



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ & ΡΟΜΠΟΤΙΚΗΣ  
ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ,  
ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ & ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ

**Ανίχνευση και Κατηγοριοποίηση Ανθρωπίνων Κινήσεων  
χρησιμοποιώντας Ιστογραφικές Μεθόδους και Μαρκοβιανά  
Μοντέλα με εφαρμογή στις Πολεμικές Τέχνες.**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Σωτήριος Δ. Στασινόπουλος**

**Επιβλέπων:** Μαραγκός Πέτρος  
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2011





.....  
**Σωτήριος Δ. Στασινόπουλος**

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

**Copyright © Σωτήριος Δ. Στασινόπουλος, 2011, Εθνικό Μετσόβιο Πολυτεχνείο.  
Με επιφύλαξη παντός δικαιώματος.All rights reserved.**

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

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

## Σύνοψη

Η Ανίχνευση και Κατηγοριοποίηση Ανθρωπίνων Κινήσεων, ως τομέας σύμπραξης της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων, χρησιμοποιείται σε ολοένα και περισσότερες εφαρμογές για την περιγραφή της ανθρώπινης δραστηριότητας. Στα πλαίσια της Διπλωματικής αυτής Εργασίας, εισάγουμε ένα νέο σύστημα τέτοιας μορφής που βασίζεται στη χρήση τοπικών χωρο-χρονικών χαρακτηριστικών από Ιστογραφικές μεθόδους. Τέτοια χαρακτηριστικά είναι αυτά που εξάγονται από Ιστογράμματα Προσανατολισμένης Κλίσης (Histograms of Oriented Gradient-HOG) και από Ιστογράμματα Οπτικής Ροής (Histograms of Optical Flow-HOF), ενώ επιχειρούμε τη μείωση της διαστασιμότητας του προβλήματος εφαρμόζοντας Ανάλυση Πρωτεύουσών Συνιστωσών (PCA) πάνω σε αυτά. Στη συνέχεια, καθώς επιθυμούμε να επικεντρώσουμε την εφαρμογή του συστήματος σε δυναμικά προβλήματα, όπως την αναγνώριση και κατηγοριοποίηση κινήσεων από Πολεμικές Τέχνες, συνδυάζουμε αυτά τα χαρακτηριστικά με χρήση Κρυφών Μαρκοβιανών Μοντέλων (HMM) για να εκπαιδεύσουμε το μοντέλο της κάθε κίνησης. Το σύστημά μας δοκιμάζεται σε μία βάση δεδομένων από ακολουθίες κινήσεων shotokan karate (katas) που κατασκευάζουμε εμείς, ύστερα από βιντεοσκόπηση παιδιών-αθλητών karate. Τα ποσοστά ακρίβειας που βρίσκουμε με την πειραματική μας διαδικασία είναι αρκετά ενθαρρυντικά για εκτενέστερη μελέτη στο αντικείμενο. Παράλληλα, επιχειρούμε να προσδώσουμε και έναν εκπαιδευτικό χαρακτήρα στην εφαρμογή μας με την εξαγωγή βαθμολογίας για την ακρίβεια εκτέλεσης της κίνησης με βάση το πρότυπο που έχουμε κατασκευάσει.

## Λέξεις Κλειδιά

αναγνώριση ανθρωπίνων κινήσεων, πολεμικές τέχνες, καράτε, ιστογραφικές μέθοδοι, ιστόγραμμα προσανατολισμένης κλίσης, ιστόγραμμα οπτικής ροής, κρυφά μαρκοβιανά μοντέλα, ανάλυση πρωτεύουσών συνιστωσών

## **Abstract**

The Detection and Classification of Human Movements, as a joint field of Computer Vision and Pattern Recognition, is used with an increasing rate in applications designed to describe human activity. With this Diploma Thesis, we are proposing a novel system of this type, based on the use of local spatio-temporal features from Histogrammic methods. Such features are those extracted by Histograms of Oriented Gradient (HOG) and Histograms of Optical Flow (HOF), while we also pursue the reduction of the problem's dimensionality by applying on them Principal Components Analysis (PCA). In continuation, since our desire is to focus the application of this system on dynamic problems, like the detection and classification of movements from Martial Arts, we combine these features with the use of Hidden Markov Models (HMM) in order to train the model for each different movement. Our system is tested upon a database, comprising sequences of shotokan karate movements (katas), that was created by us for the needs of this research after filming children-karate athletes. The accuracy rates attained with our experimental procedure are greatly encouraging for further research on this specific field. In parallel, we attempt to additionally attach an educational character to our application, with the extraction of a score for the accuracy of execution of each movement based on the prototype we have built.

## **Key Words**

human action recognition, martial arts, karate, histogrammic methods, histogram of oriented gradient, hog, histogram of optic flow, hof, hidden markov models, hmm, principal component analysis, pca

# Περιεχόμενα

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Εισαγωγή</b>  | <b>9</b>  |
| 1.1      | Στόχοι της Διπλωματικής Εργασίας                                       | 9         |
| 1.2      | Δομή Διπλωματικής Εργασίας   | 11        |
| 1.3      | Ευχαριστίες  | 13        |
| <b>2</b> | <b>Εισαγωγή στην Ανίχνευση και Κατηγοριοποίηση Ανθρωπίνων Κινήσεων</b> | <b>15</b> |
| 2.1      | Ερευνητικό Υπόβαθρο  | 15        |
| 2.1.1    | Εντοπισμός   | 16        |
| 2.1.2    | Εξαγωγή Χαρακτηριστικών  | 20        |
| 2.1.3    | Δημιουργία Μοντέλων  | 21        |
| 2.2      | Σημεία Ενδιαφέροντος-Interest Points                                   | 25        |
| 2.3      | Ιστογραφικές Μέθοδοι   | 28        |
| 2.3.1    | Ιστόγραμμα Προσανατολισμένης Κλίσης                                    | 29        |
| 2.3.2    | Ιστόγραμμα Οπτικής Ροής  | 31        |
| 2.4      | Κρυφά Μαρκοβιανά Μοντέλα   | 33        |
| 2.4.1    | Ορισμός  | 34        |
| 2.4.2    | Ανίχνευση  | 35        |
| 2.4.3    | Εκπαίδευση   | 36        |
| 2.4.4    | Επεκτάσεις   | 37        |
| <b>3</b> | <b>Εφαρμογή στις Πολεμικές Τέχνες</b>                                  | <b>39</b> |
| 3.1      | Γενικά   | 39        |
| 3.1.1    | Χαρακτηριστικά   | 40        |
| 3.1.2    | Εφαρμογές  | 42        |
| 3.2      | Καράτε   | 43        |
| 3.2.1    | Ιστορία  | 44        |
| 3.2.2    | Κινήσεις   | 46        |
| <b>4</b> | <b>Μέθοδοι και Εργαλεία</b>  | <b>53</b> |
| 4.1      | Η προσέγγισή μας   | 53        |
| 4.2      | Εξαγωγή Χαρακτηριστικών  | 54        |
| 4.3      | PCA  | 54        |
| 4.4      | HMM  | 57        |

## ΠΕΡΙΕΧΟΜΕΝΑ

---

|          |   |           |
|----------|---|-----------|
| <b>5</b> | <b>Πειραματική Διαδικασία</b>                               | <b>61</b> |
| 5.1      | Δημιουργία Βάσης Δεδομένων . . . . .                        | 61        |
| 5.2      | Επεξεργασία Δεδομένων . . . . .                             | 62        |
| 5.3      | Δοκιμές & Αποτελέσματα . . . . .                            | 67        |
| <b>6</b> | <b>Συμπεράσματα</b>   | <b>81</b> |
| 6.1      | Αξιολόγηση Αποτελεσμάτων & Συστήματος . . . . .             | 81        |
| 6.1.1    | Αξιολόγηση Αποτελεσμάτων Πειραματικής Διαδικασίας . . . . . | 81        |
| 6.1.2    | Συνολική Αξιολόγηση Συστήματος . . . . .                    | 86        |
| 6.2      | Συνεισφορές Διπλωματικής Εργασίας . . . . .                 | 88        |
| 6.3      | Μελλοντικές Επεκτάσεις . . . . .                            | 89        |
|          | <b>Βιβλιογραφία</b>   | <b>91</b> |



# Κατάλογος σχημάτων

|      |   |    |
|------|---|----|
| 1.1  | Δομή Διπλωματικής Εργασίας . . . . .  | 11 |
| 2.1  | Τεχνικές εντοπισμού άκρων από ερευνητικές προσπάθειες ανίχνευσης χειρονομιών. (α') Blobs χεριών, όπως εμφανίζεται στο σχήμα του [1]. (β') Ακριβής εντοπισμός χεριών, όπως εμφανίζεται στο σχήμα του [2]. (γ') Τρισδιάστατα (3D) blobs, όπως εμφανίζεται στο σχήμα του [3]. . . . .            | 17 |
| 2.2  | Εντοπισμός ολόκληρου ανθρώπινου σώματος. (α') Εντοπισμός ανθρώπινης φιγούρας με τη χρήση MEI και MHI, όπως εμφανίζεται στο σχήμα του [4]. (β') Εξαγωγή πληροφοριών οπτικής ροής, όπως εμφανίζεται στο σχήμα του [5]. . . . .  | 18 |
| 2.3  | Εντοπισμός σημείων με βάση μοντέλα για αναπαράσταση ανθρώπινου σώματος. (α') Αναδημιουργία ανθρώπινης φιγούρας με Pictorial Structures, όπως εμφανίζεται στο σχήμα του [6]. (β') Προσαρμογή σκελετικού μοντέλου για εντοπισμό ανθρώπινης κίνησης, όπως εμφανίζεται στο σχήμα του [7]. . . . . | 19 |
| 2.4  | Απλά τοπικά χαρακτηριστικά. (α') Χαρακτηριστικό θέσης. (β') Χαρακτηριστικό προσανατολισμού. . . . .   | 20 |
| 2.5  | Διάφορες μέθοδοι εξαγωγής χαρακτηριστικών. (α') Εξαγωγή χαρακτηριστικών από 'προεξέχοντα' ( <i>salient</i> ) σημεία, όπως εμφανίζεται στο σχήμα του [8]. (β') Χρήση κυβοειδών περιοχών ( <i>cuboids</i> ) για τοπική εξαγωγή χαρακτηριστικών, όπως εμφανίζεται στο σχήμα του [9]. . . . .     | 22 |
| 2.6  | Μοντέλα αναπαράστασης του ανθρώπινου σώματος . . . . .  | 24 |
| 2.7  | Διάφορα είδη μοντέλων . . . . .   | 25 |
| 2.8  | Σημεία ενδιαφέροντος (Interest points) αμετάβλητα στην αλλαγή κλίμακας, όπως εξάγονται από τον ανιχνευτή Harris-Laplace του [10] . . . . .  | 27 |
| 2.9  | Ο ανιχνευτής Harris3D, όπως στο σχήμα του [11]. (α') Ανίχνευση σημείων ενδιαφέροντος στο χωροχρόνο. (β') Τρισδιάστατη απεικόνιση σημείων ενδιαφέροντος κατά τη διάρκεια περπατήματος. . . . .   | 29 |
| 2.10 | Απεικόνιση ανθρώπου και υπολογισμών της προσανατολισμένης κλίσης συνολικά, αλλά και στα διαφορετικά κελιά, για την εξαγωγή των HOG, όπως στο σχήμα του [12] . . . . .   | 30 |
| 2.11 | Επέκταση των HOG στις 3 διαστάσεις με ενσωμάτωση του χρόνου από τον Klaser στο [13] . . . . .   | 32 |

|      |   |    |
|------|---|----|
| 2.12 | Ιστογράμματα Οπτικής Ροής-HOF. (α') Υπολογισμός οπτικής ροής και πλάτους ροής μεταξύ δύο διαδοχικών εικόνων με ανάδειξη ορίων κίνησης. (β') Σχήμα κωδικοποίησης IMHcd, όπου ο υπολογισμός των διαφορών γίνεται μεταξύ αντίστοιχων pixel σε γειτονικά κελιά, όπως στο σχήμα του [14]               | 33 |
| 2.13 | Αρχιτεκτονική Κρυφών Μαρκοβιανών Μοντέλων-HMM.(α') Γενική μορφή ενός HMM. (β') Καταστάσεις που αποτελούν το HMM.  | 34 |
| 2.14 | Αλληλένδετο HMM (Coupled HMM)   | 37 |
| 3.1  | Απλούστερα και πιο σύνθετα προβλήματα ανίχνευσης και κατηγοριοποίησης ανθρωπίνων κινήσεων.  | 40 |
| 3.2  | Διάφορα είδη πολεμικών τεχνών. (α') Με λακτίσματα, αερόβια στριφογυριστή κλωτσιά. (β') Με λαβές, πέταγμα <i>judo</i> . (γ') Ιαπωνική τέχνη του σπαθιού ( <i>kenjutsu</i> ). (δ') Σύγχρονη ξιφομαχία ( <i>fencing</i> ). (ε') Στρατιωτικές εφαρμογές ακινητοποίησης.                               | 41 |
| 3.3  | Εφαρμογές σε διάφορα είδη πολεμικών τεχνών. (α') Αναγνώριση ακολουθίας κινήσεων Tai chi, όπως στο σχήμα του [15]. (β') Χρήση αρχετύπων κινήσεων για κατασκευή μοντέλων, όπως στο σχήμα του [16]. (γ') Αναγνώριση πολεμικών τεχνών για χρήσεις εικονικής πραγματικότητας, όπως στο σχήμα του [17]. | 44 |
| 3.4  | Είδη εκπαίδευσης αθλητικού καράτε. (α') Kihon, ομαδική εκπαίδευση. (β') Kumite, αναμέτρηση μεταξύ δύο αθλητών.  | 45 |
| 3.5  | Ακολουθία κινήσεων 1ου επιπέδου-Heian Shodan.   | 47 |
| 3.6  | Ακολουθία κινήσεων 2ου επιπέδου-Heian Nidan.  | 48 |
| 3.7  | Ακολουθία κινήσεων 3ου επιπέδου-Heian Sandan.   | 49 |
| 3.8  | Ακολουθία κινήσεων 4ου επιπέδου-Heian Yondan.   | 50 |
| 3.9  | Ακολουθία κινήσεων 5ου επιπέδου-Heian Godan.  | 51 |
| 4.1  | Το σύστημά μας  | 53 |
| 4.2  | Στάδια εξαγωγής χαρακτηριστικών από Ιστογράμματα Προσανατολισμένης Κλίσης-HOG και Οπτικής Ροής-HOF, μετά τον εντοπισμό χωροχρονικών σημείων ενδιαφέροντος.  | 55 |
| 4.3  | Παράδειγμα HMM μοντέλου που μπορεί να χρησιμοποιηθεί στο σύστημά μας με 5 καταστάσεις και μείγμα 3 Γκουσιανών ανά κατάσταση.  | 58 |
| 5.1  | Στιγμιότυπα από τις βιντεοσκοπήσεις παιδιών-αθλητών shotokan karate κατά την εκτέλεση των 5 πρώτων κινήσεων. (α') Λήψη χωρίς οπτικό θόρυβο παρασκηνίου (κινήσεις άλλων ατόμων ή αντικειμένων). (β') Λήψη με μικρό οπτικό θόρυβο παρασκηνίου. (γ') Λήψη με μεγαλύτερο οπτικό θόρυβο παρασκηνίου.   | 63 |

|      |  |    |
|------|--|----|
| 5.2  | Στιγμιότυπα από την ανίχνευση σημείων ενδιαφέροντος κατά τη διάρκεια των εκτελέσεων των 5 πρώτων κινήσεων shotokan karate. (α'), (β'), (γ') Εντοπισμός σημείων ενδιαφέροντος στις σωστές θέσεις, παρά τις κινήσεις στο παρασκήνιο. (δ') Εντοπισμός σημείων και πάνω στον προπονητή. (ε') Λάθος εντοπισμός σημείων ενδιαφέροντος πάνω στο κινούμενο παρασκήνιο αντί για τον αθλητή. . . . . | 64 |
| 5.3  | Τα HMM που προτείνουμε προς χρήση για την πειραματική μας πριν την πραγματοποίηση των δοκιμών, με βάση μόνο την παρατήρηση των κινήσεων που εξετάζουμε. Γραμμικά left-to-right HMM 6 καταστάσεων με 4 μείγματα Γκαουσιανών ανά κατάσταση. . . . .  | 67 |
| 5.4  | Διάγραμμα για <i>porca</i> δεδομένα και εκπαίδευση με 4 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g). . . . .  | 68 |
| 5.5  | Διάγραμμα για <i>porca</i> δεδομένα και εκπαίδευση με 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g). . . . .  | 70 |
| 5.6  | Πίνακες σύγχυσης (confusion matrices) για <i>porca</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα για τις συνολικά καλύτερες ακρίβειες που βρέθηκαν. . . . .  | 70 |
| 5.7  | Διάγραμμα για <i>pcahog</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g). . . . .   | 71 |
| 5.8  | Πίνακες σύγχυσης (confusion matrices) για <i>pcahog</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα για τις συνολικά καλύτερες ακρίβειες που βρέθηκαν. . . . .   | 72 |
| 5.9  | Διάγραμμα για <i>pcahof</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g). . . . .   | 73 |
| 5.10 | Πίνακες σύγχυσης (confusion matrices) για <i>pcahof</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα για τις συνολικά καλύτερες ακρίβειες που βρέθηκαν. . . . .   | 74 |
| 5.11 | Διάγραμμα για <i>pca</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g). . . . .  | 75 |
| 5.12 | Πίνακες σύγχυσης (confusion matrices) για <i>pca</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα για τις συνολικά καλύτερες ακρίβειες που βρέθηκαν. . . . .  | 76 |
| 5.13 | Διαγράμματα με τις βαθμολογίες των δειγμάτων σύμφωνα με τις 6 συναρτήσεις αντιστοίχισης για <i>pca</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα. Η βαθμολογία του δασκάλου επισημαίνεται με κόκκινο. . . . .  | 78 |

|      |   |    |
|------|---|----|
| 5.14 | Διαγράμματα με τις βαθμολογίες των δειγμάτων σύμφωνα με τις 6 συναρτήσεις αντιστοίχισης για <i>pcahog</i> δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα. Η βαθμολογία του δασκάλου επισημαίνεται με κόκκινο. . . . .  | 79 |
| 6.1  | Συγκριτικό διάγραμμα με όλες τις κατηγορίες δεδομένων και εκπαίδευση με (α') 4 και (β') 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 s και 1-4 g. . . . .   | 82 |
| 6.2  | Συγκριτικά διαγράμματα για την κατηγορία δεδομένων <i>pcahog</i> με (α') 4 και (β') 5 δείγματα με τις καμπύλες γραμμικής παρεμβολής για τα 6 είδη συναρτήσεων αντιστοίχισης βαθμολογίας σε σχέση με την καμπύλη γραμμικής παρεμβολής για τις βαθμολογίες από τον δάσκαλο. . . . . | 85 |

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Στόχοι της Διπλωματικής Εργασίας

Μπορούν οι Υπολογιστές να δουν πραγματικά όπως οι άνθρωποι; Και αν μπορούν, γίνεται να αποκτήσουν διακριτική ικανότητα μεταξύ διαφορετικών ανθρωπίνων κινήσεων, σε βαθμό μάλιστα που να είναι σε θέση να αξιολογούν και την ποιότητα πραγματοποίησης της ίδιας της κίνησης; Τέτοια ερωτήματα έκαναν την εμφάνισή τους όταν τα υπολογιστικά συστήματα άρχισαν να χρησιμοποιούνται για την επεξεργασία εικόνων και βίντεο, με συνέπεια να συγκεντρωθεί μεγάλο μέρος του ερευνητικού ενδιαφέροντος στην εξερεύνηση των δυνατοτήτων της συγκεκριμένης περιοχής. Ως αποτέλεσμα, γεννήθηκε ο επιστημονικός κλάδος Όραση Υπολογιστών, ο οποίος μέχρι σήμερα συνεχίζει να αναπτύσσεται και να συνδράμει στην γενικότερη επιστημονική πρόοδο.

Ερχόμενος σε επαφή με το μάθημα της ‘Όρασης Υπολογιστών’ κατά τη διάρκεια των προπτυχιακών σπουδών μου, πολλά από τα προσωπικά ερωτήματα που είχα σχετικά με τις οπτικές δυνατότητες των υπολογιστικών συστημάτων άρχισαν να παίρνουν μορφή. Η ενασχόληση με την εξαγωγή οπτικών χαρακτηριστικών από εικόνες και βίντεο και η χρήση τεχνικών με τις οποίες αλγοριθμικές διαδικασίες μπορούν να αποφανθούν τόσο εύκολα όσο το ανθρώπινο μάτι για την ύπαρξη ή όχι συγκεκριμένων στοιχείων, με βοήθησε να κατανοήσω μία διαδικασία για την οποία αναρωτιόμουν επί πολλά έτη και κέντρισε το ενδιαφέρον μου, ώστε να στραφώ με μεγαλύτερη προσήλωση προς τον συγκεκριμένο τομέα. Βέβαια, ορισμένα ερωτήματα για την αναγνώριση πιο πολύπλοκων διαδικασιών και την πιθανή ομαδοποίηση των εξαγόμενων χαρακτηριστικών παρέμεναν ακόμα.

Με την προσθήκη του μαθήματος ‘Αναγνώριση Προτύπων’, τα κομμάτια του παζλ άρχισαν να μπαίνουν στη θέση τους σχετικά με την πιο εξειδικευμένη μέθοδο για την αναγνώριση συγκεκριμένων χαρακτηριστικών από οπτικοακουστικό υλικό. Αυτά τα χαρακτηριστικά με τη σειρά τους, μπορούν με την κατάλληλη επεξεργασία να χρησιμοποιηθούν για τη δημιουργία ορισμένων προτύπων και την ταυτοποίηση ειδικών ακολουθιών, όπως είναι οι ανθρώπινες κινήσεις. Η μελέτη και ανάλυση του ανθρωπίνου προτύπου καθώς και των εκφάνσεών του, όπως είναι οι σωματικές κινήσεις, αποτελούν μια ερευνητική περιοχή που βρίσκεται πολύ κοντά στα ενδιαφέροντά μου, για αυτό και όλα αυτά τα ερ-

γαλεία που μου προσφέρθηκαν με βοήθησαν να καταλάβω πώς μπορούσα να ασχοληθώ πιο ενεργά πάνω στον συγκεκριμένο τομέα.

Βέβαια, το πιο αξιοθαύμαστο φαινόμενο σχετικά με την επιστήμη είναι πως όχι μόνο σου προσφέρει απαντήσεις στα ερωτήματά σου, αλλά σου δημιουργεί καινούργια ερωτήματα και σε παρακινεί να ερευνήσεις για να ανακαλύψεις τις απαντήσεις. Όσο πιο πολύ ασχολούμουν με την Όραση Υπολογιστών και την Αναγνώριση Προτύπων, τόσο πιο εξειδικευμένα θέματα έκαναν την εμφάνισή τους. Ποιες τεχνικές εξαγωγής χαρακτηριστικών είναι συγκριτικά καλύτερες για την αναγνώριση και ταυτοποίηση ανθρωπίνων κινήσεων, οι οποίες μπορεί να χαρακτηρίζονται από ταυτόχρονη κίνηση του κορμού και διαφορετικών άκρων; Πώς μπορούν αυτά τα χαρακτηριστικά να συνδυαστούν με τον κατάλληλο τρόπο ώστε να δημιουργηθούν μοντέλα για ακολουθίες κινήσεων; Αυτά τα μοντέλα έχουν την απαραίτητη διακριτική ικανότητα για να διαχωρίσουν πολύπλοκες ακολουθίες, όπως είναι αυτές που καταγράφονται κατά την εκτέλεση σύνθετων ακολουθιών πολεμικών τεχνών; Και αν μπορούν, είναι δυνατή η διαφοροποίηση αυτών των ακολουθιών ως προς την ποιότητα εκτέλεσής τους;

Τέτοια ερωτήματα υπήρξαν η έμπνευση για την παρούσα διπλωματική εργασία. Υπό την επίβλεψη του Καθηγητή Ε.Μ.Π. Πέτρου Μαραγκού, η κατεύθυνση της εργασίας ξεκαθαρίστηκε, καθώς και τα προβλήματα που θα προσεγγίζονταν, ώστε να παραχθούν ουσιαστικά αποτελέσματα. Συνεπώς, η ερευνητική αυτή προσπάθεια θα εστιάσει στην εξεύρεση μιας ολοκληρωμένης διαδικασίας για την εξαγωγή χαρακτηριστικών από βίντεο που περιέχουν ακολουθίες ανθρωπίνων κινήσεων, έπειτα στη δημιουργία μοντέλων για συγκεκριμένες κινήσεις και τέλος στη δοκιμή της αποτελεσματικότητας αυτών των μοντέλων ως προς την κατηγοριοποίηση αγνώστων βίντεο. Σαν παράπλευρη αποστολή στο τελευταίο στάδιο, θέτουμε την εξαγωγή ενός δείκτη ποιότητας εκτέλεσης της ακολουθίας κινήσεων στο εκάστοτε βίντεο, συγκρίνοντας πόσο ομοιάζουν οι κινήσεις σε σχέση με τα εξαχθέντα μοντέλα.

Σχετικά με τη μέθοδο εξαγωγής χαρακτηριστικών από τις ακολουθίες των καρτέ που απαρτίζουν τα βίντεο, ύστερα από εμπειριστατωμένη έρευνα και εκτεταμένες συζητήσεις, επιλέχθηκαν μέθοδοι που κάνουν χρήση των τοπικών χωρο-χρονικών χαρακτηριστικών από Ιστογράμματα (Histograms) στην περιοχή σημείων των καρτέ που παρουσιάζουν οπτικό ενδιαφέρον – Σημεία Ενδιαφέροντος (Interest Points). Τις μεθόδους αυτές από εδώ και για το υπόλοιπο της εργασίας θα τις αναφέρουμε με το όνομα *'Ιστογραφικές Μέθοδοι'*, χάριν συντομίας. Η επεξεργασία των χαρακτηριστικών αυτών γίνεται με τα καθιερωμένα, πλέον, για την αποτελεσματικότητά τους σε δυναμικά μοντέλα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models-HMMs). Τέλος, για τη δοκιμή της παραπάνω διαδικασίας πάνω σε πραγματικές ακολουθίες ανθρωπίνων κινήσεων επιλέχθηκε η εφαρμογή πάνω σε ακολουθίες κινήσεων από πολεμικές τέχνες και ειδικά από Shotokan Karate. Η επιλογή των παραπάνω θα εξηγηθεί λεπτομερώς στα επόμενα κεφάλαια.

Συνοψίζοντας, τα κύρια θέματα που προσεγγισθούν και τα οποία αποτελούν ουσιαστικές συνεισφορές αυτής της διπλωματικής εργασίας είναι τα εξής:

- Χρήση Ιστογραφικών μεθόδων για την εξαγωγή οπτικών χαρακτηριστικών από βίντεο με ακολουθίες ανθρωπίνων κινήσεων

- Επιλογή των πιο αντιπροσωπευτικών ως προς τις ανθρώπινες κινήσεις χαρακτηριστικών και επεξεργασία τους για τη δημιουργία μοντέλων
- Εκτίμηση παραμέτρων των HMM μοντέλων για την καλύτερη αναπαράσταση ακολουθιών ανθρωπίνων κινήσεων
- Ανάπτυξη ολοκληρωμένου συστήματος για ανίχνευση και κατηγοριοποίηση ανθρωπίνων κινήσεων με χρήση των παραπάνω μεθόδων
- Δημιουργία βάσης δεδομένων με κινήσεις Shotokan Karate, ώστε να μπορούν να χρησιμοποιηθούν για την εκπαίδευση και δοκιμή αποτελεσματικότητας του συστήματος
- Δημιουργία και χρήση συνοδευτικών προγραμμάτων για τη βέλτιστη λειτουργία ολόκληρης τη πειραματικής διαδικασίας

## 1.2 Δομή Διπλωματικής Εργασίας

Μέσα από τη δομημένη ροή αυτής της εργασίας, ο αναγνώστης καλείται να ακολουθήσει την πορεία της ερευνητικής διαδικασίας από τη σύλληψη της ιδέας για το αντικείμενο, τη συγκεκριμενοποίηση της κατεύθυνσης και των εργαλείων που θα χρησιμοποιηθούν, το σχεδιασμό της πειραματικής διαδικασίας, μέχρι την διεκπεραίωσή της, την εξαγωγή των χαρακτηριστικών από τα βίντεο, τη δημιουργία των μοντέλων και τέλος, την αξιολόγηση της συμπεριφοράς τους. Φυσικά, η διεργασία της αναζήτησης και των συζητήσεων που καθόρισαν τελικά τη μορφή αυτής της έρευνας δεν μπορούν να παρουσιαστούν εκτενώς σε αυτή την αναφορά, αλλά μια μικρή ιδέα παρουσιάζεται μέσα από αυτό το εισαγωγικό κομμάτι, ώστε να αναδειχτεί και αυτό το καθόλου ασήμαντο μέρος της διαδικασίας, που δίνει ώθηση στην εξέλιξη της επιστημονικής έρευνας.



Σχήμα 1.1: Δομή Διπλωματικής Εργασίας

Στο **Κεφ.2** παρουσιάζεται το επιστημονικό υπόβαθρο πάνω στο οποίο βασίζεται ολόκληρη η διπλωματική εργασία. Οι αναγνώστες εισάγονται στο πεδίο της Ανίχνευσης και Κατηγοριοποίησης Ανθρωπίνων Κινήσεων με την παρουσίαση του ερευνητικού έργου και των αποτελεσμάτων σημαντικών ερευνητών του πεδίου, οι οποίοι έχουν διαμορφώσει την κατάσταση στην οποία βρίσκεται σήμερα. Ιδιαίτερη μνεία γίνεται στις μεθόδους που κάνουν χρήση χαρακτηριστικών από Ιστογράμματα, όπως είναι τα Ιστογράμματα Προσανατολισμένης Κλίσης (Histogram of Oriented Gradient-HOG) και τα Ιστογράμματα Οπτικής Ροής (Histogram of Optical Flow-HOF), τα οποία θα χρησιμοποιηθούν για τις ανάγκες αυτής της εργασίας. Τέλος, παρουσιάζονται οι παράμετροι των Κρυφών Μαρκοβιανών Μοντέλων (Hidden Markov Models-HMMs), πάνω στα οποία θα βασιστεί η δημιουργία των μοντέλων για τις διαφορετικές κινήσεις karate που θα κληθούμε να αναγνωρίσουμε.

Οι λόγοι επιλογής των κινήσεων από Πολεμικές Τέχνες θα αναλυθούν στο **Κεφ.3**. Οι ιδιαιτερότητες των Πολεμικών Τεχνών, που τις καθιστούν άριστα σύνολα κινήσεων προς μελέτη, θα παρουσιαστούν περιληπτικά, ενώ εκτενέστερη αναφορά θα γίνει στο karate, την αυστηρότητα στην τεχνική που το χαρακτηρίζει και τη μεθοδικότητα στις ακολουθίες κινήσεων που περιλαμβάνει, γεγονότα που αποτέλεσαν τον κύριο λόγο ενασχόλησης και εφαρμογής του ερευνητικού περιεχομένου αυτής της εργασίας σε αυτόν τον τομέα. Παράλληλα, θα αναλυθούν κάποιες κινήσεις που χρησιμοποιούμε μετέπειτα στην πειραματική μας διαδικασία.

Στο **Κεφ. 4** θα περιγραφούν λεπτομερώς οι τεχνικές που μελετώνται στα πλαίσια αυτής της ερευνητικής προσπάθειας. Στην **παρ. 4.1** αναλύονται οι Ιστογραφικές μέθοδοι που χρησιμοποιούνται, με όλες τις παραμέτρους που τις συνοδεύουν για την εξαγωγή των οπτικών χαρακτηριστικών από τις ακολουθίες καρέ των βίντεο. Στη συνέχεια, στην **παρ. 4.2** θα παρουσιαστούν οι μέθοδοι περαιτέρω επεξεργασίας των δεδομένων μας για να είναι έτοιμα να τροφοδοτήσουν τα μοντέλα που θα δημιουργήσουμε. Στην **παρ. 4.3** θα παρουσιαστούν οι παράμετροι των HMMs που δημιουργούνται με βάση τα δεδομένα μας, ώστε να περιγράφονται καλύτερα οι ανθρώπινες κινήσεις που εξετάζουμε, ενώ τέλος, στην **παρ. 4.4** θα γίνει ειδική αναφορά στο εργαλείο HTK (Hidden Markov Model Toolkit) που χρησιμοποιήθηκε για την εκπαίδευση και δοκιμή των μοντέλων, μαζί με όλες τις ρυθμίσεις που απαιτήθηκαν για την σωστή του λειτουργία.

Το κύριο μέρος της πειραματικής διαδικασίας θα αναλυθεί στο **Κεφ. 5**, από τον τελικό σχεδιασμό της, μέχρι την πραγματοποίηση και αξιοποίηση των αποτελεσμάτων. Αρχικά, θα περιγραφούν οι συνθήκες βιντεοσκόπησης και οι λόγοι επιλογής των συμμετεχόντων σε αυτές, ώστε να δημιουργηθεί μια αντιπροσωπευτική βάση δεδομένων ικανή να χρησιμοποιηθεί για τις ανάγκες αυτής της εργασίας. Έπειτα, θα ακολουθήσουμε στάδιο προς στάδιο την εξαγωγή των χαρακτηριστικών, την επεξεργασία τους και τη δημιουργία των μοντέλων για τις επιλεγμένες ακολουθίες κινήσεων. Τέλος, θα παρουσιαστούν τα αποτελέσματα της δοκιμής των μοντέλων και η σύγκριση αποτελεσμάτων που εξήχθησαν με χρήση διαφορετικών παραμέτρων, αξιολογώντας ποιοι συνδυασμοί ήταν οι βέλτιστοι.

Συμπερασματικά, η διπλωματική εργασία θα κλείσει με το **Κεφ. 6**, στο οποίο θα γίνει αξιολόγηση των μεθόδων που αναλύθηκαν και του ερευνητικού αποτελέσματος, ενώ παράλληλα θα γίνει και σύγκριση με κάποιες παλαιότερες προσπάθειες. Σαν επιστέγασμα



αυτής της επιστημονικής έρευνας, θα παρουσιαστούν επιγραμματικά και κάποιες πιθανές ερευνητικές επεκτάσεις αυτού του έργου που μπορεί να δώσουν ώθηση σε μελλοντικές προσπάθειες.

### 1.3 Ευχαριστίες

Πριν να περάσω στο κύριο μέρος της εργασίας, νιώθω υποχρεωμένος να αποτεινώ τις ευχαριστίες μου σε όλα τα άτομα που με στήριξαν και με βοήθησαν καθόλη τη διάρκεια της απαιτητικής ερευνητικής αυτής προσπάθειας.

Μεγάλο μέρος της ύπαρξης αυτής της εργασίας οφείλεται στον επιβλέποντα μου, καθ. Πέτρο Μαραγκό. Είναι για αρχή, ο άνθρωπος που μέσω των μαθημάτων του με εισήγαγε στον πολύπλοκο θαυμαστό κόσμο της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων που ευθυγραμμίστηκαν με τις ερευνητικές μου επιδιώξεις και έδωσαν ώθηση στη θέλησή μου να ασχοληθώ με την έρευνα. Χωρίς την καθοδήγηση και την υπομονή του να κατευνάσει το ανήσυχο πνεύμα μου, μέσα από ώρες συζητήσεων σχετικά με την κατεύθυνση της έρευνας, ίσως να μην είχα καταλήξει ακόμα στο θέμα της Διπλωματικής Εργασίας. Η κατανόηση που έδειξε στην τάση μου να θέτω μη ρεαλιστικούς ερευνητικούς στόχους κατά καιρούς, αλλά και στην παράταση της ολοκλήρωσης της εργασίας για διάφορους προσωπικούς μου λόγους, μου έδειξε τον πραγματικό ρόλο του μέντορα στην διεκπεραίωση μιας ερευνητικής προσπάθειας. Μα πέρα από τα παραπάνω, τον ευχαριστώ γιατί με βοήθησε να καλλιεργήσω την αρετή της μεθοδικής επιστημονικής αναζήτησης, να μάθω πως να σκέφτομαι και να ενεργώ στοχευμένα για την παραγωγή νέας γνώσης και όχι απλά αναμάσηση κοινοτοπιών. Μέσα από τη συνεργασία μας, μου έδειξε πόσο ακέραιος ερευνητής είναι, με γνώσεις πάνω σε πολλαπλά αντικείμενα μα και δίψα για περαιτέρω επέκταση και εμβάθυνση αυτής της γνώσης, αλλά και μέσω του ενδιαφέροντος προς τους μαθητές του, πόσο άξιος καθηγητής και άνθρωπος.

Βεβαίως, μεγάλο μέρος των ευχαριστιών μου πρέπει να αποδοθεί σε όλα τα άτομα που βοήθησαν την πρώτη μεγάλης κλίμακας ερευνητική μου προσπάθεια. Στους πρωτότερους ερευνητές του επιστημονικού αυτού πεδίου, που μέσω των έργων τους άνοιξαν τον δρόμο, μου προσέφεραν ιδέες ή ακόμα και αλγοριθμική βοήθεια, μέσω ελεύθερων εργαλείων και εφαρμογών που έχουν αναρτήσει στο Διαδίκτυο, για την ολοκλήρωση της πολυτμηματικής αυτής εργασίας. Ιδιαίτερη θέση σε αυτό το σύνολο κατέχει ο Prof. Ivan Laptev, ο οποίος με το έργο του αποτέλεσε μεγάλη έμπνευση για τη μορφή του πειραματικού τμήματος της έρευνας, αλλά και μέσω προσωπικής συνομιλίας μαζί μου και μέσω προσφοράς τμήματος κώδικα αλλά και αποτελεσμάτων της δικής του έρευνας, διευκόλυνε κατά πολύ τη δική μου. Επίσης, δεν πρέπει να παραλείψω την ομάδα διδακτορικών βοηθών του καθ. Μαραγκού και ιδιαίτερα τους Σταύρο Θεοδωράκη και Βασίλη Πιτσικάλη, που μου προσέφεραν τη συμβουλή και τη βοήθειά τους σε αρκετά στάδια της πολυδαίδαλης αυτής διαδρομής.

Οφείλω επίσης να πω ότι είμαι υπόχρεος στον karate Sensei κο. Ιωάννη Στυλιανόπουλο και τη σχολή shotokan karate του 'Yubukai Dojo'-'Η Καρδιά του Πολεμιστή' μαζί με όλους τους αθλητές της που με μύησαν στην τέχνη του shotokan karate και με βοήθη-

### 1.3. ΕΥΧΑΡΙΣΤΙΕΣ

---

σαν να συγκεντρώσω το απαραίτητο υλικό για την πειραματική διαδικασία της εργασίας αυτής. Ένα μεγάλο ευχαριστώ και στους φίλους μου Ειρήνη και Παναγιώτη, καθώς και στην αδερφή μου, Βενετία, που με βοήθησαν κατά τις βιντεοσκοπήσεις.

Τέλος, πέρα από τα άτομα που με βοήθησαν στο επιστημονικό κομμάτι αυτής της έρευνας, θα ήθελα να ευχαριστήσω τους γονείς μου και την αδερφή μου, που ήταν πάντα δίπλα μου και στήριζαν τις αποφάσεις μου, όσο τρελές κι αν ήταν κατά καιρούς, ενώ με παρότρυναν να θέτω τους στόχους μου όσο πιο ψηλά μπορούσε η φαντασία μου να με οδηγήσει. Ακόμα, τους φίλους μου που όλο αυτό τον καιρό μου προσέφεραν ένα δίκτυο υποστήριξης και κατανόησης, έτοιμο να με βοηθήσει σε ώρες ανάγκης και αμφισβήτησης. Ιδιαίτερη θέση σε αυτό τον κύκλο έχει ο φίλος μου Παναγιώτης Συμβουλίδης, που πέραν των άλλων υπήρξε μια μόνιμη τεχνική υποστήριξη και πηγή θετικής ενέργειας όποτε τον χρειάστηκα. Τέλος, ένα ευχαριστώ στους φίλους μου από την οικογένεια της EESTEC (Electrical Engineering Student's European Association) και τους υπόλοιπους φίλους μου από το εξωτερικό, που βοήθησαν στην διεύρυνση των στόχων και οριζόντων μου.

*«Ένας άνδρας είναι τόσο ισχυρός όσο οι άνθρωποι που τον στηρίζουν»*

## Κεφάλαιο 2

# Εισαγωγή στην Ανίχνευση και Κατηγοριοποίηση Ανθρώπινων Κινήσεων

Πολλά έχουν αλλάξει τις δύο τελευταίες δεκαετίες στον τομέα της Ορασης Υπολογιστών που εξετάζει την Ανίχνευση και Κατηγοριοποίηση Ανθρώπινων Κινήσεων. Ενώ οι αρχικές προσπάθειες είχαν ως στόχο την απλή αναγνώριση του ανθρώπινου περιγράμματος ή της ανθρώπινης φιγούρας, με την εξέλιξη των ερευνητικών εργαλείων (τεχνικές εντοπισμού, εξαγωγής οπτικών χαρακτηριστικών, δημιουργίας μοντέλων) αλλά και των απαιτήσεων αναγνώρισης για χρήση σε πιο πολύπλοκες εφαρμογές, οι τελευταίες προσπάθειες εστιάζουν στη βελτιστοποίηση της διακριτικής ικανότητας των υπολογιστικών συστημάτων, ώστε να αναγνωρίζονται με υψηλά ποσοστά αξιοπιστίας ακόμα και υποκατηγορίες συγκεκριμένων ανθρώπινων δράσεων. Σε αυτό το τμήμα, θα προσπαθήσουμε να παρουσιάσουμε διάφορες καθιερωμένες τεχνικές που χρησιμοποιούνται στην Ανίχνευση και Κατηγοριοποίηση, καθώς και το πώς προέκυψαν, και σταδιακά θα φτάσουμε σε πιο σύγχρονες μεθόδους, μορφές των οποίων θα χρησιμοποιηθούν στο πειραματικό κομμάτι της εργασίας.

### 2.1 Ερευνητικό Υπόβαθρο

Προσεγγίζοντας ερευνητικές προσπάθειες του παρελθόντος πάνω στο συγκεκριμένο αντικείμενο, θα επιχειρήσουμε και να αναλύσουμε τα βασικά στάδια κατά τη διαδικασία της Ανίχνευσης και Κατηγοριοποίησης Ανθρώπινων Κινήσεων, ώστε ο αναγνώστης να ακολουθήσει τις διαφορετικές μεθόδους που έχουν αναπτυχθεί για αυτά και την εξέλιξή τους στο χρόνο και να διευκολυνθεί η κατανόησή τους.

### 2.1.1 Εντοπισμός

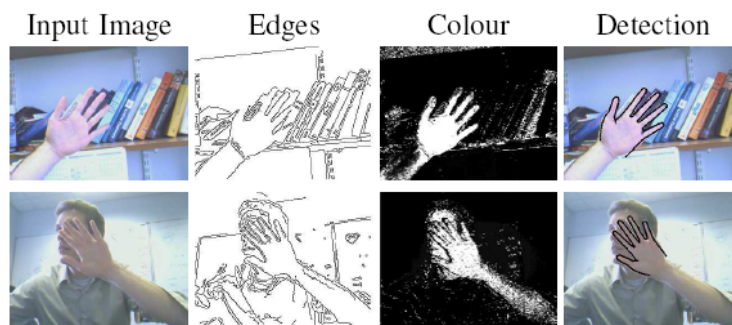
Το πρωταρχικό στάδιο για οποιαδήποτε διαδικασία ανίχνευσης από ένα βίντεο, δηλαδή μία ακολουθία από καρέ εικόνων, δεν είναι άλλο από τον εντοπισμό του αντικειμένου ενδιαφέροντος. Δεδομένου πως για την αναγνώριση ανθρωπίνων κινήσεων είναι απαραίτητος ο καθορισμός της ανθρωπίνης φιγούρας σε κάθε καρέ, συμπεριλαμβανομένων των άκρων και του κορμού, πολλές τεχνικές αναπτύχθηκαν για αυτό το σκοπό. Δεν είχαν όλες την ίδια απαρχή, αλλά συνέβαλαν εξίσου στην καλύτερη ανεύρεση της ανθρωπίνης φιγούρας και των κινήσεων της.

Σαν μια πιο μερική λύση του προβλήματος, μερικοί ερευνητές που επικεντρώθηκαν στην ανίχνευση χειρονομιών, που περιλαμβάνει τον εντοπισμό των δύο χεριών και του κεφαλιού, έκαναν χρήση της τεχνικής των *blob* (άμορφη μάζα). Μπορεί η θεωρία των *blobs* να εισήχθη αρκετά νωρίς από τους Kauth και Pentland το 1977 [18], αλλά η συστηματική χρήση τους για την ανίχνευση χειρονομιών ήλθε πολύ αργότερα. Τα *blobs*, όπως υποδεικνύει και η απόδοσή τους στην ελληνική, είναι χαλαρά καθορισμένα συμπλέγματα που χρησιμοποιούνται για να υποδείξουν τις περιοχές που μας ενδιαφέρουν. Οι περιοχές αυτές μπορούν να ευρεθούν κάνοντας χρήση απλούστερων τεχνικών αναγνώρισης ακμών, ταυτοποίησης χρώματος με το χρώμα του δέρματος, αντιδιαστολή προσκηνίου-παρασκηνίου ή ακόμα και πιο πολύπλοκων τεχνικών [19, 20, 1] (βλ. εικόνα 2.1(α')). Μάλιστα, με την εφαρμογή ορισμένων τεχνικών, όπως χρήση Bayesian φίλτρων, οι περιοχές αυτές γίνονταν πολύ συγκεκριμένες και αγκάλιαζαν στην ουσία το περίγραμμα των χεριών [2] (βλ. εικόνα 2.1(β')). Άλλες προσπάθειες ανίχνευσης χειρονομιών, στην προσπάθειά τους να εξελίσσουν τον εντοπισμό των χεριών ώστε να μπορεί να ανταπεξέλθει στην παρακολούθηση των χειρονομιών σε πραγματικό χρόνο (*real-time gesture following*) εισήγαγαν αρκετά στατιστικά στοιχεία [21, 22]. Στην πιο διαδεδομένη μορφή τους, όμως, τα *blobs* εξελίχθηκαν από δισδιάστατα (2D) σε τρισδιάστατα (3D), ώστε να δίνουν μια πιο στερεοσκοπική απεικόνιση του χώρου και να μπορέσουν πιο αποδοτικά να ακολουθήσουν τις απαιτούμενες περιοχές με την πάροδο του χρόνου (βλ. εικόνα 2.1(γ')). Με αυτή τη νέα τους μορφή χρησιμοποιήθηκαν κατά κόρον και σε εφαρμογές πραγματικού χρόνου [3, 23].

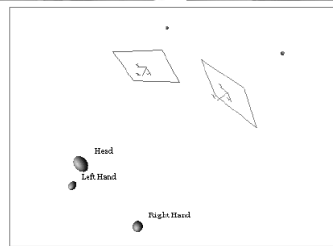
Προχωρώντας προς τον εντοπισμό κινήσεων ολόκληρου του ανθρωπίνου σώματος, μία από τις πιο γνωστές προσεγγίσεις αποτελεί αυτή των Bobick και Davis [24, 4]. Σύμφωνα με αυτούς, η ανίχνευση των ανθρωπίνων κινήσεων μπορεί να γίνει μέσω ενός χρονικού προτύπου που ξεετάζει ένα προς ένα τα *pixel* των εικόνων κατά την εξέλιξη της κίνησης και περιλαμβάνει δύο διαφορετικές μορφές, μια *εικόνα κίνησης-ενέργειας* (*motion-energy image-MEI*) και μια *εικόνα κίνησης-ιστορίας* (*motion-history image-MHI*). Αυτή η ανάλυση, σε αντιδιαστολή με την προηγούμενη, δεν αντιμετωπίζει μέρη του σώματος σαν μεμονωμένες περιοχές-*blobs*, αλλά εστιάζει στις χρονικές μεταβολές των αντικειμένων του χώρου κατά την εξέλιξη του βίντεο (βλ. εικόνα 2.2(α')). Σαν αποτέλεσμα όμως, η ανίχνευση δεν ήταν τόσο λεπτομερής ώστε να ξεπερνάει τη διάκριση απλά του περιγράμματος του ανθρωπίνου σώματος. Μια προσπάθεια που κινείται επίσης σε αυτά τα πλαίσια, εξελίσσοντας αυτό το μοντέλο για πιο τρισδιάστατες απεικονίσεις είναι αυτή των Gorelick et al. [8], στην οποία οι κινήσεις αντιμετωπίζονται σαν χωροχρονικά σχήματα και ανευρίσκονται μέσω τεχνικών εξαγωγής χωροχρονικών 'προεξοχών' από τις εικόνες,



(α')



(β')



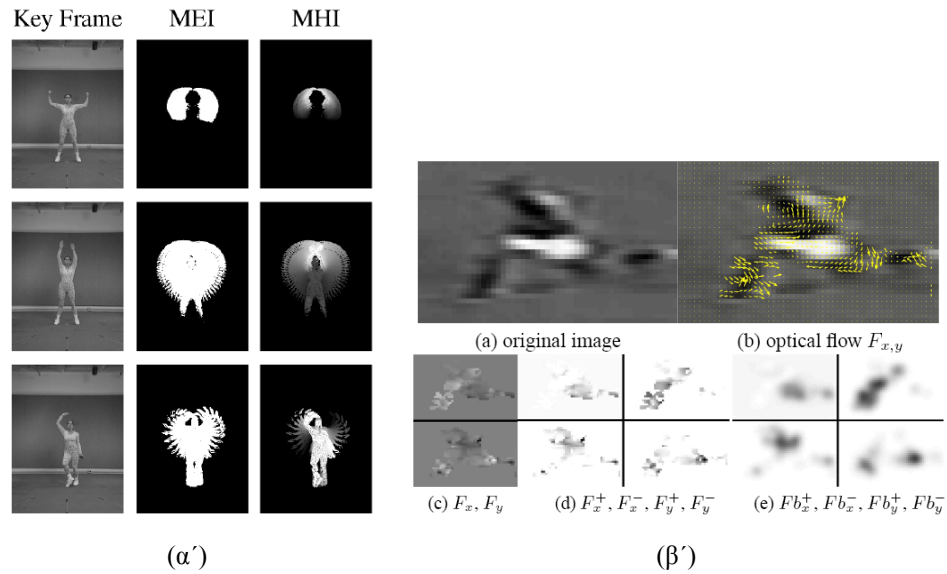
3-D estimate

(γ')

Σχήμα 2.1: Τεχνικές εντοπισμού άκρων από ερευνητικές προσπάθειες ανίχνευσης χειρονομιών. (α') Blobs χεριών, όπως εμφανίζεται στο σχήμα του [1]. (β') Ακριβής εντοπισμός χεριών, όπως εμφανίζεται στο σχήμα του [2]. (γ') Τρισδιάστατα (3D) blobs, όπως εμφανίζεται στο σχήμα του [3].

με αποτέλεσμα να ανιχνεύονται τα κινούμενα μέρη του κινούμενου ανθρώπινου σώματος. Μια ακόμα γνωστή προσέγγιση είναι αυτή των Efron et al. [5], στην οποία αντί να

## 2.1. ΕΡΕΥΝΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

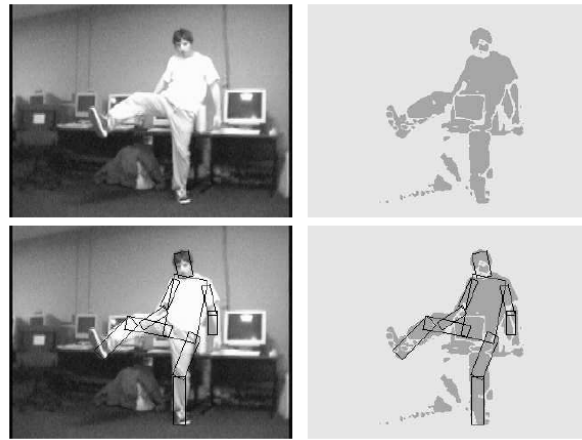


Σχήμα 2.2: Εντοπισμός ολόκληρου ανθρώπινου σώματος. (α') Εντοπισμός ανθρώπινης φιγούρας με τη χρήση MEI και MHI, όπως εμφανίζεται στο σχήμα του [4]. (β') Εξαγωγή πληροφοριών οπτικής ροής, όπως εμφανίζεται στο σχήμα του [5].

δουλεύουμε με πληροφορίες για το ανθρώπινο περίγραμμα, ο εντοπισμός των μελών του σώματος που συμμετέχουν στην κίνηση γίνεται μέσω μιας απλούστερης μεθόδου που κάνει χρήση της οπτικής ροής μεταξύ των διαδοχικών καρέ των βίντεο. Μάλιστα, η διάκριση γίνεται μεταξύ καναλιών τεσσάρων διαφορετικών κατευθύνσεων, τα οποία διαχωρίζονται για περαιτέρω επεξεργασία(βλ. και εικόνα 2.2(β')). Αυτή η τεχνική αποδείχθηκε ιδιαίτερα χρήσιμη για ανίχνευση ανθρώπινων δράσεων μεσαίας κλίμακας μεγέθους, όχι σε πραγματικό χρόνο (offline), χωρίς όμως να μπορεί να εστιάσει με μεγάλη λεπτομέρεια.

Άλλες προσπάθειες που έχουν τις ρίζες τους στην πιο γενική ανίχνευση αντικειμένων σε εικόνες, κάνουν χρήση ανίχνευσης ακμών για τη δημιουργία καμπυλών ή πολυγώνων που προσαρμόζονται γύρω από τα μέρη του ανθρώπινου σώματος. [25, 26, 27, 6]. Κυριώς στο έργο του Felzenszwalb πάνω στις *εικονογραφικές δομές (pictorial structures)*, αυτές οι τεχνικές μπόρεσαν να εφαρμοστούν και για τον εντοπισμό του ανθρώπινου σώματος κατά τη διάρκεια κινήσεων (βλ. και εικόνα 2.3(α')). Μερικές ερευνητικές προσεγγίσεις που στηρίζονται σε ένα προκαθορισμένο μοντέλο για το ανθρώπινο σώμα, μέσω ενός σκελετού συνήθως με συγκεκριμένους βαθμούς ελευθερίας (degrees of freedom-DOF), προσπαθούν κατά κύριο λόγο να εντοπίσουν μια δομή στην εικόνα που ομοιάζει με το σκελετικό τους πρότυπο και στη συνέχεια να την ταυτοποιήσουν με ένα πρότυπο πόζας που έχουν αποθηκευμένο στη βάση δεδομένων τους [28, 29, 7, 30, 31]. Αυτό βέβαια, ενέχει και κινδύνους μη αναγνώρισης του είδους της κίνησης ή αδυναμίας επαναπροσδιορισμού της ανθρώπινης φιγούρας σε περίπτωση που στιγμιαία αυτή χαθεί.

Πέρα από τις παραπάνω τεχνικές για τον εντοπισμό της ανθρώπινης μορφής, μια πιο γενική τακτική για την Όραση Υπολογιστών χρησιμοποιήθηκε για το συγκεκριμένο λόγο



(α')



(β')

Σχήμα 2.3: Εντοπισμός σημείων με βάση μοντέλα για αναπαράσταση ανθρώπινου σώματος. (α') Αναδημιουργία ανθρώπινης φιγούρας με Pictorial Structures, όπως εμφανίζεται στο σχήμα του [6]. (β') Προσαρμογή σκελετικού μοντέλου για εντοπισμό ανθρώπινης κίνησης, όπως εμφανίζεται στο σχήμα του [7].

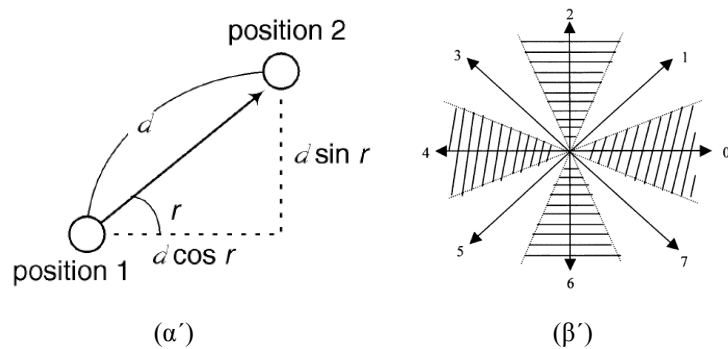
με αρκετά μεγάλη επιτυχία. Αυτή δεν είναι άλλη από τον εντοπισμό σημείων ενδιαφέροντος (*interest points*) μέσα στα καρέ τοπικά, αλλά και με τη μεταβολή στοιχείων χρονικά, την οποία θα αναλύσουμε πιο διεξοδικά παρακάτω μαζί με τους λόγους για τους οποίους επιλέγουμε να την χρησιμοποιήσουμε σε αυτή την εργασία.

### 2.1.2 Εξαγωγή Χαρακτηριστικών

Ύστερα από την ανίχνευση του ανθρωπίνου σώματος μέσα στα καρέ των βίντεο ή της περιοχής που μας ενδιαφέρει εν τέλει να εξετάσουμε, απαραίτητη είναι η εξαγωγή χαρακτηριστικών από πιθανόν πολλαπλά σημεία, ώστε η ανθρώπινη κίνηση να περιγράφεται πιο αντιπροσωπευτικά και να τροφοδοτηθούν τα μοντέλα που θα εξετάσουμε παρακάτω είτε για εκπαίδευση ή για δοκιμή. Και σε αυτό το στάδιο είχαμε την ανάπτυξη διαφορετικών προσεγγίσεων σχετικά με το ποια χαρακτηριστικά αναπαριστούν καλύτερα την κίνηση και παρέχουν την καλύτερη διακριτική ικανότητα μεταξύ διαφορετικών ειδών κινήσεων.

Για τις ερευνητικές προσπάθειες που εστίαζαν στον εντοπισμό άκρων του σώματος και την παρακολούθησή τους στο χώρο καθώς αυτά εκτελούσαν την κίνηση, όπως στην περίπτωση των χειρονομιών που είδαμε παραπάνω, η χρήση χαρακτηριστικών που σχετίζονταν με την τοπική μετακίνηση των παρακολουθούμενων άκρων κατά την πάροδο του χρόνου ήταν η προφανής λύση. Έτσι, το διάνυσμα χαρακτηριστικών (*feature vector*), που χρειαζόταν για τα επόμενα στάδια, είναι ένας συνδυασμός της θέσης (*location*), δηλαδή του πώς μεταβάλλεται η θέση των σημείων αυτών με την πάροδο του χρόνου, του προσανατολισμού (*orientation*), δηλαδή της κατεύθυνσης που έχει η κίνηση των σημείων, της ταχύτητας (*velocity*) των σημείων και άλλων [20, 32, 19] (βλ. και εικόνα 2.4). Σε κάθε περίπτωση πάντως, το ζητούμενο είναι η επιλογή χαρακτηριστικών αμετάβλητων ως προς μετατοπίσεις και περιστροφές, ώστε να αναπαρίστανται καλύτερα και οι πιο πολύπλοκες κινήσεις. Έτσι, είχαμε και τη χρησιμοποίηση πιο σύνθετων τοπικών χαρακτηριστικών, όπως λογαρίθμους πολικών συντεταγμένων, καμπυλότητα και άλλων[23]. Σε μερικές περιπτώσεις, είχαμε και την χρήση απλών στατιστικών στοιχείων που εξάγονταν από τα τοπικά αυτά χαρακτηριστικά, όπως μέσες τιμές θέσεων και ταχυτήτων, τυπικές αποκλίσεις και άλλων[21].

Παρόμοια είναι ως ένα βαθμό και η μέθοδος που ακολουθούν οι ερευνητικές προσπά-



Σχήμα 2.4: Απλά τοπικά χαρακτηριστικά. (α') Χαρακτηριστικό θέσης. (β') Χαρακτηριστικό προσανατολισμού.

θειες, οι οποίες χρησιμοποιούν αρθρωτά μοντέλα για την αναπαράσταση του ανθρωπίνου σώματος. Αφού η ανθρώπινη μορφή αναγνωριστεί παρακολουθούνται όλα τα μέλη του



μοντέλου και ανάλογα χαρακτηριστικά θέσης, ταχύτητας και προσανατολισμού εξάγονται για να γίνει σύγκριση με το μοντέλο της κίνησης που έχει αποθηκευτεί στη βάση δεδομένων [7, 30]. Για προσεγγίσεις που επικεντρώνονται στη χρονική αναπαράσταση του περιγράμματος του ανθρωπίνου σώματος, όπως αυτή των Bobick και Davis [4], τα χαρακτηριστικά που εξάγονται αφορούν κυρίως τις μεταβολές μεταξύ των διαδοχικών αναπαράστασεων αλλά πάλι σχετίζονται με τη χωροχρονικές αλλαγές. Πιο τρισδιάστατες μέθοδοι χρησιμοποιούν χαρακτηριστικά που μπορεί να προέρχονται από την χωροχρονική αλλαγή *σιλουετών* (*silhouettes*) που απαρτίζονται από πολλές μικρές *μετα-μπάλες* (*metaballs*) [29] ή από την χωροχρονική μετατόπιση 'προεξέχόντων' σημείων (*salient points*), που έχουν αναγνωριστεί από τη λύση της εξίσωσης Poisson [8] (βλ. και εικόνα 2.5(α')). Άλλες πιο κλασικές προσεγγίσεις που εντοπίζουν τα αντικείμενα ενδιαφέροντός τους μέσω οπτικής ροής, όπως αυτή του Efros [5], κάνουν χρήση των διαφορετικών κατευθύνσεων της οπτικής ροής για να σχηματίσουν το διάνυσμα χαρακτηριστικών τους. Πέρα από τα σαφώς καθορισμένα χαρακτηριστικά, υπάρχουν και ερευνητές που έχουν καταστήσει ορισμένα μικρά τμήματα κινήσεων ως τα βασικά δομικά στοιχεία για μεγαλύτερες κινήσεις [16]. Αφού γίνει ο εντοπισμός αυτών των *αρχετύπων κίνησης* (*motion primitives*), η εναλλαγή τους παράγει στην ουσία τα χαρακτηριστικά που θα χρησιμοποιηθούν στην εκπαίδευση και αναγνώριση.

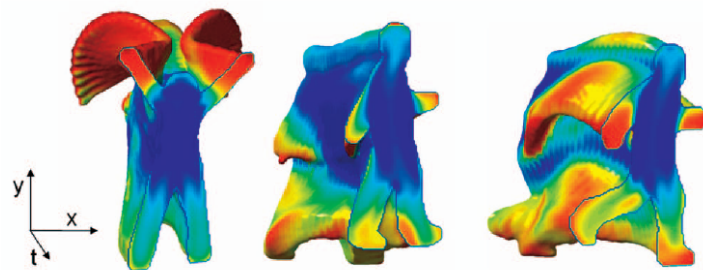
Για τις ερευνητικές προσπάθειες που διεκπεραιώνουν τον εντοπισμό μέσω της διαδομένης τεχνικής των σημείων ενδιαφέροντος (*interest points*), έχουμε εξαγωγή χαρακτηριστικών από τις συγκεκριμένες περιοχές των *interest points*, αλλά και πάλι τα χαρακτηριστικά που επιλέγονται διαφέρουν ανάλογα με την προσέγγιση. Ενώ χρησιμοποιούνται και πιο απλά τοπικά χαρακτηριστικά που αναφέρθηκαν παραπάνω, σε ορισμένες προσπάθειες το διάνυσμα χαρακτηριστικών απαρτίζεται από τοπικές μεταβολές φωτεινότητας (*brightness*), κλίσης (*gradient*) και οπτικής ροής (*optical flow*) μέσα σε κυβοειδείς περιοχές όπως στο άρθρο του Dollar (*cuboids*) [9] (βλ. και εικόνα 2.5(β')). Σε άλλες προσπάθειες, όπως του Niebles [33], τα χαρακτηριστικά έχουν τη μορφή χωροχρονικών λέξεων (*spatial-temporal words*), οι οποίες είναι ανεξάρτητες από μεταβολές κλίμακας και μετακινήσεων της κάμερας λήψης και συνιστούν λεξιλόγιο (*codebook*) ικανό να περιγράψει την εκάστοτε κίνηση.

Τα τελευταία χρόνια, παρά την εμβάθυνση για αρκετά μεγάλο χρονικό διάστημα της επιστημονικής κοινότητας στη δημιουργία λεπτομερών μοντέλων για την αναπαράσταση του ανθρωπίνου σώματος, το ενδιαφέρον στράφηκε ξανά σε απλούστερες μεθόδους που κάνουν χρήση στατιστικών στοιχείων από τις περιοχές γύρω από σημεία ενδιαφέροντος. Τέτοιες είναι οι Ιστογραφικές μέθοδοι, που θα αναλύσουμε παρακάτω καθώς θα χρησιμοποιηθούν για το πειραματικό μας μέρος.

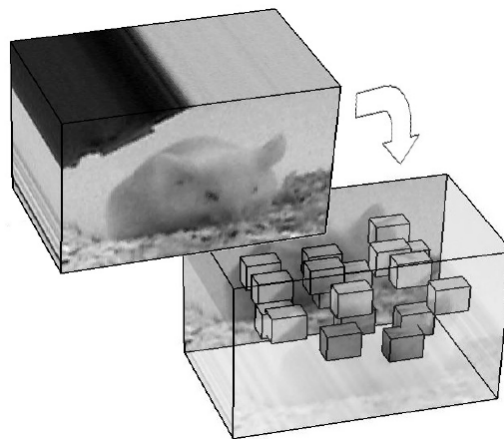
### 2.1.3 Δημιουργία Μοντέλων

Τελευταίο στάδιο για τη δημιουργία ενός συστήματος Ανίχνευσης και Κατηγοριοποίησης είναι γενικά η δημιουργία μοντέλων για τα διαφορετικά είδη που θέλουμε το σύστημά μας να μπορεί να διαχωρίζει. Στην περίπτωση των ανθρωπίνων κινήσεων, απαι-

## 2.1. ΕΡΕΥΝΗΤΙΚΟ ΥΠΟΒΑΘΡΟ



(α')



(β')

Σχήμα 2.5: Διάφορες μέθοδοι εξαγωγής χαρακτηριστικών. (α') Εξαγωγή χαρακτηριστικών από 'προεξέχοντα' (*salient*) σημεία, όπως εμφανίζεται στο σχήμα του [8]. (β') Χρήση κυβοειδών περιοχών(*cuboids*) για τοπική εξαγωγή χαρακτηριστικών, όπως εμφανίζεται στο σχήμα του [9].

τείται ένα μοντέλο για κάθε ξεχωριστή κίνηση που απαιτούμε να αναγνωρίζονται. Τα μοντέλα αυτά πρέπει να εκπαιδευτούν με τα χαρακτηριστικά που έχουμε εξάγει στο προηγούμενο στάδιο μέχρι οι παράμετροί τους να αναπαριστούν ικανοποιητικά την εκάστοτε κίνηση. Στη συνέχεια, τα ίδια μοντέλα κατά τη δοκιμή του συστήματός μας πρέπει να είναι σε θέση να αναγνωρίζουν τις διαφορετικές κινήσεις από χαρακτηριστικά αγνώστων ακολουθιών βίντεο.

Σε πιο χαμηλό επίπεδο, προσπάθειες όπως του Felzenswalb [6] που είδαμε παραπάνω, είχαν ως στόχο τη χρήση εικονογραφικών δομών για τη δημιουργία απλών αρθρωτών μοντέλων για το ανθρώπινο σώμα αποτελούμενο από ορθογωνικά μέλη, ενώ παρόμοια μοντέλα παίρνουμε και από κάποιες προσπάθειες που κάνουν χρήση περιορισμών φωτεινότητας και σταθερότητας βάθους (*depth constancy*) στις εικόνες[28]. Πιο εξελιγμένες προσπάθειες για ενιαία ανθρώπινα μοντέλα είχαν ως αποτέλεσμα τη δημιουργία σιλουετών αποτελούμενες από ελλειψοειδείς μετά-μπάλλες (*metaballs*) για την περιγραφή της

φιγούρας του ανθρώπου κατά τη διάρκεια της κίνησης [29] ή την κατασκευή σκελετικού προτύπου για τον άνθρωπο με συγκεκριμένη δομή και βαθμούς ελευθερίας(DOF) μεταξύ των αρθρώσεων του [30]. Επίσης, μοντέλα με ακόμα μεγαλύτερη έμφαση στην τρισδιάστατη απεικόνιση έκαναν την εμφάνισή τους, με ολόκληρο τον σκελετό να χαρακτηρίζεται από προσανατολισμό στο χώρο και δυνατότητα διαφορετικού προσανατολισμού του κορμού[7]. Κοινή συνιστώσα των παραπάνω είναι η εκπαίδευση των αρθρωτών μοντέλων στις συγκεκριμένες κινήσεις και η αποθήκευση των διαδοχικών ποζών σε βάσεις δεδομένων, ενώ η ανίχνευση επιχειρείται με την ταύτιση των διαδοχικών ποζών της άγνωστης κίνησης με τις αποθηκευμένες πόζες.

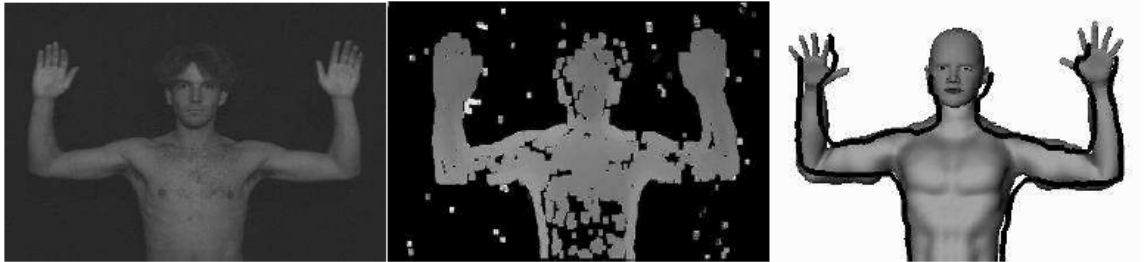
Αλλά μοντέλα ξεφεύγουν από τη λογική των σκελετών για το ανθρώπινο σώμα και επικεντρώνονται στις στη δημιουργία δομών για τον κατάλληλο συνδυασμό χαρακτηριστικών από διάφορες περιοχές ενδιαφέροντος πάνω στα καρτέ των εικόνων. Μια τέτοια προσπάθεια είναι των Niebles και Fei-Fei [34] με την εισαγωγή ενός *Ιεραρχικού μοντέλου (Hierarchical model)*. Σύμφωνα με αυτή την προσέγγιση, έχουμε διαφορετικά επίπεδα χαρακτηριστικών. Το κατώτερο επίπεδο περιέχει χαρακτηριστικά από τα σημεία ενδιαφέροντος της εικόνας και συνοδεύεται από ένα μοντέλο για το σύνολο αυτών των χαρακτηριστικών σαν μια συλλογή των πιο σημαντικών (*bag of keypoints*). Το ανώτερο επίπεδο περιλαμβάνει στοιχεία που σχετίζονται το καθένα με μια ομάδα χαρακτηριστικών του κατώτερου επιπέδου και συνοδεύεται από ένα μοντέλο αλληλοσύνδεσης των στοιχείων αυτών συσχετίζοντας τις μεταξύ τους αποστάσεις (βλ. και εικόνα 2.7(α')).

Μια πολύ μεγάλη και διαδεδομένη κατηγορία μοντέλων που χρησιμοποιούνται τα τελευταία χρόνια στην Ανίχνευση και Κατηγοριοποίηση είναι τα *SVM (Support Vector Machines)*. Αυτή η κατηγορία μοντέλων αντλεί τη δύναμή της από την απλότητα στην κατηγοριοποίηση και τη δυνατότητα να μειώνει τη διάσταση του προβλήματος όσο μεγάλη και να είναι η διάσταση του διανύσματος χαρακτηριστικών, καθώς η κατηγοριοποίηση γίνεται με καθορισμένο τρόπο στον οποίο συμμετέχουν μόνο οι οριακές παρατηρήσεις κάθε κατηγορίας. Τα SVM έχουν γίνει αρκετά δημοφιλή σε εφαρμογές που εστιάζουν στην επεξεργασία τοπικών στοιχείων και ειδικά για χαρακτηριστικά που προέρχονται από στατιστικές μεθόδους, όπως οι Ιστογραφικές, για τις οποίες έχουν χρησιμοποιηθεί σε διάφορες μορφές (γραμμικά, μη γραμμικά, πολυωνυμικά, λανθάνοντα κ.ά.) [35, 12, 14, 36, 37, 38, 39].

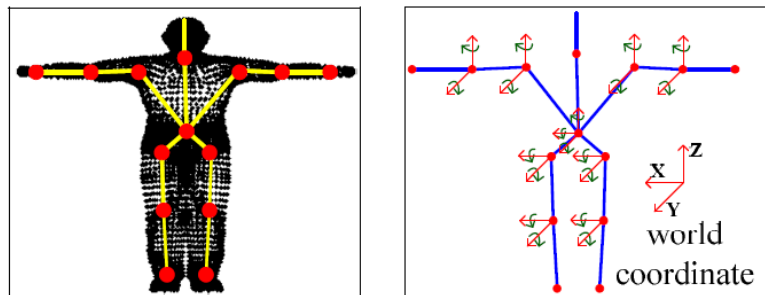
Επίσης, πιο απλοϊκές μέθοδοι δεν παύουν να χρησιμοποιούνται, όπως τεχνικές που βασίζονται στη λογική των Κοντινότερων Γειτόνων (*Nearest Neighbor-NN*), όπου έχουμε την εκπαίδευση μοντέλων να αποφαινόμενοι για την κατηγοριοποίηση ενός μη κατηγοριοποιημένου δείγματος ανάλογα με το πλησιέστερο 'μετρικά' πλήθος ήδη κατηγοριοποιημένων δειγμάτων[40]. Βέβαια, ο συνδυασμός τέτοιων πιο απλών δεικτών κατηγοριοποίησης (classifiers) έχει αποδειχθεί πολύ αποδοτικός για την κατηγοριοποίηση πιο δύσκολων προβλημάτων και παρά το υπολογιστικό 'βάρος' που τον συνοδεύει χρησιμοποιείται επίσης σε μεγάλο βαθμό. Ένα τέτοιο συνδυαστικό σύστημα είναι και το Adaboost, το οποίο μπορούμε να βρούμε σε αρκετές σύγχρονες εφαρμογές[41].

Φυσικά, η πιο διαδεδομένη κατηγορία μοντέλων για αναπαράσταση ανθρωπίνων κινήσεων και γενικότερα σκηνών που περιέχουν κίνηση είναι τα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models-HMMs). Οι κύριοι λόγοι που τα καθιστούν ιδιαίτερος

## 2.1. ΕΡΕΥΝΗΤΙΚΟ ΥΠΟΒΑΘΡΟ



(α') Μοντέλο σιλουέτας από *metaballs*, όπως εμφανίζεται στο σχήμα του [29]



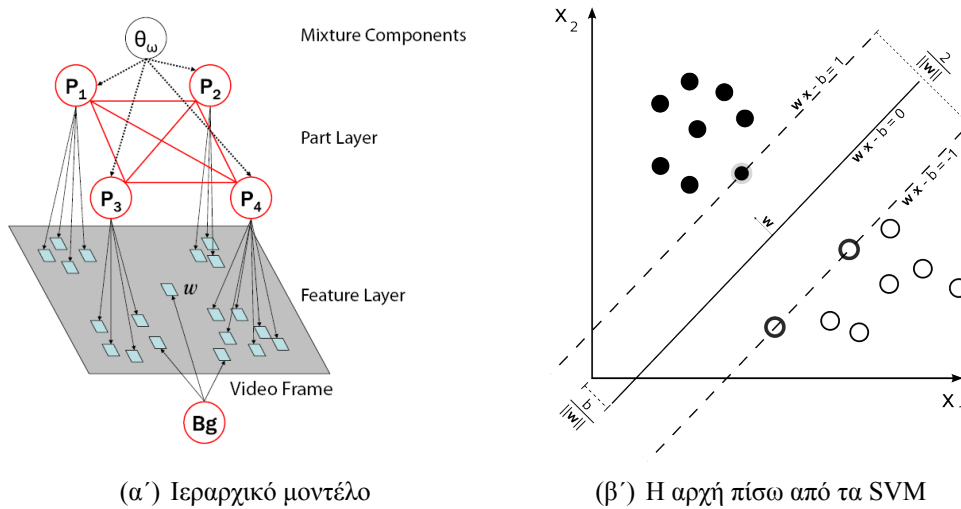
(β') DOF σκελετού, όπως εμφανίζεται στο σχήμα του [30]



(γ') Σύνθετος σκελετός ανθρωπίνου σώματος, όπως εμφανίζεται στο σχήμα του [7]

Σχήμα 2.6: Μοντέλα αναπαράστασης του ανθρωπίνου σώματος

χρήσιμα και αποτελεσματικά σε τέτοιες κατηγορίες προβλημάτων είναι η δυνατότητα μνήμης που έχουν και η καλύτερη αναπαράσταση της δυναμικής μορφής του προβλήματος. Για αυτούς τους λόγους τα επιλέξαμε για το σύστημά μας και θα τα περιγράψουμε αναλυτικά παρακάτω.



Σχήμα 2.7: Διάφορα είδη μοντέλων

## 2.2 Σημεία Ενδιαφέροντος-Interest Points

Ένας αποδεδειγμένα πολύ χρήσιμος τρόπος εντοπισμού των στοιχείων που μπορεί να μας ενδιαφέρουν μέσα σε μία εικόνα και κατ'επέκταση σε ένα βίντεο είναι η ανεύρεση σημείων ενδιαφέροντος (*interest points*). Ο κύριος λόγος για αυτό είναι πως με την εφαρμογή αυτής της τεχνικής μπορούμε, ύστερα από μία αρχική σάρωση του σήματος εισόδου μας (διαδοχικά καρέ εικόνων), να επικεντρωθούμε σε ορισμένες περιοχές των διαδοχικών εικόνων για την εξαγωγή αντιπροσωπευτικών χαρακτηριστικών που είναι εύκολα διαχωρίσιμα για διαφορετικές κατηγορίες, χωρίς να αναλωνόμαστε χρονικά και υπολογιστικά στην εξαγωγή χαρακτηριστικών από ολόκληρες τις εικόνες κάθε στιγμή. Φυσικά, η επιλογή τέτοιων σημείων δεν είναι εύκολη υπόθεση και μπορεί να διαφέρει από εφαρμογή σε εφαρμογή, αλλά κοινό χαρακτηριστικό αυτών των σημείων ανεξάρτητα από το είδος σήματος εισόδου είναι πως πρέπει να χαρακτηρίζονται από *μη-μεταβλητότητα* (*invariance*) ως προς χωροχρονικές μετατοπίσεις, περιστροφές, αλλαγές κλίμακας ή παραμορφώσεις. Με αυτή την ιδιότητα επιτρέπουν την εξαγωγή χαρακτηριστικών που δεν μεταβάλλονται μεταξύ διαφορετικών δειγμάτων μίας κατηγορίας και μπορούν να εκπροσωπήσουν καλύτερα μια κλάση αντικειμένων. Κατά αυτόν τον τρόπο, μία κατηγορία κίνησης μπορεί να αναπαρασταθεί ανεξάρτητα από τη θέση, τον προσανατολισμό του εκτελεστή ή την έκταση που αυτός δίνει στην κίνηση.

**Χωρικά Σημεία Ενδιαφέροντος αμετάβλητα στην αλλαγή κλίμακας:** Στην προσπάθεια να βρεθούν τέτοια σημεία αμετάβλητα σε αλλαγές κλίμακας των υπό εξέταση αντικειμένων ή ανθρωπίνων κινήσεων αρχικά στον χώρο (πάνω σε κάθε μία εικόνα) προτάθηκαν κατά καιρούς αρκετές μέθοδοι. Η πιο διαδεδομένη και χρηστική από αυτές, εδώ και αρκετά χρόνια, είναι ο *Ανιχνευτής Harris* (*Harris Detector*). Ο ανιχνευτής αυ-

## 2.2. ΣΗΜΕΙΑ ΕΝΔΙΑΦΕΡΟΝΤΟΣ-INTEREST POINTS

τός βασίζεται στην ιδέα ανίχνευσης περιοχών σε μια χωρική εικόνα (spatial image-sp)  $f^{sp}$ , δηλαδή χωρίς τη συμμετοχή του χρόνου, όπου οι τιμές της εικόνας έχουν σημαντικές αποκλίσεις και στις δύο κατευθύνσεις. Η μέθοδος εξετάζεται από πολλούς ερευνητές, αλλά αναλυτικά αναφέρεται από τους Mikołajczyk και Schmid [10].

Για μια δεδομένη κλίμακα διαφοροποίησης (differentiation scale)  $\sigma_l^2$ , τέτοια σημεία ενδιαφέροντος μπορούν να βρεθούν από τη μήτρα δευτέρων ροπών (second moment matrix), αφού έχει περιοριστεί από ένα παράθυρο, με κλίμακα ενσωμάτωσης (integration scale)  $\sigma_i^2 = s\sigma_l^2$ , όπου το  $s$  είναι μια σταθερά που καθορίζεται πειραματικά:

$$\mu^{sp} = g^{sp}(\cdot; \sigma_i^2) * \begin{pmatrix} (D_x^{sp})^2 & D_x^{sp} D_y^{sp} \\ D_x^{sp} D_y^{sp} & (D_y^{sp})^2 \end{pmatrix} \quad (2.1)$$

όπου τα  $D_x^{sp}$  και  $D_y^{sp}$  είναι οι Γκαουσιανές παράγωγοι (Gaussian derivatives) και ορίζονται ως:

$$\begin{aligned} D_x^{sp}(\cdot; \sigma_l^2) &= \partial_x(g^{sp}(\cdot; \sigma_l^2) * f^{sp}) \\ D_y^{sp}(\cdot; \sigma_l^2) &= \partial_y(g^{sp}(\cdot; \sigma_l^2) * f^{sp}) \end{aligned} \quad (2.2)$$

και όπου το  $g^{sp}$  είναι ο χωρικός Γκαουσιανός πυρήνας (Gaussian kernel):

$$g^{sp}(x, y; \sigma^2) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (2.3)$$

Αφού οι ιδιοτιμές  $\lambda_1, \lambda_2$ , ( $\lambda_1 \leq \lambda_2$ ) του  $\mu^{sp}$  αναπαριστούν χαρακτηριστικές μεταβολές της  $f^{sp}$  και στις δύο διευθύνσεις της εικόνας, δύο ιδιάζουσες τιμές των  $\lambda_1, \lambda_2$  αναδεικνύουν την παρουσία ενός σημείου ενδιαφέροντος. Για να ανιχνευτούν αυτά τα σημεία, γίνεται χρήση του μέτρου Harris (Harris measure)

$$H^{sp} = \det(\mu^{sp}) - k \cdot \text{trace}^2(\mu^{sp}) = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (2.4)$$

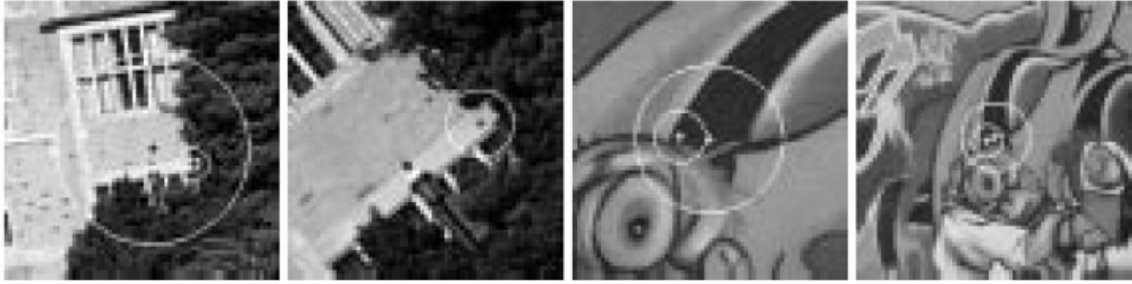
Τα ζητούμενα σημεία ενδιαφέροντος βρίσκονται στα τοπικά μέγιστα του μέτρου Harris ή αλλιώς του δείκτη γωνιότητας (cornerness).

Για την περαιτέρω βελτίωση του ανιχνευτή σε σχέση με την ανεξαρτησία από την κλίμακα, μέσω πειραμάτων εξακριβώθηκε πως με χρήση της Λαπλασιανής της Γκαουσιανής (Laplacian of Gaussian-LoG) έχουμε το υψηλότερο ποσοστό σωστής χαρακτηριστικής κλίμακας για τον ανιχνευτή μας. Η LoG για ένα σύνολο από ορισμένες κλίμακες  $\sigma_n$  ορίζεται ως:

$$|LoG(\cdot, \sigma_n)| = \sigma_n^2 |D_{xx}(\cdot, \sigma_n) + D_{yy}(\cdot, \sigma_n)| \quad (2.5)$$

Συνεπώς, κατασκευάστηκε ένας νέος ανιχνευτής, ο *Harris-Laplace*, οποίος εντοπίζει τοπικά τα σημεία μέσω του μέτρου Harris που είδαμε παραπάνω και στη συνέχεια, επιλέγει τα σημεία για τα οποία η LoG αποκτά τη μέγιστη τιμή της σε σχέση με τις εξεταζόμενες κλίμακες. Με αυτόν τον τρόπο αποκτούμε πιο αντιπροσωπευτικά σημεία ενδιαφέροντος, τα οποία συνοδεύονται και από τις συσχετιμένες τους κλίμακες (βλ. και εικόνα 2.2).

Στην ίδια προσπάθεια [10] περιγράφεται και μία ακόμα παραλλαγή για τον ανιχνευτή, ώστε να βρίσκονται σημεία αμετάβλητα προς τους αφινικούς (affine) μετασχηματισμούς, που περιλαμβάνουν διαφορετικές αλλαγές κλίμακας προς διαφορετικές κατευθύνσεις της



Σχήμα 2.8: Σημεία ενδιαφέροντος (Interest points) αμετάβλητα στην αλλαγή κλίμακας, όπως εξάγονται από τον ανιχνευτή Harris-Laplace του [10]

εικόνας. Ο Harris-Affine detector που προκύπτει μπορεί να δίνει πιο ουσιαστικά σημεία, αλλά είναι αρκετά βαρύς υπολογιστικά και στον χρόνο υπολογισμού του, λόγω των επαναληπτικών διαδικασιών που περιέχει, ώστε δεν έχει χρησιμοποιηθεί κατά κόρον σε διάφορες ερευνητικές προσπάθειες, που δείχνουν να χρησιμοποιούν τον απλούστερο Harris ανιχνευτή, με κάποιες ίσως παραλλαγές, αλλά διατηρώντας την ευκολία υπολογισμού που παρέχει.

**Χωροχρονικά Σημεία Ενδιαφέροντος:** Για προβλήματα που εμπεριέχουν κίνηση φυσικά, όπως τα προβλήματα ανθρωπίνων κινήσεων, ο εντοπισμός σημείων ενδιαφέροντος έπρεπε να εξελιχθεί ώστε να ενσωματώσει και τη διάσταση του χρόνου, αφού έχουμε διαδοχικά καρέ από εικόνες για κάθε βίντεο προς αναγνώριση. Μια τέτοια εξέλιξη του ανιχνευτή Harris βλέπουμε στην προσπάθεια των Laptev και Lindeberg [11], όπου ο εντοπισμός επεκτείνεται στο χωροχρόνο, αναζητώντας τις χωροχρονικές τιμές της εικόνας που παρουσιάζουν μεγάλες αποκλίσεις τόσο στη διάσταση του χώρου, όσο και του χρόνου. Τα σημεία αυτά θα είναι χωρικά σημεία ενδιαφέροντος, αλλά με μια χαρακτηριστική χρονική θέση, που αντιστοιχεί σε στιγμές μη συνεχόμενης κίνησης της εικόνας σε μια χωροχρονική γειτονιά.

Συνεπώς, για τη χωροχρονική ακολουθία της εικόνας,  $f : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}$  έχουμε την κατασκευή της γραμμικής χωροχρονικής αναπαράστασης  $D : \mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}_+^2$ , με συνέλιξη της  $f$  με έναν ανισοτροπικό Γκαουσιανό πυρήνα συγκεκριμένης χωρικής  $\sigma_l^2$  και χρονικής διακύμανσης  $\tau_l^2$

$$D(\cdot; \sigma_l^2, \tau_l^2) = g(\cdot; \sigma_l^2, \tau_l^2) * f(\cdot) \quad (2.6)$$

όπου ο χωροχρονικά διαχωρίσιμος Γκαουσιανός πυρήνας ορίζεται ως:

$$g(x, y, t; \sigma_l^2, \tau_l^2) = \frac{\exp(-(x^2 + y^2)/2\sigma_l^2 - t^2/2\tau_l^2)}{\sqrt{(2\pi)^3 \sigma_l^4 \tau_l^2}} \quad (2.7)$$

### 2.3. ΙΣΤΟΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΙ

Ομοίως με τη χωρική μόνο προσέγγιση, θεωρούμε τη χωροχρονική μήτρα δευτέρων ροπών (*spatio-temporal second moment matrix*), η οποία είναι ένας  $3 \times 3$  πίνακας που αποτελείται από χωρικές και χρονικές παραγώγους πρώτης τάξης, κανονικοποιημένες από μία Γκαουσιανή συνάρτηση βάρους  $g(\cdot; \sigma_i^2, \tau_i^2)$

$$\mu = g(\cdot; \sigma_i^2, \tau_i^2) * \begin{pmatrix} D_x^2 & D_x D_y & D_x D_t \\ D_x D_y & D_y^2 & D_y D_t \\ D_x D_t & D_y D_t & D_t^2 \end{pmatrix} \quad (2.8)$$

όπου οι κλίμακες ενσωμάτωσης (*integration scales*) είναι  $\sigma_i^2 = s\sigma_l^2$  και  $\tau_i^2 = s\tau_l^2$ , ενώ οι παράγωγοι πρώτης τάξης ορίζονται ως  $D_\xi(\cdot; \sigma_l^2, \tau_l^2) = \partial_\xi(g * f)$ . Η αναζήτηση για σημεία ενδιαφέροντος γίνεται πλέον στις περιοχές της  $f$  που οι ιδιοτιμές  $\lambda_1, \lambda_2, \lambda_3$  του  $\mu$  παίρνουν ιδιαίζουσες τιμές. Επεκτείνοντας πάλι τον απλό Harris ανιχνευτή, δημιουργείται ένα τρισδιάστατο μέτρο Harris/γωνιότητας

$$H = \det(\mu) - k \cdot \text{trace}^3(\mu) = \lambda_1 \lambda_2 \lambda_3 - k(\lambda_1 + \lambda_2 + \lambda_3)^3 \quad (2.9)$$

Πάλι λοιπόν, τα σημεία ενδιαφέροντος που επιλέγουμε είναι αυτά που αντιστοιχούν στα θετικά τοπικά μέγιστα της  $H$ .

Για την επιλογή της κλίμακας που συνοδεύει το κάθε σημείο ενδιαφέροντος, ακολουθούμε παρόμοια τακτική με την παραπάνω υποενότητα, όπου εμπλέκουμε την Λαπλασιανή, αλλά σε χωροχρονική μορφή πλέον και κανονικοποιημένη ως προς την κλίμακα, ώστε να ορίζεται με τον εξής τρόπο:

$$\nabla_{norm}^2 D = D_{xx,norm} + D_{yy,norm} + D_{tt,norm} \quad (2.10)$$

όπου  $D_{xx,norm} = \sigma^{2a}\tau^{2b}D_{xx}$  και  $D_{tt,norm} = \sigma^{2c}\tau^{2d}D_{tt}$ . Με την επιλογή των κατάλληλων παραμέτρων για τα  $a, b, c, d$  μπορούμε να φέρουμε τη Λαπλασιανή σε μορφή που είναι διαχειρίσιμη, ώστε τελικά, αναζητούμε τα σημεία που μεγιστοποιούν τοπικά ανάμεσα σε χωρικές και χρονικές κλίμακες τη συνάρτηση

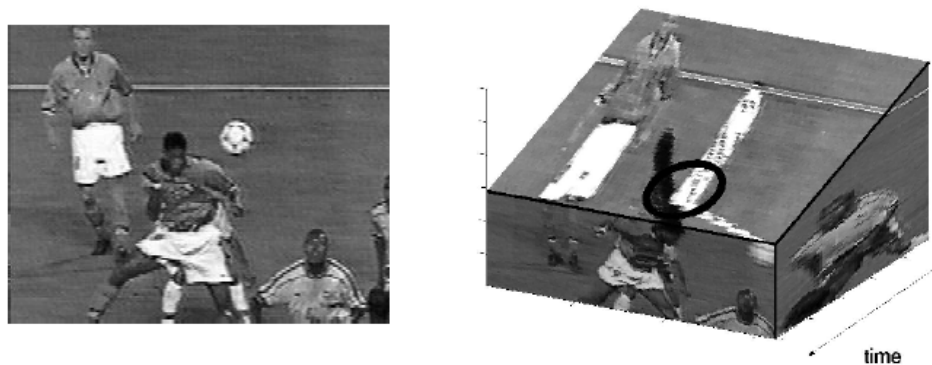
$$\nabla_{norm}^2 D = \sigma^2\tau^{1/2}(D_{xx} + D_{yy}) + \sigma\tau^{3/2}D_{tt} \quad (2.11)$$

Η τελευταία μέθοδος έχει χρησιμοποιηθεί με υψηλά ποσοστά επιτυχίας σε αρκετές σύγχρονες προσπάθειες [37, 42].

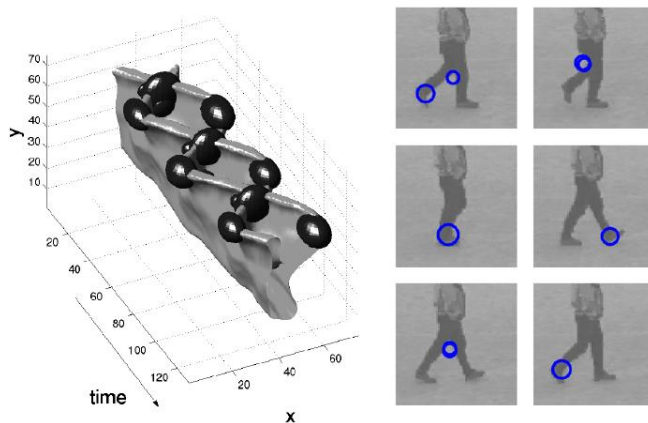
## 2.3 Ιστογραφικές Μέθοδοι

Τα τελευταία χρόνια πολλοί ερευνητές έχουν στραφεί προς απλούστερα στατιστικά στοιχεία που εξάγονται από θεμελιώδη οπτικά χαρακτηριστικά, όπως η κλίση (*gradient*) και η οπτική ροή (*optical flow*). Τέτοια στοιχεία είναι και αυτά που προέρχονται από *Ιστογράμματα* (*Histograms*), τα οποία στην ουσία παρουσιάζουν ποια χαρακτηριστικά εμφανίζονται με μεγαλύτερη συχνότητα σε συγκεκριμένες περιοχές της εικόνας που εξετάζονται. Αυτό το είδος μεθόδου βρίσκει πολύ καλή εφαρμογή σε προβλήματα που περιλαμβάνουν





(α')



(β')

Σχήμα 2.9: Ο ανιχνευτής Harris3D, όπως στο σχήμα του [11]. (α') Ανίχνευση σημείων ενδιαφέροντος στο χωροχρόνο. (β') Τρισδιάστατη απεικόνιση σημείων ενδιαφέροντος κατά τη διάρκεια περπατήματος.

κίνηση, αφού τα χαρακτηριστικά που εξάγονται μπορούν να αναπαραστήσουν περιεκτικά και με χαμηλό υπολογιστικό κόστος διαφορετικές κινήσεις που γίνονται σε διαφορετικά σημεία της εικόνας.

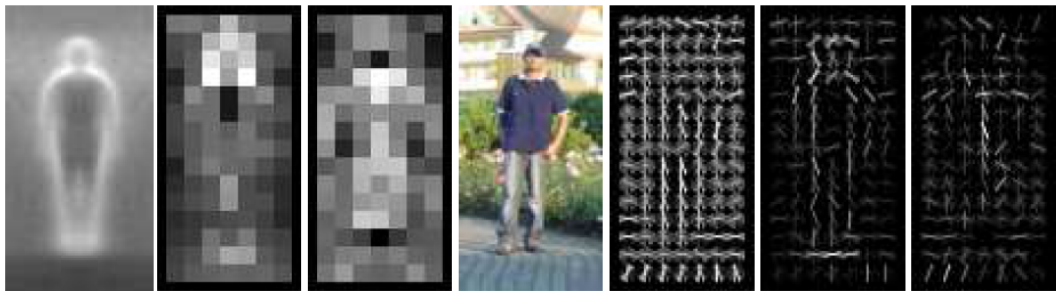
### 2.3.1 Ιστογράμματα Προσανατολισμένης Κλίσης

Οι Dalal και Triggs εισήγαγαν το 2005 [12] τα *Ιστογράμματα Προσανατολισμένης Κλίσης (Histograms of Oriented Gradient-HOG)*. Ο τρόπος περιγραφής αυτών των δεδομένων θυμίζει παλαιότερες προσεγγίσεις, όπως τα Ιστογράμματα Κατεύθυνσης Ακμών ή

### 2.3. ΙΣΤΟΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΙ

κάποιες μεθόδους από το σύνολο των Μετασχηματισμών Αμετάβλητων προς την αλλαγή Κλίμακας Χαρακτηριστικών (Scale Invariant Feature Transformation-SIFT) ή ακόμα τη μέθοδο των 'πλαισίων σχημάτων' (shape contexts). Στην προκειμένη περίπτωση βέβαια, τα Ιστογράμματα υπολογίζονται σε ένα πυκνό πλέγμα από κελιά με ομοιόμορφες αποστάσεις και χρησιμοποιούν επικαλυπτόμενες κανονικοποιήσεις τοπικής αντίθεσης για βελτιωμένη απόδοση. Ο κύριος λόγος της καλής τους απόδοσης είναι πως η εμφάνιση και το σχήμα τοπικών αντικειμένων μπορεί να χαρακτηριστεί αρκετά καλά από την κατανομή την εντάσεων των τοπικών κλίσεων ή ακμών, ακόμα χωρίς τη γνώση για την ακριβή θέση των κλίσεων ή ακμών. Στην πράξη αυτό υλοποιείται διαιρώντας την εικόνα σε μικρές τοπικές περιοχές ("κελιά") και υπολογίζοντας για κάθε κελί το τοπικό μονοδιάστατο ιστογράμμα των κατευθύνσεων των κλίσεων ή των προσανατολισμών των ακμών πάνω στα pixels του κελιού. Οι συνδυασμένες παρατηρήσεις από τα ιστογράμματα των διαφορετικών κελιών αποτελούν την αναπαράσταση των χαρακτηριστικών μας.

Για την υλοποίηση των HOG απαιτείται η ακόλουθη διαδικασία. Αρχικά, γίνεται ο



Σχήμα 2.10: Απεικόνιση ανθρώπου και υπολογισμών της προσανατολισμένης κλίσης συνολικά, αλλά και στα διαφορετικά κελιά, για την εξαγωγή των HOG, όπως στο σχήμα του [12]

υπολογισμός της κλίσης (gradient) με το φιλτράρισμα των χρωματικών τιμών ή των τιμών έντασης φωτεινότητας (αν πρόκειται για ασπρόμαυρες εικόνες) της εικόνας με τα διακριτά φίλτρα - μάσκες παραγωγίσις και για τις δύο διαστάσεις:  $[-1, 0, 1]$  και  $[-1, 0, 1]^T$ . Στη συνέχεια, πρέπει να κατασκευαστούν τα ιστογράμματα για κάθε κελί. Κάθε pixel μέσα στο κελί προσθέτει στο ιστόγραμμα τις πληροφορίες για την κατεύθυνση της κλίσης του με ένα συγκεκριμένο βάρος για την κάθε ψήφο προς μία κατεύθυνση. Συνήθως για αυτό το λόγο χρησιμοποιείται απλά σαν βάρος το πλάτος της κλίσης προς μία κατεύθυνση, αφού έχει αποδειχθεί πως δίνει τα καλύτερα αποτελέσματα. Για την καλύτερη κανονικοποίηση των πλατών της κλίσης, πολλά κελιά ομαδοποιούνται σε μεγαλύτερα χωρικά τμήματα (*blocks*). Για το είδος των τμημάτων που χρησιμοποιούνται για αυτό το σκοπό, οι Dalal και Triggs δοκίμασαν αρκετά μοντέλα, όπως το απλό ορθογώνιο (rectangular, R-HOG τμήμα) ή το κυκλικό (circular, C-HOG τμήμα). Πιο σύνηθες, λόγω απλότητας, είναι το ορθογώνιο που χαρακτηρίζεται από 3 παραμέτρους:  $\alpha, \beta, \gamma$ , όπου αναπαριστούν  $\alpha \times \alpha$  πλέγματα από  $\beta \times \beta$  κελιά pixels, που το καθένα περιέχει  $\gamma$  διαφορετικές κατευ-

θύνσεις. Για την κανονικοποίηση των τμημάτων προτείνονται επίσης διάφορα σχήματα, που περιλαμβάνουν το  $\nu$ , το μη κανονικοποιημένο διάνυσμα που περιέχει όλα τα ιστογράμματα ενός τμήματος, τις νόρμες του και μια μικρή σταθερά  $e$ . Ο υπολογισμός των κανονικοποιημένων τιμών μπορεί να γίνεται με έναν από τους παρακάτω τρόπους:

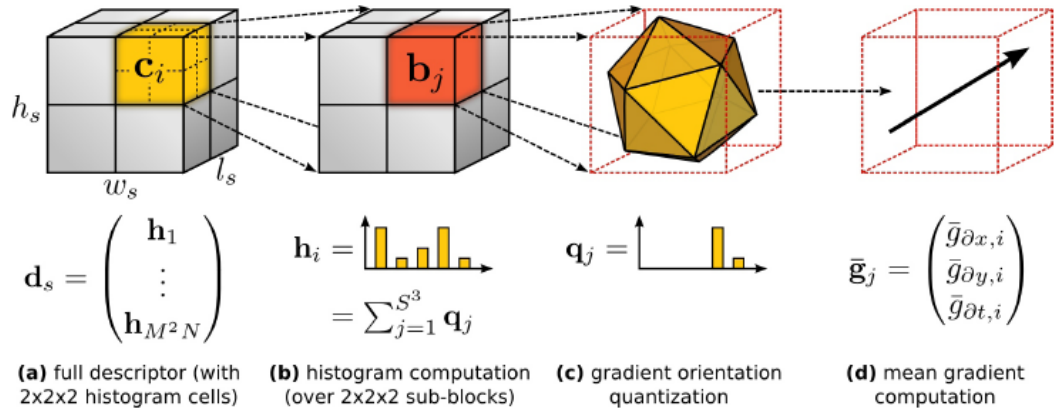
$$\begin{aligned} L2 - norm : f &= \frac{\nu}{\sqrt{\|\nu\|_2^2 + e^2}} \\ L1 - norm : f &= \frac{\nu}{\sqrt{\|\nu\|_2 + e}} \\ L1 - sqrt : f &= \sqrt{\frac{\nu}{\|\nu\|_2 + e}} \end{aligned} \quad (2.12)$$

Ύστερα από αρκετές πειραματικές εφαρμογές [14, 38, 37, 41], είναι σαφές πως τα HOG είναι αρκετά ικανά να εντοπίσουν το ανθρώπινο περίγραμμα, με μειωμένη επίδοση όμως στην ανίχνευση εσωτερικών ακμών, όπως είναι τα χέρια όταν βρίσκονται μπροστά στο σώμα. Πολύ σημαντική αποδεικνύεται πως είναι η κωδικοποίηση των κατευθύνσεων, καθώς σύμφωνα με πειραματικά δεδομένα, αυξάνοντας τον αριθμό των διαφορετικών κατευθύνσεων μέχρι 9, βελτιώνει την απόδοση σημαντικά, ενώ πέρα από αυτό λίγη διαφορά υπάρχει. Ειδικά για την ανίχνευση ανθρώπων, η σωστή δειγματοληψία των κατευθύνσεων και η ισχυρή τοπική φωτομετρική κανονικοποίηση φαίνεται να αποτελούν την καλύτερη τακτική, πιθανώς γιατί επιτρέπουν στα μέλη του σώματος να αλλάζουν εμφάνιση και να μετακινούνται ελεύθερα μέσα στην εικόνα. Σε άλλες προσπάθειες, τα HOG χαρακτηριστικά έχουν χρησιμοποιηθεί και με πιο πολύπλοκα μοντέλα ανθρωπίνου σώματος που προσπαθούν να διαχωρήσουν τα διαφορετικά μέλη από τις διαφορετικές μορφές των τμημάτων HOG, όπως στην προσπάθεια του Felzenswalb [39]. Τέλος, ο Klaser σε μία προσέγγισή του [13] προσπαθεί να επεκτείνει τα HOG σε 3 διαστάσεις υπολογίζοντας Ιστογράμματα κατευθύνσεων χωροχρονικών κλίσεων σε κυβικές περιοχές που σχηματίζονται γύρω από τρισδιάστατα σημεία ενδιαφέροντος (βλ. και εικόνα 2.3.1). Κάτι τέτοιο όμως, αυξάνει σημαντικά την πολυπλοκότητα και τη διάσταση του προβλήματος, χωρίς να προσφέρει την αντίστοιχη βελτίωση απόδοσης.

### 2.3.2 Ιστόγραμμα Οπτικής Ροής

Πέρα από την ανάγκη για αναπαράσταση περιγραμμάτων μέσω των προσανατολισμένων κλίσεων που είδαμε παραπάνω, προκύπτει και μία ανάγκη για εκπροσώπηση πιο δυναμικών στοιχείων, αφού έχουμε να κάνουμε με ανίχνευση κινήσεων. Ο Dalal το 2006 [14], συνεχίζοντας την προσέγγισή του για την χρήση Ιστογραμμάτων για εξαγωγή χαρακτηριστικών που θα τροφοδοτήσουν τα μοντέλα, εισήγαγε τα *Ιστογράμματα Οπτικής Ροής (Histograms of Optical Flow - HOF)* για να συμπληρώσει τα ήδη επιτυχημένα HOG. Ο κύριος λόγος επιτυχίας των χαρακτηριστικών που προέρχονται από οπτική ροή είναι επειδή τα τοπικά διαφορικά της οπτικής ροής έχουν την ικανότητα να παραμένουν αμετάβλητα ως προς τις μετακινήσεις της κάμερας ή του περιβάλλοντος του παρασκήνιου. Ως αποτέλεσμα, μπορούμε να αναγνωρίσουμε ανθρώπινες κινήσεις που μας ενδιαφέρουν παρά το θορυβώδες παρασκήνιο ή περιστροφές της κάμερας λήψης.

### 2.3. ΙΣΤΟΓΡΑΦΙΚΕΣ ΜΕΘΟΔΟΙ



Σχήμα 2.11: Επέκταση των HOG στις 3 διαστάσεις με ενσωμάτωση του χρόνου από τον Klaser στο [13]

Η υλοποίηση για τον υπολογισμό των Ιστογραμμάτων HOF είναι αρκετά όμοια ως προς τα στάδια της με την μέθοδο για τον υπολογισμό των Ιστογραμμάτων HOG. Για τον υπολογισμό της οπτικής ροής  $\mathbf{w}$  συγκεκριμένα, χρησιμοποιείται για κάθε pixel ανεξάρτητα η λύση της μετριασμένης γραμμική εξίσωση Ελαχίστων Τετραγώνων (damped Linear Least Squares equation)

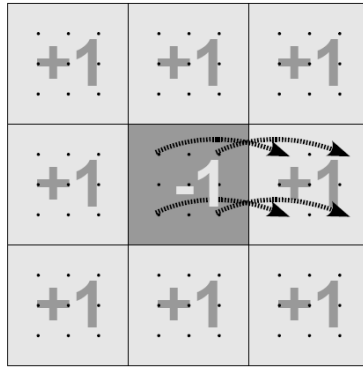
$$\mathbf{w} = (\mathbf{A}^T \mathbf{A} + \beta \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b} \quad (2.13)$$

πάνω σε μια μικρή  $N \times N$  γειτονιά, όπου το  $\mathbf{b}$  είναι ένα  $N^2$  διάνυσμα στήλης που κωδικοποιεί τις χρονικές διαφορές της εικόνας, ο  $\mathbf{A}$  είναι ένας  $N^2 \times 2$  πίνακας από χωρικές κλίσεις  $[\mathcal{I}_x, \mathcal{I}_y]$  και  $\beta$  είναι ένας μετριαστικός παράγοντας για να μειωθούν αριθμητικά θέματα από τον μοναδιαίο  $\mathbf{A}^T \mathbf{A}$ . Στη συνέχεια, για τον υπολογισμό των διαφορών των οπτικών ροών, οι οποίες θα συμμετέχουν στα Ιστογράματά μας, έχουν προταθεί αρκετοί τρόποι, αλλά ένας που αναδεικνύει καλύτερα τις σχετικές κινήσεις των άκρων είναι ο IMHcd (Internal Motion Histograms cell differences). Σύμφωνα με αυτή την τεχνική, έχουμε  $3 \times 3$  τμήματα από κελιά και σε καθένα pixel από τα 8 εξωτερικά κελιά υπολογίζονται οι διαφορές οπτικής ροής ως προς τα αντίστοιχα pixel του κεντρικού κελιού, ώστε να προκύψει ένα ιστόγραμμα προσανατολισμού. Τα 8 ιστογράμματα που προκύπτουν κανονικοποιούνται σαν ένα τμήμα (βλ. και εικόνα 2.12(β')).

Τα HOF έχουν επίσης δοκιμαστεί σε πολλές ερευνητικές προσπάθειες, κυρίως σε συνδυασμό με τα HOG, και έχουν αποδειχθεί αρκετά αξιόπιστα και αποδοτικά [37, 42]. Επίσης, πολύ σημαντικά αποδεικνύονται τα στάδια κανονικοποίησης, αλλά και η μορφή της οπτικής ροής που συμμετέχει στα ιστογράμματα, ώστε να μπορούν να αναδειχθούν καλύτερα οι σχετικές κινήσεις των άκρων των ανθρωπίνων σωμάτων.



(α')



(β')

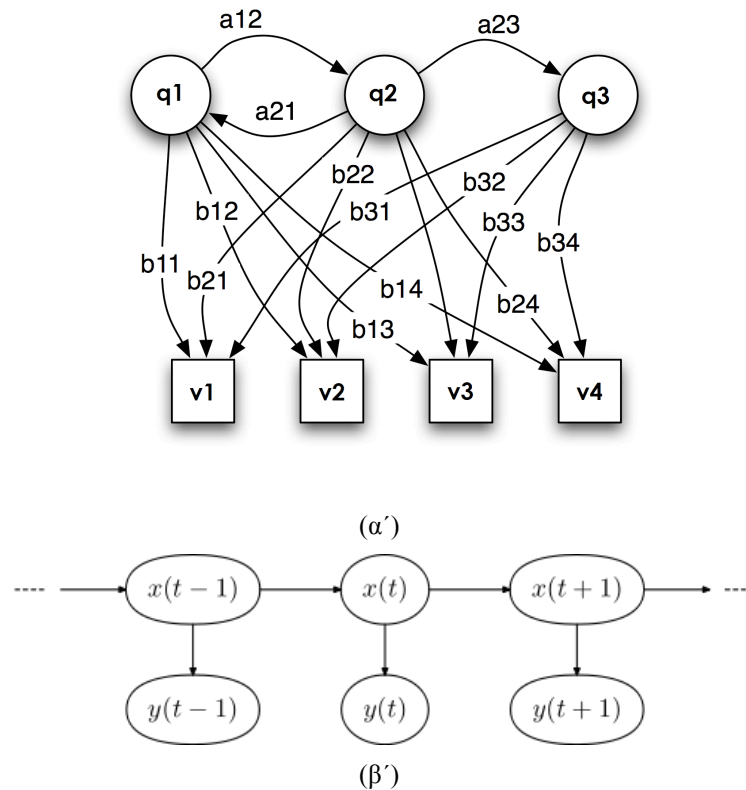
Σχήμα 2.12: Ιστογράμματα Οπτικής Ροής-HOF. (α') Υπολογισμός οπτικής ροής και πλάτους ροής μεταξύ δύο διαδοχικών εικόνων με ανάδειξη ορίων κίνησης. (β') Σχήμα κωδικοποίησης IMHcd, όπου ο υπολογισμός των διαφορών γίνεται μεταξύ αντίστοιχων pixel σε γειτονικά κελιά, όπως στο σχήμα του [14]

## 2.4 Κρυφά Μαρκοβιανά Μοντέλα

Τα *Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models)* είναι στατιστικά εργαλεία που έχουν προταθεί από πολύ παλιά, αφού εμφανίστηκαν σε μια σειρά στατιστικών μελετών από τον Leonard E. Baum και άλλους συγγραφείς στο δεύτερο μισό της δεκαετίας του 1960. Η κύρια χρήση τους σε εφαρμογές αναγνώρισης φωνής άρχισε στα μέσα της δεκαετίας του 1970 και από τότε έχουν χρησιμοποιηθεί σε πολλές ερευνητικές μελέτες αναγνώρισης. Μεγάλο ρόλο στη γενικευμένη εφαρμογή τους έπαιξε και η εκτενής ανάλυσή τους στο βιβλίο των Rabiner και Schafer, *Digital processing of speech signals* του 1978 [43]. Η κύρια χρήση τους για Ανίχνευση και Κατηγοριοποίηση ανθρωπίνων κινήσεων να αρχίζει κυρίως τη δεκαετία του 1990. Μια από τις πρώτες προσπάθειες είναι αυτή των Yamato et al. [44], στην οποία αποδεικνύονται οι μοναδικές ιδιότητες του εργαλείου αυτού στην αναπαράσταση δυναμικών προβλημάτων, όπως είναι οι ανθρώπινες κινήσεις,

κυρίως λόγω της δυνατότητας μνήμης που περιέχει και της ύπαρξης διασυνδεδεμένων καταστάσεων.

### 2.4.1 Ορισμός



Σχήμα 2.13: Αρχιτεκτονική Κρυφών Μαρκοβιανών Μοντέλων-HMM. (α') Γενική μορφή ενός HMM. (β') Καταστάσεις που αποτελούν το HMM.

Ένα HMM αποτελείται από έναν αριθμό από καταστάσεις (states) για καθεμία από τις οποίες ορίζεται μία πιθανότητα μετάβασης από τη μία κατάσταση σε μία άλλη. Με τον καιρό, οι μεταβάσεις από κατάσταση σε κατάσταση συμβαίνουν στοχαστικά. Όπως σε όλα τα μαρκοβιανά μοντέλα, οι καταστάσεις σε μία ορισμένη χρονική στιγμή εξαρτώνται μόνο από την κατάσταση κατά την προηγούμενη χρονική στιγμή. Ένα σύμβολο παράγεται σαν παρατήρηση από μία κατάσταση του HMM σύμφωνα με τις πιθανότητες που έχουν ανατεθεί στις καταστάσεις. Οι καταστάσεις του HMM δεν είναι άμεσα παρατηρήσιμα (σε αυτό οφείλεται και η ονομασία τους ως κρυφά), αλλά μπορούν να παρατηρηθούν μόνο μέσω από την ακολουθία παρατηρούμενων συμβόλων. Για την περιγραφή ενός διακριτού HMM απαιτείται ο ορισμός των παρακάτω στοιχείων:

T = το μήκος της ακολουθίας παρατηρήσεων.

$Q = \{q_1, q_2, \dots, q_N\}$  : το σύνολο των καταστάσεων.

$N$  = ο αριθμός καταστάσεων του μοντέλου.

$V = \{v_1, v_2, \dots, v_M\}$  : το σύνολο των πιθανών συμβόλων στην έξοδο.

$M$  = ο αριθμός των παρατηρούμενων συμβόλων.

$A = \{a_{ij} | a_{ij} = Pr(s_{t+1} = q_j | s_t = q_i)\}$  : η πιθανότητα μετάβασης μεταξύ καταστάσεων, όπου:

$a_{ij}$  είναι η πιθανότητα μετάβασης από την κατάσταση  $q_i$  στην κατάσταση  $q_j$ .

$B = \{b_j(k) | b_j(k) = Pr(v_k | s_t = q_j)\}$  : Πιθανότητα εμφάνισης συμβόλου στην έξοδο, όπου:

$b_j(k)$  είναι η πιθανότητα το σύμβολο  $v_k$  να είναι στην έξοδο στην κατάσταση  $q_j$ .

$\pi = \{\pi_i | \pi_i = Pr(s_1 = q_i)\}$  : Η πιθανότητα αρχικής κατάστασης.

$\lambda = \{A, B, \pi\}$  το σύνολο των παραμέτρων του μοντέλου.

Χρησιμοποιώντας αυτό το μοντέλο, οι μεταβάσεις περιγράφονται ως εξής:

$S = \{s_t\}, t = 1, 2, \dots, T$  : Η κατάσταση  $s_t$  είναι η  $t$  κατάσταση σε σειρά (μη-παρατηρήσιμο)

$O = O_1, O_2, \dots, O_T$  : Η παρατηρούμενη ακολουθία συμβόλων (μήκος=T).

Όπως φαίνεται και στο σχήμα 2.13(α'), οι μεταβάσεις γίνονται από τη μία κατάσταση στην άλλη, αλλά μπορούν να γίνουν και από την ίδια την κατάσταση στον εαυτό της, δηλαδή να παραμείνουμε στην ίδια κατάσταση. Κάθε κατάσταση στοχαστικά παράγει ένα σύμβολο, ενώ συνολικά έχουμε την συμβολοακολουθία  $O = O_1, O_2, \dots, O_T$  από τη χρονική στιγμή 1 έως την T. Συνεπώς, το μοντέλο HMM περιγράφεται πλήρως από τρεις πίνακες: τον πίνακα πιθανοτήτων μετάβασης μεταξύ καταστάσεων  $A$ , τον πίνακα πιθανότητας εμφάνισης συμβόλων στην έξοδο  $B$  και τον πίνακα πιθανοτήτων αρχικής κατάστασης  $\pi$ .

## 2.4.2 Ανίχνευση

Για την ανίχνευση των παρατηρούμενων συμβολοσειρών, απαιτείται η δημιουργία ενός HMM για κάθε κατηγορία. Για έναν ταξινομητή (classifier) των  $C$  κατηγοριών, επιλέγουμε το μοντέλο που ταιριάζει καλύτερα στις παρατηρήσεις από τα  $C$  HMMs  $\lambda_i = \{A_i, B_i, \pi_i\}, i = 1 \dots C$ . Αυτό σημαίνει ότι όταν δίνεται η ακολουθία μιας άγνωστης κατηγορίας, υπολογίζουμε την πιθανότητα  $Pr(\lambda_i | O)$  για κάθε HMM  $\lambda_i$  και επιλέγουμε το  $\lambda_{c^*}$ , όπου

$$c^* = \arg \max_i (Pr(\lambda_i | O)) \quad (2.14)$$

Δεδομένης της παρακολουθούμενης ακολουθίας  $O = O_1, O_2, \dots, O_T$  και του HMM  $\lambda_i$ , σύμφωνα με τον κανόνα του Bayes, το πρόβλημα είναι πως θα υπολογιστεί η  $Pr(O | \lambda_i)$ , η πιθανότητα η ακολουθία αυτή να παράχθηκε από το HMM  $\lambda_i$ . Αυτή η πιθανότητα υπολογίζεται με τον "εμπρόσθιο αλγόριθμο" (forward algorithm), που ορίζεται ως εξής:

$$\alpha_t(i) \equiv Pr(O_1, O_2, \dots, O_t, s_t = q_i | \lambda) \quad (2.15)$$

το  $\alpha_t(i)$  καλείται η εμπρόσθια μεταβλητή και μπορεί να υπολογισθεί αναδρομικά ως εξής:

$$\alpha_t(j) = \left\{ \sum_i \alpha_{t-1}(i) a_{ij} \right\} b_j(O_t) \quad (2.16)$$

$$\alpha_1 = \pi_j b_j(O_1) \quad (2.17)$$

Έτσι:

$$Pr(O|\lambda_i) = \sum_{i \in S_F} \alpha_T(i) \lambda_{c^*}, c^* = \arg \max_i (Pr(\lambda_i|O)) \quad (2.18)$$

Με τη βοήθεια της παραπάνω εξίσωσης μπορούμε να υπολογίσουμε την *πιθανοφάνεια* (*likelihood*) του κάθε HMM και να επιλέξουμε το πιο πιθανό HMM σαν αποτέλεσμα της αναγνώρισης. Αφού η πιθανοφάνεια υπολογίζεται από ολόκληρο το μήκος του προτύπου, η διακύμανση της χρονικής κλίμακας, οι χρονικές μετατοπίσεις και ορισμένα λάθη στην κβάντωση των διανυσμάτων έχουν μικρή επιρροή στην ακρίβεια υπολογισμού της πιθανοφάνειας. Από αυτή την ιδιότητα πηγάζει το πλεονέκτημα των HMM έναντι άλλων μεθόδων για αναγνώριση δυναμικών προτύπων και κατά συνέπεια για αναγνώριση ανθρωπίνων κινήσεων.

### 2.4.3 Εκπαίδευση

Στη φάση της εκπαίδευσης, κάθε HMM μοντέλο πρέπει να εκπαιδευτεί ώστε να είναι το πιο πιθανό να παράγει τα πρότυπα συμβόλων για την κατηγορία του. Αυτό σημαίνει να βελτιστοποιηθούν οι παράμετροι του μοντέλου ( $A, B, \pi$ ) για να μεγιστοποιηθεί η πιθανότητα παρατήρησης της ακολουθίας  $Pr(O|\lambda)$ . Ο αλγόριθμος Baum-Welch χρησιμοποιείται για αυτούς τους υπολογισμούς.

Ορίζουμε:

$$\beta_t(i) \equiv Pr(O_{t+1}, \dots, O_T | s_t = q_i, \lambda) \quad (2.19)$$

το  $\beta_t(i)$  καλείται η *οπίσθια μεταβλητή* και μπορεί επίσης να λυθεί επαγωγικά με τρόπο όμοιο με αυτό που χρησιμοποιείται για την εμπρόσθια μεταβλητή  $\alpha_t(i)$ .

$$\gamma_t(i) \equiv P(s_t = q_i | O_1, \dots, O_T, \lambda) \quad (2.20)$$

$$= \frac{\alpha_t(i) \beta_t(i)}{P(O|\lambda)} \quad (2.21)$$

$$\xi_t(i, j) \equiv P(s_t = q_i, s_{t+1} = q_j | O_1, \dots, O_T, \lambda) \quad (2.22)$$

$$= \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(i)}{P(O|\lambda)} \quad (2.23)$$

Χρησιμοποιώντας αυτές τις εξισώσεις, οι παράμετροι του HMM  $\lambda$  μπορούν βελτιωθούν στο  $\lambda$ . Η επανεκτίμηση των εξισώσεων από το  $\lambda = (A, B, \pi)$  στο  $\lambda = (\hat{A}, \hat{B}, \hat{\pi})$  είναι:

$$a_{i,j} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (2.24)$$

$$b_i(k) = \frac{\sum_{t \in \{t | O_t = v_k\}} \gamma_t(i)}{\sum_{t=1}^T \gamma_t(i)} \quad (2.25)$$

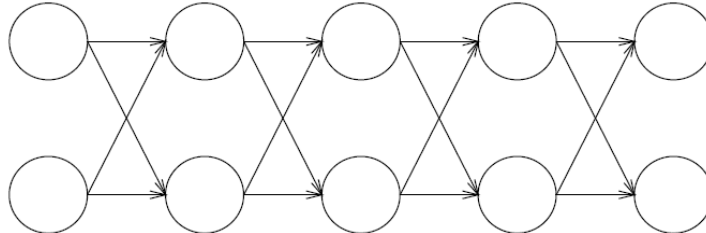
$$\pi_i = \gamma_1(i) \quad (2.26)$$



Η εκπαίδευση συγκλίνει αν  $\lambda = \lambda$ . Ο αλγόριθμος Baum–Welch είναι ένα παράδειγμα εμπρόσθιου-οπίσθιου αλγορίθμου (*forward-backward algorithm*) και είναι μια ειδική περίπτωση του αλγορίθμου *expectation-maximization*. Ο αλγόριθμος Baum–Welch δεν βρίσκει πάντα το ολικό μέγιστο, αλλά βρίσκει το τοπικό μέγιστο της  $Pr(O|\lambda)$ , χωρίς αυτό να δημιουργεί όμως προβλήματα.

#### 2.4.4 Επεκτάσεις

Τα HMM μοντέλα χρησιμοποιούνται κατά κόρον σε πολλές εφαρμογές ανίχνευσης και κατηγοριοποίησης ανθρωπίνων κινήσεων, λόγω των θετικών στοιχείων που αναφέραμε παραπάνω. Για συγκεκριμένες εφαρμογές όμως, που απαιτούν παρακολούθηση παραπάνω από ενός στοιχείων, όπως των δύο χεριών ή και του κεφαλιού του ανθρώπου ταυτόχρονα, υπήρξαν ερευνητικές προσπάθειες που προσπάθησαν να διαχωρίσουν αυτή την ξεχωριστή παρακολούθηση σε περισσότερα HMM μοντέλα που υπάρχουν ταυτόχρονα και αλληλοσυνδέονται. Μια τέτοια προσπάθεια είναι και των Brand et al. [45], στην οποία έγινε εισαγωγή της έννοιας των *Αλληλένδετων HMM (Coupled HMM)*, στα οποία οι παράμετροι του ενός μοντέλου επηρεάζουν τις παραμέτρους του άλλου με δυνατότητα εναλλαγής των καταστάσεων του ενός με του άλλου (βλ. και εικόνα 2.4.4). Παρά τα υποσχόμενα αποτελέσματα που παρήγαγε η μέθοδος αυτή για συγκεκριμένες εφαρμογές, η χρήση της δεν έγινε πολύ ευρεία κυρίως λόγω της αυξημένης πολυπλοκότητας που εισάγει.



Σχήμα 2.14: Αλληλένδετο HMM (Coupled HMM)

#### *2.4. ΚΡΥΦΑ ΜΑΡΚΟΒΙΑΝΑ ΜΟΝΤΕΛΑ*

---

## Κεφάλαιο 3

# Εφαρμογή στις Πολεμικές Τέχνες

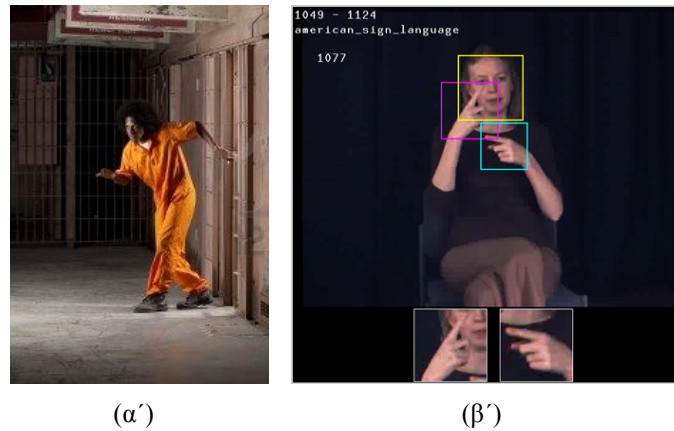
Η Ανίχνευση και Κατηγοριοποίηση Ανθρωπίνων Κινήσεων έχει πληθώρα εφαρμογών στον κόσμο γύρω μας, καθώς υπάρχουν πολλά προβλήματα στα οποία επιθυμούμε την αναγνώριση του είδους κίνησης ή γενικότερα δράσης που έχει καταγραφεί σε ένα βίντεο. Τέτοια προβλήματα μπορεί να είναι από αρκετά απλοϊκά, όπως η ανίχνευση ενός ανθρώπου που περπατά από μία κάμερα παρακολούθησης ή η διάκριση μεταξύ ενός ανθρώπου που παίζει ποδόσφαιρο από έναν που παίζει τένις, έως και πολυσύνθετα που ενέχουν μεγάλο βαθμό πολυπλοκότητας, όπως η αναγνώριση νοηματικής γλώσσας με παρακολούθηση των χεριών και του κεφαλιού ενός νοηματιστή ή η αναγνώριση εντολών που δίνονται με σωματικές κινήσεις από ένα ρομποτικό σύστημα. Ανάλογα με τη δυσκολία του προβλήματος βέβαια, καθορίζεται και η πολυπλοκότητα που πρέπει να έχουν οι μέθοδοι και τα μοντέλα που αναπτύσσονται για την ανίχνευση και κατηγοριοποίηση, αλλά σαν γενική επιδίωξη έχουμε την ανάπτυξη του απλούστερου συστήματος τεχνικών που μπορεί να εφαρμοστεί με αξιοπιστία στο μεγαλύτερο δυνατό σύνολο εφαρμογών. Μια μεγάλη κατηγορία προβλημάτων που επιδέχονται λύσεις από τέτοιου είδους τεχνικές, αλλά δεν έχουν ακόμα διερευνηθεί σε μεγάλο βαθμό, είναι τα προβλήματα που αφορούν αναγνώριση κινήσεων από πολεμικές τέχνες.

### 3.1 Γενικά

Οι Πολεμικές Τέχνες είναι εκτεταμένα συστήματα κωδικοποιημένων πρακτικών και παραδόσεων μάχης, τα οποία ακολουθούνται από ένα σύνολο ανθρώπων για μία μεγάλη ποικιλία λόγων. Τέτοιοι είναι η αυτο-άμυνα, ο διαγωνισμός, η σωματική υγεία και ευεξία, καθώς και η διανοητική και πνευματική ανάπτυξη, αφού συνήθως συνοδεύονται και από ένα σύνολο αρχών για αυτούς που τις εξασκούν. Ανάλογα με την τεχνική εστίαση οι πολεμικές τέχνες μπορεί να είναι άοπλες, οι οποίες με τη σειρά τους υποδιαιρούνται σε αυτές που χρησιμοποιούν χτυπήματα (*strikes*), όπως μπουινιές (*punching*), λακτίσματα (*kicking*) ή άλλου είδους χτυπήματα, και σε αυτές που χρησιμοποιούν λαβές (*grappling*), όπως απωθητικές (*throwing*), ακινητοποίησης/υποταγής (*locks/submission holds*) ή ενα-

### 3.1. ΓΕΝΙΚΑ

---



Σχήμα 3.1: Απλούστερα και πιο σύνθετα προβλήματα ανίχνευσης και κατηγοριοποίησης ανθρωπίνων κινήσεων.

πόθεσης (*pinning*), ή πολεμικές τέχνες που βασίζονται στη χρήση κάποιου όπλου, των οποίων οι κατηγορίες ορίζονται από το είδος του όπλου που χρησιμοποιείται, όπως σπαθιά, κοντάρια, τόξα κ.ά. Μια ξεχωριστή κατηγορία είναι επίσης και οι πολεμικές τέχνες που εκτελούνται υπο τη μορφή χορευτικών κινήσεων, κυρίως ως αποτέλεσμα πρόσμιξης και πολιτισμικών στοιχείων μέσα τους. Για την αναγνώριση ανθρωπίνων κινήσεων είναι σαφές πως μας ενδιαφέρουν κυρίως οι άοπλες πολεμικές τέχνες, αν και επεκτείνοντας τα μοντέλα είναι δυνατόν να ενταχθούν και τα όπλα στις κατηγορίες προς αναγνώριση. Επίσης, ανάλογα με την προδιάθεση ή εφαρμογή τους, μπορούν να χωριστούν αυτές με πνευματικό ή θεραπευτικό προσανατολισμό και σε αυτές με μαχητικό προσανατολισμό. Οι τελευταίες μπορεί να εξασκούνται είτε για μαχητικό αθλητισμό ή για αυτο-άμυνα, στρατιωτικές ή αστυνομικές εφαρμογές. Εμείς θα επικεντρωθούμε κυρίως σε αυτές που κινούνται στα πλαίσια του αθλητισμού, αφού είναι πιο σαφώς καθορισμένες και συναντώνται συχνότερα στην καθημερινή ζωή.

#### 3.1.1 Χαρακτηριστικά

Οι άοπλες πολεμικές τέχνες με αθλητικό προσανατολισμό θα αποτελέσουν το αντικείμενο της περαιτέρω μελέτης μας. Ο κύριος λόγος για αυτή μας την επιλογή είναι επειδή αυτού του είδους οι πολεμικές τέχνες διαθέτουν ένα σύνολο χαρακτηριστικών που τις κάνουν πολύ καλές εφαρμογές για μεθόδους ανίχνευσης και κατηγοριοποίησης ανθρωπίνων κινήσεων, με την έννοια πως οι διαφορετικές κατηγορίες κινήσεων που πρέπει να μοντελοποιηθούν έχουν τη δυνατότητα να αναπαρασταθούν με διαχωρίσιμα χαρακτηριστικά. Τέτοια χαρακτηριστικά είναι τα παρακάτω:

**Αυστηρότητα Κινήσεων** Οι κινήσεις των παραπάνω πολεμικών τεχνών που έχουν αγωνιστικό χαρακτήρα είναι απαραίτητο να εκτελούνται σύμφωνα με τους αυστηρούς κα-



(α')



(β')



(γ')



(δ')



(ε')

Σχήμα 3.2: Διάφορα είδη πολεμικών τεχνών. (α') Με λακτίσματα, αερόβια στριφογυριστή κλωτσιά. (β') Με λαβές, πέταγμα *judo*. (γ') Ιαπωνική τέχνη του σπαθιού (*kenjutsu*). (δ') Σύγχρονη ξιφομαχία (*fencing*). (ε') Στρατιωτικές εφαρμογές ακινητοποίησης.

νόνες τεχνικής που υπαγορεύει η εκάστοτε πολεμική τέχνη. Τέτοιοι κανόνες αφορούν τη σωστή θέση ολόκληρου του σώματος και των επιμέρους άκρων, τον σωστό προσανατολισμό τους, την ταχύτητα εκτέλεσης των κινήσεων και άλλες τεχνικές λεπτομέρειες. Οι αθλητές που τις εξασκούν μάλιστα, προκειμένου να προαχθούν στο επόμενο επίπεδο πρέπει να εξεταστούν μπροστά σε εγκεκριμένη επιτροπή κριτών στην εκτέλεση ακολουθιών κινήσεων ανάλογης δυσκολίας με το επίπεδο το οποίο επιθυμούν να φτάσουν. Σε αυτή τη διαδικασία, ακόμα και μικρές αποκλίσεις από το σωστό πρότυπο εκτέλεσης μπορεί να

### 3.1. ΓΕΝΙΚΑ

---

προκαλέσει την απόρριψη του εξεταζόμενου. Συνεπώς, ανεξάρτητα από τον εκτελεστή των κινήσεων, αναμένεται συνέπεια ως προς ένα ενιαίο πρότυπο, γεγονός που καθιστά εφικτή τη δημιουργία μοντέλων για τις διαφορετικές κατηγορίες κινήσεων.

**Συγκεκριμένο Εύρος Κίνησης** Κατά την εκτέλεση των ακολουθιών κινήσεων των συγκεκριμένων πολεμικών τεχνών, οι αθλητές δεν επιτρέπεται να υπερβαίνουν τα όρια μιας συγκεκριμένης περιοχής ακολουθώντας ένα συγκεκριμένο μονοπάτι. Αυτή η περιοχή είναι συνήθως τετράγωνη ή κυκλική και είναι αυστηρά οριοθετημένη. Μάλιστα, αντίστοιχης σημασίας με την αυστηρή εκτέλεση των κινήσεων σύμφωνα με τα πρότυπα τεχνικής είναι και η αυστηρή τήρηση της συγκεκριμένης διαδρομής πάνω στην οποία είναι αναγκαίο να παραμένει μια ακολουθία κινήσεων, χωρίς να ξεφεύγει από τα όρια της εν λόγω περιοχής. Ως επακόλουθο αυτού του χαρακτηριστικού, για αρχή είναι ευκολότερη η παρακολούθηση ολόκληρης της ακολουθίας κίνησης από την κάμερα λήψης, αλλά και λόγω του συγκεκριμένου μονοπατιού που πρέπει να υπακούει η καθεμία, είναι ευκολότερη η μοντελοποίηση της εξέλιξης στο χώρο της ακολουθίας.

**Εκρηκτικότητα Κινήσεων** Σε αυτά τα είδη κινήσεων έχουμε να κάνουμε με απότομες μεταβολές στη θέση και τον προσανατολισμό των άκρων και του κορμού του ανθρωπίνου σώματος. Οι αθλητές μάλιστα πρέπει να εκτελούν τις κινήσεις τους με συγκεκριμένη εκρηκτικότητα ώστε να θεωρούνται καλού επιπέδου. Με βάση αυτό το στοιχείο, η εξαγωγή χαρακτηριστικών για αυτές τις κινήσεις είναι πιο αποδοτικές, αφού απότομες μεταβολές τέτοιου είδους είναι ευκολότερο να παρατηρηθούν με οπτικά μέσα, δεδομένης μιας αρκετά καλής ταχύτητας λήψης. Ως αποτέλεσμα, μπορούν να δημιουργηθούν μοντέλα με καλύτερη αξιοπιστία που αποφαίνονται με μεγαλύτερη σιγουριά για το αποτέλεσμα.

Πέρα από τα παραπάνω χαρακτηριστικά που προωθούν τις εφαρμογές στις πολεμικές τέχνες, ένα ακόμα γεγονός που ενθαρρύνει την δημιουργία μοντέλων της σωστής εκτέλεσης των ακολουθιών κινήσεων είναι πως οι εφαρμογές αυτές μπορούν να χρησιμοποιηθούν, ύστερα ίσως από κάποια βελτιστοποίηση και τη δημιουργία ενός περιβάλλοντος αλληλεπίδρασης για πιο εύκολη χρήση, ως βοήθημα εκμάθησης της συγκεκριμένης πολεμικής. Ως συνέπεια, οι αθλητές θα μπορούν να χρησιμοποιούν τέτοιου είδους εφαρμογές για τη βελτίωση του επιπέδου της τεχνικής τους κατά τη διάρκεια των προσωπικών τους προπονήσεων σαν υποβοηθητικό εργαλείο στην καθοδήγηση από τον προπονητή.

#### 3.1.2 Εφαρμογές

Η εφαρμογή των τεχνικών της αναγνώρισης ανθρωπίνων κινήσεων σε προβλήματα που αφορούν πολεμικές τέχνες είναι αλήθεια πως δεν είναι τόσο διαδεδομένη, αν και πρόσφατες εξελίξεις στο χώρο των βιντεοπαιχνιδιών, όπως με την ευρεία χρήση της τεχνολογίας *Kinect* για τον υπολογισμό βάθους στις εικόνες και τα βίντεο με χρήση υπέρυθρης τεχνολογίας, έχουν στρέψει αρκετά το ενδιαφέρον στις εφαρμογές αυτού του πεδίου. Μια εκτεταμένη μελέτη πάνω σε ένα είδος πολεμικής τέχνης είχε γίνει από τον Becker κατά την

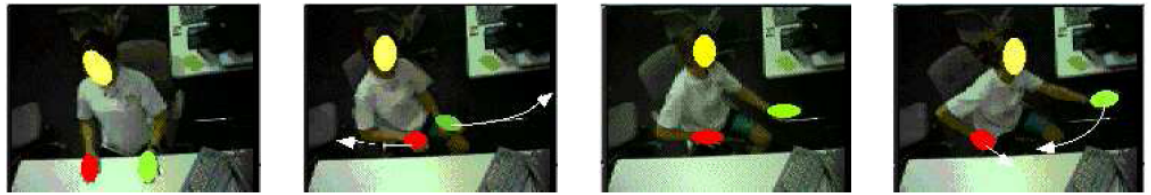
εκπόνηση της μεταπτυχιακής του εργασίας [46]. Σε αυτήν εισάγει ένα σύστημα αναγνώρισης και εκπαίδευσης με παροχή σχολίων για χειρονομίες της πολεμικής τέχνης T'ai Chi. Το σύστημα αυτό το ονομάζει *Sensei*, δηλαδή δάσκαλος στα Ιαπωνικά, και για τη λειτουργία του χρησιμοποιούνται τεχνικές *blobs* για τον εντοπισμό των χεριών. Για την εξαγωγή αμετάβλητων χαρακτηριστικών για τρισδιάστατη αναπαράσταση των εν λόγω χειρονομιών αυτός και οι ομάδα του, Campell, Becker et al., είχαν πραγματοποιήσει μια επιπλέον ερευνητική προσπάθεια [23], στην οποία κατέληξαν πως τα χαρακτηριστικά ταχύτητας, καρτεσιανής ή πολικής ανάλογα με το είδος των κινήσεων, αποτελούν τα πιο αποδοτικά για την αναπαράσταση των χειρονομιών. Για τη δημιουργία των μοντέλων επέλεξαν και αυτοί, όπως και πολλοί άλλοι ερευνητές του πεδίου τα HMMs. Της ίδιας σχολής έχουμε και άλλες ερευνητικές προσπάθειες [45, 15], στις οποίες γίνεται προσπάθεια βελτίωσης των μοντέλων για την αναγνώριση των ακολουθιών κινήσεων του T'ai Chi, όπως με τη δημιουργία αλληλένδετων HMM (coupled HMMs) ή γραφικών μοντέλων. Βέβαια πάλι, η ανίχνευση αφορά μόνο τα ανθρώπινα χέρια και όχι ολόκληρο το σώμα.

Μια προσπάθεια που ασχολήθηκε με κινήσεις από καράτε ήταν αυτή των Ig και Giese [16]. Σε αυτή την προσέγγιση, οι ακολουθίες κινήσεων μοντελοποιούνται βάσει ενός ιεραρχικού μοντέλου, στο κατώτερο επίπεδο του οποίου βρίσκονται μικρά τμήματα κινήσεων που καλούνται *αρχέτυπα κίνησης* (*motion primitives*). Με τον συνδυασμό των παραπάνω αρχετύπων δημιουργούνται στο ανώτερο επίπεδο ολοκληρωμένο μοντέλα των κινήσεων, ενώ η ανίχνευση γίνεται με βάση τις αρχές του Δυναμικού Προγραμματισμού (*Dynamic Programming*). Βέβαια, τα μοντέλα που κατασκευάζονται τελικά είναι αρκετά εξειδικευμένα, αφού τα αρχέτυπα κινήσεων που χρησιμοποιούνται εξ αρχής είναι συγκεκριμένα, με αποτέλεσμα να μην μπορούν να γενικευθούν. Τέλος, έχουν υπάρξει και μερικές ενδιαφέρουσες προσπάθειες, όπως αυτή της ομάδας Φινλανδών Hamalainen et al. [17], στην οποία αναγνωρίζεται ένα αρκετά μεγάλο πλήθος από πολεμικές τέχνες με βάση την αναπαράσταση του περιγράμματος τη ανθρώπινης φιγούρας όπως αυτή εξάγεται από τη μεταβολή της οπτικής ροής κατά τη διάρκεια της κίνησης. Το αποτέλεσμα χρησιμοποιείται σε ένα σύστημα εικονικής αναπαράστασης των ακολουθιών κινήσεων, ώστε να μπορεί να αναπαραστήσει ακόμα και εικονικές μάχες. Βέβαια, η έλλειψη λεπτομερειών αποτρέπει τη σαφή κατηγοριοποίηση των κινήσεων και απλά εστιάζει στην παρακολούθηση των κινήσεων.

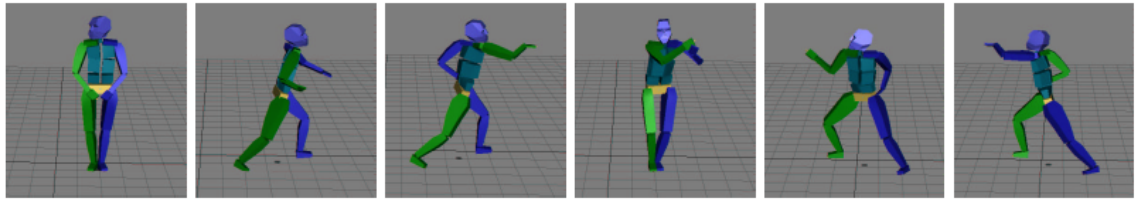
## 3.2 Καράτε

Το *καράτε* (*karate*) είναι ένα από τα πλέον γνωστά είδη πολεμικών τεχνών με μεγάλη απήχηση τόσο σε μικρότερα όσο και μεγαλύτερα ηλικιακά κοινά. Αυτή του η διάδοση ενισχύει το ενδιαφέρον μας να ασχοληθούμε με προβλήματα αναγνώρισης κινήσεων που περιέχουν κινήσεις καράτε.

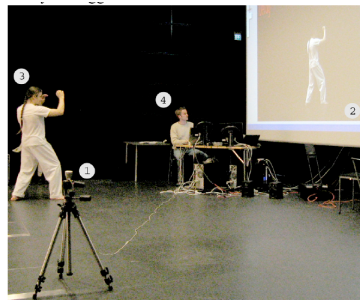
### 3.2. ΚΑΡΑΤΕ



(α')



(β')



(γ')

Σχήμα 3.3: Εφαρμογές σε διάφορα είδη πολεμικών τεχνών. (α') Αναγνώριση ακολουθίας κινήσεων Tai chi, όπως στο σχήμα του [15]. (β') Χρήση αρχετύπων κινήσεων για κατασκευή μοντέλων, όπως στο σχήμα του [16]. (γ') Αναγνώριση πολεμικών τεχνών για χρήσεις εικονικής πραγματικότητας, όπως στο σχήμα του [17].

#### 3.2.1 Ιστορία

Η τέχνη του καράτε δημιουργήθηκε στα νησιά Ryukyu (σημερινή Okinawa, Ιαπωνίας) από ενδογενείς μεθόδους πάλης νωρίτερα του 19ου αιώνα, ενώ περιλαμβάνει χτυπήματα όπως μπουνιές, λακτίσματα, χτυπήματα με τα γόνατα, τους αγκώνες, αλλά και τεχνικές ανοιχτής παλάμης, όπως τα χέρια-μαχαίρια (knife-hands). Σε κάποια στυλ διδάσκονται επίσης και λαβές, κλειδώματα, πετάγματα και χτυπήματα σε ζωτικά σημεία. Με την πάροδο του χρόνου, το καράτε έγινε ένα από τα δημοφιλέστερα είδη πολεμικών τεχνών και σήμερα εξασκείται από περισσότερους των 50 εκατομμύριων ανθρώπους παγκοσμίως, είτε για αυτο-βελτίωση, για πολιτιστικούς λόγους, για αυτο-άμυνα ή σαν άθλημα. Ο άνθρωπος που εξασκεί την πολεμική τέχνη του καράτε αποκαλείται *karateka*. Στο αγωνιστικό/αθλητικό μέρος του καράτε, η εκπαίδευση χωρίζεται σε τρία είδη:



- Το πρώτο είδος καλείται *Kihon* και αφορά τις ομαδικές ασκήσεις που γίνονται από ομάδες karateka για την εκμάθηση τεχνικών ή συνδυασμών τεχνικών.
- Το δεύτερο είδος ονομάζεται *Kata*, που κυριολεκτικά σημαίνει 'σχήμα' ή 'μοντέλο', και είναι μια αυστηρά καθορισμένη ακολουθία κινήσεων που αναπαριστά ποικίλες επιθετικές και αμυντικές στάσεις.
- Το τρίτο είδος καλείται *Kumite*, που σημαίνει 'συνάντηση των χεριών', και περιλαμβάνει στην ουσία εικονική μάχη μεταξύ δύο αθλητών, όπου οι αθλητές εκτελούν συνδυασμένες ακολουθίες κινήσεων ή μάχονται με πόντους για τη νίκη.

Οι βαθμίδες γνώσης καθορίζονται από ένα σύστημα που καλείται *Kyu/Dan*. Το πρώτο στάδιο, *Kyu*, αποτελείται από 'ζώνες', στο οποίο ανάλογα με το χρώμα της ζώνης αντιστοιχεί και το επίπεδο του karateka. Οι βαθμίδες-χρώματα από τον πιο άπειρο ως τον πιο έμπειρο έχουν ως εξής: άσπρο, κίτρινο, πορτοκαλί, μωβ, μπλε, πράσινο, καφέ και μαύρο (με μικρές παραλλαγές ίσως), ενώ μετά την επίτευξη της μαύρης ζώνης, ακολουθείται άλλο σύστημα ιεραρχίας γνώσης που περιγράφεται από το πόσα *Dan* έχει ο κάθε karateka. Για να προαχθούν σε μία επίσημη βαθμίδα-επίπεδο, οι karatekas πρέπει να επιδείξουν την ικανότητά τους στην εκτέλεση του kata που απαιτείται για το συγκεκριμένο επίπεδο. Στο *Kyu* στάδιο η προσοχή επιστάνει στη σωστή στάση του σώματος, στην ισορροπία και τον συντονισμό κινήσεων, ενώ η ταχύτητα και η δύναμη προστίθενται σε υψηλότερα επίπεδα. Οι σχολές καράτε, δηλαδή τα σύνολα ανθρώπων που ακολουθούν μια συγκεκριμένη μορφή κινήσεων καράτε και τους κανόνες που αυτή υπαγορεύει, είναι πλέον πολλές ανά τον κόσμο. Μία από τις παλαιότερες και πιο 'αυθεντικές', με την έννοια πως δεν έχει προσμίξεις από άλλα είδη πολεμικών τεχνών, σχολές είναι το *Shotokan karate*, με το οποίο θα ασχοληθούμε κατά την πειραματική διαδικασία της Διπλωματικής Εργασίας.



(α')

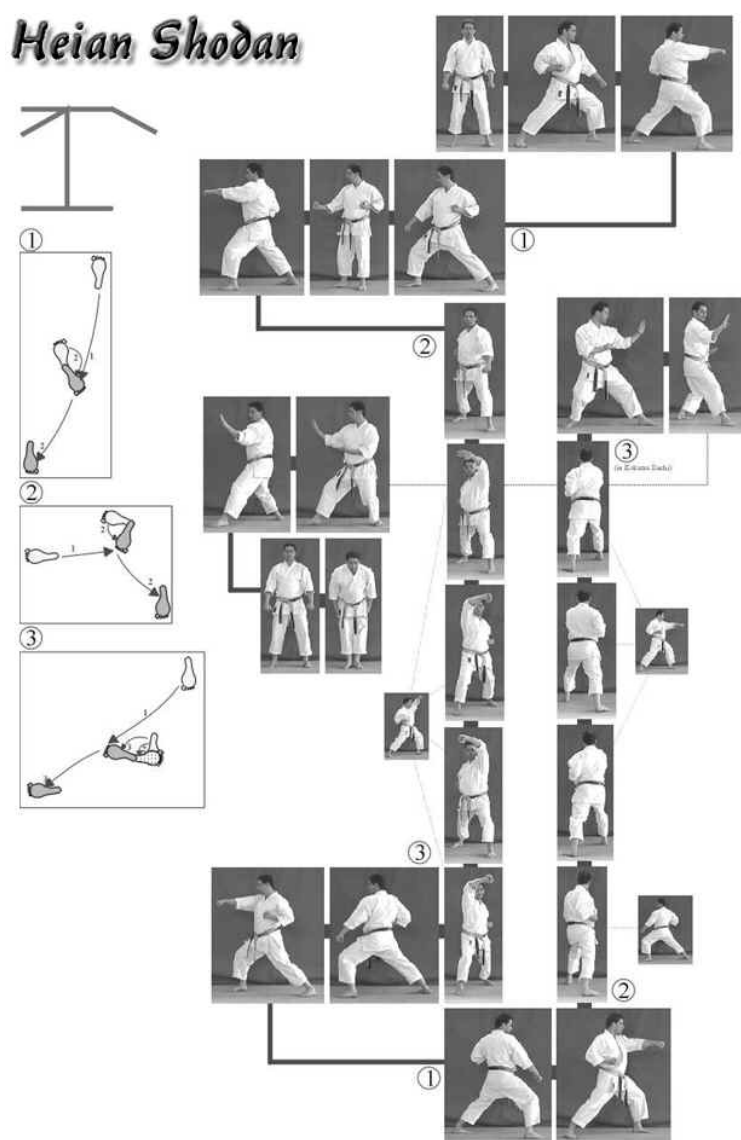


(β')

Σχήμα 3.4: Είδη εκπαίδευσης αθλητικού καράτε. (α') *Kihon*, ομαδική εκπαίδευση. (β') *Kumite*, αναμέτρηση μεταξύ δύο αθλητών.

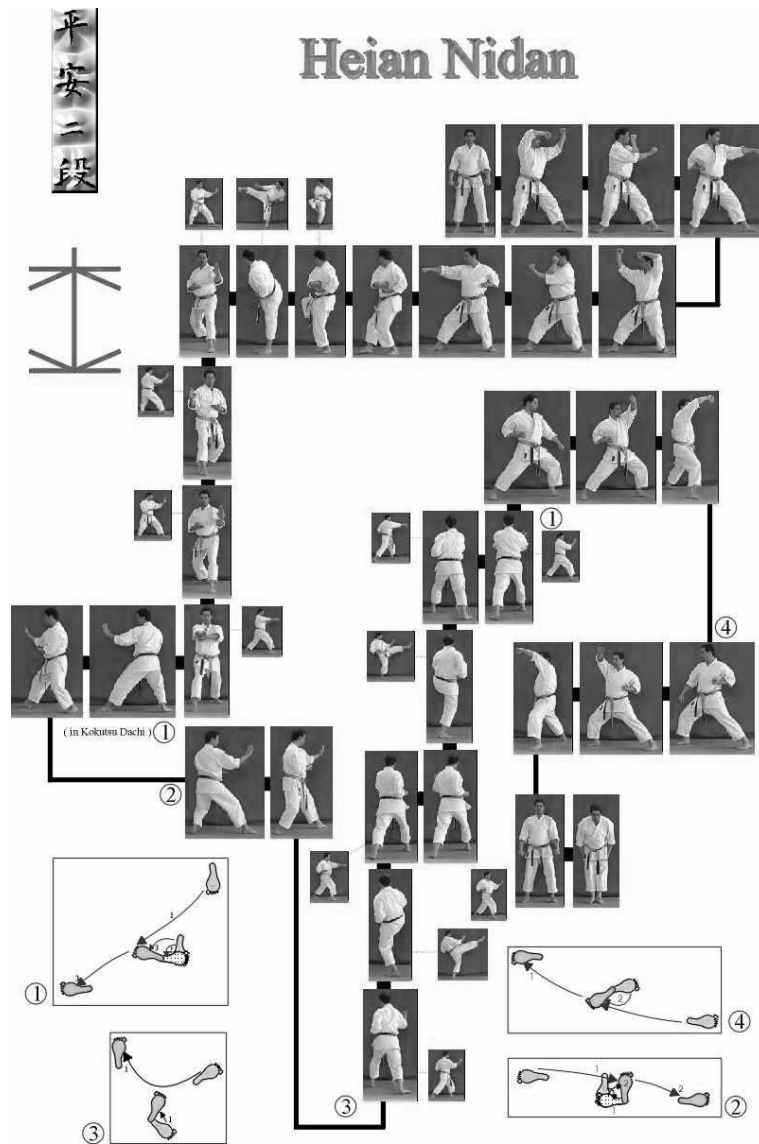
### 3.2.2 Κινήσεις

Το καράτε ανήκει στην κατηγορία των άοπλων πολεμικών τεχνών με αθλητικό προσανατολισμό που αναφέραμε στην προηγούμενη παράγραφο. Συνεπώς, οι κινήσεις του, και ειδικότερα οι ακολουθίες κινήσεων που εκτελούνται ατομικά για επίδειξη της τεχνικής του *karateka*, δηλαδή τα *katas*, έχουν όλα τα θετικά χαρακτηριστικά που τις κάνουν υποψήφιες για εφαρμογή μεθόδων από την ανίχνευση και κατηγοριοποίηση ανθρωπίνων κινήσεων, δηλαδή αυστηρότητα στην εκτέλεση των κινήσεων, συγκεκριμένο μονοπάτι για την ακολουθία κινήσεων, ορισμένο χώρο μέσα στον οποίο πρέπει να εκτελεστεί η ακολουθία και εκρηκτικότητα των κινήσεων. Συγκεκριμένα για το Shotokan karate, θα αναλύσουμε τις ακολουθίες κινήσεων για τα 5 πρώτα επίπεδα δεξιοτεχνίας. Αυτά τα *katas* έχουν το γενικό προσωνύμιο 'heian' που υποδεικνύει τον τίτλο του αρχάριου, ενώ στη συνέχεια αποτελούν αρίθμηση στα ιαπωνικά των *dan*.

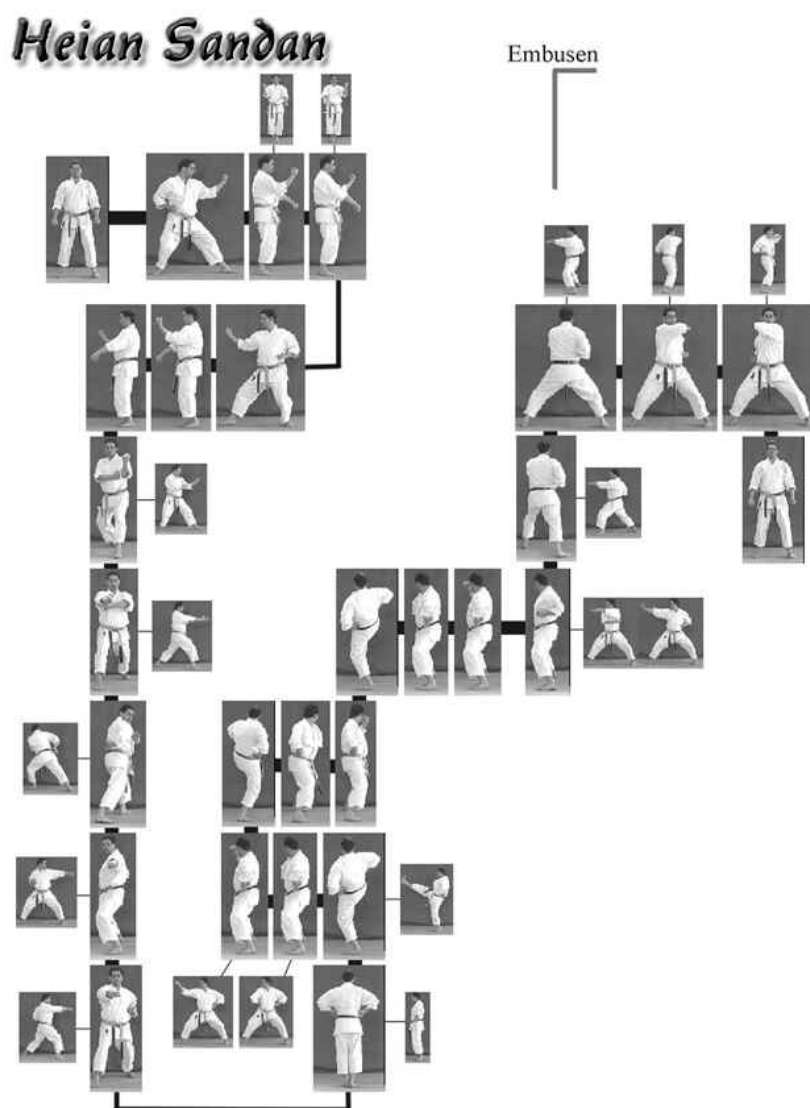


Σχήμα 3.5: Ακολουθία κινήσεων 1ου επιπέδου-Heian Shodan.

3.2. KAPATE



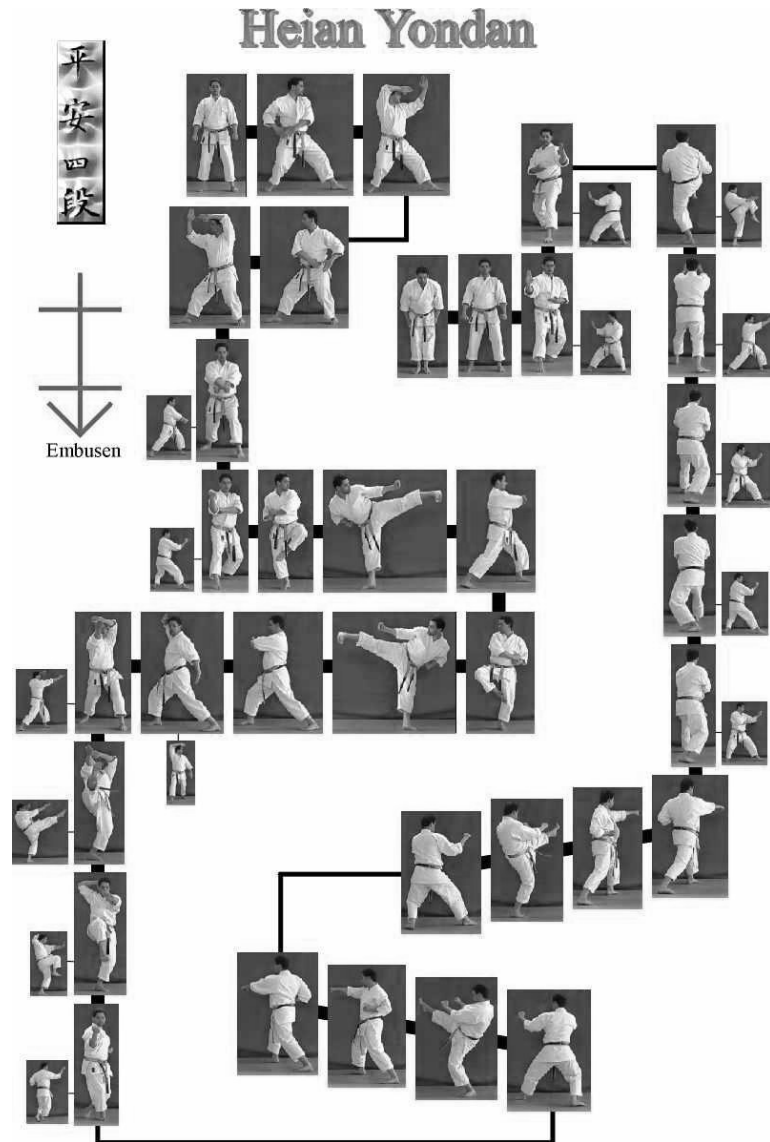
Σχήμα 3.6: Ακολουθία κινήσεων 2ου επιπέδου-Heian Nidan.



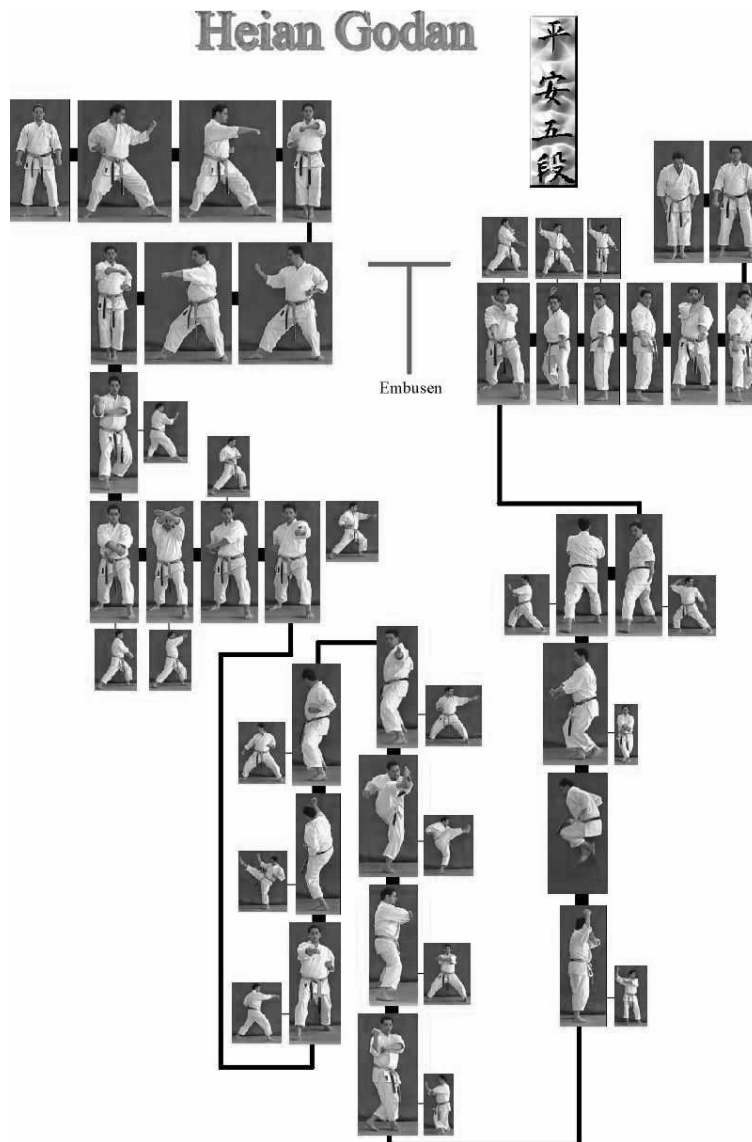
Σχήμα 3.7: Ακολουθία κινήσεων 3ου επιπέδου-Heian Sandan.

3.2. KAPATE

---



Σχήμα 3.8: Ακολουθία κινήσεων 4ου επιπέδου-Heian Yondan.



Σχήμα 3.9: Ακολουθία κινήσεων 5ου επιπέδου-Heian Godan.

### 3.2. *KAPATE*

---

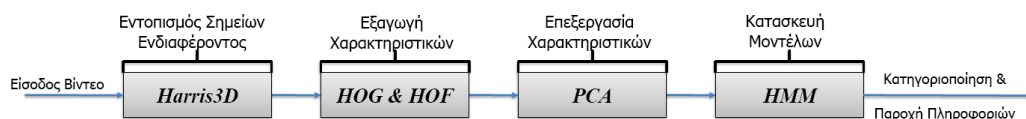


# Κεφάλαιο 4

## Μέθοδοι και Εργαλεία

### 4.1 Η προσέγγισή μας

Στα πλαίσια αυτής της ερευνητικής προσπάθειας, προτείνουμε ένα νέο σύστημα ανίχνευσης και κατηγοριοποίησης ανθρωπίνων κινήσεων, το οποίο έχει τη δυνατότητα να παρέχει πληροφορίες, ύστερα από την κατηγοριοποίηση, για την ποιότητα εκτέλεσης της κίνησης με βάση το αποθηκευμένο μοντέλο. Το σύστημα δέχεται στην είσοδό του βίντεο από μαγνητοσκοπημένες ακολουθίες ανθρωπίνων κινήσεων, εντοπίζει σε αυτή την ακολουθία από καρέ τα τρισδιάστατα σημεία ενδιαφέροντος και εξάγει από τις περιοχές που τα περιβάλλουν χαρακτηριστικά με χρήση Ιστογραφικών μεθόδων. Στη συνέχεια, το σύνολο των χαρακτηριστικών μορφοποιείται κατάλληλα με ανάλυση πρωτευουσών συνιστωσών και τροφοδοτεί HMM μοντέλα, που έχουν εκπαιδευτεί με βάση τις κατηγορίες που επιθυμούμε, ώστε τέλος, το σύστημα αποφαινεται για το είδος της κίνησης του σήματος εισόδου και παρέχει πληροφορίες για την ομοιότητα με το πρότυπο. Οι λεπτομέρειες για τις ακριβείς παραμέτρους των μεθόδων αυτών θα περιγραφούν παρακάτω.



Σχήμα 4.1: Το σύστημά μας

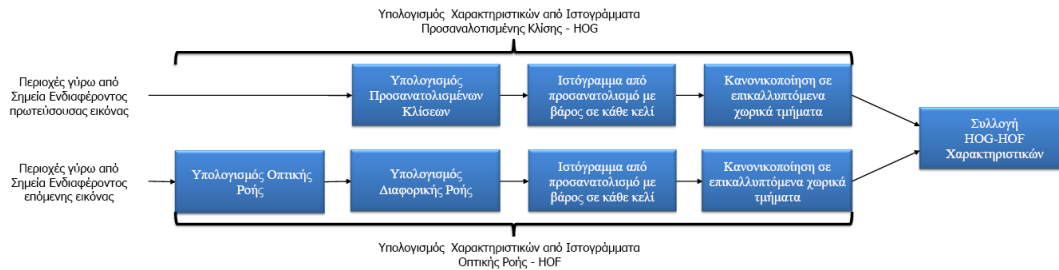
## 4.2 Εξαγωγή Χαρακτηριστικών

Με την εισαγωγή των βίντεο στην είσοδο του συστήματός μας, οι ακολουθίες από καρτέ επεξεργάζονται σύμφωνα με τον αλγόριθμο *Harris3D* που προτάθηκε από τους Laptev και Lindeberg [11] και αναλύσαμε στο Κεφ. 2 για την εξαγωγή *Χωροχρονικών Σημείων Ενδιαφέροντος (Space-Time Interest Points-STIP)*. Σε κάθε σημείο του βίντεο υπολογίζεται η μήτρα δευτέρων ροπών (second moment matrix) (2.8) με τη χρήση ανεξάρτητων τιμών για τις χωρικές και χρονικές κλίμακες  $\sigma$  και  $\tau$ , μιας διαχωρίσιμης συνάρτησης Γκαουσιανής ομαλοποίησης  $g$  και των χωροχρονικών παραγώγων  $\nabla D$ . Οι τελικές θέσεις των χωροχρονικών σημείων ενδιαφέροντος δίνονται από τα τοπικά μέγιστα της (2.9), ενώ για την επιλογή της βέλτιστης χωροχρονικής κλίμακας γίνεται χρήση της (2.11), αλλά σε μια απλοποιημένη μορφή όπου δοκιμάζονται συνδυασμοί χωροχρονικών κλιμάκων και επιλέγεται ο καλύτερος συνδυασμός. Με αυτόν τον τρόπο έχουμε ίδια ή και καλύτερα αποτελέσματα και σημαντική επιτάχυνση του αλγορίθμου σε σημείο να πλησιάζει την ταχύτητα του βίντεο. Για τις παραμέτρους της μεθόδου αυτής χρησιμοποιούμε  $k = 0.0005$  και σαν επιλογές για τις κλίμακες χώρου και χρόνου  $\sigma^2 = 4, 8, 16, 32, 64, 128$  και  $\tau^2 = 2, 4$ .

Έπειτα από την επιλογή των σημείων ενδιαφέροντος για την εξαγωγή των χαρακτηριστικών μας χρησιμοποιούμε την περιγραφή των συνδυασμένων Ιστογραμμάτων Προσανατολισμένης Κλίσης και Οπτικής Ροής (HOG - HOF descriptors) στις γειτονίες των σημείων ενδιαφέροντος, όπως εισηχθή από τους Laptev et al. στο [37]. Για αυτό τον συνδυασμό τα μεγέθη της περιγραφής της γειτονιάς του σημείου ενδιαφέροντος καθορίζονται ως  $\Delta_x(\sigma) = \Delta_y(\sigma) = 18\sigma$ ,  $\Delta_t(\tau) = 8\tau$ , όπου  $\Delta_x(\sigma)$ ,  $\Delta_y(\sigma)$  είναι τα χωρικά μεγέθη συναρτήσει της χωρικής κλίμακας  $\sigma$  και  $\Delta_t(\tau)$  είναι η χρονική έκταση συναρτήσει της χρονικής κλίμακας  $\tau$ . Κάθε όγκος χωρίζεται σε ένα  $n_x \times n_y \times n_t$  πλέγμα από κελιά, ενώ για κάθε κελί υπολογίζονται τα Ιστογράμματα Προσανατολισμένης Κλίσης 4-κατευθύνσεων (4-bin HOG) και τα Ιστογράμματα Οπτικής Ροής 5-κατευθύνσεων (5-bin HOF). Έπειτα από κανονικοποίηση τα στοιχεία των ιστογραμμάτων συνθέτουν το διάνυσμα των χαρακτηριστικών μας. Για την δική μας εφαρμογή χρησιμοποιούμε παραμέτρους  $n_x, n_y = 3, n_t = 2$ . Κατά συνέπεια, από κάθε σημείο ενδιαφέροντος έχουμε τον υπολογισμό ενός διανύσματος χαρακτηριστικών με  $3 \times 3 \times 2 \times 4 = 72$  HOG και  $3 \times 3 \times 2 \times 5 = 90$  HOF, δηλαδή ένα διάνυσμα διάστασης 3 (χωροχρονική θέση σημείου ενδιαφέροντος - x,y,t) + 2 (χωροχρονικές κλίμακες -  $\sigma^2, \tau^2$ ) + 72 + 90 = 167.

## 4.3 PCA

Λόγω της μεγάλης διάστασης του διανύσματος χαρακτηριστικών, πριν το χρησιμοποιήσουμε για την εκπαίδευση των μοντέλων μας, θα επιχειρήσουμε να μειώσουμε το μέγεθος της διάστασης με περαιτέρω επεξεργασία των χαρακτηριστικών, ώστε να μειωθεί και η πολυπλοκότητα του προβλήματος που επιβαρύνει την εφαρμογή τόσο υπολογιστικά όσο και χρονικά. Για αυτό το λόγο, θα εφαρμόσουμε *Ανάλυση Πρωτευουσών Συνιστωσών (Principal Components Analysis-PCA)* στα χαρακτηριστικά των HOG/HOF, ώστε να επιλέξουμε τις συνιστώσες που αναπαριστούν το πρόβλημα χωρίς να χάνεται σημαν-



Σχήμα 4.2: Στάδια εξαγωγής χαρακτηριστικών από Ιστογράμματα Προσανατολισμένης Κλίσης-HOG και Οπτικής Ροής-HOF, μετά τον εντοπισμό χωροχρονικών σημείων ενδιαφέροντος.

τικό κομμάτι της πληροφορίας.

Η Ανάλυση Πρωτεύουσών Συνιστωσών (PCA) είναι μια μαθηματική διαδικασία που χρησιμοποιεί έναν ορθογώνιο γραμμικό μετασχηματισμό πάνω στα δεδομένα μας, ώστε αυτά να μετασχηματιστούν σε ένα νέο σύστημα συντεταγμένων και να μετατραπεί ένα σύνολο παρατηρήσεων απο πιθανά συσχετισμένες μεταβλητές σε ένα σύνολο τιμών από ασυσχέτιστες μεταβλητές που καλούνται πρωτεύουσες συνιστώσες. Ο αριθμός των πρωτεύουσών συνιστωσών είναι μικρότερος ή ίσος με τον αριθμό των αρχικών μεταβλητών. Αυτός ο μετασχηματισμός ορίζεται κατά τέτοιο τρόπο ώστε η πρώτη πρωτεύουσα συνιστώσα να έχει τη μεγαλύτερη δυνατή διακύμανση (variance) και κάθε επόμενη συνιστώσα με τη σειρά της έχει τη μεγαλύτερη δυνατή διακύμανση με τον περιορισμό να είναι ορθογώνια (άρα και ασυσχέτιστη) προς τις προηγούμενες συνιστώσες. Συχνά, η εφαρμογή της ανάλυσης αυτής μπορεί να θεωρηθεί ως αποκαλυπτική της εσωτερικής δομής των δεδομένων με τέτοιο τρόπο ώστε να εξηγείται καλύτερα η διακύμανση μεταξύ των δεδομένων. Έτσι, για ένα σύνολο πολλών μεταβλητών που απεικονίζεται σε ένα χώρο δεδομένων υψηλής διάστασης, η PCA ανάλυση μπορεί να παράσχει στο χρήστη μια χαμηλότερης διάστασης εικόνας, μια προβολή του αντικειμένου από την οπτική γωνία που παρέχει τις περισσότερες πληροφορίες. Αυτό γίνεται με τη χρήση μόνο μερικών από τις πρώτες πρωτεύουσες συνιστώσες, ώστε να μειωθεί η διαστασιμότητα των μετασχηματισμένων δεδομένων.

Το αποτέλεσμα αυτό μπορεί να επιτευχθεί είτε με διαδικασία αποδόμησης σε ιδιοτιμές (*eigenvalue decomposition*) του πίνακα συνδιακύμανσης των δεδομένων ή με διαδικασία αποδόμησης μοναδιαίας τιμής (*singular value decomposition*) του πίνακα διακύμανσης των δεδομένων, συνήθως ύστερα από κεντράρισμα των μέσων των δεδομένων για κάθε στοιχείο. Έστω ένας πίνακας δεδομένων  $X^T$  με μηδενικό εμπειρικό μέσο, όπου η κάθε σειρά  $n$  αναπαριστά μια διαφορετική επανάληψη του πειράματος και η κάθε στήλη  $m$  αναπαριστά καταγεγραμμένες παρατηρήσεις μιας μεταβλητής. Η αποδόμηση μοναδιαίας τιμής του  $X$  είναι  $X = W\Sigma V^T$ , όπου ο  $m \times m$  πίνακας  $W$  είναι ο πίνακας των ιδιοδιανυσμάτων του  $XX^T$ , ο πίνακας  $\Sigma$  είναι ένας  $m \times n$  διαγώνιος πίνακας από μη-αρνητικούς

### 4.3. PCA

πραγματικούς αριθμούς στη διαγώνιο και ο  $n \times n$  πίνακας  $V$  είναι ο πίνακας των ιδιοδιανυσμάτων του  $X^T X$ . Ο PCA μετασχηματισμός που διατηρεί την αρχική διαστασιμότητα, δηλαδή χωρίς να αφαιρεθούν στήλες και να έχουμε απώλεια πληροφορίας από τον αρχικό πίνακα αν αντιστρέψουμε το μετασχηματισμό, δίνεται από τη σχέση:

$$\begin{aligned} Y^T &= X^T W \\ &= V \Sigma^T W^T W \\ &= V \Sigma^T \end{aligned} \quad (4.1)$$

Ο πίνακας  $W$  των μοναδιαίων διανυσμάτων του  $X$  είναι ισοδύναμα ο πίνακας  $W$  των ιδιοδιανυσμάτων του πίνακας των παρατηρούμενων συνδιακυμάνσεων  $C = X X^T$ ,

$$X X^T = W \Sigma \Sigma^T W^T \quad (4.2)$$

Εάν θέλουμε να την αναπαράσταση μειωμένης διάστασιμότητας, μπορούμε να προβάλουμε το  $X$  στον μειωμένο χώρο που ορίζεται από τα πρώτα  $L$  μοναδιαία διανύσματα,  $W_L$ :

$$Y = W_L^T X = \Sigma_L V^T \quad (4.3)$$

όπου  $\Sigma_L = I_{L \times m} \Sigma$  με  $I_{L \times m}$  τον  $L \times m$  ορθογώνιο μοναδιαίο πίνακα.

Στα πλαίσια της δικής μας εργασίας, εφαρμόζουμε την PCA ανάλυση πάνω στο σύνολο των δεδομένων που έχουμε για κάθε δείγμα, θεωρώντας τα ως τις διαφορετικές επαναλήψεις του πειράματος, με απώτερο σκοπό να ανακαλύψουμε αν η μείωση της διαστασιμότητας του προβλήματος δεν επηρεάζει την απόδοση του συστήματος. Συνεπώς, επιχειρούμε να μειώσουμε τη διάσταση του διανύσματος χαρακτηριστικών από 167, προσέχοντας όμως να μην εξαιρεθεί σημαντική πληροφορία. Για αυτό το σκοπό, αποφασίσαμε να κρατήσουμε ανέπαφα τα 5 πρώτα στοιχεία του διανύσματος χαρακτηριστικών που αναφέρονται στη χωροχρονική θέση και κλίμακα του σημείου ενδιαφέροντος και να ασχοληθούμε με την εφαρμογή της μεθόδου στα επιμέρους 72 HOG και 90 HOF χαρακτηριστικά. Ως αποτέλεσμα, σαν πίνακα δεδομένων μας  $X$  θεωρούμε τον πίνακα που έχει σαν στήλες τα HOG ή HOF χαρακτηριστικά και σαν γραμμές τα συνενωμένα το ένα κάτω από το άλλο σύνολα (ως διαδοχικές επαναλήψεις του πειράματος) από διαδοχικά σημεία ενδιαφέροντος που έχουν αναγνωριστεί σε κάθε δείγμα, δηλαδή έναν πίνακα με την εξής μορφή:

$$[(\text{samples})(\text{Int.Points/sample}) \times \text{HOG/HOF features}] \quad (4.4)$$

Για την υλοποίηση του PCA, δημιουργούμε ένα πρόγραμμα με χρήση του Matlab για πιο εύκολη επεξεργασία των μεγάλων πινάκων που δημιουργούνται. Στο κέντρο του προγράμματος αυτού βρίσκεται η εντολή *princomp*:

$$[COEFF, SCORE, latent] = \text{princomp}(X) \quad (4.5)$$

όπου  $X$  είναι ο  $n \times m$  πίνακας δεδομένων μας,  $COEFF$  είναι ο  $m \times m$  πίνακας με τους συντελεστές για τις πρωτεύουσες συνιστώσες με σειρά μειούμενης διακύμανσης,  $SCORE$  είναι ο  $n \times m$  πίνακας με τα μετασχηματισμένα δεδομένα, όπου πάλι έχουμε σειρά μειούμενης διακύμανσης και  $latent$  το διάνυσμα που περιέχει τις ιδιοτιμές του πίνακα διακύμανσης του  $X$ . Από επεξεργασία του τελευταίου πίνακα καταλαβαίνουμε τι

ποσοστό διακύμανσης-πληροφορίας κρατούμε με την αποκοπή των  $m - L$  στηλών. Για την εφαρμογή μας, θεωρήσαμε πως ένα σχετικά καλό ποσοστό διακύμανσης(*variance*)-πληροφορίας για να κρατήσουμε, ώστε να επιτευχθεί παράλληλα και κάποια μείωση της διαστασιμότητας, είναι **90%**. Συνεπώς, το πρόγραμμά μας υπολογίζει πόσες στήλες  $L$  μας εξασφαλίζουν τη διατήρηση αυτού του ποσοστού πληροφορίας και επιλέγουμε τόσες πρώτες στήλες από τον πίνακα μετασχηματισμένων δεδομένων *SCORE*. Τέλος, ξεχωρίζουμε τα δεδομένα στα ξεχωριστά δείγματα για τη χρήση τους στα μοντέλα.

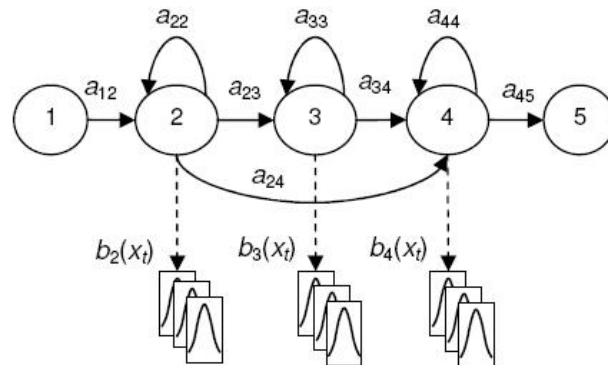
## 4.4 HMM

Τελευταίο στάδιο για την ολοκλήρωση του συστήματος αποτελεί η δημιουργία των μοντέλων μας και η εκπαίδευσή τους ώστε να μπορούν να αναγνωρίζουν τις κατηγορίες που επιθυμούμε να διαχωρίσουμε. Λόγω της δυναμικότητας των προβλημάτων που επιθυμούμε να αντιμετωπίσουμε και της συγκεκριμένης ακολουθίας από κινήσεις που περιέχει κάθε δείγμα που θα επεξεργαστούμε, αφού επεξεργαζόμαστε σκηνές από πολεμικές τέχνες, επιλέγουμε να δουλέψουμε με τα *Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models-HMM)*, τα οποία αναλύσαμε στο Κεφ. 2. Το κάθε μοντέλο που δημιουργείται έχει έναν αριθμό από καταστάσεις και σε κάθε κατάσταση έχουμε ένα μείγμα από διαφορετικές Γκαουσιανές που περιγράφουν την κάθε μεταβλητή του διανύσματος χαρακτηριστικών μας. Φυσικά, οι αριθμοί των καταστάσεων και των μειγμάτων σε κάθε κατάσταση διαφέρουν ανάλογα με το είδος της κίνησης που μοντελοποιούμε, οπότε δεν είναι κοινοί για όλα τα μοντέλα μας, αλλά γενικά χρησιμοποιούμε μοντέλα με καταστάσεις όχι λιγότερες από 3, αφού οι κινήσεις πολεμικών τεχνών σπάνια είναι απλούστερες από 3 διακριτά στάδια, και όχι περισσότερες από 7, αφού με περισσότερες καταστάσεις θα διαχωρίζονταν ακόμα και ενιαία τμήματα κινήσεων, πράγμα που δεν επιθυμούμε. Αντίστοιχα, τα μείγματα σε κάθε κατάσταση είναι 3-4, καθώς δεν έχουμε και δεν θέλουμε να έχουμε τόσο μεγάλες εσωτερικές διακυμάνσεις κατά την κατασκευή του προτύπου μιας αυστηρά καθορισμένης κίνησης.

Έπειτα από την κατασκευή των HMM είναι απαραίτητη η αρχικοποίηση των παραμέτρων των μοντέλων μας, δηλαδή η εύρεση αρχικών μέσων τιμών  $\mu$  και διακυμάνσεων  $\sigma^2$  για κάθε μεταβλητή του διανύσματος χαρακτηριστικών μας, ώστε αυτές οι τιμές να είναι κοντά στις μέσες τιμές των δειγμάτων με τα οποία θα εκπαιδευτεί. Σαν επακόλουθο βήμα έρχεται η εκπαίδευση των μοντέλων με τα διανύσματα χαρακτηριστικών που έχουν εξέλθει από κάθε σημείο ενδιαφέροντος κάθε δείγματος βίντεο που έχουμε επιλέξει για το στάδιο της εκπαίδευσης. Η επαναληπτική διαδικασία της εκπαίδευσης των παραμέτρων των HMM γίνεται μέσω του αλγορίθμου Baum Welch που αναπτύχθηκε στην Παρ.2.4.3, μέχρι να επέλθει σύγκλιση.

Μετά την εκπαίδευση, τα μοντέλα μας είναι έτοιμα για να αναγνωρίσουν τα είδη των κινήσεων από τα βίντεο στην είσοδο του συστήματος με χρήση της μεθόδου που περιγράψαμε στην Παρ.2.4.2. Μάλιστα, εφόσον κατά το στάδιο της κατηγοριοποίησης, τα εκπαιδευμένα HMM χρησιμοποιούν τον αλγόριθμο μεγίστης πιθανοφάνειας (maximum likelihood) για να αποφανθούν για την κατηγορία που ομοιάζει πιο πολύ με την κίνηση

#### 4.4. HMM



Σχήμα 4.3: Παράδειγμα HMM μοντέλου που μπορεί να χρησιμοποιηθεί στο σύστημά μας με 5 καταστάσεις και μείγμα 3 Γκουσιανών ανά κατάσταση.

στο βίντεο εισόδου, μπορούμε να τα χρησιμοποιήσουμε ώστε το σύστημά μας να χρησιμοποιείται και σαν **βοήθημα εκπαίδευσης** για τις πολεμικές τέχνες, όπως αναφέραμε στο Κεφ.3. Με την παροχή πληροφοριών προς το χρήστη που εξάγονται από την πιθανοφάνεια που υπολογίζεται, μπορεί ο χρήστης να καταλάβει πόσο κοντά είναι στην κίνηση-πρότυπο και να εξασκηθεί περισσότερο προς την κατεύθυνση της σύγκλισης με το πρότυπο. Για το σκοπό αυτό, εξάγουμε τη λογαριθμική πιθανότητα ( $\log$  probability) του κάθε βίντεο και επιχειρούμε να βαθμονομήσουμε σε μία κλίμακα 1-10 με άριστα το 10, την ομοιότητα με την κίνηση-πρότυπο, λαμβάνοντας υπ'όψιν πως η κίνηση που έχει αναγνωριστεί στην εκάστοτε κατηγορία είναι ήδη πάνω από το 0. Κατά αυτό τον τρόπο, ο προσδιορισμός της ακριβούς αντιστοιχίας γίνεται πειραματικά, δίνοντας προσοχή και στις περιπτώσεις λάθος κατηγοριοποίησης.

Για τη δημιουργία, την αρχικοποίηση, την εκπαίδευση και δοκιμή των μοντέλων μας, επιλέγουμε τη χρήση του εργαλείου HTK.

**HTK** Το HTK (Hidden Markov Model Toolkit) είναι ένα φορητό εργαλείο για την κατασκευή και διαχείριση κρυφών μαρκοβιανών μοντέλων που αναπτύχθηκε στο Εργαστήριο Μηχανικής Νοημοσύνης (Machine Intelligence Laboratory) του πανεπιστημίου του Cambridge. Ενώ κυρίως χρησιμοποιείται για έρευνα φωνητικής αναγνώρισης, χρησιμοποιείται και σε πληθώρα άλλων εφαρμογών που μπορεί να περιλαμβάνουν έρευνα σύνθεση λόγου, αναγνώρισης χαρακτήρων ή ακολουθιών DNA. Εμείς το χρησιμοποιούμε για εφαρμογές Ανίχνευσης και Κατηγοριοποίησης Ανθρωπίνων Κινήσεων καθώς είναι ένα αξιόπιστο και ευέλικτο εργαλείο με βιβλιοθήκες που διευκολύνουν κατά πολύ την εκπαίδευση και δοκιμή HMM, αλλά και την ερμηνεία των αποτελεσμάτων.

Το HTK χρησιμοποιεί ένα σύνολο από απλούστερες ή πιο σύνθετες εντολές με τη συνοδεία αρκετών παραμέτρων για την επεξεργασία των HMM μοντέλων. Τα δεδομένα πρέπει να είναι αποθηκευμένα σε ειδική μορφή για να επεξεργαστούν, αφού πρέπει να έχουν την κατάλληλη κεφαλή, που δίνει τις πληροφορίες του δείγματος, και ύστερα τα στοιχεία σε

δυναμική μορφή. Οι πιο βασικές εντολές, των οποίων γίνεται χρήση για αυτή την εφαρμογή είναι οι εξής:

- **HInit:** Με αυτή την εντολή αρχικοποιούμε τις παραμέτρους των HMM μοντέλων μας με χρήση των δειγμάτων για εκπαίδευση. Δουλεύει με επαναλαμβανόμενη χρήση της ευθυγράμμισης Viterbi για να κατατμήσει τα δείγματα εκπαίδευσης και ύστερα με επανυπολογισμό των παραμέτρων από συγκέντρωση των διανυσμάτων σε κάθε τμήμα. Για μείγματα Γκαουσιανών, όπως στην εφαρμογή μας, κάθε διάνυσμα σε κάθε τμήμα ευθυγραμμίζεται με την Γκαουσιανή με την μεγαλύτερη πιθανοφάνεια.
- **HRest:** Η συνάρτηση αυτή εκτελεί τον βασικό Baum-Welch αλγόριθμο, όπως περιγράφηκε στην Παρ.2.4.3, για τον επαναπροσδιορισμό των παραμέτρων του HMM με χρήση μιας ακολουθίας από διαφορετικά δείγματα προς εκπαίδευση.
- **HBuild:** Με αυτή τη συνάρτηση χτίζουμε το δίκτυο των μοντέλων μας, ώστε να είναι σαφής η εσωτερική δομή τους για το HTK, με τη δημιουργία ενός δένδρου για την παρατήρηση των καταστάσεων στην έξοδο.
- **HParse:** Μέσω της συνάρτησης αυτής το δίκτυο των μοντέλων μας μετασχηματίζεται σε μορφή κατάλληλη για την αναγνώριση με τη βοήθεια της γραμματικής που ορίζουμε για τα μοντέλα μας.
- **HVite:** Η συνάρτηση αυτή αποτελεί μια γενικής-χρήσης συνάρτηση αναγνώρισης με βάση τον αλγόριθμο Viterbi. Θα αντιστοιχίσει ένα δείγμα σε ένα από το σύνολο των HMM που έχουμε δημιουργήσει.
- **HResults:** Με αυτή τη συνάρτηση-εργαλείο ανάλυσης επίδοσης του HTK, μπορούμε να εξάγουμε αποτελέσματα για την επίδοση της αναγνώρισης των μοντέλων μας ύστερα από κάποια δοκιμή.





## Κεφάλαιο 5

# Πειραματική Διαδικασία

Το σύστημά μας, όπως αναλύεται στο προηγούμενο κεφάλαιο, δοκιμάζεται σε ένα σύνολο κινήσεων από την πολεμική τέχνη του καράτε και συγκεκριμένα του Shotokan karate. Η επιλογή του συγκεκριμένου ως είδος κίνησης που επιδέχεται εφαρμογή των μεθόδων που απαρτίζουν το σύστημα έγινε σαφής στο Κεφ. 3. Δεδομένου πως δεν βρέθηκε βάση δεδομένων με τις ζητούμενες ακολουθίες κινήσεων που να μπορεί να χρησιμοποιηθεί για τη δοκιμή του συστήματος, έπρεπε να δημιουργήσουμε τη δική μας βάση μέσω βιντεοσκόπησης πραγματικών αθλητών καράτε. Στη συνέχεια, είχαμε την επεξεργασία των δεδομένων που συλλέξαμε, με την εξαγωγή των χαρακτηριστικών, την επεξεργασία τους και την εκπαίδευση των μοντέλων HMM με μέρος των βίντεο που διαχωρίσαμε από το σύνολο για την εκπαίδευση. Τέλος, έγινε η δοκιμή του συστήματος με τα υπόλοιπα βίντεο και η εξαγωγή των συγκεντρωτικών αποτελεσμάτων.

### 5.1 Δημιουργία Βάσης Δεδομένων

Για τη δημιουργία της βάσης δεδομένων με τις ακολουθίες κινήσεων από Shotokan karate, επισκευθήκαμε μια πραγματική σχολή καράτε για να βιντεοσκοπήσουμε αθλητές να εκτελούν τις 5 πρώτες κινήσεις που περιγράψαμε στην Παρ.3.2.2, δηλαδή τις κινήσεις 'heian shodan', 'heian nidan', 'heian sandan', 'heian yondan' και 'heian godan'. Δεδομένου του εκπαιδευτικού χαρακτήρα που επιθυμούμε να δώσουμε στο σύστημά μας ώστε να μπορεί να χρησιμοποιηθεί σαν βοήθημα προπόνησης, επιλέγουμε οι αθλητές που θα βιντεοσκοπήσουμε να είναι παιδιά ηλικίας 7-14. Κατά αυτόν τον τρόπο, μπορούμε να δούμε πόσο εύκολη είναι η χρήση ενός τέτοιου εργαλείου σε μικρότερες ηλικίες, που όπως αποδείχτηκε είναι κάτι απολύτως εφαρμόσιμο, αφού έγινε με προθυμία αποδεκτό από τους μικρούς αθλητές. Οι βιντεοσκοπήσεις έγιναν κατά τη διάρκεια 3 διαφορετικών ημερών και σε αυτές συμμετείχαν 15 παιδιά, ενώ για την πραγματοποίησή τους χρησιμοποιήθηκαν 2 ψηφιακές κάμερες εμπορίου μέσης ανάλυσης. Συνολικά, καταφέραμε να συλλέξουμε 30 διαφορετικά δείγματα για τις κινήσεις 'heian shodan' και 'heian nidan', 24 διαφορετικές εκτελέσεις για τις 'heian sandan' και 'heian yondan', ενώ μόνο 16 διαφορετικές για την

'heian godan', δηλαδή ένα σύνολο από 124 δείγματα. Τα βίντεο έχουν διάρκεια που εξαρτάται από το είδος της κίνησης, αλλά γενικά κυμαίνονται μεταξύ 1'30" και 2'.

Η λήψη των κινήσεων έγινε από την μπροστά όψη της κίνησης, με τον αθλητή να αντικρίζει την κάμερα κατά την αρχή της ακολουθίας. Παράλληλα, είχαμε και λήψη της πλαϊνής όψης των κινήσεων, αλλά δεν χρησιμοποιούμε αυτά τα δεδομένα σε αυτή της μορφής της δοκιμής. Επίσης, η κάμερα κατά την βιντεοσκόπηση είναι σταθερή (οι κάμερες ήταν προσαρμοσμένες σε τρίποδα), χωρίς να ακολουθεί την κίνηση του αθλητή, απλά είναι προσαρμοσμένη να καλύπτει ολόκληρη την περιοχή που πρέπει να καλύπτει το μονοπάτι της κίνησης. Υπάρχουν και περιπτώσεις που κάποιοι αθλητές ξεφεύγουν από τα όρια λήψης, αλλά αυτό θεωρείται πως κάνει απλά το πρόβλημα λίγο δυσκολότερο και πιο κοντά σε μια ρεαλιστική χρήση του συστήματος από πραγματικούς αθλητές. Επίσης, επιδιώξαμε το παρασκήνιο των κινήσεων να είναι όσο πιο ομοιόμορφο και μη θορυβώδες από οπτικής πλευράς γίνεται, χωρίς κινήσεις από άλλους ανθρώπους ή αντικείμενα, αλλά σε αρκετά δείγματα δεν αποφεύγονται οι κινήσεις άλλων αθλητών, του προπονητή ή και άλλων ατόμων μέσα στο οπτικό πεδίο της λήψης. Αυτό προσθέτει έναν ακόμα παράγοντα δυσκολίας για την κατασκευή των μοντέλων και την αναγνώριση του είδους των κινήσεων.

Προκειμένου να δοκιμάσουμε τη δυνατότητα του συστήματος να χρησιμοποιηθεί ως βοήθημα εκπαίδευσης για τη συγκεκριμένη πολεμική τέχνη, ζητήσαμε από τον δάσκαλο-sensei της σχολής καράτε να βαθμολογήσει την επίδοση του κάθε αθλητή κατά την εκτέλεση της εκάστοτε κίνησης. Η βαθμολογία ήταν 1-10 με άριστα το 10, ώστε να συμβαδίζει με τις βαθμολογικές πληροφορίες που παρέχει το σύστημά μας στην έξοδο με βάση τη σύγκριση της κίνησης με το αποθηκευμένο μοντέλο-πρότυπο, όπως περιγράφεται στην Παρ.4.4. Έτσι, μετά την πραγματοποίηση της δοκιμής του συστήματος, μπορούμε να ελέγξουμε και αν η βαθμολογία που εξάγουμε συμπίπτει μέσα σε λογικά πλαίσια με τη βαθμολογία από τον δάσκαλο. Φυσικά, εφόσον οι αθλητές γνωρίζουν γενικά τις κινήσεις που εκτελούν και η βαθμολογία του δασκάλου αφορά κυρίως τις λεπτομέρειες εκτέλεσης και την ακρίβεια των κινήσεων, η αντιστοίχιση της βαθμολογία πρέπει να είναι αρκετά ικανή ώστε να μπορεί να αναγνωρίσει αυτές τις διαφορές σε λεπτομέρειες, κάτι που επίσης στοχεύουμε να εξακριβώσουμε αν ισχύει για το σύστημά μας.

## 5.2 Επεξεργασία Δεδομένων

Αφότου συλλέξαμε αρκετά δείγματα για την πραγματοποίηση της πειραματικής μας διαδικασίας, μπορούμε πλέον να περάσουμε στην επεξεργασία τους για την εξαγωγή των χαρακτηριστικών και την εκπαίδευση των μοντέλων μας. Ολόκληρη η διαδικασία που θα περιγραφεί παρακάτω γίνεται σε λειτουργικό σύστημα Linux/Ubuntu, καθώς παρέχει μεγαλύτερη ευχέρεια χειρισμού των απαραίτητων εργαλείων όπως τα έχουμε περιγράψει. Αρχικά, μετατρέπουμε τα βίντεό μας σε κατάλληλη μορφή για περαιτέρω επεξεργασία, δηλαδή τα μετατρέπουμε σε μέγεθος 120x160, σε μορφή '.avi' και με συχνότητα 25/καρέ ανά δευτερόλεπτο. Ακολούθως, χρησιμοποιούμε τους αλγορίθμους που περιγράψαμε στο προηγούμενο κεφάλαιο, Κεφ. 4 για να αναγνωριστούν σε όλα τα βίντεο τα σημεία εν-



(α')



(β')



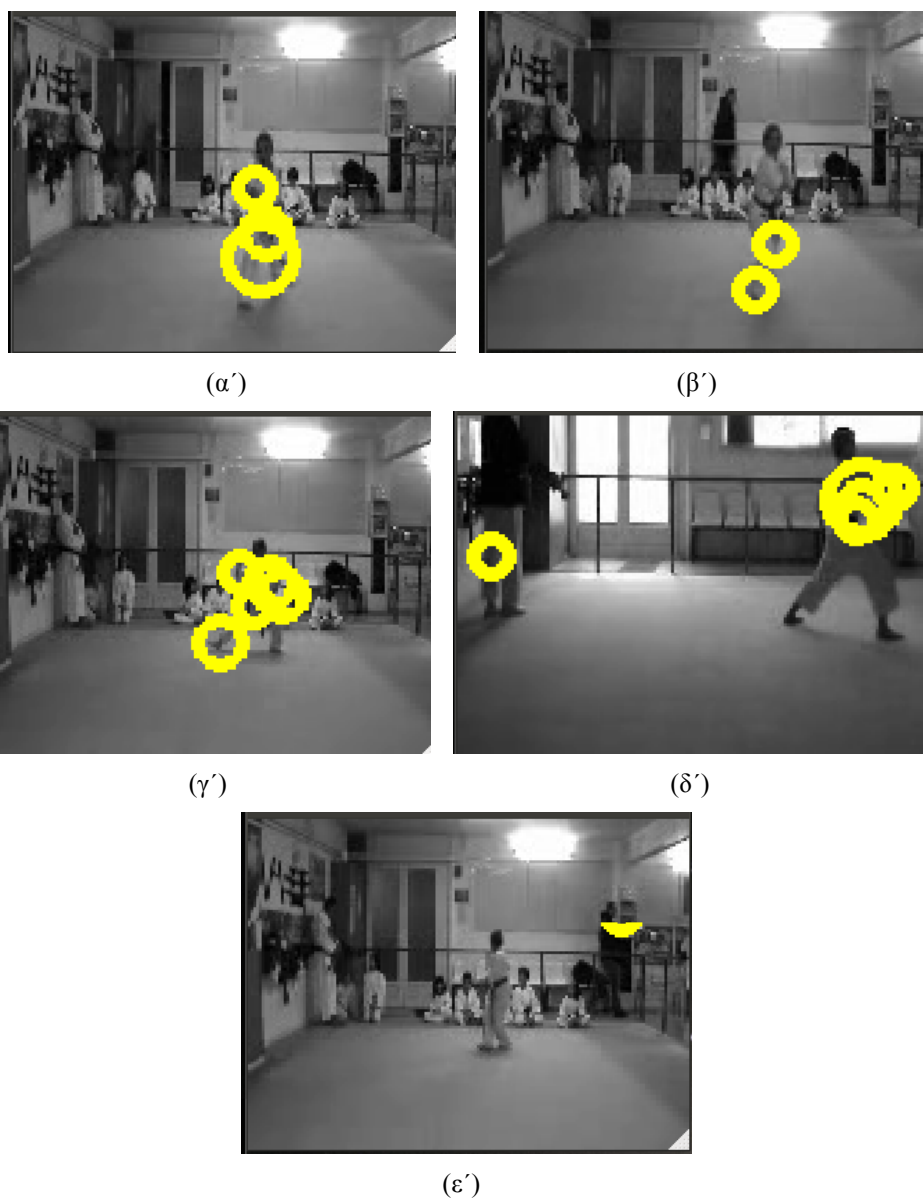
(γ')

Σχήμα 5.1: Στιγμιότυπα από τις βιντεοσκοπήσεις παιδιών-αθλητών shotokan karate κατά την εκτέλεση των 5 πρώτων κινήσεων. (α') Λήψη χωρίς οπτικό θόρυβο παρασκηνίου (κινήσεις άλλων ατόμων ή αντικειμένων). (β') Λήψη με μικρό οπτικό θόρυβο παρασκηνίου. (γ') Λήψη με μεγαλύτερο οπτικό θόρυβο παρασκηνίου.

διαφέροντος, μέσω του Harris3D στις ακολουθίες των καρτέ των βίντεο, και για κάθε σημείο ενδιαφέροντος έχουμε την εξαγωγή του διανύσματος 167 χαρακτηριστικών, που βασίζονται στην HOG/HOF περιγραφή όπως αναλύθηκε στην Παρ. 4.2. Στην εικόνα 5.2, βλέπουμε πως ενώ ο αλγόριθμός μας εντοπίζει σωστά τα σημεία ενδιαφέροντος πάνω στον αθλητή και τα κινούμενα μέλη του, ακόμα και σε κάποιες περιπτώσεις οπτικά θορυβώδους παρασκηνίου, υπάρχουν και περιπτώσεις λάθος ανίχνευσης σημείων. Για κάθε δείγμα βίντεο έχουμε την εξαγωγή 500-3.000 συνήθως σημείων ενδιαφέροντος, κάτι που είναι αρκετά λογικό αν λάβουμε υπ'όψιν τη διάρκεια των βίντεο. Συνεπώς, από κάθε βίντεο εξάγεται ένας διδιάστατος πίνακας από χαρακτηριστικά στις περιοχές γύρω από τα σημεία ενδιαφέροντος [ $Int.Points \times 167$ ].

Στη συνέχεια, επιχειρούμε τη μείωση του μεγέθους του διανύσματος χαρακτηριστικών με την PCA ανάλυση, όπως περιγράψαμε στην Παρ.4.3. Τα χαρακτηριστικά που εισάγουμε προς το μετασχηματισμό είναι τα 72 HOG και τα 90 HOF, ενώ η επεξεργασία τους γίνεται ξεχωριστά, καθώς αποτελούν εξ'αρχής διαφορετικά είδη χαρακτηρι-

## 5.2. ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ



Σχήμα 5.2: Στιγμιότυπα από την ανίχνευση σημείων ενδιαφέροντος κατά τη διάρκεια των εκτελέσεων των 5 πρώτων κινήσεων shotokan karate. (α'), (β'), (γ') Εντοπισμός σημείων ενδιαφέροντος στις σωστές θέσεις, παρά τις κινήσεις στο παρασκήνιο. (δ') Εντοπισμός σημείων και πάνω στον προπονητή. (ε') Λάθος εντοπισμός σημείων ενδιαφέροντος πάνω στο κινούμενο παρασκήνιο αντί για τον αθλητή.

στικών, τα μεν εμφάνισης και τα δε κίνησης. Δηλαδή, στον αλγόριθμό μας που υλοποιείται στο Matlab, εισάγονται οι δισδιάστατοι πίνακες που αναφέραμε παραπάνω, δια-

χωρίζονται τα HOG και HOF χαρακτηριστικά και συνενώνονται σε δύο μεγάλους πίνακες όπως στην εξίσωση (4.4)  $[(124samples)(Int.Points/sample) \times 72HOG]$  και  $[(124samples)(Int.Points/sample) \times 90HOF]$ . Οι πίνακες αυτοί επεξεργάζονται μέσω της *princomp* (4.5) και παίρνουμε τις μετασχηματισμένες στήλες δεδομένων για τα HOG και HOF χαρακτηριστικά που αντικαθιστούν τις αντίστοιχες στήλες στους αρχικούς διδιάστατους πίνακες, με αποτέλεσμα να έχουμε καινούργιους πίνακες χαρακτηριστικών μειωμένης διάστασης.

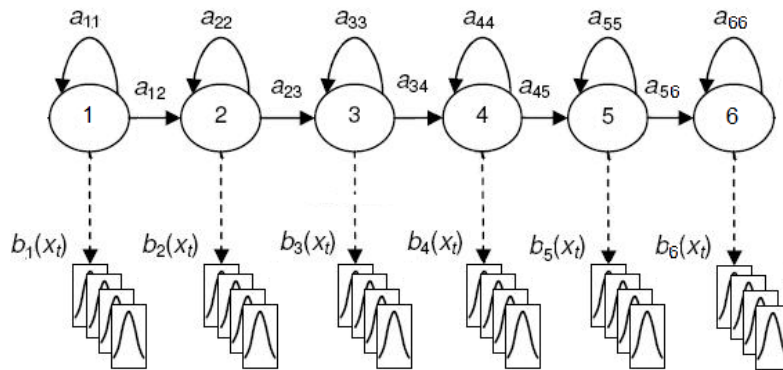
Επιλέγοντας να κρατήσουμε το 90% της πληροφορίας-διακύμανσης, έχουμε μετά την εφαρμογή της PCA ανάλυσης μείωση του διανύσματος χαρακτηριστικών από 72 σε 24 για τα HOG χαρακτηριστικά και από 90 σε 23 για τα HOF. Επιθυμώντας να εξετάσουμε την επίδραση αυτών των διαφορετικών εφαρμογών της PCA σε διαφορετικά υποσύνολα των χαρακτηριστικών, εξετάζουμε ανεξάρτητα τις περιπτώσεις αυτές, αλλά και συνδυαστικά. Ως αποτέλεσμα, δημιουργούνται 4 μεγάλες κατηγορίες για τα χαρακτηριστικά που εξετάζουμε ως προς την επεξεργασία τους. Στην πρώτη κατηγορία ανήκουν χαρακτηριστικά δειγμάτων που δεν έχουν επεξεργαστεί μέσω PCA, οπότε και οι διδιάστατοι πίνακες χαρακτηριστικών είναι οι αρχικοί  $[Int.Points \times 167]$ , ενώ αποκαλούμε την κατηγορία *porca*. Στη δεύτερη κατηγορία επιλέγουμε εφαρμογή PCA μόνο στα HOG χαρακτηριστικά, οπότε η διάσταση του διανύσματος χαρακτηριστικών γίνεται  $5 + 24 + 90 = 119$  και οι πίνακες  $[Int.Points \times 119]$ , ενώ τα καλούμε *pcahog*. Στην τρίτη, εφαρμόζουμε μόνο στα HOF, οπότε το διάνυσμα χαρακτηριστικών έχει μήκος  $5 + 72 + 23 = 100$  και οι πίνακες  $[Int.Points \times 100]$  με ονομασία *pcahof*. Τέλος, έχουμε τον συνδυασμό και των δύο μετασχηματισμών με ονομασία *pca* και πίνακες  $[Int.Points \times 52]$ .

Για την ολοκλήρωση του συστήματος πρέπει να εκπαιδύσουμε τα 5 μοντέλα μας με βάση τους διδιάστατους πίνακες χαρακτηριστικών στους οποίους έχουμε καταλήξει για κάθε διαφορετικό δείγμα. Η επιλογή των κατάλληλων δειγμάτων φυσικά αποτελεί ένα από τα σημαντικότερα βήματα αυτής της διαδικασίας. Εμείς, στην προσπάθεια να διαμορφώσουμε μοντέλα που αντιπροσωπεύουν το σωστό τρόπο εκτέλεσης των κινήσεων, από άποψη τεχνικής, τρόπου και εύρους κίνησης στο χώρο, και τα οποία μπορούν να παίζουν το ρόλο του προτύπου της κίνησης ώστε οι υπόλοιπες εκτελέσεις να μπορούν να συγκρίνονται και να βαθμολογούνται με βάση αυτά, επιλέγουμε τα δείγματα που πλησιάζουν πιο πολύ στην τέλεια εκτέλεση της κίνησης. Όπως εξηγήσαμε, το κάθε δείγμα συνοδεύεται από έναν βαθμό από τον δάσκαλο με τον οποίο αξιολογείται η ποιότητα εκτέλεσης της κίνησης. Εμείς λοιπόν θα χρησιμοποιήσουμε για την εκπαίδευση των μοντέλων μόνο τις κινήσεις με υψηλό βαθμό αξιολόγησης, δηλαδή δείγματα που κινούνται στο 8-9-10, ή και κάποια στο 7 για τις περιπτώσεις όπου έχουμε έλλειψη δειγμάτων. Επειδή, όπως προαναφέραμε, δεν έχουμε τον ίδιο αριθμό δειγμάτων για καθεμία από τις 5 κινήσεις (1η-2η:30, 3η-4η:24, 5η:16) για να είμαστε δίκαιοι προς την εκπαίδευση όλων των μοντέλων και να μην υπάρχουν μοντέλα πιο καλά εκπαιδευμένα από άλλα, ενώ παράλληλα να είναι αυστηρά ως προς το ποιος τρόπος εκτέλεσης είναι σωστός, αρχικά επιλέγουμε να χρησιμοποιήσουμε μόνο 4 δείγματα από κάθε κίνηση. Τα υπόλοιπα κάθε κατηγορίας θα χρησιμοποιηθούν για την μετέπειτα δοκιμή των μοντέλων.

Όπως αναλύσαμε στην Παρ.4.4, το είδος των μοντέλων που θα χρησιμοποιήσουμε είναι τα HMM και μάλιστα η δημιουργία και η διαχείρισή τους θα γίνει μέσω του εργα-

λείου ΗΤΚ. Για να χρησιμοποιήσουμε τα στοιχεία μας με το ΗΤΚ όμως, πρέπει να τα φέρουμε σε κατάλληλη μορφή, ώστε να μπορούν να διαβαστούν από το εργαλείο αυτό. Για το σκοπό αυτό αναπτύσσουμε ένα πρόγραμμα μετατροπής σε γλώσσα C, το οποίο παίρνει τα αρχεία με τα χαρακτηριστικά του κάθε δείγματος και επιστρέφει το αντίστοιχο δυαδικό αρχείο που περιέχει και την κατάλληλη κεφαλή με τις πληροφορίες του δείγματος. Στην περίπτωση της επεξεργασίας των δεδομένων μέσω του Matlab για την εφαρμογή της PCA ανάλυσης, όπως είδαμε παραπάνω, η μετατροπή των δεδομένων σε δυαδικά αρχεία που επιδέχονται επεξεργασία από το ΗΤΚ για τη συνέχεια, μπορεί να γίνει πέρα από το πρόγραμμα σε C, και με τη χρήση της απλούστερης εντολής *HWrite* που περιέχεται στο πακέτο *Voicebox* για το Matlab. Ύστερα, δημιουργούμε τα πρότυπα των HMM μοντέλων, ώστε να περιγραφεί η ακριβής μορφή τους, συμπεριλαμβανομένων των καταστάσεων που τα αποτελούν και των μειγμάτων Γκαουσιανών ανά κατάσταση για κάθε μεταβλητή του διανύσματος χαρακτηριστικών. Η μορφή των μοντέλων θα καθοριστεί από τις κινήσεις που θέλουμε να περιγράψουμε, δηλαδή τις 5 κινήσεις του shotokan karate. Παρατηρώντας τις κινήσεις που παρουσιάστηκαν στην Παρ.3.2.2, βλέπουμε πως αποτελούνται από έναν αριθμό διακριτών τμημάτων που για τις περισσότερες αυτών ο αριθμός είναι 6. Συνεπώς, αρχίζοντας από την πεποίθηση για την αποδοτικότητα χρήσης ενός κοινού μοντέλου για όλες τις κατηγορίες με παράλληλο σκοπό τη μείωση της πολυπλοκότητας του προβλήματος, καταλήγουμε στη χρήση HMM 6 καταστάσεων, η καθεμία από τις οποίες μπορεί να μεταβεί είτε στον εαυτό της ή στην αμέσως επόμενη, δηλαδή γραμμικών left-to-right HMM, κάτι που βέβαια δεν είναι δεσμευτικό καθώς οι κινήσεις διαφέρουν σε αρκετά σημεία. Για τον αριθμό των μειγμάτων Γκαουσιανών, επιλέγουμε τη χρήση 4 διαφορετικών Γκαουσιανών κατανομών, ώστε να έχει και κάποια ευελιξία το κάθε μοντέλο, καθώς πρωταρχικός μας στόχος είναι η σωστή αναγνώριση των κινήσεων παρά τις μικρές διαφορές στην εκτέλεση από άτομο σε άτομο. Ωστόσο, δεν επιλέγουμε μεγαλύτερο αριθμό, ώστε να μπορούμε να κρατάμε και την αυστηρότητα του εκπαιδευτικού χαρακτήρα που αναφέραμε προηγουμένως.

Στη συνέχεια, αρχικοποιούμε τα μοντέλα μας, μέσω της εντολής *HInit*, κάνοντας χρήση των προτύπων που δημιουργήσαμε, ώστε οι παράμετροι των HMM, δηλαδή οι τιμές του πίνακα μετάβασης από κατάσταση σε κατάσταση και οι μέσες τιμές και διακυμάνσεις των Γκαουσιανών κατανομών για κάθε χαρακτηριστικό, να λάβουν τιμές κοντά στις τιμές που βοηθούν για την περαιτέρω εκπαίδευση. Έπειτα, προχωρούμε στην καθαυτό εκπαίδευση των HMM, όπου τα 4 δείγματα που επιλέξαμε χρησιμοποιούνται διαδοχικά για τον επαναληπτικό επανυπολογισμό των παραμέτρων μέχρι να επέλθει σύγκλιση, μέσω της εντολής *HRest*, η οποία κάνει χρήση του 'backward-forward' Baum-Welch αλγορίθμου. Κατά την εκτέλεση των συγκεκριμένων εντολών, χρησιμοποιούμε επίσης και μία παράμετρο για να θέσουμε ένα κατώτερο όριο στις λογαριθμικές πιθανότητες που υπολογίζονται κατά την εκπαίδευση, ώστε να μην μηδενίζονται. Με την ολοκλήρωση των παραπάνω, το σύστημά μας είναι έτοιμο για τη δοκιμή, ώστε να εξακριβωθεί η αξιοπιστία του.



Σχήμα 5.3: Τα HMM που προτείνουμε προς χρήση για την πειραματική μας πριν την πραγματοποίηση των δοκιμών, με βάση μόνο την παρατήρηση των κινήσεων που εξετάζουμε. Γραμμικά left-to-right HMM 6 καταστάσεων με 4 μείγματα Γκαουσιανών ανά κατάσταση.

### 5.3 Δοκιμές & Αποτελέσματα

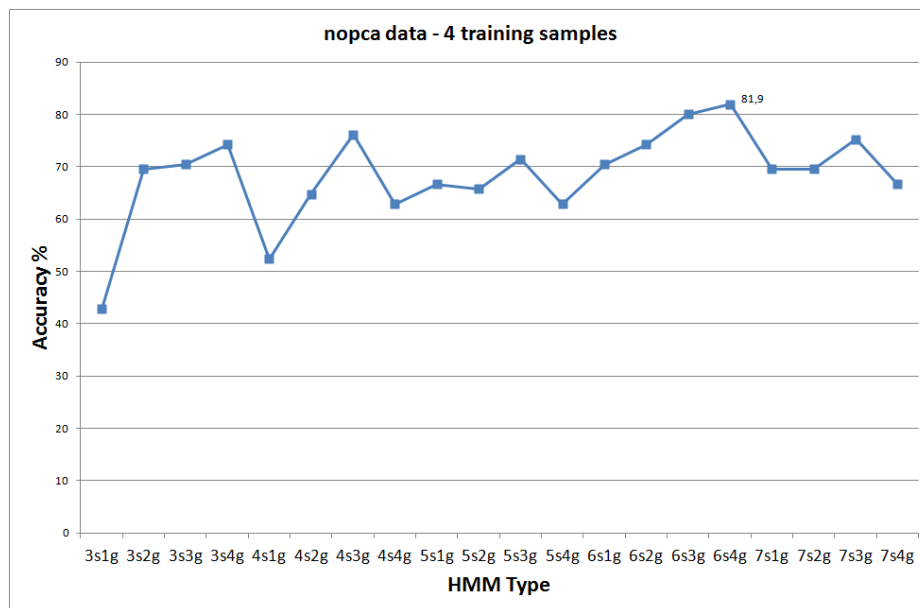
Οι δοκιμές (testing) του συστήματός μας περιλαμβάνουν την προσπάθεια αναγνώρισης της κατηγορίας όλων των δειγμάτων που συλλέξαμε με τη βοήθεια του HTK. Αφού πρώτα έχουμε δημιουργήσει το δίκτυο των μοντέλων μας με βάση το λεξιλόγιο και τη λίστα των 5 HMM που χρησιμοποιούμε με την εντολή *HBuild*, το μετασχηματίζουμε μέσω της εντολής *HParse* με βάση τη γραμματική, που στην ουσία αποτελεί απλά μια διάζευξη μεταξύ των 5 μοντέλων για ολόκληρα τα δείγματα που εισάγουμε, ώστε να προχωρήσουμε στην αναγνώριση. Η διαδικασία αυτή πραγματοποιείται με την εντολή *HVite* του HTK, που αναλαμβάνει να αντιστοιχίσει το κάθε δείγμα με την κατηγορία κίνησης που του ταιριάζει περισσότερο. Με την εκτέλεση της εντολής έχουμε και εξαγωγή κάποιων προκαταρκτικών αποτελεσμάτων σχετικά με τη λογαριθμική πιθανότητα να πρόκειται για το εκάστοτε μοντέλο, που μέσω της *HResults* θα παρουσιαστούν ακόμα καλύτερα με προβολή του πίνακα σύγχυσης (confusion matrix) και στατιστικών στοιχείων.

Έχοντας χτίσει λοιπόν την ολοκληρωμένη διαδικασία με την οποία επεξεργάζονται τα δεδομένα μέσα στο σύστημά μας, μπορούμε να προχωρήσουμε σε διαφορετικές δοκιμές για να διερευνήσουμε τα όρια αξιοπιστίας του. Έτσι, δοκιμάζουμε για καθεμία από τις 4 κατηγορίες δεδομένων που αναφέραμε παραπάνω, *porca*, *pcahog*, *pcahof* και *pca*, διαφορετικούς συνδυασμούς για τα HMM μοντέλα μας και καταγράφουμε τις επιδόσεις του συστήματος στις επιμέρους κατηγορίες αλλά και συνολικά κατά την αναγνώριση αρχικά των δειγμάτων που χρησιμοποιήθηκαν για την εκπαίδευση (4 δείγματα x 5 κατηγορίες = 20), ώστε να είμαστε σίγουροι για τη σωστή δημιουργία των HMM μας, και έπειτα, των υπόλοιπων δειγμάτων από όλες τις κατηγορίες, τα οποία όπως αναφέραμε δεν είναι ισάριθμα (26 + 26 + 20 + 20 + 12 = 104). Τα διαφορετικά είδη των HMM που εξετάζουμε σε αυτή την εργασία είναι 3 έως 7 καταστάσεων και 1 έως 4 Γκαουσιανών μειγμάτων για κάθε κατάσταση, ένα σύνολο από 20 διαφορετικά HMM. Επίσης, στο τέλος δοκιμάζουμε

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ

και μία παραλλαγή του διαχωρισμού δειγμάτων για εκπαίδευση-δοκιμή, χρησιμοποιώντας 5 δείγματα για εκπαίδευση των μοντέλων, με προσθήκη του επόμενου καλύτερου σε βαθμολογία δείγματος ώστε να μην χάνεται σε μεγάλο βαθμό η αυστηρότητα που επιζητούμε, και τα υπόλοιπα 99 (124 - 5x5 ) δείγματα για τις δοκιμές.

Αρχίζουμε εξετάζοντας τα *norca* δεδομένα που δεν έχουν μετασχηματιστεί και όπως προαναφέραμε, αρχικά δοκιμάζουμε συνδυασμούς μοντέλων HMM με τη συνθήκη όλες οι κατηγορίες να περιγράφονται από το ίδιο μοντέλο. Τα αποτελέσματα ακρίβειας για τη δοκιμή αυτών των 20 συνδυασμών φαίνονται στην παρακάτω γραφική παράσταση 5.3, όπου βλέπουμε πως ο καλύτερος συνδυασμός είναι για 6 καταστάσεις και μείγμα 4 Γκαουσιανών (που για ευκολία θα αναφέρουμε από εδώ και πέρα ως 6s4g, s:states, g:gaussians). Για αυτή την περίπτωση πετυχαίνουμε **81,9%** συνολική ακρίβεια στην αναγνώριση των νέων δειγμάτων, ενώ στις επιμέρους κατηγορίες έχουμε < heianshodan , heianndan , heiansandan , heianyondan , heiangodan > = < 88,5% , 88,5% , 50% , 85% , 100% > (που στη συνέχεια θα αναπαριστούμε μόνο από το δεύτερο σκέλος για συντομία). Θέλοντας να εξετάσουμε περαιτέρω το σύστημα, επιλέγουμε να δοκιμάσουμε και συν-



Σχήμα 5.4: Διάγραμμα για *norca* δεδομένα και εκπαίδευση με 4 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g).

δυασμούς στους οποίους δεν είναι ίδια τα μοντέλα και για τις 5 κατηγορίες, κάτι που είναι και λογικό, αφού οι ακολουθίες κινήσεων έχουν αρκετές διαφορές μεταξύ τους, όπως δείξαμε. Για αυτό το λόγο, συνεχίζουμε με 3 διαφορετικές μεθόδους, η πρώτη από τις οποίες είναι να δοκιμάσουμε να συνδυάσουμε τους τύπους HMM εκεί που εμφανίζονται οι μεγαλύτερες επιμέρους ακρίβειες για κάθε κατηγορία, κάτι που όμως δεν μας δίνει καλύτερα αποτελέσματα. Η δεύτερη μέθοδος περιλαμβάνει την εξαντλητική αναζήτηση

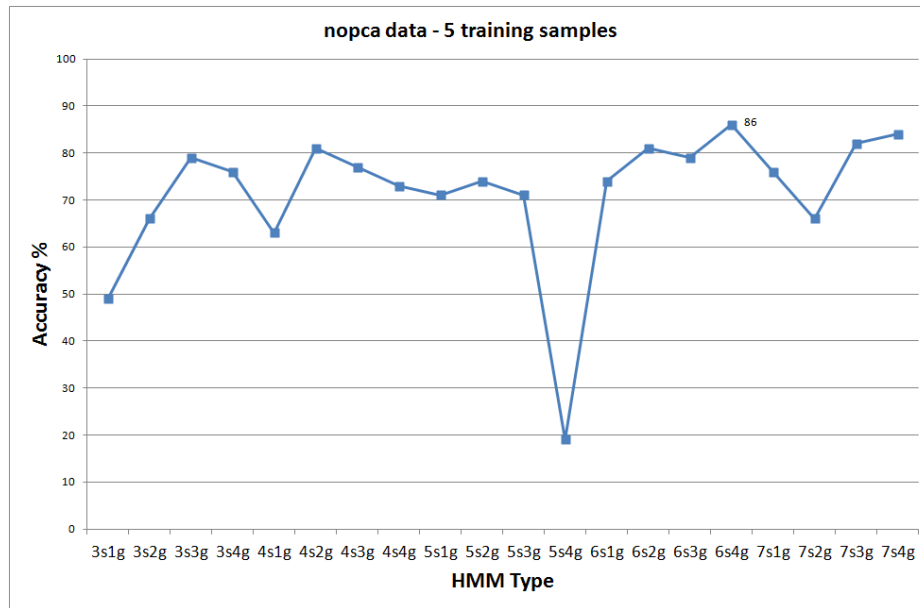


του καλύτερου συνδυασμού (τύπου 'British Museum') μέχρι να βρεθεί αυτός που μας δίνει το καλύτερο αποτέλεσμα. Δεδομένου όμως πως μιλάμε για μια αναζήτηση μεταξύ  $20 \times 20 \times 20 \times 20 \times 20 = 3.200.000$  περιπτώσεων, αυτή η μέθοδος κρίνεται ως υπολογιστικά ασύμφορη και εξαιρετικά χρονοβόρα, δεδομένου και του χρόνου που απαιτείται για την ολοκλήρωση μίας δοκιμής για όλα τα δείγματα από το ΗΤΚ, ο οποίος για τον διπύρνο επεξεργαστή στα 2,53 GHz με 4GB RAM του υπολογιστή στον οποίο έγιναν οι δοκιμές έφτανε τα 40-50sec. Συνεπώς, η τελευταία μέθοδος στην οποία καταλήγουμε περιλαμβάνει την επεξεργασία του καλύτερου συνδυασμού που βρέθηκε στο προηγούμενο στάδιο με αλλαγή των πιο 'αδύναμων' μελών, που έχουν δηλαδή το μικρότερο ποσοστό ευστοχίας, όπως το μοντέλο για το heiansandan που ήταν μόλις στο 50%. Ύστερα από εφαρμογή της εποπτικής μεθόδου, που δεν είναι τόσο αυτοματοποιημένη καθώς εξαρτάται και από την κρίση του ερευνητή, καταφέρνουμε με συνδυασμό μοντέλων [ heianshodan , heiannidan , heiansandan , heianyondan , heiangodan ] = [ 6s4g , 6s4g , 3s4g , 6s4g , 6s4g ] (που στη συνέχεια θα αναπαριστούμε μόνο από το δεύτερο σκέλος για συντομία) βελτίωση του συνολικού ποσοστού ακρίβειας από 81,90% σε **84,76%** με επιμέρους ακρίβειες < 88,5% , 88,5% , 70% , 85% , 92,3% >.

Έπειτα, δοκιμάζουμε και την παραλλαγή με τα 5 δείγματα για εκπαίδευση, όπου προσθέτουμε στα 4 ήδη επιλεγμένα, λόγω της υψηλής τους βαθμολογίας, δείγματα το αμέσως επόμενο σε υψηλή βαθμολογία (το οποία έχει βαθμολογία 7 για όλες τις κατηγορίες), ώστε να δούμε τη μεταβολή της απόδοσης του συστήματος. Ακολουθώντας την ίδια διαδικασία με πριν, εξετάζουμε τους 20 συνδυασμούς με ίδια μοντέλα για όλες τις κατηγορίες και καταλήγουμε πάλι στην περίπτωση 6s4g που μας δίνει συνολική ακρίβεια **86%** και επιμέρους < 88,5% , 88,5% , 50% , 85% , 100% > , όπως φαίνεται και στο διάγραμμα 5.3. Τέλος, με εφαρμογή της εποπτικής μεθόδου δοκιμών, καταφέρνουμε να αυξήσουμε την ακρίβεια στο **88%** με επιμέρους < 84% , 92% , 84,2% , 89,5% , 91,7% > , χρησιμοποιώντας τον συνδυασμό HMM [ 6s4g , 5s4g , 6s4g , 4s4g , 5s4g ]. Οι πίνακες σύγκυσης και για τις δύο καλύτερες περιπτώσεις για 4 ή 5 δείγματα στην εκπαίδευση φαίνονται στο σχήμα 5.6.

Στη συνέχεια, εξετάζουμε τη δεύτερη κατηγορία δεδομένων, *pcahog*, και όπως και πριν αρχίζουμε με τις δοκιμές για εκπαίδευση με 4 δείγματα. Στην περίπτωση με το ίδιο είδος για όλες τις κατηγορίες, όπως φαίνεται και στο διάγραμμα 5.7(α'), έχουμε τη μεγαλύτερη ακρίβεια για τους συνδυασμούς 6s3g και 7s4g **78,1%**, αλλά λόγω μεγαλύτερης ομοιομορφίας των επιμέρους, κρατάμε την 6s3g περίπτωση με επιμέρους ακρίβειες < 61,5% , 92,3% , 75% , 70% , 100% >. Εξετάζοντας έπειτα σύνθετες περιπτώσεις, καταφέρνουμε να αυξήσουμε την ακρίβεια στο **88,57%** με επιμέρους < 80,8% , 92,3% , 85% , 90% , 100% > , με τον συνδυασμό [ 7s4g , 6s4g , 6s4g , 6s4g , 6s4g ]. Για την περίπτωση των 5 δειγμάτων για εκπαίδευση, όπως βλέπουμε στο διάγραμμα 5.7(β'), με ίδια HMM για όλες τις κατηγορίες βρίσκουμε τις περιπτώσεις 6s2g και 7s4g με συνολική ακρίβεια **79%**, αλλά λόγω μεγαλύτερης επιμέρους αμοιομορφίας επιλέγουμε την 6s2g με επιμέρους < 80% , 64% , 84,2% , 94,7% , 75% >. Με διερεύνηση σύνθετων συνδυασμών, πετυχαίνουμε συνολική ακρίβεια **88%** για [ 5s4g , 4s4g , 4s4g , 4s4g , 4s4g ], με επιμέρους < 88% , 84% , 78,9% , 94,7% , 100% > , ενώ οι πίνακες σύγκυσης φαίνονται στην εικόνα 5.8.

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ



Σχήμα 5.5: Διάγραμμα για *norca* δεδομένα και εκπαίδευση με 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g).

|        | shodan | nidan | sandan | yondan | godan |
|--------|--------|-------|--------|--------|-------|
| shodan | 23     | 3     | 0      | 0      | 0     |
| nidan  | 3      | 23    | 0      | 0      | 0     |
| sandan | 2      | 1     | 14     | 1      | 2     |
| yondan | 0      | 2     | 1      | 17     | 0     |
| godan  | 0      | 0     | 1      | 0      | 12    |

(α')

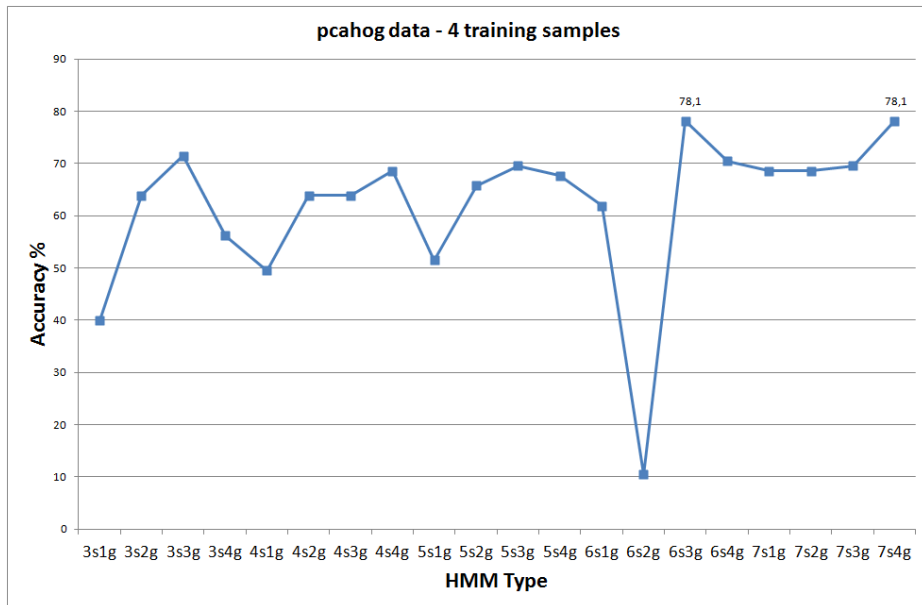
|        | shodan | nidan | sandan | yondan | godan |
|--------|--------|-------|--------|--------|-------|
| shodan | 21     | 4     | 0      | 0      | 0     |
| nidan  | 2      | 23    | 0      | 0      | 0     |
| sandan | 0      | 1     | 16     | 1      | 1     |
| yondan | 0      | 2     | 0      | 17     | 0     |
| godan  | 0      | 0     | 0      | 1      | 11    |

(β')

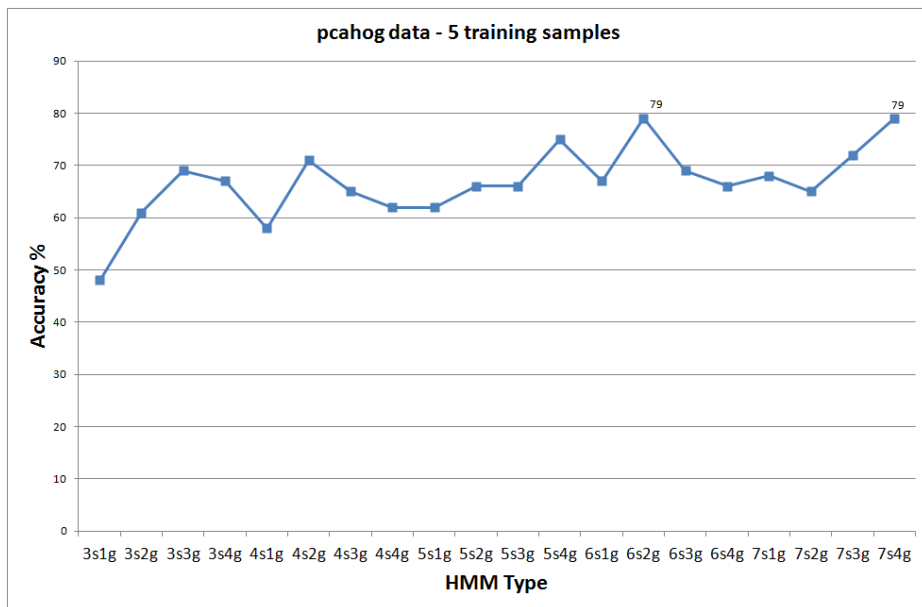
Σχήμα 5.6: Πίνακες σύγχυσης (confusion matrices) για *norca* δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα για τις συνολικά καλύτερες ακρίβειες που βρέθηκαν.

Σαν επόμενο στάδιο, εξετάζουμε την τρίτη κατηγορία δεδομένων, *pcahof*, αρχίζοντας με τις δοκιμές για εκπαίδευση με 4 δείγματα. Στην περίπτωση με το ίδιο είδος για όλες τις κατηγορίες, όπως από το διάγραμμα 5.9(α'), έχουμε τη μεγαλύτερη ακρίβεια για τον συνδυασμό 6s4g **62,86%**, με επιμέρους ακρίβειες < 61,5% , 65,4% , 35% , 80% , 76,9% >.

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ



(α')



(β')

Σχήμα 5.7: Διάγραμμα για *pcahog* δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g).

Εξετάζοντας έπειτα σύνθετες περιπτώσεις, πετυχαίνουμε την αύξηση της ακρίβειας στο

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ

|        | shodan | nidan | sandan | yondan | godan |
|--------|--------|-------|--------|--------|-------|
| shodan | 21     | 5     | 0      | 0      | 0     |
| nidan  | 2      | 24    | 0      | 0      | 0     |
| sandan | 0      | 1     | 17     | 0      | 2     |
| yondan | 0      | 1     | 0      | 18     | 1     |
| godan  | 0      | 0     | 0      | 0      | 13    |

(α')

|        | shodan | nidan | sandan | yondan | godan |
|--------|--------|-------|--------|--------|-------|
| shodan | 22     | 3     | 0      | 0      | 0     |
| nidan  | 4      | 21    | 0      | 0      | 0     |
| sandan | 0      | 0     | 15     | 1      | 3     |
| yondan | 0      | 1     | 0      | 18     | 0     |
| godan  | 0      | 0     | 0      | 0      | 12    |

(β')

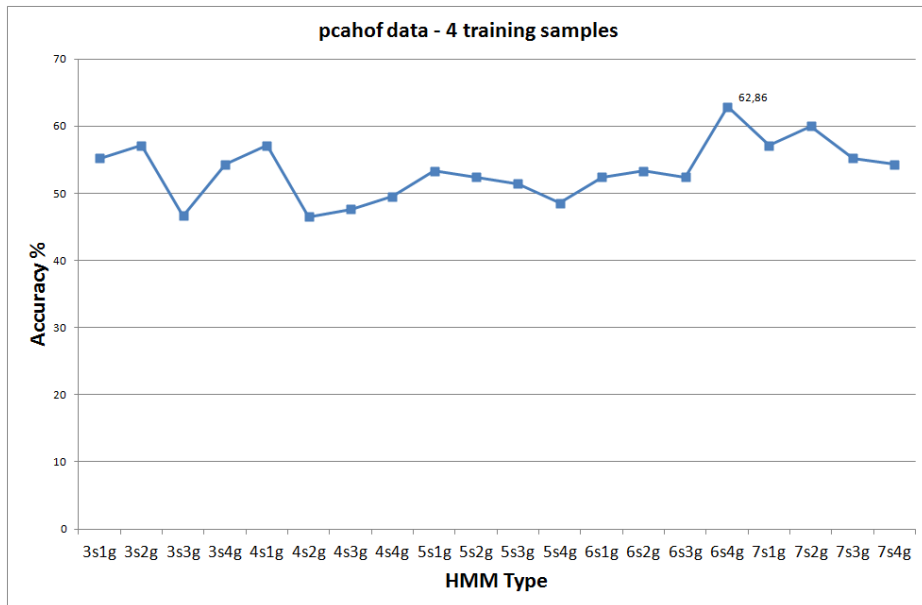
Σχήμα 5.8: Πίνακες σύγχυσης (confusion matrices) για *pcahog* δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα για τις συνολικά καλύτερες ακρίβειες που βρέθηκαν.

**64,76%** με επιμέρους  $\langle 57,7\%, 73,1\%, 40\%, 80\%, 76,9\% \rangle$ , με τον συνδυασμό [ 6s3g , 6s4g , 6s4g , 6s4g , 6s4g ]. Για την περίπτωση των 5 δειγμάτων για εκπαίδευση, όπως φαίνεται από το διάγραμμα 5.9(β'), με ίδια HMM για όλες τις κατηγορίες βρίσκουμε την περίπτωση 7s1g με συνολική ακρίβεια **68%** και επιμέρους  $\langle 80\%, 52\%, 63,2\%, 63,2\%, 91,7\% \rangle$ . Με διερεύνηση σύνθετων συνδυασμών, καταφέρνουμε αύξηση της συνολικής ακρίβειας στο **70%** για [ 6s1g , 5s1g , 6s1g , 6s1g , 6s1g ], με επιμέρους  $\langle 76\%, 64\%, 57,9\%, 73,7\%, 83,3\% \rangle$ , ενώ οι πίνακες σύγχυσης φαίνονται στην εικόνα 5.10.

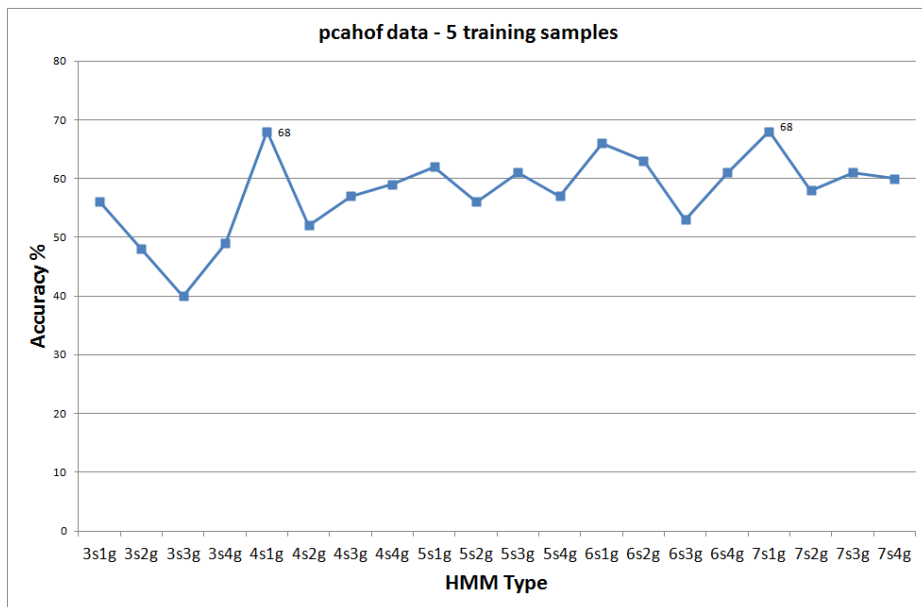
Τέλος, εξετάζουμε την κατηγορία που περιέχει δεδομένα που έχουν μετασηματιστεί και τα HOG και τα HOF χαρακτηριστικά αντικαθιστώντας με μικρότερου μήκους διανύσματα τα αρχικά δεδομένα, *pc*. Εξετάζοντας αρχικά τις δοκιμές για εκπαίδευση με 4 δείγματα, στην περίπτωση με το ίδιο είδος για όλες τις κατηγορίες, όπως φαίνεται και στο διάγραμμα 5.11(α'), έχουμε τη μεγαλύτερη ακρίβεια **67,62%** με τον συνδυασμό 3s3g, ενώ οι επιμέρους ακρίβειες βρίσκονται  $\langle 34,6\%, 92,3\%, 55\%, 75\%, 92,3\% \rangle$ . Με την εποπτική μέθοδο εξέτασης πιο σύνθετων περιπτώσεων, επιτυγχάνουμε να αυξήσουμε την ακρίβεια στο **75,24%** με επιμέρους  $\langle 53,8\%, 96,2\%, 60\%, 80\%, 92,3\% \rangle$ , με τον συνδυασμό [ 3s4g , 3s3g , 5s4g , 5s4g , 3s3g ]. Για την περίπτωση όπου χρησιμοποιούμε 5 δείγματα για την εκπαίδευση, όπως βλέπουμε και στο διάγραμμα 5.11(β'), με ίδια HMM για όλες τις κατηγορίες μεγαλύτερη συνολική ακρίβεια **70%** παρουσιάζεται για την περίπτωση 5s4g και επιμέρους  $\langle 40\%, 76\%, 73,7\%, 89,5\%, 83,3\% \rangle$ . Διερευνώντας πάλι πιο σύνθετους συνδυασμούς, αυξάνουμε τη συνολική ακρίβεια στο **76%** για [ 3s4g , 7s4g , 7s4g , 5s4g , 7s4g ], με επιμέρους  $\langle 64\%, 80\%, 63,2\%, 89,5\%, 91,7\% \rangle$ , ενώ οι πίνακες σύγχυσης φαίνονται στην εικόνα 5.12.

**Εκπαιδευτικός Χαρακτήρας:** Κατά την αναγνωριστική διαδικασία με τη χρήση της εντολής *HVite* του HTK, έχουμε εξαγωγή της λογαριθμικής πιθανότητας (*log probability*)

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ



(α')



(β')

Σχήμα 5.9: Διάγραμμα για *pcahof* δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g).

ένα δείγμα να ανήκει σε μία κατηγορία, ώστε η κατηγοριοποίηση γίνεται με βάση το ποια

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ

|        | shodan | nidan | sandan | yondan | godan |
|--------|--------|-------|--------|--------|-------|
| shodan | 15     | 11    | 0      | 0      | 0     |
| nidan  | 5      | 19    | 0      | 1      | 1     |
| sandan | 4      | 7     | 8      | 0      | 1     |
| yondan | 0      | 2     | 0      | 16     | 2     |
| godan  | 0      | 0     | 1      | 2      | 10    |

(α')

|        | shodan | nidan | sandan | yondan | godan |
|--------|--------|-------|--------|--------|-------|
| shodan | 19     | 6     | 0      | 0      | 0     |
| nidan  | 3      | 16    | 4      | 1      | 1     |
| sandan | 1      | 4     | 11     | 1      | 2     |
| yondan | 0      | 5     | 0      | 14     | 0     |
| godan  | 0      | 0     | 0      | 2      | 10    |

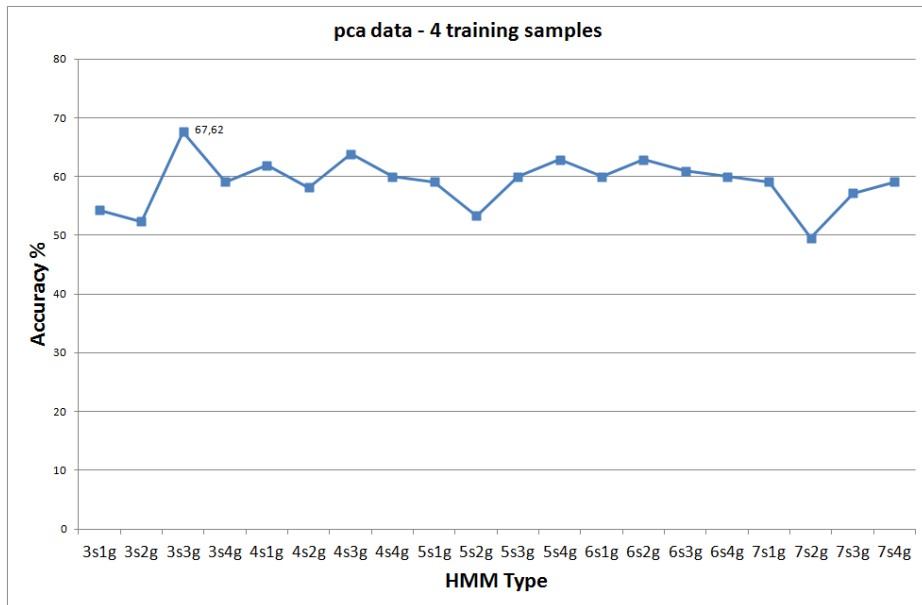
(β')

Σχήμα 5.10: Πίνακες σύγχυσης (confusion matrices) για *pcahof* δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα για τις συνολικά καλύτερες ακρίβειες που βρέθηκαν.

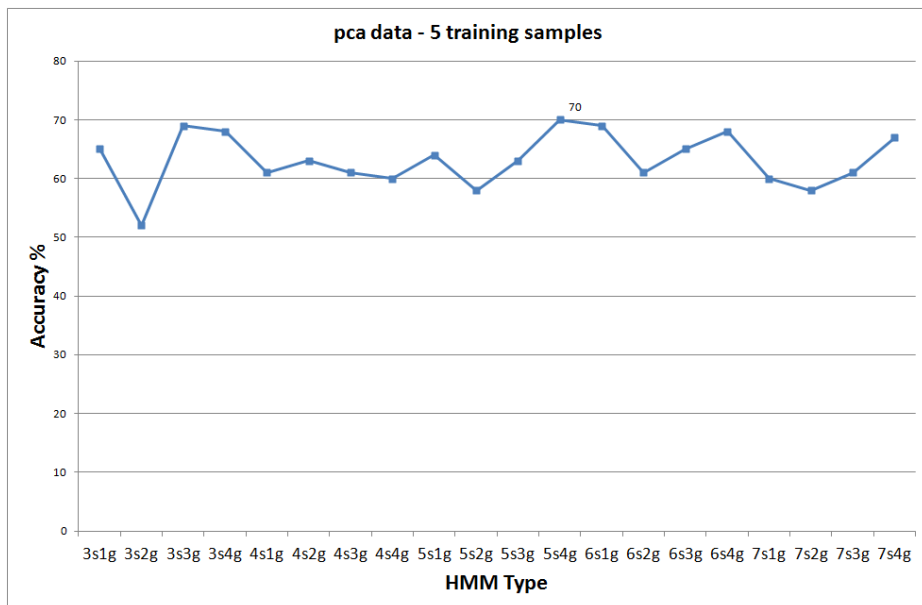
κατηγορία έχει τη μεγαλύτερη πιθανότητα. Επιθυμώντας να προσδώσουμε έναν εκπαιδευτικό χαρακτήρα στην εφαρμογή μας, όπως προαναφέραμε, επιλέγουμε να χρησιμοποιήσουμε αυτές τις εξαγόμενες πιθανότητες για να βαθμολογήσουμε την ακρίβεια εκτέλεσης της κίνησης σε σχέση με την πρότυπη. Οι βαθμοί επιλέγουμε να βρίσκονται σε κλίμακα 1-10 για να μπορούμε να έχουμε και κάποια σύγκριση με τη βαθμολογία που αντιστοιχίστηκε σε κάθε δείγμα από τον δάσκαλο (Sensei), αλλά και επειδή το είδος του προβλήματος δεν επιδέχεται και πολύ μεγαλύτερη ακρίβεια από αυτήν την κλίμακα, λόγω του πολυσύνθετου χαρακτήρα του. Συνεπώς, επιλέγουμε τους συνδυασμούς που δίνουν τις μεγαλύτερες συνολικές ακρίβειες για τις 4 κατηγορίες δεδομένων μας και για τις 2 περιπτώσεις εκπαίδευσης με 4 ή 5 δείγματα και αντιστοιχίζουμε τις λογαριθμικές πιθανότητες μέσω μίας συνάρτησης σε βαθμούς 1-10. Η συνάρτηση αντιστοίχισης βέβαια, δεν είναι κάτι τετριμμένο, οπότε μέσω ευριστικών μεθόδων προσπαθούμε να προσδιορίσουμε μια καλή μορφή για αυτήν.

Γενικά, αν τα δείγματα που έχουν χρησιμοποιηθεί για την εκπαίδευση είναι όλα ίδιας ποιότητας ακρίβειας και πλησιάζουν το 'τέλειο' 10 σε βαθμολογία, ώστε να χτιστεί ένα ισχυρό πρότυπο, η μέγιστη πιθανότητα αυτών πρέπει να χρησιμοποιηθεί ως σημείο αναφοράς για τη σύγκριση των υπολοίπων. Στην περίπτωση μας, επιλέγουμε το ολικό μέγιστο από τις παρατηρήσεις, καθώς δεν έχουμε την τέλεια εκπαίδευση. Η χαμηλότερη τιμή της κλίμακας είναι ένα εξίσου σημαντικό σημείο αναφοράς για την αντιστοίχιση, αφού καθορίζει τις τιμές που ανταποκρίνονται στο 1 της κλίμακας. Για αυτό το λόγο μπορεί να χρησιμοποιηθεί είτε η ελάχιστη τιμή πιθανοφάνειας, κάτι που όμως εξαρτάται από τα δεδομένα μας και δεν μπορεί να αποτελέσει γενική λύση για μια τέτοια εφαρμογή που πρέπει να μπορεί να δέχεται άγνωστα δεδομένα, είτε το 0 ή μία τιμή που καθορίζεται από τη μέγιστη τιμή διαιρεμένη με το 10. Επειδή επίσης, σε κάθε μια από τις 4 διαφορετικές κατηγορίες δεδομένων έχουμε διαφορετικά εύρη τιμών για τις πιθανότητες που βρίσκουμε, το μέγιστο και το ελάχιστο θα καθορίζεται διαφορετικά σε κάθε κατηγορία δεδομένων,

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ



(α')



(β')

Σχήμα 5.11: Διάγραμμα για *pca* δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 καταστάσεις (states-s) και 1-4 μείγματα Γκαουσιανών (gaussians-g).

ενώ καθώς σε κάποιες κατηγορίες έχουμε την ύπαρξη αρνητικών τιμών, θα δουλεύουμε

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ

|        | srodan | nidan | sandan | yondan | godan |
|--------|--------|-------|--------|--------|-------|
| srodan | 14     | 12    | 0      | 0      | 0     |
| nidan  | 1      | 25    | 0      | 0      | 0     |
| sandan | 4      | 1     | 12     | 0      | 3     |
| yondan | 0      | 2     | 0      | 16     | 2     |
| godan  | 0      | 0     | 0      | 1      | 12    |

(α')

|        | srodan | nidan | sandan | yondan | godan |
|--------|--------|-------|--------|--------|-------|
| srodan | 16     | 9     | 0      | 0      | 0     |
| nidan  | 4      | 20    | 0      | 1      | 0     |
| sandan | 1      | 3     | 12     | 1      | 2     |
| yondan | 1      | 1     | 0      | 17     | 0     |
| godan  | 0      | 0     | 0      | 1      | 11    |

(β')

Σχήμα 5.12: Πίνακες σύγχυσης (confusion matrices) για *pca* δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα για τις συνολικά καλύτερες ακρίβειες που βρέθηκαν.

με τις απόλυτες τιμές των πιθανοτήτων. Η ύπαρξη αυτών των αρνητικών τιμών εξηγείται κυρίως λόγω ασυμφωνίας του μοντέλου που επιθυμούμε να εξετάσουμε με τον τύπο των δεδομένων που εισάγουμε. Επιπροσθέτως, λόγω της λογαριθμικής φύσης των δεδομένων, πέρα από τη γραμμική αντιστοίχιση, θα δοκιμάσουμε και κάποιες εκθετικές μορφές για να αντιστρέψουμε τη μη γραμμικότητά τους. Συνεπώς, η γενική συνάρτηση αντιστοίχισης που δοκιμάζουμε είναι:

$$score(x) = 10 \frac{\logprob|x|^a - b^a}{\max[\logprob|x]^a - b^a} \quad (5.1)$$

όπου **a** είναι εκθετικός παράγοντας: για a=1 έχουμε γραμμική αντιστοίχιση, ενώ αν επιλέξουμε εκθετική μορφή, μέσω δοκιμών βρίσκουμε πως μια καλή τιμή είναι a=1,2 και **b** είναι η τιμή που επιλέγουμε να αντιστοιχίζεται στο κατώτερο σημείο της κλίμακας και τιμές που επιλέγουμε για αυτό, όπως είπαμε θα είναι 0,  $\min[\logprob|x]$  και  $\max[\logprob|x]/10$ .

Έτσι, προκύπτουν οι εξής 6 διαφορετικές κατηγορίες συναρτήσεων:

- I : a=1 b=0
- II : a=1,2 b=0
- III : a=1 b= $\min[\logprob|x]$
- IV : a=1,2 b= $\min[\logprob|x]$
- V : a=1 b= $\max[\logprob|x]/10$
- VI : a=1,2 b= $\max[\logprob|x]/10$

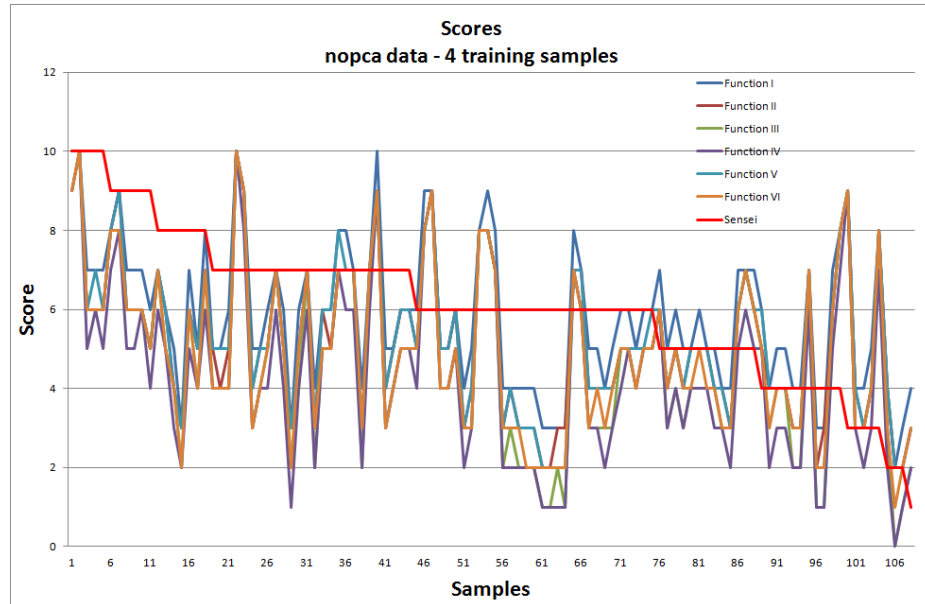


### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ

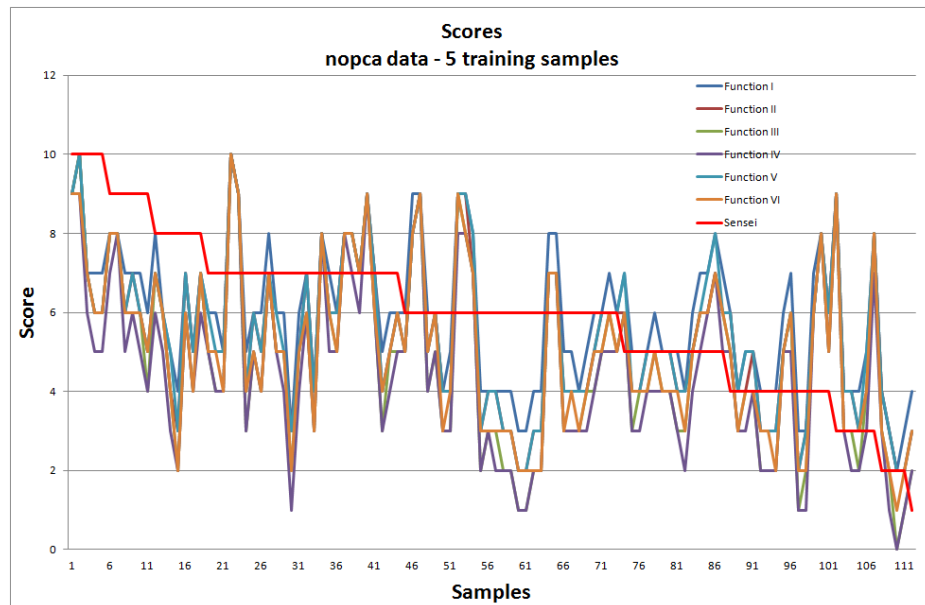
---

Παρακάτω έχουμε τις γραφικές παραστάσεις για τις 2 κατηγορίες δεδομένων που παρουσίασαν τη μεγαλύτερη συνολική ακρίβεια, *porca* και *pcahog*, και ξεχωριστά για τις περιπτώσεις εκπαίδευσης με 4 ή 5 δείγματα, παρουσιάζοντας τις βαθμολογίες που παίρνουμε χρησιμοποιώντας τις παραπάνω συναρτήσεις για το σύνολο των 124 δειγμάτων, αφαιρώντας όμως τις περιπτώσεις λανθασμένης αναγνώρισης. Παράλληλα, έχουμε τη γραφική παράσταση της βαθμολογίας που δόθηκε από τον δάσκαλο (σημειωμένη με κόκκινο χρώμα) για όλα τα δείγματα, ώστε να μπορέσουμε να συγκρίνουμε το βαθμό σύγκλισης των πειραματικών βαθμών με αυτήν.

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ



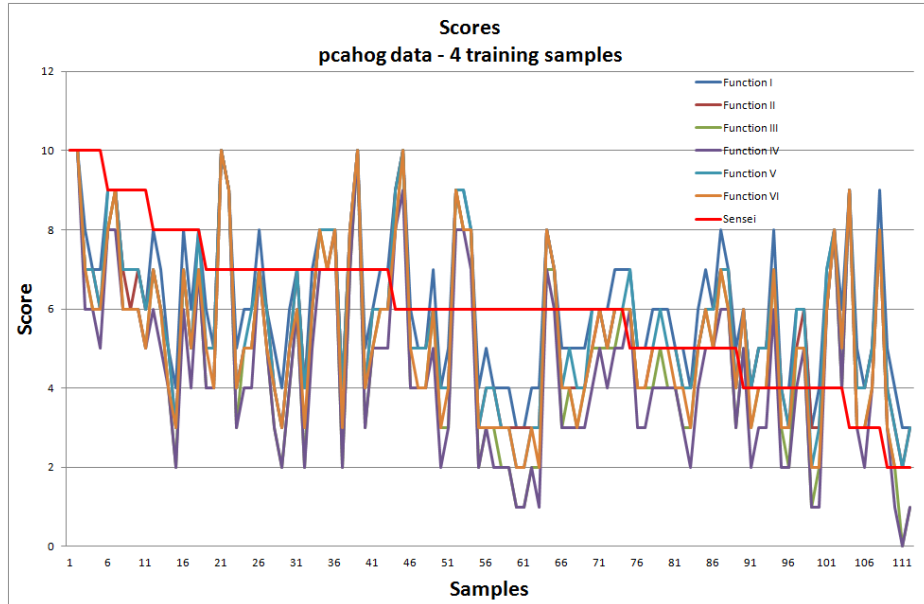
(α')



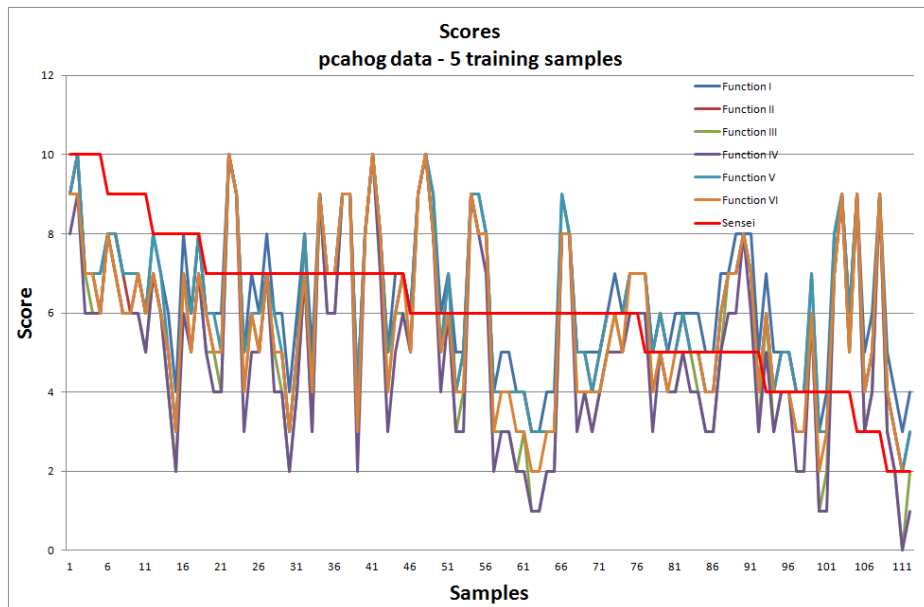
(β')

Σχήμα 5.13: Διαγράμματα με τις βαθμολογίες των δειγμάτων σύμφωνα με τις 6 συναρτήσεις αντιστοίχισης για *pca* δεδομένα και εκπαίδευση με (α') 4 και (β') 5 δείγματα. Η βαθμολογία του δασκάλου επισημαίνεται με κόκκινο.

### 5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ



( $\alpha'$ )



( $\beta'$ )

Σχήμα 5.14: Διαγράμματα με τις βαθμολογίες των δειγμάτων σύμφωνα με τις 6 συναρτήσεις αντιστοίχισης για *pcahog* δεδομένα και εκπαίδευση με ( $\alpha'$ ) 4 και ( $\beta'$ ) 5 δείγματα. Η βαθμολογία του δασκάλου επισημαίνεται με κόκκινο.

### *5.3. ΔΟΚΙΜΕΣ & ΑΠΟΤΕΛΕΣΜΑΤΑ*

---

# Κεφάλαιο 6

## Συμπεράσματα

### 6.1 Αξιολόγηση Αποτελεσμάτων & Συστήματος

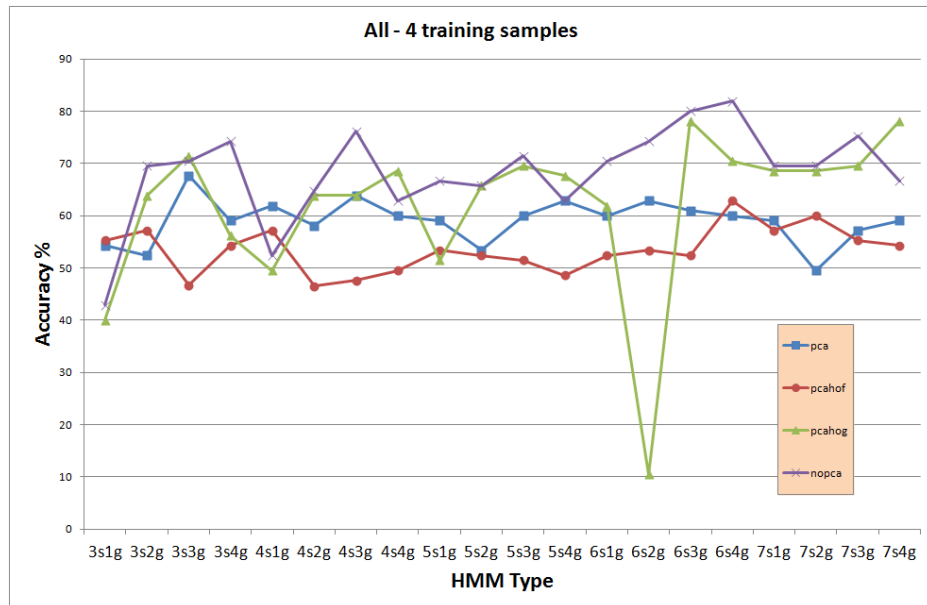
Με την ολοκλήρωση της πειραματικής μας διαδικασίας και την εξαγωγή των αποτελεσμάτων, μπορούμε να προχωρήσουμε στο σχολιασμό τους, τη σύγκρισή τους ως προς τα αναμενόμενα που αναπτύχθηκαν κατά τη θεωρητική προσέγγιση, αλλά και τέλος, την αξιολόγηση του συνολικού συστήματος που κατασκευάσαμε.

#### 6.1.1 Αξιολόγηση Αποτελεσμάτων Πειραματικής Διαδικασίας

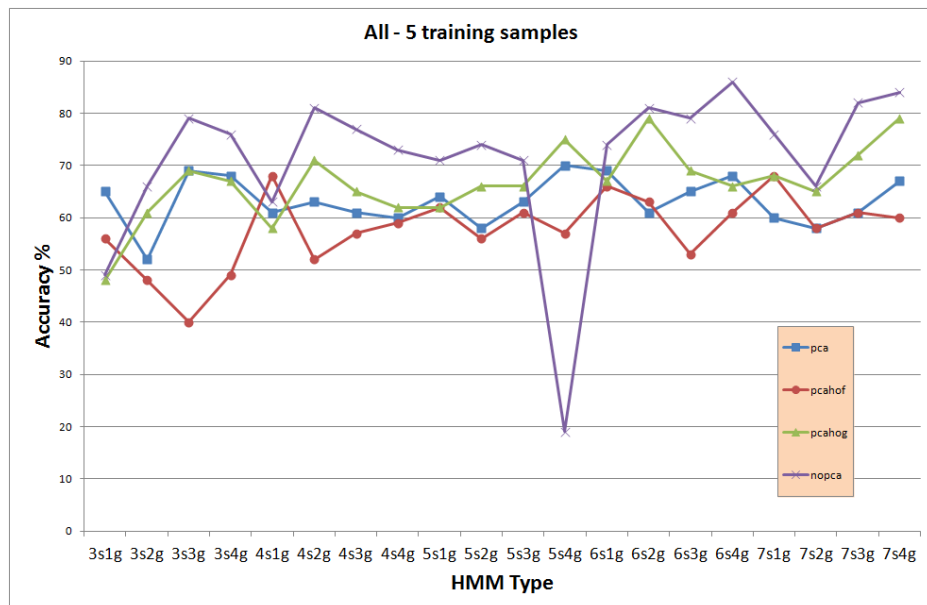
Όπως είδαμε στην Παρ.5.3, τα αποτελέσματα των δοκιμών μας δείχνουν πως το σύστημά μας αναγνωρίζει με ένα ικανοποιητικό ποσοστό ακρίβειας τα 'άγνωστα' δείγματα στην είσοδο, με ακρίβεια όμως που εξαρτάται σε μεγάλο βαθμό από την κατηγορία δεδομένων που χρησιμοποιούμε. Έτσι, ανάλογα με το αν έχουμε εφαρμόσει την PCA ανάλυση στα δεδομένα μας και ειδικότερα ανάλογα με το είδος των χαρακτηριστικών, HOG, HOF ή και στα δύο, παρουσιάζονται αξιοσημείωτες μεταβολές στη συνολική ακρίβεια, αλλά και στις επιμέρους. Αυτό μπορούμε να το διαπιστώσουμε με μεγαλύτερη ευκρίνεια στα διαγράμματα 6.1(α') και 6.1(β'), όπου παρατηρούμε πως σε γενικότερο επίπεδο, η κατηγορία δεδομένων *norca* παρουσιάζει την καλύτερη συμπεριφορά, ακολουθούμενη από την κατηγορία *pcahog*. Στη συνέχεια, κάπως χειρότερη συνολική ακρίβεια παίρνουμε με την *pca*, ενώ τη χειρότερη συμπεριφορά έχουμε χρησιμοποιώντας δεδομένα της *pcahof* κατηγορίας. Τα αποτελέσματα αυτά είναι κοινά και για τις δύο περιπτώσεις εκπαίδευσης με 4 ή 5 δείγματα, αλλά υποστηρίζονται και από τις ακρίβειες που βρίσκουμε αργότερα κατά την εποπτική μέθοδο εύρεσης συνδυασμού μοντέλων για κάθε κατηγορία κίνησης, αφού οι κατηγορίες *norca* και *pcahog* μας δίνουν την καλύτερη συνολική ακρίβεια 88-89%, ακολουθούμενες από την *pca* με 76% και από την *pcahof* με 70%.

Η καλή συμπεριφορά των δεδομένων που δεν έχουν επεξεργαστεί μέσω της PCA ανάλυσης είναι κάτι που συμπίπτει με αυτό που αναμέναμε θεωρητικά, καθώς σίγουρα αν εκπαιδευτούν τα μοντέλα με ολόκληρη την πληροφορία που περιέχεται στα δεδομένα από τα

## 6.1. ΑΞΙΟΛΟΓΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ & ΣΥΣΤΗΜΑΤΟΣ



(α')



(β')

Σχήμα 6.1: Συγκριτικό διάγραμμα με όλες τις κατηγορίες δεδομένων και εκπαίδευση με (α') 4 και (β') 5 δείγματα με τις συνολικές ακρίβειες για τη δοκιμή HMM ίδιων για όλες τις κατηγορίες με 3-7 s και 1-4 g.

Ιστογράμματα θα αναπαριστούν καλύτερα τις ακολουθίες κινήσεων. Ενδιαφέροντα είναι

η παρατήρηση της πολύ καλής συμπεριφοράς των δεδομένων στα οποία τα αρχικά χαρακτηριστικά από Ιστογράμματα Προσανατολισμένης Κλίσης, HOG, έχουν αντικατασταθεί από τα μετασχηματισμένα μέσω PCA, σε βαθμό που σε κάποιες περιπτώσεις ξεπερνά την ακρίβεια των αρχικών δεδομένων. Αυτό μας δείχνει πως ο συγκεκριμένος τύπος δεδομένων ανταποκρίνεται πολύ καλά στη μείωση της διαστασιμότητας μέσω της PCA ανάλυσης, έτσι ώστε ακόμα και μόνο το 90% της διασποράς-πληροφορίας να μεταφέρεται, δεν χάνεται ο διακριτικός χαρακτήρας που προσδίδουν στο σύστημα. Μάλιστα, σε περιπτώσεις όπου έχουμε λίγα δείγματα για εκπαίδευση, όπως στην περίπτωση με τα 4 δείγματα, μπορεί να εμφανίσει και καλύτερα αποτελέσματα από τα αρχικά δεδομένα, καθώς το μικρότερο διάνυμα χαρακτηριστικών σημαίνει πως έχουμε και λιγότερες παραμέτρους να προσδιορίσουμε κατά την εκπαίδευση των μοντέλων μας. Έτσι, στην περίπτωση των 4 δειγμάτων, βρίσκουμε συνδυασμό που φτάνει το 88,57% για την *pcahog* κατηγορία με 84,76% για την *porca*, ενώ για τα 5 δείγματα, πετυχαίνουμε και στις δύο περιπτώσεις 80%.

Στον αντίποδα, τα δεδομένα που παρουσιάζουν τη χειρότερη συμπεριφορά είναι αυτά στα οποία αντικαθίστανται κατά τον μετασχηματισμό τους με PCA τα χαρακτηριστικά από Ιστογράμματα Οπτικής Ροής, HOF. Για την κατηγορία *pcahof* λαμβάνουμε τις μικρότερες συνολικές ακρίβειες κατά τις δοκιμές μας, που σε καμία περίπτωση δεν ξεπερνούν το 70%. Συνεπώς, καταλαβαίνουμε πως τα HOF χαρακτηριστικά δεν είναι ιδανικά για τη μέθοδο των Πρωτευουσών Συνιστωσών, κάτι που μπορεί να οφείλεται στον δυναμικό χαρακτήρα που περιέχουν, αφού περιγράφουν τη σχετική κίνηση ανάμεσα στα διαδοχικά καρέ των βίντεο. Τα δεδομένα *rsa* στα οποία και τα δύο είδη χαρακτηριστικών από Ιστογράμματα, HOG και HOF, έχουν αντικατασταθεί από τους μετασχηματισμούς τους φαίνεται να αποτελεί μια ενδιάμεση κατάσταση, αφού από πλευράς επίδοσης βρίσκεται ανάμεσα στις παραπάνω περιπτώσεις. Αυτό μπορεί να εξηγηθεί αν θεωρήσουμε πως η καλή συμπεριφορά των HOG μετά από PCA εξουδετερώνεται από την κακή συμπεριφορά των HOF και το αποτέλεσμα είναι μέτριο. Βέβαια, αν σκεφτούμε πως από διάνυμα χαρακτηριστικών μήκους 167 καταλήγουμε σε μήκος 52 με αυτή τη μέθοδο, δηλαδή λιγότερο από το 1/3 των αρχικών μεταβλητών, συνολικές ακρίβειες κοντά στο 76% δεν είναι πολύ άσχημη επίδοση.

Κάνοντας μία ακόμα σύγκριση με βάση τα διαγράμματα 6.1, μπορούμε να αναλύσουμε τη διαφορετική συμπεριφορά του συστήματός μας ύστερα από την εκπαίδευση των μοντέλων μας με 4 και με 5 δείγματα. Βλέπουμε πως γενικά, η διακριτική ικανότητα του συστήματος βελτιώνεται με την προσθήκη ενός ακόμα δείγματος στη φάση της εκπαίδευσης, ενώ το συμπέρασμα αυτό ενισχύεται και από τη μέθοδο εποπτικής ανεύρεσης καλύτερου συνδυασμού που χρησιμοποιήσαμε και η οποία έδειξε για όλες τις κατηγορίες βελτίωση των αποτελεσμάτων από τα 4 στα 5 δείγματα. Αυτό συμβαίνει κατά κύριο λόγο επειδή έχουμε πολύπλοκα μοντέλα με πολλές παραμέτρους που πρέπει να προσδιοριστούν για κάθε μοντέλο και ίσως τα 4 δείγματα να είναι λίγα. Για αυτό το λόγο κιόλας, η βελτίωση στην κατηγορία *rsa* δεδομένων δεν έχουμε σημαντική βελτίωση, αφού το διάνυμα χαρακτηριστικών είναι μήκους 52, οπότε τα 4 δείγματα δεν είναι ιδιαίτερα λίγα. Φυσικά, είναι αναμενόμενη μια τέτοια συμπεριφορά αύξησης απόδοσης με περισσότερα δείγματα για εκπαίδευση, καθώς τα μοντέλα γίνονται πιο εύρωστα και αντιπροσωπευτικά ανεξάρ-

τητα από τους διαφορετικούς εκτελεστές της κίνησης. Ο λόγος που δεν χρησιμοποιήσαμε παραπάνω δείγματα για την εκπαίδευση εξαρχής ήταν πως θέλαμε να δημιουργήσουμε αυστηρά πρότυπα για την κίνηση με τα δείγματα καλύτερης βαθμολογίας κάθε είδους, ώστε να έχουμε και τον εκπαιδευτικό χαρακτήρα που αναλύσαμε. Βέβαια, πρωταρχικός στόχος μας παραμένει η καλύτερη επίδοση του συστήματος, οπότε τείνουμε προς την εκπαίδευση με περισσότερα δείγματα.

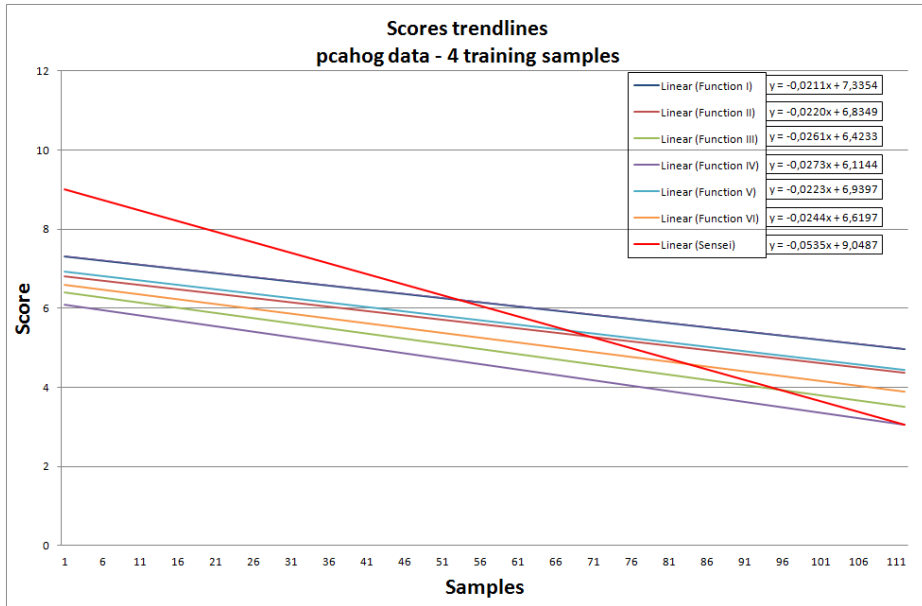
Μια επιπλέον σημαντική παρατήρηση που μπορούμε να κάνουμε με βάση τους καλύτερους συνδυασμούς HMM για τα 5 είδη κίνησης, είναι πως οι περισσότεροι συνδυασμοί που παράγουν τα καλύτερα αποτελέσματα, τόσο στην περίπτωση των ίδιων HMM για όλα τα είδη όσο και για τους πιο σύνθετους συνδυασμούς, περιέχουν τον τύπο μοντέλου 6s4g, δηλαδή με 6 καταστάσεις και μείγμα 4 Γκαουσιανών για κάθε κατάσταση. Αυτό παρατηρείται καλύτερα για την κατηγορία *porca* δεδομένων, κάτι που συμφωνεί με τη θεωρητική μας πρόβλεψη πριν την πειραματική διαδικασία όπου παρατηρήσαμε πως οι περισσότερες κινήσεις αποτελούνται από 6 διακριτά τμήματα, οπότε και υποθέσαμε μοντέλο 6 καταστάσεων. Για τα δεδομένα λοιπόν που δεν έχουν υποστεί επεξεργασία μέσω PCA, είναι λογικό να έχουμε καλύτερη συμπεριφορά μοντέλων 6 καταστάσεων. Ειδικότερα, όμως, μοντέλα με μείγμα 4 Γκαουσιανών ανά κατάσταση φαίνεται να παρουσιάζουν καλύτερη συμπεριφορά ανεξάρτητα από την κατηγορία δεδομένων, οπότε επαληθεύεται η πρόβλεψη για μεγαλύτερη αξιοπιστία μοντέλων που έχουν μεγαλύτερη συνθετότητα από αυτή την άποψη, ώστε να καλύπτουν πολλούς διαφορετικούς τρόπους εκτέλεσης των ακολουθιών κινήσεων.

Ακολούθως, αναλύοντας τα δεδομένα που παίρνουμε από τους πίνακες σύγχυσης (*confusion matrices*), μπορούμε να εξάγουμε κάποια συμπεράσματα για την ποιότητα των μοντέλων μας και τις μεταξύ τους αλληλεξαρτήσεις, που μαρτυρούν και στοιχεία για τα ίδια τα είδη κινήσεων που εξετάζουμε. Έτσι, βλέπουμε πως τα μοντέλα για τις κινήσεις "heian shodan" και "heian nidan" συνήθως συγχέονται από το σύστημα αναγνώρισης, λόγω κάποιας εγγενούς ομοιότητας. Επίσης, κατά τη φάση της δοκιμής διαφορετικών συνδυασμών, παρατηρήσαμε πως συνήθως με αύξηση της επιμέρους ακρίβειας του ενός μοντέλου είχαμε πτώση της ακρίβειας του άλλου, που σημαίνει πως αυτά τα δύο μοντέλα αλληλοεξαρτώνται. Αυτό μπορεί να εξηγηθεί αν αναλογιστούμε πως και οι ακολουθίες κινήσεων για αυτές τις δύο κινήσεις έχουν αρκετά κοινά σημεία, οπότε αν χτίζουμε ένα ισχυρό μοντέλο αναγνώρισης για το ένα είδος, τότε κάποια δείγματα του άλλου είδους είναι πιθανό να κατηγοριοποιηθούν λάθος. Μια περαιτέρω παρατήρηση που μπορούμε να κάνουμε είναι πως το μοντέλο της κίνησης "heian sandan" έχει γενικά τη χειρότερη συμπεριφορά από τα υπόλοιπα, χάνοντας αρκετά συχνά δείγματα προς τις άλλες κατηγορίες, κάτι που μπορεί να σημαίνει πως τα δείγματα για αυτό το είδος δεν ήταν τα καλύτερα δυνατά ή ακόμα πως η ίδια η κίνηση δεν έχει πολλά μοναδικά χαρακτηριστικά ώστε να διαφοροποιείται πάντα από τα άλλα είδη με βάση τα χαρακτηριστικά που χρησιμοποιούμε. Αντίθετα, οι κινήσεις "heian yondan" και "heian godan" έχουν γενικά καλή συμπεριφορά.

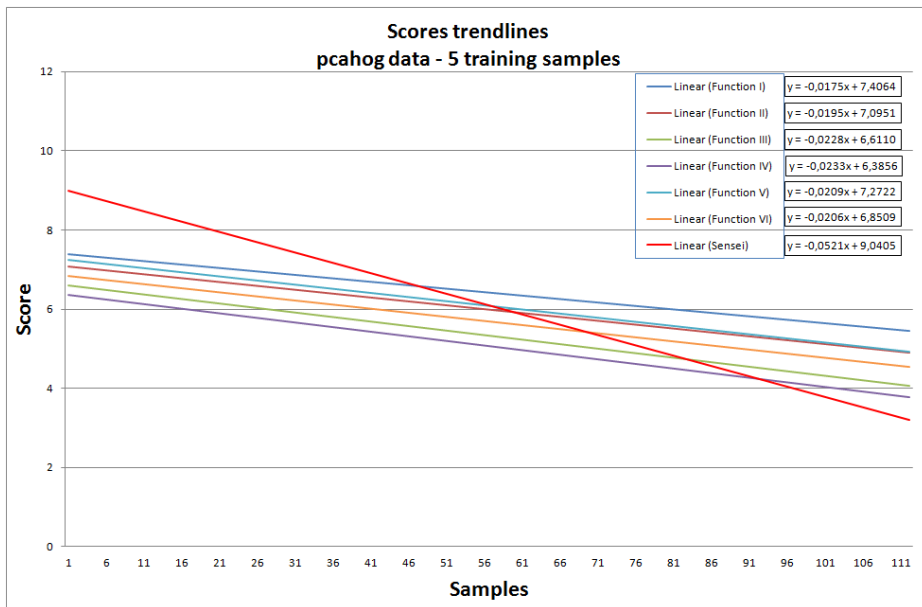
Τέλος, αναλύοντας τα διαγράμματα που προκύπτουν από τη σύγκριση της εξαγόμενης πειραματικά βαθμολογίας με τη βαθμολογία που δόθηκε από τον δάσκαλο καράτε για τις δύο καλύτερες κατηγορίες δεδομένων *porca* και *pcahog*, τόσο για την περίπτωση των 4 όσο και των 5 δειγμάτων κατά την εκπαίδευση, μπορούμε να συμπεράνουμε κάποια



## 6.1. ΑΞΙΟΛΟΓΗΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ & ΣΥΣΤΗΜΑΤΟΣ



(α')



(β')

Σχήμα 6.2: Συγκριτικά διαγράμματα για την κατηγορία δεδομένων *pcahog* με (α') 4 και (β') 5 δείγματα με τις καμπύλες γραμμικής παρεμβολής για τα 6 είδη συναρτήσεων αντιστοίχισης βαθμολογίας σε σχέση με την καμπύλη γραμμικής παρεμβολής για τις βαθμολογίες από τον δάσκαλο.

στοιχεία για τον εκπαιδευτικό χαρακτήρα της εφαρμογής. Γενικά, παρατηρούμε πως οι συναρτήσεις αντιστοίχισης μοιάζουν αρκετά μεταξύ τους ως προς τη συμπεριφορά, χωρίς όμως καμία να προσεγγίζει απόλυτα την καμπύλη που αναπαριστά τη βαθμολογία που δόθηκε από τον δάσκαλο. Παρότι οι συναρτήσεις ακολουθούν μία σχετικά καθοδική πορεία που πλησιάζει την καθοδική πορεία της καμπύλης βαθμολογίας του δασκάλου, η ομαλή τους πορεία ταρασσεται από διάφορα δείγματα που κυρίως λαμβάνουν υψηλότερη βαθμολογία από ότι θα έπρεπε. Αυτό οφείλεται κυρίως στο γεγονός ότι χρησιμοποιήθηκαν και δείγματα χαμηλότερης βαθμολογίας από '10' για την εκπαίδευση των μοντέλων-προτύπων, με αποτέλεσμα κάποια δείγματα χαμηλής ακρίβειας εκτέλεσης της κίνησης να λαμβάνουν υψηλή βαθμολογία, ακόμα και '10', σε αντίθεση με τη βαθμολογία του δασκάλου, λόγω ομοιότητας με τις μη τέλειες κινήσεις που χρησιμοποιήθηκαν στην εκπαίδευση.

Η συμπεριφορά αυτή φαίνεται ακόμα πιο έντονα στην περίπτωση που χρησιμοποιούνται 5 δείγματα κατά την εκπαίδευση, αφού εντάσσουμε ακόμα ένα μη-τέλειο, ακόμα χαμηλότερης βαθμολογίας δείγμα για εκπαίδευση του τέλειου προτύπου. Από τις διάφορες συναρτήσεις που δοκιμάζουμε, καλύτερη συμπεριφορά δείχνει να έχει η συνάρτηση τύπου V:  $a=1, b=\max[\log\text{prob}|x|]/10$ , αλλά και η VI:  $a=1,2, b=\max[\log\text{prob}|x|]/10$  καθώς πέρα από τα περιστασιακά καρφιά (spikes) ακολουθούν καλύτερα από τις υπόλοιπες τη γραμμή πρότυπο. Αυτό μπορεί να εξακριβωθεί από τα διαγράμματα 6.2, στα οποία συγκρίνουμε τις καμπύλες γραμμικής παρεμβολής, εξαγόμενες μέσω μεθόδου ελαχίστων τετραγώνων, που αντιστοιχούν στις συναρτήσεις αντιστοίχισης με την καμπύλη, εξαγόμενη με τον ίδιο τρόπο, που αντιστοιχεί στη βαθμολογία του δασκάλου. Από τη σύγκριση βλέπουμε πως η κλίση των καμπυλών αυτών είναι όντως καθοδική, αλλά δεν είναι σε τέτοιο βαθμό όσο η κλίση αυτής που αντιστοιχεί στον δάσκαλο, ενώ αυτές με τη μεγαλύτερη κλίση είναι οι συναρτήσεις III και IV, που όμως όπως είπαμε δεν μπορούν να χρησιμοποιηθούν ως γενική λύση. Έτσι, καταλήγουμε στις συναρτήσεις V και VI, που είναι οι αμέσως καλύτερες, ενώ βλέπουμε πως στην περίπτωση με 5 δείγματα προς εκπαίδευση, έχουμε μικρότερη σύγκλιση με την καμπύλη πρότυπο. Σε περίπτωση που έχουμε ιδανικότερα δείγματα, αναμένεται μεγαλύτερη προσέγγιση προς την κλίση της βαθμολογίας από τον δάσκαλο. Πάντα βέβαια κρατούμε και επιφυλάξεις για αυτή τη βαθμολογία που μας δόθηκε, καθώς εναπόκειται και σε υποκειμενικά κριτήρια.

### 6.1.2 Συνολική Αξιολόγηση Συστήματος

Στα πλαίσια της Διπλωματικής αυτής Εργασίας, κατασκευάσαμε ένα καινούργιο σύστημα για Ανίχνευση και Κατηγοριοποίηση Ανθρωπίνων Κινήσεων, που στον πυρήνα του έχει χαρακτηριστικά από Ιστογραμμικές μεθόδους, δηλαδή στατιστικά στοιχεία από Ιστογράμματα Προσανατολισμένης Κλίσης (HOG) και Ιστογράμματα Οπτικής Ροής (HOF). Τα χαρακτηριστικά αυτά εξάγονται για κάθε δείγμα κίνησης και τροφοδοτούν HMM μοντέλα για την εκπαίδευση αρχικά και στη συνέχεια για δοκιμή του συστήματος. Μέσω της πειραματικής μας διαδικασίας πάνω σε δείγματα κινήσεων που βιντεοσκοπήθηκαν για χάρη αυτής της εργασίας, δοκιμάζουμε διαφορετικούς τύπους επεξεργασίας των χαρακτηριστικών μας μέσω εφαρμογής PCA ανάλυσης, διαφορετικούς συνδυασμούς από

είδη μοντέλων για την αναπαράσταση των 5 ειδών κινήσεων που εξετάζουμε και διαφορετικούς διαχωρισμούς των δειγμάτων προς εκπαίδευση-δοκιμή. Ως αποτέλεσμα, διαπιστώνουμε την καλή συμπεριφορά των μοντέλων που δημιουργούμε, τα οποία ειδικά για τις περιπτώσεις δεδομένων που δεν έχουν επεξεργαστεί μέσω PCA ή που ο μετασχηματισμός με PCA εφαρμόζεται στα HOG χαρακτηριστικά, μας δίνουν ποσοστά συνολικής ακρίβειας της τάξης του **88-89%**. Ένα τέτοιο ποσοστό χαρακτηρίζεται ως αρκετά ικανοποιητικό για την πειραματική μας διαδικασία, αφού η δοκιμή έγινε για πολύπλοκες ακολουθίες κινήσεων, όπως είναι τα *katas* από την πολεμική τέχνη του karate.

Μάλιστα, είμαστε πεπεισμένοι πως τα ποσοστά ακρίβειας μπορούν να ξεπεράσουν σίγουρα το 90% στη μορφή του συστήματος που προτείνουμε, αν διερευνηθούν ακόμα περισσότερο κάποιες παράμετροι που επηρεάζουν άμεσα την απόδοσή του, όπως οι σύνθετοι συνδυασμοί από είδη HMM για τα είδη των κινήσεων μας που διερευνήθηκαν εποπτικά, ενώ μπορεί να γίνει εξαντλητική διερεύνησή τους με τα κατάλληλα υπολογιστικά μέσα. Επίσης, μπορούν να διερευνηθούν και εναλλακτικές μορφές εφαρμογής της PCA ανάλυσης ή χρήσης του εργαλείου HTK, το οποίο χρησιμοποιήθηκε στην πολύ βασική μορφή του, που ενδέχεται να βελτιώνουν ακόμα περισσότερο την αποδοτική λειτουργία του συστήματός μας. Επιπροσθέτως, πετυχαίνουμε αυτά τα ποσοστά ακρίβειας έχοντας μειώσει το μήκος του διανύσματος χαρακτηριστικών από 167 που ήταν αρχικά σε 119 μέσω PCA μετασχηματισμού των HOG χαρακτηριστικών που κρατάει μόνο το 90% της διακύμανσης-πληροφορίας τους, κάτι που μειώνει αρκετά την υπολογιστική πολυπλοκότητα του συστήματος, αλλά και τον χρόνο εκτέλεσης της αναγνώρισης αφού ο αριθμός των μεταβλητών που πρέπει να προσδιοριστούν σε κάθε στάδιο ελλατώνεται σημαντικά.

Συμπερασματικά, κρίνουμε πως η ερευνητική μας προσπάθεια έχει ικανοποιητικά αποτελέσματα και πως η απόδοση ως προς την ανίχνευση και κατηγοριοποίηση νέων δειγμάτων είναι αρκετά καλή με καλές προοπτικές βελτίωσης. Σημαντικό ρόλο για τον χαρακτηρισμό ως ικανοποιητικών για τα αποτελέσματα ακρίβειας διαδραματίζει το γεγονός πως τα δείγματα που συλλέχθηκαν μέσω βιντεοσκόπησης και απαρτίζουν τη βάση δεδομένων με την οποία διεξάγεται η πειραματική διαδικασία δεν είναι τέλειας ποιότητας και σε πολλές περιπτώσεις μπορούν να χαρακτηριστούν ως 'θορυβώδη', αφού δεν έχουμε ομοιόμορφο παρασκήνιο και σε πολλές περιπτώσεις υπάρχει παράλληλη κίνηση άλλων ατόμων, όπως του δασκάλου ή παρατηρητών, που χαρακτηρίζει ως σχετικά δύσκολο προς αναγνώριση το σύνολο των δειγμάτων μας. Έτσι, είμαστε σίγουροι πως το σύστημα μπορεί να ανταπεξέλθει με μεγαλύτερη αξιοπιστία σε απλούστερα προβλήματα που απαιτούν τη δημιουργία λιγότερο σύνθετων μοντέλων και για τα οποία έχουμε καλύτερη συλλογή δειγμάτων για τη φάση της εκπαίδευσης.

Αναφορικά προς τον εκπαιδευτικό χαρακτήρα και την προσπάθεια χρήσης αυτής της εφαρμογής ως βοήθημα εκμάθησης μιας πολεμικής τέχνης, που αποτελεί παράπλευρο στόχο της μελέτης αυτής, θεωρούμε πως τα προκαταρκτικά αποτελέσματα που λάβαμε μέσω της πειραματικής διαδικασίας μπορεί να μην είναι τέλεια είναι όμως ενθαρρυντικά. Παρά τα προβλήματα που αναλύσαμε πως υπήρξαν κατά την εκπαίδευση με τη χρήση μητέλειων δειγμάτων, που απέτρεψαν την κατεσκευή αυστηρών μοντέλων-προτύπων για το κάθε είδος κίνησης, βλέπουμε από τα διαγράμματα, πως γενικά έχουμε βαθμολογίες που ακολουθούν τη διάκριση του δασκάλου μεταξύ καλύτερης και χειρότερης εκτέλεσης των

κινήσεων από δείγμα σε δείγμα. Αυτή είναι και μία ένδειξη πως το σύστημά μας έχει τη διακριτική ικανότητα που απαιτείται για να λειτουργήσει ως βοήθημα εκμάθησης. Συνεπώς, με χρήση κυρίως δειγμάτων που πλησιάζουν όλες στα επιθυμητά πρότυπα των κινήσεων και ίσως με μία διερεύνηση για εύρεση καλύτερης συνάρτησης αντιστοίχισης των λογαριθμικών πιθανοτήτων σε βαθμολογίες, μπορούμε να διαπιστώσουμε και τη εκπαιδευτική λειτουργικότητα του συστήματος.

Πέρα όμως από το ίδιο το αποδοτικό σύστημα που δημιουργήσαμε, η διπλωματική αυτή εργασία συντέλεσε κινητήριο μοχλός για μία σειρά από παράπλευρα οφέλη. Αρχικά, είχαμε τη δημιουργία μίας καινούργιας Βάσης Δεδομένων με 124 δείγματα από 5 είδη είδη κινήσεων shotokan karate εμπρόσθια και πλαϊνής όψης (βιντεοσκοπήθηκαν και άλλες κινήσεις αλλά δεν χρησιμοποιήθηκαν κατά την εργασία αυτή), η οποία μπορεί να χρησιμοποιηθεί και σε μελλοντικές ερευνητικές προσπάθειες που ασχολούνται με το ίδιο αντικείμενο. Επίσης, πέρα από το κύριο κομμάτι της εξαγωγής χαρακτηριστικών, η μελέτη αυτή συνεισφέρει και με τη δημιουργία βοηθητικών κομματιών κώδικα για την ενοποίηση των διαφορετικών εργαλείων που χρησιμοποιούνται από το σύστημα. Έτσι, έχουμε ένα πρόγραμμα σε C για τη μετατροπή των αρχείων με δεδομένα σε μορφή κατάλληλη για ανάγνωση από το εργαλείο HTK, ένα πρόγραμμα για Matlab που αναλαμβάνει να πραγματοποιήσει τον μετασχηματισμό με βάση την PCA ανάλυση των χαρακτηριστικών που δέχεται και να κρατήσει ένα συγκεκριμένο ποσοστό διακύμανσης-πληροφορίας και ένα πρόγραμμα HTK γραμμένο για Linux συστήματα που εκτελεί ολόκληρη της διαδικασία εκπαίδευσης-δοκιμής των μοντέλων μας, καθώς και δοκιμές με διαφορετικούς συνδυασμούς ειδών HMM. Όλα τα παραπάνω μπορούν να χρησιμοποιηθούν ανεξάρτητα σε μελλοντικές ερευνητικές προσπάθειες. Ένα ακόμα σημαντικό στοιχείο είναι πως οι περισσότερες ερευνητικές προσπάθειες που κάνουν χρήση στατιστικών τοπικών χωροχρονικών χαρακτηριστικών (*statistical local spatio-temporal features*), όπως είναι τα χαρακτηριστικά από Ιστογράμματα Προσανατολισμένης Κλίσης (HOG) ή Οπτικής Ροής (HOF), δουλεύουν με μοντέλα τύπου SVM. Με αυτή τη μελέτη, παρουσιάσαμε ένα σύστημα που χρησιμοποιεί αυτά τα χαρακτηριστικά σε συνδυασμό με HMM που παρουσιάζει υποσχόμενα αποτελέσματα, ικανά να αποτελέσουν το έναυσμα για περαιτέρω έρευνα αυτού του συνδυασμού για τύπους προβλημάτων που χαρακτηρίζονται από δυναμικότητα.

## 6.2 Συνεισφορές Διπλωματικής Εργασίας

Συνοπτικά, μετά από την ανάλυση των παραπάνω, μπορούμε να αποφανθούμε για το σύνολο των συνεισφορών αυτής της ερευνητικής προσπάθειας:

- Απόδειξη αποδοτικότητας χρήσης Ιστογραφικών μεθόδων με την εξαγωγή τοπικών χώρο-χρονικών οπτικών χαρακτηριστικών HOG-HOF από βίντεο με ακολουθίες ανθρωπίνων κινήσεων
- Επιλογές περαιτέρω επεξεργασίας των παραπάνω χαρακτηριστικών για μείωση της διαστασιμότητας του προβλήματος και εξοικονόμηση χρόνου και υπολογιστικών πόρων

- Εκτίμηση παραμέτρων (καταστάσεις, μείγματα Γκαουσιανών) για συνεπή και ευέλικτα HMM μοντέλα που αναπαριστούν ακολουθίες ανθρωπίνων κινήσεων
- Δημιουργία νέας βάσης δεδομένων με κινήσεις Shotokan Karate, ώστε να μπορούν να χρησιμοποιηθούν για την εκπαίδευση και δοκιμή αποτελεσματικότητας του συστήματος, αλλά και για επόμενες ερευνητικές προσπάθειες
- Ανάπτυξη ενός νέου ολοκληρωμένου συστήματος για ανίχνευση και κατηγοριοποίηση ανθρωπίνων κινήσεων με χρήση των παραπάνω μεθόδων, με αρκετά αποδοτική συμπεριφορά που για τη βάση δεδομένων μας φτάνει ποσοστά ακρίβειας 89%
- Δημιουργία συνοδευτικών προγραμμάτων και ρουτινών σε προγραμματιστικά περιβάλλοντα C, Matlab, Linux Shell και HTK για την εξαγωγή και επεξεργασία χαρακτηριστικών, αλλά και τη δημιουργία, εκπαίδευση και δοκιμή των μοντέλων μας

## 6.3 Μελλοντικές Επεκτάσεις

Παρά την καλή συμπεριφορά του συστήματός μας, υπάρχει πάντα χώρος για βελτιστοποιήσεις, ώστε να επεκταθούν τα αποτελέσματα αυτής της ερευνητικής προσπάθειας και να διερευνηθεί σε μεγαλύτερη έκταση αυτός ο πραγματικά ενδιαφέρων τομέας προβλημάτων που σχετίζονται με την αναγνώριση πολεμικών τεχνών. Αρχίζοντας από αυτά που μπορούν να γίνουν εύκολα με τα στοιχεία που έχουν συγκεντρωθεί ήδη, μπορεί να γίνει χρήση της πλαϊνής λήψης που έχουμε επίσης εξασφαλίσει για κάθε δείγμα των κινήσεων. Μέσω της συνδυασμένης χρήσης εμπρόσθιας και πλαϊνής λήψης, μπορούμε να δημιουργήσουμε ένα καλύτερο και πιο αντιπροσωπευτικό μοντέλο για την αναπαράσταση των κινήσεων, αφού ακόμα και σε κινήσεις που η μπροστινή λήψη δεν μπορεί να συλλάβει το σύνολο των σχετικών κινήσεων, η πλαϊνή λήψη μπορεί να συμπληρώσει τις αναγκαίες πληροφορίες. Μάλιστα, μπορεί να χρησιμοποιηθεί ένα διασυνδεδεμένο (coupled) HMM για τις δύο όψεις, ώστε να επηρεάζουν το ένα το άλλο και να αλληλοσυμπληρώνονται.

Παράλληλα, μελλοντικές ερευνητικές προσπάθειες μπορούν να επιχειρήσουν να εμβαθύνουν ακόμα περισσότερο στη μελέτη των πολεμικών τεχνών και ιδιαίτερα του karate και να επιχειρήσουν τη δημιουργία ενός λεπτομερούς 'λεξιλογίου' για αυτή την πολεμική τέχνη. Λόγω της μορφολογίας των κινήσεων του karate, είναι δυνατή η περιγραφή μικρότερων τμημάτων των κινήσεων που εξετάσαμε σαν ενιαίες οντότητες ως διακριτές μικροκινήσεις, μία μορφή των αρχέτυπων κίνησης (*motion primitives*) που είδαμε και κατά τη θεωρητική μελέτη, για τις οποίες μάλιστα υπάρχει και ξεχωριστή ονομασία στην πολεμική αυτή τέχνη. Μάλιστα, πέρα από την ειδική ονομασία έχουμε και προσδιορισμό για τον προσανατολισμό και την ταχύτητα των μελών του ανθρωπίνων μελών κατά την εκτέλεσή τους, κάτι που μπορεί να βοηθήσει ουσιαστικά στον σαφή προσδιορισμό διακριτών μοντέλων σε συνδυασμό με τα τοπικά χωρο-χρονικά χαρακτηριστικά που χρησιμοποιούμε. Κατά αυτόν τον τρόπο, θα μπορούσαν να δημιουργηθούν σύνθετα μοντέλα αναγνώρισης

### 6.3. ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

---

αυτών μικρο-κινήσεων, τα οποία συνδυαζόμενα αποτελούν ένα είδος ακολουθίας κινήσεων, όπως οι διάφορες λέξεις που συνθέτουν μία πρόταση κατά την αναγνώριση φωνής. Μια τέτοια μελέτη βέβαια, ξεφεύγει από τα πλαίσια μιας απλής Διπλωματικής Εργασίας.

Μια επίσης προσπάθεια που μπορεί να γίνει είναι η επέκταση της βάσης δεδομένων με περισσότερα δείγματα ανά κατηγορία και αρκετά δείγματα υψηλής ποιότητας ακρίβειας, δηλαδή που συγκεντρώνουν βαθμολογία κοντά στο θεωρητικό '10', ώστε να δημιουργηθούν καλύτερα εκπαιδευμένα μοντέλα και να διερευνηθεί καλύτερα η πλευρά του μοντέλου που τροφοδοτεί το χρήστη με πληροφορίες για την πιστότητα της κίνησης που εκτελεί ως προς το πρότυπο. Προς την ίδια κατεύθυνση, για τη δοκιμή της γενικότερης χρήσης του συστήματος, άλλες ερευνητικές προσπάθειες μπορούν να εστιάσουν στη δοκιμή του για αναγνώριση περισσότερων κινήσεων καράτε, ώστε να γίνει εκτίμηση της δύναμης και προσαρμοστικότητας του συστήματος σε δυσκολότερα προβλήματα. Επίσης, πέρα από αυτή τη δοκιμή σε βάθος, μπορεί να γίνει και αντίστοιχη δοκιμή σε πλάτος, με την προσπάθεια αναγνώρισης κινήσεων από διαφορετικές πολεμικές τέχνες. Κατ' αυτόν τον τρόπο, μπορούμε να εξακριβώσουμε την ευρωστία και ευελιξία του συστήματος στην προσπάθεια να αποτελέσει ένα γενικότερο εργαλείο για τις πολεμικές τέχνες και ποιες αλλαγές είναι πιθανόν να χρειάζεται ώστε να είναι σε τέτοια θέση, όπως συμπληρωματική χρήση χαρακτηριστικών για το διάνυσμα ή χρήση πιο πολύπλοκων μοντέλων.

Τέλος, καθώς επιθυμούμε αυτό το σύστημα να αποκτήσει έναν εκπαιδευτικό χαρακτήρα, μια άμεση επέκταση αυτής της εργασίας θα ήταν η ανάπτυξη ενός διαδραστικού περιβάλλοντος, ώστε να είναι δυνατή η χρήση του κατά τη διάρκεια της προπόνησης των πολεμικών τεχνών από τον δάσκαλο ή τους μαθητές. Το περιβάλλον αυτό πρέπει να περιλαμβάνει οπτικά στοιχεία στοιχεία ελέγχου των λειτουργιών και στο τέλος, πέρα από το αποτέλεσμα της σύγκρισης με το πρότυπο, να παρουσιάζονται και στατιστικά στοιχεία για την επίδοση κατά τη διάρκεια της κίνησης, δηλαδή μια πιο τμηματική αντιμετώπιση της ακολουθίας της κίνησης, ώστε να γίνεται περισσότερο σαφές σε ποιο τμήμα ο αθλητής είχε καλύτερης ποιότητας κίνηση και σε ποιο χειρότερης. Πέρα από αυτό, παρά την ταχύτητα που έχει ο αλγόριθμός μας, μπορούν να γίνουν περαιτέρω βελτιστοποιήσεις με αποτέλεσμα το σύστημα να μπορεί να ανταποκρίνεται σε πραγματικό χρόνο (real time), δηλαδή να εμφανίζει τη σύγκριση με το πρότυπο πριν ακόμα ολοκληρωθεί η κίνηση, αλλά σταδιακά. Κάτι τέτοιο βέβαια, απαιτεί πάλι την τμηματοποίηση των κινήσεων, ώστε ιδέες ιεραρχικών μοντέλων και αρχέτυπων κινήσεων, τις οποίες είδαμε στο Κεφ.2, να μπορούν να βρουν εφαρμογή.

# Βιβλιογραφία

- [1] L. Bretzner, I. Laptev, and T. Lindeberg, "Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pp. 423--428, IEEE, 2002.
- [2] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla, "Model-based hand tracking using a hierarchical bayesian filter," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1372--1384, 2006.
- [3] A. Azarbayejani and A. Pentland, "Real-time self-calibrating stereo person tracking using 3-d shape estimation from blob features," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, vol. 3, pp. 627--632, IEEE, 1996.
- [4] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 3, pp. 257--267, 2001.
- [5] A. Efros, A. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," *ICCV*, 2003.
- [6] P. Felzenszwalb and D. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55--79, 2005.
- [7] R. Urtasun, D. Fleet, and P. Fua, "Monocular 3-d tracking of the golf swing," in *Computer Vision and Pattern Recognition, San Diego, California, USA*, IEEE Computer Society, 2005.
- [8] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2247--2253, 2007.
- [9] P. Dollar, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," in *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 65--72, IEEE, 2005.

- [10] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63--86, 2004.
- [11] I. Laptev, "On space-time interest points," *International Journal of Computer Vision*, vol. 64, no. 2, pp. 107--123, 2005.
- [12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, pp. I:886--893, IEEE Computer Society, 2005.
- [13] A. Klaser, M. Marszalek, and C. Schmid, "A spatio-temporal descriptor based on 3d-gradients," in *British Machine Vision Conference*, pp. 995--1004, 2008.
- [14] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance," *Computer Vision--ECCV 2006*, pp. 428--441, 2006.
- [15] N. Oliver, B. Rosario, and A. Pentland, "Graphical models for recognizing human interactions," *Advances in Neural Information Processing Systems*, pp. 924--930, 1999.
- [16] W. Ilg and M. Giese, "Modeling of movement sequences based on hierarchical spatial-temporal correspondence of movement primitives," in *Biologically Motivated Computer Vision*, pp. 19--22, Springer, 2010.
- [17] P. Hamalainen, T. Ilmonen, J. Hoysiemi, M. Lindholm, and A. Nykanen, "Martial arts in artificial reality," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 781--790, ACM, 2005.
- [18] R. Kauth, A. Pentland, G. Thomas, A. A. I. S. Environmental Research Inst. of Michigan, and A. Dept, "Blob: An unsupervised clustering approach to spatial preprocessing of mss imagery," in *Proc. of the 11 th Intern. Symp. on Remote Sensing of Environment*, 1977.
- [19] H. Yoon, J. Soh, Y. Bae, and H. Seung Yang, "Hand gesture recognition using combined features of location, angle and velocity," *Pattern Recognition*, vol. 34, no. 7, pp. 1491--1501, 2001.
- [20] H. Lee and J. Kim, "An hmm-based threshold model approach for gesture recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 961--973, 1999.
- [21] G. Rigoll, A. Kosmala, and S. Eickeler, "High performance real-time gesture recognition using hidden markov models," *Gesture and sign language in human-computer interaction*, pp. 69--80, 1998.
- [22] F. Bevilacqua, B. Zamborlin, A. Sypniewski, N. Schnell, F. Guédy, and N. Rasamimanana, "Continuous realtime gesture following and recognition," *Gesture in Embodied Communication and Human-Computer Interaction*, pp. 73--84, 2010.



- [23] L. Campbell, D. Becker, A. Azarbayejani, A. Bobick, and A. Pentland, "Invariant features for 3-d gesture recognition," in *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996.*, pp. 157--162, IEEE, 1996.
- [24] J. Davis and A. Bobick, "The representation and recognition of action using temporal templates," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 928--934, Citeseer, 1997.
- [25] P. Felzenszwalb, "Learning models for object recognition," in *Proceedings of the 2001 IEEE Computer Society Conference, CVPR 2001.*, vol. 1, pp. I--1056, IEEE, 2001.
- [26] P. Felzenszwalb, "Representation and detection of deformable shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 208--220, 2005.
- [27] M. Pawan Kumar, P. Torr, and A. Zisserman, "Extending pictorial structures for object recognition," *Science*, no. c, pp. 789--798, 2004.
- [28] M. Covell, A. Rahini, M. Harville, and T. Darrell, "Articulated-pose estimation using brightness-and depth-constancy constraints," in *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings*, vol. 2, pp. 438--445, IEEE, 2000.
- [29] R. Plankers and P. Fua, "Articulated soft objects for multiview shape and motion capture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1182--1187, 2003.
- [30] G. Junxia, D. Xiaoqing, W. Shengjin, and W. Youshou, "Full body tracking-based human action recognition," in *ICPR 2008.*, pp. 1--4, IEEE, 2008.
- [31] G. Taylor, L. Sigal, D. Fleet, and G. Hinton, "Dynamical binary latent variable models for 3d human pose tracking," in *Computer Vision and Pattern Recognition*, pp. 631--638, IEEE, 2010.
- [32] S. Gaffney and P. Smyth, "Trajectory clustering with mixtures of regression models," in *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 63--72, ACM, 1999.
- [33] J. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," *International Journal of Computer Vision*, vol. 79, no. 3, pp. 299--318, 2008.
- [34] J. Niebles and L. Fei-Fei, "A hierarchical model of shape and appearance for human action classification," in *IEEE Conference on CVPR'07.*, pp. 1--8, IEEE, 2007.
- [35] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local svm approach," in *Proceedings of the 17th ICPR*, vol. 3, pp. 32--36, IEEE, 2004.

- [36] I. Laptev, B. Caputo, C. Schmid, and T. Lindeberg, "Local velocity-adapted motion events for spatio-temporal recognition," *Computer Vision and Image Understanding*, vol. 108, no. 3, pp. 207--229, 2007.
- [37] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in *IEEE Conference on CVPR*, pp. 1--8, IEEE, 2008.
- [38] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2008.
- [39] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1627--1645, 2009.
- [40] T. Hastie and R. Tibshirani, "Discriminant adaptive nearest neighbor classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 607--616, 1996.
- [41] I. Laptev, "Improving object detection with boosted histograms," *Image and Vision Computing*, vol. 27, no. 5, pp. 535--544, 2009.
- [42] H. Wang, M. Ullah, A. Klaser, I. Laptev, and C. Schmid, "Evaluation of local spatio-temporal features for action recognition," in *British Machine Vision Conf. on*, 2009.
- [43] L. Rabiner and R. Schafer, *Digital processing of speech signals*, vol. 100. Prentice-hall Englewood Cliffs, NJ, 1978.
- [44] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden markov model," in *IEEE Computer Society Conference on CVPR'92, Proceedings*, pp. 379--385, IEEE, 1992.
- [45] M. Brand, N. Oliver, and A. Pentland, "Coupled hidden markov models for complex action recognition," in *cvpr*, p. 994, Published by the IEEE Computer Society, 1997.
- [46] D. Becker, *Sensei: a real-time recognition, feedback and training system for t'ai chi gestures*. PhD thesis, Citeseer, 1997.
- [47] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194--203, 2001.
- [48] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254--1259, 1998.
- [49] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83--105, 2001.

- [50] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," *Computer Vision-ECCV 2004*, pp. 228--241, 2004.
- [51] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79--116, 1998.
- [52] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," in *CVPR*, p. 465, Published by the IEEE Computer Society, 1996.
- [53] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, pp. 1615--1630, 2005.
- [54] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool, "A comparison of affine region detectors," *International journal of computer vision*, vol. 65, no. 1, pp. 43--72, 2005.
- [55] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *International journal of computer vision*, vol. 59, no. 1, pp. 61--85, 2004.