



Αναγνώριση νοηματικής. *Sign language* recognition.

Διπλωματική εργασία

Αλέξανδρος Συναδινός

1) Επιβλέπων καθηγητής :
Γεράσιμος Ποταμιάνος

2) Επιβλέπων καθηγητής :
Αντώνιος Αργυρίου

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κύριο Γεράσιμο Ποταμιάνο καθώς χωρίς την άφθονη βοήθεια, συμπαράσταση, υπομονή και επιμονή του δεν θα είχε ολοκληρωθεί η παρούσα διπλωματική εργασία. Επίσης θα ήθελα να ευχαριστήσω για τον πολύτιμο χρόνο που διέθεσε, τον επιβλέποντα καθηγητή κύριο Αντώνιο Αργυρίου. Ακόμη θέλω να δώσω τις θερμότερες ευχαριστίες μου στο Πανεπιστήμιο Θεσσαλίας και τον πρόεδρο του τμήματος κύριο Ελευθέριο Τσουκαλά, όπως επίσης και τους εργαζόμενους του τμήματος στην Γραμματεία, την φοιτητική μέριμνα, την καθαριότητα και την φύλαξη. Τέλος ευχαριστώ όλους τους φίλους μου και την οικογένειά μου αλλά ιδιαιτέρως τους Γιάννη, Μιχάλη, Νίκο, Κωνσταντίνο, Αθηνά για την στήριξη που μου παρείχαν όλα τα χρόνια των σπουδών μου.

Αναγνώριση της νοηματικής

Περίληψη

Στην εποχή μας όπου ο έλεγχος των ηλεκτρονικών υπολογιστών τείνει να πραγματοποιείται κυρίως μέσω φωνητικών εντολών, είναι αδιανόητο άνθρωποι με προβλήματα στην ακοή ή την ομιλία να μην μπορούν να αλληλεπιδράσουν με τους υπολογιστές μέσω του δικού τους οικείου τρόπου επικοινωνίας. Επίσης, αξιοσημείωτο φαινόμενο είναι αυτό του κοινωνικού αποκλεισμού της παραπάνω ομάδας ατόμων που δεν μπορεί να επικοινωνήσει με την υπόλοιπη κοινωνία εύκολα, παρά την ραγδαία ανάπτυξη τομέων της υπολογιστικής όρασης όπως η αναγνώριση χειρονομίας.

Η παρούσα διπλωματική εργασία αποτελεί μια πρακτική εφαρμογή της υπολογιστικής όρασης με σκοπό να λύσει το παραπάνω πρόβλημα. Η εργασία λαμβάνει ως δεδομένα εισόδου νοήματα που πραγματοποιούνται από άτομα με πρόβλημα στην ομιλία, τα οποία περνούν διαδικασία επεξεργασίας και ύστερα παράγεται η αντίστοιχη μετάφραση των λέξεων. Κάθε άτομο περνά από την διαδικασία εντοπισμού των χειρών και του προσώπου εκ των οποίων εξάγονται τα απαραίτητα χαρακτηριστικά. Μέσω αυτών των χαρακτηριστικών, εκπαιδεύεται ο αλγόριθμος που κατασκευάστηκε στα πλαίσια της διπλωματικής, έτσι ώστε να αναγνωρίζει τις ίδιες λέξεις πραγματοποιούμενες από κάποιο άλλο ανεξάρτητο άτομο. Οι διαφορές ανάμεσα στα χαρακτηριστικά των λέξεων, συγκρίνονται χρησιμοποιώντας τον αλγόριθμο του DTW (Dynamic Time Warping), ενώ για την εύρεση των αποστάσεων εφαρμόζεται είτε ο αλγόριθμος της Ευκλείδειας απόστασης είτε η υλοποίηση με την απόσταση Mahalanobis. Η ολοκλήρωση της διαδικασίας επέφερε ως αποτέλεσμα το ποσοστό του 60% όταν απαιτείται τέλειο ταίριασμα αναγνώρισης. Αντίθετα, όταν ο στόχος είναι η εύρεση των τριών καλύτερων επιλογών αναγνώρισης η ακρίβεια προσεγγίζει το ποσοστό του 90%.

Sign language recognition

Abstract

In our era, in which the control of personal computer tends to be accomplished mainly by voice commands, it is an inconceivable fact that people with hearing or speech problems can not interact with their computers via their own familiar communication way. Also, a noteworthy phenomenon is the social isolation of the above group of people which can not communicate with the rest of the society easily, despite of the rapid development of computer vision fields such as gesture recognition.

This thesis is a practical application of computer vision with goal to solve the above problem. The project takes as input data, signs that are being performed by people with speech problems, which are being processed and then the corresponding translation is being produced. Every signer goes through the process of detection of hand and face, from which the necessary features are being extracted. By using these features, the algorithm that was been implemented in this thesis, is being trained, so that to recognize the same signs performed by another independent person. The differences between the features of the words, are being compared using the dynamic time warping (DTW) algorithm, while for the finding of the distances it is being applied either the Euclidean distance corresponding algorithm or the Mahalanobis distance implementation. The completion of the procedure has as a result, the percentage of 60% recognition accuracy when a perfect match is required. On the contrary, when the goal is to find the 3 best recognition choices, the accuracy approaches the percentage of 90%.

Περιεχόμενα

Ευχαριστίες.....	2
Περίληψη	3
Abstract.....	4
Κατάλογος εικόνων	7
Κατάλογος πινάκων	7
1 Εισαγωγή	8
1.1 Επικοινωνία ανθρώπου-υπολογιστή	8
1.2 Επικοινωνία ανθρώπων με προβλήματα στην ομιλία	8
1.3 Νοηματική γλώσσα	9
1.4 Αναγνώριση χειρονομίας	9
1.4.1 Περιληπτικά	9
1.4.2 Εντοπισμός.....	10
1.4.3 Παρακολούθηση	10
1.4.4 Αναγνώριση	10
1.5 Επισκόπηση παρόμοιων εργασιών.....	11
1.5.1 Γράμμα Γράμμα (finger spelling)	11
1.5.2 Ολοκληρωμένες λέξεις	12
1.6 Γρήγορη επισκόπηση της εργασίας	15
1.7 Συμβιβασμοί και υποθέσεις	16
1.8 Στόχος και συνεισφορά.....	16
1.9 Οργάνωση της διπλωματικής.....	17
2 Εξαγωγή χαρακτηριστικών.....	18
2.1 Προεπεξεργασία.....	18
2.2 Μάσκα χρώματος.....	18
2.3 Εντοπισμός αντικειμένων	20
2.4 Αναγνώριση χειρών	20
2.5 Εξαγωγή σημείων κέντρου.....	21
2.6 Εξαγωγή κλίσης.....	22
2.7 Εξαγωγή μηκών αξόνων.....	22
2.8 Επεξεργασία για παραγωγή της μετατόπισης $dx-dy$	23
2.9 Αρχείο χαρακτηριστικών.....	23
3 Μηχανική μάθηση.....	24
3.1 Euclidean distance	24

3.2	<i>Mahalanobis distance</i>	24
3.3	<i>Dynamic Time Warping</i>	25
3.4	<i>Πειραματικό περιβάλλον</i>	27
3.5	<i>Κλάσεις</i>	27
4	Βάση δεδομένων και πειραματικά αποτελέσματα	29
4.1	<i>Σύνοψη</i>	29
4.2	<i>Βάση δεδομένων</i>	29
4.3	<i>Αποτελέσματα συγκρίσεων</i>	30
4.4	<i>Μεμονωμένα πειράματα με χαρακτηριστικά και αξιολόγηση αποτελεσμάτων</i>	31
4.5	<i>Πειράματα με συνδυασμό χαρακτηριστικών και αξιολόγηση αποτελεσμάτων</i>	32
5	Συμπέρασμα και επέκταση	34
5.1	<i>Συμπεράσματα</i>	34
5.2	<i>Μελλοντική εργασία και επέκταση</i>	35
Appendix		36
I.	<i>Εξαγωγή εικόνων από lossless compressed video</i>	36
II.	<i>Βιβλιοθήκη OpenCV</i>	36
III.	<i>Διαδικασία εξαγωγής χαρακτηριστικών</i>	36
IV.	<i>Διαδικασία σε Matlab</i>	37
V.	<i>Συνδυασμός χαρακτηριστικών</i>	38
Βιβλιογραφία		39

Κατάλογος εικόνων

Εικόνα 1.1: Αλφάβητο αμερικάνικης νοηματικής γλώσσας. Πηγή: [1].	9
Εικόνα 1.2: Στιγμιότυπο 1. Πηγή: [2].	12
Εικόνα 1.3: Στιγμιότυπο 2. Πηγή: [2].	13
Εικόνα 1.4: Στιγμιότυπο 3. Πηγή: [2].	13
Εικόνα 1.5: Εικόνα ιστορίας κίνησης. Πηγή: [2].	14
Εικόνα 1.6: Καταγραφή μέσω ενσωματωμένης κάμερας. Πηγή: [3].	15
Εικόνα 1.7: Διαδικασία της διπλωματικής εργασίας	16
Εικόνα 2.1: Μάσκα χρώματος έχοντας αποβάλλει τα σημεία με χρώμα δέρματος	19
Εικόνα 2.2: Τελική μάσκα	19
Εικόνα 2.3: Αναγνώριση των χεριών και εμφάνιση αξόνων κλίσης μέσα σε τετράγωνο	21
Εικόνα 2.4: Σχηματικό κλίσης ύστερα από ανάλυση κύριων συνιστωσών. Πηγή: [25].	22
Εικόνα 3.1: Ταίριασμα 2 σημάτων μέσω DTW. Πηγή: [4].	25
Εικόνα 3.2: Πίνακας μονοπατιών DTW. Πηγή: [5].	26
Εικόνα 3.3: Συγκρινόμενα στοιχεία για την εύρεση του κόστους κελιού. Πηγή: [5].	27
Εικόνα 4.1: Οι 3 όψεις της βάσης δεδομένων. Πηγή: [2].	30

Κατάλογος πινάκων

Πίνακας 1: Συγκριτικός πίνακας απόδοσης μερικών γνωστών υλοποιήσεων συστήματος για αναγνώριση του αλφάβητου της αμερικάνικης νοηματικής γλώσσας. Πηγή: [1].	12
Πίνακας 2: Παράδειγμα εξαχθέντος αρχείου χαρακτηριστικών.	23
Πίνακας 3: Τα παραγόμενα χαρακτηριστικά μετατόπισης με την σειρά με την οποία συνενώνονται στο αρχείο χαρακτηριστικών (συνέχεια του πίνακα 2).	23
Πίνακας 4: Κλάσεις-λέξεις εργασίας	28
Πίνακας 5: Σύγκριση μεμονωμένων χαρακτηριστικών	30
Πίνακας 6: Σύγκριση συνδυασμών χαρακτηριστικών	31

1 Εισαγωγή

1.1 Επικοινωνία ανθρώπου-υπολογιστή

Κατά την πρώτη φάση ανάπτυξης υπολογιστικών συστημάτων οι μηχανικοί υπολογιστών και οι προγραμματιστές δεν αφιέρωσαν πολύ χρόνο στην διευκόλυνση της χρήσης του υπολογιστή από τον άνθρωπο. Η ανάγκη λοιπόν για ευκολότερη χρήση των υπολογιστών και όσο το δυνατόν φυσικότερη διεπαφή είναι πιο εμφανής από ποτέ πλέον. Ένας από τους κυριότερους παράγοντες που κάνουν την επιστήμη της επικοινωνίας ανθρώπου-μηχανής (HCI) πολύτιμη είναι η χρήση των υπολογιστών από ανθρώπους με ποικίλες σωματικές, διανοητικές και γνωστικές δυνατότητες με εντελώς διαφορετικά περιβαλλοντικά ερεθίσματα. Ο τρόπος που ο καθένας λοιπόν θεωρεί φυσικό και οικείο του ώστε να επικοινωνεί με τον υπολογιστή διαφέρει από άνθρωπο σε άνθρωπο. Έτσι για ανθρώπους με κινητικά προβλήματα είναι ευκολότερος ο έλεγχος του υπολογιστή μέσω φωνητικών εντολών, ενώ για ανθρώπους με προβλήματα στην ομιλία, τα νοήματα αποτελούν πιο φυσικό τρόπο ελέγχου του υπολογιστή.

1.2 Επικοινωνία ανθρώπων με προβλήματα στην ομιλία

Η επικοινωνία αποτελεί ένα από τα πιο σημαντικά χαρακτηριστικά και προνόμια του ανθρώπου. Αυτή μπορεί να εκφράζεται με διάφορους τρόπους (γραφτά, ομιλία, νοήματα) αλλά πάντα έχει έναν στόχο που δεν είναι άλλος από την ανταλλαγή απόψεων, συναισθημάτων ή απλά πληροφοριών. Στη σύγχρονη κοινωνία όπου πλέον η επικοινωνία μπορεί να πραγματοποιηθεί ακόμα και ανάμεσα σε ανθρώπους που απέχουν τεράστια απόσταση είναι αδιανόητο να υπάρχουν άνθρωποι αποκλεισμένοι, που όχι μόνο δεν μπορούν να επικοινωνήσουν απομακρυσμένα αλλά δυσκολεύονται να επικοινωνήσουν με τον υπόλοιπο πληθυσμό ακόμα και κατ'ιδίαν. Αυτοί οι άνθρωποι δεν είναι άλλοι από όσους έχουν προβλήματα στην ομιλία ή/και την ακοή, οι οποίοι προκειμένου να επικοινωνούν μεταξύ τους και με τους αγαπημένους τους έχουν αναπτύξει την νοηματική γλώσσα. Ωστόσο είναι προφανές ότι όλα τα μέλη μιας κοινωνίας δεν έχουν την ανάγκη να μάθουν νοηματική καθώς μπορεί να τους χρειαστεί ελάχιστες φορές κατά τη διάρκεια της ζωής τους. Έτσι το φαινόμενο του κοινωνικού αποκλεισμού παραμένει και εντείνεται παράλληλα με την εξέλιξη της τεχνολογίας.

Ένα πρόβλημα διαφορετικής φύσεως ωστόσο αντιμετωπίζουν και τα άτομα που αναζητούν μία λέξη σε λεξικό της νοηματικής γλώσσας οι καταχωρίσεις στο οποίο γίνονται με αλφαβητική σειρά. Έτσι όταν κάποιος αναζητεί να κάνει μετάφραση μία γραπτή λέξη έχοντας στην διάθεση του μόνο την αντίστοιχη εικόνα των χειρών, η δυνατότητα αυτή δεν είναι διαθέσιμη.

1.3 Νοηματική γλώσσα

Η νοηματική γλώσσα αποτελεί έναν βασικό τρόπο επικοινωνίας για κωφούς ανθρώπους και ανθρώπους με σοβαρά προβλήματα στην ομιλία. Χρησιμοποιεί τις παλάμες, τα δάκτυλα, την κίνηση των χεριών αλλά και κινήσεις του προσώπου και των ματιών για να εκφράσει τις σκέψεις του ομιλητή χωρίς προφορικό λόγο. Πολλές χώρες έχουν την δική τους νοηματική γλώσσα η οποία σχετίζεται με την δική τους γραμματική. Οι νοηματικές γλώσσες διαφορετικών χωρών μπορεί να μοιράζονται μεταξύ τους κοινά στοιχεία όπως για παράδειγμα χρησιμοποίηση 2 χεριών για αναπαράσταση του αλφαβήτου. Τα σύμβολα της αμερικάνικης νοηματικής γλώσσας (η οποία χρησιμοποιείται στην παρούσα εργασία) είναι 26. Το αλφάβητο περιέχει 24 σταθερές φιγούρες για κάθε γράμμα εκτός των J και Z των οποίων ο σχηματισμός απαιτεί κίνηση όπως φαίνεται και στην παρακάτω εικόνα (Εικόνα 1.1). Το λεξικό της γλώσσας περιέχει 3000 περίπου νοήματα [27].



Εικόνα 1.1: Αλφάβητο αμερικάνικης νοηματικής γλώσσας. Πηγή: [1].

1.4 Αναγνώριση χειρονομίας

1.4.1 Περιληπτικά

Η αναγνώριση χειρονομίας (Hand Gesture Recognition) στην εποχή μας είναι ένας ραγδαία αναπτυσσόμενος κλάδος έρευνας της επιστήμης των υπολογιστών καθώς χρησιμοποιείται για χειρισμό των ηλεκτρονικών υπολογιστών εξ'αποστάσεως.

Όπως είναι φυσικό έχει πρακτική εφαρμογή για παράδειγμα στις κονσόλες παιχνιδιών όπου πλέον οι άνθρωποι μπορούν να παίζουν χωρίς να παρεμβάλλεται ανάμεσα σε αυτούς και το μηχάνημα κάποιο τηλεχειριστήριο αλλά και σε ιατρικές επεμβάσεις σε περιπτώσεις που χρειάζεται σταθερότητα και ο θεράπων ιατρός χειρίζεται κάποιον βραχίονα μέσω χειρονομιών. Βλέπουμε από τα παραπάνω ότι η αναγνώριση χειρονομίας πλέον έχει ενταχθεί και εξυπηρετεί

την καθημερινότητα ολοένα και μεγαλύτερου αριθμού ατόμων ενώ πλέον έχουν τεθεί οι βάσεις και για γρήγορη αναγνώριση της νοηματικής γλώσσας.

Σημαντικό ρόλο στην πρόοδο της αναγνώρισης χειρονομίας έπαιξε η χρησιμοποίηση καμερών που μπορούν να εξάγουν πληροφορίες σχετικά με το βάθος των αντικειμένων στην εικόνα όπως για παράδειγμα η κάμερα Kinect [6], [7], [10].

Επίσης πολλές προσπάθειες βασίστηκαν στην χρησιμοποίηση καμερών υπό διαφορετική γωνία [8] ή ανάγκαζαν το άτομο που πραγματοποιούσε μια χειρονομία να φοράει γάντια διαφορετικού χρώματος [9], [11] ώστε να εξάγονται οι απαραίτητες πληροφορίες.

Σε γενικές γραμμές η αναγνώριση χειρονομίας αποτελείται από 3 επιμέρους στάδια, τον εντοπισμό, την παρακολούθηση και την αναγνώριση. Το στάδιο του εντοπισμού είναι υπεύθυνο για τον καθορισμό, την ανεύρεση και την εξαγωγή οπτικών χαρακτηριστικών σχετικά με την παρουσία των χεριών στο πλάνο της κάμερας. Το στάδιο της παρακολούθησης στη συνέχεια επεμβαίνει ώστε να συσχετίζονται τα χαρακτηριστικά ανάμεσα στα διαφορετικά στιγμιότυπα (frames) με σκοπό να γνωρίζει ανά πάσα στιγμή το σύστημα που βρίσκονται τα χέρια. Έπειτα, στο επίπεδο της αναγνώρισης συντελείται η συγκέντρωση των εξαχθέντων χαρακτηριστικών τα οποία ομαδοποιούνται, τους ανατίθενται ετικέτες και συσχετίζονται με συγκεκριμένες κλάσεις.

1.4.2 Εντοπισμός

Αποτελεί το πρώτο βήμα της αναγνώρισης χειρονομίας καθώς διαδικασίες όπως η απομάκρυνση του φόντου και η ορθή απομόνωση των χρήσιμων δεδομένων είναι κρίσιμες πριν μεταφερθούν τα δεδομένα στα επόμενα βήματα. Για την εξυπηρέτηση του συγκεκριμένου σταδίου έχουν προταθεί διάφορες εφαρμογές και μέθοδοι όπως για παράδειγμα εντοπισμός με βάση το χρώμα [12], εντοπισμός σύμφωνα με το περίγραμμα ή την κίνηση των χεριών αλλά και πιο εξελιγμένες προσεγγίσεις που χρησιμοποιούν τρισδιάστατα μοντέλα αναπαράστασης των χεριών [13], [14].

1.4.3 Παρακολούθηση

Η παρακολούθηση κατά την οποία γίνεται συσχέτιση των εξαχθέντων χαρακτηριστικών ανά στιγμιότυπο (frame) είναι το δεύτερο βήμα της διαδικασίας κατανόησης των κινήσεων του χεριού. Η λειτουργία αυτού του σταδίου είναι δυσνόητη, από την μία πλευρά παρέχει πληροφορίες σχετικά με τα χαρακτηριστικά ανάμεσα στα διάφορα χρονικά στιγμιότυπα. Από την άλλη πλευρά μέσω αυτού του σταδίου υπολογίζονται και συντηρούνται οι προσεγγίσεις των παραμέτρων των τρισδιάστατων μοντέλων και των χαρακτηριστικών που δεν είναι εμφανώς παρατηρήσιμα σε κάποια χρονική στιγμή. Η υλοποίηση αυτού του σταδίου έχει υλοποιηθεί με μια ευρεία ποικιλία μεθόδων όπως για παράδειγμα χρησιμοποιώντας ταίριασμα προτύπων [15], βέλτιστες εκτιμήσεις των χαρακτηριστικών [16] ή GFD (Generic Fourier Descriptor) [23].

1.4.4 Αναγνώριση

Ο τελικός στόχος της αναγνώρισης χειρονομίας είναι η χρησιμοποίηση και μετάφραση των χαρακτηριστικών που προκύπτουν από την θέση ενός χεριού. Η αναγνώριση χρησιμοποιείται είτε για τον έλεγχο υπολογιστών οπότε αρκεί σειριακή παρακολούθηση της

θέσης του χεριού είτε στην μετάφραση σημαδιών και νοημάτων οπότε απαιτείται και γνώση σχετικά με τους γραμματικούς κανόνες. Είναι εύκολα αντιληπτό ότι όσο μεγαλύτερος είναι ο αριθμός των λέξεων (signs) προς μετάφραση τόσο αυξάνεται το επίπεδο πολυπλοκότητας και αμφιβολίας των αποτελεσμάτων. Στα πλαίσια της αναγνώρισης έχουν προταθεί μεταξύ άλλων προσεγγίσεις όπως αναγνώριση με boosting [18], [19], ταίριασμα περιγράμματος και σωματότυπου [22], κρυφά μοντέλα Markov [17], παράλληλα μοντέλα Markov [8] και μηχανές διανυσματικής στήριξης (SVM) [1].

1.5 Επισκόπηση παρόμοιων εργασιών

Η ιδιαιτερότητα που έχει η νοηματική γλώσσα να μην είναι ίδια για κάθε χώρα αλλά και η μεγάλη ανάγκη για αναγνώριση και μετάφραση των νοημάτων σε λέξεις, έχει οδηγήσει την επιστημονική κοινότητα σε εκτεταμένες προσπάθειες για απολύτως επιτυχή και γρήγορη αναγνώριση της. Είναι εύκολο να φανταστεί κανείς ότι παράγοντες επιτυχίας ενός συστήματος αναγνώρισης νοηματικής γλώσσας εκτός από το ποσοστό επιτυχίας είναι τόσο ο χρόνος απόκρισης όσο και η ανεξαρτησία από τον χρήστη, τα ρούχα του και το περιβάλλον στο οποίο πραγματοποιεί την λέξη.

Οι προσπάθειες αναγνώρισης της νοηματικής γλώσσας χωρίζονται σε 2 κατηγορίες. Στην πρώτη κατηγορία συναντώνται προσπάθειες στις οποίες αναγνωρίζεται το αλφάβητο της εκάστοτε γλώσσας όπου το κάθε γράμμα αποτελείται από ένα νόημα και έπειτα σχηματίζουν μια ολοκληρωμένη λέξη τοποθετώντας σε σειρά τα γράμματα (finger spelling). Στη δεύτερη κατηγορία βρίσκονται πιο πολύπλοκες και εξεζητημένες λύσεις καθώς αναγνωρίζονται ολόκληρες οι λέξεις ή προτάσεις που πραγματοποιεί κάποιος. Η πολυπλοκότητα της μεθόδου αυτής έγκειται στο γεγονός ότι εκτός από απλή αναγνώριση του νοήματος, αν απαιτείται επαγγελματικό αποτέλεσμα με απώτερο στόχο την παραγωγή προϊόντος λογισμικού τότε είναι αναγκαία η εφαρμογή γραμματικών και συντακτικών κανόνων παράλληλα με την διαδικασία αναγνώρισης. Παρακάτω εμφανίζονται ορισμένες χαρακτηριστικές εργασίες και δίνεται έμφαση τόσο στις μεθόδους υπολογιστικής όρασης που χρησιμοποιούνται όσο και στα αποτελέσματα που παράγονται.

1.5.1 Γράμμα Γράμμα (finger spelling)

Η επιτυχής αναγνώριση του αλφάβητου [23] της αμερικάνικης νοηματικής αποτέλεσε κινητήριο μοχλό για την αναγνώριση ολοκληρωμένων λέξεων και έπειτα προτάσεων. Τα αποτελέσματα που παρουσιάζονται σε σχέση με την ακρίβεια αναγνώρισης κυμαίνονται από 75% έως 95% όπως φαίνεται συνοπτικά και στον πίνακα 1. Έτσι ανάμεσα σε πολλές προσεγγίσεις, ξεχωρίζει για την ακρίβεια της η μέθοδος [1] που χρησιμοποιεί τον μετασχηματισμό Hough για να εντοπίσει κύκλους, γραμμές και ελλείψεις στις εικόνες και έπειτα με χρήση μηχανών διανυσματικής υποστήριξης (SVM) πραγματοποιεί την ταξινόμηση. Η ακρίβεια της προτεινόμενης αυτής μεθόδου κλιμακώνεται από 81% έως και 93.88% για τα 26 σύμβολα του αλφαβήτου πάνω σε μια βάση δεδομένων με 300 εικόνες εκπαίδευσης και 180 εικόνες δοκιμής.

<i>N.</i>	<i>Method</i>	<i>Classifier</i>	<i>Accuracy</i>
1	HOG	BPNN	91.6
2	EOH	SVM	93.75
3	SIFT	Point pattern matching	77.77
4	Geometric Features based	BPNN	80.28
5	Color based	Template matching	93
6	Hough Transform and SVM	SVM	93.88

Πίνακας 1: Συγκριτικός πίνακας απόδοσης μερικών γνωστών υλοποιήσεων συστήματος για αναγνώριση του αλφάβητου της αμερικάνικης νοηματικής γλώσσας. Πηγή: [1].

1.5.2 Ολοκληρωμένες λέξεις

Η βάση δεδομένων που χρησιμοποιήθηκε στην παρούσα εργασία συνοδεύεται από μία μέθοδο αναγνώρισής της, που στόχο είχε να δώσει ένα έναυσμα σε ερευνητές ανά τον κόσμο να ασχοληθούν με την αναγνώριση ολοκληρωμένων λέξεων της αμερικάνικης νοηματικής γλώσσας. Αυτή η μέθοδος αναγνώρισης βασίζεται σε εικόνες που απεικονίζουν την κίνηση (Motion Energy Images) [2] και οι οποίες συγκρίνονται για την παραγωγή αποτελέσματος ταξινόμησης του νοήματος στην κατάλληλη κλάση. Ένα παράδειγμα του τρόπου παραγωγής των MEI φαίνεται στην παρακάτω ακολουθία εικόνων (Εικόνα 1.2 - Εικόνα 1.5).



Εικόνα 1.2: Στιγμιότυπο 1. Πηγή: [2].



Εικόνα 1.3: Στιγμιότυπο 2. Πηγή: [2].



Εικόνα 1.4: Στιγμιότυπο 3. Πηγή: [2].

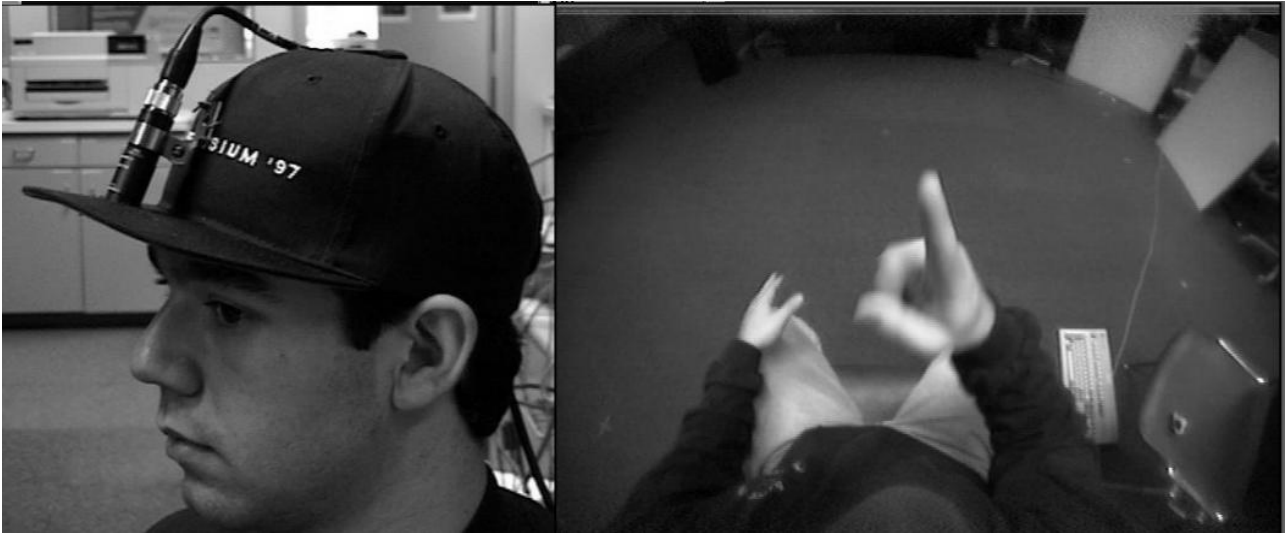


Εικόνα 1.5: Εικόνα ιστορίας κίνησης. Πηγή: [2].

Μία εναλλακτική προσέγγιση σχετικά με την αναγνώριση λέξεων της νοηματικής γλώσσας που προτάθηκε το 1999 [8] χρησιμοποιεί μία βάση δεδομένων όπου ο χρήστης βιντεοσκοπείται από 3 όψεις (μπροστά, πλάι, κάτω). Στην συγκεκριμένη εργασία χρησιμοποιούνται παράλληλα κρυφά μοντέλα Markov. Το ποσοστό επιτυχούς αναγνώρισης ανέρχεται στο 94%.

Προχωρώντας η μελέτη πάνω στα συστήματα αναγνώρισης χειρονομίας και προσπαθώντας να ενταχθεί στην καθημερινότητα έγινε αντιληπτό ότι τα αποτελέσματα των συστημάτων πρέπει να παράγονται σε πραγματικό χρόνο και σε αυτόν τον δρόμο κινείται η εργασία [3] χρησιμοποιώντας κρυφά μοντέλα Markov. Επειδή ο στόχος της συγκεκριμένης εργασίας είναι η αναγνώριση σε πραγματικό χρόνο, έγιναν κάποιοι συμβιβασμοί όπως η ύπαρξη διαφορετικών χρωματιστών γαντιών σε κάθε χέρι. Οι συμβιβασμοί βοηθάνε σε περιπτώσεις υπερκάλυψης των χεριών και στην εξαγωγή ποιοτικότερων χαρακτηριστικών σχετικά με την μορφή του χεριού. Η ακρίβεια της συγκεκριμένης μεθόδου ανέρχεται σε ποσοστό 91.3% όταν η σύγκριση γίνεται με ανεξάρτητο διάλυσμα δοκιμής και όχι με κάποιο άτομο το οποίο έχει πραγματοποιήσει χειρονομίες που έχουν χρησιμοποιηθεί στα σύνολα εκπαίδευσης.

Η αναζήτηση στο διαδίκτυο μπορεί να εμφανίσει πολλές μεθόδους ωστόσο αυτή που παρουσιάζεται στο [3] μοιάζει ιδιαίτερος καινοτόμος. Η ειδοποιός διαφορά σε σχέση με άλλες προσεγγίσεις είναι η χρήση ενός καπέλου με ενσωματωμένη κάμερα που παρέχει μία εναλλακτική οπτική γωνία πιο κοντινή στις καθημερινές συνθήκες (Εικόνα 1.6).



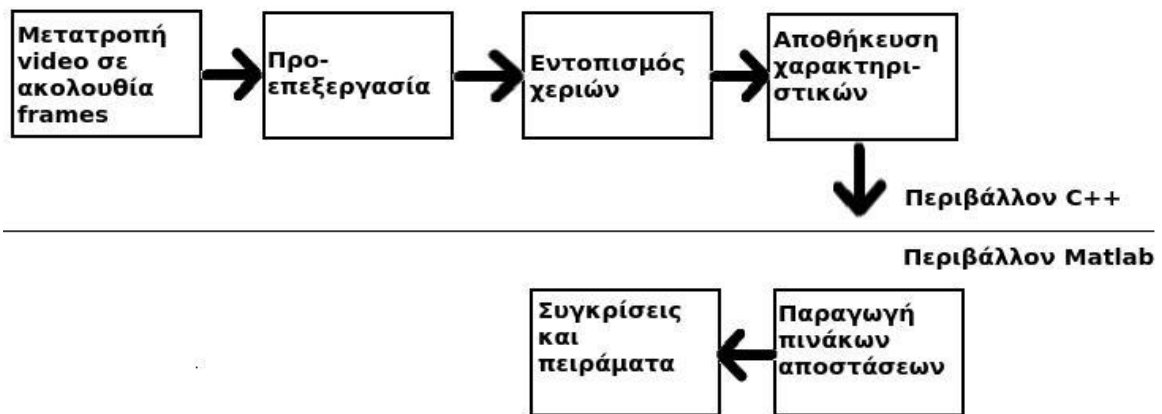
Εικόνα 1.6: Καταγραφή μέσω ενσωματωμένης κάμερας. Πηγή: [3].

Ως τελευταία μέθοδος αναγνώρισης αφέθηκε η [26] όπου χρησιμοποιείται εκ νέου η βάση δεδομένων της παρούσας διπλωματικής εργασίας σε συνδυασμό με dynamic time warping. Ως εξαχθέν χαρακτηριστικό χρησιμοποιείται το ιστόγραμμα προσανατολισμένων κλίσεων (Histogram of Oriented Gradients). Είναι εξαιρετικά σημαντικό να τονιστεί ώστε να υπάρχει ένα πλαίσιο σύγκρισης των αποτελεσμάτων ότι η συγκεκριμένη εργασία παρουσίασε ποσοστό ακρίβειας στην αναγνώριση των λέξεων στην πρώτη θέση (top 1) περίπου 57%, ενώ η αναγνώριση στις τρεις πρώτες θέσεις των επιλογών αγγίζει ποσοστό της τάξης του 74% σε ένα πλήθος 1000 λέξεων πραγματοποιημένες από τρία διαφορετικά άτομα.

1.6 Γρήγορη επισκόπηση της εργασίας

Μιλώντας γενικά η διπλωματική αυτή εργασία δεν ξεφεύγει από τα καθιερωμένα πλαίσια που έχουν τεθεί από όλες τις προαναφερθείσες προσπάθειες αλλά αντιθέτως υπακούει πιστά στις αρχές της αναγνώρισης χειρονομίας.

Έτσι σε πρώτο στάδιο γίνεται ένας διαχωρισμός των δεδομένων χειρονομιών ώστε από βίντεο να παραχθούν ακολουθίες εικόνων. Έπειτα ακολουθεί η προεπεξεργασία όπου αφαιρείται το περιβάλλον και ο θόρυβος. Στη συνέχεια εντοπίζονται τα αντικείμενα του κύριου ενδιαφέροντος μας όπως είναι το κεφάλι και τα χέρια. Ύστερα καθώς τρέχει ο αλγόριθμος εξαγωγής χαρακτηριστικών, αποθηκεύονται τα χαρακτηριστικά του κάθε χεριού αφού έχει προηγηθεί ένας συμβιβασμός σχετικά με την σειρά αποθήκευσης. Όταν ολοκληρωθεί η παραπάνω διαδικασία μεταφερόμαστε στο περιβάλλον όπου θα διεξαχθούν τα πειράματα ή θα παραχθούν τα νέα χαρακτηριστικά από την επεξεργασία των υπαρχόντων. Στα πλαίσια της αναγνώρισης της κάθε ξεχωριστής λέξης παράγονται οι πίνακες αποστάσεων ανάμεσα στους διαφορετικούς χρήστες και διεξάγονται τα πειράματα ανάμεσα σε 10 λέξεις. Η επιλογή των λέξεων δεν ήταν τυχαία, αλλά αντιθέτως επιλέχθηκαν λέξεις με σκοπό τον όσο το δυνατό πληρέστερο έλεγχο της αμερικάνικης νοηματικής γλώσσας από πλευράς πληθώρας κινήσεων και επικάλυψης χεριών. Όλη η διαδικασία μπορεί να περιγραφεί συνοπτικά από το παρακάτω διάγραμμα (Εικόνα 1.7).



Εικόνα 1.7: Διαδικασία της διπλωματικής εργασίας

1.7 Συμβιβασμοί και υποθέσεις

Η βάση δεδομένων παρουσιάζει άτομα λευκής απόχρωσης, τα οποία πραγματοποιούν τις χειρονομίες της νοηματικής γλώσσας καθήμενα μπροστά από ένα σκουρόχρωμο περιβάλλον. Είναι φυσικό λοιπόν στα πλαίσια της διπλωματικής εργασίας να γίνεται η υπόθεση ότι οι χρήστες όταν πραγματοποιούν τις χειρονομίες θα κάθονται σε καρέκλα. Επίσης, ο συμβιβασμός που προκύπτει από το χρώμα του δέρματος, των ρούχων και του περιβάλλοντος ορίζει ότι για την σωστή λειτουργία του συστήματος πρέπει ο χρήστης να πραγματοποιεί την χειρονομία με ρούχα και φόντο διαφορετικής απόχρωσης από αυτή του δέρματος του, όπου προτιμητέες αποχρώσεις είναι το άσπρο και το μαύρο. Έτσι αντίθετα με άλλες εργασίες, στην συγκεκριμένη διπλωματική δεν προκύπτουν περιορισμοί σχετικά με την απαραίτητη παρουσία μπλούζας με μακρύ μανίκι ώστε να φαίνεται αποκλειστικά η παλάμη και τα δάκτυλα, ούτε φυσικά και ύπαρξης γαντιών διαφορετικού χρώματος στα χέρια.

1.8 Στόχος και συνεισφορά

Η επιφανειακή επισκόπηση μερικών μόνο από τις προσεγγίσεις που έχουν κατά καιρούς δοκιμαστεί στον τομέα της αναγνώρισης χειρονομίας και δη στον τομέα της αναγνώρισης της αμερικάνικης νοηματικής γλώσσας δείχνει το εξαιρετικό ερευνητικό ενδιαφέρον που υπάρχει. Στόχος της παρούσας διπλωματικής εργασίας δεν είναι σε καμία περίπτωση η τετριμμένη εφαρμογή αναπτυχθέντων λύσεων αλλά η ουσιαστική συνεισφορά σε ερευνητικό επίπεδο με σκοπό να δοθεί το έναυσμα σε ακόμη περισσότερους ερευνητές να ασχοληθούν με την αναγνώριση της νοηματικής γλώσσας.

Όσον αφορά τον ιδεατό στόχο της εργασίας που είναι η προσφορά στο κοινωνικό σύνολο, αυτός θα έχει επιτευχθεί αν οποιοδήποτε μέρος της εργασίας εφαρμοστεί ή χρησιμοποιηθεί σε οποιαδήποτε άλλη προσπάθεια καταφέρει να επιλύσει το πρόβλημα με φερέγγυο τρόπο.

Λόγω του γεγονότος ότι είναι μια προσέγγιση επίλυσης του προβλήματος στα πλαίσια μιας διπλωματικής εργασίας υπάρχει το πλεονέκτημα της μη πίεσης για ευνοϊκά και άμεσα

αποτελέσματα.

Συνεπώς στα σημεία όπου οι προαναφερθείσες προσπάθειες κάνουν επίπονους για την καθημερινή χρήση συμβιβασμούς, η παρούσα εργασία επιλέγει τον δύσκολο δρόμο. Έτσι η συνεισφορά της γίνεται ουσιαστική καθώς η συγκεκριμένη υλοποίηση δεν περιέχει συμβιβασμούς με χρωματιστά γάντια, δεν χρησιμοποιεί κάποια βάση δεδομένων με τρισδιάστατες υψηλής ανάλυσης εικόνες, ούτε και κανενός άλλου είδους εξειδικευμένο εξοπλισμό. Ακόμη εξαιτίας του τρόπου επιλογής των λέξεων προς σύγκριση και έλεγχο, αναπαρίσταται μια πληθώρα κινήσεων των χεριών χωρίς να υπάρχει περιορισμός σχετικά με τον αριθμό των δρώντων σε μια χειρονομία χεριών για διευκόλυνση των αποτελεσμάτων.

Όσον αφορά τα εξαχθέντα χαρακτηριστικά, η πλειοψηφία έχει ενδεχομένως χρησιμοποιηθεί ξεχωριστά και σε εναλλακτικές προσεγγίσεις ωστόσο στα πλαίσια της εργασίας πάρθηκε η απόφαση να γίνει λεπτομερής καταγραφή των αποστάσεων και των διαφόρων πειραματικών συνδυασμών τους με απώτερο στόχο να είναι ευανάγνωστα και διαθέσιμα προς μελέτη σε κάθε ενδιαφερόμενο. Η κύρια επιστημονική συνεισφορά που προσδοκάται από την επεξεργασία των αποτελεσμάτων είναι η γνώση της θετικής συμπεριφοράς μεμονωμένων χαρακτηριστικών όπως η κλίση, η θέση και η μετατόπιση των χεριών στην αναγνώριση των λέξεων της νοηματικής. Επιπρόσθετα, κύριος στόχος είναι να παρουσιαστούν μέσα από τα επιτυχή αποτελέσματα οι δυνατότητες που προσφέρει η χρήση των χρησιμοποιηθέντων χαρακτηριστικών ώστε με την απαραίτητη επέκταση να δημιουργηθεί ένα χρήσιμο σύστημα αναγνώρισης.

1.9 Οργάνωση της διπλωματικής

Η διάρθρωση της εργασίας έγινε με τέτοιον τρόπο ώστε ο αναγνώστης να έρχεται σε επαφή με τα επιμέρους προβλήματα της αναγνώρισης χειρονομίας και έπειτα με τον τρόπο επίλυσης τους ούτως ώστε να κατανοηθεί πλήρως η διαδικασία αναγνώρισης την νοηματικής γλώσσας. Στο κεφάλαιο 2 που ακολουθεί μετά την εισαγωγή παρουσιάζεται ο τρόπος με τον οποίο εξήχθησαν τα χαρακτηριστικά που χρησιμοποιήθηκαν στην αναγνώριση, το είδος τους και τα προβλήματα που αντιμετωπίστηκαν. Στο κεφάλαιο 3 αντιμετωπίζεται το πρόβλημα της μηχανικής μάθησης και πιο συγκεκριμένα ο τρόπος με τον οποίο καταφέρνει ο υπολογιστής μέσω των εξαχθέντων χαρακτηριστικών να εντοπίσει τις λέξεις των οποίων οι αποστάσεις είναι οι μικρότερες και να τις ταξινομήσει. Στο κεφάλαιο 4 παρουσιάζονται τα πειραματικά αποτελέσματα της εργασίας ενώ στο κεφάλαιο 5 τα συμπεράσματα στα οποία οδηγούμαστε σχετικά με την αναγνώριση της νοηματικής γλώσσας και τους τομείς που χρήζουν βελτίωσης ή επέκτασης στο μέλλον.

2 Εξαγωγή χαρακτηριστικών

2.1 Προεπεξεργασία

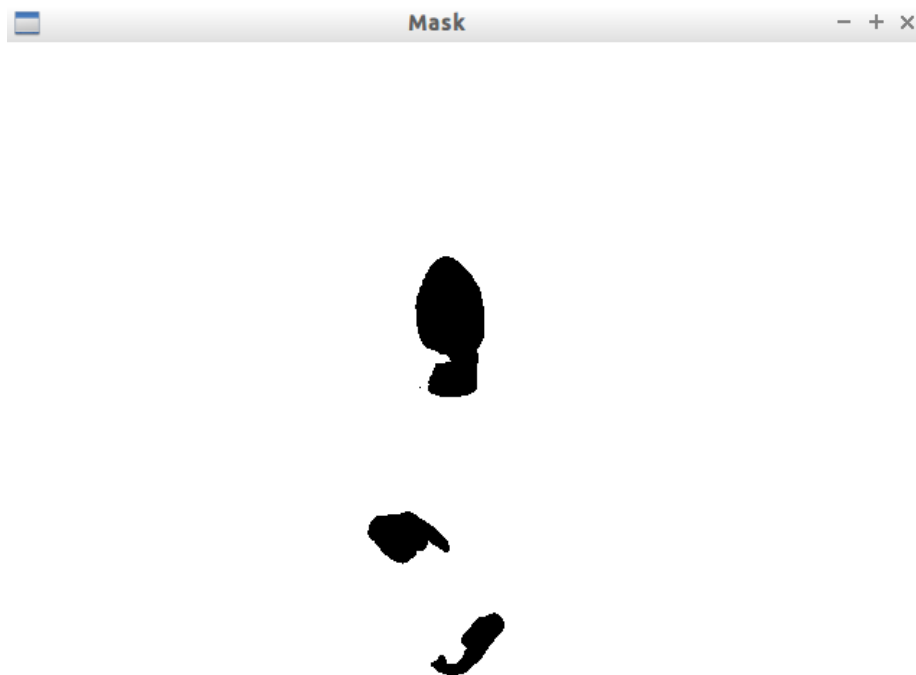
Όλη η διαδικασία της εξαγωγής των χαρακτηριστικών πραγματοποιήθηκε σε περιβάλλον προγραμματισμού C++. Προηγούμενης οποιασδήποτε προσπάθειας εντοπισμού των χειριών, ήταν αναγκαίο να διαβαστούν και να διαχωριστούν σε ακολουθίες εικόνων (frames) τα video της βάσης δεδομένων που ήταν σε συμπίεσμένη μορφή χωρίς απώλειες (lossless compressed video). Από το σημείο αυτό και έπειτα με κατάλληλους συμβιβασμούς ονοματοδοσίας κάθε χρήστης (signer) απέκτησε το δικό του σύνολο από λέξεις στο οποίο πραγματοποιείται η επεξεργασία και η αποθήκευση των εξαχθέντων χαρακτηριστικών.

2.2 Μάσκα χρώματος

Στα πλαίσια της διαδικασίας απομόνωσης των κρίσιμων αντικειμένων για την αναγνώριση της νοηματικής κρίθηκε αναγκαίο να απομακρυνθεί το φόντο της εικόνας και να διατηρηθούν μόνο τα σημεία όπου υπάρχει δέρμα. Αυτή η διαδικασία μπορεί να γίνει με πολλούς τρόπους, ωστόσο στα πλαίσια της διπλωματικής εργασίας επιχειρήθηκαν 2.

Ο πρώτος τρόπος οποίος δεν εξήγαγε τα βέλτιστα αποτελέσματα λόγω ιδιομορφιών της βάσης δεδομένων πραγματοποιείται επιλέγοντας τις περιοχές στις οποίες εμφανίζεται το χρώμα του δέρματος. Λαμβάνοντας για παράδειγμα το χρώμα από ένα pixel ανάμεσα στα μάτια του ατόμου και έπειτα αφαιρεί όλα τα pixel που απέχουν αρκετά από αυτό το χρώμα όπου το μέτρο εγγύτητας καθορίζεται εντός του αλγορίθμου.

Ο τρόπος που χρησιμοποιήθηκε και παρήγαγε τα καλύτερα αποτελέσματα προήλθε από την παρατήρηση ότι στη βάση δεδομένων τα άτομα ήταν λευκής απόχρωσης σε σκούρο μαύρο φόντο φορώντας ως επί το πλείστον ρούχα σκούρας επίσης απόχρωσης. Αυτή η παρατήρηση οδήγησε στην απόφαση να οριστεί ένα σημείο του φόντου ως χρωματικό σημείο αναφοράς και από αυτό να προκύψει μία μάσκα. Η μάσκα αυτή στην ουσία απεικονίζει την πιθανότητα που έχει κάποιο pixel της εικόνας να έχει χρώμα παρόμοιο με το χρώμα αναφοράς για αυτόν τον λόγο είναι σε αποχρώσεις του γκρι. Το άσπρο στην εικόνα αναπαριστά μεγάλη πιθανότητα ενώ μαύρο μικρή. Η μάσκα επομένως έως την συγκεκριμένη χρονική στιγμή είναι κυρίως άσπρη καθώς το φόντο αποτελεί το μεγαλύτερο μέρος της εικόνας αλλά έχουν αρχίσει εμφανώς να λείπουν (με μαύρο χρώμα) τα σημεία στα οποία υπάρχει δέρμα. Επειδή ο στόχος του συγκεκριμένου σταδίου είναι η παραγωγή εικόνας όπου λείπει το φόντο και εμφανίζονται μόνο οι περιοχές δέρματος θέτουμε μαύρα τα bits χρώματος της μάσκας που ξεπερνάνε μια συγκεκριμένη πιθανότητα (γκρι) ενώ τα bits της μάσκας που εμφανίζονταν εμφανώς μαύρα προηγουμένως, τα θέτουμε σε άσπρο ούτως ώστε να προκύψει το επιθυμητό αποτέλεσμα. Ως επιπλέον βοηθητικός παράγοντας επιτρέπεται κατά την εξαγωγή των χαρακτηριστικών να αυξομειωθεί το κατώφλι με το οποίο αποφασίζεται η εγγύτητα ενός χρώματος από ένα άλλο και κατ'επέκταση η συνένωση των αντίστοιχων pixels. Η συγκεκριμένη λειτουργία εξυπηρετεί τον στόχο για χρήστες διαφορετικών αποχρώσεων με ρούχα ποικίλων χρωματικών συνδυασμών. Στις εικόνες που ακολουθούν (Εικόνα 2.1–2.2) φαίνεται η παραγωγή της αρχικής μάσκας και ύστερα από την επεξεργασία με αντιστροφή του χρώματος και κανονικοποίηση για να εξαχθεί το περίγραμμα, η τελική μάσκα.



Εικόνα 2.1: Μάσκα χρώματος έχοντας αποβάλλει τα σημεία με χρώμα δέρματος



Εικόνα 2.2: Τελική μάσκα

2.3 Εντοπισμός αντικειμένων

Πριν αναγνωριστούν τα χέρια επιβάλλεται να εντοπιστούν τα αντικείμενα που προβάλλονται σε κάθε frame. Έχοντας ως είσοδο την εικόνα που προέκυψε έπειτα από την εφαρμογή της μάσκας χρώματος είναι εύκολα κατανοητό ότι οι μεγάλες μάζες λευκού χρώματος που ξεχωρίζουν είναι και διαφορετικά αντικείμενα. Χρησιμοποιώντας επομένως την κατάλληλη συνάρτηση της `opencv` βιβλιοθήκης ο αλγόριθμος είναι σε θέση να εντοπίσει όλα τα αντικείμενα που εμφανίζονται θέτοντας παράλληλα και ένα κατώφλι κανονικοποίησης του θορύβου για μάζες λευκού χρώματος που βρίσκονται σε κοντινές αποστάσεις. Η συνάρτηση που χρησιμοποιείται είναι η `findContours()` και στην ουσία πραγματοποιεί ανάλυση σχήματος και εντοπισμό άσπρων αντικειμένων σε μαύρο φόντο. Για να αυξηθεί η ακρίβεια των αποτελεσμάτων αυτού του σταδίου είναι απαραίτητο η είσοδος να είναι μία δυαδική εικόνα (`binary image`), προϋπόθεση που καλύπτεται από την υλοποίηση αυτής της εργασίας.

Σημαντικό εμπόδιο στον εντοπισμό των αντικειμένων αποτελούν οι εστίες θορύβου του περιβάλλοντος και της υλοποίησης της βάσης. Για παράδειγμα, ως εστία θορύβου θεωρείται η ύπαρξη μαλλιών μπροστά στο σώμα και κυρίως στις κρίσιμες περιοχές δράσης (στήθος) των χειρονομιών οι οποίες αλλοιώνουν το ακριβές περίγραμμα των αντικειμένων σε περίπτωση επικάλυψης.

2.4 Αναγνώριση χεριών

Από την επεξεργασία που έχει περιγραφεί έως τώρα έχει γίνει σαφές ότι ο υπολογιστής είναι σε θέση να εντοπίσει αντικείμενα στο πλάνο της εικόνας που έχουν το χρώμα του δέρματος. Στην φάση της αναγνώρισης των χεριών ο αλγόριθμος πρέπει έχοντας ως είσοδο την παραπάνω πληροφορία να ξεχωρίσει ποιες μάζες (`blobs`) αναπαριστούν χέρια, ποιο `blob` είναι το κεφάλι και να εκτελεστεί διαχωρισμός ανάμεσα στο αριστερό και το δεξί χέρι. Η διαδικασία αυτή μόνο ως τετριμμένη δεν μπορεί να χαρακτηριστεί καθώς η εξαιρετική ποικιλία των χειρονομιών της νοηματικής στην οποία τα χέρια μπορούν ανά πάσα στιγμή να βρίσκονται σε οποιαδήποτε θέση είτε ακόμα και να επικαλύπτονται μεταξύ τους ή με το κεφάλι, ορίζει την χρησιμοποίηση πολλών παραμέτρων καθορισμού θέσης. Επίσης, η πολυπλοκότητα της αναγνώρισης αυξάνεται από το γεγονός ότι οι χρήστες μπορεί να έχουν διαφορετικό σωματότυπο.

Εισχωρώντας βαθύτερα στην λειτουργία της εργασίας, αφού έχει τρέξει ο αλγόριθμος εντοπισμού αντικειμένων ο υπολογιστής γνωρίζει το πλήθος αυτών, τις θέσεις και το εμβαδόν τους. Πραγματοποιώντας την λογική υπόθεση ότι από τα επιστρεφόμενα αντικείμενα της εικόνας που έχουν το χρώμα του δέρματος τα 3 μεγαλύτερα θα είναι αυτά του κεφαλιού και των 2 χεριών συνεχίζουμε την διαδικασία εντοπισμού. Βέβαια σε περιπτώσεις επικάλυψης χεριών τότε τα 2 μεγαλύτερα αντικείμενα είναι κεφάλι και χέρια οπότε η απομάκρυνση των υπολοίπων γίνεται εφαρμόζοντας έναν περιορισμό σχετικά με το ελάχιστο εμβαδόν που μπορεί να καταλαμβάνει ένα αντικείμενο με χρώμα δέρματος. Επανερχόμενοι όμως στην γενική περίπτωση με τα 3 μεγαλύτερα αντικείμενα είναι εύκολο να κατανοήσει κανείς ότι το κεφάλι θα είναι το αντικείμενο του οποίου το κέντρο βρίσκεται υψηλότερα σε σχέση με τα υπόλοιπα οπότε το πρώτο κρίσιμο ερώτημα για το `blob` που αναπαριστά το κεφάλι απαντήθηκε όπως επίσης απαντήθηκε και το ερώτημα σχετικά με τα `blobs` που αναπαριστούν χέρια. Χωρίζοντας νοητά την εικόνα σε 4 τμήματα κατά τον άξονα x μπορούμε να θεωρήσουμε ότι τα κέντρα των `blobs`

που αναπαριστούν χέρια θα δρουν κυρίως γύρω από τα 2 μεσαία τμήματα οπότε τυχόν blobs από θόρυβο στα ακραία τμήματα (για παράδειγμα λόγω φωτεινότητας) απομακρύνονται (η ίδια διαδικασία ορίζεται και για τμήματα στον άξονα y). Έτσι από τα blobs που βρίσκονται προς την μέση, το blob του οποίου το κέντρο πέφτει αριστερότερα (όπως κοιτάμε την εικόνα) αναπαριστά με μεγάλη ακρίβεια το δεξί χέρι ενώ το κέντρο που βρίσκεται δεξιότερα αναπαριστά το αριστερό χέρι.

Στην παρακάτω εικόνα (Εικόνα 2.3) εμφανίζεται ένα παράδειγμα για την αναγνώριση των διαφορετικών χεριών τα οποία μπαίνουν σε διαφορετικό κουτί (bounding box) ενώ φαίνονται και ορισμένα από τα χαρακτηριστικά όπως αυτό της κλίσης και του μήκους των ανάλογων αξόνων. Ο κύκλος που εμφανίζεται με πράσινο χρώμα πάνω αριστερά αποτελεί το σημείο δειγματοληψίας χρώματος.



Εικόνα 2.3: Αναγνώριση των χεριών και εμφάνιση αξόνων κλίσης μέσα σε τετράγωνο

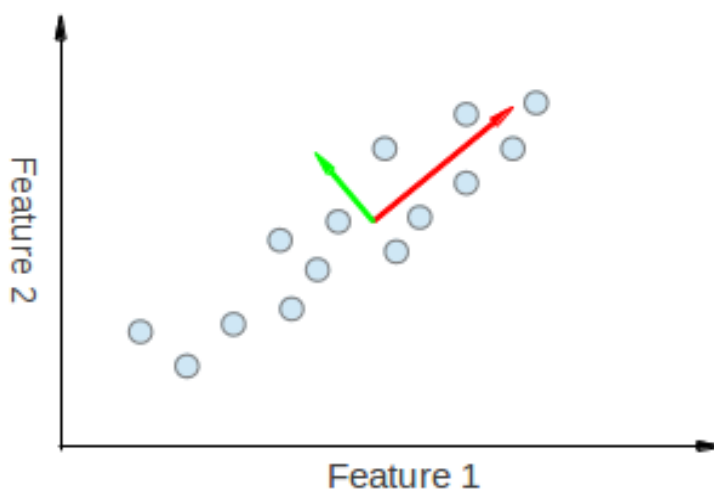
2.5 Εξαγωγή σημείων κέντρου

Από την στιγμή που αναγνωρίστηκαν οι θέσεις των χεριών σε κάθε frame είναι επόμενο τα σημεία του κέντρου αυτών των θέσεων να αποτελέσουν το πρώτο πολύ βασικό εξαχθέν χαρακτηριστικό. Το συγκεκριμένο χαρακτηριστικό θα βοηθήσει να ομαδοποιηθούν και εν τέλει να ταξινομηθούν οι χειρονομίες στις οποίες τα χέρια βρίσκονται σε διαφορετικές θέσεις πάνω στην εικόνα. Για παράδειγμα μια χειρονομία στην οποία χρησιμοποιείται μόνο το ένα χέρι θα

είναι εμφανώς διαφορετική από μία χειρονομία που χρησιμοποιούνται και τα δύο χέρια καθώς στην μία περίπτωση το αριστερό χέρι (για δεξιόχειρα χρήστη) θα βρίσκεται σταθερό χαμηλά στην εικόνα ενώ στην άλλη θα αλλάζει θέσεις.

2.6 Εξαγωγή κλίσης

Ύστερα από μελέτη διαφορετικών εργασιών αναγνώρισης χειρονομίας αλλά και παρατήρηση αρκετών λέξεων της νοηματικής προέκυψε ότι η κλίση που έχει ένα αντικείμενο αναγνωρισμένο ως χέρι σε σχέση με το ορθοκανονικό σύστημα αξόνων θα ήταν ένα βολικό χαρακτηριστικό διαφοροποίησης λέξεων. Αυτό συμβαίνει για παράδειγμα σε λέξεις που περιέχουν είτε ενός είδους κλίση κατά τη διάρκεια των χειρονομιών είτε κυκλοειδής κινήσεις που επίσης επηρεάζουν πολύ την κλίση. Για να βρεθεί και να αποθηκευτεί το συγκεκριμένο χαρακτηριστικό έγινε χρήση μιας ιδιότητας της ανάλυσης κύριων συνιστωσών (PCA) η οποία ορίζει [20] ότι εκτελώντας την συγκεκριμένη ανάλυση στα ιδιοδιανύσματα ενός συνόλου σημείων, επιστρέφεται ως αποτέλεσμα η διεύθυνση κατά την οποία αυτά τα σημεία διαφοροποιούνται περισσότερο.



Εικόνα 2.4: Σχηματικό κλίσης ύστερα από ανάλυση κύριων συνιστωσών. Πηγή: [25].

2.7 Εξαγωγή μηκών αξόνων

Έχοντας βρει την κλίση των χεριών, δηλαδή την διεύθυνση κατά την οποία τα σημεία του αντικειμένου διαφοροποιούνται περισσότερο, γεννήθηκε η ιδέα να χρησιμοποιήσουμε και τα μήκη των αξόνων που επιστρέφονται από την παραπάνω διαδικασία ώστε να δοθεί η δυνατότητα στις λέξεις να ξεχωρίζουν και να ομαδοποιούνται με κριτήριο το πόσο μεγάλη είναι η διαφοροποίηση ως προς τον κύριο άξονα διαφοροποίησης. Το συγκεκριμένο χαρακτηριστικό εξαρχής δεν αναμένεται να βοηθήσει σε όλες τις περιπτώσεις χειρονομιών καθώς από την φύση

τους τα χέρια μπορεί να έχουν μόνο ένα ορισμένο μήκος ως προς την κύρια διεύθυνση τους. Ωστόσο χρησιμοποιήθηκε καθώς σε περιπτώσεις επικάλυψης όπου τα μήκη αλλάζουν με απρόβλεπτο τρόπο είναι ένα καλό κριτήριο για να ομαδοποιηθούν λέξεις με πολλά frames επικάλυψης και να διαχωριστούν από αυτές που παρουσιάζουν μηδενικά ποσοστά επικάλυψης είτε ανάμεσα στα χέρια είτε ανάμεσα σε χέρι και κεφάλι.

2.8 Επεξεργασία για παραγωγή της μετατόπισης dx-dy

Έπειτα από ενδελεχή παρατήρηση των χειρονομιών της νοηματικής προήλθε και η ιδέα χρήσης των πρώτων παραγώγων της θέσης των χεριών. Παρατηρήθηκε ότι η πλειοψηφία των χειρονομιών εμπεριέχει πολλή κίνηση και πολλές μετατοπίσεις του κέντρου των χεριών επομένως κρίθηκε χρήσιμο να εξαχθεί ένα χαρακτηριστικό που να μπορεί να ελέγξει τέτοιου είδους διαφοροποιήσεις ανάμεσα στις λέξεις. Για να εξαχθούν οι μετατοπίσεις μεταφερόμαστε από το περιβάλλον της C++ στο περιβάλλον του matlab όπου αφού παράγονται οι μετατοπίσεις αποθηκεύονται στο ίδιο κοινό αρχείο με τα υπόλοιπα χαρακτηριστικά. Για να παραχθούν τα διανύσματα των μετατοπίσεων Dx και Dy εφαρμόστηκε ένα φίλτρο που μοιάζει με αυτό του sobel ($[-1 \ 0 \ 1]$) πάνω στα διανύσματα των κέντρων. Η προκύπτουσα πράξη στην ουσία είναι μία αφαίρεση της προηγούμενης από την επόμενη θέση για να βρούμε την παρούσα μετατόπιση όπως περιγράφεται συνοπτικά από τον τύπο:

$$Dx(n) = X(n+1) - X(n-1)$$

2.9 Αρχείο χαρακτηριστικών

Όπως προκύπτει από όλα τα παραπάνω για την ταξινόμηση των λέξεων προκύπτει ένα κοινό αρχείο χαρακτηριστικών με 14 στήλες (14-dimensional) και αριθμών γραμμών ανάλογο του μήκους των χειρονομιών (για διευκόλυνση κατά την προεπεξεργασία τέθηκε το μήκος κάθε χειρονομίας ίσο με 100 frames). Η κλίση των χεριών μετράται σε ακτίνια (rads), ενώ οι μετατοπίσεις, τα μήκη, οι μεταβλητές θέσης x και y είναι τύπου αριθμού κινητής υποδιαστολής (float). Το αρχείο που δημιουργείται για κάθε λέξη και κάθε frame επομένως είναι της μορφής που παρουσιάζεται στον παρακάτω πίνακα (Πίνακας 2).

Orientation Right Hand	Orientation Left Hand	X Right Hand	X Left Hand	Y Right Hand	Y Left Hand	X axis Right	X axis Left
------------------------------	-----------------------------	--------------------	-------------------	--------------------	-------------------	--------------------	-------------------

Πίνακας 2: Παράδειγμα εξαχθέντος αρχείου χαρακτηριστικών.

Dx Right Hand	Dy Right Hand	Dx Left Hand	Dy Left Hand
---------------------	---------------------	--------------------	--------------------

Πίνακας 3: Τα παραγόμενα χαρακτηριστικά μετατόπισης με την σειρά με την οποία συνενώνονται στο αρχείο χαρακτηριστικών (συνέχεια του πίνακα 2).

3 Μηχανική μάθηση

Το κύριο ερώτημα στο οποίο προσπαθεί να δώσει απάντηση η μηχανική μάθηση (Machine Learning) είναι το πώς είναι δυνατόν να δημιουργούνται υπολογιστικά συστήματα τα οποία θα έχουν την δυνατότητα να βελτιώνονται χρησιμοποιώντας την εμπειρία που αποκομίζουν από την καθημερινή χρήση.

3.1 Euclidean distance

Στην μηχανική μάθηση μία από τις πιο σημαντικές και βασικές λειτουργίες είναι η εύρεση αποστάσεων ανάμεσα σε διανύσματα χαρακτηριστικών με σκοπό την μετέπειτα σύγκριση τους. Ένα από τα βασικότερα εργαλεία εύρεσης τέτοιων αποστάσεων είναι η Ευκλείδεια απόσταση (Euclidean distance) [21]. Η Ευκλείδεια απόσταση ή Ευκλείδεια νόρμα ανάμεσα σε δύο σημεία του καρτεσιανού επιπέδου είναι το μήκος του ευθύγραμμου τμήματος που συνδέει αυτά τα σημεία. Όπως είναι φανερό η επέκταση του παραπάνω ορισμού σε περισσότερες διαστάσεις αποτελεί έναν ολοκληρωμένο τρόπο εύρεσης των αποστάσεων ανάμεσα σε εξαχθέντα χαρακτηριστικά. Στην παρούσα εργασία ο τύπος που χρησιμοποιήθηκε για την σύγκριση του κάθε είδους χαρακτηριστικών είναι:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + \dots + (p_n - q_n)^2}$$

Επομένως ο συγκεκριμένος τύπος χρησιμοποιείται στην εργασία για εύρεση της απόστασης μόνο χαρακτηριστικών του ίδιου είδους καθώς αν χρησιμοποιηθεί με συνδυασμό ειδών χαρακτηριστικών υπάρχει ο κίνδυνος η απόσταση του ενός χαρακτηριστικού να κυριαρχήσει στην τελική σύγκριση.

3.2 Mahalanobis distance

Με τον όρο Mahalanobis distance [21] ορίζουμε ένα μέτρο της απόστασης ανάμεσα σε ένα σημείο και μία κατανομή. Είναι στην ουσία μια πολυδιάστατη γενίκευση της ιδέας της μέτρησης της απόστασης ενός σημείου από το μέσο της κατανομής σε σχέση με την τυπική απόκλιση. Επομένως η απόσταση είναι μηδενική αν το σημείο είναι ακριβώς το μέσο της κατανομής και αυξάνεται καθώς απομακρύνεται από αυτό το σημείο.

Ο γενικός τύπος της απόστασης Mahalanobis είναι:

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}$$

Όπου D είναι η απόσταση μιας παρατήρησης x από ένα σύνολο παρατηρήσεων με μέσο μ και S το μητρώο συνδιασποράς.

Ο τύπος ωστόσο που χρησιμοποιείται στην παρούσα διπλωματική παράγεται όταν ο πίνακας συνδιασποράς είναι διαγώνιος (παντού μηδέν εκτός από την διαγώνιο) και περιγράφεται από τον κάτωθι τύπο:

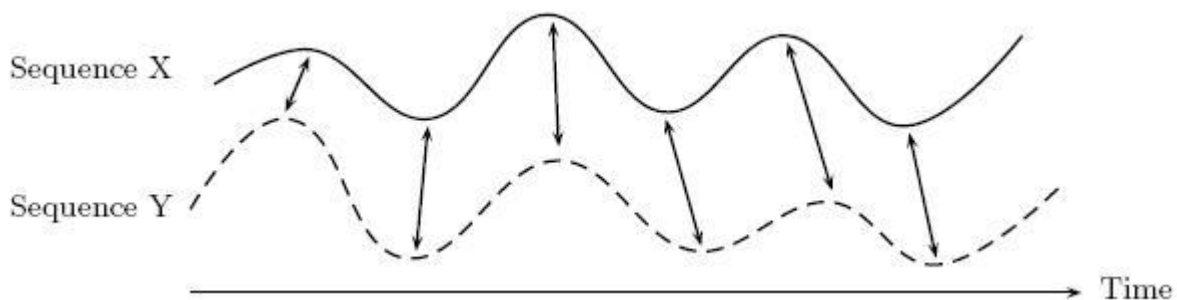
$$d(x, y) = \sqrt{\sum_{i=1}^N \left(\frac{(x_i - y_i)^2}{s_i} \right)}$$

Όπου το s_i είναι η διασπορά των x_i και y_i πάνω στο σύνολο των χαρακτηριστικών, ενώ ο δείκτης i έχει μήκος N όσο τα σήματα.

Ο παραπάνω τύπος ονομάζεται και κανονικοποιημένη Ευκλείδεια απόσταση καθώς η προκύπτουσα απόσταση δεν υποφέρει από το φαινόμενο της απλής Ευκλείδειας απόστασης όπου υπάρχει ο κίνδυνος κυριαρχίας του ενός χαρακτηριστικού έναντι των άλλων.

3.3 Dynamic Time Warping

Αποτελεί έναν βασικό αλγόριθμο μέτρησης της εγγύτητας μεταξύ 2 σημάτων που εξελίσσονται στον χρόνο. Το βασικό πλεονέκτημα του είναι ότι μπορεί να ελέγξει την ισότητα 2 σημάτων τα οποία μπορεί να εξελίσσονται με διαφορετικό τρόπο στον χρόνο όσον αφορά την ταχύτητα τους και να βρει το βέλτιστο ταίριασμα του ενός σήματος πάνω από το άλλο. Ο καλύτερος τρόπος κατανόησης των δυνατοτήτων του αλγορίθμου έρχεται μέσω ενός παραδείγματος που αφορά την εγγύτητα του τρόπου περπατήματος 2 ατόμων. Η συγκεκριμένη μέθοδος είναι ικανή να βρει τις ομοιότητες του τρόπου περπατήματος ακόμα και αν υπάρχει διαφορετική ταχύτητα, επιταχύνσεις ή επιβραδύνσεις. Η βασική λειτουργία του αλγορίθμου είναι να υπολογίζει το βέλτιστο ταίριασμα (μονοπάτι) ενός σήματος πάνω από ένα άλλο και έπειτα να υπολογίζει την απόσταση που έχουν οι 2 ακολουθίες δεδομένων όταν η σύγκριση γίνεται με το βέλτιστο ταίριασμα.

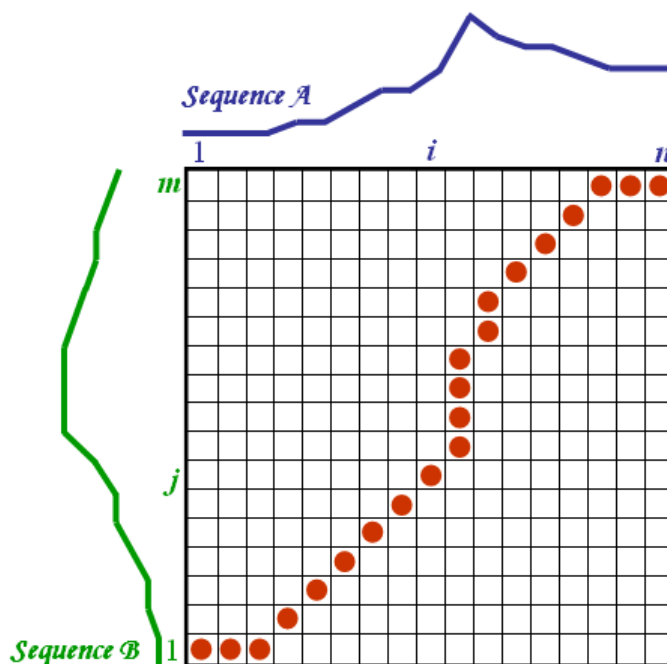


Εικόνα 3.1: Ταίριασμα 2 σημάτων μέσω DTW. Πηγή: [4].

Από τα παραπάνω είναι φανερό ότι η συγκεκριμένη μέθοδος είναι αρκετά βολική για την σύγκριση των λέξεων της νοηματικής καθώς τα δεδομένα έρχονται τμηματικά και πραγματικά υπάρχει μεγάλη πιθανότητα η ταχύτητα με την οποία πραγματοποιούνται τα νοήματα από τους χρήστες να διαφέρει. Για να βρεθεί το βέλτιστο ταίριασμα ανάμεσα σε δύο ακολουθίες από διακριτά δεδομένα όπως αυτά της εργασίας, απαιτείται ο έλεγχος κάθε δυνατού ταιριάσματος

ανάμεσα στις ακολουθίες επομένως η υπολογιστική πολυπλοκότητα αυξάνεται εκθετικά ανάλογα με το μήκος των ακολουθιών. Αντίθετα προκύπτει $O(NM)$ χρησιμοποιώντας δυναμικό προγραμματισμό με αναδρομικές κλήσεις της συνάρτησης.

Στην ουσία η εύρεση του βέλτιστου ταιριάσματος μπορεί να μεταφραστεί στην εύρεση του βέλτιστου μονοπατιού από την θέση κάτω αριστερά έως την θέση πάνω δεξιά του πίνακα που παράγεται αν τεθούν στον άξονα X τα δεδομένα της μίας ακολουθίας και στον άξονα Y τα δεδομένα της άλλης όπως φαίνεται και στην παρακάτω εικόνα.

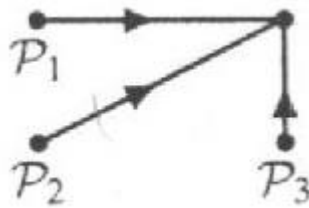


Εικόνα 3.2: Πίνακας μονοπατιών DTW. Πηγή: [5].

Το κόστος κάθε κελιού υπολογίζεται χρησιμοποιώντας κάποια μέθοδο υπολογισμού απόστασης όπως αυτές που αναφέρθηκαν σε προηγούμενα κεφάλαια. Για να βρεθεί το μονοπάτι πραγματοποιούνται συγκρίσεις του κόστους της θέσης στην οποία βρίσκεται ο αλγόριθμος με τις διπλανές θέσεις και έτσι αποφασίζεται η συνέχιση του αλγορίθμου από την θέση με το μικρότερο κόστος ώστε τελικά προσθέτοντας όλα αυτά τα μικρότερα κόστη να προκύπτει το μονοπάτι με το μικρότερο δυνατό κόστος.

Οι συγκρίσεις που πραγματοποιούνται για τον αλγόριθμο στην αναδρομική του μορφή εμφανίζονται στην παρακάτω εικόνα (Εικόνα 3.3) ενώ το τύπος που προκύπτει για το κόστος κάθε θέσης του πίνακα είναι:

$$\text{cell}(i,j) = \text{local_distance}(i,j) + \text{MIN}(\text{cell}(i-1,j), \text{cell}(i-1,j-1), \text{cell}(i, j-1))$$



Εικόνα 3.3: Συγκρινόμενα στοιχεία για την εύρεση του κόστους κελιού. Πηγή: [5].

Στην παρούσα διπλωματική εργασία δίνεται η δυνατότητα εκτέλεσης του αλγορίθμου dynamic time warping τόσο στην συμβατική του μορφή, όσο και με βήμα μεγαλύτερο του ενός δηλαδή να πραγματοποιείται επέκταση της αναζήτησης για θέσεις με την μικρότερη δυνατή απόσταση σε κάποιες θέσεις του πίνακα πιο μακριά από τις αμέσως διπλανές.

3.4 Πειραματικό περιβάλλον

Η πραγματοποίηση των πειραμάτων γίνεται σε περιβάλλον Matlab (Matrix Laboratory). Η επιλογή του συγκεκριμένου περιβάλλοντος έγινε καθώς επιτρέπει με εξαιρετικά εύκολο και γρήγορο τρόπο τον χειρισμό πολλών λειτουργιών που σχετίζονται με πίνακες αλλά έχει και την δυνατότητα εκτέλεσης κώδικα σε γλώσσα προγραμματισμού C++ μέσω του προεγκατεστημένου compiler. Έτσι λειτουργίες όπως εξαγωγή των δεδομένων από το αρχείο χαρακτηριστικών και εισαγωγή σε πίνακα, ή επιλογή και απομόνωση συγκεκριμένων στηλών του πίνακα χαρακτηριστικών με σκοπό την πραγματοποίηση διαφορετικών πειραματικών συνδυασμών διευκολύνονται. Η προαναφερθείσα διευκόλυνση έγκειται στο γεγονός ότι από την στιγμή που το Matlab είναι ένα ξεχωριστό περιβάλλον εξειδικευμένο στην διαχείριση πινάκων η εκτέλεση των πράξεων επιταχύνεται ενώ ως επιπλέον πλεονέκτημα μπορεί να θεωρηθεί ότι οποιαδήποτε ερευνητική ομάδα μπορεί να διεξάγει εκ νέου πειράματα απλά και μόνο έχοντας στην κατοχή της τα αρχεία χαρακτηριστικών.

3.5 Κλάσεις

Η επιλογή των λέξεων προς αναγνώριση δηλαδή των κλάσεων μπορεί να γίνει με διάφορους τρόπους. Ο πιο εύκολος τρόπος είναι η επιλογή πολύ συγκεκριμένων λέξεων που τα αποτελέσματά τους ευνοούν την παραγωγή επιστημονικής φαινομενικής συνεισφοράς. Τέτοιου είδους λέξεις είναι αυτές στις οποίες απαιτείται μόνο η αναγνώριση ενός χεριού για την πραγματοποίησή τους. Επίσης τέτοιου είδους λέξεις είναι αυτές οι οποίες δεν παρουσιάζουν μεγάλη κινητικότητα αλλά αντίθετα τα χέρια έπειτα από λίγο μένουν σταθερά σε κάποιο σημείο. Ένας ακόμη παράγοντας διευκόλυνσης των αποτελεσμάτων είναι η επιλογή λέξεων στις οποίες τα δάκτυλα είναι εξαιρετικά ευδιάκριτα χωρίς επικάλυψη μεταξύ τους. Για το τέλος αφέρθηκε το κριτήριο του μεγέθους επικάλυψης ανάμεσα στα χέρια και το κεφάλι το οποίο αν είναι μικρό έως ανύπαρκτο αποτελεί παράγοντα παραγωγής επιτυχών αποτελεσμάτων.

Οι λέξεις οι οποίες επιλέχθηκαν δεν χρησιμοποιούν κανέναν από τους παραπάνω τρόπους διευκόλυνσης της αναγνώρισης εκτός ίσως από το μέγεθος της επικάλυψης των χεριών. Έτσι έχουν εισαχθεί προς αναγνώριση 5 λέξεις ενός χεριού αλλά και 5 λέξεις δύο χεριών ώστε

να εξασφαλίζεται η ποικιλία και σωστή αποτύπωση των λέξεων της νοηματικής κατά τον έλεγχο. Επίσης έχουν εισαχθεί λέξεις στις οποίες τα δάκτυλα είναι από αρκετά ευδιάκριτα έως και εντελώς απόντα από το πλάνο της εικόνας. Η κινητικότητα των χεριών επίσης ποικίλει και μάλιστα για την αναγνώριση νοημάτων με μεγάλη κίνηση έχει εξαχθεί ειδικό χαρακτηριστικό. Όσον αφορά την επικάλυψη των χεριών έχουν εισαχθεί λέξεις που επικαλύπτονται τα χέρια με το κεφάλι για να εξασφαλιστεί πληρέστερος έλεγχος ωστόσο είναι γεγονός ότι αποτελεί έναν από τους παράγοντες που χρήζουν βελτίωσης, ίσως χρησιμοποιώντας την πλαϊνή κάμερα που διατίθεται με την βάση δεδομένων και θα αναλυθεί παρακάτω στο κεφάλαιο της μελλοντικής εργασίας και επέκτασης.

Το σύνολο των λέξεων που χρησιμοποιούνται είναι 10 και αποτελείται από τις λέξεις του παρακάτω πίνακα:

1 χέρι	2 χέρια
come on	adopt
dog	coat
don't like	behavior
cough	awkward
down	dress

Πίνακας 4: Κλάσεις-λέξεις εργασίας

4 Βάση δεδομένων και πειραματικά αποτελέσματα

4.1 Σύνοψη

Για τα πειραματικά αποτελέσματα εξήχθησαν 4 χαρακτηριστικά σε έναν πίνακα χαρακτηριστικών 14 στηλών. Κάθε χρήστης πραγματοποίησε από μία φορά κάθε μία από τις 10 λέξεις-κλάσεις προς διερεύνηση ενώ από τους 3 διαθέσιμους χρήστες οι 2 χρησιμοποιήθηκαν για την εκπαίδευση του αλγορίθμου και ο εναπομείναντας για την δοκιμή του. Η επιλογή της αναλογίας χρηστών εκπαίδευσης και δοκιμής δεν ήταν τυχαία καθώς όπως και ο προφορικός λόγος έτσι και η νοηματική παρουσιάζει διαφορές από άτομο σε άτομο (προφορά) οπότε πριν την δοκιμή απαιτείται κανονικοποίηση των δεδομένων σύγκρισης. Αρχικά οι συγκρίσεις διεξήχθησαν ανάμεσα σε όμοια χαρακτηριστικά για να αποκτηθεί διαισθητική γνώση για το είδος του χαρακτηριστικού που συμπεριφέρεται καλύτερα. Έστερα τα χαρακτηριστικά συνενώθηκαν με διάφορους συνδυασμούς ώστε να βρεθεί ο συνδυασμός που παράγει τα καλύτερα αποτελέσματα. Συνοψίζοντας, την καλύτερη επίδοση είχε ο συνδυασμός των χαρακτηριστικών της κλίσης, της θέσης και της μετατόπισης των χεριών σε κάθε frame με 60% ακρίβεια κατάταξης στην πρώτη θέση και 90% κατάταξης στις τρεις πρώτες θέσεις. Ο συγκεκριμένος συνδυασμός ήταν ελάχιστα καλύτερος από τον αμέσως επόμενο που χρησιμοποιούσε δύο μόνο χαρακτηριστικά, αυτά της κλίσης και της θέσης. Η χειρότερη επίδοση παρουσιάστηκε ενώνοντας τα χαρακτηριστικά της κλίσης και του μήκους των αξόνων, όπου διακρίνεται μία γενικώς δυσμενής συμπεριφορά του δεύτερου χαρακτηριστικού ενώ τα ποσοστά επιτυχούς αναγνώρισης είναι 20% για την πρώτη θέση και 40% για κατάταξη στις θέσεις της πρώτης τριάδας.

4.2 Βάση δεδομένων

Η βάση δεδομένων που χρησιμοποιήθηκε προέρχεται από το πανεπιστήμιο της Βοστώνης και είναι σε μορφή lossless compressed video όπου με χρήση κατάλληλου κώδικα μπορούν να παραχθούν οι έγχρωμες ακολουθίες εικόνων. Διαθέσιμες προς επεξεργασία είναι 3 όψεις. Αυτές αποτελούνται από μία γενική όψη που χρησιμοποιήθηκε και στην παρούσα εργασία όπου φαίνονται τα χέρια και το κεφάλι του χρήστη, μία πλαϊνή όψη και μία όψη η οποία επικεντρώνεται στο πρόσωπο του χρήστη που κάνει την χειρονομία. Παράλληλα με τα video είναι διαθέσιμη και πληροφορία για τα frames στα οποία βρίσκεται κάθε χειρονομία ώστε να καταστεί ευκολότερη η περαιτέρω έρευνα από ομάδες ανά τον κόσμο. Για κάθε μία από τις παραπάνω όψεις η λήψη έγινε με 60 frames ανά δευτερόλεπτο και ανάλυση 640x480 pixels ανά frame. Στην συγκεκριμένη βάση 3 χρήστες των οποίων η νοηματική είναι μητρική γλώσσα και 3 μαθητές της νοηματικής έχουν πραγματοποιήσει περίπου 3000 λέξεις ο καθένας. Τα άτομα είναι όλα λευκής απόχρωσης καθήμενα σε καρέκλα μπροστά από μαύρο φόντο. Ιδιαίτερη προσοχή δεν έχει δοθεί στα ρούχα που φοράνε, κάτι που δημιουργεί προβλήματα σε περιπτώσεις όπου εμφανίζονται σχέδια. Επίσης λόγω της βάσης δεδομένων προβλήματα δημιουργούνται όταν πραγματοποιούνται συγκρίσεις ανάμεσα σε άτομα που φοράνε μπλούζες με μακρύ μανίκι και άτομα που φοράνε κοντό. Παράλληλα εκτός των δεδομένων που

παρέχονται σε lossless compressed μορφή, δίνονται και video σε μορφή quicktime στα οποία η πληροφορία της έναρξης και της λήξης κάθε χειρονομίας παρουσιάζεται στο πάνω μέρος της εικόνας.



Εικόνα 4.1: Οι 3 όψεις της βάσης δεδομένων. Πηγή: [2].

4.3 Αποτελέσματα συγκρίσεων

Αποδεχόμενοι πως τα αποτελέσματα της παρούσας διπλωματικής είναι αναγκαίο να είναι σε μία μορφή η οποία θα εξυπηρετεί μελλοντικούς ερευνητές στην κατανόηση της ή κατασκευαστές συστημάτων αναγνώρισης της νοηματικής γλώσσας αποφασίστηκε να παραχθούν αποτελέσματα σχετικά με την προκύπτουσα ακρίβεια για τις 3 επικρατέστερες επιλογές (top1-top2-top3).

Στον παρακάτω πίνακα (Πίνακας 4) ακολουθούν τα πειραματικά αποτελέσματα που προέκυψαν συγκρίνοντας μόνο ένα είδος χαρακτηριστικών. Η σύγκριση έγινε με DTW, λαμβάνοντας τους πίνακες αποστάσεων χρησιμοποιώντας Ευκλείδια απόσταση. Για την δημιουργία των πινάκων αποστάσεων πραγματοποιήθηκε σύγκριση ανάμεσα σε κάθε λέξη του ατόμου (signer) του οποίου τα χαρακτηριστικά χρησιμοποιούνται στο σύνολο δοκιμής με κάθε λέξη των άλλων δύο ατόμων τα χαρακτηριστικά των οποίων χρησιμοποιούνται ως διάνυσμα εκπαίδευσης. Ύστερα από την ολοκλήρωση της διαδικασίας οι πίνακες αποστάσεων από τα δύο σύνολα δοκιμής συνενώνονται και γίνεται αναζήτηση για τις μικρότερες 3 αποστάσεις στην αντίστοιχη γραμμή που ανταποκρίνεται στην λέξη προς διερεύνηση.

<i>Feature compared</i>	<i>Top 1</i>	<i>Top 2</i>	<i>Top 3</i>
Orientation	30%	60%	80%
X,Y centre points	50%	50%	60%
Axis lengths	30%	50%	60%
Dx, Dy	20%	30%	40%

Πίνακας 5: Σύγκριση μεμονωμένων χαρακτηριστικών

Αντίθετα με την παραπάνω προσέγγιση της σύγκρισης, για να συγκριθούν οι

συνδυασμοί των χαρακτηριστικών επιβάλλεται οι πίνακες αποστάσεων να δημιουργηθούν χρησιμοποιώντας DTW με μέτρο απόστασης Mahalanobis (όπως αυτή περιγράφηκε στο αντίστοιχο κεφάλαιο) ώστε να μην κινδυνεύουν τα αποτελέσματα από την κυριαρχία της διαφοράς ενός συγκεκριμένου χαρακτηριστικού. Παρατηρούμε στον παρακάτω πίνακα (πίνακας 5) το πλήθος των χαρακτηριστικών που χρησιμοποιούνται σε κάθε σύγκριση αλλά και ποιος είναι ο συνδυασμός που συγκρίνεται.

<i>Used features</i>	<i>Orientation</i>	<i>X,Y centre points</i>	<i>Axis lengths</i>	<i>Dx, Dy</i>	<i>Top 1</i>	<i>Top 2</i>	<i>Top3</i>
4	✓	✓	✓	✓	40%	40%	60%
2	✓	✓			60%	60%	90%
2	✓		✓		20%	20%	40%
2	✓			✓	50%	60%	70%
2		✓		✓	30%	40%	50%
2		✓	✓		40%	50%	50%
2			✓	✓	30%	30%	60%
3		✓	✓	✓	20%	20%	50%
3	✓		✓	✓	30%	40%	70%
3	✓	✓		✓	60%	70%	90%
3	✓	✓	✓		30%	40%	70%

Πίνακας 6: Σύγκριση συνδυασμών χαρακτηριστικών

4.4 Μεμονωμένα πειράματα με χαρακτηριστικά και αξιολόγηση αποτελεσμάτων

Συγκρίνοντας τις λέξεις με βάση μόνο ένα χαρακτηριστικό μπορούμε να αποκομίσουμε αποτελέσματα τα οποία θα μας οδηγήσουν σε συμπεράσματα σχετικά με το είδος του χαρακτηριστικού που είναι ικανό να αναγνωρίσει καλύτερα έναν συγκεκριμένο τύπο λέξεων. Παρατηρείται ότι το καλύτερο αποτέλεσμα όσον αφορά την απολύτως επιτυχή αναγνώριση εμφανίζεται με σύγκριση της θέσης του κέντρου των χεριών όπου το ποσοστό ακρίβειας ανέρχεται στο 50%. Το συγκεκριμένο χαρακτηριστικό λειτουργεί αρκετά καλά όταν εφαρμόζεται σε συγκεκριμένο τύπο νοημάτων, τις λέξεις δηλαδή που έχουν χειρονομίες σε διακριτές θέσεις σε σχέση με άλλες λέξεις. Αντίθετα όταν 2 λέξεις πραγματοποιούνται στην ίδια περιοχή της εικόνας και το μόνο που αλλάζει είναι κάποιες κινήσεις των δακτύλων τότε η σύγκριση με βάση τη θέση κέντρου είναι επιρρεπής σε λάθη κάτι που παρατηρείται από την αύξηση μόνο κατά 10% (1 λέξη στις 10) όταν αναζητούμε την ορθή αναγνώριση στο σύνολο των 3 επικρατέστερων επιλογών.

Το δεύτερο πιο αποτελεσματικό χαρακτηριστικό είναι αυτό της κλίσης των χεριών σε κάθε frame. Από τα αποτελέσματα είναι εμφανές ότι δεν επιδεικνύει τόσο μεγάλη ακρίβεια σε σχέση με την απολύτως επιτυχή αναγνώριση όσο η σύγκριση με βάση το κέντρο. Ωστόσο είναι σαφές ότι το 80% επιτυχούς αναγνώρισης στο top 3 μόνο ευκαταφρόνητο δεν είναι καθώς δείχνει να λειτουργεί καλύτερα όταν εφαρμόζεται σε ποικιλία λέξεων, γεγονός που γεννά προσδοκίες ώστε ένας συνδυασμός με άλλα χαρακτηριστικά να αποδώσει τα μέγιστα. Επίσης, άξιο αναφοράς είναι το γεγονός ότι τα 2 προαναφερθέντα χαρακτηριστικά αναγνωρίζουν

διαφορετικό σύνολο λέξεων και όχι τις ίδιες. Η ιδιότητα τους αυτή δείχνει ότι αναγνωρίζουν διαφορετικό τύπο λέξεων και είναι απολύτως επιθυμητή για την καλή λειτουργία του συστήματος.

Το τρίτο πιο αποτελεσματικό χαρακτηριστικό είναι το μήκος των αξόνων που προέκυψαν από την ανάλυση κύριων συνιστωσών. Τα αποτελέσματα που προκύπτουν είναι παρόμοια ως ποσοστά με αυτά των υπόλοιπων χαρακτηριστικών. Παρόλα αυτά παρουσιάζει εξαιρετικό ενδιαφέρον το γεγονός πως λέξεις που αναγνωρίζονται σωστά από άλλα χαρακτηριστικά, στην προκειμένη περίπτωση δεν εμφανίζουν καλή συμπεριφορά ενώ αντιθέτως υπάρχουν νοήματα που με την εφαρμογή της σύγκρισης των αξόνων αναγνωρίζονται με την μεγαλύτερη δυνατή επιτυχία. Το παραπάνω εξηγείται με μια βαθύτερη διερεύνηση των ενδογενών χαρακτηριστικών κάθε νοήματος. Για παράδειγμα η λέξη (adopt) που αναγνωρίζεται ως πρώτη επιλογή μόνο από την συγκεκριμένη σύγκριση παρουσιάζει μικρό ποσοστό κίνησης των χεριών τα οποία έπειτα από ένα σημείο μένουν σταθερά για ορισμένα frames ενώ και η κλίση τους δεν παρουσιάζει διακυμάνσεις.

Για τέλος αφέθηκε το χαρακτηριστικό που παράχθηκε στο περιβάλλον του Matlab με σκοπό να αναγνωρίσει λέξεις οι οποίες παρουσιάζουν μεγάλη κινητικότητα. Τα ποσοστά του όταν συγκρίνεται απομονωμένο από τα άλλα χαρακτηριστικά είναι αν μη τι άλλο μη αναμενόμενα καθώς ενώ τα νοήματα εμπεριέχουν αρκετή κίνηση, απολύτως επιτυχώς αναγνωρίζονται μόνο 2 λέξεις. Παράλληλα στις υψηλότερες 3 θέσεις των επιλογών προκύπτει επίσης μόνο 40% επιτυχίας κάτι που αρχικώς οδηγεί στο συμπέρασμα ότι ενδεχομένως το συγκεκριμένο χαρακτηριστικό δεν είναι κατάλληλο για να διαχωρίσει το σύνολο των κλάσεων που διερευνούνται αλλά μπορεί μόνο να παίζει επικουρικό ρόλο στην διαδικασία αναγνώρισης.

4.5 Πειράματα με συνδυασμό χαρακτηριστικών και αξιολόγηση αποτελεσμάτων

Ο συνδυασμός των χαρακτηριστικών όπως έχει αναφερθεί πραγματοποιήθηκε χρησιμοποιώντας την κανονικοποιημένη Ευκλείδεια απόσταση. Από τα 4 χαρακτηριστικά προκύπτει 1 τετράδα, 4 τριάδες και 6 δυάδες χαρακτηριστικών προς πειραματική αξιολόγηση.

Εκκινώντας την διαδικασία αξιολόγησης από τον συνδυασμό όλων των χαρακτηριστικών δεν προκύπτουν ιδιαίτερα ευνοϊκά ποσοστά αναγνώρισης καθώς λιγότερες από τις μισές λέξεις (40%) κατατάσσονται επιτυχώς πρώτες ενώ μικρή είναι και η διαφοροποίηση σχετικά με την ύπαρξη των λέξεων στις 3 πρώτες θέσεις κατάταξης (60%).

Από τις διαθέσιμες δυάδες πριν την εκτέλεση των πειραμάτων αναμένεται ότι τα καλύτερα αποτελέσματα θα προκύπτουν από τον συνδυασμό των χαρακτηριστικών που παρουσίαζαν καλύτερη συμπεριφορά και κατά την απομονωμένη σύγκριση. Η παραπάνω προσδοκία επαληθεύεται πλήρως καθώς ο συνδυασμός της κλίσης με την θέση του κέντρου των χεριών παράγει ποσοστά αναγνώρισης 60% για κατάταξη στην πρώτη θέση και 90% για κατάταξη στις πρώτες 3 θέσεις. Αντίθετα, όπως είναι επίσης αναμενόμενο τα ποσοστά επιτυχούς αναγνώρισης όταν συνδυάζονται η θέση του κέντρου με το παράγωγο χαρακτηριστικό των μετατοπίσεων (dx,dy) είναι φτωχά καθώς μόνο 30% αναγνωρίζεται στην πρώτη θέση και μόνο το 50% των λέξεων καταφέρνει να εισαχθεί στην τριάδα της κατάταξης. Επίσης, σε αντίθεση με ότι προκύπτει από τα παραπάνω, όταν συγκρίνονται ταυτόχρονα η κλίση των χεριών και η μετατόπιση, η αναγνώριση προσεγγίζει την ακρίβεια της καλύτερης δυάδας. Το ποσοστό κατάταξης στην πρώτη θέση είναι 50% και κλιμακώνεται έως 70% για τις τρεις πρώτες θέσεις, ωστόσο παρατηρώντας ποιες λέξεις αναγνωρίζονται σε κάθε περίπτωση και βλέποντας ότι διαφέρουν δημιουργούνται προσδοκίες για τον συνδυασμό ως τριάδα των

χαρακτηριστικών κλίσης, θέσης και μετατόπισης. Επιστρέφοντας στην σύγκριση των δυνάδων, η χειρότερη συμπεριφορά που ταυτόχρονα αποτελεί και το χειρότερο πειραματικό αποτέλεσμα που προέκυψε στην παρούσα εργασία λαμβάνεται έπειτα από συνδυασμό της κλίσης των χεριών και του μήκους των αξόνων. Τα ποσοστά επιτυχίας για την παραπάνω σύγκριση είναι 20% κατάταξης στην πρώτη θέση λίγο πάνω δηλαδή από την τυχαία κατάταξη ενώ στις πρώτες τρεις θέσεις κατατάσσεται το 40% των λέξεων.

Συνεχίζοντας η διαδικασία αξιολόγησης στους συνδυασμούς που περιλαμβάνουν 3 χαρακτηριστικά ερχόμαστε αντιμέτωποι με πολύ ενδιαφέροντα αποτελέσματα καθώς προκύπτει η καλύτερη επίδοση του συστήματος αλλά και η εκ νέου επαλήθευση ποσοστών κοντά στην χειρότερη. Όπως τονίστηκε νωρίτερα, είναι διαισθητικά κατανοητό από τα αποτελέσματα ότι η κλίση και η θέση έχουν σημαίνοντα ρόλο στην απόδοση του συστήματος αλλά και το παράγωγο χαρακτηριστικό της μετατόπισης δεν έχει αμελητέα προσφορά. Έτσι προκύπτει η καλύτερη επίδοση ακρίβειας συνδυάζοντας αυτά τα τρία χαρακτηριστικά, λαμβάνοντας 60% απολύτως επιτυχή κατάταξη, 70% κατάταξη στην δυάδα (από 60% όταν συγκρίνονται μόνο κλίση και θέση) και 90% για κατάταξη στην τριάδα. Διακρίνεται για ακόμη μία φορά ότι η παρουσία των μηκών των αξόνων, επιβαρύνει δυσμενώς τα αποτελέσματα καθώς αν εφαρμοστεί στην προηγούμενη επιτυχή σύγκριση τα ποσοστά πέφτουν σε 40%, 40%, 60% όπως φαίνεται από την σύγκριση του συνόλου των χαρακτηριστικών. Η αποτυχία του συγκεκριμένου χαρακτηριστικού είναι εμφανής και από το γεγονός ότι βρίσκεται στην τριάδα με την χαμηλότερη επίδοση της οποίας τα ποσοστά είναι 20%, 20%, 50%. Μάλιστα η παραπάνω εφαρμογή των μηκών των αξόνων στην δυάδα θέσης και μετατόπισης ρίχνει επιπλέον τα ήδη χαμηλά ποσοστά τους.

5 Συμπέρασμα και επέκταση

5.1 Συμπεράσματα

Από την παρούσα διπλωματική εργασία προκύπτουν ορισμένα πολύ ενδιαφέροντα συμπεράσματα σχετικά με την χρήση ορισμένων χαρακτηριστικών κατά την διαδικασία αναγνώρισης χειρονομίας. Στα πλαίσια λοιπόν της συζήτησης για την συνεισφορά της εργασίας, προκύπτει ότι η απομονωμένη εξέταση των χαρακτηριστικών είναι σε θέση να δώσει μία αίσθηση για την καταλληλότητα του χαρακτηριστικού ακόμα και όταν αυτό συνδυάζεται μετέπειτα με άλλα. Το ιδιαίτερο αυτό εύρημα τεκμηριώνεται αν ανατρέξει κανείς στους πίνακες των συγκρίσεων όπου θα παρατηρήσει ότι τα χαρακτηριστικά των οποίων η μεμονωμένη χρήση συμβάλει στην αναγνώριση με μεγάλη ακρίβεια των λέξεων είναι και αυτά που εν τέλει συνεισφέρουν περισσότερο στο σύστημα με τον συνδυασμό τους.

Εξετάζοντας μεμονωμένα κάθε εξαχθέν χαρακτηριστικό παρατηρείται ότι η κλίση των χεριών παρόλο που οδηγεί σε επιτυχή κατάταξη στην πρώτη θέση μόλις με 30%, το γεγονός όμως ότι εμφανίζει ποσοστό 80% στην κατάταξη στην τριάδα το αναδεικνύει ως το πολυτιμότερο καθώς αναγνωρίζει την μεγαλύτερη ποικιλία κλάσεων. Η παραπάνω ιδιότητα έγκειται από την επιτυχή έως έναν βαθμό αναγνώριση λέξεων στις οποίες όλα τα άλλα χαρακτηριστικά εμφανίζουν κακή συμπεριφορά. Επομένως γίνεται κατανοητό, ότι εκτός από τα ποσοστά απόλυτης αναγνώρισης πρέπει να δίδεται σημασία και στην ποικιλία των λέξεων που αναγνωρίζονται.

Στον άτυπο διαγωνισμό για την αξία κάθε ερευνηθέντος χαρακτηριστικού στην αναγνώριση λέξεων της νοηματικής, η θέση των χεριών έρχεται δεύτερη καθώς μπορεί με μεγάλη ευκολία να διαχωρίσει τις χειρονομίες που γίνονται σε διαφορετικά σημεία της εικόνας. Έτσι εκτός του διαχωρισμού των χειρονομιών που πραγματοποιούνται με ένα χέρι από αυτές που πραγματοποιούνται με δύο χέρια, το χαρακτηριστικό της θέσης συμβάλλει και στην αναγνώριση χειρονομιών που ανήκουν στην ίδια κατηγορία.

Ως συμπληρωματική βοήθεια στα παραπάνω εμφανίζονται οι μετατοπίσεις των χεριών από frame σε frame καθώς προσφέρει αύξηση των ποσοστών της μετρούμενης ακρίβειας στην κατάταξη.

Αντίθετα όπως φαίνεται από τα αποτελέσματα, το μήκος των αξόνων δεν συμβάλλει με θετικό τρόπο σε κανένα σημείο της αναγνώρισης και σε κανέναν από τους πραγματοποιηθέντες συνδυασμούς.

Συμπερασματικά, ο καλύτερος συνδυασμός προκύπτει ότι είναι αυτός ο οποίος εξετάζει χαρακτηριστικά με διαφορετικές ιδιότητες τα οποία αναγνωρίζουν πληθώρα διαφορετικών λέξεων χωρίς να είναι απαραίτητο αυτή η αναγνώριση όταν τα χαρακτηριστικά ερευνούνται χωριστά να είναι απολύτως ακριβής.

Μία αυθαίρετη απόπειρα σύγκρισης των ποσοστών ακρίβειας της διπλωματικής εργασίας με τα ποσοστά που προκύπτουν από την [26] εργασία που χρησιμοποιεί DTW και HoG δείχνει ότι ο καλύτερος συνδυασμός της παρούσας διπλωματικής τουλάχιστον επιβεβαιώνει αν όχι ξεπερνάει τα ποσοστά ακρίβειας εκείνης της προσπάθειας. Η επιφυλακτική αυτή προσέγγιση είναι απόρροια της σύγκρισης εργασιών που πραγματεύονται την ίδια βάση δεδομένων αλλά διαφορετικά χαρακτηριστικά και διαφορετική ποικιλία κλάσεων. Ωστόσο δεν μπορεί να αγνοηθεί το γεγονός ότι η εργασία που χρησιμοποιεί HoG παρουσιάζει ποσοστά επιτυχίας στο top 1 και στο top 3 της τάξης του 57% και 74% ενώ η παρούσα διπλωματική παρουσιάζει 60% και 90% αντίστοιχα.

Επιπρόσθετα, από τα αποτελέσματα της παρούσας διπλωματικής έχει καταστεί σαφές

ότι δεν είναι δυνατόν να δημιουργηθεί ένα φερέγγυο σύστημα αναγνώρισης χωρίς περαιτέρω έρευνα ή εφαρμογή απαιτητικών όρων χρήσεως και περιοριστικών αλλά εξαιρετικά επωφελών για την ακρίβεια προϋποθέσεων. Το παραπάνω συμπέρασμα προκύπτει από το γεγονός ότι πολλά νοήματα περιέχουν επικάλυψη των χεριών μεταξύ τους με αποτέλεσμα την απώλεια σημαντικής πληροφορίας.

5.2 Μελλοντική εργασία και επέκταση

Ο συνδυασμός κλίσης, θέσης, μετατόπισης καλύπτει το κριτήριο που ορίζει ότι χαρακτηριστικά με διαφορετικά στοιχεία εστίασης συμπεριφέρονται καλύτερα στην αναγνώριση οπότε είναι αυτός πάνω στον οποίο ενδεχομένως πρέπει να βασιστεί η μελλοντική επέκταση και δημιουργία του συστήματος. Από την στιγμή που στόχος είναι η όσον το δυνατόν καλύτερη εξυπηρέτηση των ανθρώπων με προβλήματα ομιλίας, απαιτείται το σύστημα να διακρίνεται για την φερεγγυότητα του και για την πλήρη κάλυψη των λέξεων του λεξικού της νοηματικής γλώσσας.

Για να γίνει πραγματικότητα αυτό είναι κάτι παραπάνω από βέβαιο ότι απαιτείται ενσωμάτωση στην λειτουργία και της πλαϊνής όψης ή κάποιο άλλο πλάνο της χειρονομίας στο οποίο να διακρίνεται το βάθος της εικόνας. Αυτό θα βοηθήσει στην πληρέστερη αναπαράσταση των χεριών και εξαγωγή πιο προχωρημένων χαρακτηριστικών όπως θέσεις, κλίση και μετατόπιση δακτύλων. Η πληροφορία αυτή θα οδηγήσει ασφαλώς σε μεγαλύτερη πολυπλοκότητα αλλά είναι βέβαιο ότι θα βελτιώσει κατά πολύ τα αποτελέσματα καθώς μια πληθώρα λέξεων διακρίνεται μόνο από τις διαφορετικές κινήσεις των δακτύλων ή από κινήσεις που πραγματοποιούνται ενώ τα χέρια επικαλύπτονται. Ακόμη, παρατηρώντας τα αποτελέσματα παρόμοιων επιτυχημένων εργασιών είναι επιθυμητό να συνδυαστούν τα χαρακτηριστικά αυτών των εργασιών με αυτά της παρούσας ώστε να εξασφαλιστεί μεγαλύτερη ποικιλία στο σύνολο των αναγνωρισθέντων νοημάτων.

Επιπλέον προσπάθεια είναι επιθυμητό να πραγματοποιηθεί και στο πεδίο του ταξινομητή όπου η εισαγωγή κάποιας εξελιγμένης μεθόδου ενδεχομένως να αποτελέσει την κρίσιμη προσθήκη η οποία θα εκτινάξει την ακρίβεια σε επίπεδα που πραγματικά θα υπάρξει συνεισφορά στην καθημερινή χρήση ενός τέτοιου συστήματος αναγνώρισης.

Μπορεί στην παρούσα διπλωματική να έγινε προσπάθεια για έλεγχο λέξεων με όλο το φάσμα των κινήσεων της αμερικάνικης νοηματικής ωστόσο λόγω περιορισμών στην απόκτηση της βάσης δεδομένων και αποθήκευσης όλου του όγκου των χειρονομιών δεν ήταν δυνατός ένας εκτεταμένος έλεγχος του συνόλου του λεξικού 3000 λέξεων. Επομένως είναι σαφές ότι απαιτείται επέκταση σε αυτόν τον τομέα ώστε τα αποτελέσματα να ανταποκρίνονται ακόμα περισσότερο στην πραγματικότητα.

Οι δυνατότητες συνδυασμών που απορρέουν από την χρήση μόνο 3 ατόμων που κάνουν χειρονομίες είναι περιορισμένες στην απλή χρήση των 2 ως διανύσματα εκπαίδευσης και του άλλου ως διάνυσμα δοκιμής. Στην βάση δεδομένων προσφέρονται έπειτα από συνεννόηση με τους δημιουργούς της επιπλέον signers που δεν αυξάνουν μόνο αριθμητικά τις δυνατότητες αλλά και ποιοτικά καθώς πρόκειται για άτομα που είτε μαθαίνουν την νοηματική είτε είναι η μητρική τους γλώσσα. Η συγκεκριμένη επέκταση μπορεί να έχει ευεργετικά αποτελέσματα για την ακρίβεια του συστήματος αλλά και να διευρύνει τον τομέα έρευνας σε άτομα-μαθητές της νοηματικής γλώσσας.

Για τέλος αφήνεται η περαιτέρω βελτίωση της εργασίας με σκοπό να είναι δυνατόν να χρησιμοποιηθεί για αναγνώριση λέξεων σε πραγματικό χρόνο. Ο στόχος αυτός για να καλυφθεί επιβάλλεται να διεξαχθούν πειράματα σχετικά με τον χρόνο που καταναλώνεται στην δημιουργία αρχείων χαρακτηριστικών, πινάκων απόστασης και σειράς κατάταξης των λέξεων.

Appendix

I. Εξαγωγή εικόνων από lossless compressed video

Η βάση δεδομένων παρέχεται σε μορφή lossless compressed video από το http://vlm1.uta.edu/~athitsos/asl_lexicon. Για την ανάγνωση των video και δημιουργία των εικόνων προς επεξεργασία τροποποιήθηκε λίγο η παρεχόμενη από την βάση βιβλιοθήκη διαβάσματος. Το σχετικό τροποποιημένο αρχείο είναι διαθέσιμο προς χρήση. Για να χρησιμοποιηθεί ακολουθούμε τα παρακάτω βήματα:

1. make στον αντίστοιχο φάκελο vid_reader.
2. ./p1 /address_ston_sklhro_pou_apo8hkeusa_thn_vash/LIZ/scene_24-camera1.vid 525 624.

Με την παραπάνω εντολή αποκόβουμε από την σκηνή 24 τα frames από 525 έως 624.

II. Βιβλιοθήκη OpenCV

Για την εγκατάσταση της opencv βιβλιοθήκης σε περιβάλλον ubuntu ακολουθούμε τις οδηγίες που παρέχονται στον σχετικό σύνδεσμο:

http://docs.opencv.org/2.4/doc/tutorials/introduction/linux_install/linux_install.html

Σημειώνεται ότι ο αν ο φάκελος στον οποίο τρέχουμε την opencv και την εργασία δεν είναι ίδιο, πρέπει να γίνουν αλλαγές στο makefile.

III. Διαδικασία εξαγωγής χαρακτηριστικών

Για να εκτελεστεί ο αλγόριθμος ο οποίος αναγνωρίζει τα χέρια και αποθηκεύει τα χαρακτηριστικά πρέπει:

1. Να πραγματοποιηθεί compile κάνοντας make στον αντίστοιχο φάκελο test_display_image.
2. Να εκτελεστεί ο αλγόριθμος χρησιμοποιώντας ./DisplayImageTutorial down/Dana. Σημειώνεται ότι για κάθε λέξη θα πρέπει να αλλάζει η λέξη εισόδου(down) όπως και ο χρήστης εισόδου (Dana) και να υπάρχει κατάλληλος συμβιβασμός ονοματοδοσίας στην διάρθρωση των φακέλων στον δίσκο αποθήκευσης.
3. Πατώντας το πλήκτρο “V” προχωράει ο αλγόριθμος στο επόμενο frame επεξεργασίας.

4. Πατώντας το πλήκτρο “C” οπισθοχωρεί ο αλγόριθμος στο προηγούμενο frame επεξεργασίας.
5. Για κανονικοποίηση σχετικά με διαφορετικούς χρήστες μπορούν να χρησιμοποιηθούν οι μπάρες περιθωρίου (threshold) πάνω από την εικόνα ώστε να αλλάζει η επεξεργασία της ανταποκρινόμενης μάσκας.

IV. Διαδικασία σε Matlab

Χτίζοντας βήμα βήμα την διαδικασία εκτέλεσης της εργασίας για να λάβουμε τα επιθυμητά αποτελέσματα πρέπει:

1. Να διαβαστούν τα αρχεία των χαρακτηριστικών που έχουμε περάσει στο path του Matlab, φροντίζοντας να τηρούμε τους συμβιβασμούς ονοματοδοσίας δηλαδή τα αρχεία να είναι σε φάκελο με όνομα `feature_files` και κάθε αρχείο να έχει όνομα της μορφής `word_signer.txt` (π.χ `dress_liz.txt`). Αν εξασφαλίζονται οι παραπάνω προϋποθέσεις τότε εκτελώντας το αρχείο **`read_feature_files.m`** διαβάζονται και αποθηκεύονται σε πίνακες τα χαρακτηριστικά.
2. Να εκτελεστεί το αρχείο **`filecreation.m`** ώστε να παραχθεί το παράγωγο χαρακτηριστικό `dx_dy` και να συνενωθεί στους πίνακες που δημιουργήθηκαν στο προηγούμενο βήμα.
3. Να εκτελεστούν τα αρχεία τα οποία δημιουργούν τους πίνακες αποστάσεων για σύγκριση των χαρακτηριστικών μεμονωμένα **`compare_dana_lana.m`** και **`compare_dana_liz.m`**
4. Να εκτελεστούν τα αρχεία τα οποία δημιουργούν τους πίνακες αποστάσεων για σύγκριση συνδυασμών των χαρακτηριστικών **`mahal_compare_dana_lana.m`** και **`mahal_compare_dana_liz.m`**
5. Η σύγκριση όλων των από πάνω γίνεται με κάποιον αλγόριθμο όπως αυτός που δίνεται στο αρχείο **`dipwmatiki_word_recognition_1_feature_2_training_signers.m`** και περιλαμβάνει όλη την διαδικασία.
6. Για εξέταση διαφορετικών λέξεων αρκεί να θέσουμε στον κώδικα της ολοκληρωμένης διαδικασίας την λέξη `word` με την αντίστοιχη τιμή από 1 έως 10 (ή το όνομα της λέξης) όπως αυτές περιγράφονται εντός του κώδικα.
7. Για εξέταση διαφορετικών συνδυασμών ή χαρακτηριστικών αρκεί να δώσουμε ως είσοδο στις αντίστοιχες λειτουργίες ταξινόμησης (`sort`) τον κατάλληλο πίνακα αποστάσεων.
8. Για διαφορετικά πειράματα με τους ίδιους πίνακες αποστάσεων και διαφορετικές λέξεις βάζουμε σε σχόλια τις αντίστοιχες γραμμές που διαβάζουν και παράγουν τα αρχεία για πιο γρήγορη εκτέλεση του κώδικα.

V. Συνδυασμός χαρακτηριστικών

Για παραγωγή πινάκων αποστάσεων με διαφορετικούς συνδυασμούς χαρακτηριστικών πρέπει να γίνουν αλλαγές στα αρχεία **mahal_compare_dana_lana.m** και **mahal_compare_dana_liz.m** όπως αυτές προκύπτουν από τα σχόλια του κώδικα. Έτσι, αν απαιτείται το πρώτο 2-διάστατο χαρακτηριστικό και το τελευταίο 4-διάστατο χαρακτηριστικό για εξέταση τότε αρκεί να παρθούν οι αντίστοιχες στήλες από τους πίνακες χαρακτηριστικών και να συνενωθούν.

Π.χ. Η παρακάτω γραμμή συνενώνει τα orientation (2D - 1 :2), centre_points (4D - 2 :6) και axis (4D - 7:10) για εξέταση.

```
adopt_dana_features=[adopt_dana(1:100,1:6) adopt_dana(1:100,7:10)];
```

Βιβλιογραφία

- [1] S.Nagarajan, T.S.Subashini and M.Balasubramanian, “Visual Interpretation of ASL Finger Spelling using Hough Transform and Support Vector Machine”, in International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 6, June 2015.
- [2] Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Quan Yuan and Ashwin Thangali, “The American Sign Language Lexicon Video Dataset”, in Proceedings of IEEE Workshop on Computer Vision and Pattern Recognition for Human Communicative Behavior Analysis, June 2008.
- [3] Thad Starner, Joshua Weaver, Alex Pentland, “Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video”, in IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 20, no. 12, December 1998
- [4] Muller M., “Information Retrieval for Music and Motion”, in Ch. 4, 2007
- [5] “Introduction to Dynamic Time Warping ”, URL: <http://www.mblondel.org/journal/2009/08/31/dynamic-time-warping-theory>.
- [6] Yi Li, “Hand Gesture Recognition Using Kinect”, A Thesis Submitted To the Faculty of the J. B. Speed School of Engineering of the University of Louisville in Partial Fulfillment of the Requirements for the Degree of Master of Science, 2010
- [7] Valentino Frati, D.P, “Using Kinect for hand tracking and rendering in wearable haptics.”, in IEEE world Haptics Conference, 21-24, (2011).
- [8] Christian Vogler and Dimitris Metaxas, “Parallel Hidden Markov Models for American Sign Language Recognition”, in Proceedings of the International Conference on Computer Vision, September 1999
- [9] Rashmi D. Kyatanavar, P. R. Futane, “Comparative Study of Sign Language Recognition Systems”, in International Journal of Scientific and Research Publications, Volume 2, Issue 6, June 2012
- [10] Panicos Kyriacou, “Sign Language Recognition with the Kinect Sensor Based on Conditional Random Fields”, in Sensors, 2015
- [11] Ken Watanabe, Yoshio Iwai, Yashushi Yagi, Masahiko Yachida, “Recognition of sign language alphabet using colored gloves”, in Systems And Computers in Japan 30(4):51-61, April 1999
- [12] Vaishali S. Kulkarni, S.D.Lokhande, “Appearance Based Recognition of American Sign Language Using Gesture Segmentation”, in International Journal on Computer Science and Engineering, Vol. 02, No. 03, 2010

- [13] Rehg, J., Kanade, T., "DigitEyes: Vision-Based Human Hand Tracking", School of Comp. Science, Carnegie Mellon University, Pittsburgh, 1993
- [14] Mark Dilsizian, Polina Yanovich, Shu Wang, Carol Neidle, Dimitris Metaxas, "A New Framework for Sign Language Recognition based on 3D Handshape Identification and Linguistic Modeling", in LREC Conference, 2014
- [15] Karl Caezar P. Carrera, Alvin Patrick R. Erise, Eliza Marie V. Abrena, Sharmaine Joy S. Colot, Roselito E. Tolentino, "Application of Template Matching Algorithm for Dynamic Gesture Recognition of American Sign Language Finger Spelling and Hand Gesture", in Asia Pacific Journal of Multidisciplinary Research, Volume 2, No. 4, August 2014
- [16] R. E. Kalman, "A new approach to linear filtering and prediction problems.", in Transactions of the ASME-Journal of Basic Engineering, 1960.
- [17] Thad Starner and Alex Pentland, "Real-Time American Sign Language Recognition from Video Using Hidden Markov Models", in International Symposium on Computer Vision, 1995
- [18] G. Awad, "Novel boosting framework for subunit-based sign language recognition", in 16th International conference on image processing, 2009
- [19] H.Cooper and R.Bowden, "Sign Language Recognition Using Boosted Volumetric Features", in IAPR MVA, 2007
- [20] Wei Yi and S.Marshal, "Principal Component Analysis in application to object orientation", in Geo-spatial Information Science, Vol 3, Issue 3, pp 76-78, September, 2000
- [21] S.Theodoridis and K.Koutroumbas, "Pattern recognition", 4th edition, pp 36
- [22] Mika Hatano, Shinji Sako, Tadashi Kitamura, "Contour-based Hand Pose Recognition for Sign Language Recognition", in Special Interest Group on Speech and Language Processing for Assistive Technologies, 2015
- [23] Abidatul Izzah and Nanik Suciati, "Translation of sign language using generic fourier descriptor and nearest neighbour", in International Journal on Cybernetics & Informatics (IJCI) Vol.3, No.1, February, 2014
- [24] Priyanka Mekala, Ying Gao, Jeffrey Fan, Asad Davari, "Real-time Sign Language Recognition based on Neural Network Architecture",in IEEE 43rd Southeastern Symposium on System Theory, 2011
- [25] Sv. Penkov, "Object Orientation, Principal Component Analysis and OpenCV", URL: www.robospace.wordpress.com/2013/10/09/object-orientation-principal-component-analysis-opencv/, 2013
- [26] Pat Jangyodsuk, Christopher Conly, Vassilis Athitsos, "Sign Language Recognition using Dynamic Time Warping and Hand Shape Distance Based on Histogram of

Oriented Gradient Features”, in Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments, 2014

- [27] C. Valli, “The Gallaudet Dictionary of American Sign Language”, Gallaudet U. Press, 2006.