ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Μελέτη Σκελετικών Δεδομένων της Βάσης Αθλητικών Δράσεων THETIS και Εφαρμογή σε Αλγορίθμους Αναγνώρισης Κίνησης.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Χρυσούλας Α. Βαρηά

Επιβλέπων : Ανδρέας - Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2018

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Μελέτη Σκελετικών Δεδομένων της Βάσης Αθλητικών Δράσεων THETIS και Εφαρμογή σε Αλγορίθμους Αναγνώρισης Κίνησης.

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Χρυσούλας Α. Βαρηά

Επιβλέπων : Ανδρέας – Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 14η Φεβρουαρίου 2018.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Ανδρέας – Γεώργιος Σταφυλοπάτης Καθηγητής Ε.Μ.Π.	Κωνσταντίνος Καρπούζης Διευθυντής Ερευνών ΕΠΙΣΕΥ	Γεώργιος Στάμου Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2018

(Υπογραφή)

.....

Χρυσούλα Α. Βαρηά

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

Περίληψη

Στόχος της διπλωματικής εργασίας αποτελεί η μελέτη της απόδοσης των σκελετικών δεδομένων σε αλγορίθμους αναγνώρισης ανθρώπινης κίνησης και η σύγκριση με άλλες μορφές δεδομένων όπως είναι τα χωροχρονικά σημεία ενδιαφέροντος (Space Time Interest Points). Τα σκελετικά δεδομένα, τα οποία χρησιμοποιούνται για την παραπάνω μελέτη, εξάγονται από τη βάση ανθρώπινων δράσεων THETIS (Three dimensional TennIs Shots). Η συγκεκριμένη βάση περιλαμβάνει 8374 videos, τα οποία περιέχουν 12 κινήσεις του αθλήματος tennis εκτελεσμένες από 55 διαφορετικά άτομα (αρχάριους και έμπειρους παίκτες).

Η εξαγωγή των σκελετικών αρθρώσεων γίνεται με τη χρήση κατάλληλου προγράμματος, το οποίο αναπτύσσεται σε γλώσσα προγραμματισμού C++ με τη βοήθεια του λογισμικού OpenNI. Κάθε άρθρωση αποτελεί ένα τρισδιάστατο διάνυσμα συντεταγμένων στο χώρο. Συνολικά υπολογίζονται 15 αρθρώσεις του ανθρώπινου σώματος για κάθε πλαίσιο (frame) ενός video. Μετά τον υπολογισμό των 3D συντεταγμένων γίνεται οπτικοποίηση των δεδομένων μέσω κατάλληλου προγράμματος, το οποίο αναπτύσσεται σε Unity.

Στα πλαίσια της διπλωματικής εργασίας υλοποιείται, σε γλώσσα προγραμματισμού matlab, ένας δημοσιευμένος αλγόριθμος αναγνώρισης ανθρώπινης δραστηριότητας. Ο συγκεκριμένος αλγόριθμος δέχεται ως δεδομένα εισόδου τις 3D συντεταγμένες των αρθρώσεων και υπολογίζει κάποια διανύσματα ανθρώπινων στάσεων (Posture Feature Vectors). Από τις παραπάνω στάσεις επιλέγονται αυτές που θεωρούνται πιο αντιπροσωπευτικές με χρήση της μεθόδου k-means. Έπειτα εξάγονται τα διανύσματα δραστηριοτήτων (Activity Feature Vector) για κάθε κίνηση εισόδου. Αυτά χρησιμοποιούνται για την εκπαίδευση και αξιολόγηση SVM πολλαπλών κλάσεων με τα οποία γίνεται στη συνέχεια η ταξινόμηση των κινήσεων των παικτών tennis.

Τέλος τα εξαγόμενα σκελετικά δεδομένα εφαρμόζονται σε έναν αλγόριθμο, ο οποίος πραγματεύεται το επίπεδο εμπειρίας των παικτών της βάσης THETIS. Ο συγκεκριμένος αλγόριθμος χρησιμοποιείται για να αναγνωρίσει αν ένας παίκτης είναι αρχάριος ή έμπειρος, αναλύοντας τις κινήσεις εισόδου. Αυτό επιτυγχάνεται με τη χρήση διανυσμάτων απόκλισης (Variance Vector) και απόστασης συνημιτόνου (Cosine Distance Vector), καθώς αξιοποιείται επίσης η μέθοδος Dynamic Time Warping για την ευθυγράμμιση των παραγόμενων χρονικών ακολουθιών.

Λέξεις Κλειδιά

αναγνώριση ανθρώπινης κίνησης, σκελετικά δεδομένα, kinect, OpenNI, Unity, tennis, αρχάριοι παίκτες, έμπειροι παίκτες, βάση δεδομένων THETIS, ταξινόμηση κινήσεων, SVM, χωροχρονικά σημεία ενδιαφέροντος, 3D συντεταγμένες, αρθρώσεις, απόδοση, διανύσματα δραστηριοτήτων, ομαδοποίηση k-means, Dynamic Time Warping, PCA, πρωτόκολλο leave-one-person-out, 3D Cylindrical Trace Transform

Abstract

The aim of this diploma thesis is the performance analysis of skeletal data in human action recognition algorithms compared to the performance of other data types e.g. Space Time Interest Points. The skeletal data used in the above study is derived from the human action dataset THETIS (THree dimEnsional TennIs Shots). This database includes 8374 videos, which contain 12 tennis movements performed by 55 different individuals (amateurs and experts tennis players).

The skeletal joint extraction is achieved by a program coded in programming language C++ with the use of the OpenNI software. Each joint is a three dimensional coordinate vector in space. A total of 15 joints in human body are calculated for each frame of a video. After estimating the 3D coordinates, the data is visualized via a program developed in Unity.

In the context of the diploma thesis, a published activity recognition algorithm is implemented in matlab programming language. This algorithm uses the 3D joint coordinates as input data and calculates some posture feature vectors. From the above postures, those that are considered most representative are selected by the k-means method. Then an Activity Feature Vector is extracted for each input movement. These vectors are used for the training and evaluation of multi-class SVMs that are used to classify tennis player movements.

Finally, the extracted skeletal data is applied to an algorithm that examines the experience level of the THETIS players. This algorithm is used to identify whether a player is an amateur or an expert by analyzing the input movements. This is achieved by using the descriptors Variance Vector and Cosine Distance Vector. For the alignment of the generated time sequences the Dynamic Time Warping method is used.

Keywords

human action recognition, skeletal data, kinect, OpenNI, Unity, tennis, amateur players, expert players, THETIS database, movement classification, SVM, spatio-temporal interest points, 3D coordinates, skeletal joints, performance, activity feature vectors, k-means method, Dynamic Time Warping, PCA, leave-one-person-out protocol, 3D Cylindrical Trace Transform

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τον επιβλέποντα κ. Ανδρέα-Γεώργιο Σταφυλοπάτη και τον κ. Κώστα Καρπούζη για την εμπιστοσύνη που μου έδειξαν αναθέτοντας μου την παρούσα διπλωματική εργασία. Επίσης θα ήθελα να εκφράσω την ευγνωμοσύνη μου απέναντι στον υποψήφιο Δρ Γεώργιο Τσατήρη για την καθοδήγηση και τις συμβουλές του ως προς την υλοποίηση της διπλωματικής, καθώς και τη βοήθεια του για την αντιμετώπιση οποιουδήποτε προβλήματος παρουσιάστηκε. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την κατανόηση και την υπομονή τους όλα αυτά τα χρόνια.

Πίνακας Περιεχομένων

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

1.1 Αντικείμενο Διπλωματικής Εργασίας

Η αναγνώριση της ανθρώπινης κίνησης με χρήση υπολογιστών αποτελεί ένα βασικό ερευνητικό ζήτημα, η επίλυση του οποίου αποσκοπεί στη δυνατότητα ανάλυσης των ανθρώπινων δραστηριοτήτων.

Η αναγνώριση της ανθρώπινης δραστηριότητας με χρήση ηλεκτρονικού υπολογιστή και συγκεκριμένα μέσω της επεξεργασίας εικόνων και video είναι μια απαιτητική διαδικασία. Η δυσκολία του συγκεκριμένου ερευνητικού πεδίου οφείλεται σε ποικίλα προβλήματα, τα οποία παρουσιάζονται συνοπτικά παρακάτω.

Αρχικά σημειώνεται ότι υπάρχουν φυσικοί παράγοντες, οι οποίοι διαφοροποιούνται σε κάθε περίσταση και μπορεί να δυσχεραίνουν την ανάλυση της καταγεγραμμένης κίνησης. Τέτοιοι παράγοντες αποτελούν η δομή και το σχήμα του ανθρώπινου σώματος, οι ενδυματολογικές επιλογές των ατόμων, η φωτεινότητα της εικόνας, η διαμόρφωση του περιβάλλοντα χώρου και ο θόρυβος λόγω σκίασης.

Ακόμη πρέπει να σημειωθεί ότι η πολυπλοκότητα κάθε κίνησης μπορεί να διαφοροποιείται με αποτέλεσμα να μην υπάρχει ακρίβεια στην αναγνώριση των δραστηριοτήτων. Η χρονική διάρκεια των κινήσεων αλλάζει, καθώς ορισμένες κινήσεις μπορεί να είναι στιγμιαίες ή να έχουν παρατεταμένη διάρκεια. Επίσης ορισμένες δραστηριότητες είναι ατομικές, ενώ άλλες πραγματοποιούνται από περισσότερα άτομα. Επιπροσθέτως σε μία δραστηριότητα μπορεί να απαιτείται η αλληλεπίδραση με ένα ή περισσότερα αντικείμενα. Η αλλαγή στις παραπάνω συνθήκες αυξάνει την πολυπλοκότητα των αλγορίθμων αναγνώρισης, ενώ για την επιτυχή κατηγοριοποίηση μιας κίνησης μπορεί να δημιουργείται η ανάγκη πρόσθετων περιορισμών.

Τέλος ένα σημαντικό πρόβλημα, το οποίο μπορεί να αναστέλλει την ανάπτυξη αποδοτικών αλγορίθμων αναγνώρισης αφορά τη φύση των κινήσεων. Ο αριθμός των διαφορετικών κατηγοριών κίνησης αποτελεί ένα ζήτημα προς αντιμετώπιση, καθώς δεν είναι εφικτό ένας αλγόριθμος να αναγνωρίζει κινήσεις από όλες τις πιθανές κατηγορίες. Μέσα σε αυτό το πλήθος κινήσεων μπορεί να υπάρχουν ομάδες κινήσεων, οι οποίες φέρουν μεγάλο βαθμό ομοιότητας μεταξύ τους και διαφοροποιούνται μόνο μέσω ελάχιστων ή δυσδιάκριτων λεπτομερειών στο χωροχρόνο. Από την άλλη, μπορεί κινήσεις που ανήκουν στην ίδια κατηγορία να διαφοροποιούνται σε μεγάλο βαθμό λόγω της ταχύτητας, της κατεύθυνσης και του ιδιαίτερου τρόπου με τον οποίο εκτελούνται από κάθε άτομο.

Για την αντιμετώπιση των παραπάνω προβλημάτων έχει αναπτυχθεί πλήθος διαφορετικών αλγορίθμων κίνησης, οι οποίοι αξιοποιούν διαφορετικές πληροφορίες και δεδομένα εισόδου. Αυτές οι μέθοδοι μπορεί να εκμεταλλεύονται χωροχρονικά σημεία ενδιαφέροντος (STIPs) και τροχιές (space-time trajectories), σκελετικά δεδομένα, περιγράμματα και σιλουέτες, οπτική ροή και ιστογράμματα ενέργειας κ.ά. Οι παραπάνω αλγόριθμοι παρουσιάζουν διαφορετική απόδοση μεταξύ τους, η οποία μπορεί να οφείλεται τόσο στη δομή τους όσο και στη φύση των δεδομένων εισόδου που αξιοποιούν.

Σκοπός της εργασίας αποτελεί η μελέτη της αποδοτικότητας μιας συγκεκριμένης μορφής δεδομένων, των σκελετικών αρθρώσεων, σε ορισμένους αλγορίθμους αναγνώρισης και η σύγκριση της απόδοσης που προκύπτει από άλλες μορφές δεδομένων. Μέσω αυτής της μελέτης προσδίδονται περισσότερες πληροφορίες για το πρόβλημα της αναγνώρισης ανθρώπινης κίνησης και πως διαφορετικά είδη δεδομένων μπορεί να είναι πιο αποτελεσματικά από άλλα. Στη συγκεκριμένη περίπτωση λαμβάνονται πειραματικά αποτελέσματα και συμπεράσματα, τα οποία φανερώνουν πόσο αποδοτικά μπορεί να είναι τα σκελετικά δεδομένα έναντι άλλων μορφών εισόδου σε ορισμένους αλγορίθμους αναγνώρισης. Επίσης τα εξαγόμενα σκελετικά δεδομένα μπορεί να χρησιμοποιηθούν για την ανάπτυξη μελλοντικών αλγορίθμων αναγνώρισης κίνησης. Επισημαίνεται ότι τα σκελετικά δεδομένα διεξάγονται από τη βάση δράσεων THETIS, η οποία περιλαμβάνει 12 κινήσεις tennis εκτελεσμένες από αρχάριους και έμπειρους παίκτες.

Τα βασικά θέματα συνεισφοράς της συγκεκριμένης διπλωματικής εργασίας αποτελούν:

- Η εξαγωγή των 3D συντεταγμένων των σκελετικών αρθρώσεων από τα video βάθους της THETIS με χρήση του λογισμικού OpenNI και της γλώσσας προγραμματισμού C++.

- Η πρόταση για βελτίωση της ακρίβειας των σκελετικών δεδομένων με χρήση ειδικών φίλτρων (Mean Median Filter) και υλοποίηση αντίστοιχου κώδικα σε C.

- Η υλοποίηση ενός προγράμματος σε Unity για την οπτικοποίηση των σκελετικών αρθρώσεων και τον έλεγχο ορθότητας των εξαγόμενων συντεταγμένων.

- Η υλοποίηση ενός δημοσιευμένου αλγορίθμου αναγνώρισης με χρήση διανυσμάτων δραστηριότητας, σε γλώσσα προγραμματισμού matlab, και η εφαρμογή των σκελετικών δεδομένων ως είσοδος.

- Η εφαρμογή των σκελετικών συντεταγμένων ως δεδομένα εισόδου σε έναν αλγόριθμο αναγνώρισης του επιπέδου εμπειρίας των παικτών tennis της THETIS.

- Η συνοπτική αξιολόγηση και σύγκριση των πειραματικών αποτελεσμάτων με προϋπάρχοντα αποτελέσματα άλλων μορφών δεδομένων εισόδου.

Επισημαίνεται ότι η ταξινόμηση των κινήσεων σε κάθε κατηγορία είναι ήδη υλοποιημένη με χρήση SVM πολλαπλών κλάσεων.

1.2 Δομή Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία επικεντρώνεται στη μελέτη των σκελετικών δεδομένων και την εφαρμογή τους σε διάφορους υλοποιημένους αλγορίθμους αναγνώρισης κίνησης. Τα βασικά τμήματα της εργασίας μπορούν να διακριθούν στην εξαγωγή των σκελετικών δεδομένων, την παρουσίαση ορισμένων ήδη υλοποιημένων αλγορίθμων αναγνώρισης κίνησης και την απόδοση που αυτοί παρουσιάζουν με την εφαρμογή των σκελετικών αρθρώσεων ως δεδομένα εισόδου. Η δομή της εργασίας παρουσιάζεται περιληπτικά στη συνέχεια.

Στο κεφάλαιο 2 περιγράφεται το πρόβλημα της αναγνώρισης ανθρώπινης κίνησης μέσω της επεξεργασίας εικόνας και video και παρουσιάζονται τα είδη συστημάτων καταγραφής της κίνησης. Επίσης γίνεται αναφορά σε μεθόδους αναγνώρισης δραστηριοτήτων, οι οποίες έχουν προταθεί από την ερευνητική κοινότητα και αξιοποιούν διάφορες μορφές δεδομένων. Τέλος γίνεται σύντομη ανάλυση της λειτουργίας της συσκευής Kinect και του λογισμικού OpenNI, το οποίο χρησιμοποιείται για την εξαγωγή των σκελετικών δεδομένων.

Στο κεφάλαιο 3 γίνεται παρουσίαση της βάσης δεδομένων THETIS, από την οποία υπολογίζονται οι συντεταγμένες των σκελετικών αρθρώσεων των παικτών του tennis. Επίσης αναλύεται η μέθοδος 3D Cylindrical Trace Transform, της οποίας η απόδοση χρησιμοποιείται ως μέτρο σύγκρισης στα πειραματικά αποτελέσματα του 5^{ου} κεφαλαίου. Τέλος γίνεται μια εισαγωγή στα συστήματα ταξινόμησης SVM, τα οποία αξιοποιούνται στην κατηγοριοποίηση των κινήσεων των παικτών.

Στο κεφάλαιο 4 περιγράφεται η μέθοδος που αναπτύχθηκε για τον υπολογισμό των σκελετικών αρθρώσεων και προτείνεται βελτίωση των αποτελεσμάτων με χρήση των φίλτρων μέσης (Mean Filter) και διάμεσης τιμής (Median Filter). Ακόμη αναφέρονται προβλήματα, τα οποία εμφανίστηκαν κατά την επεξεργασία της βάσης THETIS. Τέλος παρουσιάζεται ένα πρόγραμμα υλοποιημένο σε Unity με το οποίο επιτυγχάνεται η οπτικοποίηση των εξαγόμενων συντεταγμένων για τον έλεγχο της ορθότητας τους.

Στο κεφάλαιο 5 γίνεται αναφορά σε έναν αλγόριθμο αναγνώρισης με χρήση διανυσμάτων δραστηριοτήτων, ο οποίος υλοποιείται σε γλώσσα C. Στον παραπάνω αλγόριθμο δίνονται ως δεδομένα εισόδου οι συντεταγμένες των σκελετικών αρθρώσεων και συγκρίνεται η απόδοση του αλγορίθμου με άλλες υλοποιημένες μεθόδους. Τα σκελετικά δεδομένα εισάγονται επίσης σε έναν αλγόριθμο, ο οποίος διαχωρίζει τους έμπειρους από τους αρχάριους παίκτες και γίνεται σύγκριση της απόδοσης των παραπάνω δεδομένων με αυτή που επιτυγχάνεται με χρήση STIPs.

ΚΕΦΑΛΑΙΟ 2

Χαρακτηριστικά και Μέθοδοι Αναγνώρισης της Ανθρώπινης κίνησης

2.1 Είδη και Συστήματα Καταγραφής Ανθρώπινης Κίνησης

2.1.1 Είδη Ανθρώπινης Κίνησης

Η αναγνώριση της ανθρώπινης κίνησης αποτελεί ένα κύριο ερευνητικό θέμα στο πεδίο της όρασης υπολογιστών (computer vision) και αποσκοπεί στη δυνατότητα ανάλυσης και αναγνώρισης ανθρώπινων δραστηριοτήτων, οι οποίες είναι καταγεγραμμένες σε κάποια μορφή (όπως video).

Η αναγνώριση της ανθρώπινης δραστηριότητας μέσω υπολογιστή αποτελεί μια δύσκολη και απαιτητική διαδικασία, καθώς περιλαμβάνει την κατανόηση της ανθρώπινης κίνησης. Η δομή και το σχήμα του ανθρώπινου σώματος ποικίλλουν, ενώ παράγοντες όπως η φωτεινότητα της εικόνας, ο θόρυβος λόγω σκίασης, οι ενδυματολογικές επιλογές και οι συνθήκες του περιβάλλοντα χώρου δυσχεραίνουν την παραπάνω διαδικασία.

Η μοντελοποίηση της ανθρώπινης συμπεριφοράς προϋποθέτει τον χαρακτηρισμό και την ταξινόμηση των διαφόρων ειδών δραστηριότητας. Η ανθρώπινη δραστηριότητα μπορεί να διακριθεί με βάση το επίπεδο της πολυπλοκότητάς της σε τέσσερις κατηγορίες: χειρονομίες (gestures), ενέργειες (actions), αλληλεπίδραση (interaction) και ομαδικές δραστηριότητες (group activities). Η πρώτη κατηγορία αναφέρεται στη μετακίνηση κάποιου μέρους του σώματος ενός ατόμου (όπως η κίνηση του χεριού κ.α.). Η δεύτερη κατηγορία απαρτίζεται από τις κινήσεις ενός μόνο ατόμου, οι οποίες αποτελούνται από ένα πλήθος χειρονομιών (όπως τρέξιμο, περπάτημα κ.α.) Η τρίτη κατηγορία αφορά την αλληλεπίδραση ενός ατόμου με άλλα άτομα ή αντικείμενα, ενώ η τελευταία κατηγορία αποτελείται από δραστηριότητες, οι οποίες πραγματοποιούνται από ομάδες ατόμων.

2.1.2 Καταγραφή Κίνησης

Η καταγραφή κίνησης (motion capture ή mocap) είναι η διαδικασία, κατά την οποία γίνεται εγγραφή της κίνησης ενός χρήστη/αντικειμένου στον πραγματικό κόσμο και στη συνέχεια τα δεδομένα αυτής της κίνησης εισάγονται σε ένα τρισδιάστατο μοντέλο ενός εικονικού περιβάλλοντος. Η ιδέα της καταγραφής κίνησης επινοήθηκε αρχικά για να διευκολυνθεί η διαδικασία παραγωγής animation στον κινηματογραφικό τομέα, ωστόσο η εφαρμογή της διευρύνεται και σε άλλους τομείς.

Βασικοί τομείς εφαρμογής της καταγραφής κίνησης είναι: η επιτήρηση (surveillance), ο έλεγχος (control) και η ανάλυση (analysis). Ο τομέας της επιτήρησης καλύπτει εφαρμογές όπου ένα ή περισσότερα αντικείμενα/χρήστες παρακολουθούνται με την πάροδο του χρόνου για την εξακρίβωση διαφόρων ενεργειών (για παράδειγμα χρήση για την παρακολούθηση κάποιου δημόσιου χώρου). Ο τομέας του ελέγχου σχετίζεται με εφαρμογές όπου η καταγεγραμμένη κίνηση χρησιμοποιείται για την παροχή λειτουργιών ελέγχου (για

παράδειγμα χρήση σε ηλεκτρονικά παιχνίδια). Ο τρίτος τομέας εφαρμογών ασχολείται με τη λεπτομερή ανάλυση των δεδομένων κίνησης που συλλαμβάνονται (για παράδειγμα χρήση σε ιατρικές μελέτες για την εξακρίβωση δυσλειτουργιών).

Τα συστήματα καταγραφής κίνησης χωρίζονται σε συστήματα εντοπισμού κίνησης βάσει δείκτη (Marker-based Motion Capture) και σε αυτά χωρίς (Markerless Motion Capture). Η ειδοποιός διαφορά των δύο κατηγοριών είναι η απαίτηση ειδικού εξοπλισμού για τον εντοπισμό της κίνησης του χρήστη/αντικειμένου στην πρώτη κατηγορία, προϋπόθεση η οποία δεν είναι αναγκαία για τη λειτουργία των συστημάτων της δεύτερης κατηγορίας. Ακολουθούν τα βασικότερα είδη συστημάτων των δύο προαναφερθέντων κατηγοριών **[1]**:

- **Ακουστικά Συστήματα (Acoustical Systems)**

Σε αυτόν τον τύπο συστήματος ένα σύνολο πομπών ήχου τοποθετείται στις κύριες αρθρώσεις του χρήστη και τρείς δέκτες τοποθετούνται στη θέση σύλληψης. Οι πομποί έπειτα ενεργοποιούνται διαδοχικά, παράγοντας ένα χαρακτηριστικό σύνολο συχνοτήτων, το οποίο λαμβάνουν οι δέκτες και υπολογίζουν τις θέσεις των πομπών στον τρισδιάστατο χώρο. Ο υπολογισμός της θέσης κάθε πομπού γίνεται με την ακόλουθη διαδικασία: Χρησιμοποιώντας ως δεδομένα το χρονικό διάστημα μεταξύ της εκπομπής του θορύβου από τον πομπό, της λήψης αυτού από τον δέκτη και της ταχύτητας κίνησης του ήχου στο περιβάλλον, μπορεί κανείς να υπολογίσει την απόσταση που διανύει το σήμα του θορύβου. Η θέση κάθε πομπού στον τρισδιάστατο χώρο υπολογίζεται με τη μέθοδο του τριγωνισμού (triangulation) των αποστάσεων του πομπού από κάθε δέκτη. Ενώ η διαδικασία της σύλληψης δεν είναι περίπλοκη, η κίνηση, η οποία εγγράφεται από τα ακουστικά συστήματα, μπορεί να μην είναι ομαλή, λόγω της διαδοχικής μετάδοσης του σήματος από κάθε πομπό, το οποίο οδηγεί σε ασυνέχειες της κίνησης σε συγκεκριμένες χρονικές στιγμές. Επίσης ο περιορισμός των κινήσεων του χρήστη και το πλήθος των πομπών μπορεί να επηρεάσει την παραπάνω κίνηση.

- **Μηχανικά Συστήματα (Mechanical Systems)**

Τα συστήματα αυτά αποτελούνται από ποτενσιόμετρα και ρυθμιστές, οι οποίοι τοποθετούνται στις επιθυμητές αρθρώσεις του χρήστη και επιτρέπουν την απεικόνιση των θέσεών τους (Εικόνα 2.1). Παρά το γεγονός ότι είναι ανεπαρκώς αναπτυγμένα, τα συστήματα μηχανικής καταγραφής κίνησης έχουν ορισμένα πλεονεκτήματα, τα οποία τα καθιστούν αρκετά ελκυστικά. Ένα πλεονέκτημα είναι ότι διαθέτουν μια διεπαφή, η οποία είναι παρόμοια με αυτή των συστημάτων stop-motion, επιτρέποντας έτσι μια εύκολη μετάβαση μεταξύ των δύο τεχνολογιών. Επίσης δεν επηρεάζονται από μαγνητικά πεδία ή ανεπιθύμητες αντανακλάσεις, με αποτέλεσμα να μην απαιτείται επαναλαμβανόμενα ο εντοπισμός της κίνησης (recalibration), γεγονός που καθιστά τη χρήση τους εύκολη και παραγωγική.

Εικόνα 2.1: Παράδειγμα μηχανικού συστήματος. *Αριστερά: Ηθοποιός χρησιμοποιώντας κοστούμι πλήρους μηχανικής κίνησης. Δεξιά: Ηθοποιός χρησιμοποιώντας μόνο μερικούς μηχανικούς αισθητήρες αντί για ένα πλήρες κοστούμι.*

- **Μαγνητικά Συστήματα (Magnetic Systems)**

Χρησιμοποιώντας ένα σύνολο δεκτών, οι οποίοι τοποθετούνται στις αρθρώσεις του χρήστη, είναι δυνατόν να υπολογιστεί η θέση και ο προσανατολισμό των αρθρώσεων σε σχέση με μια κεραία. Τα μαγνητικά συστήματα είναι οικονομικά αποδοτικότερα σε σύγκριση με άλλα συστήματα καταγραφής κίνησης, καθώς η απόκτηση και επεξεργασία των δεδομένων δεν αποτελεί ακριβή διαδικασία, ενώ η ποιότητα των επεξεργασμένων δεδομένων είναι αρκετά υψηλή. Με τυπικό ρυθμό δειγματοληψίας ~100 καρέ ανά δευτερόλεπτο, τα μαγνητικά συστήματα είναι ιδανικά για απλή σύλληψη κίνησης.

Ένα από τα μειονεκτήματα αυτής της κατηγορίας είναι η περιορισμένη κίνηση του χρήστη, λόγω του μεγάλου αριθμού καλωδίων που συνδέονται με την κεραία. Επίσης πιθανές παρεμβολές στο μαγνητικό πεδίο, οι οποίες προκαλούνται από διάφορα μεταλλικά αντικείμενα ή δομές, καθιστούν τα μαγνητικά συστήματα αρκετά ευαίσθητα και περιορίζουν τα κατάλληλα υλικά προς χρήση στον περιβάλλοντα χώρο τους.

Εικόνα 2.2: Παράδειγμα μαγνητικού συστήματος. *Αριστερά: Ηθοποιός που χρησιμοποιεί μαγνητικό κοστούμι mocap. Δεξιά: Μαγνητικός αισθητήρας mocap.*

- **Οπτικά Συστήματα (Optical Systems)**

Σε αυτά τα συστήματα ο χρήστης φοράει μια ειδική στολή, καλυμμένη με κάτοπτρα που βρίσκονται στις κύριες αρθρώσεις του. Κάμερες υψηλής ανάλυσης τοποθετούνται σε στρατηγικές θέσεις για να παρακολουθούν αυτά τα κάτοπτρα κατά τη διάρκεια της κίνησης του χρήστη. Στη συνέχεια κάθε κάμερα παράγει για κάθε κάτοπτρο τις 2D συντεταγμένες του, οι οποίες χρησιμοποιούνται για τον υπολογισμό των αντίστοιχων τρισδιάστατων συντεταγμένων μέσω κατάλληλου ιδιόκτητου λογισμικού.

Ένα από τα πλεονεκτήματα της χρήσης αυτών των συστημάτων είναι ο υψηλός ρυθμός δειγματοληψίας, ο οποίος επιτρέπει τη σύλληψη γρήγορων κινήσεων. Ο ρυθμός δειγματοληψίας εξαρτάται συνήθως από τις χρησιμοποιούμενες κάμερες, από το οποίο συνεπάγεται ότι είναι ανάλογος της ανάλυσης κάθε κάμερας. Ένα άλλο πλεονέκτημα είναι η ελευθερία που προσφέρεται από αυτά τα συστήματα, καθώς δεν υπάρχουν περιορισμοί στην κίνηση λόγω καλωδίων ή συνθηκών του περιβάλλοντα χώρου. Όμως, ενώ η ανάλυση της κίνησης είναι ακριβής, τα συστήματα αυτής της κατηγορίας είναι δαπανηρά λόγω του απαιτούμενου εξοπλισμού και παρουσιάζουν έλλειψη αλληλεπίδρασης, καθώς τα δεδομένα που συλλέγονται πρέπει να επεξεργαστούν (και ορισμένες φορές να υποβληθούν σε φιλτράρισμα για μείωση του θορύβου) πριν χρησιμοποιηθούν.

Εικόνα 2.3: Παράδειγμα οπτικού συστήματος. *Πάνω αριστερά: Ένα mocap studio. Κάτω δεξιά: Το ίδιο mocap studio κατά τη διάρκεια της καταγραφής κίνησης. Κάτω αριστερά: Κάμερα λήψης κίνησης. Κάτω δεξιά: Κάμερα λήψης κίνησης σε κοντινή απόσταση.*

- **Ενεργά Οπτικά Συστήματα - Δείκτες (Active Markers)**

Αυτό το είδος οπτικής καταγραφής της κίνησης αξιοποιεί LEDs, τα οποία αντί να αντανακλούν το φως που εκπέμπεται από τις κάμερες υψηλής ανάλυσης, εκπέμπουν το δικό τους φως, τροφοδοτούμενα από κάποια μπαταρία.

- **Παθητικά Οπτικά Συστήματα – Δείκτες (Passive Markers)**

Σε αντίθεση με την προηγούμενη κατηγορία, οι συγκεκριμένοι δείκτες επικαλύπτονται με ένα αντανακλαστικό υλικό, το οποίο αντανακλά το φως πίσω στις

κάμερες. Απαιτείται να έχει προηγηθεί η βαθμονόμηση του φωτός, ώστε να αναγνωρίζονται μόνο οι δείκτες, αγνοώντας άλλα υλικά.

Εικόνα 2.4: *Παθητικοί δείκτες τοποθετημένοι στο πρόσωπο ενός ηθοποιού*.

- **Συστήματα Βασισμένα στην Όραση (Markerless Motion Capture - Vision Based)**

Τα συστήματα αυτά, όπως προαναφέρθηκε, δεν απαιτούν ειδικό εξοπλισμό για την παρακολούθηση της κίνησης του χρήστη. Η κίνηση καταγράφεται μέσω video, τα οποία αναλύονται από αλγορίθμους υπολογιστικής όρασης προκειμένου να αναγνωριστούν οι ανθρώπινες μορφές και τμήματα της κίνησης τους. Η διαδικασία καταγραφής κινήσεων γίνεται εξολοκλήρου μέσω λογισμικού, αφαιρώντας όλους τους φυσικούς περιορισμούς, αλλά εισάγοντας υπολογιστικούς περιορισμούς. Ο τρόπος λειτουργίας τους περιγράφεται αναλυτικά στην επόμενη ενότητα. Παράδειγμα/εφαρμογή ενός τέτοιου συστήματος είναι το Kinect της Microsoft.

2.2 Συστήματα Βασισμένα στην Όραση (vision-based)

Για τη σωστή λειτουργία αυτής της κατηγορίας συστημάτων, ορισμένες φορές απαιτούνται κάποιες παραδοχές/προϋποθέσεις. Οι τυπικές παραδοχές μπορούν να χωριστούν σε δύο κατηγορίες: υποθέσεις σχετικά με την κίνηση και υποθέσεις σχετικά με την εμφάνιση. Η πρώτη κατηγορία αφορά τους περιορισμούς στις κινήσεις του χρήστη, ενώ η δεύτερη αφορά περιορισμούς ως προς τις συνθήκες του περιβάλλοντος. Επίσης τα συστήματα αυτά διακρίνονται σε δύο ομάδες ανάλογα με την αξιοποίηση κάποιου μοντέλου ανάλυσης κίνησης ή όχι.

2.2.1 Μέθοδοι Χωρίς Αξιοποίηση Μοντέλων (Model-free Methods)

Τεχνικές, οι οποίες δεν χρησιμοποιούν κάποιο μοντέλο, διακρίνονται σε δύο κατηγορίες, σε αυτές που στηρίζονται στη μάθηση (learning-based) και σε αυτές που στηρίζονται στο παράδειγμα (example-based). Στις learning-based προσεγγίσεις, μια συνάρτηση αναγνώρισης κίνησης δημιουργείται μέσω δεδομένων εκπαίδευσης και έπειτα χρησιμοποιείται για την εξακρίβωση διάφορων κινήσεων. Στις example-based προσεγγίσεις, μια συλλογή υποδειγμάτων κίνησης αποθηκεύεται σε μια βάση δεδομένων, μαζί με τις

αντίστοιχες περιγραφές τους. Για μια δεδομένη εικόνα εισόδου, πραγματοποιείται μια αναζήτηση στη βάση δεδομένων και οι κινήσεις/στάσεις με τη μεγαλύτερη ομοιότητα συνδυάζονται για τη λήψη της εκτιμώμενης στάσης.

2.2.2 Μέθοδοι Βασισμένες σε Μοντέλα (Model-based Methods)

Οι μέθοδοι, οι οποίες αξιοποιούν κάποιο μοντέλο, περιλαμβάνουν το στάδιο μοντελοποίησης. Ο στόχος του σταδίου μοντελοποίησης είναι η δημιουργία μιας συνάρτησης, η οποία αναπαριστά με κάποια πιθανότητα τα γεγονότα της εικόνας, λαμβάνοντας υπόψη ένα σύνολο παραμέτρων (παράμετροι διαμόρφωσης/ σχήμα σώματος, παράμετροι θέσης και όψης κάμερας κτλ.). Ορισμένες από αυτές τις παραμέτρους μπορεί να είναι γνωστές εκ των προτέρων όπως η περιστροφή και η θέση της κάμερας.

Τα μοντέλα, τα οποία χρησιμοποιούν οι παραπάνω μέθοδοι, βασίζονται στην κινηματική δομή και τις διαστάσεις του ανθρώπινου σώματος και μπορούν να διακριθούν σε δύο κατηγορίες:

- **Κινηματικά Μοντέλα (kinematic models)**

Σε αυτή την κατηγορία, τα μοντέλα περιγράφουν το ανθρώπινο σώμα ως μια κινηματική δομή, η οποία αποτελείται από επιμέρους τμήματα που συνδέονται μεταξύ τους μέσω αρθρώσεων. Κάθε άρθρωση περιέχει ένα πλήθος από βαθμούς ελευθερίας (DOF), υποδεικνύοντας σε ποιες κατευθύνσεις μπορεί να κινηθεί η άρθρωση. Επίσης αυτά τα μοντέλα μπορούν να είναι 2D ή 3D. Τα 2D μοντέλα είναι κατάλληλα για κίνηση παράλληλη με το επίπεδο εικόνας, όμως δε μπορούν να περιγράψουν πολύπλοκες κινήσεις, οι οποίες περιέχουν άλλη μορφή προσανατολισμού. Τα 3D μοντέλα δίνουν τη δυνατότητα περιστροφής κάθε άρθρωσης σε τρεις διαστάσεις. Κάθε περιστροφή μπορεί να ικανοποιεί διαφορετικούς περιορισμούς/προϋποθέσεις, οι οποίες αποκλείουν την αναπαράσταση ανέφικτων ανθρώπινων κινήσεων. Το κινηματικό μοντέλο μπορεί να ανακτηθεί δυναμικά με επίλυση απλών γραμμικών συστημάτων χρησιμοποιώντας τη γραμμική ορθογώνια προβολή.

- **Σχηματικά Μοντέλα (Shape Models)**

Τα σχηματικά μοντέλα στηρίζονται στην ιδέα της αναπαράστασης του ανθρώπινου σώματος με χρήση σχημάτων. Τα 2D μοντέλα αξιοποιούν σχήματα όπως τα τετράγωνα , ορθογώνια, τραπεζοειδή κτλ., ενώ τα 3D μοντέλα τρισδιάστατα σχήματα όπως ο κύβος. Τα παραπάνω σχήματα εξαρτώνται από παραμέτρους όπως το μήκος και το πλάτος, τα οποία λόγω της μεταβλητότητας των διαστάσεων των ανθρώπων, μπορεί να καθορίζονται δυναμικά στο στάδιο της αρχικοποίησης.

2.2.3 Περιγραφείς Εικόνας (Image Descriptors)

Η εμφάνιση των ανθρώπων στις εικόνες ποικίλλει λόγω των διαφορετικών συνθηκών ένδυσης και φωτισμού. Επειδή κύριος στόχος αποτελεί η ανάκτηση της κίνησης του ατόμου και όχι τα επιμέρους χαρακτηριστικά της εμφάνισης του, έχουν αναπτυχθεί ορισμένα περιγραφικά στοιχεία εικόνας (image descriptors), τα οποία παρέχουν τις απαραίτητες πληροφορίες για την κατανόηση της κίνησης **[2]**. Τέτοια στοιχεία είναι τα εξής:

- **Σιλουέτες και Περιγράμματα**

Οι σιλουέτες και τα περιγράμματα των ατόμων μπορούν να εξαχθούν με σχετικά μεγάλη ακρίβεια από τις εικόνες, αν το περιβάλλον είναι στατικό και διαφοροποιείται ως προς την εμφάνιση από το άτομο που κινείται. Αυτό εξαλείφει την ανάγκη εκτίμησης περιβαλλοντικών παραμέτρων. Οι σιλουέτες, ωστόσο, δεν είναι ευαίσθητες σε αλλαγές που αφορούν την επιφάνεια (όπως το χρώμα και η υφή), ενώ κωδικοποιούν επαρκές πλήθος πληροφοριών για την ανάκτηση κάποιας στάσης/κίνησης.

- **Ακμές**

Οι ακμές εμφανίζονται στην εικόνα όταν υπάρχει σημαντική διαφορά στην ένταση σε διαφορετικές πλευρές της ίδιας θέσης. Μπορούν να υπολογιστούν με χαμηλό κόστος και είναι, σε κάποιο βαθμό, αμετάβλητες στις συνθήκες φωτισμού. Ωστόσο η χρήση ακμών δεν ενδείκνυται για περιβάλλοντα με έντονα μοτίβα ή πολλά αντικείμενα.

- **3D Ανακατασκευές**

Τα περιγράμματα και οι ακμές δεν αποδίδουν πληροφορίες βάθους, τουλάχιστον όταν χρησιμοποιείται μόνο μία κάμερα. Αυτό δυσχεραίνει την ανίχνευση προβληματικών κινήσεων (όπως αν το ίδιο το άτομο καλύπτει με τη στάση του κάποιο μέρος του σώματος του κτλ.). Όταν χρησιμοποιούνται πολλές κάμερες, μπορεί να δημιουργηθεί μια 3D ανακατασκευή από τις σιλουέτες, οι οποίες εξάγονται σε κάθε όψη ξεχωριστά. Ένας άλλος τρόπος λήψης πληροφοριών βάθους είναι η χρήση στερεομετρίας και τριγωνισμού των αποστάσεων σημείων μέσα στην εικόνα.

- **Χρώμα και Υφή**

Η μοντελοποίηση του ατόμου με βάση το χρώμα ή την υφή εμπνέεται από την παρατήρηση ότι η εμφάνιση μεμονωμένων τμημάτων του ανθρώπινου σώματος παραμένει ουσιαστικά αμετάβλητη, ανεξαρτήτως των διαφορετικών στάσεων. Σε αυτή την κατηγορία αξιοποιούνται Gaussian κατανομές χρώματος ή χρωματικά ιστογράμματα.

- **Κίνηση**

Η κίνηση μπορεί να εντοπιστεί λαμβάνοντας τη διαφορά μεταξύ δύο διαδοχικών πλαισίων. Η φωτεινότητα των pixels, τα οποία είναι μέρος του ατόμου στην εικόνα, θεωρείται σταθερή. Η μετατόπιση των pixels στην εικόνα ονομάζεται οπτική ροή.

Συνδυασμός των παραπάνω στοιχείων, μπορεί να χρησιμοποιηθεί για την αναγνώριση της ανθρώπινης κίνησης.

2.2.4 Λειτουργικά Στάδια Vision-Based Συστημάτων

Οι λειτουργικότητες ενός συστήματος βασισμένου στην όραση μπορούν να διακριθούν στα στάδια: αρχικοποίηση (initialization), παρακολούθηση (tracking), εκτίμηση στάσης (pose estimation) και αναγνώριση (recognition). Ένα σύστημα, ωστόσο, δεν είναι αναγκαίο να περιλαμβάνει όλα τα στάδια επεξεργασίας **[3].**

- **Αρχικοποίηση (Initialization)**

Η αρχικοποίηση καλύπτει τις ενέργειες που απαιτούνται για να εξασφαλιστεί ότι ένα σύστημα αρχίζει τη λειτουργία του με μια σωστή ερμηνεία της τρέχουσας σκηνής. Μερικές φορές ο όρος αρχικοποίηση χρησιμοποιείται επίσης για την προεπεξεργασία δεδομένων. Ενέργειες που αποτελούν τμήμα της αρχικοποίησης είναι η βαθμονόμηση της κάμερας, η προσαρμογή στα χαρακτηριστικά της σκηνής και η αρχικοποίηση μοντέλου (προσδιορισμός της αρχικής θέσης του αντικειμένου/χρήστη και ρύθμιση των παραμέτρων του εικονικού μοντέλου ανάλυσης).

- **Παρακολούθηση (Tracking)**

Η παρακολούθηση αναφέρεται στη δημιουργία συνεκτικών σχέσεων της κίνησης του χρήστη και των μελών του μεταξύ των χρονικών πλαισίων/καρέ. Μπορεί να θεωρηθεί ως μια ξεχωριστή διαδικασία ή μέρος της προετοιμασίας των δεδομένων για την εκτίμηση της στάσης και την αναγνώριση της κίνησης. Εάν θεωρηθεί ξεχωριστή διαδικασία, το υποκείμενο συνήθως παρακολουθείται ως ένα ενιαίο αντικείμενο (χωρίς κανένα άκρο) και δεν χρησιμοποιείται γνώση υψηλού επιπέδου. Εάν αποτελεί μέρος της προετοιμασίας των δεδομένων, τότε ο σκοπός της είναι να εξάγει συγκεκριμένες πληροφορίες για την εικόνα, είτε χαμηλού επιπέδου (όπως ακμές) είτε υψηλού επιπέδου (όπως τμήματα του ανθρώπινου σώματος). Σε κάθε περίπτωση ακολουθούνται τρία στάδια: Αρχικά γίνεται κατάτμηση της ανθρώπινης φιγούρας από την υπόλοιπη εικόνα, έπειτα γίνεται μετασχηματισμός της φιγούρας σε αναπαράσταση με λιγότερες πληροφορίες και τέλος ορίζεται ο τρόπος εντοπισμού του υποκειμένου σε κάθε χρονικό πλαίσιο.

- **Εκτίμηση Στάσης (Pose estimation)**

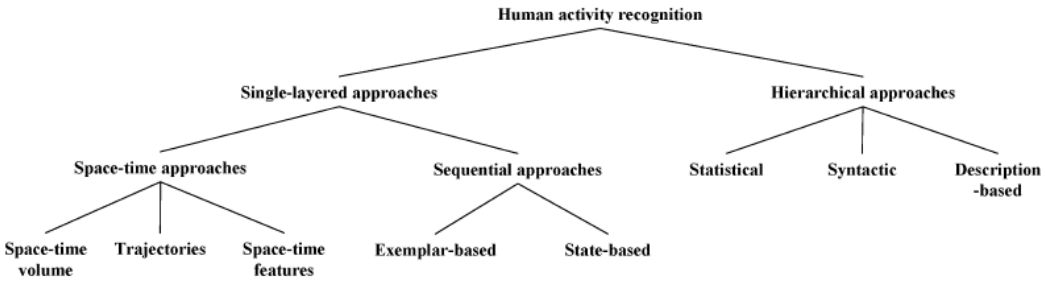
Η εκτίμηση στάσης είναι η διαδικασία προσδιορισμού του τρόπου με τον οποίο ένα ανθρώπινο σώμα και/ή μεμονωμένα άκρα του διαμορφώνονται σε μια δεδομένη σκηνή. Η συγκεκριμένη λειτουργία μπορεί να αποτελεί στάδιο μετά την επεξεργασία των δεδομένων σε έναν αλγόριθμο παρακολούθησης ή μπορεί να είναι ενεργό μέρος της διαδικασίας παρακολούθησης. Η εκτίμηση στάσης μπορεί να παράγει γενικές πληροφορίες σχετικά με τα χέρια και το κεφάλι του χρήστη (ή απλώς το κέντρο μάζας του σώματος) ή λεπτομερείς περιγραφές των θέσεων, του προσανατολισμού, του πλάτους κ.α. κάθε άκρου του σώματος. Αναλόγως με τον τρόπο εξαγωγής των πληροφοριών, αυτό το στάδιο μπορεί να διακριθεί σε τρεις κατηγορίες. Η πρώτη κατηγορία, ανάλυση χωρίς μοντέλο, καλύπτει προσεγγίσεις όπου δεν χρησιμοποιείται κάποιο πρότυπο-μοντέλο εκ των προτέρων. Οι άλλες δύο κατηγορίες, έμμεση χρήση μοντέλου και άμεση χρήση μοντέλου, χαρακτηρίζονται από την ύπαρξη ενός ανθρώπινου μοντέλου από την αρχή της διαδικασίας. Στην έμμεση περίπτωση, το μοντέλο χρησιμοποιείται στατικά ως αναφορά για τον περιορισμό και την ερμηνεία των μετρημένων δεδομένων. Στην άμεση περίπτωση, το μοντέλο διατηρείται και ενημερώνεται από τα παρατηρούμενα δεδομένα και ως εκ τούτου, μεταβάλλεται.

- **Αναγνώριση (Recognition)**

Η αναγνώριση συνήθως πραγματοποιείται με την ταξινόμηση της καταγεγραμμένης κίνησης ως κάποιο είδος ενέργειας (όπως το περπάτημα, το τρέξιμο ή πιο περίπλοκες κινήσεις όπως διαφορετικά βήματα χορού). Υπάρχουν δύο διαφορετικά είδη λειτουργίας: αναγνώριση με ανακατασκευή και άμεση αναγνώριση. Το πρώτο είδος βασίζεται στην ιδέα της ανακατασκευής της σκηνής, ενώ το δεύτερο είδος αναγνωρίζει άμεσα την ενέργεια με χρήση δεδομένων χαμηλού επιπέδου όπως η κίνηση, χωρίς (ή ελάχιστη) προεπεξεργασία. Επίσης η αναγνώριση μπορεί να είναι στατική ή δυναμική, αναλόγως με το αν τα δεδομένα που αξιοποιούνται ανήκουν στο ίδιο χρονικό πλαίσιο ή σε διαφορετικά πλαίσια αντίστοιχα.

2.3 Κατηγορίες Μεθόδων Αναγνώρισης Ανθρώπινης Κίνησης

Ανάλογα με τον τρόπο προσέγγισης της αναγνώρισης μιας κίνησης, έχουν αναπτυχθεί διάφορα συστήματα κατηγοριοποίησης των μεθόδων αναγνώρισης. Ενδεικτικά γίνεται μια σύντομη περιγραφή του συστήματος **[4]**, το οποίο παρουσιάζεται στην εικόνα 2.5. Οι μέθοδοι αναγνώρισης της ανθρώπινης δραστηριότητας μπορούν αρχικά να κατηγοριοποιηθούν σε τεχνικές μονής στιβάδας (single-layered approaches) και ιεραρχικές τεχνικές (hierarchical approaches).



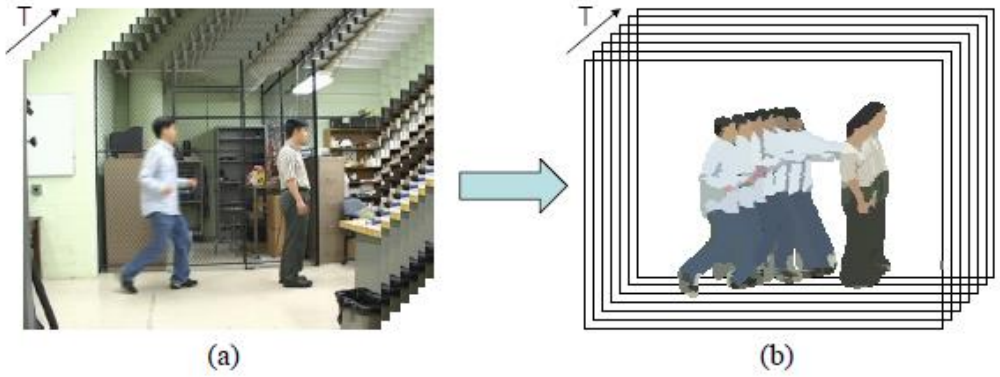
Εικόνα 2.5: *Ταξινόμηση μεθόδων αναγνώρισης με βάση τον τρόπο προσέγγισης του προβλήματος.*

2.3.1 Μέθοδοι Μονής Στιβάδας (single-layered methods)

Η συγκεκριμένη κατηγορία επεξεργάζεται και αναγνωρίζει απλές ενέργειες μέσω των εγγεγραμμένων δεδομένων του video. Κάθε δραστηριότητα αντιπροσωπεύει μια συγκεκριμένη κλάση από ακολουθίες εικόνων. Η αναγνώριση μιας δράσης, η οποία λαμβάνεται από μια άγνωστη ακολουθία εικόνων, γίνεται μέσω της ταξινόμησης της άγνωστης ακολουθίας σε κάποια από τις δεδομένες κλάσεις του συστήματος, με χρήση αλγορίθμων κατηγοριοποίησης. Οι μέθοδοι μονής στιβάδας διακρίνονται σε μεθόδους χωροχρόνου (space-time approaches) και ακολουθιακές μέθοδους (sequential approaches).

- **Μέθοδοι Χωροχρόνου (space-time approaches)**

Η λειτουργία αυτής της κατηγορίας μεθόδων στηρίζεται στην αντίληψη του video ως μια ακολουθία χρονικά διαδοχικών 2D εικόνων. Οι εικόνες αποτελούν την προβολή της τρισδιάστατης σκηνής σε δύο διαστάσεις, με αποτέλεσμα η συνένωση τους να δημιουργεί την αναπαράσταση του video με τη μορφή όγκου στο χωροχρόνο (3D XYT space-time volume).



Εικόνα 2.6: *Παραδείγματα τρισδιάστατων όγκων XYT κατασκευασμένα από : (a) ολόκληρες εικόνες και (b) blob εικόνων από ακολουθία αναπαράστασης της κίνησης «γροθιά».*

Μια τυπική μεθοδολογία, η οποία στηρίζεται στην αναπαράσταση γεγονότων μέσω όγκου στο χωροχρόνο αναλύεται στη συνέχεια. Αρχικά με βάση τα δεδομένα εκπαίδευσης δημιουργείται ένα 3D XYT space-time μοντέλο για κάθε δραστηριότητα. Λαμβάνοντας ένα video εισόδου, το σύστημα δημιουργεί έναν όγκο χωροχρόνου για την άγνωστη ακολουθία εικόνων και έπειτα, με τη χρήση κατάλληλου αλγορίθμου (template matching), ο νέος όγκος συγκρίνεται με τα υπάρχοντα μοντέλα δραστηριοτήτων του συστήματος. Η άγνωστη δραστηριότητα συνδέεται με το μοντέλο, στο οποίο παρουσιάστηκε μεγαλύτερος δείκτης ομοιότητας. Για την αναπαράσταση των δεδομένων υπάρχουν τρείς διαφορετικοί τρόποι:

- **Όγκος Χωροχρόνου (space-time volume):**

Αυτή η τεχνική χρησιμοποιείται όταν η κίνηση μπορεί να διασπαστεί χωρικά στις επιμέρους εικόνες και βασίζεται στην ιδέα της σύγκρισης ομοιότητας μεταξύ των όγκων διαφορετικών ακολουθιών εικόνων. Για την εξακρίβωση της ομοιότητας τους έχουν αναπτυχθεί διαφορετικοί αλγόριθμοι ταιριάσματος και τύποι αναπαράστασης όπως είναι η χρήση motion-energy images (MEI), motion-history images (MHI) και οπτικής ροής (optical flow).

- **Τροχιές Χωροχρόνου (space-time trajectories):**

Αυτή η κατηγορία μεταφράζει κάθε δραστηριότητα ως ένα σύνολο αποτελούμενο από τροχιές στο χωροχρόνο. Συγκεκριμένα ένα άτομο αναπαριστάται ως ένα σύνολο από 2D (XY) ή 3D (XYZ) σημεία, τα οποία αντιστοιχούν στις αρθρώσεις του σώματος του. Καθώς το άτομο εκτελεί κάποια κίνηση, οι θέσεις των αρθρώσεων μεταβάλλονται και καταγράφονται ως τροχιές στο χωροχρόνο, οι οποίες κατασκευάζουν μια αναπαράσταση στις τρεις (XYT) ή στις τέσσερις (XYZT) διαστάσεις. Έπειτα μια κίνηση εισόδου κατηγοριοποιείται μέσω της σύγκρισης της τροχιάς της με τις τροχιές του συστήματος.

- **Χαρακτηριστικά Χωροχρόνου (space-time local features):**

Η αναπαράσταση μέσω χαρακτηριστικών χωροχρόνου (space-time features) στηρίζεται στην εξαγωγή τοπικών χαρακτηριστικών από 3D όγκους στο χωροχρόνο. Για την υλοποίηση ενός αποδοτικού συστήματος αναγνώρισης πρέπει να καθοριστούν τα τοπικά χαρακτηριστικά που εξάγονται, ο τρόπος με τον οποίο αυτά αξιοποιούνται για την αναπαράσταση της κίνησης (χρήση appearance descriptors, interest points, local intensity gradient κ.α) και η τεχνική με την οποία γίνεται η ταξινόμηση των κινήσεων (χρήση

clustering, nearest neighbor κ.α). Τα τοπικά χαρακτηριστικά δεν απαιτούν το διαχωρισμό του ατόμου από τον περιβάλλοντα χώρο και είναι ανεξάρτητα της κλίμακας και της περιστροφής.

- **Ακολουθιακές Μέθοδοι (Sequential approaches)**

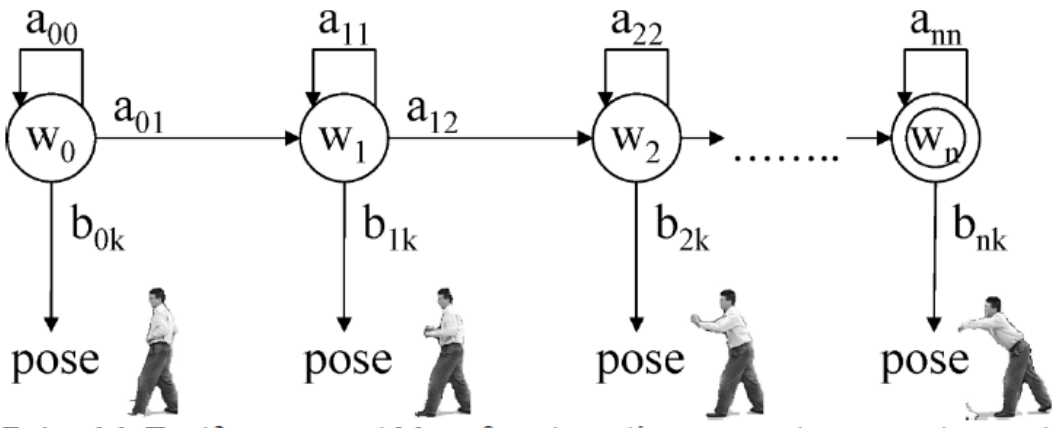
Σε αυτή την περίπτωση γίνεται η παραδοχή ότι ένα video μπορεί να αναπαρασταθεί ως ένα σύνολο από ακολουθίες παρατηρήσεων (πχ. διανύσματα χαρακτηριστικών), ενώ κάθε δραστηριότητα αντιστοιχεί σε μια συγκεκριμένη ακολουθία χαρακτηριστικών. Κάθε ακολουθία εικόνων μετατρέπεται σε μια ακολουθία από διανύσματα χαρακτηριστικών. Τα χαρακτηριστικά αυτά περιγράφουν την κίνηση του ατόμου σε κάθε καρέ. Έπειτα η παραγόμενη ακολουθία χαρακτηριστικών της εισόδου αναλύεται και υπολογίζεται η πιθανότητα τα διανύσματα της ακολουθίας να αντιστοιχούν σε κάποια συγκεκριμένη δραστηριότητα. Η ακολουθία εισόδου αντιστοιχίζεται σε κάποια κλάση δραστηριοτήτων όταν ο δείκτης ομοιότητας τους είναι αρκετά υψηλός. Οι ακολουθιακές μέθοδοι χωρίζονται σε τεχνικές αναγνώρισης βάσει προτύπων και σε τεχνικές αναγνώρισης βάσει μοντέλων.

- **Μέθοδοι βασισμένες σε πρότυπα (Exemplar-based approaches):**

Οι μέθοδοι βασισμένες σε πρότυπα αντιπροσωπεύουν την ανθρώπινη δραστηριότητα μέσω μιας ακολουθίας προτύπων ή ενός συνόλου ακολουθιών δειγμάτων εκτέλεσης δράσης. Όταν ένα νέο video εισόδου δίνεται, γίνεται σύγκριση της ακολουθίας διανυσμάτων χαρακτηριστικών που εξάγονται από το video με την αποθηκευμένη ακολουθία προτύπων (ή τις ακολουθίες δειγμάτων). Εάν η ομοιότητά τους είναι αρκετά υψηλή, το σύστημα είναι σε θέση να αναγνωρίσει την εκτέλεση μιας δραστηριότητας μέσα στην ακολουθία. Τα ποσοστά ομοιότητας και οι μορφές κάθε δραστηριότητας ποικίλλουν, αναλόγως με τον τρόπο που εκτελείται η κίνηση. Για την αντιμετώπιση των παραπάνω διαφορών, χρησιμοποιούνται τεχνικές δυναμικής ευθυγράμμισης / σύγκρισης προτύπων (Dynamic Time Warping- DTW) για το ταίριασμα δυο ακολουθιών με χρονικές αποκλίσεις.

- **Μέθοδοι βασισμένες σε μοντέλα κατάστασης (State model-based approaches):**

Σε αυτή την κατηγορία, η ανθρώπινη δραστηριότητα αντιμετωπίζεται σαν μοντέλο, το οποίο αποτελείται από ένα σύνολο καταστάσεων. Κάθε μοντέλο εκπαιδεύεται στατιστικά ώστε να ανταποκρίνεται σε ακολουθίες χαρακτηριστικών, οι οποίες ανήκουν στην κλάση της συγκεκριμένης δραστηριότητας. Το στατιστικό μοντέλο σχεδιάζεται με τρόπο ώστε να παράγει μια ακολουθία με καθορισμένη πιθανότητα. Δεδομένης μιας ακολουθίας εισόδου, για κάθε μοντέλο της βάσης υπολογίζεται η πιθανότητα να παραχθεί η παραπάνω ακολουθία από αυτό. Αυτή η πιθανότητα αποτελεί το δείκτη ομοιότητας, με τον οποίο ταξινομείται η ακολουθία εισόδου σε κάποια κλάση δραστηριοτήτων. Τέλος για την αναγνώριση κάποιας δραστηριότητας κατασκευάζεται ένας ταξινομητής - εκτιμητής μέγιστης πιθανοφάνειας (maximum likelihood estimation – MLE) ή ένας ταξινομητής μέγιστης εκ των υστέρων πιθανότητας (maximum a posteriori probability – MAP). Τα κρυφά Μαρκοβιανά μοντέλα (hidden Markov models – HMMs) και τα δυναμικά δίκτυα Bayes (dynamic Bayesian networks – DBNs) έχουν χρησιμοποιηθεί ευρέως στις μεθόδους βασισμένες σε μοντέλα κατάστασης.



Εικόνα 2.7: *Παράδειγμα κρυφού Μαρκοβιανού μοντέλου*

2.3.2 Ιεραρχικές μέθοδοι (Hierarchical approaches)

Οι ιεραρχικές μέθοδοι στηρίζονται στην ιδέα της αναγνώρισης υψηλού επιπέδου δραστηριοτήτων, μέσω της αναγνώρισης δραστηριοτήτων χαμηλότερου επιπέδου. Με άλλα λόγια, η συγκεκριμένη κατηγορία μεθόδων αναλαμβάνει την ανάλυση μιας σύνθετης δραστηριότητας σε επιμέρους συμβάντα (sub-events) μέχρι να προκύψουν πολύ απλές κινήσεις (atomic or primitive actions). Οι παραπάνω κινήσεις μπορούν να επεξεργαστούν με μεθόδους αναγνώρισης μονής στιβάδας. Πλεονέκτημα των ιεραρχικών προσεγγίσεων αποτελεί η δυνατότητα αναγνώρισης σύνθετων δραστηριοτήτων, το οποίο τις καθιστά κατάλληλες για την ανάλυση ομαδικών ενεργειών και κινήσεων αλληλεπίδρασης μεταξύ ατόμων ή/και αντικειμένων. Η παραπάνω κατηγορία διακρίνεται σε τρείς επιμέρους τεχνικές, οι οποίες παρουσιάζονται συνοπτικά στη συνέχεια.

- **Στατιστικές μέθοδοι (Statistical approaches)**

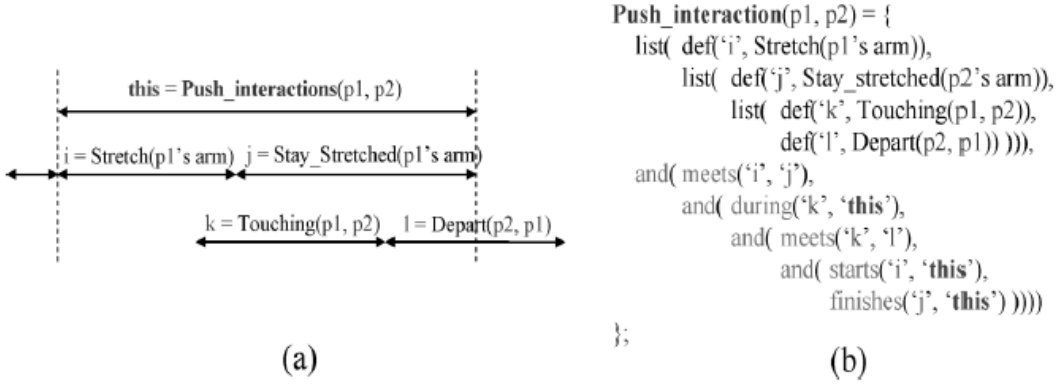
Κάνουν χρήση στατιστικών μοντέλων κατάστασης για την αναγνώριση της ανθρώπινης δραστηριότητας, όπως είναι τα κρυφά Μαρκοβιανά μοντέλα (hidden Markov models – HMMs) και τα δυναμικά δίκτυα Bayes (dynamic Bayesian networks – DBNs). Στα χαμηλότερα επίπεδα, οι απλές ή ατομικές κινήσεις αναγνωρίζονται από ακολουθίες χαρακτηριστικών διανυσμάτων, όπως γίνεται στα μοντέλα μονής στιβάδας. Τα μοντέλα δευτέρου επιπέδου αντιμετωπίζουν τις ακολουθίες ατομικών κινήσεων ως παρατηρήσεις, τις οποίες δημιουργούν τα ίδια. Για κάθε μοντέλο, υπολογίζεται η πιθανότητα αυτό να δημιουργεί μια ακολουθία από παρατηρήσεις και έπειτα μετράται η ομοιότητα ανάμεσα σε μια δραστηριότητα και την είσοδο του συστήματος.

- **Συντακτικές μέθοδοι (Syntactic approaches)**

Η συγκεκριμένη κατηγορία αντιμετωπίζει τις ανθρώπινες δραστηριότητες ως συμβολοσειρές, των οποίων κάθε σύμβολο ανταποκρίνεται σε μια απλή ατομική κίνηση. Η αναγνώριση των απλών κινήσεων γίνεται όπως και στην προηγούμενη κατηγορία. Η ανθρώπινη δραστηριότητα παρουσιάζεται σαν ένα σύνολο από κανόνες παραγωγής, οι οποίοι δημιουργούν μια συμβολοσειρά από ατομικές ενέργειες, και αναγνωρίζεται αξιοποιώντας τεχνικές συντακτικής ανάλυσης από το πεδίο των γλωσσών προγραμματισμού.

- **Περιγραφικές μέθοδοι (Description-based approaches)**

Σε αυτή την κατηγορία η ανθρώπινη δραστηριότητα υψηλού επιπέδου αντιμετωπίζεται ως ένα σύνολο από πιο απλές κινήσεις (π.χ. υπο-γεγονότα), οι οποίες περιγράφουν τις χωρικές, χρονικές και λογικές τους συσχετίσεις. Η αναγνώριση κάποιας δραστηριότητας επιτυγχάνεται αναζητώντας όλα τα υπο-γεγονότα, τα οποία έχουν προσδιοριστεί κατά την αναπαράστασή της. Οι περιγραφικές μέθοδοι έχουν τη δυνατότητα να χειρίζονται ενέργειες, οι οποίες γίνονται παράλληλα. Γι' αυτό το λόγο θεμελιώδες στοιχείο τους αποτελούν τα χρονικά διαστήματα (time intervals), τα οποία καθορίζουν τις χρονικές σχέσεις μεταξύ υπο-συμβάντων. Για τον καθορισμό αυτών των σχέσεων, έχουν οριστεί τα χρονικά κατηγορήματα του Allen **[5].**



Εικόνα 2.8: Παράδειγμα περιγραφικής μεθόδου. (*a) Χρονικά διαστήματα της κίνησης «σπρώξιμο» και των υπο-ενεργειών και (b) η αναπαράσταση τους σε γλώσσα προγραμματισμού*

Περισσότερες πληροφορίες για την έρευνα στις μεθόδους αναγνώρισης παρουσιάζονται στο **[4].** Ιδιαίτερο ενδιαφέρον εμφανίζουν οι μέθοδοι αναγνώρισης, οι οποίες επεξεργάζονται δεδομένα βάθους. Οι παραπάνω μέθοδοι περιγράφονται αναλυτικότερα στη συνέχεια, από μια διαφορετική οπτική σκοπιά.

2.4 Αναγνώριση Ανθρώπινης Κίνησης από Δεδομένα Βάθους (Action Recognition from 3D data)

Σε μια εικόνα βάθους, η τιμή κάθε pixel αντιστοιχεί στην απόσταση μεταξύ ενός σημείου του πραγματικού κόσμου και του αισθητήρα, ο οποίος παρέχει τις δομικές πληροφορίες της 3D σκηνής. Οι τεχνικές ανάλυσης κίνησης οι οποίες αξιοποιούν δεδομένα βάθους, ανάλογα με τα χαρακτηριστικά επεξεργασίας που χρησιμοποιούν, μπορούν να διακριθούν στις παρακάτω κατηγορίες **[6]**:

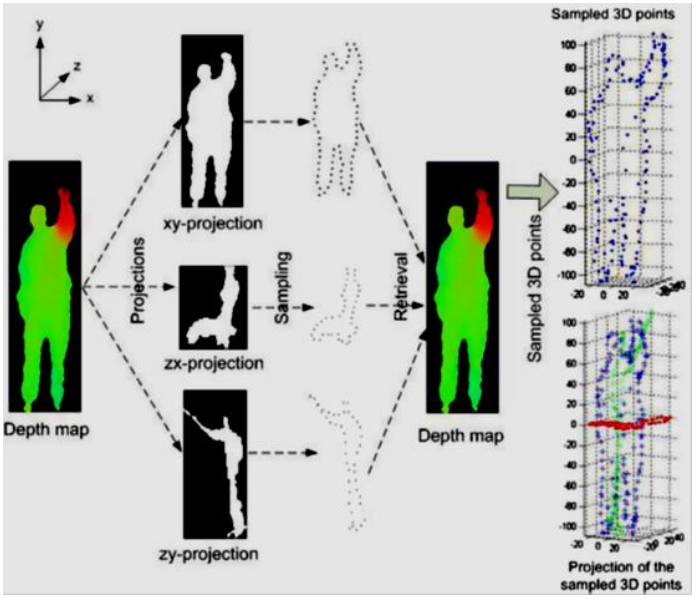
2.4.1 Αναγνώριση με Χρήση Τρισδιάστατης Σιλουέτας (3D Silhouettes)

Στα αρχικά στάδια αναγνώρισης της κίνησης, αναπτύχθηκαν και αξιοποιήθηκαν 2D σιλουέτες για την απλή αναπαράσταση του ανθρώπινου σχήματος από RGB εικόνες. Η χρονική εξέλιξη των παραπάνω σιλουετών αντιπροσώπευε την κίνηση του χρήστη και απέδιδε αρκετές πληροφορίες για το ανθρώπινο σώμα. Ωστόσο, η εξαγωγή των παραπάνω

φιγούρων σχετίζεται άμεσα με το σημείο προβολής και απαιτείται οποιαδήποτε κίνηση να είναι παράλληλη με την κάμερα. Επίσης η συγκεκριμένη τεχνική παρουσιάζει δυσκολίες αν οι συνθήκες του περιβάλλοντος και του φωτισμού δεν είναι καλές.

Σε μια εικόνα βάθους, η 3D σιλουέτα ενός ατόμου μπορεί να εξαχθεί με μεγαλύτερη ευκολία και ακρίβεια, ενώ παρέχονται πληροφορίες σχετικά με το σχήμα του σώματος όχι μόνο κατά μήκος της σιλουέτας, αλλά ολόκληρης της πλευράς που αντικρίζει την κάμερα. Οι τρέχοντες αλγόριθμοι, οι οποίοι χρησιμοποιούν 3D σιλουέτες, είναι κατάλληλοι για την αναγνώριση της δράσης ενός ατόμου και λειτουργούν καλύτερα σε απλές ατομικές ενέργειες. Υπάρχει, όμως, δυσκολία στην αναγνώριση πολύπλοκων δραστηριοτήτων λόγω των πληροφοριών που χάνονται κατά τον υπολογισμό της 2D προβολής των τρισδιάστατων δεδομένων. Επίσης ο χάρτης βάθους στην πραγματικότητα δίνει μόνο τις 3D σιλουέτες του ατόμου, το οποίο βλέπει προς την κάμερα, με αποτέλεσμα οι εξαγόμενες πληροφορίες να εξαρτώνται άμεσα από τη θέση προβολής.

Η συσχέτιση μεταξύ 3D και 2D σιλουετών παρουσιάζεται με λεπτομέρειες στα **[7]** και **[8]**.

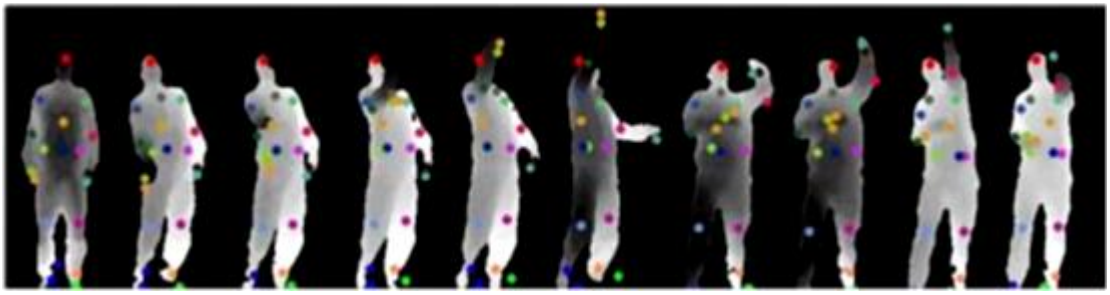


Εικόνα 2.9: *Παράδειγμα εξαγωγής χαρακτηριστικών με χρήση 3D silhouettes.*

2.4.2 Αναγνώριση με Χρήση Σκελετικών Δεδομένων (Skeletal Data)

Στα πλαίσια της ανάπτυξης των τεχνικών καταγραφής κίνησης, οι ερευνητές προσπάθησαν να εξάγουν «σκελετούς» από σιλουέτες ή να επισημάνουν τα κύρια μέρη του σώματος για την αναγνώριση κάποιας δραστηριότητας. Η παραπάνω έρευνα διευρύνθηκε με τον ορισμό των αρθρώσεων του ανθρώπινου σώματος από εικόνες βάθους, χρησιμοποιώντας μεθόδους αναγνώρισης αντικειμένων. Η κύρια ιδέα των αλγορίθμων, οι οποίοι επεξεργάζονται και υπολογίζουν σκελετικά δεδομένα, είναι ο υπολογισμός της διαφοράς θέσης μιας άρθρωσης μεταξύ του τρέχοντος χρονικού πλαισίου και το προηγούμενου. Με αυτό τον τρόπο ορίζεται η κίνηση που διαγράφει το ανθρώπινο σώμα, μέσω των σχετικών μετατοπίσεων στις θέσεις των αρθρώσεων. Η παρατηρούμενη σχετική μετατόπιση μπορεί επίσης να είναι καθοριστική για την εύρεση συγκεκριμένων τύπων κίνησης, ενώ τα σκελετικά δεδομένα μπορούν επίσης να καθορίσουν τον προσανατολισμό του σώματος.

Σε σχέση με τις τεχνικές χρήσης 3D σιλουέτας, τα χαρακτηριστικά των σκελετικών δεδομένων είναι αμετάβλητα ως προς τη θέση της κάμερας. Οι αντίστοιχοι αλγόριθμοι σκελετικής ανίχνευσης μπορούν να εξάγουν με ακρίβεια τις θέσεις των αρθρώσεων από διάφορες όψεις και μπορούν να επεκταθούν για την αναγνώριση ενεργειών αλληλεπίδρασης μεταξύ ατόμων. Επιπλέον, ο προσανατολισμός των δεδομένων είναι ανεξάρτητος του μεγέθους του ανθρώπινου σώματος. Ένας περιορισμός όμως των παραπάνω αλγορίθμων είναι ότι δεν δίνουν πληροφορίες για τα αντικείμενα του περιβάλλοντα χώρου. Κατά τη μοντελοποίηση της αλληλεπίδρασης χρήστη-αντικειμένου, η ανίχνευση και η παρακολούθηση των αντικειμένων πρέπει να συνδυαστούν. Επιπροσθέτως, δυσκολίες ανίχνευσης μπορεί να παρουσιαστούν αν η γωνία/κλίση καταγραφής της κίνησης δεν είναι κατάλληλη.



Εικόνα 2.10: *Παράδειγμα εξαγωγής σκελετικών δεδομένων από μια ακολουθία εικόνων βάθους, η οποία απεικονίζει την κίνηση 'tennis serve' στο MSRAction3D dataset*

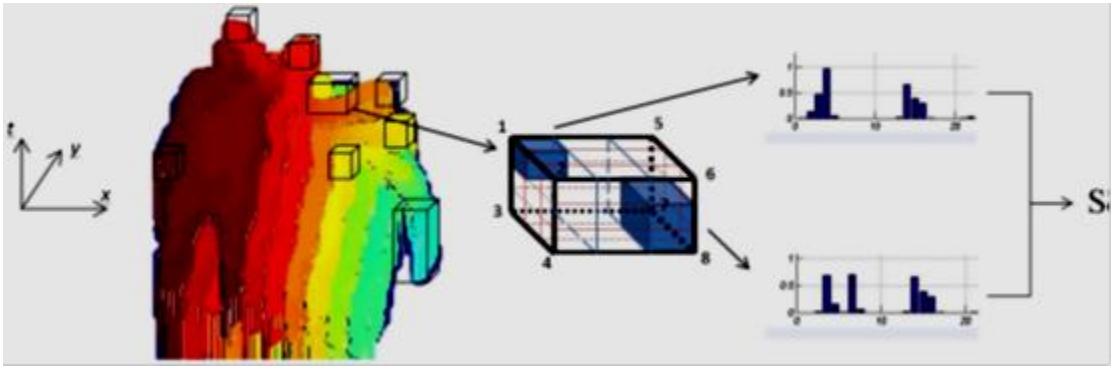
.

2.4.3 Αναγνώριση με Χρήση Τοπικών Χωροχρονικών Χαρακτηριστικών (Local Spatio-Temporal Features)

Το τοπικά χωροχρονικά χαρακτηριστικά γνωρίσματα είναι μια δημοφιλής μέθοδο για την αναγνώριση της δράσης από ένα video ή εικόνα βάθους. Εντοπίζουν σχήματα και κινήσεις ενός video και παρέχουν μια ανεξάρτητη αναπαράσταση των γεγονότων σε σχέση με το φόντο, τις πολλαπλές κινήσεις, τις κλίμακες και τις χωροχρονικές μετατοπίσεις. Το video θεωρείται ως τρισδιάστατος όγκος κατά μήκος του διαστήματος (x, y) και του χρονικού άξονα t. Στη γενική διαδικασία επεξεργασίας, τα τοπικά σημεία χωροχρονικού ενδιαφέροντος (STIPs) ανιχνεύονται αρχικά, και στη συνέχεια περιγραφικοί δείκτες (descriptors) ορίζονται με βάση αυτά. Για την ανίχνευση των τοπικών χωροχρονικών χαρακτηριστικών έχουν αναπτυχθεί διάφορα μαθηματικά μοντέλα όπως είναι οι Harris3D, Cuboid, Hessian και Dense Sampling detector, ενώ μερικοί διαδεδομένοι περιγραφικοί δείκτες είναι οι Cuboid descriptor (Dollár), HOG/HOF descriptor (Laptev), HOG3D descriptor (Kläser) και Extended Surf (ESURF) descriptor (Willems). Περισσότερες πληροφορίες για τη λειτουργία των παραπάνω μαθηματικών σχημάτων υπάρχουν στο **[9]**. Έπειτα η ταξινόμηση μπορεί να γίνει από τους περιγραφείς (με μεθόδους όπως η bag-of-words).

Τα τοπικά χωροχρονικά χαρακτηριστικά αποτελούν μια ικανοποιητική προσέγγιση για την αναγνώριση ενός πλήθους κατηγοριών δράσης με ποικίλες δυσκολίες. Όπως αναφέρθηκε, είναι αμετάβλητα στις χωροχρονικές μετατοπίσεις ή κλίμακες και αντιμετωπίζουν συστήματα πολλαπλών κινήσεων και αλληλεπιδράσεων μεταξύ ατόμων με άλλα άτομα ή αντικείμενα. Επειδή αυτά τα χαρακτηριστικά συνήθως εξάγονται άμεσα χωρίς την ανάγκη κατάτμησης και παρακολούθησης της κίνησης, οι αντίστοιχοι αλγόριθμοι είναι διαδεδομένοι σε ένα ευρύ φάσμα εφαρμογών. Ωστόσο, όπως οι προηγούμενες κατηγορίες, η χρήση των τοπικών χωροχρονικών χαρακτηριστικών έχει ορισμένα μειονεκτήματα. Ένα από αυτά είναι ότι τα χαρακτηριστικά εξαρτώνται από τη θέση/προβολή της κάμερας, καθώς

εξάγονται απευθείας από την περιοχή των {x, y, t.} Επίσης η τρέχουσα μέθοδος απαιτεί ολόκληρο το video ως είσοδο και οι αντίστοιχοι αλγόριθμοι επεξεργασίας δεν είναι αρκετά γρήγοροι, καθιστώντας τους ακατάλληλους για εφαρμογές σε πραγματικό χρόνο.

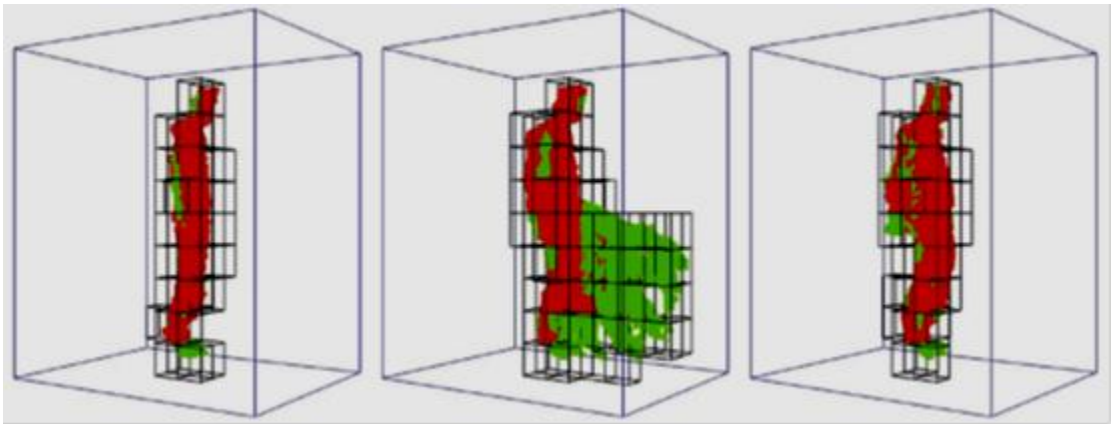


Εικόνα 2.11: *Παράδειγμα εξαγωγής τοπικών χωροχρονικών χαρακτηριστικών από ένα video βάθους.*

2.4.4 Αναγνώριση με Χρήση 3D Τοπικών Χαρακτηριστικών Πληρότητας (3D Occupancy Features)

Σε αυτή την κατηγορία, τα σημεία μπορεί να προβάλλονται στον 4D χώρο (x, y, z, t). Σε αυτόν το χώρο, κάποιες θέσεις θα καταλαμβάνονται από τα σημεία των δεδομένων, τα οποία καταγράφτηκαν στο video (δηλαδή τα σημεία, τα οποία ο αισθητήρας παρατήρησε στον πραγματικό κόσμο). Οι θέσεις αυτές παίρνουν τιμή 0 ή 1. Γενικά τα τοπικά χαρακτηριστικά πληρότητας μπορεί να ορίζονται στο χώρο (x, y, z) ή (x, y, z, t). Το πρώτο πεδίο περιγράφει τα δεδομένα μιας συγκεκριμένης χρονικής στιγμής, ενώ το δεύτερο πεδίο περιγράφει τα ατομικά γεγονότα σε ένα συγκεκριμένο χρονικό διάστημα. Στα πλαίσια έρευνας διαμορφώθηκε μια λειτουργία 3D Τοπικής Συμπεριφοράς (Local Occupancy Patterns) για την περιγραφή δεδομένων βάθους σε αλληλεπιδράσεις χρήστη-αντικειμένου. Ένα παράδειγμα είναι όταν ο χρήστης κρατάει ένα φλιτζάνι, ο χώρος γύρω από το χέρι του καταλαμβάνεται από το αντικείμενο. Η περιοχή (x, y, z) γύρω από την άρθρωση του χεριού χωρίζεται σε ένα χωρικό πλέγμα (Nx, Ny, Nz) και έπειτα υπολογίζεται και ομαλοποιείται το πλήθος των σημείων που εμπίπτουν σε κάθε ομάδα (χέρι-φλιτζάνι) για να οριστεί το χαρακτηριστικό πληρότητας της ομάδας.

Τα χαρακτηριστικά τοπικής πληρότητας που ορίζονται στο χώρο (x, y, z, t) είναι όμοια με τα τοπικά χωροχρονικά χαρακτηριστικά της προηγούμενης κατηγορίας, καθώς και τα δύο περιγράφουν τα τοπικά δεδομένα στο χωροχρόνο. Τα τοπικά χωροχρονικά χαρακτηριστικά αντιμετωπίζουν τη διάσταση z ως τιμές pixels στο πεδίο (x, y, t), ενώ τα τοπικά χαρακτηριστικά πληρότητας προβάλλουν τα δεδομένα στον 4D χώρο (x, y, z, t) με τιμές 0-1. Επίσης τα τελευταία μπορεί να είναι πολύ αραιά (πλειονότητα μηδενικών τιμών), ενώ τα χωροχρονικά χαρακτηριστικά δεν είναι. Τέλος τα χωροχρονικά χαρακτηριστικά περιέχουν πληροφορίες σχετικά με το περιβάλλον, ενώ τα τοπικά χαρακτηριστικά πληρότητας περιέχουν μόνο πληροφορίες γύρω από ένα συγκεκριμένο σημείο σε ένα (x, y, z, t) διάστημα.



Εικόνα 2.12: *Παράδειγμα χρήσης 3D τοπικών χαρακτηριστικών πληρότητας για την αναγνώριση μιας κίνησης.*

2.4.5 Αναγνώριση με χρήση 3D οπτικής ροής (3D optical flow)

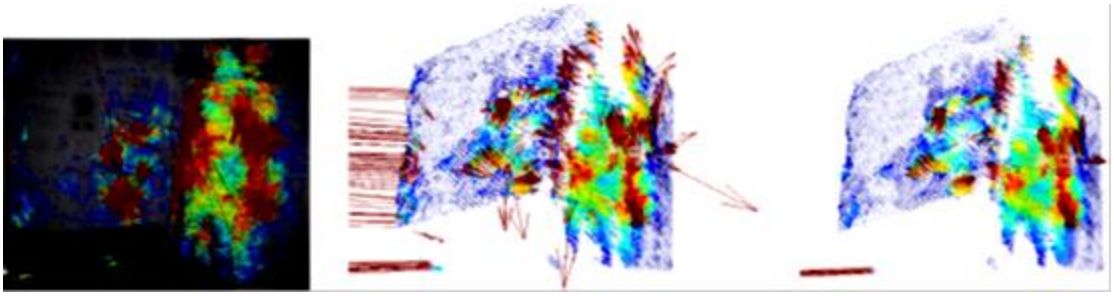
Η οπτική ροή είναι η κατανομή των φαινομενικών ταχυτήτων κίνησης των μοτίβων φωτεινότητας σε μια εικόνα, οι οποίες προκύπτουν τόσο από την κίνηση των αντικειμένων όσο και από την κίνηση του ατόμου. Συγκεκριμένα στις δύο διαστάσεις η οπτική ροή καθορίζει τη μετακίνηση κάθε pixel μεταξύ χρονικά διαδοχικών εικόνων, ενώ στις τρεις διαστάσεις καθορίζει τη χωρική μετατόπιση κάθε voxel μεταξύ γειτονικών χωρικών περιοχών/όγκων **[10]**, **[11]**. Χρησιμοποιείται ευρέως σε εικόνες για ανίχνευση κίνησης, κατάτμηση αντικειμένων και μέτρηση στερεοφωνικών ανωμαλιών. Επίσης αποτελεί ένα δημοφιλές χαρακτηριστικό για την αναγνώριση δραστηριοτήτων σε video. Όταν είναι διαθέσιμες πολλές κάμερες, η ενσωμάτωση τους σε διαφορετικές οπτικές γωνίες επιτρέπει την παρατήρηση/ορισμό ενός πεδίου 3D κίνησης (ροή σκηνής). Ωστόσο, οι διακυμάνσεις της έντασης δεν επαρκούν για την εκτίμηση της κίνησης και πρόσθετοι περιορισμοί όπως η ομαλότητα πρέπει να εισαχθούν στις περισσότερες περιπτώσεις. Οι κάμερες βάθους παρέχουν χρήσιμες γεωμετρικές πληροφορίες από τις οποίες μπορούν να προκύψουν σταθεροί περιορισμοί ως προς την ομαλότητα της κίνησης.

Μερικοί αλγόριθμοι υπολογίζουν τη ροή της 3D σκηνής μέσω συνόλων σημείων, χρησιμοποιώντας τη μέθοδο αναζήτησης πλησιέστερου γείτονα, η οποία προσαρμόζεται με βάση τις 3D γεωμετρικές συντεταγμένες και τα RGB χρωματικά δεδομένα. Επιπροσθέτως, η οπτική ροή υπολογίζεται μόνο για ορισμένες περιοχές της τρισδιάστατης σκηνής, οι οποίες αντιπροσωπεύουν το χρήστη με ένα σύμπλεγμα 4D σημείων. Ωστόσο, η χρήση της για την αναγνώριση ενεργειών είναι αρκετά περιορισμένη και ο υπολογισμός της αποτελεί χρονοβόρα διαδικασία. Από μαθηματικής πλευράς, ο υπολογισμός της οπτικής ροής βασίζεται στη χρήση τοπικών παραγώγων μιας δεδομένης σειράς εικόνων και διακρίνεται στην εκτέλεση δύο βημάτων:

- τον υπολογισμό των παραγώγων χωροχρονικής έντασης (η οποία ισοδυναμεί με τη μέτρηση των ταχυτήτων στις τοπικές δομές έντασης) και

- την ολοκλήρωση των ταχυτήτων σε πλήρεις/τελικές ταχύτητες (για παράδειγμα, τοπικά αυτό γίνεται μέσω ενός υπολογισμού των ελαχίστων τετραγώνων).

Το μαθηματικό υπόβαθρο και ο υπολογισμός της οπτικής ροής στις δύο και τρεις διαστάσεις αναλύονται περαιτέρω στο **[12]**.



Εικόνα 2.13: Παράδειγμα 3D οπτικής ροής. (α) 2D διανύσματα ταχύτητας υπολογιζόμενα με χρήση οπτικής ροής, (β) 3D διανύσματα ταχύτητας, (γ) το τελικό ομαλοποιημένο αποτέλεσμα από ένα ενδιάμεσο φίλτρο. Κάθε τρίτο διάνυσμα ταχύτητας εμφανίζεται και χρωματίζεται σε σχέση με το μήκος του: το κόκκινο δηλώνει ένα μεγάλο διάνυσμα κίνησης και ένα μπλε ένα μικρό.

Εφαρμογές και παραδείγματα των παραπάνω κατηγοριών παρουσιάζονται αναλυτικά στο **[13]**.

2.5 Hardware - Συσκευή Kinect

2.5.1 Εισαγωγή

Η συσκευή Kinect είναι ένας RGB-D αισθητήρας, ο οποίος παρέχει συγχρονισμένες εικόνες χρώματος και βάθους. Αρχικά χρησιμοποιήθηκε ως συσκευή εισόδου από τη Microsoft για την κονσόλα παιχνιδιών Xbox. Με έναν αλγόριθμο καταγραφής ανθρώπινης 3D κίνησης, επιτρέπει αλληλεπιδράσεις μεταξύ χρηστών και παιχνιδιού χωρίς να είναι απαραίτητη η χρήση κάποιου χειριστηρίου.

Εικόνα 2.14: *Συσκευή Kinect. Αριστερά: Διαμόρφωση υλικού του Kinect, στην οποία επισημαίνεται η θέση κάθε αισθητήρα. Δεξιά: Δύο δείγματα εικόνων, οι οποίες έχουν ληφθεί από την κάμερα RGB και την κάμερα βάθους.*

2.5.2 Λειτουργία και Χαρακτηριστικά του Kinect

Η διάταξη ενός αισθητήρα Kinect αποτελείται από έναν υπέρυθρο προβολέα (IR), μια κάμερα IR και μια έγχρωμη κάμερα (Εικόνα 2.14). Ο αισθητήρας βάθους της συσκευής περιλαμβάνει τον προβολέα IR και την κάμερα IR. Η γενική ιδέα της λειτουργίας του Kinect, η οποία στηρίζεται στο σύστημα Light Coding, είναι η εξής: Αρχικά ο υπέρυθρος προβολέας (IR) εκπέμπει ένα καθορισμένο σύνολο/μοτίβο σημείων φωτός IR στην 3D σκηνή,

στην οποία γίνεται η καταγραφή της κίνησης. Το εκπεμπόμενο φως δε γίνεται αντιληπτό από την έγχρωμη κάμερα, ωστόσο η κάμερα IR αντιλαμβάνεται και καταγράφει τις αντανακλάσεις IR.

Η γεωμετρική σχέση μεταξύ του προβολέα υπέρυθρων ακτινών (IR) και της κάμερας IR είναι γνωστή και επιτυγχάνεται μέσω μιας διαδικασίας βαθμονόμησης (off-line calibration). Επειδή κάθε τοπικό σύνολο/μοτίβο σημείων φωτός IR είναι μοναδικό, είναι δυνατή η αντιστοίχιση μεταξύ των παρατηρούμενων τοπικών σημείων IR στην εικόνα με τα προϋπολογισμένα σημεία του προβολέα. Η διαφορά της απόστασης μεταξύ των παραπάνω σημείων υποδεικνύει την αντίστοιχη θέση κάποιου αντικειμένου στη 3D σκηνή.

Τα χαρακτηριστικά του hardware, το οποίο εμπλέκεται στην παραπάνω λειτουργία, περιγράφονται στη συνέχεια:

- **Κάμερα RGB:** Παρέχει τρία βασικά χρωματικά συστατικά στο video. Η κάμερα λειτουργεί σε 30Hz και μπορεί να προσφέρει εικόνες ανάλυσης 640 x 480 pixels με 8-bit ανά κανάλι. Το Kinect έχει επίσης τη δυνατότητα να παράγει εικόνες υψηλότερης ανάλυσης 1280 x 1024 pixels με ρυθμό 10 καρέ ανά δευτερόλεπτο.

- **3D αισθητήρας βάθους:** Αποτελείται από έναν προβολέα laser IR και μια κάμερα IR.Ο προβολέας και η κάμερα δημιουργούν ένα χάρτη βάθους, ο οποίος παρέχει πληροφορίες για την απόσταση ενός αντικειμένου από την κάμερα. Ο αισθητήρας έχει ένα πρακτικό όριο διακύμανσης από απόσταση 0.8 έως 3.5 μέτρα και εξάγει video με ρυθμό 30 καρέ ανά δευτερόλεπτο και ανάλυση 640 x 480 pixels. Το γωνιακό πεδίο θέασης είναι 57° οριζόντια και 43° κατακόρυφα.

- **Μηχανική κλίση:** Είναι ένας άξονας για τη ρύθμιση του αισθητήρα. Ο αισθητήρας μπορεί να βρίσκεται σε κλίση έως και 27° προς τα πάνω ή προς τα κάτω.

Επισημαίνεται ότι για τη σωστή αξιοποίηση και λειτουργία της συσκευής Kinect είναι απαραίτητο το αντίστοιχο λογισμικό. Το λογισμικό του Kinect αναφέρεται στις βιβλιοθήκες ανάπτυξης του Kinect, καθώς και στο σύνολο των αλγορίθμων, οι οποίοι διευκολύνουν τη χρήση της συσκευής. Διαθέσιμα εργαλεία για τη διαχείριση της συσκευής Kinect είναι τα OpenNI , Microsoft Kinect SDK και OpenKinect (LibFreeNect). Στα πλαίσια της διπλωματικής εργασίας γίνεται αναφορά στη λειτουργία του OpenNI σε συνδυασμό με το middleware NITE.

Με τον παραπάνω εξοπλισμό, η συσκευή Kinect παρέχει πλήθος δυνατοτήτων όπως είναι ο εντοπισμός και η αναγνώριση αντικειμένων, η ανάλυση και αναγνώριση ανθρώπινης κίνησης (εντοπισμός κίνησης, αναγνώριση χειρονομιών, ανθρώπινων εκφράσεων κτλ.) και η 3D χαρτογράφηση εσωτερικών χώρων (indoor 3D mapping). Περισσότερες πληροφορίες για τις δυνατότητες του Kinect παρουσιάζονται στο **[14].**

2.5.3 Εξαγωγή Σκελετικών Δεδομένων

Το Kinect χρησιμοποιεί δύο μεθόδους για την εκτίμηση της στάσης του χρήστη και της θέσης των σκελετικών του αρθρώσεων: Body Part Classification (BPC) και Offset Joint Regression(OJR).

Η ιδέα της εύρεσης των σκελετικών δεδομένων στηρίζεται στο διαχωρισμό όλων των pixels της εικόνας βάθους με βάση το μέλος του σώματος c, το οποίο πιθανά απεικονίζουν. Η κύρια διαφορά των δύο αλγορίθμων είναι ότι ο BPC αλγόριθμος, όταν διαχωρίζει όλα τα

pixels, χρησιμοποιεί μια ενδιάμεση αναπαράσταση για τις πιθανές θέσεις των αρθρώσεων και έπειτα υπολογίζει τις πραγματικές 3D συντεταγμένες μέσω αυτής της αναπαράστασης. Στον OJR αλγόριθμο ο εντοπισμός των 3D αρθρώσεων γίνεται με άμεσο τρόπο.

Εικόνα 2.15: *Διαδικασία Εξαγωγής της θέσης των αρθρώσεων.*

- **Γενικά χαρακτηριστικά εικόνων βάθους (Depth Image Features)**

Για ένα δεδομένο pixel u μιας εικόνας βάθους, η συνάρτηση απόστασης ορίζεται ως:

$$f(u|\varphi) = z\left(u + \frac{\delta_1}{z(u)}\right) - z\left(u + \frac{\delta_2}{z(u)}\right) \qquad (2.1)$$

όπου οι παράμετροι φ = (δ_1, δ_2) περιγράφουν τις 2D μετατοπίσεις του pixel u και η συνάρτηση z(u) δίνει την τιμή βάθους του pixel. Η κανονικοποίηση κατά 1/z(u) καθιστά την απόσταση ανεξάρτητη από το βάθος της εικόνας, δηλαδή η θέση της κάμερας δεν επηρεάζει το αποτέλεσμα. Αν ένα pixel u' βρίσκεται στον περιβάλλοντα χώρο ή εκτός των ορίων της κίνησης, η z(u') λαμβάνει μια πολύ μεγάλη θετική τιμή.

- **Randomized Decision Trees**

Ένα δάσος αποτελείται από ένα σύνολο δέντρων απόφασης T (Decision Trees). Κάθε δέντρο αποτελείται από διαχωριστικούς ενδιάμεσους κόμβους n και κόμβους-φύλλα l. Κάθε κόμβος n περιλαμβάνει ένα σύνολο παραμέτρων θ = (φ,τ), όπου φ είναι η προαναφερθείσα παράμετρος μετατόπισης και τ είναι ένα κατώφλι. Για να γίνει μια πρόβλεψη για κάποιο pixel u μιας εικόνας, ξεκινώντας από τη ρίζα γίνεται διάσχιση του δέντρου T μέχρι να βρεθεί κάποιος κόμβος-φύλλο. Η διάσχιση του δέντρου επιτυγχάνεται με τον υπολογισμό, σε κάθε κόμβο n, της συνάρτησης:

$$h(u, \theta_n) = [f(u, \varphi_n) \geq T_n] \qquad (2.2)$$

Αν η h(u, θ_n) παίρνει τιμή 1, το μονοπάτι διάσχισης συνεχίζει προς το αριστερό παιδί του κόμβου n, διαφορετικά συνεχίζει προς το δεξί παιδί. Η διαδικασία επαναλαμβάνεται μέχρι το μονοπάτι να καταλήξει σε κάποιο φύλλο l(u) (για το συγκεκριμένο pixel u). Η παραπάνω διάσχιση γίνεται για κάθε pixel u, για όλα τα δέντρα T, σχηματίζοντας έτσι σύνολα κόμβων-φύλλων L(u) = [$l(u)_{t=1..t=T}$], ένα για κάθε pixel.

Στη συγκεκριμένη περίπτωση σε κάθε κόμβο-φύλλο l είναι επίσης αποθηκευμένη μια συνάρτηση πρόβλεψης, η οποία χρησιμοποιείται για την ταξινόμηση κάθε pixel. Στον BPC αλγόριθμο, ως μοντέλο πρόβλεψης, χρησιμοποιείται μια πυκνότητα πιθανότητας pl(c) στα μέλη του σώματος c.

- **Body Part Classification (BPC)**

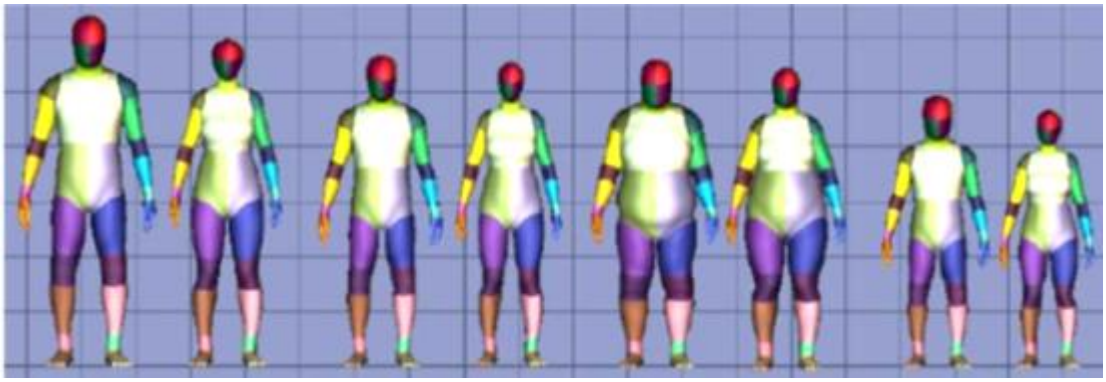
Ο BPC αλγόριθμος προβλέπει μια ετικέτα μέλους c για κάθε pixel, ως ενδιάμεσο βήμα του υπολογισμού των θέσεων των αρθρώσεων του σώματος. Αυτό επιτυγχάνεται με τη χρήση των Randomized Decision Trees, τα οποία αποθηκεύουν την κατανομή pl(c) κάθε μέλους του σώματος c σε κάθε φύλλο l. Για ένα δεδομένο pixel εισόδου, η διάσχιση του δέντρου οδηγεί στο φύλλο l = l(u), όπου ανακτάται η τιμή της κατανομής pl(c). Για την τελική ταξινόμηση του pixel, υπολογίζεται ο μέσος όρος των κατανομών όλων των δέντρων του δάσους:

$$p(c|u) = \frac{1}{T}\sum_{l \in L(u)} pl(c) \qquad (2.3)$$

Στη συνέχεια, με βάση την ταξινόμηση των pixels, υπολογίζεται η θέση x(u) κάθε pixel u στις 3D διαστάσεις, προσδιορίζονται οι πραγματικές συντεταγμένες των αρθρώσεων όλων των μελών του σώματος και δημιουργείται ο γενικός σκελετός. Οι τιμές της παραμέτρου φ(u) λαμβάνονται μέσω δειγματοληψίας στη γειτονιά του u.
Η σωστή λειτουργία του παραπάνω αλγορίθμου απαιτεί κάποια διαδικασία εκπαίδευσης. Αυτή επιτυγχάνεται με τη χρήση παραδειγμάτων/εικόνων βάθους, στις οποίες οι θέσεις των αρθρώσεων του σκελετού και η ταξινόμηση των pixels είναι δεδομένα.

- **Offset Joint Regression (OJR)**

Η προσέγγιση του OJR αλγορίθμου αποσκοπεί στην άμεση πρόβλεψη της θέσης των αρθρώσεων, χωρίς τη χρήση κάποιας ενδιάμεσης αναπαράστασης των συντεταγμένων των αρθρώσεων στις δύο διαστάσεις. Για το σκοπό αυτό χρησιμοποιείται ένα δάσος παλινδρόμησης (Regression Forest) (του οποίου οι κόμβοι-φύλλα κάνουν συνεχείς προβλέψεις). Σε κάθε κόμβο-φύλλο l αποθηκεύεται μια κατανομή, η οποία αφορά τη σχετική 3D μετατόπιση της προβολής x(u) καθενός pixel u από τη θέση της ζητούμενης άρθρωσης. Με αυτόν τον τρόπο, οι προβλέψεις των θέσεων των αρθρώσεων γίνονται άμεσα στις 3D συντεταγμένες του πραγματικού κόσμου.



Εικόνα 2.16: *Σύνολο εικόνων βάθους για την εκπαίδευση των αλγορίθμων.*

Περισσότερες πληροφορίες για τη λειτουργία και την εκπαίδευση των δύο αλγορίθμων παρουσιάζονται στο **[15]**.

2.6 Software – OpenNI/NiTE

2.6.1 Φυσική Αλληλεπίδραση

Ο όρος Φυσική Αλληλεπίδραση (NI) αναφέρεται στην αλληλεπίδραση Ανθρώπου-Συσκευής, η οποία βασίζεται στις ανθρώπινες αισθήσεις, επικεντρωμένες κυρίως στην ακοή και την όραση. Ορισμένα παραδείγματα χρήσης NI περιλαμβάνουν:

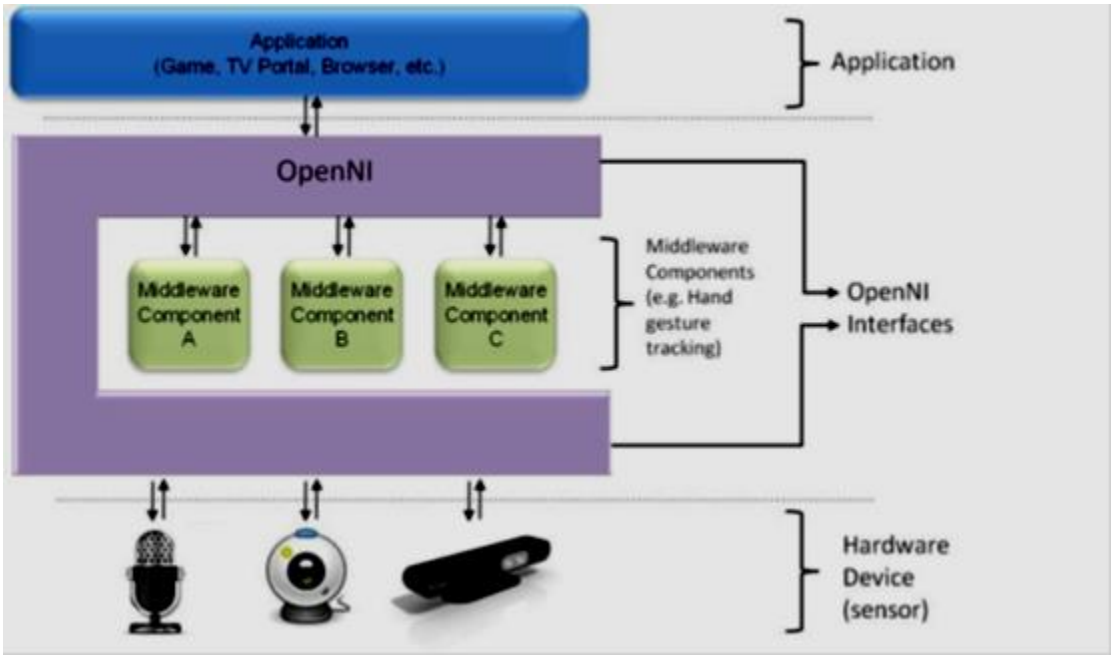
- Αναγνώριση ομιλίας μέσω συσκευών, οι οποίες λαμβάνουν οδηγίες μέσω φωνητικών εντολών.
- Χειρονομίες, με τις οποίες αναγνωρίζονται εντολές για τον έλεγχο των συσκευών.
- Παρακολούθηση και ερμηνεία της κίνησης σώματος για εφαρμογές σε παιχνίδια κτλ.

2.6.2 OpenNI

Το OpenNI (Open Natural Interaction) είναι μια πλατφόρμα, η οποία καθορίζει APIs για την ανάπτυξη εφαρμογών που χρησιμοποιούν τη φυσική αλληλεπίδραση **[16]**. Ο κύριος σκοπός του OpenNI είναι να διαμορφώσει ένα τυπικό API, το οποίο επιτρέπει την επικοινωνία με:

- Αισθητήρες εικόνας και ήχου .
- Middleware οπτικής και ακουστικής σύλληψης (τα οποία αναλύουν και επεξεργάζονται οπτικά και ακουστικά δεδομένα στη σκηνή καταγραφής).

Συγκεκριμένα το OpenNI παρέχει ένα σύνολο API, τα οποία υλοποιούνται από τις συσκευές αισθητήρων και ένα σύνολο API, τα οποία υλοποιούνται από το middleware. Ο διαχωρισμός των αισθητήρων από το middleware προσφέρει ευελιξία στην ανάπτυξη και μεταφορά προγραμμάτων σε διαφορετικές πλατφόρμες χωρίς να είναι απαραίτητες επιμέρους αλλαγές, με αποτέλεσμα οι εφαρμογές να μπορούν να αναπτυχθούν ανεξάρτητα από τους παρόχους των αισθητήρων ή του middleware. Η ιεραρχία του OpenNI μπορεί να διακριθεί σε τρία στρώματα, όπως φαίνεται στην εικόνα 2.17:



Εικόνα 2.17: *Ιεραρχία λειτουργικών στρωμάτων σε σύστημα καταγραφής δεδομένων.*

- Ανώτερο στρώμα: Αντιπροσωπεύει το λογισμικό, το οποίο υλοποιεί εφαρμογές φυσικής αλληλεπίδρασης πάνω από το OpenNI.

- Μεσαίο στρώμα: Αντιπροσωπεύει το OpenNI, παρέχοντας διασυνδέσεις επικοινωνίας για την αλληλεπίδραση τόσο με τους αισθητήρες όσο και με τα συστατικά του middleware, τα οποία λαμβάνουν τα δεδομένα από τον αισθητήρα.

- Κατώτερο στρώμα: Εμφανίζει τις συσκευές υλικού/hardware, οι οποίες καταγράφουν τις οπτικές και ηχητικές πληροφορίες της σκηνής.

2.6.3 Δυνατότητες του OpenNI

Το OpenNI υποστηρίζει την ανάλυση και επεξεργασία δεδομένων, τα οποία μπορεί να παρέχονται από έναν 3D αισθητήρα, μια RGB κάμερα, μια IR κάμερα υπερύθρων ή μια ακουστική συσκευή/μικρόφωνο. Οι λειτουργικότητες του μπορούν να διακριθούν σε:

- Ανάλυση σώματος: Περιλαμβάνει το λογισμικό, το οποίο εξάγει πληροφορίες σχετικά με το σώμα του χρήστη (αρθρώσεις, κέντρο βάρους κτλ.) μέσω επεξεργασίας των καταγεγραμμένων δεδομένων.
- Ανάλυση χεριών.
- Αναγνώριση χειρονομιών: Αφορά την αναγνώριση συγκεκριμένων χειρονομιών.
- Ανάλυση της σκηνής: Περιλαμβάνει το λογισμικό, το οποίο αναλύει την καταγεγραμμένη σκηνή και εξάγει πληροφορίες σχετικά με το διαχωρισμό προσκηνίου/παρασκηνίου, την αναγνώριση μεμονωμένων ατόμων στο χώρο και την κάτοψη του χώρου.

Ορισμένες από τις επιμέρους υλοποιημένες δυνατότητες του OpenNI, οι οποίες επιτελούν τα παραπάνω, είναι οι εξής:

- Alternative View: Επιτρέπει σε κάθε τύπο γεννήτριας χάρτη (βάθους, εικόνας, IR) να μετασχηματίσει τα παρατηρούμενα δεδομένα, ώστε να εμφανίζονται σα να έχει μετατοπιστεί ο αισθητήρας σε άλλη θέση.

- Cropping: Ενεργοποιεί μια γεννήτρια χάρτη (βάθους, εικόνας, IR) για την έξοδο μιας επιλεγμένης μόνο περιοχής του πλαισίου. Όταν είναι ενεργοποιημένη αυτή η λειτουργία, το μέγεθος του χάρτη που δημιουργείται μειώνεται για να χωρέσει σε χαμηλότερη ανάλυση (λιγότερα pixels).

- Pose Detection: Δίνει τη δυνατότητα σε μια γεννήτρια χρήστη να αναγνωρίζει πότε ο χρήστης βρίσκεται σε μια συγκεκριμένη στάση.

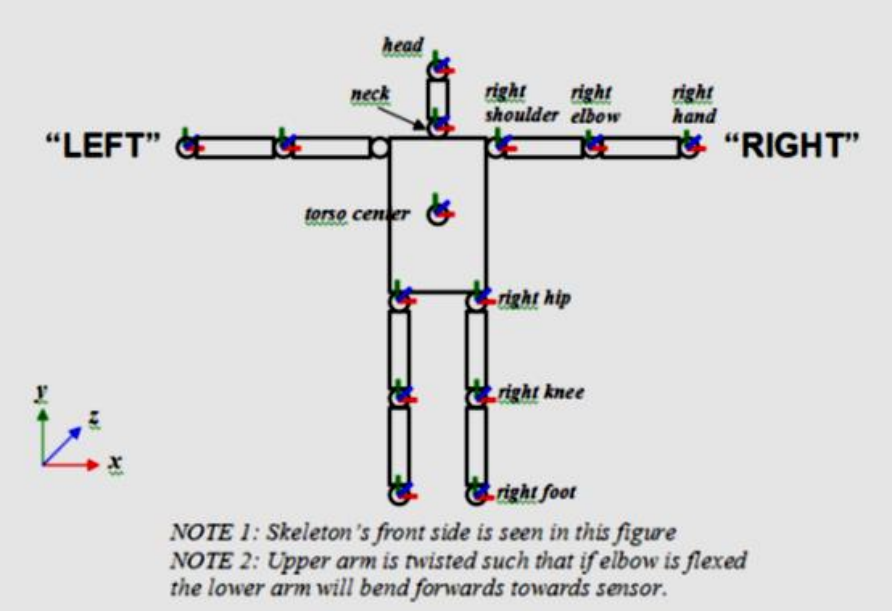
- Skeleton: Επιτρέπει σε μια γεννήτρια χρήστη να εξάγει τα δεδομένα του σκελετού του. Αυτά τα δεδομένα περιλαμβάνουν τη θέση των σκελετικών αρθρώσεων, την ικανότητα εντοπισμού του σκελετού και τις δυνατότητες βαθμονόμησης του χρήστη.

- User Position: Ενεργοποιεί μια γεννήτρια βάθους για τη βελτιστοποίηση του χάρτη βάθους, ο οποίος δημιουργείται για μια συγκεκριμένη περιοχή της σκηνής.

2.6.4 Middleware NiTE

Το NiTE (PrimeSense's Natural Interaction Technology for End-User) , το οποίο αναπτύχθηκε από την εταιρεία PrimeSense, αποτελεί middleware υποστήριξης πολλαπλών πλατφόρμων για την επεξεργασία 3D δεδομένων στο πεδίο της όρασης υπολογιστών **[17].** Παρέχει στην εφαρμογή ένα API ελέγχου χρήστη (είτε πρόκειται για έλεγχο κίνησης χεριών είτε για έλεγχο κίνησης όλου του σώματος), καθώς ερμηνεύει τα δεδομένα σύλληψης του 3D αισθητήρα στις τρεις διαστάσεις μέσω των εικόνων βάθους. Οι αλγόριθμοι ελέγχου/όρασης του NiTE αξιοποιούν πληροφορίες βάθους, χρώματος, IR και ήχου, οι οποίες λαμβάνονται από τους αισθητήρες και εκτελούν λειτουργίες όπως είναι η παρακολούθηση/αναγνώριση κίνησης χεριών, ο εντοπισμός του σκελετού χρήστη, η ανάλυση της σκηνής (διαχωρισμός χρήστη από το περιβάλλον) κτλ.

Ως προς την εξαγωγή σκελετικών δεδομένων, οι αλγόριθμοι όρασης εντοπίζουν και αναγνωρίζουν ένα πλήθος βασικών αρθρώσεων του ανθρώπινου σώματος, όπως παρουσιάζονται στην ακόλουθη εικόνα.



Εικόνα 2.18: *Σύστημα αναπαράστασης σκελετικών δεδομένων του NiTE.*

ΚΕΦΑΛΑΙΟ 3

Παρουσίαση της Βάσης Δράσεων THETIS και του Αλγορίθμου 3D Cylindrical Trace Transform

3.1 Εισαγωγή

Η αναγνώριση της ανθρώπινης δράσης, όπως αναφέρθηκε, αποτελεί ένα σημαντικό πεδίο της όρασης των υπολογιστών, το οποίο περιλαμβάνει μεγάλο πλήθος εφαρμογών (συστήματα επιτήρησης κτλ.). Επειδή το ερευνητικό ενδιαφέρον για το συγκεκριμένο πεδίο είναι ευρύ, υπάρχει ένας μεγάλος αριθμός προτεινόμενων τεχνικών, οι οποίες προσπαθούν να παρέχουν λύσεις σε διάφορα προβλήματα σχετικά με την αυτοματοποιημένη αναγνώριση κινήσεων. Σε αυτή την προσπάθεια, έχουν αναπτυχθεί σύνολα δεδομένων, τα οποία χρησιμοποιούνται για την ανάλυση διάφορων δραστηριοτήτων/σεναρίων.

Με βάση τα παραπάνω σενάρια, τα σύνολα δεδομένων μπορούν να διακριθούν σε τρεις κατηγορίες. Η πρώτη κατηγορία περιλαμβάνει σύνολα που έχουν σχεδιαστεί για την αξιολόγηση συστημάτων γενικής αναγνώρισης δράσης. Τέτοια σύνολα είναι το KHT **[18]** και το Weizmann **[4]**, τα οποία περιγράφονται στη συνέχεια. Τα σύνολα δεδομένων της δεύτερης κατηγορίας προσανατολίζονται κυρίως σε εφαρμογές, οι οποίες προκύπτουν από ρεαλιστικά περιβάλλοντα (όπως αεροδρόμια, χώροι στάθμευσης κτλ.). Ένα παράδειγμα αυτής της κατηγορίας παρουσιάζεται ως πρόκληση στο **[19]**. Η πρόκληση έχει ως στόχο να παρακινήσει τον ερευνητή να μελετήσει τεχνικές, οι οποίες αντιμετωπίζουν ζητήματα αναγνώρισης ενεργειών σε video χαμηλής ανάλυσης. Τέλος, η τρίτη κατηγορία περιλαμβάνει μια σειρά δεδομένων που έχουν συγκεντρωθεί από τη συλλογή video, τα οποία προέρχονται από μέσα ενημέρωσης, τηλεοπτικά προγράμματα και ταινίες. Ένα τέτοιο παράδειγμα είναι το σύνολο δεδομένων που παρουσιάζεται στο **[20]**.

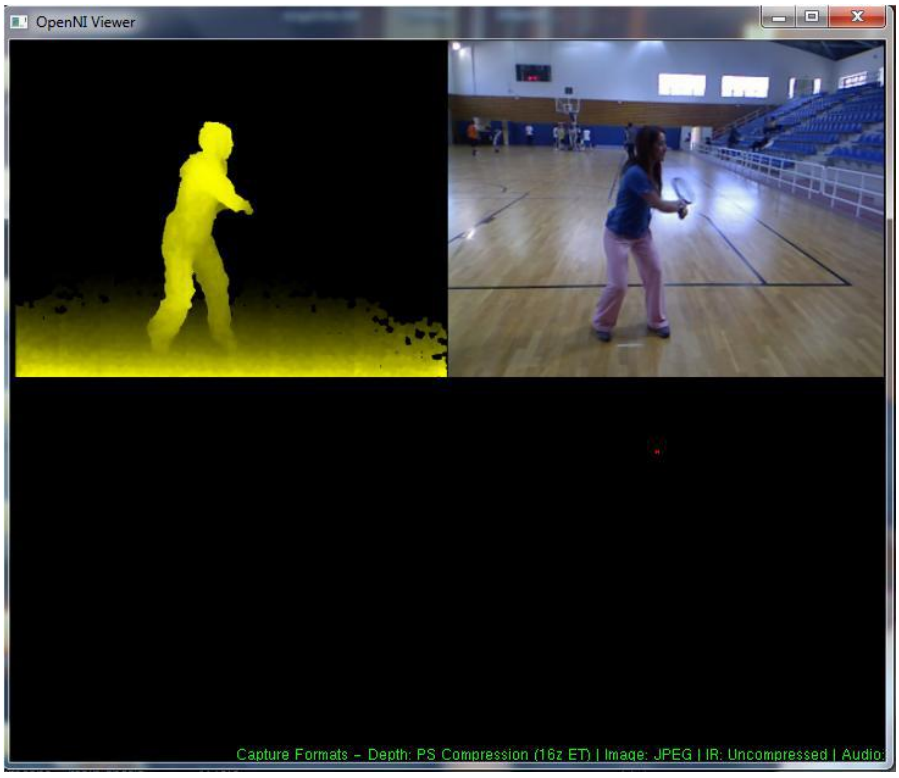
Στα πλαίσια της διπλωματικής εργασίας έγινε χρήση της βάσης δεδομένων THETIS **[21].** Σκοπός της προτεινόμενης βάσης είναι να προσφέρει στην ερευνητική κοινότητα ένα επιπλέον σύνολο δεδομένων από καταγεγραμμένες κινήσεις για την ανάπτυξη εφαρμογών αναγνώρισης της ανθρώπινης δραστηριότητας (όπως gaming, αυτοματοποιημένου σχολιασμού αθλητικών γεγονότων κ.α.). Η βάση αποτελείται από ένα σύνολο 12 βασικών κινήσεων tennis, οι οποίες εκτελούνται από 31 ερασιτέχνες και 24 έμπειρους παίκτες. Κάθε λήψη πραγματοποιείται παραπάνω από μία φορά, δημιουργώντας συνολικά 8734 video AVI μορφής. Η συνολική διάρκεια των videos είναι 7 ώρες και 15 λεπτά. Οι κινήσεις που αναλύονται είναι οι εξής:

Backhand with two hands
Backhand
Backhand slice
Backhand volley
Forehand flat
Forehand open stands
Forehand slice
Forehand volley
Service flat

Service kick
Service slice
Smash

3.2 Συνθήκες Καταγραφής Κινήσεων

Ως συσκευή εγγραφής για το σύνολο δεδομένων της THETIS χρησιμοποιείται το Kinect της Microsoft, σε συνδυασμό με το opensource λογισμικό OpenNI και το middleware του PrimeSence Nite. Τα δεδομένα, τα οποία καταγράφει ο αισθητήρας βάθους του Kinect (depth map), καθώς και τα δεδομένα, τα οποία καταγράφει η RGB κάμερα (image map), συνδυάζονται και αποθηκεύονται σε ένα αρχείο τύπου ONI που είναι συμβατό με το OpenNI.



Εικόνα 3.1: *Απεικόνιση του depth map και του image map από το NiViewer του OpenNI κατά τη διαδικασία καταγραφής.*

Η καταγραφή των κινήσεων tennis γίνεται σε δύο διαφορετικές εσωτερικές τοποθεσίες. Η εγγραφή της κίνησης σε εσωτερικό χώρο είναι απαραίτητη για τη σωστή λειτουργία της συσκευής Kinect. Συγκεκριμένα, δεδομένου ότι το Kinect επεξεργάζεται ένα υπέρυθρο πλέγμα για την κατασκευή πληροφοριών 3D, το αποτέλεσμα επηρεάζεται ιδιαίτερα από την υπέρυθρη ακτινοβολία του αντίστοιχου φάσματος ηλιακού φωτός των εξωτερικών χώρων. Επομένως η διεξαγωγή των καταγραφών σε εσωτερικό χώρο κρίνεται αναγκαία για την αποφυγή παρεμβολών, τις οποίες μπορεί να προκαλεί το ηλιακό φως.

Η συσκευή Kinect τοποθετείται σε ύψος 1.6m από το έδαφος και η κάμερα παραμένει στατική κατά τη διάρκεια κάθε λήψης. Το σημείο εκτέλεσης κάθε κίνησης tennis ορίζεται περίπου 1.5m από την κάμερα λήψης, ενώ η κάθε κίνηση επαναλαμβάνεται αρκετές φορές.

Σημειώνεται επίσης ότι πριν την εκτέλεση των κινήσεων, οι αρχάριοι παίκτες παρακολουθούν μια επίδειξη για κάθε δράση από έναν εκπαιδευτή tennis και ακολουθούν κατάλληλες οδηγίες.

Το παρασκήνιο (background) δεν παραμένει στατικό. Συγκεκριμένα μπορεί να διαφοροποιείται από άτομο σε άτομο, από κίνηση σε κίνηση για το ίδιο άτομο και τέλος, μπορεί να διαφοροποιείται κατά τη διάρκεια καταγραφής μιας κίνησης του ίδιου ατόμου. Ακόμη αναφέρεται ότι σε ένα video μπορεί να εμφανίζονται περισσότερα άτομα εκτός από τον παίκτη, τα οποία ενεργούν με ποικίλους τρόπους πίσω από τη σκηνή δράσης (περπατώντας, παίζοντας μπάσκετ κλπ.).

Τέλος παρατηρούνται διαφοροποιήσεις ως προς την απόσταση που χωρίζει τους παίκτες από τη συσκευή Kinect, καθώς δεν πραγματοποιούνται όλες οι λήψεις στον ίδιο χώρο. Πιθανές διαφορές υπάρχουν και ως προς τη γωνία λήψης της κάμερας, οι οποίες όμως θεωρούνται αμελητέες.

Εικόνα 3.2: *Διαφοροποιήσεις στο background κατά τη διάρκεια διαφορετικών λήψεων*.

3.3 Δομή της Βάσης THETIS

Κάθε παίκτης επαναλαμβάνει κάθε μια από τις παραπάνω 12 κινήσεις του tennis από δύο έως τέσσερις φορές. Αυτό έχει ως αποτέλεσμα τη δημιουργία 660 αρχείων τύπου ONI. Για τη χρήση των δεδομένων από διάφορες εφαρμογές είναι απαραίτητη η μετατροπή των αρχείων ONI σε μια μορφή ευρέως διαδεδομένη. Για τον παραπάνω λόγο τα δεδομένα έχουν μετατραπεί σε αρχεία μορφής AVI χρησιμοποιώντας έναν αλγόριθμο βασισμένο στο OpenNI. Η παραπάνω εφαρμογή προσφέρει τα ακόλουθα χαρακτηριστικά:

- Απομόνωση 3D δεδομένων βάθους που καταγράφονται από τον αισθητήρα βάθους του Kinect.

- Εξαγωγή της σιλουέτας του χρήστη που εκτελεί τη δράση.

- Εξαγωγή του σκελετού του ανθρώπινου σώματος μέσω ανίχνευσης των αρθρώσεων του.

- Αναπαράσταση των σκελετικών δεδομένων σε δύο και τρεις διαστάσεις.

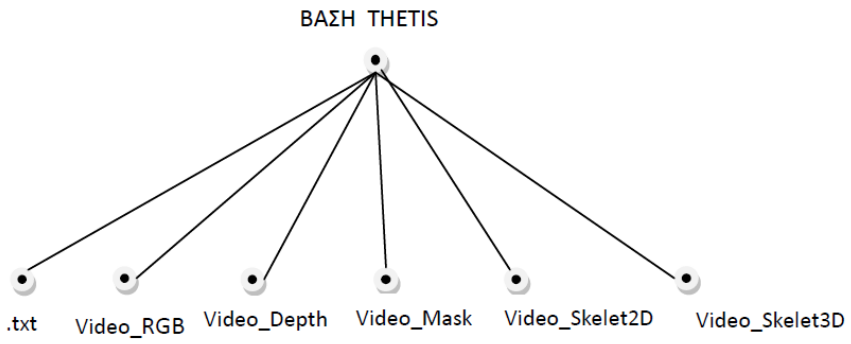
Συγκεκριμένα για κάθε αρχείο ONI δημιουργούνται πέντε AVI αρχεία, ίσης διάρκειας:

Ένα αρχείο AVI που περιέχει τις πληροφορίες RGB.
Ένα αρχείο AVI που περιέχει τις πληροφορίες βάθους.
Ένα αρχείο AVI που περιέχει τη σιλουέτα του ατόμου.
Ένα αρχείο AVI που περιέχει την κίνηση του σκελετού σε 2 διαστάσεις.
Ένα αρχείο AVI που περιέχει την κίνηση του σκελετού σε 3 διαστάσεις.

Αυτό έχει ως αποτέλεσμα τη δημιουργία 3300 αρχείων AVI, τα οποία στη συνέχεια διαχωρίζονται σε επιμέρους τμήματα μεμονωμένων ενεργειών. Στόχος της παραπάνω διαδικασίας κατάτμησης είναι η λήψη από κάθε αρχικό video, τριών νέων που περιέχουν μόνο μία πλήρη επανάληψη κάθε δράσης tennis. Με αυτόν τον τρόπο παράγονται 1980 RGB videos, 1980 videos βάθους και 1980 video-σιλουέτες (mask/silhouette).

Ωστόσο, στα videos των σκελετικών δεδομένων, οι 2D και 3D σκελετοί δεν παρέχονται πάντα για όλες τις επαναλήψεις. Αυτό οφείλεται σε περιορισμούς, οι οποίοι εμφανίζονται κατά την ανάκτηση των σκελετών από τα αρχεία ONI. Πιο συγκεκριμένα, κάθε παίκτης πρέπει να λάβει αρχικά μια καθορισμένη στάση κατά την έναρξη της εγγραφής, η οποία ονομάζεται calibration pose. Αν αποτύχει, η ανάκτηση του σκελετού δεν είναι δυνατή. Τέτοιες περιπτώσεις παρουσιάζονται στη βάση, καθώς ο έλεγχος δεν είναι εφικτός από την αρχή της καταγραφής των κινήσεων. Επιπροσθέτως, ορισμένοι από τους συμμετέχοντες πραγματοποιούν μερικές από τις κινήσεις με μεγάλη ταχύτητα, με αποτέλεσμα η εξαγωγή των σκελετικών δεδομένων να επιτυγχάνεται μόνο στις τελευταίες επαναλήψεις. Για τους παραπάνω λόγους, τα videos με τα σκελετικά δεδομένα είναι συνολικά 1217 για τις δύο διαστάσεις και 1217 για τις τρεις διαστάσεις αντίστοιχα.

Το σύνολο των δεδομένων της THETIS χωρίζεται σε έξι υποφακέλους, όπως φαίνεται στην Εικόνα 3.3.



Εικόνα 3.3: *Δομή της βάσης δεδομένων THETIS*.

Συνοπτικά τα δεδομένα της THETIS οργανώνονται ως εξής:

- **Txt**: περιλαμβάνει λεπτομερή περιγραφή των περιεχομένων της βάσης δεδομένων.

- **Video RGB**: περιέχει 1980 αρχεία AVI, σε 12 υποφακέλους (ανά κλάση). Σε κάθε φάκελο, υπάρχουν 3 επαναλήψεις από κάθε άτομο για την κίνηση αυτή.

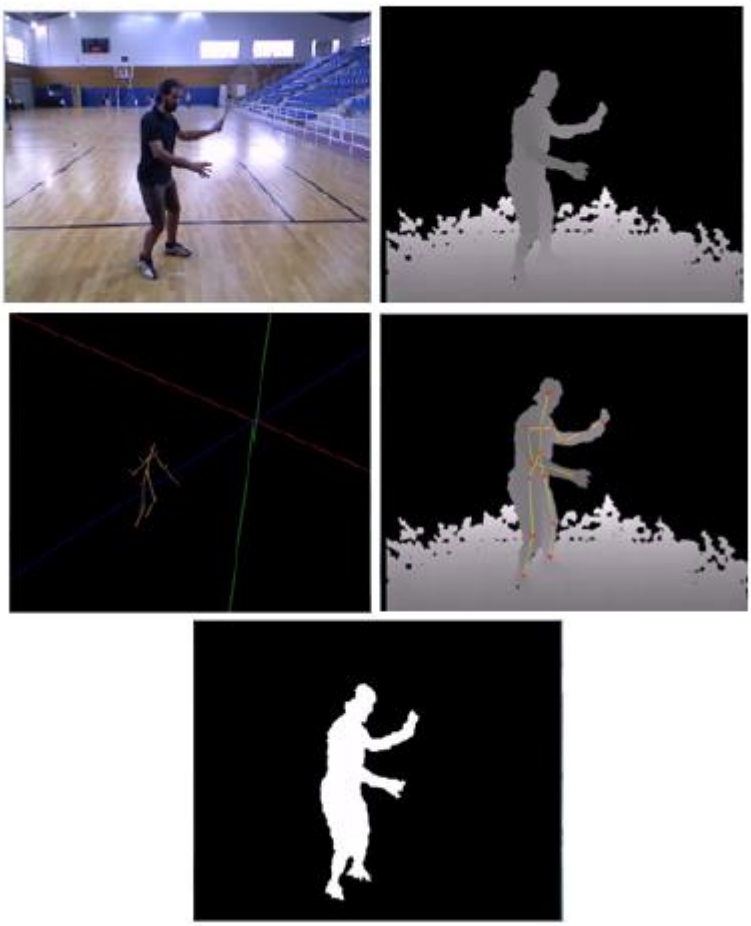
- **Video Depth**: περιέχει 1980 αρχεία AVI, σε 12 υποφακέλους (ανά κλάση). Σε κάθε φάκελο, υπάρχουν 3 επαναλήψεις από κάθε άτομο για την κίνηση αυτή.

- **Video Mask**: περιέχει 1980 αρχεία AVI, σε 12 υποφακέλους (ανά κλάση). Σε κάθε φάκελο, υπάρχουν 3 επαναλήψεις από κάθε άτομο για την κίνηση αυτή.

- **Video Skelet2D**: περιέχει 1217 αρχεία AVI, σε 12 υποφακέλους (ανά κλάση).

- **Video Skelet3D**: περιέχει 1217 αρχεία AVI, σε 12 υποφακέλους (ανά κλάση).

Στην εικόνα 3.4 παρουσιάζονται τα στιγμιότυπα όλων των τύπων video ενός παίκτη, ο οποίος πραγματοποιεί την κίνηση forehand slice.



Εικόνα 3.4: *Στιγμιότυπα όλων των τύπων video ενός παίκτη που εκτελεί την κίνηση forehand slice*.

3.4 Πειραματικά Αποτελέσματα

3.4.1 Παρουσίαση Μεθόδου Αξιολόγησης

Για την αξιολόγηση της THETIS πραγματοποιείται ένα σύνολο πειραμάτων, στα οποία εφαρμόζονται δύο μέθοδοι αναγνώρισης δραστηριότητας. Ως δεδομένα εισόδου χρησιμοποιούνται τα videos με τους σκελετούς στις τρεις διαστάσεις Skelet3D Video και τα videos βάθους Depth Video. Επίσης αναφέρεται ότι εφαρμόζεται ο ίδιος τύπος πειραμάτων στο σύνολο δεδομένων KTH, με σκοπό να μελετηθούν οι προκλήσεις της αναγνώρισης κίνησης στην προτεινόμενη βάση δεδομένων.

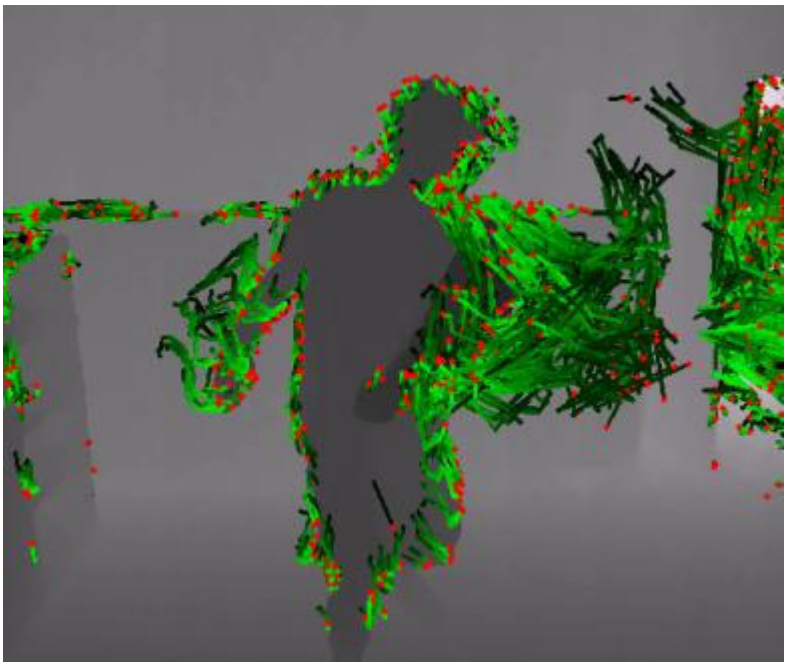
Επισημαίνεται ότι στο σύνολο δεδομένων KTH περιλαμβάνονται 6 τύποι διαφορετικών ενεργειών: walking, jogging, running, boxing, hand waving και hand clapping. Αυτές οι ενέργειες εκτελούνται αρκετές φορές, από 25 διαφορετικά άτομα σε τέσσερα διαφορετικά σενάρια, υπό διαφορετικές συνθήκες φωτισμού. Η καταγραφή των ενεργειών γίνεται μέσω στατικής κάμερας ρυθμού λήψης 25 καρέ ανά δευτερόλεπτο.

Ένας αποτελεσματικός τρόπος για την ανίχνευση των σημαντικών σημείων σε μια εικόνα και κατ' επέκταση σε ένα video είναι ο εντοπισμός των σημείων ενδιαφέροντος (spatio-temporal points). Διάφορες τεχνικές ανίχνευσης τέτοιων σημείων έχουν προταθεί **[22]**.

Στην περιγραφείσα πειραματική διαδικασία, για την ανίχνευση των χωροχρονικών σημείων ενδιαφέροντος ακολουθείται η μέθοδος Space-Time Interest Points (STIP), η οποία παρουσιάζεται στο **[23]**. Η συγκεκριμένη μέθοδος αποτελεί μια επέκταση του προτεινόμενου ανιχνευτή Harris 3D, η οποία εντοπίζει τα χωροχρονικά σημεία ενδιαφέροντος και υπολογίζει τους αντίστοιχους περιγραφείς (descriptors), τα ιστογράμματα προσανατολισμένης κλίσης/Histograms Oriented Gradient (HOG) και τα ιστογράμματα οπτικής ροής/ Histograms of Optical Flow (HOF).

Η δεύτερη μέθοδος αναγνώρισης ανθρώπινης δράσης, η οποία εφαρμόστηκε στα δεδομένα της βάσης THETIS, ονομάζεται τεχνική Dense Trajectories ή πυκνές τροχιές κίνησης και περιγράφεται αναλυτικά στο **[24]**.Η συγκεκριμένη μέθοδος βασίζεται στην πυκνή δειγματοληψία σημείων που εντοπίζονται σε κάθε πλαίσιο και παρακολουθεί τη μετατόπισή τους σύμφωνα με πληροφορίες, οι οποίες διεξάγονται από τα πεδία οπτικής ροής. Ο αριθμός των σημείων που παρακολουθούνται μπορεί να πολλαπλασιαστεί εύκολα, εφόσον υπολογιστούν τα πεδία οπτικής ροής χωρίς κόστος. Οι τροχιές των παραπάνω σημείων περιγράφουν τελικά την κίνηση μέσα στο video. Ένα παράδειγμα τέτοιων τροχιών, οι οποίες εξάγονται από ένα video της βάσης δεδομένων THETIS, απεικονίζεται στο σχήμα 3.5.

Επιπλέον για την αντιμετώπιση προβλημάτων, τα οποία οφείλονται στην κίνηση της κάμερας, στο **[24]** οι συγγραφείς εισάγουν ένα νέο τοπικό περιγραφέα που επικεντρώνεται στην κίνηση του προσκηνίου. Ο περιγραφέας αυτός αποτελεί επέκταση του τρόπου κωδικοποίησης της κίνησης με Ιστογράμματα Ορίων Κίνησης (Motion Boundary Histograms)**[25]**, τα οποία χρησιμοποιούνται ως ένα μέσο αξιολόγησης της THETIS.



Εικόνα 3.5: *Εφαρμογή της μεθόδου Dense Trajectories σε video βάθους της βάσης THETIS.*

Οι τεχνικές εξαγωγής χαρακτηριστικών κίνησης από εικονοσειρές είναι διεργασίες με υψηλό υπολογιστικό κόστος, οι οποίες παράγουν δεδομένα μεγάλου όγκου. Τα ακατέργαστα εξαγόμενα δεδομένα δε μπορούν να χρησιμοποιηθούν αποδοτικά σε αλγορίθμους ταξινόμησης (classification) λόγω της υπολογιστικής πολυπλοκότητας που επιβάλλουν. Επομένως κρίνεται απαραίτητη η εφαρμογή κάποιας τεχνικής "bag-of-words". Συγκεκριμένα δημιουργείται ένα οπτικό λεξικό με χρήση αλγορίθμων ομαδοποίησης δεδομένων (clustering) για την ποσοτικοποίηση των περιγραφικών δεικτών. Η έννοια της ομαδοποίησης αφορά την κατάταξη των δεδομένων σε επιμέρους σύνολα, με τέτοιο τρόπο, ώστε τα δεδομένα της ίδιας ομάδας να παρουσιάζουν μεγάλη ομοιότητα μεταξύ τους και διακριτή ετερότητα με αυτά των άλλων ομάδων. Το οπτικό λεξικό δημιουργείται χρησιμοποιώντας τον αλγόριθμο k-means**[26]**, με τον οποίο κατασκευάζονται k διαμερίσεις του δοθέντος συνόλου, όπου κάθε διαμέριση αναπαριστά μια ομάδα που περιέχει τουλάχιστον ένα αντικείμενο. Στη συγκεκριμένη περίπτωση κάθε οπτικό λεξικό δημιουργείται για k = 500 διαμερίσεις.

3.4.2 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines – SVMs)

Για την ταξινόμηση χρησιμοποιούνται 12 μη γραμμικά SVMs (όσες είναι και οι κατηγορίες κινήσεων tennis).

Οι μηχανές διανυσμάτων υποστήριξης (support vector machines - SVMs) αποτελούν συστήματα εκμάθησης με χώρο υποθέσεων που περιλαμβάνει γραμμικές συναρτήσεις. Ως είσοδο δέχονται δεδομένα από χώρους χαρακτηριστικών πολλών διαστάσεων και εκπαιδεύονται με βάση κάποιον αλγόριθμο εκμάθησης. Η θεωρία ανάπτυξης των SVMs αναλύεται στην έρευνα των **[27].**

Βασική λειτουργία των SVMs είναι η εύρεση ενός υπολογιστικά αποδοτικού τρόπου εκμάθησης υπερεπιπέδων (hyperlines) για τον διαχωρισμό δεδομένων σε χώρους χαρακτηριστικών πολλών διαστάσεων. Επομένως η ταξινόμηση ενός συνόλου δεδομένων σε κάποια κατηγορία ανάγεται στην εύρεση υπερεπιπέδων διαχωρισμού, τα οποία βελτιστοποιούν τα όρια γενίκευσης (generalization bounds). Η σωστή λειτουργία ενός

συστήματος SVM περιλαμβάνει τα στάδια εκπαίδευσης (training) και έλεγχου (testing) των δεδομένων.

Τα δεδομένα προς ταξινόμηση μπορεί να ανήκουν σε δύο ή περισσότερες κλάσεις διαχωρισμού. Για τον παραπάνω διαχωρισμό έχουν αναπτυχθεί διάφορες μέθοδοι, οι οποίες μπορεί να συνδυάζουν περισσότερα από ένα SVM για την ταξινόμηση των δεδομένων σε πολλαπλές κλάσεις. Οι πιο διαδεδομένες από τις παραπάνω μεθόδους είναι οι εξής:

- **One-against-all**

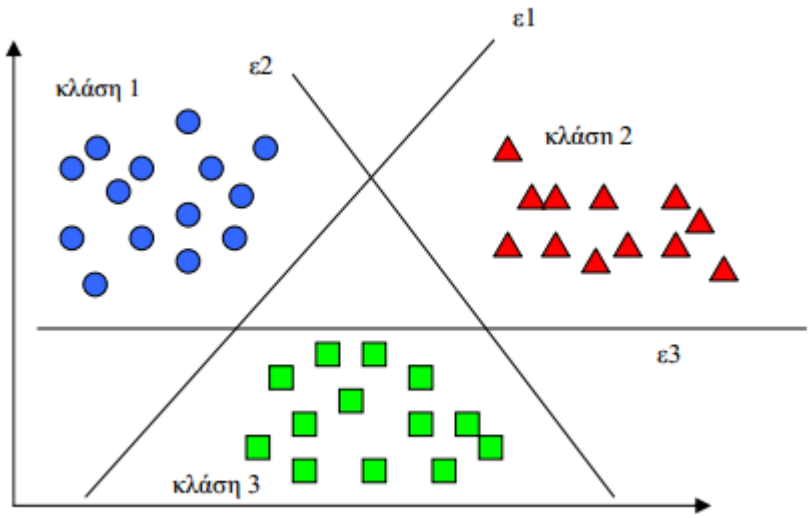
Σε αυτή τη μέθοδο επιδιώκεται ο υπολογισμός των ορίων μιας κλάσης για τον διαχωρισμό των δεδομένων της από τα δεδομένα όλων των υπόλοιπων κλάσεων. Επομένως απαιτούνται k συναρτήσεις απόφασης για την εύρεση k διαχωριστικών υπερεπιπέδων των κλάσεων. Τα όρια της κάθε κλάσης προκύπτουν με συνδυασμό των αποτελεσμάτων των k SVMs. Η ταξινόμηση των δεδομένων γίνεται με τον υπολογισμό των τιμών των k συναρτήσεων απόφασης και την κατάταξη κάθε σημείου στην κλάση, η οποία αντιστοιχεί στη συνάρτηση απόφασης με τη μεγαλύτερη τιμή. Η διαδικασία ταξινόμησης παρουσιάζεται στο σχήμα 3.6, στο οποίο απεικονίζεται η κατάταξη των δεδομένων σε τρεις κλάσεις.

- **All-together**

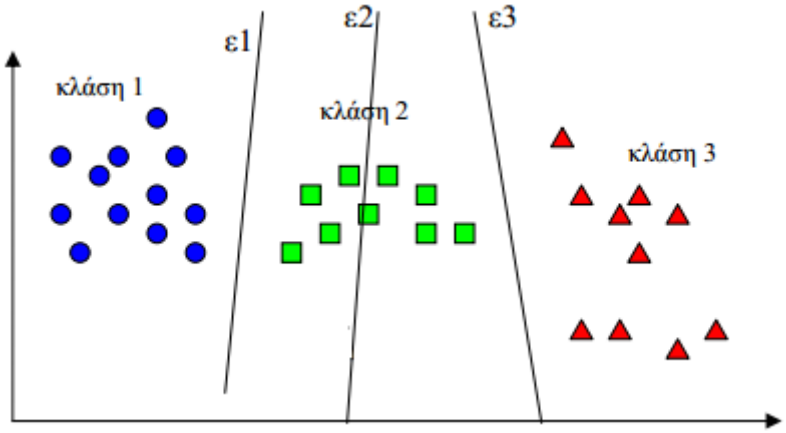
Η συγκεκριμένη προσέγγιση είναι όμοια με τη μέθοδο one-against-all. Αρχικά δημιουργούνται k συναρτήσεις απόφασης για τις k κλάσεις, όπου η i-οστή συνάρτηση απόφασης διαχωρίζει την i-οστή κλάση δεδομένων από τις υπόλοιπες. Σε αντίθεση με την τεχνική one-against-all, στην οποία δημιουργούνται k SVMs για να βρεθούν τα όρια της κάθε κλάσης, σε αυτή την περίπτωση αρκεί η δημιουργία ενός συστήματος. Η ταξινόμηση των σημείων γίνεται με τον τρόπο της προηγούμενης προσέγγισης.

- **One-against-one**

Στην One-against-one τεχνική λαμβάνονται υπ' όψη όλα τα πιθανά ζεύγη των κλάσεων και λύνεται το πρόβλημα του μεταξύ τους διαχωρισμού. Επομένως για k κλάσεις δημιουργούνται k!/[2(k-2)!] δυνατοί συνδυασμοί ζευγών κλάσεων και k!/[2(k-2)!] συναρτήσεις απόφασης. Για κάθε πιθανό ζεύγος κλάσεων, εφαρμόζεται κάποια από τις προηγούμενες μεθοδολογίες για την εύρεση του διαχωριστικού τους υπερεπιπέδου. Στη εικόνα 3.7 παρουσιάζεται η ταξινόμηση δεδομένων τριών κλάσεων μαζί με τις διαχωριστικές τους ευθείες.



Εικόνα 3.6: *Διαχωρισμός τριών κλάσεων με τη μέθοδο one-against-all.*



Εικόνα 3.7: *Διαχωρισμός τριών κλάσεων με τη μέθοδο one-against-one.*

Στην προτεινόμενη βάση για την εκπαίδευση και επαλήθευση των SVMs χρησιμοποιείται το πρωτόκολλο cross-one-person-out ή leave-one-person-out. Η παραπάνω τεχνική διατηρεί τα δείγματα (videos) ενός συγκεκριμένου ατόμου ως σύνολο επαλήθευσης (testing set), ενώ τα υπόλοιπα δείγματα χρησιμοποιούνται για την εκπαίδευση (training set). Η διαδικασία επαναλαμβάνεται Ν φορές, όπου Ν είναι ο αριθμός των υποκειμένων μέσα στο σύνολο δεδομένων. Επομένως, για το σύνολο KTH που περιέχει 25 άτομα, η διαδικασία επαναλαμβάνεται 25 φορές, ενώ για το σύνολο των 3D σκελετών και των δεδομένων βάθους της THETIS, η διαδικασία επαναλαμβάνεται 55 φορές. Το τελικό αποτέλεσμα σε κάθε πείραμα προκύπτει από τον συνδυασμό των Ν επιμέρους αποτελεσμάτων.

Για την απόδοση των μεθόδων χρησιμοποιείται ο μέσος όρος της ορθότητας ταξινόμησης των Ν επαναλήψεων (average accuracy) :

$$average\ accuracy = \frac{correctly\ classified\ videos}{set\ of\ videos} \qquad (3.1)$$

Τα αποτελέσματα που παράγονται για τα διαφορετικά είδη δεδομένων των προαναφερόμενων μεθόδων αναγνώρισης παρουσιάζονται στους Πίνακες 3.1 και 3.2.

Dataset	Average Accuracy (%)
THETIS DEPTH	60.23
THETIS Skelet3D	54.40
KTH	92.99

Πίνακας 3.1: *Εφαρμογή τεχνικής STIP με χρήση περιγραφέων HOG και HOF σε 3 διαφορετικά σύνολα δεδομένων.*

Dataset	Average Accuracy (%)		
Descriptor	**Trajectory**	**MBH**	**Combination**
THETIS DEPTH	51.59	54.32	57.50
THETIS Skelet3D	46.84	50.78	53.08
KTH	86.98	92.32	90.65

Πίνακας 3.2: *Αποτελέσματα εφαρμογής Dense Trajectory με χρήση συνδυασμού περιγραφέων: a) Trajectory, b) MBH, c) συνδυασμός Trajectory, HOG, HOF και MBH αντίστοιχα.*

3.5 Προβλήματα Εξαγωγής Σκελετικών Δεδομένων από τη THETIS

Στα πλαίσια της διπλωματικής εργασίας, γίνεται νέα εξαγωγή και επεξεργασία των σκελετικών δεδομένων της βάσης THETIS. Ωστόσο κατά την παραπάνω διαδικασία εξαγωγής αντιμετωπίζονται ορισμένες δυσκολίες:

- Σε μερικά videos, η διαδικασία εντοπισμού των αρθρώσεων του σκελετού δεν είναι εφικτή. Αυτό μπορεί να οφείλεται σε πιθανές παρεμβολές, οι οποίες προκύπτουν κατά την εγγραφή των κινήσεων. Επίσης σε ορισμένες περιπτώσεις, οι παίκτες δε στέκονται στην κατάλληλη απόσταση, με αποτέλεσμα να μην εντοπίζονται καθόλου οι ίδιοι ή να εντοπίζονται μόνο συγκεκριμένα σημεία του σώματός τους.

- Προβλήματα εντοπισμού οφείλονται επίσης στον τρόπο με τον οποίο οι παίκτες εκτελούν τις κινήσεις. Ορισμένοι παίκτες, όπως αναφέρθηκε, εκτελούν τις κινήσεις με μεγάλη ταχύτητα. Αυτό δυσχεραίνει τη διαδικασία εύρεσης των αρθρώσεων, ενώ σε αρκετές περιπτώσεις μπορεί να μη γίνεται εντοπισμός του σκελετού σε διαδοχικά πλάνα και η κίνηση να μην παρουσιάζεται ομαλή.

- Τέλος σε ορισμένες εκτελέσεις, άτομα του περιβάλλοντα χώρου επηρεάζουν τη διαδικασία εγγραφής, καθώς αποσπούν την προσοχή του παίκτη ή άθελα τους μετατοπίζουν τη συσκευή λήψης κατά τη διάρκεια της πραγματοποίησης της κίνησης.

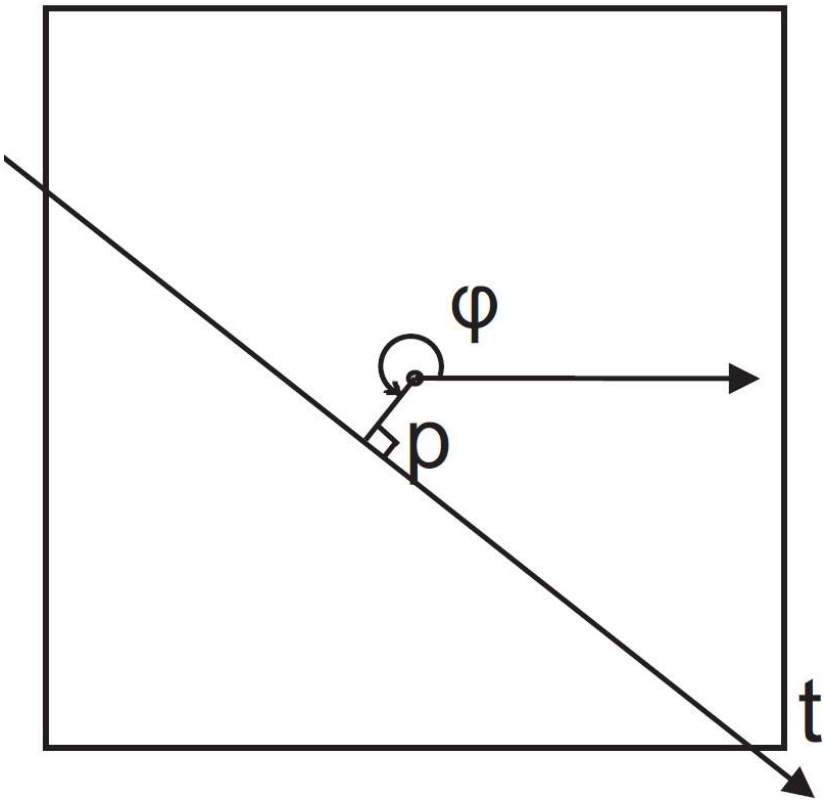
3.6 Αναγνώριση Ανθρώπινης Δραστηριότητας με χρήση του 3D Cylindrical Trace Transform (3D CTT)

3.6.1 Εισαγωγή

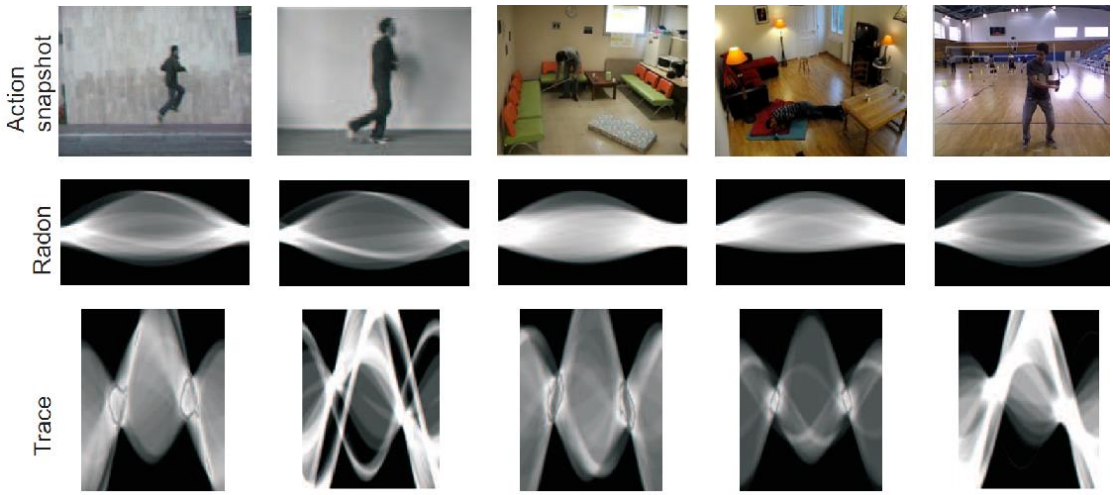
Διάφορες τεχνικές αναγνώρισης, οι οποίες αξιοποιούν 3D δεδομένα, έχουν αναπτυχθεί. Σε αυτό το σημείο γίνεται αναφορά σε μια μέθοδο αναγνώρισης, η οποία στηρίζεται στον μετασχηματισμό Trace και αποτελεί επέκταση της μελέτης των **[28]**. Η μέθοδος βασίζεται στην αξιοποίηση ενός νέου μετασχηματισμού, ο οποίος ονομάζεται 3D Cylindrical Trace Transform.

3.6.2 Συσχετίσεις του 3D Cylindrical Trace Transform

Επισημαίνεται ότι ο μετασχηματισμός Trace μπορεί να θεωρηθεί γενίκευση του μετασχηματισμού Radon **[29]**, καθώς ο μετασχηματισμός Radon αποτελεί μια ειδική περίπτωση του Trace. Συγκεκριμένα ο μετασχηματισμός Radon μιας εικόνας αναφέρεται στην 2D αναπαράσταση της σε συντεταγμένες ϕ και p (με την τιμή του ολοκληρώματος της εικόνας υπολογιζόμενο κατά μήκος της αντίστοιχης γραμμής στα [ϕ, p]). Αντίστοιχα ο μετασχηματισμός Trace υπολογίζει ένα συναρτησιακό T κατά μήκος της γραμμής ανίχνευσης, το οποίο μπορεί να μην είναι απαραίτητα ολοκληρώσιμο. Ο τελικός μετασχηματισμός δημιουργείται με τον εντοπισμό μιας εικόνας με χρήση ευθειών γραμμών και τον υπολογισμό διαφόρων συναρτησιακών χαρακτηριστικών της εικόνας κατά μήκος αυτών των γραμμών. Με αυτόν τον τρόπο πλήθος μετασχηματισμών με διαφορετικές ιδιότητες μπορούν να εξαχθούν από την ίδια εικόνα. Ο παραγόμενος μετασχηματισμός αποτελεί μια δισδιάστατη συνάρτηση των παραμέτρων (ϕ, p) των γραμμών ανίχνευσης. Ο ορισμός αυτών των παραμέτρων δίνεται στην εικόνα 3.8 ενώ παραδείγματα μετασχηματισμών Radon και Trace παρουσιάζονται στην εικόνα 3.9. Μια λεπτομερής ανάλυση της θεμελιώδους θεωρίας του μετασχηματισμού Trace περιγράφεται στα **[28]** και **[30]**.



Εικόνα 3.8: *Ορισμός παραμέτρων μιας γραμμής ανίχνευσης μιας εικόνας.*



Εικόνα 3.9: *Παραδείγματα των μετασχηματισμών Radon και Trace που δημιουργούνται από τις σιλουέτες διαφορετικών δραστηριοτήτων σε διάφορα σύνολα δεδομένων.*

3.6.3 Γενικευμένος Μετασχηματισμός 3D Radon

Ο νέος προτεινόμενος μετασχηματισμός εμπνέεται από τον μετασχηματισμό 3D Radon **[31]**. Ο τελευταίος καθορίζεται χρησιμοποιώντας μονοδιάστατες προβολές ενός τρισδιάστατου αντικειμένου f(x,y,z), οι οποίες λαμβάνονται από την ολοκλήρωση του f(x,y,z) σε ένα επίπεδο. Ο προσανατολισμός του παραπάνω επιπέδου περιγράφεται από ένα

μοναδιαίο διάνυσμα $\bar{a}$. Γεωμετρικά, ο συνεχής μετασχηματισμός 3D Radon περιγράφει μια αντιστοίχιση μεταξύ μιας συνάρτησης και των ολοκληρωμάτων της σε ένα επίπεδο στον R^3. Για την ευκολότερη κατανόηση του προτεινόμενου μετασχηματισμού, παρέχεται μια σύντομη επισκόπηση του μετασχηματισμού 3D Radon, όπως διατυπώνεται στο **[32].**

Έστω M ένα τρισδιάστατο μοντέλο και f(x) μια συνάρτηση όγκου του M, η οποία ορίζεται ως:

$$f(x) = \begin{cases} 1 & \text{when x lies within the 3D model's volume} \\ 0 & otherwise \end{cases} \qquad (3.2)$$

Ορίζεται επίσης η το μοναδιαίο διάνυσμα του R^3. Ο διακριτός 3D Radon μετασχηματισμός του μοντέλου της f(x) δίνεται από τον τύπο:

$$T_f(\eta, \rho) = \sum_{n=1}^{N} f(x_n)\delta(x_n^T \eta - \rho) \qquad (3.3)$$

όπου η είναι το μοναδιαίο διάνυσμα του R^3, ρ είναι ένας πραγματικός αριθμός και δ(.) η συνάρτηση Dirac. Το διάνυσμα η εκφράζεται με σφαιρικές συντεταγμένες ως η = [cosϕ sinθ, sinϕ sinθ, cosθ]. Επομένως ο τύπος (3.3) μετασχηματίζεται ως εξής:

$$T_f(\rho, \theta, \varphi) = \sum_{n=1}^{N} f(x_n, y_n, t_n)\delta(x_n \cos\varphi \sin\theta + y_n \sin\varphi \sin\theta + t_n \cos\theta - \rho) \qquad (3.4)$$

Ο παραπάνω τύπος υπολογίζεται εύκολα, δεν είναι όμως ανεξάρτητος της περιστροφής και της αλλαγής κλίμακας.

3.6.4 Μετασχηματισμός 3D CTT

Ο προτεινόμενος 3D Cylindrical Trace Transform αποτελεί επέκταση του μετασχηματισμού Trace στον τρισδιάστατο χώρο. Ο CTT_f μετασχηματισμός μιας συνάρτησης f(x) ενός 3D μοντέλου M, συσχετίζει ένα συναρτησιακό T με κάθε ανιχνεύσιμη γραμμή της περιοχής (p, φ, θ), όπου τα p και φ προσδιορίζουν μοναδικά κάθε γραμμή και θ είναι η γωνία του επιπέδου με το αρχικό (σχήμα 3.10). Οι μετασχηματισμοί Trace υπολογίζονται συνεχώς σε επίπεδα, τα οποία περιστρέφονται προς την κατεύθυνση ενός πολικού άξονα A, «κόβοντας» το 3D πλέγμα M. Ο εικονικός κύλινδρος που δημιουργείται από τη συνεχή περιστροφή των επιπέδων κατά την πολική κατεύθυνση έχει ακτίνα ρ και μήκος l, με αρχή O:(0,0,0). Η ακτίνα ρ ορίζεται ως η απόσταση του πιο απομακρυσμένου σημείου $x(\rho_{max})$ του M από τον άξονα L του κυλίνδρου. Το μήκος l ορίζεται ως η απόσταση, η οποία είναι παράλληλη προς τον άξονα L του κυλίνδρου και ενώνει τα δύο πιο απομακρυσμένα σημεία του τρισδιάστατου μοντέλου M. Κάθε μετασχηματισμός Trace $\breve{g}$ υπολογίζεται σε σχέση με το κέντρο του 3D μοντέλου, το οποίο συμπίπτει με το σημείο K του άξονα L του εικονικού κυλίνδρου και έχει συντεταγμένες (0, 0, l/2). Επομένως, μια περιστροφή 180° οδηγεί στον υπολογισμό ενός μετασχηματισμού $\breve{g}$, ο οποίος είναι ίσος με τον μετασχηματισμό που προκύπτει για επίπεδο με θ = 0.

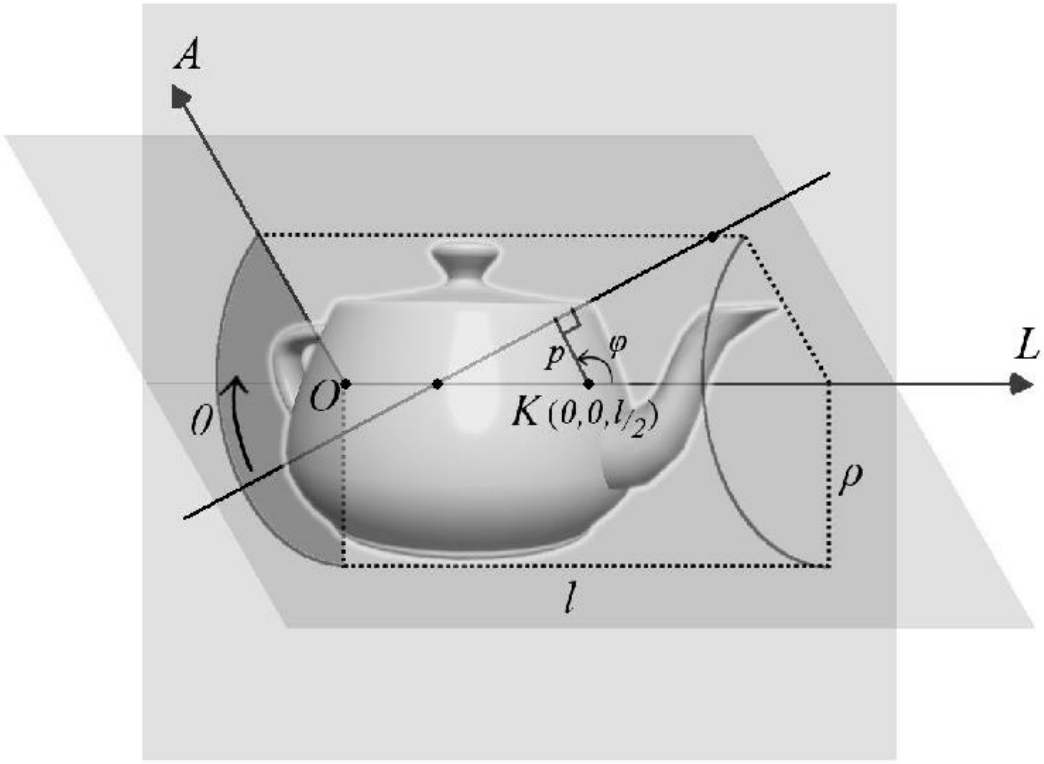
Θεωρώντας κάθε επίπεδο ως μια 2D συνάρτηση ξ(x,y), η οποία σχηματίζεται από την προβολή του πλέγματος σε αυτό το επίπεδο, ο αντίστοιχος μετασχηματισμός Trace $\breve{g}$(p,φ) μπορεί να δοθεί από τον υπολογισμό ενός συναρτησιακού T κατά μήκος όλων των γραμμών ανίχνευσης (p,φ) της ξ:

$$\breve{g}(p, \varphi) = T(\xi(x, y)\delta(p - x\cos\varphi - y\sin\varphi)) \qquad (3.5)$$

Η τελική αναπαράσταση του 3D μοντέλου είναι ο προτεινόμενος 3D CTT. Ο παραπάνω μετασχηματισμός αποτελεί επίσης μια συνάρτηση 2D των παραμέτρων (p,φ) και προκύπτει από το άθροισμα των μεμονωμένων μετασχηματισμών Trace των επιπέδων, τα οποία περιστρέφονται κατά θ σε σχέση με το αρχικό (στην κατεύθυνση των γωνιακών συντεταγμένων που ορίζονται από το θ):

$$CTT_f(p,\varphi) = \sum_{n=1}^{N} \breve{g}_n(p,\varphi) \qquad (3.6)$$

όπου $\breve{g}_n$ είναι ο n-οστός μετασχηματισμός Trace, δηλαδή ο μετασχηματισμός που υπολογίζεται για την 2D επίπεδη προβολή του M πάνω στο επίπεδο που σχηματίζει γωνία θ_n με το αρχικό. Επισημαίνεται ότι $N \geq 2$, $0 < \theta_n \leq \theta_{max}$ και $\theta_{max} = 180°$. Μια απεικόνιση του μετασχηματισμού 3D CTT δίνεται στο σχήμα 3.10. Για να γίνει ο μετασχηματισμός ανεξάρτητος της κλίμακας, υπολογίζεται η μέγιστη απόσταση d_{max} μεταξύ του κέντρου μάζας και του πιο απομακρυσμένου voxel του 3D όγκου. Στη συνέχεια η f(x) μετασχηματίζεται έτσι ώστε $d_{max} = 1$. Το κέντρο μάζας συμπίπτει με το σημείο K.



Εικόνα 3.10: *Απεικόνιση του 3D CTT μετασχηματισμού*.

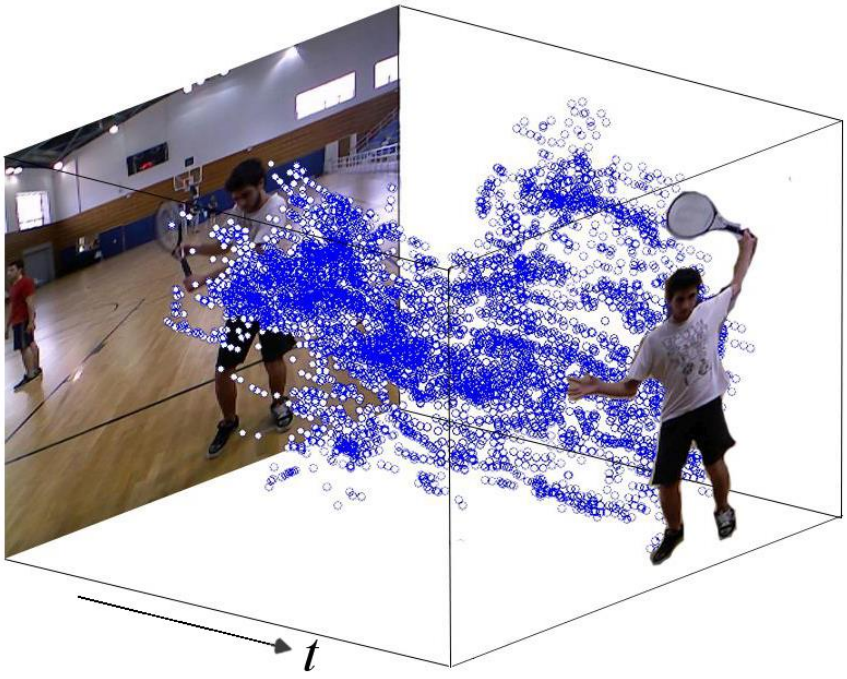
3.6.5 Σύστημα Αναγνώρισης Κίνησης με Χρήση του Μετασχηματισμού 3D CTT

Η προτεινόμενη μέθοδος αναγνώρισης ανθρώπινης δραστηριότητας συνδυάζει την τεχνική 3D CTT με τον αλγόριθμο εύρεσης χωροχρονικών σημείων ενδιαφέροντος Selective Spatio-Temporal Interest Points (SSTIPs) **[33]**. Αρχικά ένας τρισδιάστατος χωροχρονικός όγκος σχηματίζεται από το πλέγμα των SSTIPs και έπειτα υπολογίζονται διάφοροι 3D CTT

μετασχηματισμοί από αυτόν. Τα αποτελέσματα των μετασχηματισμών συνδυάζονται, δημιουργώντας διανύσματα χαρακτηριστικών. Επισημαίνεται ότι τεχνικές όπως STIPs ή Bag-of-Visual-Words (BOVW), ενώ είναι διαδεδομένες, αγνοούν πιθανές πληροφορίες της ολικής κατανομής των σημείων ενδιαφέροντος στο χωροχρόνο. Η προτεινόμενη τεχνική αξιοποιεί την ολική γεωμετρική κατανομή των σημείων ενδιαφέροντος, παράγοντας πληροφορίες για τις συνθήκες λήψης και περιβάλλοντος. Περισσότερες λεπτομέρειες παρουσιάζονται στη συνέχεια.

- **Selective Spatio-Temporal Interest Points (SSTIPs)**

Η τεχνική SSTIP επικεντρώνεται στη μελέτη των ολικών (αντί τοπικών) μεταβολών κίνησης των χωροχρονικών σημείων ενδιαφέροντος, αποτρέποντας με αυτόν τον τρόπο την εσφαλμένη ανίχνευση σημείων ενδιαφέροντος, τα οποία προκαλούνται λόγω της κίνησης της κάμερας ή του πολύπλοκου περιβάλλοντα χώρου. Η SSTIP μέθοδος λειτουργεί αποτελεσματικά για την παραγωγή σταθερών, επαναλαμβανόμενων σημείων ενδιαφέροντος, ενώ διακρίνεται σε τρεις επιμέρους διαδικασίες: α) ανιχνεύει τα χωρικά σημεία ενδιαφέροντος, β) εξαλείφει ανεπιθύμητα σημεία του περιβάλλοντα χώρου και γ) επιβάλλει τοπικούς και χρονικούς περιορισμούς στο αποτέλεσμα. Η πρώτη διαδικασία επιτυγχάνεται με χρήση του ανιχνευτή Harris. Η δεύτερη διαδικασία στηρίζεται στην ιδέα της παρατήρησης ότι τα γωνιακά σημεία που ανιχνεύονται στο παρασκήνιο ακολουθούν ένα συγκεκριμένο γεωμετρικό μοτίβο, ενώ αυτά που αντιστοιχούν στους ανθρώπους δε φέρουν τη συγκεκριμένη ιδιότητα. Τέλος χωροχρονικοί περιορισμοί επιβάλλονται, καθώς για να θεωρηθεί ένα σημείο ενδιαφέροντος ως επιθυμητό STIP, πρέπει να παρουσιάζει καθορισμένη αλλαγή θέσης κατά την ακολουθία κίνησης. Ένα παράδειγμα των εξαγόμενων SSTIPs από ένα δείγμα του συνόλου δεδομένων THETIS δίνεται στο σχήμα 3.11.



Εικόνα 3.11: *Εξαγόμενα σημεία SSTIPs μιας κίνησης backhand από video της βάσης δεδομένων THETIS. Το t υποδηλώνει την κατεύθυνση του χρόνου.*

- **Εφαρμογή Μετασχηματισμού 3D CTT στα Selective Spatio-temporal Interest Points**

Σύμφωνα με το **[30]**, μέσω της διαμόρφωσης του μετασχηματισμού Trace, εκτός από τα ολοκληρώματα (μετασχηματισμός Radon) υπάρχουν λειτουργικότητες (όπως η διάμεσος ή ο μέσος όρος), οι οποίες έχουν την ικανότητα να αποκαταστήσουν πλήρως μια δισδιάστατη συνάρτηση κατά μήκος των ευθειών ανίχνευσης του πεδίου της. Όπως προαναφέρθηκε, διαφορετικές λειτουργικότητες/συναρτησιακά οδηγούν στον υπολογισμό διαφορετικών μετασχηματισμών Trace μιας συνάρτησης. Κάθε τέτοιος μετασχηματισμός αποτελεί μια δισδιάστατη συνάρτηση παραμέτρων (p, φ) της αντίστοιχης γραμμής ανίχνευσης. Από τα παραπάνω προκύπτει ότι ο 3D CTT μετασχηματισμός παράγεται με την ανίχνευση διαδοχικών επιπέδων, περιστρεφόμενων κατά γωνία θ, που ανήκουν στο ίδιο ελάχιστο παράθυρο (κύλινδρο) ενός 3D μοντέλου και έχουν την ίδια αρχή Κ. Ο 3D CTT μετασχηματισμός αποτελεί το άθροισμα των επιμέρους μετασχηματισμών επίπεδων και οδηγεί στη δημιουργία μιας νέας 2D συνάρτησης παραμέτρων (p, φ). Διαφορετικοί 3D CTT μετασχηματισμοί προκύπτουν από τον υπολογισμό διαφορετικών συναρτησιακών. Ο κύριος στόχος είναι η λήψη της δυναμικής πληροφορίας και της δομής μιας κίνησης στο μεγαλύτερο δυνατό βαθμό. Ο μετασχηματισμός Trace είναι κατάλληλος για την παραπάνω απαίτηση, ωστόσο ο 3D CTT μετασχηματισμός επεκτείνει τις δυνατότητές του στις τρεις διαστάσεις (χωροχρόνος).

Έστω Μ ένα 3D μοντέλο, το οποίο σχηματίζεται από το πλέγμα SSTIP που δημιουργείται από ένα video ανθρώπινης δράσης. Ορίζονται ρ και l η ακτίνα και το μήκος αντίστοιχα του μικρότερου κυλίνδρου που οριοθετεί το πλέγμα και z ορίζεται ένα τυχαίο επίπεδο με μέγεθος 2ρ x l. Ο μετασχηματισμός Trace $\breve{g}_z(p, \varphi)$, είναι μια συνάρτηση που ορίζεται στο χώρο των ευθειών ανίχνευσης του z. Οι παράμετροι p και φ καθορίζουν τη θέση της ευθείας στο σχετικό σύστημα συντεταγμένων 2D του z, το οποίο σχηματίζει γωνία θ με το αρχικό επίπεδο. Σύμφωνα με τον τύπο (3.5) το επίπεδο z μπορεί να εκφραστεί ως 2D προβολή ξ(x,y) των κοντινών σημείων, τα οποία συμπίπτουν στο z, ενώ $\breve{g}_z(p, \varphi)$ είναι ο μετασχηματισμός που προκύπτει από τον υπολογισμό του συναρτησιακού Τ πάνω στη γραμμή ανίχνευσης $p = x \cos\varphi + y \sin\varphi$ του z. Το σημείο αναφοράς καθορίζεται από το κέντρο του όγκου των SSTIPs. Το άθροισμα όλων των επίπεδων μετασχηματισμών δίνει το τελικό 3D CTT του πλέγματος Μ, όπως εξηγείται στην εξίσωση (3.6).

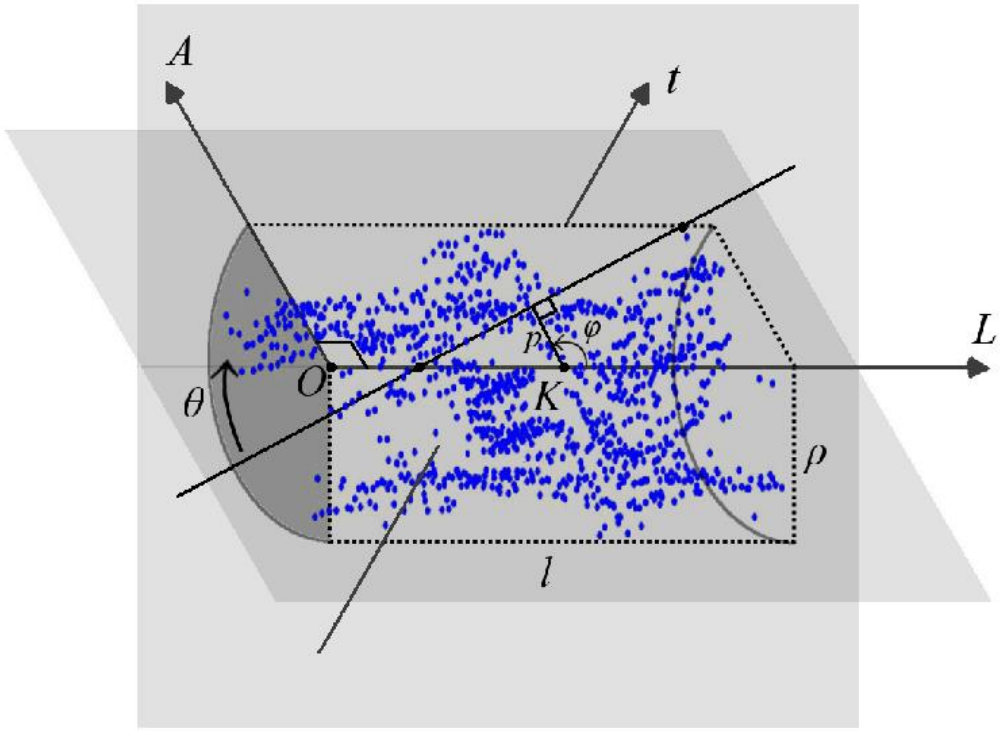
Με την εφαρμογή διαφορετικών συναρτησιακών στο πλέγμα Μ των SSTIPs, παράγεται ένα σύνολο μετασχηματισμών CTT_{f_i} (p, φ), όπου i = 1...I και I είναι ο αριθμός των υπολογιζόμενων μετασχηματισμών. Το τελικό σύνολο των 3D CTT αποτελεί την είσοδο του προγράμματος για την εξαγωγή των χαρακτηριστικών, τα οποία περιγράφονται στη συνέχεια.

Με την παραπάνω μεθοδολογία δεν απαιτείται εξαγωγή σιλουέτας ή χρονική ευθυγράμμιση των αλληλουχιών. Τα σημεία ενδιαφέροντος εξάγονται βάσει της χωρικής και χρονικής αξιολόγησης των αλλαγών κίνησης και έντασης. Με τον υπολογισμό του CTT σε ένα τέτοιο πλέγμα, η γεωμετρική κατανομή των σημείων ενσωματώνεται στον τελικό μετασχηματισμό, ο οποίος ανιχνεύει χωροχρονικές πληροφορίες που χαρακτηρίζουν κάθε ενέργεια. Επομένως, δεν απαιτείται προσωρινή ευθυγράμμιση. Μια απεικόνιση του τρόπου με τον οποίο υπολογίζεται ο 3D CTT σε SSTIP δίνεται στο σχήμα 3.12.

Ένα άλλο πλεονέκτημα είναι η ικανότητα αυτής της διαδικασίας να κωδικοποιεί παραλλαγές στο μήκος των αλληλουχιών δράσης. Πιο συγκεκριμένα, αν η ταχύτητα με την οποία εκτελείται μια ενέργεια διαδραματίζει σημαντικό ρόλο στην ταξινόμηση αυτής της ενέργειας, η παραπάνω διαδικασία ενσωματώνει κατάλληλες πληροφορίες στον παραγόμενο διάνυσμα χαρακτηριστικών. Αυτή η ιδιότητα αποκαλείται χρονική ευαισθησία (time-

sensitivity). Τεχνικές, οι οποίες εφαρμόστηκαν στο παρελθόν, βασίζονται στην εξαγωγή χαρακτηριστικών ανά πλαίσιο και δε διαθέτουν αυτή την ιδιότητα.

Τέλος πρέπει να σημειωθεί ότι ο 3D CTT μετασχηματισμός διαφοροποιείται από τον 3D Radon με έναν ακόμη τρόπο, καθώς ο πρώτος σχεδιάστηκε λαμβάνοντας υπ' όψη την εξαγωγή χαρακτηριστικών από χωροχρονικές αλληλουχίες. Δεδομένου ότι τα μοντέλα 3D στα οποία εφαρμόζεται ο CTT είναι στην πραγματικότητα χωροχρονικοί όγκοι, η διαδικασία επεξεργασίας του μετασχηματισμού εκτελείται πάντοτε με την κατεύθυνση του χρόνου να είναι κάθετη στον άξονα περιστροφής. Με αυτόν τον τρόπο, γίνεται η παραδοχή ότι η δημιουργία ενός χωροχρονικού πλέγματος είναι σταθερή και ευθυγραμμισμένη με τον χρονικό άξονα. Καθώς η χρονική διάσταση δεν περιστρέφεται ποτέ σε άλλη κατεύθυνση μέσα στον τρισδιάστατο χώρο, ο CTT μπορεί να πάρει τη μορφή ενός κυλίνδρου και όχι απαραίτητα σφαίρας. Επομένως η απόδοση του προτεινόμενου μετασχηματισμού αυξάνεται.



Εικόνα 3.12: *Εφαρμογή του 3D CTT μετασχηματισμού σε SSTIPs σημεία μιας ακολουθίας δράσης. Το t υποδηλώνει την κατεύθυνση του χρόνου.*

- **Εφαρμογή Τριπλών Χαρακτηριστικών σε Όγκο (Volumetric Triple Feature - VTF)**

Τα αποτελέσματα του μετασχηματισμού Trace μπορούν να εκφραστούν με χρήση ενός συγκεκριμένου τύπου χαρακτηριστικών, ο οποίος ονομάζεται αναπαράσταση των τριπλών χαρακτηριστικών (triple features). Ένα τριπλό χαρακτηριστικό κατασκευάζεται με τον ακόλουθο τρόπο:

- Αρχικά δημιουργείται ο μετασχηματισμός Trace μιας δισδιάστατης συνάρτησης με τη χρήση ενός συναρτησιακού T.

- Έπειτα υπολογίζεται ένα διαμετρικό συναρτησιακό P κατά μήκος των στηλών του μετασχηματισμού Trace της δισδιάστατης συνάρτησης, από το οποίο λαμβάνεται μια κυκλική συνάρτηση.

- Το τελικό τριπλό χαρακτηριστικό παράγεται με την εφαρμογή μιας κυκλικής συνάρτησης Φ κατά μήκος του προκύπτοντος διανύσματος του βήματος 2.

Στο **[28]** επισημαίνεται ότι ο συνδυασμός των ζευγών τριπλών χαρακτηριστικών, τα οποία έχουν κατασκευαστεί με τη χρήση διαφορετικών συναρτησιακών T, P και Φ στα πλαίσια ενός video, μπορεί να δημιουργήσουν διανύσματα χαρακτηριστικών που αντιπροσωπεύουν αποτελεσματικά την ακολουθία δράσης του video. Η παραπάνω τεχνική εξαγωγής διανυσμάτων χαρακτηριστικών ονομάζεται Ιστορικό Τριπλών Χαρακτηριστικών (HTFs) και είναι κατάλληλη για την αναγνώριση της ανθρώπινης δραστηριότητας. Ωστόσο αυτή η μέθοδος παρουσιάζει ορισμένα ελαττώματα, τα οποία επιδιώκει να αντιμετωπίσει η προτεινόμενη τεχνική διαμόρφωσης τριπλών χαρακτηριστικών σε όγκους (Volumetric Triple Features).

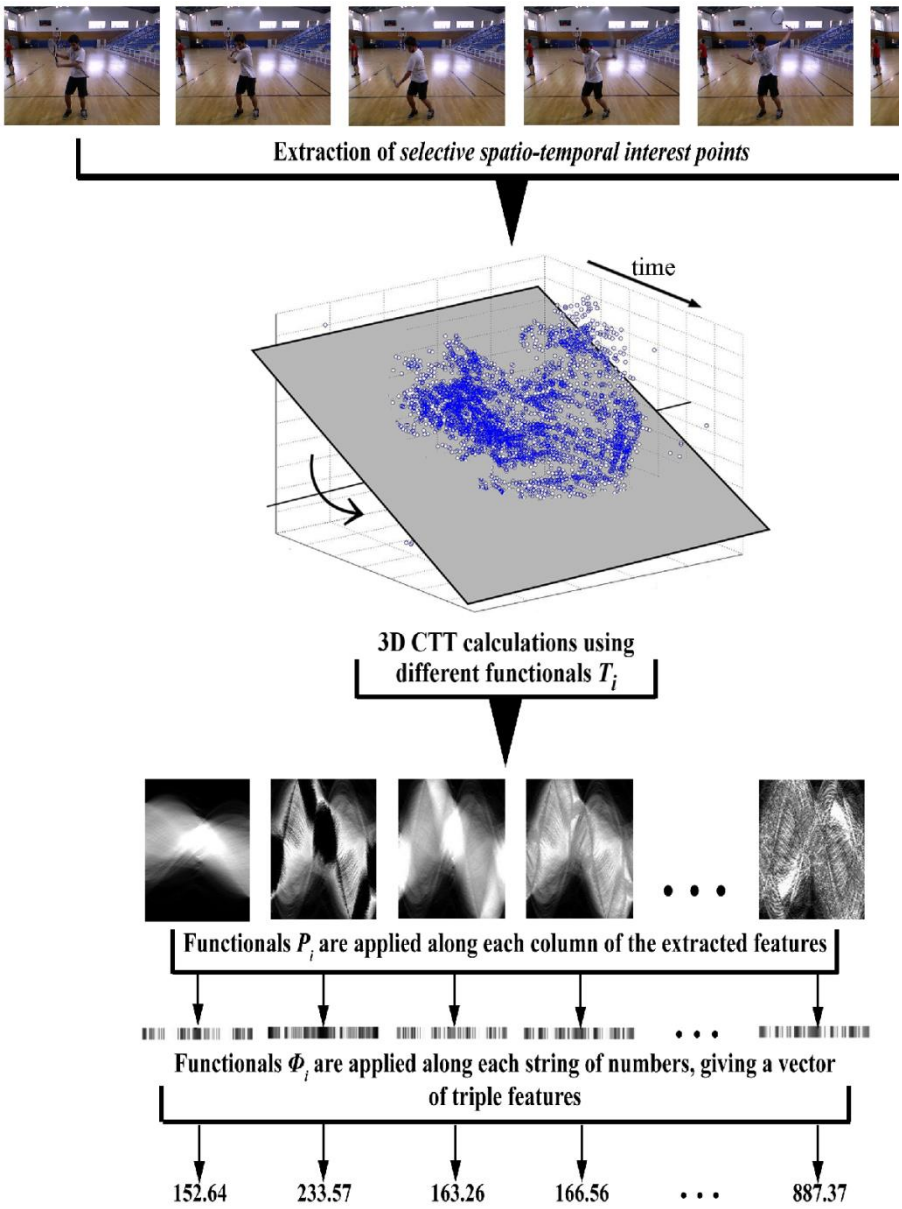
Η προτεινόμενη μέθοδος VTF αξιοποιεί τα αποτελέσματα του 3D CTT, τα οποία προκύπτουν από την εφαρμογή του μετασχηματισμού σε χωροχρονικούς όγκους που παράγονται από SSTIPs. Το σχήμα εξαγωγής των VTF διανυσμάτων ακολουθεί το προαναφερθέν πρότυπο για την εξαγωγή τριπλών χαρακτηριστικών. Συγκεκριμένα για κάθε $CTT_{f_i}(\rho,\varphi)$ μετασχηματισμό, ο οποίος υπολογίζεται μέσω του συναρτησιακού T_i, εφαρμόζεται μια διαμετρική συνάρτηση P_i κατά μήκος των στηλών του μετασχηματισμού. Στη συνέχεια υπολογίζεται μια κυκλική συνάρτηση Φ_i κατά μήκος της προκύπτουσας συμβολοσειράς. Με αυτόν τον τρόπο, ένα σύνολο Π από τριπλά χαρακτηριστικά δημιουργείται.

Επειδή δεν έχουν όλα τα χαρακτηριστικά του διανύσματος την ίδια διακριτική ισχύ, η χρήση μιας τεχνικής μείωσης των διαστάσεων θεωρείται κατάλληλη για την επιλογή των πιο σημαντικών διακριτών χαρακτηριστικών. Αυτή η διαδικασία διευκολύνει την αναγνώριση της ανθρώπινης δραστηριότητας. Στο συγκεκριμένο σύστημα αναγνώρισης χρησιμοποιείται η τεχνική βασικής ανάλυσης συνιστωσών (PCA), η οποία εφαρμόζεται στους φορείς VTF προκειμένου να κατασκευαστεί ένα κατάλληλο υποσύνολο των χαρακτηριστικών ταξινόμησης.

Η μέθοδος PCA μπορεί να θεωρηθεί ότι αναπαριστά ένα n διαστάσεων ελλειψοειδές, του οποίου κάθε άξονας αντιπροσωπεύει ένα κύριο χαρακτηριστικό ενός συνόλου δεδομένων. Αν κάποιος άξονας του ελλειψοειδούς είναι μικρός τότε η διακύμανση κατά μήκος αυτού του άξονα είναι επίσης μικρή. Επομένως παραλείποντας αυτόν τον άξονα και το αντίστοιχο κύριο χαρακτηριστικό του από την αναπαράστασή του συνόλου δεδομένων, χάνεται ένα σχετικά μικρό πλήθος πληροφοριών.

Για την εύρεση των αξόνων του ελλειψοειδούς, πρώτα πρέπει να αφαιρεθεί ο μέσος όρος κάθε μεταβλητής από το σύνολο δεδομένων για να συγκεντρωθούν τα δεδομένα γύρω από την αρχή των αξόνων. Στη συνέχεια, υπολογίζεται ο πίνακας συνδιακύμανσης (covariance matrix) των δεδομένων και τα αντίστοιχα ιδιοδιανύσματα και ιδιοτιμές. Έπειτα γίνεται κανονικοποίηση των ιδιοδιανυσμάτων, τα οποία αντιπροσωπεύουν τους άξονες του ελλειψοειδούς που είναι προσαρμοσμένοι στα δεδομένα. Το ποσοστό της διακύμανσης που αντιπροσωπεύει κάθε ιδιοδιάνυσμα μπορεί να υπολογιστεί διαιρώντας την αντίστοιχη ιδιοτιμή του με το άθροισμα όλων των ιδιοτιμών.

Περισσότερες πληροφορίες παρουσιάζονται στην έρευνα των **[34], [35]** και **[36].**



Εικόνα 3.13: *Παράδειγμα εξαγωγής τριπλών χαρακτηριστικών στις τρεις διαστάσεις (χωροχρόνος).*

3.6.6 Πειραματικά Αποτελέσματα

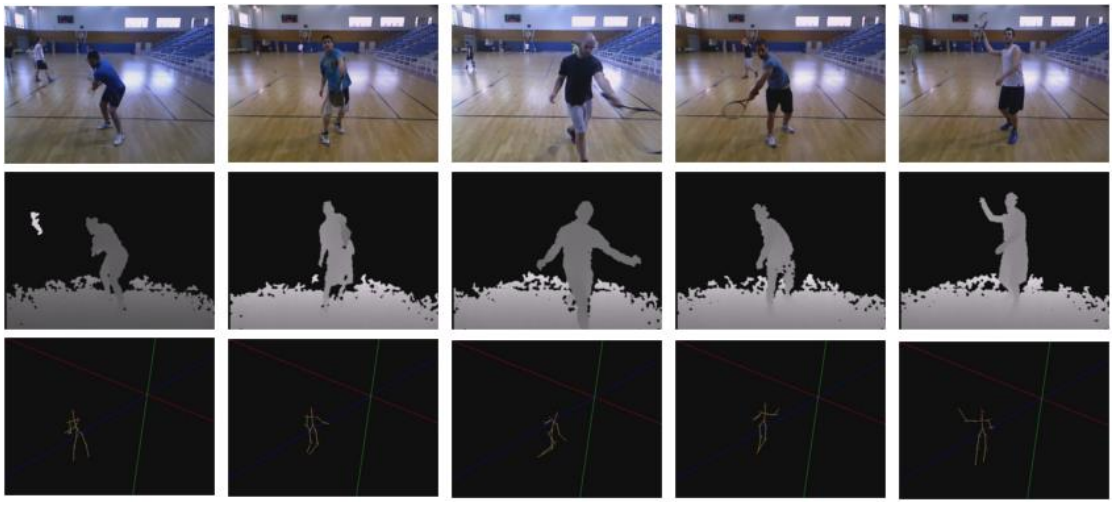
Για την πειραματική αξιολόγηση της προτεινόμενης μεθόδου χρησιμοποιούνται τα σύνολα KTH, Weizmann και THETIS.

Διευκρινίζεται ότι η βάση δεδομένων Weizmann αποτελείται από ένα σύνολο 90 ακολουθιών video χαμηλής ανάλυσης που δείχνουν εννέα διαφορετικά θέματα. Κάθε υποκείμενο/χρήστης εκτελεί 10 φυσικές ενέργειες όπως: running, walking, skipping, jumping-jack (jack), jumping forward on two legs (jump), jumping in place on two legs (pjump), galloping sideways (or side), waving with two hands (wave2), waving using one hand (wave1) και bending.

Στα ακόλουθα πειράματα, χρησιμοποιήθηκε το πρωτόκολλο leave-one-person-out cross validation μαζί με SVM για την ταξινόμηση των ακολουθιών δράσης. Για το σύνολο KTH έγινε εκπαίδευση 6 SVMs, ένα για κάθε κατηγορία δράσης, με βάση το πρωτόκολλο one-against-all. Ομοίως για τα σύνολα Weizmann και THETIS χρησιμοποιήθηκαν 10 και 12 SVMs αντίστοιχα. Τα πειράματα εκτελέστηκαν για διαφορετικές τιμές περιστροφής θ του επιπέδου βήματος 6° και 9°. Επισημαίνεται ότι όσο μικρότερο είναι το βήμα, τόσο περισσότερο τείνει να γίνει συνεχόμενη η κίνηση των ακολουθιών δράσης. Τέλος για τη μείωση των διαστάσεων του παραγόμενου διανύσματος έγινε χρήση της PCA τεχνικής.

Εικόνα 3.14: *Βάση δεδομένων Weizmann. Παραδείγματα των κινήσεων wave1, wave2, walk, pjump, side,run, skip, jack, jump και bend.*

Εικόνα 3.15: *Βάση δεδομένων KTH. Παραδείγματα των κινήσεων walking, jogging, running, boxing, hand waving και hand clapping.*



Εικόνα 3.16: *Βάση δεδομένων THETIS. Παραδείγματα των κινήσεων backhand, flat service, forehand flat, slice service και smash. Πρώτη γραμμή: RGB δεδομένα. Δεύτερη γραμμή: Δεδομένα βάθους. Τρίτη γραμμή: Σκελετικά δεδομένα.*

Τα αποτελέσματα αναγνώρισης δράσης της προτεινόμενης μεθόδου περιγράφονται στον πίνακα 3.5. Επιπροσθέτως, οι πίνακες 3.3 και 3.4 παρουσιάζουν τα αποτελέσματα της εφαρμογής της μεθόδου στα σύνολα δεδομένων KTH και Weizmann. Οι γραμμές

απεικονίζουν την επιτευχθείσα ακρίβεια (σωστές απαντήσεις αναγνώρισης / σύνολο απαντήσεων) από όλα τα εκπαιδευμένα SVMs για μια συγκεκριμένη κατηγορία δράσης. Οι στήλες απεικονίζουν την απόδοση των επιμέρους SVMs σε κάθε κατηγορία δράσης.

Σύμφωνα με τον πίνακα 3.5, η διαδικασία εξαγωγής χαρακτηριστικών με βάση το 3D CTT και τα επιλεκτικά STIPs επιτυγχάνει μεγάλη ακρίβεια σε όλα τα εξεταζόμενα σύνολα δεδομένων. Για παράδειγμα στη βάση δεδομένων KTH, επιτυγχάνεται ακρίβεια 99.98%, η οποία είναι μεγαλύτερη από αυτή ορισμένων τεχνικών **[32]** στο ίδιο σύνολο δεδομένων. Τα αποτελέσματα για το Weizmann υποδηλώνουν την ύπαρξη μιας μικρής σύγχυσης στην κατηγοριοποίηση διαφόρων ενεργειών όπως η δράση "jump".

Αναφέρεται επίσης ότι υπάρχει σημαντική διαφορά στην απόδοση μεταξύ των τριών τύπων δεδομένων στο σύνολο THETIS. Στα συγκριτικά αποτελέσματα του πίνακα 3.5, υποδηλώνεται ότι η προτεινόμενη τεχνική ξεπερνά όλες τις άλλες μεθόδους, οι οποίες δοκιμάστηκαν σε αυτό το σύνολο δεδομένων, δηλαδή τη μέθοδο HOG-HOF με βάση τα STIPs **[23]**, τη προσέγγιση πυκνών τροχιών **[24]** για τα υποσύνολα Depth και Skelet3D και την προσέγγιση των δυναμικών φάσεων **[37]** στο υποσύνολο RGB. Στο υποσύνολο Skelet3D παρατηρείται κάποια πτώση της ακρίβειας, η οποία όμως οφείλεται στη φύση των δεδομένων. Οι σκελετοί αποτελούν ένα υπεραπλουστευμένο μοντέλο του κινούμενου ανθρώπινου σώματος με ελάχιστη επιφάνεια, με αποτέλεσμα να εμποδίζεται η ακριβή απόκτηση του STIP. Ωστόσο, σε σύγκριση με τα αποτελέσματα μιας άλλης μεθόδου που βασίζεται σε STIP **[23]**, η προτεινόμενη μέθοδος φαίνεται να είναι μια τελική βελτίωση. Για τα δεδομένα βάθους και RGB, η απόδοση είναι πολύ υψηλή, γεγονός το οποίο υποδεικνύει την καταλληλότητα της συγκεκριμένης μεθόδου σε απεικόνιση RGB και γκρι κλίμακας.

	boxing	Handclapping	handwaving	jogging	running	Walking
boxing	1	0	0	0	0	0
Handclapping	0.0009	0.9991	0	0	0	0
Handwaving	0	0	1	0	0	0
Jogging	0	0	0	1	0	0
Runing	0	0	0.01	0	1	0
Walking	0	0	0	0	0	1

Πίνακας 3.3: *Αποτελέσματα απόδοσης της προτεινόμενης μεθόδου στο σύνολο δεδομένων KTH*.

	Bend	Jack	Jump	Pjump	Run	side	Skip	walk	wave1	wave2
bend	1	0	0.1111	0	0	0	0	0	0	0
jack	0	1	0.0455	0	0	0	0	0	0	0
jump	0.0588	0.0294	0.9706	0.0294	0.1176	0.0294	0.0588	0.0294	0.0588	0.0294
pjump	0.0303	0.0303	0.0303	0.9697	0.0303	0.0303	0.0303	0.0303	0.0303	0.0303
run	0.1053	0.1579	0.2105	0.1053	0.9474	0.0526	0.1579	0.0526	0.1053	0.0526
side	0	0	0	0	0	1	0.0357	0	0	0
skip	0.0526	0.0526	0.1053	0.0526	0.0789	0.0526	0.9474	0.0526	0.0526	0.0526
walk	0.0455	0.0455	0.1364	0	0.0455	0	0.0455	1	0	0
wave1	0.0476	0.0476	0.0476	0	0.0952	0.0476	0.0952	0.0476	0.9524	0.0476
wave2	0	0	0	0	0	0	0	0	0	1

Πίνακας 3.4: *Αποτελέσματα απόδοσης της προτεινόμενης μεθόδου στο σύνολο δεδομένων WEIZMANN.*

Method	Dataset				
	KTH	Weizmann	THETIS- Skelet3D	THETIS-Depth	THETIS-RGB
3D CTT - SSTIP based VTFs	99.98	96.34	86.06	98.03	100
Selective STIPs + BoVW (B. Chakraborty, 2012)	96.35	99.5	-	-	-
Dense Trajectories: MBH (H.Wang, 2011)	92.32	-	46.84	51.59	-
Dense Trajectories: Combination (H.Wang, 2011)	90.65	-	50.78	54.32	-
Dense Trajectories: Trajectory (H.Wang, 2011)	86.98	-	53.08	57.5	-
History Triple Features (G. Goudelis, 2013)	93.14	95.42	-	-	-
Yuan et al. (C. Yuan, 2013)	95.49	-	-	-	-
Vainstein et al. (J. Vainstein, 2014)	-	-	-	-	86.44
Kumar and John (Kumar and John, 2016)	94.62	95.69	-	-	-
Liu et al. (A.A. Liu, 2015)	96.7	-	-	-	-
Laptev et al. (I. Laptev, 2008)	92.99	-	54.4	60.23	-

Πίνακας 3.5: *Ποσοστά απόδοσης ταξινόμησης(%) για διαφορετικές τεχνικές εφαρμοσμένες στα σύνολα δεδομένων KTH, WEIZMANN,THETIS.*

ΚΕΦΑΛΑΙΟ 4

Εξαγωγή και Επεξεργασία Σκελετικών Δεδομένων από τη THETIS

4.1 Εισαγωγή

Η εξαγωγή των σκελετικών δεδομένων των παικτών επιτυγχάνεται μέσω προγράμματος γραμμένου σε C++ με τη χρήση της βιβλιοθήκης των OpenNI 2.2 και NiTE 2.0. Συγκεκριμένα η εύρεση του σκελετού κάθε παίκτη γίνεται με την αξιοποίηση του ενδεικτικού προγράμματος UserViewer του NiTE 2.0. Το παραπάνω πρόγραμμα επεκτείνεται για την κατάλληλη επεξεργασία των δεδομένων.

4.2 Παραδοχές Επεξεργασίας Δεδομένων

Επισημαίνεται ότι για την επεξεργασία των σκελετικών δεδομένων θεωρείται ότι καθ' όλη τη διάρκεια του video αναγνωρίζεται ένας κύριος χρήστης (prominent user), ο οποίος είναι ο αντίστοιχος παίκτης του tennis. Στα περισσότερα videos της βάσης εντοπίζεται μόνο ο παίκτης του tennis, καθώς δεν παρουσιάζονται άλλα υποκείμενα στη 3D σκηνή ή αυτά βρίσκονται σε απόσταση αρκετά μεγάλη από τον αισθητήρα του Kinect. Σε μερικές περιπτώσεις όμως, μπορεί να αναγνωρίζονται άτομα εκτός των παικτών, αλλά αυτά εντοπίζονται σε λιγότερα καρέ, με αποτέλεσμα κύριος χρήστης να παραμένει ο παίκτης. Τέλος σε ορισμένα videos υπάρχουν επιμέρους παρεμβολές – ατυχήματα (μετακίνηση της κάμερας από λάθος, διακοπή της εκτέλεσης κάποιας κίνησης tennis στη μέση κτλ.), τα οποία προαναφέρονται στο Κεφάλαιο 2. Τα τελευταία δε λαμβάνονται στο σύνολο των videos για την επεξεργασία των σκελετικών δεδομένων.

4.3 Περιγραφή Προγράμματος Εξαγωγής Σκελετικών Δεδομένων

Το πρόγραμμα της επεξεργασίας των σκελετικών δεδομένων διακρίνεται σε επιμέρους στάδια, τα οποία επαναλαμβάνονται σε κάθε frame κατά τη διάρκεια του video (σε κάθε εκτέλεση του). Τα βασικά στάδια της λειτουργίας του προγράμματος παρουσιάζονται στη συνέχεια μαζί με τις κυριότερες δομές που χρησιμοποιήθηκαν.

4.3.1 Περιγραφή Δομών

- **FrameList[ID][FrameNo][Joint][Dimension]**

Το διάνυσμα FrameList χρησιμοποιείται για την προσωρινή αποθήκευση των 3D συντεταγμένων όλων των αρθρώσεων κάθε ατόμου που εντοπίζεται στη 3D σκηνή σε μία

επανάληψη του ONI video. Το διάνυσμα αποτελείται από τέσσερις διαστάσεις, οι οποίες περιγράφονται παρακάτω.

ID: Αναφέρεται στο πλήθος των ατόμων, τα οποία ανιχνεύονται σε κάποιο frame κατά τη διάρκεια εκτέλεσης του video. Κάθε ένα από αυτά τα άτομα διαθέτει ένα δικό του αναγνωριστικό ID, το οποίο χρησιμοποιείται στη συγκεκριμένη περίπτωση ως δείκτης για το διάνυσμα FrameList. Επισημαίνεται ότι σε διαφορετικές εκτελέσεις του video μπορεί να ανατίθενται διαφορετικά ID στα άτομα, ωστόσο για μια μεμονωμένη εκτέλεση αυτά παραμένουν σταθερά για κάθε χρήστη και είναι καθοριστικά για την αναγνώριση του.

FrameNo: Αναφέρεται στο πλήθος των συνολικών καρέ, τα οποία αποτελούν το video. Σε κάθε frame ανατίθεται ένας αριθμός FrameNo, ο οποίος αποτελεί το αναγνωριστικό του καρέ και χρησιμοποιείται για δεικτοδότηση στο FrameList διάνυσμα.

Joint: Αναφέρεται στο πλήθος των αρθρώσεων, οι οποίες είναι συνολικά δεκαπέντε. Για την κάθε άρθρωση έχει οριστεί μια αντιστοίχιση με έναν αριθμό (0–14), προκειμένου να γίνεται κατάλληλα η αποθήκευση των συντεταγμένων της στο διάνυσμα. Συγκεκριμένα:

Index0 = HEAD
Index1 = LEFT_ELBOW
Index2= LEFT_FOOT
Index3= LEFT_HAND
Index4= LEFT_HIP
Index5= LEFT_KNEE
Index6= LEFT_SHOULDER
Index7 = NECK
Index8= RIGHT_ELBOW
Index9= RIGHT_FOOT
Index10= RIGHT_HAND
Index11= RIGHT_HIP
Index12= RIGHT_KNEE
Index13= RIGHT_SHOULDER
Index14= TORSO

Dimension: Αναφέρεται στις τρεις διαστάσεις (x,y,z), οι οποίες καθορίζουν τη θέση μιας άρθρωσης στη 3D σκηνή. Επισημαίνεται ότι το διάνυσμα αρχικοποιείται όταν το video τελειώσει, για την επόμενη εκτέλεση του.

- **ProminentUser[FrameNo][Joint][Dimension]**

Το διάνυσμα ProminentUser χρησιμοποιείται για την αποθήκευση των επεξεργασμένων συντεταγμένων κάθε frame του κύριου χρήστη (παίκτη του tennis) για όλες τις εκτελέσεις του video. Αποτελείται από τρεις διαστάσεις, οι οποίες καθορίζουν το καρέ, την άρθρωση και τη διάσταση στην οποία γίνεται αναφορά. Η επεξεργασία των συντεταγμένων περιγράφεται στη συνέχεια και έχει στόχο την ομαλοποίηση των συντεταγμένων από πιθανό θόρυβο και την καλύτερη προσέγγιση της θέσης των αρθρώσεων.

- **trackingCounter[ID]**

Το διάνυσμα trackingCounter αποτελεί έναν μετρητή, στον οποίο αποθηκεύεται το πλήθος εμφανίσεων (ανίχνευσης) κάθε χρήστη στη 3D σκηνή κατά τη διάρκεια μιας εκτέλεσης του video. Σε κάθε νέα επανάληψη ενός video απαιτείται αρχικοποίηση του διανύσματος.

- **Factors[FrameNo]**

Το διάνυσμα Factors αποτελεί έναν μετρητή, ο οποίος αποθηκεύει για κάθε frame πόσες φορές εντοπίστηκε Prominent User σε όλες τις επαναλήψεις του video. Τα περιεχόμενα του διανύσματος χρησιμοποιούνται για την επεξεργασία των συντεταγμένων του παίκτη.

4.3.2 Ανάλυση Βημάτων Επεξεργασίας των Σκελετικών Δεδομένων

Ακολουθούν τα κυριότερα στάδια της εξαγωγής και επεξεργασίας των σκελετικών δεδομένων, τα οποία αξιοποιούν τις προαναφερόμενες δομές.

- **Εντοπισμός Χρήστη και Σκελετικών Δεδομένων**

Ο εντοπισμός των σκελετικών δεδομένων αφορά την ανίχνευση του παίκτη tennis στη 3D σκηνή και κατ' επέκταση την εύρεση των σκελετικών του αρθρώσεων (skeletal joints). Όπως αναφέρθηκε μπορεί να εντοπιστούν περισσότερα από ένα άτομα, κάθε ένα από τα οποία διαθέτει δικό του αναγνωριστικό ID. Κατά την ανίχνευση κάθε χρήστη σε ένα frame, εντοπίζονται οι συντεταγμένες των αρθρώσεων του (μέσω συναρτήσεων του OpenNI) και αποθηκεύονται προσωρινά στο 4D διάνυσμα FrameList. Η διαδικασία της ανίχνευσης ενός ατόμου επαναλαμβάνεται σε κάθε καρέ για όλες τις εκτελέσεις του video. Αν δεν εντοπιστεί κάποιος χρήστης στη 3D σκηνή δε γίνεται καμία διαδικασία επεξεργασίας, καθώς δεν είναι διαθέσιμα σκελετικά δεδομένα.

- **Αναγνώριση Κύριου Χρήστη (Prominent User)**

Ως Prominent User θεωρείται ο χρήστης, ο οποίος εντοπίζεται στα περισσότερα frames του video. Καθώς κάθε παίκτης του tennis βρίσκεται στη 3D σκηνή εγγραφής καθ' όλη τη διάρκεια του video, ανιχνεύεται σε περισσότερα καρέ και αναγνωρίζεται ως το κεντρικό υποκείμενο – στόχος, ανεξαρτήτως άλλων ατόμων που μπορεί να παρουσιαστούν στη σκηνή για κάποιο χρονικό διάστημα. Για την εύρεση του Prominent User χρησιμοποιείται το μονοδιάστατο διάνυσμα trackingCounter[ID], το οποίο ενημερώνεται κάθε φορά που εντοπίζεται ένας χρήστης. Συγκεκριμένα αν σε ένα frame ανιχνευθεί ένα άτομο, το διάνυσμα trackingCounter αυξάνεται στη θέση με ID αυτό του παραπάνω ατόμου. Στο τέλος κάθε εκτέλεσης ενός video, συγκρίνονται τα περιεχόμενα του διανύσματος και εντοπίζεται το ID του χρήστη που ανιχνεύθηκε τις περισσότερες φορές (Prominent User).

- **Ενημέρωση Σκελετικών Δεδομένων του Prominent User**

Μετά την ολοκλήρωση μιας εκτέλεσης του video εισόδου, υπολογίζεται ο κύριος χρήστης και έπειτα γίνεται ενημέρωση των συντεταγμένων του. Για την επεξεργασία των συντεταγμένων του χρησιμοποιείται το 3D διάνυσμα ProminentUser, στο οποίο

αποθηκεύονται οι συντεταγμένες (x,y,z) κάθε άρθρωσης (joint) για όλα τα καρέ του video. Η ενημέρωση των συντεταγμένων των σκελετικών δεδομένων γίνεται με τον τύπο:

$$ProminentUser[FrameNo][Joint][Dimentsion] = \frac{Factors[FrameNo]-1}{Factors[FrameNo]} * ProminentUser[FrameNo][Joint][Dimentsion] + \frac{1}{Factors[FrameNo]} * FrameList[ID][FrameNo][Joint][Dimension] \quad (4.1)$$

ή πιο απλά

$$New_{Joint} = \frac{Factors-1}{Factors} * Old_{Joint} + \frac{1}{Factors} * Current_{Joint} \quad (4.2)$$

όπου το ID αναφέρεται στο αναγνωριστικό του ατόμου, το οποίο ανιχνεύεται ως κύριος χρήστης στη συγκεκριμένη επανάληψη του video. Ο συγκεκριμένος τύπος εφαρμόζεται για όλα τα frames του video, για όλες τις αρθρώσεις του παίκτη στις τρεις διαστάσεις. Η διαδικασία της ενημέρωσης των συντεταγμένων επαναλαμβάνεται μετά την ολοκλήρωση μιας εκτέλεσης του video. Από τα παραπάνω προκύπτει ότι η επαναλαμβανόμενη εκτέλεση κάθε video, βελτιώνει/ομαλοποιεί τις συντεταγμένες των αρθρώσεων του παίκτη σε κάθε καρέ. Επισημαίνεται ότι η χρήση του διανύσματος Factors είναι απαραίτητη, καθώς δεν ανιχνεύεται πάντοτε κάποιο υποκείμενο στη 3D σκηνή για όλα τα frames. Επομένως σε διαφορετικά καρέ ο Prominent User μπορεί να έχει εντοπιστεί διαφορετικό πλήθος φορών.

- **Έναρξη και Τερματισμός Προγράμματος**

Κατά την έναρξη του προγράμματος γίνονται όλες οι απαραίτητες αρχικοποιήσεις των πινάκων/διανυσμάτων και των συσκευών εισόδου δεδομένων / devices. Αντίστοιχα ο τερματισμός του προγράμματος γίνεται με τη χρήση του πλήκτρου esc ή όταν κάποιος από τους συντελεστές ενημέρωσης (1/Factors[FrameNo]) γίνει μικρότερος ενός καθορισμένου κατωφλίου (threshold). Πριν την εκτέλεση του τερματισμού, οι επεξεργασμένες συντεταγμένες του Prominent User αποθηκεύονται σε κατάλληλο αρχείο, το οποίο ορίζεται από το χρήστη του προγράμματος.

4.3.3 Σύνοψη της Λειτουργίας του Προγράμματος

Βήμα 1: Αρχικοποίηση κατάλληλων devices του OpenNI/NiTE.

Βήμα 2: Αρχικοποίηση των δομών ProminentUser, Factors.

Βήμα 3: Αρχικοποίηση των δομών FrameList, trackingCounter.

Βήμα 4: Ανάγνωση/Λήψη επόμενου frame:

- **Βήμα 4.1:** Εντοπισμός ενός ή περισσότερων ατόμων (αν ανιχνεύονται).

- **Βήμα 4.2:** Ενημέρωση των διανυσμάτων FrameList και trackingCounter (αν εντοπίζεται κάποιο υποκείμενο).
- **Βήμα 4.3:** Έλεγχος αν εξετάζεται το τελευταίο frame. Αν ναι, εκτέλεση του Βήματος 5. Αν όχι, επανάληψη του Βήματος 4.

Βήμα 5: Επεξεργασία συντεταγμένων του Prominent User:

- **Βήμα 5.1:** Εύρεση του Prominent User με χρήση του trackingCounter.
- **Βήμα 5.2:** Ενημέρωση του διανύσματος ProminentUser με χρήση του τύπου (4.1).
- **Βήμα 5.3:** Ενημέρωση του διανύσματος Factors, με βάση τις νέες εμφανίσεις του κύριου χρήστη στα καρέ του video.
- **Βήμα 5.4:** Έλεγχος τερματισμού προγράμματος. Αν πατήθηκε esc ή ικανοποιείται η συνθήκη του threshold, εκτέλεση του Βήματος 6. Αν όχι, εκτέλεση του Βήματος 3.

Βήμα 6: Αποθήκευση συντεταγμένων του Prominent User σε αρχείο.

Βήμα 7: Τερματισμός προγράμματος.

4.4 Αναπαράσταση Σκελετικών Δεδομένων στο Περιβάλλον του Unity

Για τον έλεγχο της ορθότητας των εξαγόμενων σκελετικών δεδομένων, αναπτύχθηκε πρόγραμμα σε C# με τη χρήση του Unity. Το παραπάνω χρησιμοποιήθηκε για την οπτικοποίηση των αρθρώσεων του χρήστη, με σκοπό να ανιχνευθούν πιθανές ανωμαλίες στις συντεταγμένες τους.

Τα βασικά στάδια της λειτουργίας του προγράμματος περιλαμβάνουν την ανάγνωση των συντεταγμένων από τα εξαγόμενα αρχεία και την αναπαράσταση τους στο Unity με χρήση κατάλληλων αντικειμένων. Συγκεκριμένα :

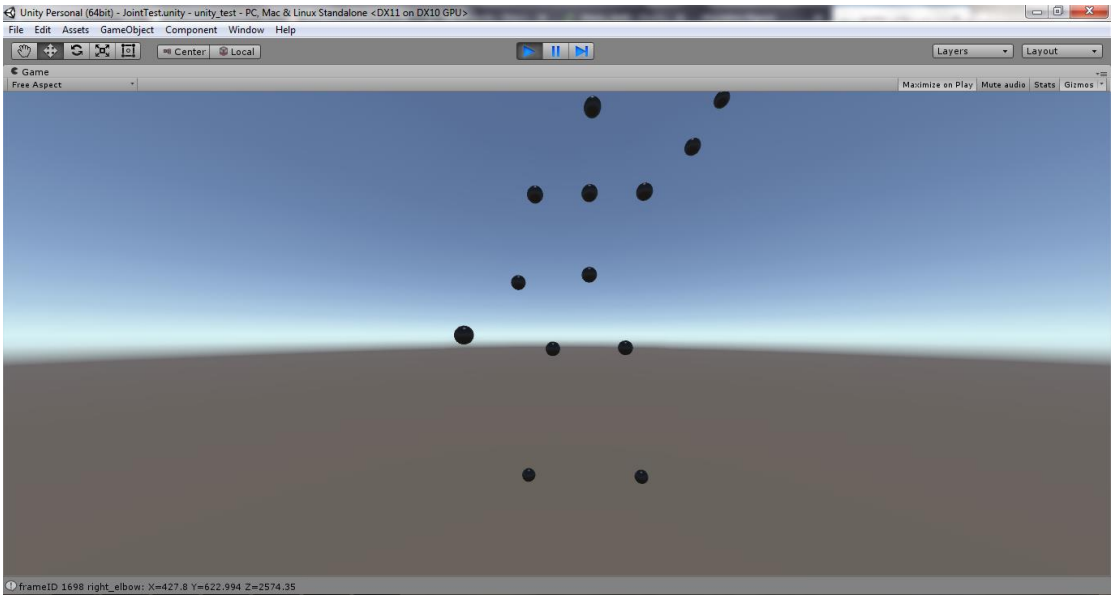
Κάθε άρθρωση του σκελετού αντιπροσωπεύεται από ένα αντικείμενο (Object) σχήματος σφαίρας. Κάθε σφαίρα περιέχει τρία πεδία (Transform Position (x, y, z) , τα οποία περιγράφουν τη θέση της στο περιβάλλον του Unity. Επίσης κάθε αντικείμενο σφαίρα συνδέεται με ένα script (JointPositionController.cs), το οποίο αναλαμβάνει την ανανέωση των συντεταγμένων του στο χώρο.

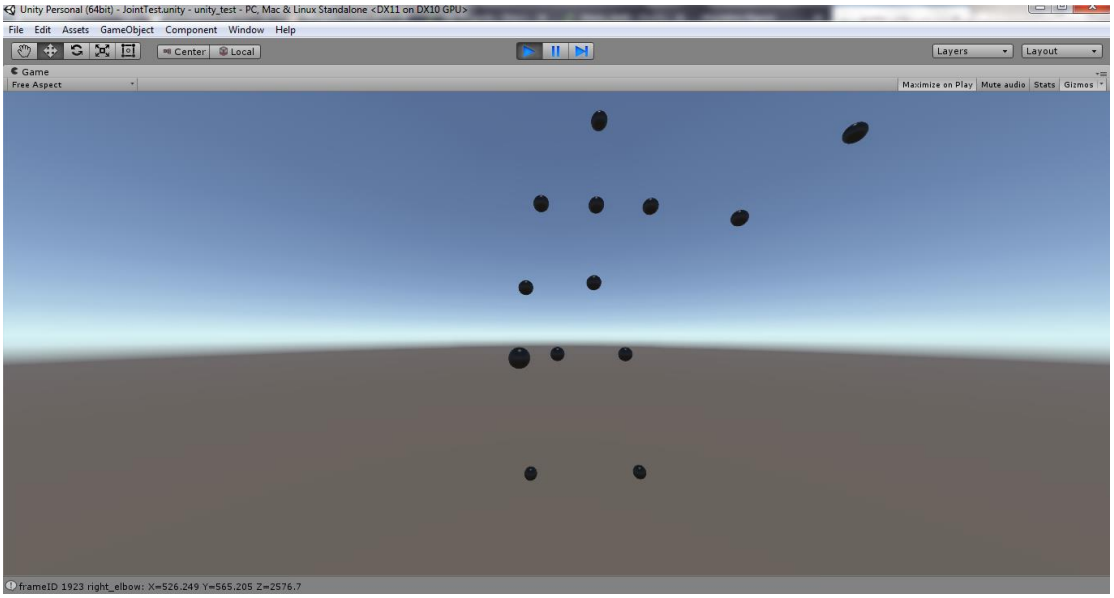
Κατά την έναρξη του προγράμματος, εκτελείται ένα script (OniFileReader.cs), στο οποίο γίνεται ανάγνωση και αποθήκευση των συντεταγμένων των αρθρώσεων για κάθε frame του αρχείου εισόδου σε μια λίστα.

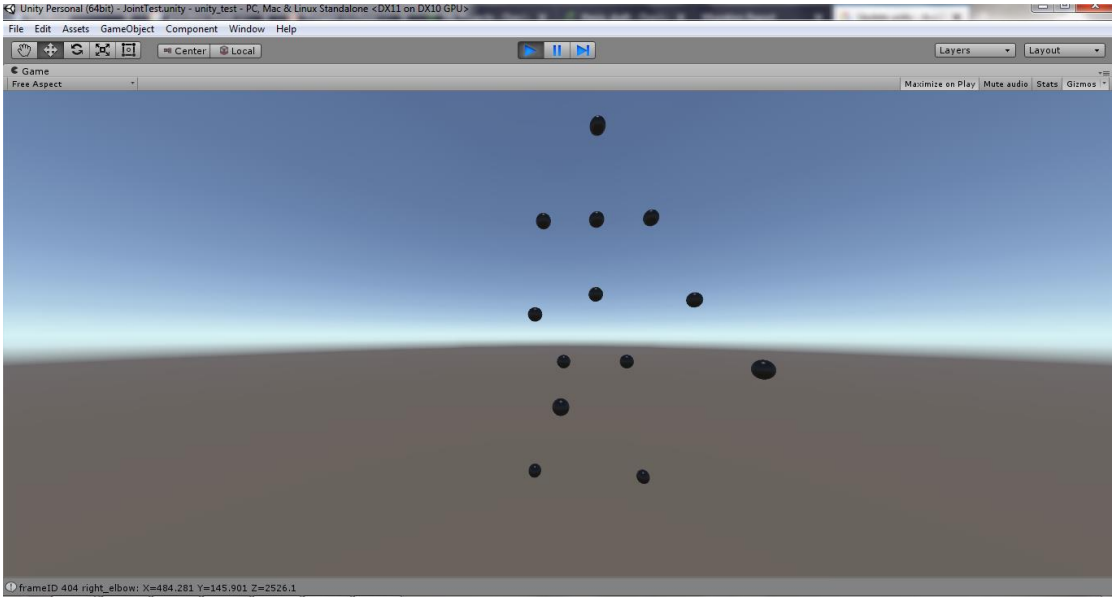
Στη συνέχεια, οι αποθηκευμένες συντεταγμένες χρησιμοποιούνται για την ανανέωση των x, y, z πεδίων των αντικειμένων μέσω του JointPositionController.cs. Επισημαίνεται ότι

η ανανέωση των συντεταγμένων εκτελείται σε κάθε frame μέσω της συνάρτησης Monobehaviour.update( ). Επίσης οι συντεταγμένες που περιέχονται στο αρχείο εισόδου αντιπροσωπεύουν τις πραγματικές 3D συντεταγμένες στο χώρο.

Στιγμιότυπα της αναπαράστασης των σκελετικών δεδομένων στο περιβάλλον του Unity παρουσιάζονται παρακάτω.







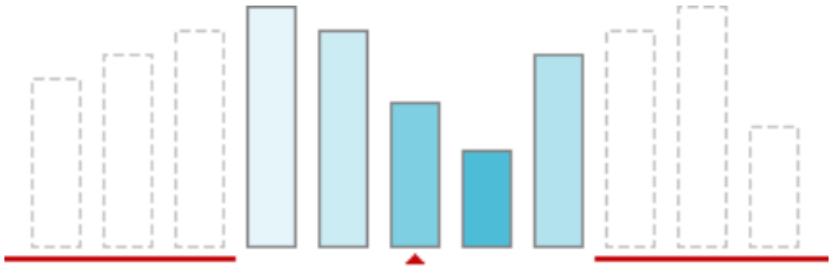
Εικόνα 4.1: *Στιγμιότυπο αναπαράστασης σκελετικών δεδομένων της κίνησης backhand στο Unity.*

4.5 Περιγραφή Φίλτρων Ομαλοποίησης των Σκελετικών Δεδομένων

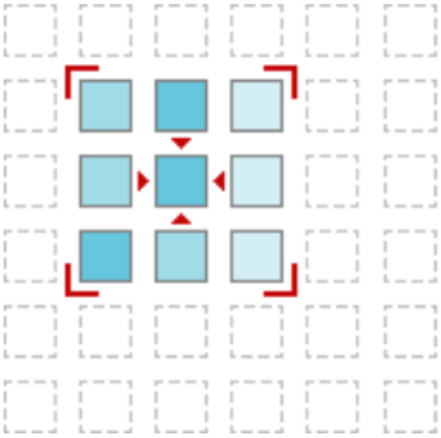
4.5.1 Φίλτρο της Διάμεσης Τιμής (Median Filter)

Το φίλτρο της διάμεσης τιμής είναι μια μη γραμμική τεχνική ψηφιακού φιλτραρίσματος, η οποία χρησιμοποιείται για την εξάλειψη/μείωση του θορύβου σε μια εικόνα ή ένα σήμα. Η εφαρμογή του φίλτρου αποτελεί συνήθως ένα στάδιο προεπεξεργασίας των δεδομένων εισόδου. Το συγκεκριμένο φίλτρο χρησιμοποιείται ευρέως στην ψηφιακή επεξεργασία εικόνων, καθώς μπορεί σε κάποιο βαθμό να διακρίνει τον απομονωμένο θόρυβο από χαρακτηριστικά της εικόνας (ακμές, γραμμές κτλ.) χωρίς να τα επηρεάσει.

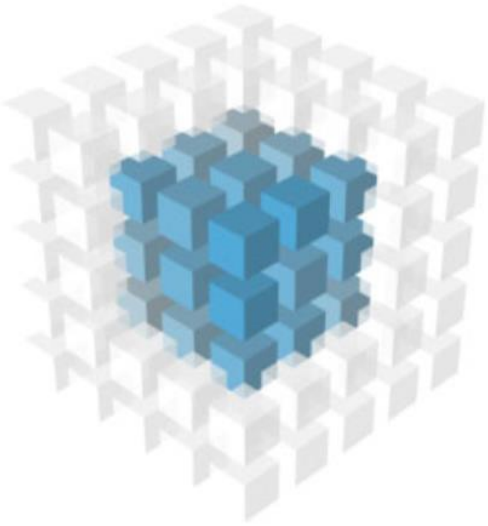
Η κύρια ιδέα του συγκεκριμένου φίλτρου στηρίζεται στην αντικατάσταση της τιμής κάθε pixel της εικόνας με το διάμεσο του συνόλου τιμών των γειτονικών pixels. Το μοτίβο της γειτονιάς ονομάζεται "παράθυρο" (window) και ολισθαίνει από pixel σε pixel σε όλη την εικόνα. Για τα μονοδιάστατα σήματα, το παράθυρο αποτελείται από προηγούμενα και επόμενα γειτονικά pixels από αυτό, το οποίο εξετάζεται τη δεδομένη χρονική στιγμή. Για τα δισδιάστατα ή πολυδιάστατα σήματα τα μοτίβα της γειτονιάς μπορεί να είναι πιο πολύπλοκα (όπως ένα παράθυρο σε σχήμα σταυρού ή κύβου κτλ.). Παραδείγματα "παραθύρου" γειτονιάς παρουσιάζονται στις εικόνες 4.2, 4.3 και 4.4. Επισημαίνεται ότι αν το παράθυρο γειτονιάς έχει περιττό αριθμό καταχωρήσεων, τότε η διάμεσος τιμή είναι η μεσαία τιμή, η οποία προκύπτει όταν όλες οι καταχωρίσεις/pixels της γειτονιάς ταξινομηθούν. Ωστόσο αν το παράθυρο γειτονιάς έχει άρτιο αριθμό εγγραφών, τότε υπάρχουν περισσότερες από μια πιθανές διάμεσες τιμές.



Εικόνα 4.2: *Παράδειγμα "παραθύρου" γειτονιάς σε μια διάσταση.*



Εικόνα 4.3: *Παράδειγμα "παραθύρου" γειτονιάς σε δύο διαστάσεις.*



Εικόνα 4.4: *Παράδειγμα "παραθύρου" γειτονιάς σε τρεις διαστάσεις.*

Τα παραπάνω περιγράφονται από τον τύπο:

$$y[m,n] = median\ x[i,j], (i,j) \in w \qquad (4.3)$$

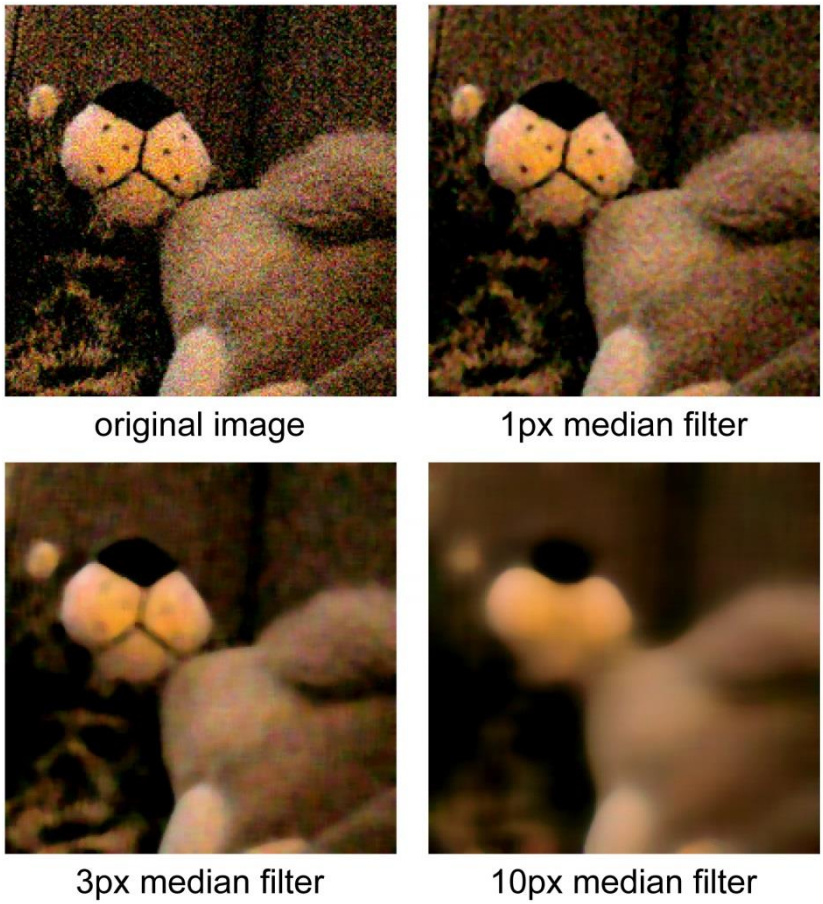
όπου w είναι η γειτονιά (η οποία καθορίζεται από το χρήστη) με κέντρο το σημείο της εικόνας [m,n] και x[i,j] είναι οι τιμές των pixels που ανήκουν στη γειτονιά w.

Ένα παράδειγμα εφαρμογής του φίλτρου παρουσιάζεται στη συνέχεια για μονοδιάστατο σήμα x και παράθυρο γειτονιάς μεγέθους 3 καταχωρήσεων:

x= (2, 80, 6, 3)
y1 = med (2, 2, 80) = 2
y2 = med (2, 80, 6) = med (2, 6, 80) = 6
y3 = med (80, 6, 3) = med (3, 6, 80) = 6
y4 = med (6, 3, 3) = med (3, 3, 6) = 3

Επομένως το σήμα εξόδου μετά την εφαρμογή του φίλτρου είναι y = (2,6, 6,3).

Εικόνα 4.5: *Παράδειγμα εφαρμογής φίλτρου διάμεσης τιμής για την εξάλειψη απομονωμένου θορύβου*.

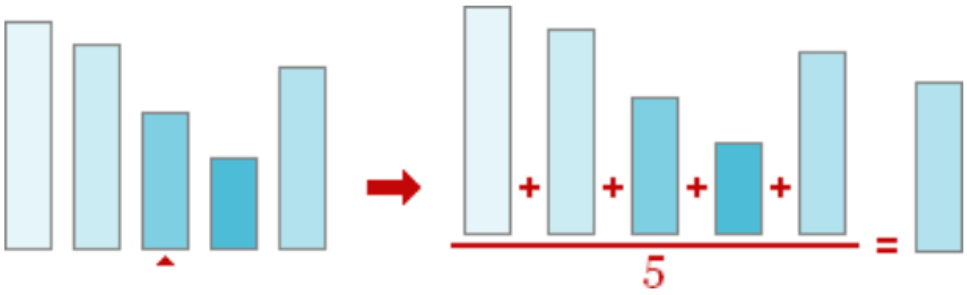


Εικόνα 4.6: *Παράδειγμα πολλαπλής εφαρμογής φίλτρου διάμεσης τιμής*.

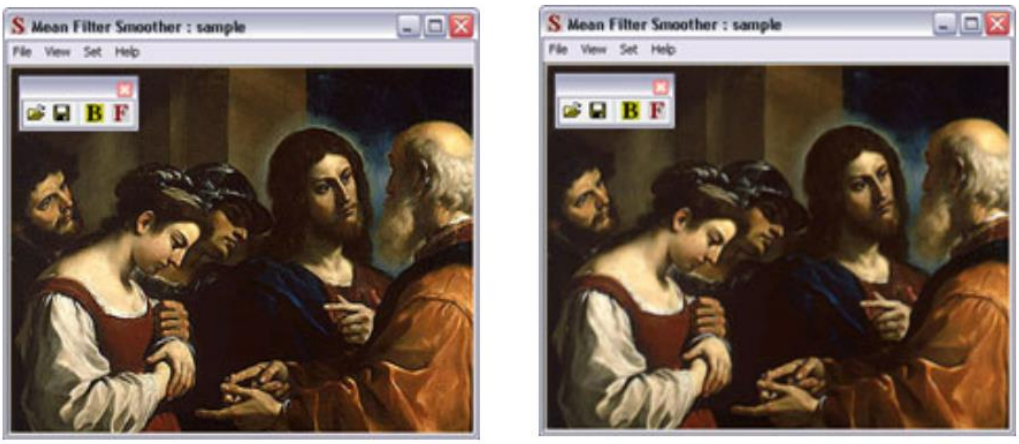
Περισσότερες πληροφορίες παρουσιάζονται στα **[38]**, **[39]** και **[40]**.

4.5.2 Φίλτρο της Μέσης Τιμής (Mean Filter)

Το φίλτρο του μέσου είναι μια μέθοδος εξομάλυνσης του θορύβου στις εικόνες, με την οποία μειώνεται η μεταβολή της έντασης μεταξύ γειτονικών pixels. Αποτελεί μια μέθοδος φιλτραρίσματος κυλιόμενου παραθύρου, όπως αυτή που παρουσιάστηκε παραπάνω. Η διαφορά των median και mean φίλτρων έγκειται στην τιμή, με την οποία αντικαθιστάται το κεντρικό pixel της γειτονιάς. Συγκεκριμένα στο φιλτράρισμα του μέσου, η τιμή του κεντρικού pixel αντικαθιστάται με τη μέση τιμή των pixels της γειτονιάς (συμπεριλαμβανομένης της ίδιας). Αυτό έχει ως αποτέλεσμα την εξάλειψη των τιμών, οι οποίες δεν είναι αντιπροσωπευτικές του περιβάλλοντα χώρου της εικόνας.



Εικόνα 4.7: *Παράδειγμα λειτουργίας φίλτρου μέσης τιμής.*



Εικόνα 4.8: *Παράδειγμα εφαρμογής φίλτρου μέσης τιμής.*

Το φίλτρο μέσης τιμής περιγράφεται αναλυτικά στα **[41]**, **[42]**, **[43]** και **[44]**.

ΚΕΦΑΛΑΙΟ 5

Διεξαγωγή και Αξιολόγηση Πειραματικών Αποτελεσμάτων

5.1 Αλγόριθμος Αναγνώρισης Ανθρώπινης Κίνησης με Χρήση Διανυσμάτων Δραστηριοτήτων (Activity Feature Vectors)

5.1.1 Εισαγωγή

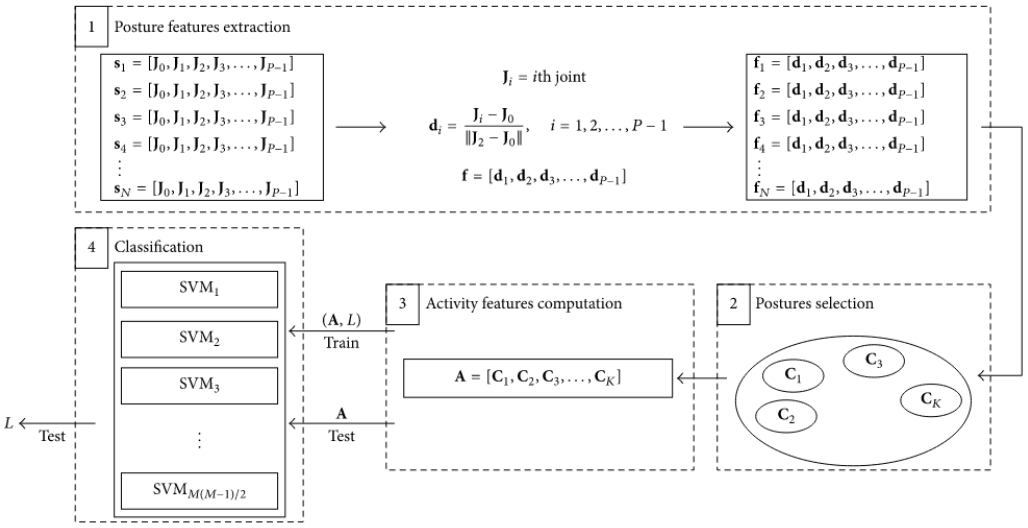
Στα πλαίσια της αξιολόγησης των σκελετικών δεδομένων, τα οποία διεξήχθησαν από τη βάση THETIS, υλοποιήθηκε και εφαρμόστηκε ο αλγόριθμος που περιγράφεται στο **[45]**. Ο συγκεκριμένος αλγόριθμος αναγνώρισης κίνησης στηρίζεται στην εύρεση κατάλληλων διανυσμάτων χαρακτηριστικών, τα οποία αντιπροσωπεύουν κάθε δραστηριότητα κίνησης. Τα παραπάνω διανύσματα χαρακτηριστικών χρησιμοποιούνται για την ταξινόμηση κάθε κίνησης εισόδου σε μια από τις υπάρχουσες κατηγορίες. Η ταξινόμηση γίνεται με χρήση SVM πολλαπλών κλάσεων. Η λειτουργία του αλγορίθμου μπορεί να διαχωριστεί σε τέσσερα βασικά στάδια:

- Εξαγωγή Χαρακτηριστικών Ανθρώπινων Στάσεων (Posture Features Extraction): Οι συντεταγμένες των σκελετικών αρθρώσεων χρησιμοποιούνται για την αξιολόγηση των διανυσμάτων χαρακτηριστικών που αντιπροσωπεύουν τις ανθρώπινες στάσεις.

- Επιλογή Ανθρώπινων Στάσεων (Postures Selection): Οι σημαντικότερες και πιο αντιπροσωπευτικές ανθρώπινες στάσεις επιλέγονται για κάθε δραστηριότητα.

- Υπολογισμός Χαρακτηριστικών Δραστηριοτήτων (Activity Features Computation): Για την ταξινόμηση κάθε κίνησης εισόδου χρησιμοποιούνται κατάλληλα διανύσματα χαρακτηριστικών, τα οποία αντιπροσωπεύουν κάθε κατηγορία δραστηριοτήτων.

- Ταξινόμηση: Το στάδιο της ταξινόμησης πραγματοποιείται χρησιμοποιώντας ένα SVM πολλαπλών κλάσεων με την προσέγγιση της μεθόδου one-versus-one.



Εικόνα 5.1: *Σύνοψη των τεσσάρων βασικών σταδίων του αλγορίθμου ανθρώπινης κίνησης με χρήση διανυσμάτων δραστηριότητας.*

5.1.2 Περιγραφή Αλγορίθμου Αναγνώρισης Κίνησης με Χρήση Διανυσμάτων Δραστηριοτήτων

Δεδομένα εισόδου του αναφερόμενου αλγορίθμου αποτελούν οι συντεταγμένες των σκελετικών αρθρώσεων του αντικειμένου που κινείται στο χώρο. Σύμφωνα με τους συγγραφείς του **[45]**, η παραπάνω μορφή δεδομένων προτιμήθηκε, διότι διευκολύνει την εύρεση μιας συμπαγής αναπαράστασης του ανθρώπινου σώματος. Κάθε άρθρωση του 3D σκελετού εισόδου αντιπροσωπεύεται από ένα τρισδιάστατο διάνυσμα Ji στο χώρο συντεταγμένων του Kinect, μέσω του οποίου γίνεται η λήψη των δεδομένων. Ο εξεταζόμενος αλγόριθμος υπολογίζει τα κατάλληλα χωρικά χαρακτηριστικά μέσω των προαναφερόμενων συντεταγμένων. Σε αυτόν τον υπολογισμό δε συμπεριλαμβάνονται οι σχετιζόμενες με το χρόνο πληροφορίες, προκειμένου το σύστημα να είναι ανεξάρτητο από την ταχύτητα με την οποία εκτελείται κάθε κίνηση. Επισημαίνεται ότι ένα άτομο μπορεί να βρεθεί σε οποιοδήποτε σημείο της περιοχής κάλυψης του Kinect, με αποτέλεσμα μια άρθρωση να λαμβάνει πολλές διαφορετικές συντεταγμένες. Μια μέθοδος αντιμετώπισης αυτού του φαινομένου είναι η αναπροσαρμογή των σκελετικών συντεταγμένων, θέτοντας ως σημείο αρχής μία από τις αρθρώσεις του δεδομένου σκελετού.

Θεωρώντας ότι ένας σκελετός αποτελείται από P αρθρώσεις, αν J0 είναι το διάνυσμα συντεταγμένων του κορμού (torso) και J2 το διάνυσμα συντεταγμένων του λαιμού, για κάθε άρθρωση i μπορεί να υπολογιστεί η απόσταση di ως έξης:

$$d_i = \frac{J_i - J_0}{||J_2 - J_0||}, i = 1,2, \ldots, P-1 \qquad (5.1)$$

Σύμφωνα με τον παραπάνω τύπο, κάθε απόσταση di μετράται από το σημείο του κορμού και ομαλοποιείται από την απόσταση των J0, J2 διανυσμάτων. Με αυτόν τον τρόπο η υπολογιζόμενη απόσταση δεν επηρεάζεται από τη θέση του σκελετού μέσα στην περιοχή κάλυψης του Kinect, ενώ η δομή του σκελετού παραμένει σταθερή λόγω της κανονικοποίησης. Τα μεγέθη di αποτελούν ένα σύνολο διανυσμάτων που συνδέουν κάθε άρθρωση με την άρθρωση του κορμού του σκελετού. Για κάθε πλαίσιο (frame) του σκελετού δημιουργείται ένα διάνυσμα χαρακτηριστικών στάσης (posture feature vector) f:

$$f = [d_1, d_2, d_3, \dots, d_{P-1}] \qquad (5.2)$$

Επομένως, για μια δραστηριότητα αποτελούμενη από N σκελετικά πλαίσια υπολογίζεται ένα σύνολο από N διανύσματα χαρακτηριστικών στάσης.

Μετά τον υπολογισμό των παραπάνω διανυσμάτων είναι απαραίτητη η επιλογή των πιο αντιπροσωπευτικών ανθρώπινων στάσεων, με στόχο τη μείωση της πολυπλοκότητας και την εκπροσώπηση της δραστηριότητας μέσω μόνο ενός υποσυνόλου στάσεων, χωρίς τη χρήση όλων των πλαισίων. Για την κατάλληλη επιλογή των στάσεων μπορεί να αξιοποιηθεί ένας αλγόριθμος ομαδοποίησης, ο οποίος ταξινομεί τα διανύσματα χαρακτηριστικών σε ομάδες. Στη συγκεκριμένη περίπτωση επιλέγεται ο αλγόριθμος των μέσων (k-means), ο οποίος, υπολογίζοντας την τετραγωνική ευκλείδεια απόσταση, μπορεί να χρησιμοποιηθεί για την ομαδοποίηση των πλαισίων που αντιπροσωπεύουν παρόμοιες στάσεις. Λαμβάνοντας υπόψη μια δραστηριότητα που αποτελείται από διανύσματα χαρακτηριστικών στάσεων, ο αλγόριθμος των μέσων δίνει σαν έξοδο αναγνωριστικά ID (ένα για κάθε διάνυσμα χαρακτηριστικών) και K διανύσματα [C1,C2,C3,...,CK] που αντιπροσωπεύουν τα κέντρα κάθε ομάδας. Τα διανύσματα χαρακτηριστικών χωρίζονται σε συστοιχίες S1, S2,..., SK έτσι ώστε να ικανοποιούν τη σχέση:

$$\arg min_s \ \sum_{j=1}^{K} \sum_{f_i \in S_j} ||f_i - C_j||^2 \qquad (5.3)$$

Τα K κέντρα μπορούν να θεωρηθούν οι κύριες στάσεις της δραστηριότητας, οι οποίες αποτελούν τα σημαντικότερα διανύσματα. Επομένως ο αλγόριθμος ομαδοποίησης εκτελείται για κάθε ακολουθία.

Το επόμενο στάδιο του αλγορίθμου σχετίζεται με τον υπολογισμό ενός διανύσματος χαρακτηριστικών, το οποίο περιγράφει ολόκληρη τη δραστηριότητα (Activity Feature Vector). Το παραπάνω διάνυσμα διαμορφώνεται με την κατάλληλη ταξινόμηση των K κέντρων, τα οποία υπολογίστηκαν κατά την ομαδοποίηση του προηγούμενου σταδίου. Συγκεκριμένα, τα K διανύσματα C1,C2,C3,...,CK ταξινομούνται λαμβάνοντας υπόψη τη σειρά με την οποία εμφανίζονται τα στοιχεία του συμπλέγματος κατά τη διάρκεια της δραστηριότητας. Το διάνυσμα κάθε δραστηριότητας αποτελείται από τη σύζευξη των ταξινομημένων κέντρων Ci. Για παράδειγμα, λαμβάνοντας υπόψη μια δραστηριότητα που χαρακτηρίζεται από N=10 και K=4, μετά την εκτέλεση του αλγορίθμου ομαδοποίησης, μία από τις πιθανές εξόδους μπορεί να είναι η ακολουθία αναγνωριστικών ID: [2,2,2,3,3,1,1,4,4,4]. Από την παραπάνω ακολουθία γίνεται κατανοητό ότι τα πρώτα τρία διανύσματα στάσης ανήκουν στο σύμπλεγμα 2, τα επόμενα δύο διανύσματα σχετίζονται με το σύμπλεγμα 3 κτλ. Σε αυτή την περίπτωση, το διάνυσμα δραστηριοτήτων είναι το A=[C2,C3,C1,C4]. Ένα διάνυσμα δραστηριοτήτων έχει μια διάσταση μεγέθους 3K(P-1), η οποία μπορεί να αντιμετωπιστεί χωρίς να χρησιμοποιηθούν αλγόριθμοι μείωσης διαστάσεων (για παράδειγμα PCA) εάν το K είναι μικρό.

Πρέπει να σημειωθεί ότι στα πλαίσια της διπλωματικής εργασίας, κατά την υλοποίηση του συγκεκριμένου σταδίου του αλγορίθμου έγινε μια παραδοχή, η οποία δε διευκρινίζεται στο **[45]**. Συγκεκριμένα κατά τη δημιουργία ενός διανύσματος δραστηριοτήτων, για την ταξινόμηση των κέντρων Ci λαμβάνεται υπόψη εκτός από τη σειρά που εμφανίζονται τα αναγνωριστικά, και το πλήθος τους, αν αυτά εμφανίζονται παραπάνω από μια φορά σε μη διαδοχικές θέσεις. Για παράδειγμα, για μια δραστηριότητα με N=10 και K=4, αν ο αλγόριθμος ομαδοποίησης δίνει αποτέλεσμα την ακολουθία αναγνωριστικών ID [1,2,2,1,1,1,3,4,4,4], τότε το διάνυσμα δραστηριοτήτων είναι το [2,1,3,4] και όχι το [1,2,3,4]. Επομένως κριτήριο της ταξινόμησης των κέντρων Ci είναι το πλήθος των διαδοχικών θέσεων που εμφανίζεται ένα αναγνωριστικό και η θέση στην οποία εμφανίζεται σε σχέση με τα υπόλοιπα αναγνωριστικά.

Το τελευταίο βήμα του αλγορίθμου αφορά την ταξινόμηση των παραγόμενων διανυσμάτων δραστηριοτήτων στη σωστή κατηγορία κινήσεων. Η παραπάνω λειτουργία επιτυγχάνεται με τη χρήση ενός SVM. Αν υπάρχουν l διανύσματα εκπαίδευσης xi∈Rn και ένα διάνυσμα y∈Rl με ετικέτες yi∈[-1,1], τότε ένα SVM μπορεί να διατυπωθεί ως :

$$min_{w,b.\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \qquad (5.4)$$

$$y_i(w^T \varphi(x_i) + b) \geq 1 - \xi_i \qquad (5.5)$$

$$\xi_i \geq 0, i = 1, \dots, l \qquad (5.6)$$

όπου

$$w^T \varphi(x) + b = 0 \qquad (5.7)$$

είναι το βέλτιστο υπερεπίπεδο που επιτρέπει τον διαχωρισμό μεταξύ τάξεων στο χώρο των χαρακτηριστικών, C είναι μια σταθερά και ξi είναι μη αρνητικές μεταβλητές που σχετίζονται με σφάλματα εκπαίδευσης. Η συνάρτηση φ επιτρέπει τη μετατροπή του χώρου χαρακτηριστικών σε ένα χώρο υψηλότερων διαστάσεων, όπου τα δεδομένα είναι διαχωρίσιμα. Λαμβάνοντας υπόψη δύο διανύσματα εκπαίδευσης xi και xj, η συνάρτηση πυρήνα μπορεί να οριστεί ως:

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \qquad (5.8)$$

Στη συγκεκριμένη περίπτωση χρησιμοποιείται η συνάρτηση πυρήνα Radial Basis (RBF):

$$K(x_i, x_j) = e^{-\gamma \left|\left|x_i - x_j\right|\right|^2}, \gamma = \frac{1}{2\sigma^2} > 0 \qquad (5.9)$$

Οι παράμετροι C και γ πρέπει να εκτιμηθούν πριν από τη χρήση του SVM.

Όπως αναφέρθηκε, ο αλγόριθμος αξιοποιεί ένα SVM πολλαπλών κλάσεων, όπου κάθε κλάση αντιπροσωπεύει μια δραστηριότητα του συνόλου δεδομένων. Προκειμένου να επεκταθεί ο ρόλος του SVM από δυαδικό ταξινομητή σε ταξινομητή πολλαπλών κλάσεων χρησιμοποιείται η προσέγγιση one-versus-one. Αυτή η μέθοδος βασίζεται στην κατασκευή πολλών δυαδικών ταξινομητών SVM. Συγκεκριμένα M(M-1)/2 δυαδικά SVMs είναι απαραίτητα για την ταξινόμηση σε ένα σύνολο δεδομένων M κατηγοριών. Αυτό συμβαίνει επειδή κάθε SVM εκπαιδεύεται για να διακρίνει μεταξύ δύο τάξεων και η τελική απόφαση λαμβάνεται μέσω μιας στρατηγικής ψηφοφορίας μεταξύ όλων των δυαδικών ταξινομητών. Κατά τη διάρκεια της φάσης εκπαίδευσης, τα διανύσματα δραστηριότητας μαζί με τις αντίστοιχες ετικέτες L δραστηριοτήτων δίνονται ως είσοδοι στο SVM. Στη φάση της αξιολόγησης, η ετικέτα δραστηριότητας υπολογίζεται από το ίδιο το SVM.

5.1.3 Υλοποίηση Αλγορίθμου Αναγνώρισης με Χρήση Διανυσμάτων Δραστηριότητας

Ο προαναφερόμενος αλγόριθμος υλοποιήθηκε σε γλώσσα προγραμματισμού matlab. Επισημαίνεται ότι κατά την υλοποίηση του αλγορίθμου το βασικό τμήμα της επεξεργασίας

των σκελετικών δεδομένων δε διαφοροποιείται από αυτό που περιγράφεται παραπάνω, ωστόσο ακολουθείται διαφορετική διαδικασία ταξινόμησης και αξιολόγησης των κινήσεων εισόδου.

Πιο αναλυτικά κάθε στάδιο της επεξεργασίας των διανυσμάτων αναπτύσσεται ως διαφορετική συνάρτηση:

- **ReadClassFiles.m:** Διαβάζει τα δεδομένα από όλα τα αρχεία ενός φακέλου/κίνησης και επιστρέφει τη λίστα με τα δεδομένα και τις αντίστοιχες ετικέτες (labels). Το παραπάνω script καλεί τη συνάρτηση ReadData, η οποία αναλύεται στη συνέχεια.

- **ReadData.m:** Παίρνει το path και το όνομα ενός αρχείου και διαβάζει τα δεδομένα του, τα οποία επιστρέφει σε έναν cell array μαζί με το πλήθος των frames του συγκεκριμένου αρχείου. Χρησιμοποιείται για να διαβάζει κάθε αρχείο των εξαγόμενων σκελετικών δεδομένων.

- **JointFeatures.m:** Για δεδομένες 3D συντεταγμένες (x,y,z) μιας άρθρωσης Ji, υπολογίζει την απόσταση di της άρθρωσης από τον κορμό (torso).

- **PostureVector.m:** Βρίσκει τις αποστάσεις di όλων των αρθρώσεων ενός frame και τις αποθηκεύει σε έναν πίνακα.

- **FeatureVectors.m:** Υπολογίζει τα διανύσματα χαρακτηριστικών fi (Feature Vectors) των αρθρώσεων για όλα τα frames, χρησιμοποιώντας το script PostureVector.m.

- **ActivityFeature.m:** Δημιουργεί τα διανύσματα δραστηριότητας. Αρχικά υπολογίζει τα fi καλώντας το FeatureVector.m και κάνει ομαδοποίηση μέσω του αλγορίθμου k-means. Το πλήθος των κέντρων Ci ορίζεται από τον χρήστη μέσω μιας παραμέτρου NoClusters.

- **sortCenters.m:** Για μια δεδομένη ακολουθία από διανύσματα στάσεων (Posture Feature Vectors) δημιουργεί το διάνυσμα δραστηριοτήτων της δεδομένης κίνησης υπολογίζοντας τη σειρά ταξινόμησης των κέντρων Ci.

- **split_input.m:** Διαχωρίζει τα δεδομένα μαζί με τις αντίστοιχες ετικέτες σε σύνολα training set και test set για την εκπαίδευση των SVMs.

Κατά την πειραματική διαδικασία εφαρμόζεται το πρωτόκολλο αξιολόγησης leave-one-person-out σε 12 SVMs. Κάθε SVM αντιστοιχεί σε μια κατηγορία κινήσεων της βάσης δεδομένων THETIS. Η εκπαίδευση των SVMs γίνεται με χρήση συνάρτησης πυρήνα Radial Basis (RBF), ενώ οι παράμετροι c και σ έχουν υπολογιστεί δοκιμαστικά. Οι διαστάσεις των διανυσμάτων δραστηριότητας περιορίζονται μέσω της τεχνικής PCA και υπολογίζονται πειραματικά για την καλύτερη απόδοση της ταξινόμησης. Περισσότερες πληροφορίες για το σύστημα πειραματικής αξιολόγησης αναφέρονται στο κεφάλαιο 2.

5.1.4 Πειραματική αξιολόγηση

Στον πίνακα 5.1 παρουσιάζονται τα αποτελέσματα της πειραματικής εφαρμογής του αλγορίθμου. Σημειώνονται οι τιμές των παραμέτρων c, Sigma, Length στις οποίες παρατηρήθηκε η καλύτερη απόδοση ταξινόμησης. Το πεδίο Length αναφέρεται στο τελικό μήκος των διανυσμάτων δραστηριότητας (activity feature vectors) μετά από την εφαρμογή της τεχνικής PCA.

Movement	c	Sigma	Length	Performance
Backhand	10	4	23	0,9549
Backhand with two hands	10	5	27	0,9775
Backhand slice	1	3	26	0,9275
Backhand volley	1	3	26	0,9388
Forehand flat	10	6	57	0,9308
Forehand open	1	4	41	0,9678
Forehand slice	1	2	22	0,9147
Service flat	1	2	21	0,9275
Service kick	1	2	29	0,9275
Service slice	1	2	29	0,9163
Smash	10	3	31	0,9195
Volley	10	6	45	0,9356

Πίνακας 5.1: *Αποτελέσματα απόδοσης του εξεταζόμενου αλγορίθμου στο σύνολο των εξαγόμενων σκελετικών δεδομένων.*

Ο μέσος όρος καλύτερης απόδοσης ταξινόμησης για όλες τις κλάσεις είναι 93.65%.

Ο πίνακας 5.2 περιλαμβάνει τις αστοχίες στην πρόβλεψη των SVMs για κάθε κατηγορία κίνησης.

	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Class10	Class11	Class12
Class1	19	0	3	1	2	0	0	0	0	0	0	0
Class2	2	12	0	1	2	0	0	0	0	0	0	0
Class3	2	0	30	0	1	0	0	0	0	0	0	14
Class4	0	2	0	33	0	1	0	0	0	0	0	0
Class5	1	0	0	1	15	0	0	0	0	0	0	0
Class6	0	0	0	6	0	42	0	0	0	0	13	0
Class7	0	0	0	0	0	0	44	0	0	0	0	0
Class8	1	0	0	0	0	0	0	45	0	2	0	0
Class9	0	0	0	0	0	0	0	0	52	1	0	0
Class10	1	0	0	0	0	0	1	0	0	47	0	0
Class11	1	0	0	1	0	10	0	0	0	0	26	0
Class12	1	0	12	0	0	0	0	0	0	0	1	24

Πίνακας 5.2: *Πίνακας αστοχιών πρόβλεψης SVM των 12 κινήσεων του tennis (Class1: Backhand, Class2: Backhand with two hands, Class3: Backhand slice, Class4: Forehand flat, Class5: Forehand Open, Class6: Forehand slice, Class7: Service flat, Class8: Service kick, Class9: Service slice, Class10: Smash, Class11: Volley, Class12: Backhand volley).*

Τέλος γίνεται σύγκριση των παραπάνω αποτελεσμάτων με την επίδοση των αλγορίθμων (πίνακας 5.3), οι οποίοι αναφέρονται στα κεφάλαια 2 και 3 (3D CTT, HOG, HOF, MBH). Επισημαίνεται ότι οι συγκεκριμένες μέθοδοι εφαρμόζονται στα υπάρχοντα σκελετικά δεδομένα της THETIS, ενώ ο υλοποιημένος αλγόριθμος αναγνώρισης αυτού του

κεφαλαίου (Method based on Activity Feature Vectors) εφαρμόζεται στα εξαγόμενα σκελετικά δεδομένα.

Όπως παρατηρείται, υψηλότερη απόδοση ταξινόμησης παρουσιάζει η μέθοδος των διανυσμάτων δραστηριότητας (93.65%), ενώ ακολουθεί η μέθοδος του 3D CTT. Οι υλοποιήσεις των HOG/ HOF/ MBH περιγραφέων έχουν τη χαμηλότερη απόδοση που κυμαίνεται από 46.84% έως 54.40%.

Algorithm	Average Performance (%)
STIP based method with HOG/HOF descriptors	54.40
Dense Trajectory method with Trajectory Descriptor	46.84
Dense Trajectory method with MBH Descriptor	50.78
Dence Trajectory with HOG/HOF/MBH/Trajectory Descriptor	53.08
3D CTT - SSTIP based VTFs	86.06
Method Based on Activity Feature Vectors	93.65

Πίνακας 5.3: *Απόδοση αλγορίθμων πάνω σε σκελετικά δεδομένα.*

5.2 Προσδιορισμός Επιπέδου Εμπειρίας Παικτών Tennis με Χρήση Περιγραφέων Απόκλισης.

5.2.1 Εισαγωγή

Τα εξαγόμενα σκελετικά δεδομένα εφαρμόζονται επίσης σε έναν αλγόριθμο **[46]**, ο οποίος πραγματεύεται το επίπεδο εμπειρίας των παικτών της βάσης THETIS. Ο συγκεκριμένος αλγόριθμος χρησιμοποιείται για να αναγνωρίσει αν ένας παίκτης είναι αρχάριος ή έμπειρος, αναλύοντας τις κινήσεις εισόδου. Τα βασικότερα τμήματα της μεθόδου περιγράφονται στη συνέχεια.

5.2.2 Εύρεση Σημείων Ενδιαφέροντος (Selective Spatio-Temporal Interest Points)

Ο εξεταζόμενος αλγόριθμός χρησιμοποιεί επιλεκτικά χωροχρονικά σημεία ενδιαφέροντος (Selective Spatio-Temporal Interest Points / SSTIPs). Αυτή η επιλογή είναι επιθυμητή, καθώς αξιοποιούνται πληροφορίες της κίνησης σε όλο τον χώρο και όχι τοπικά. Με αυτόν τον τρόπο αποφεύγεται η λανθασμένη ανίχνευση σημείων ενδιαφέροντος του περιβάλλοντα χώρου, η οποία οφείλεται στην κίνηση της κάμερας. Η διαδικασία εξαγωγής των SSTIPs περιλαμβάνει:

- Τον εντοπισμό των STIPs μέσω κατάλληλου περιγραφέα.
- Την απόρριψη ανεπιθύμητων σημείων του περιβάλλοντα χώρου.
- Την επιβολή χωρικών και χρονικών περιορισμών για τη δημιουργία του τελικού συνόλου των STIPs.

Η ιδέα της εξαγωγής των SSTIPs στηρίζεται στη παρατήρηση ότι τα σημεία του περιβάλλοντα χώρου ακολουθούν γεωμετρικά μοτίβα, ενώ τα σημεία που ανήκουν στο κινούμενο υποκείμενο δε φέρουν αυτή την ιδιότητα.

5.2.3 Υπολογισμός Περιγραφέων (Variance-based Descriptors on STIPs)

Το τελικό σύνολο των επιλεγμένων σημείων χρησιμοποιείται ως είσοδο για τον υπολογισμό των διανυσμάτων απόκλισης (Variance Vector) και απόστασης συνημιτόνου (Cosine Distance Vector). Πιο αναλυτικά, για μια ακολουθία κίνησης μήκους N πλαισίων τα παραπάνω ορίζονται ως εξής:

Variance Vector $V=[V_i \,|i=1..N]$:

$$V_i = \frac{1}{M_i}\sum_{j=1}^{M_i}\left(p_i^j - \mu_i\right)\left(p_i^j - \mu_i\right)^T \qquad (5.10)$$

Cosine Distance Vector $D=[D_i \,|i=1..N]$:

$$D_i = \frac{1}{M_i}\sum_{j=1}^{M_i} p_i^j \mu_i^T \qquad (5.11)$$

όπου i είναι ο αριθμός του πλαισίου, M_i είναι το πλήθος των σημείων ενδιαφέροντος του i πλαισίου, p_i^j είναι το j σημείο ενδιαφέροντος και μ_i είναι το μέσο σημείο. Ο υπολογισμός των δύο παραπάνω μεγεθών έγινε για να εξεταστεί η απόδοση της απόκλισης και της απόστασης συνημιτόνου ως περιγραφείς.

5.2.4 Μέθοδος Dynamic Time Warping (DTW)

Για τη συσχέτιση και την ευθυγράμμιση των χρονικών ακολουθιών των σημείων ενδιαφέροντος χρησιμοποιείται η μέθοδος Dynamic Time Wrapping (DTW). Η παραπάνω τεχνική αποτελεί έναν αλγόριθμο δυναμικού προγραμματισμού, ο οποίος υπολογίζει την ομοιομορφία μεταξύ δύο χρονικών αλληλουχιών που μπορεί να διαφέρουν σε ταχύτητα και μήκος. Για παράδειγμα ομοιότητες μπορεί να εντοπιστούν στο περπάτημα δύο ατόμων, τα οποία όμως κινούνται με διαφορετική ταχύτητα. Επισημαίνεται ότι ο αλγόριθμος μπορεί να χρησιμοποιηθεί γενικά για την εύρεση της βέλτιστης αντιστοιχίας μεταξύ δύο οποιοδήποτε δεδομένων ακολουθιών, θέτοντας κάποιους περιορισμούς. Στη συγκεκριμένη περίπτωση χρησιμοποιείται, γιατί το μήκος των ακολουθιών STIPs ποικίλλει, ενώ κατά την πειραματική αξιολόγηση χρησιμοποιούνται k-NN ταξινομητές που απαιτούν σταθερό μήκος διανυσμάτων. Στην ταξινόμηση k-NN, η έξοδος είναι μέλος μιας κατηγορίας κινήσεων. Ένα αντικείμενο ταξινομείται σε μια ομάδα δραστηριοτήτων με βάση την πλειοψηφία των k γειτόνων του. Συνεπώς το αντικείμενο ανατίθεται στην κατηγορία που είναι πιο συνηθισμένη στους k πλησιέστερους γείτονές του. Αν k = 1, τότε το αντικείμενο απλώς αποδίδεται στην κλάση του μοναδικού πλησιέστερου γείτονα του.

5.2.5 Πειραματική Αξιολόγηση

Κατά τη διεξαγωγή των πειραμάτων χρησιμοποιείται το πρωτόκολλο leave-one-person-out. Σε ένα υποθετικό σενάριο πραγματικής λειτουργίας, η δραστηριότητα ενός άγνωστου ατόμου καταγράφεται από ένα σύστημα αναγνώρισης κίνησης. Στη συνέχεια, η

δραστηριότητα εισόδου επεξεργάζεται και συγκρίνεται με ένα προκατασκευασμένο σύνολο δεδομένων, το οποίο έχει χρησιμοποιηθεί για την εκπαίδευση του συστήματος αναγνώρισης. Η ταξινόμηση της καταγεγραμμένης δραστηριότητας καθορίζεται με βάση τη συνάφεια της σε σύγκριση με οποιοδήποτε δείγμα δεδομένων που περιλαμβάνει το σύνολο εκπαίδευσης, σύμφωνα με τους ειδικούς κανόνες του συστήματος. Συνεπώς, το πρωτόκολλο leave-one-person-out χρησιμοποιεί δείγματα δράσης ενός ατόμου για έλεγχο, ενώ τα υπόλοιπα δείγματα αποτελούν το σύνολο εκπαίδευσης. Αυτή η διαδικασία επαναλαμβάνεται N φορές, όπου N είναι ο αριθμός των ατόμων, τα οποία περιλαμβάνονται στο σύνολο των δεδομένων. Η μέση απόδοση A μετριέται ως:

$$A = \frac{1}{N}\sum_{i=1}^{N} A(s_i) \qquad (5.12)$$

όπου A(si) είναι η απόδοση των κινήσεων ενός ατόμου si, όταν αυτό επιλέγεται για έλεγχο.

Τα δεδομένα της βάσης THETIS χρησιμοποιούνται για την εξαγωγή των διανυσμάτων V και D. Επειδή το μήκος των videos και των αντίστοιχων χρονικών ακολουθιών ποικίλλει, εφαρμόζεται η μέθοδος DTW και τα αποτελέσματα προωθούνται σε k-NN ταξινομητές. Κατά τη διαδικασία αξιολόγησης εφαρμόζονται δύο πειραματικά σενάρια. Στο πρώτο σενάριο υπολογίζεται η ελάχιστη απόσταση μεταξύ δύο ακολουθιών μέσω του DTW και έπειτα οι ταξινομητές συγκρίνουν τις ακολουθίες μέσω αυτού του μεγέθους. Στο δεύτερο σενάριο, γίνεται τροποποίηση της διαδικασίας για να γίνει ευθυγράμμιση των χρονικών ακολουθιών. Έπειτα οι ταξινομητές συγκρίνουν τις ακολουθίες χρησιμοποιώντας ως μετρική την Ευκλείδεια απόσταση.

Καθώς ο εξεταζόμενος αλγόριθμος μελετά το επίπεδο εμπειρίας των παικτών, το πρόβλημα της κατηγοριοποίησης μετατρέπεται σε διαδικασία δυαδικής ταξινόμησης μιας κίνησης ανάμεσα σε έμπειρους και ερασιτέχνες παίκτες. Όπως προαναφέρεται, για κάθε πείραμα εκπαιδεύεται ένας k-NN ταξινομητής χρησιμοποιώντας ως σύνολο εκπαίδευσης όλες τις κινήσεις της συγκεκριμένης κατηγορίας, εκτός από αυτές που εκτελεί το άτομο προς εξέταση. Η ταξινόμηση επαναλαμβάνεται για όλους τους παίκτες.

Στους πίνακες 5.4 και 5.5 γίνεται σύγκριση των πειραματικών αποτελεσμάτων της μεθόδου, η οποία εφαρμόστηκε στα SSTIPs καθώς και στα σκελετικά δεδομένα, τα οποία διεξήχθησαν στα πλαίσια της διπλωματικής εργασίας. Η σύγκριση αναφέρεται μόνο στα διανύσματα χαρακτηριστικών V για τα δύο προαναφερόμενα πειραματικά σενάρια αξιολόγησης. Παρατηρείται ότι η απόδοση είναι παρόμοια στα δύο είδη δεδομένων εισόδου.

Movement	STIPS		Skeletal Joints	
	Accuracy (%)	Neighbors	Accuracy (%)	Neighbors
Backhand	74.55	4	79.59	2
Backhand with two hands	70.91	19	61.11	6
Backhand slice	64.85	2	68.52	2
Backhand volley	69.09	8	60.42	15
Forehand flat	69.09	43	53.57	18
Forehand open stands	68.48	26	65.45	8
Forehand slice	66.06	9	70.91	3
Forehand volley	66.06	25	71.7	1
Service flat	63.64	33	63.83	28
Service kick	72.73	30	62.22	14
Service slice	67.88	24	61.54	6
Smash	64.85	23	58.49	2

Πίνακας 5.4: *Απόδοση διανυσμάτων απόκλισης V με χρήση DTW.*

Movement	STIPS		Skeletal Joints	
	Accuracy (%)	Neighbors	Accuracy (%)	Neighbors
Backhand	64.85	12	77.55	5
Backhand with two hands	66.06	28	62.96	9
Backhand slice	69.09	43	55.56	42
Backhand volley	63.64	60	64.58	20
Forehand flat	58.18	4	53.57	42
Forehand open stands	68.48	1	67.27	8
Forehand slice	61.21	48	54.55	32
Forehand volley	60.61	68	62.26	4
Service flat	63.03	17	61.7	4
Service kick	70.91	3	64.44	2
Service slice	66.67	2	59.62	2
Smash	63.03	15	62.26	2

Πίνακας 5.5: *Απόδοση διανυσμάτων απόκλισης V σε χρονικά ευθυγραμμισμένες ακολουθίες με χρήση Ευκλείδειας απόστασης.*

ΒΙΒΛΙΟΓΡΑΦΙΑ

[1] Nogueira, Pedro (2011). "Motion Capture Fundamentals: A Critical and Comparative Analysis on Real-World Applications". In: Programa Doutoral em Engenharia Informática, Instituto de Telecomunicações.

[2] Poppe, Ronald (2007). "Vision-based human motion analysis: An overview". In: Computer Vision and Image Understanding 108.1-2, pp. 4–18.

[3] Moeslund, Thomas B. and Erik Granum (2001). "A Survey of Computer Vision- Based Human Motion Capture". In: Computer Vision and Image Understanding - Modeling people toward vision-based underatanding of a person's shape, appearance, and movement 81.3, pp. 231–268.

[4] Aggarwal, J. K. and M. S. Ryoo (2011). "Human activity analysis: A review". In: ACM Computing Surveys 43.3.

[5] Allen, J. F. (1983). "Maintaining knowledge about temporal intervals". In: Communications of the ACM 26.11, 832–843.

[6] Aggarwal, J.K. and Lu Xia (2014). "Human activity recognition from 3D data: A review". In: Pattern Recognition Letters 48, pp. 70–80.

[7] Laurentini, Aldo (1995). "How far 3D shapes can be understood from 2D silhouettes". In: IEEE Transactions on pattern analysis and machine intelligence 17.2.

[8] Napoleon, Thibault and Hichem Sahbi (2010). "From 2D Silhouettes to 3D Object Retrieval: Contributions and Benchmarking". In: EURASIP Journal on Image and Video Processing.

[9] WANG, ULLAH, KLÄSER, LAPTEV, SCHMID (2009). "Evaluation of local spatio-temporal features for action recognition". In: British Machine Vision Conference, BMVC.

[10] Warren, David H. and Edward R. Strelow (1985). In: Electronic Spatial Sensing for the Blind: Contributions from Perception.

[11] Burton, Andrew and John Radford (1978). In: Thinking in Perspective: Critical Essays in the Study of Thought Processes.

[12] J.L.Barron and N.A.Thacker (2004). "Tutorial: Computing 2D and 3D Optical Flow". In: Features and Measurement 2004-012.

[13] Mao Ye Qing Zhang, Liang Wang-Jiejie Zhu Ruigang Yang Juergen Gall (2013). "A Survey on Human Motion Analysis from Depth Data". In: Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications, pp. 149–187.

[14] Jungong Han Ling Shao, Dong Xu-Jamie Shotton (2013). "Enhanced Computer Vision with Microsoft Kinect Sensor: A Review". In: IEEE TRANSACTIONS ON CYBERNETICS 43.5.

[15] Jamie Shotton Ross Girshick, Andrew Fitzgibbon-Toby Sharp Mat Cook Mark Finocchio Richard Moore Pushmeet Kohli Antonio Criminisi Alex Kipman Andrew Blake (2012).

"Efficient Human Pose Estimation from Single Depth Images". In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013 35.12, pp. 2821 –2840.

[16] OpenNI TM. "OpenNI User Guide". URL: https://github.com/OpenNI/OpenNI/blob/master/Documentation/OpenNI_UserGuide.pdf

[17] Prime Sensor TM. "NITE 1.3 Algorithms Notes". URL: http://pr.cs.cornell.edu/humanactivities/data/NITE.pdf.

[18] C. Schuldt I. Laptev, B. Caputo (2004). "Recognizing human actions: A local svm approach". In: In Proc. ICPR, pp. 32–36.

[19] V. Bloom, D. Makris and V. Argyriou (2012). "G3d: A gaming action dataset and real time action recognition evaluation framework". In: Computer Vision and Pattern RecognitionWorkshops (CVPRW), 2012 IEEE Computer Society Conference on, pp. 7– 12.

[20] M. Marszałek I. Laptev, C. Schmid (2009). "Actions in context". In: IEEE Conference on Computer Vision and Pattern Recognition.

[21] Sofia Gourgari Georgios Goudelis, Konstantinos Karpouzis-Stefanos Kollias (2013). "THETIS: Three Dimensional Tennis Shots a Human Action Dataset". In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on.

[22] OpenNI TM (2013). "The standard framework for 3D sensing". In: http://www.openni.org

[23] I. Laptev M. Marszałek, C. Schmid-B. Rozenfeld (2008). "Learning realistic human actions from movies". In: Conference on Computer Vision and Pattern Recognition (CVPR).

[24] H.Wang A. Klaser, C. Schmid C.-L. Liu (2011). "Action recognition by dense trajectories". In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE ,Colorado Springs, US, 3169–3176.

[25] N. Dalal B. Triggs, C. Schmid (2006). "Human detection using oriented histograms of flow and appearance". In: Computer Vision–ECCV, pp. 428–441.

[26] Sivic, J. and A. Zisserman (2003). "Video Google: A text retrieval approach to object matching in videos". In: Proceedings of the International Conference on Computer Vision 2, 1470–1477.

[27] Cortes, C. and V. Vapnik (1995). "Support-vector network". In: Machine Learning 20, pp. 273–297.

[28] G. Goudelis K. Karpouzis, S. Kollias (2013). "Exploring trace transform for robust human action recognition". In: Pattern Recognition 46.12, pp. 3238–3248.

[29] Deans, S. R. (1983). "The Radon Transform and Some of Its Applications". In: Krieger Publishing Company.

[30] Kadyrov, A. and M. Petrou (2001). "The Trace transform and its applications". In: IEEE Trans. Pattern Anal. Mach. Intell. 23, pp. 811–828.

[31] Averbuch, A. and Y. Shkolnisky (2003). "3d fourier based discrete radon transform". In: Applied and Computational Harmonic Analysis 15.1, pp. 33–69.

[32] C. Yuan X. Li,W. Hu-H. Ling S. Maybank (2013). “3d R transform on spatiotemporal interest points for action recognition”. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 724–730.

[33] B. Chakraborty M. Holte, T. Moeslund-J. Gonzalez (2012). “Selective spatiotemporal interest points”. In: Computer Vision and Image Understanding 116.3, pp. 396–410.

[34] Pearson, K. (1901). “On Lines and Planes of Closest Fit to Systems of Points in Space”. In: Philosophical Magazine 2.11, pp. 559–572.

[35] Hotelling, H. (1933). “Analysis of a complex of statistical variables into principal components”. In: Journal of Educational Psychology 24, 417–441 and 498–520.

[36] Hotelling, H. (1936). “Relations between two sets of variates”. In: Biometrika 28, pp. 321–377.

[37] J. Vainstein J. Manera, P. Negri-C. Delrieux A. Maguitman (2014). “Modeling video activity with dynamic phrases and its application to action recognition in tennis videos”. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, Springer, 909–916.

[38] T. Huang G. Yang, G. Tang (1979). “A fast two-dimensional median filtering algorithm”. In: IEEE Trans. Acoust., Speech, Signal Processing 27.1, pp. 13–18.

[39] Arias-Castro, E. and D. L. Donoho (2009). “Does median filtering truly preserve edges better than linear filtering?” In: Annals of Statistics 37.3, pp. 1172–1206.

[40] Arce, G. R. (2005). “Nonlinear Signal Processing: A Statistical Approach”. In: Wiley: New Jersey, USA.

[41] Hall, M. (2007). “Smooth operator: smoothing seismic horizons and attributes.” In: The Leading Edge 26.1, pp. 16–20.

[42] Boyle, R. and R. Thomas (1988). In: Computer Vision: A First Course, Blackwell Scientific Publications, pp. 32–34.

[43] Davies, E. (1990). “Machine Vision: Theory, Algorithms and Practicalities.” In: Academic Press 3.

[44] Vernon, D. (1991). “Machine Vision”. In: Prentice-Hall 4.

[45] Enea Cippitelli Samuele Gasparrini, Ennio Gambi-Susanna Spinsante (2016). “A Human Activity Recognition System Using Skeleton Data from RGBD Sensors”. In: Computational Intelligence and Neuroscience.

[46] Georgios Tsatiris Kostas Karpouzis, Stefanos Kollias (2017). “Variance-based shape descriptors for determining the level of expertise of tennis players”. In: Virtual Worlds and Games for Serious Applications (VS-Games), 9th International Conference.

[47] Georgios Goudelis Georgios Tsatiris Kostas Karpouzis, Stefanos Kollias (2017). “3D cylindrical trace transform based feature extraction for effective human action classification”. In: Computational Intelligence and Games (CIG), 2017 IEE Conference.