



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

**Αναγνώριση Δυναμικών Χειρονομιών σε Στατικό
Εγωκεντρικό Περιβάλλον με χρήση Αισθητήρα Βάθους**

Προσέγγιση σε Πραγματικό Χρόνο με χρήση ελάχιστων Υπολογιστικών Πόρων

Διπλωματική Εργασία

του

Βασιλείου Α. Λεμονίδη

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2017



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Αναγνώριση Δυναμικών Χειρονομιών σε Στατικό Εγωκεντρικό Περιβάλλον με χρήση Αισθητήρα Βάθους

Προσέγγιση σε Πραγματικό Χρόνο με χρήση ελάχιστων
Υπολογιστικών Πόρων

του

Βασιλείου Α. Λεμονίδη

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή στις 18 Ιουλίου 2017.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Πέτρος Μαραγκός
Καθηγητής
Ε.Μ.Π

.....
Γεράσιμος Ποταμιάνος
Αναπληρωτής
Καθηγητής
Παν/μίου Θεσσαλίας

.....
Κων/νος Τζαφέστας
Επίκουρος Καθηγητής
Ε.Μ.Π

Αθήνα, Ιούλιος 2017

(Υπογραφή)

.....
Βασίλειος Α. Λεμονίδης

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

© 2017– All rights reserved



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής

Copyright ©–All rights reserved Βασίλειος Α. Λεμονίδης, 2017.

Με επιφύλαξη παντός δικαιώματος.

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.

Περίληψη

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η ανάλυση κινούμενης εικόνας, με στόχο την εξαγωγή χρήσιμων πληροφοριών για την ανίχνευση, κατάτμηση και αναγνώριση δυναμικών χειρονομιών. Τα δεδομένα τα οποία χρησιμοποιούνται προέρχονται εξ ολοκλήρου από έναν αισθητήρα βάθους. Στόχος αποτελεί η κατασκευή ενός αποδοτικού αλγορίθμου, ο οποίος επιτρέπει στην ταχεία αναγνώριση των δράσεων του χρήστη, αποσκοπώντας ταυτόχρονα στην ελάχιστη επιβάρυνση του. Προς αυτήν την κατεύθυνση, χρησιμοποιείται για την προεπεξεργασία ένας συνδυασμός γκαουσιανών μοντέλων και ανάλυσης περιγράμματος για την ανίχνευση και κατάτμηση του χεριού, με τη βοήθεια ενός πρότυπου αλγορίθμου ταχείας σκελετοποίησης. Οι περιγραφητές οι οποίοι χρησιμοποιούνται αποτελούν μία μίξη προϋπαρχόντων και πρότυπων ταχέων περιγραφητών, εξειδικευμένων στην εξαγωγή πληροφορίας για τη μεταβολή του σχήματος του υπό εξέταση αντικειμένου στο χρόνο. Το στάδιο της ταξινόμησης χρησιμοποιεί διανύσματα υποστήριξης μηχανής (γνωστά ως SVM) και τυχαία δάση (γνωστά στη διεθνή βιβλιογραφία ως Random Forests), τα οποία συνδυάζονται με τη χρήση ενός πρότυπου πιθανοτικού μοντέλου για να καταστεί δυνατή η αναγνώριση της χειρονομίας. Ο τρόπος με τον οποίο παρουσιάζεται ο προτεινόμενος αλγόριθμος φανερώνει την ανάγκη ανάδειξης της πρακτικότητας του και τη δυνατότητα άμεσης εφαρμογής του.

Λέξεις-Κλειδιά

Όραση Υπολογιστών, Μηχανική Μάθηση, Αναγνώριση Χειρονομιών, Εικόνα Βάθους, Πραγματικού Χρόνου, Αποδοτικός Αλγόριθμος, Εφαρμογή

Abstract

The objective of the current Thesis is the extraction of valuable information from a video stream, relative to the detection, segmentation and recognition of the enrolling dynamic human gestures. The data being used is fully produced with a depth sensor. In order to accomplish this task, we propose an efficient algorithm, which allows fast recognition of a user's actions. To this end, in the preprocessing stage we manage to detect and segment the user's hand, making use of a combination of Gaussian Mixture Models and Contour Analysis, with the addition of an innovative fast skeletonization subsystem. In order to describe efficiently the actions' nature, we deploy a mixture of existing and original descriptors, specialized in extracting information about the shape alteration of the object of interest. In the classification stage we use Support Vector Machines (SVM) and Random Forests, which are joined together, with the deployment of an innovative probability model, in order to perform successful gesture recognition. The way with which the proposed algorithm is being presented exposes its applicability and its potential of immediate implementation.

Keywords

Computer Vision, Machine Learning, Gesture Recognition, Depth Image, Realtime, Efficient Algorithm, Implementation

Ευχαριστίες

Στην παρούσα σελίδα, θα ήθελα να ευχαριστήσω τον καθηγητή Κ. Πέτρο Μαραγκό, για την ευκαιρία που μου έδωσε να ασχοληθώ με ένα εξαιρετικά ανερχόμενο και ενδιαφέρον αντικείμενο, το οποίο αντλεί την ισχύ του από διάφορους ερευνητικούς τομείς, το ενδιαφέρον μου για τους οποίους δημιουργήθηκε κατά την παρακολούθηση των μαθημάτων που διδάσκει στο Εθνικό Μετσόβιο Πολυτεχνείο. Στα μαθήματά του δεν παρέλειπε να παρουσιάζει τις σύγχρονες ερευνητικές τάσεις και να καλλιεργεί στο ακροατήριό του το ζήλο για εκ βάθους ενασχόληση και έρευνα, ενώ είχε πάντοτε τη διάθεση να απαντήσει σε ερωτήσεις. Αυτό που πραγματικά με ώθησε να ασχοληθώ με το κομμάτι της Όρασης Υπολογιστών ήταν ο τρόπος με τον οποίο ο καθηγητής κατάφερε να συνδυάσει αρμονικά πολλά γνωσιακά αντικείμενα, ενοποιώντας τα και προσφέροντας ένα πανεπιστημονικό τρόπο προσέγγισης των θεμάτων που ανέλυε.

Ακόμα, θα ήθελα να ευχαριστήσω τον καθηγητή κ. Αλέξανδρο Ποταμιάνο για παρόμοιους λόγους με τους παραπάνω, ιδίως επειδή μου επέτρεψε να γνωρίσω εις βάθος το αντικείμενο της Μηχανικής Μάθησης και να αντιληφθώ τη σπουδαιότητα αυτού στη διαμόρφωση της σύγχρονης τεχνολογίας και στην αλληλεπίδραση διαφόρων επιστημονικών τομέων. Επιπλέον, πρέπει να αποδώσω τις θερμές μου ευχαριστίες στον επιστημονικό συνεργάτη δρ. Βασίλη Πιτσικάλη, τους βοηθούς του εργαστηρίου CVSP Πέτρο Κούτρα, Ισίδωρο Ροδομαγουλάκη, καθώς επίσης την επιστημονική συνεργάτη δρ. Νάνσυ Ζλατίντση για τον χρόνο που διέθεσαν σε όλα τα κομβικά στάδια της παρούσας εργασίας. Θα ήθελα ακόμα να ευχαριστήσω τον επιστημονικό ερευνητή του ISLAB δρ. Γιώργο Σιόλα για την πολύτιμη συμβολή του στην περάτωση της εργασίας αυτής.

Τέλος, δεν θα μπορούσα να μην ευχαριστήσω τα φιλικά και συγγενικά μου πρόσωπα, που με χαρά με βοήθησαν και συνέτρεξαν στην εκπλήρωση αυτής της εργασίας, περισσότερο από κάθε άλλον. Μέσα από την συζήτηση μαζί τους κατάφερα να αντιμετωπίσω και να λύσω πολλά προβλήματα που προέκυπταν καθ' όλη την πορεία της εργασίας, ενώ μου προσέφεραν το ψυχολογικό κίνητρο και την υποστήριξη που χρειαζόμουν σε αυτό το στάδιο της ζωής μου.

Περιεχόμενα

Ευχαριστίες	x
Κατάλογος Εικόνων	xv
Κατάλογος Πινάκων	xix
Συνοτμεύσεις	xxv
1 Εισαγωγή	1
1.1 Γενικά	1
1.2 Αυτόματη Αναγνώριση Χειρονομιών και η Χρησιμότητά της	1
1.3 Τρέχουσα Έρευνα	2
1.3.1 Γενικά	2
1.3.2 Ανάλυση	2
1.3.2.1 Γενικευμένη Προσέγγιση	3
1.3.2.2 Εξειδικευμένη Προσέγγιση	3
1.3.3 Γενικό Μοντέλο Μεθόδων Αναγνώρισης Χειρονομιών	5
1.3.4 Ανάπτυξη Νέας Μεθοδολογίας	6
1.4 Σχετική Θεματολογία	6
1.4.1 Εξελιξίεις στην Αλληλεπίδραση Ανθρώπου-Μηχανής μέσω κίνησης	7
1.4.1.1 Ιστορική Ανασκόπηση	7
1.4.1.2 Τρέχουσα Ανάπτυξη	8
1.4.2 Πρόσβαση Ευπαθών Ομάδων στη Σύγχρονη Τεχνολογία	8
1.5 Περιγραφή Βασικών Εννοιών	9
1.6 Στόχοι Εργασίας και Επιβαλλόμενοι Περιορισμοί	9
1.6.1 Επιθυμητές Ιδιότητες Προτεινόμενου Συστήματος από την Άποψη της Αλληλεπίδρασης Ανθρώπου-Μηχανής	9
1.6.2 Επιβαλλόμενοι Περιορισμοί	11
1.7 Χρησιμοποιούμενο Υλισμικό	12
1.7.1 Kinect	12
1.7.2 Άλλο υλικό	14
1.7.3 Χρησιμοποιούμενο Λογισμικό	14
2 Θεωρητικό Υπόβαθρο	17
2.1 Όραση Υπολογιστών	17
2.1.1 Γενικά	17
2.1.2 Ανάλυση Συνδεδεμένων Συνιστωσών Connected Components Analysis	18

2.1.2.1	Ανίχνευση Συνδεδεμένων Συνιστωσών	18
2.1.2.2	Ανίχνευση Περιγράμματος	19
2.1.2.3	Ανάλυση Χαρακτηριστικών Περιγράμματος	19
2.1.3	Οπτική Ροή	20
2.1.4	Περιγραφητές	22
2.1.4.1	Histogram of Oriented Optical Flow (HOOF)	22
2.1.4.2	Histogram of Oriented Gradient(HOG)	24
2.2	Μηχανική Μάθηση	24
2.2.1	Support Vector Machine (SVM)	24
2.2.1.1	Μαθηματική Περιγραφή Γραμμικού SVM	24
2.2.1.2	Επίλυση προβλήματος ελαχιστοποίησης	26
2.2.2	Random Decision Forest (RDF)	27
2.2.2.1	Decision Trees (DT)	27
2.2.2.1.1	Hunt Algorithm	28
2.2.2.1.2	Gini Impurity	29
2.2.2.2	Bootstrapping Aggregating (Bagging)	29
2.2.3	Sparse Coding	30
2.2.3.0.1	Επίλυση του Προβλήματος 2.10a	31
2.2.3.0.2	Επίλυση του Προβλήματος 2.10b	31
2.3	Μικτά Θέματα	32
2.3.1	Ανίχνευση και Οριοθέτηση Κινούμενου Αντικειμένου σε Στατικό Υπό- βαθρο	32
2.3.1.1	Gaussian Mixture Model (GMM)	33
2.3.1.1.1	Εκπαίδευση	34
2.3.1.2	Ανίχνευση αντικειμένου	35
2.3.1.3	Ανίχνευση Σκιών	35
3	Περιγραφή Μεθοδολογίας	37
3.1	Γενικά	37
3.2	Προεπεξεργασία	38
3.2.1	Εικόνα βάθους (Depth Image)	38
3.2.2	Αρχική Προεπεξεργασία (Raw Preprocessing)	38
3.2.3	Ανίχνευση Σκιών (MOG2 Shadows Retrieval)	39
3.2.4	Υπολογισμός Περιγράμματος (Contours Calculation/Filtering)	39
3.2.5	Ταχεία Σκελετοποίηση (Fast Skeletonization)	40
3.2.6	Προσέγγιση Περιοχής Ενδιαφέροντος του Χεριού (Hand ROI Approxi- mation)	41
3.3	Εξαγωγή Χαρακτηριστικών	42
3.3.1	Περιγραφητής GHOG	43
3.3.2	Περιγραφητής 3DHOF	43
3.3.3	Περιγραφητής ZHOF	44
3.3.4	Περιγραφητής 3DXYPKA	46
3.3.5	Τεχνική Sparse Coding	47
3.4	Κατασκευή Buffer	49
3.5	Post Time PCA	49
3.6	Ταξινόμηση	49
3.6.1	Γενικά	49

3.6.2	Επεξεργασία των Scores	50
3.6.2.1	Μέθοδος CProb	50
3.6.2.2	Μέθοδος CSTD	51
3.6.3	Combined Dynamic Bayesian Mixture Model (CDBIMM)	52
3.6.3.1	Εκπαίδευση (Training)	52
3.6.3.2	Δοκιμή (Testing)	53
3.6.3.2.1	Κανονικοποίηση των SVMs scores	55
3.6.3.3	Διαισθητική Επεξήγηση Χρήσης του CDBIMM	55
4	Πειραματική Διαδικασία	57
4.1	Δεδομένα	57
4.1.1	Γενική Περιγραφή	57
4.1.2	Δεδομένα Training	58
4.1.3	Δεδομένα Validation	61
4.1.4	Δεδομένα Testing	65
4.2	Πειραματικό Πρωτόκολλο	66
4.2.1	Λήψη και Αρχική Προεπεξεργασία Δεδομένων	66
4.2.2	Εξαγωγή Μάσκας Χεριού	66
4.2.3	Επεξεργασία Χαρακτηριστικών	67
4.2.4	Ταξινόμηση και Επεξεργασία των Scores	67
4.3	Τρόπος Χρήσης της Επισημάνσης των Δεδομένων (Ground Truth)	68
4.4	Περιγραφή Χρησιμοποιούμενων Μετρικών	68
4.4.1	Είδη Μετρικών	68
4.4.2	Χώροι Υπολογισμού Μετρικών	69
4.5	Εκτέλεση Πειραμάτων Ταξινόμησης	69
4.6	Πειραματισμός πάνω στην Ταξινόμηση Δυναμικών Χειρονομιών χωρίς το μοντέλο CDBIMM	70
4.6.1	Αποτελέσματα Δοκιμών	70
4.7	Πειραματισμός πάνω στην Ταξινόμηση Παθητικών Χειρονομιών	74
4.7.1	Αποτελέσματα Δοκιμών	75
5	Χρήση Αποτελεσμάτων Πειραματικής Διαδικασίας - Επέκταση	79
5.1	Γενικά	79
5.2	Εφαρμογή Μοντέλου CDBIMM	80
5.3	Διαχείριση Εξόδου Ταξινόμησης	83
5.3.1	Σύγχρονη Ταξινόμηση (In Sync)	83
5.3.2	Συνδυασμένη Ταξινόμηση (Combined)	86
6	Συμπεράσματα και Παρατηρήσεις	89
6.1	Γενικά	89
6.2	Προεπεξεργασία/Εξαγωγή Χαρακτηριστικών	89
6.2.1	Παρατηρήσεις	89
6.2.2	Συμπεράσματα	90
6.3	Ταξινόμηση	90
6.3.1	Παρατηρήσεις	90
6.3.2	Συμπεράσματα	91
6.4	Συνολικά	92

7	Μελλοντικές Επεκτάσεις- Περιγραφή Πρακτικών Υλοποιήσεων	95
7.1	Γενικά	95
7.2	Εφαρμογή Επεξεργασίας Δεδομένων	95
7.2.1	Παρουσίαση και Περιληπτική Περιγραφή	95
7.3	Εφαρμογή Σχεδίασης	97
7.3.1	Σύντομη Περιγραφή	97
7.4	Πραγματικού Χρόνου Αναγνώριση	97
7.5	Σύντομη Περιγραφή Protractor	98
7.6	Περιγραφή Εφαρμογής Σχεδίασης	99
7.7	Παρουσίαση Διαγράμματος Ροής Εφαρμογής	100
	Παραρτήματα	103
A	Αλγόριθμοι	105
A.1	Αλγόριθμος Ανίχνευσης και Χαρακτηρισμού Συνδεδεμένων Στοιχείων	106
A.2	Αλγόριθμος Feature Sign Search	107
B	Ανάλυση Μαθηματικών Θεμάτων	109
B.1	Θεώρημα του Green	110
B.2	Θεώρημα Taylor	110
B.3	Σταθμισμένη Μέθοδος Διακριτών Ελαχίστων Τετραγώνων	111
B.3.1	Πρόβλημα Ελαχίστων Τετραγώνων	111
B.3.2	Επίλυση	112
B.3.3	Προσθήκη Βαρών	112
B.4	Αφινικός (ή Ομοπαράλληλος) Μετασχηματισμός	113
C	Αποτελέσματα στα Δεδομένα Validation	115
C.1	Αποτελέσματα Ταξινόμησης Δυναμικών Χειρονομιών	116
C.1.1	Αποτελέσματα Validation0	116
C.1.2	Αποτελέσματα Validation1	118
C.1.3	Αποτελέσματα Validation2	121
C.2	Αποτελέσματα Ταξινόμησης Παθητικών Χειρονομιών	123
C.2.1	Αποτελέσματα Validation0	123
C.2.2	Αποτελέσματα Validation1	125
C.2.3	Αποτελέσματα Validation2	127
	Βιβλιογραφία	129

Κατάλογος σχημάτων

1.1	Είδη Χειρονομιών, όπως περιγράφονται από τους Escallera et al. [10]. Ο διαχωρισμός καθορίζεται από τη χρηστικότητα και τη χρονική μεταβλητότητα της χειρονομίας, καθώς επίσης και τη συσχέτιση αυτής με το υπόλοιπο ανθρώπινο σώμα.	5
1.2	Γενικευμένο Μοντέλο Μεθόδων Αναγνώρισης Χειρονομιών	5
1.3	Προτεινόμενη Τοποθέτηση Αισθητήρα [45]	12
1.4	Περιγραφή Αισθητήρων της συσκευής Kinect [44]	13
3.1	Στάδιο Προεπεξεργασίας (Preprocessing)	38
3.2	‘Derotated’ Οπτική Ροή για την εξαγωγή των χαρακτηριστικών από τον 3DHOF Περιγραφητή. Το κάθε εικονοπλαίσιο έχει το ίδιο μέγεθος με το αρχικό frame βάθους, με μηδενισμένα τα εικονοστοιχεία. Τα χαρακτηριστικά της κάμερας είναι το Σημείο Εστίασης (Focal Point) και το Πρωτεύον Κύριο Σημείο (Principal Point) και χρησιμοποιούνται για την εφαρμογή του μοντέλου Pinhole Camera. Το υποσύστημα ‘Get mesh points’ δεν παράγει τίποτα άλλο από δύο πίνακες με τις οριζόντιες και κάθετες συντεταγμένες των εικονοστοιχείων των frames. Το υποσύστημα ‘Farneback Dense Optical Flow’ δίνει έξοδο ίδιας μορφής. Επειδή τα εικονοπλάγια έχουν το ίδιο μέγεθος, αυτοί οι πίνακες δεν αλλάζουν ανά frame. Το υποσύστημα ‘map’ αντιστοιχίζει τους πίνακες συντεταγμένων με τις τιμές των εικονοπλαισίων, παράγοντας έτσι έναν διδιάστατο πίνακα.	44
3.3	‘Derotated’ Οπτική Ροή για την εξαγωγή των χαρακτηριστικών από τον ZHOF Περιγραφητή.	45
3.4	Εξαγωγή 3DXYPKA Χαρακτηριστικών	46
4.1	Διάγραμμα Απεικόνισης των διαφόρων τύπων χειρονομιών, με αντιστοίχιση εμφανίσεων αυτών στο πεδίο των frames. Μπορεί να παρατηρήσει κανείς την έντονη χρονική μεταβλητότητα στα frames που αφορούν μία εμφάνιση δυναμικής χειρονομίας και τη στατικότητα στα frames που αφορούν μια εμφάνιση παθητικής χειρονομίας. Με κόκκινο νόυμερο εμφανίζονται οι σχετικοί δείκτες του frame κάθε χειρονομίας, με αρχή το σημείο έναρξής της. Για λόγους ευκρίνειας, φαίνονται λιγότερα frames για εμφανίσεις με πάνω από 40 frames.	59
4.2	Αριθμός Εμφανίσεων των Χειρονομιών ανά dataset εκπαίδευσης. Οι παθητικές χειρονομίες έχουν μία εμφάνιση στα αντίστοιχα dataset.	60
4.3	Δεδομένα Validation0, όπως έχουν επισημανθεί στο βάθος του χρόνου (που μετράται σε frames). Δεν φαίνεται η πληροφορία ότι μετά το 1500ο frame τα χέρια εναλλάσσονται. Μπορεί να παρατηρήσει κανείς ότι δεν υπάρχουν πολλές εμφανίσεις δυναμικών χειρονομιών.	62
4.4	Δεδομένα Validation1	63
4.5	Δεδομένα Validation2	64
4.6	Αριθμός Εμφανίσεων Χειρονομιών ανά validation dataset.	64

4.7	Δεδομένα Δοκιμής	65
4.8	Αριθμός Εμφανίσεων Χειρονομιών στο test dataset.	66
4.9	Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών , ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.	73
4.10	Δεδομένα Test: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances)των δυναμικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών	74
4.11	Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών , ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.	77
4.12	Δεδομένα Test: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances)των παθητικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών	77
5.1	Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον CDBIMM Ταξινομητή (των δυναμικών χειρονομιών), ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων.	81
5.2	Σύγχρονη Ταξινόμηση (In Sync). Με ελλείψεις φαίνονται η είσοδος και η έξοδος του συστήματος.	84
5.3	Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τους συγχρονισμένους Ταξινομητές Cl_{pas} και CDBIMM, ως προς τα frames.Οι δύο συνεχείς γραμμές δείχνουν την προβλεφθείσα κλάση ανά frame από κάθε ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων.	85
5.4	Συνδυασμένη Ταξινόμηση (Combined). Με ελλείψεις φαίνονται η είσοδος και η έξοδος του συστήματος.	86
5.5	Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον CRDF Ταξινομητή, ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων.	88
6.1	Συνολικός Προτεινόμενος Αλγόριθμος. Με ελλείψεις φαίνονται η είσοδος και οι εξοδοί του συστήματος.	93
7.1	Στιγμιότυπο πρότυπης εφαρμογής εξαγωγής δεδομένων από αρχεία rosbag, για annotation χειρονομιών.	96

7.2	Διάγραμμα Ροής Εφαρμογής. Το μωβ χρώμα υποδηλώνει τα διαφορετικά νήματα, το κίτρινο επιλογείς, το πράσινο καταχωρητές, το ροζ πράξεις χρονόμετρησης και το πορτοκαλί γενικές δράσεις. Οι ρόμβοι υποδηλώνουν σύγκριση. Τα βέλη με τις διακεκομμένες γραμμές υποδηλώνουν ως επί το πλείστον φόρτωση δεδομένων από καταχωρητή ή μεταφορά αποτελέσματος από ένα νήμα σε ένα άλλο.	100
7.3	Ανίχνευση ζωγραφισμένου Τριγώνου από τον Protractor	101
7.4	Ανίχνευση ζωγραφισμένου Τετραγώνου από τον Protractor	101
7.5	Η φράση 'Hi!' σχεδιασμένη στο Free Mode. Μπορεί να παρατηρήσει κανείς ότι είναι δυνατή η εναλλαγή των χειρονομιών, για να μπορέσει ο χρήστης να σχεδιάσει μη συνεχόμενες καμπύλες. Ακόμα, τα λανθασμένα σχεδιασμένα σημεία οφείλονται στο γεγονός ότι η χειρονομία έχει μια μικρή καθυστέρηση στην ανίχνευση, αφού έχει μειωθεί ο συντελεστής ευαισθησίας της αναγνώρισης.	101
C.1	Δεδομένα Validation0: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών , ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.	117
C.2	Δεδομένα Validation0: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances)των δυναμικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών	118
C.3	Δεδομένα Validation1: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών , ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.	120
C.4	Δεδομένα Validation1: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances)των δυναμικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών	120
C.5	Δεδομένα Validation2: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών , ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.	122
C.6	Δεδομένα Validation2: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances)των δυναμικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών	122

- C.7 Δεδομένα Validation0: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών , ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές. 124
- C.8 Δεδομένα Validation0: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances)των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών 124
- C.9 Δεδομένα Validation1: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών , ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές. 126
- C.10 Δεδομένα Validation1: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances)των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών 126
- C.11 Δεδομένα Validation2: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών , ως προς τα frames.Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές. 128
- C.12 Δεδομένα Validation2: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances)των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών 128

Κατάλογος πινάκων

3.1	Παράδειγμα Πίνακα Συνεκτικότητας Χειρονομιών	53
4.1	Λεξικό Δυναμικών Χειρονομιών Δεδομένων Εκπαίδευσης. Φαίνονται 10 frames από μία εμφάνιση (utterance) κάθε χειρονομίας.	60
4.2	Λεξικό Παθητικών Χειρονομιών Δεδομένων Εκπαίδευσης. Φαίνεται ότι η μορφή της χειρονομίας παραμένει σχεδόν σταθερή και μεταβάλλεται σε περιορισμένο εύρος ο προσανατολισμός της ως προς την κάμερα.	61
4.3	Ακρίβειες των 3 καλύτερων δοκιμασμένων διατάξεων στα δεδομένα Validation για την ταξινόμηση των δυναμικών χειρονομιών, σε επίπεδο frames και εμφανίσεων. Οι στήλες <i>PTPCA</i> και <i>Sparsecoding Level</i> αναφέρονται για πληρότητα, αν και όπως φαίνεται δεν χρησιμοποιούνται από κανένα από τα καλύτερα συστήματα.	71
4.4	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των δυναμικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πως έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	71
4.5	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πως έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	72
4.6	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των δυναμικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών ως προς τα frames.	72
4.7	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των δυναμικών χειρονομιών ως προς τα frames.	72
4.8	Ακρίβειες των 3 καλύτερων δοκιμασμένων διατάξεων στα δεδομένα Validation για την ταξινόμηση των παθητικών χειρονομιών, σε επίπεδο frames και εμφανίσεων. Η στήλη <i>Sparsecoding Level</i> αναφέρεται για πληρότητα, αν και όπως φαίνεται δεν χρησιμοποιείται από κανένα από τα καλύτερα συστήματα.	75

4.9	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των παθητικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	76
4.10	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	76
4.11	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των παθητικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τα frames.	76
4.12	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τα frames.	76
5.1	Οι καλύτεροι επιλεγμένοι ταξινομητές, ο ταξινομητής δυναμικών χειρονομιών CI_{pas} και ο ταξινομητής δυναμικών χειρονομιών CI_{dyn} βάσει του πειραματισμού που έλαβε χώρα. Με μαύρο αναγράφεται το κελί του PTPCA, αφού δεν εφαρμόζεται στην περίπτωση του πειραματισμού του ταξινομητή CI_{pas} .	79
5.2	Πίνακας Συνεκτικότητας (CM) Χειρονομιών, προερχόμενος από τις δοκιμές του Επιλεγμένου Ταξινομητή Παθητικών Χειρονομιών στα Δεδομένα Εκπαίδευσης των Δυναμικών Χειρονομιών.	80
5.3	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των δυναμικών χειρονομιών από τον CDBIMM Ταξινομητή (των δυναμικών χειρονομιών) ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	81
5.4	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον CDBIMM Ταξινομητή (των δυναμικών χειρονομιών) ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	82
5.5	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των δυναμικών χειρονομιών από τον CDBIMM Ταξινομητή (των δυναμικών χειρονομιών) ως προς τα frames.	82
5.6	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον CDBIMM Ταξινομητή (των δυναμικών χειρονομιών) ως προς τα frames.	82
5.7	Σύγκριση Επιδόσεων των Ταξινομητών CI_{dyn} και CDBIMM πάνω στα δεδομένα Test, ως προς τις εμφανίσεις των χειρονομιών.	82
5.8	Σύγκριση Επιδόσεων των Ταξινομητών CI_{dyn} και CDBIMM πάνω στα δεδομένα Test, ως προς τα frames.	82

5.9	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση όλων των χειρονομιών από τους συγχρονισμένους Ταξινομητές CI_{pas} και CD-BIMM ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από κάποιον από τους ταξινομητές ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. Λόγω αυτού του ορισμού, μια εμφάνιση μπορεί να αποκτήσει δύο προβλεφθείσες τιμές από τους δύο συγχρονισμένους ταξινομητές. Τότε, αν κάποια από τις τιμές αυτές αντιστοιχεί στην πραγματική επισήμανση της εμφάνισης, η εμφάνιση θεωρείται πως έχει ανιχνευθεί σωστά.	85
5.10	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τους συγχρονισμένους Ταξινομητές CI_{pas} και CDBIMM ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από κάποιον από τους ταξινομητές ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. Λόγω αυτού του ορισμού, μια εμφάνιση μπορεί να αποκτήσει δύο προβλεφθείσες τιμές από τους δύο συγχρονισμένους ταξινομητές. Τότε, αν κάποια από τις τιμές αυτές αντιστοιχεί στην πραγματική επισήμανση της εμφάνισης, η εμφάνιση θεωρείται πως έχει ανιχνευθεί σωστά.	85
5.11	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση όλων των χειρονομιών από τους συγχρονισμένους Ταξινομητές CI_{pas} και CD-BIMM ως προς τα frames.	86
5.12	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τους συγχρονισμένους Ταξινομητές CI_{pas} και CDBIMM ως προς τα frames.	86
5.13	Δεδομένα Test: Ακρίβειες Συγχρονισμένων Ταξινομητών όλων των χειρονομιών.	86
5.14	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση όλων των χειρονομιών από τον CRDF Ταξινομητή ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	87
5.15	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον CRDF Ταξινομητή ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	87
5.16	Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση όλων των χειρονομιών από τον CRDF Ταξινομητή ως προς τα frames.	87
5.17	Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον CRDF Ταξινομητή ως προς τα frames.	88
7.1	Αντιστοίχιση Χειρονομιών με Λειτουργίες της Εφαρμογής Σχεδίασης	99

- C.1 Δεδομένα Validation0: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 116
- C.2 Δεδομένα Validation0: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 116
- C.3 Δεδομένα Validation0: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames. 116
- C.4 Δεδομένα Validation0: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames. 117
- C.5 Δεδομένα Validation1: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 118
- C.6 Δεδομένα Validation1: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 119
- C.7 Δεδομένα Validation1: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames. 119
- C.8 Δεδομένα Validation1: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames. 119
- C.9 Δεδομένα Validation2: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 121

- C.10 Δεδομένα Validation2: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 121
- C.11 Δεδομένα Validation2: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames. 121
- C.12 Δεδομένα Validation2: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames. 121
- C.13 Δεδομένα Validation0: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 123
- C.14 Δεδομένα Validation0: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 123
- C.15 Δεδομένα Validation0: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames. 123
- C.16 Δεδομένα Validation0: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames. 123
- C.17 Δεδομένα Validation1: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 125
- C.18 Δεδομένα Validation1: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. 125

C.19	Δεδομένα Validation1: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των παθητικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τα frames.	125
C.20	Δεδομένα Validation1: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τα frames.	125
C.21	Δεδομένα Validation2: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των παθητικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	127
C.22	Δεδομένα Validation2: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.	127
C.23	Δεδομένα Validation2: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των παθητικών χειρονομιών από τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τα frames.	127
C.24	Δεδομένα Validation2: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των παθητικών χειρονομιών ως προς τα frames.	127

Συντομεύσεις

RGBD	Red/Green/Blue/Depth Channels Κανάλια Κόκκινου/Πράσινου/Μπλε/Βάθους
FOV	Field Of View Πεδίο Θέασης
ROI	Region Of Interest Περιοχή Ενδιαφέροντος
PCA	Principal Components Analysis Ανάλυση Πρωτευουσών Συνιστωσών
GHOg	Global Histogram of Oriented Gradients Παγκόσμιο Ιστόγραμμα Κατευθυνόμενων Βαθμίδων
3D	Three Dimensional Τρισδιάστατος
3DHOF	3D Histogram of Optical Flow 3D Ιστόγραμμα Οπτικής Ροής
BG	BackGround Υπόβαθρο
FG	ForeGround Προσκήνιο
GMM	Gaussian Mixture Model Μοντέλο Γκαουσιανής Μίξης

Κεφάλαιο 1

Εισαγωγή

1.1 Γενικά

Με την εξέλιξη της τεχνολογίας, η ανάπτυξη εφαρμογών, οι οποίες προσδίδουν τη δυνατότητα σε ένα άψυχο αντικείμενο, μια υπολογιστική μηχανή, να αναγνωρίζει το περιβάλλον όχι μόνο σαν ένας αισθητήρας αλλά σαν ένα έμβιο ον με την κατάλληλη εκπαίδευση, βρίσκεται στο επίκεντρο του επιστημονικού ενδιαφέροντος τα τελευταία χρόνια. Μέθοδοι και δομές, οι οποίες έβρισκαν παλιότερα μόνο θεωρητικό πεδίο εφαρμογής, αποτελούν τα βασικά εργαλεία αυτών των εφαρμογών, ενώ η τυποποίηση τους σε υπολογιστικές πλατφόρμες επιτρέπει σε οποιονδήποτε να τις χρησιμοποιήσει με τις ελάχιστες απαιτήσεις.

Συγκεκριμένα, η επιστήμη της Όρασης Υπολογιστών, η οποία γνώρισε ιδιαίτερη άνθιση τις δεκαετίες πριν το 2000, αναδεικνύεται ξανά πλέον σε πρακτικό επίπεδο, μέσω της συνεισφοράς της στον τομέα της Μηχανικής Μάθησης και της Αναγνώρισης Προτύπων. Η κύρια επιδίωξη αυτής της ιδιαίτερα υποσχόμενης συνεργασίας είναι η δημιουργία έξυπνων μηχανών, οι οποίες θα αντιλαμβάνονται το περιβάλλον με υψηλό επίπεδο νοημοσύνης, προσφέροντας μια πληθώρα εφαρμογών στην καθημερινότητα του ανθρώπου.

1.2 Αυτόματη Αναγνώριση Χειρονομιών και η Χρησιμότητά της

Η παρούσα εργασία είναι μία προσπάθεια παραγωγής ενός μοντέλου αυτόματης Αναγνώρισης Χειρονομιών (γνωστή στη βιβλιογραφία ως *gesture recognition*) και βρίσκει εφαρμογή κυρίως

στον τομέα της Αλληλεπίδρασης Ανθρώπου-Μηχανής (Human-Computer Interaction). Οι χειρονομίες αποτελούν κατά βάση εργαλείο ανθρώπινης επικοινωνίας. Προσδίδουν φυσικότητα και αμεσότητα στο λόγο και διευκολύνουν συγχρόνως την κατανόηση [20]. Η αναγωγή της επικοινωνίας αυτής στο επίπεδο της σχέσης ανθρώπου και υπολογιστή επιφέρει τα προαναφερθέντα προτερήματα. Παράλληλα με αυτά, επιτρέπει:

- την ανάπτυξη εφαρμογών για ανθρώπους με χαμηλά επίπεδα όρασης.
- τον έλεγχο συσκευών, χωρίς χρήση συμβατικών μεθόδων εισαγωγής.
- την πλοήγηση σε περιβάλλοντα εικονικής πραγματικότητας.
- τη δημιουργία εφαρμογών που προϋποθέτουν έλεγχο σε τρισδιάστατο επίπεδο.

1.3 Τρέχουσα Έρευνα

1.3.1 Γενικά

Όντας ένα ανοικτό ζήτημα, το οποίο δεν επιδέχεται, λόγω αυξημένης διαστατικότητας, μοναδική ντετερμινιστική λύση, το πρόβλημα της Αναγνώρισης Χειρονομιών μπορεί να προσεγγιστεί με διάφορους τρόπους. Το υπάρχον εύρος των διαθέσιμων εφαρμογών έχει ωθήσει πολλές ερευνητικές ομάδες να ασχοληθούν με το πρόβλημα αυτό με ανταγωνιστικό πνεύμα χωρίς να υπάρχει κάποια κοινή γραμμή, αφού η αντιμετώπιση γίνεται σε προσεγγιστικό επίπεδο, ενώ τίθεται και θέμα υποκειμενικότητας της μέτρησης της ακρίβειας των αντιστοιχών εργασιών. Αποτέλεσμα αυτής της έρευνας είναι η ανάπτυξη ποικιλίας αλγορίθμων, οι οποίοι προσεγγίζουν το πρόβλημα με διάφορες οπτικές γωνίες. Η αποτελεσματικότητα αυτών είναι συχνά αντιστρόφως ανάλογη των βαθμών ελευθερίας κίνησης του χρήστη και της ταχύτητας εκτέλεσης. Παράλληλα, επιβάλλονται περιορισμοί στο περιβάλλον, στο οποίο βρίσκεται ο χρήστης, καθώς επίσης και στον αισθητήρα που χρησιμοποιείται για την αναγνώριση.

1.3.2 Ανάλυση

Όπως προαναφέρθηκε, η σκοπιά με την οποία προσεγγίζεται το πρόβλημα της Αναγνώρισης Χειρονομιών δεν είναι αυστηρώς ορισμένη. Υπάρχουν δύο κυρίαρχες σχολές, αυτή που προσπαθεί να επιλύσει το ζήτημα μέσω της γενικευμένης ανάλυσης του χεριού και αυτή που επιδιώκει

να μειώσει το υπολογιστικό κόστος, εκπαιδεύοντας μοντέλα να αναγνωρίζουν διακριτό αριθμό χειρονομιών.

1.3.2.1 Γενικευμένη Προσέγγιση

Η χειρονομία μπορεί να περιγραφεί ως ένα σύνολο παραμέτρων, που περιγράφουν την κινηματική κατάσταση του χεριού, δηλαδή τις γωνίες και τις θέσεις των αρθρώσεων, στο βάθος του χρόνου. Υπό αυτήν την σκοπιά, υπάρχει ενεργό ερευνητικό ενδιαφέρον, στην εφαρμογή της μεθόδου model fitting, για να εξαχθεί πλήρως το μοντέλο του χεριού και να προσδιορισθεί η χειρονομία με απόλυτη ακρίβεια. Τέτοιες state-of-the-art εργασίες παρουσιάζονται παρακάτω:

Οι Shridar et. al [41] χρησιμοποιούν Gaussian Mixture μοντέλα (GMM) σε συνδυασμό με Random Forests (RF), συνδυάζοντας μη εποπτευόμενους (unsupervised) και εποπτευόμενους (supervised) αλγορίθμους μάθησης.

Οι Tan et. al [17] χρησιμοποιούν επαναληπτικές μεθόδους ελαχιστοποίησης συναρτησιακών, για να ταιριάζουν τις παραμέτρους μιας προκατασκευασμένης παραμορφώσιμης επιφάνειας, στις παραμέτρους της επιφάνειας του χεριού.

Οι Oberweger et al. [34] κάνουν χρήση συνελκτικού νευρωνικού δικτύου (CNN). Αυτό αποτελείται από νευρώνες με συνάρτηση ενεργοποίησης ReLU και δρα στην είσοδο πολυκλιμακωτά, ενώ ένα στρώμα (layer) έχει λιγότερους νευρώνες από αυτούς που χρειάζονται έτσι, ώστε να γίνουν γνωστοί οι φυσικοί περιορισμοί των αρθρώσεων και να εξαχθεί μια πρότερη εκτίμηση της πόζας του χεριού.

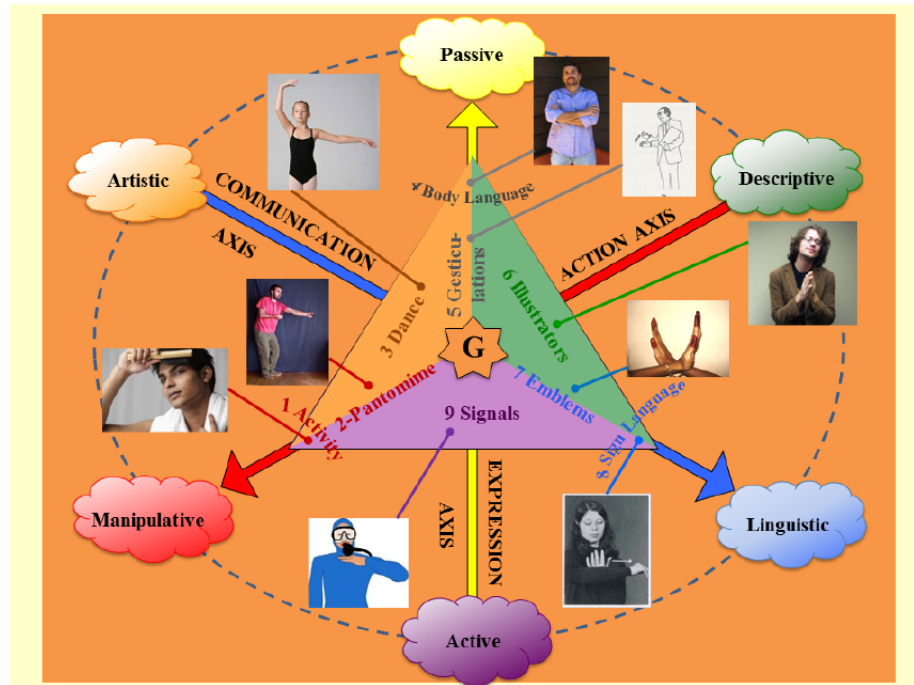
Οι Α.Μακρής et. al. [26] χρησιμοποιούν particle filtering. Βασιζόμενοι σε ένα προκατασκευασμένο μοντέλο του χεριού, κάνουν χρήση ενός δυναμικού Bayesian δικτύου, για να εντοπίσουν τοπικά μέγιστα της a posteriori πιθανότητας να βρίσκεται το χέρι σε κάποια συγκεκριμένη κατάσταση.

1.3.2.2 Εξειδικευμένη Προσέγγιση

Η χειρονομία μπορεί να θεωρηθεί ότι περιγράφεται πλήρως από ένα σύνολο προβολών, οι οποίες προέρχονται από την έξοδο ενός αισθητήρα, από συγκεκριμένη γωνία λήψης. Ερευνητές που προσεγγίζουν τοιούτω τρόπο το πρόβλημα είναι οι:

- Molchanov et al. [28]. Χρησιμοποιώντας ένα τρισδιάστατο συνελκτικό δίκτυο (3D-CNN), επιτυγχάνουν state-of-the-art αποτελέσματα στην αναγνώριση δυναμικά μεταβαλλόμενων χειρονομιών, χωρίς να περιέχεται το στάδιο της λήψης και προεπεξεργασίας των εισαγόμενων εικονοσειρών και με χρήση ισχυρής GPU.
- Guler et al. [13]. Χρησιμοποιούν πυκνές τροχιές (dense trajectories), με τις οποίες κατασκευάζεται, μέσω τυχαίου clustering, μια υπερπλήρης περιγραφή της κίνησης. Αυτή στη συνέχεια χρησιμεύει ως ένα λεξικό οπτικών λέξεων (BoVW) για την ταξινόμηση της δυναμικής χειρονομίας.
- Fanello et al. [11]. Κάνουν χρήση αραιοποίησης (sparse coding) για να αναγνωρίσουν και να κατατμήσουν χρονικά δυναμικές χειρονομίες.
- Esfahani et al. [31]. Εκπαιδεύουν ένα συνελκτικό νευρωνικό δίκτυο και εφαρμόζουν δυαδικοποίηση (binarization), για ταχύτερη λειτουργία του δικτύου.
- Kui Liu et al. [23]. Χρησιμοποιούν την τεχνική DTW, σε συνδυασμό με την αναγνώριση βάθους μέσω 2 αισθητήρων με διαφορετική γωνία λήψης, για να παραγάγουν έναν εύρωστο αλγόριθμο αναγνώρισης μη δυναμικά μεταβαλλόμενων χειρονομιών.

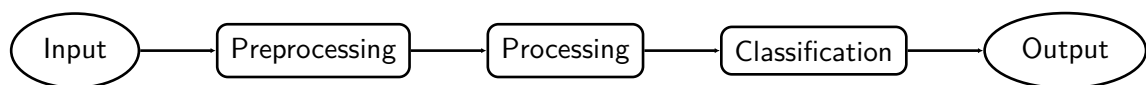
Μια ενδελεχής έρευνα, η οποία πραγματοποιήθηκε από τους Escallera et al. [10], περιγράφει εκτενώς μια πληθώρα μεθόδων αντιμετώπισης του προβλήματος της αναγνώρισης χειρονομιών στο διάστημα των ετών 2011 ως 2015. Εκεί γίνεται αυστηρότερη διάκριση των μεθόδων, τόσο σε επίπεδο προεπεξεργασίας της εικόνας όσο και σε επίπεδο επιλογής κατάλληλων χαρακτηριστικών περιγραφής της αλλά και στο τι είδους χειρονομία ζητείται να ανιχνευθεί και σε ποιο σημασιολογικό σύνολο ανήκει, όπως φαίνεται στο Σχήμα 1.1.



Σχήμα 1.1: Είδη Χειρονομιών, όπως περιγράφονται από τους Escallera et al. [10]. Ο διαχωρισμός καθορίζεται από τη χρηστικότητα και τη χρονική μεταβλητότητα της χειρονομίας, καθώς επίσης και τη συσχέτιση αυτής με το υπόλοιπο ανθρώπινο σώμα.

1.3.3 Γενικό Μοντέλο Μεθόδων Αναγνώρισης Χειρονομιών

Όπως φαίνεται από τα παρατιθέμενα παραδείγματα εργασιών, δεν υπάρχει κάποιος βασικός άξονας στην έρευνα για την βέλτιστη μέθοδο Αναγνώρισης, πέρα από το ακόλουθο γενικευμένο μοντέλο:



Σχήμα 1.2: Γενικευμένο Μοντέλο Μεθόδων Αναγνώρισης Χειρονομιών

Στα παραπάνω στάδια η είσοδος διαφοροποιείται μεταξύ των αλγορίθμων, όντας είτε έγχρωμη εικόνα είτε εικόνα βάθους ή και τα δύο μαζί, ενώ η έξοδος είναι η πόζα του χεριού ή μια φράση (δηλ. το αποτέλεσμα της αναγνώρισης). Λαμβάνοντας υπόψη το εύρος της παραμετροποίησης που υφίσταται ο κάθε αλγόριθμος αναγνώρισης χειρονομιών, κάποιος θα αντιμετώπιζε δυσκολία στη σύγκριση των μεθόδων καθώς επίσης και των ποσοστών επιτυχίας αυτών.

1.3.4 Ανάπτυξη Νέας Μεθοδολογίας

Στην εργασία αυτή βασιζόμαστε σε προηγούμενες προσπάθειες, για να γεφυρώσουμε το χάσμα μεταξύ θεωρίας και εφαρμογής, προτείνοντας έναν εύρωστο αλγόριθμο, με τον οποίο ελαχιστοποιούμε τους επιβαλλόμενους περιορισμούς, τόσο σε επίπεδο περιβάλλοντος όσο και σε επίπεδο χρήστη, και τις υπολογιστικές απαιτήσεις, με κάποια απώλεια ακρίβειας. Συγχρόνως, επιτρέπουμε την επέκταση των δεδομένων μάθησης του αλγορίθμου με τις ίδιες απαιτήσεις και την προσθήκη αυτού ως υπομονάδα ενός μεγαλύτερου συστήματος, αφού υπάρχει προσπάθεια γενίκευσης των πεδίων εφαρμογής του.

Ενδεικτικά, αναφέρονται οι καινοτομίες που έγιναν, για να χαρακτηρίσουν την εργασία αυτή μερικώς πρωτότυπη:

- Στο στάδιο της προεπεξεργασίας ενσωματώνουμε έναν πρότυπο αλγόριθμο ταχείας Σκελετοποίησης (βλ. Εν. 3.2.5).
- Στο στάδιο της εξαγωγής χαρακτηριστικών κατασκευάστηκαν και δοκιμάστηκαν δύο πρότυποι περιγραφητές, ο 3DXYPKA (βλ. Εν. 3.3.4) και ο ZHOF (βλ. Εν. 3.3.3).
- Κατά την εφαρμογή της ταξινόμησης, κάνουμε χρήση του πρωτότυπου μοντέλου CD-BIMM (βλ. Εν. 3.6.3).

Πέρα από τις παραπάνω πρωτοτυπίες, χρησιμοποιούμε τροποποιήσεις προϋπάρχοντων μοντέλων αναγνώρισης, όπως του μοντέλου Αναγνώρισης Κινούμενου Αντικειμένου με χρήση Μίξης Γκαουσιανών (βλ. Εν. 3.2.3), για να καταστούν συμβατά με τη φύση των χρησιμοποιούμενων δεδομένων (βάθους).

1.4 Σχετική Θεματολογία

Η επικοινωνία ανθρώπου και υπολογιστή διερευνάται ενδελεχώς τα τελευταία χρόνια έτσι, ώστε να αυξηθεί το διαθέσιμο εύρος εφαρμογών, να μειωθεί το κόστος των σχετικών εφευρέσεων και να απελευθερωθεί η αγορά σε όλες τις κοινωνικές ομάδες, επιφέροντας ισότιμα προνόμια πληροφόρησης και υπηρεσιών σε όλους, ανεξαρτήτως οικονομικής κατάστασης, ηλικιακής διαφοράς ή προβλημάτων υγείας. Λόγω του ότι η παρούσα εργασία σχετίζεται άμεσα με την επικοινωνία αυτή και ακόμα επειδή θεωρείται χρήσιμο να αναδειχθεί η προσφορά της στην τεχνολογία και στην κοινωνία, αναφέρονται περιληπτικά οι εξής άξονες:

- Εξελίξεις στην αλληλεπίδραση ανθρώπου-μηχανής μέσω κίνησης.
- Πρόσβαση ευπαθών κοινωνικών ομάδων στην σύγχρονη τεχνολογία.

1.4.1 Εξελίξεις στην Αλληλεπίδραση Ανθρώπου-Μηχανής μέσω κίνησης

1.4.1.1 Ιστορική Ανασκόπηση

Αφ' ενός η επανάσταση στον τομέα των παράγωγων Πυριτίου και των έξυπνων υλικών και αφετέρου η διαρκής μείωση κόστους παραγωγής οδήγησαν στην εμπορική απελευθέρωση εφευρέσεων, που αντικαθιστούν τις συμβατικές μεθόδους εισαγωγής.

Μια από αυτές τις τεχνολογίες ήταν η οθόνες αφής (touch screens). Με την πολυαπτική τεχνολογία (multi-touch), η οποία, αν και γνωστή από το 1982 [22], ξεκίνησε να είναι δημοφιλής από το 2007, με την εισαγωγή της στις κινητές συσκευές από την Apple [48] η αλληλεπίδραση με τον υπολογιστή έπαψε να είναι μονοδιάστατη.

Παράλληλα, μια απομακρυσμένη μέθοδος εισαγωγής, η οποία συμπεριλάμβανε την κίνηση του χεριού και κατά την οποία χρησιμοποιείτο ένα ειδικά διαμορφωμένο γάντι, είχε εφευρεθεί ήδη από το 1987 [55], αλλά έγινε αρκετά δημοφιλής μέσω του Power Glove από την Nintendo το 1989 ως μέθοδος εισαγωγής σε βιντεοπαιχνίδια. Από τότε, πληθώρα εφαρμογών και εφευρέσεων βασίστηκε στην εν λόγω ευρεσιτεχνία, χωρίς όμως να αποκτήσει κυρίαρχη θέση στην αγορά, λόγω της δυσκολίας στη χρήση της.

Με την εμπορική διάθεση στο απλό κοινό αισθητήρων βάθους, όπως είναι το Kinect, το οποίο βγήκε στο εμπόριο το 2010, κατέστη δυνατή η αποστασιοποίηση της Αλληλεπίδρασης Ανθρώπου-Υπολογιστή από την ανάγκη κατασκευής καινοτόμου υλισμικού, για την υποστήριξη μιας νέας μεθόδου επικοινωνίας. Έτσι, το κύριο επιστημονικό ενδιαφέρον στράφηκε στην κατασκευή πρωτοπόρων αλγορίθμων, οι οποίοι χρησιμοποιούν μεθόδους παραλληλοποίησης και την συνεχώς εξελισσόμενη τεχνολογία των GPU, για να προσφέρουν στο χρήστη μια εμπειρία πραγματικού χρόνου. Η εξέλιξη αυτή έφερε τον τρισδιάστατο έλεγχο του υπολογιστή στο προσκήνιο, επιτρέποντας την επέκταση των ορίων ευρηματικότητας των εφαρμογών.

1.4.1.2 Τρέχουσα Ανάπτυξη

Το μέλλον της επικοινωνίας του Ανθρώπου και Υπολογιστή αποτελεί η Εικονική Πραγματικότητα. Νέες συσκευές, όπως το Oculus Rift [25], προσφέρουν μια άνευ προηγούμενου εμπειρία, κατά την οποία ο χρήστης, με τη βοήθεια κάποιων ειδικά διαμορφωμένων γυαλιών και κάποιων αισθητήρων στο περιβάλλον του, διεισδύει σε μια ψηφιακά παραμορφωμένη πραγματικότητα, μέσα στην οποία η επαφή του με τον υπολογιστή είναι πιο στενή από ποτέ, αφού αναγνωρίζονται όλες οι κινήσεις και περιστροφές του κεφαλιού, αλλά και οι κινήσεις των υπόλοιπων άκρων. Η διασύνδεση αυτή, η οποία παρεμβάλλεται ακόμα και στην αίσθηση προσανατολισμού του χρήστη, αλλά και στην αντίληψή του [51], είναι η πιο ακραία μορφή Αλληλεπίδρασης με τον Υπολογιστή αυτή τη στιγμή, με πλήθος διαθέσιμων εφαρμογών, που δεν περιορίζονται σε εγωκεντρικές ενέργειες, αλλά μπορούν και να συμβάλλουν στην εξέλιξη της διαπροσωπικής επικοινωνίας και του τρόπου με τον οποίο εφαρμόζονται πολλές επιστήμες, όπως η Χειρουργική [40].

1.4.2 Πρόσβαση Ευπαθών Ομάδων στη Σύγχρονη Τεχνολογία

Οι κοινωνικές ομάδες δεν τυγχάνουν ισότιμης μεταχείρισης, όσον αφορά στην εξέλιξη της Τεχνολογίας. Υπάρχουν πολλοί παράγοντες, που επιβάλλουν αυτήν την κατάσταση, πχ. πολιτικο-οικονομικοί, και δεν εμπίπτουν στη σφαίρα επιρροής της διεπιστημονικής κοινότητας. Κάποιες κοινωνικές ομάδες, για τις οποίες παρατηρείται άμβλυνση αυτών των διακρίσεων, λόγω τεχνολογικής υποβοήθησης, είναι οι άνθρωποι μεγάλης ηλικίας και άτομα που πάσχουν από διάφορα ψυχοσωματικά προβλήματα υγείας.

Ειδικότερα οι εξελίξεις στην διεπαφή με τον υπολογιστή προσφέρουν τη δυνατότητα σε άτομα με χαμηλή κινητικότητα να δίνουν φωνητικές εντολές ή μέσω χειρονομιών να επικοινωνούν με τη μηχανή, με στόχο την παροχή βοήθειας σε αυτούς και την κάλυψη κάποιων βασικών αναγκών τους. Χαρακτηριστικό παράδειγμα αποτελεί το σύστημα υποβοήθησης MOBOT [37], το οποίο επιτρέπει σε ηλικιωμένους να εκτελούν τις παραπάνω εργασίες και να αυτοεξυπηρετούνται σε έναν ικανοποιητικό βαθμό.

Αντίστοιχα, άτομα που αντιμετωπίζουν προβλήματα ομιλίας μπορούν να αλληλεπιδράσουν με τον υπολογιστή μέσω χειρονομιών Νοηματικής Γλώσσας (Sign Language), χρησιμοποιώντας προγράμματα αναγνώρισης, τα οποία προσφέρουν πραγματικού χρόνου εμπειρία [19]. Ομοίως, για τα άτομα με δυσκολία όρασης, επιστρατεύεται η επιστήμη της Επεξεργασίας Εικόνας και η

Μηχανική Μάθηση στη δημιουργία αλγορίθμων, που τους επιτρέπουν, σε συνδυασμό με συμβατές συσκευές, όπως είναι το OrCam [29], να ενημερώνονται για το περιβάλλον τους. Αυτά αποτελούν μόνο κάποια παραδείγματα από το μεγάλο πλήθος εργασιών, που επιδιώκουν να υποβοηθήσουν τις ευπαθείς πληθυσμιακές ομάδες.

1.5 Περιγραφή Βασικών Εννοιών

Παρακάτω αναφέρονται κάποιες έννοιες που συναντιούνται συχνά μέσα στο κείμενο και η τυποποίηση της περιγραφής τους μπορεί να υποβοηθήσει την κατανόηση από τον αναγνώστη.

- Χειρονομία ή Χειρομορφή: Ως χειρονομία στην εργασία αυτή θα ορίζουμε μία πράξη, η οποία εκτελείται από κάποιο άτομο με τη χρήση του μέρους του χεριού από τον καρπό μέχρι τα δάχτυλα. Η θέση της πράξης αυτής ως προς την κάμερα δεν διαδραματίζει ρόλο στον χαρακτηρισμό της χειρονομίας.
- Δυναμική Χειρονομία: Ως **Δυναμική** ορίζουμε τη χειρονομία, η οποία χαρακτηρίζεται τόσο από τη σχετική θέση των αρθρώσεων όσο και από τη χρονική μεταβολή αυτών ως προς το σύστημα συντεταγμένων *CS*. Το *CS* έχει κέντρο τον **καρπό** του χεριού και ο προσανατολισμός του μπορεί να τεθεί αυθαίρετα σταθερός ή συσχετισμένος με τον προσανατολισμό της παλάμης.
- Παθητική Χειρονομία: Ως **Παθητική** ορίζουμε τη χειρονομία, η οποία χαρακτηρίζεται μόνο από τη σχετική θέση των αρθρώσεων.

Χρονικά μεταβαλλόμενες παθητικές χειρονομίες συνθέτουν μια δυναμική χειρονομία. Από το γεγονός αυτό υπονοείται η δυνατότητα **συνύπαρξης** και των δύο τύπων χειρονομιών σε ένα συγκεκριμένο χρονικό σημείο.

1.6 Στόχοι Εργασίας και Επιβαλλόμενοι Περιορισμοί

1.6.1 Επιθυμητές Ιδιότητες Προτεινόμενου Συστήματος από την Άποψη της Αλληλεπίδρασης Ανθρώπου-Μηχανής

Η παρούσα Διπλωματική Εργασία καλείται να αντιμετωπίσει έναν συνδυασμό προβλημάτων, τα οποία παρουσιάζονται μαζί με τον τρόπο που επιλύονται ακολούθως:

- Επίτρεψη αναγνώρισης χειρονομιών σε περιβάλλον με χαμηλό ή μη ομοιόμορφο μεταβαλλόμενο φωτισμό, όπου μια προσέγγιση χρησιμοποιώντας έναν αισθητήρα χρώματος (RGB) θα αποτύγχανε. Αυτό είναι εφικτό από τη στιγμή που καθίσταται δυνατή η χρήση μόνο του αισθητήρα βάθους για την αναγνώριση των χειρονομιών, πράγμα που επιτυγχάνεται σε έναν ικανοποιητικό βαθμό στην παρούσα εργασία.
- Αναγνώριση προηγμένων, υψηλά συσχετισμένων παθητικών και δυναμικών χειρονομιών. Ζητείται να ανιχνεύονται οι χειρονομίες, ενώ και οι θέσεις των αρθρώσεων αλλάζουν δυναμικά ως προς το σημείο αναφοράς του χεριού (τον καρπό) και η θέση του καρπού μεταβάλλεται ως προς το σύστημα αναφοράς της κάμερας. Αυτό σημαίνει ότι η προβολή του χεριού στο πεδίο της κάμερας είναι σε υψηλό βαθμό παραμορφώσιμη και η θέση του καρπού δεν είναι αυστηρώς ορισμένη. Πέρα από αυτό, υπονοείται ότι η αναγνώριση μιας μόνης χειρονομίας, δυναμικής ή στατικής, είναι ανεξάρτητη της θέσης του καρπού. Το γεγονός αυτό επιτρέπει τη χρήση της εφαρμογής από άτομα που πάσχουν από κάποια ασθένεια στη κινητική μυική μονάδα, η οποία προκαλεί ασταθείς ή τρεμάμενες κινήσεις (πχ. Parkinson), ή άτομα τα οποία βρίσκονται σε κάποιο κινούμενο περιβάλλον ως προς το στατικό περιβάλλον της κάμερας και του υποβάθρου, για παράδειγμα σε ένα κινούμενο όχημα. Η γρήγορη σκελετοποίηση του χεριού που επιτυγχάνεται κατά τη διάρκεια της προεπεξεργασίας επιτρέπει την αποσυνέλιξη της χειρονομίας από την θέση και τη γωνία του ως προς το σύστημα συντεταγμένων του αισθητήρα.
- Επεκτασιμότητα στις αναγνωρίσιμες χειρονομίες. Το σύστημα εκπαίδευσης που χρησιμοποιείται από τον αλγόριθμο απαιτεί λίγους σχετικά υπολογιστικούς πόρους, ενώ η προσθήκη δεδομένων εκπαίδευσης είναι πολύ απλή. Παρόντες αλγόριθμοι, επιδιώκοντας να είναι συγκρίσιμοι με ήδη υπάρχοντες, εκπαιδεύονται με δεδομένα προεπεξεργασμένα από κάποια άλλη ερευνητική ομάδα, των οποίων η επισήμανση (annotation) επιτυγχάνεται με χρήση διαφόρων μεθόδων, όπως για παράδειγμα αισθητήρες μαγνητικής επαγωγής στα δάκτυλα [49]. Η παρούσα εργασία, λόγω των ιδιότυπων περιορισμών που έχουν επιβληθεί ως προς τη θέση και την κίνηση του χρήστη (βλ. Εν. 1.6.2, χρησιμοποιεί ένα σύνολο δεδομένων (dataset), το οποίο έχει προέλθει από την προεπεξεργασία που περιέχεται στο αλγοριθμικό κομμάτι. Σε αυτή την κατεύθυνση, έχει δημιουργηθεί και μία πλατφόρμα με γραφικό περιβάλλον (GUI), η οποία δίνει τη δυνατότητα εύκολης προσθήκης δεδομένων εκπαίδευσης ανά πάσα στιγμή από οποιονδήποτε, η οποία περιγράφεται περιληπτικά στο Κεφάλαιο. Το annotation εν χρήσει είναι απλό, αφού αρκεί να χαρακτηριστεί κάθε frame με τη χειρονομία (ή τις χειρονομίες) που βρίσκεται σε εξέλιξη.

- Εκτέλεση σε πραγματικό χρόνο, με μικρό υπολογιστικό κόστος. Ιδιαίτερη μέριμνα δόθηκε στο να καταστεί ο προτεινόμενος αλγόριθμος τόσο αποδοτικός, ώστε να μπορεί να εκτελεστεί σε χρόνο χαμηλότερο αυτού του πραγματικού σε φυσιολογικές συνθήκες, δηλαδή σε συνθήκες οι οποίες είναι προδιαγεγραμμένες από τους επιβαλλόμενους περιορισμούς της υποενότητας 1.6.2. Πολλοί state-of-the-art αλγόριθμοι προϋποθέτουν την ύπαρξη ισχυρών υπολογιστικών πόρων, όπως υψηλών προδιαγραφών κάρτα γραφικών, ιδίως κατά τη διάρκεια της εκπαίδευσης [28, 34], ενώ η προτεινόμενη μέθοδος εκτελείται, χωρίς μέθοδο πολυεπεξεργασίας (multiprocessing), με τη χρήση ενός πυρήνα της διαθέσιμης CPU.

1.6.2 Επιβαλλόμενοι Περιορισμοί

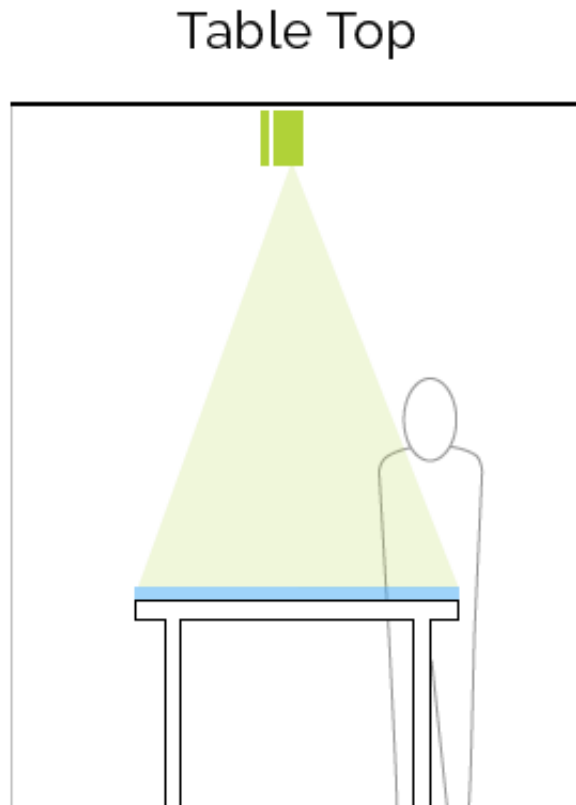
Οι ακόλουθες υποθέσεις έγιναν, με σκοπό τη μείωση της διαστατικότητας του προβλήματος που επιδιώκουμε να επιλύσουμε:

- Ο αισθητήρας παραμένει σταθερός κατά τη διάρκεια της εκτέλεσης.
- Το οπτικό πεδίο του αισθητήρα (FOV) δεν περιέχει χρονικές μεταβολές, πέρα από αυτές που προκαλούνται από το κινούμενο αντικείμενο, το οποίο δεν μπορεί να επηρεάσει το αρχικό υπόβαθρο.
- Το χέρι, δηλ. ολόκληρο το ανθρώπινο χέρι μέχρι τον ώμο, είναι το μόνο κινούμενο ορατό ανθρώπινο μέλος στην εικόνα.
- Το χέρι κινείται με ταχύτητα, τέτοια ώστε το ανθρώπινο μάτι να διακρίνει συνέχεια στην κίνησή του.
- Το χέρι δεν ανήκει μέσα στο FOV του αισθητήρα στην εκκίνηση της λήψης (για τουλάχιστον ένα δευτερόλεπτο).
- Ένα μόνο χέρι μπορεί να ανιχνευθεί. Αυτός ο περιορισμός μπορεί να αρθεί σε περίπτωση χρήσης μεθόδου πολυεπεξεργασίας (multiprocessing).

Πέρα από τις παραπάνω υποθέσεις, μερικοί περιορισμοί παράγονται από τον χρησιμοποιούμενο αισθητήρα:

- Ο αισθητήρας βάθους επιβάλλει κάποια όρια, ως προς το πόσο κοντά ή πόσο μακριά βρίσκεται το αντικείμενο υπό επίβλεψη.

- Το FOV του αισθητήρα περιέχει κυρίως αντικείμενα, των οποίων η επιφάνεια είναι αρκετά τραχεία (πχ. ξύλο) και δεν είναι διαπερατή (όπως πχ. γυαλί). Σε περίπτωση που αυτές οι προϋποθέσεις δεν ισχύουν, μπορεί να δημιουργηθεί πρόβλημα απώλειας πληροφορίας βάθους και μείωση ποιότητα αναγνώρισης. Για παράδειγμα, μια ιδανική τοποθέτηση του αισθητήρα θα ήταν πάνω από ένα τραπέζι (το οποίο δεν έχει γυάλινη επιφάνεια ούτε είναι πολύ γυαλιστερό) και να το ‘κοιτάζει’ κατακόρυφα:

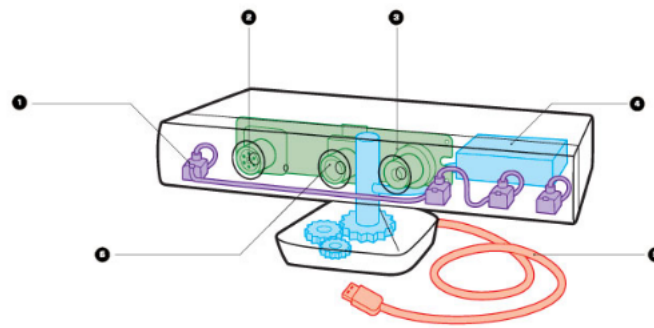


Σχήμα 1.3: Προτεινόμενη Τοποθέτηση Αισθητήρα [45]

1.7 Χρησιμοποιούμενο Υλισμικό

1.7.1 Kinect

Για την παρούσα εργασία χρησιμοποιήθηκε η συσκευή Kinect v2 της Microsoft για τη λήψη δεδομένων από το περιβάλλον. Το Kinect είναι ένα σύστημα, εξοπλισμένο με αισθητήρες εικόνας, χρώματος και βάθους, και ήχου. Παρακάτω φαίνονται τα συστατικά στοιχεία του Kinect v1, το οποίο περιέχει παρόμοια χαρακτηριστικά:



<p>1 Microphone array</p> <p>Four mics pinpoint where voices or sounds are coming from while filtering out background noise.</p>	<p>2 IR emitter</p> <p>Projects a pattern of infrared light into a room. As the light hits a surface, the pattern becomes distorted, and the distortion is read by the depth camera.</p>	<p>3 Depth camera</p> <p>Analyzes IR patterns to build a 3-D map of the room and all objects and people within it.</p>	<p>4 Tilt motor</p> <p>Automatically adjusts based on the object in front of it. If you're tall, it tilts the box up. If you're short, it knows to angle down.</p>	<p>5 USB cable</p> <p>Transmits data to the Xbox via an unencrypted feed, which makes it relatively easy to use the Kinect with other devices.</p>	<p>6 Color camera</p> <p>Like a webcam, this captures a video image. The Kinect uses that information to get details about objects and people in the room.</p>
--	--	--	--	--	--

Σχήμα 1.4: Περιγραφή Αισθητήρων της συσκευής Kinect [44]

Όπως έχει προαναφερθεί, ο μοναδικός αισθητήρας που χρησιμοποιήθηκε είναι ο αισθητήρας βάθους, ο οποίος είναι ένα σύστημα, αρχικά κατασκευασμένο από την PrimeSense, με έναν υπερπομπό, δηλ. ένα υποσύστημα πομπών, χωρικά (στο διδιάστατο επίπεδο) μεταβαλλόμενων υπέρυθρων ακτίνων (IR) και μία κάμερα, η οποία ανιχνεύει την ανακλώμενη ακτινοβολία. Η συσκευή εκμεταλλεύεται δύο κατά βάση φαινόμενα για την εξαγωγή του βάθους [39], τα οποία είναι :

- Η παράλλαξη. Η κάμερα απέχει από τη θέση του υπερπομπού. Όταν γνωστή η αρχική διάταξη των εκπεμπομένων ακτίνων, οι τιμές τις οποίες λαμβάνει κατωφλιοποιημένα η IR κάμερα αντιστοιχίζονται στις νέες θέσεις των ακτίνων. Με άλλα λόγια, με γνωστή την απόσταση της κάμερας από τον πομπό, καθώς επίσης τη διαφορά γωνίας μετάδοσης και λήψης, με χρήση απλής τριγωνομετρίας, αναγνωρίζεται η απόσταση κάποιου σημείου στο χώρο.

- Το φαινόμενο της εστίασης του φακού. Αν και η παράλλαξη αποτελεί το κύριο εργαλείο τρισδιάστατης απεικόνισης, η εστίαση του φακού δίνει επιπλέον χρήσιμη πληροφορία για το βάθος, αν είναι γνωστές οι παράμετροι της κάμερας [52].

Το Kinect περιέχει στο εσωτερικό του ακόμα ένα ηλεκτρονικό κύκλωμα, το οποίο υλοποιεί ταχύτατα έναν αλγόριθμο σκελετοποίησης και αναγνώρισης βασικών χειρονομιών. Παρ'όλ'αυτά, προαπαιτείται από τον αλγόριθμο να φαίνεται ολόκληρο το σώμα του χρήστη για να λειτουργήσει. Σε συνδυασμό με το γεγονός ότι ήταν επιθυμητό να μην περιορίσουμε την εφαρμογή του παραγόμενου αλγορίθμου σε μία συγκεκριμένη συσκευή, κρίθηκε σωστό να μην χρησιμοποιηθεί αυτή η υπομονάδα.

1.7.2 Άλλο υλικό

Πέρα από το Kinect, χρησιμοποιήθηκε ο υπολογιστής-notebook 'ASUS N550JK', ο οποίος έχει τις ακόλουθες, διαμορφωμένες προδιαγραφές:

- Επεξεργαστή Intel® Core™ i7 4700HQ
- Κάρτα Γραφικών NVIDIA® GeForce® GTX850M με μνήμη 4GB DDR3
- Μνήμη RAM DDR3L 1600 MHz SDRAM, DIMM, χωρητικότητας 16GB
- Λειτουργικό Σύστημα Linux Ubuntu 16.04
- Σκληρό Δίσκο SSD

1.7.3 Χρησιμοποιούμενο Λογισμικό

Η εργασία στηρίχτηκε στο ακόλουθο software:

- OpenCV: Συνθέτει τον πυρήνα της διπλωματικής εργασίας, αφού μέσω αυτής χρησιμοποιήσαμε αλγορίθμους Όρασης Υπολογιστών, υλοποιημένους σε γλώσσα C, χωρίς να χρειαστεί προσωπικός μόχθος υλοποίησής αυτών εξ αρχής. [3]
- Επιστημονικές Βιβλιοθήκες της Python: Λόγω του ότι η εργασία έγινε εξ ολοκλήρου στη γλώσσα Python, χρησιμοποιείται η βιβλιοθήκη Μηχανικής Μάθησης Scikit Learn, καθώς επίσης οι Scipy και NumPy βιβλιοθήκες για πράξεις πινάκων και επίλυση συστημάτων με αριθμητικές μεθόδους.

- ROS, το οποίο είναι συμβατό μόνο με Linux και χρησιμοποιήθηκε για την επικοινωνία των διεργασιών και τη διασύνδεση με το Kinect. [36]
- Kinect iai : πρόγραμμα που χρησιμοποιήθηκε σε συνεργασία με το ROS, για την λήψη δεδομένων από το Kinect.[50]

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Η εργασία αυτή στηρίχτηκε σε μια εκτενή θεωρητική υποδομή, η οποία προκύπτει κυρίως από δύο επιστημονικούς κλάδους, τον τομέα της Μηχανικής Μάθησης και αυτόν της Όρασης Υπολογιστών. Για το λόγο αυτό, κρίθηκε σωστό να διακριθεί η αναφορά της κατ' αρχήν σε αυτές τις δύο κύριες ενότητες. Μία τρίτη ενότητα συνθέτουν γνώσεις, οι οποίες συνδέονται και με τους δύο τομείς, ενώ ο αναγνώστης παραπέμπεται σε παραρτήματα στο τέλος της εργασίας, για να ενημερωθεί για βασικά θεωρητικά θέματα.

2.1 Όραση Υπολογιστών

Στην παρούσα ενότητα περιγράφονται θεωρητικά προαπαιτούμενα για να κατανοήσει κανείς το μέρος της προεπεξεργασίας και της εξαγωγής χαρακτηριστικών του κατασκευασθέντος αλγορίθμου.

2.1.1 Γενικά

Σε αντίθεση με την επιστήμη της Επεξεργασίας Εικόνας, η επιστήμη της Όρασης Υπολογιστών επιδιώκει να εξαγάγει από μία εικόνα χαρακτηριστικά, τα οποία εκφράζουν τα πρόσωπα και τα αντικείμενα μέσα σε αυτή, δηλαδή επιδιώκει να συσχετίσει τοπική πληροφορία, η οποία είναι σε μορφή εικονοστοιχείων (pixels), για να περιγράψει οντότητες του φυσικού κόσμου. Μέσω του μαθηματικού φορμαλισμού και της φυσικής μοντελοποίησης, προσεγγίζονται συνήθως με μη γραμμικό διακριτό τρόπο φαινόμενα τα οποία εκτυλίσσονται μέσα στην εικόνα (και σε βάθος

χρόνου, αν πρόκειται για εικονοσειρά (video)), εκ των οποίων εξάγεται μία συνεχής περιγραφή του σχήματος, της υφής, της στατικής και της δυναμικής των εν λόγω αντικειμένων, καθώς επίσης και της ετεροσυσχέτισής τους. Με τη χρήση στατιστικής απεικόνισης, η περιγραφή αυτή γενικεύεται και χρησιμοποιείται ως πρότερη γνώση για τη μελέτη άλλων σχετικών οντοτήτων. Βασικά εργαλεία για την εν λόγω επεξεργασία αποτελούν η επιστήμη των Μαθηματικών, τη Φυσικής και των Υπολογιστών, για την κατασκευή αποδοτικών αλγορίθμων.

2.1.2 Ανάλυση Συνδεδεμένων Συνιστωσών Connected Components Analysis

2.1.2.1 Ανίχνευση Συνδεδεμένων Συνιστωσών

Το πρόβλημα συνδετικότητας γράφων, με το οποίο ασχολείται η Επιστήμη Υπολογιστών, βρίσκει εφαρμογή στην ανίχνευση των αντικειμένων σε μία εικόνα. Κατωφλιώνοντας μια εικόνα εντός κάποιου εύρους τιμών, τα εικονοστοιχεία των οποίων ενδιαφερόμαστε να επεξεργαστούμε, μπορούμε να εξαγάγουμε μία ασπρόμαυρη εικόνα. Έστω 0 η τιμή των μαύρων αποκλεισμένων pixel και 1 η τιμή των λευκών pixel. Η εικόνα που προκύπτει είναι στην πραγματικότητα ένας πίνακας γειτνίασης, στον οποίο τα ενεργά στοιχεία (τα λευκά pixel) υποδεικνύονται από μία θέση, που, στη περίπτωση διδιάστατης εικόνας, αντιπροσωπεύεται από X-Y συντεταγμένες, ως προς κάποιο σημείο αναφοράς. Σε αυτήν τη μορφή, αλγόριθμοι συνδετικότητας, που συναντιούνται στην Πληροφορική, μπορούν να εφαρμοσθούν για τη διασύνδεση των γειτονικών ενεργών εικονοστοιχείων και την εξαγωγή των συνδεδεμένων στοιχείων (connected components). Στις διδιάστατες εικόνας, η συνδετικότητα εμφανίζεται συνήθως σε δύο μορφές, στην 4-συνδετικότητα (4-connectivity), η οποία αναφέρεται στην κατακόρυφη και οριζόντια διασύνδεση, και στην 8-συνδετικότητα (8-connectivity), η οποία αναφέρεται στην διασύνδεση σε όλες τις κατευθύνσεις. Για να μην υπάρξει ασάφεια, τα ενεργά pixel διασυνδέονται με αντίθετη συνδετικότητα από αυτή των ανενεργών [43]. Ένας από τους πιο διαδεδομένους αλγορίθμους είναι αυτός που εισηγήθηκαν επίσημα οι Vincent et al. [46], ο οποίος είναι πολυπλοκότητας $O(N \cdot C)$, όπου N ο αριθμός των pixel της εικόνας και C ο βαθμός συνεκτικότητας, και περιγράφεται στο Παράρτημα A.1.

2.1.2.2 Ανίχνευση Περιγράμματος

Προς μία παρεμφερή κατεύθυνση με αυτήν της Ανίχνευσης Συνδεδεμένων Στοιχείων κινείται και η Ανίχνευση Περιγράμματος. Στόχος στην προκειμένη περίπτωση είναι να βρεθεί το περίγραμμα των σχημάτων σε μία κατωφλιωμένη εικόνα και να χαρακτηριστεί ως εσωτερικό ή εξωτερικό, ανάλογα με το αν η κλειστή καμπύλη περιβάλλει ή περιβάλλεται από το εν λόγω αντικείμενο. Οι Suzuki et. al [43] έχουν δημιουργήσει έναν από τους πιο διαδεδομένους αλγορίθμους, βασισμένους στην Επιστήμη Υπολογιστών, για την ανίχνευση του περιγράμματος σε δισδιάστατες εικόνες, ο οποίος παραμετροποιείται έτσι, ώστε να είναι δυνατή και η περαιτέρω ταξινόμηση των ανιχνευθέντων περιγραμμάτων σε μία λογική γενεαλογικού δέντρου. Πιο συγκεκριμένα, ως ‘πατέρας’ ενός περιγράμματος ορίζεται αυτό που το περιβάλλει. Ακόμα, ένα ισχυρό πλεονέκτημα του αλγορίθμου αυτού είναι ότι εξάγει προσανατολισμένα περιγράμματα, δηλαδή καμπύλες με φορά τέτοια, ώστε το περιγεγραμμένο αντικείμενο να βρίσκεται στα δεξιά της κίνησης τους. Το αποτέλεσμα του αλγορίθμου λοιπόν είναι ένα δέντρο, του οποίου τα φύλλα είναι λίστες των συντεταγμένων των περιγραμμάτων, με σειρά αναφοράς τέτοια, ώστε να τηρείται η προαναφερθείσα φορά κίνησης.

2.1.2.3 Ανάλυση Χαρακτηριστικών Περιγράμματος

Το εξαγόμενο περίγραμμα μπορεί να δώσει ενδιαφέρουσες πληροφορίες σχετικά με το αντικείμενο που περιβάλλει. Χρησιμοποιώντας την ειδικευση του θεωρήματος Kelvin-Stokes στο δισδιάστατο χώρο, με άλλα λόγια το θεώρημα του Green και ενώνοντας το περίγραμμα και τα περιγράμματα που αυτό περιβάλλει, μπορούμε να προσεγγίσουμε τις ροπές που περιγράφουν το αντικείμενο [60]. Ως ροπές χαρακτηρίζουμε τις ποσότητες:

$$M_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

με $p, q = 0, 1, 2, \dots$ και f τη συνάρτηση που περιγράφει το αντικείμενο, η οποία, στην περίπτωση της κατωφλιωμένης εικόνας, έχει σύνολο τιμών

$$f \rightarrow \{0, 1\}$$

Το θεώρημα ή τύπος του Green περιγράφεται στο Παράρτημα B.1 . Αυτός ο τύπος δίνει τη δυνατότητα, θέτοντας τις συναρτήσεις P και Q έτσι, ώστε

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1$$

να βρεθεί το εμβαδόν του αντικειμένου, χρησιμοποιώντας επικαμπύλια ολοκληρώματα που περιγράφουν το περίγραμμα. Τα ολοκληρώματα αυτά ανάγονται σε αθροίσματα στο διακριτό επίπεδο της εικόνας. Ομοίως, θέτοντας τις παραπάνω συναρτήσεις ώστε στο δεξί μέλος του τύπου Green να προκύψουν οι σχέσεις των ροπών ανώτερης τάξης, δύναται κανείς να υπολογίσει άλλα γεωμετρικά χαρακτηριστικά του περιγεγραμμένου αντικειμένου, όπως το κέντρο μάζας, τα οποία παραμένουν αναλλοίωτα υπό την δράση ενός αφινικού μετασχηματισμού.

2.1.3 Οπτική Ροή

Όταν ένα φαινόμενο εξελίσσεται χρονικά μέσα σε μία εικονοσειρά, είναι επιθυμητό να αναλυθεί η κινητική του κατάσταση, ώστε να εξαχθούν χρήσιμα συμπεράσματα σχετικά με το ίδιο αλλά και την δράση του στο περιβάλλον. Η Οπτική Ροή (Optical Flow) αποτελεί ένας από τους βασικούς τρόπους προσέγγισης τέτοιων προβλημάτων. Από τους πρώτους ερευνητές που τυποοίησαν αυτήν την ιδέα, βασίζόμενοι στην τοπική ομοιότητα των εικονοστοιχείων που αναπαριστούν τη μελετώμενη κίνηση, ήταν οι Bruce D. Lucas και Takeo Kanade [24]. Προσεγγίζοντας το πρόβλημα του μετασχηματισμού εικόνων σε ένα κοινό σύστημα αναφοράς (Image Registration), εφηύραν μία μέθοδο η οποία βασίζεται στις σειρές Taylor (βλ. Παρ. B.2).

Ο αλγόριθμος της Οπτικής Ροής βασίζεται στην εξής προσέγγιση Taylor:

$$F(\vec{x} + \vec{h}) \approx F(\vec{x}) + \vec{h} \nabla_{\vec{x}} F(\vec{x}) \quad (2.1)$$

όπου $\nabla_{\vec{x}}$ είναι η πράξη gradient, ως προς το διάνυσμα $\vec{x} \in \mathbb{R}^n$

$$\nabla_{\vec{x}} = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \dots, \frac{\partial}{\partial x_n} \right]^T$$

και $\vec{h} \in \mathbb{R}^n$ είναι το διάνυσμα μετατόπισης. Στόχος είναι να βρεθεί η μετατόπιση \vec{h} για έναν πίνακα στοιχείων \vec{x} :

$$X = [\vec{x}_i], i = 1, 2, \dots, m$$

έτσι, ώστε να ελαχιστοποιηθεί η ενέργεια L_2 της ομοιότητας (Similarity) δύο συνεχόμενων εικόνων F και G :

$$E = \left\| G(X + \vec{h}) - F(X) \right\|_F^2 \quad (2.2)$$

όπου $\|(\cdot)\|_F$ είναι η νόρμα κατά Frobenius:

$$\|A\|_F := \text{tr}(A^T A)$$

Χρησιμοποιώντας την προσέγγιση της σχέσης 2.1 και απαιτώντας οι μερικοί παράγωγοι της E ως προς x_{ij} , $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$ να είναι μηδενικές, λαμβάνουμε $n \times m$ εξισώσεις, οι οποίες πλαισιώνουν το γραμμικό σύστημα προς επίλυση. Υποθέτοντας ότι τα m στοιχεία είναι διανύσματα θέσης μιας γειτονιάς ενός κεντρικού σημείου \vec{x}_c και επιλύοντας το σύστημα για κάθε \vec{x}_c που μας ενδιαφέρει, υλοποιούμε τον αλγόριθμο των Lucas Kanade (LK), εξάγοντας την εκτίμηση της μετατόπισης κάθε σημείου ενδιαφέροντος. Ο παραπάνω αλγόριθμος γενικεύεται και για την περίπτωση όπου οι μεταβολές στην εικονοσειρά περιέχουν περιστροφές των αντικειμένων ή άλλου είδους γραμμικούς μετασχηματισμούς, με αυξανόμενη βέβαια την εν λόγω πολυπλοκότητα.

Η υπόθεση ότι η μετατόπιση σε μία γειτονιά είναι ανεξάρτητη της χρονικής μεταβολής της έντασης αυτής, η οποία γίνεται στον παραπάνω αλγόριθμο, δεν είναι απολύτως σωστή, επιφέροντας συχνά προσεγγιστικά λάθη. Μία καλύτερη εκτίμηση μπορεί να γίνει, αν προσεγγιστεί η ένταση της γειτονιάς των σημείων από μια πολυωνυμική σχέση, έτσι ώστε να γίνει συνεχής και διαφορίσιμη. Αυτή η σκέψη προήλθε από τον Gunnar Farneback [12]. Η προσέγγιση γίνεται με χρήση της Σταθμισμένης μεθόδου Ελαχίστων Τετραγώνων (βλ. Παρ. B.3). Η μορφή της πολυωνυμικής προσέγγισης σε μια γειτονιά σημείων \mathcal{N}_x δίνεται από τη σχέση:

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c$$

Με τη μέθοδο αυτή, η ένταση σε μία γειτονιά σημείων γίνεται διαφορίσιμη ως προς \mathbf{x} και η συνολική μετατόπιση, η οποία βρίσκεται με παρόμοιο τρόπο όπως αυτή στον αλγόριθμο LK, εξαρτάται πλέον και από την τοπική χωροχρονική μεταβολή της έντασης. Ο αλγόριθμος αυτός προϋποθέτει ότι οι μετατοπίσεις στις εικόνες μεταβάλλονται αργά στο χρόνο, αλλιώς εισάγεται μεγάλο σφάλμα προσέγγισης σε χρονικές ασυνέχειες. Με άλλα λόγια, ο ρυθμός δειγματοληψίας της εικονοσειράς θα πρέπει να είναι σχετικά μεγάλος.

2.1.4 Περιγραφητές

Για να καταστεί δυνατή η περιγραφή ενός αντικειμένου της εικόνας, η οποία δεν θα μεταβάλλεται σε περίπτωση δράσης κάποιου αφινικού μετασχηματισμού (βλ. Παρ. B.4) ή γενικής παραμόρφωσης (πχ. αλλαγή φωτεινότητας ή αντίθεσης), η εικόνα πρέπει να υποστεί κάποια συγκεκριμένη μη γραμμική επεξεργασία, η οποία κατόπιν συνήθως υπόκειται σε πολλαπλές κανονικοποιήσεις και στατιστικές προσεγγίσεις. Το αποτέλεσμα αυτής της διαδικασίας ονομάζεται περιγραφητής.

Υπάρχουν διάφοροι τύποι περιγραφητών και χωρίζονται σε δύο μεγάλες κατηγορίες, τους **περιγραφητές γενικής πληροφορίας**, οι οποίοι χρησιμοποιούνται για να αποφανθούν για το χρώμα, το σχήμα, την υφή και την κίνηση σε μία σκηνή, λαμβάνοντας υπόψη την τοπικότητα της πληροφορίας, και τους **περιγραφητές ειδικής εφαρμογής**, οι οποίοι συντίθενται συνήθως από τους προηγούμενους και χρησιμοποιούνται για να περιγράψουν συγκεκριμένου είδους σύνθετα αντικείμενα, όπως για παράδειγμα πρόσωπα.

Η εργασία αυτή χρησιμοποιεί περιγραφητές από την πρώτη κατηγορία. Πιο συγκεκριμένα, έγινε χρήση παραλλαγής των περιγραφητών HOOF (και της γενίκευσης 3DHOF) και HOG, οι οποίοι περιγράφονται ακολούθως.

2.1.4.1 Histogram of Oriented Optical Flow (HOOF)

Το Ιστόγραμμα Προσανατολισμένης Οπτικής Ροής (HOOF) αποτελεί έναν περιγραφητή γενικής πληροφορίας κίνησης, ο οποίος χρησιμοποιεί την κατεύθυνση της Οπτικής Ροής (βλ. Υποενότητα 2.1.3) μιας περιοχής ενδιαφέροντος (ROI, από το Region of Interest), για να μοντελοποιήσει τα χαρακτηριστικά της κίνησης [6]. Βασική απαίτηση για την περιγραφή της κίνησης είναι αυτή να είναι ανεξάρτητη από το μέτρο της ταχύτητας της κίνησης, καθώς επίσης του μεγέθους της περιοχής ενδιαφέροντος. Αφού υπολογιστεί η πυκνή (σε όλα τα εικονοστοιχεία της ROI) Οπτική Ροή, προκύπτουν δύο πίνακες στον χώρο των δύο διαστάσεων, οι πίνακες εκτίμησης της οριζόντιας και της κάθετης μετατόπισης. Από τους πίνακες αυτούς εξάγεται ο πίνακας κατευθύνσεων της κίνησης, ο οποίος κατόπιν χρησιμοποιείται για να παραχθεί ένα ιστόγραμμα κατευθύνσεων, με τα όρια των κάδων των ψήφων να υποδηλώνουν εύρη γωνιών. Κατόπιν, το ιστόγραμμα κανονικοποιείται, ώστε το άθροισμα των ψήφων να δίνει μονάδα, για να μην λαμβάνεται υπόψη το μέγεθος της περιοχής ενδιαφέροντος.

Μια γενίκευση αυτής της διαδικασίας στο χώρο των τριών διαστάσεων δίνεται από τους Fanello et. al [11]. Αρχικά υποτίθεται ότι διαθέτουμε ένα depth stream εικόνων, οπότε η πληροφορία της θέσης των πραγματικών σημείων του χώρου προκύπτει και από αυτή των εικονοστοιχείων, αλλά και από την έντασή τους. Η θέση των εικονοστοιχείων προκύπτει από έναν μη γραμμικό διακριτό μετασχηματισμό των θέσεων από το τρισδιάστατο χώρο στο επίπεδο της κάμερας και εξαρτάται από πληθώρα παραγόντων. Κάποιοι από αυτούς είναι η γωνία λήψης, η απόσταση του αντικειμένου από την κάμερα, το επίπεδο εστίασης, ο πίνακας χαρακτηριστικών του φακού της κάμερας, οι αποκλίσεις του φακού και το είδος της κάμερας. Απλουστεύοντας αυτήν την περίπλοκη κατά τ'άλλα σχέση μεταξύ των δύο χώρων, μπορεί κανείς να εφαρμόσει μια προσέγγιση, γνωστή ως Pinhole Camera Approximation. Εφαρμόζοντας αυτό το μοντέλο, μπορεί κάποιος, χρησιμοποιώντας μόνο την πληροφορία της θέσης (x, y) των εικονοστοιχείων, της απόστασης z του αντίστοιχου σημείου από την κάμερα, το μήκος εστίασης f (Focal Length) της κάμερας και το κέντρο της εικόνας (x_0, y_0) (Principal Point), να αντιπροβάλλει τα σημεία του επιπέδου της κάμερας στα σημεία του χώρου (X, Y, Z) :

$$\begin{bmatrix} X_t \\ Y_t \\ Z_t \end{bmatrix} \approx \begin{bmatrix} \frac{(x_t - x_0)z_t}{f} \\ \frac{(y_t - y_0)z_t}{f} \\ z_t \end{bmatrix} \quad (2.3)$$

Έτσι, προσθέτοντας τους πίνακες οριζόντιας και κάθετης μετατόπισης με τον πίνακα της αρχικής θέσης των εικονοστοιχείων και εφαρμόζοντας τον μετασχηματισμό της σχέσης 2.3, μπορούμε να υπολογίσουμε την εκτιμητέα θέση των σημείων του χώρου. Υπολογίζοντας και την αρχική θέση των σημείων με την ίδια προσέγγιση, μπορούμε τελικά να υπολογίσουμε την τρισδιάστατη Οπτική Ροή.

Το αποτέλεσμα είναι πλέον τρεις πίνακες, ο καθένας από τους οποίους περιγράφει τη μετατόπιση κάθε σημείου της ROI σε κάθε διάσταση. Κανονικοποιώντας τις αντίστοιχες θέσεις των πινάκων, ώστε το μέτρο της ταχύτητας να είναι μοναδιαίο, λαμβάνουμε τους πίνακες κατευθύνσεως της κίνησης. Κατόπιν, κατασκευάζεται το τρισδιάστατο ιστόγραμμα κατευθύνσεων, του οποίου κάθε κάδος, ο οποίος απεικονίζεται πλέον ως ένα ορθογώνιο παραλληλεπίπεδο, έχει ακμή σε κάθε διάσταση ένα εύρος της προβολής του διανύσματος κατευθύνσεως στο χώρο. Κανονικοποιώντας ξανά ως προς τον αριθμό των ψήφων, λαμβάνουμε έναν περιγραφητή, ο οποίος ονομάστηκε 3DHOF (3D Histogram of Optical Flow) από τους Fanello et al. [11], ανεξάρτητο της κατεύθυνσης της κίνησης και του μεγέθους της αντιπροβολής της μελετώμενης δισδιάστατης

ROI στον τρισδιάστατο χώρο.

2.1.4.2 Histogram of Oriented Gradient(HOG)

Το Ιστόγραμμα Προσανατολισμένων Βαθμίδων είναι ένας περιγραφητής του σχήματος και της υφής, ο οποίος είχε παρουσιαστεί αρχικά με διαφορετικό όνομα από τους McConnel et. al[27]. Προσεγγίζουμε καταρχάς τη βαθμίδα (gradient) της εικόνας με χρήση πεπερασμένων διαφορών. Ομαδοποιώντας τη σε γειτονιές εικονοστοιχείων, συνήθως με μεγάλη αλληλοεπικάλυψη, και κατασκευάζοντας για την κάθε γειτονιά το αντίστοιχο ιστόγραμμα κατευθύνσεων, με βάρη ψήφων το μέτρο της βαθμίδας κάθε εικονοστοιχείου, καταλήγουμε με ένα σύνολο ιστογραμμάτων διευθύνσεων, ονομαζόμενων **κυττάρων** (cells). Για να καταστήσουμε τον παρόντα περιγραφητή αναλλοίωτο στη μεταβολή φωτεινότητας ή αντίθεσης της εικόνας, σχηματίζονται ομάδες γειτονικών κυττάρων, γνωστές ως **blocks**, οι οποίες υφίστανται κανονικοποίηση ως προς το σύνολο των ψήφων που συνθέτουν κάθε block. Το σύνολο όλων των κανονικοποιημένων ιστογραμμάτων, τα οποία έχουν συνδεθεί μεταξύ τους έτσι, ώστε να σχηματίσουν ένα διάνυσμα $H \in \mathbb{R}^{NM}$ όπου N ο αριθμός των ιστογραμμάτων και M ο αριθμός των κάδων ανά ιστόγραμμα, είναι ο *HOG* περιγραφητής.

2.2 Μηχανική Μάθηση

Αυτή η ενότητα αφορά κυρίως το κομμάτι του αλγορίθμου, στο οποίο γίνεται η αναγνώριση της κίνησης και η ταξινόμηση.

2.2.1 Support Vector Machine (SVM)

Η Μηχανή Διανυσμάτων Υποστήριξης (SVM) αποτελεί έναν από τους πιο διαδεδομένους ταξινομητές. Η λογική που διέπει αυτήν την υπολογιστική οντότητα είναι κυρίως εργασία των Vapnik et. al [8]. Ο πιο απλός τύπος SVM είναι ο γραμμικός.

2.2.1.1 Μαθηματική Περιγραφή Γραμμικού SVM

Για χάριν απλότητας, χρησιμοποιούμε τον συντελεστή i για να απαριθμήσουμε τα δείγματα και τον συντελεστή j για τα χαρακτηριστικά. Ορίζουμε ως ποσοτικό χώρο χαρακτηριστικών

τον δειγματικό υπερχώρο Ω , ο οποίος αποτελεί το καρτεσιανό γινόμενο των δειγματικών χώρων $\Omega_j, j = 1 \dots n$, ποσοτικών μεταβλητών. Εξ ορισμού, ο χώρος αυτός περιέχει τα δείγματα $x_i, i = 1 \dots m$, τα οποία χαρακτηρίζονται από τα χαρακτηριστικά $x_j, j = 1 \dots n$, με $x_j \in \Omega_j$. Σε κάθε δείγμα από αυτά ανατίθεται μία τιμή $y_k, k = 1, 2$ (label), η οποία χαρακτηρίζει την κατηγορία (κλάση) στην οποία ανήκει το δείγμα αυτό. Η διαδικασία αυτή χωρίζει τα δείγματα σε δύο κατηγορίες $Y_k, k = 1, 2$. Ο στόχος του γραμμικού SVM είναι η εύρεση του υπερεπιπέδου $S \subset \Omega$, το οποίο περιγράφεται από την ακόλουθη εξίσωση :

$$S : \mathbf{w}\mathbf{x} - b = 0$$

και διαχωρίζει βέλτιστα τα δείγματα των δύο κατηγοριών. Ο όρος ‘βέλτιστα’ μπορεί να τυποποιηθεί μαθηματικά, αν θεωρήσουμε ότι 2 υπερεπίπεδα S^+ και S^- , με την ίδια κλίση με το S , αντιπαράλληλα όμως ως προς αυτό, σε κάποια δεδομένη απόσταση d , η οποία συνήθως τίθεται ίση με $\frac{1}{\|\mathbf{w}\|}$, δηλαδή:

$$S^+ : \mathbf{w}\mathbf{x} - b = 1$$

$$S^- : \mathbf{w}\mathbf{x} - b = -1$$

περιγράφουν τα όρια των δειγμάτων της αντίστοιχης κατηγορίας και απαιτήσουμε να βρεθούν οι όροι w και b έτσι, ώστε να μεγιστοποιείται η απόσταση μεταξύ τους. Αν υποθέσουμε ότι οι τιμές του Y είναι οι $\{-1, 1\}$, τότε η συνθήκη που πρέπει να ικανοποιούν τα επίπεδα εκφράζεται ως:

$$\left. \begin{array}{l} \mathbf{w}\mathbf{x}_i - b \geq 1, \quad y_i = 1 \\ \mathbf{w}\mathbf{x}_i - b \leq -1, \quad y_i = -1 \end{array} \right\} \Leftrightarrow y_i (\mathbf{w}\mathbf{x}_i - b) \geq 1, \forall i \in [1, m] \quad (2.4)$$

Η μεγιστοποίηση της απόστασης των δύο επιπέδων μεταφράζεται σε ελαχιστοποίηση του $\|\mathbf{w}\|$:

$$\text{minimize } \|\mathbf{w}\| \quad \text{subject to } y_i (\mathbf{w}\mathbf{x}_i - b) \geq 1 \quad \text{for } i = 1 \dots n \quad (2.5)$$

Τα διανύσματα x_i τα οποία βρίσκονται σε μικρότερη απόσταση από τα υπερεπίπεδα και πρακτικά συμμετέχουν στον καθορισμό των παραμέτρων τους ονομάζονται διανύσματα υποστήριξης (support vectors).

Σε περίπτωση που τα δείγματα των δύο κατηγοριών δεν είναι γραμμικά διαχωρίσιμα, τότε η συνθήκη (2.4) δεν μπορεί να ισχύσει. Για αυτό το λόγο, στη γενική περίπτωση θεωρούμε ως ακόμα μία συνθήκη (εισάγοντας σε αυτήν τους περιορισμούς) είτε τη μη γραμμική συνάρτηση

‘hinge loss’:

$$\max(0, 1 - y_i(\mathbf{w}\mathbf{x}_i - b))$$

είτε άλλου τύπου συναρτήσεις, όπως την ‘square-loss’:

$$\|\mathbf{y}\mathbf{w} - \mathbf{x}\|^2$$

στη σχέση (2.5).

2.2.1.2 Επίλυση προβλήματος ελαχιστοποίησης

Το πρόβλημα που καλείται να επιλύσει το SVM βάσει των παραπάνω, με χρήση της συνάρτησης ‘hinge loss’, είναι το ακόλουθο:

$$\text{minimize } \frac{1}{m} \sum_{i=1}^m \{\max(0, 1 - y_i(\mathbf{w}\mathbf{x}_i - b))\} + C\|\mathbf{w}\|^2 \quad (2.6)$$

όπου ο συντελεστής C καθορίζει το πού εστιάζει η ελαχιστοποίηση, στο να βρίσκονται όλα τα δείγματα της κάθε κατηγορίας στη σωστή πλευρά των επιπέδων, είτε στο να μεγιστοποιείται η απόσταση των επιπέδων. Το πρόβλημα (2.6) επιδέχεται πολλούς τρόπους λύσεων, αφού αποδεικνύεται ότι είναι κυρτό, δηλαδή υπάρχει λύση που οδηγεί σε τοπικό ελάχιστο και είναι μοναδική, δηλαδή το τοπικό ελάχιστο είναι και ολικό. Ένας από τους πιο διαδεδομένους τρόπους επίλυσης είναι αυτός στον οποίο χρησιμοποιούνται οι **Πολλαπλασιαστές Lagrange** (Lagrangian Multipliers). Θεωρώντας ως συνθήκη τη ‘hinge loss’, αν θέσουμε για κάθε $i = 1 \dots m$ μια μεταβλητή $\zeta_i = \max(0, 1 - y_i(\mathbf{w}\mathbf{x}_i - b))$, τότε το πρόβλημα ελαχιστοποίησης γίνεται:

$$\begin{aligned} \text{minimize } & \frac{1}{m} \sum_{i=1}^m \zeta_i + C\|\mathbf{w}\|^2 \\ \text{subject to } & y_i(\mathbf{w}\mathbf{x}_i - b) \geq 1 - \zeta_i \text{ and } \zeta_i \geq 0, \quad \forall i \in \{1, \dots, m\} \end{aligned} \quad (2.7)$$

Οι πολλαπλασιαστές Lagrange, αυξάνοντας τις μεταβλητές του προβλήματος και άρα τη διαστατικότητα του, το κάνουν υπολογιστικά πιο απλό. Η απλή μορφή, η οποία προκύπτει από την εφαρμογή των πολλαπλασιαστών Lagrange, ονομάζεται δυαδικό πρόβλημα (Lagrange Dual) και

είναι η ακόλουθη:

$$\begin{aligned} & \text{maximize} && f(\lambda_1, \dots, \lambda_m) = \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i=1}^m \sum_{k=1}^m \{y_i \lambda_i (\mathbf{x}_i \mathbf{x}_k) y_k \lambda_k\} \\ & \text{subject to} && \sum_{i=1}^m \{\lambda_i y_i\} = 0 \text{ and } 0 \leq \lambda_i \leq \frac{1}{2mC} \quad \forall i \in \{1, \dots, m\} \end{aligned} \quad (2.8)$$

Το παραπάνω πρόβλημα μπορεί να λυθεί ταχύτερα σε σχέση με το πρόβλημα (2.7) με χρήση αριθμητικών μεθόδων επίλυσης διαφορικών εξισώσεων πρώτου βαθμού, όπως η ‘**Truncated Conjugate Gradient Method**’, η οποία αναλύεται στο [54].

2.2.2 Random Decision Forest (RDF)

Το Τυχαίο Δάσος Απόφασης (Random Decision Forest) έχει πάρει το όνομά του από το γεγονός ότι χρησιμοποιεί ένα πλήθος **Δέντρων Απόφασης** (Decision Trees), το καθένα από τα οποία εκπαιδεύεται με ένα τυχαίο υποσύνολο δειγμάτων εκπαίδευσης. Το υποσύνολο αυτό έχει υποστεί τη δράση της μεθόδου **Bootstrapping**, η οποία αναλύεται στη συνέχεια, ενώ το υποσύνολο αυτό έχει λιγότερο αριθμό χαρακτηριστικών από το αρχικό σύνολο, με τα χαρακτηριστικά να έχουν επίσης δειγματοληπτηθεί τυχαία. Κατόπιν, η απόφαση του συνολικού δάσους είναι ίση με την απόφαση πλειοψηφίας των δέντρων. Ο ταξινομητής αυτός εφευρέθηκε από τον Tim Kam Ho [14], ενώ έχει βασιστεί ιδιαίτερος στην εργασία πάνω στα Δέντρα Απόφασης των Breiman et. al [5].

2.2.2.1 Decision Trees (DT)

Τα Δέντρα Απόφασης (DT) αποτελούν ένα δημιουργήμα fractal μη γραμμικής στατιστικής ανάλυσης. Ο στόχος είναι να κατασκευαστεί ένα μοντέλο, το οποίο, χρησιμοποιώντας συγκριτικές μεθόδους, να μπορεί να αποφανθεί για την κατηγορία, στην οποία ανήκει ένα δείγμα, βασισμένο σε κάποια άλλα δείγματα, τα οποία ονομάζονται δείγματα εκπαίδευσης. Στη συνέχεια θα ασχοληθούμε με μία υποκατηγορία των DT, η οποία αφορά δείγματα με το πολύ δύο κατηγορίες.

Η περιγραφή της γεωμετρίας του δέντρου μπορεί να γίνει με χρήση αναδρομής: κάθε δέντρο υπηρετεί συνήθως την top-down λογική, στην οποία υπάρχει ένας αρχικός κόμβος, ο οποίος λειτουργεί ως είσοδος για τα δεδομένα. Από αυτόν εκκινούν 2 κλάδοι, οι οποίοι συνδέονται σε 2 νέες οντότητες. Η νέα οντότητα μπορεί να είναι είτε νέος κόμβος είτε ένα *σημείο τερματισμού*,

στο οποίο λαμβάνεται μια απόφαση. Η διαδικασία αυτή επαναλαμβάνεται αναδρομικά για κάθε κόμβο. Το *βάθος τερματισμού*, δηλαδή το πόσοι κόμβοι υπάρχουν ανά μονοπάτι, δεν είναι σταθερό. Συνέπεια αυτής της αναδρομικής διαδικασίας είναι η δημιουργία ενός μη ισορροπημένου δυαδικού δέντρου.

Κάθε κόμβος του δέντρου αποτελεί και ένα *σημείο σύγκρισης*, στο οποίο γίνεται σύγκριση κάποιου χαρακτηριστικού f του εισερχόμενου δείγματος με μία τιμή v και, αναλόγως το αποτέλεσμα, το δείγμα ‘μεταβαίνει’ στον αντίστοιχο κλάδο, όπου είτε ‘φτάνει’ σε ένα σημείο τερματισμού, ή σε ένα νέο σημείο σύγκρισης. Για να κατασκευαστεί αρχικά το δέντρο, λαμβάνονται κάποια κριτήρια, τα οποία καθορίζουν το βάθος τερματισμού και το ποια χαρακτηριστικά θα συγκριθούν με ποια τιμή ανά κόμβο. Γενικά το πρόβλημα της κατασκευής του βέλτιστου δέντρου έχει την υπολογιστική δυσκολία των **NP-complete** προβλημάτων. Με την εφαρμογή της δημοφιλούς μεθόδου CART (Classification And Regression Tree), χρησιμοποιείται για τη προσέγγιση του βέλτιστου διαχωρισμού ανά κόμβο (splitting) το κριτήριο προσμίξεων Gini (Gini Impurity). Για να καθοριστεί πότε και αν χρειάζεται να γίνει splitting χρησιμοποιούνται διάφοροι αλγόριθμοι, εκ των οποίων αναλύεται ο πλέον χρησιμοποιήσιμος, γνωστός ως Αλγόριθμος του Hunt (Hunt Algorithm), εφεύρεση των Hunt et. al [15], ο οποίος ανήκει στην κατηγορία των **greedy search** αλγορίθμων, που ευριστικά ανιχνεύουν μια μη μοναδική υποβέλτιστη λύση σε κάποιο πρόβλημα αναζήτησης [7].

2.2.2.1.1 Hunt Algorithm Έστω D_t ένας πίνακας με γραμμές τα δείγματα εκπαίδευσης τα οποία ‘προσεγγίζουν’ το σημείο σύγκρισης t . Η αναδρομική διαδικασία που ακολουθείται ορίζεται ως εξής:

- Αν το D_t περιέχει δείγματα τα οποία ανήκουν στην ίδια κλάση y_t , τότε το σημείο σύγκρισης t γίνεται σημείο απόφασης για την κατηγορία y_t .
- Αν το D_t είναι ένα κενό σύνολο, τότε το t είναι ένα σημείο απόφασης για μία μη προσδιορισμένη κλάση y_d .
- Αν το D_t περιέχει δείγματα που ανήκουν σε παραπάνω από μία κλάσεις, τότε χρησιμοποιείται μια ευριστική μέθοδος με κάποιο κριτήριο splitting για να δημιουργηθεί ένα νέο σημείο σύγκρισης και να χωριστεί το D_t σε μικρότερα υποσύνολα.

2.2.2.1.2 Gini Impurity Το κριτήριο αυτό εκφράζεται από την ακόλουθη σχέση:

$$G = \sum_{k=1}^{n_c} p_k(1 - p_k)$$

όπου G είναι ο δείκτης Gini, p_k το ποσοστό των δειγμάτων του D_t , όπως αυτό ορίζεται παραπάνω, που ανήκει στην κλάση k και n_c ο αριθμός των κλάσεων.

Χρησιμοποιώντας το κριτήριο Gini, η πιο ευρέως χρησιμοποιούμενη ευριστική μέθοδος για το splitting είναι η ακόλουθη:

1. Απομονώνουμε με τη σειρά τις στήλες του D_t και χωρίζουμε σε όλα τα πιθανά σημεία την στήλη υπό εξέταση, λαμβάνοντας ζεύγη υποσυνόλων.
2. Εφαρμόζουμε το κριτήριο Gini και εξάγουμε το μέσο όρο σε κάθε ζεύγος.
3. Επιλέγουμε ως καλύτερο ζεύγος αυτό με το μικρότερο δείκτη Gini και αποθηκεύουμε τον δείκτη ως b_j , συγκρατώντας την τιμή στην οποία χωρίσαμε το D_t για να προκύψει το εν λόγω ζεύγος ως v_j .
4. Όταν ολοκληρώσουμε την προσπέλαση όλων των στηλών, λαμβάνουμε ως συνθήκη για τον κόμβο t την ακόλουθη:

$$f_t(x_k) = x_k > v_k, \quad k = \operatorname{argmin} b_j$$

με το x_k να υποδηλώνει το k -οστό χαρακτηριστικό του εισερχόμενου δείγματος.

Το πιο συνηθισμένο κριτήριο τερματισμού της παραπάνω διαδικασίας είναι να σταματήσει όταν το D_t έχει λιγότερα από μ δείγματα.

Αφού ολοκληρωθεί η εκπαίδευση, χρησιμοποιείται συνήθως και η μέθοδος **pruning**, κατά την οποία αφαιρούνται σημεία απόφασης, τα οποία αυξάνουν τον συνολικό δείκτη Gini του δέντρου απόφασης.

2.2.2.2 Bootstrapping Aggregating (Bagging)

Δεδομένου ενός συνόλου εκπαίδευσης D_t , μεγέθους m , ορίζουμε ως Bagging (**B**ootstrap **a**ggregating) τη μέθοδο, η οποία παράγει d καινούργια σύνολα εκπαίδευσης D_{t_k} , $k = 1, \dots, d$

μεγέθους m' , δειγματοληπτώντας το σύνολο D_t ομοιόμορφα, με επανατοποθέτηση. Δειγματοληπτώντας με επανατοποθέτηση, οι παρατηρήσεις του D_t μπορούν να επαναληφθούν σε κάθε σύνολο D_{t_i} . Πρωτοπόρος αυτής της μεθόδου είναι ο Breiman [4].

2.2.3 Sparse Coding

Η Αραιή Κωδικοποίηση (Sparse Coding) είναι μια διαδικασία, κατά την οποία ένα σήμα ξ διάστασης N μετατρέπεται σε ένα αραιό (δηλαδή με μεγάλο πλήθος από μηδενικά) σήμα διαστάσεων M , με $M > N$, χρησιμοποιώντας ένα **λεξικό** B , διάστασης $N \times M$, με την ελάχιστη απώλεια της αρχικής πληροφορίας, με την εξής σχέση:

$$\xi = Bs$$

Το B είναι μία υπερπλήρης βάση ως προς την είσοδο, δηλαδή οι στήλες του είναι γραμμικώς ανεξάρτητες και περισσότερες από τη διάσταση της εισόδου. Μέσω της αραιής κωδικοποίησης, επιδιώκεται να βρεθούν σχέσεις υψηλότερων διαστάσεων μεταξύ των χαρακτηριστικών της εισόδου, οι οποίες να προσομοιάζουν την αναπαράσταση αυτών από το ανθρώπινο αισθητήριο σύστημα. Με την αύξηση της διάστασης, τα δείγματα απεικονίζονται σε έναν νέο δειγματικό υπερχώρο, στον οποίο υπάρχει μεγαλύτερη πιθανότητα οι κατηγορίες αυτών να είναι γραμμικώς διαχωρίσιμες [21].

Η μαθηματική τυποποίηση του παραπάνω προβλήματος είναι η ελαχιστοποίηση του ακόλουθου τετραγωνικού προβλήματος:

$$\text{minimize } \|\xi - Bs\|_F^2 \quad (2.9)$$

Στόχος είναι να βρεθεί και το αραιό διάνυσμα s αλλά και το λεξικό (ή κώδικας) B . Λόγω της ύπαρξης μη επαρκών περιορισμών για το πρόβλημα, δεν υπάρχει κυρτή λύση. Παρ'όλ'αυτά, αν επιλέξουμε να λύσουμε το πρόβλημα επαναληπτικά, κρατώντας σταθερή τη μία μεταβλητή και λύνοντας ως προς την άλλη, αποδεικνύεται ότι μπορούμε να προσεγγίσουμε το ολικό ελάχιστο. Η απαίτηση να είναι το s αραιό επιβάλλει την ελαχιστοποίηση της ενέργειάς του (L_2 νόρμας) ή του αθροίσματος των απολύτων των στοιχείων του (L_1 νόρμας). Έχει αποδειχτεί στο [32] ότι η επιλογή της L_1 νόρμας οδηγεί σε αραιότερες λύσεις. Από την άλλη πλευρά, για να είναι ο B ένας αραιός πίνακας με στοιχεία με μικρή απόλυτη τιμή και οι στήλες του να παραμένουν γραμμικώς ανεξάρτητες, το προτεινόμενο μοτίβο είναι η χρήση συστηματικοποίησης με την L_2 νόρμα (L_2 -regularization), διαμήκους του χώρου των γραμμών (row space) του πίνακα. Με μαθηματικούς

όρους, τα δύο προαναφερόμενα υποπροβλήματα αναφέρονται παρακάτω:

$$\begin{aligned} &\text{Solve for } \mathbf{s} \text{ while } \mathbf{B} \text{ is constant:} \\ &\text{Minimize } \|\boldsymbol{\xi} - \mathbf{B}\mathbf{s}\|_F^2 + \gamma\|\mathbf{s}\|_1, \text{ with } \gamma \text{ constant} \end{aligned} \quad (2.10a)$$

$$\begin{aligned} &\text{Solve for } \mathbf{B} \text{ while } \mathbf{s} \text{ is constant:} \\ &\text{Minimize } \|\boldsymbol{\xi} - \mathbf{B}\mathbf{s}\|_F^2 \\ &\text{subject to } \sum_{i=1}^k \mathbf{B}_{i,j}^2 \leq c, \text{ for } j = 1, \dots, n \end{aligned} \quad (2.10b)$$

Το γ , το οποίο εισάγεται στη σχέση 2.10a είναι μία σταθερά, της οποίας η τιμή καθορίζει έμμεσα το πλήθος των μηδενικών του \mathbf{s} . Όσο πιο αραιό είναι το διάνυσμα, τόσο πιο αυξημένο αναμένεται να είναι το σφάλμα ανακατασκευής.

2.2.3.0.1 Επίλυση του Προβλήματος 2.10a Το πρόβλημα αυτό ανήκει στην κατηγορία των προβλημάτων ελάχιστων τετραγώνων (βλ. Παρ. B.3), με την προσθήκη L_1 -**συστηματικοποίησης** (L_1 -regularization). Για να επιλυθεί το πρόβλημα αυτό, χρησιμοποιούμε τον ευριστικό αλγόριθμο Feature-sign search [21, 53](βλ. Παρ. A.2). Ο αλγόριθμος αυτός μπορεί να παραγάγει με υψηλή ταχύτητα λύσεις, ‘μαντεύοντας’ τα πρόσημα των στοιχείων του \mathbf{s} και ελέγχοντας αν αυτή η εκτίμηση μειώνει ή όχι την αντικειμενική συνάρτηση, ή αλλιώς το σφάλμα ανακατασκευής. Η ταχύτητα του αλγορίθμου είναι από κατασκευής εξαρτημένη από την αραιότητα του διανύσματος \mathbf{s} , οπότε, αν μόνο κάποιες τιμές του \mathbf{s} απαιτούνται για να ελαχιστοποιήσουν το τετραγωνικό πρόβλημα, ο αλγόριθμος εκτελείται ταχύτατα.

2.2.3.0.2 Επίλυση του Προβλήματος 2.10b Το πρόβλημα της εκπαίδευσης του πίνακα \mathbf{B} είναι ουσιαστικά και το μη επιτηρούμενο (non-supervised) στάδιο εκπαίδευσης του αλγορίθμου Sparse Coding. Υποθέτοντας ότι διαθέτουμε έναν πίνακα $\boldsymbol{\Xi}$ με γραμμές τα δείγματα εισόδου, τότε με χρήση του αλγορίθμου Feature-sign search για κάθε γραμμή του $\boldsymbol{\Xi}$ λαμβάνουμε, χρησιμοποιώντας τυχαία αρχικοποίηση για τον πίνακα \mathbf{B} , τον πίνακα εξόδου \mathbf{S} . Η σχέση 2.10b μπορεί να αναπτυχθεί έτσι στην διάσταση των στηλών, αποκτώντας την ακόλουθη μορφή:

$$\begin{aligned} &\text{Solve for } \mathbf{B} \text{ while } \mathbf{S} \text{ is constant:} \\ &\text{Minimize } \|\boldsymbol{\Xi} - \mathbf{B}\mathbf{S}\|_F^2 \\ &\text{Subject to } \sum_{i=1}^k \mathbf{B}_{i,j}^2 \leq c, \text{ for } j = 1, \dots, n \end{aligned} \quad (2.11)$$

Εφαρμόζοντας την μέθοδο των **Πολλαπλασιαστών Lagrange**, μπορούμε να εξαγάγουμε το δυικό πρόβλημα (Lagrange Dual), το οποίο είναι ευκολότερο να λυθεί, λόγω αύξησης της διαστατικότητας των μεταβλητών:

$$D(\vec{\lambda}) = \text{trace}(\Xi^T \Xi - \Xi S^T (S S^T + \Lambda)^{-1} (\Xi S^T)^T - c \Lambda) \quad (2.12)$$

όπου $\vec{\lambda}$ είναι το διάνυσμα των πολλαπλασιαστών Lagrange και $\Lambda = I_N \vec{\lambda}$, με I_N τον μοναδιαίο πίνακα διαστάσεων $N \times N$. Η Συνάρτηση 2.12 έχει μερικές παραγώγους πρώτου και δευτέρου βαθμού ως προς $\lambda_i, i = 1, \dots, N$, οι οποίες μπορούν να υπολογιστούν. Από αυτές συμπεραίνεται ότι το πρόβλημα ελαχιστοποίησης είναι όντως κυρτό, για δεδομένο S . Για να επιλυθεί αριθμητικά το πρόβλημα ελαχιστοποίησης μπορούν να χρησιμοποιηθούν διάφορες αριθμητικές μέθοδοι, όπως η ‘Nonlinear Conjugate Gradient’, επίσης γνωστή ως ‘Truncated Newton’ [30, 33].

2.3 Μικτά Θέματα

Στην ενότητα αυτή αναλύεται μία συστάδα θεμάτων, τα οποία ανήκουν και στους δύο ερευνητικούς τομείς, τη Μηχανική Μάθηση και την Όραση Υπολογιστών.

2.3.1 Ανίχνευση και Οριοθέτηση Κινούμενου Αντικειμένου σε Στατικό Υπόβαθρο

Μία από τις πιο πρακτικές εργασίες στην ανάλυση μιας εικονοσειράς (video) αποτελεί η ανίχνευση κινούμενου αντικειμένου. Το πρόβλημα αυτό είναι ιδιαίτερα απαιτητικό, γιατί το αντικείμενο, το οποίο είναι επιθυμητό να ανιχνευθεί, κινείται σε έναν μη στατικό χρωματικά χώρο, με ιδιάζουσες μεταβολές αντίθεσης και φωτεινότητας, ακόμα και όταν το σημείο λήψης παραμένει στατικό. Αυτές οι μη γραμμικές μεταβολές της έντασης των εικονοστοιχείων είναι αποτέλεσμα διαφόρων παραγόντων, όπως είναι η ίδια κίνηση του αντικειμένου, το οποίο μπορεί να προκαλεί σκίαση στο χώρο ή μεταβολή της συνολικής έντασης του φωτός που δέχεται ο αισθητήρας, ή ο θόρυβος από ατέλειες στη λήψη.

Λόγω των παραπάνω, μία πληθώρα αλγορίθμων έχει παραχθεί, για να προσεγγίσει αυτό το ζήτημα. Μία από τις πιο χρήσιμες μεθόδους, εμπνευσμένη από τον Zoran Zunic [56], η οποία έχει πολύ μικρή υπολογιστική πολυπλοκότητα σε σχέση με αντίστοιχες μεθόδους, ονομάζεται

Improved Adaptive Gaussian Mixture Model for Background Subtraction και για χάρην συντομίας θα καλείται **MOG2**, όπως είναι η κωδική ονομασία της στη βιβλιοθήκη της OpenCV.

2.3.1.1 Gaussian Mixture Model (GMM)

Η βασική λειτουργία που διέπει τη MOG2 είναι η κατασκευή ενός **Μοντέλου Μίξης Γκαουσιανών Κατανομών** (GMM). Με αυτό επιδιώκεται να μοντελοποιηθεί η πιθανότητα να ανήκει ένα εικονοστοιχείο ενός frame στο προσκήνιο μιας εικονοσειράς, δεδομένης της εκθετικά σταθμισμένης χρονικής μέσης έντασης των εικονοστοιχείων σε ένα τρέχον χρονικό παράθυρο. Στο παράθυρο αυτό ως προσκήνιο θεωρείται ένα χωρικό μέρος αυτού, χαρακτηριζόμενο από μεταβολές έντασης, οι οποίες συνήθως προκαλούνται από κίνηση αντικειμένων της σκηνής.

Ένα GMM με M στοιχεία είναι ουσιαστικά ο γραμμικός συνδυασμός M κανονικών κατανομών και έχει την ακόλουθη μορφή στο χώρο \mathbb{R}^N :

$$G(\mathbf{x}) = \sum_{m=1}^M \pi_m \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_m; \boldsymbol{\Sigma}_m) \quad (2.13)$$

όπου:

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}; \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{N/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

είναι η N -διάστατη κανονική κατανομή (με N τη διάσταση του δείγματος), με μέση τιμή $\boldsymbol{\mu} \in \mathbb{R}^N$ και **Πίνακα Συνδιακύμανσης** $\boldsymbol{\Sigma} \in S^{N \times N}$ και π_m είναι οι συντελεστές του γραμμικού συνδυασμού. Υποθέτοντας για απλούστευση ότι ο πίνακας συνδιακύμανσης κάθε κανονικής κατανομής $\boldsymbol{\Sigma}_m$ είναι διαγώνιος, δηλαδή ότι τα στοιχεία του διανύσματος x είναι στατιστικώς ανεξάρτητα, και ακόμη ότι η τυχαία απόκλιση σ_i είναι σταθερή ανά στοιχείο x_i του x , τότε καταλήγουμε στην απλοποιημένη μορφή:

$$G(\mathbf{x}) = \sum_{m=1}^M \pi_m \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_m; \sigma_m^2 \mathbf{I}_N) \quad (2.14)$$

¹Το σύμβολο $S^{n \times n}$ χρησιμοποιείται για να αναφερθεί ο χώρος των συμμετρικών και θετικά ορισμένων $n \times n$ πινάκων, οι οποίοι ορίζονται ως:

$$S^{n \times n} = \left\{ \mathbf{A} \in \mathbb{R}^{n \times n} : \mathbf{A} = \mathbf{A}^T \text{ και } \mathbf{x}^T \mathbf{A} \mathbf{x} > 0 \quad \forall \mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\} \right\}$$

Στην εν λόγω περίπτωση ο πίνακας συνδιακύμανσης $\boldsymbol{\Sigma}_m$ συνδέεται με τη μέση τιμή ως εξής:

$$\boldsymbol{\Sigma}_m = E[\mathbf{x} \mathbf{x}^T] - \boldsymbol{\mu}_m \boldsymbol{\mu}_m^T$$

όπου \mathbf{I}_N είναι ο μοναδιαίος N -διάστατος πίνακας.

2.3.1.1.1 Εκπαίδευση Η μη επιβλεπόμενη εκπαίδευση του GMM εκτελείται με χρήση μιας επαναληπτική μεθόδου, κατά την οποία γίνεται εκτίμηση των παραμέτρων του μοντέλου μέσω των εισόδων. Έστω η μορφή του μοντέλου 2.13 με τη χρήση εκτιμήσεων των παραμέτρων:

$$G(\mathbf{x}) = \sum_{m=1}^M \hat{\pi}_m \mathcal{N}(\mathbf{x}; \hat{\boldsymbol{\mu}}_m; \hat{\sigma}_m \mathbf{I}_N)$$

Οι εκτιμήσεις ανανεώνονται βάσει του ακόλουθου σχήματος, με είσοδο το δείγμα $\vec{x}^{(t)}$:

$$\begin{aligned} \hat{\pi}_m &\leftarrow \hat{\pi}_m + \alpha \left(o_m^{(t)} - \hat{\pi}_m \right) - \alpha c_T \\ \hat{\boldsymbol{\mu}}_m &\leftarrow \hat{\boldsymbol{\mu}}_m + o_m^{(t)} \left(\frac{\alpha}{\hat{\pi}_m} \right) \boldsymbol{\delta}_m \\ \hat{\sigma}_m^2 &\leftarrow \hat{\sigma}_m^2 + o_m^{(t)} \left(\frac{\alpha}{\hat{\pi}_m} \right) (\boldsymbol{\delta}_m^T \boldsymbol{\delta}_m - \hat{\sigma}_m^2) \end{aligned} \quad (2.15)$$

όπου:

- $\boldsymbol{\delta}_m = \mathbf{x}^{(t)} - \hat{\boldsymbol{\mu}}_m$,
- α είναι η σταθερά που περιγράφει τα βάρη της εκθετικής περιβάλλουσας του χρονικού παραθύρου μήκους T , με $w_{t_i} = \exp(-\alpha(T - t_i))$ το βάρος του δείγματος $x^{(t_i)}$ της t_i -στης θέσης του παραθύρου,
- $c_T = c/T$ είναι μία αυθαίρετη σταθερά η οποία εκφράζει την πρότερη πιθανότητα (a priori) να ανήκει ένα δείγμα $\vec{x}^{(t)}$ σε κάποια από τις κατανομές του μοντέλου,
- $o_m^{(t)}$ είναι μία μεταβλητή, η οποία παίρνει τις τιμές $\{0, 1\}$, βάσει της απόστασης Mahalanobis από το κέντρο της κατανομής $\mathcal{N}(\mathbf{x}; \hat{\boldsymbol{\mu}}_m; \hat{\sigma}_m \mathbf{I}_N)$ αλλά και του μεγέθους του αντίστοιχου π_m . Πιο συγκεκριμένα, ισχύει το παρακάτω:

$$o_m^{(t)} = \begin{cases} 1, & \text{if and only if } m = \operatorname{argmax}_j (\hat{\pi}_j) \forall j \in S, S = \left\{ j : D_j > r \hat{\sigma}_j, D_j(\mathbf{x}^{(t)}) = \sqrt{\frac{\boldsymbol{\delta}_j^T \boldsymbol{\delta}_j}{\hat{\sigma}_j^2}} \right\} \\ 0, & \text{otherwise} \end{cases}$$

όπου το r τίθεται αυθαίρετα και δείχνει την ανοχή στην απόσταση.

Μετά από κάθε ανανέωση, τα π_m κανονικοποιούνται για να έχουν άθροισμα μονάδα. Σε περίπτωση που το σύνολο S είναι κενό, τότε προστίθεται μια καινούργια κανονική κατανομή στο

άθροισμα, με $\hat{\pi}_{M+1} = \alpha$, $\hat{\boldsymbol{\mu}}_{M+1} = \mathbf{x}^{(t)}$ και $\hat{\sigma}_{M+1} = \sigma_0$, όπου σ_0 είναι η αρχική αυθαίρετη τυχαία απόκλιση της κατανομής. Σε περίπτωση που ο αριθμός των κατανομών του μοντέλου έχει υπερβεί ένα μέγιστο πλήθος, βάσει αυθαίρετου περιορισμού, τότε αφαιρείται από το άθροισμα η κατανομή με το μικρότερο π_m .

Η πιθανότητα ενός εικονοστοιχείου $\mathbf{x}^{(t)}$ να ανήκει στο υπόβαθρο, η οποία εκφράζεται από το GMM, συμβολίζεται ως $p(\mathbf{x}|\mathcal{X}_T, BG)$, με \mathcal{X}_T να είναι ένας πίνακας $N \times T$ με στήλες τις τιμές του $\mathbf{x}^{(\tau)}$, $t - T < \tau < t$ και BG να συμβολίζει την έννοια του υποβάθρου (**B**ack**G**round).

2.3.1.2 Ανίχνευση αντικειμένου

Χρησιμοποιώντας το μοντέλο GMM μπορούμε να εξαγάγουμε το σύνολο των εικονοστοιχείων $\mathbf{x}^{(t)}$ που ανήκουν στο προσκήνιο, θεωρώντας ότι είναι αυτά για τα οποία ισχύει:

$$p(\mathbf{x}|\mathcal{X}_T, BG) \leq c_{thr}$$

όπου $c_{thr} \in [0, 1]$ μία αυθαίρετη τιμή. Τα εικονοστοιχεία αυτά τα ονομάζουμε για χάριν συντομίας *ενεργοποιημένα*, σε αντίθεση με τα εικονοστοιχεία του υποβάθρου, που είναι *απενεργοποιημένα*.

2.3.1.3 Ανίχνευση Σκιών

Τα απενεργοποιημένα pixel χρησιμοποιούνται για να εξαχθεί το μοντέλο χαρακτηρισμού της σκίασης, βάσει της μεθόδου των KaewTraKulPong et al. [18]. Η συγκεκριμένη μέθοδος χρησιμοποιεί τη μέση τιμή \mathbf{E} των απενεργοποιημένων pixel, η οποία μπορεί να εξαχθεί από το GMM, για να κατασκευάσει μία μετρική, με την οποία μπορεί κανείς να αποφανθεί αν κάποιο ενεργοποιημένο pixel \mathbf{x}_a της εικόνας είναι σκιά ή όχι. Η μετρική χαρακτηρίζεται από δύο μεγέθη, την *φωτεινή παραμόρφωση* α και την *χρωματική παραμόρφωση* c , οι οποίες ορίζονται ως ακολούθως:

$$\begin{aligned} \alpha &= \operatorname{argmin}_z \|\mathbf{x}_a - z\mathbf{E}\|^2 \\ c &= \|\mathbf{x}_a - \alpha\mathbf{E}\| \end{aligned} \quad (2.16)$$

Ένα ενεργοποιημένο pixel θεωρείται σκιά όταν το $\alpha \in [0, 2.5\sigma]$ και $\tau < c < 1$, όπου σ είναι η μέση τυπική απόκλιση των συνιστωσών κατανομών του GMM και τ αυθαίρετη τιμή η οποία ονομάζεται *κατώφλι φωτεινότητας*.

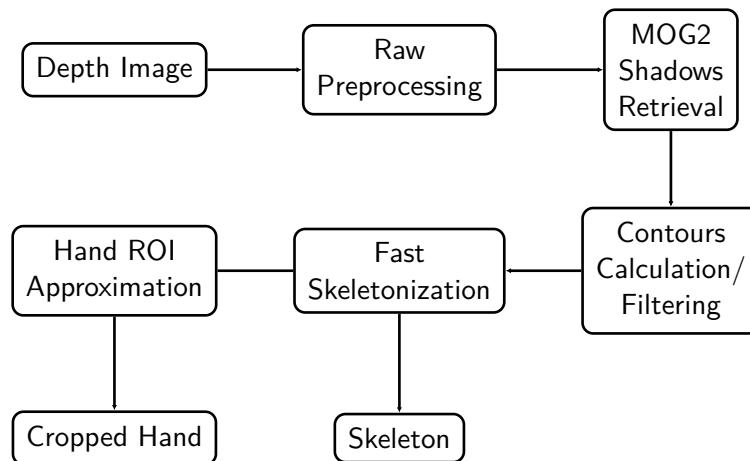
Κεφάλαιο 3

Περιγραφή Μεθοδολογίας

3.1 Γενικά

Όπως αναφέρθηκε και στην Εισαγωγή, η εργασία αυτή είναι μια προσπάθεια να προσεγγιστεί η δοκιμασία αναγνώρισης χειρονομιών, δυναμικών και παθητικών, χρησιμοποιώντας εξ ολοκλήρου το κανάλι βάθους ενός RGBD αισθητήρα, που στην προκείμενη περίπτωση είναι το Kinect v2. Ο προτεινόμενος αλγόριθμος περιέχει δύο στάδια, αυτό της εκπαίδευσης (το οποίο είναι *off-line*, δηλαδή γίνεται εκτός πραγματικού χρόνου) και αυτό της δοκιμής (*testing*). Κατά τη διάρκεια του σταδίου εκπαίδευσης εκπαιδεύονται με χρήση συνδυασμού μη επιβλεπόμενης και επιβλεπόμενης μάθησης τα χρησιμοποιούμενα μοντέλα. Στις επόμενες ενότητες αναλύονται οι υπομονάδες του αλγορίθμου με τρόπο έτσι, ώστε να μπορεί να παρακολουθήσει κανείς την πορεία της πληροφορίας από την αρχή ως το τέλος. Κάποιες από αυτές τις υπομονάδες, όπως η Αραιή Κωδικοποίηση ή η τεχνική PTPCA, παραλείπονται στην τελική υλοποίηση, όμως υποβάλλονται σε πειραματισμούς στο Κεφάλαιο 4, για αυτό και αναφέρονται στη συνέχεια του κεφαλαίου.

3.2 Προεπεξεργασία



Σχήμα 3.1: Στάδιο Προεπεξεργασίας (Preprocessing)

3.2.1 Εικόνα βάθους (Depth Image)

Η είσοδος του συστήματος είναι μια σειρά από εικόνες βάθους, λαμβανόμενες από τον αισθητήρα με ρυθμό δειγματοληψίας f_s . Η εικόνα βάθους θυμίζει ασπρόμαυρη δισδιάστατη εικόνα, μόνο που η ένταση κάθε εικονοστοιχείου έχει τιμή την υπολογιζόμενη απόσταση του προβελλόμενου σημείου από τον αισθητήρα. Στην περίπτωση του Kinect η απόσταση εκφράζεται σε χιλιοστά με ακέραια τιμή, δηλαδή η διακριτική ικανότητα του αισθητήρα είναι το 1 χιλιοστό. Σε περίπτωση απουσίας πληροφορίας χρησιμοποιείται η τιμή 0 από τον αισθητήρα. Η απουσία πληροφορίας προκαλείται είτε από λάθος της προσεγγιστικής μεθόδου που περιγράφεται στην Ενότητα 1.7.1 ή από το γεγονός ότι το σημείο ανίχνευσης βρίσκεται έξω από το πεδίο ανίχνευσης του αισθητήρα, το οποίο είναι περίπου το διάστημα $[400, 8000]$ mm για τον αισθητήρα του Kinect v2.

3.2.2 Αρχική Προεπεξεργασία (Raw Preprocessing)

Η εικόνα βάθους αναλύεται αρχικά έτσι, ώστε να ανιχνευθούν εικονοστοιχεία τα οποία είναι θορυβώδη και δεν συντρέχουν στη αποκομιδή κάποιας χρήσιμης πληροφορίας. Ως θορυβώδη χαρακτηρίζονται όσα pixels x , των οποίων η median ένταση $I_{med}(x)$ για τα πρώτα n_{init} frames ικανοποιεί την ακόλουθη συνθήκη:

$$|I_{med}(x) - I_{nnz}(x)| \geq th_{noise} \vee I_{nnz} = 0$$

όπου I_{nnz} η τελευταία ένταση του εικονοστοιχείου μέσα στο χρονικό παράθυρο των n_{init} frames και th_{noise} ένα κατώφλι, με το οποίο φιλτράρεται ο θόρυβος. Εικονοστοιχεία θορυβωδών περιοχών μικρής επιφάνειας λαμβάνουν πληροφορία από τα γειτονικά με επαναλαμβανόμενη χρήση ενός φίλτρου median, με συγκεκριμένο αριθμό επαναλήψεων n_{med} και με μέγεθος μάσκας s_{med} .

3.2.3 Ανίχνευση Σκιών (MOG2 Shadows Retrieval)

Η μέθοδος που περιγράφεται στην Ενότητα 2.3.1 για την ανίχνευση εφαρμόζεται βέλτιστα από τη βιβλιοθήκη της OpenCV, η οποία την ονοματίζει ως *MOG2*. Τα πρώτα n_{init} εικονοπλαίσια χρησιμοποιούνται για την αρχική ρύθμιση των βαρών της μεθόδου. Για το μοντέλο χρησιμοποιούμε M_{GMM} γκαουσιανές συνιστώσες και θέτουμε r_{GMM} κατώφλι για την Mahalanobis απόσταση, $c_{thr_{GMM}}$ κατώφλι για να βρούμε τα ενεργοποιημένα pixels και συγκρατούμε ένα σχετικά μεγάλο παράθυρο n_{hist} frames για την ανίχνευση, υποθέτοντας ότι το background είναι στατικό. Το μοντέλο ανανεώνεται από κάθε είσοδο με μη επιβλεπόμενο τρόπο.

Λόγω του ότι η είσοδος είναι εικόνα βάθους, μας ενδιαφέρει από τα ενεργοποιημένα εικονοστοιχεία να εξαγάγουμε αυτά τα οποία έχουν μικρή σχετικά διαφορά φωτεινότητας σε σχέση με το υπόβαθρο, θεωρώντας τα υπόλοιπα εικονοστοιχεία ως θόρυβο. Για αυτόν το λόγο, χρησιμοποιούμε τη μέθοδο ανίχνευσης σκιών και θεωρούμε ως κινούμενο αντικείμενο την ίδια την υπολογιζόμενη σκιά. Η προσέγγιση αυτή φέρει καλύτερα αποτελέσματα ανίχνευσης, από το αν θεωρούσαμε ως κινούμενο αντικείμενο τη μάσκα που προτείνει ο αλγόριθμος MOG2. Η εξαγόμενη μάσκα υφίσταται μια μετέπειτα επεξεργασία με τη δράση του μορφολογικού τελεστή **Opening**, τετραγωνικού πυρήνα μεγέθους s_{opker} .

3.2.4 Υπολογισμός Περιγράμματος (Contours Calculation/Filtering)

Η επεξεργασία της μάσκας των κινούμενων αντικειμένων με χρήση των αλγοριθμικών μεθόδων που περιγράφονται στην Ενότητα 2.1.2. Η ανίχνευση του περιγράμματος της Ενότητας 2.1.2.2 έχει υλοποιηθεί με τη συνάρτηση *findContours* της OpenCV. Κατόπιν της ανίχνευσης όλων των περιγραμμάτων, επιλέγεται ως μάσκα χεριού αυτή με το μεγαλύτερο εμβαδόν, με χρήση του τύπου Green.

3.2.5 Ταχεία Σκελετοποίηση (Fast Skeletonization)

Στόχος του βήματος αυτού είναι να βρεθεί ο βασικός άξονας των μερών του χεριού, δηλαδή ένα σύνολο συνδεδεμένων ακυκλικά ευθυγράμμων τμημάτων, τα οποία να εκκινούν από το κέντρο του σημείου εισόδου του χεριού στην εικόνα και να καταλήγουν στο ακρότατο σημείο του χεριού. Το βήμα αυτό γίνεται για τους ακόλουθους λόγους:

- Να απομονωθεί η παλάμη από το υπόλοιπο χέρι έτσι, ώστε να καθοριστεί όσο πιο επιτυχώς γίνεται η περιοχή ενδιαφέροντος (ROI) και να βελτιωθεί η ταχύτητα του αλγορίθμου.
- Να βρεθεί η κατεύθυνση της παλάμης.
- Να βρεθεί το τελικό σημείο δράσης του χεριού, για πρακτική εφαρμογή.

Σε αυτήν την κατεύθυνση επιστρατεύουμε έναν πρότυπο αλγόριθμο ταχείας σκελετοποίησης. Αυτός βρίσκει με επαναληπτικό τρόπο τον βασικό άξονα, χρησιμοποιώντας μια μέθοδο, η οποία, δεδομένου ενός σημείου εκκίνησης μέσα στο περιγράμμα, βρίσκει το σημείο του περιγράμματος, με το οποίο μπορούμε να κατασκευάσουμε το μέγιστου μήκους ευθύγραμμο τμήμα, το οποίο να μην παρεμβάλλεται του περιγράμματος και να βρίσκεται μέσα σε αυτό. Η μέθοδος είναι η ακόλουθη:

Algorithm 1 Εύρεση μακρύτερης, μη παρεμβαλλόμενης γραμμής με το εξωτερικό περιγράμμα, δεδομένου ενός αρχικού σημείου εντός του περιγράμματος

```

1: procedure FLLIC(cnt, p)                                ▷ cnt : contour points  $\in \mathbb{R}^{Nx2}$ , p : point  $\in \mathbb{R}^2$ 
2:   polar  $\leftarrow$  cart2polar(cnt, p)                    ▷ Polar transformation with p as reference point
3:   H  $\leftarrow$  hist2d(polar, R, PHI)                     ▷ 2D Histogram of polar coordinates
4:   H  $\leftarrow$  H > 0
5:   a  $\leftarrow$  ( $\sum_{r-axis} H$ ) > 2 ▷ Find the digitized angles where line intersects multiple times
   with contour
6:   H[:, a] = 0                                           ▷ Zero out all corresponding columns
7:   candindsr  $\leftarrow$  argmaxr(H)
8:   indphi  $\leftarrow$  argmax(candindsr) ▷ Find the longest intersection point of the nonzero
   columns
9:   indr  $\leftarrow$  candindsr[indphi]
10:  pf  $\leftarrow$  [R[indr], PHI[indphi]]
11:  lineres  $\leftarrow$  [p, pf]
12:  return lineres
13: end procedure

```

R and *PHI* in lines 3 and 10 are constants describing the edges of the 2D histogram bins. They should be dense enough and the resolution of the phi and r variables should be considered.

Χρησιμοποιώντας τη μέθοδο 1 επαναληπτικά και εναλλάσσοντας το αρχικό σημείο, είμαστε ικανοί να εξαγάγουμε έναν προσεγγιστικό σκελετό του χεριού. Για να γίνει αυτό, χρειάζονται και τα **διανύσματα υποστύλωσης** (pile vectors) v_p , τα οποία ορίζονται εκ κατασκευής στη συνέχεια.

Όπως αναφέρθηκε προηγουμένως, ως πρωταρχικό σημείο $x^{(0)}$ θεωρούμε το κέντρο του σημείου εισόδου του χεριού. Ως αντίστοιχο διάνυσμα υποστύλωσης $v_p^{(0)}$ χαρακτηρίζεται αυτό με αρχή και πέρας τα άκρα του τμήματος εισόδου. Μετά από κάθε i -οστή επανάληψη, αρχικό σημείο $x_{skel}^{(i)}$ τίθεται το κέντρο του τμήματος, το οποίο ορίζεται από το υπολογιζόμενο σημείο p_f και το σημείο το οποίο απέχει το ίδιο από το αρχικό σημείο $x_{skel}^{(i-1)}$ και βρίσκεται στην απέναντι πλευρά του περιγράμματος. Τα σημεία αυτά χρησιμοποιούνται και για την κατασκευή του αντίστοιχου διανύσματος υποστύλωσης $v_p^{(i)}$, το οποίο έχει αρχή και πέρας το σημείο που είναι κοντύτερα στην αρχή και το πέρας του διανύσματος υποστύλωσης $v_p^{(i-1)}$ αντίστοιχα. Ο πίνακας $[x_{skel}^{(i-1)}, x_{skel}^{(i)}]$ αποτελεί τον i -οστό **σύνδεσμο** (link) l_i του σκελετού. Επιπροσθέτως, το μέρος του περιγράμματος που περιφράσσεται από τα δύο αντίστοιχα διανύσματα υποστύλωσης $v_p^{(i-1)}$ και $v_p^{(i)}$ είναι το i -οστό **μέρος** (part) p_i του σκελετού και αφαιρείται από το περίγραμμα υπό εξέταση στις επόμενες επαναλήψεις. Ο αλγόριθμος σταματάει όταν το εναπομείναν περίγραμμα έχει λιγότερα από $th_{n_{skel}}$ σημεία, όταν το τελευταίο διάνυσμα υποστύλωσης έχει μικρότερο μήκος από $th_{l_{skel}}$ εικονοστοιχεία ή όταν ο μέγιστος αριθμός επαναλήψεων $th_{i_{skel}}$ έχει εγγραφεί. Αν ο ευρεθής σύνδεσμος l_i έχει μήκος λιγότερο από $th_{w_{skel}}$ εικονοστοιχεία κατά τη διάρκεια της επανάληψης, συγχωνεύεται με τον l_{i-1} , οπότε κατασκευάζεται ένας νέος σύνδεσμος στη θέση τους, με διανύσματα υποστύλωσης τα $v_p^{(i-2)}$ και $v_p^{(i)}$. Συνέπεια αυτού του βήματος είναι ότι ένας η περισσότεροι σύνδεσμοι μπορεί να μην βρίσκονται εντός του περιγράμματος του χεριού.

Εφόσον η μάσκα του χεριού δεν είναι θορυβώδης και το τμήμα εισόδου δεν έχει ασυνέχειες, το αποτέλεσμα του παραπάνω αλγορίθμου είναι ο προσεγγιστικός σκελετός \mathcal{S} , ο οποίος περιέχει τα μέρη p και τους συνδέσμους l του χεριού. Αν δεν ισχύουν οι παραπάνω συνθήκες, ο αλγόριθμος δεν είναι εγγυημένο ότι θα λειτουργήσει όπως αναμένεται. Η υπολογιστική πολυπλοκότητά του είναι ανάλογη του αριθμού των σημείων του περιγράμματος και του βαθμού κυρτότητάς του.

3.2.6 Προσέγγιση Περιοχής Ενδιαφέροντος του Χεριού (Hand ROI Approximation)

Υποθέτοντας ότι ο σκελετός έχει ληφθεί επιτυχώς από το προηγούμενο βήμα, η ROI του χεριού μπορεί να προσεγγιστεί, χρησιμοποιώντας τον τελικό σύνδεσμο και μέρος του σκελετού.

Η ROI υποτίθεται ακόμα ότι έχει μορφή ορθογώνιου παραλληλογράμμου. Για το σκοπό αυτό, βρίσκεται το ελάχιστο περιστραμμένο ορθογώνιο παραλληλόγραμμο, το οποίο περικλείει το τελευταίο μέρος του σκελετού (bounding box). Η μία πλευρά της ROI του χεριού ορίζεται να είναι αυτή του bounding box, η οποία είναι κοντύτερα στο τελευταίο διάλυσμα υποσύτλωσης του σκελετού. Το μήκος της άλλης πλευράς περιοχής ενδιαφέροντος ορίζεται ανάλογα με το πλάτος αυτής, πολλαπλασιασμένο με έναν συντελεστή r_{bskel} . Η έξοδος αυτού του υποσυστήματος είναι η προσεγγισμένη ROI του χεριού, η οποία επιτρέπει την αποκοπή της παλάμης και των δαχτύλων (cropped hand) από το υπόλοιπο χέρι, δηλαδή μέσω αυτής λαμβάνουμε ένα ορθογώνιο παραλληλόγραμμο, του οποίου τα μη μηδενικά pixels εκφράζουν το βάθος των σημείων που μας ενδιαφέρουν.

3.3 Εξαγωγή Χαρακτηριστικών

Οι περιγραφητές που χρησιμοποιούνται για την εξαγωγή των χαρακτηριστικών της μορφής και της κίνησης τη παλάμης είναι παραλλαγές εν μέρει αυτών που αναφέρονται στην ενότητα 2.1.4 και βασίζονται σε ένα βαθμό στην εργασία των Fanello et. al [11]. Κύριο μέλημα είναι να επιτευχθεί μία ισορροπία μεταξύ του υπολογιστικού κόστους και της ακρίβειας ταξινόμησης, για αυτό και επιλέγονται ‘οικονομικές’ παραλλαγές, οι οποίες αφορούν όμως τον τρισδιάστατο χώρο. Όπως και άλλοι, οι Yang et al. [53] και οι Lee et al. [21] αναφέρουν ότι η Αραιή Κωδικοποίηση ενός συνόλου διανυσμάτων χαρακτηριστικών μη γραμμικά διαχωρίσιμων κατηγοριών το ανάγει σε ένα σύνολο χαρακτηριστικών γραμμικά διαχωρίσιμων κατηγοριών. Η υπόθεση αυτή βασίζεται στο γεγονός ότι έχει αποδειχτεί ότι ο ανθρώπινος εγκέφαλος αποθηκεύει την πληροφορία σε αραιή μορφή υπό τη μορφή νευρικών συνάψεων, η οποία κατόπιν είναι ευκολότερα προσπελάσιμη και συσχετίσιμη με άλλες πληροφορίες. Παρά την υποστηριζόμενη άποψη, η αραιή κωδικοποίηση δεν φάνηκε να βελτιώνει ιδιαίτερα την ακρίβεια της ταξινόμησης, ενώ αύξανε ιδιαίτερος πολύ το χρόνο επεξεργασίας, οπότε καθιστούσε τον αλγόριθμο μη εκτελέσιμο σε πραγματικό χρόνο. Για τους δύο αυτούς λόγους κρίθηκε πρέπον να μην χρησιμοποιηθεί sparse coding στην τελική υλοποίηση. Χρησιμοποιήθηκε βέβαια στον πειραματισμό, κατά τον οποίο και εμφανίζονται τα παραπάνω προβλήματα.

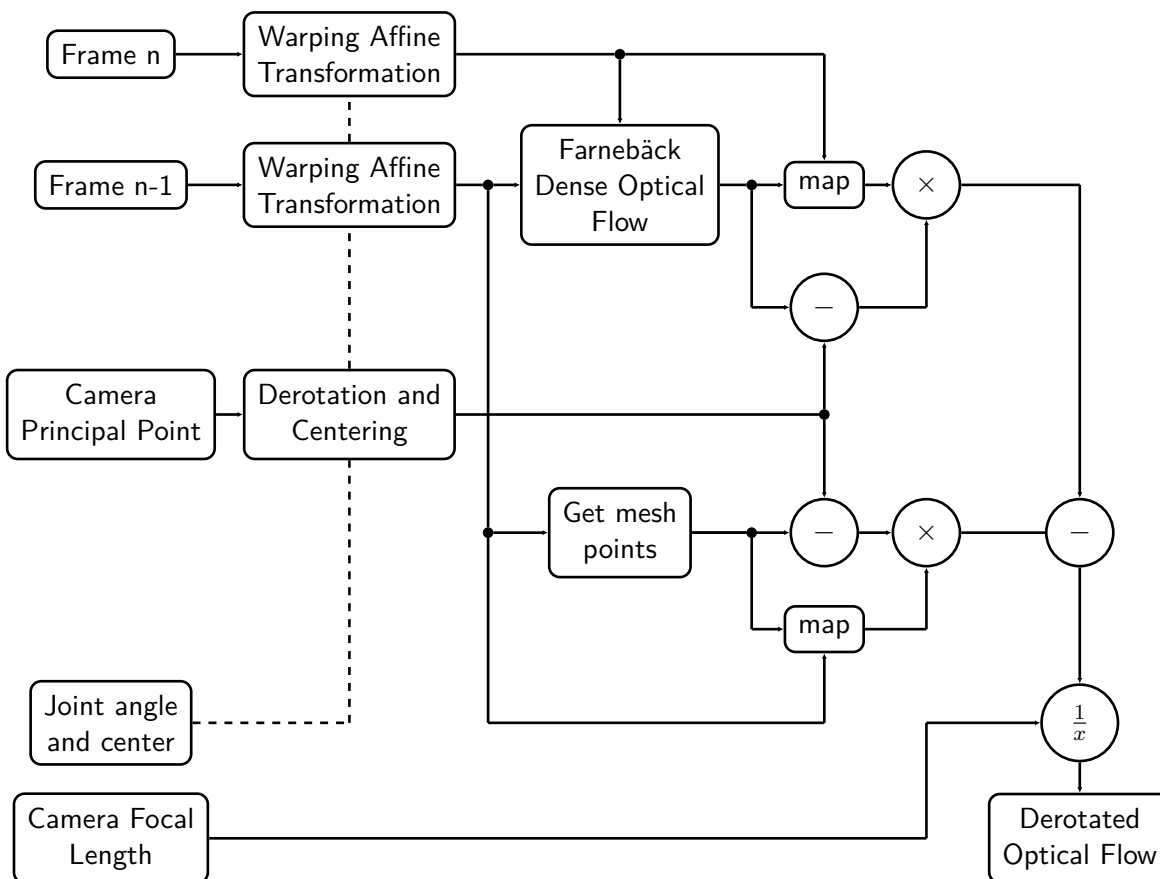
Οι περιγραφητές, με τους οποίους εκτελέστηκαν τα πειράματα, είναι οι GHOG, 3DHOF και 3DXYPEA και αναλύονται στη συνέχεια.

3.3.1 Περιγραφητής GHOG

Ο περιγραφητής **GHOG** αποτελεί μια απλούστευση, την οποία έχουν υιοθετήσει και οι Fanello et al. [11], του περιγραφητή HOG, ο οποίος περιγράφεται στην Εν. 2.1.4.2. Αντί να υπολογίζεται το ιστόγραμμα των βαθμίδων σε κύτταρα της περιοχής ενδιαφέροντος, υπολογίζεται σε όλη τη ROI, με είσοδο την τιμή του βάθους κάθε εικονοστοιχείου. Αποτελεί μία ασθενή αλλά ταχεία αναπαράσταση, που χρησιμοποιείται για να δώσει μία γενική εκτίμηση της τραχύτητας και της καμπυλότητας της ROI. Θετική συνεισφορά στη μετέπειτα αναγνώριση είναι εμφανής από το συγκεκριμένο περιγραφητή, εφόσον αυτός λάβει μια αυστηρά προσδιορισμένη και φιλτραρισμένη είσοδο, ιδιότητες που κατέχει η έξοδος του συστήματος προεπεξεργασίας. Ο αριθμός των κάδων του ιστογράμματος του GHOG ορίστηκε ως n_{GHOG} .

3.3.2 Περιγραφητής 3DHOF

Ο περιγραφητής **3DHOF** είναι μία γενίκευση του περιγραφητή κίνησης HOF στο τρισδιάστατο χώρο (βλ. Εν. 2.1.4.1). Όπως έχει αναφερθεί, η ιδιαιτερότητά του έγκειται στο γεγονός ότι, αντί να υπολογίζεται η Οπτική Ροή (βλ. Εν. 2.1.3) χρησιμοποιώντας τη χρωματική πληροφορία του δισδιάστατου επιπέδου, αυτή βρίσκεται συσχετίζοντας τα προβαλλόμενα σημεία της εικόνας βάθους στον τρισδιάστατο χώρο, όντας τελικά ένας πενταδιάστατος τένσορας, στον οποίο περιγράφεται η τρισδιάστατη μετακίνηση κάθε εικονοστοιχείου. Κατασκευάζοντας ένα τρισδιάστατο ιστόγραμμα της κατεύθυνσης αυτής της μετακίνησης, με τη χρήση αυτού του τένσορα, και κανονικοποιώντας, λαμβάνουμε το τρισδιάστατο ιστόγραμμα κίνησης, του οποίου οι κάδοι (bins) τέθηκαν ίσοι με n_{3DHOF} ανά διάσταση (συνολικά n_{3DHOF}^3 κάδοι). Ο περιγραφητής αυτός δεν είναι αναλλοίωτος όσον αφορά την περιστροφή του πεδίου της κίνησης, δηλαδή σε περίπτωση που το χέρι εκτελέσει την ίδια κίνηση σε σχέση με τον χρήστη αλλά με διαφορετική γωνία ως προς την κάμερα, ο περιγραφητής θα δώσει διαφορετικά αποτελέσματα. Για να επιλύσουμε αυτό το ζήτημα, προαπαιτείται να κανονικοποιήσουμε τη θέση και τη γωνία του χεριού. Αυτό καθίσταται δυνατό με το να βρούμε τη γωνία του τελευταίου συνδέσμου $l^{(-1)}$ του σκελετού S ως προς τον οριζόντιο άξονα και να περιστρέψουμε τη μάσκα του χεριού με την αντίθετη γωνία ως προς το προτελευταίο αρχικό σημείο $x_{skel}^{(-2)}$, εφαρμόζοντας έτσι μία διαδικασία, που ονοματίστηκε ‘derotation’. Επειδή συμμετέχουν δύο frame στην εξαγωγή των χαρακτηριστικών, απαιτείται το derotation να γίνει με τη χρήση του σκελετού του χεριού του πρώτου από αυτά. Αυτή η διαδικασία μπορεί να εφαρμοστεί βέλτιστα, αν την εισαγάγουμε στο σύστημα εξαγωγής της Οπτικής Ροής όπως φαίνεται ακολούθως:



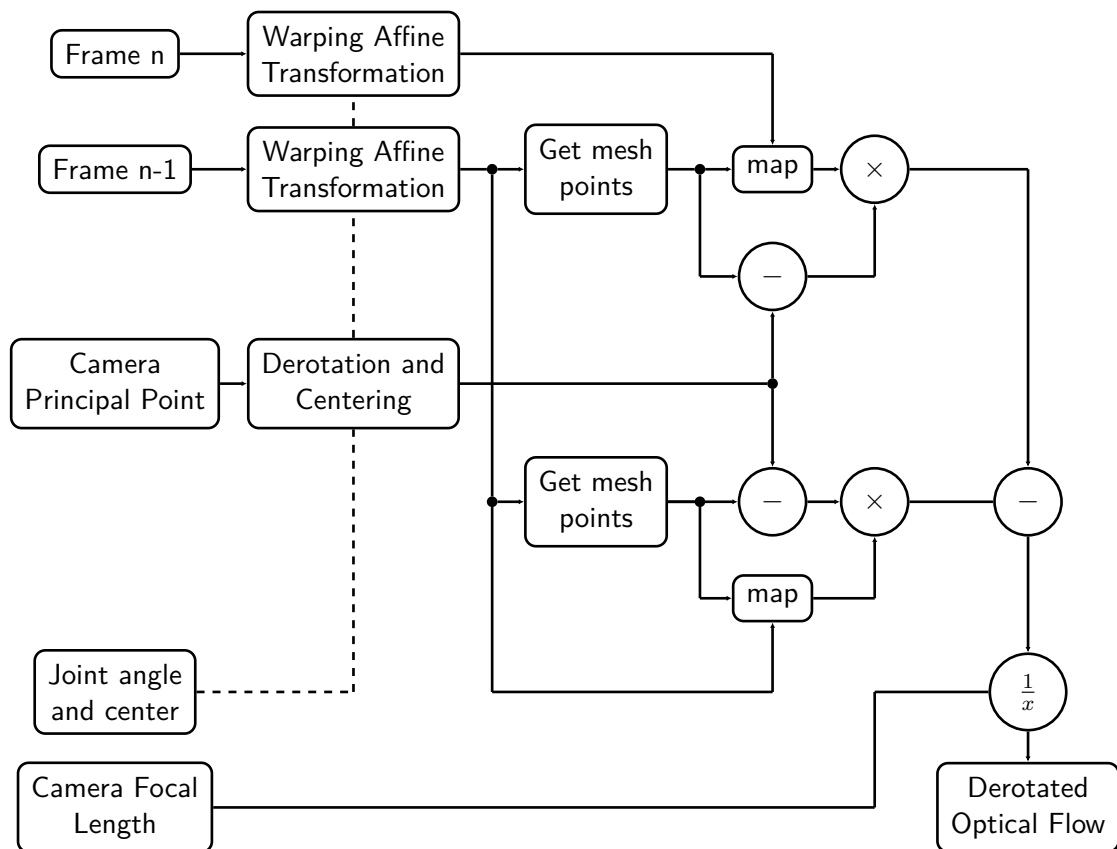
Σχήμα 3.2: ‘Derotated’ Οπτική Ροή για την εξαγωγή των χαρακτηριστικών από τον 3DHOF Περιγραφητή. Το κάθε εικονοπλαίσιο έχει το ίδιο μέγεθος με το αρχικό frame βάθους, με μη-δενισμένα τα εικονοστοιχεία. Τα χαρακτηριστικά της κάμερας είναι το Σημείο Εστίασης (Focal Point) και το Πρωτεύον Κύριο Σημείο (Principal Point) και χρησιμοποιούνται για την εφαρμογή του μοντέλου Pinhole Camera. Το υποσύστημα ‘Get mesh points’ δεν παράγει τίποτα άλλο από δύο πίνακες με τις οριζόντιες και κάθετες συντεταγμένες των εικονοστοιχείων των frames. Το υποσύστημα ‘Farneback Dense Optical Flow’ δίνει έξοδο ίδιας μορφής. Επειδή τα εικονοπλαίσια έχουν το ίδιο μέγεθος, αυτοί οι πίνακες δεν αλλάζουν ανά frame. Το υποσύστημα ‘map’ αντιστοιχίζει τους πίνακες συντεταγμένων με τις τιμές των εικονοπλαισίων, παράγοντας έτσι έναν δισδιάστατο πίνακα.

3.3.3 Περιγραφητής ZHOF

Μία απλούστευση του περιγραφητή 3DHOF είναι ο πρότυπος περιγραφητής **ZHOF**. Η πυκνή Οπτική Ροή, που λαμβάνει κανείς από την εφαρμογή του αλγορίθμου του Farneback [12] όπως φαίνεται στο Διάγραμμα 3.2, δρα σε εικονοστοιχεία που εκφράζουν βάθος. Το βάθος κάθε σημείου όμως μεταβάλλεται χωρίς να υπάρχει πάντοτε γειτονικά χωρική ομοιογένεια και χρονική συνέχεια σε αυτή τη μεταβολή, όπως γίνεται στην περίπτωση του χρώματος. Καθώς έχουμε λάβει ως προϋπόθεση τη μη χρήση χρώματος για τον αλγόριθμο, δεν είναι απολύτως δόκιμη η

ανίχνευση της πυκνής Οπτικής Ροής με τις παρούσες διδιάστατες μεθόδους. Ο αριθμός των κάδων (bins) του ιστογράμματος ανά διάσταση τέθηκε ίσος με n_{ZHOF} .

Θα μπορούσαμε να αυξήσουμε πολύ περισσότερο το υπολογιστικό κόστος εισάγοντας μεθόδους ανίχνευσης Scene Flow, όπως αυτή των Vogel et al. [47], οι οποίοι παρουσιάζουν μία state-of-the-art μέθοδο μοντελοποίησης του χώρου με χρήση δυναμικών στερεών σωμάτων και επιφανειών. Επειδή όμως στοχεύουμε στην πραγματικού χρόνου εκτέλεση, επιλέξαμε ως ‘φτηνή’ εναλλακτική τον περιγραφητή ZHOF, ο οποίος υπολογίζει μία προσεγγιστική οπτική ροή, όπως φαίνεται στο ακόλουθο διάγραμμα:

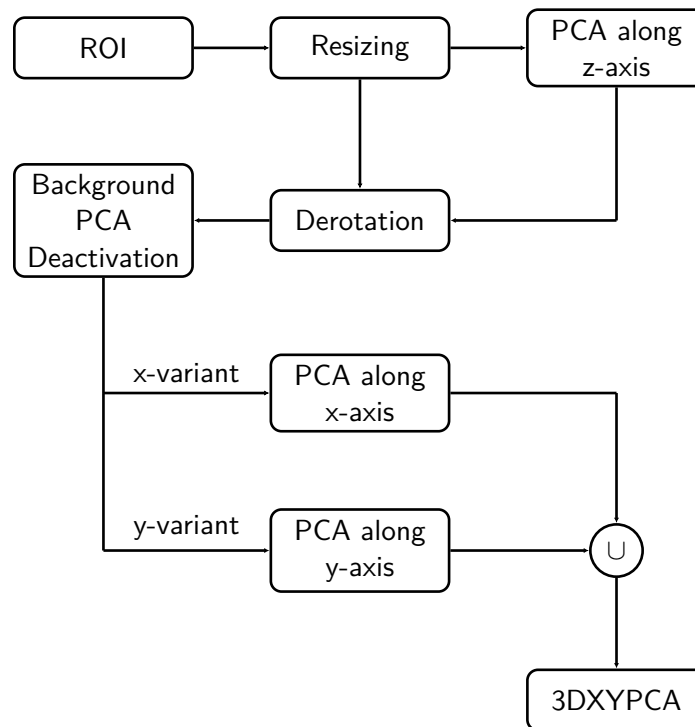


Σχήμα 3.3: ‘Derotated’ Οπτική Ροή για την εξαγωγή των χαρακτηριστικών από τον ZHOF Περιγραφητή.

Όπως μπορεί να παρατηρήσει κανείς, αφαιρέσαμε το στάδιο υπολογισμού της διδιάστατης πυκνής Οπτικής Ροής. Με τον τρόπο αυτό, αφαιρώντας μέρος πληροφορίας για την εγκάρσια κατεύθυνση στον άξονα του αισθητήρα, επιτυγχάνουμε και την αφαίρεση του θορύβου που εισάγεται από την ελαττωματική ανίχνευση του Optical Flow. Οι δύο περιγραφητές ελέγχονται στο πειραματικό μέρος για την ικανότητά τους στην αποτύπωση της κίνησης του χεριού.

3.3.4 Περιγραφητής 3DXYPCA

Για να αναπαρασταθεί η μεταβλητότητα των διαστάσεων του χεριού και η στατικότητα του, κρίθηκε βέλτιστο να χρησιμοποιηθεί ένας άλλος πρότυπος περιγραφητής, ονόματι **3DXYPCA**. Η δράση αυτού στην προεπεξεργασμένη είσοδο φαίνεται ακολούθως:



Σχήμα 3.4: Εξαγωγή 3DXYPCA Χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών συνίσταται αρχικά από έναν μετασχηματισμό μεγέθους (Resizing) μέσω παρεμβολικών μεθόδων, συγκεκριμένα μέσω απλής διγραμμικής παρεμβολής (bilinear interpolation), της περιοχής ενδιαφέροντος, δίνοντας ως έξοδο μία τετραγωνική εικόνα μεγέθους $n_{PCA} \times n_{PCA}$, με n_{PCA} , έτσι, ώστε να κανονικοποιηθούν οι διαστάσεις της και να μην επηρεάζεται η εξαγωγή χαρακτηριστικών από μικρές μεταβολές βάθους. Κατόπιν το αποτέλεσμα διαμορφώνεται από τη διαδικασία του derotation (βλ. Εν. 3.3.2), χρησιμοποιώντας αυτήν τη φορά τον άξονα που προκύπτει εκτελώντας **Ανάλυση Κύριων Συνιστωσών (PCA)**[16], μια στατιστική μέθοδο κατά την οποία επιλέγεται ο άξονας στην εικόνα, όπου υπάρχει η μεγαλύτερη διασπορά των μη μηδενικών εικονοστοιχείων και άρα η μεγαλύτερη ποσότητα πληροφορίας, για να προσεγγιστεί η invariance των χαρακτηριστικών στην περιστροφή ως προς τον άξονα Z, δηλαδή αυτόν του βάθους.

Στη συνέχεια, υπάρχει ένα στάδιο το οποίο ονομάστηκε 'Απενεργοποίηση PCA Υποβάθρου' (Background PCA Deactivation), κατά το οποίο τα μηδενισμένα εικονοστοιχεία του υποβάθρου της ROI λαμβάνουν τέτοιες τιμές, ώστε να μηδενίζεται η διασπορά τους ως προς τις κατευθύνσεις X και Y, παράγοντας έτσι δύο εξόδους x-variant και y-variant αντίστοιχα. Τελικά εκτελείται PCA στην έξοδο x-variant ως προς τον άξονα X και στην έξοδο y-variant ως προς τον άξονα Y, λαμβάνοντας έτσι δύο διανύσματα διάστασης n_{PCA} , με το κάθε διάνυσμα να αναφέρεται στο πρωταρχικό στοιχείο του PCA, τα οποία συνενώνονται παράγοντας τα χαρακτηριστικά 3DXYPEA, διάστασης $2n_{PCA}$.

Η απενεργοποίηση PCA Υποβάθρου είναι απαραίτητη, γιατί αλλιώς η μέγιστη διασπορά θα εκφυλιστεί στην απόσταση του υποδειγματοληπτημένου χεριού από τον αισθητήρα στην εκάστοτε διεύθυνση. Ένας ακόμα λόγος για τον οποίο μετασχηματίζεται η εικόνα είναι για να αυξηθεί η ταχύτητα εκτέλεσης του PCA, αφού το συνολικό υπολογιστικό κόστος του και στις δύο διευθύνσεις είναι $O(4n_{PCA}^3)$. Το κόστος αυτό προκύπτει από το γεγονός ότι το PCA υπολογίζεται μέσω της εύρεσης ιδιοτιμών (eigenvalues) του πίνακα συνδιακύμανσης (covariance matrix), του οποίου το κόστος υπολογισμού είναι $O(n^3)$ για έναν τετραγωνικό πίνακα $n \times n$, ενώ το κόστος της εύρεσης ιδιοτιμών είναι το ίδιο, δηλαδή $O(n^3)$, αν θεωρήσει κανείς ότι εφαρμόζεται η τεχνική SVD [2].

Εκ κατασκευής, τα χαρακτηριστικά 3DXYPEA είναι αναλλοίωτα σε περιστροφές ως προς τον άξονα z και στην αλλαγή της κλίμακας, λόγω της τεχνικής του Resizing. Παραμένουν όμως εξαρτημένα από την γωνία του χεριού ως προς τον X και Y άξονα αντίστοιχα, οπότε η μεταβολή της εγκάρσιας θέσης του χεριού στο πεδίο του αισθητήρα μπορεί να επηρεάσει σημαντικά τη μορφή τους.

3.3.5 Τεχνική Sparse Coding

Αν και η Αραιή Κωδικοποίηση αυξάνει ιδιαιτέρως την υπολογιστική πολυπλοκότητα του αλγορίθμου, καθιστώντας τον αδύνατο να εκτελεστεί σε πραγματικό χρόνο με χαμηλές απαιτήσεις υλισμικού, κρίθηκε σωστό να χρησιμοποιηθεί σε κάποια από τα πειράματα, αν και απέχει από την τελική μορφή του αλγορίθμου, για να φανεί το ποσοστό επιρροής που ασκεί στην ακρίβεια των αποτελεσμάτων. Ο αλγόριθμος Feature sign-search χρησιμοποιήθηκε αυτούσιος στο κομμάτι των δοκιμών, λόγω του ότι φάνηκε να επιδεικνύει ταχύτερη σύγκλιση από άλλες μεθόδους και είναι απλός στην κατανόηση και στην υλοποίηση, ενώ στο offline στέλεχος της εκπαίδευσης

των λεξικών η επίλυση της Lagrangian Dual, όπως περιγράφεται από τη Σχέση 2.8, δεν συνέκλινε, λόγω κακών μεθόδων αρχικοποίησης και μη καλά ορισμένου πίνακα συνδιασποράς του λεξικού, ο οποίος είχε πολύ χαμηλό βαθμό (rank). Επιπλέον, ο συντελεστής γ που εισάγεται στη σχέση 2.10a δεν έχει κάποια εμφανή συσχέτιση με το συντελεστή c της Σχέσης 2.10b, οπότε η απαίτηση του να είναι αραιό το διάνυσμα εξόδου δεν λαμβάνεται υπόψη κατά τη διάρκεια εκπαίδευσης του λεξικού. Αυτό έχει ως αποτέλεσμα το ότι η συνθήκη στο βήμα 2 του αλγορίθμου FSS (βλ. Παρ. A.2) μπορεί να ισχύει εξ αρχής για δεδομένο γ , οπότε το αραιό διάνυσμα εκφυλίζεται στο μηδενικό. Για να μην ισχύει σε καμία περίπτωση αυτό, θα έπρεπε το γ να είναι εξαρτημένη μεταβλητή από το διάνυσμα εισόδου, βάσει της ακόλουθης σχέσης, η οποία προκύπτει από τη προαναφερόμενη συνθήκη:

$$\gamma \leq \max (|-2\mathbf{A}^T \mathbf{y}|)$$

όπου \mathbf{y} το διάνυσμα εισόδου και \mathbf{A} το λεξικό. Το να απαιτεί κανείς όμως να μεταβάλλεται η συνάρτηση ελαχιστοποίησης ανά δείγμα εισόδου καθιστά την εκπαίδευση του λεξικού ιδιαίτερα απαιτητική, κυρίως από την άποψη ότι ο αριθμός των δειγμάτων εκπαίδευσης θα πρέπει να είναι ιδιαίτερα μεγάλος, για να προσαρμοστεί το λεξικό σε κάθε πιθανή μεταβολή της παραμέτρου γ , ενώ δεν μπορεί να εγγυηθεί κανείς ότι η εκπαίδευση εντέλει θα συγκλίνει σε μία υποβέλτιστη λύση.

Για τους παραπάνω λόγους αποφασίστηκε να αντικατασταθεί η επίλυση της Lagrange Dual από την εφαρμογή της μεθόδου **Least Angle Regression** των Effron et al. [9], μέσω του εργαλείου Scikit Learn. Αυτή η μέθοδος, αν και χαρακτηρίζεται από σχετικά αργή σύγκλιση, μπορεί και συγκλίνει με τα παρόντα δεδομένα. Επιλύει ένα σχετικά διαφορετικό πρόβλημα βελτιστοποίησης, κάνοντας χρήση διαφορετικών συνθηκών για τις στήλες του λεξικού:

$$(\mathbf{U}^*, \mathbf{V}^*) = \operatorname{argmin}_{(\mathbf{U}, \mathbf{V})} \left(\frac{1}{2} \|\mathbf{Y} - \mathbf{UV}\|_F^2 + \alpha * \|\mathbf{U}\|_1 \right) \quad (3.1)$$

with $\|\mathbf{V}_j\|_F = 1$ for $j = 1, \dots, n_{sparse}$

με \mathbf{V} τον κώδικα και \mathbf{U} την ένωση διάστασης $n_{sparse} \times l_{tr}$ των εξόδων του συστήματος Αραιής Κωδικοποίησης, στο οποίο έχουν εισαχθεί l_{tr} διανύσματα εκπαίδευσης διάστασης $m \times 1$.

3.4 Κατασκευή Buffer

Χρησιμοποιώντας την ένωση των παραπάνω χαρακτηριστικών, μπορούμε να εξαγάγουμε έναν Buffer, συνδέοντας χαρακτηριστικά συνεχόμενων frames, ο οποίος χρησιμεύει στην αναγνώριση δυναμικών χειρονομιών. Για να μην υπάρξει πρόβλημα ελλιπούς συνοχής του buffer, θέτουμε ένα όριο μεταξύ της απόστασης των frames που μπορούν να τον συνθέσουν. Πιο συγκεκριμένα, ένας buffer είναι έγκυρος όταν:

$$\left| \text{diff}(f_i) \right| < th_{conf} * l_b \forall i = 1 \dots l_b$$

όπου f_i είναι ο χρονικός δείκτης του i -οστού frame του buffer μήκους l_b , και th_{conf} . Επιπλέον, δεν μπορεί να σχηματιστεί buffer με λιγότερα από l_b στοιχεία, γεγονός που και αυτό με τη σειρά του περιορίζει τον διαθέσιμο αριθμό από έγκυρους buffers, ιδίως σε ένα θορυβώδες πεδίο λήψης, στο οποίο δεν υπάρχει έξοδος από το στάδιο της προεπεξεργασίας ή το στάδιο των περιγραφητών.

3.5 Post Time PCA

Ως ένα ακόμα στάδιο επεξεργασίας που δοκιμάστηκε είναι η εφαρμογή της μεθόδου PCA στον παραγόμενο Buffer στον άξονα του χρόνου, με στόχο τη μείωση των χαρακτηριστικών με την ελάχιστη απώλεια πληροφορίας και ταυτόχρονα την αύξηση της ταχύτητας της ταξινόμησης και της μάθησης των ταξινομητών. Από τη διαδικασία αυτή συγκρατούμε τα πρώτα n_{PTPCA} ιδιοδιανύσματα που παράγονται από τη μέθοδο της Ανάλυσης Πρωτευόντων Συνιστωσών.

3.6 Ταξινόμηση

3.6.1 Γενικά

Η διαδικασία της ταξινόμησης ενέχει δύο ταξινομητές, οι οποίοι λειτουργούν συγχρόνως και χρησιμοποιούνται για να ταξινομήσουν διαφορετικού είδους χειρονομίες:

- Ο Cl_{pas} , δέχεται τα χαρακτηριστικά των παθητικών χειρονομιών, για τις οποίες η χρονική μεταβολή δεν είναι ιδιαίτερου ενδιαφέροντος, ενώ απαιτείται κάποιου είδους αδράνεια

στην ταξινόμηση, για να αποφευχθούν ταχείες μεταβολές. Για το λόγο αυτό, επιλέγονται περιγραφητές μορφής.

- Ο ταξινομητής Cl_{dyn} χρησιμοποιείται για την αναγνώριση δυναμικών χειρονομιών, για τις οποίες η κίνηση παίζει σημαντικό ρόλο στην ταξινόμηση, οπότε δέχεται περιγραφητές μορφής και κίνησης. Τα χαρακτηριστικά προέρχονται από l_b frames και έχουν ενωθεί μεταξύ τους για να δημιουργήσουν ένα διάνυσμα $l_b n_{feat_{dyn}} \times 1$, με $n_{feat_{dyn}}$ το άθροισμα των διαστάσεων των χαρακτηριστικών ανά frame.

Ο κάθε ταξινομητής από τους Cl_{pas} και Cl_{dyn} είναι είτε ένας ταξινομητής RDF που περιέχει n_{RDF} δέντρα απόφασης ή ένα σύνολο από υποταξινομητές, n_{dyn} δυαδικές μηχανές διανυσμάτων υποστήριξης. Οι υποταξινομητές τύπου SVM εκπαιδεύονται με χρήση της μεθόδου **OneVsRest**, δηλαδή κάθε ‘υποταξινομητής’ C_{S_i} εκπαιδεύεται με θετικά δείγματα τα δείγματα της κλάσης i και με αρνητικά δείγματα τα δείγματα των υπολοίπων. Σε κάθε SVM χρησιμοποιείται η παράμετρος C_{SVM} , η οποία ρυθμίζει την ευαισθησία του SVM στα outliers και το ελάχιστο μέγεθος της βέλτιστης απόστασης μεταξύ των κατηγοριών, με σταθερή τιμή. Αυτή η μέθοδος εκπαίδευσης μπορεί να επιφέρει προβλήματα μη σταθμισμένων δεδομένων εκπαίδευσης και ελαττωματικής τελικής ταξινόμησης, τα οποία δεν παρατηρήθηκαν σε μεγάλο βαθμό με τα δεδομένα δείγματα.

3.6.2 Επεξεργασία των Scores

Στη συνέχεια αναλύονται οι διαφορετικοί τρόποι χειρισμού των αποτελεσμάτων (scores). Πριν από την είσοδό τους στα ακόλουθα συστήματα, τα scores ομαλοποιούνται με τη χρήση ενός φίλτρου μέσης τιμής, μεγέθους $n_{f_{sc}}$ frames.

3.6.2.1 Μέθοδος CProb

Η μέθοδος *CProb* (Check **Probability**) αποτελεί μια απλοϊκή προσέγγιση επεξεργασίας των εξαγόμενων scores. Για να λειτουργήσει, απαιτείται τα scores να εκφραστούν υπό τη μορφή πιθανοτήτων, αν δεν είναι ήδη, όπως συμβαίνει στην περίπτωση αυτών που προκύπτουν από τον RDF ταξινομητή.

- Αν όλες οι πιθανότητες είναι κατώτερες από κάποιο κατώφλι th_{RDF} , τότε απονέμεται στη δείγμα η προηγούμενη προβλεφθείσα κλάση.

- Σε περίπτωση που δεν ισχύει η παραπάνω συνθήκη ή αν δεν υπάρχει προηγούμενη προβλεφθείσα κλάση, απονέμεται στο δείγμα η κλάση με τη μέγιστη πιθανότητα.

Υπό μαθηματική μορφή, η επεξεργασία εκφράζεται ως:

$$C^{(t)} = \begin{cases} \operatorname{argmax} \left(\mathbf{S}_{RDF}^{(t)} \right) & \text{if } \max \left(\mathbf{S}_{RDF}^{(t)} \right) \geq th_{RDF} \vee \#C^{(t-1)} \\ C^{(t-1)} & \text{otherwise} \end{cases} \quad (3.2)$$

Με τη μέθοδο αυτή, αποκλείονται scores με μικρή ισχύ και προσδίδεται μία χρονική εξάρτηση των κλάσεων στο χρόνο, όσον αφορά την κάθε χειρονομία, αποτρέποντας πιθανές ασυνέχειες λόγω λανθασμένης αναγνώρισης. Εκ κατασκευής, αυτή η μέθοδος φαίνεται να αποδίδει καλύτερα στην αναγνώριση των παθητικών χειρονομιών, λόγω μικρής χρονικής εξάρτησης.

3.6.2.2 Μέθοδος CSTD

Η μέθοδος που περιγράφεται αποτελεί παραλλαγή της εργασίας των Fanello et al. [11]. Η μέθοδος *CSTD* (Check **STD**) καθιστά δυνατή τη χρονική τοποθέτηση των δράσεων, δηλαδή επιτρέπει την εύρεση των αρχικών και τελικών χρονικών σημείων αυτών:

1. Υπολογίζεται η τυπική απόκλιση των scores ανά χρονική στιγμή, η οποία ομαλοποιείται χρησιμοποιώντας ένα φίλτρο μέση τιμής μεγέθους n_{sstd} frames.
2. Υπολογίζεται η μέση τιμή των τυπικών αποκλίσεων σε ένα σχετικά μικρό χρονικό παράθυρο, μεγέθους n_{bstd} frames.
3. Σε περίπτωση που η τυπική απόκλιση των φιλτραρισμένων scores είναι μικρότερη από τη υπολογιζόμενη μέση τιμή, τότε αναγνωρίζεται ότι υπάρχει εναλλαγή δράσεων.

Η υπόθεση του τελευταίου βήματος έγκειται στο γεγονός ότι τη χρονική στιγμή που σταματά μια δράση να εκτελείται, κανένας από τους ταξινομητές δεν μπορεί να αποφανθεί με βεβαιότητα για την κατηγορία που ανήκει το εκάστοτε δείγμα, οπότε όλα τα scores είναι περίπου ίδια και η τυπική απόκλιση αποκτά χαμηλή τιμή. Η αναγνώριση της περιεχόμενης δράσης γίνεται είτε στην ανίχνευση της ολοκλήρωσής της, είτε κατά τη διάρκεια της εκτέλεσής της και επιτυγχάνεται βρίσκοντας την κλάση με τη μέγιστη χρονική μέση τιμή των scores, από το σημείο εκκίνησής

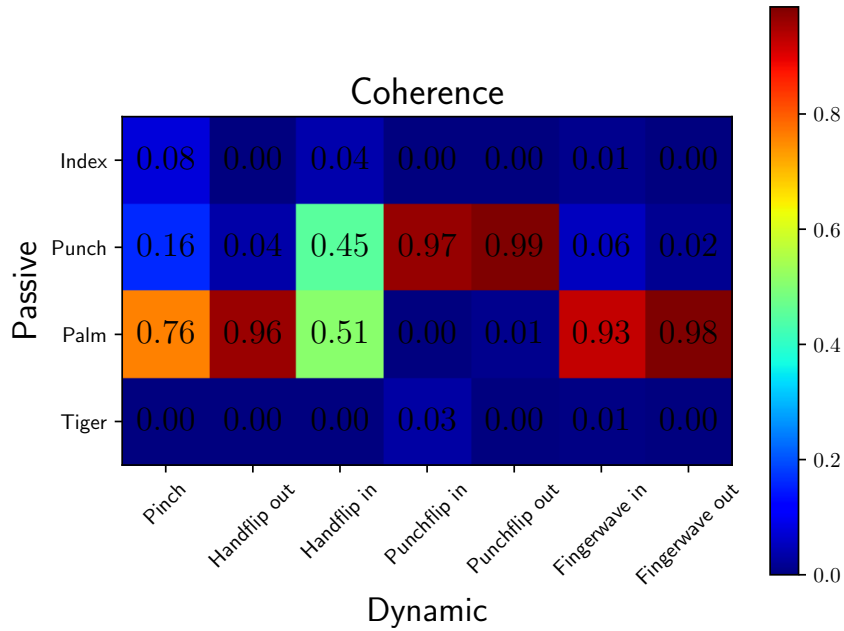
της. Με τη λογική αυτή, πέρα από την αναγνώριση της δράσης, επιτυγχάνεται θεωρητικά και η αναγνώριση των ορίων αυτής.

3.6.3 Combined Dynamic Bayesian Mixture Model (CDBIMM)

Εξ ορισμού οι δυναμικές χειρονομίες αποτελούν υπέρθεση παθητικών χειρονομιών στο χρόνο. Για το λόγο αυτό, είναι λογικό κανείς να υποθέσει ότι υπάρχει συσχέτιση μεταξύ των αποτελεσμάτων των δύο ταξινομητών. Η συσχέτιση αυτή ενδεχομένως μπορεί να βοηθήσει στη βελτίωση της ταξινόμησης των δυναμικών χειρονομιών, οι οποίες είναι και οι πιο απαιτητικές στην αναγνώριση. Με αυτό το σκεπτικό, κατασκευάσαμε έναν ενισχυμένο ταξινομητή δυναμικών δράσεων, ο οποίος δέχεται ως είσοδο την έξοδο των προαναφερόμενων ταξινομητών και παράγει μια βελτιωμένη πρόβλεψη για τις δυναμικές δράσεις που παρατηρούνται. Το συνολικό αυτό σύστημα ονομάστηκε Combined Dynamic Bayesian Mixture Model (CDBIMM) και αναλύεται στη συνέχεια.

3.6.3.1 Εκπαίδευση (Training)

Η εκπαίδευση του CDBIMM γίνεται με τη χρήση του υποσυστήματος Cl_{pas} , το οποίο έχει ήδη εκπαιδευτεί με τα δείγματα των **παθητικών** χειρονομιών, όπως περιγράφεται στην Ενότητα 3.6.1. Στο στάδιο της εκπαίδευσης του συστήματος CDBIMM, ο ταξινομητής Cl_{pas} δοκιμάζεται στα δεδομένα εκπαίδευσης των **δυναμικών** χειρονομιών, από όπου και εξάγονται κάποιες προβλέψεις, χρησιμοποιώντας κάποια από τις μεθόδους CProb και CSTD. Οι προβλέψεις αυτές, οι οποίες αφορούν παθητικές χειρονομίες, συσχετίζονται με τις πραγματικές κλάσεις των δυναμικών χειρονομιών, από όπου και παράγεται ένας πίνακας ονόματι **CM**, ο οποίος έχει τη μορφή ενός πίνακα συνεκτικότητας (Coherence Matrix) και φαίνεται ακολούθως:



Πίνακας 3.1: Παράδειγμα Πίνακα Συνεκτικότητας Χειρονομιών

Οι στήλες του πίνακα, οι οποίες έχουν κανονικοποιηθεί, εκφράζουν την πιθανότητα να γίνεται κάποια δυναμική χειρονομία ($Action, a_j$) i , δεδομένου ότι προβλέφθηκε η παθητική χειρονομία j ($Pose, p_j$) από τον ταξινομητή Cl_{pas} . Με άλλα λόγια, κάθε στοιχείο c του πίνακα εκφράζει το ακόλουθο likelihood:

$$CM = [c[i, j]] = [P_{est}[p_i|a_j]] \quad (3.3)$$

Ο περιορισμός που τίθεται για να λειτουργήσει σωστά αυτή η μέθοδος είναι τα scores των ταξινομητών Cl_{pas} και Cl_{dyn} να είναι εκφρασμένα υπό τη μορφή πιθανοτήτων. Αν $S_p(t)$ και $S_d(t)$ τα επεξεργασμένα υπό μορφή πιθανοτήτων scores από τον Cl_{pas} και Cl_{dyn} αντίστοιχα, τότε βάσει των παραπάνω έχουμε:

$$\begin{aligned} S_p &= [P_{est}[p_i|t]] \\ S_d &= [P_{est}[a_j|t]] \end{aligned} \quad (3.4)$$

3.6.3.2 Δοκιμή (Testing)

Κάποιος μπορεί να εξαγάγει από τα παραπάνω μία πιθανοτική συσχέτιση, η οποία βασίζεται στο απλό Bayesian μοντέλο. Το πρώτο βήμα είναι να θεωρήσουμε την ‘αντίστροφη’ likelihood

πιθανότητα από αυτή που διαθέτουμε, χρησιμοποιώντας τις εκτιμήσεις των ταξινομητών και υποθέτοντας, όπως σιωπηλά είχαμε υποθέσει και παραπάνω, ότι δεν είναι χρονικά εξαρτημένη:

$$P[a_j|p_i]_{est} = \frac{c[i, j]S_{dj}}{S_{p_i}} \quad (3.5)$$

Η πιθανότητα της δράσης a_j να έχει συμβεί στο χρόνο t βάσει των πιθανοτήτων των παθητικών χειρονομιών δίνεται από την ακόλουθη σχέση:

$$P[a_j|t] = \sum_i P[a_j|p_i]P[p_i] \quad (3.6)$$

Μέχρι αυτό το σημείο έχουμε κατασκευάσει ένα μοντέλο, με το οποίο γίνεται να συσχετιστούν με διάφορους τρόπους οι δύο τύποι χειρονομιών. Επιλέγουμε έναν τρόπο, στον οποίο ασκεί μεγαλύτερη επιρροή η εκτίμηση του Cl_{dyn} , αφού περιέχει τη μεγαλύτερη πληροφορία στο χώρο των δυναμικών κλάσεων. Επεκτείνοντας αυτό το σκεπτικό, θεωρούμε ότι η πιθανότητα $P[p_i]$ θα εκτιμηθεί ως ακολούθως:

$$P_{comb}[p_i|t] = \sum_j c[i, j]S_{dj} \quad (3.7)$$

ενώ η likelihood πιθανότητα $P[a_j|p_i]$ υπολογίζεται χρησιμοποιώντας τη σχέση (3.5). Αν θέσουμε $S_p^- = [\frac{1}{S_{p_i}}]$, τότε καταλήγουμε στο εξής ανάμεικτο μοντέλο εκτίμησης:

$$P_{comb}[a_j|t] = S_{dj} \sum_i \{c[i, j]S_{p_i}^- \sum_k c[i, k]S_{dk}\} \quad (3.8)$$

Προβάλλοντας τη σχέση (3.8) στο διανυσματικό χώρο των κλάσεων, λαμβάνουμε τελικά:

$$\tilde{S}_d = S_d \times [C^T(S_p^- \times (CS_d))] \quad (3.9)$$

όπου \tilde{S}_d είναι τα νέα scores για τις δυναμικές χειρονομίες και \times είναι το βαθμωτό γινόμενο, ή αλλιώς γινόμενο Hadamard.

Όπως αναφέρθηκε παραπάνω, για να λειτουργήσει η περιγραφόμενη διαδικασία στην περίπτωση που κάποιος ταξινομητής από τους Cl_{pas} και Cl_{dyn} αποτελείται από SVM υποταξινομητές, πράγμα που μελετάται στο κεφάλαιο της Πειραματικής Διαδικασίας, χρειάζεται να μετασχηματιστούν τα scores αυτών από αποστάσεις από τα υπερεπίπεδα απόφασης σε πιθανότητες. Υπάρχουν διάφορες μέθοδοι που μπορούν να εφαρμοστούν για τη συγκεκριμένη εργασία, όπως είναι η μέθοδος **Platt Scaling** από τους Platt et al. [35]. Αντί αυτού επιλέγουμε να χρησιμοποιήσουμε μία πρότυπη μέθοδο κατωφλιοποίησης.

3.6.3.2.1 Κανονικοποίηση των SVMs scores Ο προτεινόμενος τρόπος κανονικοποίησης βασίζεται σε μία σύνθετη κατωφλιοποίηση. Πιο συγκεκριμένα, για κάθε ‘υποταξινομητή’ i διαχωρίζουμε τον χώρο των χαρακτηριστικών σε δύο τμήματα, το θετικό και το αρνητικό τμήμα, αναλόγως με το αν τα δείγματα στα τμήματα αυτά ανήκουν ή όχι στην κατηγορία i , χρησιμοποιώντας το υπερεπίπεδο απόφασης. Τότε τα θετικά scores θα βρίσκονται στο θετικό τμήμα και τα αρνητικά στο αρνητικό. Η a priori πιθανότητα των κλάσεων υποτίθεται ότι έχει ομοιόμορφη κατανομή. Τότε, αν l_s είναι ο αριθμός των κλάσεων, ορίζουμε το μη γραμμικό μετασχηματισμό P_s , για τον οποίο ισχύει το ακόλουθο:

$$P_s(score) = \begin{cases} T \frac{score_{\min} - score}{score_{\min}}, & \text{if } score \leq 0 \\ (1 - T) \frac{score}{score_{\max}}, & \text{otherwise} \end{cases} \quad \text{with } T = \frac{1}{l_s} \quad (3.10)$$

Όπως παρατηρεί κανείς, ο αριθμός των κλάσεων καθορίζει το λόγο των διαστημάτων κανονικοποίησης. Λόγω αυτού του μη γραμμικού μετασχηματισμού, ο οποίος εξ ορισμού εκφράζει ‘πιθανότητες’ βάσει των ακραίων τιμών ενός και μόνου δείγματος, η μετρική που χρησιμοποιείται από του Fanello et al. δεν μπορεί να εφαρμοστεί για να ανιχνευθεί η αρχή και το τέλος της δράσης. Παρ’όλ’αυτά, μπορούμε να χρησιμοποιήσουμε ως σημεία εκκίνησης και λήξης των δράσεων αυτά που υπολογίζονται από τον αρχικό Cl_{dyn} ταξινομητή.

3.6.3.3 Διαισθητική Επεξήγηση Χρήσης του CDBIMM

Ο CDBIMM χρησιμοποιήθηκε για να υποβοηθηθεί η αναγνώριση δυναμικών χειρονομιών μέσα από την πιο εύρωστη αναγνώριση των παθητικών χειρονομιών. Η σύγκριση των δύο αυτών χώρων κλάσεων δεν μπορεί να γίνει με χρήση κάποιας άμεσης απεικόνισης ή απλής γραμμικής υπέρθεσης, λόγω του ότι εκ κατασκευής ο ένας χώρος είναι εικόνα ενός χώρου χαρακτηριστικών συνεχούς χρόνου ενώ ο άλλος χώρος είναι εικόνα ενός χώρου στατικών χαρακτηριστικών, για τα οποία έχει υποτεθεί ότι δεν υφίσταται χρονική εξάρτηση. Αυτή η λογική κατασκευής είναι άμεση συνέπεια των διαφορετικών απαιτήσεων αναγνώρισης που υπάρχουν για τους δύο τύπους χειρονομιών, οι οποίες και περιγράφονται στην Ενότητα 1.5.

Κεφάλαιο 4

Πειραματική Διαδικασία

Το κεφάλαιο αυτό περιλαμβάνει τα πειράματα που έγιναν, χρησιμοποιώντας διαφορετικούς συνδυασμούς χαρακτηριστικών, με ή χωρίς χρήση Αραιής Κωδικοποίησης, διαφορετικές παραμέτρους για τους ταξινομητές που περιγράφηκαν και διαφορετικές συνδεσμολογίες. Ο τρόπος με τον οποίο έλαβαν χώρα τα πειράματα ήταν καθολικός για κάθε ταξινομητή (χωρίς τη χρήση του CDBIMM), δηλαδή δοκιμάστηκαν όλοι οι παραπάνω συνδυασμοί μεθόδων έτσι, ώστε να υπάρξει και μία επεξήγηση για το λόγο που αποφασίστηκε να χρησιμοποιηθεί μία τόσο σύνθετη, και μάλλον υπερπλήρης ως προς τη βάση του χώρου των δράσεων, δομή ταξινόμησης όπως ο CDBIMM.

4.1 Δεδομένα

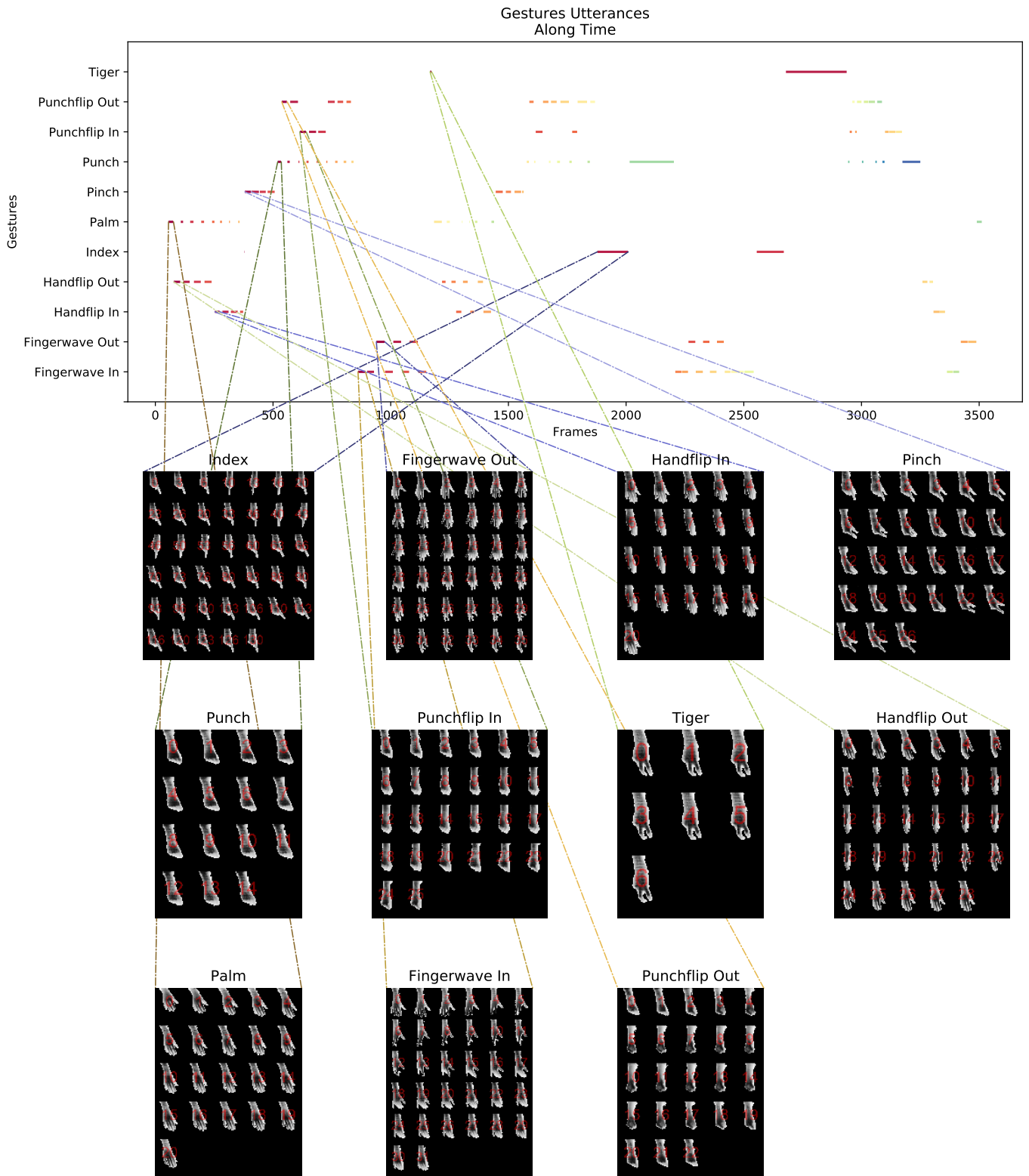
4.1.1 Γενική Περιγραφή

Λόγω των ιδιότυπων περιορισμών που επιβλήθηκαν, αλλά και επειδή υπάρχει ανάγκη να επιτρέψουμε τη χρήση και την εκπαίδευση του αλγορίθμου από οποιονδήποτε, τα χρησιμοποιούμενα δεδομένα προήλθαν από προσωπική εργασία. Σε αυτήν την κατεύθυνση, δημιουργήθηκε μία εφαρμογή με γραφικό περιβάλλον, η οποία επιτρέπει την ανάλυση αρχείων μορφής `rosbag`, τα οποία στην πράξη είναι βιντεοσκοπημένα δεδομένα από το Kinect μέσω του Kinect iai, την εξαγωγή της μάσκας του χεριού και την ετικετοποίηση των δεδομένων ανά εικόνα. Η δεικτοδότηση αποθηκεύεται σε μία λίστα (μέσα σε ένα αρχείο `csv`), μέσω της οποίας ο χρήστης μπορεί

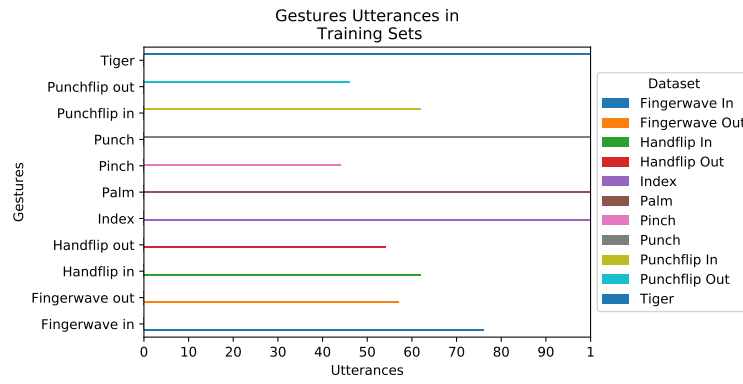
να διαχωρίσει το βίντεο στις διάφορες χειρονομίες, αποθηκεύοντας την κάθε μία σε έναν ειδικά διαμορφωμένο χώρο φακέλων μέσα στη μνήμη ως ένα σύνολο από εικόνες και ένα σύνολο από μάσκες επί αυτών (που αποκόπτουν το χέρι). Προς υποβοήθηση του χρήστη, φαίνεται και το μέρος του χεριού που έχει εξαχθεί, κατόπιν της προεπεξεργασίας, για να διαπιστώσει αυτός αν είναι κατάλληλη ή όχι η προσθήκη της εκάστοτε χειρονομίας μέσα στην κατασκευασθείσα βάση των χειρονομιών. Τα παραγόμενα csv αρχεία μπορούν να χρησιμοποιηθούν για την κατασκευή του ground truth. Γενικά επιτρέπεται η δεικτοδότηση ενός frame παραπάνω από μία φορές στην παραγόμενη λίστα, επιτρέποντας έτσι την ‘υπέρθεση’ πολλών χειρονομιών την ίδια στιγμή, αλλά αυτό δεν έγινε κατά τη διάρκεια της επισήμανσης, για χάρη απλότητας. Μια εκτενέστερη περιγραφή αυτής της εφαρμογής δίνεται στην Εν. 7.2. Για κάθε χειρονομία ελήφθη και ένα διαφορετικό βίντεο, το οποίο καταγράφει τις χειρονομίες από έναν χρήστη, μέσω της συσκευής του Kinect, με την βοήθεια του ROS και της βιβλιοθήκης Kinect iai.

4.1.2 Δεδομένα Training

Δημιουργήθηκε μια αρχική βάση δεδομένων εκπαίδευσης από εικόνες βάθους, η οποία περιέχει ένα λεξικό 11 διαφορετικών χειρονομιών, 4 παθητικών και 7 δυναμικών (βλ. Σχ. 4.1, 4.2). Για λόγους καλύτερης επεξήγησης, παρουσιάζονται στο Σχ. 4.1 οι διάφοροι τύποι χειρονομιών χρησιμοποιώντας το validation0 dataset, το οποίο αναφέρεται στην Ενότητα 4.1.3, με επισήμανση αυτών στο βάθος του χρόνου (ως προς τα frames). Οι χειρονομίες εκτελέστηκαν από έναν χρήστη. Κάθε δυναμική χειρονομία εκτελέστηκε περισσότερες από 40 φορές, με την κάθε εκτέλεση να αποτελείται από 20 περίπου στιγμιότυπα (frames). Ο ρυθμός δειγματοληψίας ήταν περίπου 30 FPS. Στο Σχήμα 4.2 παρουσιάζεται ο αριθμός των εμφανίσεων κάθε χειρονομίας μέσα στα δεδομένα εκπαίδευσης. Επειδή τα δεδομένα δεν είναι εξ αρχής απολύτως ισορροπημένα ως προς τον αριθμό των εμφανίσεων, πριν την εκπαίδευση οι συχνότητες των εμφανίσεων εξισώνονται στον αρχικά μικρότερο αριθμό εμφανίσεων που παρατηρείται, αφαιρώντας τα αντίστοιχα δείγματα.



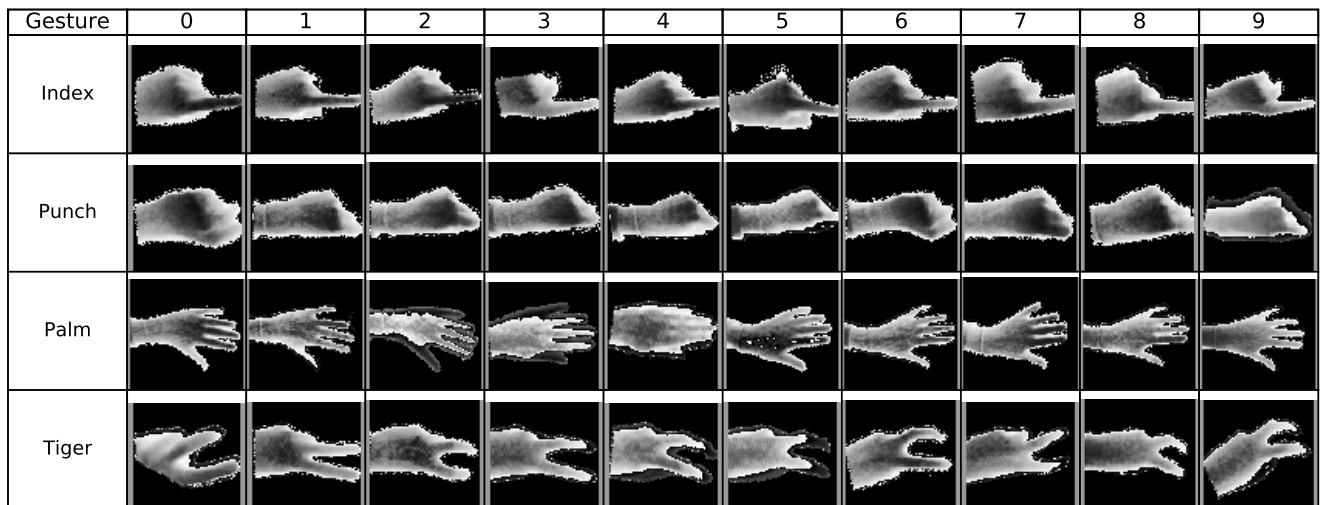
Σχήμα 4.1: Διάγραμμα Απεικόνισης των διαφόρων τύπων χειρονομιών, με αντιστοίχιση εμφανίσεων αυτών στο πεδίο των frames. Μπορεί να παρατηρήσει κανείς την έντονη χρονική μεταβλητότητα στα frames που αφορούν μία εμφάνιση δυναμικής χειρονομίας και τη στατικότητα στα frames που αφορούν μια εμφάνιση παθητικής χειρονομίας. Με κόκκινο νούμερο εμφανίζονται οι σχετικοί δείκτες του frame κάθε χειρονομίας, με αρχή το σημείο έναρξής της. Για λόγους ευκρίνειας, φαίνονται λιγότερα frames για εμφανίσεις με πάνω από 40 frames.



Σχήμα 4.2: Αριθμός Εμφάνσεων των Χειρονομιών ανά dataset εκπαίδευσης. Οι παθητικές χειρονομίες έχουν μία εμφάνιση στα αντίστοιχα dataset.

Gesture	0	1	2	3	4	5	6	7	8	9
Punchflip out										
Punchflip in										
Fingerwave out										
Fingerwave in										
Handflip out										
Handflip in										
Pinch										

Πίνακας 4.1: Λεξιικό Δυναμικών Χειρονομιών Δεδομένων Εκπαίδευσης. Φαίνονται 10 frames από μία εμφάνιση (utterance) κάθε χειρονομίας.

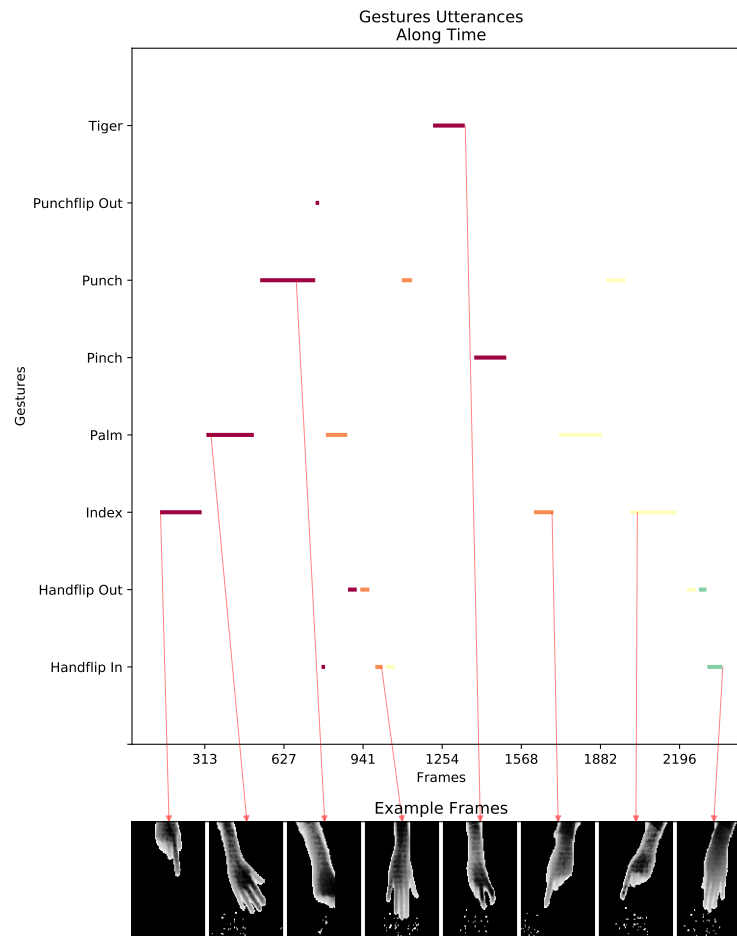


Πίνακας 4.2: Λεξικό **Παθητικών** Χειρονομιών Δεδομένων Εκπαίδευσης. Φαίνεται ότι η μορφή της χειρονομίας παραμένει σχεδόν σταθερή και μεταβάλλεται σε περιορισμένο εύρος ο προσανατολισμός της ως προς την κάμερα.

4.1.3 Δεδομένα Validation

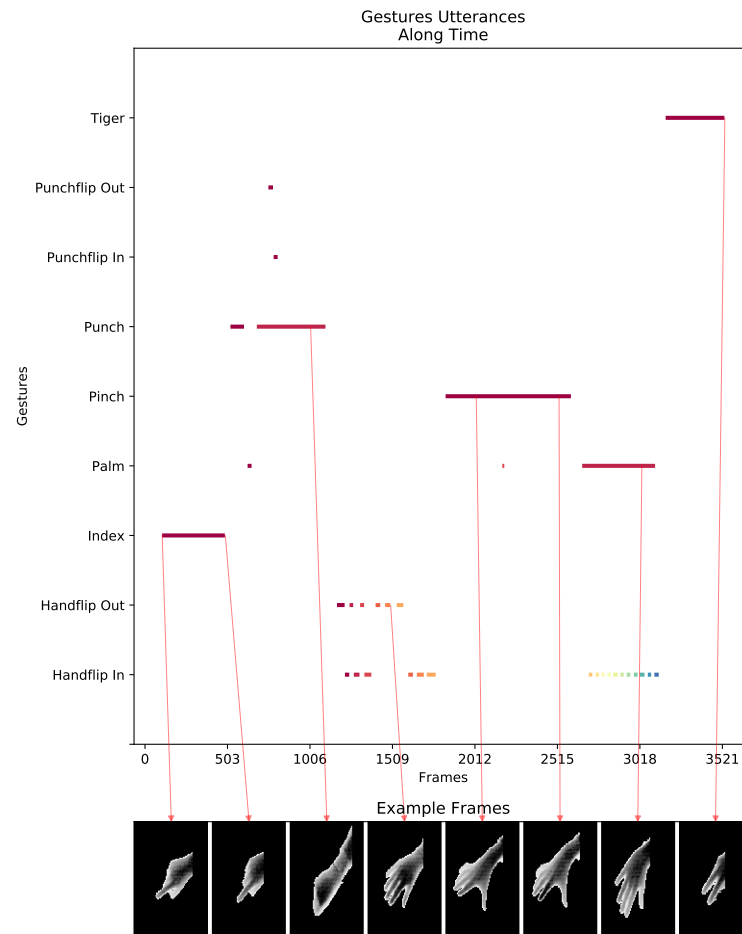
Για την επιλογή των υπερπαραμέτρων του αλγορίθμου χρησιμοποιήθηκε πάλι η αναπτυγμένη εφαρμογή για την ετικετοποίηση τριών(3) βίντεο δοκιμής, καθένα από τα οποία παρουσιάζει κάποιες ιδιομορφίες που αναφέρονται στη συνέχεια. Η επισήμανση των τριών βίντεο παρουσιάζεται στα Διαγράμματα 4.3, 4.4 και 4.5, όπου με διαφορετικό χρώμα απεικονίζονται οι διαφορετικές εμφανίσεις κάθε χειρονομίας, ενώ παράλληλα δίνεται ένα μικρό δείγμα από τα προεπεξεργασμένα δεδομένα και η αντιστοίχιση αυτών με χρήση βελών πάνω στα διαγράμματα.

- ‘Validation0’: Περιέχει σχεδόν εξ ολοκλήρου παθητικές χειρονομίες. Μετά από κάποια στιγμή (1500ο frame), τα χέρια εναλλάσσονται:



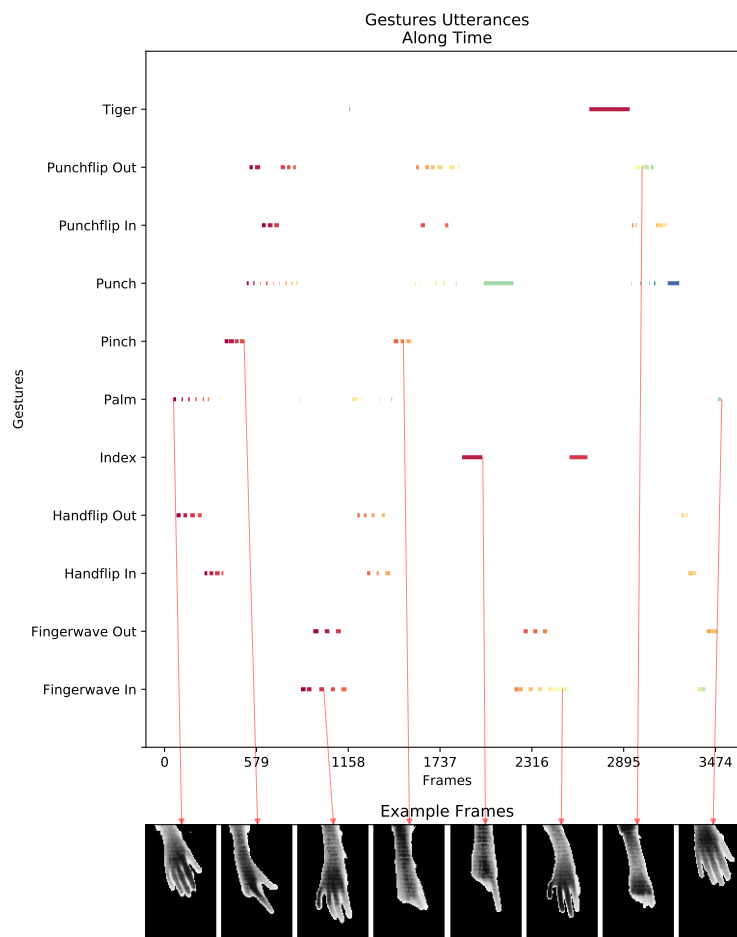
Σχήμα 4.3: Δεδομένα Validation0, όπως έχουν επισημανθεί στο βάθος του χρόνου (που μετράται σε frames). Δεν φαίνεται η πληροφορία ότι μετά το 1500ο frame τα χέρια εναλλάσσονται. Μπορεί να παρατηρήσει κανείς ότι δεν υπάρχουν πολλές εμφανίσεις δυναμικών χειρονομιών.

- ‘Validation1’: Περιέχει και παθητικές και κάποιες δυναμικές χειρονομίες. Μία περιοχή, η οποία παρουσιάζει υψηλή ανακλαστική ιδιότητα με κλίση, που κρίνεται απαγορευτική για τον αισθητήρα βάθους του Kinect, βρίσκεται στο πεδίο κίνησης του χεριού:



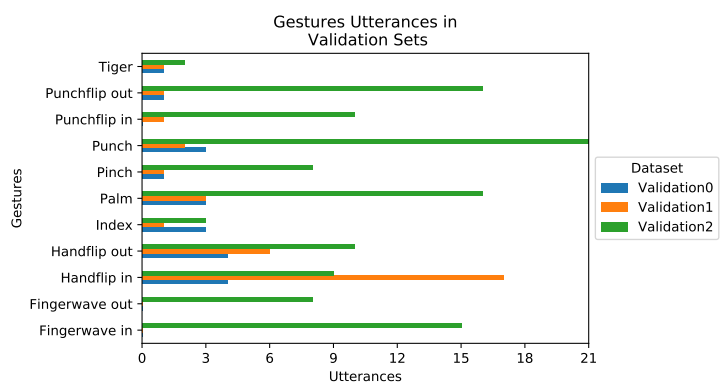
Σχήμα 4.4: Δεδομένα Validation1

- 'Validation2': Περιέχει όλες τις παθητικές και δυναμικές χειρονομίες. Το υπόβαθρο, το οποίο βρίσκεται πίσω από το κινούμενο χέρι, παρουσιάζει μικρές διαφορές βάθους:



Σχήμα 4.5: Δεδομένα Validation2

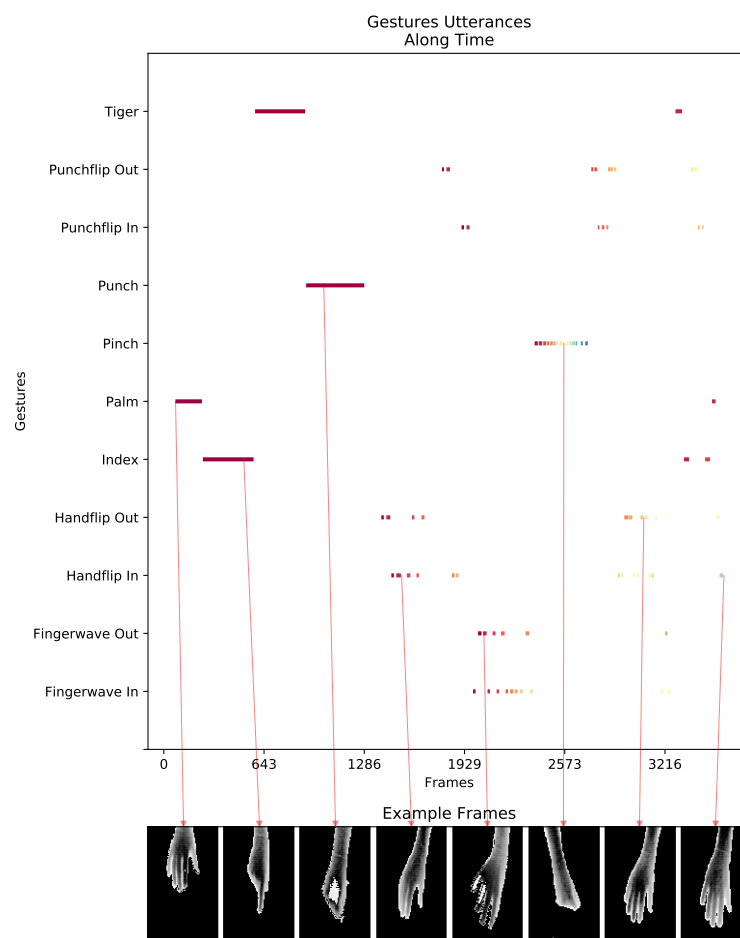
Τα βίντεο αυτά ελήφθησαν με παρόμοια μέση απόσταση από την κάμερα (80 εκατοστών) και παρόμοια γωνία λήψης ως προς το κατακόρυφο επίπεδο (15-20 μοίρες). Στη συνέχεια, φαίνεται η συχνότητα των χειρονομιών ανά Validation βίντεο.



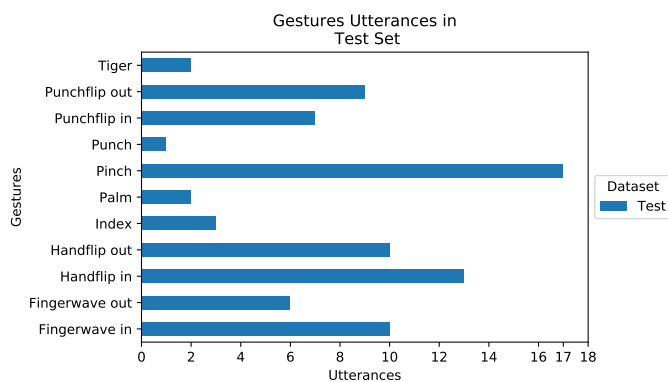
Σχήμα 4.6: Αριθμός Εμφανίσεων Χειρονομιών ανά validation dataset.

4.1.4 Δεδομένα Testing

Για λόγους δοκιμής κατασκευάστηκε ένα βίντεο, το οποίο χαρακτηρίζεται από αυξημένο βαθμό δυσκολίας, τόσο από την άποψη του υποβάθρου, το οποίο παρουσιάζει μεγάλες υψομετρικές διαφορές στο εύρος κίνησης του χεριού, όσο και από τη μεγάλη μέση απόσταση του χεριού από την κάμερα, η οποία είναι περίπου 1.5 μέτρα. Η γωνία λήψης παραμένει στα ίδια περίπου επίπεδα με αυτή στα βίντεο του validation. Στο Σχήμα 4.7 παρουσιάζεται το annotation των δεδομένων test, με μερικά παραδείγματα, ενώ στο Σχήμα 4.8 αναφέρονται οι συχνότητες εμφάνισης κάθε χειρονομίας.



Σχήμα 4.7: Δεδομένα Δοκιμής



Σχήμα 4.8: Αριθμός Εμφανίσεων Χειρονομιών στο test dataset.

4.2 Πειραματικό Πρωτόκολλο

Στην ενότητα αυτή αναφέρονται οι παράμετροι, οι οποίες χρησιμοποιούνται στα διάφορα στάδια του αλγορίθμου αλλά παρέμειναν ίδιες καθ' όλη τη διάρκεια του πειραματισμού. Ο τρόπος με τον οποίο βρέθηκαν οι ακόλουθες παράμετροι, οι οποίες απευθύνονται στο μέρος της λήψης και της προεπεξεργασίας των δεδομένων, είναι κατά βάσει ευριστικός, ενώ για τον καθορισμό των παραμέτρων της ταξινόμησης χρησιμοποιήθηκε η τεχνική **Cross Validation** με 3 folds, χρησιμοποιώντας τα δεδομένα εκπαίδευσης.

4.2.1 Λήψη και Αρχική Προεπεξεργασία Δεδομένων

Ο ρυθμός δειγματοληψίας των εικονοσειρών ήταν $f_s = 30$ FPS (*Frames/second*). Τα frames που χρησιμοποιήθηκαν για την αρχικοποίηση του αλγορίθμου ήταν $n_{init} = 30$. Το th_{noise} επιλέχτηκε έτσι, ώστε στην περίπτωση του Kinect να εκφράζει το 1 εκατοστό, οπότε τέθηκε ίσο με 10. Για το φίλτρο median που χρησιμοποιήθηκε για την προσθήκη απύσας πληροφορίας στην εικόνα χρησιμοποιήθηκε αριθμός επαναλήψεων $n_{med} = 2$ και μέγεθος μάσκας $s_{med} = 5$.

4.2.2 Εξαγωγή Μάσκας Χεριού

Για το GMM μοντέλο χρησιμοποιήθηκαν $M_{GMM} = 10$ γκαουσιανές συνιστώσες και $r_{GMM} = 20$ κατώφλι για την Mahalanobis απόσταση. Για να βρεθούν τα ενεργοποιημένα εικονοστοιχεία

έγινε χρήση κατωφλίου $c_{thrGMM} = 0.2$ και ο αριθμός των frames, μέσα από τον οποίο ενημερώνεται η μάσκα υποβάθρου, τέθηκε $n_{hist} = 4000$. Ο τετραγωνικός πυρήνας, ο οποίος χρησιμοποιήθηκε για την εκτέλεση του τελεστή Opening στο αποτέλεσμα του αλγορίθμου MOG2, είχε μέγεθος $s_{opker} = 5$.

Για την εφαρμογή της ταχείας σκελετοποίησης, χρησιμοποιήθηκαν οι τιμές για τις παραμέτρους τερματισμού $th_{nskel} = 50$, $th_{lskel} = 10$ και $th_{iskel} = 5$, ενώ η συγχώνευση δύο συνδέσμων επιβλήθηκε να γίνεται όταν ο ευρεθής σύνδεσμος έχει μήκος μικρότερο από $th_{wskel} = 50$ εικονοστοιχεία. Για την εξαγωγή της μάσκας του χεριού μέσα από τον ευρεθέντα σκελετό, το μήκος της άγνωστης πλευράς προσδιορίζεται πολλαπλασιάζοντας το γνωστό πλάτος με τον συντελεστή $r_{bskel} = 1.7$.

4.2.3 Επεξεργασία Χαρακτηριστικών

Ο αριθμός των bins των χρησιμοποιούμενων περιγραφητών τέθηκε ίσος με $n_{GHOG} = 64$, $n_{3DHOFF} = 4$ και $n_{ZHOFF} = 4$. Το μήκος της πλευράς της τετραγωνικής εισόδου στον περιγραφητή $3DXYPCA$ τέθηκε ίσο με $n_{PCA} = 32$. Η διάσταση n_{sparse} της εισόδου στον Αραιό Κωδικοποιητή τέθηκε ανάλογη της εισόδου, με συντελεστή αναλογίας $r_{sparse} = 2$. Η αρχικοποίηση των λεξικών του Sparse Coding έγινε τυχαία, με μηδενική μέση τιμή και μοναδιαία τυπική απόκλιση. Οι καταχωρητές που χρησιμοποιήθηκαν για τη συνένωση της εισόδου στον ταξινομητή των δυναμικών χειρονομιών έχουν μέγεθος $l_b = 10$, ενώ τέθηκε κατώφλι $th_{conf} = 0.25$. Από την μέθοδο PTPCA συγκρατούμε τα πρώτα $n_{PTPCA} = 3$ ιδιοδιανύσματα.

4.2.4 Ταξινόμηση και Επεξεργασία των Scores

Οι υπερπαραμέτροι $n_{RDF} = 29$ και $C_{SVM} = 83$ βρέθηκαν χρησιμοποιώντας 3-fold cross validation σε μία τυχαία από τις προτεινόμενες πειραματικές διάταξεις. Τα scores ομαλοποιήθηκαν με ένα φίλτρο μεγέθους $n_{fsc} = 5$ frames. Για τη μέθοδο CSTD χρησιμοποιήθηκαν $n_{sstd} = 20$ και $n_{bstd} = 40$, ενώ για τη μέθοδο CProb τέθηκε $th_{RDF} = 0.7$.

4.3 Τρόπος Χρήσης της Επισήμανσης των Δεδομένων (Ground Truth)

Όπως φαίνεται από την περιγραφή των δεδομένων, οι τύποι χειρονομιών έχουν αναμειχθεί εσκεμμένως στα βίντεο του Validation και του Test, έτσι ώστε να υπάρχει στο τέλος συγχρονισμένη και ενοποιημένη αναγνώριση των δύο αυτών τύπων. Ανάλογα με την ταξινόμηση που λαμβάνει χώρα, η επισήμανση των δεδομένων (**Ground Truth**) χρησιμοποιείται αφού έχει αφαιρεθεί από αυτή οι χειρονομίες που δεν ανήκουν στον τύπο που αναγνωρίζεται και οποιαδήποτε μέτρηση της ακρίβειας του αλγορίθμου γίνεται στις χειρονομίες που απομένουν στο Ground Truth.

4.4 Περιγραφή Χρησιμοποιούμενων Μετρικών

4.4.1 Είδη Μετρικών

Για να αποφανθούμε για την ποιότητα των αποτελεσμάτων χρησιμοποιούμε κατά βάση μία μετρική, την ακρίβεια (accuracy) η οποία εκφράζει τον λόγο των σωστά ταξινομημένων δειγμάτων προς τον συνολικό αριθμό των δειγμάτων που ταξινομήθηκαν.

Επιπλέον, υπολογίζουμε τον Πίνακα Σύγχυσης (Confusion Matrix). Το στοιχείο c_{ij} αυτού δείχνει το πλήθος των δειγμάτων που ανήκουν στην κατηγορία i αλλά ταξινομήθηκαν στην κατηγορία j . Συνήθως αυτός ο πίνακας είναι τετραγωνικός, όμως υπάρχει και η δυνατότητα προσθήκης μίας στήλης για την τοποθέτηση εκεί του πλήθους δειγμάτων που ταξινομήθηκαν σε μια κατηγορία που δεν υπάρχει στα δεδομένα, ενώ ανήκουν σε κάποια που υπάρχει. Η στήλη αυτή χρησιμοποιείται ακολούθως με την ετικέτα *Undefined*. Όσο περισσότερο διαγώνιος είναι αυτός ο πίνακας, τόσο καλύτερη είναι και η ταξινόμηση.

Σε περίπτωση που ο αριθμός των δειγμάτων στις κατηγορίες δεν είναι ισορροπημένος, τότε η μετρική της ακρίβειας δεν δίνει καλό αποτέλεσμα. Σε αυτήν την κατεύθυνση, χρησιμοποιήσαμε για την υποστήριξη των πειραμάτων και τη μετρική ονόματι F1-Scores. Αυτή εκφράζει τον λόγο:

$$F_1 = 2 \frac{precision \times recall}{precision + recall}$$

Οι όροι **precision** και **recall** απευθύνονται κατά βάση σε δυαδική ταξινόμηση, έστω θετικών και αρνητικών δειγμάτων. Ως precision ορίζεται ο λόγος του πλήθους των πραγματικά θετικών δειγμάτων που βρέθηκαν θετικά προς τον συνολικό αριθμό των δειγμάτων. Ως recall ορίζεται ο λόγος του αριθμού των πραγματικών θετικών δειγμάτων προς τον αριθμό των συνολικών ταξινομημένων ως θετικά δειγμάτων. Λόγω της φύσης αυτών των μετρικών, η F1-Scores είναι ένα διάλυμα στον χώρο που μελετάμε και χρησιμοποιείται για την επίβλεψη της ικανότητας της ταξινόμησης για κάθε χειρονομία ξεχωριστά.

4.4.2 Χώροι Υπολογισμού Μετρικών

Ο υπολογισμός των παραπάνω μετρικών έγινε σε δύο χώρους ταξινόμησης, για να καταστεί αντιληπτή η ικανότητα ταξινόμησης τόσο στο **επίπεδο των frames**, όσο και στο **επίπεδο εμφανίσεων των χειρονομιών**. Στο επίπεδο των frames (θεώρηση **micro**), υπολογίστηκαν οι μετρικές προσμετρώντας κάθε στιγμιότυπο ξεχωριστά, βάσει του ορισμένου Ground Truth. Στο επίπεδο των εμφανίσεων (utterances) (θεώρηση **macro**), θεωρήθηκε ως κατηγορία ταξινόμησης εμφάνισης μιας χειρονομίας, της οποίας τα όρια είναι γνωστά από το Ground Truth, αυτή που εμφανίζεται παραπάνω από κάποιο ποσοστό frames στα γνωστά όρια. Το ποσοστό των frames τεθηκε στο 50%. Σε περίπτωση που σε μία εμφάνιση του Ground Truth δεν υπάρχει κάποια κατηγορία που να ικανοποιεί αυτή τη συνθήκη, η εμφάνιση θεωρείται ότι είναι αρνητικό δείγμα για όλες τις υπάρχουσες χειρονομίες. Η μέθοδος αυτή καλύπτεται από τους Ruiloba et al. [38]

4.5 Εκτέλεση Πειραμάτων Ταξινόμησης

Τα ακόλουθα πειράματα διεξήχθησαν πάνω στα Validation δεδομένα που έχουμε κατασκευάσει και αναφέρονται στην Ενότητα 4.1.3. Στόχος των πειραμάτων αυτών είναι ο καθορισμός της καλύτερης διάταξης, που παράγεται από τα εργαλεία και τις μονάδες που αναφέρονται στο Κεφάλαιο 3.

4.6 Πειραματισμός πάνω στην Ταξινόμηση Δυναμικών Χειρονομιών χωρίς το μοντέλο CDBIMM

Παρακάτω αναφέρονται οι παράμετροι και οι αντίστοιχες τιμές που έλαβαν, δημιουργώντας το πλήρες συνδεδεμένο πειραματικό δίκτυο, για την περίπτωση της ταξινόμησης των δυναμικών χειρονομιών, οι οποίες αποτέλεσαν και το πιο σύνθετο και απαιτητικό πρόβλημα ταξινόμησης:

- **Περιγραφητές:** {GHOG, ZHOF}, {3DHOF, GHOG}, {3DXYPKA, ZHOF}, {3DHOF, 3DXYPKA}, {3DXYPKA, GHOG, ZHOF} και {3DHOF, 3DXYPKA, GHOG}
- **Sparse Coding:** Στο επίπεδο των χαρακτηριστικών, στο επίπεδο των buffers και καθόλου
- **Post Time PCA:** Με ή χωρίς χρονικό PCA μετά την εξαγωγή των buffers
- **Ταξινομητές:** RDF και Linear SVM
- **Μέθοδοι Επεξεργασίας Scores:** CProb και CSTD

Στους πειραματισμούς δεν συμμετείχε το μοντέλο CDBIMM.

4.6.1 Αποτελέσματα Δοκιμών

Οι διατάξεις με τις καλύτερες μέσες μετρικές στα δεδομένα Validation φαίνονται επιγραμματικά στον Πίνακα 4.3. Ως πρωτεύουσα μετρική για την ταξινόμηση χρησιμοποιήθηκε η **Micro Accuracy**, η ακρίβεια στο επίπεδο των frames δηλαδή. Στο Παρ. C.1, παρουσιάζονται τα αποτελέσματα της καλύτερης ευρεθείσας διάταξης για τα δεδομένα Validation.

	Classifiers Used	Descriptors	Post Scores Processing Method	PTPCA	Sparsecoding Level	Mean Micro Accuracy	Mean Macro Accuracy
0	SVM	['GHOG', 'ZHOF']	CSTD	Not Used	None	0.685829	0.582336
1	RDF	'3DXYPCA', '3DHOF']	CSTD	Not Used	None	0.683028	0.502849
2	RDF	['GHOG', '3DHOF']	CSTD	Not Used	None	0.665245	0.495252

Πίνακας 4.3: Ακρίβειες των 3 καλύτερων δοκιμασμένων διατάξεων στα δεδομένα Validation για την ταξινόμηση των **δυναμικών** χειρονομιών, σε επίπεδο frames και εμφανίσεων. Οι στήλες *PTPCA* και *Sparsecoding Level* αναφέρονται για πληρότητα, αν και όπως φαίνεται δεν χρησιμοποιούνται από κανένα από τα καλύτερα συστήματα.

Λαμβάνοντας από τον πειραματισμό την καλύτερη διάταξη που παρατηρήθηκε, τη δοκιμάσαμε στα δεδομένα του test. Στο Σχήμα 4.9 φαίνεται το αποτέλεσμα της δοκιμής, με διαφορετικά χρώματα να υποδεικνύουν τη μεταβολή των εμφανίσεων μέσα στο Ground Truth. Στο Διάγραμμα 4.10 δείχνουμε τη μετρική που χρησιμοποιήθηκε από τη μέθοδο CSTD για να εξάγει τα σημεία έναρξης και λήξης των χειρονομιών. Στους Πίνακες 4.6 και 4.4 παρουσιάζεται ο πίνακας σύγχυσης ως προς τα frames και ως προς τα utterances αντίστοιχα. Ομοίως, στους Πίνακες 4.7 και 4.5 φαίνονται η Ακρίβεια και τα F1-Scores ανά χειρονομία, για τον χώρο των frames και των εμφανίσεων αντίστοιχα.

Macro Metrics

		Predicted							Undefined
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	
Actual	Pinch	11	0	0	5	0	0	1	0
	Handflip out	0	0	9	0	2	0	0	0
	Handflip in	0	0	10	0	3	0	0	0
	Punchflip in	0	0	0	2	3	0	0	0
	Punchflip out	1	0	1	0	11	0	0	0
	Fingerwave in	0	1	1	0	2	0	6	0
	Fingerwave out	0	1	0	0	1	0	4	0

Πίνακας 4.4: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics							
Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
0.759	0.000	0.588	0.333	0.629	0.000	0.471	0.507

Πίνακας 4.5: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

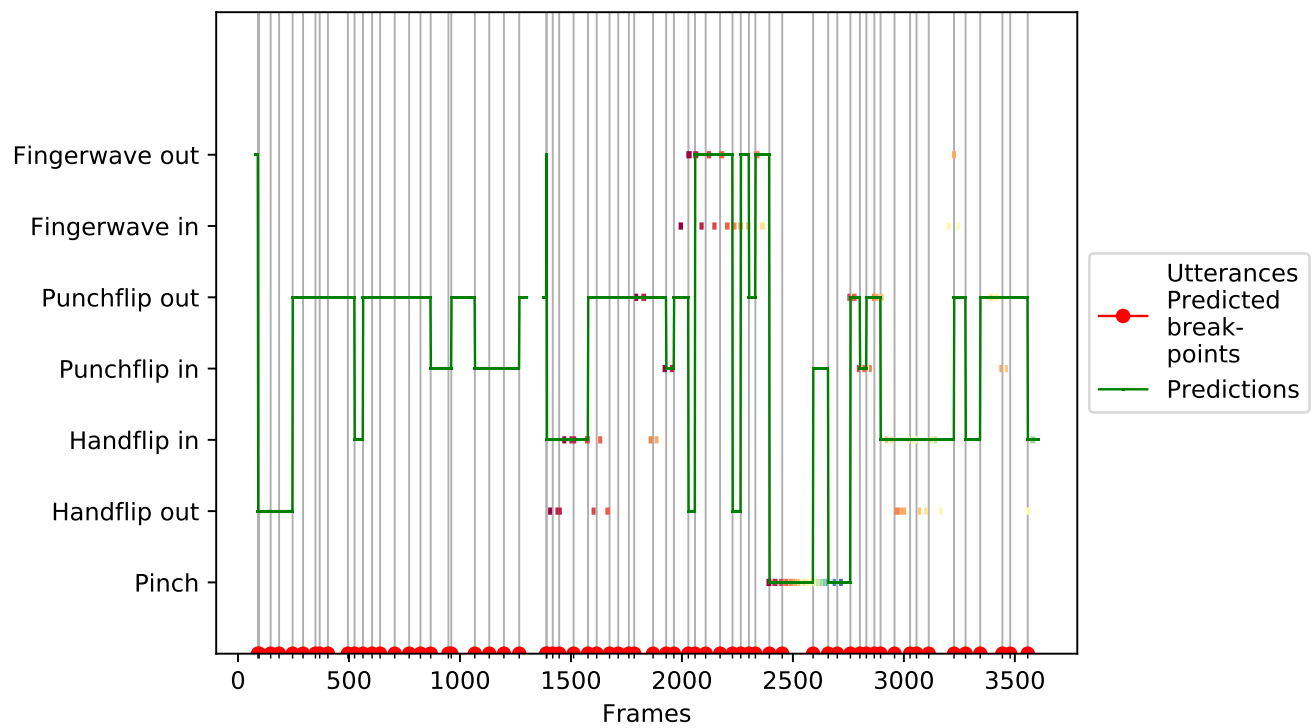
Micro Metrics

		Predicted							
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Undefined
Actual	Pinch	216	0	0	67	0	0	13	0
	Handflip out	0	0	172	0	52	0	0	0
	Handflip in	0	0	196	0	69	0	0	0
	Punchflip in	0	0	0	41	77	0	0	0
	Punchflip out	14	0	16	0	145	0	0	0
	Fingerwave in	0	31	19	0	49	0	119	0
	Fingerwave out	0	25	9	0	26	0	85	0

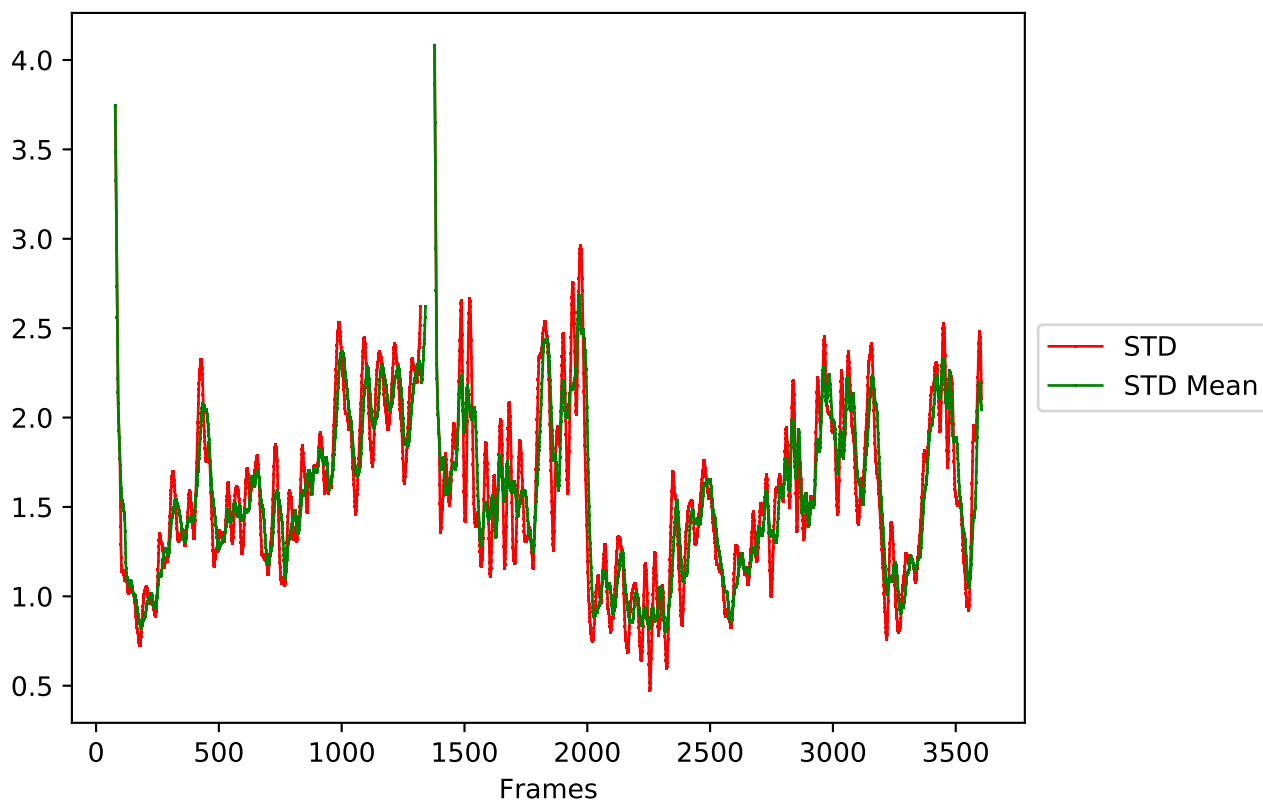
Πίνακας 4.6: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames.

Micro Metrics							
Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
0.821	0.000	0.579	0.363	0.489	0.000	0.470	0.474

Πίνακας 4.7: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames.



Σχήμα 4.9: Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών , ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.



Σχήμα 4.10: Δεδομένα Test: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances) των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών .

4.7 Πειραματισμός πάνω στην Ταξινόμηση Παθητικών Χειρονομιών

Όπως και με τις δυναμικές χειρονομίες, τα πειράματα της ταξινόμησης των παθητικών χειρονομιών διεξήχθησαν πάνω στις τρεις Validation εικονοσειρές που αναφέρονται στην Ενότητα 4.1.3. Για την περίπτωση των παθητικών, ο αριθμός των πειραμάτων ήταν μικρότερος, με τους συνδυασμούς που δοκιμάστηκαν να αναφέρονται στη συνέχεια:

- **Περιγραφητές:** {GHOG}, {3DXYPEA} και {3DXYPEA, GHOG}
- **Sparse Coding:** Στο επίπεδο των χαρακτηριστικών και καθόλου.
- **Ταξινομητές:** RDF και Linear SVM
- **Μέθοδοι Επεξεργασίας Scores:** CProb και CSTD

Εδώ είναι προφανές ότι δεν χρησιμοποιήθηκε το μοντέλο CDBIMM, αφού, όπως έχει αναφερθεί στην Ενότητα 3.6.3, αυτό έχει κατασκευαστεί για τη βελτίωση της αναγνώρισης των δυναμικών χειρονομιών.

4.7.1 Αποτελέσματα Δοκιμών

Οι διατάξεις που επέδειξαν τις καλύτερες μέσες μετρικές στα δεδομένα Validation φαίνονται αναφορικά στον Πίνακα 4.8. Ως μετρική για την ταξινόμηση χρησιμοποιήθηκε η **Macro Accuracy**, η ακρίβεια στο επίπεδο των frames δηλαδή. Στο Παρ. C.2, παρουσιάζονται τα αποτελέσματα της καλύτερης ευρεθείσας διάταξης για τα δεδομένα Validation.

	Classifiers Used	Descriptors	Post Scores Processing Method	Sparsecoding Level	Mean Micro Accuracy	Mean Macro Accuracy
0	RDF	3DXYPKA	CSTD	None	0.866902	0.876190
1	RDF	['GHOG', '3DXYPKA']	CSTD	None	0.851716	0.819048
2	SVM	3DXYPKA	CSTD	None	0.843460	0.877778

Πίνακας 4.8: Ακρίβειες των 3 καλύτερων δοκιμασμένων διατάξεων στα δεδομένα Validation για την ταξινόμηση των **παθητικών** χειρονομιών, σε επίπεδο frames και εμφανίσεων. Η στήλη *Sparsecoding Level* αναφέρεται για πληρότητα, αν και όπως φαίνεται δεν χρησιμοποιείται από κανένα από τα καλύτερα συστήματα.

Όπως και στις δυναμικές, λάβαμε από τον πειραματισμό την καλύτερη διάταξη που παρατηρήθηκε και τη δοκιμάσαμε στα δεδομένα του test. Στο Σχήμα 4.11 φαίνεται το αποτέλεσμα της δοκιμής, με διαφορετικά χρώματα να υποδεικνύουν τη μεταβολή των εμφανίσεων μέσα στο Ground Truth. Στο Διάγραμμα 4.12 δείχνουμε τη μετρική που χρησιμοποιήθηκε από τη μέθοδο CSTD για να εξάγει τα σημεία έναρξης και λήξης των χειρονομιών. Στους Πίνακες 4.11 και 4.9 παρουσιάζεται ο πίνακας σύγχυσης ως προς τα frames και ως προς τα utterances αντίστοιχα. Ομοίως, στους Πίνακες 4.12 και 4.10 φαίνονται η Ακρίβεια και τα F1-Scores ανά χειρονομία, για τον χώρο των frames και των εμφανίσεων αντίστοιχα.

Macro Metrics

		Predicted				
		Index	Punch	Palm	Tiger	Undefined
Actual	Index	2	1	0	0	0
	Punch	0	1	0	0	0
	Palm	0	0	2	0	0
	Tiger	0	0	0	2	0

Πίνακας 4.9: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics				
Index	Punch	Palm	Tiger	Accuracy
0.800	0.667	1.000	1.000	0.875

Πίνακας 4.10: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

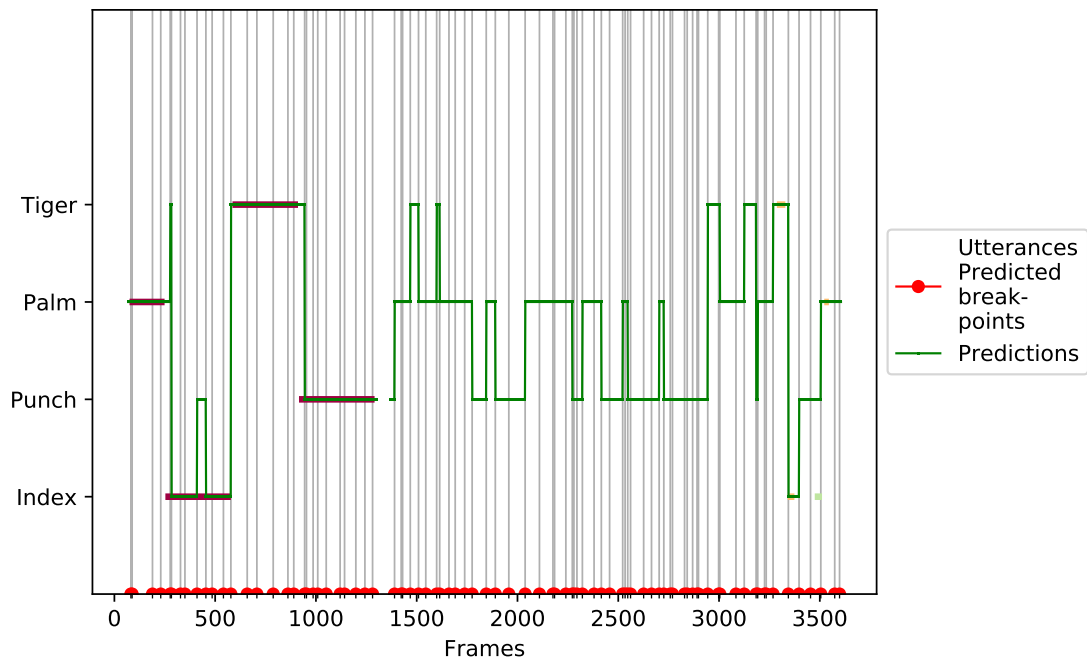
Micro Metrics

		Predicted				
		Index	Punch	Palm	Tiger	Undefined
Actual	Index	283	74	32	11	0
	Punch	0	349	0	28	0
	Palm	0	0	201	0	0
	Tiger	0	0	0	371	0

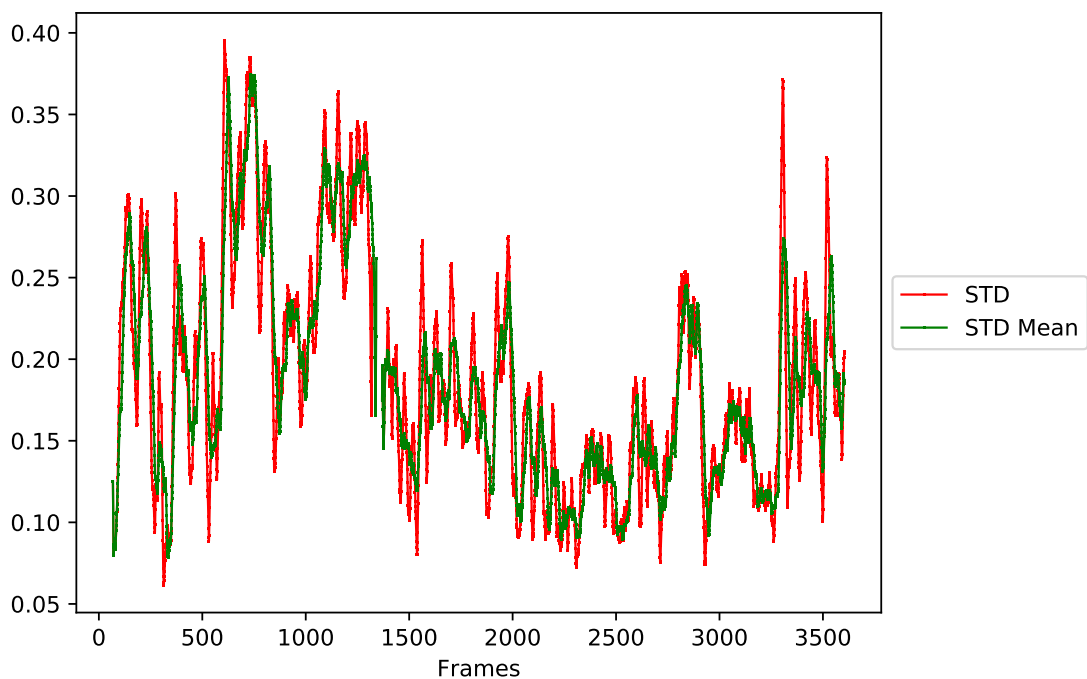
Πίνακας 4.11: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames.

Micro Metrics				
Index	Punch	Palm	Tiger	Accuracy
0.829	0.873	0.926	0.950	0.893

Πίνακας 4.12: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames.



Σχήμα 4.11: Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών, ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.



Σχήμα 4.12: Δεδομένα Test: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances) των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών.

Κεφάλαιο 5

Χρήση Αποτελεσμάτων Πειραματικής Διαδικασίας - Επέκταση

5.1 Γενικά

Το κεφάλαιο αυτό αποτελεί συνέχεια του πειραματισμού που εκτελέστηκε στο Κεφάλαιο 4. Γνωρίζοντας πλέον την καλύτερη διάταξη για την ταξινόμηση, η οποία ενέχει τα υποσυστήματα του Πίνακα 5.1, μπορούμε να περάσουμε στο επόμενο, πιο σύνθετο στάδιο, που αφορά τη βελτίωση των παρατηρούμενων αποτελεσμάτων, μέσω της χρήσης του CDBIMM, τον πρότυπο ταξινομητή δυναμικών χειρονομιών που έχει περιγραφεί στην Ενότητα 3.6.3.

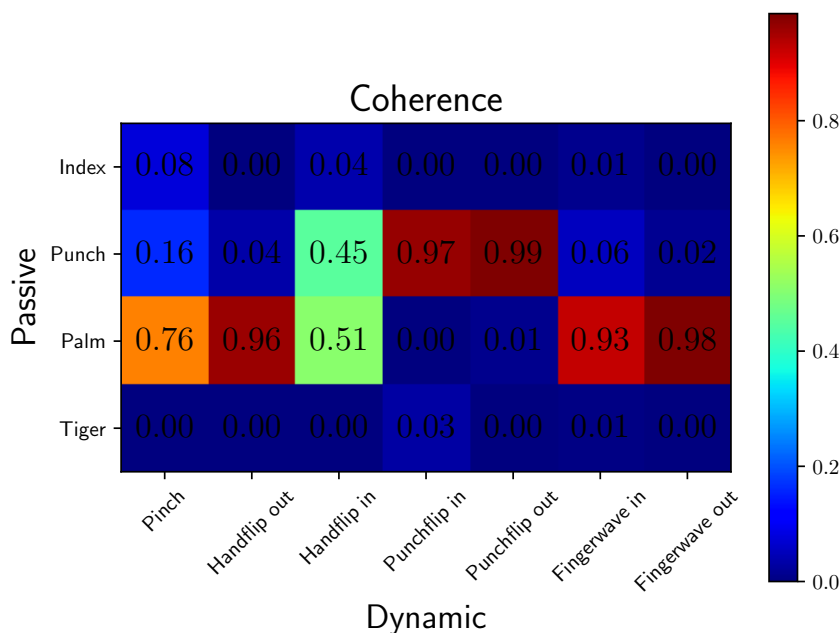
	Classifiers Used	Descriptors	PTPCA	Post Scores Processing Method	Sparsecoding Level	Mean Micro Accuracy	Mean Macro Accuracy
CI_{dyn}	SVM	['GHOG', 'ZHOF']	Not Used	CSTD	None	0.686	0.582
CI_{pas}	RDF	3DXYPECA		CSTD	None	0.867	0.876

Πίνακας 5.1: Οι καλύτεροι επιλεγμένοι ταξινομητές, ο ταξινομητής δυναμικών χειρονομιών CI_{pas} και ο ταξινομητής δυναμικών χειρονομιών CI_{dyn} βάσει του πειραματισμού που έλαβε χώρα. Με μαύρο αναγράφεται το κελί του PTPCA, αφού δεν εφαρμόζεται στην περίπτωση του πειραματισμού του ταξινομητή CI_{pas}

Κατόπιν, χρησιμοποιώντας το μοντέλο αυτό στη θέση του ταξινομητή δυναμικών χειρονομιών Cl_{dyn} , εκτελούμε δύο διαφορετικές προσεγγίσεις διαχείρισης της εξόδου της ταξινόμησης, οι οποίες περιγράφονται σε επόμενη ενότητα.

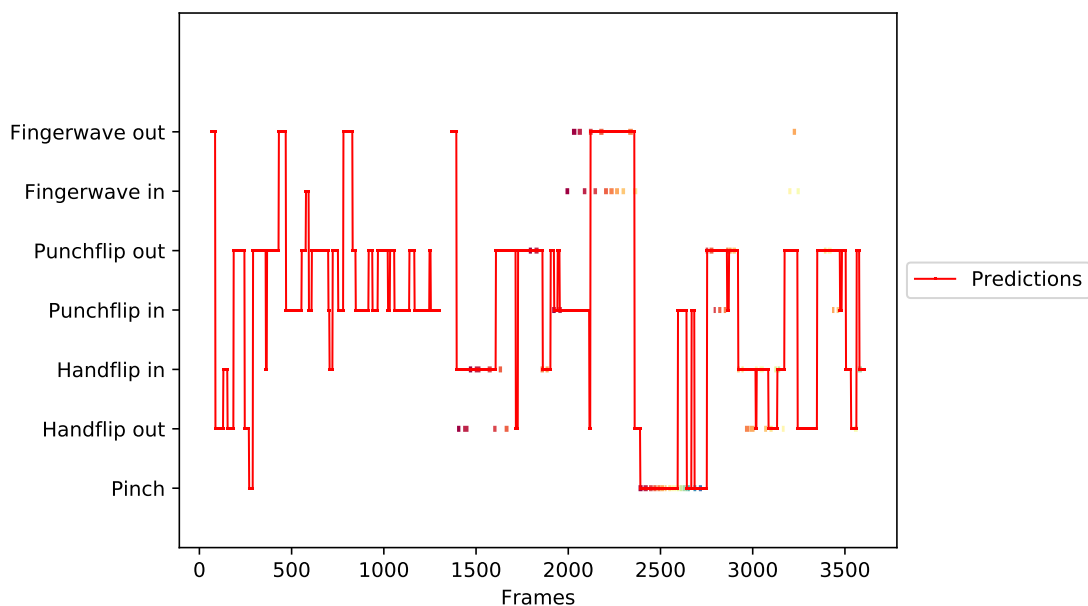
5.2 Εφαρμογή Μοντέλου CDBIMM

Επιλέγοντας τους δύο καλύτερους ταξινομητές από τους παραπάνω, έναν από τις δυναμικές χειρονομίες και έναν από τις παθητικές όπως φαίνεται στον Πίνακα 5.1, οδηγήσαμε τα scores που δίνουν αυτοί στο μοντέλο CDBIMM. Αρχικά, για την εκπαίδευση του μοντέλου και την παραγωγή του Πίνακα Συνεκτικότητας CM (βλ. Πιν. 5.2) βάσει του αλγορίθμου της Ενότητας 3.6.3.1, χρησιμοποιήσαμε τα δεδομένα Training (βλ. Εν. 4.1.2). Ο Cl_{dyn} είναι SVM ταξινομητής, οπότε και χρειάστηκε να προβούμε στη μετατροπή των scores αυτού σε πιθανότητες, όπως περιγράφεται στην Ενότητα 3.6.3.2. Εξαιτίας του γεγονότος αυτού, κατέστη απαγορευτική η χρήση της τεχνικής CSTD, αφού η τυπική απόκλιση των scores του SVM έπαψε να είναι χρονικά εξαρτημένη μετά την μη γραμμική μετατροπή τους. Αντί αυτής, χρησιμοποιήθηκε η μέθοδος **CProb**.



Πίνακας 5.2: Πίνακας Συνεκτικότητας (CM) Χειρονομιών, προερχόμενος από τις δοκιμές του Επιλεγμένου Ταξινομητή Παθητικών Χειρονομιών στα Δεδομένα Εκπαίδευσης των Δυναμικών Χειρονομιών.

Αφού εκπαιδεύσαμε το σύστημα, εφαρμόσαμε το εκπαιδευμένο μοντέλο CDBIMM απευθείας στα δεδομένα Testing. Στο Σχήμα 5.1 φαίνεται το αποτέλεσμα της δοκιμής, με διαφορετικά χρώματα να υποδεικνύουν τη μεταβολή των εμφανίσεων μέσα στο Ground Truth. Στους Πίνακες 5.5 και 5.3 παρουσιάζεται ο πίνακας σύγκρισης ως προς τα frames και ως προς τα utterances αντίστοιχα. Ομοίως, στους Πίνακες 5.6 και 5.4 φαίνονται η Ακρίβεια και τα F1-Scores ανά χειρονομία, για τον χώρο των frames και των εμφανίσεων αντίστοιχα.



Σχήμα 5.1: Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον CDBIMM Ταξινομητή (των **δυναμικών** χειρονομιών), ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων.

Macro Metrics

		Predicted							
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Undefined
Actual	Pinch	13	0	0	4	0	0	0	0
	Handflip out	0	2	7	0	2	0	0	0
	Handflip in	0	1	10	0	2	0	0	0
	Punchflip in	0	0	0	2	3	0	0	0
	Punchflip out	0	0	0	1	12	0	0	0
	Fingerwave in	0	2	0	2	1	0	5	0
	Fingerwave out	0	0	0	2	1	0	3	0

Πίνακας 5.3: Δεδομένα Test: Πίνακας Σύγκρισης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον CDBIMM Ταξινομητή (των **δυναμικών** χειρονομιών) ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics							
Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
0.867	0.250	0.667	0.250	0.706	0.000	0.429	0.560

Πίνακας 5.4: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον CDBIMM Ταξινομητή (των **δυναμικών** χειρονομιών) ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

		Micro Metrics							
		Predicted							
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Undefined
Actual	Pinch	235	9	0	52	0	0	0	0
	Handflip out	0	33	151	0	40	0	0	0
	Handflip in	0	21	200	0	44	0	0	0
	Punchflip in	0	0	0	25	93	0	0	0
	Punchflip out	7	0	0	10	158	0	0	0
	Fingerwave in	0	31	0	43	24	0	120	0
	Fingerwave out	0	6	0	53	20	0	66	0

Πίνακας 5.5: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον CDBIMM Ταξινομητή (των **δυναμικών** χειρονομιών) ως προς τα frames.

Micro Metrics							
Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
0.874	0.204	0.649	0.166	0.570	0.000	0.399	0.498

Πίνακας 5.6: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον CDBIMM Ταξινομητή (των **δυναμικών** χειρονομιών) ως προς τα frames.

Για καλύτερη επισκόπηση συνενώνουμε τα αποτελέσματα του CDBIMM με αυτά του χρησιμοποιούμενου Cl_{dyn} στα δεδομένα Test στους ακόλουθους δύο πίνακες:

Macro Metrics CDBIMM- Cl_{dyn} comparison								
	Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
Cl_{dyn}	0.759	0.000	0.588	0.333	0.629	0.000	0.471	0.507
CDBIMM	0.867	0.250	0.667	0.250	0.706	0.000	0.429	0.560

Πίνακας 5.7: Σύγκριση Επιδόσεων των Ταξινομητών Cl_{dyn} και CDBIMM πάνω στα δεδομένα Test, ως προς τις εμφανίσεις των χειρονομιών.

Micro Metrics CDBIMM- Cl_{dyn} comparison								
	Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
Cl_{dyn}	0.821	0.000	0.579	0.363	0.489	0.000	0.470	0.474
CDBIMM	0.874	0.204	0.649	0.166	0.570	0.000	0.399	0.498

Πίνακας 5.8: Σύγκριση Επιδόσεων των Ταξινομητών Cl_{dyn} και CDBIMM πάνω στα δεδομένα Test, ως προς τα frames.

Όπως μπορεί να δει κανείς, το μοντέλο CDBIMM φαίνεται να είναι πιο εύρωστο στην ανίχνευση εμφανίσεων σε σχέση με το Cl_{dyn} , ενώ καταφέρνει να ανιχνεύσει μία ακόμα χειρονομία, την ‘Handflip out’.

5.3 Διαχείριση Εξόδου Ταξινόμησης

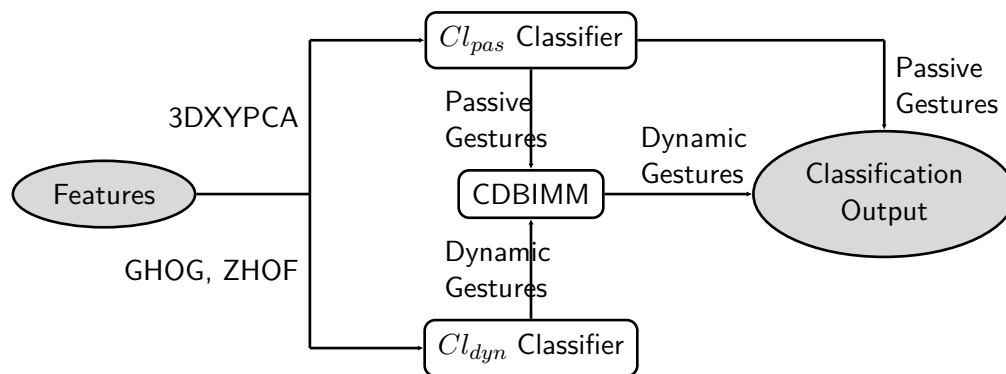
Μέχρι αυτό το σημείο έχουμε περιγράψει μία δομή ταξινόμησης, της οποίας δεν είναι εντελώς προφανές ποια ακριβώς είναι η έξοδος. Προς αυτήν την κατεύθυνση, προτείνουμε και εφαρμόζουμε τις ακόλουθες δύο προσεγγίσεις ταξινόμησης, οι οποίες εκφράζουν πλήρως και τη μορφή της εξόδου:

1. **Σύγχρονη Ταξινόμηση (In Sync)**. Κατά τη διάρκεια αυτής, το σύστημα ταξινόμησης εξάγει δύο προβλέψεις από την είσοδο. Η πρώτη πρόβλεψη χαρακτηρίζει την παθητική χειρονομία που παρατηρείται (προερχόμενη από τον ταξινομητή Cl_{pas}), ενώ η δεύτερη πρόβλεψη χαρακτηρίζει την δυναμική χειρονομία που παρατηρείται (προερχόμενη από το μοντέλο CDBIMM) σε βάθος χρόνου $l_b = 10$ frames.
2. **Συνδυασμένη Ταξινόμηση (Combined)**. Αποτελεί μια επέκταση της **σύγχρονης ταξινόμησης**. Επιστρατεύουμε έναν νέο ταξινομητή RDF, τον οποίο και ονομάζουμε, για να αποφύγουμε τη σύγχυση, **CRDF** και ο οποίος έχει $n_{RDF} = 29$ παραμέτρους. Η έξοδος που προκύπτει από τον Cl_{pas} και το CDBIMM υπό μορφή scores (πριν την εφαρμογή κάποιας τεχνικής επεξεργασίας των scores για εξαγωγή των προβλέψεων) χρησιμεύει ως είσοδος για τον ταξινομητή CRDF, ο οποίος και αναλαμβάνει να δώσει την τελική πρόγνωση του συστήματος. Αυτή η πρόγνωση εκφράζει πλέον όλο τον χώρο των χειρονομιών.

Γενικά, θεωρούμε ότι, λόγω του ότι η δυναμική χειρονομία αποτελεί υπέρθεση παθητικών χειρονομιών, η πρώτη προσέγγιση ταξινόμησης είναι αυτή που αποτυπώνει καλύτερα την πραγματικότητα. Παρ’όλ’αυτά, η συνδυασμένη ταξινόμηση προσφέρει μονοσήμαντη σχέση μεταξύ του συνολικού χώρου των κλάσεων και του χώρου των χαρακτηριστικών, οπότε και είναι πιο ‘συμβατική’ από την σύγχρονη ταξινόμηση.

5.3.1 Σύγχρονη Ταξινόμηση (In Sync)

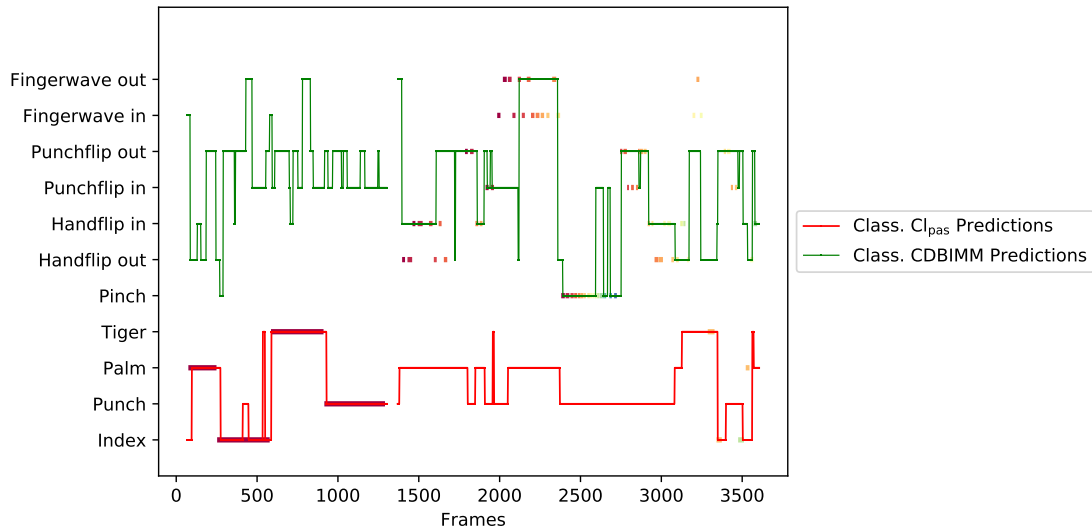
Το σχεδιάγραμμα του συστήματος της σύγχρονης ταξινόμησης παρουσιάζεται στη συνέχεια:



Σχήμα 5.2: Σύγχρονη Ταξινόμηση (In Sync). Με ελλείψεις φαίνονται η είσοδος και η έξοδος του συστήματος.

Αυτή η προσέγγιση δεν χρειάζεται κάποια περαιτέρω επεξεργασία της εξόδου των ήδη υπάρχοντων σταδίων ταξινόμησης, αρκεί να υπερθέσουμε τα αποτελέσματα των δύο ταξινομητών (του Cl_{pas} και του CDBIMM). Εφαρμόζοντας αυτήν την απλή προσέγγιση δοκιμάζοντας το σύστημα στα δεδομένα Test, λαμβάνουμε τα ακόλουθα αποτελέσματα.

Στο Σχήμα 5.3 φαίνεται το αποτέλεσμα της δοκιμής, με διαφορετικά χρώματα να υποδεικνύουν τη μεταβολή των εμφανίσεων μέσα στο Ground Truth. Στους Πίνακες 5.11 και 5.9 παρουσιάζεται ο πίνακας σύγκρισης ως προς τα frames και ως προς τα utterances αντίστοιχα. Ομοίως, στους Πίνακες 5.12 και 5.10 φαίνονται η Ακρίβεια και τα F1-Scores ανά χειρονομία, για τον χώρο των frames και των εμφανίσεων αντίστοιχα. Ακόμα στον Πίνακα 5.13 εμφανίζονται διαχωρισμένες οι ακρίβειες των συμμετεχόντων ταξινομητών.



Σχήμα 5.3: Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τους συγχρονισμένους Ταξινομητές Cl_{pas} και CDBIMM, ως προς τα frames. Οι δύο συνεχείς γραμμές δείχνουν την προβλεφθείσα κλάση ανά frame από κάθε ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων.

Macro Metrics

		Predicted											
		Index	Punch	Palm	Tiger	Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Undefined
Actual	Index	2	1	0	0	0	0	0	0	0	0	0	0
	Punch	0	1	0	0	0	0	0	0	0	0	0	0
	Palm	1	0	1	0	0	0	0	0	0	0	0	0
	Tiger	0	0	0	2	0	0	0	0	0	0	0	0
	Pinch	0	0	0	0	13	0	0	4	0	0	0	0
	Handflip out	0	0	0	0	0	3	6	0	2	0	0	0
	Handflip in	0	0	0	0	0	2	9	0	2	0	0	0
	Punchflip in	0	0	0	0	0	0	0	2	3	0	0	0
	Punchflip out	0	0	0	0	0	0	0	1	12	0	0	0
	Fingerwave in	0	0	0	0	0	2	0	2	1	0	5	0
	Fingerwave out	0	0	0	0	0	0	0	2	1	0	3	0

Πίνακας 5.9: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση όλων των χειρονομιών από τους συγχρονισμένους Ταξινομητές Cl_{pas} και CDBIMM ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από κάποιον από τους ταξινομητές ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. Λόγω αυτού του ορισμού, μια εμφάνιση μπορεί να αποκτήσει δύο προβλεφθείσες τιμές από τους δύο συγχρονισμένους ταξινομητές. Τότε, αν κάποια από τις τιμές αυτές αντιστοιχεί στην πραγματική επισήμανση της εμφάνισης, η εμφάνιση θεωρείται πώς έχει ανιχνευθεί σωστά.

Macro Metrics

Index	Punch	Palm	Tiger	Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
0.667	0.667	0.667	1.000	0.867	0.333	0.643	0.250	0.706	0.000	0.429	0.633

Πίνακας 5.10: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τους συγχρονισμένους Ταξινομητές Cl_{pas} και CDBIMM ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από κάποιον από τους ταξινομητές ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες. Λόγω αυτού του ορισμού, μια εμφάνιση μπορεί να αποκτήσει δύο προβλεφθείσες τιμές από τους δύο συγχρονισμένους ταξινομητές. Τότε, αν κάποια από τις τιμές αυτές αντιστοιχεί στην πραγματική επισήμανση της εμφάνισης, η εμφάνιση θεωρείται πώς έχει ανιχνευθεί σωστά.

Micro Metrics

		Predicted											
		Index	Punch	Palm	Tiger	Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Undefined
Actual	Index	294	64	21	21	0	0	0	0	0	0	0	0
	Punch	0	364	0	13	0	0	0	0	0	0	0	0
	Palm	46	0	155	0	0	0	0	0	0	0	0	0
	Tiger	2	0	0	369	0	0	0	0	0	0	0	0
	Pinch	0	0	0	0	235	9	0	52	0	0	0	0
	Handflip out	0	0	0	0	0	47	137	0	40	0	0	0
	Handflip in	0	0	0	0	0	33	188	0	44	0	0	0
	Punchflip in	0	0	0	0	0	0	0	25	93	0	0	0
	Punchflip out	0	0	0	0	7	0	0	10	158	0	0	0
	Fingerwave in	0	0	0	0	0	31	0	43	24	0	120	0
	Fingerwave out	0	0	0	0	0	6	0	53	20	0	66	0

Πίνακας 5.11: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση όλων των χειρονομιών από τους συγχρονισμένους Ταξινομητές Cl_{pas} και CDBIMM ως προς τα frames.

Micro Metrics

Index	Punch	Palm	Tiger	Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
0.792	0.904	0.822	0.953	0.874	0.269	0.637	0.166	0.570	0.000	0.399	0.644

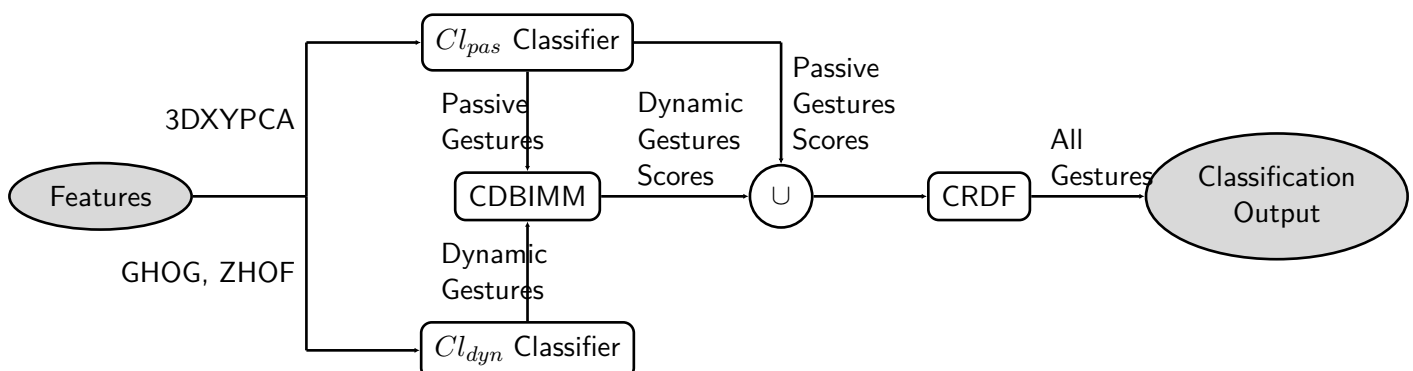
Πίνακας 5.12: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τους συγχρονισμένους Ταξινομητές Cl_{pas} και CDBIMM ως προς τα frames.

	Accuracy		
	Class. Cl_{pas}	Class. CDBIMM	Total Mean
Micro	0.876	0.499	0.644
Macro	0.750	0.560	0.633

Πίνακας 5.13: Δεδομένα Test: Ακρίβειες Συγχρονισμένων Ταξινομητών όλων των χειρονομιών.

5.3.2 Συνδυασμένη Ταξινόμηση (Combined)

Επιδιώκοντας να γεφυρώσουμε το χάσμα μεταξύ των δύο χώρων χειρονομιών, με στόχο να λαμβάνουμε τελικά μία πρόβλεψη ως έξοδο από το σύστημα ταξινόμησης, περιγράφουμε ακολούθως τη Συνδυασμένη Ταξινόμηση. Η προσέγγιση αυτή προσθέτει ένα ακόμα στάδιο στο ‘cascaded’ σύστημα ταξινομητών, όπως φαίνεται στο ακόλουθο σχεδιάγραμμα:



Σχήμα 5.4: Συνδυασμένη Ταξινόμηση (Combined). Με ελλείψεις φαίνονται η είσοδος και η έξοδος του συστήματος.

Προσθέτοντας τον ταξινομητή CRDF, χρειάζεται αρχικά να τον εκπαιδεύσουμε, χρησιμοποιώντας τα συνενωμένα scores των CDBIMM και CI_{pas} , δίνοντας ως είσοδο τα χαρακτηριστικά που προέρχονται από τα δεδομένα Training. Ο ταξινομητής CRDF δίνει έξοδο, η οποία ανήκει και στις δυναμικές και στις παθητικές χειρονομίες. Στη συνέχεια, δοκιμάζουμε το σύστημα με τη χρήση των δεδομένων Test.

Στο Σχήμα 5.5 φαίνεται το αποτέλεσμα της δοκιμής, με διαφορετικά χρώματα να υποδεικνύουν τη μεταβολή των εμφανίσεων μέσα στο Ground Truth. Στους Πίνακες 5.16 και 5.14 παρουσιάζεται ο πίνακας σύγχυσης ως προς τα frames και ως προς τα utterances αντίστοιχα. Ομοίως, στους Πίνακες 5.17 και 5.15 φαίνονται η Ακρίβεια και τα F1-Scores ανά χειρονομία, για τον χώρο των frames και των εμφανίσεων αντίστοιχα.

Macro Metrics

		Predicted											
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Index	Punch	Palm	Tiger	Undefined
Actual	Pinch	12	0	0	4	0	0	0	0	0	0	0	1
	Handflip out	0	0	7	0	1	0	1	0	0	0	0	1
	Handflip in	0	0	11	0	1	0	0	0	0	0	0	1
	Punchflip in	0	0	0	3	0	0	0	0	2	0	0	0
	Punchflip out	0	0	0	4	1	0	0	1	6	0	0	1
	Fingerwave in	0	2	0	1	1	0	6	0	0	0	0	0
	Fingerwave out	0	0	0	0	1	0	5	0	0	0	0	0
	Index	0	0	0	1	0	0	0	1	0	0	0	1
	Punch	0	0	0	0	0	0	0	0	1	0	0	0
	Palm	0	1	0	0	0	0	0	0	0	1	0	0
	Tiger	0	0	0	0	0	0	0	0	0	0	2	0

Πίνακας 5.14: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση όλων των χειρονομιών από τον CRDF Ταξινομητή ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics

Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Index	Punch	Palm	Tiger	Accuracy
0.828	0.000	0.710	0.333	0.111	0.000	0.556	0.400	0.200	0.667	0.800	0.446

Πίνακας 5.15: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον CRDF Ταξινομητή ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

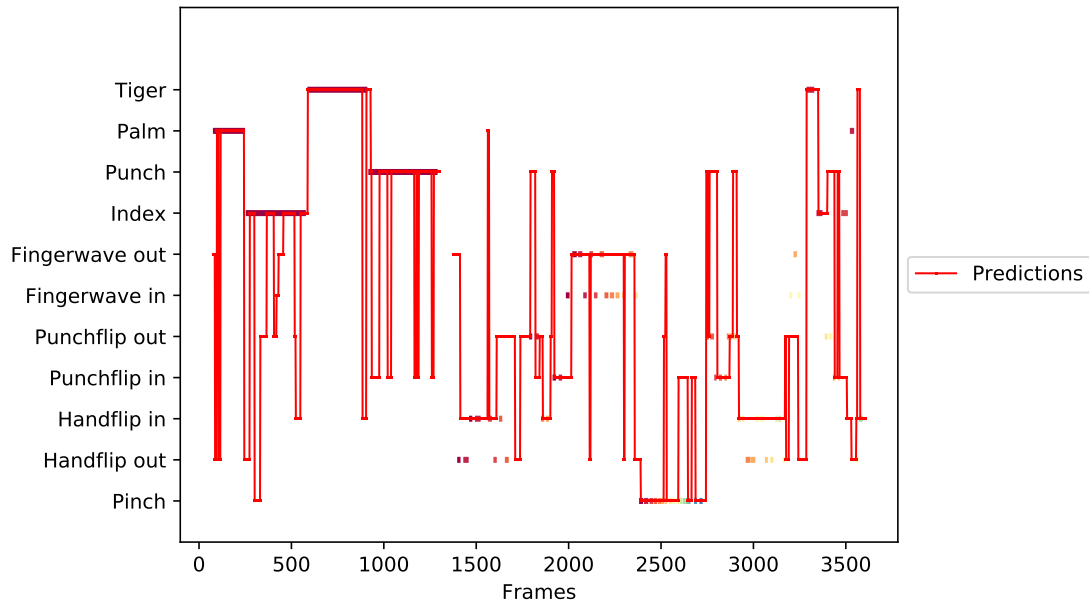
Micro Metrics

		Predicted											
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Index	Punch	Palm	Tiger	Undefined
Actual	Pinch	216	9	0	57	8	0	6	0	0	0	0	
	Handflip out	0	7	162	0	30	0	16	0	0	0	9	
	Handflip in	0	0	216	0	36	0	0	0	0	7	6	
	Punchflip in	0	0	0	77	0	0	0	0	41	0	0	
	Punchflip out	0	0	0	36	33	0	0	14	92	0	0	
	Fingerwave in	0	38	0	22	23	0	135	0	0	0	0	
	Fingerwave out	0	6	0	0	20	0	119	0	0	0	0	
	Index	31	21	33	31	55	11	26	182	0	0	10	
	Punch	0	0	0	91	0	0	0	0	273	0	13	
	Palm	0	44	8	0	0	0	8	0	0	138	0	
	Tiger	0	0	21	0	0	0	0	2	0	0	348	

Πίνακας 5.16: Δεδομένα Test: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση όλων των χειρονομιών από τον CRDF Ταξινομητή ως προς τα frames.

Micro Metrics											
Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Index	Punch	Palm	Tiger	Accuracy
0.796	0.040	0.613	0.356	0.174	0.000	0.523	0.609	0.697	0.798	0.919	0.577

Πίνακας 5.17: Δεδομένα Test: F-Scores και Ακρίβεια της ταξινόμησης για τον CRDF Ταξινομητή ως προς τα frames.



Σχήμα 5.5: Δεδομένα Test: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον CRDF Ταξινομητή, ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων.

Κεφάλαιο 6

Συμπεράσματα και Παρατηρήσεις

6.1 Γενικά

Κατά την κατασκευή του αλγορίθμου και εξαγωγή των πειραματικών αποτελεσμάτων, φάνηκε ότι το πρόβλημα της αναγνώρισης χειρονομιών μπορεί να αντιμετωπιστεί επαρκώς ικανοποιητικά σε πραγματικό χρόνο χωρίς τη χρήση εξαιρετικά περίπλοκων εργαλείων. Στη συνέχεια αναφέρονται περαιτέρω συμπεράσματα και παρατηρήσεις, όσον αφορά στα διάφορα στάδια του αλγορίθμου.

6.2 Προεπεξεργασία/Εξαγωγή Χαρακτηριστικών

6.2.1 Παρατηρήσεις

Κατά τη διάρκεια της προεπεξεργασίας με την προτεινόμενη μέθοδο παρατηρήθηκε η ανάγκη στάθμισης πολλών παραμέτρων ευριστικά. Αυτό είχε ως συνέπεια να μη βρεθεί η βέλτιστη λύση στο πρόβλημα του διαχωρισμού του χεριού από το υπόβαθρο και της σκελετοποίησής του. Μία ακόμα παρατήρηση είναι ότι όσο μεγάλωνε η απόσταση από τον αισθητήρα τόσο αυξανόταν ο θόρυβος της εξαγόμενης μάσκας από το κανάλι βάρους, ενώ λόγω της μεταβολής της διαμήκου διακριτικής ικανότητάς του οι χρησιμοποιούμενοι περιγραφητές, οι οποίοι βάσει κατασκευής θα έπρεπε να είναι αναλλοίωτοι στην απόσταση από την κάμερα, είχαν μικρή τελικά εξάρτηση από αυτήν. Επιπλέον, παρατηρήθηκε σε κάποιες δυναμικές χειρονομίες, όπως στις Fingerwave

in/out, ότι ο αλγόριθμος προεπεξεργασίας δεν ήταν αρκετά εύρωστος στις μικρού εύρους κινήσεις, αφού, λόγω της σχετικά μεγάλης απόστασης από την κάμερα και της χαμηλής ποιότητας της εικόνας, αυτές άγγιζαν τα όρια θορύβου. Πέρα από τα παραπάνω, η χρήση μόνο του καναλιού βάθους στην αναγνώριση και συγχρόνως η σχετικά κακή ποιότητα της εικόνας βοήθησαν ιδιαίτερος στο να επιταχύνουν τους υπολογισμούς και να μειώσουν τον υπολογιστικό φόρτο, καθιστώντας τον **μέσο χρόνο της προεπεξεργασίας** ίσο με **15 ms**.

6.2.2 Συμπεράσματα

Από τους πειραματισμούς με τις διάφορες προτεινόμενες μεθόδους τμηματοποίησης εικόνας στη βιβλιογραφία, καταλήξαμε στο συμπέρασμα ότι απαιτείται μεγάλη παραμετροποίηση, για να μπορέσει να ταιριάζει ένα ιδεατό μαθηματικό μοντέλο στο πραγματικό πρόβλημα που μελετάει κανείς. Οι διάφορες ατέλειες, που ενισχύουν το σφάλμα της τμηματοποίησης, για να μπορέσουν να αντιμετωπιστούν χρειάζονται υπολογιστικά και χρονικά εξαντλητικές μεθόδους, καθιστώντας το πρόβλημα της τμηματοποίησης της εικόνας δύσκολο να περατωθεί σε πραγματικό χρόνο. Εκτιμούμε λοιπόν, ότι με την αύξηση του κόστους του χρησιμοποιούμενου εξοπλισμού αλλά και με την πάροδο του χρόνου και την τεχνολογική εξέλιξη, η προτεινόμενη μέθοδος προεπεξεργασίας θα μπορέσει να αποδώσει καλύτερα ποιοτικά αποτελέσματα από αυτά που καταφέραμε να εξαγάγουμε.

6.3 Ταξινόμηση

6.3.1 Παρατηρήσεις

Όπως μπορεί να παρατηρήσει κανείς, ο ταξινομητής CDBIMM δίνει κατά κράτος καλύτερα αποτελέσματα στην αναγνώριση δυναμικών χειρονομιών, προσθέτοντας περίπου 2% στην ακρίβεια που μπορεί να ληφθεί από τον καλύτερο δυναμικό ταξινομητή για το Test dataset, άρα όντως υποβοηθάει την ταξινόμηση των δυναμικών χειρονομιών η a priori γνώση των παθητικών.

Ακόμα, φαίνεται από το πρώτο dataset ότι τα χαρακτηριστικά που χρησιμοποιήθηκαν δεν είναι invariant στον κατεπτρισμό (mirroring) της εικόνας, αφού, χρησιμοποιώντας δεδομένα εκπαίδευσης που έχουν ληφθεί από το δεξί χέρι, δεν γίνεται σωστή ταξινόμηση του αριστερού

χειριού (από το frame 1500 και μετά) και για τους δύο τύπους δράσεων. Παράλληλα, αυτό σημαίνει ότι μπορεί να εκπαιδευτεί ο αλγόριθμος να αναγνωρίζει δράσεις που εκτελούνται από διαφορετικά χέρια, ώστε να υπάρχει μεγαλύτερος έλεγχος κάποιας εργασίας από τον χρήστη.

Ένα ζήτημα που παρατηρήθηκε κατά τη διάρκεια της εκπαίδευσης είναι ότι τα δεδομένα που διαθέτουμε δεν είναι επαρκή για να εκπαιδευτούν οι ταξινομητές των δυναμικών χειρονομιών, αφού ο λόγος αριθμού χαρακτηριστικών ανά δείγμα προς τον αριθμό δειγμάτων ήταν σχετικά υψηλός, της τάξης του 0.3. Αυτό είχε ως συνέπεια να μην μπορέσει να γίνει επαρκής εκπαίδευση των λεξικών της Αραιής Κωδικοποίησης, των οποίων οι επιδόσεις στην ταξινόμηση μεταβάλλονταν σε κάθε εκπαίδευση, φανερόντας τη σημαντική επιρροή της αρχικοποίησης τους και την τελικά ασθενή εκπαίδευσή τους. Με την ίδια λογική παρατηρείται σημαντική απόκλιση της ακρίβειας του $C_{l_{pas}}$ σε διαφορετικές εκτελέσεις της εκπαίδευσης του ταξινομητή με ίδιες παραμέτρους.

Μία ακόμα παρατήρηση είναι το ότι ο αλγόριθμος δεν είναι ιδιαίτερα εύρωστος όσον αφορά στον διαχωρισμό δυναμικών χειρονομιών, που διαφέρουν στη φορά εκτέλεσης της κίνησης, όπως για παράδειγμα το ζεύγος (Handflip in, Handflip out) ή το ζεύγος (Fingerwave in, Fingerwave out). Αυτό ενδεχομένως να συμβαίνει για άλλη μία φορά λόγω μη επαρκών δειγμάτων εκπαίδευσης.

Όσον αφορά στο χρονικό κομμάτι, η συλλογή των χρησιμοποιούμενων χαρακτηριστικών διήρκεσε κατά μέσο όρο 4 ms, ενώ η ταξινόμηση 1.2 ms.

6.3.2 Συμπεράσματα

Λόγω των παραπάνω παρατηρήσεων εκτιμούμε ότι περισσότερα δεδομένα εκπαίδευσης από διαφορετικούς χρήστες θα βελτίωναν την ποιότητα της απόκρισης των ταξινομητών.

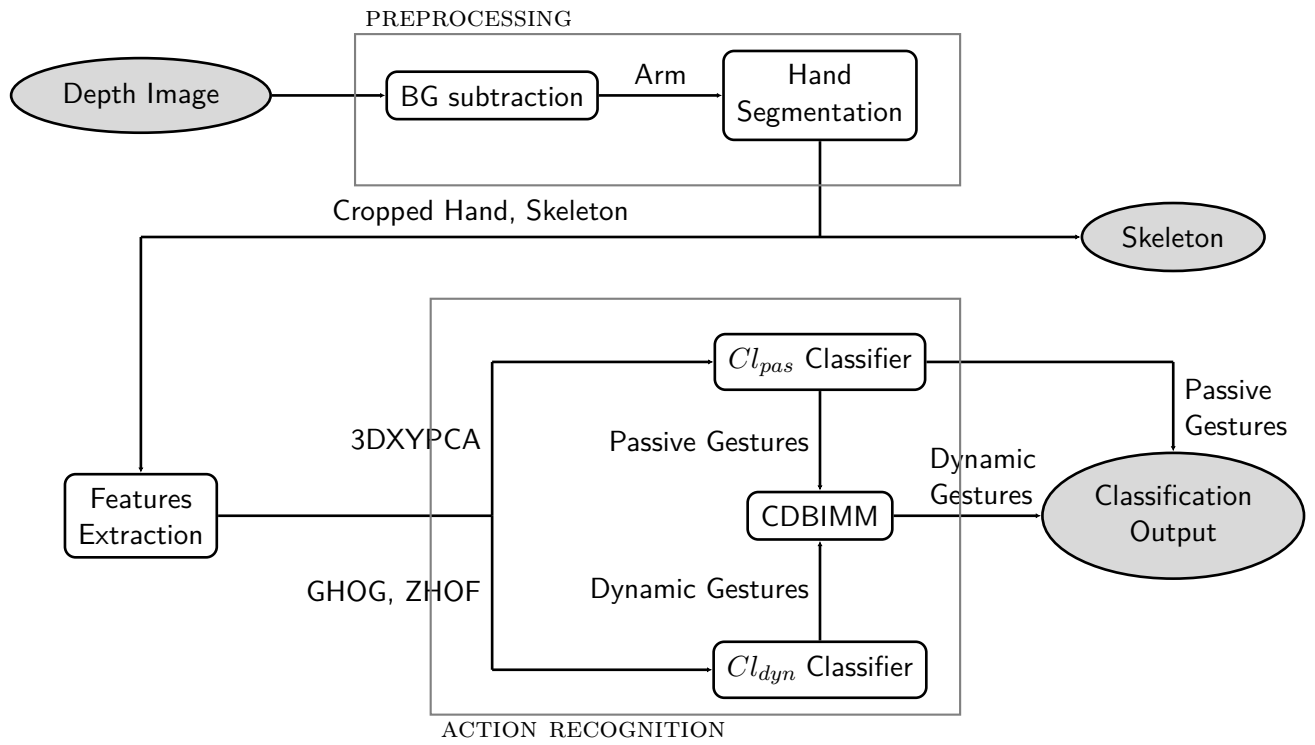
Ακόμα, η εξειδίκευση του προτεινόμενου αλγορίθμου ταξινόμησης, τόσο από πλευράς χαρακτηριστικών, για τα οποία απαιτείτο γνώση του σκελετού της παλάμης και το βάθος της εικόνας, όσο και από πλευράς προσέγγισης, μιας και δεν βρέθηκαν datasets για τα οποία να είναι δυνατός ο διαχωρισμός σε δύο ‘υπερκατηγορίες’ δυναμικών και παθητικών δράσεων, κατέστησε πιθανόν αδύνατη την χρήση ήδη υπαρχόντων datasets.

Από τα αποτελέσματα της ταξινόμησης συμπεραίνουμε ότι το μοντέλο ταξινόμησης με τις καλύτερες επιδόσεις από αυτά που δοκιμάστηκαν είναι αυτό με τη μορφή In Sync.

6.4 Συνολικά

Κατόπιν των διαφόρων θεωρητικών και πρακτικών μελετών, καταλήγουμε στο ότι ποιοτικά καλύτερη ταξινόμηση επιτυγχάνεται με το συνολικό μοντέλο του Σχήματος 6.1. Πιο συγκεκριμένα, από το στάδιο της προεπεξεργασίας (Preprocessing), του οποίου οι υπερπαραμέτροι έχουν βρεθεί ευριστικά και το οποίο περιέχει μη επιβλεπόμενη (unsupervised) εκπαίδευση του μοντέλου παρασκηνίου GMM, λαμβάνουμε τη μάσκα του χεριού (Cropped Hand), καθώς επίσης το σκελετό του χεριού. Ο σκελετός δίνεται στην έξοδο, αφού αυτός περιέχει τις θέσεις των τμημάτων του χεριού ως προς το σύστημα συντεταγμένων του αισθητήρα. Κατά την εξαγωγή των χαρακτηριστικών (Features Extraction) αφαιρείται η πληροφορία της θέσεως και του προσανατολισμού του χεριού ως προς την κάμερα, για να γίνουν τα εξαγόμενα χαρακτηριστικά invariant σε αυτά τα στοιχεία. Κατόπιν, βάσει του πειραματισμού που έλαβε χώρα στο Κεφ. 4, εξάγονται τα χαρακτηριστικά 3DXYPECA, από το κάθε frame, και χρησιμοποιούνται ως είσοδος στον ταξινομητή Cl_{pas} , ο οποίος βρέθηκε ότι είναι τύπου RDF. Επιπλέον, εξάγονται τα χαρακτηριστικά GHOG και ZHOF, από τον κάθε buffer $l_b = 10$ frames, τα οποία εισάγονται στον ταξινομητή των δυναμικών χειρονομιών Cl_{dyn} , ο οποίος αποτελείται από SVMs. Στη συνέχεια, τα scores του Cl_{dyn} μετατρέπονται σε πιθανότητες και μαζί με τα scores του Cl_{pas} μεταφέρονται στο μοντέλο CDBIMM, ενώ τα scores του Cl_{pas} με τη χρήση της μεθόδου CSTD δίνουν ως έξοδο την πρόβλεψη των παθητικών χειρονομιών. Το μοντέλο CDBIMM με τη χρήση της τεχνικής CProb δίνει ως έξοδο την πρόβλεψη των δυναμικών χειρονομιών. Για να λειτουργήσει το παραπάνω σύστημα, επειδή περιέχει σε σειρά δύο ταξινομητές, χρειάζεται δύο στάδια επιβλεπόμενης (supervised) εκπαίδευσης, το στάδιο εκπαίδευσης των Cl_{pas} και Cl_{dyn} και το στάδιο εκπαίδευσης του CDBIMM.

Από τα παραπάνω, φαίνεται ότι το προτεινόμενο μοντέλο αναγνώρισης χειρονομιών αποτελεί σύστημα αναγνώρισης ημιεπιβλεπόμενης μάθησης (semisupervised learning), ενώ μπορεί και εκτελείται εξ ολοκλήρου σε χρόνο κάτω από 30 ms (περίπου 25ms), δηλαδή σε πραγματικό χρόνο.



Σχήμα 6.1: Συνολικός Προτεινόμενος Αλγόριθμος. Με ελλείψεις φαίνονται η είσοδος και οι έξοδοι του συστήματος.

Κεφάλαιο 7

Μελλοντικές Επεκτάσεις- Περιγραφή Πρακτικών Υλοποιήσεων

7.1 Γενικά

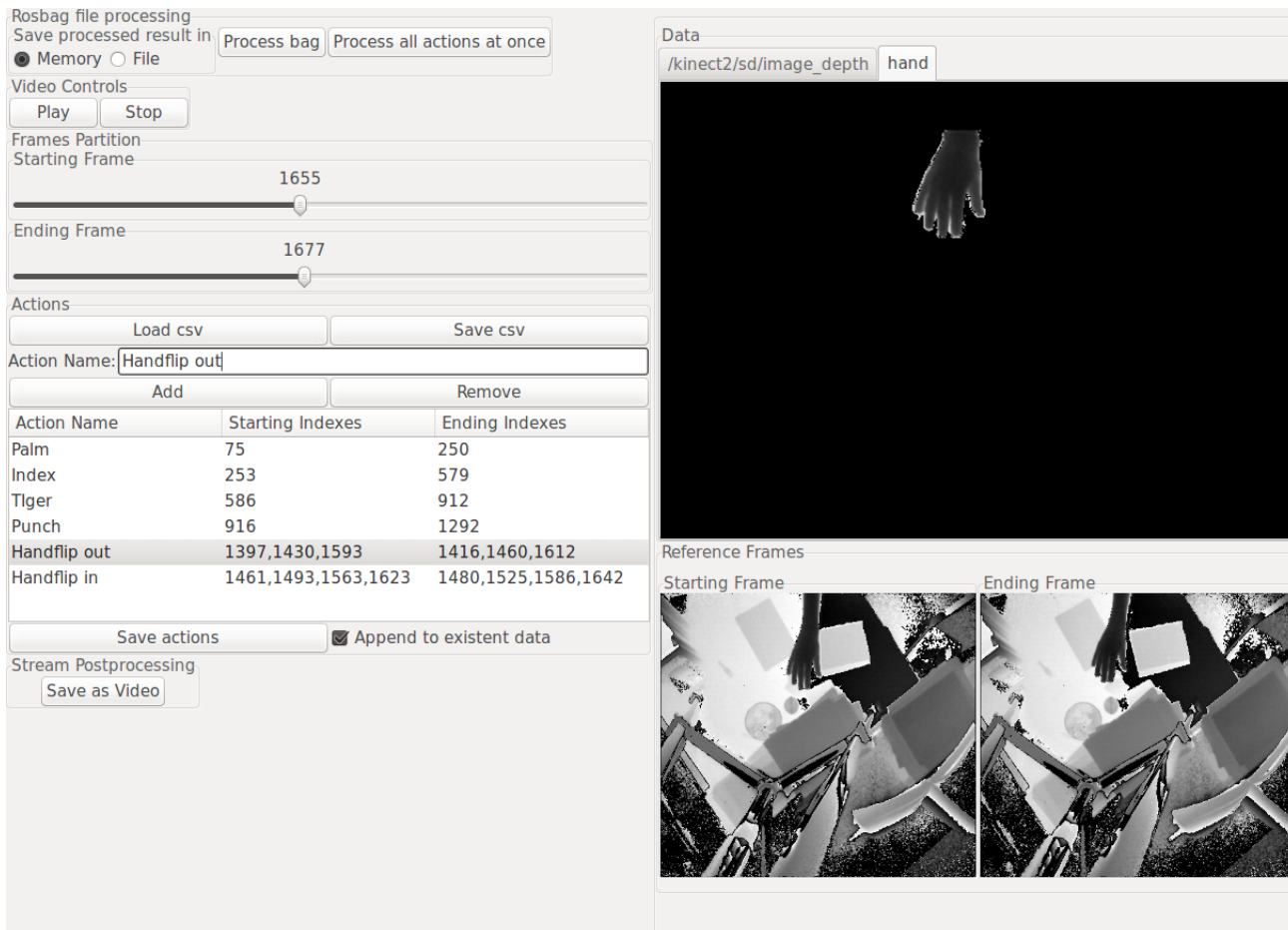
Στο κεφάλαιο αυτό περιγράφονται δύο υλοποιήσεις σχετικές με τον περιγραφέντα αλγόριθμο. Αρχικά, περιγράφεται εν συντομία η GUI εφαρμογή που αναπτύχθηκε για την επεξεργασία των δεδομένων και το annotation αυτών. Στη συνέχεια, φανερώνεται η δυνατότητα μελλοντικής επέκτασης της εργασίας, μέσω της κατασκευής και της περιγραφής μιας εφαρμογής σχεδίασης, κατά την οποία μέρος του προτεινόμενου αλγορίθμου γίνεται υπομονάδα ενός απλού **συστήματος σχεδίασης σε οποιαδήποτε επιφάνεια με τη χρήση του αισθητήρα Kinect**.

7.2 Εφαρμογή Επεξεργασίας Δεδομένων

7.2.1 Παρουσίαση και Περιληπτική Περιγραφή

Όπως αναφέρθηκε στην Εισαγωγή, για την επισήμανση των δεδομένων, που λήφθησαν μέσω του Kinect, κατασκευάστηκε μία εφαρμογή με γραφικό περιβάλλον, η οποία εκτελεί την προεπεξεργασία της εικόνας μέχρι το σημείο εξαγωγής της μάσκας του χεριού και επιτρέπει στον χρήστη να επιθεωρήσει το αποτέλεσμα και να επισημάνει τα είδη χειρονομιών ανά frame. Λόγω του ότι μία δράση μπορεί να αποτελεί υποσύνολο μιας άλλης, κρίθηκε σωστό να επιτραπεί η

επισήμανση περισσότερων από μία χειρονομιών ανά frame, αν και αυτό αποφεύγεται στην κατασκευή και annotation των εικονοσειρών που χρησιμοποιούνται για το training, validation και testing στην προκειμένη εργασία. Ένα στιγμιότυπο αυτής της εφαρμογής είναι το ακόλουθο:



Σχήμα 7.1: Στιγμιότυπο πρότυπης εφαρμογής εξαγωγής δεδομένων από αρχεία rosbag, για annotation χειρονομιών.

Η εν λόγω εφαρμογή λαμβάνει ως είσοδο ένα rosbag αρχείο (το οποίο περιέχει το stream από το Kinect και παράγεται μέσω της αλληλεπίδρασης του ROS και του Kinect iai framework), το επεξεργάζεται ανά frame, κρατώντας μόνο το κανάλι της εικόνας βάθους που ενδιαφέρει, εφαρμόζει την προεπεξεργασία σε κάθε frame, απομονώνοντας το χέρι, και εμφανίζει το αποτέλεσμα στην καρτέλλα 'hand'. Κατόπιν ο χρήστης έχει τη δυνατότητα να ανατρέξει σε κάθε διάστημα από frames, το οποίο ορίζεται μέσα στο 'πλαίσιο' *Frames Partition*, και να επισημάνει τη δράση που παρατηρεί ότι εκτελεί το χέρι στο διάστημα αυτό στο πλαίσιο *Actions*, ενώ μπορεί να παρατηρήσει και πού ακριβώς έχει τοποθετήσει τους δείκτες *Starting Frame* και *Ending Frame* βλέποντας τα αντίστοιχα, μη επεξεργασμένα frames στο πλαίσιο *Reference Frames*. Η απάντησή του αποθηκεύεται σε μία λίστα, η οποία εμφανίζεται στην οθόνη και ο χρήστης έχει τη δυνατότητα να

την επεξεργαστεί in place. Η λίστα αυτή, όταν ολοκληρωθεί το annotation, μπορεί να αποθηκευτεί σε ένα αρχείο csv σε μια θέση στον δίσκο ορισμένη από τον χρήστη. Ακόμα, μέσω αυτής της εφαρμογής μπορούν να εξαχθούν τα επεξεργασμένα δεδομένα σε μορφή φακέλων εικόνων, με τις μάσκες του αποκομμένου χεριού να αποθηκεύονται ξεχωριστά από το αντίστοιχο frame.

7.3 Εφαρμογή Σχεδίασης

7.3.1 Σύντομη Περιγραφή

Η πρακτικότητα του αλγορίθμου παρουσιάζεται μέσα από την υλοποίηση μιας απλής εφαρμογής, μέσω της οποίας παρέχεται η δυνατότητα στον χρήστη να σχεδιάσει σε οποιαδήποτε επιφάνεια με χρήση των παθητικών χειρονομιών. Παράλληλα, παρουσιάζεται και η ικανότητα του αλγορίθμου να προσαρμοστεί αρμονικά σε ένα μεγαλύτερο σύστημα αναγνώρισης, με πρόσθετες δυνατότητες. Συγκεκριμένα, στην εν λόγω εφαρμογή έχει προστεθεί η δυνατότητα αναγνώρισης ακόμα πιο εξελιγμένων χειρονομιών, οι οποίες προϋποθέτουν μετατόπιση του χεριού στο απόλυτο σύστημα συντεταγμένων του χεριού, μέσω του αλγορίθμου **Protractor**, γνωστού και ως **\$N** [1]. Η προσθήκη αυτή επιτρέπει την ταύτιση της τροχιάς του ακρότατου σημείου του χεριού με διάφορα, προκαθορισμένα γεωμετρικά σχήματα. Με τον τρόπο αυτό, ο συνολικός αλγόριθμος αποκτά την πλήρη μορφή του διαγράμματος 6.1, αν και οι ταξινομητές των δυναμικών χειρονομιών παραμένουν απενεργοποιημένοι. Το αποτέλεσμα της αναγνώρισης και σχεδίασης φαίνεται σε ένα πρότυπο γραφικό περιβάλλον στην οθόνη του υπολογιστή, το οποίο δεν παρουσιάζει κάποια άλλη δυνατότητα ελέγχου έτσι, ώστε να φανεί ότι ένα απλοϊκό πρόβλημα σχεδίασης μπορεί να προσεγγιστεί εξ ολοκλήρου με την προτεινόμενη μέθοδο αναγνώρισης, χωρίς κάποια άλλη υποβοήθηση μεθόδου εισαγωγής.

7.4 Πραγματικού Χρόνου Αναγνώριση

Ο χρόνος που απαιτείται για να λειτουργήσει το σύστημα αναγνώρισης των παθητικών χειρονομιών ανά frame, συμπεριλαμβανομένης της προεπεξεργασίας, είναι περίπου 18ms. Η απόκριση του συστήματος στην αλλαγή των χειρονομιών, δεδομένης σωστής διάταξης και κατάλληλης εκπαίδευσης, είναι μικρότερη από 1 sec, ενώ ως επί το πλείστον γίνεται στιγμιαία. Η απόκριση

θα μπορούσε τεχνικά να είναι ταχύτερη, αλλά αποφασίστηκε να μειωθεί η ευαισθησία του αλγορίθμου στις εναλλαγές, ώστε να γίνει πιο εύχρηστος και να μην επηρεάζεται από μικρές τοπικές διαφορές των scores του RDF ταξινομητή.

7.5 Σύντομη Περιγραφή Protractor

Ο Protractor είναι ένας εύρωστος και ταχύς αλγόριθμος ταιριάσματος μη κλειστής ένωσης σύνθετων καμπυλών, χρησιμοποιώντας ως κριτήριο ομοιότητας (similarity) την **απόσταση συνημιτόνου**. Η απόσταση αυτή επιτρέπει τη σύγκριση δύο n -διάστατων κανονικοποιημένων διανυσμάτων v_t και v_g και εκφράζεται ως:

$$S(t, g) = \frac{1}{\arccos\left(\frac{v_t \cdot v_g}{\|v_t\| \|v_g\|}\right)}$$

Καθένα από αυτά τα διανύσματα αναπαριστά την κατά προσέγγιση παρεμβολή 16 σημείων μιας καμπύλης (η οποία ορίζεται ως stroke), υπό τη μορφή $[x_1, y_1, x_2, y_2, \dots, x_{16}, y_{16}]$. Ως v_t ορίζουμε το διάνυσμα 32×1 το οποίο εκφράζει το πρότυπο αντιστοίχισης (template), ενώ ως v_g το διάνυσμα που χαρακτηρίζει τη μετατόπιση του χεριού. Για να γίνει η παραπάνω μετρική ανεξάρτητη της σχετικής γωνίας θ του template με το σύνολο της μετατόπισης του χεριού, υπολογίζεται η γωνία που πρέπει να στραφεί η καμπύλη μετατόπισης του χεριού ώστε να ελαχιστοποιηθεί η παραπάνω απόσταση και κατόπιν η απόσταση των παραγόμενων διανυσμάτων. Εξ ορισμού λοιπόν, η απόσταση αυτή είναι ανεξάρτητη του μέτρου των διανυσμάτων (δηλαδή της κλίμακας της καμπύλης που αυτά περιγράφουν) και της περιστροφής των καμπυλών ως προς το απόλυτο σύστημα συντεταγμένων του ‘καμβά’ σχεδίασης, εξαρτάται όμως από την φορά διαγραφής της κάθε καμπύλης. Ακόμα, απαιτείται ένα σύνολο ‘εκπαίδευσης’, ή αλλιώς ένα λεξικό χειρονομιών, το οποίο να επιτρέπει το template matching. Ένα ακόμα χαρακτηριστικό του Protractor είναι και το ότι δεν εξαρτάται από το πλήθος των strokes που συμμετάσχουν στην κατασκευή της χειρονομίας, αρκεί η παρεμβολή της ένωσης των template strokes να είναι ισομεγέθης με την αυτή της ένωσης των gesture strokes.

7.6 Περιγραφή Εφαρμογής Σχεδίασης

Για να καταστεί δυνατή η σχεδίαση μόνο με τη χρήση χειρονομιών, ανατέθηκε σε κάθε αναγνωριζόμενη χειρονομία και μία λειτουργία, ανάλογα με το πόσο ‘εύκολο’ και ‘φυσικό’ είναι να εκτελεστεί από κάποιον χρήστη. Η αντιστοίχιση φαίνεται στον ακόλουθο πίνακα:

Χειρονομία	Λειτουργία
Palm	Ανάπαυση / Εκτέλεση Αναγνώρισης Protractor
Tiger	Εναλλαγή Μεθόδου Σχεδίασης (Free/Guided)
Index	Σχεδίαση
Punch	Σβήσιμο

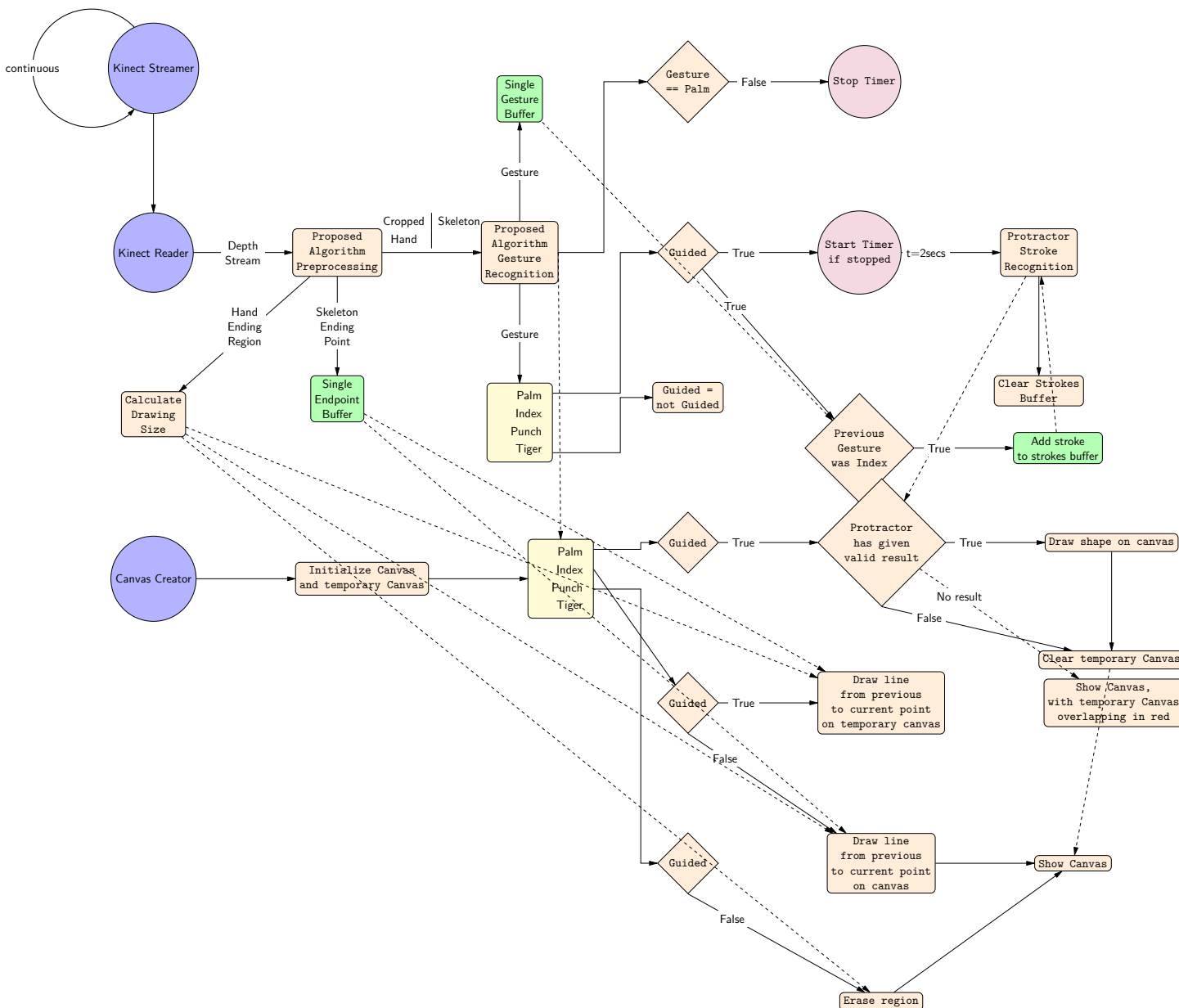
Πίνακας 7.1: Αντιστοίχιση Χειρονομιών με Λειτουργίες της Εφαρμογής Σχεδίασης

Η πληροφορία του σημείου σχεδίασης παρέχεται από μια μικρή περιοχή γύρω από το τελικό σημείο του σκελετού του χεριού, ενώ το μέγεθος γραφής είναι αντιστρόφως ανάλογο της απόστασης του χεριού από τον αισθητήρα. Ο καμβάς σβήνεται εντελώς, όταν το χέρι του χρήστη πάει πολύ κοντά στον αισθητήρα βάθους. Όπως φαίνεται στον πίνακα, υπάρχουν δύο διαφορετικές μέθοδοι εισαγωγής:

- *Guided*: Κατά τη μέθοδο αυτή χρησιμοποιείται ο αλγόριθμος Protractor, με τον οποίο γίνεται template matching όλων των strokes (μονοκοντυλιές) του χρήστη με τα σχήματα του λεξικού. Για να καταστεί δυνατή η σύγκριση, θα πρέπει ο προσδιορισμός των σχημάτων του λεξικού να γίνει με επαρκή δείγματα, που παρουσιάζουν διαφορετική φορά σχεδίασης. Ο χρήστης με τη χειρονομία Index μπορεί να σχεδιάσει ένα stroke, ενώ με τη χειρονομία Palm μπορεί είτε να αναμένει κάποιο χρόνο ($t_{pg} = 2\text{secs}$) για να ενημερώσει το σύστημα να εκκινήσει τη διαδικασία του matching με τα gestures του λεξικού, ή να δημιουργήσει πολλά strokes για το ίδιο επιθυμητό σχήμα με την αλληλουχία Index-Palm-Index-. . . Σε περίπτωση μη ταιριάσματος με κάποιο σχήμα του λεξικού, όλα τα strokes διαγράφονται από τον καμβά σχεδίασης. Η χειρονομία του σβησίματος Punch είναι απενεργοποιημένη κατά τη διάρκεια αυτής της μεθόδου εισαγωγής.
- *Free*: Σε αυτή τη μέθοδο εισαγωγής, ο χρήστης μπορεί να σχεδιάσει ελεύθερα, χωρίς την επίβλεψη του Protractor.

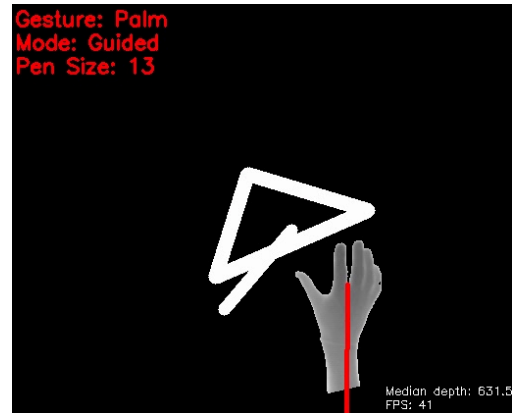
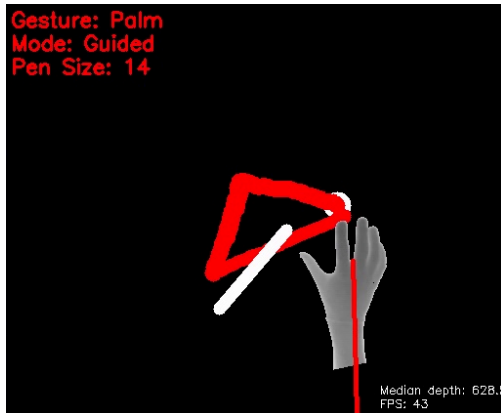
7.7 Παρουσίαση Διαγράμματος Ροής Εφαρμογής

Στην παρούσα ενότητα απεικονίζονται κάποιες τεχνικές λεπτομέρειες της υλοποίησης της εφαρμογής έτσι, ώστε να φανεί η απλότητα χρήσης του προτεινόμενου αλγορίθμου ως υποσυστήματος αναγνώρισης. Ενδεικτικά, το διάγραμμα ροής της εφαρμογής φαίνεται ακολούθως:

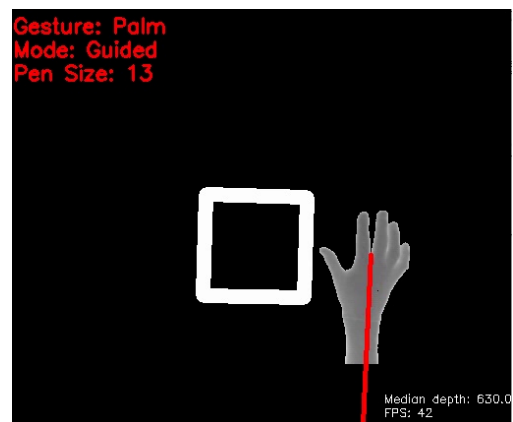
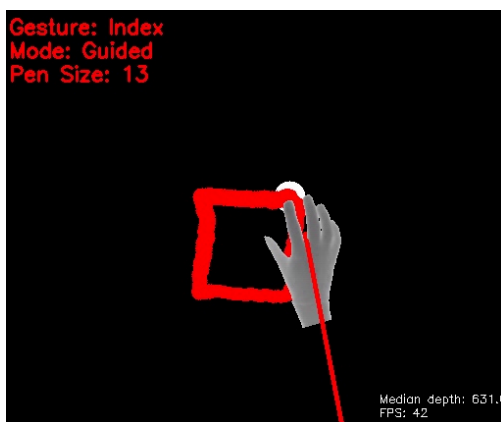


Σχήμα 7.2: Διάγραμμα Ροής Εφαρμογής. Το μωβ χρώμα υποδηλώνει τα διαφορετικά νήματα, το κίτρινο επιλογείς, το πράσινο καταχωρητές, το ροζ πράξεις χρονομέτρησης και το πορτοκαλί γενικές δράσεις. Οι ρόμβοι υποδηλώνουν σύγκριση. Τα βέλη με τις διακεκομμένες γραμμές υποδηλώνουν ως επί το πλείστον φόρτωση δεδομένων από καταχωρητή ή μεταφορά αποτελέσματος από ένα νήμα σε ένα άλλο.

Κάποια στιγμιότυπα της εφαρμογής για επίδειξη εμφανίζονται παρακάτω:



Σχήμα 7.3: Ανίχνευση ζωγραφισμένου Τριγώνου από τον Protractor



Σχήμα 7.4: Ανίχνευση ζωγραφισμένου Τετραγώνου από τον Protractor



Σχήμα 7.5: Η φράση 'Hi!' σχεδιασμένη στο Free Mode. Μπορεί να παρατηρήσει κανείς ότι είναι δυνατή η εναλλαγή των χειρονομιών, για να μπορέσει ο χρήστης να σχεδιάσει μη συνεχόμενες καμπύλες. Ακόμα, τα λανθασμένα σχεδιασμένα σημεία οφείλονται στο γεγονός ότι η χειρονομία έχει μια μικρή καθυστέρηση στην ανίχνευση, αφού έχει μειωθεί ο συντελεστής ευαισθησίας της αναγνώρισης.

Παραρτήματα

Παράρτημα Α

Αλγόριθμοι

A.1 Αλγόριθμος Ανίχνευσης και Χαρακτηρισμού Συνδεδεμένων Στοιχείων

Ο αλγόριθμος αυτός αποτελεί τμήμα του αλγορίθμου ‘Fast Watersheds’[46].

Algorithm 2 Connected Components Labeling

```

1: procedure CCL(inp) ▷ inp: binary input image
2:   Init 0: Define detection queue fifo
3:   Init 1: Create a zero label image L
4:   Init 2: currentlabel ← 0 ▷ Current Label name
5:   Init 3: p ← -1 ▷ p is assumed to be  $y \times \text{size}(\textit{inp})_{x\text{-axis}} + x$ 
6:   p ← p + 1
7:   currentlabel ← currentlabel + 1
8:   if AND (L(p) = 0, inp(p) = 1) then
9:     I(p) ← currentlabel
10:    fifo_add(p)
11:    go to 23
12:  else
13:    p ← p + 1
14:    if p > size(inp) then
15:      if fifo_empty() then
16:        exit()
17:      end if
18:      go to 23
19:    else
20:      go to 8
21:    end if
22:  end if
23:  while fifo_not_empty() do
24:    p' ← fifo_pop()
25:    for all p'' ∈  $N_G(p')$  do ▷  $N_G(x)$  are the neighbors of x
26:      if AND (L(p'') = 0, inp(p'') = 1) then
27:        I(p'') ← currentlabel
28:        fifo_add(p'')
29:      end if
30:    end for
31:  end while
32:  go to 6
33: end procedure

```

A.2 Αλγόριθμος Feature Sign Search

Ο αλγόριθμος αυτός αποτελεί εργασία των Lee et. al [21].

Algorithm 3 Feature-sign search Algorithm

- 1: Initialize $x := \vec{0}$, $\theta := \vec{0}$, and *active set* := $\{\}$, where $\theta_i \in \{-1, 0, 1\}$ denotes $\text{sign}(x_i)$
 - 2: From zero coefficients of x , select $i = \text{argmax}_i \left| \frac{\partial \|y - Ax\|^2}{\partial x_i} \right|$.
 Activate x_i (add i to the *active set*) only if it locally improves the objective, namely:
 If $\text{der}_i := \left| \frac{\partial \|y - Ax\|^2}{\partial x_i} \right| > \gamma$ then set $\theta_i := -\text{sign}(\text{der}_i)$,
active set := $\{i\} \cup \text{active set}$
 - 3: Feature-sign step:
 Let \hat{A} be a submatrix of A that contains only the columns corresponding to the *active set*.
 Let \hat{x} and $\hat{\theta}$ be subvectors of x and θ corresponding to *active set*.
 Compute the analytical solution to the resulting unconstrained QP:
 (minimize $_x \|y - \hat{A}\hat{x}\|^2 + \gamma \hat{\theta}^T \hat{x}$):
 $\hat{x}_{\text{new}} := (\hat{A}^T \hat{A})^{-1} (\hat{A}^T y - \gamma \hat{\theta} / 2)$
 Perform a discrete line search on the closed line segment from \hat{x} to \hat{x}_{new} .
 Check the objective value at \hat{x}_{new} and all points where any coefficient changes sign.
 Update \hat{x} (and the corresponding entries in x) to the point with the lowest objective value.
 Remove zero coefficients of \hat{x} from the *active set* and update $\theta := \text{sign}(x)$.
 - 4: Check the optimality conditions:
 - (a) Optimality condition for nonzero coefficients:
 $\frac{\partial \|y - Ax\|^2}{\partial x_j} + \gamma \text{sign}(x_j) = 0, \forall x_j \neq 0$
 If condition (a) is not satisfied, go to Step 3 (without any new activation), else check condition (b).
 - (b) Optimality condition for zero coefficients:
 $\left| \frac{\partial \|y - Ax\|^2}{\partial x_j} \right| \leq \gamma, \forall x_j = 0$
 If condition (b) is not satisfied, go to Step 2; otherwise return x as the solution.
-

Παράρτημα Β

Ανάλυση Μαθηματικών Θεμάτων

B.1 Θεώρημα του Green

Θεώρημα 2.1. [60] Έστω c μια απλή κλειστή και τμηματικά λεία καμπύλη που είναι θετικά προσανατολισμένη ως προς το κανονικό χωρίο $D, D \subset \mathbb{R}^2$, το οποίο είναι απλά συνεκτικό και έχει σύνορο δοθείσα καμπύλη c , δηλαδή $c = \partial D$. Υποθέτουμε επίσης ότι δίνονται δύο συνεχείς πραγματικές συναρτήσεις:

$$P, Q : D \rightarrow \mathbb{R}$$

των πραγματικών μεταβλητών x, y με κοινό πεδίο ορισμού το D , τέτοιες ώστε οι μερικές παράγωγοι

$$\frac{\partial Q}{\partial x} \text{ και } \frac{\partial P}{\partial y}$$

υπάρχουν και είναι συνεχείς συναρτήσεις στο D . Τότε

$$\oint_{\partial D} P dx + Q dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy$$

B.2 Θεώρημα Taylor

Το Θεώρημα του Taylor δίνει μια πολυωνυμική έκφραση μιας κ -φορές παραγωγίσιμης συνάρτησης γύρω από ένα δοσμένο σημείο. Ακολουθεί η δισδιάστατη περίπτωση του θεωρήματος, η οποία μπορεί εύκολα να γενικευτεί και σε ανώτερες διαστάσεις:

Θεώρημα 2.2. [59] Έστω μία συνάρτηση $f : A \rightarrow \mathbb{R}$, όπου A ένα ανοικτό υποσύνολο του \mathbb{R}^2 , των ανεξάρτητων μεταβλητών x, y , που έχει συνεχείς μερικές παραγώγους μέχρι και $(\nu + 1)$ -τάξης στο A . Αν $P_0(x_0, y_0)$ και $P(x_0 + h, y_0 + k)$ είναι δύο σημεία του A , όπου $(h, k) \in \mathbb{R}^2 - \{(0, 0)\}$ τέτοια, ώστε το ευθύγραμμο τμήμα με άκρα τα σημεία P_0, P να περιέχεται στο A , τότε υπάρχει θετικός πραγματικός αριθμός $\theta, 0 < \theta < 1$, τέτοιος, ώστε να ισχύει:

$$f(x_0 + h, y_0 + k) = f(x_0, y_0) + \sum_{i=1}^{\nu} \left[\frac{1}{i!} \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^{(i)} f(x_0, y_0) \right] + R_{\nu} \quad (\text{B.1})$$

όπου το υπόλοιπο (remainder) R_{ν} λαμβάνει την μορφή

$$R_{\nu}((x_0, y_0), (h, k)) = \frac{1}{(\nu + 1)!} \left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y} \right)^{(\nu+1)} f(x_0 + \theta h, y_0 + \theta k). \quad (\text{B.2})$$

Ο διαφορικός τελεστής που χρησιμοποιείται στους τύπους (B.1) και (B.2) έχει την εξής έννοια:

$$\left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y}\right) f(x_0, y_0) = h \frac{\partial f}{\partial x}(x_0, y_0) + k \frac{\partial f}{\partial y}(x_0, y_0)$$

και εν γένει, η δυναμική έκφραση

$$\left(h \frac{\partial}{\partial x} + k \frac{\partial}{\partial y}\right)^{(i)} f(x_0, y_0)$$

αναπτύσσεται με τον διωνυμικό τύπο, όπου ό,τι αφορά τα h και k είναι δυνάμεις πραγματικών αριθμών, ενώ όποια μερική παράγωγος είναι υψωμένη σε δύναμη i είναι η i -οστή μερική παράγωγος.

B.3 Σταθμισμένη Μέθοδος Διακριτών Ελαχίστων Τετραγώνων

Η Μέθοδος Ελαχίστων Τετραγώνων (Least Squares Fitting) είναι μία από τις βασικές μεθόδους στην Ανάλυση Παλινδρόμησης (Regression Analysis). [57] Παρουσιάζεται ακολούθως η μορφή του προβλήματος που αυτή η μέθοδος επιλύει, η οποία συναντάται ως επί το πλείστον σε φυσικά προβλήματα.

B.3.1 Πρόβλημα Ελαχίστων Τετραγώνων

Δίνονται m σημεία $t_i \in \mathbb{R}, i = 1, \dots, m$ και m πραγματικοί αριθμοί $f_i, i = 1, \dots, m$, οι οποίοι παριστάνουν δεδομένα, φερ' ειπείν τιμές μιας μετρήσιμης ποσότητας στα αντίστοιχα σημεία t_i . Θεωρούμε n πραγματικές συναρτήσεις ϕ_1, \dots, ϕ_n , οι οποίες ορίζονται τουλάχιστον στα σημεία t_1, \dots, t_m . Τότε η ελαχιστοποίηση του τετραγώνου της ποσότητας:

$$\rho_i := f_i - \sum_{j=1}^n c_j \phi_j(t_i), \quad i = 1, \dots, m$$

δηλαδή, το πρόβλημα ελαχιστοποίησης:

$$\min_{\mathbf{c} \in \mathbb{R}^n} E(\mathbf{c}), \quad E(\mathbf{c}) = \sum_{i=1}^m \left[f_i - \sum_{j=1}^n c_j \phi_j(t_i) \right]^2$$

έχει λύση τα c_j , τα οποία ελαχιστοποιούν το σφάλμα μεταξύ της παρατήρησης και της τιμής του γραμμικού συνδυασμού των συναρτήσεων προσέγγισης. Ελαχιστοποιώντας λοιπόν το σφάλμα,

πετυχαίνουμε να ‘ταιριάξουμε’ το σύνολο συναρτήσεων που έχουμε θεωρήσει στα δεδομένα f_i που διαθέτουμε, τα οποία προέρχονται συνήθως από πειραματική μέτρηση.

B.3.2 Επίλυση

Ορίζουμε ως A τον πίνακα:

$$A := [\alpha_{ij}] \in \mathbb{R}^{m \times n}, \quad \alpha_{ij} = \phi_j(t_i)$$

Το παραπάνω πρόβλημα αποδεικνύεται ότι έχει πάντοτε λύση, η οποία είναι και μοναδική αν και μόνο αν οι στήλες του A είναι γραμμικώς ανεξάρτητα διανύσματα. Για να ελαχιστοποιήσουμε την ενέργεια του σφάλματος, επιβάλλουμε ως αναγκαίες συνθήκες τον μηδενισμό των μερικών παραγώγων ως προς c_j , δηλαδή:

$$0 = \frac{\partial E}{\partial c_k} = 2 \sum_{i=1}^m \left[f_i - \sum_{j=1}^n c_j \phi_j(t_i) \right] \phi_k(t_i), \quad 1 \leq k \leq n$$

οι οποίες οδηγούν σε ένα σύστημα της μορφής:

$$A^T A x = A^T b \tag{B.3}$$

με $x := c \in \mathbb{R}^n$ και $b := [f_i], i = 1, \dots, m$. Σε περίπτωση που ο $A^T A$ είναι θετικά ορισμένος (αν και μόνο αν ο A έχει γραμμικώς ανεξάρτητες στήλες), τότε είναι και αντιστρέψιμος, με συνέπεια η μοναδική λύση του προβλήματος να δίνεται από τη σχέση:

$$x = (A^T A)^{-1} A^T b$$

Σε αντίθετη περίπτωση οι λύση δεν είναι μοναδική και μπορεί να βρεθεί παραμετρικά, αφαιρώντας τις γραμμικά εξαρτημένες στήλες από τον πίνακα A .

B.3.3 Προσθήκη Βαρών

Αν αντί για την Ευκλείδεια απόσταση στον υπολογισμό του σφάλματος, χρησιμοποιηθεί η απόσταση Mahalanobis και υποτεθεί για απλούστευση ότι ο πίνακας συνδιακύμανσης των στοιχείων του πίνακα A είναι κατά προσέγγιση διαγώνιος[42], τότε το πρόβλημα ελαχιστοποίησης γίνεται

σταθμισμένο και αποκτά την ακόλουθη μορφή:

$$\min_{\mathbf{c} \in \mathbb{R}^n} E_w(\mathbf{c}), \quad E_w(\mathbf{c}) = \sum_{i=1}^m \left[w_i \left(f_i - \sum_{j=1}^n c_j \phi_j(t_i) \right) \right]^2$$

Τα βάρη $w_i, i = 1, \dots, m$ είναι ίσα με τις αντίστοιχες τιμές της διαγωνίου του πίνακα συνδιακύμανσης του . Έστω W ο διαγώνιος πίνακας των βαρών w_i . Τότε το σύστημα προς επίλυση, κατ'αντιστοιχία με το σύστημα της σχέσης (B.3), είναι το ακόλουθο:

$$A^T W A x = A^T W b$$

Αν δεν υπάρχουν μηδενικά βάρη, τότε οι συνθήκες για τη λύση παραμένουν ίδιες με αυτές που αναφέρθηκαν στην Υποενότητα B.3.2.

B.4 Αφινικός (ή Ομοπαράλληλικός) Μετασχηματισμός

Ορισμός 2.1. Μια απεικόνιση $T : U \rightarrow V$ είναι **γραμμική** αν και μόνο αν για κάθε $x, y \in U$ και $\lambda, \mu \in K$ ισχύει:

$$T(\lambda x + \mu y) = \lambda T(x) + \mu T(y)$$

Ορισμός 2.2. [58] Αν $T : U \rightarrow U$ είναι μία γραμμική απεικόνιση από ένα διανυσματικό χώρο U μέσα στον ίδιο το χώρο τότε η T λέγεται **γραμμικός μετασχηματισμός**

Ορισμός 2.3. [58] Έστω $X = (x, y) \in \mathbb{R}^2$ και $X_0 = (x_0, y_0) \in \mathbb{R}^2$. Η σύνθεση $S := L \circ T_0$ ενός γραμμικού μετασχηματισμού $L : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ με $L(x, y) = (x', y') = (a_{11}x + a_{12}y, a_{21}x + a_{22}y)$ και με πίνακα ως προς την κανονική βάση τον

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

με έναν μη γραμμικό μετασχηματισμό $T_0 : \mathbb{R}^n \rightarrow \mathbb{R}^n$, με $T_0(x, y) = (x + x_0, y + y_0)$, ονομάζεται **δισδιάστατος ομοπαράλληλικός μετασχηματισμός**:

$$S = AX + X_0$$

Η βασική γεωμετρική ιδιότητα του αφινικού μετασχηματισμού είναι ότι απεικονίζει, εν γένει, ένα ευθύγραμμο τμήμα σε ευθύγραμμο τμήμα με άκρα τις εικόνες των άκρων του αρχικού. Σε αντίθεση με το γραμμικό μετασχηματισμό, δεν διατηρεί σταθερή την αρχή του συστήματος, προσφέροντας τη δυνατότητα να περιγραφούν στροφές γύρω από σημεία διαφορετικά της αρχής, ή συμμετρίες ως προς ευθείες που δεν διέρχονται από την αρχή.

Παράρτημα C

Αποτελέσματα στα Δεδομένα

Validation

Στη συνέχεια παρουσιάζονται τα αποτελέσματα των καλύτερων, όπως βρέθηκαν από την πειραματική διαδικασία, διατάξεων για την ταξινόμηση των δυναμικών και των παθητικών χειρονομιών.

C.1 Αποτελέσματα Ταξινόμησης Δυναμικών Χειρονομιών

C.1.1 Αποτελέσματα Validation0

Macro Metrics

		Predicted				
		Pinch	Handflip out	Handflip in	Punchflip out	Undefined
Actual	Pinch	1	0	0	0	0
	Handflip out	0	2	2	0	0
	Handflip in	0	0	3	1	0
	Punchflip out	0	0	0	1	0

Πίνακας C.1: Δεδομένα Validation0: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics				
Pinch	Handflip out	Handflip in	Punchflip out	Accuracy
1.000	0.667	0.667	0.667	0.700

Πίνακας C.2: Δεδομένα Validation0: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

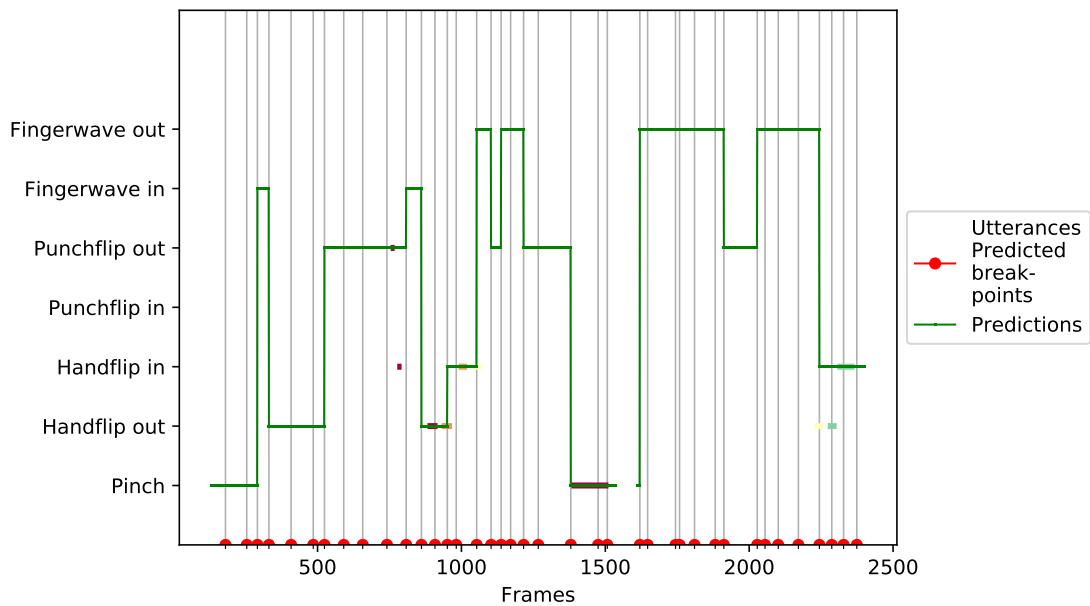
Micro Metrics

		Predicted				
		Pinch	Handflip out	Handflip in	Punchflip out	Undefined
Actual	Pinch	131	0	0	0	0
	Handflip out	0	57	70	0	18
	Handflip in	0	0	114	17	18
	Punchflip out	0	0	0	16	0

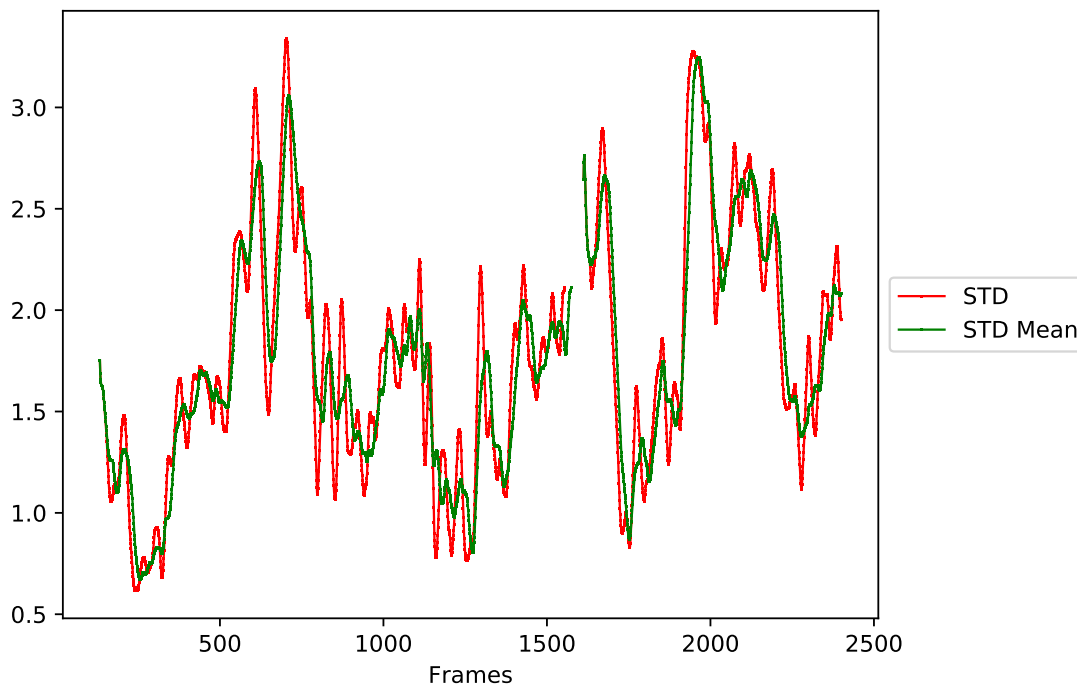
Πίνακας C.3: Δεδομένα Validation0: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames.

Micro Metrics				
Pinch	Handflip out	Handflip in	Punchflip out	Accuracy
1.000	0.564	0.685	0.653	0.721

Πίνακας C.4: Δεδομένα Validation0: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames.



Σχήμα C.1: Δεδομένα Validation0: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών , ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.



Σχήμα C.2: Δεδομένα Validation0: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances) των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών .

C.1.2 Αποτελέσματα Validation1

Macro Metrics

		Predicted					
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Undefined
Actual	Pinch	1	0	0	0	0	0
	Handflip out	0	3	4	0	0	0
	Handflip in	1	0	7	0	1	8
	Punchflip in	0	0	0	0	1	0
	Punchflip out	0	0	0	0	1	0

Πίνακας C.5: Δεδομένα Validation1: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics					
Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Accuracy
0.667	0.600	0.500	0.000	0.500	0.444

Πίνακας C.6: Δεδομένα Validation1: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πως έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

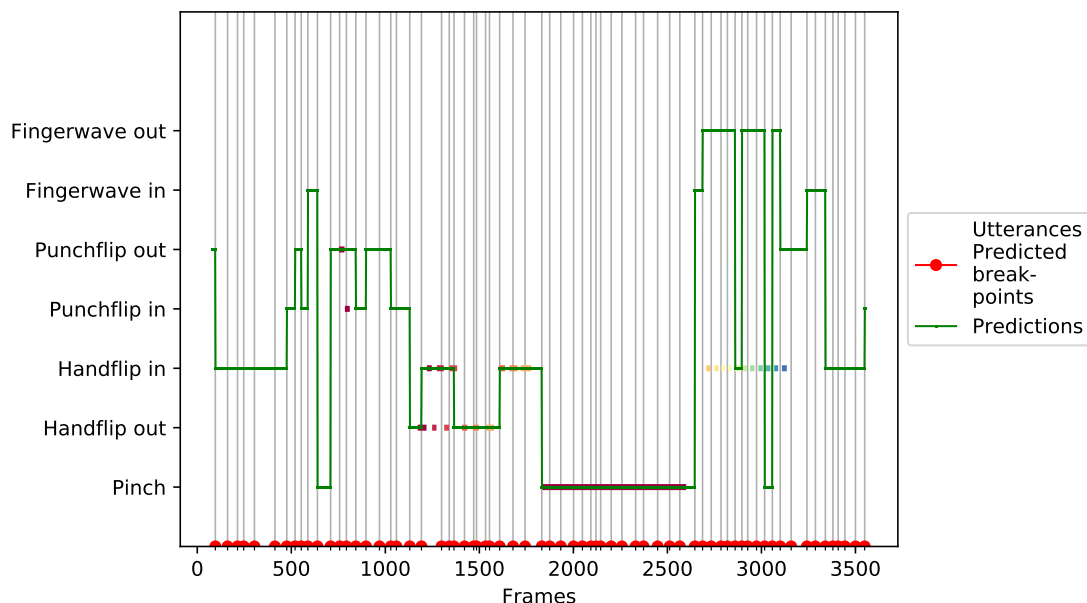
Micro Metrics

		Predicted					
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Undefined
Actual	Pinch	767	0	0	0	0	0
	Handflip out	0	127	80	0	0	0
	Handflip in	30	21	253	0	28	195
	Punchflip in	0	0	0	0	28	0
	Punchflip out	0	0	0	0	31	0

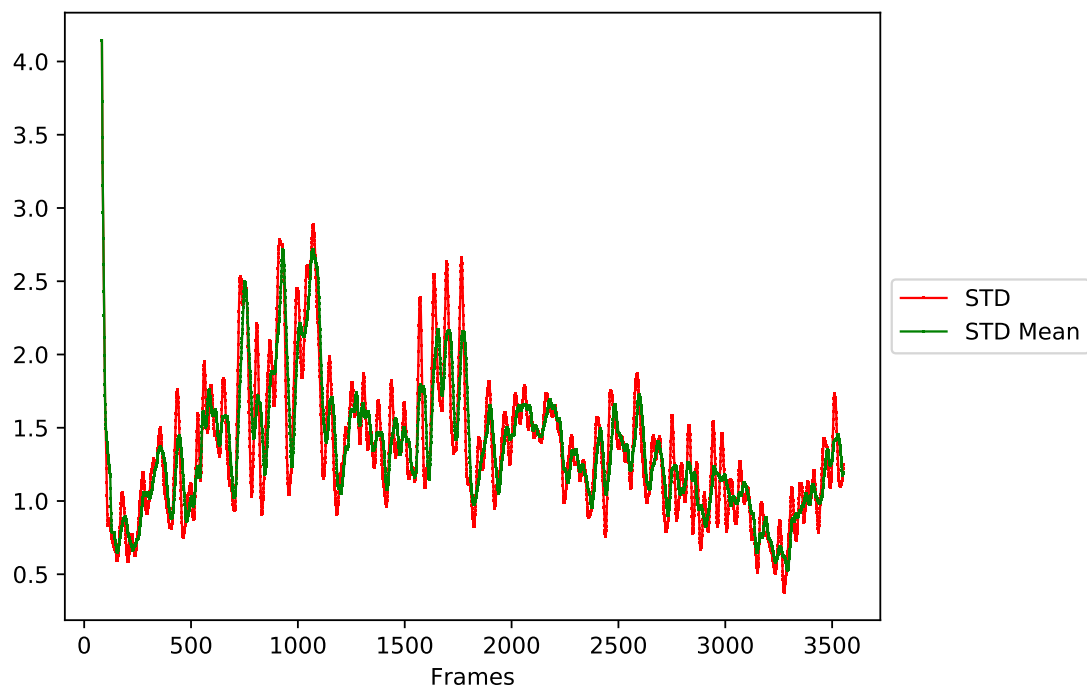
Πίνακας C.7: Δεδομένα Validation1: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames.

Micro Metrics					
Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Accuracy
0.981	0.715	0.588	0.000	0.525	0.755

Πίνακας C.8: Δεδομένα Validation1: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames.



Σχήμα C.3: Δεδομένα Validation1: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών, ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.



Σχήμα C.4: Δεδομένα Validation1: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances) των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών.

C.1.3 Αποτελέσματα Validation2

Macro Metrics

		Predicted							Undefined
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	
Actual	Pinch	7	0	1	0	0	0	0	0
	Handflip out	0	6	3	0	1	0	0	0
	Handflip in	0	0	9	0	0	0	0	0
	Punchflip in	0	0	0	0	9	0	0	0
	Punchflip out	1	0	0	0	17	0	0	0
	Fingerwave in	3	4	0	0	0	5	4	0
	Fingerwave out	0	4	0	0	0	1	3	0

Πίνακας C.9: Δεδομένα Validation2: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics

Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
0.737	0.500	0.818	0.000	0.756	0.455	0.400	0.603

Πίνακας C.10: Δεδομένα Validation2: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Micro Metrics

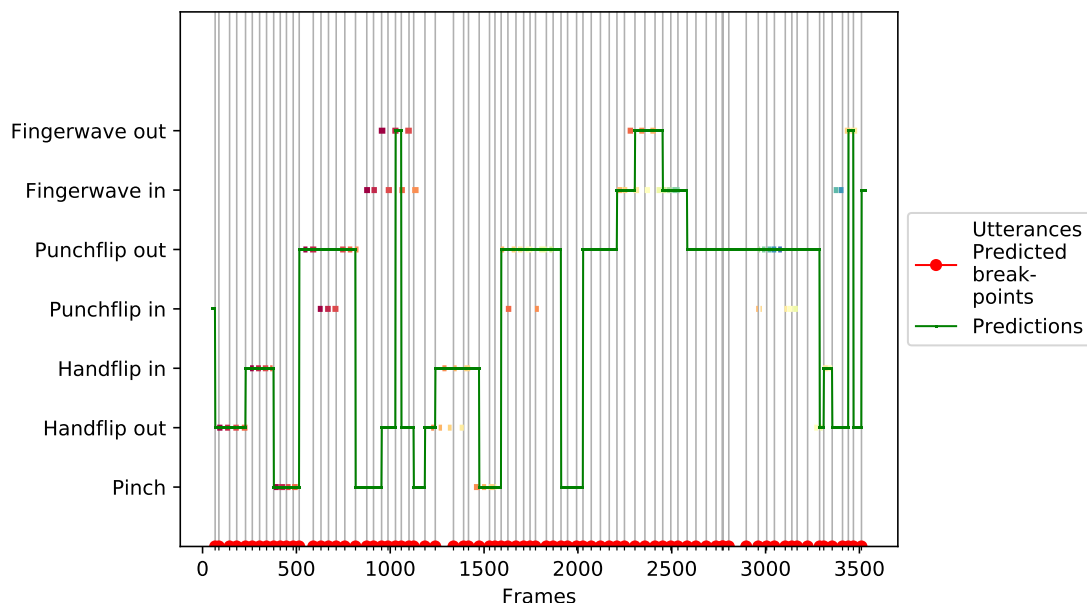
		Predicted							Undefined
		Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	
Actual	Pinch	188	0	28	0	0	0	0	0
	Handflip out	0	142	78	0	23	0	0	0
	Handflip in	0	6	217	0	0	0	0	0
	Punchflip in	0	0	0	0	240	0	0	0
	Punchflip out	20	0	0	0	393	0	0	0
	Fingerwave in	92	117	0	0	0	155	96	0
	Fingerwave out	15	117	0	0	0	31	103	0

Πίνακας C.11: Δεδομένα Validation2: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames.

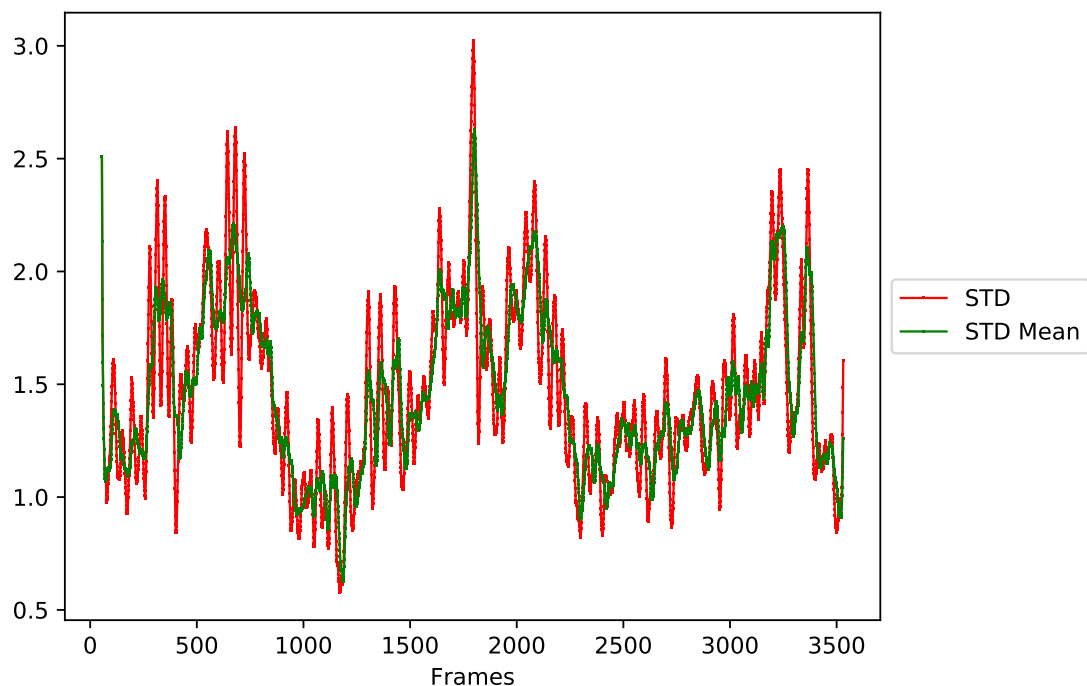
Micro Metrics

Pinch	Handflip out	Handflip in	Punchflip in	Punchflip out	Fingerwave in	Fingerwave out	Accuracy
0.708	0.454	0.795	0.000	0.735	0.480	0.443	0.581

Πίνακας C.12: Δεδομένα Validation2: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών ως προς τα frames.



Σχήμα C.5: Δεδομένα Validation2: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών, ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.



Σχήμα C.6: Δεδομένα Validation2: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances) των **δυναμικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **δυναμικών** χειρονομιών.

C.2 Αποτελέσματα Ταξινόμησης Παθητικών Χειρονομιών

C.2.1 Αποτελέσματα Validation0

Macro Metrics

		Predicted				
		Index	Punch	Palm	Tiger	Undefined
Actual	Index	1	0	2	0	0
	Punch	0	2	1	0	0
	Palm	0	0	3	0	0
	Tiger	0	0	0	1	0

Πίνακας C.13: Δεδομένα Validation0: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics				
Index	Punch	Palm	Tiger	Accuracy
0.500	0.800	0.667	1.000	0.700

Πίνακας C.14: Δεδομένα Validation0: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

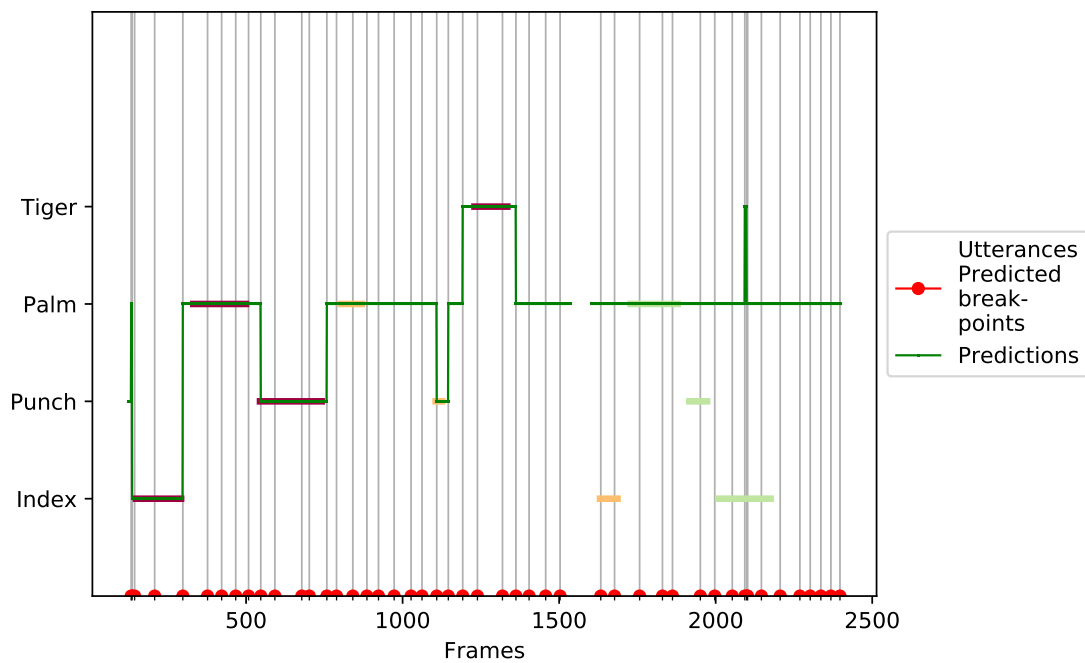
Micro Metrics

		Predicted				
		Index	Punch	Palm	Tiger	Undefined
Actual	Index	160	0	269	6	0
	Punch	0	237	107	0	0
	Palm	0	0	451	0	0
	Tiger	0	0	0	128	0

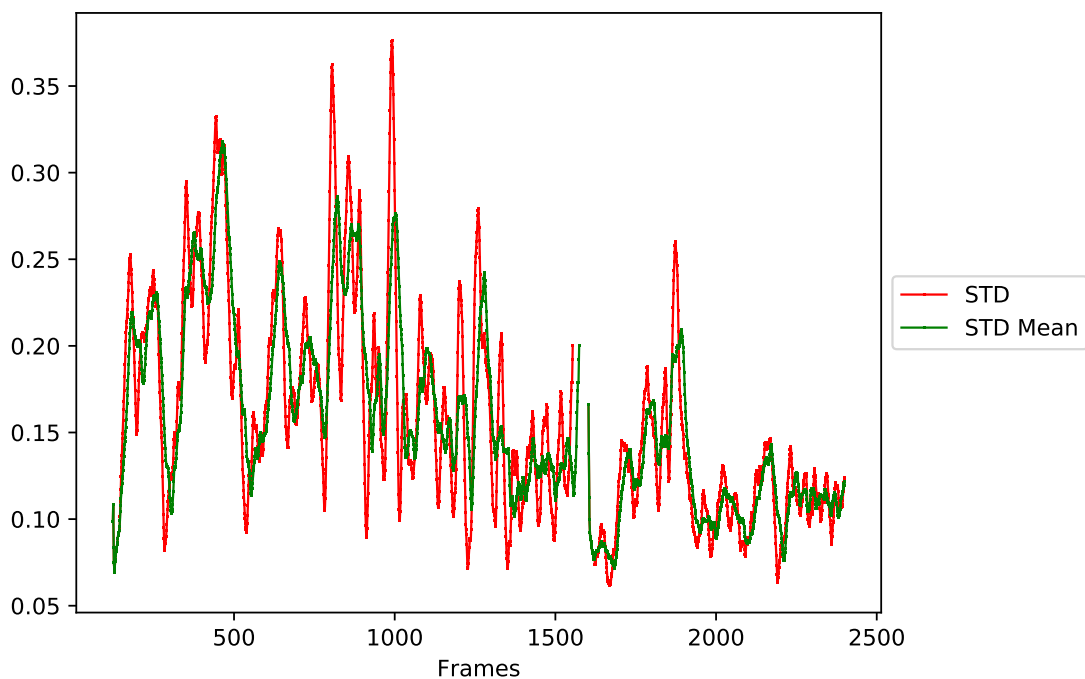
Πίνακας C.15: Δεδομένα Validation0: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames.

Micro Metrics				
Index	Punch	Palm	Tiger	Accuracy
0.538	0.816	0.706	0.977	0.719

Πίνακας C.16: Δεδομένα Validation0: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames.



Σχήμα C.7: Δεδομένα Validation0: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών, ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.



Σχήμα C.8: Δεδομένα Validation0: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances) των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών.

C.2.2 Αποτελέσματα Validation1

Macro Metrics

		Predicted				
		Index	Punch	Palm	Tiger	Undefined
Actual	Index	1	0	0	0	0
	Punch	0	2	0	0	0
	Palm	0	0	3	0	0
	Tiger	0	0	0	1	0

Πίνακας C.17: Δεδομένα Validation1: Πίνακας Σύγκρισης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics				
Index	Punch	Palm	Tiger	Accuracy
1.000	1.000	1.000	1.000	1.000

Πίνακας C.18: Δεδομένα Validation1: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

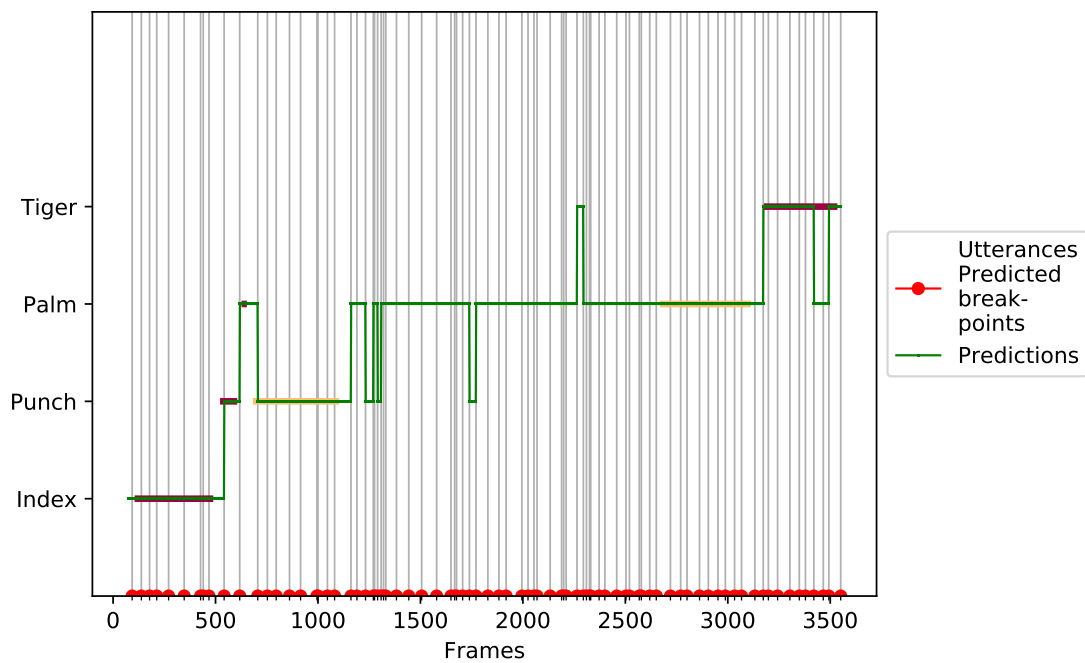
Micro Metrics

		Predicted				
		Index	Punch	Palm	Tiger	Undefined
Actual	Index	386	0	0	0	0
	Punch	20	465	23	0	0
	Palm	0	0	489	0	0
	Tiger	0	0	73	288	0

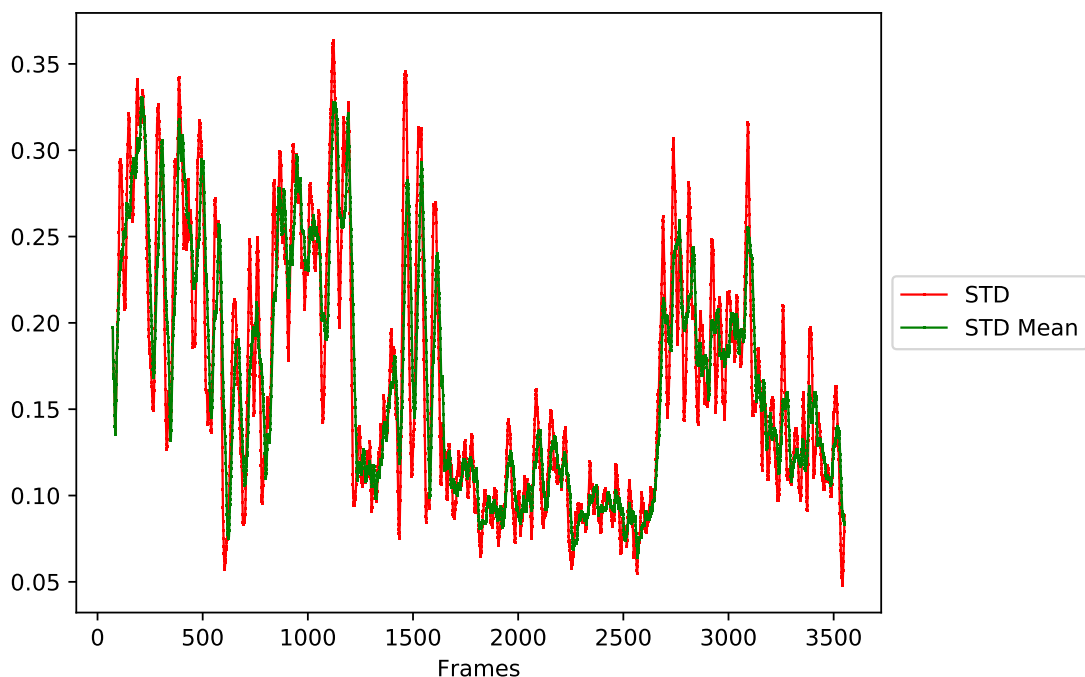
Πίνακας C.19: Δεδομένα Validation1: Πίνακας Σύγκρισης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames.

Micro Metrics				
Index	Punch	Palm	Tiger	Accuracy
0.975	0.956	0.911	0.888	0.933

Πίνακας C.20: Δεδομένα Validation1: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames.



Σχήμα C.9: Δεδομένα Validation1: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών, ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.



Σχήμα C.10: Δεδομένα Validation1: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances) των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών.

C.2.3 Αποτελέσματα Validation2

Macro Metrics

		Predicted				
		Index	Punch	Palm	Tiger	Undefined
Actual	Index	2	0	1	0	0
	Punch	0	18	2	1	0
	Palm	0	0	16	0	0
	Tiger	0	0	1	1	0

Πίνακας C.21: Δεδομένα Validation2: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

Macro Metrics				
Index	Punch	Palm	Tiger	Accuracy
0.800	0.923	0.889	0.500	0.881

Πίνακας C.22: Δεδομένα Validation2: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τις εμφανίσεις των χειρονομιών. Μια χειρονομία θεωρείται πώς έχει εντοπιστεί, αν το 50% των frames στη διάρκεια μιας εμφάνισης που ορίζεται από το Ground Truth των δεδομένων έχει χαρακτηριστεί από τον ταξινομητή ότι ανήκει σε αυτή. Αν δεν υπάρχει τέτοια χειρονομία για κάποια εμφάνιση, η εμφάνιση αυτή θεωρείται αρνητικό δείγμα για όλες τις χειρονομίες.

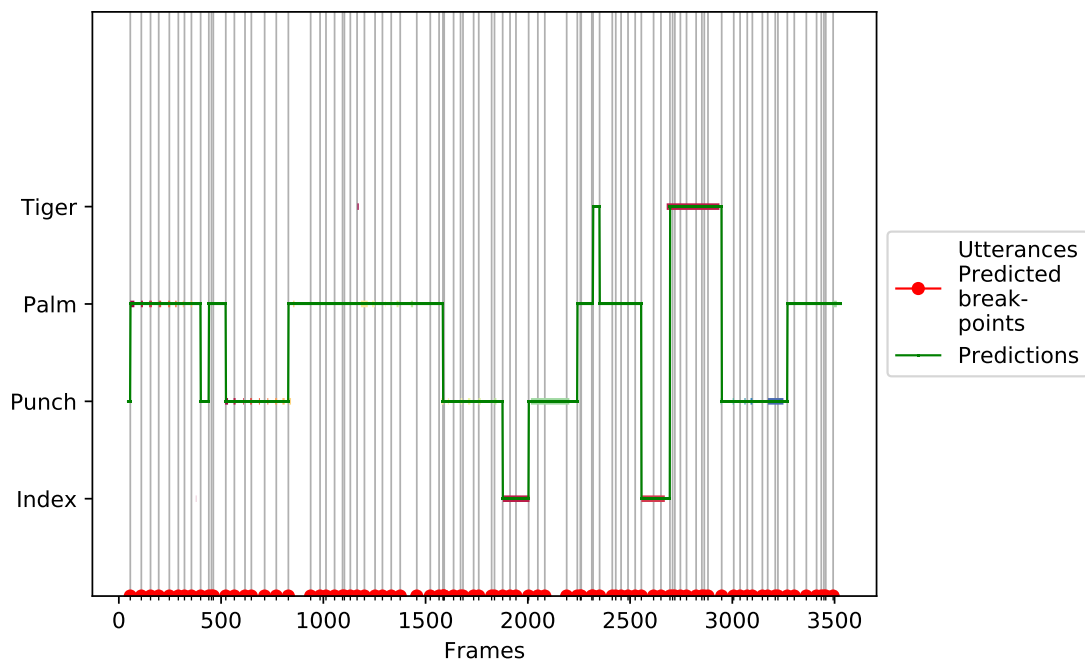
Micro Metrics

		Predicted				
		Index	Punch	Palm	Tiger	Undefined
Actual	Index	243	6	3	0	0
	Punch	0	438	21	5	0
	Palm	0	0	216	0	0
	Tiger	16	0	8	244	0

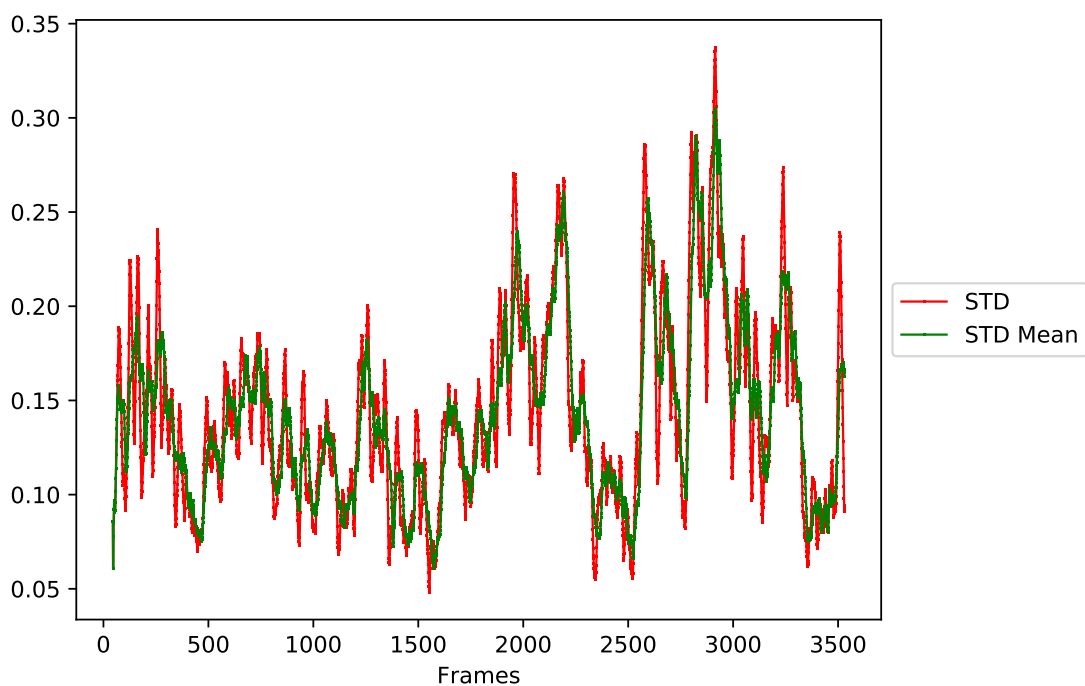
Πίνακας C.23: Δεδομένα Validation2: Πίνακας Σύγχυσης (Confusion Matrix) από την ταξινόμηση των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames.

Micro Metrics				
Index	Punch	Palm	Tiger	Accuracy
0.951	0.965	0.931	0.944	0.951

Πίνακας C.24: Δεδομένα Validation2: F-Scores και Ακρίβεια της ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών ως προς τα frames.



Σχήμα C.11: Δεδομένα Validation2: Διάγραμμα Αποτελεσμάτων Ταξινόμησης για τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών, ως προς τα frames. Ως συνεχής με σταθερό χρώμα απεικονίζεται η προβλεφθείσα κλάση ανά frame από τον ταξινομητή, ενώ με μεταβαλλόμενα χρώματα παρουσιάζεται η επισήμανση των δεδομένων. Η τεχνική CSTD επιτρέπει την εύρεση των ορίων αρχής και τέλους κάθε εμφάνισης χειρονομίας, γεγονός που υποδεικνύουν οι κατακόρυφες συνεχείς γραμμές.



Σχήμα C.12: Δεδομένα Validation2: Μετρική Τυπικής Απόκλισης Scores για τον καθορισμό της έναρξης και της λήξης των εμφανίσεων (utterances) των **παθητικών** χειρονομιών από τον τελικό καλύτερο Ταξινομητή των **παθητικών** χειρονομιών.

Βιβλιογραφία

- [1] Lisa Anthony και Jacob Wobbrock. “\$N-protractor: a fast and accurate multistroke recognizer”. Στο: *Proceedings of Graphics Interface 2012*. GI 2012. Canadian Human-Computer Communications Society, 2012, σσ. 117–120. isbn: 978-1-4503-1420-6. doi: [10.20380/GI2012.15](https://doi.org/10.20380/GI2012.15).
- [2] S. Banerjee και A. Roy. *Linear Algebra and Matrix Analysis for Statistics*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2014, σσ. 371–398. isbn: 9781420095388.
- [3] G. Bradski. Στο: *Dr. Dobb's Journal of Software Tools* (2000).
- [4] Leo Breiman. “Bagging Predictors”. Στο: *Maching Learning* 24.2 (1996), σσ. 123–140. issn: 0885-6125. doi: [10.1023/A:1018054314350](https://doi.org/10.1023/A:1018054314350).
- [5] Leo Breiman. *Classification and regression trees*. Wadsworth International Group, 1984, σ. 358. isbn: 0412048418.
- [6] Rizwan Chaudhry κ.ά. “Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions”. Στο: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2009.
- [7] Thomas H. Cormen κ.ά. *Introduction to Algorithms, Third Edition*. 2009.
- [8] Corinna Cortes και Vladimir Vapnik. “Support-Vector Networks”. Στο: *Mach. Learn.* 20.3 (1995), σσ. 273–297. issn: 0885-6125. doi: [10.1023/A:1022627411411](https://doi.org/10.1023/A:1022627411411).
- [9] B. Efron κ.ά. “Least Angle Regression”. Στο: *ArXiv Mathematics e-prints* (2004).
- [10] Sergio Escalera, Vassilis Athitsos και Isabelle Guyon. “Challenges in Multimodal Gesture Recognition”. Στο: *J. Mach. Learn. Res.* 17.1 (2016), σσ. 2549–2602. issn: 1532-4435.
- [11] Sean Ryan Fanello κ.ά. “Keep It Simple And Sparse: Real-Time Action Recognition”. Στο: *Journal of Machine Learning Research* 14 (2013), σσ. 2617–2640.

- [12] Gunnar Farneback. “Two-Frame Motion Estimation Based on Polynomial Expansion”. Στο: *Image Analysis: 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29 – July 2, 2003 Proceedings*. Επιμέλεια υπό Josef Bigun και Tomas Gustavsson. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, σσ. 363–370. isbn: 978-3-540-45103-7. doi: [10.1007/3-540-45103-X_50](https://doi.org/10.1007/3-540-45103-X_50).
- [13] Alp Guler κ.ά. “Human Joint Angle Estimation and Gesture Recognition for Assistive Robotic Vision”. Στο: *Computer Vision – ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II*. Επιμέλεια υπό Gang Hua και Hervé Jégou. Cham: Springer International Publishing, 2016, σσ. 415–431. isbn: 978-3-319-48881-3. doi: [10.1007/978-3-319-48881-3_29](https://doi.org/10.1007/978-3-319-48881-3_29).
- [14] Tin Kam Ho. “Random Decision Forests”. Στο: *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1*. ICDAR '95. Washington, DC, USA: IEEE Computer Society, 1995, σσ. 278–. isbn: 0-8186-7128-9.
- [15] Earl B. Hunt, Janet Marín και Philip J. Stone. *Experiments in induction*. Academic Press, 1966.
- [16] I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002. isbn: 9780387954424.
- [17] David Joseph Tan κ.ά. “Fits like a glove: Rapid and reliable hand shape personalization”. Στο: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, σσ. 5610–5619.
- [18] P. KaewTraKulPong και R. Bowden. “An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection”. Στο: *Video-Based Surveillance Systems: Computer Vision and Distributed Processing*. Επιμέλεια υπό Paolo Remagnino κ.ά. Boston, MA: Springer US, 2002, σσ. 135–144. isbn: 978-1-4615-0913-4. doi: [10.1007/978-1-4615-0913-4_11](https://doi.org/10.1007/978-1-4615-0913-4_11).
- [19] Byeongkeun Kang, Subarna Tripathi και Truong Q. Nguyen. “Real-time Sign Language Fingerspelling Recognition using Convolutional Neural Networks from Depth map”. Στο: *CoRR* abs/1509.03001 (2015).
- [20] Robert M. Krauss κ.ά. “The Communicative Value of Conversational Hand Gesture”. Στο: *Journal of Experimental Social Psychology* 31.6 (1995), σσ. 533 –552. issn: 0022-1031. doi: <http://dx.doi.org/10.1006/jesp.1995.1024>.

- [21] Honglak Lee κ.ά. “Efficient sparse coding algorithms”. Στο: *Advances in Neural Information Processing Systems 19*. Επιμέλεια υπό P. B. Schölkopf, J. C. Platt και T. Hoffman. MIT Press, 2007, σσ. 801–808.
- [22] SK Lee, William Buxton και K. C. Smith. “A Multi-touch Three Dimensional Touch-sensitive Tablet”. Στο: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '85. ACM, 1985, σσ. 21–25. isbn: 0-89791-149-0. doi: [10.1145/317456.317461](https://doi.org/10.1145/317456.317461).
- [23] Kui Liu και Nasser Kehtarnavaz. “Real-time robust vision-based hand gesture recognition using stereo images”. Στο: *Journal of Real-Time Image Processing* 11.1 (2016), σσ. 201–209. issn: 1861-8219. doi: [10.1007/s11554-013-0333-6](https://doi.org/10.1007/s11554-013-0333-6).
- [24] Bruce D. Lucas και Takeo Kanade. “An Iterative Image Registration Technique with an Application to Stereo Vision”. Στο: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'81. Morgan Kaufmann Publishers Inc., 1981, σσ. 674–679.
- [25] P. Luckey κ.ά. *Virtual reality headset*. US Patent D701,206. 2014.
- [26] A. Makris, N. Kyriazis και A. A. Argyros. “Hierarchical particle filtering for 3D hand tracking”. Στο: *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2015, σσ. 8–17. doi: [10.1109/CVPRW.2015.7301343](https://doi.org/10.1109/CVPRW.2015.7301343).
- [27] R.K. McConnell. *Method of and apparatus for pattern recognition*. US Patent 4,567,610. 1986.
- [28] P. Molchanov κ.ά. “Hand gesture recognition with 3D convolutional neural networks”. Στο: *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2015, σσ. 1–7. doi: [10.1109/CVPRW.2015.7301342](https://doi.org/10.1109/CVPRW.2015.7301342).
- [29] E. Na’aman, A. Shashua και Y. Wexler. *A user wearable visual assistance system*. EP Patent App. EP20,120,275,016. 2012.
- [30] Stephen G. Nash. “Newton-Type Minimization via the Lanczos Method”. Στο: *SIAM Journal on Numerical Analysis* 21.4 (1984), σσ. 770–788. doi: [10.1137/0721052](https://doi.org/10.1137/0721052).
- [31] Ebrahim Nasr-Esfahani κ.ά. “Hand Gesture Recognition for Contactless Device Control in Operating Rooms”. Στο: *CoRR* abs/1611.04138 (2016).

- [32] Andrew Y. Ng. “Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance”. Στο: *Proceedings of the Twenty-first International Conference on Machine Learning*. ICML '04. ACM, 2004, σσ. 78–. isbn: 1-58113-838-5. doi: [10 . 1145 / 1015330 . 1015435](https://doi.org/10.1145/1015330.1015435).
- [33] Jorge Nocedal και Stephen J. Wright. “Conjugate Gradient Methods”. Στο: *Numerical Optimization*. New York, NY: Springer New York, 2006, σσ. 101–134. isbn: 978-0-387-40065-5. doi: [10.1007/978-0-387-40065-5_5](https://doi.org/10.1007/978-0-387-40065-5_5).
- [34] Markus Oberweger, Paul Wohlhart και Vincent Lepetit. “Hands Deep in Deep Learning for Hand Pose Estimation”. Στο: *CoRR abs/1502.06807* (2015).
- [35] John C. Platt. “Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods”. Στο: *Advances in Large Margin Classifiers*. MIT Press, 1999, σσ. 61–74.
- [36] Morgan Quigley κ.ά. “ROS: an open-source Robot Operating System”. Στο: *ICRA Workshop on Open Source Software*. 2009.
- [37] I. Rodomagoulakis κ.ά. “Multimodal human action recognition in assistive human-robot interaction”. Στο: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2016, σσ. 2702–2706. doi: [10.1109/ICASSP.2016.7472168](https://doi.org/10.1109/ICASSP.2016.7472168).
- [38] Rosa Ruiloba, Stephane March και Georges Quenot. “Towards a Standard Protocol for the Evaluation of Video-to-Shots Segmentation Algorithms”. Στο: *First European Workshop on Content-Based Multimedia Indexing*. 1999.
- [39] A. Shpunt. *Optical designs for zero order reduction*. US Patent App. 12/330,766. 2009.
- [40] Khor Wee Sim κ.ά. “Augmented and virtual reality in surgery—the digital surgical environment: applications, limitations and legal pitfalls”. Στο: *Annals of Translational Medicine* 4.23 (2016). issn: 2305-5847.
- [41] Srinath Sridhar κ.ά. “Fast and Robust Hand Tracking Using Detection-Guided Optimization”. Στο: *Proceedings of Computer Vision and Pattern Recognition (CVPR)*. 2015.
- [42] Tilo Strutz. Στο: *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond*. Germany: Vieweg και Teubner, 2010, σσ. 47–103. isbn: 3834810223, 9783834810229.
- [43] S. Suzuki και K. Abe. “Topological Structural Analysis of Digitized Binary Images by Border Following”. Στο: *Computer Vision, Graphics, and Image Processing* 30 (1985), σσ. 32–46.

- [44] Jason Tanz. *Kinect Hackers Are Changing the Future of Robotics*. 2011.
- [45] *Ubi-Interactive*. url: <http://forum.ubi-interactive.com/>. 2017.
- [46] L. Vincent και P. Soille. “Watersheds in digital spaces: an efficient algorithm based on immersion simulations”. Στο: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13.6 (1991), σσ. 583–598. issn: 0162-8828. doi: [10.1109/34.87344](https://doi.org/10.1109/34.87344).
- [47] C. Vogel, K. Schindler και S. Roth. “Piecewise Rigid Scene Flow”. Στο: *2013 IEEE International Conference on Computer Vision*. 2013, σσ. 1377–1384. doi: [10.1109/ICCV.2013.174](https://doi.org/10.1109/ICCV.2013.174).
- [48] W.C. Westerman και J.G. Elias. *System and method for packing multitouch gestures onto a hand*. US Patent 7,705,830. 2010.
- [49] Aaron Wetzler, Ron Slossberg και Ron Kimmel. “Rule Of Thumb: Deep derotation for improved fingertip detection”. Στο: *CoRR abs/1507.05726* (2015).
- [50] Thiemo Wiedemeyer. *IAI Kinect2*. Accessed June 12, 2015. University Bremen: Institute for Artificial Intelligence, 2014 – 2015.
- [51] Christopher J. Wilson και Alessandro Soranzo. “The Use of Virtual Reality in Psychology: A Case Study in Visual Perception”. Στο: *Comp. Math. Methods in Medicine*. 2015.
- [52] Y. Xiong και S. A. Shafer. “Depth from focusing and defocusing”. Στο: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 1993, σσ. 68–73. doi: [10.1109/CVPR.1993.340977](https://doi.org/10.1109/CVPR.1993.340977).
- [53] Jianchao Yang κ.ά. “Linear spatial pyramid matching using sparse coding for image classification”. Στο: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, σσ. 1794–1801. doi: [10.1109/CVPR.2009.5206757](https://doi.org/10.1109/CVPR.2009.5206757).
- [54] Y. Yuan. “On the truncated conjugate gradient method”. Στο: *Mathematical Programming* 87.3 (2000), σσ. 561–573. issn: 1436-4646. doi: [10.1007/s101070050012](https://doi.org/10.1007/s101070050012).
- [55] T.G. Zimmerman και J.Z. Lanier. *Computer data entry and manipulation apparatus and method*. US Patent 4,988,981. 1991.
- [56] Zoran Zuvkovic. “Improved adaptive Gaussian mixture model for background subtraction”. Στο: *Proceedings of the 17th International Conference on*. Τόμ. 2. 2004, σσ. 1236–1239.
- [57] Γ.Δ Ακρίβης και Β.Α. Δούγαλης. Στο: *Εισαγωγή στην Αριθμητική Ανάλυση*. 2011, σσ. 249–255. isbn: 978-960-524-022-6.

-
- [58] Σ. Καρανάσιος. Στο: *Γραμμική Άλγεβρα, Αναλυτική Γεωμετρία και Εφαρμογές*. 2011, σσ. 277–295. isbn: 960-91725-0-4.
- [59] Θ. Ρασσιάς. Στο: *Μαθηματική Ανάλυση II*. Τόμ. 1. 2007, σσ. 332–335. isbn: 9607888820.
- [60] Θ. Ρασσιάς. Στο: *Μαθηματική Ανάλυση II*. Τόμ. 2. 2007, σσ. 332–344. isbn: 960940006X.