έξοδο σε κάθε κατάσταση του μοντέλου. Το πρόβλημα με αυτή την προσέγγιση , τουλάχιστον σε μερικές εφαρμογές είναι ότι οι παρατηρήσεις (έξοδοι) είναι συνεχή σήματα. Αν και είναι δυνατόν αυτά τα σήματα να υποστούν κβαντισμό και να γίνουν διακριτά αυτό μπορεί να οδηγήσει σε σφάλμα και αλλοίωση της μορφής τους. Για το λόγο αυτό θα ήταν πολύ χρήσιμη η ύπαρξη HMM με συνεχείς κατανομές εξόδων.

Προκειμένου να χρησιμοποιηθούν συνεχείς κατανομές για τις εξόδους, πρέπει να τεθούν κάποιοι περιορισμοί στην μορφή του μοντέλου της συνάρτησης πυκνότητας πιθανότητας, ώστε να εξασφαλιστεί ότι θα πετύχουμε μια αξιόπιστη εκτίμηση των παραμέτρων της. Η πιο γενική παρουσίαση της συνάρτησης πυκνότητας πιθανότητας , για την οποίαν έχει αναπτυχθεί και η διαδικασία εκτίμησης, είναι ένα μείγμα της μορφής

$$b_i(\mathbf{O}) = \sum_{m=1}^M c_{jm} \prod [O, \mu_{jm}, U_{jm}], \quad 1 \leq j \leq N$$

όπου \mathbf{O} είναι το διάνυσμα που μοντελοποιείται, c_{jm} είναι το ο συντελεστής μείγματος για το m-οστό μείγμα στην κατάσταση j και \mathcal{O}_t είναι μια οποιαδήποτε λογαριθμική ή συμμετρική κατανομή (πχ. Γκαουσιανή) με μέση τιμή μ_{jm} και πίνακα αυτοσυσχέτισης U_{jm} για το M-οστο μείγμα στην κατάσταση j. Συνήθως η κατανομή που χρησιμοποιείται είναι η Γκαουσιανή. Οι συντελεστές μείγματος c_{jm} ικανοποιούν τους στοχαστικούς περιορισμούς

$$\begin{aligned} \sum_{m=1}^M c_{jm} &= 1, & 1 \leq j \leq N \\ c_{jm} &\geq 0, & 1 \leq j \leq N, 1 \leq m \leq M \end{aligned}$$

έτσι ώστε να ισχύει για την κατανομή η γνωστή κανονικοποίηση

$$\int_{-\infty}^{\infty} b_i(x) dx = 1, \quad 1 \leq j \leq N$$

Η παραπάνω συνάρτηση πυκνότητας πιθανότητας μπορεί να χρησιμοποιηθεί για την προσέγγιση σχεδόν οποιασδήποτε κατανομής και άρα είναι ιδιαίτερος πολλές οι εφαρμογές.

Μπορεί να αποδειχτεί ότι οι τύποι για την προσέγγιση των συντελεστών των μειγμάτων της κατανομής δηλ. για τα c_{jm} , μ_{jk} , U_{jk} είναι οι ακόλουθοι

$$c_{jk} = \frac{\sum_{t=1}^T \gamma(i, k)}{\sum_{t=1}^T \sum_{k=1}^M \gamma(i, k)} \quad (8)$$

$$\mu_{jk} = \frac{\sum_{t=1}^T \gamma(i, k) O_t}{\sum_{t=1}^T \gamma(i, k)}$$

$$U_{jk} = \frac{\sum_{t=1}^T \gamma(i, k) (O_t - \mu_{\xi k})(O_t - \mu_{jk})'}{\sum_{t=1}^T \gamma(i, k)}$$

όπου $\gamma_t(\mathbf{j}, \mathbf{k})$ είναι η πιθανότητα την χρονική στιγμή t να έχουμε την κατάσταση j και το k μείγμα να συμβάλει στην δημιουργία του \mathbf{O}_t δηλ.

$$\gamma_t(i,k) = \left[\frac{a_t(j)\beta_t(j)}{\sum_{j=1}^N a_t(j)\beta_t(j)} \right] \left[\frac{c_{jk} \prod (O_t, \mu_{jk}, u_{jk})}{\sum_{m=1}^M c_{jk} \prod (O_t, \mu_{jk}, u_{jk})} \right]$$

Ο όρος $\gamma_t(j,k)$ είναι η γενίκευση του όρου $\gamma_t(j)$ ο οποίος χρησιμοποιείται όταν έχουμε ένα μόνο μείγμα ή όταν η κατανομή είναι διακριτή. Η διαδικασία εκτίμησης του πίνακα μετάβασης (των συντελεστών a_{ij}) είναι η ίδια όπως και στα διακριτά HMM. Η εκτίμηση για το c_{jk} είναι ο λόγος του αριθμού που το σύστημα βρίσκεται στην κατάσταση j χρησιμοποιώντας το k μείγμα προς τον συνολικό αριθμό των φορών που το σύστημα βρίσκεται στην j κατάσταση. Παρομοίως, η διαδικασία εκτίμησης για το διάλυμα μέσης τιμής μ_{jk} χρησιμοποιεί ως βάρη στον αριθμητή της (8) τις παρατηρήσεις O_t, \dots, \dots . Παρόμοια είναι η ερμηνεία και για την διαδικασία εκτίμησης του πίνακα ετεροσυσχέτισης U_{jk}

ΚΕΦΑΛΑΙΟ 6

ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ

6.1 Γενική Περιγραφή του Συστήματος

Στα πλαίσια της παρούσας διπλωματικής εργασίας υλοποιήθηκε ένα σύστημα αναγνώρισης επτά διαφορετικών χειρονομιών. Η αναγνώριση των χειρονομιών, με τον τρόπο που υλοποιήθηκε, δεν γίνεται σε πραγματικό χρόνο. Το σύστημα όμως θα μπορούσε να επεκταθεί, με τέτοιο τρόπο, ώστε να αναγνωρίζει, σε πραγματικό χρόνο τις χειρονομίες που κάνει ο χρήστης ο οποίος βρίσκεται καθισμένος μπροστά από μια αναλογική κάμερα που είναι συνδεδεμένη με τον υπολογιστή.

Ένα τέτοιο σύστημα μπορεί να έχει πολλές εφαρμογές που έχουν να κάνουν τόσο με την αναγνώριση της συναισθηματικής κατάστασης του χρήστη όσο και με την αντικατάσταση κάποιων εντολών του υπολογιστή, οι οποίες αυτή τη στιγμή δίνονται από το χρήστη με περιφερειακές συσκευές (mouse, joystick, keyboard), από χειρονομίες. Με την ανάπτυξη τέτοιων εφαρμογών δίνεται η δυνατότητα σε ανθρώπους οι οποίοι δεν είναι απαραίτητως, ιδιαιτέρως εξοικειωμένοι με τους υπολογιστές, καθώς και σε άτομα με ειδικές ανάγκες, να επικοινωνούν εύκολα και αποδοτικά με τον υπολογιστή.

Προφανώς, όταν η εφαρμογή εντάσσεται στο γενικότερο πλαίσιο της εκτίμησης της συναισθηματικής κατάστασης του χρήστη, η αναγνώριση των χειρονομιών αποτελεί μόνο μια επιμέρους ένδειξη, αφού μεγαλύτερη έμφαση δίνεται στις εκφράσεις του προσώπου. Επίσης, σε τέτοιες εφαρμογές ο χρήστης είναι ελεύθερος να κάνει χειρονομίες οι οποίες δεν έχουν τόσο αυστηρή δομή, όσο θα έπρεπε να έχουν σε εφαρμογές στις οποίες οι χειρονομίες ερμηνεύονται ως συγκεκριμένες εντολές από τον υπολογιστή ή σε εφαρμογές που οι χειρονομίες αποτελούν λέξεις κάποιας νοηματικής γλώσσας. Για τους παραπάνω λόγους σε κάθε εφαρμογή απαιτείται διαφορετική ακρίβεια στην αναγνώριση, γίνεται με διαφορετικό τρόπο η εκπαίδευση του μοντέλου της κάθε χειρονομίας αλλά και εισάγονται περιορισμοί στις χειρονομίες αλλά και στον τρόπο με τον οποίον αυτές πρέπει να εκτελούνται από τον χρήστη.

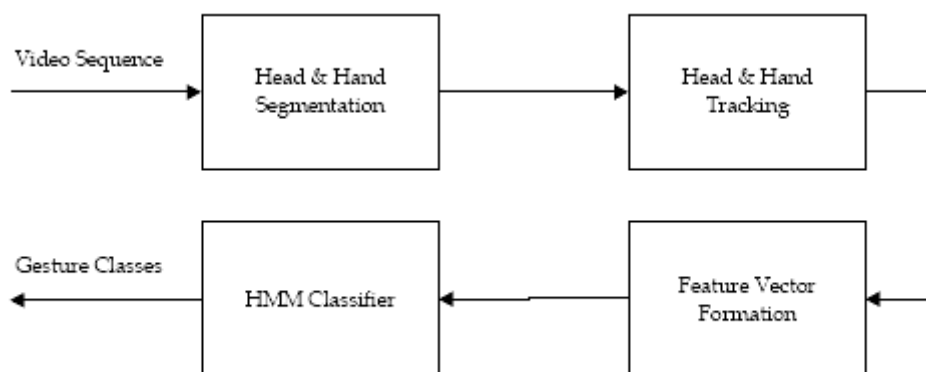
Όπως προαναφέραμε, το σύστημα αναγνώρισης χειρονομιών που υλοποιήθηκε, αναγνωρίζει επτά χειρονομίες και χρησιμοποιεί τεχνολογία όρασης υπολογιστών. Αυτό σημαίνει πως για εντοπισμό της θέσης των χεριών και του κεφαλιού και την εξαγωγή τροχιάς που ακολουθούν και των χαρακτηριστικών γνωρισμάτων της κάθε χειρονομίας, χρησιμοποιείται μόνο το σήμα το οποίο προέρχεται από την αναλογική κάμερα που βιντεοσκοπεί τον χρήστη. Δεν χρησιμοποιούνται άλλου είδους αισθητήρες, όπως ειδικά γάντια ή στολές, οι οποίοι μπορεί μεν να παρέχουν πιο ακριβή δεδομένα αλλά μειώνουν αισθητά την ευχρηστία του συστήματος αφού μπορεί να εμποδίζουν την ευχέρεια των κινήσεων του χρήστη αλλά απαιτούν και εξειδικευμένες γνώσεις για την βαθμονόμησή τους και την εν γένει χρήση τους.

Βεβαία ένα μειονέκτημα που σχετίζεται με την τεχνολογία που χρησιμοποιείται για τον εντοπισμό των χεριών, έχει να κάνει με το γεγονός πως χρησιμοποιείται μόνο μία κάμερα για την βιντεοσκόπηση του χρήστη, πράγμα το οποίο σημαίνει πως ορισμένες παράμετροι των κινήσεων δεν μπορούν να καθοριστούν πλήρως. Παρόλα αυτά, τις περισσότερες φορές, αρκούν ορισμένες μόνο

παράμετροι για την αποδοτική μοντελοποίηση της κάθε κίνησης, οπότε η απώλεια ορισμένων άλλων, δεν δημιουργεί ιδιαίτερα προβλήματα.

Για την ταξινόμηση των χειρονομιών χρησιμοποιούνται κρυφά μαρκοβιανά μοντέλα. Η αντίστοιχη θεωρία και το κατάλληλο μαθηματικό υπόβαθρο που αφορά τα HMM παρουσιάστηκαν σε προηγούμενο κεφάλαιο, ενώ ο τρόπος με τον οποίο έγινε η εκπαίδευσή τους, καθώς και η ταξινόμηση των χειρονομιών με χρήση του προγραμματιστικού περιβάλλοντος του MATLAB, θα αναλυθεί στη συνέχεια. Η χρήση των HMM σε εφαρμογές που αφορούν τόσο αναγνώριση φωνής όσο και αναγνώριση χειρονομιών είναι πολύ διαδεδομένη. Το κοινό στοιχείο της φωνής και των ανθρώπινων χειρονομιών είναι ο στοχαστικός τους χαρακτήρας. Δεν υπάρχει δηλαδή κάποια συγκεκριμένη συνάρτηση που να περιγράφει την κάθε χειρονομία είτε αυτή γίνεται από τον ίδιο χρήστη και ακόμη περισσότερο όταν αυτή γίνεται από διαφορετικούς χρήστες. Οι ανθρώπινες χειρονομίες είναι εκφραστικές κινήσεις του ανθρωπίνου σώματος οι οποίες, γενικά, περιέχουν τόσο χρονικές όσο και χωρικές μεταβολές. Υπό αυτή την έννοια, όταν αναφερόμαστε σε ανθρώπινες χειρονομίες, αναφερόμαστε σε « διπλά στοχαστικές διαδικασίες». Οι «διπλά στοχαστικές διαδικασίες» είναι μαρκοβιανές διαδικασίες των οποίων οι εσωτερικές καταστάσεις δεν μπορούν να παρατηρηθούν άμεσα και ακριβώς τέτοιες διαδικασίες μοντελοποιούνται από τα HMM. Σε μια τέτοια διαδικασία, για κάθε μετάβαση από κατάσταση σε κατάσταση, το σύστημα παράγει κάποια έξοδο, η οποία εξαρτάται από την συνάρτηση κατανομής της πιθανότητας που αντιστοιχεί σε κάθε εσωτερική κατάσταση του μοντέλου. Μοντελοποιώντας λοιπόν τις ανθρώπινες χειρονομίες με χρήση HMM μπορούμε να αναπαραστήσουμε αποδοτικά αυτή τη στοχαστική φύση των χειρονομιών, να απαλλαγούμε από την χρονική εξάρτηση που παρουσιάζουν και να δώσουμε έμφαση στον δυναμικό χαρακτήρα τους.

Ένα γενικό μοντέλο που παρουσιάζει την δομή του συστήματος που υλοποιήσαμε είναι το ακόλουθο



Όπως φαίνεται και από το παραπάνω σχήμα, τα βήματα που ακολουθούνται για την αναγνώριση των χειρονομιών είναι τα εξής :

- Το αναλογικό βίντεο, που παρέχει ως έξοδο η κάμερα, χωρίζεται σε frames (30 frames /sec)
- Στην ακολουθία των frames, ξεχωρίζουμε τα τμήματα της εικόνας που περιλαμβάνουν το κεφάλι και τα χέρια. Η τμηματοποίηση αυτή επιτυγχάνεται με εφαρμογή масκών χρώματος (με βάση το χαρακτηριστικό χρώμα που έχει το δέρμα) και αλγόριθμους εκτίμησης κίνησης. Με τον τρόπο αυτό επιτυγχάνεται η εξαγωγή των

χαρακτηριστικών γνωρισμάτων κάθε χειρονομίας κατά την εξέλιξή της.

- Με βάση τα χαρακτηριστικά αυτά, δημιουργούνται κατάλληλα διάνυσματα τα οποία θα αποτελέσουν την είσοδο των HMM που μοντελοποιούν την κάθε χειρονομία.
- Το τελευταίο στάδιο είναι η ταξινόμηση των χειρονομιών. Αυτή επιτυγχάνεται με τον υπολογισμό της πιθανότητας που προκύπτει, το διάνυσμα εισόδου να έχει παραχθεί από καθένα από τα HMM. Αφού, όπως είπαμε, κάθε HMM μοντελοποιεί και μία χειρονομία, το HMM το οποίο θα παρουσιάσει την μεγαλύτερη πιθανότητα να έχει δημιουργήσει την ακολουθία εισόδου, αντιπροσωπεύει την χειρονομία που είναι πιθανότερο να αναπαριστά το συγκεκριμένο διάνυσμα εισόδου.

Σημειώνουμε τέλος, πως το τελευταίο στάδιο του συστήματος μας, η ταξινόμηση δηλαδή με χρήση των HMM πραγματοποιήθηκε με δύο διαφορετικούς τρόπους. Ο πρώτος είναι με χρήση διακριτών HMM. Τα διακριτά HMM έχουν ως κύριο χαρακτηριστικό ότι οι παρατηρήσιμες έξοδοι τους είναι στοιχεία ενός πεπερασμένου συνόλου συμβόλων. Για το λόγο αυτό απαιτείται και η αντίστοιχη μετατροπή των χαρακτηριστικών γνωρισμάτων που έχουν εξαχθεί από την διαδικασία εντοπισμού των κινήσεων των χεριών σε σύμβολα αυτού του συνόλου. Για την ταξινόμηση με βάση διακριτά HMM χρησιμοποιήθηκε το stats4.1 toolbox του MATLAB.

Ο δεύτερος τρόπος είναι με χρήση πολυδιάστατων συνεχών HMM. Τα συνεχή HMM είναι HMM με συνεχείς κατανομές εξόδων. Σε αυτού του είδους τα μοντέλα η έξοδος από κάθε κατάσταση εξαρτάται από τη μορφή της συνάρτησης πυκνότητας πιθανότητας που έχει οριστεί για κάθε κατάσταση. Η πιο συνηθισμένη προσέγγιση είναι η χρήση κ-μειγμάτων γκαουσιανών κατανομών. Με τον τρόπο αυτό τα παρατηρήσιμα σύμβολα μοντελοποιούνται από κ γκαουσιανές συνιστώσες (όπου κ είναι η διάσταση του χώρου των παρατηρούμενων διανυσμάτων). Για την υλοποίηση της ταξινόμησης με βάση πολυδιάστατα συνεχή HMM χρησιμοποιήθηκε το toolbox “HMM” του Kevin Murphy για το MATLAB. Η καθεμία από τις δύο αυτές υλοποιήσεις παρουσιάζει πλεονεκτήματα και μειονεκτήματα που έχουν να κάνουν με την ακρίβεια, τον υπολογιστικό φόρτο, τον τρόπο υπολογισμού της πιθανότητας για κάθε μοντέλο κατά την αποκωδικοποίηση μιας ακολουθίας κ.α. τα οποία παρουσιάζονται και εξηγούνται αναλυτικά στη συνέχεια.

6.2 Ο Αλγόριθμος Εντοπισμού του Δέρματος και των κινούμενων Μελών

Η κίνηση των χεριών, γενικά, αποτελείται από την καθολική κίνηση των χεριών και την τοπική κίνηση των δακτύλων. Ο εντοπισμός της κίνησης των χεριών έχει να κάνει τόσο με τον εντοπισμό της ολικής κίνησης, όσο και με τον εντοπισμό της τοπικής κίνησης. Υπάρχουν δύο ειδών χαρακτηριστικά που βοηθούν στην διαδικασία εντοπισμού : χρωματικά στοιχεία και στοιχεία κίνησης.

Ο εντοπισμός των χεριών αφορά την εξαγωγή των περιοχών των εικόνων της ακολουθίας, οι οποίες περιέχουν κάθε φορά τα χέρια. Το χρώμα του δέρματος αποτελεί ένα αποδοτικό και γρήγορο τρόπο να επιτευχθεί κάτι τέτοιο. Ανάλογα με

την αναπαράσταση του χρώματος στους διάφορους χρωματικούς χώρους (RGB, normalized RGB, YCbCr, HSV, YUV), υπάρχουν δύο ειδών προσεγγίσεις οι μη παραμετρικές και οι παραμετρικές.

Για τον εντοπισμό της αρθρωτής κίνησης των χεριών, σε όλους τους βαθμούς ελευθερίας, πρέπει να καθοριστεί τόσο η τοπική κίνηση των δακτύλων όσο και η καθολική κίνηση των χεριών. Μια προσέγγιση για τη λύση του παραπάνω προβλήματος είναι μια προσέγγιση με βάση την «εμφάνιση του αντικειμένου» κατά την οποία δισδιάστατα πρότυπα των χεριών χρησιμοποιούνται για τον εντοπισμό του κινούμενου χεριού. Μια άλλη προσέγγιση είναι αυτή, της χρήσης τρισδιάστατων μοντέλων, η οποία πλεονεκτεί στο γεγονός ότι εκμεταλλεύεται την α priori γνώση που υπάρχει στα τρισδιάστατα μοντέλα.

Η διαδικασία που ακολουθήσαμε εμείς για την ανίχνευση των κινούμενων χεριών και / η κεφαλιού βασίστηκε στη χρήση μάσκων με την εφαρμογή των οποίων απομονώνονται, αρχικά οι περιοχές της εικόνας στις οποίες βρίσκονται τα χέρια και το κεφάλι και στη συνέχεια, μόνο τα μέλη τα οποία κινούνται. Δηλαδή με τις μάσκες αυτές εντοπίζονται οι περιοχές οι οποίες είχαν το χαρακτηριστικό χρώμα του δέρματος και κινούνταν από frame σε frame.

Αρχικά δημιουργείται ένας πίνακας, που ουσιαστικά, περιέχει έναν αριθμό, που αντιπροσωπεύει την πιθανότητα που έχει κάθε pixel της εικόνας, να ανήκει σε κάποια δερματική περιοχή (σχήμα 1). Ο πίνακας αυτός δημιουργείται με τον υπολογισμό του λόγου των πινάκων Cr/Cb (τα frame της ακολουθίας βρίσκονται στο χώρο YCbCr). Κατόπιν, με χρήση κάποιου κατωφλίου (οι τιμές που θα έχει ο πίνακας που δημιουργείται θα βρίσκονται, μετά από κάποια διαδικασία κανονικοποίησης, μεταξύ του 0 και του 1. Όσο πιο κοντά στο 1 βρίσκεται μια τιμή, τόσο πιο πιθανό είναι το συγκεκριμένο pixel να ανήκει σε κάποια δερματική περιοχή. Επομένως με την επιλογή ενός κατωφλίου πχ. 0.7 ή 0.8 κρατάμε όλα τα pixel, τα οποία παρουσιάζουν αυξημένη πιθανότητα να ανήκουν σε δερματική περιοχή) , δημιουργείται η μάσκα που απομονώνει το δέρμα (σχήμα 2).



σχήμα 1
Ο πίνακας πιθανότητας, Cr/Cb



σχήμα 2
κατωφλίωση της
προηγούμενης εικόνας με $\rho > 0.8$

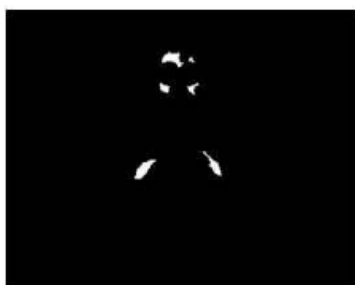
Η μάσκα που εντοπίζει τα τμήματα της εικόνας που πιθανώς κινούνται, δημιουργείται με κατωφλίωση της εικόνας, ανάλογα με τη διαφορά που υπάρχει στις τιμές των pixel από frame σε frame (σχήμα 3). Σημειώνουμε εδώ, πως αυτή η μάσκα κίνηση δεν περιέχει πληροφορίες που αφορούν την κατεύθυνση ή το πλάτος της κίνησης, αλλά απομονώνει απλά τα κινούμενα μέρη της εικόνας ώστε να επιταχύνει

τον αλγόριθμο, με το να επικεντρώνει την διαδικασία του εντοπισμού μόνο στις κινούμενες περιοχές.

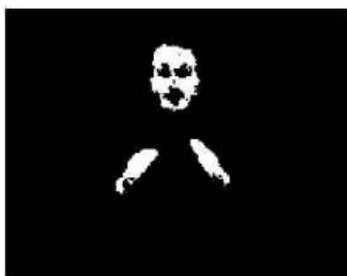


Σχήμα 3
Μάσκα κινούμενων περιοχών

Η εικόνα που προκύπτει μετά την εφαρμογή των δύο αυτών масκών κανονικά πρέπει να έχει άσπρο χρώμα στις δερματικές περιοχές που κινούνται και μαύρο για όλα τα υπόλοιπα αντικείμενα. Παρόλα αυτά, είναι δυνατόν, να υπάρχουν στο background της αρχικής εικόνας, αντικείμενα τα οποία έχουν παρόμοιο χρώμα με αυτό του δέρματος. Αυτά τα αντικείμενα, μαζί με την παρουσία θορύβου, έχουν ως αποτέλεσμα την ύπαρξη άσπρων περιοχών στην φιλτραρισμένη εικόνα οι οποίες δεν αντιπροσωπεύουν κάποια κινούμενη δερματική περιοχή. Για την αντιμετώπιση του παραπάνω προβλήματος γίνεται μορφολογικό φιλτράρισμα στις εικόνες που προκύπτουν μετά την εφαρμογή των δύο масκών. Με το μορφολογικό αυτό φιλτράρισμα απομακρύνονται από τη εικόνα όλα τα στοιχεία, των οποίων το μέγεθος είναι πολύ μικρότερο σε σχέση με το μέγεθος που αναμένεται να έχουν οι κινούμενες δερματικές περιοχές. Κατόπιν τα αντικείμενα τα οποία έχουν απομείνει χρησιμοποιούνται ως δείκτες (σχήμα 4) για την ανακατασκευή της εικόνας (σχήμα 5). Τέλος, γίνεται κλείσιμο της εικόνας, ώστε να επιτευχθεί καλύτερος υπολογισμός των κεντροειδών των κινούμενων δερματικών περιοχών (σχήμα 6).



Σχήμα 4
Δείκτες για την
ανακατασκευή της εικόνας



Σχήμα 5
Ανακατασκευασμένη
εικόνα



Σχήμα 6
Κλείσιμο της εικόνας

Κατόπιν γίνεται η κατασκευή της ολικής μάσκας ανίχνευσης των κινούμενων δερματικών περιοχών. Η μάσκα αυτή δημιουργείται με την μορφολογική ανακατασκευή της μάσκας ανίχνευσης του δέρματος, χρησιμοποιώντας όμως ως δείκτες εκείνους, οι οποίοι έχουν προκύψει από την μάσκα κίνησης. Το αποτέλεσμα της παραπάνω διαδικασίας, μετά την αφαίρεση του κεφαλιού από την εικόνα,

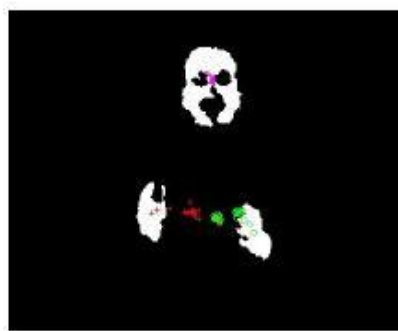
φαίνεται στο σχήμα 7. Η μάσκα των κινούμενων δερματικών περιοχών αποτελείται από πολλές μεγάλες ενιαίες περιοχές (με άσπρο χρώμα). Όταν γίνεται η επεξεργασία του επόμενου frame, δημιουργείται μια νέα ολική μάσκα και γίνεται μια «ένα προς ένα» αντιστοίχιση στα αντικείμενα των εικόνων. Η αντιστοίχιση των αντικειμένων ανάμεσα στα δύο frame, γίνεται στην εικόνα, η οποία έχει προέλθει μετά την εφαρμογή της χρωματικής μάσκας και γίνεται με τον υπολογισμό της απόστασης των κεντροειδών των αντικειμένων, για τα αντικείμενα εκείνα, τα οποία επικαλύπτονται ανάμεσα στα δύο frame σε κάποιο ποσοστό.

Στο σχήμα 8 φαίνονται οι τροχιές των κεντροειδών των δύο χεριών και του κεφαλιού κατά την χειρονομία του «χειροκρότηματος». Το κεντροειδές του δεξιού χεριού του χρήστη απεικονίζεται με ένα κόκκινο σταυρό ενώ το κεντροειδές του αριστερού με ένα πράσινο κυκλάκι. Το κεντροειδές του κεφαλιού απεικονίζεται με μια μωβ τελεία.



Σχήμα 7

Τα χέρια που κινούνται, όπως φαίνονται μετά την μορφολογική ανακατασκευή της εικόνας



Σχήμα 8

Η τροχιά των κεντροειδών των χεριών και του κεφαλιού κατά το «χειροκρότημα».

Ο αλγόριθμος που περιγράφηκε για τον εντοπισμό των κεντροειδών των χεριών και του κεφαλιού είναι σχετικά μικρής υπολογιστικής πολυπλοκότητας και μας δίνει αρκετά ικανοποιητικά αποτελέσματα σχετικά γρήγορα (οι ταχύτητα του είναι ικανοποιητική και για real time εφαρμογές). Η α priori γνώση της θέσης του χρήστη μπροστά στον υπολογιστή και των χειρονομιών που πρόκειται να αναγνωριστούν από το μοντέλο, υπήρξε μια σημαντική βοήθεια, ώστε, αφενός να μειωθεί ο υπολογιστικός φόρτος που προέρχεται από τον αλγόριθμο εντοπισμού των κινούμενων δερματικών περιοχών και αφετέρου να επιτυγχάνεται πιο ακριβής εντοπισμός. Για παράδειγμα, γνωρίζουμε ότι το κεφάλι βρίσκεται στο άνω μισό της εικόνας και, συνήθως, στο μέσο αυτής, όσον αφορά την οριζόντια διάσταση. Έτσι το αντικείμενο που εντοπίζεται σε κάθε frame μέσα σε αυτό το πλαίσιο θεωρούμε πως είναι το κεφάλι. Επίσης, γνωρίζουμε πως η αρχική θέση των χεριών είναι στα δύο κάτω άκρα της εικόνας, καθώς επίσης, και ότι κατά τη διάρκεια όλων των χειρονομιών, το δεξί χέρι παραμένει εντός του αριστερού μισού της εικόνας και το αριστερό εντός του δεξιού μισού της εικόνας (όσον αφορά την οριζόντια διάσταση). Με τον τρόπο αυτό αποφεύγονται τα λάθη στον εντοπισμό, σε χειρονομίες όπως πχ. το χειροκρότημα, όπου για κάποια δέκατα του δευτερολέπτου τα δύο χέρια ενώνονται και κατόπιν χωρίζονται πάλι. Γνωρίζοντας ότι το αριστερό χέρι θα βρίσκεται πάντα πιο δεξιά από το δεξί χέρι του χρήστη, μπορούμε γρήγορα και σωστά

να τοποθετήσουμε ξανά τα κεντροειδή των δύο χεριών, μόλις αυτά αρχίσουν και πάλι να διακρίνονται ως δύο διαφορετικά αντικείμενα.

Πρέπει, τέλος, να διευκρινίσουμε τον τρόπο λειτουργίας του παραπάνω αλγόριθμου σε ορισμένες οριακές περιπτώσεις. Όταν δύο δερματικές περιοχές επικαλύπτονται (πχ. όταν ενώνονται τα δύο χέρια ή όταν το χέρι περνάει μπροστά από το κεφάλι) τότε δεν είναι δυνατόν να εντοπιστούν τα κεντροειδή κανενός από τα δύο αντικείμενα. Το γεγονός αυτό, μπορεί να έχει πολύ δυσμενή αποτελέσματα στη σωστή δημιουργία των διανυσμάτων με τα χαρακτηριστικά γνωρίσματα των χειρονομιών, καθώς είναι δυνατόν να μη γίνει καθόλου εντοπισμός ή να γίνει λανθασμένος εντοπισμός, των αντικειμένων για ορισμένα frame. Δεδομένου ότι τα χαρακτηριστικά που χρησιμοποιήσαμε για την εκπαίδευση των HMM βασίζονται στην τροχιά και στις συντεταγμένες των αντικειμένων σε κάθε frame, γίνονται κατανοητά τα προβλήματα που δημιουργούνται όταν λείπουν μερικά από τα σημεία της τροχιάς. Για παράδειγμα, αν στην χειρονομία “attention” (σήκωμα του δεξιού χεριού), ο χρήστης σηκώσει το χέρι του με τέτοιο τρόπο ώστε να η τροχιά του χεριού να περνάει μπροστά από το πρόσωπο, τότε οι θέσεις του χεριού στα ενδιάμεσα frame δεν θα εντοπιστούν και θα εντοπιστούν μόνο οι θέσεις του χεριού στα frame που αυτό δεν περνούσε μπροστά από το πρόσωπο, δηλαδή κατά την αρχή και το τέλος της χειρονομίας.

Επειδή η σωστή εξαγωγή των χαρακτηριστικών γνωρισμάτων, είναι ιδιαίτερα σημαντική για τη δημιουργία ενός συστήματος που θα αναγνωρίζει με ακρίβεια της διάφορες χειρονομίες, πρέπει να δοθεί ιδιαίτερη βαρύτητα στην επίλυση του παραπάνω προβλήματος. Ένας πρώτος τρόπος, είναι να τεθούν περιορισμοί στον τρόπο με τον οποίο θα πραγματοποιεί τις χειρονομίες ο χρήστης. Με τον τρόπο αυτό θα εξασφαλιστεί ότι οι χειρονομίες θα εκτελούνται με τρόπο σαφή και με τρόπο ο οποίος δεν θα δημιουργεί προβλήματα στην διαδικασία εντοπισμού. Βεβαίως η ευχρηστία του συστήματος περιορίζεται με αυτό τον τρόπο αλλά όχι σε τέτοιο βαθμό που να αντισταθμίζει το πλεονέκτημα της ακριβούς αναγνώρισης που αποκτάμε, θέτοντας αυτούς τους περιορισμούς. Ένας δεύτερος τρόπος, αναπτύχθηκε για την αντιμετώπιση των προβλημάτων τα οποία δημιουργήθηκαν, κατά την εκπαίδευση των συνεχών HMM. Το toolbox που χρησιμοποιήθηκε, απαιτούσε ίδιο μήκος για τις ακολουθίες εκπαίδευσης της κάθε χειρονομίας. Όταν λοιπόν σε κάποια ακολουθία, έλειπαν ορισμένα σημεία από το μέσο της τροχιάς, λόγω του παραπάνω προβλήματος, δεν ήταν δυνατόν να γίνει σωστά η εκπαίδευση. Για το λόγο αυτό έπρεπε όλες οι ακολουθίες να διαμορφωθούν με τρόπο τέτοιο ώστε να έχουν το ίδιο μήκος. Επιστρέφοντας στο προηγούμενο παράδειγμα της χειρονομίας “attention”, αυτό επιτεύχθηκε με γραμμική παρεμβολή των σημείων τα οποία δεν είχε καταφέρει να αναγνωρίσει ο αλγόριθμος εντοπισμού. Αυτή η λύση εφαρμόστηκε, χρησιμοποιώντας την a priori γνώση που είχαμε για τις ακολουθίες εκπαίδευσης (από ποια χειρονομία προέρχονταν και την πραγματική τροχιά που είχε ακολουθηθεί και ο αλγόριθμος εντοπισμού δεν είχε καταφέρει να ανιχνεύσει). Βεβαίως κάτι αντίστοιχο δεν μπορούσε να γίνει στις ακολουθίες των χειρονομιών που έρχονταν προς αναγνώριση, αφού προφανώς, δεν υπήρχε η γνώση από ποια χειρονομία προέρχονταν, προτού αναγνωριστούν. Παρόλα αυτά, τα αποτελέσματα των πειραμάτων έδειξαν, πως τα ποσοστά αναγνώρισης ήταν υψηλά, ακόμα και σε ακολουθίες οι οποίες παρουσίαζαν το παραπάνω πρόβλημα και είχαν εν γένει διαφορετικό μήκος από το μήκος που είχαν οι ακολουθίες που χρησιμοποιήθηκαν για την εκπαίδευση των αντίστοιχων HMM.

Μία δεύτερη ιδιαιτερότητα που έχει να κάνει με αλγόριθμο εντοπισμού, είναι ο τρόπος με τον οποίον αντιμετωπίζει τα ακίνητα αντικείμενα. Όπως εξηγήσαμε

προηγούμενως, εφαρμόζεται αρχικά μια χρωματική μάσκα και κατόπιν μια μάσκα κίνηση. Με τον τρόπο αυτό εντοπίζονται τα κεντροειδή των δερματικών περιοχών οι οποίες κινούνται. Όταν λοιπόν πχ. το δεξί χέρι είναι ακίνητο, ο πίνακας με τις συντεταγμένες του θα είναι αρχικοποιημένος και θα παραμείνει στην τιμή 0 για όλα τα frame, κάτι το οποίο είναι θεμιτό. Ας υποθέσουμε τώρα πως ένα από τα χαρακτηριστικά που χρησιμοποιούμε για την εκπαίδευση των μοντέλων είναι η απόσταση άμεσα στο δεξί χέρι και το κεφάλι. Αν λοιπόν το κεφάλι κάνει μικρές κινήσεις (όπως συμβαίνει στις περισσότερες χειρονομίες που αναγνωρίζει το σύστημά μας), και το δεξί χέρι τυχαίνει στην συγκεκριμένη χειρονομία να παραμένει ακίνητο, η απόσταση τους θα πρέπει να είναι περίπου σταθερή και ίση με τις συντεταγμένες του κεφαλιού σε κάθε frame. Υπάρχει όμως η περίπτωση ο χρήστης να κινήσει ελαφρώς το δεξί του χέρι σε κάποιο frame, χωρίς βεβαίως η κίνηση αυτή να σχετίζεται με την χειρονομία την οποία πραγματοποιεί. Αυτό το οποίο θα συμβεί στη συγκεκριμένη περίπτωση, είναι ότι το σύστημα θα εντοπίσει το δεξί χέρι, στο συγκεκριμένο frame που το κίνησε ο χρήστης, και θα θέσει τις συντεταγμένες του στον πίνακα συντεταγμένων, ο οποίος σε όλα τα άλλα frame, που το χέρι παρέμενε ακίνητο, είχε μόνο μηδενικά στοιχεία. Έτσι η απόσταση του δεξιού χεριού από το κεφάλι θα παρουσιάσει μια ξαφνική διακύμανση. Η διακύμανση αυτή είναι δυνατόν να δημιουργήσει προβλήματα κατά την διαδικασία αναγνώρισης, αφού γίνεται σε τυχαίο χρονικό σημείο και το εύρος της κίνησης είναι επίσης τυχαίο και δεν σχετίζεται με την πραγματοποιούμενη χειρονομία. Ένας τρόπος αντιμετώπισης αυτού του προβλήματος είναι ο, εκ νέου, περιορισμός και η θέσπιση αυστηρότερων ορίων στις κινήσεις του χρήστη. Ένας δεύτερος τρόπος αντιμετώπισης του προβλήματος είναι να θέσουμε πιο αυστηρή μάσκα κίνησης, δηλαδή να απαιτούμε μεγαλύτερη κίνηση των αντικειμένων από frame σε frame, προκειμένου να εντοπιστούν. Με τον τρόπο αυτό οι τυχαίες μικροκινήσεις μελών που δεν συμμετέχουν στην πραγματοποιούμενη χειρονομία, αγνοούνται. Σε περίπτωση εφαρμογής αυτής της λύσεως απαιτείται ιδιαίτερη προσοχή, καθώς σε περίπτωση που τεθεί πολύ αυστηρή μάσκα κίνησης, είναι δυνατόν να αγνοηθούν και κινήσεις που δεν επιθυμούμε να αγνοήσουμε.

Από τα παραπάνω καταλαβαίνουμε πόσο δύσκολο είναι να επιτευχθεί απόλυτα ακριβής εντοπισμός των κινούμενων μελών. Άλλα προβλήματα που μπορεί να παρουσιαστούν έχουν να κάνουν με το χρώμα που μπορεί να έχει το δέρμα σε κάθε ακολουθία frames και με την ακριβή θέση του χρήστη σε κάθε ακολουθία.

Όσον αφορά το χρώμα του δέρματος, είναι δυνατόν, λόγω του διαφορετικού φωτισμού και των διαφορετικών αποχρώσεων του δέρματος του κάθε χρήστη, η εφαρμογή της χρωματικής μάσκας να οδηγήσει σε ανεπιθύμητα αποτελέσματα και να μην γίνει σωστά ο εντοπισμός των δερματικών περιοχών. Ένας τρόπος αντιμετώπισης του προβλήματος αυτού είναι η μείωση της πιθανότητας που χρησιμοποιείται για την κατωφλίωση της αρχικής εικόνας. Ουσιαστικά με τον τρόπο αυτό εφαρμόζουμε μια πιο χαλαρή χρωματική μάσκα. Το πρόβλημα το οποίο είναι δυνατόν να προκύψει με τον τρόπο αυτό, είναι ότι όσο πιο χαλαρή είναι η χρωματική μάσκα, τόσο πιο πιθανό είναι να θεωρηθούν ως δερματικές περιοχές, αντικείμενα του background τα οποία τυχαίνει να έχουν παραπλήσιο χρώμα με αυτό του δέρματος. Τα αντικείμενα αυτά θα πρέπει να αφαιρεθούν στη συνέχεια κατά την μορφολογική ανακατασκευή της εικόνας, κάτι το οποίο, όμως, δεν είναι πάντα εφικτό. Ένας δεύτερος τρόπος, είναι η αλλαγή των παραμέτρων της συνάρτησης με τον οποίο γίνεται η κανονικοποίηση του πίνακα που περιέχει τις πιθανότητες κάθε pixel να ανήκει σε κάποια δερματική περιοχή (πίνακας Cr/Cb). Με την αλλαγή των παραμέτρων αυτών, επιτυγχάνεται ουσιαστικά η μετατόπιση της χρωματικής περιοχής του δέρματος έτσι ώστε να

προσαρμοστεί στην περιοχή που εκτείνεται το χρώμα του δέρματος στην δεδομένη ακολουθία.

Τέλος, όσον αφορά την θέση του χρήστη στην εικόνα τα προβλήματα που μπορεί να προκύψουν έχουν να κάνουν με το γεγονός πως, όπως προαναφέρθη, στον αλγόριθμο εντοπισμού χρησιμοποιείται η *a priori* γνώση της θέσης του προσώπου και των χεριών στην εικόνα. Όταν, λοιπόν, ο χρήστης κάτσει απέναντι από τον υπολογιστή σε θέση που δεν ικανοποιούνται οι χωρικοί περιορισμοί που έχουν τεθεί στον αλγόριθμο εντοπισμού (πχ. η το κεφάλι αναμένεται να βρεθεί στο άνω μισό και στη μέση της εικόνας), δεν θα υπάρξει σωστή αναγνώριση. Για την αντιμετώπιση αυτού του φαινομένου θα πρέπει ή η θέση του χρήστη να είναι αυστηρώς ορισμένη ή να αλλάζουν οι χωρικοί περιορισμοί ανάλογα με την τρέχουσα θέση του χρήστη.

Η σωστή ανίχνευση της θέσης των κινούμενων μελών και η εξαγωγή των χαρακτηριστικών γνωρισμάτων των χειρονομιών με ακρίβεια, είναι πολύ σημαντικοί παράγοντες και αποτελούν την βάση για την ανάπτυξη ενός ακριβούς συστήματος αναγνώρισης χειρονομιών. Όταν ο εντοπισμός αυτός δεν γίνεται σωστά, τότε δεν θα είναι πετυχημένη ούτε η διαδικασία εκπαίδευσης, ούτε η διαδικασία αναγνώρισης, αφού τα χαρακτηριστικά γνωρίσματα των χειρονομιών με τα οποία θα τροφοδοτούνται τα HMM, θα είναι αλλοιωμένα. Δεδομένων των προβλημάτων που είναι δυνατόν να υπάρξουν κατά την διαδικασία εντοπισμού και τα οποία περιγράφηκαν παραπάνω, καταλαβαίνουμε πόσο σημαντικό είναι να γίνει σωστή ρύθμιση των παραμέτρων του αλγορίθμου και ενημέρωση του χρήστη για τους περιορισμούς που πρέπει να «σεβαστεί» κατά την εκτέλεση των χειρονομιών, προκειμένου να έχουμε ικανοποιητικά αποτελέσματα.

6.3 Επιλογή των Χαρακτηριστικών Γνωρισμάτων

Πριν περιγράψουμε τα χαρακτηριστικά που επιλέξαμε να χρησιμοποιήσουμε για την εκπαίδευση των HMM, είναι χρήσιμο να γίνει μια ανάλυση των, προς αναγνώριση, χειρονομιών, ώστε να αναδειχθούν τα ιδιαίτερα χαρακτηριστικά της καθεμιάς και να εξηγηθεί, πώς μπορούν αυτά τα χαρακτηριστικά, να χρησιμοποιηθούν, ώστε να υπάρξει επιτυχημένος διαχωρισμός των χειρονομιών αυτών.

Όπως προαναφέραμε, οι χειρονομίες που αναγνωρίζονται από το σύστημα που υλοποιήθηκε είναι επτά. Αναλυτικότερα, είναι : η “attention” (σήκωμα του δεξιού χεριού μέχρι το ύψος το κεφαλιού), η “wave” (χαιρετισμός με το δεξί χέρι), η “explain” (σχεδόν περιοδικές, κυκλικές κινήσεις και των δύο χεριών, στο ύψος σχεδόν του στήθους), η “oh my god” (κράτημα του κεφαλιού με τα δύο χέρια), η “clapping” (χειροκρότημα), η “one hand” (στήριγμα του κεφαλιού με το δεξί χέρι) και η “go away” (τίναγμα προς τα πάνω και δεξιά του δεξιού χεριού).

Όλες οι χειρονομίες ξεκινάνε από το ίδιο σημείο, αφού ο χρήστης έχει ενωμένα τα χέρια του και ακουμπισμένα στο γραφείο που βρίσκεται ακριβώς εμπρός του. Με μια εποπτική ματιά, διαπιστώνουμε πως μερικές από τις παραπάνω χειρονομίες έχουν χαρακτηριστικά γνωρίσματα με παραπλήσιες τιμές (πχ. η wave, μέχρι ενός σημείου, είναι σχεδόν ίδια με την attention, αφού και στις δύο έχουμε σήκωμα το δεξιού χεριού μέχρι το ύψος του κεφαλιού περίπου). Δεδομένου ότι τα γνωρίσματα αυτά χρησιμοποιούνται για την εκπαίδευση των HMM, υπάρχει η πιθανότητα μειωμένης ακρίβειας κατά την αναγνώριση τέτοιων χειρονομιών. Αν για παράδειγμα, η προς αναγνώριση ακολουθία προέρχεται από μία χειρονομία “attention”, τότε είναι πιθανόν δύο μοντέλα, ίσως και τρία (της wave, της attention

και της “go away”), να μας δώσουν αυξημένη πιθανότητα η προς αναγνώριση ακολουθία να έχει προέλθει από αυτά.

Μία δεύτερη διαπίστωση είναι ότι υπάρχουν κάποιες χειρονομίες, οι οποίες παρουσιάζουν μεγαλύτερη στοχαστικότητα από τις υπόλοιπες. Αυτές είναι η “explain”, η “wave” και η “clapping”. Αυτές οι χειρονομίες είναι δύσκολο να επαναλαμβάνονται με παρόμοιο τρόπο από τους χρήστες, αφού είναι κινήσεις πιο σύνθετες και ταυτόχρονα με λιγότερο τυποποιημένη μορφή, από το απλό σήκωμα του χεριού. Για τους λόγους αυτούς, αφενός μεν είναι δύσκολη η εκπαίδευση των HMM και απαιτεί μεγάλο πλήθος δεδομένων εκπαίδευσης, ώστε το μοντέλο που θα δημιουργηθεί να μπορεί να αναγνωρίζει τέτοιου είδους χειρονομίες από οποιοδήποτε χρήστη και αφετέρου η πιθανότητα που θα προκύπτει κατά την διαδικασία της ταξινόμησης θα είναι πολύ μικρή. Αν για παράδειγμα η προς αναγνώριση ακολουθία προέρχεται από μια “wave” το HMM που μοντελοποιεί την “wave” θα δώσει πολύ μικρή πιθανότητα να έχει προέλθει από αυτό. Αυτό οφείλεται, όπως είπαμε στον σύνθετο χαρακτήρα της χειρονομίας και στην μεγάλη στοχαστικότητα που την διέπει. Τα υπόλοιπα όμως μοντέλα θα δώσουν πολλές τάξεις μεγέθους μικρότερες πιθανότητες, καθότι η “wave” είναι ξεχωριστή και δεν μπορεί να ταυτιστεί με καμία από της υπόλοιπες χειρονομίες. Αντιστοίχως, παρόμοια προβλήματα παρουσιάζονται τόσο στην “clapping” όσο και στην “explain”. Όμως και αυτές οι χειρονομίες είναι αρκετά διαφορετικές από όλες τις υπόλοιπες, οπότε μπορεί μεν τα HMM που τις μοντελοποιούν, κατά την διαδικασία της αποκωδικοποίησης μιας τέτοιας ακολουθίας να δώσουν πολύ μικρές πιθανότητες, τα HMM όμως που μοντελοποιούν τις υπόλοιπες χειρονομίες θα δώσουν ακόμα μικρότερες πιθανότητες, οπότε η αναγνώριση αυτών των χειρονομιών θα είναι επιτυχημένη.

Τα σημαντικότερα χαρακτηριστικά που χρησιμοποιούνται για την αναγνώριση χειρονομίας είναι τα πρότυπα, οι σφαιρικοί μετασχηματισμοί, οι ζώνες, και τα γεωμετρικά χαρακτηριστικά. Μεταξύ αυτών των χαρακτηριστικών, τα πρότυπα είναι τα απλούστερα κατά τον υπολογισμό τους, καθώς είναι τα ίδια τα δεδομένα εισόδου σε ακατέργαστη μορφή. Για μια τροχιά, ένα πρότυπο αποτελείται από τις συντεταγμένες των σημείων που αποτελούν την τροχιά. Το βασικότερο πλεονέκτημα των προτύπων είναι, ότι τα χαρακτηριστικά γνωρίσματα είναι εύκολο να υπολογιστούν. Το σημαντικότερο μειονέκτημα τους είναι, ότι το μέγεθος των δεδομένων αυξάνεται καθώς αυξάνεται το μέγεθος της εισόδου, καθιστώντας τα χαρακτηριστικά αυτά ακατάλληλα ως είσοδο, σε ορισμένα είδη συστημάτων αναγνώρισης.

Οι σφαιρικοί μετασχηματισμοί, (πχ.η περιστροφή) μπορούν να υιοθετηθούν για να μειώσουν τον αριθμό των χαρακτηριστικών γνωρισμάτων στα πρότυπα. Οι μετασχηματισμοί επιλέγονται συχνά με τέτοιο τρόπο, ώστε να παραμένουν αναλλοίωτοι κατά τη περιστροφή των δεδομένων εισόδου. Παραδείγματος χάριν, ο μετασχηματισμός Φουριέ μπορεί να οδηγήσει στη σταθερότητα των χαρακτηριστικών κατά την περιστροφή δεδομένων εισόδου. Οι σφαιρικοί μετασχηματισμοί παράγουν γενικά έναν σταθερό αριθμό χαρακτηριστικών, γεγονός το οποίο παρέχει μεγαλύτερη ευελιξία στην επιλογή του συστήματος αναγνώρισης. Ο υπολογισμός του μετασχηματισμού, όμως, κοστίζει σε υπολογιστική ισχύ, και τα προκύπτοντα χαρακτηριστικά είναι συνήθως δύσκολο να ερμηνευθούν άμεσα.

Στην εξαγωγή χαρακτηριστικών με το χωρισμό της εικόνας σε ζώνες, το διάστημα μιας πορείας διαιρείται σε έναν αριθμό από ζώνες, και η πορεία μετασχηματίζεται στην ακολουθία ζωνών τις οποίες διαπερνά. Είναι επίσης δυνατό να κωδικοποιηθεί και η κατεύθυνση από την οποία το αντικείμενο εισχωρεί σε κάθε ζώνη. Τα σημαντικότερα πλεονεκτήματα της τακτικής αυτής είναι η απλότητα και η

αποδοτικότητα της. Τα μειονεκτήματα είναι παρόμοια με εκείνα για τα χαρακτηριστικά γνωρίσματα προτύπων.

Τα γεωμετρικά χαρακτηριστικά μιας διαδρομής, όπως το συνολικό μήκος της, η συνολική γωνιακή μετατόπιση, ο αριθμός των διασταυρώσεων, κ.λπ., μπορούν να χρησιμοποιηθούν για να απεικονίσουν τις γενικές ιδιότητες της πορείας. Γεωμετρικά χαρακτηριστικά χρησιμοποιούνται ευρέως στην αναγνώριση γραφής και φυσικά στην αναγνώριση χειρονομίας. Τα γεωμετρικά χαρακτηριστικά μπορούν να φέρουν ποικίλες χρήσιμες πληροφορίες, όπως το συνολικό μήκος της διαδρομής. Επίσης, μπορούν να εφαρμοστούν σε διάφορα συστήματα αναγνώρισης που απαιτούν έναν σταθερό αριθμό χαρακτηριστικών γνωρισμάτων, ή σε συστήματα που αναμένουν μια ακολουθία χαρακτηριστικών. Τα γεωμετρικά χαρακτηριστικά τείνουν να είναι πιο σύνθετα στον υπολογισμό από τους άλλους τύπους χαρακτηριστικών που αναφέρονται παραπάνω.

Τα χαρακτηριστικά τα οποία επιλέξαμε να χρησιμοποιήσουμε στο σύστημα αναγνώρισης χειρονομιών το οποίο υλοποιήσαμε, είναι κυρίως της πρώτης κατηγορίας, δηλαδή της κατηγορίας των προτύπων, καθώς αφορούν την τροχιά της κίνησης. Σημειώνουμε εδώ, πως για την εκπαίδευση των διακριτών HMM, χρησιμοποιήσαμε και μία παραλλαγή της τεχνικής του χωρισμού σε ζώνες, η οποία θα αναλυθεί στο κεφάλαιο, στο οποίο περιγράφεται η υλοποίηση με βάση τα διακριτά HMM. Όταν αναφερόμαστε σε «τροχιά», στις συγκεκριμένες χειρονομίες τις οποίες καλείται να αναγνωρίσει το σύστημά μας, αναφερόμαστε στην τροχιά του κεφαλιού και των δύο χεριών, αφού αυτά είναι τα μοναδικά μέλη που βρίσκονται σε κίνηση και στις επτά χειρονομίες. Τα χαρακτηριστικά, λοιπόν, τα οποία θα προκύψουν είναι οι συντεταγμένες των χεριών και του κεφαλιού σε κάθε frame. Θα έχουμε δηλαδή έξι αριθμούς (τετμημένη και τεταγμένη του καθενός από τα τρία σημεία), με τους οποίους θα καθορίζονται πλήρως οι θέσεις που θα έχουν στην εικόνα, σε κάθε frame, το κεφάλι και το καθένα από τα χέρια.

Η χρήση όμως απλά αυτών των έξι αριθμών για την εκπαίδευση των HMM δεν θα είχε ικανοποιητικά αποτελέσματα, αφού οι συντεταγμένες των χεριών και του κεφαλιού σε κάθε ακολουθία βίντεο, εξαρτώνται από την θέση που έχει ο χρήστης στην εικόνα. Εάν λοιπόν, τα HMM είχαν εκπαιδευθεί από ακολουθίες στις οποίες οι χρήστες κάθονταν κατά κύριο λόγο στην μέση της εικόνας (οπότε για παράδειγμα η αρχική θέση του κεφαλιού θα ήταν (200,78)) και στην προς αναγνώριση ακολουθία, ο νέος χρήστης, βρισκόταν μετατοπισμένος από το κέντρο της εικόνας, προς κάποια κατεύθυνση (είτε οριζόντια είτε κάθετη), γίνεται κατανοητό πως οι συντεταγμένες των μελών του θα ήταν αρκετά διαφορετικές απ' ότι στις ακολουθίες εκπαίδευσης και έτσι θα οδηγούμασταν σε λανθασμένα αποτελέσματα κατά την αναγνώριση.

Για το λόγο αυτό, τα χαρακτηριστικά, τα οποία αποφασίσαμε να χρησιμοποιήσουμε για την εκπαίδευση των HMM αποτελούνται από τις αποστάσεις, και πιο συγκεκριμένα, από τις διαφορές των τετμημένων και των τεταγμένων των δύο χεριών και του κεφαλιού. Με τον τρόπο αυτό εξασφαλίζεται η ανεξαρτησία των χαρακτηριστικών από την σχετική θέση του χρήστη μέσα στην εικόνα. Πιο συγκεκριμένα, σε κάθε frame, εντοπίζονται, με τον αλγόριθμο που περιγράψαμε στην προηγούμενη ενότητα, οι θέσεις των κεντροειδών των χεριών και του κεφαλιού. Οι τιμές αυτές αποθηκεύονται και, κατόπιν, δημιουργείται ένας πίνακας (με τόσες στήλες όσες είναι τα frame της ακολουθίας και τόσες γραμμές όσες είναι τα χαρακτηριστικά) με τα επιθυμητά χαρακτηριστικά. Τα χαρακτηριστικά αυτά αναλυτικότερα είναι :

$$X_L-X_R, X_H-X_R, X_H-X_L, Y_L-Y_R, Y_H-Y_R, Y_H-Y_L$$

Όπου X_L είναι η τεταγμένη του αριστερού χεριού ,
 X_R είναι η τεταγμένη του δεξιού χεριού,
 X_H είναι η τεταγμένη του κεφαλιού,
 Y_L είναι η τετμημένη του αριστερού χεριού ,
 Y_R είναι η τετμημένη του δεξιού χεριού,
 Y_H είναι η τετμημένη του κεφαλιού.

Μπορούμε εύκολα να παρατηρήσουμε πως καθεμιά από τις παραπάνω χειρονομίες, παρουσιάζει ένα τουλάχιστον χαρακτηριστικό, το οποίο την κάνει να ξεχωρίζει από όλες τις υπόλοιπες. Πιο συγκεκριμένα, η “attention” έχει ως βασικό χαρακτηριστικό την ανύψωση του δεξιού χεριού. Η ανύψωση του δεξιού χεριού κάνει την “attention” να ξεχωρίζει από την “clapping”, την “explain”, την “wave” και την “one hand”, αφού, οι μεν δύο πρώτες δεν περιλαμβάνουν καν ανύψωση του δεξιού χεριού, ενώ στην “wave” η ανύψωση του δεξιού χεριού ακολουθείται από άλλη κίνηση αλλά και στην “one hand” η τελική θέση του χεριού είναι σε ύψος πολύ μικρότερο από αυτό που έχει το χέρι, στην τελική του θέση, στην “attention”.

Όμως, ανύψωση του δεξιού χεριού, με τρόπο παρόμοιο, με αυτόν που γίνεται στην “attention”, περιλαμβάνουν τόσο η “oh my god” όσο και η “go away” . Για να επιτευχθεί, λοιπόν, ο διαχωρισμός της “attention” από τις δύο αυτές χειρονομίες, είναι απαραίτητο να χρησιμοποιηθούν και άλλα χαρακτηριστικά εκτός από αυτό του σηκώματος του δεξιού χεριού. Ένα χαρακτηριστικό το οποίο ξεχωρίζει την “attention” από την “oh my god” είναι η τροχιά του αριστερού χεριού, το οποίο στην μεν “attention” παραμένει ακίνητο, πάνω στο τραπέζι, στη δε “oh my god” σηκώνεται με την ίδια ταχύτητα με το δεξί χέρι.

Ο διαχωρισμός της “attention” από την “go away” δεν είναι τόσο ξεκάθαρος καθώς οι δύο χειρονομίες είναι από τη φύση τους παρόμοιες. Παρόλα αυτά, εάν τα HMM εκπαιδευτούν με ικανοποιητικό (πλήθος) ακολουθιών εκπαίδευσης , είναι δυνατόν να επιτευχθεί ικανοποιητική αναγνώριση, η οποία θα βασίζεται στις μικροδιαφορές που παρουσιάζουν οι δύο χειρονομίες. Μία από αυτές είναι το γεγονός πως, ενώ στην “attention”, το δεξί χέρι ανυψώνεται σχεδόν κάθετα, στην “ go away” το δεξί χέρι σηκώνεται ακολουθώντας μια λοξή τροχιά, απομακρυνόμενο ελαφρώς από το κεφάλι.

Από τα παραπάνω, συμπεραίνουμε πως για το διαχωρισμό της “attention” από τις υπόλοιπες χειρονομίες, είναι απαραίτητο να συμπεριληφθούν στα χαρακτηριστικά, οι τροχιές του δεξιού και του αριστερού χεριού. Αυτό επιτυγχάνεται με την χρήση του χαρακτηριστικού της διαφοράς των τεταγμένων των δύο χεριών X_L-X_R . Όπως εξηγήσαμε , ο αλγόριθμος εντοπισμού, όταν κάποιο αντικείμενο δεν κινείται καθόλου κατά την εξέλιξη μιας χειρονομίας, θεωρεί πως έχει συντεταγμένες (0,0) καθ’ όλη την διάρκεια της χειρονομίας. Έτσι, η παραπάνω διαφορά, στις χειρονομίες που έχουμε κίνηση μόνο του δεξιού χεριού , θα ισούται με την τεταγμένη του δεξιού χεριού, ενώ στις υπόλοιπες θα ισούται με την διαφορά των τεταγμένων των χεριών (η οποία, για παράδειγμα, στην “oh my god” θα πρέπει να είναι περίπου 0 , αφού τα δύο χέρια ανεβαίνουν σχεδόν ταυτόχρονα).

Όσον αφορά την “oh my god” το χαρακτηριστικό της γνώρισμα είναι ότι η διαφορά των τεταγμένων των χεριών είναι καθ’ όλη τη διάρκεια της χειρονομίας, κοντά στο 0, αφού έχουμε ανύψωση τόσο του δεξιού όσο και του αριστερού χεριού. Αυτό το χαρακτηριστικό την κάνει να ξεχωρίζει εύκολα από όλες τις χειρονομίες στις

οποιές δεν έχουμε κίνηση του αριστερού χεριού. Στην “clapping”, που έχουμε κίνηση και των δύο χεριών, η διαφορά των τεταγμένων των δύο χεριών είναι επίσης σχεδόν στο 0, αφού κατά το χειροκρότημα, τα δύο χέρια βρίσκονται στο ίδιο ύψος. Το χαρακτηριστικό το οποίο χρησιμοποιούμε για να διαχωρίσουμε την “oh my god” από την “clapping” είναι η διαφορά των τετμημένων των χεριών, η οποία στη μεν “oh my god” είναι σταθερή, αφού τα χέρια εκτελούν κατακόρυφη κίνηση και δεν αλλάζει η απόσταση που έχουν κατά τον οριζόντιο άξονα, ενώ στην “clapping” η διαφορά αυτή αυξομειώνεται με τρόπο σχεδόν περιοδικό (η οριζόντια απόσταση μηδενίζεται όταν τα δύο χέρια έρχονται σε επαφή ενώ αυξάνεται μέχρι ενός σημείου όταν τα δύο χέρια απομακρύνονται). Επίσης, ένα ακόμα ιδιαίτερο χαρακτηριστικό της “oh my god” είναι το γεγονός ότι τελικά μηδενίζονται οι αποστάσεις (τόσο κατά τον οριζόντιο όσο και κατά τον κατακόρυφο άξονα) των χεριών από το κεφάλι. Αυτό συμβαίνει διότι στο τέλος της κίνησης, τα χέρια ακουμπούν στο κεφάλι. Το γεγονός αυτό, έχει σαν αποτέλεσμα ο αλγόριθμος εντοπισμού να θεωρεί τα χέρια και το κεφάλι ως ένα, ενιαίο, αντικείμενο και για το λόγο αυτό υπολογίζει μόνο ένα κεντροειδές και για τα τρία αντικείμενα, μηδενίζοντας με αυτό τον τρόπο τις διαφορές των αποστάσεών τους. Επομένως οι διαφορές που είναι απαραίτητες για τον διαχωρισμό της “oh my god” και της “clapping” από τις υπόλοιπες είναι οι $Y_L - Y_R$ και $X_L - X_R$ αλλά και οι αποστάσεις (οριζόντιες και κάθετες) των χεριών από το κεφάλι δηλαδή $X_H - X_R$, $X_H - X_L$, $Y_H - Y_R$, $Y_H - Y_L$.

Η “explain” μπορεί εύκολα να διαχωριστεί από τις περισσότερες από τις υπόλοιπες χειρονομίες, αφού αφενός μεν έχουμε κίνηση και των δύο χεριών και αφετέρου, το δεξί χέρι κινείται σε ύψος αρκετά μικρότερο από το ύψος στο οποίο ολοκληρώνονται οι χειρονομίες που περιλαμβάνουν ανύψωση του δεξιού χεριού. Η χειρονομία, η οποία εμφανίζει ορισμένες ομοιότητες με την “explain” είναι η “clapping”. Τα παρόμοια χαρακτηριστικά των δύο αυτών χειρονομιών είναι α) ότι η διαφορά των τεταγμένων των χεριών στις δύο χειρονομίες είναι σχεδόν 0, αφού τα χέρια βρίσκονται σχεδόν στο ίδιο ύψος κατά την εξέλιξη των χειρονομιών και β) ότι η διαφορά των τετμημένων των δύο χεριών παρουσιάζει μια σχεδόν περιοδική αυξομείωση. Ένα χαρακτηριστικό το οποίο βοηθάει στο διαχωρισμό των δύο αυτών χειρονομιών, είναι το γεγονός ότι στην “explain” η διαφορά των τεταγμένων των χεριών από το κεφάλι παρουσιάζει και αυτή μια, σχεδόν, περιοδική, αυξομείωση ενώ στην “clapping” – υπό την προϋπόθεση ότι ο χρήστης χειροκροτεί έχοντας τα χέρια του οριζόντια – η διαφορά αυτή παραμένει σταθερή. Επίσης, οι χειρονομίες αυτές διαχωρίζονται ακόμα και στην περίπτωση που ο χρήστης χειροκροτεί σε, ελαφρώς, διαγώνια διεύθυνση, αφού, τόσο το εύρος της κίνησης όσο και η συχνότητα των κινήσεων παρουσιάζουν μεγάλες διαφοροποιήσεις. Συνεπώς, τα χαρακτηριστικά τα οποία χρειαζόμαστε για την αποτελεσματική αναγνώριση της “explain” είναι οι διαφορές $X_L - X_R$, $X_H - X_R$, $X_H - X_L$, $Y_L - Y_R$.

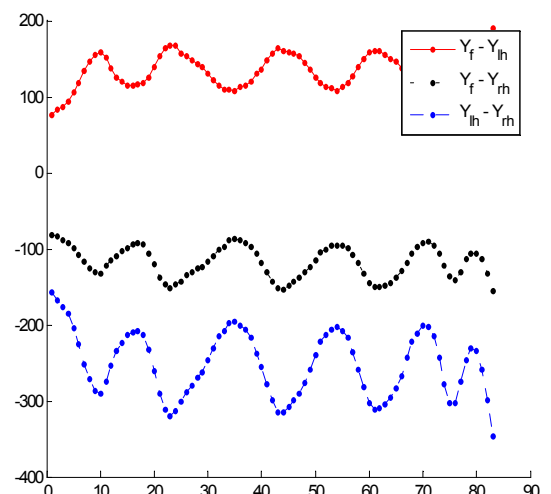
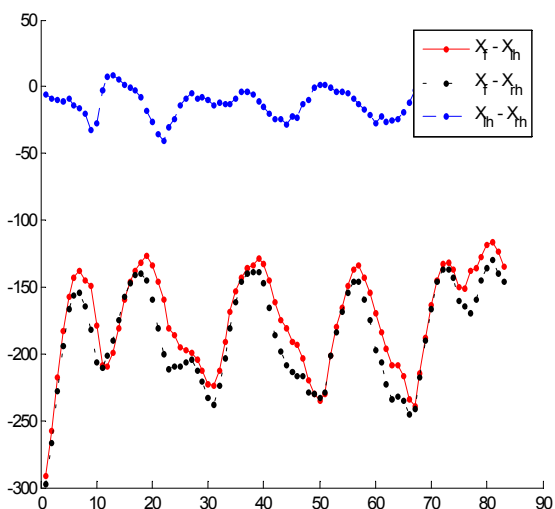
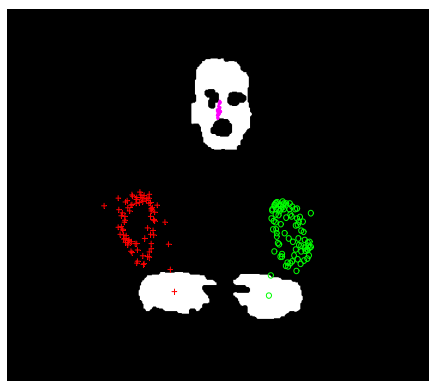
Η “wave”, όπως αναφέραμε, ξεχωρίζει από τις “explain” και “clapping” λόγω της ακινησίας του αριστερού χεριού. Από τις υπόλοιπες χειρονομίες, στις οποίες έχουμε επίσης ακινησία του αριστερού χεριού, ξεχωρίζει, λόγω της κίνησης του δεξιού χεριού, η οποία ακολουθεί την ανύψωση αυτού. Η κίνηση αυτή, η οποία έχει τη διεύθυνση του οριζόντιου άξονα, έχει ως αποτέλεσμα την σχεδόν περιοδική αυξομείωση της διαφοράς των τετμημένων του δεξιού χεριού και του κεφαλιού. Συνεπώς, τα χαρακτηριστικά τα οποία είναι απαραίτητα για το διαχωρισμό της “wave” από τις υπόλοιπες χειρονομίες, είναι οι διαφορές $Y_H - Y_R$ και $X_L - X_R$.

Τέλος, η “one hand”, έχει το ιδιαίτερο χαρακτηριστικό ότι στο τέλος της κίνησης το χέρι ακουμπάει στο κάτω μέρος του κεφαλιού, συνεπώς, όπως συμβαίνει και την “oh my god”, ο αλγόριθμος εντοπισμού, στα τελευταία frame της

ακολουθίας, δεν εντοπίζει το κεφάλι και το χέρι ως δύο ξεχωριστά αντικείμενα αλλά ως ένα. Με τον τρόπο αυτό μηδενίζονται οι διαφορές των τετμημένων και των τεταγμένων του δεξιού χεριού και του κεφαλιού, διαφοροποιώντας την χειρονομία αυτή από τις υπόλοιπες που περιλαμβάνουν, επίσης, ανύψωση του δεξιού χεριού. Από τις υπόλοιπες χειρονομίες διαφοροποιείται λόγω της ακινησίας του αριστερού χεριού. Άρα τα χαρακτηριστικά τα οποία είναι απαραίτητα για την αναγνώριση αυτής της χειρονομίας είναι οι διαφορές $Y_H - Y_R$ και $X_L - X_R$.

Από τα παραπάνω, διαπιστώνουμε, πως με την χρήση των διαφορών των τετμημένων και των τεταγμένων των χεριών και του κεφαλιού, που αναφέραμε παραπάνω, ως χαρακτηριστικών γνωρισμάτων για την εκπαίδευση των HMM, είναι δυνατόν να υπάρξει σαφής διαχωρισμός και των επτά χειρονομιών μεταξύ τους, καθώς, σε κάθε περίπτωση, υπάρχει τουλάχιστον ένα χαρακτηριστικό (και τις περισσότερες φορές δύο, τρία ή και παραπάνω) το οποίο διαφοροποιεί την κάθε χειρονομία από τις υπόλοιπες. Η ανάλυση που προηγήθηκε συμπυκνώνεται και γίνεται εύκολα αντιληπτή στον παρακάτω πίνακα. Οι γραμμές του πίνακα αυτού αποτελούνται από τις επτά χειρονομίες τις οποίες αναγνωρίζει το σύστημά μας ενώ οι στήλες του από τις έξι διαφορές που χρησιμοποιούνται ως χαρακτηριστικά γνωρίσματα εκπαίδευσης. Σε κάθε κελί του πίνακα υπάρχει μια σύντομη περιγραφή για το πώς μεταβάλλεται το χαρακτηριστικό αυτό κατά την χρονική εξέλιξη καθεμιάς από τις χειρονομίες. Με προσεκτική παρατήρηση του πίνακα προκύπτει το συμπέρασμα το οποίο αναφέραμε πριν, ότι, δηλαδή, με χρήση των διαφορών αυτών ως χαρακτηριστικών γνωρισμάτων, επιτυγχάνεται ο διαχωρισμός των επτά χειρονομιών.

Αποτελέσματα του αλγορίθμου εντοπισμού των κεντροειδών για την κίνηση «explain» και γραφικές παραστάσεις των έξι χαρακτηριστικών γνωρισμάτων.



	X_L-X_R (διαφορά των τεταγμένων των δύο χεριών)	X_H-X_R (διαφορά των τεταγμένων του κεφαλιού και του δεξιού χεριού)	X_H-X_L (διαφορά των τεταγμένων του κεφαλιού και του αριστερού χεριού)	Y_L-Y_R (διαφορά των τετμημένων των δύο χεριών)	Y_H-Y_R (διαφορά των τετμημένων του κεφαλιού και του δεξιού χεριού)	Y_H-Y_L (διαφορά των τετμημένων του κεφαλιού και του αριστερού χεριού)
Attention	Αυξάνεται (το δεξί χέρι ανυψώνεται ενώ το αριστερό όχι)	Μειώνεται (το δεξί χέρι ανυψώνεται και φτάνει στο ύψος του κεφαλιού)	Σταθερή (τόσο το κεφάλι όσο και το αριστερό χέρι παραμένουν ακίνητα)	Σταθερή (η κίνηση του δεξιού χεριού είναι σχεδόν κατακόρυφη)	Σταθερή (η κίνηση του δεξιού χεριού είναι σχεδόν κατακόρυφη)	Σταθερή (τόσο το κεφάλι όσο και το αριστερό χέρι παραμένουν ακίνητα)
Clapping	Σταθερή (και περίπου ίση με 0) (υποθέτοντας πως ο χρήστης χειροκροτεί οριζόντια)	Σταθερή (το χέρι παραμένει στο ίδιο ύψος κατά τη διάρκεια της χειρονομίας)	Σταθερή (το χέρι παραμένει στο ίδιο ύψος κατά τη διάρκεια της χειρονομίας)	Σχεδόν περιοδική αυξομείωση	Σχεδόν περιοδική αυξομείωση (με μικρότερο πλάτος από αυτή των τετμημένων των χεριών)	Σχεδόν περιοδική αυξομείωση (με μικρότερο πλάτος από αυτή των τετμημένων των χεριών)
Explain	Σταθερή (και περίπου ίση με 0) (η κατακόρυφη κίνηση των δύο χεριών είναι ίδια)	Σχεδόν περιοδική αυξομείωση (λόγω της κατακόρυφης συνιστώσας της κίνησης)	Σχεδόν περιοδική αυξομείωση (λόγω της κατακόρυφης συνιστώσας της κίνησης)	Σχεδόν περιοδική αυξομείωση	Σχεδόν περιοδική αυξομείωση (λόγω της οριζόντιας συνιστώσας της κίνησης)	Σχεδόν περιοδική αυξομείωση (λόγω της οριζόντιας συνιστώσας της κίνησης)
One hand	Αυξάνεται (το δεξί χέρι ανυψώνεται ενώ το αριστερό όχι)	Μειώνεται και τελικά μηδενίζεται (το δεξί χέρι στο τέλος της κίνησης ενώνεται με το κεφάλι)	Σταθερή (το αριστερό χέρι παραμένει ακίνητο. Ίσως υπάρχει μια μικρή μείωση λόγω της μικρής κίνησης του κεφαλιού προς τα κάτω)	Σταθερή (η κίνηση του δεξιού χεριού είναι σχεδόν κατακόρυφη)	Μειώνεται και τελικά μηδενίζεται (το δεξί χέρι και το κεφάλι πλησιάζουν και στο τέλος της κίνησης ενώνονται)	Σταθερή (το αριστερό χέρι παραμένει ακίνητο. Ίσως υπάρχει μια μικρή αύξηση λόγω της μικρής κίνησης του κεφαλιού προς τα δεξιά)

	X_L-X_R (διαφορά των τεταγμένων των δύο χεριών)	X_H-X_R (διαφορά των τεταγμένων του κεφαλιού και του δεξιού χεριού)	X_H-X_L (διαφορά των τεταγμένων του κεφαλιού και του αριστερού χεριού)	Y_L-Y_R (διαφορά των τετμημένων των δύο χεριών)	Y_H-Y_R (διαφορά των τετμημένων του κεφαλιού και του δεξιού χεριού)	Y_H-Y_L (διαφορά των τετμημένων του κεφαλιού και του αριστερού χεριού)
Oh my god	Σταθερή (και περίπου ίση με 0) (και τα δύο χέρια ανυψώνονται ταυτόχρονα)	Μειώνεται και τελικά μηδενίζεται (το δεξί χέρι στο τέλος της κίνησης ενώνεται με το κεφάλι)	Μειώνεται και τελικά μηδενίζεται (το αριστερό χέρι στο τέλος της κίνησης ενώνεται με το κεφάλι)	Σταθερή (η κίνηση των χεριών γίνεται μόνο στον κατακόρυφο άξονα. Ίσως υπάρχει μια μικρή μείωση καθώς τα χέρια συγκλίνουν προς το κεφάλι)	Μειώνεται και τελικά μηδενίζεται (το δεξί χέρι στο τέλος της κίνησης ενώνεται με το κεφάλι)	Μειώνεται και τελικά μηδενίζεται (το αριστερό χέρι στο τέλος της κίνησης ενώνεται με το κεφάλι)
Wave	Αυξάνεται και κατόπιν παραμένει σταθερή (το δεξί χέρι ανυψώνεται ενώ το αριστερό όχι. Κατόπιν η κίνηση του δεξιού χεριού είναι στον οριζόντιο άξονα μόνο)	Μειώνεται (το δεξί χέρι ανυψώνεται και φτάνει στο ύψος του κεφαλιού)	Σταθερή (τόσο το κεφάλι όσο και το αριστερό χέρι παραμένουν ακίνητα)	Σταθερή και κατόπιν αυξομειώνεται περιοδικά (είναι σταθερή κατά την διάρκεια της ανύψωσης και αυξομειώνεται κατά την διάρκεια της οριζόντιας κίνησης του δεξιού χεριού)	Σταθερή και κατόπιν αυξομειώνεται περιοδικά (είναι σταθερή κατά την διάρκεια της ανύψωσης και αυξομειώνεται κατά την διάρκεια της οριζόντιας κίνησης του δεξιού χεριού)	Σταθερή (τόσο το κεφάλι όσο και το αριστερό χέρι παραμένουν ακίνητα)
Go away	Αυξάνεται (το δεξί χέρι ανυψώνεται ενώ το αριστερό όχι)	Μειώνεται (το δεξί χέρι ανυψώνεται και φτάνει στο ύψος του κεφαλιού)	Σταθερή (τόσο το κεφάλι όσο και το αριστερό χέρι παραμένουν ακίνητα)	Αυξάνεται (η κίνηση του δεξιού χεριού είναι προς τα πάνω και προς τα δεξιά)	Αυξάνεται (η κίνηση του δεξιού χεριού είναι προς τα πάνω και προς τα δεξιά)	Σταθερή (τόσο το κεφάλι όσο και το αριστερό χέρι παραμένουν ακίνητα)

6.4 Ταξινόμηση με Χρήση HMM

Διάφορες μέθοδοι έχουν χρησιμοποιηθεί για την αναγνώριση χειρονομιών. Μερικές από αυτές είναι : η αντιστοίχιση προτύπων, η αναζήτηση σε λεξικό, η στατιστική αντιστοίχιση, το γλωσσικό ταίριασμα, τα νευρωνικά δίκτυα και κάποιες άλλες εξειδικευμένες μέθοδοι.

Τα συστήματα που χρησιμοποιούν τη μέθοδο αντιστοίχισης προτύπων, είναι εύκολο να εκπαιδευθούν, επειδή τα πρότυπα των συστημάτων είναι απλά παραδείγματα προτύπων. Εντούτοις, όταν χρησιμοποιείται μεγάλος αριθμός προτύπων μπορεί να καταστήσουν τη χρήση της μεθόδου αυτής μη οικονομική.

Όταν τα χαρακτηριστικά γνωρίσματα κωδικοποιούνται σαν μια ακολουθία χαρακτήρων με βάση ένα μικρό αλφάβητο, οι τεχνικές αναζήτησης είναι αποδοτικές για την αναγνώριση. Το μειονέκτημα της τεχνικής αυτής είναι ότι το σύστημα δεν είναι αρκετά αποδοτικό.

Οι μέθοδοι στατιστικής αντιστοίχισης υιοθετούν τη λήψη στατιστικών στοιχείων από τα διανύσματα των χαρακτηριστικών για να πραγματοποιήσουν την ταξινόμηση. Τα τυπικά στατιστικά στοιχεία που λαμβάνονται είναι το μέσο διάνυσμα του χαρακτηριστικού ανά κατηγορία, η διαφορά ενός συγκεκριμένου χαρακτηριστικού ανάμεσα σε δύο κατηγορίες κ.α. Μερικές μέθοδοι στατιστικού ταιριάσματος, κάνουν υποθέσεις που έχουν να κάνουν με την κατανομή των χαρακτηριστικών γνωρισμάτων μέσα σε μια κατηγορία. Η απόδοση τέτοιων συστημάτων, τείνει να είναι μικρή όταν παραβιάζονται οι υποθέσεις ή, στην περίπτωση που δεν γίνονται τέτοιες υποθέσεις, όταν τα δεδομένα εκπαίδευσης δεν είναι αρκετά, ώστε να γίνει καλή εκτίμηση της κατανομής.

Τα νευρωνικά δίκτυα έχουν εφαρμοστεί επιτυχώς για την επίλυση προβλημάτων που αφορούν την αναγνώριση προτύπων. Το σημαντικό πλεονέκτημά τους είναι ότι δημιουργούνται από έναν μεγάλο αριθμό απλών στοιχείων και είναι σε θέση να λύνουν, συλλογικά, τα περίπλοκα και διαφορούμενα προβλήματα. Δυστυχώς, τα νευρωνικά δίκτυα, απαιτούν μεγάλη υπολογιστική ισχύ, ειδικά για το κομμάτι της εκπαίδευσής τους.

Η μέθοδος την οποία επιλέξαμε να χρησιμοποιήσουμε για την υλοποίηση του συστήματός μας, είναι αυτή των Κρυφών Μαρκοβιανών Μοντέλων (Hidden Markov Model). Τα HMM, όπως έχουμε προαναφέρει, είναι διπλά στοχαστικά μοντέλα και για το λόγο αυτό είναι κατάλληλα για την μοντελοποίηση των στοχαστικών ιδιοτήτων των χειρονομιών. Τα HMMs υιοθετούνται για να αναπαραστήσουν τις χειρονομίες, ενώ οι παράμετροί τους εκτιμούνται με βάση τα δεδομένα εκπαίδευσης. Με βάση το κριτήριο της πιο πιθανής απόδοσης, οι χειρονομίες μπορούν να αναγνωριστούν με την αξιολόγηση των εκπαιδευμένων HMMs.

6.4.1 Τα Κρυφά Μαρκοβιανά Μοντέλα

Αν και η θεωρία των HMM αναπτύχθηκε σε προηγούμενο κεφάλαιο αναλυτικά, επαναλαμβάνουμε σε αυτό το σημείο κάποια χαρακτηριστικά τους, που είναι απαραίτητο να τα θυμάται ο αναγνώστης.

Ένα HMM είναι μια συλλογή πεπερασμένων καταστάσεων που συνδέονται με μεταβάσεις. Κάθε κατάσταση χαρακτηρίζεται από δύο σύνολα πιθανοτήτων: μια πιθανότητα μετάβασης, και μια κατανομή πιθανότητας εξόδου, η οποία μπορεί να είναι είτε μια διακριτή κατανομή, είτε μια συνεχής συνάρτηση πυκνότητας

πιθανότητας. Η κατανομή αυτή, καθορίζει, δεδομένης της κατάστασης, την πιθανότητα εμφάνισης κάθε συμβόλου εξόδου ή κάποιου συνεχούς τυχαίου διανύσματος, αντιστοίχως.

Ένα HMM μπορεί να καθοριστεί από:

- $\{S\}$ - Ένα σύνολο καταστάσεων, συμπεριλαμβανομένης της αρχικής κατάστασης S_I και της τελικής κατάστασης S_F
- A - Ο πίνακας πιθανοτήτων μετάβασης, $A = \{a_{ij}\}$, όπου a_{ij} είναι η πιθανότητα μετάβασης από την κατάσταση I στην κατάσταση j .
- B - Ο πίνακας πιθανοτήτων εξόδου, $B = \{b_j(O_k)\}$ για διακριτά HMM, $B = \{b_j(x)\}$ για συνεχή HMM, όπου O_k παριστάνει ένα διακριτό σύμβολο παρατήρησης, και το x παριστάνει k -διαστάσεων τυχαία διανύσματα συνεχών παρατηρήσεων.

Ειδικά για τα διακριτά HMM, τα a_{ij} και b_{ij} έχουν τις ακόλουθες ιδιότητες:

$$a_{ij} \geq 0, b_j(O_k) \geq 0, \forall i, j, k$$

$$\sum_j a_{ij} = 1, \forall i$$

$$\sum_k b_j(O_k) = 1, \forall j.$$

Γενικά, υπάρχουν τρία βασικά προβλήματα τα οποία σχετίζονται με τα HMM: α)το πρόβλημα αξιολόγησης, β)το πρόβλημα αποκωδικοποίησης, και γ)το πρόβλημα εκμάθησης.

Στο πρόβλημα εκμάθησης, γίνεται εκτίμηση των παραμέτρων του μοντέλου, με χρήση επαναληπτικών αλγορίθμων, κατά τέτοιο τρόπο, ώστε η πιθανότητα η ακολουθία παρατήρησης που να έχει παραχθεί από το δεδομένο μοντέλο, να μεγιστοποιείται. Επομένως, η διαδικασία εκμάθησης είναι ,ουσιαστικά, η δημιουργία των μοντέλων των χειρονομιών από τα δεδομένα εκπαίδευσης.

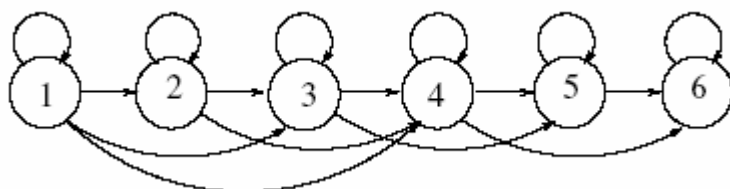
Στο πρόβλημα αποκωδικοποίησης, υπολογίζεται η πιθανότητα του μοντέλου, με τις δεδομένες παραμέτρους, να έχει γεννήσει την παρατηρούμενη ακολουθία.

6.4.2 Χρήση των HMM στο Σύστημα μας

Τα HMM τα οποία χρησιμοποιήσαμε για την μοντελοποίηση των δυναμικών χαρακτηριστικών των χειρονομιών που αναγνωρίζει το σύστημά μας είναι τυπικά forward HMM. Δημιουργήθηκε ένα μοντέλο για κάθε χειρονομία και οι παράμετροι των μοντέλων εκτιμήθηκαν από τις ακολουθίες εκπαίδευσης που ήταν διαθέσιμες για κάθε χειρονομία, με χρήση του αλγόριθμου Baum-Welch. Η πιθανότητα που εμφανίζει κάθε HMM, κατά την αποκωδικοποίηση, να έχει γεννήσει την ακολουθία παρατήρησης, αποτελεί το κριτήριο για την αναγνώριση. Η ακολουθία θεωρείται πως ανήκει στην χειρονομία, της οποίας το HMM μας δίνει την μεγαλύτερη πιθανότητα. Όταν κάποια ακολουθία δίνει πάρα πολύ μικρές πιθανότητες (μικρότερες από ένα κατώφλι) σε όλα μοντέλα τότε αυτή η ακολουθία θεωρείται πως δεν ανήκει σε καμία από τις χειρονομίες τις οποίες μπορεί να αναγνωρίσει το σύστημά μας.

Όπως είπαμε τα μοντέλα που χρησιμοποιήθηκαν, είναι αριστερά προς τα δεξιά μοντέλα (bakis), αφού δεν είναι δυνατόν να υπάρξουν μεταβάσεις σε καταστάσεις οι

οποίες αναπαριστούν στάδια της χειρονομίας τα οποία προηγούνται χρονικά. Οι καταστάσεις που επιλέξαμε να έχουν τα μοντέλα των χειρονομιών μας, εξαρτώνται από την κάθε χειρονομία. Οι πιο σύνθετες χειρονομίες, απαιτούν και περισσότερες καταστάσεις για την μοντελοποίησή τους. Πάντως, για τις περισσότερες χειρονομίες χρησιμοποιήθηκαν μοντέλα με 6 καταστάσεις. Η επιλογή του αριθμού των καταστάσεων για κάθε μοντέλο, βασίστηκε στην ανάλυση των χαρακτηριστικών της κάθε χειρονομίας στο χρόνο. Αυτό που παρατηρήσαμε, είναι, ότι καθεμία από τις χειρονομίες που επιλέξαμε να αναγνωρίζει το σύστημά μας, μπορεί να χωριστεί σε 6 το πολύ ξεχωριστά στάδια. Για παράδειγμα η attention μπορεί να χωριστεί σε τρία στάδια : στο στάδιο κατά το οποίο τα χέρια βρίσκονται στην αρχική θέση, στο στάδιο κατά το οποίο εκτελείται η ανύψωση του δεξιού χεριού και στο τελευταίο στάδιο, στο οποίο το χέρι έχει σχεδόν φτάσει στο τελικό του ύψος και παραμένει εκεί. Ιδανικά, καθένα από τα στάδια, στα οποία μπορούμε εμείς, χρησιμοποιώντας την παραπάνω ανάλυση, να χωρίσουμε τις χειρονομίες, αντιστοιχεί και σε μία κατάσταση για το HMM που μοντελοποιεί την συγκεκριμένη χειρονομία. Παρόλα αυτά, διαπιστώθηκε πως το μοντέλο το οποίο αναπαριστούσε καλύτερα μια χειρονομία μπορεί να διέφερε κατά μία κατάσταση από τα στάδια στα οποία είχαμε χωρίσει εμείς την χειρονομία. Η τοπολογία ενός από τα μοντέλα που χρησιμοποιήσαμε φαίνεται στο παρακάτω σχήμα.



Όσον αφορά την επιλογή των εξόδων, όπως προαναφέραμε, χρησιμοποιήθηκαν δύο τρόποι. Ο πρώτος ήταν η χρήση διακριτών HMM, στα οποία έχουμε διακριτή κατανομή πιθανότητας εξόδου και οι έξοδοι αποτελούνται από στοιχεία που ανήκουν σε ένα πεπερασμένο σύνολο συμβόλων. Στην περίπτωση αυτή, τα διανύσματα των χαρακτηριστικών που είχαν δημιουργηθεί, κβαντίστηκαν με τέτοιο τρόπο, ώστε τα στοιχεία που περιέχουν, να ανήκουν σε αυτό το πεπερασμένο σύνολο συμβόλων. Το σύνολο των συμβόλων επιλέγεται με διαφορετικό τρόπο για την κάθε χειρονομία, ώστε να μην υπάρχει απώλεια κρίσιμων πληροφοριών που αφορούν την συγκεκριμένη κίνηση. Συνεπώς, όταν μια ακολουθία πρόκειται να αποκωδικοποιηθεί από μοντέλα των οποίων οι έξοδοι προέρχονται από διαφορετικά σύνολα συμβόλων, κβαντίζεται και με διαφορετικό τρόπο πριν γίνει η αποκωδικοποίηση με το κάθε μοντέλο, ώστε να υπάρχει αντιστοιχία στα σύμβολα που περιέχει το σύνολο συμβόλων του συγκεκριμένου μοντέλου και στα σύμβολα που περιέχονται στα διανύσματα με τα χαρακτηριστικά της ακολουθίας. Τα παραπάνω θα περιγραφούν αναλυτικότερα κατά την περιγραφή της υλοποίησης του διακριτού μοντέλου.

Η δεύτερη προσέγγιση είναι αυτή των HMM με συνεχή κατανομή εξόδου. Στα συνεχή HMM, η παρατηρήσιμη έξοδος, που προκύπτει από κάθε κατάσταση, καθορίζεται από τη συνάρτηση πυκνότητας πιθανότητας που έχει επιλεγεί για την συγκεκριμένη κατάσταση. Η πιο συνηθισμένη μορφή της συνάρτησης πυκνότητας πιθανότητας είναι η γκαουσιανή κατανομή. Επειδή, όμως, η χρήση μίας μόνο γκαουσιανής κατανομής μπορεί να μην επαρκεί για να μοντελοποιήσει ικανοποιητικά τις εξόδους που υπάρχουν στις ακολουθίες εκπαίδευσης, ειδικά σε συστήματα στα οποία είναι επιθυμητό να αναγνωρίζονται χειρονομίες που γίνονται από διαφορετικούς χρήστες (πράγμα το οποίο συνεπάγεται μεγαλύτερη διακύμανση

των τιμών των χαρακτηριστικών που παρουσιάζουν διαφορετικές ακολουθίες - της ίδιας όμως χειρονομίας - , χρησιμοποιούνται ως κατανομή εξόδου, μείγματα (δηλαδή γραμμικοί συνδυασμοί) από γκαουσιανές κατανομές για την μοντελοποίηση της εξόδου. Αυτές οι κατανομές εξόδου ενός Gaussian Mixture HMM περιγράφονται από την σχέση

$$b_j = \sum_{m=1}^M c_{jm} N(x_t; \mu_{jm}, \Sigma_{jm})$$

για την οποία ισχύει

$$\sum_{m=1}^M c_{jm} = 1$$

με $c_{jm} = P(m|s_t = j)$ έτσι ώστε

$$\int b_j(x_t) dx_t = 1$$

Παρόλο που ο υπολογισμός των πιθανοτήτων με χρήση διακριτών HMM είναι ταχύτερος από τον υπολογισμό με χρήση συνεχών HMM, τα τελευταία επιλέγονται λόγω της καλύτερης επίδοσης τους (σε εφαρμογές στις οποίες οι έξοδοι παίρνουν συνεχείς τιμές και πρέπει να υποστούν διαδικασία κβαντισμού).

Ένα άλλο σημαντικό πρόβλημα, είναι η επιλογή της αρχικής κατάστασης του μοντέλου, οι τιμές δηλαδή που θα έχουν ο πίνακας μετάβασης και ο πίνακας εξόδου (εάν πρόκειται για διακριτά HMM) ή οι παράμετροι του μείγματος των γκαουσιανών κατανομών (εάν πρόκειται για συνεχή HMM). Υπάρχουν τρεις τρόποι για την αρχικοποίηση ενός μοντέλου HMM. Ο πρώτος είναι με τη χρήση απλών τυχαίων αρχικών μοντέλων, ο δεύτερος με τη χρήση πολλαπλών τυχαίων αρχικών μοντέλων και ο τρίτος με τον απευθείας υπολογισμό του αρχικού μοντέλου με βάση τις φυσικές ιδιότητες της κίνησης, που θέλουμε να μοντελοποιήσουμε.

Η χρήση ενός απλού τυχαίου μοντέλου για την αρχικοποίηση του HMM συναντάται πολύ συχνά στις περιπτώσεις, που ο επαναληπτικός αλγόριθμος που χρησιμοποιείται για την εκτίμηση των παραμέτρων είναι ο Baum – Welch. Όταν χρησιμοποιούμε πολλές ακολουθίες για την εκπαίδευση του HMM, η μέθοδος του απλού τυχαίου μοντέλου, προϋποθέτει ότι η αρχική κατάσταση του HMM θα οριστεί μία φορά και κατόπιν η ίδια αρχικοποίηση θα χρησιμοποιηθεί κατά την εκπαίδευση με όλες τις ακολουθίες. Το μειονέκτημα αυτής της μεθόδου αρχικοποίησης είναι ότι η τελική τιμή των παραμέτρων του HMM παρουσιάζουν μεγάλη εξάρτηση από τις αρχικές τιμές που τους είχαν δοθεί. Η απόδοση που παρουσιάζουν μοντέλα που εκπαιδεύονται με αυτό τον τρόπο ποικίλλει. Το γεγονός αυτό οφείλεται στο ότι ο Baum – Welch μπορεί να προσεγγίσει μόνο ένα τοπικό μέγιστο και όχι πάντα το ολικό. Συνεπώς, εάν το μέγιστο που βρίσκεται κοντά στις τιμές με τις οποίες έχει αρχικοποιηθεί το μοντέλο είναι τοπικό, η απόδοσή του δεν είναι καλή, ενώ εάν τύχει να έχουν δοθεί τιμές που βρίσκονται κοντά στο ολικό μέγιστο, ο αλγόριθμος θα συγκλίνει σε αυτό και η απόδοση του μοντέλου θα είναι καλή.

Η αρχικοποίηση που επιλέξαμε εμείς για τα HMM, που μοντελοποιούν τις χειρονομίες, που αναγνωρίζει το σύστημά μας, είναι μεν αυτή του απλού τυχαίου μοντέλου, έχει παρόλα αυτά κάποια στοιχεία από την τρίτη μέθοδο, αυτή της απευθείας εκτίμησης των παραμέτρων. Δηλαδή μπορεί μεν να χρησιμοποιήθηκε η ίδια αρχική κατάσταση για την εκτίμηση των παραμέτρων από όλες τις ακολουθίες, όμως η αρχική αυτή κατάσταση δεν είναι εντελώς τυχαία αλλά ήταν προσαρμοσμένη κάθε φορά στην φύση της χειρονομίας. Με βάση την ανάλυση της κάθε χειρονομίας και την υπόθεση που κάναμε για την φυσική ερμηνεία και την αντιστοιχία των καταστάσεων του κάθε μοντέλου με τα επιμέρους στάδια κάθε κίνησης, αρχικοποιήθηκαν οι πίνακες μεταβάσεως και εξόδου έτσι, ώστε να μην έχουμε μεταβάσεις ανάμεσα σε καταστάσεις που δεν είναι δυνατόν να υπάρχουν κατά την εξέλιξη της χειρονομίας καθώς επίσης και συγκεκριμένες εξόδους από καταστάσεις που δεν είναι δυνατόν να τις γεννήσουν.

Για παράδειγμα, αν υποθέσουμε ότι η attention, με βάση την ανάλυσή της σε επιμέρους στάδια μοντελοποιείται ικανοποιητικά με τρεις καταστάσεις για το αντίστοιχο HMM, μία για την περίοδο που τα χέρια βρίσκονται στο τραπέζι, μία για το στάδιο που έχουμε την ανύψωση του χεριού και μία για το στάδιο που το δεξί χέρι φτάνει στο τελικό του ύψος, τότε η πιθανότητα μετάβασης από την πρώτη κατάσταση στη δεύτερη καθώς και η πιθανότητα παραμονής στην ίδια κατάσταση και στο επόμενο frame, θα είναι σαφώς μεγαλύτερη από την απευθείας μετάβαση από τη πρώτη στην τρίτη κατάσταση.

Επίσης, αν συμβολίσουμε την έξοδο που παρατηρείται όταν ο δεξί χέρι βρίσκεται στο κάτω μέρος της εικόνας, με «Α», την έξοδο, που παρατηρείται όταν ο δεξί χέρι βρίσκεται στο μέση της εικόνας, με «Β» και την έξοδο, που παρατηρείται όταν ο δεξί χέρι βρίσκεται στο πάνω μέρος της εικόνας, με «Γ», τότε, η πιθανότητα να έχουμε την έξοδο «Α», όταν το μοντέλο βρίσκεται στην πρώτη κατάσταση είναι σαφώς μεγαλύτερη από την πιθανότητα εμφάνισης του «Α» ενόσω το μοντέλο βρίσκεται σε κάποια από τις άλλες δύο καταστάσεις. Αντιστοίχως, η πιθανότητα εμφάνισης του «Β» είναι μεγαλύτερη όταν το μοντέλο βρίσκεται στην δεύτερη κατάσταση και του «Γ», όταν το μοντέλο βρίσκεται στην τρίτη κατάσταση.

Με τον τρόπο αυτό αυξάνονται οι πιθανότητες οι τελικές παράμετροι του μοντέλου να έχουν προέλθει από σύγκλιση του Baum-Welch σε ολικό και όχι σε τοπικό μέγιστο. Όμως, δεν είναι δυνατόν η παραπάνω ανάλυση να γίνει για όλες τις χειρονομίες, καθώς μερικές είναι αρκετά πιο πολύπλοκες από την “attention”. Στην περίπτωση λοιπόν, που οι αρχική κατάσταση του HMM ορίζεται με εντελώς τυχαίο τρόπο, είναι πολύ σημαντικό να υπάρχουν όσο το δυνατόν περισσότερες ακολουθίες για την εκπαίδευση του μοντέλου. Ο μεγάλος όγκος των δεδομένων εκπαίδευσης έχει πολύ ευεργετικά αποτελέσματα στη σωστή εκτίμηση των παραμέτρων του μοντέλου, ακόμα και εάν οι αρχικές τιμές δεν έχουν επιλεγεί με τον καλύτερο δυνατό τρόπο.

ΚΕΦΑΛΑΙΟ 7

ΥΛΟΠΟΙΗΣΗ ΜΕ ΔΙΑΚΡΙΤΑ HMM

7.1 Κβαντισμός των Χαρακτηριστικών

Στην περίπτωση των διακριτών HMM, είναι απαραίτητο να μετατραπούν τα διανύσματα των χαρακτηριστικών που έχουν εξαχθεί από τα frame εισόδου σε σύμβολα που ανήκουν σε ένα πεπερασμένο σύνολο. Η ανεξάρτητη κβαντοποίηση κάθε τιμής ή παραμέτρου του σήματος, μπορεί να πραγματοποιηθεί με κλιμακωτή κβαντοποίηση. Αντίθετα, η κοινή κβαντοποίηση ενός συνόλου παραμέτρων μπορεί να πραγματοποιηθεί μέσω της διανυσματικής κβαντοποίησης (VQ).

Η αντιπροσώπευση της κωδικής λέξης της VQ στο χώρο του δείγματος, μπορεί να είναι το κεντροειδές του αντίστοιχου block της εικόνας, όπως συμβαίνει στη συμβατική διανυσματική κβαντοποίηση, ή μπορεί να υπολογιστεί ως η συνάρτηση πυκνότητας πιθανότητας για το αντίστοιχο block. Οι τεχνικές VQ έχουν επιτυχώς χρησιμοποιηθεί για να λύσουν προβλήματα κβαντοποίησης ή συμπίεσης δεδομένων. Σε μια υλοποίηση με HMM, η VQ μπορεί να παίξει έναν σημαντικό ρόλο στη μετατροπή των συνεχών σημάτων σε διακριτά σύμβολα για διακριτά HMMs όπως εκείνα που χρησιμοποιούνται στην αναγνώριση φωνής ή αυτά που έχουν χρησιμοποιηθεί στην συγκεκριμένη εφαρμογή.

Ένας διανυσματικός κβαντιστής καθορίζεται πλήρως από ένα codebook, το οποίο αποτελείται από ένα σύνολο σταθερών, πρότυπων διανυσμάτων. Η VQ μειώνει τον πλεονασμό δεδομένων, και αυτό προκαλεί, αναπόφευκτα, την παραμόρφωση μεταξύ των αρχικών σημάτων. Ένα βασικό πρόβλημα στην τεχνική VQ είναι να ελαχιστοποιηθεί η παραμόρφωση που εισάγεται. Μια περιγραφή της διαδικασίας VQ περιλαμβάνει: (1) την εκτίμηση της παραμόρφωσης, και (2) την παραγωγή συγκεκριμένου αριθμού πρότυπων διανυσμάτων. Δύο τυπικές τεχνικές VQ που ελαχιστοποιούν την παραμόρφωση που δημιουργείται, αναφέρονται παρακάτω.

Έστω $x = (x_1, x_2, \dots, x_d) \in R^d$ ένα d-διάστατο διάνυσμα, όπου $\{x_k, 1 \leq k \leq d\}$ πραγματικές τυχαίες μεταβλητές. Στη διαδικασία VQ, το διάνυσμα x αντιστοιχίζεται σε ένα άλλο d-διάστατο διάνυσμα z με πραγματικές τιμές $z = q(x)$, όπου q είναι ο παράγοντας κβαντοποίησης. Γενικά, το z είναι ένα στοιχείο ενός πεπερασμένου συνόλου τιμών $Z = \{z_i, 1 \leq i \leq L\}$, όπου $z_i = (z_{i1}, z_{i2}, \dots, z_{id})$. Το σύνολο Z αναφέρεται ως “codebook”, L είναι το μέγεθός του, και $\{z_i\}$ είναι το σύνολο των κωδικών-λέξεων.

Η διαδικασία δημιουργίας του κβαντιστή είναι επίσης γνωστή ως διαδικασία εκπαίδευσης. Σε αυτήν, ο d-διάστατος χώρος του αρχικού τυχαίου διανύσματος x χωρίζεται σε L περιοχές ή block $\{C_i, 1 \leq i \leq L\}$ και το κάθε block αντιστοιχίζεται με ένα διάνυσμα z_i . Ο κβαντιστής τότε αναθέτει την κωδική-λέξη z_i αν το x ανήκει στο C_i , ώστε, $q(x) = z_i$, αν το x ανήκει στο C_i .

Οποιοδήποτε διάνυσμα εισόδου X , που ανήκει στο block C κβαντοποιείται ως z_i . Η μορφή κάθε block μπορεί να είναι διαφορετική, και οι θέσεις των κωδικών-λέξεων που αντιστοιχούν στα block, καθορίζεται από την ελαχιστοποίηση της μέσης παραμόρφωσης που εισάγεται από τα αντίστοιχα block.

Για να αξιολογηθεί το σφάλμα κβαντοποίησης μεταξύ του x και του z , είναι απαραίτητο να καθοριστεί ένα μέτρο παραμόρφωσης $d(x, z)$, ώστε να μετρηθεί η ποιότητα κβαντοποίησης. Το μέτρο παραμόρφωσης μεταξύ του x και του z είναι

επίσης γνωστό, ως μέτρο απόστασης. Το μέτρο πρέπει να είναι εντοπίσιμο και υπολογίσιμο ώστε να είναι δυνατή η ανάλυσή του, και πρέπει επίσης να είναι υποκειμενικά σχετικό έτσι, ώστε διαφορές στις τιμές της παραμόρφωσης να συσχετίζονται με την ποιότητα της κβαντοποίησης.

Γενικά, το σταθμισμένο μέσο τετραγωνικό σφάλμα της παραμόρφωσης μπορεί να χρησιμοποιηθεί έτσι, ώστε οι παραμορφώσεις που εισάγονται με την κβαντοποίηση διαφορετικών παραμέτρων να είναι ίσες. Τα άνισα βάρη μπορούν να χρησιμοποιηθούν για να επιτρέψουν σ' ορισμένα στοιχεία να είναι σημαντικότερα από άλλα. Μια δημοφιλής επιλογή για άνισα βάρη είναι να χρησιμοποιηθεί ο αντίστροφος του πίνακα διακύμανσης του z ,

$$d(x, z) = (x - z)' \Sigma^{-1} (x - z),$$

όπου Σ ο πίνακας διακύμανσης του z . Η μέτρηση της παραπάνω παραμόρφωσης μπορεί να απλοποιηθεί σε τετραγωνικό σφάλμα παραμόρφωσης,

$$d(x, \hat{x}) = \|x, \hat{x}\| = \sum_{i=0}^{d-1} (x_i - z_i)^2 .$$

Δεν είναι γνωστή καμία αναλυτική λύση που να εγγυάται τη γενική ελαχιστοποίηση του μέτρου της μέσης παραμόρφωσης για ένα δεδομένο σύνολο διανυσμάτων. Εντούτοις, επαναληπτικοί αλγόριθμοι, οι οποίοι μπορούν να εγυθηθούν ένα τοπικό ελάχιστο, είναι διαθέσιμοι και λειτουργούν καλά στην πράξη.

7.2 Υλοποίηση

7.2.1 Εκπαίδευση των Μοντέλων

Όπως προαναφέρθη, η υλοποίηση των διακριτών HMM, έγινε με τη χρήση του toolbox του MATLAB stats 4.1 . Το toolbox αυτό, περιλαμβάνει πέντε συναρτήσεις, οι οποίες χρησιμοποιούνται για την δημιουργία και επεξεργασία HMM. Εμείς χρησιμοποιήσαμε μόνο τις δύο από αυτές. Οι συναρτήσεις που χρησιμοποιήσαμε είναι η `hmmtrain` και η `hmmdecode`. Οι υπόλοιπες τρεις συναρτήσεις είναι η `hmmdecode`, η `hmmestimate` και η `hmmviterbi` .

Η σύνταξη της `hmmtrain` είναι η ακόλουθη :

$$[ESTTR, ESTEMIT] = \text{hmmtrain}(\text{seq}, \text{TRGUESS}, \text{EMITGUESS})$$

Η `hmmtrain` χρησιμοποιείται για την εκτίμηση των παραμέτρων μετάβασης και εκπομπής του HMM με βάση την ακολουθία `seq` και με χρήση του αλγόριθμου Baum – Welch. Οι πίνακες `TRGUESS` και `EMITGUESS` είναι οι αρχικές εκτιμήσεις για τους πίνακες μετάβασης και εκπομπής (εξόδου) του HMM και ο υπολογισμός τους γίνεται με τον τρόπο που περιγράφηκε στην προηγούμενη ενότητα. Το στοιχείο `TRGUESS(i,j)` του πίνακα μετάβασης είναι η εκτιμώμενη πιθανότητα ώστε να έχουμε μετάβαση από την κατάσταση i στην κατάσταση j του μοντέλου. Επίσης, το στοιχείο `EMITGUESS(i,k)` του πίνακα εκπομπής είναι η εκτιμώμενη πιθανότητα ώστε να έχουμε την εμφάνιση του συμβόλου k , όταν το μοντέλο βρίσκεται στην κατάσταση i .

Το MATLAB μας παρέχει την δυνατότητα να ορίσουμε τον επαναληπτικό αλγόριθμο, που επιθυμούμε να χρησιμοποιηθεί από την `hmmtrain`, ώστε να γίνει η

εκτίμηση των παραμέτρων του μοντέλου, ανάμεσα στον Baum – Welch (που είναι και αυτός που χρησιμοποιείται ως default) και στον Viterbi. Ο αλγόριθμος που εμείς επιλέξαμε για την εκτίμηση των παραμέτρων των μοντέλων μας ήταν ο Baum – Welch .

Δύο ακόμα παράμετροι που μπορούν να οριστούν κατά την κλήση της `hmmtrain` είναι ο μέγιστος αριθμός επαναλήψεων που επιτρέπεται να εκτελέσει ο αλγόριθμος ώσπου να επιτύχει την επιθυμητή ακρίβεια και η επιθυμητή ακρίβεια. Με τον όρο επιθυμητή ακρίβεια, εννοούμε την ελάχιστη ποσότητα κατά την οποία θέλουμε να διαφέρουν οι πιθανότητες που προκύπτουν (οι πιθανότητες αυτές είναι η πιθανότητα που προκύπτει, η ακολουθία εκπαίδευσης να έχει γεννηθεί από το εκπαιδευόμενο μοντέλο με τις τρέχουσες – για την συγκεκριμένη επανάληψη - τιμές των συντελεστών), κατά την εκτέλεση δύο διαδοχικών επαναλήψεων του αλγορίθμου. Για την επίτευξη της δεδομένης ακρίβειας που έχει οριστεί, είναι απαραίτητο να γίνει ένας συγκεκριμένος αριθμός επαναλήψεων. Σε περίπτωση που η ακρίβεια είναι πολύ μικρή, είναι δυνατόν, ο αλγόριθμος να χρειαστεί πάρα πολλές επαναλήψεις για να την πετύχει. Για το λόγο αυτό, ορίζουμε και τον μέγιστο αριθμό επαναλήψεων, που επιθυμούμε να εκτελέσει ο αλγόριθμος, ώστε στην περίπτωση που δεν επιτευχθεί η ζητούμενη ακρίβεια μετά από δεδομένο αριθμό επαναλήψεων, να τερματιστεί η εκτέλεση του αλγορίθμου.

Τα δεδομένα εκπαίδευσης εισάγονται στην `hmmtrain`, μέσω της `seq`. Η `seq` μπορεί να είναι ένα διάνυσμα (πίνακας στήλη) ή μπορεί να είναι μια δομή με πολλούς πίνακες. Ο καθένας από αυτούς τους πίνακες αντιστοιχεί σε μια ακολουθία (διάνυσμα) εκπαίδευσης. Με τον τρόπο αυτό μας παρέχεται η δυνατότητα να γίνεται η εκπαίδευση του μοντέλου, με όλες τις ακολουθίες εισόδου, με μία μόνο κλήση της `hmmtrain`.

Το πλεονέκτημα που μας προσφέρει η `hmmtrain` είναι η δυνατότητα που μας παρέχει, οι πίνακες οι οποίοι αποτελούν τη δομή `seq` να μην έχουν απαραίτητα το ίδιο μήκος. Δηλαδή, οι ακολουθίες οι οποίες έχουν προέλθει από τα διαφορετικά `video`, της ίδιας χειρονομίας, να μην έχουν αναγκαστικά τον ίδιο αριθμό `frame` (πράγμα το οποίο, πρακτικά, σημαίνει διαφορετικό χρόνο εκτέλεσης της χειρονομίας σε κάθε `video`).

Το σημαντικότερο μειονέκτημα, που παρουσιάζει η `hmmtrain` είναι ότι δεν υποστηρίζει την εκπαίδευση πολυδιάστατων HMM. Τα χαρακτηριστικά γνωρίσματα τα οποία επιλέξαμε, όπως εξηγήσαμε σε προηγούμενη ενότητα, για την εκπαίδευση των HMM κάθε χειρονομίας είναι έξι. Πιο συγκεκριμένα είναι οι διαφορές στις τεταγμένες και στις τετημένες, των δύο χεριών και του κεφαλιού. Δυστυχώς η `hmmtrain`, δεν μπορεί να δεχθεί ως όρισμα μέσω του `cell array`, `seq`, πίνακες οι οποίοι αποτελούνται από περισσότερες από μία γραμμές. Ως εκ τούτου, δεν μπορεί να δημιουργηθεί ένας `cell array seq`, ο οποίος να αποτελείται από τόσους πίνακες όσες είναι οι διαθέσιμες ακολουθίες εκπαίδευσης (τα `videos`) για κάθε χειρονομία, και ο καθένας από αυτούς τους πίνακες να έχει τόσες στήλες όσα τα `frame` (μήκος) της ακολουθίας και τόσες γραμμές όσα τα χαρακτηριστικά με τα οποία επιθυμούμε να εκπαιδύσουμε τα μοντέλα. Για τον λόγο αυτό αναγκαστήκαμε αντί να έχουμε ένα μοντέλο για κάθε χειρονομία, να έχουμε ένα μοντέλο για κάθε χαρακτηριστικό της κάθε χειρονομίας. Κατόπιν, κατά τη διαδικασία αποκωδικοποίησης η συνολική πιθανότητα που παρουσιάζει μια ακολουθία να έχει προέλθει από ένα μοντέλο που αναπαριστά μία χειρονομία, υπολογίζεται, ως το γινόμενο των πιθανοτήτων που προκύπτουν για καθένα από τα έξι χαρακτηριστικά της συγκεκριμένης χειρονομίας.

Όπως αναφέρθηκε και προηγουμένως, τα διανύσματα των χαρακτηριστικών, για να μπορέσουν να χρησιμοποιηθούν για την εκπαίδευση διακριτών μοντέλων

HMM, πρέπει να υποστούν διαδικασία κβαντισμού. Η διαδικασία κβαντισμού που ακολουθήσαμε στην υλοποίησή μας, ήταν προσαρμοσμένη στις ανάγκες της συγκεκριμένης εφαρμογής, έτσι ώστε να εξασφαλιστεί, ότι θα υπάρξει η μικρότερη δυνατή απώλεια πληροφορίας και θα μεγιστοποιηθεί η απόδοση του συστήματος.

Αρχικά ορίστηκε το codebook, το σύνολο δηλαδή των πεπερασμένων συμβόλων που χρησιμοποιήθηκαν. Το σύνολο αυτό πρέπει να είναι ενιαίο για τα HMM που μοντελοποιούν όλες τις χειρονομίες, ώστε τα αποτελέσματα, τα οποία προκύπτουν κατά την αποκωδικοποίηση, να είναι συγκρίσιμα μεταξύ τους. Ο αριθμός των συμβόλων, που διαπιστώθηκε, ότι μπορεί να κωδικοποιήσει ικανοποιητικά όλες τις χειρονομίες είναι έξι. Κατόπιν, παρατηρήσαμε ότι οι διαφορές των τετμημένων και των τεταγμένων των χεριών και του κεφαλιού (δηλαδή οι τιμές των χαρακτηριστικών) δεν ξεπερνούν μια μέγιστη τιμή, η οποία είναι μικρότερη από την μεγαλύτερη διάσταση της εικόνας, η οποία είναι το μήκος (που είναι 720 pixels). Το γεγονός αυτό οφείλεται τόσο στο ότι ο χρήστης, που εκτελεί τις χειρονομίες βρίσκεται στο μέσο περίπου της εικόνας και δεν καταλαμβάνει εύρος μεγαλύτερο από 300 pixels αλλά και στο ότι οι χειρονομίες που εκτελούνται, έχουν, ως κύρια διεύθυνση κίνησης τον κατακόρυφο άξονα, ενώ η κίνηση κατά τον οριζόντιο άξονα δεν ξεπερνάει τα όρια (η τα ξεπερνάει για μερικά pixels) του σώματος του χρήστη (που όπως είπαμε είναι περίπου 300 pixels).

Το κατακόρυφο μέγεθος της εικόνας είναι 576 pixels και είναι μεγαλύτερο από το εύρος των 300 περίπου pixels, που καταλαμβάνει, όπως προείπαμε, η κίνηση στον οριζόντιο άξονα. Ως εκ τούτου, η μέγιστη τιμή που μπορεί να παρατηρηθεί στα χαρακτηριστικά, είναι αυτή των 576 pixels. Επειδή όμως, δεν υπάρχει καμία κίνηση η οποία να εκτείνεται σε ολόκληρο τον κατακόρυφο άξονα της εικόνας (πράγμα το οποίο θα σήμαινε ότι το ένα χέρι θα έπρεπε να βρίσκεται στο άνω άκρο της εικόνας και το άλλο στο κάτω άκρο), αλλά η μέγιστη διαφορά ανάμεσα στις τεταγμένες των χεριών είναι περίπου 500 pixels, χωρίσαμε την εικόνα σε blocks των 82 pixels. Η τιμή αυτή προέκυψε με διαίρεση της μέγιστης τιμής που είναι δυνατόν να παρατηρηθεί σε κάποιο από τα χαρακτηριστικά, με βάση την παραπάνω ανάλυση, με τον αριθμό των συμβόλων, που αποφασίσαμε να χρησιμοποιήσουμε για την κωδικοποίηση των κινήσεων, που όπως είπαμε είναι έξι.

Με τον τρόπο που μόλις περιγράφηκε, δημιουργούνται οι πίνακες που περιέχουν τα χαρακτηριστικά γνωρίσματα για κάθε χειρονομία. Οι πίνακες αυτοί περιέχουν τις τιμές των χαρακτηριστικών, για όλα τα frame της ακολουθίας. Συγκεκριμένα, επειδή όπως είπαμε δεν είναι δυνατόν να εκπαιδύσουμε το μοντέλο μας με πολλά χαρακτηριστικά ταυτόχρονα, ο κάθε πίνακας έχει μια γραμμή (οι τιμές για ένα χαρακτηριστικό) και τόσες στήλες όσες τα frame της ακολουθίας εκπαίδευσης. Επίσης, οι τιμές που υπάρχουν στους πίνακες, σύμφωνα με την παραπάνω διαδικασία κβαντισμού, θα πρέπει να έχουν τιμές από 1 έως 6. Κατόπιν, δημιουργείται η **seq**, η οποία είναι ένας cell array και αποτελείται από τους πίνακες που έχουν τις τιμές του συγκεκριμένου χαρακτηριστικού για όλες τις ακολουθίες εκπαίδευσης.

Επισημαίνουμε εδώ, πως οι πίνακες με τις τιμές των χαρακτηριστικών επιτρέπεται να έχουν, μόνον, θετικές τιμές, από 1 έως 6. Επειδή όμως, πολλά χαρακτηριστικά παίρνουν κανονικά, σε ορισμένα frames, αρνητικές τιμές, πριν τη διαδικασία κβαντισμού, υπολογίζεται η απόλυτη τιμή των στοιχείων αυτών. Η διαδικασία αυτή, δεν οδηγεί σε απώλεια πληροφορίας, αφού ουσιαστικά είναι θέμα επιλογής - η οποία μπορεί να γίνει αυθαίρετα - τι σύμβολο θα χρησιμοποιηθεί για την κωδικοποίηση της τιμής του κάθε χαρακτηριστικού. Συνεπώς, μία αρνητική τιμή πχ. -246, επιλέγουμε να κωδικοποιείται με το σύμβολο '3' και όχι με το '-3'. Σε κάθε

ακολουθία, είτε εκπαίδευσης είτε αναγνώρισης, γίνεται η ίδια επεξεργασία και οι ίδιες τιμές αντιστοιχούν στα ίδια σύμβολα, οπότε, δεν υπάρχει περίπτωση να προκύψει λάθος κατά την εκπαίδευση ή κατά την αναγνώριση, εξ αιτίας της συγκεκριμένης διαδικασίας.

Αφού, πραγματοποιηθεί η παραπάνω διαδικασία προεπεξεργασίας των δεδομένων, καλείται η `hmmtrain` και δημιουργούνται για κάθε χειρονομία έξι μοντέλα, όσα είναι και τα χαρακτηριστικά γνωρίσματα, που χρησιμοποιούνται για την εκπαίδευση. Όπως είπαμε και προηγουμένως, η `hmmtrain`, υπολογίζει τις παραμέτρους του μοντέλου με χρήση του αλγόριθμου Baum – Welch. Ο κώδικας που χρησιμοποιήθηκε για την εκπαίδευση των έξι μοντέλων για την χειρονομία “wave” φαίνεται παρακάτω. Ο κώδικας για τις υπόλοιπες χειρονομίες είναι παρόμοιος και δεν παρατίθεται εδώ για λόγους συντομίας.

% αρχικοποίηση του πίνακα μεταβάσεων

```
wav_TRANS_GUESS=[0.4  0.3  0.1  0.1  0.06  0.04
                  0   0.4  0.3  0.1  0.1  0.1
                  0   0   0.5  0.3  0.1  0.1
                  0   0   0   0.5  0.3  0.2
                  0   0   0   0   0.7  0.3
                  0   0   0   0   0   1];
```

% αρχικοποίηση του πίνακα εξόδου

```
wav_EMIS_GUESS=ones(6,6)/6
```

% κβαντισμός των χαρακτηριστικών

```
wav_XL_XR1=ceil((abs(PAR1(1,:))+0.1)/82);
wav_XL_XR2=ceil((abs(PAR3(1,:))+0.1)/82);
wav_XL_XR3=ceil((abs(PAR4(1,:))+0.1)/82);
wav_XL_XR4=ceil((abs(PAR6(1,:))+0.1)/82);
wav_XL_XR5=ceil((abs(PAR7(1,:))+0.1)/82);
wav_XL_XR6=ceil((abs(PAR10(1,:))+0.1)/82);
wav_XL_XR7=ceil((abs(PAR13(1,:))+0.1)/82);
wav_XL_XR8=ceil((abs(PAR14(1,:))+0.1)/82);
```

```
wav_XR_XH1=ceil((abs(PAR1(2,:))+0.1)/82);
wav_XR_XH2=ceil((abs(PAR3(2,:))+0.1)/82);
wav_XR_XH3=ceil((abs(PAR4(2,:))+0.1)/82);
wav_XR_XH4=ceil((abs(PAR6(2,:))+0.1)/82);
wav_XR_XH5=ceil((abs(PAR7(2,:))+0.1)/82);
wav_XR_XH6=ceil((abs(PAR10(2,:))+0.1)/82);
wav_XR_XH7=ceil((abs(PAR13(2,:))+0.1)/82);
wav_XR_XH8=ceil((abs(PAR14(2,:))+0.1)/82);
```

```
wav_XH_XL1=ceil((abs(PAR1(3,:))+0.1)/82);
wav_XH_XL2=ceil((abs(PAR3(3,:))+0.1)/82);
wav_XH_XL3=ceil((abs(PAR4(3,:))+0.1)/82);
wav_XH_XL4=ceil((abs(PAR6(3,:))+0.1)/82);
```

```
wav_XH_XL5=ceil((abs(PAR7(3,:))+0.1)/82);
wav_XH_XL6=ceil((abs(PAR10(3,:))+0.1)/82);
wav_XH_XL7=ceil((abs(PAR13(3,:))+0.1)/82);
wav_XH_XL8=ceil((abs(PAR14(3,:))+0.1)/82);
```

```
wav_YR_YL1=ceil((abs(PAR1(4,:))+0.1)/82);
wav_YR_YL2=ceil((abs(PAR3(4,:))+0.1)/82);
wav_YR_YL3=ceil((abs(PAR4(4,:))+0.1)/82);
wav_YR_YL4=ceil((abs(PAR6(4,:))+0.1)/82);
wav_YR_YL5=ceil((abs(PAR7(4,:))+0.1)/82);
wav_YR_YL6=ceil((abs(PAR10(4,:))+0.1)/82);
wav_YR_YL7=ceil((abs(PAR13(4,:))+0.1)/82);
wav_YR_YL8=ceil((abs(PAR14(4,:))+0.1)/82);
```

```
wav_YH_YR1=ceil((abs(PAR1(5,:))+0.1)/82);
wav_YH_YR2=ceil((abs(PAR3(5,:))+0.1)/82);
wav_YH_YR3=ceil((abs(PAR4(5,:))+0.1)/82);
wav_YH_YR4=ceil((abs(PAR6(5,:))+0.1)/82);
wav_YH_YR5=ceil((abs(PAR7(5,:))+0.1)/82);
wav_YH_YR6=ceil((abs(PAR10(5,:))+0.1)/82);
wav_YH_YR7=ceil((abs(PAR13(5,:))+0.1)/82);
wav_YH_YR8=ceil((abs(PAR14(5,:))+0.1)/82);
```

```
wav_YH_YL1=ceil((abs(PAR1(6,:))+0.1)/82);
wav_YH_YL2=ceil((abs(PAR3(6,:))+0.1)/82);
wav_YH_YL3=ceil((abs(PAR4(6,:))+0.1)/82);
wav_YH_YL4=ceil((abs(PAR6(6,:))+0.1)/82);
wav_YH_YL5=ceil((abs(PAR7(6,:))+0.1)/82);
wav_YH_YL6=ceil((abs(PAR10(6,:))+0.1)/82);
wav_YH_YL7=ceil((abs(PAR13(6,:))+0.1)/82);
wav_YH_YL8=ceil((abs(PAR14(6,:))+0.1)/82);
```

%Δημιουργία των cell arrays με τους οποίους θα γίνει η εκπαίδευση

```
wav_XL_XR={wav_XL_XR7,wav_XL_XR1,wav_XL_XR2,wav_XL_XR3,wav_XL_XR4,wav_XL_XR5,wav_XL_XR6};
```

```
wav_XR_XH={wav_XR_XH7,wav_XR_XH1,wav_XR_XH2,wav_XR_XH3,wav_XR_XH4,wav_XR_XH5,wav_XR_XH6};
```

```
wav_XH_XL={wav_XH_XL7,wav_XH_XL1,wav_XH_XL2,wav_XH_XL3,wav_XH_XL4,wav_XH_XL5,wav_XH_XL6};
```

```
wav_YR_YL={wav_YR_YL7,wav_YR_YL1,wav_YR_YL2,wav_YR_YL3,wav_YR_YL4,wav_YR_YL5,wav_YR_YL6};
```

```
wav_YH_YR={wav_YH_YR7,wav_YH_YR1,wav_YH_YR2,wav_YH_YR3,wav_YH_YR4,wav_YH_YR5,wav_YH_YR6};
```



```
wav_YH_YL={wav_YH_YL7,wav_YH_YL1,wav_YH_YL2,wav_YH_YL3,wav_YH_YL4,wav_YH_YL5,wav_YH_YL6};
```

%εκπαίδευση των μοντέλων των χαρακτηριστικών

```
[wav_TRANS1,wav_EMIS1]=hmmtrain(wav_XL_XR,wav_TRANS_GUESS,wav_EMIS_GUESS,'Tolerance',0.01);
```

```
[wav_TRANS2,wav_EMIS2]=hmmtrain(wav_XR_XH,wav_TRANS_GUESS,wav_EMIS_GUESS,'Tolerance',0.01);
```

```
[wav_TRANS3,wav_EMIS3]=hmmtrain(wav_XH_XL,wav_TRANS_GUESS,wav_EMIS_GUESS,'Tolerance',0.01);
```

```
[wav_TRANS4,wav_EMIS4]=hmmtrain(wav_YR_YL,wav_TRANS_GUESS,wav_EMIS_GUESS,'Tolerance',0.01);
```

```
[wav_TRANS5,wav_EMIS5]=hmmtrain(wav_YH_YR,wav_TRANS_GUESS,wav_EMIS_GUESS,'Tolerance',0.01);
```

```
[wav_TRANS6,wav_EMIS6]=hmmtrain(wav_YH_YL,wav_TRANS_GUESS,wav_EMIS_GUESS,'Tolerance',0.01);
```

Ο πίνακας PAR είναι ένας δισδιάστατος πίνακας με έξι γραμμές και τόσες στήλες, όσες τα frame της αντίστοιχης ακολουθίας. Ο αριθμός δίπλα στον PAR είναι το νούμερο της ακολουθίας από την επεξεργασία της οποίας, προήλθαν οι τιμές των χαρακτηριστικών ,που υπάρχουν στον PAR. Από τους πίνακες PAR, δημιουργούνται, σύμφωνα με τη διαδικασία, που περιγράψαμε παραπάνω (απόλυτη τιμή, κβαντισμός), οι πίνακες, που χρησιμοποιούνται, ως συστατικά στοιχεία της **seq**. Κατόπιν, δημιουργούνται οι **seq** για κάθε χαρακτηριστικό και, τέλος, καλείται η **hmmtrain** έξι φορές και δημιουργούνται τα μοντέλα των χαρακτηριστικών για τη χειρονομία “wave”.

7.2.2 Αποκωδικοποίηση

Αφού γίνει η εκπαίδευση των μοντέλων όλων των χειρονομιών, είμαστε έτοιμοι για την διαδικασία της αποκωδικοποίησης. Κατά την διαδικασία της αποκωδικοποίησης, μας δίνεται μια ακολουθία, η οποία δεν γνωρίζουμε εάν έχει προέλθει από κάποια από τις επτά χειρονομίες, των οποίων τα μοντέλα εκπαιδεύσαμε αλλά και σε περίπτωση, που ανήκει σε κάποια από αυτές, δεν γνωρίζουμε σε ποια. Σκοπός της διαδικασίας αποκωδικοποίησης είναι να αναγνωρίσει εάν η χειρονομία, που αντιπροσωπεύεται από την προς αναγνώριση ακολουθία, είναι δυνατόν να έχει προέλθει από κάποιο από τα εκπαιδευμένα μοντέλα και εάν ναι, από ποιο.

Τα χαρακτηριστικά γνωρίσματα της προς αναγνώριση ακολουθίας, βρίσκονται στη μορφή που προκύπτουν από την εφαρμογή του αλγόριθμου εντοπισμού στο προς αναγνώριση video. Για το λόγο αυτό, υφίστανται την προεπεξεργασία που υπέστησαν και οι ακολουθίες που χρησιμοποιήθηκαν για την

εκπαίδευση των μοντέλων. Με τον τρόπο αυτό εξασφαλίζεται ότι τα χαρακτηριστικά θα είναι στην ίδια μορφή, ώστε να είναι συγκρίσιμα και σωστά τα αποτελέσματα.

Ο υπολογισμός των πιθανοτήτων, η προς αναγνώριση ακολουθία, να έχει προέλθει από καθένα από τα επτά μοντέλα, γίνεται με χρήση της `hmmdecode`.

Η `hmmdecode` υπολογίζει την πιθανότητα η ακολουθία εισόδου `seq` να έχει προέλθει από το συγκεκριμένο μοντέλο με πίνακες μετάβασης και εξόδου TRANS και EMIS αντίστοιχα και συντάσσεται ως εξής

$$[\text{PSTATES}, \text{logpseq}] = \text{hmmdecode}(\text{seq}, \text{TRANS}, \text{EMIS})$$

Η `hmmdecode` υπολογίζει τις εκ των υστέρων πιθανότητες για τις καταστάσεις, PSTATES, της ακολουθίας `seq`. Οι εκ των υστέρων πιθανότητες είναι οι δεσμευμένες πιθανότητες, το μοντέλο να βρίσκεται στην κατάσταση k στο i βήμα, δεδομένης της συμβολοακολουθίας εισόδου. Το μοντέλο καθορίζεται, όπως είπαμε, από τον πίνακα μετάβασης TRANS και εκπομπής EMIS. TRANS(i,j) είναι η πιθανότητα να έχουμε μετάβαση από την κατάσταση i στην κατάσταση j . EMIS(k, sym) είναι η πιθανότητα να έχουμε παρατήρηση του συμβόλου `sym` ενώ το μοντέλο βρίσκεται στην κατάσταση k . Ο PSTATES είναι ένας πίνακας, που έχει τις ίδιες στήλες με την `seq` και μία γραμμή για κάθε κατάσταση του μοντέλου. Το (i, j) στοιχείο του PSTATES αντιπροσωπεύει την πιθανότητα το μοντέλο να βρίσκεται στην κατάσταση i στο j βήμα, δεδομένης της ακολουθίας `seq`.

Σημειώνουμε εδώ ότι η `hmmdecode` ξεκινάει στο βήμα 0 από την κατάσταση 1 του μοντέλου, χωρίς να έχει γίνει ακόμα η εκπομπή του πρώτου συμβόλου παρατήρησης. Η μεταβλητή `logpseq`, που επιστρέφει η `hmmdecode` αποτελεί τον λογάριθμο της πιθανότητας η ακολουθία `seq`, να έχει προέλθει από το μοντέλο με πίνακα μετάβασης TRANS και πίνακα εκπομπής EMIS. Ουσιαστικά ο λογάριθμος της πιθανότητας αυτής αποτελεί το κριτήριο για την ταξινόμηση της ακολουθίας εισόδου και την κατάταξή της σε μία από τις υπάρχουσες χειρονομίες ή μη. Οι τιμές που παίρνει θεωρητικά η μεταβλητή αυτή είναι από $-\text{Inf}$ έως 0. Το $-\text{Inf}$ αντιστοιχεί σε μηδενική πιθανότητα, δηλαδή είναι ουσιαστικά αδύνατο το συγκεκριμένο HMM να έχει γεννήσει την ακολουθία `seq` ενώ το 0 αντιστοιχεί σε πιθανότητα 1, δηλαδή βεβαιότητα πως η `seq` έχει προέλθει από το συγκεκριμένο HMM.

Βεβαίως, πιθανότητα 1, δεν υπάρχει περίπτωση να συναντήσουμε για οποιαδήποτε ακολουθία, ακόμα και αν ακολουθία αναγνώρισης είχε προηγουμένως χρησιμοποιηθεί και για την εκπαίδευση του μοντέλου, πράγμα το οποίο δεν πρόκειται βεβαίως να συμβεί κατά την πραγματική λειτουργία του συστήματος. Το γεγονός αυτό οφείλεται στον αριθμό των καταστάσεων και των εξόδων που είναι και στις δύο περιπτώσεις έξι, ενώ και στις ακολουθίες που χρησιμοποιούνται για την εκπαίδευση του μοντέλου, δεν έχουμε αναγκαστικά ακριβώς την ίδια αλληλουχία εξόδων, αλλά παραπλήσιες. Έτσι, στους πίνακες μετάβασης και εκπομπής των μοντέλων, δεν συναντάμε ποτέ την τιμή 1 και στους δύο μαζί, πράγμα το οποίο θα οδηγούσε και σε μοναδιαία πιθανότητα, για κάποια `seq`, η οποία θα περιγραφόταν ακριβώς από αυτούς τους πίνακες (πχ. μία ακολουθία για ένα χαρακτηριστικό που έχει συνεχώς την τιμή '5', θα έδινε πιθανότητα 1, εάν το HMM είχε στον πίνακα μετάβασης 1 στην πρώτη στήλη και 0 σε όλες τις άλλες και έξοδο με πιθανότητα 1, το σύμβολο '5', όταν το HMM βρίσκεται στην κατάσταση 1. Συγκεκριμένα, σε μια τέτοια περίπτωση, η οποία όμως δεν πρόκειται να υπάρξει για τους λόγους που αναφέραμε, η `hmmdecode` θα ξεκινούσε στο βήμα 0 από την κατάσταση 1, από την οποία θα είχαμε πάντα έξοδο '5' –αφού στον EMIS η πιθανότητα να έχουμε έξοδο '5' είναι 1 – ενώ και στα

επόμενα βήματα το μοντέλο θα παρέμενε στην κατάσταση 1 –αφού η πιθανότητα να παραμείνει στην κατάσταση 1 στον TRANS είναι 1-).

Όπως είπαμε, για κάθε χειρονομία έχουμε ένα HMM το οποίο αντιπροσωπεύει καθένα από τα έξι χαρακτηριστικά γνωρίσματα της χειρονομίας. Συνεπώς για να συγκρίνουμε κάθε ακολουθία με τη συγκεκριμένη χειρονομία πρέπει να γίνει κλήση της `hmmdecode` έξι φορές – μία για κάθε χαρακτηριστικό-. Η `seq` σε κάθε κλήση της `hmmdecode` αντιστοιχεί στο χαρακτηριστικό της συγκεκριμένης ακολουθίας, με το μοντέλο της οποίας γίνεται η αποκωδικοποίηση κάθε φορά. Ο λογάριθμος της συνολικής πιθανότητας για την κάθε ακολουθία να ανήκει σε κάποια χειρονομία, υπολογίζεται με πρόσθεση των λογαρίθμων που προκύπτουν από την `hmmdecode` για καθένα από τα έξι χαρακτηριστικά της χειρονομίας (η συνολική πιθανότητα, ως γνωστόν, υπολογίζεται με πολλαπλασιασμό των επιμέρους πιθανοτήτων και άρα ο συνολικός λογάριθμος υπολογίζεται με πρόσθεση των αντίστοιχων λογαρίθμων). Η χειρονομία της οποία ο συνολικός λογάριθμος βρίσκεται πιο κοντά στο 0 (δηλαδή η συνολική πιθανότητα βρίσκεται πιο κοντά στο 1) είναι η χειρονομία στην οποία θεωρεί το σύστημα πως ανήκει η ακολουθία εισόδου. Σε περίπτωση που το αποτέλεσμα είναι `-inf` για όλες τις χειρονομίες, θεωρούμε πως η ακολουθία εισόδου δεν ανήκει σε καμία από τις χειρονομίες τις οποίες έχει εκπαιδευθεί να αναγνωρίζει το σύστημα.

Ο κώδικας με τον οποίον γίνεται η αποκωδικοποίηση μιας ακολουθίας εισόδου, με όλα τα μοντέλα χειρονομιών, που υπάρχουν, είναι ο ακόλουθος

```
function [logprob]= ges_decode(PAR)
```

```
%xorismos tis dosmenis akolouthias stis akolouthies ton parametron kai  
%prosartisi tous stous kanones tou codebook
```

```
XL_XR=ceil((abs(PAR(1,:))+0.1)/82);  
XR_XH=ceil((abs(PAR(2,:))+0.1)/82);  
XH_XL=ceil((abs(PAR(2,:))+0.1)/82);  
YR_YL=ceil((abs(PAR(4,:))+0.1)/82);  
YH_YR=ceil((abs(PAR(5,:))+0.1)/82);  
YH_YL=ceil((abs(PAR(6,:))+0.1)/82);
```

```
seq1=XL_XR;  
seq2=XR_XH;  
seq3=XH_XL;  
seq4=YR_YL;  
seq5=YH_YR;  
seq6=YH_YL;
```

```
%ipologismos tis pithanotitas i akolouthia na einai i "1-hand"
```

```
[one_PSTATES1, one_logpseq1] = hmmdecode(seq1, one_TRANS1,  
one_EMIS1);  
[one_PSTATES2, one_logpseq2] = hmmdecode(seq2, one_TRANS2,  
one_EMIS2);  
[one_PSTATES3, one_logpseq3] = hmmdecode(seq3, one_TRANS3,  
one_EMIS3);
```

```

    [one_PSTATES4, one_logpseq4] = hmmdecode(seq4, one_TRANS4,
one_EMIS4);
    [one_PSTATES5, one_logpseq5] = hmmdecode(seq5, one_TRANS5,
one_EMIS5);
    [one_PSTATES6, one_logpseq6] = hmmdecode(seq6, one_TRANS6,
one_EMIS6);

    A=[one_logpseq1 one_logpseq2 one_logpseq3 one_logpseq4 one_logpseq5
one_logpseq6];
    one_NaN=isnan(A);
    for i=1:6
        if (one_NaN(i)==1) A(i)=-500; end
    end

one_logprop=sum(A);

```

%ipologismos tis pithanotitas i akolouthia na einai i "attention"

```

[at_PSTATES1, at_logpseq1] = hmmdecode(seq1, at_TRANS1, at_EMIS1);
[at_PSTATES2, at_logpseq2] = hmmdecode(seq2, at_TRANS2, at_EMIS2);
[at_PSTATES3, at_logpseq3] = hmmdecode(seq3, at_TRANS3, at_EMIS3);
[at_PSTATES4, at_logpseq4] = hmmdecode(seq4, at_TRANS4, at_EMIS4);
[at_PSTATES5, at_logpseq5] = hmmdecode(seq5, at_TRANS5, at_EMIS5);
[at_PSTATES6, at_logpseq6] = hmmdecode(seq6, at_TRANS6, at_EMIS6);

B=[at_logpseq1 at_logpseq2 at_logpseq3 at_logpseq4 at_logpseq5 at_logpseq6];
at_NaN=isnan(B);
for i=1:6
    if (at_NaN(i)==1) B(i)=-500; end
end

at_logprop=sum(B);

```

%ipologismos tis pithanotitas i akolouthia na einai i "clapping"

```

[cl_PSTATES1, cl_logpseq1] = hmmdecode(seq1, cl_TRANS1, cl_EMIS1);
[cl_PSTATES2, cl_logpseq2] = hmmdecode(seq2, cl_TRANS2, cl_EMIS2);
[cl_PSTATES3, cl_logpseq3] = hmmdecode(seq3, cl_TRANS3, cl_EMIS3);
[cl_PSTATES4, cl_logpseq4] = hmmdecode(seq4, cl_TRANS4, cl_EMIS4);
[cl_PSTATES5, cl_logpseq5] = hmmdecode(seq5, cl_TRANS5, cl_EMIS5);
[cl_PSTATES6, cl_logpseq6] = hmmdecode(seq6, cl_TRANS6, cl_EMIS6);

C=[cl_logpseq1 cl_logpseq2 cl_logpseq3 cl_logpseq4 cl_logpseq5 cl_logpseq6];
cl_NaN=isnan(C);
for i=1:6
    if (cl_NaN(i)==1) C(i)=-500; end
end

```

```
cl_logprop=sum(C);
```

```
%ipologismos tis pithanotitas i akolouthia na einai i "ai sto diaolo"
```

```
[dia_PSTATES1, dia_logpseq1] = hmmdecode(seq1, dia_TRANS1, dia_EMIS1);  
[dia_PSTATES2, dia_logpseq2] = hmmdecode(seq2, dia_TRANS2, dia_EMIS2);  
[dia_PSTATES3, dia_logpseq3] = hmmdecode(seq3, dia_TRANS3, dia_EMIS3);  
[dia_PSTATES4, dia_logpseq4] = hmmdecode(seq4, dia_TRANS4, dia_EMIS4);  
[dia_PSTATES5, dia_logpseq5] = hmmdecode(seq5, dia_TRANS5, dia_EMIS5);  
[dia_PSTATES6, dia_logpseq6] = hmmdecode(seq6, dia_TRANS6, dia_EMIS6);
```

```
D=[dia_logpseq1 dia_logpseq2 dia_logpseq3 dia_logpseq4 dia_logpseq5  
dia_logpseq6];  
dia_NaN=isnan(D);  
for i=1:6  
    if (dia_NaN(i)==1) D(i)=-500; end  
end
```

```
dia_logprop=sum(D);
```

```
%ipologismos tis pithanotitas i akolouthia na einai i "explain"
```

```
[exp_PSTATES1, exp_logpseq1] = hmmdecode(seq1, exp_TRANS1,  
exp_EMIS1);  
[exp_PSTATES2, exp_logpseq2] = hmmdecode(seq2, exp_TRANS2,  
exp_EMIS2);  
[exp_PSTATES3, exp_logpseq3] = hmmdecode(seq3, exp_TRANS3,  
exp_EMIS3);  
[exp_PSTATES4, exp_logpseq4] = hmmdecode(seq4, exp_TRANS4,  
exp_EMIS4);  
[exp_PSTATES5, exp_logpseq5] = hmmdecode(seq5, exp_TRANS5,  
exp_EMIS5);  
[exp_PSTATES6, exp_logpseq6] = hmmdecode(seq6, exp_TRANS6,  
exp_EMIS6);
```

```
E=[exp_logpseq1 exp_logpseq2 exp_logpseq3 exp_logpseq4 exp_logpseq5  
exp_logpseq6];  
exp_NaN=isnan(E);  
for i=1:6  
    if (exp_NaN(i)==1) E(i)=-500; end  
end
```

```
exp_logprop=sum(E);
```

```
%ipologismos tis pithanotitas i akolouthia na einai i "oh my god"
```

```

    [god_PSTATES1, god_logpseq1] = hmmdecode(seq1, god_TRANS1,
god_EMIS1);
    [god_PSTATES2, god_logpseq2] = hmmdecode(seq2, god_TRANS2,
god_EMIS2);
    [god_PSTATES3, god_logpseq3] = hmmdecode(seq3, god_TRANS3,
god_EMIS3);
    [god_PSTATES4, god_logpseq4] = hmmdecode(seq4, god_TRANS4,
god_EMIS4);
    [god_PSTATES5, god_logpseq5] = hmmdecode(seq5, god_TRANS5,
god_EMIS5);
    [god_PSTATES6, god_logpseq6] = hmmdecode(seq6, god_TRANS6,
god_EMIS6);

    F=[god_logpseq1 god_logpseq2 god_logpseq3 god_logpseq4 god_logpseq5
god_logpseq6];
    god_NaN=isnan(F);
    for i=1:6
        if (god_NaN(i)==1) F(i)=-500; end
    end

god_logprop=sum(F);

```

%ipologismos tis pithanotitas i akolouthia na einai i "wave"

```

    [wav_PSTATES1, wav_logpseq1] = hmmdecode(seq1, wav_TRANS1,
wav_EMIS1);
    [wav_PSTATES2, wav_logpseq2] = hmmdecode(seq2, wav_TRANS2,
wav_EMIS2);
    [wav_PSTATES3, wav_logpseq3] = hmmdecode(seq3, wav_TRANS3,
wav_EMIS3);
    [wav_PSTATES4, wav_logpseq4] = hmmdecode(seq4, wav_TRANS4,
wav_EMIS4);
    [wav_PSTATES5, wav_logpseq5] = hmmdecode(seq5, wav_TRANS5,
wav_EMIS5);
    [wav_PSTATES6, wav_logpseq6] = hmmdecode(seq6, wav_TRANS6,
wav_EMIS6);

    G=[wav_logpseq1 wav_logpseq2 wav_logpseq3 wav_logpseq4 wav_logpseq5
wav_logpseq6];
    wav_NaN=isnan(G);
    for i=1:6
        if (wav_NaN(i)==1) G(i)=-500; end
    end

wav_logprop=sum(G);

```

```

logprob(1)=one_logprop;
logprob(2)=at_logprop;
logprob(3)=cl_logprop;
logprob(4)=dia_logprop;
logprob(5)=exp_logprop;
logprob(6)=god_logprop;
logprob(7)=wav_logprop;

```

```

sprintf('I pithanotita i dosmeni akolouthia na einai i "1-hand" einai %d',logprob(1))
sprintf('I pithanotita i dosmeni akolouthia na einai i "attention" einai %d',logprob(2))
sprintf('I pithanotita i dosmeni akolouthia na einai i "clapping" einai %d',logprob(3))
sprintf('I pithanotita i dosmeni akolouthia na einai i "ai sto diaolo" einai
%d',logprob(4))
sprintf('I pithanotita i dosmeni akolouthia na einai i "explain" einai %d',logprob(5))
sprintf('I pithanotita i dosmeni akolouthia na einai i "oh my god" einai
%d',logprob(6))
sprintf('I pithanotita i dosmeni akolouthia na einai i "wave" einai %d',logprob(7))

```

7.3 Πειραματικά Αποτελέσματα

Μετά την εκπαίδευση των μοντέλων, πραγματοποιήθηκαν πειράματα για να διαπιστωθεί το ποσοστό επιτυχία της υλοποίησης. Για κάθε χειρονομία υπήρχε ένας αριθμός διαθέσιμων ακολουθιών (διαφορετικός για κάθε χειρονομία). Οι ακολουθίες αυτές προέκυψαν από τα video, στα οποία συμμετείχαν έξι διαφορετικοί χρήστες οι οποίοι πραγματοποιούσαν όλες τις χειρονομίες, με εφαρμογή του αλγόριθμου εντοπισμού και την επεξεργασία (κβαντισμός) που περιγράφηκε σε προηγούμενη ενότητα. Τα αποτελέσματα παρατίθενται στον παρακάτω πίνακα :

Χειρονομία	<i>Attention</i>	<i>Clapping</i>	<i>Explain</i>	<i>One hand</i>	<i>Oh my god</i>	<i>Wave</i>	<i>Go away</i>	<i>Unclassified</i>
<i>Attention</i>	10	0	0	0	0	0	1	0
<i>Clapping</i>	0	7	0	0	0	0	0	1
<i>Explain</i>	0	0	4	0	0	0	0	1
<i>One hand</i>	0	0	0	6	0	0	0	0
<i>Oh my god</i>	0	0	0	0	10	0	0	1
<i>Wave</i>	2	0	0	0	0	6	0	0
<i>Go away</i>	1	0	0	0	0	1	6	0
Ποσοστό αναγνώρισης (%)	90,9	87,5	80	100	90,9	75	75	

Όπως παρατηρούμε τα πιο ικανοποιητικά αποτελέσματα αναγνώρισης υπάρχουν στην attention ενώ λιγότερο ικανοποιητικά είναι τα αποτελέσματα που προκύπτουν για τις wave και explain. Οι παράγοντες, οι οποίοι συντελούν σε αυτό, αναλύονται στην αντίστοιχη ενότητα της συνεχούς υλοποίησης. Επειδή η ανάλυση είναι κοινή και για τις δύο περιπτώσεις (συνεχές και διακριτό σύστημα) , στο σημείο αυτό δεν θα επεκταθούμε περαιτέρω.

Η μοναδική παρατήρηση, η οποία αφορά μόνο την υλοποίηση με διακριτά HMM είναι, ότι οι περισσότερες ακολουθίες, που δεν αναγνωρίζονται επιτυχώς, το πιθανότερο είναι να τεθούν στην κατηγορία unclassified. Σε αντίθεση με το συνεχές σύστημα, στο διακριτό, όταν μια ακολουθία παρουσιάσει χαμηλή πιθανότητα στο μοντέλο από το οποίο έχει στην πραγματικότητα προέλθει, είναι πολύ δύσκολο να παρουσιάσει μεγαλύτερη πιθανότητα σε κάποιο άλλο μοντέλο, οπότε και να ταξινομηθεί λανθασμένα σε κάποια άλλη χειρονομία. Αυτό συμβαίνει διότι στην διακριτή υλοποίηση τα χαρακτηριστικά γνωρίσματα αποτελούνται από ακολουθίες συμβόλων. Μάλιστα, άπαξ και στην προς αναγνώριση ακολουθία υπάρχει έστω και ένα σύμβολο, σε οποιοδήποτε από τα έξι χαρακτηριστικά, το οποίο δεν υπήρχε σε καμία από τις ακολουθίες που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου, η πιθανότητα που προκύπτει είναι πάρα πολύ μικρή (ουσιαστικά 0). Συνεπώς, είναι σχεδόν απίθανο, κάποια ακολουθία που έχει προέλθει από κάποια συγκεκριμένη χειρονομία, να μην περιέχει ούτε ένα σύμβολο διαφορετικό από αυτά με βάση τα οποία εκτιμήθηκαν οι παράμετροι των υπόλοιπων μοντέλων (εκτός αυτού που αντιπροσωπεύει τη χειρονομία από την οποία όντως προήλθε η ακολουθία).

Βεβαίως, υπάρχουν κάποιες εξαιρέσεις, όπως στην περίπτωση της wave, η οποία μέχρι ενός σημείου (ανύψωση του δεξιού χεριού) είναι σχεδόν ίδια με την attention. Αυτό έχει σαν αποτέλεσμα να περιέχουν τα ίδια σύμβολα και συνεπώς το μοντέλο της attention να δίνει αυξημένη πιθανότητα όταν αποκωδικοποιείται μια ακολουθία που έχει προέλθει από χειρονομία wave .

ΚΕΦΑΛΑΙΟ 8

ΥΛΟΠΟΙΗΣΗ ΜΕ ΣΥΝΕΧΗ HMM

8.1 HMM με συνεχείς Κατανομές Εξόδου και Γκαουσιανά Μείγματα

Ο δεύτερος τρόπος με τον οποίο έγινε η ταξινόμηση – αναγνώριση των χειρονομιών βασίστηκε, όπως είπαμε, σε συνεχή HMM. Τα συνεχή HMM είναι HMM με συνεχείς συναρτήσεις κατανομής των εξόδων τους. Με τη χρήση συνεχών κατανομών εξόδου, μας παρέχεται η δυνατότητα, τα χαρακτηριστικά που προκύπτουν από την διαδικασία εντοπισμού της τροχιάς, να τα παράσχουμε ως είσοδο στα HMM χωρίς καμία διαδικασία προεπεξεργασίας και κβαντοποίησης, αντίστοιχη με αυτή που χρειάστηκε κατά την εκπαίδευση αλλά και αναγνώριση με διακριτά HMM. Με αυτό τον τρόπο αποφεύγονται τυχόν απώλειες χρήσιμων δεδομένων. Βεβαίως, για την εκπαίδευση συνεχών HMM απαιτείται μεγαλύτερος αριθμός ακολουθιών, προκειμένου να μπορέσει ο αλγόριθμος Εκτίμησης – Μεγιστοποίησης να δώσει ικανοποιητικά αποτελέσματα για τις παραμέτρους του μοντέλου.

Η μορφή που έχει η κατανομή εξόδου $b_j(o_t)$ στα συνεχή HMM είναι είτε αυτή μιας συνεχούς συνάρτησης πυκνότητας πιθανότητας, είτε ένα μείγμα συνεχών συναρτήσεων πυκνότητας πιθανότητας, δηλαδή:

$$b_j(o_t) = \sum_{k=1}^M w_{jk} b_{jk}(o_t), \quad j=1, \dots, N$$

όπου M είναι ο αριθμός των μειγμάτων και w είναι ο συντελεστής βάρους για το κάθε μείγμα. Για τους συντελεστές βαρών των μειγμάτων ισχύουν οι ακόλουθοι περιορισμοί

$$\sum_{k=1}^M w_{jk} = 1, \quad j=1, \dots, N; \quad w_{jk} \geq 0, \quad j=1, \dots, N; \quad k=1, \dots, M$$

κάθε μια από τις συναρτήσεις $b_{jk}(o_t)$ είναι μια κατανομή λογαριθμικής ή ελλειπτικά συμμετρικής μορφής, D διαστάσεων, με διάνυσμα μέσης τιμής μ_{jk} και πίνακα συσχετιστικότητας Σ_{jk} :

$$b_{jk}(o_t) = N(o_t, \mu_{jk}, \Sigma_{jk})$$

Κατά την διαδικασία της εκπαίδευσης ενός συνεχούς μοντέλου, πρέπει να γίνει η εκτίμηση των παραμέτρων αυτών των συναρτήσεων πυκνότητας πιθανότητας. Για το λόγο αυτό, οι ακολουθίες εισόδου, πρέπει αφενός μεν να είναι όσο το δυνατόν περισσότερες και αφετέρου, πρέπει να παρουσιάζουν όσο το δυνατόν μεγαλύτερη συσχέτιση.

Ένα ακόμα πρόβλημα στα συνεχή HMM είναι αυτό της επιλογής της κατάλληλης συνάρτησης πυκνότητας πιθανότητας για κάθε μείγμα. Οι πιο συνηθισμένες κατανομές που χρησιμοποιούνται είναι η Γκαουσιανή και η κανονική κατανομή.

Η μορφή που έχει η συνάρτηση μιας απλής Γκαουσιανής κατανομής είναι η ακόλουθη :

$$b_{jk}(o_t) = N(o_t, \mu_{jk}, \Sigma_{jk}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

και το o_t είναι πραγματικός αριθμός.

8.1.1 Ο Αλγόριθμος Baum – Welch για Γκαουσιανά Μείγματα

Όπως είπαμε, η πιο ευρέως χρησιμοποιούμενη συνεχής συνάρτηση πυκνότητας πιθανότητας, είναι το μείγμα Γκαουσιανών κατανομών της μορφής

$$b_j(o_t) = \sum_{k=1}^M w_{jk} b_{jk}(o_t), \quad j=1, \dots, N$$

όπου o_t είναι το διάνυσμα (ακολουθία παρατήρησης) και w_{im} είναι ο συντελεστής βάρους του m μείγματος στην i κατάσταση. Οι συντελεστές βαρών των μειγμάτων υπόκεινται στους ακόλουθους περιορισμούς

$$\begin{aligned} \sum_{k=1}^M w_{jk} &= 1, \quad j=1, \dots, N \\ w_{jk} &\geq 0, \quad j=1, \dots, N \\ & \quad k=1, \dots, M \end{aligned}$$

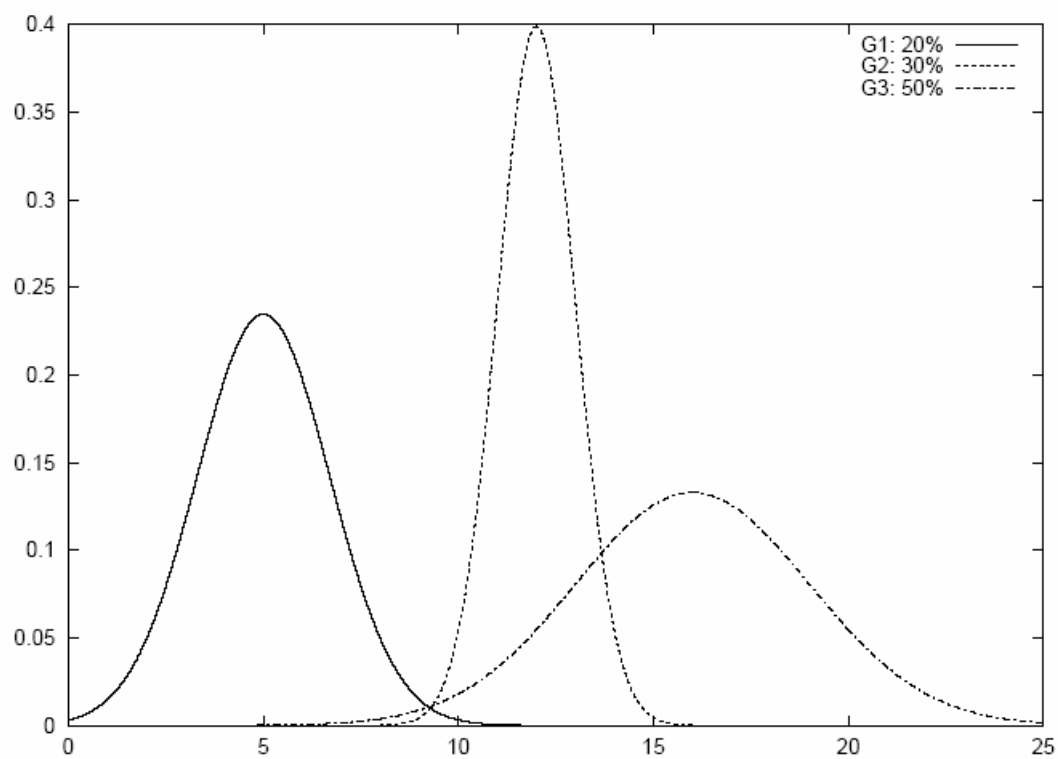
Έστω $N(O, \mu_{im}, \Sigma_{im})$ μια πολυμεταβλητή Γκαουσιανή κατανομή με διάνυσμα μέσης τιμής μ' και πίνακα συμμεταβλητότητας Σ' :

$$N(O, \mu_{im}, \Sigma_{im}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma'|}} \exp\left(-\frac{1}{2}(o - \mu')\Sigma'^{-1}(o - \mu')\right)$$

Στην περίπτωση αυτή, η εκτίμηση των παραμέτρων των HMM συνεπάγεται την εκτίμηση των παραμέτρων (της μέσης τιμής και της μεταβλητότητας) κάθε Γκαουσιανού μείγματος. Η ενδιάμεση μεταβλητή δ ορίζεται ως εξής:

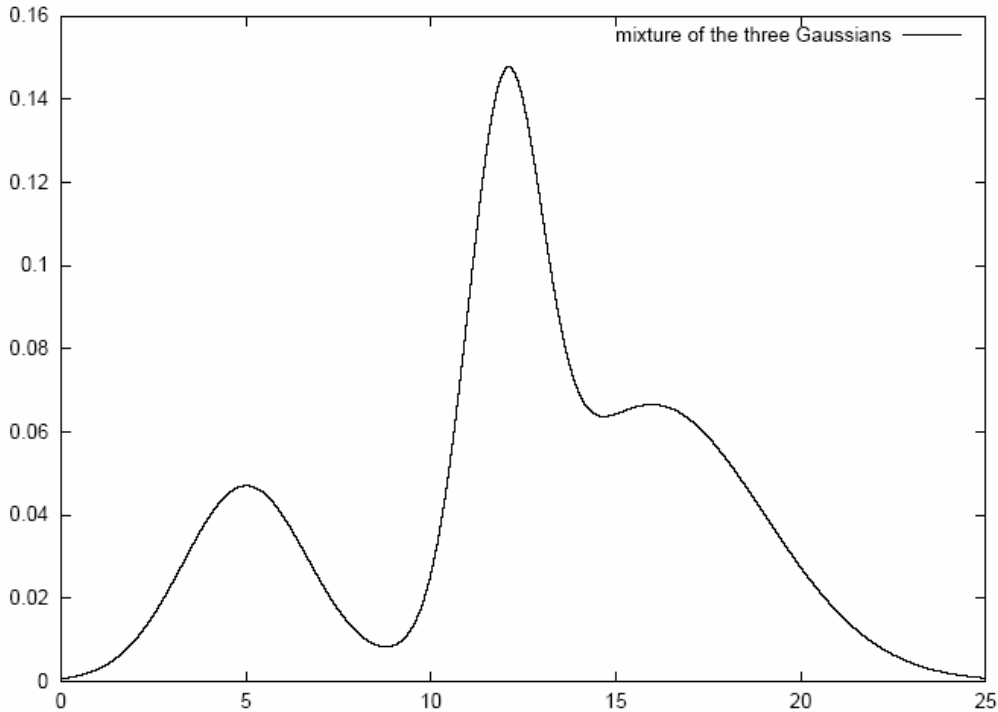
$$\delta_i(t) = \sum_{m=1}^M \delta_{im}(t) = \sum_{m=1}^M \frac{1}{p} a_j(t-1) a_{ji} w_{im} b_{im}(o_t) \beta_i(t)$$

όπως προαναφέραμε M είναι ο συνολικός αριθμός των κατανομών που χρησιμοποιούνται για το σχηματισμό του μείγματος στην i κατάσταση και N είναι ο συνολικός αριθμός των καταστάσεων. Η διαφορά με την προηγούμενη περίπτωση είναι ότι κάθε παρατήρηση στην ακολουθία εξόδου, δεν σχετίζεται με μία μόνο συνάρτηση πυκνότητας πιθανότητας αλλά στη δημιουργία της συνεισφέρουν, όλες οι κατανομές που αποτελούν το γκαουσιανό μείγμα, η κάθε μια με το δικό της συντελεστή βάρους.



Σχημα

Στο παραπάνω σχήμα απεικονίζονται τρεις απλές γκαουσιανές κατανομές, κάθε μια από τις οποίες θα συνεισφέρει κατά ένα ποσοστό (ανάλογα με το συντελεστή βάρους που της αντιστοιχεί) στη δημιουργία της ακολουθίας εξόδου.



Σχήμα

Ένα Γκαουσιανό μείγμα, το οποίο προέκυψε με χρήση των τριών Γκαουσιανών κατανομών του προηγούμενου σχήματος. Για κάθε μικρή περιοχή δ , γύρω από κάθε σημείο χ του άξονα των χ , υπάρχει πιθανότητα 20% η τυχαία μεταβλητή χ , να δημιουργείται με χρήση της πρώτης Γκαουσιανή κατανομή του μείγματος, 30% με χρήση της δεύτερης και 50% από με χρήση της τρίτης.

Μια ακόμα μεταβλητή, που είναι απαραίτητη για τον πλήρη ορισμό της μεθόδου προσδιορισμού της ακολουθίας εξόδου με γκαουσιανά μείγματα, είναι η μεταβλητή γ , η οποία εκφράζει την πιθανότητα το μοντέλο να βρίσκεται στην κατάσταση j την χρονική στιγμή t και η παρατήρηση o_t να δημιουργείται με συνεισφορά του k μείγματος. Η μεταβλητή γ ορίζεται ως :

$$\gamma_{j,k}(t) = \frac{a_j(t)\beta_j(t)}{\sum_{j=1}^N a_j(t)\beta_j(t)} \cdot \frac{w_{jk}N(o_t, \mu_{jk}, \Sigma_{jk})}{\sum_{k=1}^M w_{jk}N(o_t, \mu_{jk}, \Sigma_{jk})}$$

Με χρήση της ενδιάμεσης μεταβλητής δ , οι τύποι που χρησιμοποιούνται για τον επαναληπτικό υπολογισμό των παραμέτρων του μείγματος, παίρνουν την ακόλουθη μορφή:

$$\mu'_{im} = \frac{\sum_{t=1}^T \delta_{im}(t)o_t}{\sum_{t=1}^T \delta_{im}(t)}$$

$$\sigma'_{im} = \frac{\sum_{t=1}^T \delta_{im}(t)(o_t - \mu'_{im})(o_t - \mu'_{im})}{\sum_{t=1}^T \delta_{im}(t)}$$

$$w_{im} = - \frac{\sum_{t=1}^T \delta_{im}(t)}{\sum_{t=1}^T \delta_i(t)}$$

$$\alpha'_{ij} = \frac{\sum_{t=1}^T a_i(t)a_{ij}b_j(o_{t+1})\beta_j(t+1)}{\sum_{t=1}^T a_i(t)\beta_i(t)}$$

8.2 Υλοποίηση

Για την υλοποίηση με χρήση συνεχών HMM χρησιμοποιήσαμε το toolbox του Kevin Merhy για το Matlab. Το σημαντικότερο πλεονέκτημα της υλοποίησης των HMM με χρήση του συγκεκριμένου toolbox σε σχέση με την υλοποίηση, που πραγματοποιήσαμε με χρήση του toolbox του Matlab, Stats 4.1, είναι η δυνατότητα που παρέχει να γίνεται η εκπαίδευση του HMM με περισσότερα του ενός χαρακτηριστικά, απευθείας, χωρίς να χρειάζεται να δημιουργηθεί διαφορετικό HMM για το καθένα από τα χαρακτηριστικά της ίδιας χειρονομίας.

Συγκεκριμένα, σχηματίζεται ο πίνακας *data*. Ο πίνακας αυτός είναι τρισδιάστατος και περιλαμβάνει όλα τα χαρακτηριστικά (και τα έξι), όλων των ακολουθιών εξόδου, κάθε χειρονομίας. Ο αριθμός των γραμμών του *data* είναι έξι, όσα δηλαδή και τα χαρακτηριστικά. Ο αριθμός των στηλών του είναι ίσος με τον αριθμό των *frame*, που χρησιμοποιήθηκαν για την εξαγωγή των χαρακτηριστικών κάθε χειρονομίας, ίσος δηλαδή με το μήκος των ακολουθιών.

Σημειώνουμε εδώ, πώς για να είναι ο *data* ένας τρισδιάστατος πίνακας, πρέπει οι ακολουθίες εκπαίδευσης να έχουν το ίδιο μήκος. Όμως αυτό δεν είναι εφικτό, αφού οι ακολουθίες προέρχονται από χειρονομίες, που πραγματοποιούνται από διαφορετικούς χρήστες, πράγμα το οποίο οδηγεί σε διαφορετικούς χρόνους, που χρειάζονται για την ολοκλήρωση διαφορετικών επαναλήψεων της ίδιας χειρονομίας. Αν, για παράδειγμα ένα χρήστης πραγματοποιήσει μια χειρονομία “attention” γρηγορότερα από κάποιον άλλον, η ακολουθία χαρακτηριστικών, που αντιστοιχεί στον πρώτο χρήστη, θα έχει μικρότερο μέγεθος από αυτή που θα προκύψει από την επεξεργασία του *video* με την χειρονομία του δεύτερου χρήστη.

Για την αντιμετώπιση αυτού του προβλήματος χρησιμοποιούμε *cell arrays*. Κατασκευάζουμε λοιπόν ένα *cell array*, που αποτελείται από πίνακες με έξι γραμμές (όσα τα χαρακτηριστικά) και αριθμό στηλών ίσο με το μήκος της κάθε ακολουθίας.

Ο κώδικας με τον οποίον γίνεται η εκπαίδευση του HMM μιας χειρονομίας και εν συνεχεία η αποκωδικοποίηση μιας ακολουθίας εισόδου παρατίθεται και αναλύεται στη συνέχεια.

```
O = 6;      %Number of coefficients in a vector
           %T = 23;      %Number of vectors in a sequence
nex = 7;    %Number of sequences
M = 1;     %Number of mixtures
Q = 6;     %Number of states
cov_type = 'full';

           %data = randn(O,T,nex);

data{1}(:,:)=PAR1;
data{2}(:,:)=PAR;
data{3}(:,:)=PAR2;
data{4}(:,:)=PAR4;
data{5}(:,:)=PAR5;

%initial guess of parameters

prior0 = [0.6;0.2;0.05;0.05;0.05;0.05];
          %transmat0 = at_TRANS_GUESS;
transmat0 = mk_lefttright_transmat(6,0.3);
[mu0, Sigma0] = mixgauss_init(Q*M, data, cov_type);
mu0 = reshape(mu0, [O Q M]);
Sigma0 = reshape(Sigma0, [O O Q M]);
mixmat0 = mk_stochastic(rand(Q,M));

[LL, prior1, transmat1, mu1, Sigma1, mixmat1] = ...
  mhmm_em(data, prior0, transmat0, mu0, Sigma0, mixmat0, 'max_iter',
  10,'thresh',0.001);

loglik2 = mhmm_logprob(PAR3, prior1, transmat1, mu1, Sigma1, mixmat1)
```

Αρχικά ορίζουμε τον αριθμό των χαρακτηριστικών, δηλαδή των γραμμών που θα έχουν οι πίνακες μας, το μήκος τους (εάν είναι ίδιο και δεν κάνουμε χρήση cell array) καθώς και το πόσες ακολουθίες υπάρχουν. Κατόπιν, γίνονται οι αρχικοποιήσεις του μοντέλου. Ο prior περιέχει τις πιθανότητες να βρεθεί σε κάθε μια από τις έξι καταστάσεις το μοντέλο, στην πρώτη επανάληψη του αλγορίθμου. Επειδή τα μοντέλα που εκπαideύσαμε ξεκινάνε από την πρώτη κατάσταση, έχουμε ορίσει την πιθανότητα να ξεκινάει από την πρώτη κατάσταση πολύ μεγαλύτερη. Ο πίνακας που περιέχει την αρχικοποίηση για τις πιθανότητες μεταβάσεων είναι ο transmat0 και δημιουργείται με χρήση της συνάρτησης mk_lefttright_transmat(6,0.3), η οποία δέχεται ως ορίσματα το μέγεθος του πίνακα (αριθμός καταστάσεων) και την πιθανότητα που επιθυμούμε να έχουν τα διαγώνια στοιχεία του πίνακα. Ο πίνακας που δημιουργεί η mk_lefttright_transmat είναι τύπου “bakis”, δηλαδή δεν έχουμε μεταβάσεις προς

προηγούμενες καταστάσεις(οι αντίστοιχες πιθανότητες είναι 0). Τέλος, δημιουργούνται τα γκαουσιανά μείγματα που θα χρησιμοποιηθούν για την προσέγγιση των εξόδων. Η συνάρτηση `mixgauss_init` δημιουργεί την αρχική εκτίμηση για τις παραμέτρους ενός γκαουσιανού μείγματος . Τα ορίσματα που δέχεται είναι ο συνολικός αριθμός των συστατικών των μειγμάτων, τα δεδομένα εκπαίδευσης και ο τύπος του πίνακα Sigma (πλήρης, διαγώνιος, σφαιρικός).

Μετά τις αρχικοποιήσεις καλείται η συνάρτηση `mhmm_em(data, prior0, transmat0, mu0, Sigma0, mixmat0, 'max_iter', 10, 'thresh', 0.001)` η οποία εκτελεί ουσιαστικά την εκπαίδευση του μοντέλου με χρήση του αλγόριθμου εκτίμησης-μεγιστοποίησης (εκτιμώντας τις παραμέτρους της γκαουσιανής κατανομής με βάση τα δεδομένα εισόδου και υπολογίζοντας τις πιθανότητες του πίνακα μετάβασης) . Όπως και στην περίπτωση της εκπαίδευσης των διακριτών HMM, μπορούμε και στην συνάρτηση `mhmm_em` να ορίζουμε την επιθυμητή ακρίβεια, αλλά και το μέγιστο αριθμό επαναλήψεων που θα διεκπεραιώνει ο αλγόριθμος σε περίπτωση που η επιθυμητή ακρίβεια δεν επιτυγχάνεται.

Τέλος, η συνάρτηση `mhmm_logprob(PAR3, prior1, transmat1, mu1, Sigma1, mixmat1)` χρησιμοποιείται κατά την αποκωδικοποίηση μιας ακολουθίας και μας επιστρέφει το λογάριθμο της πιθανότητας η ακολουθία που εξετάζουμε να έχει δημιουργηθεί από το συγκεκριμένο μοντέλο. Τα ορίσματα που δέχεται είναι η ακολουθία εισόδου (PAR3) καθώς και οι παράμετροι του HMM.

8.3 Πειραματικά Αποτελέσματα

Για την εκτίμηση της απόδοσης του συστήματος η διαδικασία που ακολουθήθηκε είναι η εξής: για κάθε χειρονομία τραβήχτηκαν video στα οποία συμμετείχαν έξι διαφορετικοί χρήστες. Ο κάθε χρήστης πραγματοποιούσε όλες τις χειρονομίες μερικές φορές. Κατόπιν, τα video χωρίστηκαν σε frame (30 f/s) και από την επεξεργασία των frames με τον αλγόριθμο εντοπισμού προέκυψαν οι ακολουθίες εκπαίδευσης αλλά και οι ακολουθίες που χρησιμοποιήθηκαν για αναγνώριση. Τα χαρακτηριστικά με τα οποία εκπαιδεύτηκαν τα συνεχή μοντέλα, είναι οι διαφορές που περιγράψαμε στο κεφάλαιο της υλοποίησης του διακριτού συστήματος, δηλαδή οι διαφορές, X_L-X_R , X_H-X_R , X_H-X_L , Y_L-Y_R , Y_H-Y_R , Y_H-Y_L . Στο συνεχές μοντέλο, δεν χρειάζεται να προβούμε σε περαιτέρω επεξεργασία τους , οπότε οι πίνακες με τα χαρακτηριστικά περιέχουν ουσιαστικά την χρονική εξέλιξη των παραπάνω διαφορών.

Για κάθε χειρονομία είχαμε στη διάθεσή μας διαφορετικό αριθμό ακολουθιών. Ορισμένες από τις ακολουθίες για κάθε χειρονομία, που είχαμε στη διάθεσή μας, χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων ενώ οι υπόλοιπες, δεν περιλαμβάνονταν στο σύνολο εκπαίδευσης και χρησιμοποιήθηκαν για αποκωδικοποίηση. Τα αποτελέσματα φαίνονται στον παρακάτω πίνακα

Χειρονομία	<i>Attention</i>	<i>Clapping</i>	<i>Explain</i>	<i>One hand</i>	<i>Oh my god</i>	<i>Wave</i>	<i>Go away</i>	<i>Unclassified</i>
<i>Attention</i>	10	0	0	0	0	0	1	0
<i>Clapping</i>	0	6	2	0	0	0	0	0
<i>Explain</i>	0	0	3	0	0	0	0	2
<i>One hand</i>	0	0	0	5	0	1	0	0
<i>Oh my god</i>	0	0	1	0	10	0	0	1
<i>Wave</i>	2	0	0	0	0	5	1	0
<i>Go away</i>	2	0	0	0	0	0	6	0
Ποσοστό αναγνώρισης (%)	90,2	75	60	83,3	90,2	62,5	75	

Από τον παραπάνω πίνακα παρατηρούμε ότι οι χειρονομίες που παρουσίασαν τα μικρότερα ποσοστά αναγνώρισης είναι η wave και explain. Τα χαμηλά ποσοστά ερμηνεύονται με διαφορετικό τρόπο για την κάθε χειρονομία. Ο λόγος που η explain δεν έδωσε ικανοποιητικά αποτελέσματα είναι ο έντονα στοχαστικός της χαρακτήρας καθώς και ο περιορισμένος αριθμός ακολουθιών που διαθέταμε. Ο συνδυασμός αυτών των δύο παραγόντων, έχει ως αποτέλεσμα η εκτίμηση των παραμέτρων των HMM να μην είναι τέτοια που να επιτρέπει στο μοντέλο να αναγνωρίζει κάποια άλλη ακολουθία, η οποία μπορεί μεν να έχει προέλθει και αυτή από χειρονομία explain, αλλά λόγω της φύσης της χειρονομίας, διαφέρει αρκετά από τις ακολουθίες που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου. Παρατηρούμε πως, καθώς η explain δεν μοιάζει ιδιαιτέρως με κάποια από τις υπόλοιπες κινήσεις, οι ακολουθίες που δεν αναγνωρίζονται ότι έχουν προέλθει από explain, δίνουν τόσο μικρή πιθανότητα σε όλα τα μοντέλα που κατατάσσονται στις unclassified.

Αντιθέτως, οι ακολουθίες που έχουν προέλθει από την wave και δεν ήταν δυνατόν να αναγνωριστούν, δεν κατατάσσονται στις unclassified αλλά γίνεται λανθασμένη αναγνώριση και κατατάσσονται στην χειρονομία attention. Ο λόγος είναι ο εξής : η τροχιά που ακολουθεί το χέρι σε αυτές τις δύο χειρονομίες , είναι μέχρι ενός σημείου παρόμοια (στο κομμάτι της ανύψωσης μέχρι το ύψος του κεφαλιού). Κατόπιν, ενώ η attention ουσιαστικά τερματίζεται μερικά frame μετά το σημείο που το χέρι φτάνει στο ύψος του κεφαλιού, η wave συνεχίζεται με την οριζόντια παλινδρόμηση του χεριού. Η οριζόντια αυτή κίνηση, παρουσιάζει μεγάλη στοχαστικότητα.. Συνεπώς, όταν γίνεται η αποκωδικοποίηση της ακολουθίας εισόδου(που έχει προέλθει από χειρονομία wave) με το μοντέλο της attention, γίνεται η σύγκριση του μέρους που αντιστοιχεί στην ανύψωση του χεριού και προκύπτει αυξημένη πιθανότητα να έχει προέλθει από το συγκεκριμένο μοντέλο. Αντιθέτως, όταν γίνει η αποκωδικοποίηση με το μοντέλο wave, επειδή συγκρίνεται και το στοχαστικό κομμάτι της χειρονομίας η πιθανότητα μικραίνει και μπορεί να μικραίνει σε σημείο να είναι μικρότερη από την πιθανότητα που προέκυψε κατά την αποκωδικοποίηση με το μοντέλο της attention.

Το μεγαλύτερο ποσοστό αναγνώρισης επιτυγχάνεται στην attention η οποία αφενός μεν είναι ίσως η πιο απλή και πιο «σαφής» χειρονομία από τις υπόλοιπες, και αφετέρου, κατά την εκπαίδευση του μοντέλου της attention, χρησιμοποιήθηκαν οι περισσότερες ακολουθίες συγκριτικά με τα υπόλοιπα μοντέλα. Το γεγονός αυτό πιστοποιεί την κρισιμότητα της ύπαρξης ικανού αριθμού δεδομένων εκπαίδευσης.

Τα λάθη που συμβαίνουν στην αναγνώριση των υπολοίπων χειρονομιών οφείλονται είτε στο γεγονός ότι οι χειρονομίες παρουσιάζουν κοινά στοιχεία – ή παρουσιάζουν τμήματα στα οποία τα χαρακτηριστικά τους γνωρίσματα λαμβάνουν παραπλήσιες τιμές – είτε στο γεγονός ότι τα δεδομένα εκπαίδευσης είναι λίγα ώστε να μοντελοποιηθούν ικανοποιητικά οι διαφοροποιήσεις που υπάρχουν κατά την εκτέλεση της χειρονομίας από διαφορετικούς χρήστες, οπότε οδηγούμαστε σε unclassified.

Τέλος, σημειώνουμε πως ο λογάριθμος της πιθανότητας που προέκυπτε κατά την αποκωδικοποίηση ήταν περίπου μία τάξη μεγέθους μικρότερος από την τιμή που προέκυπτε κατά την αποκωδικοποίηση κάποιας ακολουθίας στο διακριτό μοντέλο. Παρόλα αυτά, σε σχέση με τις τιμές που προέκυπταν από τα υπόλοιπα μοντέλα (ήταν πολύ μικρότερες) ήταν –τις περισσότερες φορές – σαφές από ποιο μοντέλο είχε «γεννηθή» η ακολουθία που αποκωδικοποιούσαμε κάθε φορά.

ΚΕΦΑΛΑΙΟ 9

ΣΥΝΟΨΗ

Αυτό το κεφάλαιο συνοψίζει τα σημαντικά εμπειρικά αποτελέσματα και περιγράφει τα περαιτέρω πειράματα και τις μελλοντικές βελτιώσεις στο υπάρχον σύστημα αναγνώρισης χειρονομίας

Στην εργασία αυτή έχει σχεδιαστεί και έχει τεθεί σε εφαρμογή ένα σύστημα για την αναγνώριση χειρονομιών από ηλεκτρονικό υπολογιστή που διαθέτει κάμερα. Η αναγνώριση γίνεται με τη βοήθεια στατιστικών μοντέλων, των HMM. Αρχικά το σύστημα εκπαιδεύεται με ένα συγκεκριμένο σύνολο κινήσεων, τις οποίες θα μπορεί να αναγνωρίσει. Ο χρήστης καλείται να πραγματοποιήσει μια κίνηση και το σύστημα υπολογίζει την πιθανότητα η κίνηση που δόθηκε ως είσοδος να ανήκει σε καθεμία από τις εκπαιδευμένες κινήσεις. Η εργασία αυτή επικεντρώθηκε σε αναγνώριση χειρονομιών από ακολουθίες βίντεο που ήδη είχαν τραβηχτεί και μάλιστα υπό συγκεκριμένες συνθήκες.

Για την αναγνώριση εφαρμόστηκαν 2 μέθοδοι: διακριτό και συνεχές κρυφό μοντέλο Markov, και οι δύο με την ίδια περίπου επιτυχία. Στο διακριτό μοντέλο οι παράμετροι που λαμβάνονται από την ακολουθία κωδικοποιούνται έτσι ώστε να αντιστοιχηθούν σε ένα από τα συγκεκριμένα διακριτά σύμβολα τα οποία δέχεται σαν είσοδο το διακριτό HMM. Στο συνεχές μοντέλο οι τιμές των παραμέτρων δεν υφίστανται καμία μεταβολή, αλλά γίνεται προσπάθεια να προσεγγιστούν από το σύστημα με μείγματα κατανομών Gauss, όπως έχει ήδη αναφερθεί παραπάνω. Η διακριτή προσέγγιση από τη μία πλευρά, έχει μεγάλα ποσοστά επιτυχίας ακόμα και με λίγα δεδομένα εκπαίδευσης εφόσον οι πιθανές τιμές των παραμέτρων είναι πεπερασμένες και μπορούν να μοντελοποιηθούν σχετικά εύκολα. Όμως, απαιτεί προεπεξεργασία των δεδομένων εισόδου ώστε να κωδικοποιηθούν, γεγονός που το καθιστά λιγότερο ικανό να λειτουργήσει σε πραγματικό χρόνο. Το συνεχές μοντέλο, από την άλλη, έχει τη δυνατότητα να λαμβάνει συνεχείς τιμές ώστε να γίνεται πιο ευέλικτο για χρήση σε διαφορετικού μεγέθους εικόνες χωρίς την ανάγκη να οριστεί νέο codebook. Το σημαντικότερο μειονέκτημα του είναι ότι απαιτεί πολλά δεδομένα εκπαίδευσης για να λειτουργήσει επιτυχώς. Εκτίμησή μας είναι ότι σε μελλοντικά συστήματα όπου οι αναγνωριζόμενες κινήσεις θα είναι πολύ περισσότερες και πιο εφάμιλλες μεταξύ τους, το διακριτό μοντέλο θα παρουσιάσει αδυναμία αναγνώρισης και ανάγκη αύξησης των διακριτών συμβόλων που χρησιμοποιεί, ενώ τελικά θα προτιμηθεί η λύση του συνεχούς.

Παρά τα ενθαρρυντικά αποτελέσματα που παρήγαγε το σύστημα αυτό, εύκολα συνειδητοποιεί κανείς, ότι αυτά προήλθαν από εφαρμογή του συστήματος σε αυστηρά προκαθορισμένες συνθήκες και ελεγχόμενες μεταβολές στις παραμέτρους του. Για να χρησιμοποιηθεί το σύστημα σε πρακτικές εφαρμογές απαιτείται ανασχεδιασμός του ώστε να γίνει πιο ευέλικτο και ανθεκτικό σε μεταβολές του περιβάλλοντος και να μπορεί να εφαρμοστεί σε πραγματικό χρόνο. Από τη φύση του το ανθρώπινο σώμα είναι ευκίνητο με δυνατότητα πραγματοποίησης μεγάλου εύρους κινήσεων, οι οποίες μάλιστα διαφέρουν σημαντικά από άτομα σε άτομο. Το γεγονός αυτό καθιστά τη μοντελοποίηση και την αναγνώριση των κινήσεων του ανθρώπου ένα πολύ δύσκολο εγχείρημα. Το συγκεκριμένο σύστημα προϋποθέτει ότι οι κινήσεις του χρήστη γίνονται σύμφωνα με τα αρκετά αυστηρά όρια που αυτό ορίζει. Κατά τη γνώμη μας θα πρέπει, αρχικά, να αναπτυχθεί πιο εξειδικευμένος αλγόριθμος εύρεσης των σημείων του σώματος χωρίς εξάρτηση από το μέγεθος της εικόνας, το χρώμα

δέρματος του χρήστη και το φόντο. Επίσης, το μοντέλο θα εμπλουτιστεί με περισσότερες και πιο συγκεκριμένες κινήσεις που θα μπορεί να αναγνωρίσει με τη χρήση επιπλέον υλικού, όπως περισσότερες από μία κάμερες, αισθητήρες, ειδικά γάντια κτλ. Τέλος, είναι απαραίτητο να υπάρξει η κατάλληλη υπολογιστική δύναμη που θα υποστηρίξει τη λειτουργία του συστήματος σε πραγματικό χρόνο.

Η έρευνα για την αναγνώριση χειρονομίας έχει πολλά κίνητρα, τα οποία συσχετίζονται με τη βελτίωση της επαφής μεταξύ των ανθρώπων και των υπολογιστών. Εάν ένας υπολογιστής μπορεί να ανιχνεύσει και να αναγνωρίσει ένα σύνολο χειρονομιών, μπορεί να συμπεράνει το μήνυμα του αποστολέα και να αποκριθεί κατάλληλα. Υπάρχει ένα μεγάλο εύρος εφαρμογών που μπορεί να στηριχθεί σε συστήματα αναγνώρισης χειρονομιών. Αυτές ξεκινούν από απλούς προσωπικούς υπολογιστές όπου ο χρήστης μπορεί να εισάγει γράμματα χωρίς τη χρήση πληκτρολογίου, κάμερες ασφαλείας που να αναγνωρίζουν παράνομες κινήσεις, και μπορούν να φτάσουν μέχρι εικονικές ορχήστρες, συστήματα εκπαίδευσης πολεμικών τεχνών και πολλές άλλες εφαρμογές που ακόμα και η φαντασία μας δεν μπορεί να προβλέψει. Μόνο το μέλλον θα μας δείξει τι άλλο μπορεί να παραγάγει ο ανθρώπινος νους.

BIBΛΙΟΓΡΑΦΙΑ

1. Lawrence R. Rabiner, Fellow, IEE , “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”
2. Yang Liu & Yunde Jia, “A Robust Hand Tracking for Gesture-Based Interaction of Wearable Computers”
3. Nianjun Liu, Richard I.A. Davis, Brian C. Lovell and Peter J. Kootsookos, “Effect of Initial HMM Choices in Multiple Sequence Training for Gesture Recognition”
4. Guangqi Ye, Jason J. Corso, Gregory D. Hager, “Gesture Recognition and Motion Features”
5. Kostas Karpouzis, Amaryllis Raouzaïou, Athanasios Drosopoulos, Spiros Ioannou, Themis Balomenos, Nicolas Tsapatsoulis and Stefanos Kollias, “Gesture Analysis for Emotionally-Rich Man-Machine Interaction”
6. Feng Wang, Chong-Wah Ngo, Ting-Chuen Pong, “Gesture Tracking and Recognition for Lecture Video Editing”
7. Dorthe Meyer, “Human Gait Classification Based on Hidden Markov Models”
8. J.K. Aggarwal, Sangho Park, “Human Motion: Modeling and Recognition of Actions and Interactions”
9. Tian-Shu Wang, Heung-Yeung Shum, Ying-Quing Xu, Nan-Ning Zheng, “Unsupervised Analysis of Human Gestures”
10. Christopher Lee, Yangsheng Xu, “Online, Interactive Learning of Gestures for Human/Robot Interfaces”
11. Cìan W. Shaffrey, Nick G. Kingsbury, Ian H. Jermyn, “Unsupervised Image Segmentation via Markov Trees and Complex Wavelets”
12. Jie Yang, Yangsheng Xu, “Hidden Markov Model for Gesture Recognition”
13. Arnab Ghoshal, Pavel Ircing, Sanjeev Khudanpur, “Hidden Markov Models for Image and Video Retrieval Using Textual Queries”
14. Olivier Cappé, “Ten years of HMM’s”
15. Jeff Bilmes, “What HMMs Can Do”
16. Eric Fosler-Lussier, “Markov Models and Hidden Markov Models: A Brief Tutorial”
17. Jeff A. Bilmes, “A Gentle Tutorial of the EM Algorithm and its Applications to Parameter Estimation for Gaussian Mixture and Hidden Markov Models”
18. Steve Young, Gunnar Evermann, Thomas Hain, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, Phil Woodland,”The HTK Book”
19. Jenny Preece, “Human Computer Interaction”
20. Andrew D. Wilson, Aaron F. Bobick, “Recognition and Interpretation of Parametric Gesture”
21. Nianjun Liu, Brian C. Lovell, Peter J. Kootsookos, Richard I.A. Davis, “Understanding HMM Training for Video Gesture Recognition”
22. Andrew D. Wilson, Aaron F. Bobick, “Nonlinear PHMMs for the Interpretation of Parameterized Gesture”
23. Jan Mäntyjärvi, Juha Kela, Panu Korpipää, Sanna Kallio, “Enabling fast and effortless customization in accelerometer based gesture interaction”
24. Donald O. Tanguay Jr., “Hidden Markov Models for Gesture Recognition”
25. Shi Zhong, “Probabilistic Model-based Clustering of Time Series”
26. Yingjian Zhang, “Prediction of Financial Time Series with Hidden Markov Models”

