ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΚΑΙ ΔΙΚΤΥΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«HMM background removal for event localization»

ΕΝΑΛΛΑΚΤΙΚΑ

«Αφαίρεση παρασκηνίου με χρήση Κρυφών Μαρκοβιανών Μοντέλων για εντοπισμό τοπικών γεγονότων»

Κουτσαυτίκης Δημήτριος

e-mail : dikoutsa@inf.uth.gr

Επιβλέπουσα : Μπριασούλη Αλεξία

Επιτροπή : Χούστης Ηλίας

Βόλος , Μάρτιος 2008

ΕΥΧΑΡΙΣΤΙΕΣ

Καταρχήν, θα ήθελα να ευχαριστήσω την Καθηγήτριά μου κ. Αλεξία Μπριασούλη που με εμπιστεύθηκε με την παρούσα διπλωματική εργασία. Με την στάση της, με βοήθησε να εργαστώ απρόσκοπτα για την ολοκλήρωσή της. Οι γνώσεις και οι εμπειρίες που αποκόμισα μέσα από την διαδρομή αυτής της εργασίας είναι πολύ σημαντικές και σίγουρα θα φανούν χρήσιμες στο μέλλον.

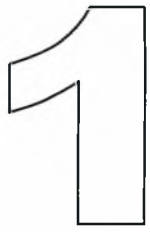
Στη συνεχεία, θα ήθελα να ευχαριστήσω τον κ. Ηλία Χούστη για την πολύτιμη καθοδήγηση, συνεργασία και βοήθεια που μου προσέφερε, στην επίλυση των προβλημάτων που ανέκυπταν κατά διάρκεια της διπλωματικής εργασίας.

Παράλληλα, θα ήθελα να ευχαριστήσω τη μητέρα μου, Πηνελόπη, για την αμέριστη συμπαράστασή της, ηθική και υλική, την υπομονή και την καθοδήγηση της, καθόλη τη διάρκεια των σπουδών μου. Το μεγαλύτερο μερίδιο της επιτυχούς ολοκλήρωσης της φοίτησης μου, το οφείλω σε εκείνη.

Τέλος, θέλω να ευχαριστήσω, τον αδερφό μου και συνάδελφο Γιάννη, όπως και δύο φίλους και συμφοιτητές, που πιστεύω ότι χωρίς τη συμπαράστασή τους τα πράγματα θα ήταν πολύ διαφορετικά . Ήταν πάντα δίπλα μου, όταν τους χρειάστηκα. Γιάννη, Νίκο και Στέφανε, σας ευχαριστώ.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1
ΕΙΣΑΓΩΓΗ



1.1 Γενικά

Στα πλαίσια της παρούσας διπλωματικής εργασίας μελετάται η χρήση των Κρυφών Μαρκοβιανών Μοντέλων(Hidden Markov Models) στην αναγνώριση τοπικών γεγονότων που προκύπτουν σε μια σειρά συνεχόμενων χρονικά εικόνων(frames). Ειδικότερα, εξετάζεται η μοντελοποίηση του παρασκηνίου (background) σε μια σειρά από tracking frames, τεχνική που χρησιμοποιείται ευρέως στην παρακολούθηση εικόνας-βίντεο μέσω υπολογιστών. Αυτή έχει ως σκοπό την ανίχνευση αντικειμένων στο προσκήνιο (foreground objects) ή την παρακολούθηση κινούμενων αντικείμενων σε μια ακολουθία βίντεο.

Επισημαίνεται ότι όσο πιο ακριβής είναι η μοντελοποίηση του παρασκηνίου τόσο πιο αξιόπιστη θα είναι η ανίχνευση και η περαιτέρω ταξινόμηση των διαφόρων τοπικών γεγονότων στο προσκήνιο. Συνηθισμένες χρήσεις της μεθόδου είναι περιπτώσεις αυτόματων συστημάτων οπτικής παρακολούθησης, περιήγησης βίντεο, καθώς επίσης και ταξινόμησης βίντεο βάσει του περιεχομένου. Αξίζει να σημειωθεί πως θα ήταν πολύ χρήσιμο αν τόσο οι χρήστες όσο και τα ερευνητικά μέλη συμφωνούσαν σε μια κοινή θεώρηση της περιγραφής ενός τοπικού γεγονότος, αφού αυτό θα βοηθούσε στην εύκολη ανταλλαγή επισημάνσεων πάνω σε βίντεο και στην υιοθέτηση κοινών μοντέλων ανίχνευσης βίντεο, από τους ερευνητές.

Η ανίχνευση κινούμενων αντικειμένων-στόχων στο προσκήνιο, με ψηφιακά δεδομένα τύπου βίντεο, έχει πολλές εφαρμογές τα τελευταία χρόνια. Σε αυτές τις εφαρμογές οι στόχοι είναι άνθρωποι, οχήματα ή άλλα διάφορα αντικείμενα. Η κοινή ιδιότητα των στόχων αυτών είναι πως αργά ή γρήγορα θα παρουσιάσουν κάποια κίνηση, η οποία τα ξεχωρίζει από τα αντικείμενα του παρασκηνίου.

Προκειμένου να γίνουν πλήρως κατανοητές οι εφαρμογές των HMM στο αντικείμενο της αναγνώρισης τοπικών γεγονότων, η διπλωματική αυτή εργασία περιγράφει το σχεδιασμό, την υλοποίηση και τα πειράματα που διενεργήθηκαν πάνω

σε διαφορετικές ακολουθίες βίντεο, ώστε με χρήση συστημάτων εκπαίδευσης HMM να γίνει ταξινόμηση και συσχετισμός παρόμοιων τοπικών γεγονότων.

1.2 Κίνητρα

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

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

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

Η διπλωματική αυτή εργασία, παρουσιάζει μια μέθοδο ανίχνευσης και ταξινόμησης γεγονότων που προκύπτουν εντός ενός προσκηνίου που απαρτίζεται από

κινούμενα αντικείμενα, χρησιμοποιώντας, όπως είπαμε, Hidden Markov Models. Τα HMM είναι ένα διπλά στοχαστικά μοντέλα και είναι κατάλληλα για αναπαράσταση των στοχαστικών ιδιοτήτων των κινούμενων αντικείμενων. Τα HMM υιοθετούνται, λοιπόν, για την αναπαράσταση των κινούμενων αντικείμενων, και οι παράμετροί τους εκτιμώνται από τα δεδομένα εκπαίδευσης. Τα γεγονότα αναγνωρίζονται μετά την αξιολόγηση των εκπαιδευμένων HMM, με βάση το κριτήριο της πιο πιθανής απόδοσης.

Τα βασικά πεδία εφαρμογής της μεθόδου είναι η υιοθέτηση κοινών μοντέλων ανίχνευσης βίντεο, περιπτώσεις αυτόματων συστημάτων οπτικής παρακολούθησης, περιήγηση καθώς επίσης και ταξινόμηση βίντεο βάσει του περιεχομένου τους. Μερικά ενδεικτικά παραδείγματα είναι:

- εντοπισμός διάβασης πεζών σε ένα βίντεο όπου υπάρχει συσσωρευμένη αμφίδρομη κίνηση ανθρώπων, τοπικά, σε μια λωρίδα των αντίστοιχων frame
- ανίχνευση των κινήσεων ενός χορού μπαλέτου και ταξινόμηση της προβαλλόμενης φιγούρας στα tracking frames ως χορευτή μπαλέτου
- καταγραφή όλων των στιγμιότυπων ενός αγώνα τένις όπου οι παίκτες κάνουν την κίνηση του σερβίς.

Επιπρόσθετα, αναφέρεται πως η προτεινόμενη μέθοδος μπορεί να έχει επίσης πιθανές εφαρμογές σε ποικίλα προβλήματα επικοινωνίας ανθρώπου-μηχανής, όπως αναγνώριση συναισθημάτων ή ανάπτυξη συστημάτων καθορισμού εντολών με βάση χειρονομίες.

1.3 Οι Μέθοδοι, τα Βήματα και οι Ιδιότητες τους

Σε αυτήν την ενότητα καταθέτουμε συνοπτικά τη διαδικασία που ακολουθείται ξεκινώντας από μια αρχική επιλογή ενός video sequence που θα αποτελέσει το σύνολο εκπαίδευσης των HMM, μέχρι την τελική κατηγοριοποίηση των γεγονότων που προκύπτουν σε αυτό.

Αρχικά εντοπίζουμε ένα βίντεο που περιέχει τα επιθυμητά γεγονότα που θέλουμε να ταξινομήσουμε(π.χ. όπως είδαμε προηγουμένων την κίνηση του σερβίς στο τένις ή τις χορευτικές κινήσεις μπαλέτου). Έπειτα υπάρχει μια κατηγορία μεθόδων για να διαχωρίσουμε αντικείμενα του background από αυτά του foreground,

χρησιμοποιώντας το Optical flow*. Για παράδειγμα, έχοντας μια εικόνα ενός κινούμενου αυτοκινήτου αποφασίζοντας ποια pixel στην εικόνα αντιπροσωπεύουν κίνηση μπορεί να μας βοηθήσει να αποφασίσουμε ποια pixels ανήκουν στο αυτοκίνητο, και ποια στο background.

Τυπικά, η κίνηση αντιπροσωπεύεται από διανύσματα που δημιουργούνται ή τερματίζονται σε pixels, ως αποτέλεσμα του όλου frame sequence. Οι μέθοδοι αυτοί εκμεταλλεύονται την συνοχή και την συνάφεια που έχει το optical flow σε ένα μικρό χρονικό διάστημα. Έχει παρατηρηθεί ότι αυτός ο τρόπος αποδίδει ικανοποιητικά αποτελέσματα όσον αφορά αντικείμενα του προσκηνίου. Εάν πάρουμε μια σειρά εικόνων και υπάρχουν κινούμενα αντικείμενα στη σκηνή, μπορούμε να πάρουμε χρήσιμες πληροφορίες για την εικόνα αναλύοντας και κατανοώντας τις διαφορές μεταξύ των εικόνων, που προκαλούνται από την κίνηση.

Υπάρχουν όμως δύο σοβαρά μειονεκτήματα που αξίζει να σημειωθούν. Καταρχάς, είναι δύσκολο να παράγουμε το optical flow σε περιοχές της εικόνας με λίγα χαρακτηριστικά γνωρίσματα «υφής» (texture), ή στα όρια των ασυνεχειών της εικόνας. Τέτοια λάθη και ασάφειες μπορούν να οδηγήσουν σε προβλήματα στην ανίχνευση αντικείμενων του προσκηνίου. Το δεύτερο βασικό μειονέκτημα είναι η χρονική καθυστέρηση που εισάγεται κατά την ανίχνευση της συνοχής του optical flow, το οποίο με τη σειρά του οδηγεί σε καθυστέρηση στην ανίχνευση foreground αντικειμένων.

Η «αφαίρεση» παρασκηνίου(background subtraction) προϋποθέτει τον υπολογισμό μιας εικόνας αναφοράς (reference image), από την οποία θα «αφαιρείται» κάθε νέο frame και έπειτα θα «φιλτράρεται» με βάση κάποιο κατώφλι για να μας δώσει την δυαδική τμηματοποίηση της εικόνας, η οποία ξεχωρίζει τις περιοχές με μη στάσιμα αντικείμενα. Αυτό είναι και το βασικό μειονέκτημα του μη προσαρμόσιμου background. Το κατώφλι είναι μια σταθερή παράμετρος και δεν υποστηρίζεται από προσαρμοστικές μεθόδους. Τα διάφορα βίντεο απαιτούν διαφορετικά κατώφλια. Το παρασκήνιο αλλάζει και μάλιστα με διαφορετικούς ρυθμούς κάθε φορά. Οι αλλαγές αφορούν τα αντικείμενα, τον φωτισμό αλλά και τον θόρυβο (π.χ. από την κάμερα λήψης, jitter, κλπ.). Μια απλοϊκή μορφή υπολογισμού της εικόνας αναφοράς του παρασκηνίου είναι ο μέσος όρος όλων των εικόνων-frames στη πάροδο ενός συγκεκριμένου χρονικού διαστήματος. Όμως και αυτή η μέθοδος

Optical flow : *Η προσέγγιση της κίνησης των αντικειμένων μέσω μίας οπτικής αναπαράστασης*

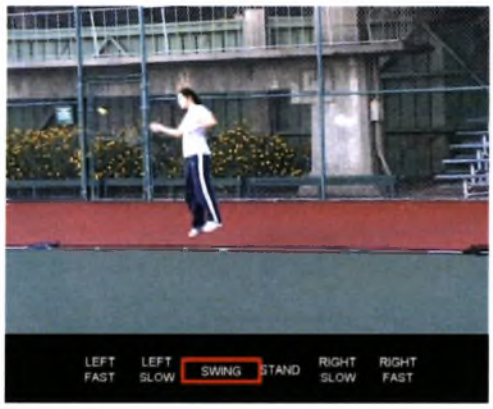
έχει διάφορα προβλήματα και απαιτεί μια περίοδο «εκπαίδευσης» υπό την απουσία αντικειμένων στο προσκήνιο. Η κίνηση αντικειμένων στο παρασκήνιο μετά την περίοδο της εκπαίδευσης και η στασιμότητα αντικειμένων στο προσκήνιο κατά τη διάρκεια της περιόδου εκπαίδευσης θα θεωρηθούν από το σύστημα ως μόνιμα foreground αντικείμενα. Επιπροσθέτως, η προσέγγιση αυτή δεν μπορεί να χειριστεί κατάλληλα βαθμιαίες αλλαγές στον φωτισμό του σκηνικού. Όλα αυτά οδηγούν στην απαίτηση της συνεχούς επανεκτίμησης του background μοντέλου και έτσι αναπτύσσονται συνεχώς νέες μέθοδοι προσαρμόσιμων background models.

Επιγραμματικά, η διαδικασία ανίχνευσης κίνησης βασισμένης σε γεγονότα(event detection) περιλαμβάνει τα εξής βήματα :

- Εντοπισμό και επιλογή ενός βίντεο με τις επιθυμητές ιδιότητες
- Εξαγωγή των διανυσμάτων κίνησης μέσω της οπτικής ροής
- Ορισμός των καταστάσεων του συστήματος
- Δημιουργία του HMM συνόλου εκπαίδευσης
- Εξαγωγή των πινάκων μετάβασης στις καταστάσεις τις οποίες αρχικά θέσαμε
- Εφαρμογή των HMM σε εναλλακτικά βίντεο για αναγνώριση παρόμοιων χαρακτηριστικών με το αρχικό video sequence

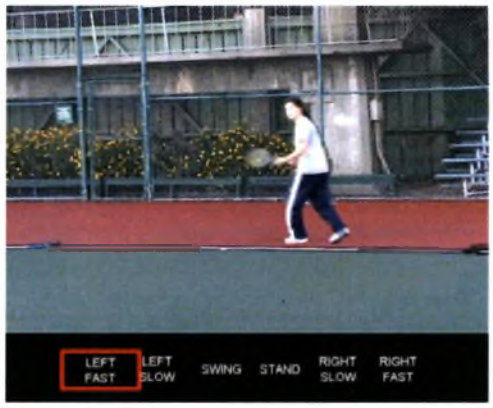
Εφόσον έχουμε επιλέξει ένα αρχικό-πρότυπο βίντεο κατηγοριοποιούμε τις καταστάσεις στις οποίες εντάσσουμε τα εκάστοτε τοπικά γεγονότα που ανιχνεύονται. Για παράδειγμα, στο τένις η κίνηση με του παίκτη με τη ρακέτα για να χτυπήσει τη μπάλα θα μπορούσε να είναι μία κατάσταση, ενώ το γρήγορο βάδισμά του στα αριστερά μία άλλη. Στην παρούσα εργασία θα ασχοληθούμε ιδιαίτερα με την ανίχνευση και κατηγοριοποίηση των ανθρώπινων δραστηριοτήτων και αλληλεπιδράσεων που προκύπτουν σε ένα βίντεο.

*State **swing***



Εικόνα 1 : Χτύπημα μπάλας με ρακέτα [3]

*State **left fast***

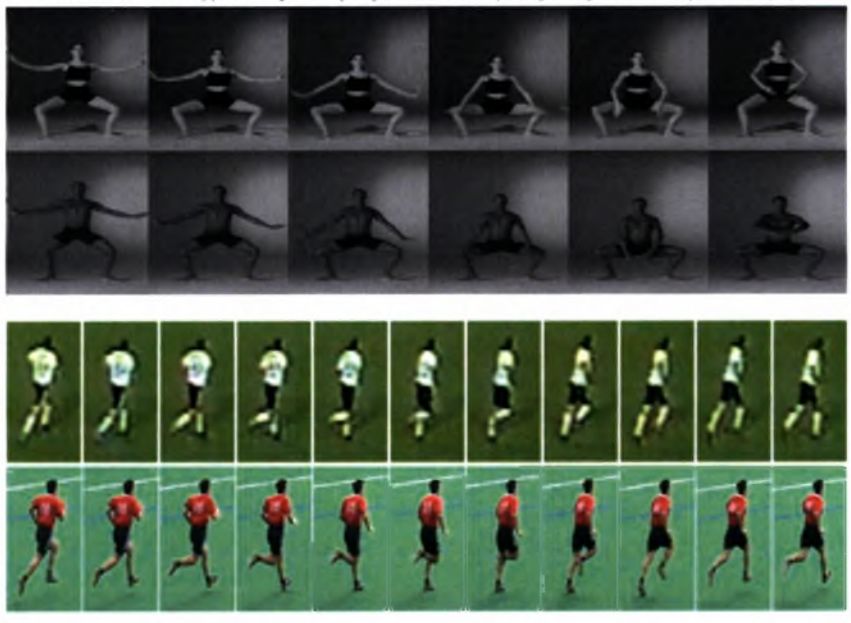


Εικόνα 2 : Γρήγορο τρέξιμο στα αριστερά [3]

Χρησιμοποιώντας την οπτική ροή(Optical Flow) εξάγουμε τα διανύσματα κίνησης από το frame sequence, εστιάζοντας πάνω στα «ενεργά» αντικείμενα. Ενδεικτικά, αναφέρουμε ότι τα Hidden Markov Models απαιτούν τον ορισμό κάποιων καταστάσεων, παρόμοιων με αυτές που ορίσαμε πιο πάνω για την περίπτωση του τένις. Τα HMM χρησιμοποιούν το σύνολο καταστάσεων και τα διανύσματα κίνησης του optical flow για να κατασκευάσουν πίνακες μετάβασης στις καταστάσεις τις οποίες αρχικά θέσαμε. Από αυτό το σημείο και μετά, έχοντας ανακτήσει τους πίνακες μετάβασης μπορούμε να επιλέξουμε ένα οποιοδήποτε εναλλακτικό βίντεο και με χρήση των Κρυφών Μαρκοβιανών Μοντέλων να αποφασίσουμε αν κάποιο τοπικό γεγονός αντιστοιχεί σε μια από τις καταστάσεις που ορίσαμε.

Περιπτώσεις ανίχνευσης τοπικών γεγονότων βάσει ενός αρχικού-πρότυπου frame sequence, απεικονίζονται παρακάτω :

Εικόνα 3 : Ανίχνευση κίνησης και αναγνώριση αντικειμένου [2]



Εικόνα 4 : Βέλτιστα ταιριάσματα για κατηγοριοποίηση κινήσεων χορού μπαλέτου και αγώνα ποδοσφαίρου. Στην πάνω σειρά έχουμε μια ακολουθία frame εισόδου και στην κάτω τα βέλτιστα ταιριάσματα για κάθε frame. [3]

ΚΕΦΑΛΑΙΟ 2
ΑΝΙΧΝΕΥΣΗ ΚΙΝΗΣΗΣ ΚΑΙ ΟΠΤΙΚΗ ΡΟΗ

ΕΙΣΑΓΩΓΗ

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

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

Η κατανόηση της ανθρώπινης κίνησης, μπορεί να προσεγγιστεί με διάφορα επίπεδα λεπτομερειών, ανάλογα με την πολυπλοκότητα της εκάστοτε κίνησης. Η μοντελοποίηση και η αναγνώριση της ανθρώπινης συμπεριφοράς προϋποθέτει τον χαρακτηρισμό και την ταξινόμηση των διαφόρων ειδών κίνησης. Μια ιδέα που εφαρμόστηκε αρχικά για την επίλυση αυτού του ζητήματος ήταν η ταξινόμηση της κίνησης σε « αλλαγή, γεγονός, επεισόδιο και ιστορία» ώστε να υπάρξει αποτύπωση των διαφορετικών διαστάσεων του προβλήματος. Η κάθε διάσταση σχετίζεται και με

διαφορετικό όγκο πληροφορίας που απαιτείται για την επίτευξη αναγνώρισης. Μια διαφορετική προσέγγιση είναι ο διαχωρισμός της κίνησης σε «κινήσεις, δραστηριότητα, ενέργεια». Σε αυτού του είδους την ταξινόμηση οι *κινήσεις* είναι εξατομικευμένες στοιχειώδεις κινήσεις οι οποίες δεν απαιτούν την συλλογή δεδομένων από κάποια ακολουθία για να αναγνωριστούν. Αντίθετα η *δραστηριότητα* αναφέρεται σε μια ακολουθία κινήσεων ή καταστάσεων όπου η μόνη πραγματική γνώση είναι τα στατιστικά χαρακτηριστικά της ακολουθίας. Μεγάλο μέρος της πρόσφατης έρευνας πάνω στην αναγνώριση χειρονομιών εμπίπτει σε αυτή την κατηγορία αναγνώρισης κινήσεων. Τέλος, οι *ενέργειες* είναι γεγονότα μεγαλύτερης διάρκειας τα οποία συνήθως περιλαμβάνουν αλληλεπιδράσεις με το περιβάλλον.

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

2.1 Αντικείμενα, καταστάσεις και γεγονότα

Σε ένα δεδομένο frame μπορούμε να εντοπίσουμε *αντικείμενα(objects)*. Ως αντικείμενο ορίζουμε τα τμήματα της αντίστοιχης εικόνας που έχουν κάποιες χαρακτηριστικές ιδιότητες/ γνωρίσματα. Ένα αντικείμενο μπορεί να είναι «ενεργό», δηλαδή κινούμενο ή «ανενεργό», χωρίς κίνηση. Τα «ενεργά» αντικείμενα συνήθως αλληλεπιδρούν μεταξύ τους ή με ακίνητα τμήματα ενός frame. Εναλλακτικά, μπορούμε να τα θεωρήσουμε ως συναρτήσεις πιθανοτήτων που έχουν δύο ή παραπάνω ορίσματα.

Οι ιδιότητες, τα γνωρίσματα, οι συσχετίσεις σε ένα βίντεο μπορούν επίσης να θεωρηθούν ως *καταστάσεις*. Ειδικότερα, ο όρος *κατάσταση*(state) στον τομέα της πληροφορικής αναφέρεται ως το άθροισμα όλων των ιδιοτήτων και συσχετίσεων

μεταξύ των εμπλεκόμενων οντοτήτων σε μία δεδομένη χρονική στιγμή. Με βάση αυτό τον ορισμό μια τέτοια κατάσταση την αποκαλούμε καθολική(world state).

Ένα *γεγονός(event)* ορίζεται ως η αλλαγή στην κατάσταση ενός αντικειμένου. Χαρακτηριστικό παράδειγμα, η πτώση ενός βράχου από ένα λόφο. Η κατάσταση του μεταβάλλεται συνεχώς. Γενικότερα, τα γεγονότα προκύπτουν εντός μιας δεδομένης χρονικής στιγμής ή κατά τη διάρκεια ενός χρονικού διαστήματος. Επίσης συνήθως έχουν συγκεκριμένη θέση μέσα στο frame την οποία «κληρονομούν» από τα διάφορα «ενεργά» αντικείμενα που υπάρχουν στην εικόνα. Συνεχίζοντας το παραπάνω σενάριο, ο βράχος που πέφτει ίσως τελικά προσκρούσει πάνω σε ένα παράθυρο, θρυμματίζοντάς το. Συνεπώς η εμφάνιση ενός γεγονότος μπορεί να προκαλέσει τη δημιουργία άλλων σε ένα δεδομένο video sequence.

Εάν πάρουμε μια σειρά εικόνων και υπάρχουν κινούμενα αντικείμενα στη σκηνή, μπορούμε να πάρουμε χρήσιμες πληροφορίες για την εικόνα αναλύοντας και κατανοώντας τις διαφορές μεταξύ των εικόνων, που προκαλούνται από την κίνηση. Για παράδειγμα, έχοντας μια εικόνα ενός κινούμενου αυτοκινήτου αποφασίζοντας ποια pixel στην εικόνα αντιπροσωπεύουν κίνηση μπορεί να μας βοηθήσει να αποφασίσουμε ποια pixels ανήκουν στο αυτοκίνητο, και ποια στο background.

Εικόνα 5 : Ανίχνευση κίνησης αντικειμένου [2]

2.2 Ανίχνευση κίνησης

Στο δικό μας σύστημα η εφαρμογή των προτεινόμενων μεθόδων γίνεται μέσω της ειδικής βιβλιοθήκης της OpenCV (**Open** Source **C**omputer **V**ision Library) της Intel. Σε ένα αντίστοιχο σύστημα, ο στόχος της διαδικασίας ανίχνευσης κίνησης είναι

να αποφασιστεί αν στο δεδομένο προσκήνιο(foreground) ενός video sequence υπάρχουν κινούμενα αντικείμενα και να μαρκαριστούν τα περιγράμματα(contours) των «ενεργών» αντικειμένων με καμπύλες.
Για το σκοπό αυτό χρησιμοποιείται η συνάρτηση:

int cvFindContours(image, storage, first_contour, header_size,CV_RETR_LIST, ALGORITHM, offset)

της OpenCV,η οποία υλοποιείται με χρήση της κωδικοποίησης Freeman για εντοπισμό του περιγράμματος.

Η συνάρτηση αυτή ανακτά τα περιγράμματα μιας δυαδικά τμηματοποιημένης εικόνας και επιστρέφει τον αριθμό των περιγραμμάτων που ανακτήθηκαν. Μετρώντας τον αριθμό των σημείων σε κάθε περίγραμμα, μπορούμε να αποφασίσουμε για το αν το παρών προσκήνιο είναι ένα «ενεργό» αντικείμενο. Για παράδειγμα, αν ο αριθμός είναι μεγαλύτερος από μια τιμή κατωφλίου (threshold value), η πιθανότητα να υπάρχει κίνηση εξαιτίας ενός πραγματικού αντικειμένου είναι μεγάλη.

Σε αντίθετη περίπτωση, αν η κινούμενη περιοχή είναι σχετικά μικρή, η πιθανότητα η κίνηση που εντοπίστηκε να αντιστοιχεί σε θόρυβο είναι μεγάλη, ενώ να αντιστοιχεί σε αντικείμενο, μικρή. Συνεπώς, αυτό το βήμα μπορεί να μας βοηθήσει να απομακρύνουμε το θόρυβο που μπορεί να εμφανίζεται σε μια εικόνα και να αναγνωρίζουμε τα βασικά σημεία κίνησης στο προσκήνιο όπως φαίνεται και στις παρακάτω εικόνες.

Εικόνα 6 : Προσεγγιστικά περιγράμματα κίνησης πεζών [2]

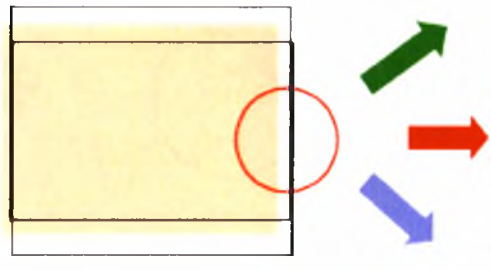
Ωστόσο, το περίγραμμα που ανακτάται με την παραπάνω μέθοδο δεν είναι το ακριβές περιθώριο ενός κινούμενου αντικειμένου. Το τμηματοποιημένο προσκήνιο έχει «απλωθεί» από την εφαρμογή του φίλτρου *dilation**. Για να αντιμετωπιστεί αυτό το πρόβλημα, ένα εναλλακτικό μοντέλο «ενεργού» περιγράμματος θα πρέπει να εφαρμοστεί ώστε να εντοπιστεί το ακριβές περιθώριο της κίνησης.

Dilation: Φίλτρο επεξεργασίας θορύβου σε μια εικόνα. Εφαρμόζεται σαν συνέλιξη και είναι χρήσιμο στο να γεμίζει κενά σε εικόνες.

2.3 Προκλήσεις στην εκτίμηση κίνησης

Θα πρέπει να υπογραμμιστεί ότι η εκτίμηση κίνησης είναι ένα πολυσύνθετο πρόβλημα. Οι αλληλεπιδράσεις κίνησης είναι τοπικά γεγονότα και ποικίλουν από frame σε frame. Η χρήση δισδιάστατης απεικόνισης ενός τρισδιάστατου κόσμου, για τον υπολογισμό της κίνησης, είναι ημιτελής και μη ιδανική. Υπάρχουν προβλήματα που είναι δύσκολο να αποφευχθούν. Ειδικότερα, προσωρινές μεταβολές στο περιεχόμενο των frames μπορεί να προκύπτουν εξαιτίας υπερφωτισμού ή σκιών και όχι λόγω κάποιας κίνησης αντικειμένου.

Κανένας αλγόριθμος από τους ήδη υπάρχοντες δεν αντιμετωπίζει καταλυτικά τα παραπάνω προβλήματα. Η χρήση της οπτικής ροής εξάγει μία πυκνή κατανομή της κίνησης, ωστόσο αυτή η τεχνική σχετίζεται με ένα σημαντικό πρόβλημα που ονομάζεται πρόβλημα διαφράγματος(aperture problem). Εξαιτίας του ότι ο αλγόριθμος βασίζεται σε αποκλίσεις τοπικής φωτεινότητας, προκαλεί τοπική «σύγχυση» στα ανακτώμενα διαγράμματα κίνησης(motion fields)., όπως φαίνεται και στην παρακάτω εικόνα.



Εικόνα 7 : Το πρόβλημα διαφράγματος(aperture problem) [8]

2.4 Οπτική ροή(Optical Flow)

Από μια ακολουθία εικόνων, υπολογίζουμε την οπτική ροή. Για κάθε pixel υπολογίζουμε το διάνυσμα ταχύτητας που μας δίνει:

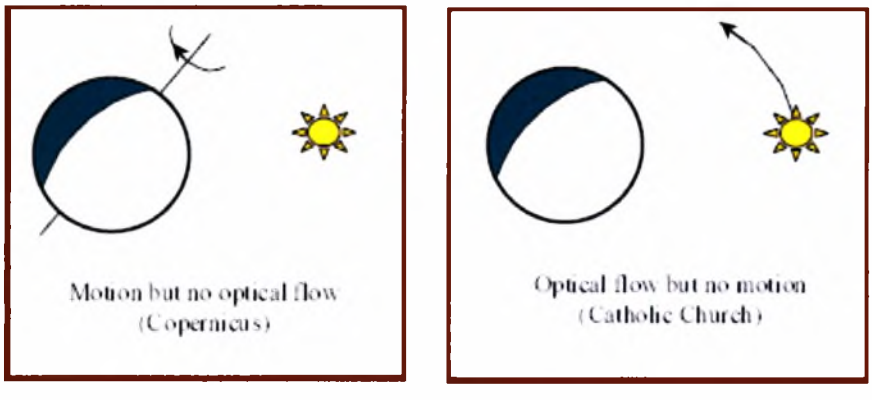
$$V=(u,v)$$

- πόσο γρήγορα κινείται στην εικόνα ό,τι βρίσκεται σε εκείνο το pixel.
- σε ποια κατεύθυνση κινείται.

Το διάνυσμα οπτικής ροής w(x,y) έχει δύο συστατικά: το u(x,y) και το v(x,y) που περιγράφουν την κίνηση ενός σημείου στην x και την y κατεύθυνση στην εικόνα. Για να μπορούμε να μετρήσουμε την οπτική ροή πρέπει να βρούμε τα αντίστοιχα σημεία μεταξύ των δύο εικόνων.

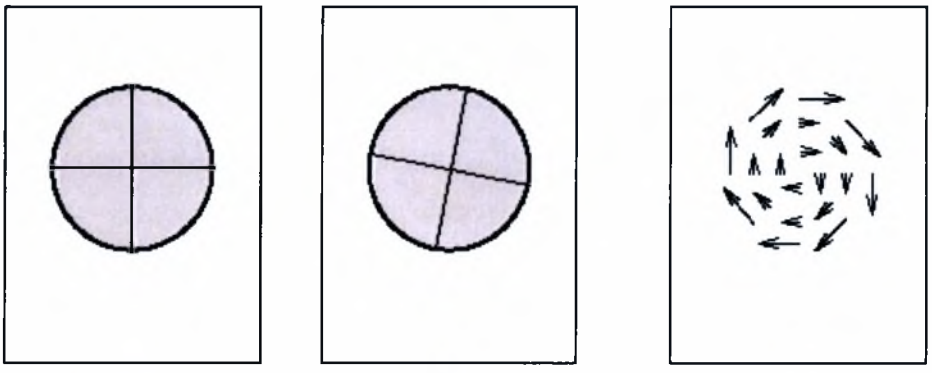
Ένα σημείο βρίσκεται σε μια θέση (x0,y0) στο χρόνο t0 και υπάρχουν διάφορες άλλες πιθανές θέσεις (x0+δx,y0+δy) στις οποίες εκείνο το σημείο μπορεί να έχει κινηθεί μεταξύ δύο εικόνων στο χρόνο t0 και t0+δt. Η φωτεινότητα της εικόνας σε ένα σημείο (x0,y0) στο χώρο και το χρόνο είναι I(x0,y0,t0). Μετά από κάποιο χρόνο δt η φωτεινότητα στο (x0,y0) θα κινηθεί στο σημείο (x0+δx,y0+δy). Ουσιαστικά θεωρούμε ότι η φωτεινότητα δεν αλλάζει αλλά αλλάζει μόνο η θέση της.

Οι δυσκολίες του ταιριάσματος σημείων πάνω σε κινούμενα αντικείμενα, σε συνδυασμό με τη μέτρηση των θέσεων και των ταχυτήτων τους με επαρκή ακρίβεια, μας ώθησε σε μία τοπική προσέγγιση του προβλήματος όπου λαμβάνεται υπόψη η αλλαγή φωτεινότητας σε ένα pixel. Το ακόλουθο διάγραμμα δείχνει πως μία περιστρεφόμενη σφαίρα, με σταθερές θέσεις στο φως και στην κάμερα δεν έχει μεταβολή στη φωτεινότητα των pixel ενώ, αντίθετα, μία σφαίρα που είναι σχετικά στατική με βάση μια κινούμενη πηγή φωτός θα προκαλέσει μεταβολές στη φωτεινότητα. Ομοίως, μία κινούμενη σφαίρα γύρω από μία στατική πηγή φωτός θα παράγει μεταβολές στη φωτεινότητα των pixel.



Εικόνα8: Optical flow - local approach [8]

Συνήθως, διαφορετικά αντικείμενα κάνουν διαφορετικές κινήσεις στο προσκήνιο, γεγονός που οδηγεί σε διαφορετικές οπτικές ροές. Χρησιμοποιώντας τις ασυνέχειες μιας οπτικής ροής, μπορούμε να τμηματοποιήσουμε εικόνες σε περιοχές. Για παράδειγμα,, στην εικόνα 14, ο κύκλος που περιστρέφεται στη μέση μπορεί να απομονωθεί ως προσκήνιο από το στατικό περιβάλλον όπου η ροή σταματάει.



(a) Αντικείμενο τη στιγμή 1
(b) Αντικείμενο τη στιγμή 2
(c) Οπτική ροή – τα βέλη δείχνουν την κατεύθυνση της κίνησης- τα μήκη τους αναπαριστούν την ταχύτητα της κίνησης.

Εικόνα 9 :Optical flow - background subtraction [2]

Το μαθηματικό μοντέλο αυτής της διαδικασίας δίνεται παρακάτω. Ας υποθέσουμε ότι η φωτεινότητα σε ένα pixel P(x, y) τη χρονική στιγμή t συμβολίζεται ως $I(x, y,t)$ και το αντίστοιχο διάνυσμα κίνησης $V(u,v)$. Σε μία τοπική «γειτονιά» pixel η ένταση της φωτεινότητας δίνεται από τη σχέση:

$I(x, y,t) = I(x + u*t, y + v*t,0)$

Καθώς η φωτεινότητα σε ένα συγκεκριμένο pixel είναι σταθερή, προκύπτει ότι :

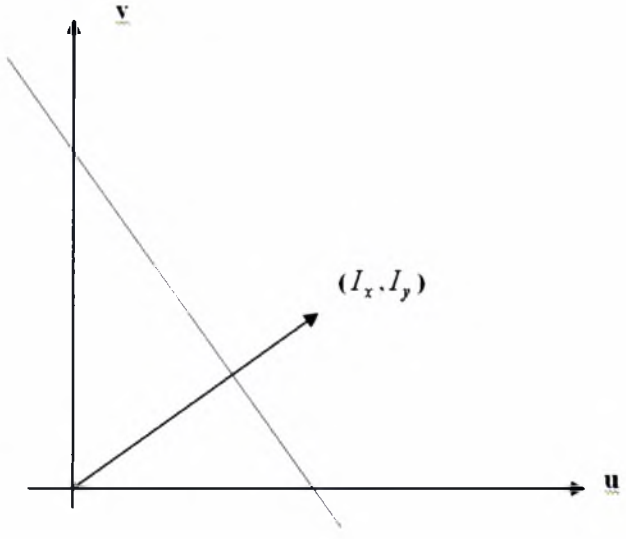
$I(x, y,t) = I(x + u*t, y + v*t,0)$ = σταθερά(constant),

$$\frac{dI}{dt}=\frac{\partial I}{\partial x}\frac{dx}{dt}+\frac{\partial I}{\partial y}\frac{dy}{dt}+\frac{\partial I}{\partial t}=0$$

Έστω $\frac{dI}{dt}=I_t, \frac{\partial I}{\partial x}=I_x, \frac{\partial I}{\partial y}=I_y, u=\frac{dx}{dt}, v=\frac{dy}{dt}$

Έτσι η εξίσωση γράφεται ως :

$$I_x u + I_y v + I_t = 0$$



Εικόνα 10 : Γραφική παράσταση της απλής γραμμικής συνάρτησης που ορίζεται από τα δύο άγνωστα στοιχεία της ταχύτητας(u,v) [2]

Μία εναλλακτική μορφή της εξίσωσης είναι :

$(I_x, I_y) \bullet (u,v) = I_t$

Επομένως, η κίνηση του Pixel P(x,y) στην διεύθυνση του διανύσματος της φωτεινότητας μπορεί να αναπαρασταθεί ως :

$$-\frac{I_y}{\sqrt{I_x^2 + I_y^2}}$$

Ως τώρα, έχουμε βρει τη βασική συνάρτηση οπτικής ροής. Ωστόσο, μια συνάρτηση δε μπορεί να επιλύσει δύο άγνωστα ορίσματα. Επιπρόσθετες υποθέσεις πρέπει να γίνουν ώστε να υπολογιστεί η πραγματική ροή ταχύτητας. Σύμφωνα με το σε ποια μέτρηση έχει βασιστεί η εκάστοτε υπόθεση, οι αλγόριθμοι γενικά κατατάσσονται σε τέσσερις ομάδες – τη διαφορική τεχνική, την τεχνική των περιοχών, την ενεργειακή τεχνική και την τεχνική που βασίζεται στη φάση.

Στο σύστημα μας, υλοποιείται η πιο δημοφιλής διαφορική μέθοδος δύο frame των Lucas-Kanade, γιατί είναι η πιο ανθεκτική στην παρουσία θορύβου. Ειδικότερα, χρησιμοποιήθηκε η συνάρτηση :

cvCalcOpticalFlowLK(srca, srcb, winsize, velx, vely) της OpenCV

Σε αυτή τη μέθοδο, θεωρείται ότι η οπτική ροή είναι τοπικά σταθερή. Επομένως, σε ένα μικρό παράθυρο μεγέθους w x w(w>1, με κέντρο το pixel P(x,y)), μία σειρά εξισώσεων για το κάθε pixel μέσα στο παράθυρο είναι η παρακάτω :

$$
\begin{aligned}
&I_{x1}u + I_{y1}v = -I_{t1},\\
&I_{x2}u + I_{y2}v = -I_{t2},\\
&\vdots\\
&I_{xn}u + I_{yn}v = -I_{tn}
\end{aligned}
$$

Τα pixel αριθμούνται ως 1…n.
Επομένως,

$$
\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \vdots & \vdots \\ I_{xn} & I_{yn} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -I_{t1} \\ -I_{t2} \\ \vdots \\ -I_{tn} \end{bmatrix}
$$

Ή αλλιώς :

$$I\vec{v} = -I_t$$

Τότε η μέθοδος των ελαχίστων τετραγώνων μπορεί να χρησιμοποιηθεί για να λύσει αυτή την εξίσωση. Επομένως, έχουμε :

$$I^T I\vec{v} = I^T(-I_t),$$

και

$$\vec{v} = (I^T I)^{-1} I^T(-I_t),$$

Επίσης η μέθοδος Lucas – Kanade μπορεί να υλοποιηθεί με ένα επαναληπτικό τρόπο διεξάγοντας τους υπολογισμούς για την οπτική ροή στην πυραμίδα της εικόνας(εξομαλυμμένο αντίγραφό της). Έτσι το πρόβλημα ανάγεται στον εντοπισμό της ταχύτητας $\vec{V}$,σε μια μικρή γειτονιά pixel Ω, η οποία ελαχιστοποιεί την ποσότητα:

$$\sum_{(x,y\in\Omega)} W^2(x,y)[\nabla I(x,y,t)\cdot\vec{V}+I_t(x,y,t)]^2$$

όπου W(x,y) υποδεικνύει μία συνάρτηση «παραθύρου» (μεγέθους 3 x 3) η οποία δίνει μεγάλα βάρη γύρω από το κεντρικό Pixel και όχι γύρω από την περιφέρεια. Η λύση της παραπάνω εξίσωσης για n σημεία $(x_i, y_i)\in\Omega$ τη στιγμή t, δίνεται από τη σχέση :

$$A^T W^2 A\vec{V} = A^T W^2 b \quad ,$$

όπου

$A=[\nabla I(x_1),\nabla I(x_2),...\nabla I(x_n)]^T$,

$W = diag[W(x_1),W(x_2),...W(x_n)]$,

$b=-(I_t(x_1),I_t(x_2),...I_t(x_n)^T$

Λύνοντας ως προς την ταχύτητα έχουμε :

$$\vec{V} = [A^T W^2 A]^{-1} A^T W^2 b$$

2.5 Εφαρμογές οπτική ροής

Καθώς υπολογίστηκε το διάνυσμα $\vec{V}$ της οπτικής ροής με βάση τα παραπάνω, μπορούμε να απομονώσουμε κινούμενα αντικείμενα από ένα στατικό παρασκήνιο περιορίζοντας το πλάτος της ταχύτητας ροής. Έτσι λοιπόν, το αποτέλεσμα της τμηματοποίησης μέσω οπτικής ροής ενός παρασκηνίου ή προσκηνίου αντίστοιχα(background/foreground segmentation) εξαρτάται σε μεγάλο βαθμό από την επιλογή της τιμής κατωφλίου. Αν η τιμή είναι υψηλή τότε ο αλγόριθμος αποτυγχάνει να εντοπίσει το μεγαλύτερο τμήμα της πληροφορίας. Αν η τιμή είναι πολλή μικρή, ο αλγόριθμος θα είναι πολύ ευαίσθητος σε κάθε παραμικρή αλλαγή της

φωτεινότητας της εικόνας. Από την άλλη πλευρά, η μέθοδος της οπτικής ροής δεν είναι καλή στο να ανιχνεύει αργή κίνηση, ειδικά όταν το «ενεργό» αντικείμενο είναι μεγάλο.

Για να αντιμετωπιστεί το συγκεκριμένο πρόβλημα, η μέθοδος που υλοποιήθηκε έγκειται σε ένα εναλλακτικό τρόπο υπολογισμού των διανυσμάτων ταχύτητας. Ειδικότερα, η εικόνα χωρίζεται σε μικρές περιοχές μεγέθους 3 x 3 (pixel window), και όχι για κάθε pixel χωριστά. Στη συνέχεια, το διάνυσμα της ταχύτητας υπολογίζεται χωριστά για κάθε περιοχή αθροίζοντας την τιμή για το κάθε pixel. Αν το πλάτος της ταχύτητας είναι μεγαλύτερο από το προτεινόμενο κατώφλι τότε αυτή η περιοχή ανήκει στο προσκήνιο, σε αντίθεση περίπτωση ανήκει στο παρασκήνιο.

Εικόνα 11 : Οπτική ροή με χρήση του αλγορίθμου Lucas –Kanade στην OpenCV

ΚΕΦΑΛΑΙΟ 3

ΤΑ ΚΡΥΦΑ ΜΑΡΚΟΒΙΑΝΑ ΜΟΝΤΕΛΑ

3

ΕΙΣΑΓΩΓΗ

ΙΣΤΟΡΙΑ ΤΩΝ ΜΑΡΚΟΒΙΑΝΩΝ ΑΛΥΣΙΔΩΝ

- Η θεωρία των Μαρκοβιανών αλυσίδων αναπτύχθηκε γύρω στο 1900.
- Τα Κρυφά Μαρκοβιανά Μοντέλα αναπτύχθηκαν στο τέλος του 1960.
- Χρησιμοποιήθηκαν εκτενώς στην αναγνώριση ομιλίας το 1960-70.
- Εφαρμόστηκαν στο τομέα της πληροφορικής το 1989.

Εικόνα 12 : Andrei Andreyevich Markov [6]

ΕΦΑΡΜΟΓΕΣ

- Βιοπληροφορική
- Ψηφιακή Επεξεργασία Σήματος
- Ανάλυση δεδομένων
- Ανίχνευση προτύπων(Pattern recognition)

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

3.1 Μοντελοποίηση των Φυσικών Διεργασιών

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

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

Υπάρχουν πολλές δυνατότητες επιλογής για τον τύπο του μοντέλου που θα χρησιμοποιηθεί για να περιγράψει τις ιδιότητες ενός φυσικού σήματος. Μια πρώτη διάκριση είναι ο διαχωρισμός τους σε ντετερμινιστικά και στατιστικά. Τα ντετερμινιστικά μοντέλα χρησιμοποιούν κάποιες γνωστές ιδιότητες του σήματος και

η εξαγωγή του μοντέλου γίνεται απευθείας (με εκτίμηση μόνο μερικών παραμέτρων). Τα στατιστικά μοντέλα μπορούν να μας δώσουν πληροφορίες που αφορούν μόνο τις στατιστικές ιδιότητες του σήματος. Στα στατιστικά μοντέλα περιλαμβάνονται οι Γκαουσσιανές διαδικασίες, οι διαδικασίες Poisson όπως και τα Κρυφά Μαρκοβιανά Μοντέλα (HMM). Η βασική θεώρηση στα στατιστικά μοντέλα είναι ότι το σήμα μπορεί να χαρακτηριστεί σαν μια παραμετρική τυχαία διαδικασία, της οποίας οι παράμετροι μπορούν να καθοριστούν με σαφές τρόπο.

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

3.2 Τα HMM

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

Κάθε κατάσταση του μοντέλου χαρακτηρίζεται από δύο σετ πιθανοτήτων. Την πιθανότητα μετάβασης και είτε μια διακριτή κατανομή πιθανότητας εξόδου είτε μια συνεχή συνάρτηση πυκνότητας πιθανότητας εξόδου. Έπειτα αυτές, με δεδομένη την κατάσταση καθορίζουν την δεσμευμένη πιθανότητα εκπομπής κάποιου από τα σύμβολα εξόδου (που υπάρχουν σε ένα πεπερασμένο αλφάβητο) ή κάποιου συνεχούς τυχαίου διανύσματος. Η ικανότητα των HMM να χειρίζονται ακολουθιακά δεδομένα, η ανεξαρτησία τους από μεταβολές της κλίμακας του χρόνου–η διάρκεια των ανθρώπινων κινήσεων δεν πρέπει να θεωρείται ως χαρακτηριστικό τους αντίθετα με τις καταστάσεις από τις οποίες διέρχονται- καθώς και η δυνατότητα μάθησης, τα καθιστούν κατάλληλα για την ταξινόμηση αγνώστων ακολουθιών από διανύσματα

Εικόνα 13 : Μαθηματικό μοντέλο μιας «υφής» βίντεο(video texture). [6]

Ένα HMM έχει τα ακόλουθα χαρακτηριστικά :

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

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

3) Την κατανομή πιθανότητας μετάβασης των καταστάσεων $A = \{a_{ij}\}$ όπου το a_{ij} ορίζεται ως εξής : $a_{ij} = P[q_{t+1}= S_j \mid q_t = S_i]$, $1\leq i, j\leq N$.Για την ειδική περίπτωση που κάθε κατάσταση μπορεί να οδηγήσει σε οποιαδήποτε από τις υπόλοιπες σε ένα μόνο

βήμα ισχύει $a_{ij}>0$ για όλα τα i,j. Για άλλους τύπους HMM θα έχουμε $a_{ij}=0$ για παραπάνω από ένα ζευγάρια (i ,j)

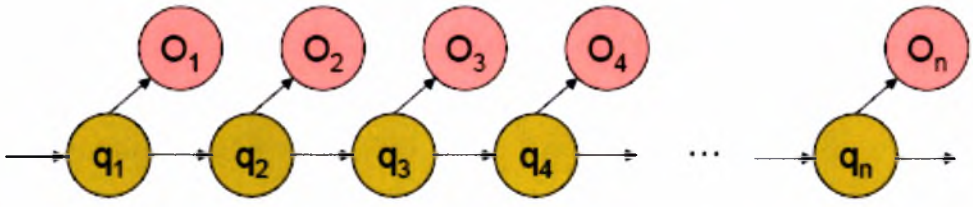
4) Την κατανομή πιθανότητας των συμβόλων παρατήρησης στην κατάσταση j , $B=\{b_j(k)\}$ όπου $b_j(k) = P[v_k \text{ όταν } t \mid q_t = S_j]$, $1\leq j \leq N$ και $1\leq k \leq M$.

5) Την κατανομή της αρχικής κατάστασης $\pi = \{\pi_i\}$ όπου
$\pi_i = P[q_1 = S_i]$, $1\leq j \leq N$

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

$$O = O_1O_2\ldots O_T$$

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



Εικόνα 14 : Γεννήτρια για ακολουθία εξόδων O_t ***[6]***

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

3.3 Τρία Βασικά Προβλήματα

Δοθείσας της μορφής του HMM που περιγράφτηκε προηγουμένως, υπάρχουν τρία βασικά προβλήματα στα οποία πρέπει να δοθεί ικανοποιητική λύση προτού το μοντέλο μπορέσει να χρησιμοποιηθεί σε εφαρμογές του πραγματικού κόσμου. Τα προβλήματα αυτά είναι τα ακόλουθα :

Α) Με δεδομένη την ακολουθία εξόδων $O = O_1O_2...O_T$ και το μοντέλο $\lambda=(A,B,\pi)$ πώς μπορεί να υπολογιστεί αποδοτικά η πιθανότητα $P(O|\lambda)$, δηλαδή η πιθανότητα η συγκεκριμένη ακολουθία να έχει «γεννηθεί» από το συγκεκριμένο μοντέλο.

Β) Με δεδομένη την ακολουθία εξόδων $O = O_1O_2...O_T$ και το μοντέλο $\lambda=(A,B,\pi)$ πως μπορεί να προσδιοριστεί μία ακολουθία καταστάσεων $Q = q_1q_2...q_T$ η οποία είναι η βέλτιστη κατά τρόπο που να έχει κάποια φυσική σημασία. (πχ. να εξηγεί την ακολουθία των εξόδων).

Γ) Με δεδομένη την ακολουθία εξόδων $O = O_1O_2...O_T$ πώς μπορούν να προσδιοριστούν οι βέλτιστες παράμετροι $\lambda=(A,B,\pi)$ του μοντέλου ώστε να μεγιστοποιείται η πιθανότητα $P(O|\lambda)$.

ΠΡΟΒΛΗΜΑ Α

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

ΠΡΟΒΛΗΜΑ Β

Εδώ ουσιαστικά πρέπει να αποκαλυφθεί το «κρυφό» μέρος του μοντέλου (π.χ. να βρεθεί η σωστή ακολουθία καταστάσεων). Είναι προφανές ότι για όλα τα μοντέλα, πλην ορισμένων εκφυλισμένων καταστάσεων, δεν υπάρχει «σωστή» ακολουθία. Αντί αυτού επιχειρείται η βελτιστοποίηση κάποιου κριτηρίου ώστε να δοθεί η καλύτερη δυνατή λύση. Δυστυχώς, υπάρχουν αρκετά κριτήρια βελτιστοποίησης τα οποία μπορούν να χρησιμοποιηθούν και έτσι σε κάθε περίπτωση , αναλόγως με την εφαρμογή και το σκοπό που έχει η «αποκάλυψη» της ακολουθίας των καταστάσεων, επιλέγεται το πιο κατάλληλο.

ΠΡΟΒΛΗΜΑ Γ

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

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

ΛΥΣΗ ΣΤΟ ΠΡΟΒΛΗΜΑ Α

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

$Q = q_1 q_2 \ldots q_T$ ***(1)***

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

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

η οποία γράφεται και $P(O|Q,\lambda) = b_{q1}(O_1) \cdot b_{q2}(O_2) \ldots b_{qT}(O_T)$ ***(3)***

Η πιθανότητα να προκύψει μια τέτοια ακολουθία Q είναι :

$P(Q|\lambda) = \pi_{q1} a_{q1q2} a_{q2q3} \ldots a_{qT-1qT}$ ***(4)***

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

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

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

$$P(O|\lambda) = \sum_{allQ} P(O|Q,\lambda)P(Q,\lambda) = \sum_{q1,q2,\ldots,qT} \pi_{q1} b_{q2}(O_1) a_{q1q2} b_{q2}(O_2) \ldots a_{qT\text{-}1qT} b_{qT}(Q_T) \qquad (6)$$

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

Η ΕΜΠΡΟΣ – ΠΙΣΩ ΔΙΑΔΙΚΑΣΙΑ

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

$$\alpha_t(i) = P(O_1O_2\ldots O_t, q_t = S_i | \lambda)$$

η οποία είναι η πιθανότητα να έχει προκύψει, με δεδομένο το μοντέλο, το μέρος της ακολουθίας $O_1O_2\ldots O_t$ (μέχρι την χρονική στιγμή t) και κατάσταση την χρονική στιγμή t την Si . Μπορούμε να υπολογίσουμε επαγωγικά το αt(i) ως εξής:

1) Αρχικοποίηση

$\alpha 1(i) = \pi i bi(O1)$, $1 \le i \le N$

2) Επαγωγή

$$a_{t+1}(j) = \left[\sum_{j=1}^{N} a_t(i) aij\right] bj(O_{t+1}), \qquad 1 \le j \le N, \qquad 1 \le t \le T-1$$

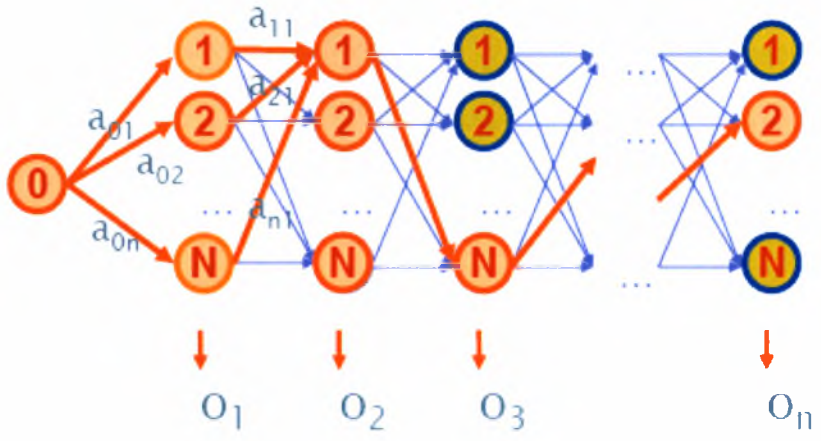
3) Τερματισμός

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

Το βήμα *1* αρχικοποιεί την προς τα εμπρός πιθανότητα ως την δεσμευμένη πιθανότητα της κατάστασης Si όταν η πρώτη παρατήρηση είναι η O_1. Το βήμα της επαγωγής αποτελεί και την «καρδία» της διαδικασίας . Η κατάσταση Sj ,την χρονική στιγμή t+1 ,μπορεί να προέλθει από τις N δυνατές καταστάσεις Si , $1 \leq i \leq N$ της χρονικής στιγμής t. Αφού η μεταβλητή $\alpha_t(i)$ είναι η δεσμευμένη πιθανότητα να έχουμε ακολουθία εξόδου $O_1O_2\ldots O_t$ όταν η κατάσταση την χρονική στιγμή t είναι η S_i , το γινόμενο $\alpha_t(i) * a_{ij}$ είναι η πιθανότητα να έχουμε ακολουθία εξόδου $O_1O_2\ldots O_t$ και η κατάσταση S_j την χρονική στιγμή t+1 έχει προέλθει από την κατάσταση S_i της χρονικής στιγμής t. Αθροίζοντας αυτό το γινόμενο πάνω στις N πιθανές καταστάσεις S_i την χρονική στιγμή t , με $1 \leq i \leq N$, μας δίνει την πιθανότητα , την χρονική στιγμή t+1, να προκύψει η κατάσταση S_j με δεδομένη τη μέχρι εκείνη τη στιγμή ακολουθία εξόδου. Αφού υπολογιστεί το S_j , το $\alpha_{t+1}(j)$ υπολογίζεται εύκολα με πολλαπλασιασμό του αθροίσματος που έχει προκύψει με την πιθανότητα στην κατάσταση j να έχουμε την έξοδο O_{t+1} δηλ. $b_j(O_{t+1})$. Ο παραπάνω υπολογισμός γίνεται για όλες τις καταστάσεις j ,με $1 \leq j \leq N$, για δεδομένο t και εν συνεχεία επαναλαμβάνεται για t=1,2,…,T-1 . Τέλος, το τρίτο βήμα μας δίνει την επιθυμητή τιμή της $P(O \mid \lambda)$ σαν το άθροισμα όλων των τελικών προς τα εμπρός μεταβλητών $\alpha_T(i)$.

Η αποδοτικότητα του συγκεκριμένου αλγόριθμου είναι σαφώς καλύτερη από αυτή του προηγούμενου. Οι υπολογισμοί που χρειάζονται είναι της τάξεως του N2*T, πράγμα το οποίο σημαίνει ότι για μοντέλο ίδιο με το προηγούμενο (N=5 και Tα=100) χρειάζονται 3000 υπολογισμοί (αντί για 10^{72}). Το γεγονός που αυξάνει την αποδοτικότητα του αλγόριθμου είναι ότι αφού υπάρχουν μόνο N καταστάσεις, όλες οι δυνατές ακολουθίες θα προκύπτουν από τις ίδιες N καταστάσεις κάθε χρονική στιγμή. Έτσι την χρονική στιγμή t=1 υπολογίζουμε τις τιμές $\alpha_t(i)$ με $1 \leq i \leq N$. Τις χρονικές στιγμές t=2,3,…,T χρειάζεται να υπολογίσουμε μόνο τις τιμές $\alpha_t(j)$ με $1 \leq j \leq N$, και ο κάθε υπολογισμός περιλαμβάνει μόνο N προηγούμενες τιμές του $\alpha_{t-1}(i)$ γιατί καθεμιά από τις N καταστάσεις είναι η επόμενη κάποιας από τις N ,ίδιες , καταστάσεις της προηγούμενης χρονικής στιγμής.

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



Εικόνα 15 : Η εμπρός διαδικασία(forward algorithm) . [6]

ΛΥΣΗ ΣΤΟ ΠΡΟΒΛΗΜΑ Β

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

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

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

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

αφού το $\alpha_t(i)$ περιλαμβάνει το μέρος της ακολουθίας μέχρι την χρονική στιγμή t ($O_1O_2\ldots O_t$) ενώ το $\beta_t(i)$ περιλαμβάνει την εναπομένουσα ακολουθία μέχρι το T ($O_{t+1}O_{t+2}\ldots O_T$) , με δεδομένη την κατάσταση S_i την χρονική στιγμή t. Ο παράγοντας κανονικοποίησης $\sum_{i=1}^{N} a_t(i)\beta_t(i)$ κάνει την $\gamma_t(i)$ μετρήσιμη, έτσι ώστε $\sum_{i=1}^{N} \gamma_t(i) = 1$. Χρησιμοποιώντας το $\gamma_t(i)$ μπορούμε να λύσουμε το πρόβλημα εύρεσης της πιο πιθανής κατάστασης q_t την χρονική στιγμή t ως εξής :

$q_t = \text{argmax}\,[\gamma_t(i)], \qquad 1 \le i \le N, \qquad 1 \le t \le T \qquad$ *(7)*

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

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

ΑΛΓΟΡΙΘΜΟΣ VITERBI

Ορίζουμε την ποσότητα: $\delta_t(i) = \max_{q1,q2,\ldots,qt-1} P[\, q1q2\ldots qt = i\, ,O_1O_2\ldots O_t|\lambda]$ δηλαδή $\delta_{t+1}(i)$ είναι η μεγαλύτερη πιθανότητα κατά μήκος ενός απλού μονοπατιού την χρονική στιγμή t, που περιλαμβάνει τις πρώτες t παρατηρήσεις και τελειώνει στην κατάσταση S_i. Με επαγωγή έχουμε :

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

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

1) *Αρχικοποίηση*

$\delta_1(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N$

$\psi_1(i) = 0$

2) *Επανάληψη*

$\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i)a_{ij}] \cdot b_j(O_t), \qquad 2 \le t \le T$

$1 \le j \le T$

$\psi_t(j) = \text{argmax}_{1 \le j \le N} [\delta_{t-1}(i)a_{ij}], \qquad 2 \le t \le T$

$1 \le j \le T$

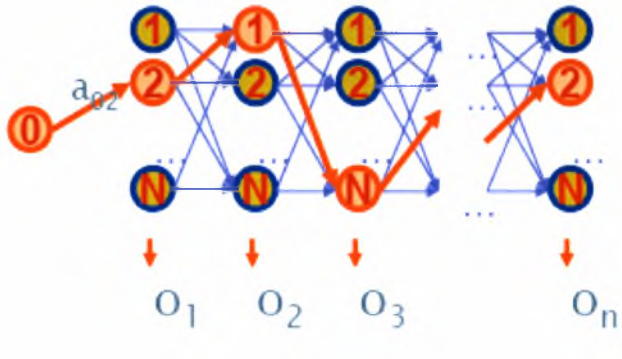
3)*Τερματισμός*

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

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

4)Εξαγωγή του μονοπατιού

$q^*_T = \psi_{t+1}(q^*_{t+1}), \; t = T-1, T-2, \ldots, 1.$



Εικόνα 16 : Ο αλγόριθμος VITERBI με δυναμικό προγραμματισμό . [6]

ΛΥΣΗ ΣΤΟ ΠΡΟΒΛΗΜΑ Γ

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

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

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

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

$$\xi_t(i,j) = \frac{a_t(i)a_{ij}b_j(o_{t+1})b_{t+1}(j)}{P(O \mid \lambda)}$$

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

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

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

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

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

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

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

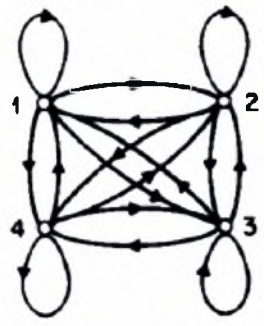
$$b'_j(k) = \frac{\text{αναμενόμενη πιθανότητα της κατάστασης } s_j \text{ και ταυτόχρονη παρατήρηση του συμβόλου } v_k}{\text{αναμενόμενη πιθανότητα της } s_j}$$

π'_i = αναμενόμενη συχνότητα στην κατάσταση s_i τη χρονική στιγμή t=1

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

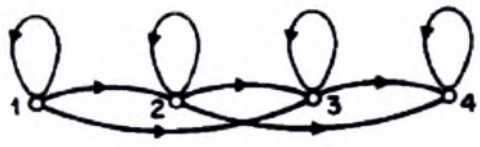
3.4 Τύποι HMM

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



Εικόνα 17 : Εργοδικό Μοντέλο HMM με 4 καταστάσεις . [7]

Για μερικές εφαρμογές , όπως είναι η αναγνώριση φωνής καθώς και η αναγνώριση χειρονομιών, έχει βρεθεί ότι άλλοι τύποι HMM μοντελοποιούν καλύτερα το φυσικό σήμα απ' ότι ένα εργοδικό μοντέλο. Ένα τέτοιο μοντέλο φαίνεται στην παρακάτω εικόνα.



Εικόνα 18 : Μοντέλο Bakis με 4 καταστάσεις . [7]

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

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

καθώς η ακολουθία πρέπει να ξεκινάει από την κατάσταση 1 (και να τελειώνει στην κατάσταση N).Πολλές φορές στα Bakis μοντέλα τίθενται επιπρόσθετοι περιορισμοί στους συντελεστές μετάβασης ώστε να εξασφαλιστεί ότι δεν θα συμβαίνουν μεγάλες

μεταβάσεις. Ένας τέτοιος περιορισμός είναι πχ. $a_{ij} = 0$, $j > i + \Delta$. Για το μοντέλο της εικόνας 11 , είναι $\Delta=2$ και ο πίνακας μετάβασης είναι ο παρακάτω :

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

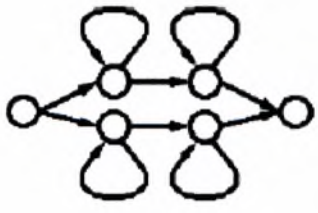
Εικόνα 19 : Πίνακας μετάβασης για μοντέλο Bakis με 4 καταστάσεις. [8]

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

$a_{NN} = 1$

$a_{Ni} = 0$, για $i < N$

Παρόλο που διαχωρίσαμε τα HMM σε εργοδικά και δεξιά – αριστερά μοντέλα, υπάρχουν ακόμα πολλοί άλλοι τύποι και συνδυασμοί. Όπως το HMM της εικόνας 24, που δείχνει ένα HMM το οποίο δημιουργείται με την παράλληλη σύνδεση δύο Bakis μοντέλων. Το μοντέλο αυτό υπακούει μεν στους περιορισμούς των Bakis μοντέλων για τους συντελεστές μετάβασης, αλλά έχει επιπλέον χαρακτηριστικά που δεν συναντώνται στα Bakis μοντέλα.



Εικόνα 20 : Παράλληλη σύνδεση δύο μοντέλων Bakis. [7 Fo5]

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

3.5 Συνεχείς Ποσότητες Παρατήρησης στα HMM

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

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

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

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

$$\sum_{m=1}^{M} c_{jm} = 1, \qquad 1 \le j \le N$$

$$c_{jm} \ge 0, \qquad 1 \le j \le N,\ 1 \le m \le M$$

έτσι ώστε να ισχύει για την κατανομή η γνωστή κανονικοποίηση

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

Η παραπάνω συνάρτηση πυκνότητας πιθανότητας μπορεί να χρησιμοποιηθεί για την προσέγγιση σχεδόν οποιασδήποτε κατανομής και άρα είναι ιδιαιτέρως πολλές οι εφαρμογές. Μπορεί να αποδειχτεί ότι οι τύποι για την προσέγγιση των συντελεστών των μειγμάτων της κατανομής δηλ. για τα $\mathbf{c_{jm}}$, $\boldsymbol{\mu_{jk}}$, $\mathbf{U_{jk}}$ είναι οι ακόλουθοι:

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

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

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

όπου $\boldsymbol{\gamma_t(j,k)}$ είναι η πιθανότητα τη χρονική στιγμή t να έχουμε την κατάσταση j και το **κ** μείγμα να συμβάλει στην δημιουργία του $\mathbf{O_t}$ δηλαδή :

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

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

ΚΕΦΑΛΑΙΟ 4

ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ

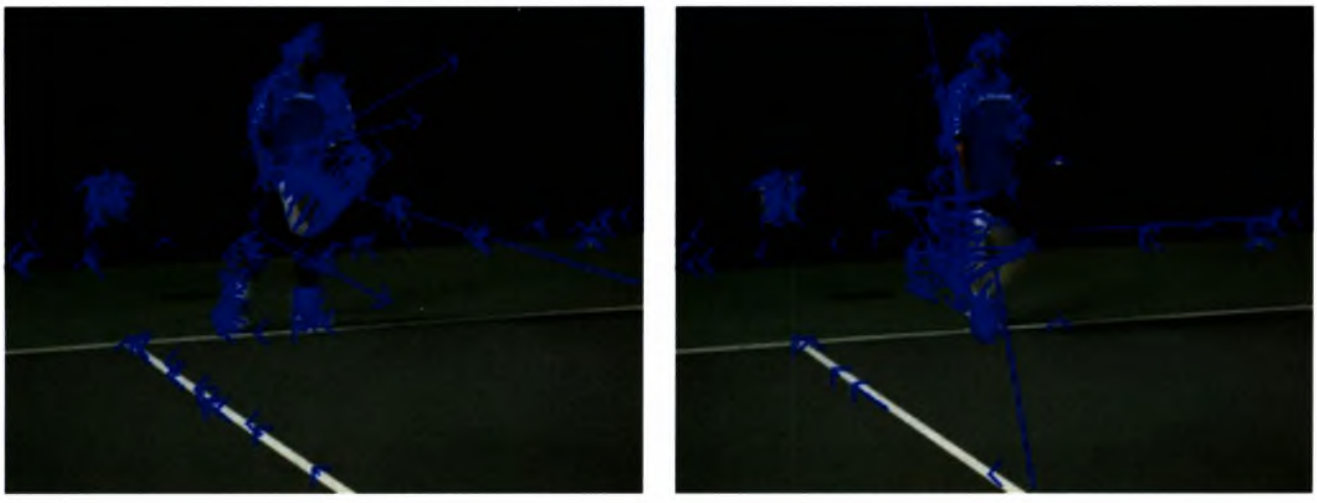
4.1 Γενικά χαρακτηριστικά του συστήματος

Στα πλαίσια της παρούσας διπλωματικής εργασίας υλοποιήθηκε ένα σύστημα αναγνώρισης των ανθρώπινων κινήσεων σε μια κοντινή απόσταση, όπου κάθε άτομο μπορεί να έχει «ύψος», ας πούμε, περίπου στα 300 pixel. Εισάγουμε τη μέθοδο εξαγωγής διανυσμάτων κίνησης με χρήση πεδίων οπτικής ροής(optical flow fields), που εφαρμόζονται σε κάθε ακολουθία βίντεο με ανθρώπινη φιγούρα, και συσχετίζουμε ένα κριτήριο ομοιότητας με βάσει τα Hidden Markov Models. Το κριτήριο αυτό προέρχεται από εκτίμηση των παραμέτρων των HMM, δοθέντων κάποιων αρχικά ορισμένων καταστάσεων και των διανυσμάτων κίνησης. Οι πίνακες μετάβασης μεταξύ των καταστάσεων του video sequence είναι και το κύριο χαρακτηριστικό που χρησιμοποιείται για την ταξινόμηση των κινήσεων.

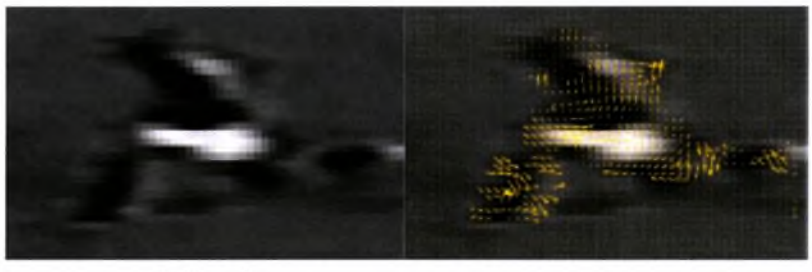
Το στοιχείο-κλειδί στην περίπτωσή μας είναι να περιορίσουμε την ανάκτηση μετρήσεων οπτικής ροής με αυξημένο θόρυβο. Στο σύστημά μας, για να αντιμετωπιστεί το συγκεκριμένο πρόβλημα, η μέθοδος που υλοποιήθηκε έγκειται σε ένα εναλλακτικό τρόπο υπολογισμού των διανυσμάτων ταχύτητας. Ειδικότερα, η εικόνα χωρίζεται σε μικρές περιοχές μεγέθους 3 x 3 (pixel window), και όχι για κάθε pixel χωριστά. Στη συνέχεια, το διάνυσμα της ταχύτητας υπολογίζεται χωριστά για κάθε περιοχή αθροίζοντας την τιμή για το κάθε pixel. Αν το πλάτος της ταχύτητας είναι μεγαλύτερο από το προτεινόμενο κατώφλι τότε αυτή η περιοχή ανήκει στο προσκήνιο, σε αντίθεση περίπτωση ανήκει στο παρασκήνιο.

Ξεκινάμε εντοπίζοντας και απομονώνοντας την ανθρώπινη φιγούρα. Για το λόγο αυτό είναι το αρχικό-πρότυπο βίντεο να έχει εξομαλυμένες κινήσεις, όσο πιο σταθερή κάμερα γίνεται και επίσης η κινούμενη φιγούρα να είναι όσο το δυνατότερο πιο κεντρικά στο συνολικό πλάνο. Κάθε σχετική κίνηση που προκύπτει εντός του πεδίου ανίχνευσης μέσα στο βίντεο αντιστοιχεί σε φυσικές κινήσεις των άκρων, του

κεφαλιού και του κορμού της φιγούρας. Θα αναλύσουμε αυτή την κίνηση μέσω του υπολογισμού της οπτικής ροής και προβολής της μέσα από ένα αριθμό διανυσμάτων. Η αναγνώριση θα γίνει λαμβάνοντας υπόψη τα τμήματα των εκάστοτε frame όπου το φάσμα των διανυσμάτων κίνησης είναι πιο έντονο.



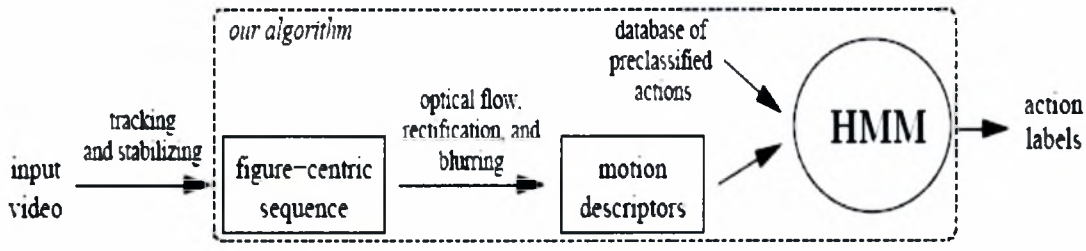
Εικόνα 21 : Optical flow με χρήση του αλγορίθμου Lucas – Kanade στην OpenCV. Παρατηρείται έντονο φάσμα των διανυσμάτων κίνησης γύρω από τη φιγούρα του παίκτη, καθώς επίσης και η ύπαρξη θορύβου.



(a) original image (b) optical flow $F_{x,y}$

Εικόνα 22 : Εξαγωγή των διανυσμάτων κίνησης: (a) Original image,(b) Optical flow

Έχοντας μία συγκεκριμένη βάση με πρότυπα βίντεο «εκπαίδευσης», καταγράφουμε και ονοματίζουμε τις διακεκριμένες καταστάσεις. Δε χρησιμοποιούμε την κίνηση που εμφανίζεται σε όλη την εικόνα παρά μόνο σε ένα κεντρικό «παράθυρο» (central window) όπου δραστηριοποιείται κυρίως η κινούμενη φιγούρα. Οποιοσδήποτε αριθμός πρότυπων βίντεο ανίχνευσης είναι επιθυμητός. Η μοναδική απαίτηση του συστήματος είναι ένα ανθρώπινο σώμα με συγκεκριμένη δομή να ταιριάζει σχετικά πάντα με τη μεμονωμένη φιγούρα που μας ενδιαφέρει.



Εικόνα 23 : Ροή δεδομένων για το σύστημα [3]

Μετά τη δημιουργία της βάσης των πρότυπων-βίντεο ανίχνευσης που θα χρησιμοποιηθούν στα HMM, το πιο σημαντικό που πρέπει να ληφθεί υπόψη είναι ποια στοιχεία – κλειδιά θα λάβουμε υπόψη στον υπολογισμό των διανυσμάτων κίνησης.

4.2 Συναρτήσεις και Υλοποίηση

Στην παρούσα εργασία χρησιμοποιήθηκε αποκλειστικά η βιβλιοθήκη της OpenCV (**Open** Source **C**omputer **V**ision Library) της Intel. Όπως αναφέραμε και στο κεφάλαιο 2, στο σύστημα μας, υλοποιείται η πιο δημοφιλής διαφορική μέθοδος δύο frame των Lucas-Kanade, γιατί είναι η πιο ανθεκτική στην παρουσία θορύβου. Ειδικότερα, χρησιμοποιήθηκε η συνάρτηση :

cvCalcOpticalFlowLK(srca, srcb, winsize, velx, vely)

Σε αυτή τη μέθοδο, θεωρείται ότι η οπτική ροή είναι τοπικά σταθερή. Επομένως, μέσα σε ένα μικρό παράθυρο μεγέθους w x w(w>1, με κέντρο το pixel P(x,y)),υπολογίζεται μία σειρά εξισώσεων για το κάθε pixel.

Για την δημιουργία των HMM χρησιμοποιήθηκε η συνάρτηση :

void cvCreate2DHMM(CvEHMM** hmm, int* state_number, int* num_mix, int obs_size)

Για την «εκπαίδευση» του συνόλου HMM (υπολογισμός παραμέτρων) η :

void cvEstimateHMMStateParams(CvImgObsInfo** obs_info_array,int num_img, CvEHMM* hmm)

Και τέλος για την υλοποίηση του αλγορίθμου Viterbi για τον υπολογισμό των καταστάσεων των βίντεο προς ταίριασμα, χρησιμοποιήθηκε η :

Float cvEViterbi(CvImgObsInfo* obs_info,CvEHMM* hmm)

Εικόνα 24 : Ιδανικό ταίριασμα για ακολουθία βίντεο τένις [3]

ΚΕΦΑΛΑΙΟ 5

ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΤΩΝ ΑΝΘΡΩΠΙΝΩΝ ΔΡΑΣΤΗΡΙΟΤΗΤΩΝ

ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο αυτό θα γίνει προσπάθεια καταγραφής των κύριων ζητημάτων που ανακύπτουν κατά τη μοντελοποίηση των ανθρώπινων δραστηριοτήτων, θέμα που απασχόλησε σημαντικά τόσο την ανάπτυξη των μεθόδων μας, όσο και τα αποτελέσματα εκτέλεσης τους. Συνοπτικά αναφέρουμε τι έχει ήδη δοκιμαστεί όσων αφορά την ακριβή καταγραφή και αναπαράσταση των ανθρώπινων κινήσεων, όπως επίσης και πού έγκειται η αδυναμία υλοποίησης μεθόδων εντοπισμού τους.

5.1 Μοντελοποίηση του Ανθρωπίνου Σώματος

Η ανάλυση της κίνησης του ανθρώπινου σώματος αποτελεί μέρος της αναγνώρισης κίνησης στα πλαίσια της όρασης υπολογιστών. Πολλές μελέτες στην ανάλυση κίνησης χρησιμοποιούν τον διαχωρισμό της κίνησης σε άκαμπτες και μη, βασιζόμενες στο κατά πόσο το αντικείμενο που κινείται είναι στερεό ή όχι. Μια άλλη προσέγγιση είναι η θεώρηση της ανθρώπινης κίνησης σαν αρθρωτή κίνηση, η οποία βεβαίως μπορεί να θεωρηθεί υποκατηγορία της μη άκαμπτης κίνησης. Η αρθρωτή κίνηση του ανθρώπινου σώματος αποτελείται από μικρότερες άκαμπτες κινήσεις των ξεχωριστών μελών του σώματος, η ολική όμως κίνηση είναι μη άκαμπτη. Μπορούμε να κατατάξουμε τις μελέτες που χρησιμοποιούν την προσέγγιση της αρθρωτής κίνησης σε αυτές που χρησιμοποιούν μια a priori μορφή του μοντέλου και σε αυτές που δεν χρησιμοποιούν κάποιο μοντέλο αλλά βασίζονται στην εικόνα του αντικειμένου. Ο διαχωρισμός βασίζεται στο κατά πόσο χρησιμοποιείται στην ανάλυση της κίνησης a priori γνώση για τη μορφή του αντικειμένου. Και οι δύο προσεγγίσεις έχουν πλεονεκτήματα και μειονεκτήματα.

Οι προσεγγίσεις που βασίζονται στην εικόνα του αντικειμένου μπορούν να έχουν εφαρμογή σε πιο ιδιαίτερες περιπτώσεις αφού δεν απαιτούν κάποιο συγκεκριμένο μοντέλο για το αντικείμενο. Παρόλα αυτά η προσέγγιση με βάση την εικόνα είναι πιο ευαίσθητη στο θόρυβο αφού δεν διαθέτει κάποιο μηχανισμό

διαχωρισμού του θορύβου από το οπτικό σήμα εισόδου. Από την άλλη η προσέγγιση που βασίζεται σε γνώση του μοντέλου του κινούμενου αντικειμένου μπορεί να συνδυάσει αποδοτικά τη γνώση του σχήματος και την είσοδο του οπτικού σήματος, κάνοντάς την καλύτερη να αναγνωρίζει υψηλού – επιπέδου, πολύπλοκες κινήσεις. Το βασικό μειονέκτημα αυτού του είδους προσέγγισης είναι ότι προϋποθέτει επιπλέον βήματα επεξεργασίας και επιλογής του μοντέλου και εκτίμηση των παραμέτρων του μοντέλου ώστε να ταιριάξει στην εκάστοτε οπτική είσοδο. Επίσης, για την προσθήκη κάποιας νέας ενέργειας ή κίνησης μπορεί να χρειαστούν σημαντικής δυσκολίας υπολογισμοί για την αναβάθμιση του μοντέλου.

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

Σε κάθε προσέγγιση, το ανθρώπινο σώμα μπορεί να αναπαρασταθεί με διάφορα επίπεδα λεπτομερειών, τα οποία περιλαμβάνουν ευλύγιστα κουτιά, ραβδόμορφα σχήματα, δισδιάστατα χωρία ή τρισδιάστατους όγκους, ανάλογα με την πολυπλοκότητα του μοντέλου που απαιτεί η συγκεκριμένη εφαρμογή. Η αναπαράσταση με βάση τα ευλύγιστα κουτιά είναι ένα από τα πιο απλά μοντέλα του ανθρωπίνου σώματος. Η ικανότητα αναπαράστασης που προσφέρει είναι περιορισμένη. Το μοντέλο των ευλύγιστων κουτιών είναι χρήσιμο κυρίως σε ακολουθίες εικόνων που το ανθρώπινο σώμα είναι τόσο μικρό ώστε να καταλαμβάνει μόνο μερικά pixels. Η αναπαράσταση με βάση ραβδόμορφα σχήματα θεωρεί το σώμα σαν μια σύνθεση από παραλληλόγραμμα (ραβδιά) και τις μεταξύ τους αρθρώσεις. Η θεώρηση αυτή βασίζεται στην παρατήρηση ότι η κίνηση του ανθρώπινου σώματος προέρχεται κυρίως από την κίνηση των οστών (τα οποία ουσιαστικά αναπαριστούνται με τα ραβδιά). Η αναπαράσταση με δισδιάστατα χωρία βασίζεται στην ιδέα ότι το ανθρώπινο σώμα που εμφανίζεται στην εικόνα είναι η προβολή του τρισδιάστατου ανθρωπίνου σώματος σε ένα δισδιάστατο χώρο. Για το λόγο αυτό προσεγγίζει το σώμα με χωρία χωρίς καθορισμένα όρια. Τα τρισδιάστατα, ογκομετρικά μοντέλα προσπαθούν να περιγράψουν με λεπτομέρειες το τρισδιάστατο ανθρώπινο σώμα με

χρήση πολύεδρων στερεών όπως ελλειπτικούς κυλίνδρους, γενικευμένους κώνους ή σφαίρες.

Εικόνα 25 : Αποτυχία αναγνώρισης ανθρώπινης δραστηριότητας από μοντέλα με πολλές κάμερες και επιτυχής ανίχνευση της με χρήση HMM μοντέλων [1].

Τα παραπάνω μοντέλα αναφέρθηκαν με σειρά αυξανόμενης πολυπλοκότητας αλλά και αυξανόμενου επιπέδου λεπτομερειών. Τα πιο λεπτομερή μοντέλα μπορούν να αναπαραστήσουν πιο πολύπλοκες πλευρές της ανθρώπινης δραστηριότητας, αλλά απαιτούν σημαντικά μεγαλύτερη υπολογιστική πολυπλοκότητα. Για παράδειγμα, τα τρισδιάστατα μοντέλα απαιτούν στερεομετρικά χαρακτηριστικά τα οποία για να αποκτηθούν χρειάζονται πολλές κάμερες.

Το επίπεδο των λεπτομερειών που χρειάζεται για την αναπαράσταση του ανθρωπίνου σώματος, εξαρτάται από την εφαρμογή. Για παράδειγμα, σε μερικές εφαρμογές δεν υπάρχει ανάγκη για αναπαράσταση ολόκληρου του σώματος ή για αναπαράσταση πολλών λεπτομερειών σε ορισμένα σημεία αυτού. Σε τέτοιες περιπτώσεις μια πιο απλή προσέγγιση μπορεί να δώσει ικανοποιητικά αποτελέσματα. Ένα τέτοιο παράδειγμα υλοποίησης είναι η μοντελοποίηση κάθε παίκτη ενός αγώνα

αμερικάνικου ποδοσφαίρου με χρήση ενός ευλύγιστου κουτιού. Ο εντοπισμός γινόταν με το να διατηρείται το κουτί στην εικόνα διαμέσου των frames.

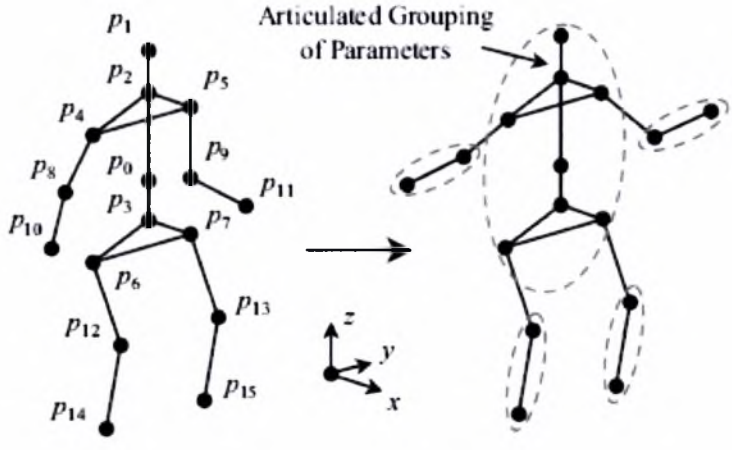
Εικόνα 26 : Εικόνα από αγώνα Παγκόσμιου Πρωταθλήματος ποδοσφαίρου. Είναι εύκολο να ξεχωρίσει κανείς τους παίχτες παρόλο που είναι μικρή η ανάλυση(κουτί με ζουμ στην αριστερή κάτω γωνία). [3]

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

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

ταχύτητα των κινούμενων pixel. Όσα Pixel κινούνται με παρόμοια ταχύτητα θεωρείται πως ανήκουν στο ίδιο μέλος του ανθρωπίνου σώματος.

Επίσης σε διάφορες μελέτες έχουν χρησιμοποιηθεί χαρακτηριστικά όπως ο όγκος, οι ακμές, ο ήχος, το χρώμα, η ταχύτητα της οπτικής ροής και η ένταση για την εξαγωγή της δραστηριότητας. Ένα τέτοιο παράδειγμα είναι ο συνδυασμός της έντασης των pixel με τον ήχο της μουσικής για την αναγνώριση των διάφορων ειδών χορού από ζευγάρια ανθρώπων σε κοινωνικές εκδηλώσεις. Σε μια πρόσφατη μελέτη χρησιμοποιήθηκαν πολλαπλά επίπεδα επεξεργασίας της εικόνας (επίπεδο pixel, μικρών μερών του σώματος, μελών, ακολουθία εικόνων) για τη δημιουργία ενός μοντέλου του ανθρωπίνου σώματος, που βασίζεται στην «εικόνα του αντικειμένου». Τα pixel ομαδοποιούνται ανάλογα με την ομοιομορφία στο χρώμα τους και στη συνέχεια οι ομάδες που δημιουργούνται, ενώνονται για να αποτελέσουν κάποιο μέλος (πχ. τα κεφάλι, το στήθος, τη λεκάνη, το πρόσωπο, τα μαλλιά, τα χέρια και τα πόδια). Τέλος τα μέλη του σώματος εντοπίζονται κατά μήκος της ακολουθίας των εικόνων.



Εικόνα 27 : Δομική αναπαράσταση του ανθρώπινου σώματος. [1]

5.2 Επίπεδα Λεπτομερειών

Μεγάλο μέρος της έρευνας πάνω στην αναγνώριση των πράξεων έχει επικεντρωθεί στην ανάλυση των δραστηριοτήτων ενός μόνο ατόμου. Η αναγνώριση όμως των αλληλεπιδράσεων περιλαμβάνει αλληλεπιδράσεις μεταξύ δύο ατόμων, μεταξύ μιας ομάδας τριών ή περισσοτέρων ατόμων, μεταξύ ανθρώπου και υπολογιστή, μεταξύ

ανθρώπου και κάποιου αντικειμένου. Γενικά, κάθε μία από τις παραπάνω περιπτώσεις απαιτεί διαφορετικό επίπεδο ανάλυσης της εικόνας και διαφορετικό τρόπο αναπαράστασης του προς αναγνώριση γεγονότος. Όσο περισσότεροι άνθρωποι περιλαμβάνονται στην εικόνα τόσο λιγότερα pixel θα καταλαμβάνει ο καθένας από αυτούς, συντελώντας έτσι σε μια εικόνα χαμηλής ανάλυσης. Για το λόγο αυτό κάθε περίπτωση χρειάζεται και διαφορετικές μεθόδους επεξεργασίας. Η αναγνώριση των πράξεων και των αλληλεπιδράσεων μπορεί να επιτευχθεί με διαφορετικά επίπεδα λεπτομερειών κατά την ανάλυση. Ειδικότερα, τα επίπεδα αυτά είναι το *γενικό*, το *ενδιάμεσο* και το *λεπτομερές* επίπεδο.

Στο γενικό επίπεδο, κάθε άτομο αναπαρίσταται με ένα ξεχωριστό κινούμενο, ευλύγιστο κουτί ή έλλειψη. Σε αυτό το επίπεδο, η αναγνώριση των αλληλεπιδράσεων του ανθρώπου, περιορίζεται στο γενικό επίπεδο κατανόησης των προτύπων κίνησης αυτών των κουτιών ή ελλείψεων. Οι εφαρμογές παρακολούθησης χρησιμοποιούν πολλές φορές αυτό, το γενικό, επίπεδο. Ένα τέτοιο σύστημα, είναι ένα σύστημα αναγνώρισης των αλληλεπιδράσεων (συνάντηση, αποχαιρετισμός κτλ) δύο πεζών σε ένα δρόμο .Το σύστημα αναγνωρίζει κάθε έναν από τους πεζούς σαν ένα ευλύγιστο κουτί και κατόπιν κατατάσσει την κίνηση των δύο κουτιών σε κάποιο από τα υπάρχοντα πρότυπα. Ένα άλλο σύστημα που κάνει χρήση αυτού του γενικού επιπέδου λεπτομερειών, είναι ένα σύστημα το οποίο αναλύει μια ακολουθία εικόνων ανθρώπων, που παίζουν αμερικάνικο ποδόσφαιρο(βλ. εικόνα 6). Κάθε παίκτης αναπαριστάται με ένα ευλύγιστο ορθογώνιο κουτί. Η αναγνώριση επιτυγχάνεται μέσω της μελέτης της αλληλεπίδρασης των κουτιών, έχοντας υπ' όψιν τους κανόνες του αμερικάνικου ποδοσφαίρου, που αφορούν τις επιτρεπτές αλληλεπιδράσεις μεταξύ των παικτών.

Στο ενδιάμεσο επίπεδο λεπτομερειών, τα άτομα αναπαρίστανται με διακριτά τα βασικά μέρη του σώματός τους, όπως το κεφάλι, τον κορμό, τα χέρια και τα πόδια. Πολλές μέθοδοι έχουν προταθεί για την τμηματοποίηση των βασικότερων μελών του ανθρώπινου σώματος. Ένα σύστημα ενδιάμεσου επιπέδου λεπτομερειών είναι αυτό της αναγνώρισης πολλών ατόμων σε κάποιο χώρο. Εφαρμόζοντας διαδοχικές αφαιρέσεις του παρασκηνίου, επιτυγχάνεται η τμηματοποίηση των περιοχών του προσκήνιου στις οποίες βρίσκεται μια ομάδα ατόμων. Κατόπιν προβάλλεται μια δυαδική εικόνα του προσκήνιου προκειμένου να εντοπιστούν τα κεντροειδή των κεφαλιών του κάθε ατόμου της ομάδας.

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

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

Συνήθως η αλληλεπίδραση ανθρώπου και αντικειμένων απαιτεί ένα μόνο φορέα σε αντίθεση με την αλληλεπίδραση μεταξύ ανθρώπων, η οποία, προφανώς,

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

5.3 Μέθοδοι Αναγνώρισης Ανθρωπίνων Ενεργειών

Η αναγνώριση των ανθρώπινων ενεργειών επιτυγχάνεται με την ταξινόμηση των δεδομένων που προέρχονται από κάποιο βίντεο. Υπάρχουν δύο τρόποι για να επιτευχθεί η ταξινόμηση. Η απευθείας αναγνώριση και η αναγνώριση μέσω της ανακατασκευής. Στην περίπτωση της απευθείας αναγνώρισης οι ανθρώπινες ενέργειες αναγνωρίζονται απευθείας από τα δεδομένα της εικόνας, χωρίς την ανακατασκευή στιγμιότυπων του ανθρωπίνου σώματος. Για παράδειγμα, ένα σύστημα αναγνώρισης της συμπεριφοράς των πεζών (περπάτημα, τρέξιμο, συνάντηση δύο ανθρώπων). Το σύστημα αυτό δεν χρησιμοποιεί κάποιο μοντέλο για το ανθρώπινο σώμα αλλά την περιοδικότητα της κίνησης των ανθρώπων. Η περιοδικότητα αυτή είναι αποτέλεσμα της αιώρησης των χεριών και των ποδιών κατά το βάδισμα. Κάθε ακολουθία εικόνων θεωρείται πως είναι ένας χώρος με τρεις διαστάσεις, δύο χωρικές και μία χρονική. Οι επαναλαμβανόμενες κινήσεις των πεζών στην εικόνα, γεννούν κάποιες καμπύλες στον τρισδιάστατο χώρο. Η αναγνώριση των ενεργειών των ανθρώπων επιτυγχάνεται με σύγκριση των καμπυλών που προέρχονται από μια ακολουθία εισόδου με τις καμπύλες αναφοράς που υπάρχουν για την κάθε ανθρώπινη δραστηριότητα. Από την άλλη, στην περίπτωση της αναγνώρισης μέσω

της ανακατασκευής , πρώτα γίνεται η ανακατασκευή του αντικειμένου με βάση την εικόνα και μετά γίνεται η αναγνώριση των ανθρωπίνων ενεργειών. Παράδειγμα χρησιμοποίησης της παραπάνω μεθόδου, είναι ένα σύστημα αναγνώρισης του γεγονότος «ψηλά τα χέρια» κατά την διάρκεια μιας ληστείας. Η αναγνώριση επιτυγχάνεται αφού πρώτα γίνει η τμηματοποίηση του σώματος σε επιμέρους μέλη και κατόπιν εξεταστεί η σχετική θέση των χεριών.

Εικόνα 28 : Υφή Κίνησης (Motion Texture.) Ένα στατιστικό μοντέλο δύο επιπέδων για τη σύνθεση της κίνησης ενός χαρακτήρα [6]

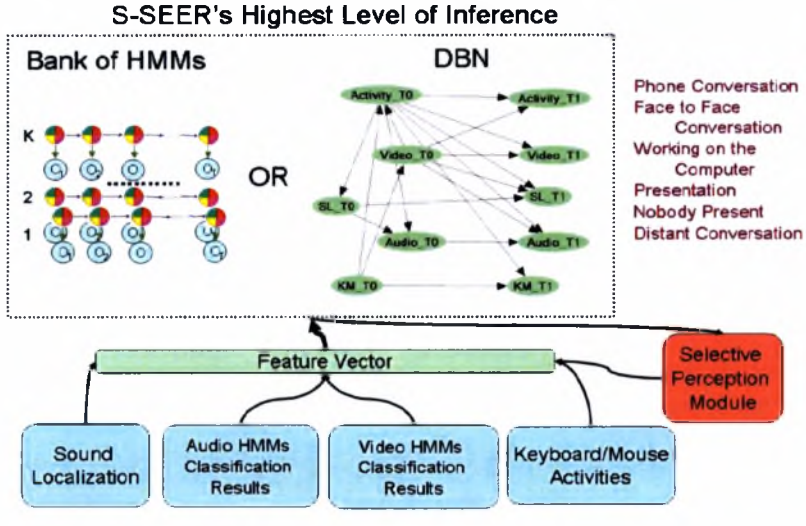
Υπάρχουν όμως και κάποια υβριδικά συστήματα, τα οποία δεν υιοθετούν αυστηρά την μια ή την άλλη μέθοδο αλλά δανείζονται χαρακτηριστικά και των δύο. Ένας άλλος τρόπος διαχωρισμού των μεθόδων αναγνώρισης είναι σε στατικές και δυναμικές. Η αναγνώριση της ανθρώπινης κίνησης προϋποθέτει την ανάλυση μιας σειράς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

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

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

αναπαράσταση ενεργειών από την DTW λόγω της ικανότητας που τους παρέχει η στοχαστική τους φύση, να συμπεριλάβουν και να διαχειριστούν την αβεβαιότητα η οποία υπάρχει σε όλες τις κινήσεις, ενέργειες και χειρονομίες μίας ακολουθίας εικόνων. Τα μοντέλα HMM έχουν την ικανότητα να χειρίζονται χρονικά μεταβαλλόμενα δεδομένα ανεξάρτητα από την κλίμακα του χρόνου. Η ιδιότητα τους αυτή είναι χρήσιμη στην ταξινόμηση εκφράσεων από ακολουθίες βίντεο. Για το σκοπό αυτό μπορούν να τροφοδοτηθούν με κάποιο διάνυσμα αναπαράστασης το οποίο εξάγεται από την κίνηση των διαφόρων σημείων του προσώπου.

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



Εικόνα 29 : Διαβάθμιση HMM επιπέδων και έμφαση στο ανώτερο επίπεδο [4]

ΚΕΦΑΛΑΙΟ 6
ΑΝΘΡΩΠΙΝΗ ΚΙΝΗΣΗ ΚΑΙ ΧΕΙΡΟΝΟΜΙΕΣ

ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο καταθέτουμε μια εξειδικευμένη εκδοχή των ανθρώπινων δραστηριοτήτων, τις χειρονομίες. Ειδικότερα, αναλύεται η φύση των ανθρωπίνων κινήσεων και χειρονομιών, πράγμα απαραίτητο για την σωστή σχεδίαση ενός συστήματος αναγνώρισης χειρονομιών. Δίνεται ένας «ορισμός» του όρου «ανθρώπινη χειρονομία» και επιχειρείται ο διαχωρισμός τους και η ταξινόμησή της ανάλογα με τα ιδιαίτερα χαρακτηριστικά που εμφανίζουν ορισμένες ομάδες χειρονομιών. Τέλος, παρατίθενται κάποια πρώτα στοιχεία που αφορούν τους τρόπους αλλά και τα προβλήματα που παρουσιάζονται κατά την αναγνώριση των χειρονομιών, συμπληρωματικά με το θέμα μου μελετάται στην παρούσα διπλωματική εργασία.

6.1 Αναγνώριση της ανθρώπινης κίνησης

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

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

Το πρόβλημα της ερμηνείας των ανθρώπινων κινήσεων με χρήση της όρασης υπολογιστών είναι σύνθετο αλλά και πολύ ενδιαφέρον. Επίσης, ένα ακόμα δύσκολο σημείο είναι ότι η ίδια χειρονομία μπορεί να διαφοροποιείται σημαντικά από πρόσωπο σε πρόσωπο, από πολιτισμό σε πολιτισμό και μερικές φορές ακόμη και όταν πραγματοποιείται πολλές φορές από το ίδιο άτομο. Η πολυπλοκότητα του προβλήματος έγκειται κυρίως στην τετρα-διάστατη φύση της ανθρώπινης κίνησης(δηλαδή χωρική και χρονική). Τρεις από τις διαστάσεις βρίσκονται στο χώρο, ενώ η τέταρτη είναι ο χρόνος και συσχετίζεται επομένως με τη δυναμική της κάθε κίνησης . Οι πληροφορίες που αφορούν την χρονική εξέλιξη μιας χειρονομίας είναι ιδιαιτέρως σημαντικές, δεδομένου ότι δείχνουν πού αρχίζει και πού τελειώνει κάθε χειρονομία. Χωρίς αυτές τις πληροφορίες, θα ήταν δύσκολο να διακριθεί η μία χειρονομία από την άλλη, δεδομένου ότι πολλές χειρονομίες περιλαμβάνουν σχεδόν όμοια κομμάτια ακολουθιών στις τροχιές τους. Παραδείγματος χάριν, οι χειρονομίες : «δείχνοντας προς τα δεξιά» και «χαιρετώντας με το δεξί χέρι», αρχίζουν με την ίδια κίνηση. Την ανύψωση δηλαδή του χεριού πάνωαπό το ύψος της μέσης και κατόπιν την κίνηση του χεριού και του μπράτσου πίσω στην αρχικές θέσεις τους. Τα ιδιαίτερα χαρακτηριστικά γνωρίσματα αυτών των δύο χειρονομιών, που τις διαφοροποιούν, μπορούν να εντοπισθούν μόνο στη μέση της τροχιάς των χειρονομιών αυτών. Ως εκ

τούτου μια μέθοδος τμηματοποίησης των χειρονομιών σύμφωνα με το που αρχίζουν και που τελειώνουν και ο διαχωρισμός τους στις επιμέρους φάσεις τις κινήσεώς τους , μπορεί να βοηθήσει σημαντικά τη διαδικασία αναγνώρισης .

Εικόνα 30 : Προσπάθεια ταξινόμησης δυναμικών χειρονομιών [5]

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

Ένα ακόμη χρήσιμο οπτικό στοιχείο θα μπορούσε να είναι το γύρισμα του κεφαλιού προς την κατεύθυνση που δείχνει η χειρονομία υπόδειξης. Η αναγνώριση χειρονομιών με βάση το περιεχόμενό τους έχει μελετηθεί από τους Sherrah και Gong, που χρησιμοποίησαν την θέση του κεφαλιού για να θέσουν περιορισμούς όσον αφορά την σημασία της χειρονομίας, σε ένα σύστημα αναγνώρισης χειρονομίας για τηλεδιάσκεψη. Τέτοιες λεπτομέρειες στις κινήσεις του ανθρώπινου σώματος μπορούν να εξαχθούν μόνο με τον εντοπισμό των συνδυασμών των κινήσεων των άκρων. Συχνά, οι μικρές διαφορές ανάμεσα στις χειρονομίες μπορούν να εντοπιστούν μόνο με βάση αυτούς τους συνδυασμούς κινήσεων των άκρων. Προφανώς, εάν εντοπιστούν όλες αυτές οι μικρές κινήσεις που λαμβάνουν χώρα κατά την εξέλιξη της χειρονομίας, είναι πολύ πιθανόν να έχουμε μια αποτελεσματικότερη αναγνώριση. Το ιδανικό θα ήταν η δημιουργία ενός συστήματος που θα αναγνωρίζει τις χειρονομίες που γίνονται σε πραγματικό χρόνο αλλά κάτι τέτοιο περιορίζει σημαντικά το χρόνο υπολογισμού που έχει στην διάθεσή του ο υπολογιστής για να διεκπεραιώσει την διαδικασία αναγνώρισης. Επομένως, είναι προτιμότερο να βρεθεί η χρυσή τομή ανάμεσα στα αντιμαχόμενα χαρακτηριστικά της καλής ακρίβειας, του μικρού χρόνου επεξεργασίας καθώς επίσης και της δυνατότητας του συστήματος να προσαρμοστεί στις αλλαγές στο θέμα ή το περιεχόμενο. Η μείωση του χρόνου υπολογισμού επιτυγχάνεται συχνά με κάποια μορφή εξαγωγής ή αναπαράστασης χαρακτηριστικών γνωρισμάτων των χειρονομιών, η οποία έχει ως στόχο την συμπύκνωση των πληροφοριών των κινήσεων και την απαλλαγή από δεδομένα τα οποία δεν χρησιμεύουν στην διαδικασία αναγνώρισης.

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

πως το πρόβλημα της ευφυούς ερμηνείας της ανθρώπινης συμπεριφοράς μπορεί να χωριστεί σε τρία στάδια. Την εξαγωγή και την αναπαράσταση των χαρακτηριστικών των κινήσεων, τον εντοπισμό της τροχιάς και το τελικό στάδιο της αναγνώρισης.

6.2 Οι Χειρονομίες

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

6.2.1 Βιολογικός και Κοινωνιολογικός Ορισμός

Ταξινόμηση των Χειρονομιών

Από βιολογική και κοινωνιολογική σκοπιά δεν μπορεί να υπάρξει κάποιος αυστηρός ορισμός για τις χειρονομίες.. Για το λόγο αυτό οι ερευνητές έχουν την δυνατότητα να απεικονίζουν και να ταξινομούν τις χειρονομίες σύμφωνα με τα υποκειμενικά τους κριτήρια. Η έρευνα πάνω στην αναγνώριση ομιλίας και γραφής έχει αποδώσει χρήσιμες μεθόδους σχεδίασης συστημάτων αναγνώρισης και κριτηρίων για την ταξινόμηση τέτοιων συστημάτων. Για το λόγο αυτό μελετώνται επίσης και συστήματα αναγνώρισης χειρονομιών, τα οποία χρησιμοποιούνται για τον έλεγχο συσκευών αποθήκευσης και απεικόνισης. Οι συσκευές αυτές μπορεί να βρίσκονται είτε στον ίδιο χώρο με τον χρήστη είτε σε κάποιο απομακρυσμένο περιβάλλον.

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

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

6.2.2 Ταξινόμηση χειρονομιών

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

Ο πρώτος διαχωρισμός είναι σε χειρονομίες ενέργειας και συμβολικές χειρονομίες. Οι χειρονομίες ενέργειας, όπως υποδηλώνει και το όνομά τους , δεν έχουν κάποια μεταφορική / συμβολική σημασία . Παραδείγματα τέτοιων χειρονομιών είναι το κόψιμο ξύλων ή το μέτρημα χρημάτων. Περιπτώσεις συμβολικών χειρονομιών είναι το σήμα του «ΟΚ» ή η το νόημα για οτοστόπ. Υπάρχουν όμως και

χειρονομίες ενέργειας στις οποίες είναι δυνατόν να αποδοθεί κάποιο συμβολικό νόημα(σημειογέννεση) όπως για παράδειγμα, σε μια κατασκοπική νουβέλα , όπου το αντικείμενο που κρατάει ένας πράκτορας στο χέρι του μπορεί να έχει πολύ σημαντικό νόημα. Ο διαχωρισμός αυτός δίνει την δυνατότητα στους ερευνητές να χρησιμοποιήσουν για τον έλεγχο συσκευών χειρονομίες οι οποίες ουσιαστικά αναπαριστούν κάποια πραγματική κίνηση.

Ο διαχωρισμός σε προφανείς και μη χειρονομίες αναφέρεται στην ευκολία με την οποία είναι δυνατόν οι υπόλοιποι να ερμηνεύσουν κάποια χειρονομία.. Ο όρος «προφανής χειρονομία» έχει συσχετιστεί με την έννοια της οικουμενικότητας, σύμφωνα με την οποία ορισμένες χειρονομίες έχουν καθορισμένο, διαπολιτισμικό νόημα. Στην πραγματικότητα το νόημα κάθε χειρονομίας εξαρτάται σε πολύ μεγάλο βαθμό από τον πολιτισμό μέσα στον οποίο αναπτύσσεται. Μέσα σε μια κοινωνία οι χειρονομίες έχουν καθορισμένα νοήματα, όμως δεν υπάρχει καμία γνωστή χειρονομία ή κίνηση του σώματος η οποία να έχει το ίδιο νόημα σε όλες τις κοινωνίες. Ακόμα και στην ASL μερικά νοήματα έχουν τόσο ξεκάθαρη σημασία που μπορεί να τα αναγνωρίσει και κάποιος ο οποίος δεν γνωρίζει την ASL. Αυτό σημαίνει πως οι χειρονομίες οι οποίες θα καθορίσουν την λειτουργία μιας συσκευής μπορούν να επιλεχθούν σχεδόν αυθαίρετα.

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

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

6.2.3 Τυπολογία Χειρονομιών

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

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

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

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

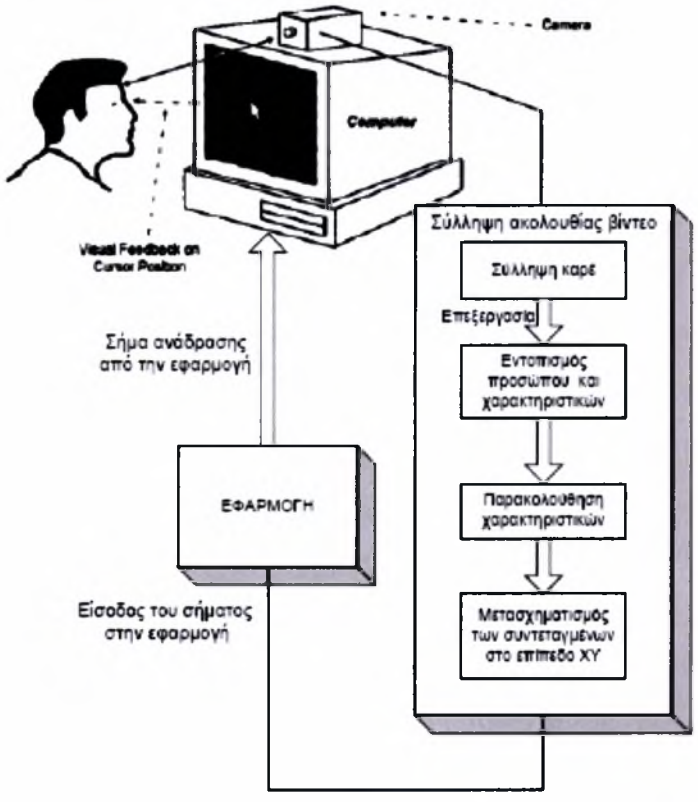
6.3 Αναγνώριση Φωνής και Κειμένου: Θέματα παράλληλα με την Αναγνώριση Χειρονομιών

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

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

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

- Βαθμός αναγνώρισης: Τα περισσότερα εμπορικά προϊόντα ισχυρίζονται πως επιτυγχάνουν βαθμούς αναγνώρισης τουλάχιστον 95%. Αν και το ποσοστό αυτό φαίνεται ιδιαιτέρως υψηλό, είναι μετρημένο σε συνθήκες εργαστηρίου. Χαρακτηριστικό είναι πως ο ίδιος ο άνθρωπος, παρουσιάζει ποσοστό αναγνώρισης περίπου 99.2%



Εικόνα 31 :

Μπλοκ διάγραμμα της επικοινωνίας ανθρώπου μηχανής μέσω της οπτικής πληροφορίας [5]

Τα συστήματα τα οποία έχουν την δυνατότητα να αναγνωρίζουν λέξεις από μεγάλο λεξιλόγιο χρησιμοποιούν τα Κρυφά Μαρκοβιανά Μοντέλα (HMM). Τα HMM χρησιμοποιούνται επίσης και από αρκετά συστήματα αναγνώρισης χειρονομιών. Σε μερικά συστήματα αναγνώρισης φωνής, οι καταστάσεις του HMM αντιστοιχούν στις φωνητικές μονάδες. Ο πίνακας με τις πιθανότητες μετάβασης καθορίζει ποια θα είναι η επόμενη κατάσταση. Ο όρος «κρυφό» αναφέρεται στο είδος του Μαρκοβιανού μοντέλου που χρησιμοποιείται. Σε αυτού του είδους τα μοντέλα οι παρατηρήσεις εξόδου αποτελούν μια στοχαστική διαδικασία που εξαρτάται από την κάθε κατάσταση. Ο πλήρης καθορισμός ενός Μαρκοβιανού μοντέλου απαιτεί τις παρακάτω πληροφορίες : την κατανομή πιθανότητας ανάμεσα στις καταστάσεις, την κατανομή πιθανότητας εξόδου κάθε κατάστασης, την κατανομή πιθανότητας της αρχικής κατάστασης. Κατά την διαδικασία αναγνώρισης φωνής, δημιουργείται ένα

HMM για κάθε λέξη η οποία περιέχεται στο λεξικό. Κατόπιν, για κάθε ακολουθία φωνητικών μονάδων η οποία αποτελεί την είσοδο, υπολογίζεται η πιθανότητα ,αυτή, να έχει προέλθει από το κάθε ένα από τα HMM του λεξικού.

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

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

6.4 Αναγνώριση Χειρονομιών με HMM

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

Κατ' αρχάς, οι χειρονομίες, όπως και ο προφορικός λόγος, ποικίλλουν ανάλογα με την τοποθεσία, το χρόνο, και τους κοινωνικούς παράγοντες. Δεύτερον, οι

κινήσεις του σώματος, όπως και οι διάφοροι ήχοι, κωδικοποιούν κάποιο συγκεκριμένο νόημα. Τρίτον, η διαδοχή των κινήσεων που κάνει κάποιος καθώς μιλάει, μπορεί να παρομοιαστεί με τους συντακτικούς κανόνες. Επομένως, οι γλωσσικές μέθοδοι μπορούν να χρησιμοποιηθούν στην αναγνώριση χειρονομίας.

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

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

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

ΚΕΦΑΛΑΙΟ 7

ΣΥΝΟΨΗ

7.1 Γενικά

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

Στην εργασία αυτή έχει σχεδιαστεί και έχει τεθεί σε εφαρμογή ένα σύστημα για την αναγνώριση των κινήσεων που κάνει ένας παίκτης τένις σε ένα video sequence. Η αναγνώριση γίνεται με τη βοήθεια στατιστικών μοντέλων, των HMM. Αρχικά ο χρήστης καλείται να επιλέξει ένα αρχικό, πρότυπο βίντεο και το σύστημα εκπαιδεύεται με ένα συγκεκριμένο σύνολο κινήσεων, τις οποίες θα μπορεί να αναγνωρίσει. Έπειτα το σύστημα υπολογίζει την οπτική ροή που εμφανίζεται στα αντίστοιχα frames και εξάγει τα διανύσματα ταχύτητας. Δοθέντος ενός νέου video sequence και βάσει των διανυσμάτων αυτών ταχύτητας που υπολογίστηκαν στο προηγούμενο βήμα, εκτιμάται η πιθανότητα η κίνηση που δόθηκε ως είσοδος να ανήκει σε καθεμία από τις εκπαιδευμένες κινήσεις. Η εργασία αυτή επικεντρώθηκε σε αναγνώριση κινήσεων από ακολουθίες βίντεο που ήδη είχαν τραβηχτεί και μάλιστα υπό συγκεκριμένες συνθήκες.

Για την αναγνώριση εφαρμόστηκε η μέθοδος του συνεχούς Κρυφού Μοντέλου Markov. Το συνεχές μοντέλο, έχει τη δυνατότητα να λαμβάνει συνεχείς τιμές ώστε να γίνεται πιο ευέλικτο για χρήση σε διαφορετικού μεγέθους εικόνες χωρίς την ανάγκη να οριστεί νέο codebook. Το σημαντικότερο μειονέκτημα του είναι ότι απαιτεί πολλά δεδομένα εκπαίδευσης για να λειτουργήσει επιτυχώς.

7.2 Μελλοντική Έρευνα

Εκτίμησή μας είναι ότι σε μελλοντικά συστήματα όπου οι αναγνωριζόμενες κινήσεις θα είναι πολύ περισσότερες και πιο εφάμιλλες μεταξύ τους, το διακριτό μοντέλο θα παρουσιάσει αδυναμία αναγνώρισης και ανάγκη αύξησης των διακριτών συμβόλων που χρησιμοποιεί, ενώ τελικά θα προτιμηθεί η λύση του συνεχούς.

Παρά τα ενθαρρυντικά αποτελέσματα που παρήγαγε το σύστημα αυτό, εύκολα συνειδητοποιεί κανείς, ότι αυτά προήλθαν από εφαρμογή του συστήματος σε αυστηρά προκαθορισμένες συνθήκες και ελεγχόμενες μεταβολές στις παραμέτρους του. Για να χρησιμοποιηθεί το σύστημα σε πρακτικές εφαρμογές απαιτείται ανασχεδιασμός του ώστε να γίνει πιο ευέλικτο και ανθεκτικό σε μεταβολές του περιβάλλοντος και να μπορεί να εφαρμοστεί σε πραγματικό χρόνο.

Θα πρέπει να υπογραμμιστεί ότι η εκτίμηση κίνησης είναι ένα πολυσύνθετο πρόβλημα. Οι αλληλεπιδράσεις κίνησης είναι τοπικά γεγονότα και ποικίλουν από frame σε frame. Η χρήση δισδιάστατης απεικόνισης ενός τρισδιάστατου κόσμου, για τον υπολογισμό της κίνησης, είναι ημιτελής και μη ιδανική. Υπάρχουν προβλήματα που είναι δύσκολο να αποφευχθούν. Ειδικότερα, προσωρινές μεταβολές στο περιεχόμενο των frames μπορεί να προκύπτουν εξαιτίας υπερφωτισμού ή σκιών και όχι λόγω κάποιας κίνησης αντικειμένου. Κανένας αλγόριθμος από τους ήδη υπάρχοντες δεν αντιμετωπίζει καταλυτικά τα παραπάνω προβλήματα. Η χρήση της οπτικής ροής εξάγει μία πυκνή κατανομή της κίνησης, ωστόσο αυτή η τεχνική σχετίζεται με ένα σημαντικό πρόβλημα που ονομάζεται πρόβλημα διαφράγματος(aperture problem). Εξαιτίας του ότι ο αλγόριθμος βασίζεται σε αποκλίσεις τοπικής φωτεινότητας, προκαλεί τοπική «σύγχυση» στα ανακτώμενα διαγράμματα κίνησης(motion fields).

Ταυτόχρονα αξίζει να σημειωθεί Από τη φύση του το ανθρώπινο σώμα είναι ευκίνητο με δυνατότητα πραγματοποίησης μεγάλου εύρους κινήσεων, οι οποίες μάλιστα διαφέρουν σημαντικά από άτομα σε άτομο. Το γεγονός αυτό καθιστά τη μοντελοποίηση και την αναγνώριση των κινήσεων του ανθρώπου ένα πολύ δύσκολο εγχείρημα. Το συγκεκριμένο σύστημα προϋποθέτει ότι οι κινήσεις που εμφανίζονται στην ακολουθία του βίντεο γίνονται σύμφωνα με τα αρκετά αυστηρά όρια, που αυτό ορίζει. Κατά τη γνώμη μας θα πρέπει, αρχικά, να αναπτυχθεί πιο εξειδικευμένος

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

Η αναγνώριση της ανθρώπινης δραστηριότητας σε ένα βίντεο παρέχει δυνατότητα για ανάπτυξη πολλών εφαρμογών όπως η αυτόματη παρακολούθηση, ιατρικές διαγνώσεις, ανάλυση βίντεο από διάφορα σπορ, και επικοινωνία ανθρώπου-υπολογιστή. Ταυτόχρονα, λόγω της προόδου της τεχνολογίας υπάρχει πλέον η δυνατότητα για επεξεργασία εικόνων και βίντεο σε πραγματικό χρόνο. Με βάση τα παραπάνω, γίνεται κατανοητό, πως η αναγνώριση των ανθρώπινων δραστηριοτήτων σε ακολουθίες βίντεο θα είναι μία από τις βασικότερες εφαρμογές του μέλλοντος.

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

[1] "Markov-Based Failure Prediction for Human Motion Analysis",Shiloh L. Dockstader, Nikita S. Imennov, and A. Murat Tekalp,Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627,College of Engineering, Koη University, Istanbul, Turkey,Dept. of Biomedical Engineering, University of Rochester, Rochester, NY 14627

[2] CRANFIELD UNIVERSITY School of Engineering MSc THESIS ACADEMIC YEAR 2006-2007 Li Xiang "Visual Surveillance – Scene Inventories" Supervisor: Toby Breckon September 2007

[3] "Recognizing Action at a Distance" Alexei A. Efros, Alexander C. Berg, Greg Mori, Jitendra Malik, Computer Science Division, UC Berkeley Berkeley, CA 94720, USA http://www.cs.berkeley.edu/~efros/research/action/

[4] "Learning Dynamic Bayesian Networks",Ashley Mills,ashley@igi.tugraz.at

[5] ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ, Νικόλας Α. Τσαπατσούλης,Διπλ. Ηλεκτρολόγος Μηχανικός ΕΜΠ, "ΠΡΟΗΓΜΕΝΕΣ ΤΕΧΝΙΚΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ ΚΑΙ ΑΝΑΛΥΣΗΣ ΕΚΦΡΑΣΕΩΝ"

[6] "Hidden Markov Models",Dr. Zhang Hongx in State key lab of CAD&CG, 2005-06-30

[7] "Introduction into Hidden Markov Models(HMMs)",• X. Wang, A. Acero, H-W. Hon: "Spoken Language Processing", Chapter 8, pp 374-409, Tapas Kanungo, Uni Maryland, "HMM Tutorial Slides"

[8] "Motion Picture Processing" Mulitmedia- Department of Computing, Imperial College, Professor GZ Yang, http://www.doc.ic.ac.uk/~gzy

[9] L.R. Rabiner: "A Tutorial on HMM and Selected,Applications in Speech Recognition", In: [WL], pp 267-296

[10] "An Ontology for Video Event Representation" Ram Nevatia, Jerry Hobbs and Bob Bolles,Institute for Robotics and Intelligent Systems, University of Southern California,Los Angeles, CA 90089-0273,nevatia@iris.usc.edu,USC Information Sciences Institute,4676 Admiralty Wa,Marina del Rey, CA 90292,hobbs@isi.edu,SRI International,333 Ravenswood Avenue,Menlo Park, CA 94040,bolles@ai.sri.com

[11] Feng Wang, Chong-Wah Ngo, Ting-Chuen Pong, "Gesture Tracking and Recognition for Lecture Video Editing"

[12] Dorthe Meyer, "Human Gait Classification Based on Hidden Markov Models"

[13] J.K. Aggarwal, Sangho Park, "Human Motion: Modeling and Recognition of Actions and Interactions"

[14] Tian-Shu Wang, Heung-Yeung Shum, Ying-Quing Xu, Nan-Ning Zheng, "Unsupervised Analysis of Human Gestures"

[15] Christopher Lee, Yangsheng Xu, "Online, Interactive Learning of Gestures for Human/Robot Interfaces"

[16] Cïön W. Shaffrey, Nick G. Kingsbury, Ian H. Jermyn, "Unsupervised Image Segmentation via Markov Trees and Complex Wavelets"

[17] Jie Yang, Yangsheng Xu, "Hidden Markov Model for Gesture Recognition"

[18] Arnab Ghoshal, Pavel Ircing, Sanjeev Khudanpur, "Hidden Markov Models for Image and Video Retrieval Using Textual Queries"

[19] Steve Young, Gunnar Evermann, Thomas Hain, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, Phil Woodland,"The HTK Book"

[20] Nianjun Liu, Brian C. Lovell, Peter J. Kootsookos, Richard I.A. Davis, "Understanding HMM Training for Video Gesture Recognition"

[22] Jeff A. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Applications to Parameter Estimation for Gaussian Mixture and Hidden Markov Models"

[23] Olivier Cappe "Ten years of HMM's"

*[24] R. Doraiswami, C. P. Diduch, and J. Kuehner, "Failure detection and isolation: A new paradigm," Proc. of the American Control Conf., Arlington, VA, 25-27 June 2001,*vol. 1, pp. 470-475.