



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Παρακολούθηση και Αναγνώριση
Ανθρώπινων Χειρονομιών με Τεχνικές
Υπολογιστικής Όρασης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ευαγγελία Α. Κατσαούνου

Επιβλέπων: Κωνσταντίνος Τζαφέστας
Λέκτορας Ε.Μ.Π.

Αθήνα, Οκτώβριος 2006



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Παρακολούθηση και Αναγνώριση Ανθρώπινων Χειρονομιών με Τεχνικές Υπολογιστικής Όρασης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ευαγγελία Α. Κατσαούνου

Επιβλέπων: Κωνσταντίνος Τζαφέστας
Λέκτορας Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26^η Οκτωβρίου 2006.

.....
Κων/νος Τζαφέστας
Λέκτορας Ε.Μ.Π.

.....
Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2006

.....

Ευαγγελία Α. Κατσαούνου

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

Copyright © Ευαγγελία Κατσαούνου, 2006

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Η ραγδαία εξέλιξη της επιστήμης των υπολογιστών τις τελευταίες δεκαετίες είχε σαν φυσικό επακόλουθο τη διάδοση της χρήσης των υπολογιστών σε όλο και περισσότερους τομείς της καθημερινής μας ζωής, από όλο και περισσότερους ανθρώπους. Υπάρχουν ωστόσο ακόμα σημαντικές δυσκολίες στη χρήση διάφορων υπολογιστικών συστημάτων, με αποτέλεσμα η έρευνα να προσανατολίζεται στην ανάπτυξη συστημάτων η χρήση των οποίων βασίζεται στα "φυσικά" μέσα που χρησιμοποιούν οι άνθρωποι για να επικοινωνήσουν μεταξύ τους, όπως για παράδειγμα στις χειρονομίες. Αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η μελέτη και υλοποίηση ενός συστήματος αυτόματης ανίχνευσης, παρακολούθησης και αναγνώρισης ανθρώπινων χειρονομιών (συγκεκριμένα, στατικών χειρομορφών) σε ακολουθίες εικόνων (video), με χρήση τεχνικών υπολογιστικής όρασης. Μετά από μια σύντομη ανασκόπηση της γενικής θεωρίας γύρω από την παρακολούθηση κινούμενων αντικειμένων (tracking of moving objects) σε ακολουθία εικόνων (video), μελετώνται διεξοδικά τρία πεδία της όρασης υπολογιστών: το χρώμα, και ειδικότερα η χρήση του χρώματος στην ταξινόμηση των εικονοστοιχείων (pixels) μιας εικόνας, η μέθοδος particle filtering για την παρακολούθηση κινούμενων αντικειμένων και η τεχνική της σύμπτωσης προτύπων (template matching) για αναζήτηση πρωτοτύπων εικόνων σε άλλες εικόνες. Οι μεθοδολογίες αυτές συνδυάζονται για την ανάπτυξη του συστήματος παρακολούθησης και αναγνώρισης χειρονομιών, το οποίο μπορεί να χρησιμοποιηθεί σε εφαρμογές τηλεχειρισμού ηλεκτρονικών υπολογιστών, ρομποτικών οχημάτων και άλλων συσκευών.

Λέξεις - Κλειδιά

Αναγνώριση χειρομορφών, παρακολούθηση χειρονομιών, particle filtering, σύμπτωση προτύπων, ανίχνευση χρώματος δέρματος, παρακολούθηση κίνησης, μοντέλο χρώματος, επικοινωνία ανθρώπου-υπολογιστή, όραση υπολογιστών

Abstract

The rapid development of computer science during the last decades had as a natural consequence the wide use of computers in a variety of sectors of our daily life, by a constantly increasing number of people. However there are still several important difficulties in the use of various computer systems; therefore the research is directed towards the development of systems the use of which is based on more "natural" means of communication used by humans, like hand gestures for example. The objective of the present diploma thesis is the study and development of a system for automatic detection, tracking and recognition of human gestures (more specifically, static hand postures) in video sequences, using computer vision techniques. After a short review of the general theory of tracking of moving objects in video sequences, three fields of computer vision are studied: colour, and specifically the use of colour for the classification of pixels in a picture, the particle filtering method for tracking moving objects and the template matching technique used to detect and recognize templates in pictures. These methodologies are combined to develop a hand gesture tracking and recognition system, which can be used in several applications, including teleoperation of computers, robotic vehicles and other devices.

Keywords

Hand posture recognition, gesture tracking, particle filtering, template matching, skin colour detection, motion tracking, colour model, human-computer interaction, computer vision

Ευχαριστίες

Σε αυτές τις λίγες γραμμές θα ήθελα να εκφράσω τις θερμότερες ευχαριστίες μου στον επιβλέποντα καθηγητή μου, κ. Κωνσταντίνο Τζαφέστα, για τη δυνατότητα που μου έδωσε να ασχοληθώ με ένα τόσο ενδιαφέρον θέμα, αλλά και για την αμέριστη βοήθεια και συμπαράστασή του σε όλες τις φάσεις εκπόνησης της παρούσας διπλωματικής εργασίας. Θα ήθελα επίσης να ευχαριστήσω όλους τους καθηγητές μου για τις πολύτιμες γνώσεις που απέκτησα καθ' όλη τη διάρκεια των σπουδών μου στο Εθνικό Μετσόβιο Πολυτεχνείο.

Περιεχόμενα

1	Εισαγωγή.....	19
1.1.	Αντικείμενο της εργασίας	20
1.2.	Δομή της εργασίας.....	20
2	Όραση Υπολογιστών και Επικοινωνία Ανθρώπου - Ρομπότ.....	23
2.1.	Εισαγωγή.....	23
2.2.	Όραση Υπολογιστών.....	23
2.2.1.	Εισαγωγή.....	23
2.2.2.	Κύρια προβλήματα και περιοχές εφαρμογών	24
2.3.	Επικοινωνία Ανθρώπου - Μηχανής.....	25
2.3.1.	Εισαγωγή.....	25
2.3.2.	Στόχοι	26
2.3.3.	Χαρακτηριστικά και δομή συστήματος ΕΑΥ	26
2.3.4.	Πραγματοποίηση επικοινωνίας	27
2.3.5.	Επικοινωνία Ανθρώπου - Μηχανής με χρήση χειρονομιών	28
2.3.6.	Επικοινωνία ανθρώπου - ρομπότ.....	31
3	Ανίχνευση και Παρακολούθηση Ανθρώπινων Χειρονομιών.....	33
3.1.	Εισαγωγή.....	33
3.2.	Παρακολούθηση κινούμενων αντικειμένων	33
3.2.1.	Τοποθέτηση προβλήματος.....	33
3.2.2.	Ανίχνευση κινούμενου αντικειμένου.....	34
3.2.3.	Παρακολούθηση του αντικειμένου.....	34
3.2.4.	Σύνοψη.....	36
3.3.	Ανίχνευση και παρακολούθηση χειρονομιών.....	36
3.3.1.	Εισαγωγή.....	36
3.3.2.	Ταξινόμηση εικονοστοιχείων (pixels) με βάση το χρώμα	37
3.3.3.	Particle filtering	37
3.3.4.	Σύμπτωση Προτύπων (Template Matching)	38
4	Χρώμα.....	39
4.1.	Εισαγωγή.....	39
4.2.	Αναπαράσταση χρώματος	39
4.2.1.	Πρωταρχικά χρώματα	39

4.2.2. Γραμμικοί χώροι χρώματος.....	41
4.2.3. Μη γραμμικοί χώροι χρώματος.....	43
4.3. Χρώμα και Όραση Υπολογιστών.....	49
4.3.1. Κατάτμηση εικόνων με βάση το χρώμα.....	49
4.3.2. Ανίχνευση περιοχών ανθρώπινου δέρματος.....	50
4.3.3. Παρακολούθηση αντικειμένων με βάση το χρώμα τους.....	51
4.3.4. Κατασκευή μοντέλου χρώματος.....	51
5 Παρακολούθηση Κίνησης με Particle Filtering.....	57
5.1. Εισαγωγή.....	57
5.2. Χρήσιμες έννοιες.....	58
5.2.1. Μοντέλο Πιθανότητας.....	58
5.2.2. Κανόνας του Bayes, εκ των προτέρων και εκ των υστέρων πιθανότητες.....	59
5.3. Μη γραμμική Bayesian παρακολούθηση στόχου.....	60
5.3.1. Εισαγωγή.....	60
5.3.2. Καθορισμός προβλήματος και εύρεση λύσης.....	61
5.3.3. Βέλτιστοι αλγόριθμοι.....	63
5.3.4. Σχεδόν βέλτιστοι (suboptimal) αλγόριθμοι.....	65
5.4. Μέθοδοι Particle Filtering.....	67
5.4.1. Sequential Importance Sampling (SIS) Algorithm.....	67
5.4.2. Πρόβλημα εκφυλισμού και τεχνικές αντιμετώπισής του.....	70
5.4.3. Άλλες μέθοδοι particle filtering.....	76
5.4.4. Επίδραση του αριθμού των particles.....	78
5.5. Εφαρμογή particle filtering στην παρακολούθηση χειρονομιών.....	79
5.5.1. Εισαγωγή.....	79
5.5.2. Αλγόριθμος particle filtering για παρακολούθηση κίνησης.....	79
5.5.3. Παρακολούθηση κίνησης χεριού με particle filtering.....	81
6 Σύμπτωση Προτύπων (Template Matching).....	83
6.1. Εισαγωγή.....	83
6.2. Σύμπτωση προτύπων και επεξεργασία εικόνων.....	83
6.2.1. Το πρόβλημα της σύμπτωσης προτύπων.....	83
6.2.2. Τεχνικές σύμπτωσης προτύπων.....	84
6.2.3. Σύμπτωση προτύπων και αναγνώριση χειρονομιών.....	89
7 Υλοποίηση Συστήματος Ανίχνευσης, Παρακολούθησης και Αναγνώρισης Ανθρώπινων Χειρονομιών.....	91
7.1. Εισαγωγή.....	91
7.2. Περιγραφή συστήματος.....	91
7.2.1. Γενική περιγραφή του συστήματος.....	91

7.2.2. Δημιουργία μοντέλου χρώματος	93
7.2.3. Χρήση του particle filtering στην παρακολούθηση του χεριού	96
7.2.4. Αναγνώριση χειρομορφών με σύμπτωση προτύπων	98
7.3. Πλεονεκτήματα και περιορισμοί συστήματος.....	100
7.3.1. Πλεονεκτήματα συστήματος	100
7.3.2. Περιορισμοί συστήματος	101
7.4. Αποτελέσματα και παρατηρήσεις.....	101
7.4.1. Αποτελέσματα	101
7.4.2. Παρατηρήσεις.....	117
8 Σύνοψη και Μελλοντικές Επεκτάσεις.....	119
8.1. Σύνοψη της εργασίας και συμπεράσματα	119
8.2. Μελλοντικές επεκτάσεις	121
9 Βιβλιογραφία.....	123

Κατάλογος Εικόνων

Εικόνα 3.1: Βασικά βήματα της μεθόδου particle filtering.....	38
Εικόνα 4.1: RGB κύβος χρώματος	40
Εικόνα 4.2: Προσθετικά χρώματα (τηλεόραση, οθόνες H/Y)	40
Εικόνα 4.3: Αφαιρετικά χρώματα (ζωγραφική, εκτύπωση)	40
Εικόνα 4.4: Χρωματικές συνιστώσες RGB εικόνας.....	41
Εικόνα 4.5: RGB κύβος χρώματος: Στην εικόνα αριστερά φαίνονται τα βασικά και δευτερεύοντα χρώματα. Τα σημεία πάνω στην κύρια διαγώνιο ανήκουν στην κλίμακα του γκρι. Στην εικόνα δεξιά φαίνεται ο RGB κύβος χρώματος με όλα τα χρώματα	41
Εικόνα 4.6: CIE-xy χρωματικό διάγραμμα: Παρουσιάζεται ολόκληρη η κλίμακα των ορατών χρωμάτων. Ο RGB χώρος χρώματος περιορίζεται μέσα στο κόκκινο τρίγωνο	43
Εικόνα 4.7: Απόχρωση και κορεσμός στον HSI χώρο χρώματος. Η κουκκίδα είναι ένα τυχαίο χρωματικό σημείο. Η γωνία από τον κόκκινο άξονα δίνει την απόχρωση και το μήκος του διανύσματος τον κορεσμό. Η φωτεινότητα των χρωμάτων σε κάθε σχήμα δίνεται από τη θέση του σχήματος στον κάθετο άξονα φωτεινότητας.....	44
Εικόνα 4.8: Ο HSI χώρος χρώματος βασισμένος σε χρωματικά τρίγωνα και κύκλους. Τα τρίγωνα και οι κύκλοι είναι κάθετα στον κάθετο άξονα φωτεινότητας	45
Εικόνα 4.9: Εικόνα αριστερά: Σχέση μεταξύ RGB και HSI χώρων χρώματος. Εικόνα δεξιά: HSI μοντέλο χρώματος.....	46
Εικόνα 4.10: Συνιστώσες HSI στον RGB κύβο χρώματος: απόχρωση, κορεσμός, φωτεινότητα.....	46
Εικόνα 4.11: Εικόνα αριστερά: HSV εξαγώνιο χρωμάτων. Εικόνα δεξιά: HSV εξαγωνικός κώνος βασικών και δευτερευόντων χρωμάτων	47
Εικόνα 4.12: Αντιλήψεις χρώματος για το χρωματικό χώρο CIE-Lab	48
Εικόνα 4.13: Σχέση μεταξύ των ζευγών χρώματος	48
Εικόνα 4.14: Διαφορά φωτεινότητας - φωτοβολίας	49
Εικόνα 4.15: Εικόνα εισόδου, που περιέχει την περιοχή ενδιαφέροντος.....	52
Εικόνα 4.16: Απομόνωση περιοχής ενδιαφέροντος I.....	53
Εικόνα 4.17: Μετατροπή από RGB σε Lab.....	53
Εικόνα 4.18: Γραφική απεικόνιση των ζευγών (a_i, b_i) για το pixel p_i και για όλα τα pixels	53
Εικόνα 4.19: Γραφική απεικόνιση για $n = 240$	54

Εικόνα 4.20: Τοποθέτηση συσσωρευτή στις περιοχές των a, b στο χώρο ab	54
Εικόνα 4.21: Το προκύπτον μοντέλο χρώματος πριν και μετά τη συνέλιξη με τη Gaussian. Κάθε σημείο της δεύτερης γραφικής παράστασης δίνει την πιθανότητα το αντίστοιχο ζεύγος τιμών (a, b) να είναι δέρμα.....	54
Εικόνα 4.22: Εικόνα εισόδου	55
Εικόνα 4.23: Μετατροπή από RGB σε Lab.....	55
Εικόνα 4.24: Αντιστοίχιση του κάθε ζεύγους (a_i, b_i) σε πιθανότητα μέσω του μοντέλου.....	55
Εικόνα 4.25: Εικόνα εισόδου και εικόνα πιθανοτήτων.....	56
Εικόνα 5.1: Βασική ιδέα του particle filtering.....	57
Εικόνα 5.2: Αλγόριθμος 1	70
Εικόνα 5.3: Αλγόριθμος 2	74
Εικόνα 5.4: Αλγόριθμος 3	74
Εικόνα 5.5: Κλασσικό particle filtering και Augmented particle filtering.....	77
Εικόνα 5.6: Αλγόριθμος particle filtering.....	80
Εικόνα 6.1: Παράδειγμα εφαρμογής της σύμπτωσης προτύπων.....	84
Εικόνα 7.1: Χειρομορφές που αναγνωρίζονται από το σύστημα.....	92
Εικόνα 7.2: Απεικόνιση των ζευγών (a_{ind}, b_{ind}) στο επίπεδο ab	94
Εικόνα 7.3: Απεικόνιση των ζευγών (a_{ind}, b_{ind}) στο επίπεδο ab (λεπτομέρεια).....	94
Εικόνα 7.4: Το μοντέλο χρώματος μετά τη συνέλιξη με Gaussian. Κάθε σημείο της γραφικής παράστασης δίνει την πιθανότητα το αντίστοιχο ζεύγος τιμών (a, b) να είναι δέρμα.	95
Εικόνα 7.5: Το μοντέλο χρώματος μετά τη συνέλιξη με Gaussian. Κάθε σημείο της γραφικής παράστασης δίνει την πιθανότητα το αντίστοιχο ζεύγος τιμών (a, b) να είναι δέρμα. (λεπτομέρεια).....	95
Εικόνα 7.6: Εικόνα εισόδου	96
Εικόνα 7.7: Εικόνα πιθανοτήτων.....	96
Εικόνα 7.8: Χειρομορφές που διαφέρουν κατά 180°	101
Εικόνα 7.9: Στιγμιότυπα από τα αποτελέσματα λειτουργίας του συστήματος. Στην εικόνα 7.9.1 παρουσιάζεται η αρχικοποίηση του συστήματος, όπου τοποθετούνται τυχαία τα particles στο επίπεδο της εικόνας. Στις υπόλοιπες εικόνες φαίνονται τα αποτελέσματα της παρακολούθησης και αναγνώρισης των χειρονομιών.....	104
Εικόνα 7.10: Απαιτούμενος χρόνος επεξεργασίας για κάθε frame	104
Εικόνα 7.11: Στιγμιότυπα από τον έλεγχο λειτουργίας του συστήματος. Στην εικόνα 7.11.1 παρουσιάζεται η αρχικοποίηση του συστήματος, όπου τοποθετούνται τυχαία τα particles στο επίπεδο της εικόνας. Στις υπόλοιπες εικόνες φαίνονται τα αποτελέσματα της παρακολούθησης και αναγνώρισης των χειρονομιών.....	108

Εικόνα 7.12: Απαιτούμενος χρόνος επεξεργασίας για κάθε frame	108
Εικόνα 7.13: Στιγμιότυπα από τον έλεγχο λειτουργίας του συστήματος. Στην εικόνα 7.13.1 παρουσιάζεται η αρχικοποίηση του συστήματος, όπου τοποθετούνται τυχαία τα particles στο επίπεδο της εικόνας. Στις υπόλοιπες εικόνες φαίνονται τα αποτελέσματα της παρακολούθησης και αναγνώρισης των χειρονομιών.....	113
Εικόνα 7.14: Απαιτούμενος χρόνος επεξεργασίας για κάθε frame	113
Εικόνα 7.15: Στιγμιότυπα από τον έλεγχο λειτουργίας του συστήματος. Στην εικόνα 7.15.1 παρουσιάζεται η αρχικοποίηση του συστήματος, όπου τοποθετούνται τυχαία τα particles στο επίπεδο της εικόνας. Στις υπόλοιπες εικόνες φαίνονται τα αποτελέσματα της παρακολούθησης του ενός και των δύο χεριών.....	117
Εικόνα 7.16: Απαιτούμενος χρόνος επεξεργασίας για κάθε frame	117

1

Εισαγωγή

Η τεχνολογία, και ιδιαίτερα η τεχνολογία των υπολογιστών, αποτελεί πλέον κομμάτι της καθημερινής ζωής όλων μας. Ωστόσο, παρά τη ραγδαία ανάπτυξη και διάδοσή της, υπάρχουν ακόμα σημαντικές δυσκολίες στη χρήση πολλών τεχνολογικών εφαρμογών, με αποτέλεσμα να καταβάλλεται προσπάθεια ώστε ο χειρισμός των νέων συστημάτων να γίνεται με όσο το δυνατόν πιο φυσικό και διαισθητικά απλό για τον άνθρωπο τρόπο. Είναι λοιπόν αναμενόμενο η έρευνα να στρέφεται στον πλέον απλό και φυσικό τρόπο επικοινωνίας μεταξύ των ανθρώπων, τις χειρονομίες. Οι χειρονομίες κατέχουν σημαντική θέση στην επικοινωνία μεταξύ των ανθρώπων, αφού αποτελούν φυσικό μέσο έκφρασης της διάθεσης και των συναισθημάτων των συνομιλητών, αλλά και βασικό εργαλείο της νοηματικής γλώσσας. Έτσι, είναι απαραίτητη η ανάπτυξη συστημάτων παρακολούθησης και αναγνώρισης χειρονομιών. Τα συστήματα αυτά προσδιορίζουν τις ανθρώπινες χειρονομίες και τις χρησιμοποιούν για να μεταβιβάσουν πληροφορίες όπως τα δεδομένα εισόδου σε ένα υπολογιστή ή για να ελέγξουν διάφορες συσκευές, όπως τηλεόραση, κινητό τηλέφωνο, οικιακά ρομπότ, αυτοκινούμενα οχήματα ακόμα και ιατρικά εργαλεία. Είναι βεβαίως αναγκαία η ανάπτυξη γρήγορων και αξιόπιστων αλγορίθμων παρακολούθησης και αναγνώρισης χειρονομιών για την αποφυγή ανεπιθύμητων σφαλμάτων.

Εντούτοις, η τεχνολογία αντιμετωπίζει ακόμα σημαντικές προκλήσεις. Για παράδειγμα, οι πρώτες συσκευές αναγνώρισης χειρονομιών που έχουν αναπτυχθεί δεν είναι κατάλληλες για χρήση σε ευρύ πεδίο εφαρμογών, αφού περιλαμβάνουν ειδικό εξοπλισμό (όπως ειδικά γάντια), ο οποίος δυσκολεύει αντί να διευκολύνει το φυσικό τρόπο "επικοινωνίας" μεταξύ ανθρώπου και μηχανής. Στις αρχές της δεκαετίας του 1990 η έρευνα στράφηκε στη χρήση τεχνικών της Όρασης Υπολογιστών για την παρακολούθηση και αναγνώριση των χειρονομιών. Στα συστήματα αυτού του είδους, οι χρήστες εκτελούν τις χειρονομίες, οι οποίες καταγράφονται από μία ή περισσότερες κάμερες. Μέσω κατάλληλου λογισμικού (software) παρακολουθείται η κίνηση του χεριού, οι χειρονομίες αναγνωρίζονται και μετατρέπονται σε γράμματα και λέξεις ή σε απλές ή πιο σύνθετες εντολές. Στη συνέχεια η μηχανή δρα σύμφωνα με αυτά τα δεδομένα εισόδου. Τα συστήματα αυτά έχουν επωφεληθεί ιδιαίτερα από τη μείωση της ταχύτητας επεξεργασίας σε συνδυασμό με τη μείωση του κόστους του υλικού (hardware).

1.1. Αντικείμενο της εργασίας

Έχοντας ως εφιαλτήριο τα όσα ήδη αναφέρθηκαν, κρίθηκε σκόπιμη η μελέτη στην παρούσα εργασία ορισμένων τεχνικών υπολογιστικής όρασης με στόχο την υλοποίηση ενός συστήματος ανίχνευσης, παρακολούθησης και αναγνώρισης ανθρώπινων χειρονομιών. Αναλυτικότερα, αντικείμενο της εργασίας είναι η αναγνώριση έξι προκαθορισμένων στατικών χειρονομιών που εκτελούνται από ένα κινούμενο χέρι, η κίνηση του οποίου καταγράφεται σε ακολουθίες εικόνων (video). Το αποτέλεσμα της αναγνώρισης προκύπτει μετά από επεξεργασία κάθε εικόνας (frame). Με χρήση της στατιστικής μεθόδου particle filtering εντοπίζεται και παρακολουθείται η κατάσταση του χεριού και προκύπτει μια προσεγγιστική εικόνα του. Με βάση την εικόνα αυτή και χρησιμοποιώντας την τεχνική της σύμπτωσης προτύπων (template matching) αναγνωρίζεται η εκάστοτε διάταξη του χεριού (χειρομορφή).

Το σύστημα αυτό, που υλοποιήθηκε στο Matlab, δε λειτουργεί σε πραγματικό χρόνο, αλλά το video εισόδου θα πρέπει να έχει ήδη καταγραφεί πριν την έναρξη λειτουργίας του συστήματος. Δεν είναι δυνατή η άμεση χρήση του στην πραγματικού χρόνου αναγνώριση των χειρονομιών, αποτελεί όμως το ενδιάμεσο στάδιο της μετατροπής της κίνησης του χεριού σε μορφή κατάλληλη για μορφοποίηση εντολών από κάποιο υπολογιστικό σύστημα που μπορούν να χρησιμοποιηθούν σε ποικίλες εφαρμογές. Οι εφαρμογές αυτές περιλαμβάνουν τον τηλεχειρισμό ρομποτικών οχημάτων, ηλεκτρονικών υπολογιστών, ακόμα και οικιακών συσκευών. Μια ακόμα πολύ σημαντική εφαρμογή είναι η αναγνώριση της νοηματικής γλώσσας και η μετατροπή της σε κείμενο μέσω ηλεκτρονικού υπολογιστή.

1.2. Δομή της εργασίας

Στην ενότητα αυτή παρουσιάζονται συνοπτικά τα περιεχόμενα των κεφαλαίων που ακολουθούν.

Στο 2ο κεφάλαιο γίνεται μια σύντομη εισαγωγή στο πεδίο της Όρασης Υπολογιστών και στην επικοινωνία ανθρώπου-υπολογιστή και ανθρώπου-ρομπότ. Περιγράφονται τα βασικότερα προβλήματα της Όρασης Υπολογιστών καθώς και οι κυριότερες περιοχές εφαρμογών της. Επίσης περιγράφονται τα βασικά στοιχεία της επικοινωνίας ανθρώπου-μηχανής και δίνεται ιδιαίτερη έμφαση στην πραγματοποίηση της επικοινωνίας αυτής με χρήση χειρονομιών.

Στο 3ο κεφάλαιο παρουσιάζεται η γενική θεωρία γύρω από την παρακολούθηση κινούμενων αντικειμένων (tracking of moving objects) σε ακολουθία εικόνων (video). Ιδιαίτερη βαρύτητα δίνεται στην ανίχνευση και παρακολούθηση ανθρώπινων χειρονομιών και περιγράφονται συνοπτικά τα βασικά βήματα που ακολουθήθηκαν για το σχεδιασμό του συστήματος ανίχνευσης και παρακολούθησης χειρονομιών της παρούσας εργασίας.

Στο 4ο κεφάλαιο παρουσιάζονται τα βασικά στοιχεία των χώρων χρώματος και αναλύεται η χρήση του χρώματος στην Όραση Υπολογιστών και συγκεκριμένα στην αναγνώριση αντικειμένων σε εικόνες.

Στο 5ο κεφάλαιο περιγράφεται η μέθοδος του particle filtering, που χρησιμοποιήθηκε στο σύστημα που υλοποιήθηκε για την παρακολούθηση των χειρονομιών. Δίνονται γενικά θεωρητικά στοιχεία, αλγόριθμοι καθώς και εφαρμογές.

Στο 6ο κεφάλαιο εξετάζεται η τεχνική της σύμπτωσης προτύπων (template matching). Παρουσιάζονται οι βασικοί αλγόριθμοι σύμπτωσης προτύπων και δίνεται έμφαση στη χρήση της τεχνικής αυτής στην αναγνώριση χειρονομιών.

Στο 7ο κεφάλαιο δίνεται η περιγραφή του συστήματος που υλοποιήθηκε και αναλύονται διεξοδικά τα επιμέρους στάδια. Επιπλέον, παρατίθενται και σχολιάζονται τα πειραματικά αποτελέσματα.

Στο τελευταίο κεφάλαιο περιλαμβάνεται η ανακεφαλαίωση της εργασίας και προτάσεις για μελλοντικές επεκτάσεις.

2

Όραση Υπολογιστών και Επικοινωνία Ανθρώπου - Ρομπότ

2.1. Εισαγωγή

Στο εισαγωγικό αυτό κεφάλαιο παρουσιάζεται σύντομα το επιστημονικό πεδίο της Όρασης Υπολογιστών και μελετάται η επικοινωνία ανθρώπου-μηχανής, με έμφαση στην επικοινωνία ανθρώπου-ρομπότ μέσω χειρονομιών. Αφού δοθεί ένας σύντομος ορισμός του πεδίου της Όρασης Υπολογιστών, παρουσιάζονται οι κύριες κατευθύνσεις αλλά και τα κυριότερα προβλήματα και εφαρμογές του. Στη συνέχεια του κεφαλαίου δίνεται μια σύντομη περιγραφή της έννοιας της επικοινωνίας ανθρώπου-μηχανής και αναλύονται οι στόχοι και οι τρόποι πραγματοποίησης της επικοινωνίας αυτής. Τέλος, εξετάζεται η χρήση χειρονομιών ως μέσο για την πραγματοποίηση της επικοινωνίας ανθρώπου-μηχανής γενικότερα και ανθρώπου-ρομπότ ειδικότερα.

2.2. Όραση Υπολογιστών

2.2.1. Εισαγωγή

Όραση είναι η αισθητηριακή αντίληψη των αντικειμένων του κόσμου και των ιδιοτήτων τους με βάση τη φωτεινή ενέργεια που εκπέμπεται από αυτά και συλλέγεται από οπτικούς αισθητήρες. Η σημασία της όρασης για τον άνθρωπο και για όσα έμβια όντα διαθέτουν αυτή την αίσθηση επαυξάνεται από το γεγονός ότι το μεγαλύτερο ποσό της πληροφορίας που εισέρχεται στο ανθρώπινο σύστημα αντίληψης μέσω των αισθήσεων είναι οπτικό. Η είσοδος στο ανθρώπινο (ή κάθε άλλο βιολογικό) σύστημα όρασης είναι οι εικόνες. Για να δούμε όμως πρέπει να συνδέσουμε τα χωριστά συμπεράσματα για την κίνηση, το χρώμα, το σχήμα, την υφή και το βάθος σε μια ενοποιημένη εξήγηση. Αυτή η ενοποιημένη πληροφορία χρησιμοποιείται για την εκτέλεση διάφορων σύνθετων εργασιών, όπως για παράδειγμα η αναγνώριση και παρακολούθηση ενός κινούμενου αντικειμένου σε μια ακολουθία εικόνων.

Σε αντίθεση όμως με τα βιολογικά συστήματα, όπου αυτή η δυνατότητα της ενοποίησης και επεξεργασίας των οπτικών πληροφοριών είναι αυτονόητη, αυτόματη και έμφυτη, στα μηχανικά συστήματα είναι απαραίτητο να προηγηθεί επεξεργασία μιας πλειάδας χαρακτηριστικών και πληροφοριών, που προέρχονται από την απεικόνιση του πραγματικού κόσμου σε μορφή ακολουθίας εικόνων, ώστε να καταλήξουμε σε δυνατότητες αντίστοιχες αυτών των βιολογικών συστημάτων. Η επεξεργασία αυτή γίνεται με μεθόδους που υποδεικνύει η Όραση Υπολογιστών, της οποίας βασικός στόχος είναι, δεδομένης μιας ή περισσότερων διαστάσεων (2Δ) εικόνων, η εύρεση μιας όσο το δυνατόν πληρέστερης συμβολικής περιγραφής των αντικειμένων του τρισδιάστατου (3Δ) κόσμου που περιέχονται στη σκηνή αυτή.

Η Όραση Υπολογιστών άρχισε ως επιστημονικό πεδίο τη δεκαετία του 1960 από προσπάθειες ερευνητών της Τεχνητής Νοημοσύνης να δημιουργήσουν υπολογιστές που βλέπουν. Τα επιστημονικά πεδία που συνέβαλαν στη δημιουργία και ανάπτυξη της Όρασης Υπολογιστών είναι η Τεχνητή Νοημοσύνη, η Επεξεργασία Σημάτων και η Αναγνώριση Προτύπων. Οι τρεις κύριες κατευθύνσεις της Όρασης Υπολογιστών είναι:

- ✓ Τεχνητή (Μηχανική) Όραση: εφοδιασμός μηχανών ή ρομπότ ή αισθητήρων με δυνατότητες ή στοιχεία βιολογικής όρασης.
- ✓ Βιολογική Όραση: ανάλυση και κατανόηση βιολογικής όρασης με υπολογιστικά μοντέλα και χρησιμοποίηση των ενδείξεων και παραδειγμάτων για δημιουργία μηχανικής όρασης.
- ✓ Μοντελοποίηση Εγκεφάλου: όραση είναι η σημαντικότερη αίσθηση για αντίληψη του κόσμου. Κατανόηση του πώς σκεπτόμαστε εξαρτάται κυρίως από την κατανόηση του πώς βλέπουμε.

2.2.2. Κύρια προβλήματα και περιοχές εφαρμογών

Τα προβλήματα της όρασης υπολογιστών μπορούν να ταξινομηθούν στις ακόλουθες περιοχές:

- ✓ Σχηματισμός εικόνων: προβολή, φωτισμός, οπτικά συστήματα, αισθητήρες
- ✓ Ανάλυση εικόνων, δυαδικών ή γκριζών, με αλγεβρικές, γεωμετρικές ή στατιστικές μεθόδους
- ✓ Ανάλυση σε πολλαπλές κλίμακες: χώροι κλίμακας
- ✓ Βελτίωση ποιότητας και εξαγωγή χαρακτηριστικών: χαρακτηριστικό παράδειγμα η ανίχνευση ακμών
- ✓ Υφή: μοντελοποίηση και ανάλυση
- ✓ Χρώμα: μοντελοποίηση και ανάλυση
- ✓ Σχήμα: μοντελοποίηση και ανάλυση
- ✓ Αποκατάσταση χαμένης πληροφορίας: παραδείγματα η παρεμβολή και η πρόβλεψη
- ✓ Οπτική κίνηση, 2Δ και 3Δ: ανίχνευση, εκτίμηση
- ✓ Ανακατασκευή 3Δ δομής: στερέωση, γεωμετρία πολλαπλών όψεων, σχήμα από σκίαση ή κίνηση ή υφή
- ✓ Κατάτμηση: διαχωρισμός εικόνων σε ομογενείς περιοχές
- ✓ Αναγνώριση και κατηγοριοποίηση οπτικών αντικειμένων

Στην παρούσα εργασία ασχολούμαστε με πέντε από τα παραπάνω προβλήματα: ανάλυση εικόνων, χρώμα, σχήμα, κατάτμηση και αναγνώριση και κατηγοριοποίηση αντικειμένων.

Οι σημαντικότερες περιοχές εφαρμογών της όρασης υπολογιστών είναι: επεξεργασία πληροφορίας σε εικόνες, μοντελοποίηση εγκεφάλου και τεχνητή νοημοσύνη, αυτοματισμός και ρομποτική, βιοϊατρική έρευνα και τεχνολογία, ιατρική απεικόνιση, περιβάλλον και γεωεπιστήμες, δορυφορική όραση και τηλεπισκόπηση, ευφυή συστήματα ψηφιοποίησης και ανάλυσης βιβλίων, ευφυείς αισθητήρες και ψηφιακές κάμερες, οπτική παρακολούθηση οχημάτων, τεχνολογία video, τεχνολογία διαδικτύου και αυτόματη αναζήτηση, επικοινωνία ανθρώπου και υπολογιστή και ψηφιακές εικαστικές τέχνες.

2.3. Επικοινωνία Ανθρώπου - Μηχανής

2.3.1. Εισαγωγή

Η ραγδαία εξέλιξη της επιστήμης των υπολογιστών τις τελευταίες δεκαετίες είχε σαν φυσικό επακόλουθο τη διάδοση της χρήσης των υπολογιστών από όλο και περισσότερους ανθρώπους, σε όλο και περισσότερους τομείς της καθημερινής μας ζωής. Σε αντίθεση με τα πρώτα συστήματα υπολογιστών, των οποίων η χρήση απαιτούσε ειδικές γνώσεις και ήταν σχεδόν απαγορευτική για το ευρύ κοινό, τα σημερινά συστήματα είναι ολοένα και πιο εύχρηστα και η χρήση τους έχει διαδοθεί σε ολόκληρο τον κόσμο. Σε αυτό το γεγονός συνετέλεσαν πολλοί παράγοντες μεταξύ των οποίων η μείωση του μεγέθους των υπολογιστών, η αύξηση της υπολογιστικής ισχύος των νέων συστημάτων, η ραγδαία μείωση του κόστους τους και η διεύρυνση του πεδίου εφαρμογών τους. Ωστόσο, υπάρχουν ακόμα σημαντικές δυσκολίες στη χρήση διάφορων συστημάτων, με αποτέλεσμα η έρευνα να προσανατολίζεται στην ανάπτυξη συστημάτων που είναι ανθρωποκεντρικά και φιλικά προς το χρήστη.

Το επιστημονικό πεδίο που ασχολείται με αυτόν ακριβώς το σχεδιασμό, την αξιολόγηση και την υλοποίηση των διαλογικών υπολογιστικών συστημάτων καθώς και με τη μελέτη σημαντικών φαινομένων που τα περιβάλλουν είναι το πεδίο της διάδρασης/αλληλεπίδρασης/επικοινωνίας ανθρώπου-υπολογιστή (human-computer interaction).

Όταν μιλάμε για "επικοινωνία ανθρώπου-υπολογιστή" (EAY), δεν αναφερόμαστε απαραίτητα σε ένα μεμονωμένο χρήστη και έναν προσωπικό υπολογιστή. Με τον όρο χρήστη (user) μπορεί να εννοούμε ένα μεμονωμένο χρήστη, μια ομάδα χρηστών που δουλεύουν μαζί ή μια αλληλουχία χρηστών, καθένας εκ των οποίων ασχολείται με κάποιο επιμέρους θέμα μιας εργασίας. Με τον όρο υπολογιστής (computer) αναφερόμαστε σε οποιαδήποτε τεχνολογία υπολογιστών - από τους προσωπικούς υπολογιστές γενικής χρήσης, μέχρι τα μεγάλης κλίμακας υπολογιστικά συστήματα, τα συστήματα ελέγχου διεργασιών (process control systems), ή τα ενσωματωμένα συστήματα (embedded systems). Επίσης το σύστημα μπορεί να περιλαμβάνει και άλλα μέρη εκτός από τους υπολογιστές. Τέλος με τον όρο διάδραση (interaction) εννοούμε οποιαδήποτε επικοινωνία μεταξύ ενός χρήστη και ενός υπολογιστή, άμεση ή έμμεση. Η άμεση επικοινωνία συνεπάγεται ένα διάλογο με ανάδραση και

έλεγχο καθ' όλη τη διάρκεια της εκτέλεσης της εργασίας. Η έμμεση επικοινωνία μπορεί να περιλαμβάνει μαζική επεξεργασία ή επεξεργασία στο παρασκήνιο. Το βασικότερο χαρακτηριστικό όμως είναι ότι ο χρήστης αλληλεπιδρά / επικοινωνεί με τον υπολογιστή προκειμένου να πετύχει κάποιο στόχο.

2.3.2. Στόχοι

Βασικός στόχος της ΕΑΥ είναι η βελτίωση της αλληλεπίδρασης μεταξύ των χρηστών και των υπολογιστών, κατασκευάζοντας υπολογιστές που είναι περισσότερο φιλικό προς το χρήστη και δεκτικοί στις ανάγκες του. Συγκεκριμένα, η ΕΑΥ ασχολείται με:

- ✓ μεθοδολογίες και διαδικασίες για το σχεδιασμό διεπαφών (interfaces): λαμβάνοντας υπόψη έναν στόχο και μια κατηγορία χρηστών, να σχεδιαστεί η καλύτερη δυνατή διεπαφή υπό δεδομένους περιορισμούς, όπου η βελτιστοποίηση γίνεται για μια επιθυμητή ιδιότητα όπως η δυνατότητα εκμάθησης ή η αποδοτικότητα της χρήσης
- ✓ μέθοδοι για την υλοποίηση διεπαφών (όπως για παράδειγμα εργαλεία και βιβλιοθήκες λογισμικού, αποδοτικοί αλγόριθμοι)
- ✓ τεχνικές για αξιολόγηση και σύγκριση διεπαφών
- ✓ ανάπτυξη νέων διεπαφών και τεχνικών διάδρασης
- ✓ ανάπτυξη περιγραφικών και προβλεπτικών μοντέλων και θεωριών της διάδρασης

Ένας μακροπρόθεσμος στόχος της ΕΑΥ είναι ο σχεδιασμός συστημάτων που ελαχιστοποιούν την απόσταση μεταξύ του γνωστικού μοντέλου του ανθρώπου για το σκοπό που θέλει να επιτύχει και της κατανόησης του στόχου του χρήστη από τον υπολογιστή.

Όσοι ασχολούνται επαγγελματικά με την ΕΑΥ ενδιαφέρονται για την πρακτική εφαρμογή των μεθοδολογιών σχεδιασμού στα πραγματικά προβλήματα. Η εργασία τους περιστρέφεται συχνά γύρω από το σχεδιασμό των γραφικών διεπαφών χρήστη (graphical user interfaces) και των διεπαφών ιστού (web interfaces). Οι ερευνητές της ΕΑΥ ενδιαφέρονται για την ανάπτυξη νέων μεθοδολογιών σχεδιασμού, πειραματίζονται με νέες συσκευές υλικού (hardware devices), δημιουργούν πρωτότυπα συστήματα λογισμικού, εξερευνούν τα νέα παραδείγματα αλληλεπίδρασης, και αναπτύσσουν πρότυπα και θεωρίες της αλληλεπίδρασης.

Για την σχεδίαση ενός διαδραστικού συστήματος απαιτούνται γνώσεις από πολλά επιστημονικά πεδία: επιστήμη υπολογιστών, τεχνολογία πληροφοριών, αισθητική, ανθρωπολογία, τεχνητή νοημοσύνη, γνωστική επιστήμη (cognitive science), όραση υπολογιστών, σχέδιο (design), εργονομία, βιβλιοθήκη και επιστήμη των πληροφοριών (library and information science), φιλοσοφία, φαινομενολογία (phenomenology), ψυχολογία, κοινωνική ψυχολογία, κοινωνιολογία, λεκτική-γλωσσική παθολογία (speech-language pathology), ψυχοφυσική (psychophysics).

2.3.3. Χαρακτηριστικά και δομή συστήματος ΕΑΥ

Αν και δεν υπάρχει κάποια ενοποιημένη θεωρία πάνω στην ΕΑΥ, υπάρχει μια θεμελιώδης αρχή πάνω στην οποία πρέπει να βασίζεται η σχεδίαση οποιουδήποτε

διαδραστικού συστήματος και αυτή είναι ότι οι άνθρωποι χρησιμοποιούν τους υπολογιστές για να εκτελέσουν κάποια εργασία. Έτσι, το σύστημα είναι απαραίτητο να χαρακτηρίζεται από λειτουργικότητα, αφού εάν το σύστημα υποχρεώνει τον χρήστη να υιοθετήσει ένα πολύπλοκο και δύσκολο τρόπο εργασίας τότε δεν είναι εύχρηστο. Καταλήγουμε λοιπόν στα εξής χαρακτηριστικά για το σύστημα:

- ✓ εύκολο στην εκμάθηση της λειτουργίας του
- ✓ εύκολο στην απομνημόνευση του τρόπου λειτουργίας του
- ✓ αποδοτικό
- ✓ ασφαλές
- ✓ ενδιαφέρον / διασκεδαστικό κατά την χρήση

Είναι αυτονόητο ότι η επικοινωνία ανθρώπου-υπολογιστή πρέπει να είναι αμφίδρομη. Έτσι, πρέπει και οι δύο πλευρές να είναι σε θέση να επεξεργάζονται και να ανταποκρίνονται στα ερεθίσματα και στις πληροφορίες που δέχονται. Βασικά δομικά στοιχεία ενός συστήματος διάδρασης είναι:

- ✓ το υλικό: μετατρέπει τα σήματα που αποστέλλονται αμφίδρομα ανάμεσα στον άνθρωπο και τον υπολογιστή
- ✓ το λογισμικό: ο υπολογιστής πρέπει να είναι σε θέση να ανταποκρίνεται ακόμα και σε δεδομένα που δεν έχουν κάποιο ιδιαίτερο νόημα χωρίς να διακόπτει την επικοινωνία
- ✓ η μεθοδολογία επικοινωνίας μεταξύ ανθρώπου και μηχανής: υπάρχουν διάφοροι τρόποι με τους οποίους πραγματοποιείται η επικοινωνία, όπως οι παραδοσιακές συσκευές εισόδου (πληκτρολόγιο, ποντίκι), οι λεκτικές εντολές και η χρήση χειρονομιών.

2.3.4. Πραγματοποίηση επικοινωνίας

Με την ογκώδη εισροή των υπολογιστών στην κοινωνία, η ΕΑΥ γίνεται όλο και πιο σημαντικό μέρος της καθημερινής μας ζωής. Ακόμα όμως και μετά από περισσότερο από δύο δεκαετίες ανάπτυξης συσκευών εισόδου, πολλοί άνθρωποι βρίσκουν ακόμα την αλληλεπίδραση με υπολογιστές μια δυσάρεστη εμπειρία. Η επικοινωνία ανθρώπου-μηχανής αποτελεί καθοριστικό παράγοντα στη δημιουργία αξιόπιστων και αποδοτικών συστημάτων ελέγχου που λειτουργούν με ηλεκτρονικούς υπολογιστές και για το λόγο αυτό θα πρέπει να δοθεί ιδιαίτερη προσοχή στον τρόπο πραγματοποίησης της επικοινωνίας αυτής.

Επικρατεί η άποψη ότι όσο οι τεχνολογίες υπολογισμού, επικοινωνιών και απεικόνισης προοδεύουν όλο και περισσότερο, οι υπάρχουσες τεχνικές ΕΑΥ μπορεί να αποδειχτούν εμπόδιο στην αποτελεσματική χρησιμοποίηση της διαθέσιμης ροής πληροφοριών. Ο δημοφιλέστερος τρόπος επικοινωνίας σήμερα είναι βασισμένος στις απλές μηχανικές συσκευές: πληκτρολόγια και ποντίκια. Μπορεί βέβαια με αυτές τις συσκευές να έχουμε εξοικειωθεί, αλλά αναπόφευκτα περιορίζουν την ταχύτητα και τη φυσικότητα της αλληλεπίδρασης με τον υπολογιστή. Αυτός ο περιορισμός έχει γίνει ακόμα φανερότερος με την εμφάνιση νέων τεχνολογιών απεικόνισης όπως η εικονική πραγματικότητα. Κατά συνέπεια τα τελευταία χρόνια έχει υπάρξει μια τεράστια ώθηση στην έρευνα για νέες συσκευές και τεχνικές που θα διευθετήσουν τη δυσχέρεια στην ΕΑΥ.

Συγκεκριμένα, η ΕΑΥ εξελίσσεται προς την κατεύθυνση των συσκευών που δεν απαιτούν επαφή, χρησιμοποιώντας αντιληπτικές και πολύμορφες διεπαφές με το χρήστη που βασίζονται στα "φυσικά" μέσα που χρησιμοποιούν οι άνθρωποι για να επικοινωνήσουν μεταξύ τους. Αυτό σημαίνει ότι το σύστημα επιτρέπει στο χρήστη να αλληλεπιδράσει χωρίς φυσική επαφή, χρησιμοποιώντας φωνή ή/και τη γλώσσα του σώματος (κίνηση μερών του σώματος, χρήση χειρονομιών). Έχοντας ως κίνητρο τα παραπάνω, η αυτόματη αναγνώριση φωνής αποτελεί θέμα έρευνας για δεκαετίες. Τεράστια πρόοδος έχει σημειωθεί στο πεδίο αυτό και έχουν δημιουργηθεί διάφορες εμπορικά επιτυχείς φωνητικές διεπαφές. Εντούτοις, τα τελευταία χρόνια υπάρχει ένα αυξανόμενο ενδιαφέρον για την προσπάθεια εισαγωγής και άλλων ανθρώπινων μορφών επικοινωνίας στην ΕΑΥ. Στην προσπάθεια αυτή περιλαμβάνεται μια κατηγορία τεχνικών που βασίζονται στην κίνηση του ανθρώπινου βραχίονα και της παλάμης, ή σε χειρονομίες, αφού όπως η προφορική επικοινωνία, έτσι και οι χειρονομίες είναι ένας φυσικός τρόπος να πραγματοποιηθεί η επικοινωνία ανθρώπου-μηχανής.

Αυτή η μορφή αλληλεπίδρασης είναι αντικείμενο των ευφυών διεπαφών (Intelligent User Interfaces - IUI), κλάδο της ΕΑΥ. Στόχος είναι η βελτίωση της επικοινωνίας ανθρώπου-υπολογιστή χρησιμοποιώντας νέες τεχνολογίες και τεχνικές της τεχνητής νοημοσύνης. Ειδικότερα, η όραση υπολογιστών παρέχει νέους τρόπους αλληλεπίδρασης IUI χρησιμοποιώντας την κίνηση του ανθρώπινου σώματος. Κατ' αυτό τον τρόπο, παρέχοντας νέες μορφές αλληλεπίδρασης βασισμένες στην επεξεργασία εικόνας ή φωνής αντί των παραδοσιακών συσκευών όπως το πληκτρολόγιο ή το ποντίκι, θα είναι δυνατό να ελέγχονται από απόσταση διάφορες συσκευές. Επιπλέον, οι ανάπηροι άνθρωποι μπορούν να έχουν ευκολότερα πρόσβαση στους υπολογιστές. Είναι φυσικά απαραίτητο οι νέες αυτές μεθοδολογίες ΕΑΥ να είναι σε θέση να λειτουργήσουν αποδοτικά και με αξιοπιστία.

Τα κύρια πλεονεκτήματα των νέων αυτών μεθόδων ΕΑΥ έναντι των παραδοσιακών συσκευών εισόδου είναι τα ακόλουθα:

- ✓ Με συσκευές όπως το ποντίκι ή το joystick που έχουν μόνο δύο βαθμούς ελευθερίας, απαιτούνται διάφορες διαδοχικές ενέργειες για τον έλεγχο περισσότερων από δύο παραμέτρων. Τέτοιος έλεγχος δεν είναι πάντα φυσικός και απαιτεί εκπαίδευση των χρηστών. Στα νέα συστήματα, μια εντολή που έχει μέχρι 10 παραμέτρους μπορεί να αντικατασταθεί με μόλις μια φωνητική εντολή ή μια χειρονομία. Παραδείγματος χάριν, μπορεί κάποιος να χρησιμοποιήσει απολύτως φυσικά το χέρι του για να κινηθεί σε μια εικονική τρισδιάστατη σκηνή, ενώ παράλληλα ελέγχει τη θέση (X, Y, Z) και τον προσανατολισμό (γωνίες στροφής roll, pitch, yaw).
- ✓ Δεν υπάρχει ανάγκη για οποιαδήποτε συσκευή στο εγγύς περιβάλλον του χρήστη εκτός από ένα σύστημα καταγραφής της φωνής ή/και εικόνας της του χρήστη που μπορούν να είναι αρκετά μέτρα μακριά.

2.3.5. Επικοινωνία Ανθρώπου - Μηχανής με χρήση χειρονομιών

Σήμερα υπάρχουν πολλές διαφορετικές συσκευές διαθέσιμες για επικοινωνία ανθρώπου-υπολογιστή βασισμένη στα χέρια. Μερικά παραδείγματα είναι το πληκτρολόγιο, το ποντίκι, το track-ball, το track-pad, το joystick, τα ηλεκτρονικά στυλό και τα τηλεχειριστήρια. Πιο περίπλοκα παραδείγματα περιλαμβάνουν ειδικά

γάντια (cyber-gloves), ποντίκια τριών διαστάσεων και μαγνητικές συσκευές παρακολούθησης. Ωστόσο παρά την ποικιλία των νέων συσκευών, η αλληλεπίδραση ανθρώπου-υπολογιστή διαφέρει ακόμα από πολλές απόψεις από την αλληλεπίδραση ανθρώπου με άνθρωπο. Η φυσική αλληλεπίδραση μεταξύ των ανθρώπων δεν περιλαμβάνει συσκευές επειδή έχουμε τη δυνατότητα να αισθανόμαστε το περιβάλλον με τα μάτια και τα αυτιά μας. Σε γενικές γραμμές, ο υπολογιστής πρέπει να είναι σε θέση να μιμηθεί εκείνες τις δυνατότητες με τις κάμερες και τα μικρόφωνα.

Η έρευνα λοιπόν προσανατολίζεται, όπως έχουμε ήδη αναφέρει, στη χρήση χειρονομιών για ΕΑΥ, και μάλιστα χωρίς τη χρήση ειδικών συσκευών (bare hand human-computer interaction). Οι ανθρώπινες χειρονομίες είναι μέσα μη λεκτικής αλληλεπίδρασης μεταξύ των ανθρώπων. Κυμαίνονται από τις απλές ενέργειες του χεριού για την κατάδειξη και κίνηση των αντικειμένων γύρω μας μέχρι τις πιο σύνθετες που εκφράζουν τα συναισθήματά μας και μας επιτρέπουν να επικοινωνήσουμε με άλλους.

Για να εκμεταλλευτούμε τη χρήση των χειρονομιών στην ΕΑΥ είναι απαραίτητο να παρέχουμε τα μέσα με τα οποία μπορούν να αναπαρασταθούν από τους υπολογιστές. Αυτή η αναπαράσταση των χειρονομιών απαιτεί οι διάφοροι δυναμικοί ή/και στατικοί σχηματισμοί του ανθρώπινου χεριού, του βραχίονα και άλλων τμημάτων του ανθρώπινου σώματος να είναι με κάποιο τρόπο μετρήσιμοι από τον υπολογιστή. Οι πρώτες προσπάθειες για την επίλυση αυτού του προβλήματος οδήγησαν στις μηχανικές συσκευές που μετρούν άμεσα τις γωνίες των αρθρώσεων και τη θέση του χεριού στο χώρο. Αυτές οι συσκευές αποτελούν τμήμα των συσκευών που βασίζονται σε ειδικά γάντια (glove-based devices). Τα αντίστοιχα συστήματα διεπαφών απαιτούν από το χρήστη τη χρήση μιας δυσκίνητης συσκευής που συνοδεύεται από πληθώρα καλωδίων για τη σύνδεσή της με τον υπολογιστή. Το γεγονός αυτό εμποδίζει την ευκολία και τη φυσικότητα με την οποία ο χρήστης μπορεί να αλληλεπιδράσει με το ελεγχόμενο από τον υπολογιστή περιβάλλον. Αν και η χρήση τέτοιων συσκευών μπορεί να δικαιολογηθεί από το ότι η περιοχή εφαρμογής είναι ιδιαίτερα εξειδικευμένη, όπως για παράδειγμα η προσομοίωση μιας χειρουργικής επέμβασης σε ένα περιβάλλον εικονικής πραγματικότητας, ένας απλός χρήστης θα διστάσει να χρησιμοποιήσει τέτοιου είδους δυσκίνητα εργαλεία διεπαφών. Αυτό έχει ωθήσει την έρευνα προς περισσότερο "φυσικές" τεχνικές ΕΑΥ. Οποιαδήποτε πιθανή αδεξιότητα στη χρησιμοποίηση των γαντιών και άλλου εξοπλισμού μπορεί να υπερνικηθεί με τη χρησιμοποίηση τεχνικών αλληλεπίδρασης που βασίζονται στο βίντεο και δεν απαιτούν κανενός είδους φυσική επαφή μεταξύ του χρήστη και της μηχανής (video-based non-contact interaction techniques).

Αυτή η προσέγγιση προτείνει τη χρήση ενός συνόλου από κάμερες και τεχνικές υπολογιστικής όρασης για την ερμηνεία των χειρονομιών. Η ευκολία χρήσης του συστήματος αυτού έχει οδηγήσει σε άνθηση της έρευνας σε αυτήν την περιοχή. Άλλοι παράγοντες που συνέβαλαν σε αυτό το αυξανόμενο ενδιαφέρον περιλαμβάνουν τη διαθεσιμότητα γρήγορων υπολογισμών, που καθιστούν εφικτή την επεξεργασία σε πραγματικό χρόνο, και την πρόσφατη πρόοδο στις τεχνικές της όρασης υπολογιστών. Πολυάριθμες προσεγγίσεις έχουν εφαρμοστεί στο πρόβλημα της οπτικής αναπαράστασης χειρονομιών για ΕΑΥ, πολλές από τις οποίες έχουν επιλεγεί και εφαρμοστεί έτσι ώστε να εστιάζουν σε μια συγκεκριμένη όψη των χειρονομιών, όπως παρακολούθηση χεριών (hand tracking), εκτίμηση χειρομορφών (hand posture estimation), ή ταξινόμηση της στάσης του χεριού (hand pose

classification). Πολλές μελέτες έχουν διεξαχθεί στο πλαίσιο μιας ιδιαίτερης εφαρμογής, όπως η χρησιμοποίηση ενός δάχτυλου ως δείκτη για τον έλεγχο μιας τηλεόρασης, ή η ερμηνεία της νοηματικής γλώσσας.

Βασικό στοιχείο για την ανάπτυξη των συστημάτων αναγνώρισης δυναμικών χειρονομιών χωρίς χρήση ειδικών συσκευών αποτελούν τα συστήματα αναγνώρισης στατικών χειρομορφών που βασίζονται στην υπολογιστική όραση. Τα συστήματα αυτά μπορούν να ταξινομηθούν σε αυτά που βασίζονται σε τρισδιάστατα μοντέλα και σε αυτά που βασίζονται στην εμφάνιση. Τα συστήματα της πρώτης κατηγορίας προσπαθούν να προσαρμόσουν ένα τρισδιάστατο μοντέλο του χεριού στην εικόνα που δέχονται ως είσοδο, παρέχοντας μια πλήρη περιγραφή της διάταξης (configuration) του χεριού. Συχνά τα συστήματα αυτά είναι περιορισμένα σχετικά με το σχηματισμό, τη θέση ή τον προσανατολισμό του χεριού και είναι υπολογιστικά δαπανηρά. Τα συστήματα που ανήκουν στη δεύτερη κατηγορία στηρίζονται σε χαρακτηριστικά γνωρίσματα που περιγράφουν τη 2Δ-προβολή του χεριού στο επίπεδο της εικόνας. Εφαρμόζονται εάν δεν μας ενδιαφέρει ο ακριβής σχηματισμός του χεριού, αλλά οι είσοδοι πρέπει να ταξινομηθούν σε διαφορετικές χειρομορφές. Ακολουθείται συχνά η κλασσική διαδικασία επεξεργασίας εικόνας: απόκτηση εικόνας - κατάτμηση - εξαγωγή χαρακτηριστικών γνωρισμάτων - ταξινόμηση.

Στη συνέχεια παρατίθενται παραδείγματα εφαρμογών όπου το γυμνό χέρι είναι πιο πρακτικό από τις παραδοσιακές συσκευές εισόδου:

- ✓ Τα τηλεχειριστήρια για την τηλεόραση, τα στερεοφωνικά συγκροτήματα και τα φώτα μπορούν να αντικατασταθούν με το χέρι, με χρήση κατάλληλων χειρονομιών.
- ✓ Τα οικιακά ρομπότ μπορούν να ελεγχθούν με χειρονομίες.
- ✓ Οι κινητές συσκευές με πολύ περιορισμένο χώρο για σύστημα διεπαφής με το χρήστη μπορούν να λειτουργούν με χειρονομίες.
- ✓ Διευκολύνεται η επικοινωνία μεταξύ κωφών και μη κωφών ανθρώπων, αφού καθίσταται δυνατή η αναγνώριση της νοηματικής γλώσσας μέσω ανάλυσης των χειρονομιών.
- ✓ Κατά τη διάρκεια τηλεδιασκέψεων, ο εκάστοτε ομιλητής θα στρέφει την κάμερα προς το μέρος του απλά τεντώνοντας ένα δάχτυλό του, όπως γίνεται και σε μια τάξη.
- ✓ Κατά τη διάρκεια μιας παρουσίασης, ο παρουσιαστής δεν είναι απαραίτητο να κινείται συνεχώς μεταξύ του υπολογιστή και της οθόνης για να επιλέξει την επόμενη διαφάνεια.

Επιπλέον, οι αντιληπτικές (perceptual) διεπαφές επιτρέπουν τη δημιουργία υπολογιστών που δεν γίνονται αντιληπτοί υπό αυτήν τη μορφή. Χωρίς οθόνη, ποντίκι και πληκτρολόγιο, ένας υπολογιστής μπορεί να κρυφτεί σε οικιακές συσκευές, αυτοκίνητα, μηχανές πώλησης και παιχνίδια. Τα κύρια πλεονεκτήματα των διεπαφών αυτών σε σχέση με τα παραδοσιακά κουμπιά και τους διακόπτες περιγράφονται στη συνέχεια:

- ✓ Τα συστήματα μπορούν να ενσωματωθούν σε πολύ μικρές επιφάνειες.
- ✓ Τα συστήματα μπορούν να χρησιμοποιηθούν από ορισμένη απόσταση.
- ✓ Ο αριθμός μηχανικών μερών μέσα σε ένα σύστημα μπορεί να μειωθεί, καθιστώντας το ανθεκτικότερο.

- ✓ Είναι εφικτός ο σχεδιασμός κορυφών συσκευών (όπως για παράδειγμα ένα CD-player χωρίς κανένα πλήκτρο).
- ✓ Τα συστήματα μπορούν να προστατευθούν από την καταστροφή με την εισαγωγή μιας απόστασης ασφάλειας μεταξύ του χρήστη και της συσκευής.
- ✓ Σε συνδυασμό με την αναγνώριση φωνής, η αλληλεπίδραση ανθρώπου-μηχανής μπορεί να απλοποιηθεί πολύ.

Τέλος, υπάρχει μια κατηγορία εφαρμογών, η οποία μπορεί να ενσωματωθεί σε συνδυασμό με μια μηχανή προβολής (projector). Εικονικά αντικείμενα που προβάλλονται επάνω σε έναν τοίχο ή ένα τραπέζι μπορούν να διαχειρίζονται άμεσα με τα δάχτυλα. Αυτή η δομή μπορεί να είναι χρήσιμη με διάφορους τρόπους:

- ✓ Διάφορα πρόσωπα μπορούν ταυτόχρονα να εργαστούν με τα προβαλλόμενα αντικείμενα.
- ✓ Τα φυσικά συστήματα μπορούν να αντικατασταθούν με τα αντίστοιχα ψηφιακά. Η ψηφιακή έκδοση μπορεί εύκολα να αποθηκεύεται, να τυπώνεται και να αποστέλλεται μέσω του Διαδικτύου.
- ✓ Εάν η μηχανή προβολής και η κάμερα τοποθετηθούν σε μια θέση που δεν είναι προσιτή στο χρήστη, δημιουργείται μια σχεδόν ακατάλυτη διεπαφή. Για το χρήστη, η φυσική υπόσταση του υπολογιστή είναι μόνο ο τοίχος στον οποίο προβάλλεται η διεπαφή.

Πρέπει να σημειωθεί, ότι τα συστήματα λεκτικής αναγνώρισης είναι επίσης ικανά να παρέχουν μερικές από τις παραπάνω ιδιότητες. Οι τεχνικές που βασίζονται όμως στην όραση υπολογιστών έχουν το πλεονέκτημα ότι δεν ενοχλούν τη ροή της συνομιλίας (για παράδειγμα κατά τη διάρκεια μιας παρουσίασης) και λειτουργούν σε θορυβώδη περιβάλλοντα (για παράδειγμα στις εφαρμογές δημόσιων χώρων).

Σημειώνεται επίσης ότι συναντώνται και αρκετές δυσκολίες στην ανάπτυξη συστημάτων ΕΑΥ που βασίζονται σε αναγνώριση και παρακολούθηση χειρονομιών, αφού τα συστήματα αυτά θα πρέπει να είναι σε θέση να λειτουργήσουν επιτυχώς υπό συνθήκες μη ελεγχόμενου φωτισμού, ανεξάρτητα από το φόντο μπροστά από το οποίο στέκεται ο χρήστης. Επιπλέον, αντικείμενα παραμορφώσιμα και αρθρωτά όπως τα χέρια οδηγούν σε αυξανόμενη δυσκολία όχι μόνο της διαδικασίας κατάτμησης αλλά και της αναγνώρισης της μορφής τους.

2.3.6. Επικοινωνία ανθρώπου - ρομπότ

Ένας κλάδος της επικοινωνίας ανθρώπου-μηχανής είναι και η επικοινωνία ανθρώπου-ρομπότ. Με αυτή τη μορφή επικοινωνίας η χρήση των ρομπότ μπορεί να γενικευτεί. Η αλληλεπίδραση ανθρώπου-ρομπότ μπορεί να είναι φυσικού, λειτουργικού και διανοητικού είδους.

Η αλληλεπίδραση φυσικού είδους αποτελείται από ανταλλαγή επαφής και δυνάμεων και απαιτεί κοινό περιβάλλον εργασίας και ύπαρξη επαφής μεταξύ του ρομπότ και του ανθρώπου. Κυμαίνεται από διαχωρισμό του χώρου εργασίας, όπου δεν επιτρέπεται καμία φυσική αλληλεπίδραση, μέχρι από κοινού χρήση του χώρου εργασίας, με φυσική επαφή. Αλληλεπίδραση λειτουργικού είδους περιλαμβάνει συνεργασία στην εκτέλεση των εργασιών, και κυμαίνεται από ένα κατώτατο επίπεδο όπου το ρομπότ και ο άνθρωπος εκτελούν ανεξάρτητα την εργασία τους

(για παράδειγμα ο άνθρωπος υποδεικνύει και το ρομπότ εκτελεί), μέχρι τη συνεργασία ανθρώπου-ρομπότ στην εκτέλεση των εργασιών, και τον τηλεχειρισμό του ρομποτικού συστήματος από το χρήστη. Αλληλεπίδραση διανοητικού είδους σημαίνει ανταλλαγή των πληροφοριών και της γνώσης, η οποία απεικονίζεται για παράδειγμα, στην εκμάθηση και τη διδασκαλία ικανοτήτων. Σε αυτήν την περίπτωση, δεν μπορούμε να έχουμε καμία ανταλλαγή πληροφοριών, έμμεση (για παράδειγμα στην αυτόματη προσαρμοστικότητα του συστήματος στις προτιμήσεις του χρήστη) ή άμεση (για παράδειγμα στη ρητή διδασκαλία από το χρήστη).

Ο εντοπισμός και η παρακολούθηση των χρηστών είναι μια βασική εργασία για κάθε ρομπότ που προορίζεται για την εξυπηρέτηση των ανθρώπων σε διάφορους τομείς της καθημερινής ζωής. Στα πρώτα τηλερομποτικά συστήματα, χρησιμοποιήθηκαν ανθρώπινες χειρονομίες για να ελέγξουν τις κινήσεις του ρομπότ υπό την έννοια της δομής αφέντη-σκλάβου (master-slave), όπου ο άνθρωπος χειριζόταν το ρομπότ-αφέντη και το ρομπότ-σκλάβος προσπαθούσε να μιμηθεί τις κινήσεις του ρομπότ-αφέντη. Επομένως, το ρομπότ πρέπει να ανιχνεύει τους χρήστες σε μια ευρεία περιοχή λειτουργίας ενώ συγχρόνως είναι επιθυμητό να συγκεντρώνει πληροφορίες όπως η ταυτότητα, το γένος και η ηλικία του χρήστη για να προσαρμόσει αναλόγως το διάλογο. Αυτοί οι δύο στόχοι είναι λίγο πολύ αντιφατικοί: για τον πρώτο πρέπει να αναλυθούν πλήρως τα περιχώρα του ρομπότ, το οποίο μπορεί να επιτευχθεί με τη χρήση μιας πανοραμικής εικόνας με οπτικό πεδίο 360° και χαμηλή ανάλυση, ενώ για το δεύτερο απαιτείται μια εικόνα του προσώπου του χρήστη υψηλής ανάλυσης.

Η χρήση χειρονομιών για την επικοινωνία ανθρώπου-ρομπότ παρέχει μια φυσική και ισχυρή διεπαφή, η οποία μπορεί να χρησιμοποιηθεί για να δώσει σε ένα ρομπότ εντολές για την εκτέλεση ορισμένων εργασιών ή για τον έλεγχο των λειτουργιών του. Εντούτοις, η αυτόματη αναγνώριση χειρονομιών είναι ιδιαίτερα ενδιαφέρουσα λόγω της παρουσίας ασαφών τοποθετήσεων του χεριού, που δεν αποτελούν χειρονομίες, καθώς και της κίνησής του. Όπως έχει αναφερθεί, χρησιμοποιούνται συνήθως δύο προσεγγίσεις για την αναπαράσταση των χειρονομιών στην ΕΑΥ. Η πρώτη προσέγγιση βασίζεται στη χρήση ειδικών γαντιών (gloved based approach) και περιλαμβάνει χρήση δυσκίνητων συσκευών και ένα πλήθος καλωδίων που συνδέουν τις συσκευές αυτές σε έναν υπολογιστή. Μια δεύτερη προσέγγιση βασίζεται στην όραση υπολογιστών (vision based technique) και δεν απαιτεί τη χρήση κάποιας συσκευής, αλλά τη χρήση ενός συνόλου καταγραφικών συσκευών (video cameras) και τεχνικών υπολογιστικής όρασης για την αναπαράσταση των χειρονομιών.

Αυτού του είδους τα συστήματα προσανατολίζονται σε εφαρμογές τηλεχειρισμού, που απαιτούν λειτουργία με ελεύθερα χέρια (hands-free operation). Όπως η προφορική επικοινωνία, έτσι και οι χειρονομίες είναι ένας φυσικός τρόπος να πραγματοποιηθεί η επικοινωνία ανθρώπου-μηχανής. Παρά τη χρήση των γλωσσών προγραμματισμού, ο χειροκίνητος έλεγχος είναι ο πιο αξιόπιστος τρόπος να εκτελεστούν σύνθετες εργασίες σε μη δομημένα περιβάλλοντα. Σε αυτές τις καταστάσεις, ένα παθητικό σύστημα που λειτουργεί από απόσταση και δεν απαιτεί φυσική επαφή (non-contact, passive and remote system), μπορεί να χρησιμοποιηθεί για τον τηλεχειρισμό ενός ρομπότ με χειρονομίες.

3

Ανίχνευση και Παρακολούθηση Ανθρώπινων Χειρονομιών

3.1. Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται αρχικά η γενική θεωρία γύρω από το πρόβλημα της παρακολούθησης κινούμενων αντικειμένων (tracking of moving objects) σε ακολουθίες εικόνων (video). Στη συνέχεια του κεφαλαίου η προσοχή επικεντρώνεται σε μια εφαρμογή του προβλήματος αυτού, την παρακολούθηση ανθρώπινων χειρονομιών, που αποτελεί και το αντικείμενο της παρούσας εργασίας. Γίνεται μια εισαγωγή στη δομή του συστήματος ανίχνευσης, παρακολούθησης και αναγνώρισης χειρονομιών που υλοποιήθηκε, και συγκεκριμένα περιγράφονται συνοπτικά οι τρεις βασικές τεχνικές που χρησιμοποιήθηκαν: χρήση χρώματος για ταξινόμηση των εικονοστοιχείων (pixels) της εικόνας, particle filtering και σύμπτωση προτύπων (template matching). Οι μέθοδοι αυτές εξετάζονται εκτενέστερα στα κεφάλαια 4, 5 και 6 αντίστοιχα, ενώ αναλυτικά το σύστημα που υλοποιήθηκε περιγράφεται στο κεφάλαιο 7.

3.2. Παρακολούθηση κινούμενων αντικειμένων

3.2.1. Τοποθέτηση προβλήματος

Η παρακολούθηση κινούμενων αντικειμένων (tracking of moving objects) σε ακολουθία εικόνων (video) αποτελεί πρόβλημα της Όρασης Υπολογιστών και περιλαμβάνει δύο υπο-προβλήματα. Το πρώτο είναι ο εντοπισμός του κινούμενου αντικειμένου στην εικόνα, ενώ το δεύτερο είναι η παρακολούθησή του. Η παρακολούθηση πραγματικών τρισδιάστατων αντικειμένων σε ακολουθίες εικόνων βρίσκει εφαρμογή σε περιοχές όπως η κωδικοποίηση video, η παρακολούθηση διάφορων χώρων (surveillance) και η ανάκτηση δεδομένων.

3.2.2. Ανίχνευση κινούμενου αντικειμένου

Το πρόβλημα της ανίχνευσης ενός αντικειμένου σε μια εικόνα αποτελεί εφαρμογή του γενικότερου προβλήματος της κατάτμησης εικόνας (image segmentation). Ως κατάτμηση εικόνας μπορούμε να ορίσουμε τη διαδικασία απομόνωσης των αντικειμένων της εικόνας από το φόντο, δηλαδή το χωρισμό της εικόνας σε ανεξάρτητες περιοχές, με βάση χαρακτηριστικά όπως φωτεινότητα, υφή, χρώμα, κίνηση, βάθος. Λόγω της φύσης του προβλήματος η διαδικασία της κατάτμησης δεν είναι μονοσήμαντη, αλλά εξαρτάται τόσο από τα χαρακτηριστικά που θα επιλεγούν για να εκτελεστεί, όσο και από την εφαρμογή για την οποία θα χρησιμοποιηθεί.

Στην κατάτμηση παρουσιάζεται μια δυϊκότητα μεταξύ μεθόδων που βασίζονται στα όρια των αντικειμένων (boundary based) και μεθόδων που βασίζονται σε περιοχές (region based). Στις τεχνικές που ανήκουν στην πρώτη κατηγορία περιλαμβάνονται μέθοδοι ανίχνευσης ακμών και μέθοδοι ενεργών περιγραμμάτων. Οι δημοφιλέστερες τεχνικές αυτή τη στιγμή είναι αυτές των Γεωδαιτικών Ενεργών Περιγραμμάτων (Geodesic Active Contours) που υλοποιούνται με χρήση επιπεδοσυνόλων (level sets) και εξέλιξη καμπύλης βάσει μη γραμμικών γεωμετρικά διαχεόμενων μερικών διαφορικών εξισώσεων (nonlinear geometric-diffusion partial differential equations). Στην κατηγορία μεθόδων που βασίζονται σε περιοχές υπάρχουν τρεις υπο-κατηγορίες μεθόδων: μέθοδοι επέκτασης περιοχής (region growing methods) όπως οι μέθοδοι split-merge και κατάτμηση watershed, στατιστικές μέθοδοι που βασίζονται σε Markov Random Fields και διαφορικές μέθοδοι (variational methods) που βασίζονται στην ελαχιστοποίηση συναρτησιακών ενέργειας.

Ένα από τα χαρακτηριστικά που χρησιμοποιείται αρκετά συχνά στην κατάτμηση εικόνων είναι το χρώμα των περιοχών ενδιαφέροντος, ιδιαίτερα όταν πρόκειται για περιοχές με ιδιαίτερα χαρακτηριστικά χρώματος, όπως για παράδειγμα το ανθρώπινο δέρμα. Η χρήση του χρώματος είναι ιδιαίτερα διαδεδομένη στην αναγνώριση ανθρώπινων χειρονομιών.

3.2.3. Παρακολούθηση του αντικειμένου

Αφού έχει εντοπιστεί το αντικείμενο ενδιαφέροντος στην εικόνα, το επόμενο βήμα είναι η παρακολούθησή του (tracking). Η "παρακολούθηση" ενός αντικειμένου σχετίζεται με την εκτίμηση της θέσης και του προσανατολισμού (αλλά και άλλων παραμέτρων) του αντικειμένου αυτού σε κάθε εικόνα. Έχουν αναπτυχθεί διάφορες μέθοδοι για την παρακολούθηση αντικειμένων, με κυρίαρχες αυτές που βασίζονται σε τεχνικές υπολογιστικής όρασης.

Στις πρώτες προσπάθειες ανήκει η εφαρμογή κατάλληλων κατωφλίων στη διαφορά μεταξύ των διαδοχικών εικόνων. Σε άλλες μεθόδους χρησιμοποιείται η έννοια της κίνησης και συγκεκριμένα της οπτικής ροής (optical flow): όταν η κάμερα και το περιβάλλον βρίσκονται σε σχετική κίνηση το πεδίο φωτεινότητας που καταγράφεται από την κάμερα μεταβάλλεται και το προκύπτον 2D διανυσματικό πεδίο ταχύτητας της προβολής των αντικειμένων στο επίπεδο της εικόνας ονομάζεται οπτική ροή. Γενικά, ο υπολογισμός της οπτικής ροής γίνεται σε δύο στάδια: εκτίμηση της οπτικής ροής, δηλαδή της 2D κίνησης στο επίπεδο της εικόνας και εκτίμηση της 3D κίνησης και πιθανόν του 3D σχήματος των

αντικειμένων. Υπάρχουν επίσης στατιστικές μέθοδοι για την παρακολούθηση αντικειμένων. Μια μέθοδος που ανήκει στην τελευταία κατηγορία και έχει εξελιχθεί αρκετά τα τελευταία χρόνια ονομάζεται *particle filtering*. Το *particle filter* εκτελεί μια τυχαία αναζήτηση οδηγούμενο από ένα στοχαστικό μοντέλο κίνησης για την εκτίμηση της εκ των υστέρων κατανομής (*posterior distribution*) που περιγράφει το σχηματισμό του αντικειμένου. Με τη μέθοδο αυτή θα ασχοληθούμε εκτενέστερα σε επόμενο κεφάλαιο.

Η έμφυτη δυσκολία της παρακολούθησης αντικειμένων είναι ότι κατά την κίνηση ενός συμπαγούς αντικειμένου σε μια τρισδιάστατη σκηνή, το σχήμα του είναι μια προοπτική προβολή στο επίπεδο της εικόνας. Γενικά, για να επιτευχθεί εύρωστη παρακολούθηση, χρησιμοποιούνται γεωμετρικά μοντέλα για αυτόν το μετασχηματισμό σχήματος. Στην απλούστερη περίπτωση, το σχήμα του αντικειμένου υποβάλλεται σε ένα μετασχηματισμό συντεταγμένων από τη μια εικόνα στην επόμενη της ακολουθίας, για τον οποίο αρκεί μια απλή τεχνική συσχέτισης. Το μοντέλο του μετασχηματισμού μπορεί να επεκταθεί και να συμπεριλάβει περιστροφή, αλλαγή κλίμακας (*scaling*), και γενικότερα έναν αφινικό μετασχηματισμό (*affine transformation*). Όταν είναι διαθέσιμα επαρκή χαρακτηριστικά γνωρίσματα (όπως οι ακμές των αντικειμένων), τα αντικείμενα μπορούν να παρακολουθηθούν αξιόπιστα με αυτόν τον τρόπο. Εντούτοις, για μη συμπαγή αντικείμενα, όπως οι άνθρωποι, η κίνηση των αντικειμένων είναι πιο σύνθετη από τον αφινικό μετασχηματισμό. Για την παρακολούθηση μη συμπαγών αντικειμένων έχει προταθεί το φίλτρο Kalman, μέθοδος που αναφέρεται σε επόμενο κεφάλαιο.

Τα διαφορετικά χαρακτηριστικά παρέχουν διαφορετικά στοιχεία για αλγορίθμους παρακολούθησης. Οι μέθοδοι που βασίζονται στις ακμές (*edge-based approaches*) βρίσκουν τις ακμές των αντικειμένων στις εικόνες, και οι μέθοδοι που βασίζονται σε περιοχές (*region-based approaches*) βρίσκουν δεδομένα πρωτότυπα αντικειμένων (*templates*) στις εικόνες. Υποθέτοντας και εδώ ότι η κίνηση των αντικειμένων μεταξύ δύο διαδοχικών εικόνων είναι μικρή, αυτές οι μέθοδοι θα μπορούσαν να επιτύχουν ικανοποιητικά αποτελέσματα. Εντούτοις, όταν αυτή η υπόθεση δεν ισχύει, το οποίο στην πράξη είναι πολύ πιθανό να συμβεί, αυτοί οι αλγόριθμοι θα έχαναν πιθανώς την τροχιά του αντικειμένου, και θα ήταν απαραίτητες κατάλληλες διορθώσεις. Επιπλέον, αυτές οι μέθοδοι απαιτούν συνήθως χειροκίνητη αρχικοποίηση. Μια εναλλακτική λύση είναι η *blob-based* προσέγγιση, η οποία δεν χρησιμοποιεί τοπικές πληροφορίες όπως οι ακμές και οι περιοχές της εικόνας. Αντ' αυτού, ο στόχος αντιπροσωπεύεται με το χρώμα και την κίνησή του έτσι ώστε ο εντοπισμός και η παρακολούθηση να μπορούν να πραγματοποιηθούν με κατάτμηση του στόχου στις εικόνες. Για παράδειγμα, όταν μας ενδιαφέρει ο εντοπισμός του χεριού σε *video*, είναι πολύ δύσκολο να παρασταθεί το χέρι βάσει μόνο των ακμών λόγω των ιδιαίτερα αρθρωτών κινήσεων των δάχτυλων. Επιπλέον, υπάρχουν πολλές παραλλαγές στην εμφάνιση του χεριού από διαφορετικές κατευθύνσεις παρατήρησης. Επίσης, αν παρατηρήσουμε τη μοναδικότητα του χρωματικού τόνου του δέρματος, καταλήγουμε στο ότι οι μέθοδοι κατάτμησης που βασίζονται στο χρώμα θα μπορούσαν να οδηγήσουν σε αποδοτικό και σθεναρό οπτικό εντοπισμό. Βεβαίως, ο συνδυασμός των ανωτέρω δύο προσεγγίσεων με την ενσωμάτωση των πολλαπλών οπτικών στοιχείων θα οδηγούσε σε ακόμα πιο ισχυρά συστήματα.

Υπάρχουν δύο βασικές προσεγγίσεις για το πρόβλημα της παρακολούθησης. Σύμφωνα με την πρώτη, διατηρείται στη μνήμη η τελευταία γνωστή θέση του

αναγνωρισμένου αντικειμένου. Λαμβάνοντας υπόψη μερικούς γνωστούς περιορισμούς για τις πιθανές μετακινήσεις του αντικειμένου μεταξύ δύο διαδοχικών εικόνων, ένας αλγόριθμος παρακολούθησης μπορεί να παρακολουθήσει το αντικείμενο κατά τη διάρκεια του χρόνου. Η δεύτερη δυνατότητα είναι η επανάληψη της αναγνώρισης για κάθε εικόνα. Παρόλο που αυτή η προσέγγιση φαίνεται μάλλον απλοϊκή και απαιτεί έναν πολύ γρήγορο αλγόριθμο ανίχνευσης, είναι ίσως η μόνη εφικτή τεχνική παρακολούθησης για ελεύθερη κίνηση χεριών για δύο λόγους [von Hardenberg, Bérard, 2001]:

- ✓ Οι μετρήσεις δείχνουν ότι τα χέρια μπορούν να φθάσουν σε ταχύτητες 5 m/s, οπότε σε ένα ρυθμό 25 frames/sec το χέρι εκτελεί "άλματα" με βήμα μέχρι 20cm ανά πλαίσιο.
- ✓ Γρήγορες κινήσεις των δάχτυλων οδηγούν σε θόλωση της εικόνας (motion blurring). Τα θολωμένα δάχτυλα είναι σχεδόν αδύνατο να αναγνωριστούν και επομένως οι θέσεις των δαχτύλων κατά τη διάρκεια μιας γρήγορης κίνησης είναι άγνωστες.

Αν τα παραπάνω συνδυαστούν, τότε θα προκύψουν αποστάσεις της τάξης μεγέθους του ενός μέτρου μεταξύ δύο προσδιορισμένων θέσεων δάχτυλων, για γρήγορη κίνηση του χεριού. Αυτό απαιτεί αναζήτηση στο μεγαλύτερο μέρος της εικόνας από τον αλγόριθμο παρακολούθησης. Με την παράλειψη του σταδίου παρακολούθησης δεν αποδεσμεύεται πολλή ταχύτητα επεξεργασίας, αλλά εξασφαλίζεται αρκετή σταθερότητα.

3.2.4. Σύνοψη

Ανακεφαλαιώνοντας, το πρόβλημα της παρακολούθησης ενός κινούμενου αντικειμένου αφορά στη διατήρηση μιας εκτίμησης για τη διάταξη του αντικειμένου στο χώρο και στο χρόνο. Το βασικό πλεονέκτημα της παρακολούθησης έναντι της απλής ανίχνευσης ενός αντικειμένου είναι ότι σε ασαφείς/διφορούμενες περιπτώσεις η ιστορία, δηλαδή οι προηγούμενες εκτιμήσεις, οδηγεί στην εύρεση της σωστής εκτίμησης για τη παρούσα κατάσταση. Ένα μειονέκτημα είναι ότι είναι απαραίτητο να γίνουν υποθέσεις για τη συμπεριφορά του υπό παρακολούθηση αντικειμένου. Επίσης, είναι πολύ εύκολο να οδηγηθούμε σε λανθασμένη εκτίμηση και πολλές φορές είναι δύσκολη η επαναφορά στη σωστή εκτίμηση χωρίς επανεκκίνηση της διαδικασίας παρακολούθησης.

3.3. Ανίχνευση και παρακολούθηση χειρονομιών

3.3.1. Εισαγωγή

Όπως έχει ήδη αναφερθεί, η αναγνώριση και παρακολούθηση χειρονομιών από συστήματα υπολογιστών αποτελεί εφαρμογή του γενικότερου προβλήματος της παρακολούθησης αντικειμένων και έχει ευρύ πεδίο εφαρμογών που περιλαμβάνουν κυρίως τηλεχειρισμό απλών ή σύνθετων συσκευών και εκτέλεση εργασιών χωρίς ανάγκη ύπαρξης επαφής με τη μηχανή που τις εκτελεί. Στις επόμενες

παραγράφους δίνεται μια συνοπτική περιγραφή των μεθόδων που ακολουθήθηκαν για το σχεδιασμό των διάφορων σταδίων του συστήματος ανίχνευσης και παρακολούθησης χειρονομιών που υλοποιήθηκε για τις ανάγκες της παρούσας εργασίας:

- ✓ Χρήση του χρώματος του δέρματος με σκοπό τη δημιουργία ενός χρωματικού μοντέλου για την ταξινόμηση των εικονοστοιχείων (pixels) της εικόνας σύμφωνα με το χρώμα τους.
- ✓ Παρακολούθηση του κινούμενου χεριού με τη μέθοδο particle filtering.
- ✓ Αναγνώριση χειρομορφών με χρήση της τεχνικής της σύμπτωσης προτύπων (template matching).

3.3.2. Ταξινόμηση εικονοστοιχείων (pixels) με βάση το χρώμα

Έχουν αναπτυχθεί πολλά διαφορετικά συστήματα αναγνώρισης χειρονομιών και στα περισσότερα από αυτά βασικό χαρακτηριστικό για την ανίχνευση των χεριών αποτελεί το χρώμα. Μια απλή προσέγγιση είναι να συλλεχθούν δείγματα pixels από το στόχο που ανήκουν σε χρώμα δέρματος για την εκπαίδευση ενός ταξινομητή χρώματος, έτσι ώστε να μπορεί να γίνει κατάτμηση των περιοχών δέρματος με ταξινόμηση και ομαδοποίηση των pixels της εικόνας εισόδου. Μια από τις λύσεις για να αντιμετωπιστεί η δυσκολία που παρουσιάζεται λόγω της ύπαρξης μεγάλης ποικιλίας στο χρωματικό τόνο του δέρματος μεταξύ διαφορετικών ανθρώπων είναι να ρυθμιστεί ο ταξινομητής χρώματος χρησιμοποιώντας ένα εκτεταμένο σύνολο εκπαίδευσης που περιέχει δείγματα από πολλούς ανθρώπους. Δυστυχώς, στην πράξη, υπάρχουν ακόμα μερικές δυσκολίες. Μια από αυτές είναι ότι οι κατανομές χρώματος μπορούν να αλλάξουν με τις συνθήκες φωτισμού. Ακόμη και στην περίπτωση σταθερών πηγών φωτός, το χρώμα του στόχου μπορεί ακόμα να είναι διαφορετικό σε μια ακολουθία εικόνων, αφού ο στόχος μπορεί να σκιαστεί από άλλα αντικείμενα. Κατά συνέπεια, ένα σταθερό μοντέλο χρώματος μπορεί να μην λειτουργήσει καλά υπό όλες τις συνθήκες. Μια ακόμα δυσκολία είναι ότι η συλλογή ενός εκτεταμένου συνόλου εκπαίδευσης δεν είναι τετριμμένη διαδικασία. Μια αποτελεσματική προσέγγιση σε αυτό το πρόβλημα είναι η δημιουργία ενός προσαρμοστικού ταξινομητή, δηλαδή η προσαρμογή του μοντέλου χρώματος σε διαφορετικές συνθήκες φωτισμού και σε διαφορετικούς ανθρώπους.

3.3.3. Particle filtering

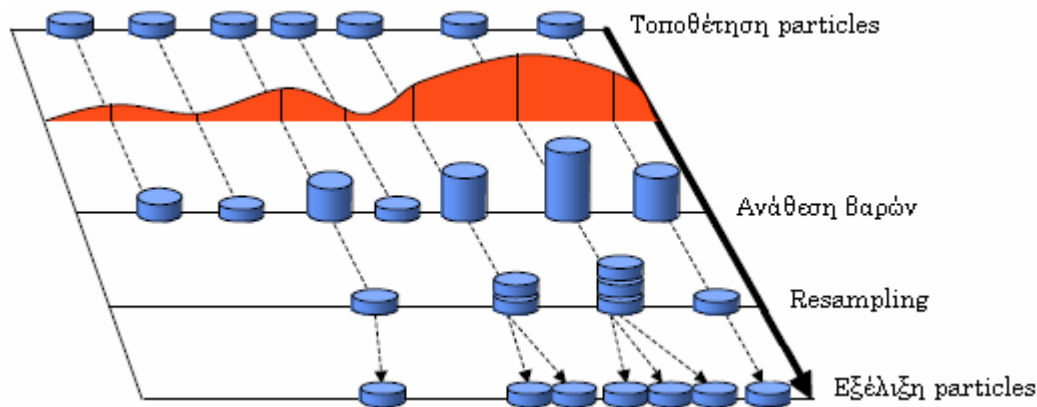
Μια από τις μεθόδους για την παρακολούθηση ενός κινούμενου αντικειμένου είναι αυτή του particle filtering. Πρόκειται για στατιστική μέθοδο που βασίζεται στον κανόνα του Bayes και έχει χρησιμοποιηθεί σε πολλά συστήματα παρακολούθησης αντικειμένων.

Η βασική ιδέα της μεθόδου είναι η χρήση δειγμάτων (particles) για την αναπαράσταση του μοντέλου του συστήματος, αντί για χρήση Gaussian κατανομών ή οποιουδήποτε άλλου μοντέλου, και η διάδοση μόνο των επικρατέστερων δειγμάτων. Αναλυτικότερα η θεωρία της μεθόδου περιγράφεται στο τέταρτο κεφάλαιο.

Τα βήματα που ακολουθούνται για την παρακολούθηση ενός αντικειμένου με particle filtering είναι:

- ✓ Τοποθέτηση n δειγμάτων (particles) στο χώρο κατάστασης (state-space) του συστήματος.
- ✓ Ανάθεση βαρών στα δείγματα με βάση τις παρατηρήσεις (μετρήσεις) που είναι διαθέσιμες για το σύστημα.
- ✓ Επαναδειγματοληψία (resampling): δίνεται περισσότερη βαρύτητα στα δείγματα που έχουν μεγαλύτερο βάρος και "ακυρώνονται" τα δείγματα με μικρά βάρη. Αυτό γίνεται με επαναδειγματοληψία της κατανομής.
- ✓ Εξέλιξη δειγμάτων.
- ✓ Επανάληψη των παραπάνω από το δεύτερο βήμα.

Στο παρακάτω σχήμα φαίνονται αυτά τα βήματα:



Εικόνα 3.1: Βασικά βήματα της μεθόδου particle filtering

3.3.4. Σύμπτωση Προτύπων (Template Matching)

Το τελευταίο στάδιο στην αναγνώριση των χειρονομιών είναι η αναγνώριση της συγκεκριμένης τοποθέτησης του χεριού (χειρομορφής), έτσι ώστε να είναι δυνατός ο διαχωρισμός μεταξύ διαφορετικών χειρονομιών και η αντιστοιχησή τους σε διαφορετικές λειτουργίες όταν πρόκειται για εφαρμογή επικοινωνίας ανθρώπου-μηχανής. Ένας απλός τρόπος για την αναγνώριση των χειρονομιών είναι η τεχνική της σύμπτωσης προτύπων (template matching).

Η σύμπτωση προτύπων αποτελεί την πιο απλοϊκή μορφή αναγνώρισης προτύπων (pattern recognition). Στην τεχνική αυτή αποθηκεύεται για κάθε διαφορετικό πρότυπο, εν προκειμένω για κάθε διαφορετική χειρομορφή, ένα ίχνος ή μάσκα ή πρωτότυπο. Το πρότυπο εισόδου συγκρίνεται με καθένα από τα πρωτότυπα και η ταξινόμηση βασίζεται σε ένα προαποφασισμένο κριτήριο ομοιότητας. Η απόφαση λαμβάνεται για αντιστοιχησή της εισόδου σε εκείνο το πρωτότυπο για το οποίο η ομοιότητα είναι μεγαλύτερη.

Η τεχνική αυτή παράγει τα καλύτερα αποτελέσματα όταν οι τάξεις που αντιπροσωπεύουν τα πρωτότυπα είναι ισχυρά διαφοροποιημένες και υπάρχει μεγάλη συνέπεια στα πρωτότυπα.

4

Χρώμα

4.1. Εισαγωγή

Στο παρόν κεφάλαιο εξετάζονται οι διαφορετικοί τρόποι αναπαράστασης του χρώματος και η χρήση του χρώματος στα προβλήματα της Όρασης Υπολογιστών. Έμφαση δίνεται στη χρήση του χρώματος για την ανίχνευση περιοχών δέρματος (όπως για παράδειγμα τα χέρια) και για την κατασκευή χρωματικών μοντέλων που χρησιμοποιούνται στην ταξινόμηση των pixels μιας εικόνας. Σημειώνεται ότι η μεθοδολογία για την κατασκευή του χρωματικού μοντέλου με βάση το χρώμα του δέρματος χρησιμοποιήθηκε στην υλοποίηση του συστήματος ανίχνευσης, παρακολούθησης και αναγνώρισης χειρονομιών αυτής της εργασίας.

Οι εικόνες που γίνονται αντιληπτές από το ανθρώπινο μάτι αντιστοιχούν σε φασματική κατανομή ενέργειας σε μήκη κύματος περίπου από 400nm μέχρι 700nm. Η όραση με χρώμα βασίζεται στην ενεργοποίηση φωτοευαίσθητων περιοχών στους φωτο-αισθητήρες που διαθέτει το ανθρώπινο μάτι. Ο αμφιβληστροειδής χιτώνας του ματιού έχει τριών ειδών αισθητήρες, που ονομάζονται κωνία, με διαφορετικές φασματικές αποκρίσεις. Η αίσθηση του χρώματος δημιουργείται από αυτή τη διαφορά στην απόκριση των αισθητήρων σε διαφορετικά μήκη κύματος του ορατού φωτός καθώς και από τη χωρική επεξεργασία πληροφορίας από επόμενα στάδια του συστήματος όρασης. Οι άνθρωποι είναι σε θέση να αντιληφθούν εκατοντάδες χιλιάδες διαφορετικών χρωματικών τόνων και εντάσεων, και μόλις περίπου εκατό επίπεδα του γκρι (επίπεδα φωτεινότητας).

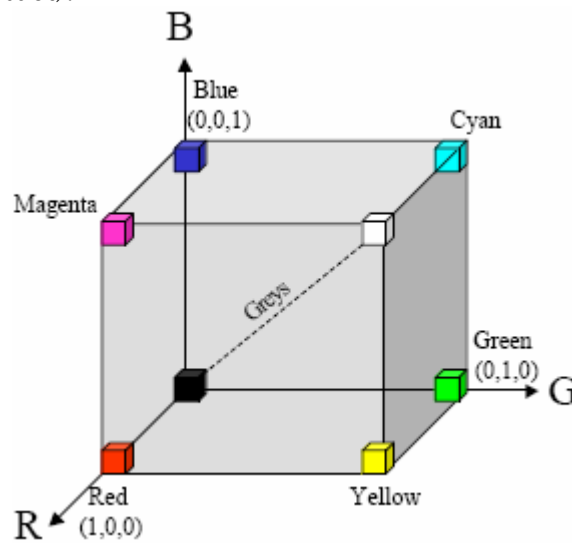
4.2. Αναπαράσταση χρώματος

4.2.1. Πρωταρχικά χρώματα

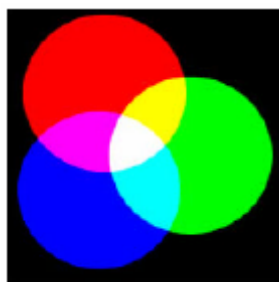
Κάθε χρώμα σε μια οθόνη υπολογιστή είναι γραμμικός συνδυασμός τριών πρωταρχικών χρωμάτων. Τα πρωταρχικά αυτά χρώματα δημιουργούν μια παλέτα

χρωμάτων που καλύπτει σχεδόν όλα τα πιθανά χρώματα. Υπάρχουν δύο σύνολα πρωταρχικών χρωμάτων: προσθετικά και αφαιρετικά.

Τα προσθετικά χρώματα είναι το κόκκινο, το πράσινο και το μπλε (red, green, blue RGB). Τα αφαιρετικά (ή συμπληρωματικά) χρώματα είναι το κυανό, η ματζέντα και το κίτρινο (cyan, magenta, yellow CMY). Οι τηλεοράσεις και οι οθόνες υπολογιστών παράγουν φως, το οποίο προστίθεται για την απεικόνιση του επιθυμητού χρώματος. Το χρώμα παράγεται από πρόσθεση κόκκινων, πράσινων και μπλε πηγών φωτός. Οι καλλιτέχνες στην ανάμιξη των χρωστικών ουσιών χρησιμοποιούν τα αφαιρετικά πρωταρχικά χρώματα, επειδή οι χρωστικές ουσίες που χρησιμοποιούν απορροφούν όλα τα χρώματα του φωτός εκτός από εκείνα που ανακλούν, και είναι έτσι μια πηγή αφαιρετικού φωτός. Το χρώμα απεικονίζεται στο λευκό χαρτί με την προσθήκη κυανής, ματζέντα και κίτρινης χρωστικής ουσίας. Για να είναι ορατή η εικόνα θα πρέπει να αντανακλάται φως από αυτή (δηλαδή δεν είναι ορατή στο σκοτάδι).



Εικόνα 4.1: RGB κύβος χρώματος¹



Εικόνα 4.2:
Προσθετικά χρώματα
(τηλεόραση, οθόνες Η/Υ)



Εικόνα 4.3:
Αφαιρετικά χρώματα
(ζωγραφική, εκτύπωση)

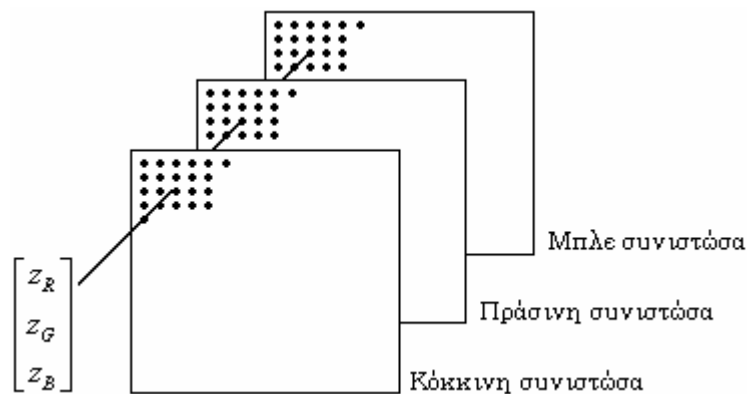
Δεν υπάρχει τρόπος να δημιουργηθεί μαύρο χρώμα με την προσθήκη κόκκινων, πράσινων και μπλε πηγών φωτός, οπότε θα πρέπει η οθόνη να είναι αρχικά μαύρη.

¹ Οι εικόνες 4.1, 4.2, 4.3 έχουν ληφθεί από τις διαφάνειες παρουσίασης του Luke Fletcher για το μάθημα Όραση Υπολογιστών στο Australia National University.
URL: <http://users.rsise.anu.edu.au/~luke/cvcourse.htm>

Όμοια, δεν υπάρχει τρόπος δημιουργίας λευκού με το συνδυασμό κυανής, ματζέντα και κίτρινης χρωστικής ουσίας, οπότε το χαρτί πρέπει να είναι άσπρο.

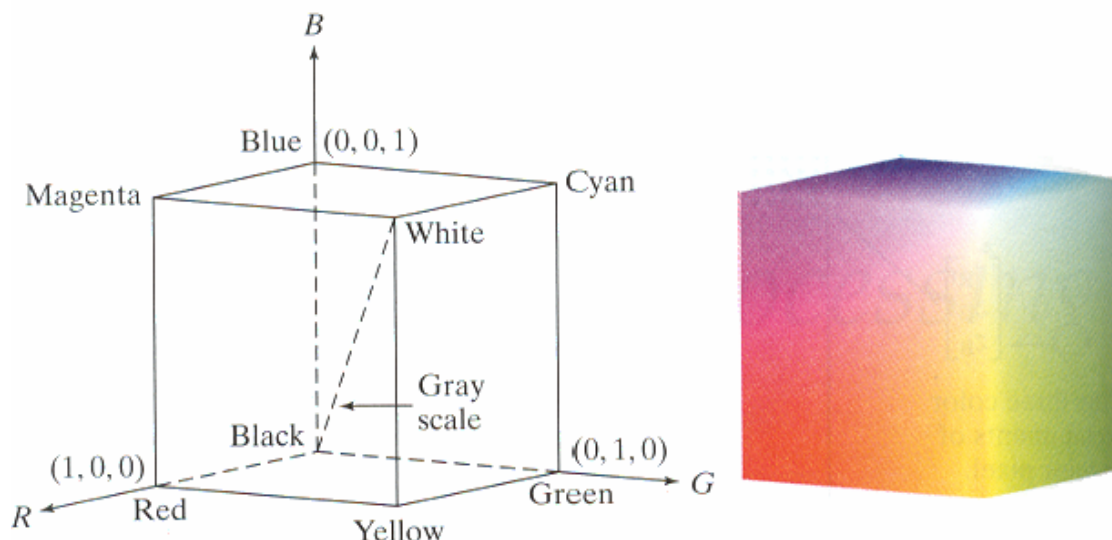
4.2.2. Γραμμικοί χώροι χρώματος

RGB χώρος χρώματος: Ο χώρος RGB προκύπτει επίσημα χρησιμοποιώντας μονοχρωματικά πρωταρχικά χρώματα: κόκκινο R στα 645,16nm, πράσινο G στα 526,32nm και μπλε B στα 444,44nm. Μια εικόνα RGB είναι ένας $M \times N \times 3$ πίνακας έγχρωμων pixels, όπου κάθε pixel αποτελείται από τρεις συνιστώσες (κόκκινη, πράσινη, μπλε) σε μια συγκεκριμένη θέση, όπως φαίνεται στο παρακάτω σχήμα:



Εικόνα 4.4: Χρωματικές συνιστώσες RGB εικόνας (Εικόνα από το βιβλίο "Digital Image Processing using MATLAB" των R. C. Gonzalez, R. E. Woods, S. L. Eddins)

Αυτές οι τρεις γκριζες εικόνες όταν τροφοδοτηθούν στην κόκκινη, πράσινη και μπλε είσοδο μιας έγχρωμης οθόνης παράγουν μια έγχρωμη εικόνα. Ο RGB χώρος χρώματος συνήθως απεικονίζεται γραφικά σαν ένας χρωματικός κύβος, όπως φαίνεται στο παρακάτω σχήμα:



Εικόνα 4.5: RGB κύβος χρώματος: Στην εικόνα αριστερά φαίνονται τα βασικά και δευτερεύοντα χρώματα. Τα σημεία πάνω στην κύρια διαγώνιο ανήκουν στην κλίμακα του γκρι. Στην εικόνα δεξιά φαίνεται ο RGB κύβος χρώματος με όλα τα χρώματα. (από το "Digital Image Processing using MATLAB", R. C. Gonzalez, R. E. Woods, S. L. Eddins)

Αξιζει να σημειωθεί ότι στον RGB χώρο χρώματος δεν μπορούν να αναπαρασταθούν όλα τα ορατά χρώματα.

NTSC ή ΥΙQ χώρος χρώματος: Ο NTSC χώρος χρώματος χρησιμοποιείται στην τηλεόραση στις Ηνωμένες Πολιτείες. Ένα από τα κύρια πλεονεκτήματά του συστήματος είναι ότι η πληροφορία της φωτεινότητας είναι χωρισμένη από την πληροφορία χρώματος και έτσι το ίδιο σήμα μπορεί να χρησιμοποιηθεί και για έγχρωμες και για ασπρόμαυρες τηλεοράσεις. Με το σύστημα NTSC, τα δεδομένα της εικόνας αποτελούνται από τρεις συνιστώσες: φωτεινότητα (Y), χροιά (I) και κορεσμός (Q), όπου η επιλογή των γραμμάτων ΥΙQ είναι συμβατική. Η συνιστώσα φωτεινότητας αντιπροσωπεύει τις πληροφορίες γκριζας-κλίμακας και οι άλλες δύο συνιστώσες φέρουν τις πληροφορίες χρώματος ενός τηλεοπτικού σήματος. Οι συνιστώσες ΥΙQ λαμβάνονται από τις RGB συνιστώσες μιας εικόνας χρησιμοποιώντας το μετασχηματισμό:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

YCbCr χώρος χρώματος: Ο χώρος χρώματος YCbCr χρησιμοποιείται ευρέως στο ψηφιακό βίντεο. Σε αυτόν το χώρο, η πληροφορία φωτεινότητας αντιπροσωπεύεται από μια συνιστώσα Y και η πληροφορία χρώματος αποθηκεύεται σαν δύο συνιστώσες χρωματικής διαφοράς, το Cb και το Cr. Η συνιστώσα Cb είναι η διαφορά μεταξύ της μπλε συνιστώσας και μιας σταθεράς αναφοράς, ενώ η συνιστώσα Cr είναι η διαφορά μεταξύ της κόκκινης συνιστώσας και μιας σταθεράς αναφοράς. Ο μετασχηματισμός για τη μετατροπή από το χώρο RGB στον YCbCr είναι:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

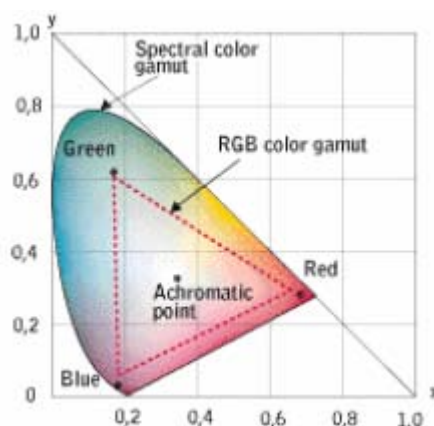
XYZ χώρος χρώματος: Το 1931 η CIE (Commission Internationale de l'Éclairage) καθόρισε τρία πρωταρχικά χρώματα X, Y, Z με βάση τα οποία μπορούν να δημιουργηθούν όλα τα ορατά χρώματα. Ο χώρος αυτός προκύπτει μέσω του παρακάτω γραμμικού μετασχηματισμού των R, G, B συνιστωσών:

$$\begin{bmatrix} 0.950456 \cdot X \\ Y \\ 1.088754 \cdot Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Τα πρωταρχικά αυτά χρώματα δεν υπάρχουν στην πραγματικότητα, είναι δηλαδή φανταστικά. Συνήθως χρησιμοποιούνται οι κανονικοποιημένες συνιστώσες:

$$(x, y, z) = \left(\frac{X}{X+Y+Z}, \frac{Y}{X+Y+Z}, \frac{Z}{X+Y+Z} \right)$$

οι οποίες διαγράφουν ένα επίπεδο, αφού $x + y + z = 1$. Παράσταση του χρώματος στο κανονικοποιημένο επίπεδο των ανεξάρτητων χρωματικών συνιστωσών (x, y) είναι ευρέως διαδεδομένη ως CIE-xy χρωματικό διάγραμμα, το οποίο φαίνεται στο παρακάτω σχήμα:



Εικόνα 4.6: CIE-xy χρωματικό διάγραμμα: Παρουσιάζεται ολόκληρη η κλίμακα των ορατών χρωμάτων. Ο RGB χώρος χρώματος περιορίζεται μέσα στο κόκκινο τρίγωνο. (από τις διαφάνειες παρουσίασης του Luke Fletcher για το μάθημα Όραση Υπολογιστών στο Australia National University)

Τα μειονεκτήματα του χώρου XYZ είναι τα ακόλουθα: είναι δύσκολο να συμπεριλάβει τη φωτεινότητα και η αντιληπτή διαφορά χρώματος δεν συσχετίζεται καλά με τη χωρική διάταξη στο χώρο χρώματος.

CMY χώρος χρώματος: Πρόκειται για το χώρο χρώματος που ορίζεται από τα συμπληρωματικά χρώματα κυανό (C), ματζέντα (M) και κίτρινο (Y). Ορίζονται ως εξής:

$$(C, M, Y) = (1 - R, 1 - G, 1 - B)$$

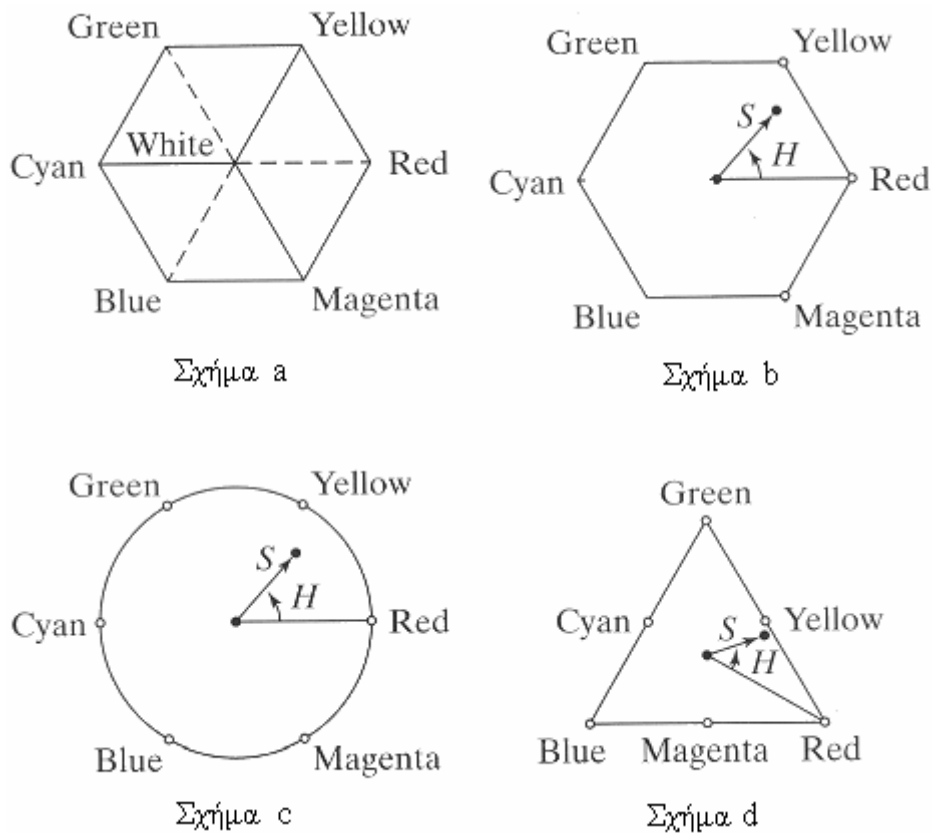
Είναι τα δευτερεύοντα χρώματα του φωτός, αλλά τα πρωταρχικά χρώματα των χρωστικών ουσιών, όπως έχει ήδη αναφερθεί.

4.2.3. Μη γραμμικοί χώροι χρώματος

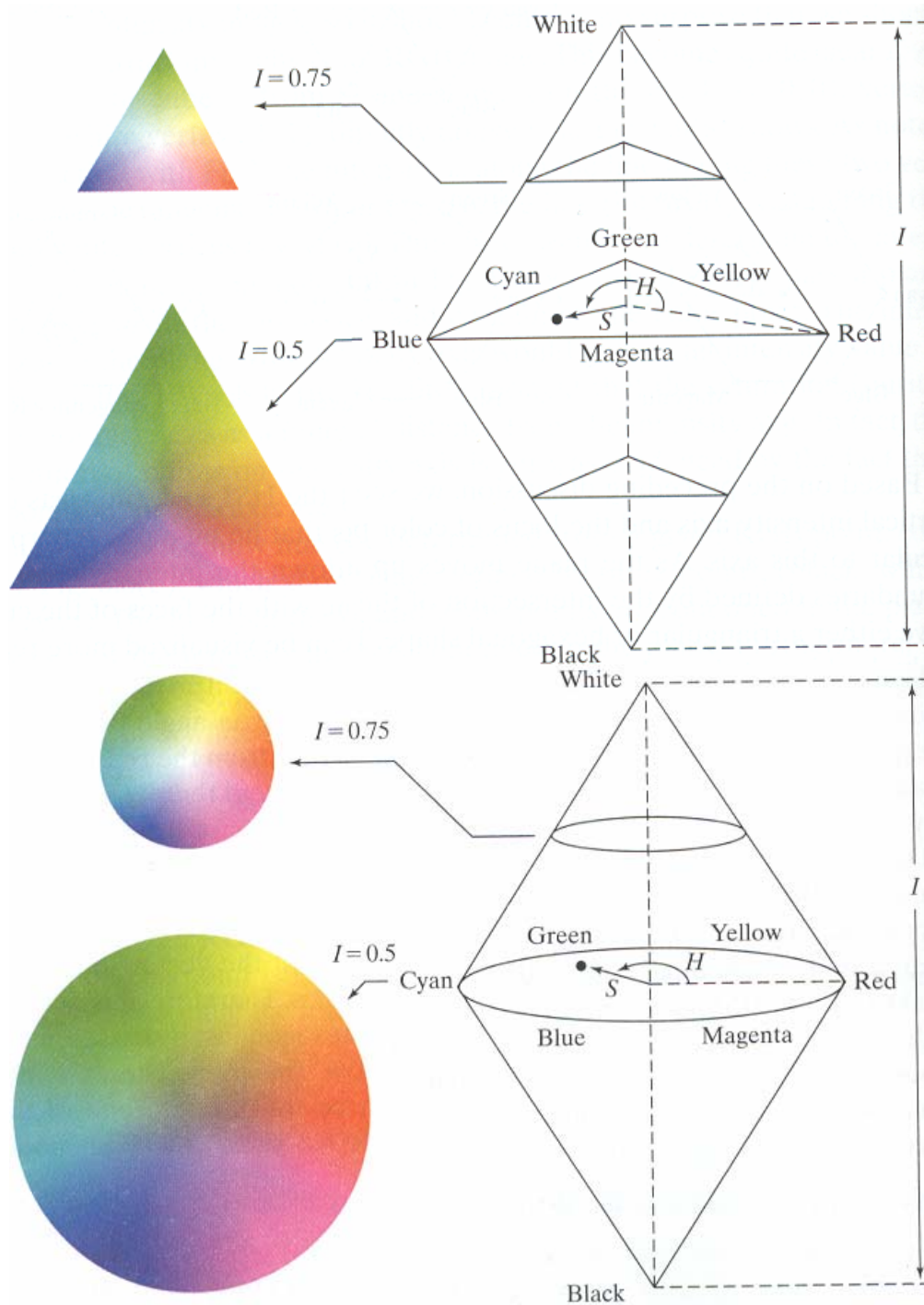
Οι γραμμικοί χώροι χρώματος έχουν κυρίως μαθηματικά πλεονεκτήματα λόγω της αλγεβρικής τους δομής. Δεν αντιστοιχούν καλά όμως στην ανθρώπινη αντίληψη χρώματος, σε αντίθεση με τους μη γραμμικούς χώρους.

HSI χώρος χρώματος: Όταν περιγράφουμε το χρώμα ενός αντικειμένου χρησιμοποιούμε την απόχρωση (hue), τον κορεσμό (saturation) και τη φωτεινότητά του (intensity) και όχι το ποσοστό καθενός από τα βασικά χρώματα που το συνθέτουν. Η απόχρωση είναι η ιδιότητα του χρώματος που μεταβάλλεται πηγαίνοντας σε διαφορετικά κύρια χρώματα του φάσματος και σχετίζεται με το μήκος κύματος του φωτός. Ο κορεσμός είναι η ιδιότητα του χρώματος που

μεταβάλλεται με το βαθμό απόχρωσης σε κάποιο χρώμα και σχετίζεται με την καθαρότητα του χρώματος. Προσθέτοντας λευκό και μονοχρωματικό φως δημιουργεί μείγματα που μπορούν να ταιριαχθούν με τα περισσότερα χρώματα. Η καθαρότητα τέτοιων χρωμάτων σχετίζεται με το ποσοστό του λευκού που χρειάζεται να προστεθεί γι' αυτό το ταιριασμα. Ένα μονοχρωματικό φως έχει 100% καθαρότητα, ενώ το λευκό φως έχει μηδέν. Η φωτεινότητα, το κύριο οπτικό ερέθισμα, είναι η ιδιότητα του χρώματος που μεταβάλλεται πηγαινόντας από το μαύρο στο λευκό και σχετίζεται με την εκπεμπόμενη φωτοβολία (luminance). Το σημαντικό πλεονέκτημα του χώρου HSI είναι ότι αποσυνδέει τη συνιστώσα φωτεινότητας από τις χρωματικές πληροφορίες (απόχρωση και κορεσμός) σε μια έγχρωμη εικόνα. Κατά συνέπεια, είναι ιδανικό εργαλείο για τους αλγόριθμους επεξεργασίας εικόνας που είναι βασισμένοι στις περιγραφές χρώματος που είναι φυσικές και διαιοθητικές στους ανθρώπους. Έτσι, οι αλγόριθμοι τελικά είναι πιο εύχρηστοι.



Εικόνα 4.7: Απόχρωση και κορεσμός στον HSI χώρο χρώματος. Η κουκκίδα είναι ένα τυχαίο χρωματικό σημείο. Η γωνία από τον κόκκινο άξονα δίνει την απόχρωση και το μήκος του διανύσματος τον κορεσμό. Η φωτεινότητα των χρωμάτων σε κάθε σχήμα δίνεται από τη θέση του σχήματος στον κάθετο άξονα φωτεινότητας. (από το "Digital Image Processing using MATLAB", R. C. Gonzalez, R. E. Woods, S. L. Eddins)



Εικόνα 4.8: Ο HSI χώρος χρώματος βασισμένος σε χρωματικά τρίγωνα και κύκλους. Τα τρίγωνα και οι κύκλοι είναι κάθετα στον κάθετο άξονα φωτεινότητας (από το "Digital Image Processing using MATLAB", R. C. Gonzalez, R. E. Woods, S. L. Eddins)

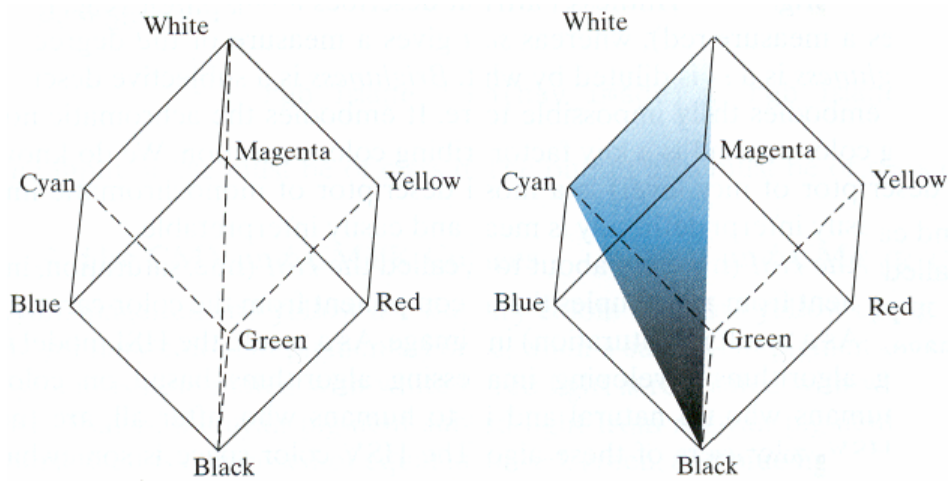
Η μετατροπή από το χώρο RGB στον HSI γίνεται με χρήση των παρακάτω εξισώσεων:

$$H = \begin{cases} \theta, & B \leq G \\ 360 - \theta, & B > G \end{cases}, \text{ όπου } \theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\left[(R-G)^2 + (R-B)(G-B) \right]^{\frac{1}{2}}} \right\}$$

$$S = 1 - \frac{3}{R+G+B} \cdot [\min(R, G, B)]$$

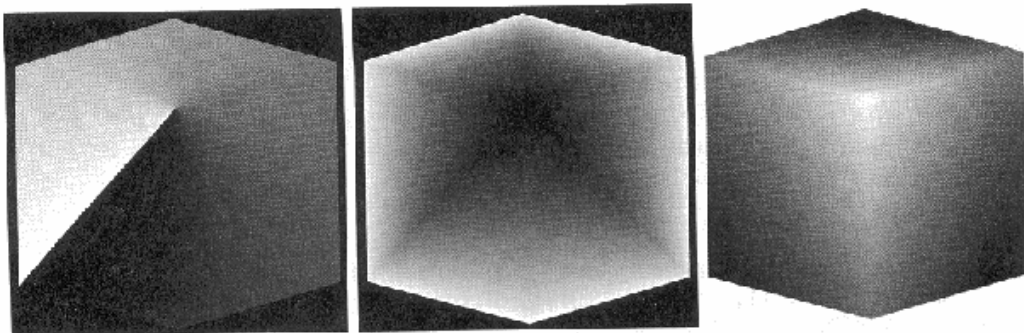
$$I = \frac{1}{3}(R+G+B)$$

Αν οι συνιστώσες (R, G, B) έχουν κανονικοποιηθεί στο διάστημα $[0,1]$, τότε και οι συνιστώσες (H, S, I) κυμαίνονται στο ίδιο διάστημα.



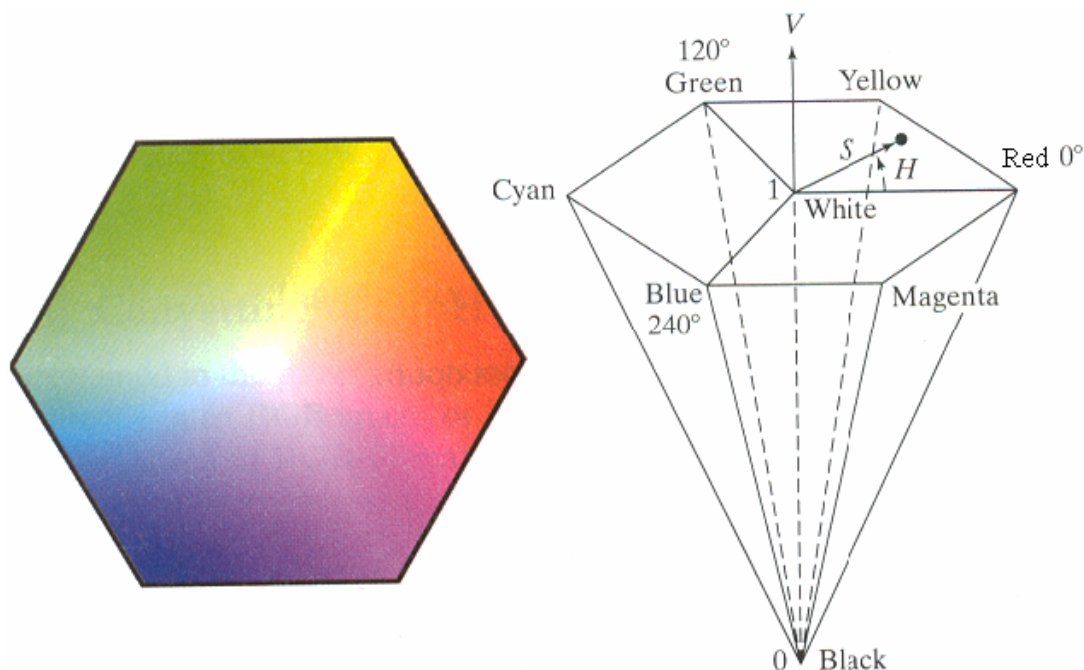
Εικόνα 4.9: Εικόνα αριστερά: Σχέση μεταξύ RGB και HSI χώρων χρώματος. Εικόνα δεξιά: HSI μοντέλο χρώματος. (από το "Digital Image Processing using MATLAB", R. C. Gonzalez, R. E. Woods, S. L. Eddins)

Η παρακάτω εικόνα απεικονίζει τις HSI συνιστώσες μιας εικόνας στον RGB κύβο χρώματος:



Εικόνα 4.10: Συνιστώσες HSI στον RGB κύβο χρώματος: απόχρωση, κορεσμός, φωτεινότητα (από το "Digital Image Processing using MATLAB", R. C. Gonzalez, R. E. Woods, S. L. Eddins)

HSV χώρος χρώματος: Ο HSV χώρος χρώματος διαμορφώνεται εξετάζοντας τον RGB κύβο χρώματος κατά μήκος του γκριζου άξονά του (ο άξονας που ενώνει τις κορυφές του μαύρου και του άσπρου), ο οποίος οδηγεί στην εξαγωνικά διαμορφωμένη παλέτα χρώματος που παρουσιάζεται στο σχήμα που ακολουθεί.



Εικόνα 4.11: Εικόνα αριστερά: HSV εξάγωνο χρωμάτων.

Εικόνα δεξιά: HSV εξαγωνικός κώνος βασικών και δευτερευόντων χρωμάτων.

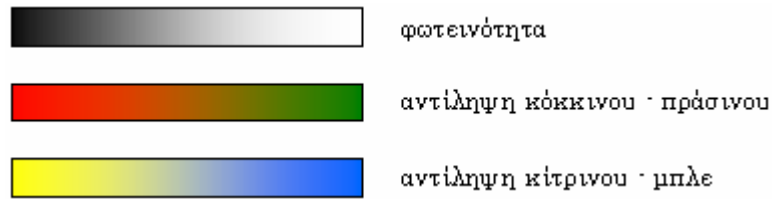
(από το "Digital Image Processing using MATLAB", R. C. Gonzalez, R. E. Woods, S. L. Eddins)

Καθώς κινούμαστε κατά μήκος του κάθετου (γκριζου) άξονα στο παρακάτω σχήμα, το μέγεθος του εξαγώνου που είναι κάθετο στον άξονα αλλάζει, δίνοντας τον όγκο που απεικονίζεται στην εικόνα. Η απόχρωση (H-hue) εκφράζεται ως η γωνία γύρω από το εξάγωνο των χρωμάτων, χρησιμοποιώντας συνήθως τον άξονα του κόκκινου ως άξονα αναφοράς 0°. Η φωτεινότητα (V-value) μετριέται κατά μήκος του άξονα του κώνου. Το τέλος $V = 0$ του άξονα είναι το μαύρο. Το τέλος $V = 1$ του άξονα είναι το άσπρο, και βρίσκεται στο κέντρο του πλήρους εξαγώνου στο σχήμα A. Κατά συνέπεια, αυτός ο άξονας αντιπροσωπεύει όλες τις αποχρώσεις του γκριζου. Ο κορεσμός (S-saturation) μετριέται ως η απόσταση από τον άξονα V.

Ο HSV χώρος χρώματος είναι βασισμένος στις κυλινδρικές συντεταγμένες. Η μετατροπή από RGB σε HSV είναι απλά ένα θέμα διαμόρφωσης των εξισώσεων ώστε να απεικονίζουν τις RGB τιμές (που είναι στις καρτεσιανές συντεταγμένες) σε κυλινδρικές συντεταγμένες.

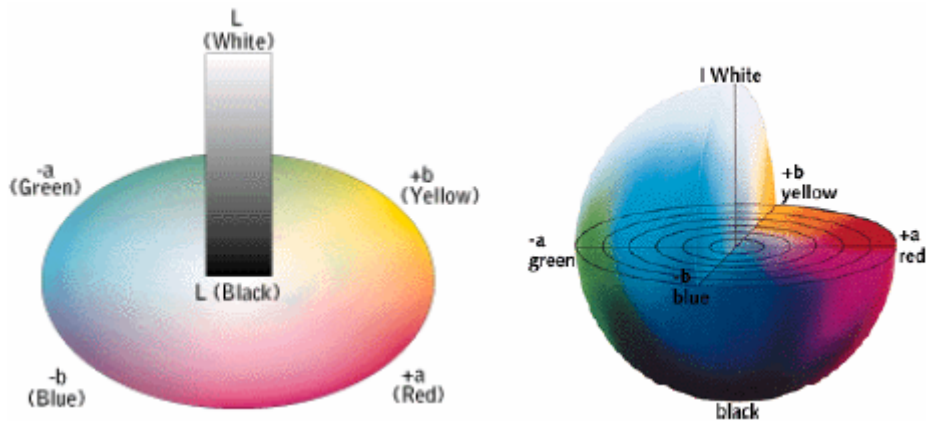
CIE-Lab χώρος χρώματος: Το χρωματικό σύστημα CIE-xy έχει το μεγάλο μειονέκτημα ότι η Ευκλείδεια απόσταση μεταξύ δύο σημείων δεν αντιστοιχεί στην ανθρώπινη αντίληψη για πιθανή διαφορά των δυο χρωμάτων. Επομένως, για σύγκριση χρωμάτων, υπάρχει η ανάγκη κατασκευής ομοιόμορφων ισοτροπικών χρωματικών χώρων στους οποίους η απόσταση μεταξύ δύο σημείων να είναι μια πιστή ένδειξη για την ανθρώπινη σημασία της διαφοράς μεταξύ των δυο χρωμάτων. Ο πιο δημοφιλής καθιερωμένος ομοιόμορφος χρωματικός χώρος είναι σήμερα ο

CIE-Lab, ο οποίος δημιουργήθηκε το 1976 και βασίζεται σε τρεις αντιλήψεις για τα χρώματα, όπως αυτές προκύπτουν από την επεξεργασία των εισόδων των φωτοαισθητήρων του ματιού. Οι αντιλήψεις αυτές είναι η φωτεινότητα, η σχέση μεταξύ κόκκινου και πράσινου και η σχέση μεταξύ κίτρινου και μπλε.



Εικόνα 4.12: Αντιλήψεις χρώματος για το χρωματικό χώρο CIE-Lab (από τις διαφάνειες παρουσίασης του Luke Fletcher για το μάθημα Όραση Υπολογιστών στο Australia National University)

Ο χώρος CIE Lab είναι βασισμένος στις διαφορές τριών στοιχειωδών ζευγών χρώματος: άσπρο - μαύρο, κόκκινο - πράσινο, κίτρινο - μπλε.



Εικόνα 4.13: Σχέση μεταξύ των ζευγών χρώματος (από τις διαφάνειες παρουσίασης του Luke Fletcher για το μάθημα Όραση Υπολογιστών στο Australia National University)

Ο μετασχηματισμός που χρησιμοποιείται για μετατροπή από τον XYZ χώρο χρώματος στον CIE-Lab είναι ο ακόλουθος:

$$L = 116 \cdot (Y/Y_w)^{1/3} - 16$$

$$a = 500 \cdot \left[(X/X_w)^{1/3} - (Y/Y_w)^{1/3} \right]$$

$$b = 200 \cdot \left[(Y/Y_w)^{1/3} - (Z/Z_w)^{1/3} \right]$$

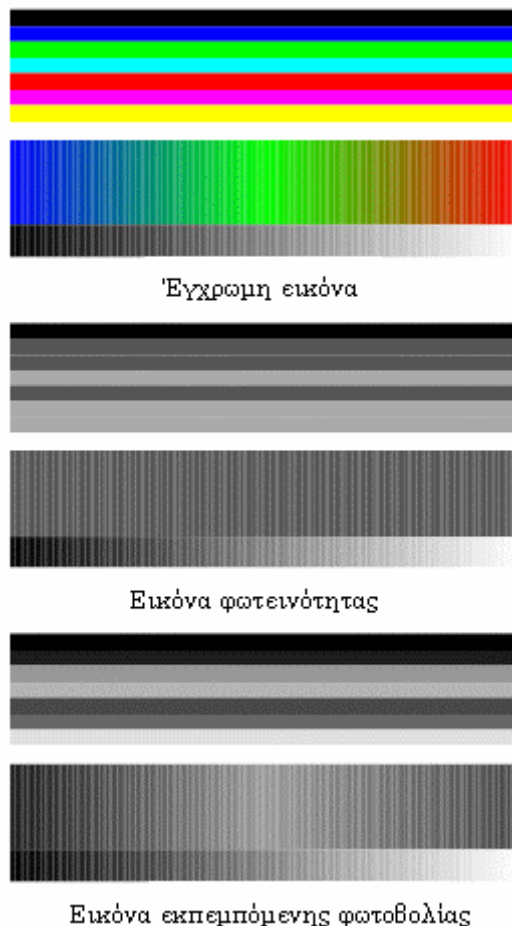
Σημειώνεται ότι για την υλοποίηση του συστήματος ανίχνευσης, παρακολούθησης και αναγνώρισης χειρονομιών της παρούσας εργασίας χρησιμοποιήθηκε ο χώρος CIE-Lab.

4.3. Χρώμα και Όραση Υπολογιστών

4.3.1. Κατάτμηση εικόνων με βάση το χρώμα

Όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, μια από τις δημοφιλέστερες μεθόδους κατάτμησης εικόνων βασίζεται στο χρώμα των αντικειμένων που μας ενδιαφέρουν. Το σημαντικότερο πρόβλημα που παρουσιάζεται είναι ότι η μέθοδος αυτή είναι πολύ ευαίσθητη όσον αφορά τις αλλαγές στο φωτισμό, αφού όσο αλλάζει ο φωτισμός μεταβάλλεται και η θέση στο χώρο χρώματος.

Η λύση για το πρόβλημα αυτό είναι η κανονικοποίηση του χρωματικού χώρου ως προς τη φωτεινότητα (intensity) ή την εκπεμπόμενη φωτοβολία (luminance). Η διαφορά μεταξύ των δύο αυτών μεγεθών εξηγείται στη συνέχεια: Η φωτεινότητα/ένταση (intensity) είναι μια φυσική ποσότητα που καθορίζεται από την ενέργεια του σήματος, ενώ η εκπεμπόμενη φωτοβολία (luminance) σχετίζεται με την αντίληψη χρώματος και είναι καθαρά ψυχολογικό φαινόμενο. Η διαφορά τους γίνεται πιο εύκολα αντιληπτή στην παρακάτω εικόνα:



Εικόνα 4.14: Διαφορά φωτεινότητας - φωτοβολίας (από τις διαφάνειες παρουσίασης του Luke Fletcher για το μάθημα Όραση Υπολογιστών στο Australia National University)

Η κανονικοποίηση ενός χρωματικού χώρου ως προς τη φωτεινότητα/φωτοβολία έχει ως αποτέλεσμα ένα χώρο χρωματικότητας (chrominance space). Παραδείγματα κανονικοποιημένων χρωματικών χώρων είναι:

- ✓ rg χώρος: προκύπτει από τις RGB συνιστώσες ως εξής:

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B}$$

- ✓ HS: οι πρώτες δύο συνιστώσες του χώρου HSV
- ✓ ab: η δεύτερη και τρίτη συνιστώσα του χώρου CIE-Lab

Η κατάτμηση της εικόνας με βάση το χρώμα γίνεται με εφαρμογή κατάλληλων κατωφλίων στις συνιστώσες της και συνδυασμό των αποτελεσμάτων. Η μέθοδος αυτή χρησιμοποιείται σε πολλά συστήματα ανίχνευσης και παρακολούθησης περιοχών ανθρώπινου δέρματος, όπως το πρόσωπο ή τα χέρια.

4.3.2. Ανίχνευση περιοχών ανθρώπινου δέρματος

Η κατάτμηση περιοχών δέρματος είναι εφικτή επειδή το ανθρώπινο δέρμα έχει μια χρωματική κατανομή που διαφέρει σημαντικά, αν και όχι εξ ολοκλήρου, από εκείνων των αντικειμένων του φόντου. Η ανίχνευση περιοχών δέρματος έχει υιοθετηθεί κυρίως στην αναγνώριση και παρακολούθηση προσώπων σε εικόνες και βίντεο. Η κατάτμηση εκτελείται συνήθως χρησιμοποιώντας τις χρωματικές συνιστώσες της εικόνας και όχι τη συνιστώσα φωτεινότητας. Υπάρχουν δύο λόγοι που δικαιολογούν το γεγονός αυτό: 1) με χρήση μόνο των χρωματικών συνιστωσών, οι αλγόριθμοι κατάτμησης παραμένουν σχετικά ανεπηρέαστοι στις αλλαγές φωτεινότητας, και 2) έχει αναφερθεί ευρέως ότι οι εμφανείς διαφορές στο χρώμα του δέρματος μεταξύ διαφορετικών φυλών χαρακτηρίζονται από τη διαφορά στη φωτεινότητα του χρώματος, το οποίο κυριαρχείται από τη συνιστώσα φωτεινότητας και όχι από τις χρωματικές συνιστώσες. Ένας άλλος λόγος είναι ότι χρησιμοποιώντας μόνο τις χρωματικές συνιστώσες, ο χώρος των χαρακτηριστικών γνωρισμάτων μειώνεται από τρισδιάστατος σε διδιάστατος, μειώνοντας κατά συνέπεια την υπολογιστική πολυπλοκότητα του αλγορίθμου κατάτμησης.

Υπάρχουν ορισμένοι περιορισμοί που πρέπει να ληφθούν υπόψη στους αλγορίθμους κατάτμησης δέρματος. Ακριβή αποτελέσματα επιτυγχάνονται μόνο εάν υπάρχει σημαντική αντίθεση μεταξύ του δέρματος και του φόντου. Φυσικά, στα πλαίσια της κατάτμησης χεριών και προσώπου, άλλα μέρη του σώματος, συμπεριλαμβανομένου του ρουχισμού, θεωρούνται επίσης ως φόντο. Οι στατικές περιοχές του φόντου με παρόμοιο χρώμα με αυτό του δέρματος δεν δημιουργούν σοβαρό πρόβλημα, δεδομένου ότι μπορούν να προσδιοριστούν μέσω της ανίχνευσης κίνησης. Εντούτοις, τα μέρη του ρουχισμού που έχουν παρόμοιο χρώμα με αυτό του δέρματος και κινούνται, μπορούν να δημιουργήσουν προβλήματα. Υπάρχουν επίσης και οι περιορισμοί που σχετίζονται με το φωτισμό κατά τη διάρκεια καταγραφής της εικόνας εισόδου.

Είναι σημαντικό να επιλεγεί ο κατάλληλος χώρος χρώματος για την εκτέλεση της κατάτμησης περιοχών δέρματος. Οι χώροι χρώματος που έχουν χρησιμοποιηθεί σε υπάρχοντα συστήματα περιλαμβάνουν τους χώρους YCbCr, HSV, CIE-Lab,

κανονικοποιημένο RGB (rg), και RGB. Σημειώνουμε ότι στο χώρο RGB, οι χρωματικές συνιστώσες και η συνιστώσα φωτεινότητας δεν αποσυνδέονται. Επίσης, ο χώρος χαρακτηριστικών γνωρισμάτων στον RGB χώρο είναι τριδιάστατος. Έτσι, δεν οδηγούμαστε σε πολύ ακριβή αποτελέσματα.

Μια απλή μέθοδος κατάτμησης περιοχών δέρματος βασίζεται στη συλλογή δειγμάτων pixels δέρματος από το στόχο για την εκπαίδευση ενός ταξινομητή χρώματος. Έτσι, η κατάτμηση περιοχών που έχουν χρώμα δέρματος πραγματοποιείται με την ταξινόμηση και την ομαδοποίηση των pixels της εικόνας εισόδου. Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, μια από τις λύσεις για να αντιμετωπιστεί η δυσκολία που παρουσιάζεται λόγω της ύπαρξης μεγάλης ποικιλίας στο χρωματικό τόνο του δέρματος μεταξύ διαφορετικών ανθρώπων είναι να ρυθμιστεί ο ταξινομητής χρώματος χρησιμοποιώντας ένα εκτεταμένο σύνολο εκπαίδευσης που περιέχει δείγματα από πολλούς ανθρώπους, υπό διάφορες συνθήκες φωτισμού.

4.3.3. Παρακολούθηση αντικειμένων με βάση το χρώμα τους

Εφόσον είναι δυνατή η ανίχνευση των αντικειμένων σε μια εικόνα με βάση το χρώμα τους είναι αναμενόμενο ότι θα είναι δυνατή και η παρακολούθησή τους στο χώρο και στο χρόνο χρησιμοποιώντας αυτό το χαρακτηριστικό. Τα πλεονεκτήματα που προκύπτουν από τη χρήση του χρώματος σε συστήματα παρακολούθησης είναι ότι είναι στοιχείο ανεξάρτητο από τον προσανατολισμό του αντικειμένου και μια από τις ταχύτερες μεθόδους εντοπισμού περιοχών όπως το πρόσωπο και τα χέρια.

Η παρακολούθηση του επιθυμητού αντικειμένου μπορεί να γίνει είτε μόνο με βάση το χρώμα, είτε σε συνδυασμό με κάποιο άλλο χαρακτηριστικό όπως για παράδειγμα η κίνηση, είτε τέλος με χρήση του χρώματος ως παράμετρο κάποιου αλγορίθμου παρακολούθησης αντικειμένου, συνήθως για να περιοριστεί η περιοχή αναζήτησης του αντικειμένου ενδιαφέροντος. Μια από τις μεθόδους που χρησιμοποιούν το χρώμα ως παράμετρο είναι και αυτή του particle filtering, που χρησιμοποιήθηκε στην παρούσα εργασία. Σε κάθε περίπτωση πάντως είναι απαραίτητη η δημιουργία ενός μοντέλου χρώματος για να είναι δυνατή η ταξινόμηση των pixels εισόδου.

4.3.4. Κατασκευή μοντέλου χρώματος

Στη συνέχεια εξηγείται συνοπτικά η διαδικασία δημιουργίας ενός μοντέλου χρώματος, όπως αυτό που χρησιμοποιήθηκε στην υλοποίηση του συστήματος παρακολούθησης και αναγνώρισης χειρονομιών της διπλωματικής εργασίας. Πρώτο βήμα είναι η συλλογή ικανού αριθμού δειγμάτων pixels του χρώματος που μας ενδιαφέρει, από διάφορες εικόνες (σε μορφή RGB). Οι RGB συνιστώσες μετατρέπονται σε χρωματικές συνιστώσες, αντιστοιχίζονται δηλαδή στο διδιάστατο χώρο χρωματικότητας. Τέλος επιλέγεται ένα μοντέλο για την περιγραφή της κατανομής των δειγμάτων στο 2Δ χώρο χρωματικότητας.

Για παράδειγμα, για τη δημιουργία ενός μοντέλου για ανίχνευση περιοχών δέρματος ακολουθούμε τα παρακάτω βήματα:

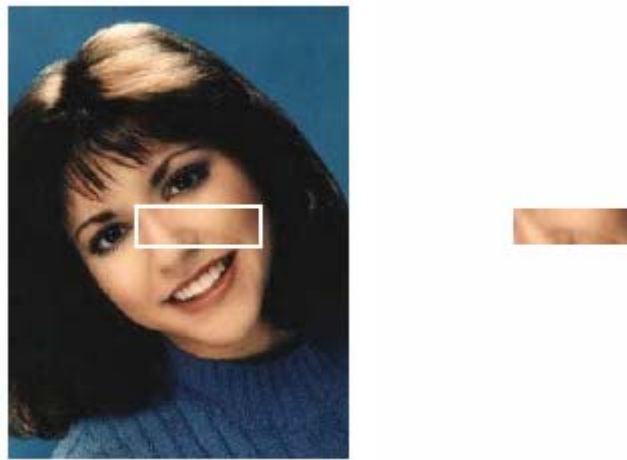
- ✓ Σε μια εικόνα που περιέχει την περιοχή που μας ενδιαφέρει απομονώνουμε ένα τμήμα της περιοχής αυτής.
- ✓ Μετατρέπουμε τα pixels της περιοχής που απομονώσαμε από τον RGB χώρο χρώματος σε ένα χώρο που διαχωρίζει τις χρωματικές συνιστώσες, έστω τον CIE-Lab.
- ✓ Τοποθετούμε κάθε ζεύγος τιμών των χρωματικών συνιστωσών (a, b) στο χώρο χρωματικότητας, χρησιμοποιώντας έναν πίνακα-συσσωρευτή όπου αποθηκεύεται στην αντίστοιχη θέση του ο αριθμός των εμφανίσεων ενός ζεύγους (a, b) για όλα τα pixels της περιοχής που απομονώσαμε.
- ✓ Τελικά, όλα τα σημεία της γραφικής παράστασης συγκεντρώνονται σε μια μικρή περιοχή.

Οι [Cai et al., 1998] χρησιμοποίησαν 2300 δείγματα δέρματος από 80 εικόνες και ανακάλυψαν ότι όλα βρίσκονται σε μια μικρή περιοχή: $a \in [-10, 60]$ και $b \in [-10, 40]$. Αν θεωρήσουμε ότι γενικά οι τιμές των a και b κυμαίνονται στα παραπάνω διαστήματα για τις περιοχές δέρματος, μπορούμε να τοποθετήσουμε τον συσσωρευτή σε αυτό το διάστημα το χώρου χρωματικότητας, και να καταλήξουμε έτσι σε πολύ καλύτερη ανάλυση του μοντέλου. Έτσι, μπορούμε εύκολα να καταλήξουμε στην πιθανότητα κάθε pixel της εισόδου να είναι δέρμα αν ομαλοποιήσουμε τα δεδομένα του συσσωρευτή μέσω συνέλιξης με μια Gaussian συνάρτηση και κανονικοποίησης του αποτελέσματος στο διάστημα $[0, 1]$. Τα παραπάνω βήματα παρουσιάζονται στα σχήματα που ακολουθούν.

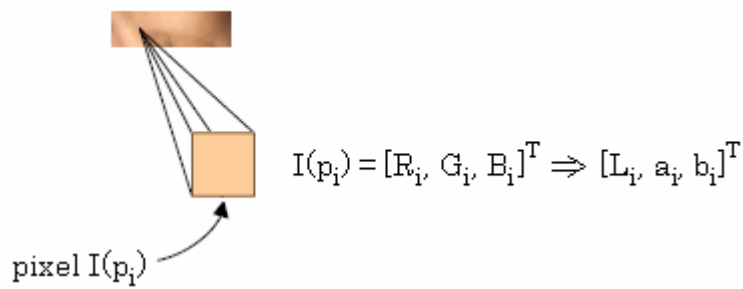


Εικόνα 4.15: Εικόνα εισόδου, που περιέχει την περιοχή ενδιαφέροντος²

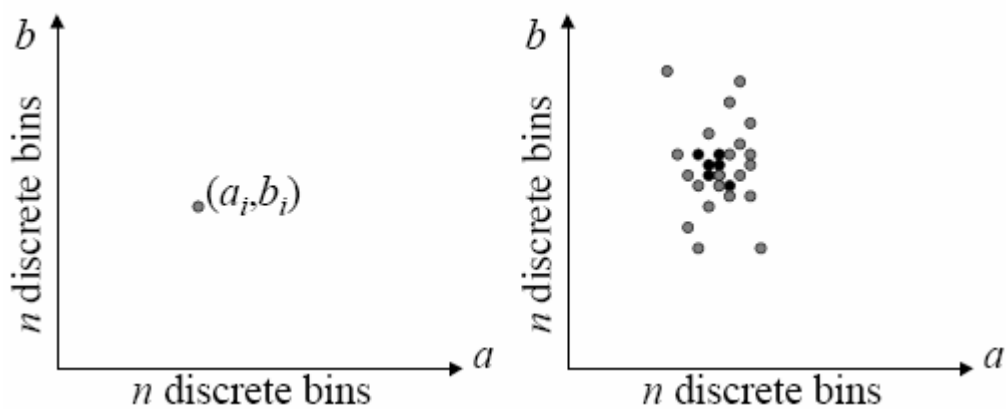
² Οι εικόνες 4.15 έως 4.25 έχουν ληφθεί από τις διαφάνειες παρουσίασης του Luke Fletcher για το μάθημα Όραση Υπολογιστών στο Australia National University.



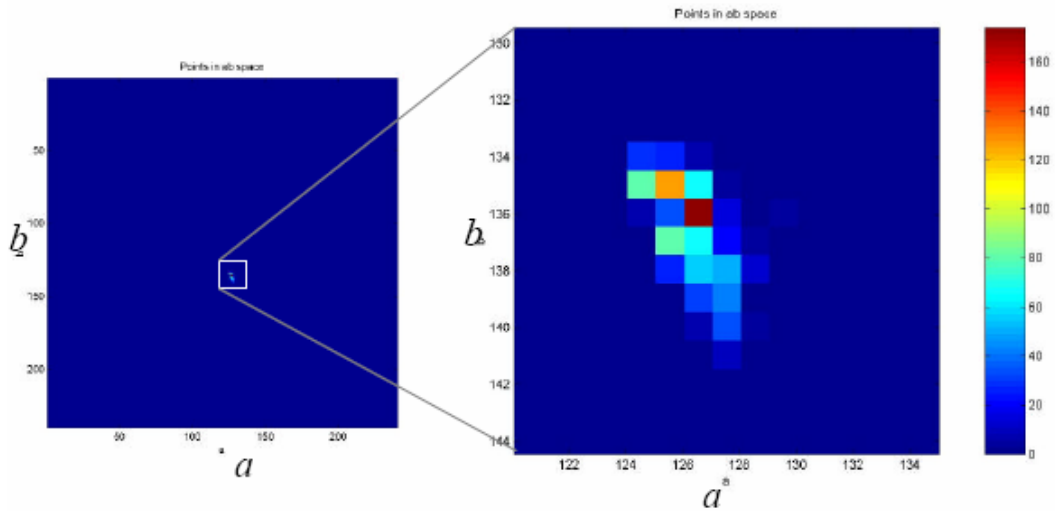
Εικόνα 4.16: Απομόνωση περιοχής ενδιαφέροντος I



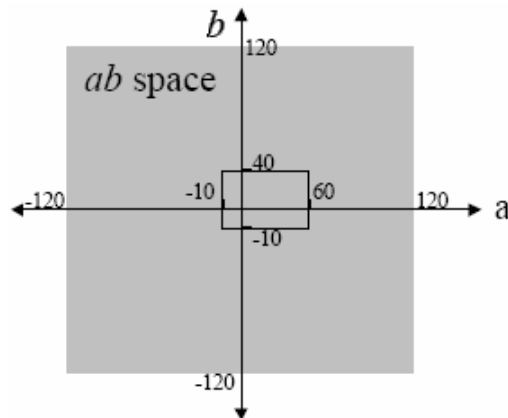
Εικόνα 4.17: Μετατροπή από RGB σε Lab



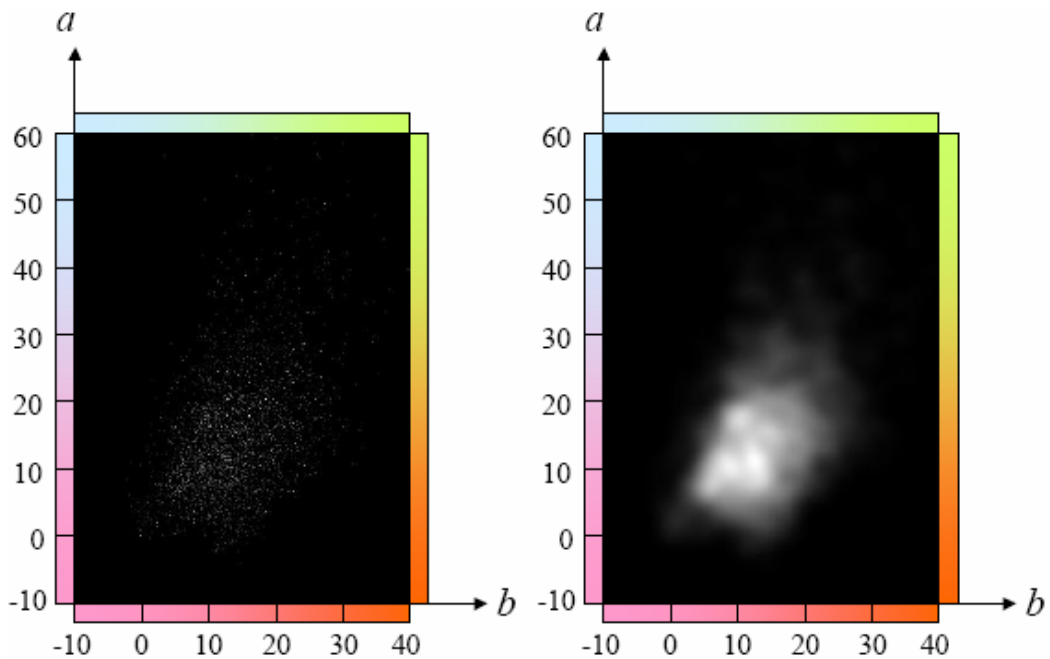
Εικόνα 4.18: Γραφική απεικόνιση των ζευγών (a_i, b_i) για το pixel p_i και για όλα τα pixels



Εικόνα 4.19: Γραφική απεικόνιση για $n = 240$



Εικόνα 4.20: Τοποθέτηση συσσωρευτή στις περιοχές των a, b στο χώρο ab

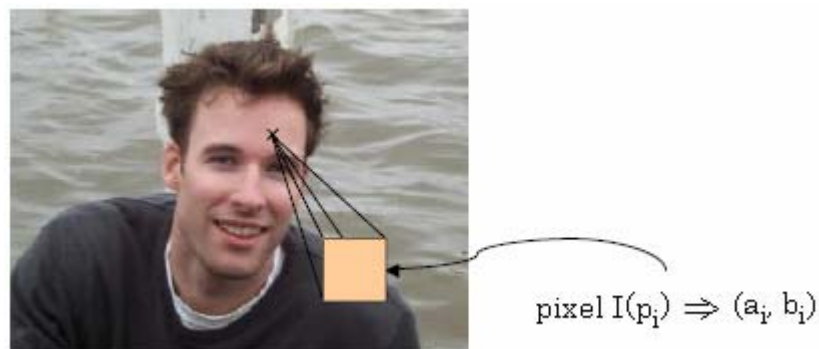


Εικόνα 4.21: Το προκύπτον μοντέλο χρώματος πριν και μετά τη συνέλιξη με τη Gaussian. Κάθε σημείο της δεύτερης γραφικής παράστασης δίνει την πιθανότητα το αντίστοιχο ζεύγος τιμών (a, b) να είναι δέρμα.

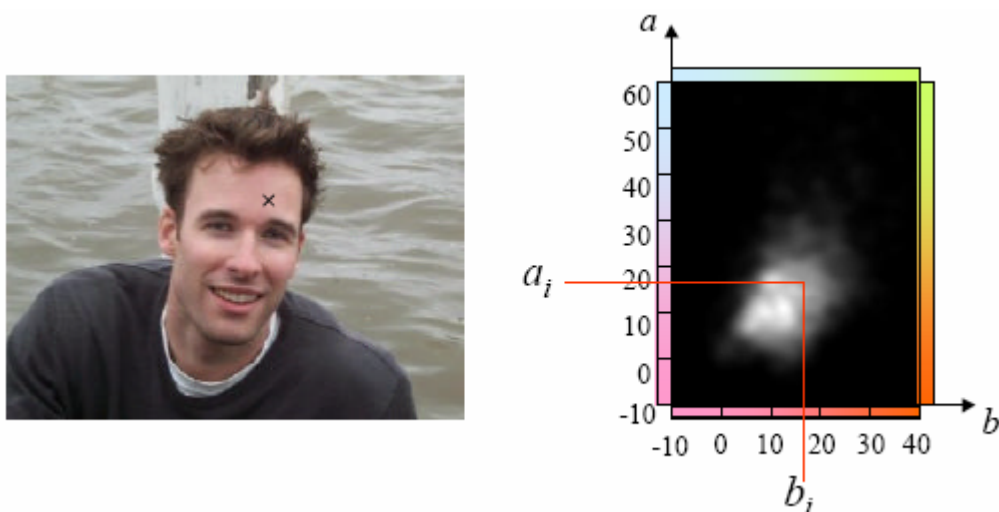
Έχοντας καθορίσει το μοντέλο χρώματος, μπορούμε εύκολα να το χρησιμοποιήσουμε για την ανίχνευση περιοχών δέρματος. Σε μια τυχαία εικόνα, καθορίζουμε τις τιμές (a_i, b_i) για κάθε pixel p_i . Σε κάθε ζεύγος αντιστοιχούμε την πιθανότητα το ζεύγος αυτό να είναι pixel δέρματος σύμφωνα με το μοντέλο που κατασκευάστηκε προηγουμένως. Με την εφαρμογή του κατάλληλου κατωφλίου μπορούμε να απομονώσουμε στη συνέχεια τις περιοχές δέρματος. Στις εικόνες που ακολουθούν φαίνεται ένα παράδειγμα εφαρμογής.



Εικόνα 4.22: Εικόνα εισόδου



Εικόνα 4.23: Μετατροπή από RGB σε Lab



Εικόνα 4.24: Αντιστοίχιση του κάθε ζεύγους (a_i, b_i) σε πιθανότητα μέσω του μοντέλου

πιθανότητα του pixel p_i



Εικόνα 4.25: Εικόνα εισόδου και εικόνα πιθανοτήτων

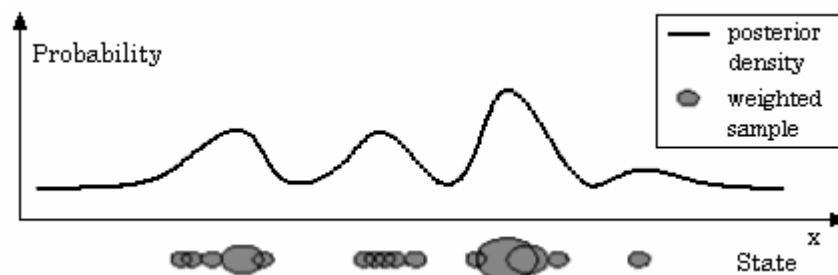
5

Παρακολούθηση Κίνησης με Particle Filtering

5.1. Εισαγωγή

Η αναδρομική υλοποίηση της στατιστικής επεξεργασίας σήματος που βασίζεται σε μεθόδους Monte Carlo είναι γνωστή ως particle filtering και αποτελεί το αντικείμενο με το οποίο ασχολείται αυτό το κεφάλαιο. Πρόκειται για στατιστική μέθοδο που βασίζεται στον κανόνα του Bayes. Η έρευνα γύρω από τη μέθοδο αυτή έχει προχωρήσει αρκετά τα τελευταία χρόνια. Το particle filtering πιθανόν να είναι μια αξιόλογη εναλλακτική λύση για τις εφαρμογές πραγματικού χρόνου που παραδοσιακά προσεγγίζονται από βασισμένες σε μοντέλα τεχνικές Kalman φίλτρων, όπου απαιτούνται γραμμικά μοντέλα συστήματος και θόρυβος Gaussian κατανομής. Το κύριο πλεονέκτημα του particle filtering είναι η δυνατότητα χειρισμού οποιασδήποτε μη γραμμικότητας και οποιασδήποτε κατανομής θορύβου. Όσο πιο μη γραμμικό είναι το μοντέλο, ή όσο πιο μη-Gaussian είναι ο θόρυβος, τόσο μεγαλύτερες δυνατότητες έχει το particle filtering. Επίσης, είναι σε θέση να αντιμετωπίσει σύνθετα περιβάλλοντα και είναι πολύ εύκολο να υλοποιηθεί.

Η βασική ιδέα της μεθόδου είναι η χρήση δειγμάτων (particles) για την αναπαράσταση του μοντέλου του συστήματος αντί για χρήση Gaussian κατανομών ή οποιουδήποτε άλλου μοντέλου και η διάδοση μόνο των επικρατέστερων δειγμάτων.



Εικόνα 5.1: Βασική ιδέα του particle filtering (από τις διαφάνειες παρουσίασης του Luke Fletcher για το μάθημα Όραση Υπολογιστών στο Australia National University)

Η μέθοδος του particle filtering έχει χρησιμοποιηθεί σε ποικίλες εφαρμογές. Οι περιοχές εφαρμογών είναι:

- ✓ Εκτίμηση θέσης (Positioning), όπου το πρόβλημα είναι ο προσδιορισμός της θέσης κάποιου αντικειμένου. Πρόκειται μάλλον για ένα πρόβλημα φιλτραρίσματος παρά ένα στατικό πρόβλημα εκτίμησης, όταν χρησιμοποιείται ένα σύστημα πλοήγησης για να παρέχει τις μετρήσεις της κίνησης.
- ✓ Πλοήγηση (Navigation), όπου, εκτός από τη θέση, συμπεριλαμβάνονται στο πρόβλημα η γραμμική και γωνιακή ταχύτητα, η στάση και η κατεύθυνση, καθώς και η επιτάχυνση.
- ✓ Παρακολούθηση στόχου (Target Tracking), όπου το πρόβλημα είναι η εκτίμηση της θέσης ενός αντικειμένου βάσει μετρήσεων απόστασης και γωνιών από μια θέση αναφοράς.

Μια άλλη σχετική εφαρμογή είναι ο εντοπισμός θέσης κινούμενου ρομπότ (mobile robot localization). Τα παραπάνω προβλήματα έχουν κοινά σημεία, δεδομένου ότι μπορούν να περιγραφούν με παρόμοια μοντέλα του χώρου κατάστασης (state space models), όπου το διάνυσμα κατάστασης περιέχει τη θέση και τις παραγώγους της θέσης. Οι τυπικές μέθοδοι είναι βασισμένες σε γραμμικοποιημένα μοντέλα και Gaussian προσεγγίσεις θορύβου έτσι ώστε να μπορεί να εφαρμοστεί το φίλτρο Kalman. Η έρευνα στρέφεται στον τρόπο με τον οποίο μπορούν να χρησιμοποιηθούν διαφορετικές συντεταγμένες κατάστασης ή πολλαπλά μοντέλα για τον περιορισμό των προσεγγίσεων. Αντίθετα, το particle filtering προσεγγίζει τη βέλτιστη λύση αριθμητικά βάσει ενός φυσικού μοντέλου, αντί να εφαρμόζει ένα βέλτιστο φίλτρο σε ένα προσεγγιστικό μοντέλο. Ένα γνωστό πρόβλημα του particle filtering είναι ότι η απόδοσή του μειώνεται γρήγορα όταν αυξάνεται η διάσταση του χώρου κατάστασης.

5.2. Χρήσιμες έννοιες

5.2.1. Μοντέλο Πιθανότητας

Η πιθανότητα είναι ο κατάλληλος μηχανισμός για εκτίμηση της αβεβαιότητας. Τα μοντέλα πιθανοτήτων συγκρίνουν τα αποτελέσματα διάφορων πειραμάτων, που ονομάζονται ενδεχόμενα. Αυτά τα αποτελέσματα αντιπροσωπεύονται από το σύνολο όλων των δυνατών υποσυνόλων κάποιου διαστήματος, το οποίο πρέπει να έχει ορισμένες ιδιότητες. Αφού οριστεί ένα τέτοιο σύνολο μπορεί στη συνέχεια να οριστεί μια συνάρτηση πιθανότητας. Υπάρχουν διάφορες μέθοδοι για τον τρόπο επιλογής μιας συνάρτησης πιθανότητας για μια συγκεκριμένη εφαρμογή.

Αν μπορούμε να πούμε αν ένα ενδεχόμενο πραγματοποιήθηκε, τότε μπορούμε να πούμε και αν δεν πραγματοποιήθηκε. Επίσης αν μπορούμε να πούμε αν δύο ενδεχόμενα πραγματοποιήθηκαν ανεξάρτητα, τότε μπορούμε να πούμε αν πραγματοποιήθηκαν ταυτόχρονα. Θεωρούμε λοιπόν ένα διακριτό διάστημα D , το οποίο μπορεί να είναι άπειρο και αντιπροσωπεύει το περιβάλλον διεξαγωγής των πειραμάτων. Κατασκευάζουμε στη συνέχεια ένα σύνολο F από τα υποσύνολα του D ,

καθένα από τα οποία αντιπροσωπεύει ένα ενδεχόμενο. Το σύνολο αυτό πρέπει να έχει τις ακόλουθες ιδιότητες:

- ✓ Στο F περιλαμβάνονται το κενό σύνολο και το σύνολο D .
- ✓ Κλείσιμο υπό συμπληρώματα (Closure under complements): αν $S_1 \in F$ τότε $\overline{S_1} = D - S_1 \in F$, δηλαδή αν μπορούμε να πούμε αν ένα ενδεχόμενο πραγματοποιήθηκε, τότε μπορούμε να πούμε και αν δεν πραγματοποιήθηκε.
- ✓ Κλείσιμο υπό τομή (Closure under intersection): Αν $S_1 \in F$ και $S_2 \in F$, τότε $S_1 \cap S_2 \in F$, δηλαδή αν μπορούμε να πούμε αν δύο ενδεχόμενα πραγματοποιήθηκαν ανεξάρτητα, τότε μπορούμε να πούμε αν πραγματοποιήθηκαν ταυτόχρονα.

Στη συνέχεια κατασκευάζουμε μια συνάρτηση P , η οποία περνάει τα στοιχεία του F στο μοναδιαίο διάστημα. Απαιτείται η P να έχει τις ακόλουθες ιδιότητες:

- ✓ Η P ορίζεται για κάθε στοιχείο του F
- ✓ $P(\emptyset) = 0$
- ✓ $P(D) = 1$
- ✓ Για $A \in F$ και $B \in F$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

οι οποίες ονομάζονται αξιωματικές ιδιότητες της πιθανότητας. Σημειώνουμε ότι $0 \leq P(A) \leq 1$ για όλα τα $A \in F$, επειδή η συνάρτηση περνάει τα στοιχεία του F στο μοναδιαίο διάστημα. Το σύνολο των D, P και F αποτελεί ένα μοντέλο πιθανότητας. Η $P(A)$ ονομάζεται πιθανότητα του γεγονότος A .

5.2.2. Κανόνας του Bayes, εκ των προτέρων και εκ των υστέρων πιθανότητες

Η απλή παρατήρηση ότι $P(A, B) = P(A|B)P(B)$ οδηγεί σε μια ταυτότητα για την αναστροφή της σειράς σε μια υπό συνθήκη πιθανότητα:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

που αποτελεί τον κανόνα του Bayes.

Η ενδιαφέρουσα ιδιότητα του κανόνα του Bayes είναι ότι κρίνει ποια επιλογή παραμέτρων είναι καταλληλότερη, δεδομένου του μοντέλου και των εκ των προτέρων γνώσεών μας. Ξαναγράφοντας τον κανόνα του Bayes προκύπτει:

$$P(\text{parameters} | \text{data}) = \frac{P(\text{data} | \text{parameters})P(\text{parameters})}{P(\text{data})}$$

Ο όρος $P(\text{parameters})$ ονομάζεται εκ των προτέρων πιθανότητα (prior) και περιγράφει τη γνώση του περιβάλλοντος πριν τη λήψη μετρήσεων. Ο όρος

$P(\text{parameters} | \text{data})$ ονομάζεται εκ των υστέρων πιθανότητα (posterior) και περιγράφει την πιθανότητα διάφορων μοντέλων μετά τη λήψη μετρήσεων. Η φιλοσοφία του κανόνα Bayes είναι ότι ολόκληρη η πληροφορία για τον κόσμο περιέχεται στην posterior.

5.3. Μη γραμμική Bayesian παρακολούθηση στόχου

5.3.1. Εισαγωγή

Πολλά προβλήματα απαιτούν την εκτίμηση της κατάστασης ενός συστήματος που αλλάζει με την πάροδο του χρόνου, χρησιμοποιώντας μια ακολουθία θορυβωδών μετρήσεων που πραγματοποιούνται στο σύστημα. Επικεντρωνόμαστε στην προσέγγιση του χώρου κατάστασης στη μοντελοποίηση δυναμικών συστημάτων, και ιδιαίτερα στη διατύπωση του προβλήματος στο διακριτό χρόνο. Κατά συνέπεια, χρησιμοποιούνται εξισώσεις διαφορών για να μοντελοποιήσουν την εξέλιξη του συστήματος με το χρόνο, και οι μετρήσεις υποτίθεται ότι είναι διαθέσιμες σε διακριτές χρονικές στιγμές. Για την εκτίμηση δυναμικής κατάστασης, είναι κατάλληλη και ευρέως διαδεδομένη η προσέγγιση διακριτού χρόνου. Δίνεται προσοχή στο διάνυσμα κατάστασης ενός συστήματος. Το διάνυσμα κατάστασης περιέχει όλες τις σχετικές πληροφορίες που απαιτούνται για να περιγράψουν το υπό εξέταση σύστημα. Παραδείγματος χάριν, σε προβλήματα παρακολούθησης, αυτές οι πληροφορίες θα μπορούσαν να αφορούν τα κινηματικά χαρακτηριστικά του στόχου. Το διάνυσμα μετρήσεων αντιπροσωπεύει τις (θορυβώδεις) παρατηρήσεις που συσχετίζονται με το διάνυσμα κατάστασης και είναι γενικά (αλλά όχι απαραίτητα) μικρότερης διάστασης από το διάνυσμα κατάστασης. Η προσέγγιση χώρου κατάστασης είναι κατάλληλη για το χειρισμό των δεδομένων πολλών μεταβλητών και των μη γραμμικών/μη Gaussian ανεξίτηλων, και παρέχει ένα σημαντικό πλεονέκτημα έναντι των παραδοσιακών time-series τεχνικών για αυτά τα προβλήματα.

Προκειμένου να αναλυθεί ένα δυναμικό σύστημα και να εξαχθούν συμπεράσματα για αυτό, απαιτούνται τουλάχιστον δύο μοντέλα: Κατ' αρχάς, ένα μοντέλο που περιγράφει την εξέλιξη της κατάστασης με το χρόνο (το μοντέλο συστήματος) και, δεύτερον, ένα μοντέλο που σχετίζει τις θορυβώδεις μετρήσεις με την κατάσταση (το μοντέλο μετρήσεων). Θα υποθέσουμε ότι αυτά τα μοντέλα είναι διαθέσιμα σε μια πιθανολογική μορφή. Η πιθανολογική διατύπωση του χώρου κατάστασης και η απαίτηση για την ενημέρωση των πληροφοριών μετά τη λήψη των νέων μετρήσεων είναι απόλυτα κατάλληλες για την Bayesian προσέγγιση. Αυτό παρέχει ένα αυστηρό γενικό πλαίσιο για τα δυναμικά προβλήματα εκτίμησης κατάστασης.

Στη Bayesian προσέγγιση στη δυναμική εκτίμηση κατάστασης, γίνεται προσπάθεια για κατασκευή της εκ των υστέρων συνάρτησης πυκνότητας πιθανότητας (posterior probability density function - pdf) της κατάστασης, που βασίζεται σε όλες τις διαθέσιμες πληροφορίες, συμπεριλαμβανομένου του συνόλου των λαμβανόμενων μετρήσεων. Δεδομένου ότι η pdf ενσωματώνει όλες τις διαθέσιμες στατιστικές πληροφορίες, μπορεί να θεωρηθεί ως η πλήρης λύση στο πρόβλημα εκτίμησης. Γενικά, από την pdf μπορεί να ληφθεί μια βέλτιστη (ως προς

οποιοδήποτε κριτήριο) εκτίμηση της κατάστασης. Μπορεί επίσης να ληφθεί ένα μέτρο της ακρίβειας της εκτίμησης. Για πολλά προβλήματα απαιτείται μια εκτίμηση κάθε φορά που παραλαμβάνεται μια μέτρηση. Σε αυτήν την περίπτωση, μια κατάλληλη λύση είναι ένα αναδρομικό φίλτρο. Μια προσέγγιση αναδρομικού φιλτραρίσματος σημαίνει ότι τα λαμβανόμενα στοιχεία μπορούν να υποβληθούν σε επεξεργασία ακολουθιακά παρά ως σύνολο, ώστε να μην είναι απαραίτητο να αποθηκευτεί το σύνολο όλων των στοιχείων ούτε να επανεπεξεργάζονται τα υπάρχοντα στοιχεία όταν γίνεται διαθέσιμη μια νέα μέτρηση. Ένα τέτοιο φίλτρο αποτελείται από δύο στάδια: πρόβλεψη και ενημέρωση. Το στάδιο πρόβλεψης χρησιμοποιεί το μοντέλο του συστήματος για να προβλέψει την pdf της κατάστασης από έναν χρόνο μέτρησης στον επόμενο. Δεδομένου ότι η κατάσταση υπόκειται συνήθως σε άγνωστες διαταραχές (που μοντελοποιούνται ως τυχαίος θόρυβος), η πρόβλεψη γενικά μετασχηματίζει, παραμορφώνει, και διαδίδει την pdf της κατάστασης. Η λειτουργία της ενημέρωσης χρησιμοποιεί την πιο πρόσφατη μέτρηση για να τροποποιήσει την πρόβλεψη της pdf. Αυτό επιτυγχάνεται χρησιμοποιώντας το θεώρημα Bayes, το οποίο είναι ο μηχανισμός για την ανανέωση της γνώσης για την κατάσταση του στόχου λαμβάνοντας υπόψη τις πρόσθετες πληροφορίες από τα νέα στοιχεία.

5.3.2. Καθορισμός προβλήματος και εύρεση λύσης

Για τον καθορισμό του προβλήματος παρακολούθησης, θεωρούμε ότι η εξέλιξη της ακολουθίας κατάστασης ενός στόχου δίνεται από:

$$x_k = f_k(x_{k-1}, v_{k-1}) \quad (1)$$

όπου η $f_k : \mathcal{R}^{n_x} \times \mathcal{R}^{n_v} \rightarrow \mathcal{R}^{n_x}$ είναι μια πιθανώς μη γραμμική συνάρτηση της κατάστασης x_{k-1} , η $\{v_{k-1}, k \in \mathbb{N}\}$ είναι μια ανεξάρτητη, ομοιόμορφα κατανεμημένη ανέλιξη (independent, identically distributed (i.i.d.) process) ακολουθία θορύβου, οι n_x και n_v είναι οι διαστάσεις των διανυσμάτων κατάστασης και της διαδικασίας θορύβου αντίστοιχα, και \mathbb{N} είναι το σύνολο των φυσικών αριθμών.

Ο στόχος της παρακολούθησης είναι να βρεθεί αναδρομικά μια εκτίμηση για το x_k από τις μετρήσεις

$$z_k = h_k(x_k, u_k) \quad (2)$$

όπου η $h_k : \mathcal{R}^{n_x} \times \mathcal{R}^{n_u} \rightarrow \mathcal{R}^{n_z}$ είναι μια πιθανώς μη γραμμική συνάρτηση, η $\{u_k, k \in \mathbb{N}\}$ είναι μια i.i.d. ανέλιξη ακολουθία θορύβου και οι n_z και n_u είναι οι διαστάσεις των διανυσμάτων μετρήσεων και μετρήσεων θορύβου αντίστοιχα. Ειδικότερα, αναζητούμε τις φιλτραρισμένες εκτιμήσεις x_k βασισμένοι στο σύνολο όλων των διαθέσιμων μετρήσεων $z_{1:k} = \{z_i, i = 1, \dots, k\}$ μέχρι τη στιγμή k .

Από μια Bayesian προοπτική, το πρόβλημα παρακολούθησης είναι ο αναδρομικός υπολογισμός μιας εκτίμησης της κατάστασης x_k τη στιγμή k , λαμβάνοντας υπόψη τα δεδομένα $z_{1:k}$ μέχρι τη στιγμή k . Κατά συνέπεια, πρέπει

να κατασκευαστεί η pdf $p(x_k | z_{1:k})$. Υποτίθεται ότι η αρχική pdf $p(x_0 | z_0) \equiv p(x_0)$ του διανύσματος κατάστασης, που είναι γνωστή ως εκ των προτέρων συνάρτηση πυκνότητας πιθανότητας (prior pdf), είναι διαθέσιμη (z_0 είναι το σύνολο απουσίας μετρήσεων). Κατόπιν, η pdf $p(x_k | z_{1:k})$ μπορεί να ληφθεί, αναδρομικά, σε δύο στάδια: πρόβλεψη και ενημέρωση.

Υποθέτουμε ότι η απαιτούμενη pdf $p(x_{k-1} | z_{1:k-1})$ είναι διαθέσιμη τη στιγμή $k-1$. Το στάδιο πρόβλεψης περιλαμβάνει τη χρήση του μοντέλου (1) του συστήματος για τον υπολογισμό της εκ των προτέρων pdf της κατάστασης τη στιγμή k , μέσω της εξίσωσης Chapman–Kolmogorov:

$$p(x_k | z_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | z_{1:k-1}) dx_{k-1} \quad (3)$$

Στην (3) έχει γίνει χρήση του γεγονότος ότι $p(x_k | x_{k-1}, z_{1:k-1}) = p(x_k | x_{k-1})$, αφού η (1) περιγράφει μια Markov διαδικασία πρώτης τάξης. Το πιθανολογικό πρότυπο της εξέλιξης κατάστασης καθορίζεται από την εξίσωση συστήματος (1) και τις γνωστές στατιστικές για το v_{k-1} .

Τη χρονική στιγμή k γίνεται διαθέσιμη μια μέτρηση z_k , και αυτό μπορεί να χρησιμοποιηθεί για την ανανέωση της εκ των προτέρων pdf (στάδιο ενημέρωσης) μέσω του κανόνα Bayes

$$p(x_k | z_{1:k}) = \frac{p(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})} \quad (4)$$

όπου η σταθερά κανονικοποίησης

$$p(z_k | z_{1:k-1}) = \int p(z_k | x_k) p(x_k | z_{1:k-1}) dx_k \quad (5)$$

εξαρτάται από τη συνάρτηση πιθανότητας $p(z_k | x_k)$ που καθορίζεται από το μοντέλο μέτρησης (2) και τις γνωστές στατιστικές για το u_k . Στο στάδιο ενημέρωσης (4), η μέτρηση z_k χρησιμοποιείται για να τροποποιήσει την εκ των προτέρων πυκνότητα για να υπολογιστεί η απαραίτητη εκ των υστέρων πυκνότητα της τρέχουσας κατάστασης.

Οι αναδρομικές σχέσεις (3) και (4) αποτελούν τη βάση για τη βέλτιστη Bayesian λύση. Αυτή η αναδρομική διάδοση της εκ των υστέρων πυκνότητας είναι μόνο μια θεμελιώδης λύση η οποία γενικά, δεν μπορεί να καθοριστεί αναλυτικά. Λύσεις υπάρχουν σε ένα περιορισμένο σύνολο περιπτώσεων, συμπεριλαμβανομένου του φίλτρου Kalman και των βασισμένων σε πλέγματα φίλτρων, που περιγράφονται στην επόμενη ενότητα. Περιγράφεται επίσης πώς, όταν η αναλυτική λύση δεν είναι δυνατή, τα εκτεταμένα φίλτρα Kalman, τα κατά προσέγγιση βασισμένα σε πλέγματα φίλτρα, και το particle filtering προσεγγίζουν τη βέλτιστη Bayesian λύση.

5.3.3. Βέλτιστοι αλγόριθμοι

A. Φίλτρο Kalman

Το φίλτρο Kalman υποθέτει ότι η εκ των υστέρων πυκνότητα σε κάθε χρονική στιγμή είναι Gaussian και, ως εκ τούτου, περιγράφεται με δύο παραμέτρους: μέση τιμή και συνδιακύμανση. Εάν η $p(x_{k-1} | z_{1:k-1})$ είναι Gaussian, μπορεί να αποδειχθεί ότι η $p(x_k | z_{1:k})$ είναι επίσης Gaussian, δεδομένου ότι ισχύουν ορισμένες υποθέσεις:

- ✓ οι v_{k-1} και u_k προέρχονται από Gaussian κατανομές γνωστών παραμέτρων
- ✓ η $f_k(x_{k-1}, v_{k-1})$ είναι γνωστή και είναι γραμμική συνάρτηση των x_{k-1} και v_{k-1}
- ✓ η $h_k(x_k, u_k)$ είναι γνωστή και είναι γραμμική συνάρτηση των x_k και u_k

Έτσι, οι (1) και (2) μπορούν να ξαναγραφούν ως εξής:

$$x_k = F_k x_{k-1} + v_{k-1} \quad (6)$$

$$z_k = H_k x_k + u_k \quad (7)$$

όπου οι F_k και H_k είναι γνωστοί πίνακες που ορίζουν τις γραμμικές συναρτήσεις. Οι συνδιακυμάνσεις των v_{k-1} και u_k είναι αντίστοιχα Q_{k-1} και R_k . Εδώ θεωρούμε την περίπτωση που οι v_{k-1} και u_k έχουν μηδενική μέση τιμή και είναι στατιστικά ανεξάρτητες. Οι πίνακες συστήματος και μετρήσεων F_k και H_k , όπως και οι παράμετροι θορύβου Q_{k-1} και R_k , μπορεί αν είναι χρονικά μεταβαλλόμενα μεγέθη.

Ο αλγόριθμος του φίλτρου Kalman, ο οποίος παρήχθη χρησιμοποιώντας τις (3) και (4), περιγράφεται από την παρακάτω αναδρομική σχέση:

$$p(x_{k-1} | z_{1:k-1}) = N(x_{k-1}; m_{k-1|k-1}, P_{k-1|k-1}) \quad (8)$$

$$p(x_k | z_{1:k-1}) = N(x_k; m_{k|k-1}, P_{k|k-1}) \quad (9)$$

$$p(x_k | z_{1:k}) = N(x_k; m_{k|k}, P_{k|k}) \quad (10)$$

όπου

$$m_{k|k-1} = F_k m_{k-1|k-1} \quad (11)$$

$$P_{k|k-1} = Q_{k-1} + F_k P_{k-1|k-1} F_k^T \quad (12)$$

$$m_{k|k} = m_{k|k-1} + K_k (z_k - H_k m_{k|k-1}) \quad (13)$$

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} \quad (14)$$

και η $N(x; m, P)$ είναι Gaussian πυκνότητα με όρισμα x , μέση τιμή m και συνδιακύμανση P και οι όροι

$$S_k = H_k P_{k|k-1} H_k^T + R_k \quad (15)$$

$$K_k = P_{k|k-1} H_k^T S_k^{-1} \quad (16)$$

είναι η συνδιακύμανση του όρου $z_k - H_k m_{k|k-1}$ και το κέρδος Kalman αντίστοιχα. Στις παραπάνω εξισώσεις M^T είναι ο ανάστροφος πίνακας του M .

Αυτή είναι η βέλτιστη λύση του προβλήματος παρακολούθησης, αν ισχύουν οι κατάλληλες υποθέσεις. Η έννοια της βέλτιστης λύσης είναι ότι κανείς αλγόριθμος δε θα παράγει καλύτερο αποτέλεσμα από ένα φίλτρο Kalman σε αυτό το γραμμικό, Gaussian περιβάλλον.

B. Μέθοδοι πλεγμάτων (Grid-Based Methods)

Οι βασισμένες σε πλέγματα μέθοδοι παρέχουν τη βέλτιστη αναδρομή της φιλτραρισμένης πυκνότητας $p(x_k | z_{1:k})$ αν ο χώρος κατάστασης είναι διακριτός και αποτελείται από πεπερασμένο αριθμό καταστάσεων. Υποθέτουμε ότι ο χώρος κατάστασης τη χρονική στιγμή $k-1$ αποτελείται από διακριτές καταστάσεις $x_{k-1}^i, i=1, \dots, N_s$. Για κάθε κατάσταση x_{k-1}^i συμβολίζουμε την υπό συνθήκη πιθανότητα της κατάστασης, δεδομένων των μετρήσεων μέχρι τη στιγμή $k-1$ ως $w_{k-1|k-1}^i$, δηλαδή $\Pr(x_{k-1} = x_{k-1}^i | z_{1:k-1}) = w_{k-1|k-1}^i$. Έτσι, η εκ των υστέρων pdf τη στιγμή $k-1$ μπορεί να γραφτεί ως

$$p(x_{k-1} | z_{1:k-1}) = \sum_{i=1}^{N_s} w_{k-1|k-1}^i \delta(x_{k-1} - x_{k-1}^i) \quad (17)$$

όπου $\delta(\cdot)$ είναι η συνάρτηση Dirac. Αντικαθιστώντας τη (17) στις (3) και (4) προκύπτουν οι εξισώσεις πρόβλεψης και ενημέρωσης αντίστοιχα:

$$p(x_k | z_{1:k-1}) = \sum_{i=1}^{N_s} w_{k|k-1}^i \delta(x_k - x_k^i) \quad (18)$$

$$p(x_k | z_{1:k}) = \sum_{i=1}^{N_s} w_{k|k}^i \delta(x_k - x_k^i) \quad (19)$$

όπου

$$w_{k|k-1}^i \triangleq \sum_{j=1}^{N_s} w_{k-1|k-1}^j p(x_k^i | x_{k-1}^j) \quad (20)$$

$$w_{k|k}^i \triangleq \frac{w_{k|k-1}^i p(z_k | x_k^i)}{\sum_{j=1}^{N_s} w_{k|k-1}^j p(z_k | x_k^j)} \quad (21)$$

Υποτίθεται ότι οι $p(x_k^i | x_{k-1}^j)$ και $p(z_k | x_k^j)$ είναι γνωστές, χωρίς όμως να περιορίζεται η μορφή αυτών των διακριτών πυκνοτήτων. Όπως και στην περίπτωση του φίλτρου Kalman, αυτή είναι η βέλτιστη λύση, αν ισχύουν οι απαραίτητες υποθέσεις.

5.3.4. Σχεδόν βέλτιστοι (suboptimal) αλγόριθμοι

Σε πολλές ενδιαφέρουσες περιπτώσεις οι υποθέσεις που αναφέρθηκαν προηγουμένως δεν ισχύουν. Τότε, το φίλτρο Kalman και οι βασισμένες σε πλέγματα μέθοδοι δεν μπορούν να χρησιμοποιηθούν όπως περιγράφηκαν, αλλά είναι απαραίτητες κάποιες προσεγγίσεις. Σε αυτή την ενότητα μελετώνται τρία προσεγγιστικά, μη γραμμικά Bayesian φίλτρα:

- ✓ εκτεταμένο φίλτρο Kalman (extended Kalman filter EKF)
- ✓ προσεγγιστικές μέθοδοι βασισμένες σε πλέγματα
- ✓ particle filters

A. Extended Kalman Filter

Αν οι (1) και (2) δεν μπορούν να ξαναγραφούν στη μορφή των (6) και (7) επειδή οι συναρτήσεις είναι μη γραμμικές, τότε μια τοπική γραμμικοποίηση των εξισώσεων μπορεί να είναι μια επαρκής περιγραφή της μη γραμμικότητας. Το EKF βασίζεται σε αυτή την προσέγγιση. Η $p(x_k | z_{1:k})$ προσεγγίζεται από μια Gaussian:

$$p(x_{k-1} | z_{1:k-1}) \approx N(x_{k-1}; m_{k-1|k-1}, P_{k-1|k-1}) \quad (22)$$

$$p(x_k | z_{1:k-1}) \approx N(x_k; m_{k|k-1}, P_{k|k-1}) \quad (23)$$

$$p(x_k | z_{1:k}) \approx N(x_k; m_{k|k}, P_{k|k}) \quad (24)$$

όπου

$$m_{k|k-1} = f_k(m_{k-1|k-1}) \quad (25)$$

$$P_{k|k-1} = Q_{k-1} + \hat{F}_k P_{k-1|k-1} \hat{F}_k^T \quad (26)$$

$$m_{k|k} = m_{k|k-1} + K_k(z_k - h_k(m_{k|k-1})) \quad (27)$$

$$P_{k|k} = P_{k|k-1} - K_k \hat{H}_k P_{k|k-1} \quad (28)$$

όπου οι $f_k(\cdot)$ και $h_k(\cdot)$ είναι μη γραμμικές συναρτήσεις και οι \hat{F}_k και \hat{H}_k είναι τοπικές γραμμικοποιήσεις αυτών των μη γραμμικών συναρτήσεων (δηλαδή πίνακες):

$$\hat{F}_k = \left. \frac{df_k(x)}{dx} \right|_{x=m_{k-1|k-1}} \quad (29)$$

$$\hat{H}_k = \left. \frac{dh_k(x)}{dx} \right|_{x=m_{k|k-1}} \quad (30)$$

$$S_k = \hat{H}_k P_{k|k-1} \hat{H}_k^T + R_k \quad (31)$$

$$K_k = P_{k|k-1} \hat{H}_k^T S_k^{-1} \quad (32)$$

Το EKF όπως περιγράφεται παραπάνω χρησιμοποιεί τον πρώτο όρο αναπτύγματος Taylor της μη γραμμικής συνάρτησης. Υπάρχει και EKF

μεγαλύτερης τάξης, που διατηρεί περισσότερους όρους του αναπτύγματος Taylor αλλά η πρόσθετη πολυπλοκότητά του έχει απαγορεύσει τη διαδεδομένη χρήση του.

Εντούτοις, το EKF προσεγγίζει πάντα την $p(x_k | z_{1:k})$ με Gaussian. Εάν η πραγματική πυκνότητα είναι μη Gaussian, τότε μια Gaussian δεν μπορεί να την περιγράψει καλά. Σε τέτοιες περιπτώσεις, τα προσεγγιστικά φίλτρα που βασίζονται σε πλέγματα και το particle filtering θα οδηγήσουν σε βελτίωση της απόδοσης σε σύγκριση με το EKF.

B. Προσεγγιστικές μέθοδοι πλεγμάτων (Approximate Grid-Based Methods)

Αν ο χώρος κατάσταση είναι συνεχής, αλλά μπορεί να χωριστεί σε N_s "κελιά", $x_k^i, i = 1, \dots, N_s$, τότε μια βασισμένη σε πλέγματα μέθοδος μπορεί να χρησιμοποιηθεί για την προσέγγιση της εκ των υστέρων πυκνότητας. Συγκεκριμένα, υποθέτουμε ότι η προσέγγιση της εκ των υστέρων pdf τη στιγμή $k-1$ δίνεται από:

$$p(x_{k-1} | z_{1:k-1}) \approx \sum_{i=1}^{N_s} w_{k-1|i}^i \delta(x_{k-1} - x_{k-1}^i) \quad (33)$$

Τότε οι εξισώσεις πρόβλεψης και ενημέρωσης γράφονται αντίστοιχα:

$$p(x_k | z_{1:k-1}) \approx \sum_{i=1}^{N_s} w_{k|k-1}^i \delta(x_k - x_k^i) \quad (34)$$

$$p(x_k | z_{1:k}) \approx \sum_{i=1}^{N_s} w_{k|k}^i \delta(x_k - x_k^i) \quad (35)$$

όπου

$$w_{k|k-1}^i \triangleq \sum_{j=1}^{N_s} w_{k-1|j}^j \int_{x \in x_k^i} p(x | \bar{x}_{k-1}^j) dx \quad (36)$$

$$w_{k|k}^i \triangleq \frac{w_{k|k-1}^i \int_{x \in x_k^i} p(z_k | x) dx}{\sum_{j=1}^{N_s} w_{k|k-1}^j \int_{x \in x_k^j} p(z_k | x) dx} \quad (37)$$

Εδώ, ο όρος \bar{x}_{k-1}^j αποτελεί το κέντρο του j -κελιού τη χρονική στιγμή $k-1$. Τα ολοκληρώματα στις (36) και (37) προκύπτουν από το γεγονός ότι τα σημεία του πλέγματος $x_k^i, i = 1, \dots, N_s$ αναπαριστούν περιοχές του συνεχούς χώρου κατάσταση, και έτσι οι πιθανότητες πρέπει να ολοκληρωθούν σε αυτές τις περιοχές. Πρακτικά, για την απλοποίηση των υπολογισμών, γίνεται μια περαιτέρω προσέγγιση στην εκτίμηση του $w_{k|k}^i$. Συγκεκριμένα, αυτά τα βάρη υπολογίζονται στο κέντρο του αντίστοιχου στο x_k^j κελιού.

$$w_{k|k-1}^i \approx \sum_{j=1}^{N_s} w_{k-1|j}^j p(\bar{x}_k^i | \bar{x}_{k-1}^j) \quad (38)$$

$$w_{k|k}^j \approx \frac{w_{k|k-1}^j p(z_k | \bar{x}_k^j)}{\sum_{j=1}^{N_s} w_{k|k-1}^j p(z_k | \bar{x}_k^j)} \quad (39)$$

Το πλέγμα πρέπει να είναι αρκετά πυκνό για να προκύψει μια καλή προσέγγιση του συνεχούς χώρου κατάστασης. Όσο αυξάνεται η διάσταση του χώρου κατάστασης, τόσο περισσότερο αυξάνεται το υπολογιστικό κόστος της προσέγγισης. Εάν ο χώρος κατάστασης δεν είναι πεπερασμένης έκτασης, τότε η χρήση μιας βασισμένης σε πλέγμα προσέγγισης απαιτεί κάποια αποκοπή του χώρου κατάστασης. Ένα άλλο μειονέκτημα των βασισμένων σε πλέγματα μεθόδων είναι ότι ο χώρος κατάστασης πρέπει να προκαθοριστεί και, επομένως, δεν μπορεί να χωριστεί άνισα για να δώσει μεγαλύτερη ανάλυση σε περιοχές υψηλής πυκνότητας πιθανότητας, εκτός αν χρησιμοποιηθεί η προγενέστερη γνώση. Μια εφαρμογή τέτοιων μεθόδων είναι τα Hidden Markov model (HMM) φίλτρα σε fixed-interval smoothing context και έχουν χρησιμοποιηθεί ευρέως στην επεξεργασία φωνής.

Γ. Μέθοδοι particle filtering

Ο βασικός αλγόριθμος του particle filtering, καθώς και διάφορες παραλλαγές και επεκτάσεις του περιγράφονται στην επόμενη ενότητα.

5.4. Μέθοδοι Particle Filtering

5.4.1. Sequential Importance Sampling (SIS) Algorithm

Ο αλγόριθμος δειγματοληψίας ακολουθιακής σημασίας (Sequential Importance Sampling SIS) είναι μια μέθοδος Monte Carlo (MC) που αποτελεί τη βάση για τα περισσότερα ακολουθιακά φίλτρα MC που αναπτύχθηκαν κατά τη διάρκεια των προηγούμενων δεκαετιών. Αυτή η ακολουθιακή MC (Sequential MC SMC) προσέγγιση είναι ευρέως γνωστή ως bootstrap filtering, αλγόριθμος condensation, particle filtering, και survival of the fittest. Η βασική ιδέα είναι η αναπαράσταση της απαιτούμενης εκ των υστέρων συνάρτησης πυκνότητας (posterior density function) με ένα σύνολο τυχαίων δειγμάτων με αντίστοιχα βάρη και ο υπολογισμός εκτιμήσεων που βασίζονται σε αυτά τα δείγματα και τα βάρη. Όσο ο αριθμός των δειγμάτων γίνεται πολύ μεγάλος, αυτός ο χαρακτηρισμός MC γίνεται μια ισοδύναμη αναπαράσταση της συνήθους περιγραφής της εκ των υστέρων pdf, και το φίλτρο SIS προσεγγίζει τη βέλτιστη Bayesian εκτίμηση.

Προκειμένου να μελετηθούν οι λεπτομέρειες του αλγορίθμου, θεωρούμε μια τυχαία ποσότητα $\{x_{0:k}^i, w_k^i\}_{i=1}^{N_s}$ που χαρακτηρίζει την εκ των υστέρων pdf $p(x_{0:k} | z_{1:k})$, όπου $\{x_{0:k}^i, i = 0, \dots, N_s\}$ είναι ένα σύνολο σημείων στήριξης (support points) με τα αντίστοιχα βάρη $\{w_k^i, i = 1, \dots, N_s\}$ και $x_{0:k} = \{x_j, j = 0, \dots, k\}$ είναι το σύνολο όλων των καταστάσεων μέχρι τη χρονική στιγμή k . Τα βάρη

κανονικοποιούνται, έτσι ώστε $\sum_i w_k^i = 1$. Έτσι, η εκ των υστέρων pdf μπορεί να προσεγγιστεί ως:

$$p(x_{0:k} | z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_{0:k} - x_{0:k}^i) \quad (40)$$

Επομένως έχουμε μια διακριτή σταθμισμένη προσέγγιση της πραγματικής posterior pdf, $p(x_{0:k} | z_{1:k})$. Τα βάρη επιλέγονται χρησιμοποιώντας την αρχή της δειγματοληψίας σπουδαιότητας (importance sampling). Αυτή η αρχή στηρίζεται στα ακόλουθα. Υποθέτουμε ότι $p(x) \propto \pi(x)$ είναι μια πυκνότητα πιθανότητας από την οποία είναι δύσκολο να πάρουμε δείγματα αλλά για την οποία μπορεί να εκτιμηθεί το $\pi(x)$ (όπως και το $p(x)$ ως προς την αναλογικότητα). Επιπλέον, θεωρούμε $x^i \sim q(x), i=1, \dots, N_s$ δείγματα που παράγονται εύκολα από μια proposal αποκαλούμενη πυκνότητα σπουδαιότητας (importance density). Έτσι, η σταθμισμένη προσέγγιση στην πυκνότητα $p(\cdot)$ δίνεται από:

$$p(x) \approx \sum_{i=1}^{N_s} w^i \delta(x - x^i) \quad (41)$$

όπου

$$w^i \propto \frac{\pi(x^i)}{q(x^i)} \quad (42)$$

είναι το κανονικοποιημένο βάρος του particle i .

Επομένως, εάν τα δείγματα προήλθαν από μια importance density $q(x_{0:k} | z_{1:k})$, τότε τα βάρη στη (40) καθορίζονται από τη (42) και είναι:

$$w_k^i \propto \frac{p(x_{0:k}^i | z_{1:k})}{q(x_{0:k}^i | z_{1:k})} \quad (43)$$

Επιστρέφοντας στην ακολουθιακή περίπτωση, σε κάθε επανάληψη θα μπορούσαμε να έχουμε δείγματα που αποτελούν μια προσέγγιση στην $p(x_{0:k-1} | z_{1:k-1})$ και να θέλουμε να προσεγγίσουμε την $p(x_{0:k} | z_{1:k})$ με ένα νέο σύνολο δειγμάτων. Εάν η importance density επιλέγεται έτσι ώστε

$$q(x_{0:k} | z_{1:k}) = q(x_k | x_{0:k-1}, z_{1:k}) q(x_{0:k-1} | z_{1:k-1}) \quad (44)$$

Τότε τα νέα δείγματα $x_{0:k}^i \sim q(x_{0:k} | z_{1:k})$ μπορούν να προκύψουν προσαυξάνοντας τα υπάρχοντα δείγματα $x_{0:k-1}^i \sim q(x_{0:k-1} | z_{1:k-1})$ με τη νέα κατάσταση $x_k^i \sim q(x_k | x_{0:k-1}, z_{1:k})$. Για να παραχθεί η εξίσωση ανανέωσης των βαρών, η

$p(x_{0:k} | z_{1:k})$ εκφράζεται αρχικά σε όρους $p(x_{0:k-1} | z_{1:k-1})$, $p(z_k | x_k)$ και $p(x_k | x_{k-1})$. Σημειώνουμε ότι η (4) μπορεί να προκύψει ολοκληρώνοντας τη (45):

$$\begin{aligned} p(x_{0:k} | z_{1:k}) &= \frac{p(z_k | x_{0:k} | z_{1:k-1}) p(x_{0:k} | z_{1:k-1})}{p(z_k | z_{1:k-1})} \\ &= \frac{p(z_k | x_{0:k} | z_{1:k-1}) p(x_k | x_{0:k-1} | z_{1:k-1})}{p(z_k | z_{1:k-1})} \times p(x_{0:k-1} | z_{1:k-1}) \end{aligned} \quad (45)$$

$$\begin{aligned} &= \frac{p(z_k | x_k) p(x_k | x_{k-1})}{p(z_k | z_{1:k-1})} p(x_{0:k-1} | z_{1:k-1}) \\ &\propto p(z_k | x_k) p(x_k | x_{k-1}) p(x_{0:k-1} | z_{1:k-1}) \end{aligned} \quad (46)$$

Με την αντικατάσταση των (44) και (46) στη (43), η εξίσωση ενημέρωσης των βαρών προκύπτει ως:

$$\begin{aligned} w_k^i &\propto \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i) p(x_{0:k-1}^i | z_{1:k-1})}{q(x_k^i | x_{0:k-1}^i, z_{1:k}) q(x_{0:k-1}^i | z_{1:k-1})} \\ &= w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{0:k-1}^i, z_{1:k})} \end{aligned} \quad (47)$$

Επιπλέον, αν $q(x_k | x_{0:k-1}, z_{1:k}) = q(x_k | x_{k-1}, z_k)$, τότε η importance density εξαρτάται μόνο από τα x_{k-1} και z_k . Αυτό είναι ιδιαίτερα χρήσιμο στην κοινή περίπτωση όπου απαιτείται μόνο μια φιλτραρισμένη εκτίμηση της $p(x_k | z_{1:k})$ σε κάθε χρονική στιγμή. Στο εξής, αν δεν αναφέρεται διαφορετικά, υποθέτουμε αυτή την περίπτωση. Σε τέτοια σενάρια, χρειάζεται να αποθηκεύεται μόνο το x_k^i , επομένως, η ακολουθία $x_{0:k-1}^i$ και η ιστορία των παρατηρήσεων $z_{1:k-1}$ μπορούν να απορριφθούν. Το τροποποιημένο βάρος είναι τότε

$$w_k^i \propto w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, z_k)} \quad (48)$$

και η φιλτραρισμένη posterior density μπορεί να προσεγγιστεί ως:

$$p(x_k | z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i) \quad (49)$$

όπου τα βάρη καθορίζονται από τη (48). Αποδεικνύεται ότι όσο $N_s \rightarrow \infty$, η προσέγγιση της (49) πλησιάζει την πραγματική posterior density $p(x_k | z_{1:k})$.

Ο SIS αλγόριθμος αποτελείται λοιπόν από αναδρομική διάδοση των βαρών και των σημείων στήριξης (recursive propagation of the weights and support points) όταν παραλαμβάνεται ακολουθιακά μια νέα μέτρηση.

Μια περιγραφή ψευδοκώδικα αυτού του αλγορίθμου δίνεται από τον αλγόριθμο 1.

Αλγόριθμος 1: SIS Particle Filter

$$\left[\left\{ x_k^i, w_k^i \right\}_{i=1}^{N_s} \right] = \text{SIS} \left[\left\{ x_{k-1}^i, w_{k-1}^i \right\}_{i=1}^{N_s}, z_k \right]$$

- FOR $i = 1 : N_s$
 - Λήψη $x_k^i \sim q(x_k | x_{k-1}^i, z_k)$
 - Ανάθεση βάρους w_k^i στο particle, σύμφωνα με τη (48)
- END FOR

Εικόνα 5.2: Αλγόριθμος 1

5.4.2. Πρόβλημα εκφυλισμού και τεχνικές αντιμετώπισής του

Πρόβλημα εκφυλισμού (degeneracy problem): Ένα κοινό πρόβλημα με το SIS particle filtering είναι το φαινόμενο εκφυλισμού, όπου μετά από μερικές επαναλήψεις όλα τα δείγματα (particles) εκτός από ένα, θα έχουν αμελητέο βάρος. Έχει αποδειχθεί ότι η διακύμανση των βαρών μπορεί μόνο να αυξάνεται στο χρόνο και έτσι είναι αδύνατο να αποφευχθεί το φαινόμενο εκφυλισμού. Αυτός ο εκφυλισμός υπονοεί ότι αφιερώνεται πολύς υπολογιστικός χρόνος στην ενημέρωση των δειγμάτων, των οποίων η συμβολή στην προσέγγιση της $p(x_k | z_{1:k})$ είναι σχεδόν μηδενική. Ένα κατάλληλο μέτρο του εκφυλισμού του αλγορίθμου είναι το μέγεθος των αποτελεσματικών δειγμάτων N_{eff} που ορίζεται ως:

$$N_{eff} = \frac{N_s}{1 + \text{var}(w_k^{*i})} \quad (50)$$

όπου ο όρος $w_k^{*i} = p(x_k^i | z_{1:k}) / q(x_k^i | x_{k-1}^i, z_k)$, αναφέρεται σαν το "πραγματικό βάρος". Αυτό το μέγεθος δεν μπορεί να υπολογιστεί ακριβώς, αλλά μπορεί να δοθεί για αυτό μια εκτίμηση \hat{N}_{eff} , μέσω της παρακάτω εξίσωσης:

$$\hat{N}_{eff} = \frac{1}{\sum_{i=1}^{N_s} (w_k^i)^2} \quad (51)$$

όπου w_k^i είναι το κανονικοποιημένο βάρος που υπολογίζεται από την (47). Σημειώνουμε ότι $N_{eff} \leq N_s$ και ότι μικρή τιμή του N_{eff} δείχνει ισχυρό εκφυλισμό, δηλαδή υποδεικνύει ότι οι εκτιμήσεις που βασίζονται στο σύνολο των particles μπορεί να μην είναι αξιόπιστες και υπάρχει κίνδυνος να χαθεί ο στόχος που

παρακολουθείται. Ουσιαστικά, το μέγεθος αυτό υποδεικνύει τον αριθμό των particles που θα "επιζούσαν" σε μια διαδικασία επαναδειγματοληψίας (resampling), διαδικασία που θα περιγραφεί στη συνέχεια του κεφαλαίου.

Σαφώς, το πρόβλημα εκφυλισμού είναι μια ανεπιθύμητη επίδραση στο particle filtering. Μια δραστική λύση για τη μείωση της επίδρασής του είναι η χρήση πολύ μεγάλου N_s . Αυτό είναι συχνά μη πρακτικό, επομένως στηριζόμαστε σε δύο άλλες μεθόδους, που περιγράφονται στη συνέχεια:

- ✓ καλή επιλογή της importance density και
- ✓ χρήση επαναδειγματοληψίας (resampling).

Καλή επιλογή importance density: Η πρώτη μέθοδος περιλαμβάνει επιλογή της importance density $q(x_k | x_{k-1}^i, z_k)$ ώστε να ελαχιστοποιείται η $\text{var}(w_k^{*i})$ και να μεγιστοποιείται το N_{eff} . Η βέλτιστη συνάρτηση importance density που ελαχιστοποιεί τη διακύμανση των πραγματικών βαρών w_k^{*i} υπό τους όρους των x_{k-1}^i και z_k έχει αποδειχθεί ότι είναι:

$$\begin{aligned} q(x_k | x_{k-1}^i, z_k)_{opt} &= p(x_k | x_{k-1}^i, z_k) \\ &= \frac{p(z_k | x_k | x_{k-1}^i) p(x_k | x_{k-1}^i)}{p(z_k | x_{k-1}^i)} \end{aligned} \quad (52)$$

Αντικατάσταση της (52) στη (48) δίνει:

$$\begin{aligned} w_k^i &\propto w_{k-1}^i p(z_k | x_{k-1}^i) \\ &= w_{k-1}^i \int p(z_k | x_k^i) p(x_k^i | x_{k-1}^i) dx_k^i \end{aligned} \quad (53)$$

Αυτή η επιλογή της importance density είναι βέλτιστη αφού για δεδομένο x_{k-1}^i , το w_k^i παίρνει την ίδια τιμή, οτιδήποτε δείγμα και να ληφθεί από $q(x_k | x_{k-1}^i, z_k)_{opt}$. Ως εκ τούτου, υπό τους όρους του x_{k-1}^i , είναι $\text{var}(w_k^{*i}) = 0$. Αυτή είναι η διακύμανση του διαφορετικού w_k^i που προκύπτει από διαφορετικό δείγμα του x_k^i .

Αυτή η βέλτιστη importance density πάσχει από δύο σημαντικά μειονεκτήματα. Απαιτεί τη δυνατότητα δειγματοληψίας από την $p(x_k | x_{k-1}^i, z_k)$ και την εκτίμηση του ολοκληρώματος στη νέα κατάσταση. Στη γενική περίπτωση, μπορεί να μην είναι απλό να γίνει κανένα από αυτά τα δύο. Υπάρχουν δύο περιπτώσεις όπου είναι δυνατή η χρήση της βέλτιστης importance density.

Η πρώτη περίπτωση είναι όταν το x_k είναι μέλος ενός πεπερασμένου συνόλου. Σε τέτοιες περιπτώσεις, το ολοκλήρωμα στην (53) γίνεται άθροισμα, και είναι δυνατή η δειγματοληψία από την $p(x_k | x_{k-1}^i, z_k)$.

Αναλυτική εκτίμηση είναι δυνατή για μια δεύτερη κατηγορία μοντέλων για τα οποία η $p(x_k | x_{k-1}^i, z_k)$ είναι Gaussian. Αυτό μπορεί να συμβεί αν η δυναμική του συστήματος είναι μη γραμμική και οι μετρήσεις γραμμικές. Ένα τέτοιο σύστημα δίνεται από τις:

$$x_k = f_k(x_{k-1}) + v_{k-1} \quad (54)$$

$$z_k = H_k x_k + n_k \quad (55)$$

όπου

$$v_{k-1} \sim N(v_{k-1}; 0_{n_v \times 1}, Q_{k-1}) \quad (56)$$

$$n_k \sim N(n_k; 0_{n_u \times 1}, R_k) \quad (57)$$

και η $f_k : \mathcal{R}^{n_x} \rightarrow \mathcal{R}^{n_x}$ είναι μια μη γραμμική συνάρτηση, $H_k \in \mathcal{R}^{n_z \times n_x}$ είναι η μήτρα παρατήρησης, και οι v_{k-1} και n_k είναι αμοιβαία ανεξάρτητες i.i.d. Gaussian ακολουθίες με $Q_{k-1} > 0$ και $R_k > 0$. Ορίζοντας

$$\Sigma_k^{-1} = Q_{k-1}^{-1} + H_k^T R_k^{-1} H_k \quad (58)$$

$$m_k = \Sigma_k \left(Q_{k-1}^{-1} f_k(x_{k-1}) + H_k^T R_k^{-1} z_k \right) \quad (59)$$

παίρνουμε

$$p(x_k | x_{k-1}, z_k) = N(x_k; m_k, \Sigma_k) \quad (60)$$

και

$$p(z_k | x_{k-1}) = N(z_k; H_k f_k(x_{k-1}), Q_{k-1} + H_k R_k H_k^T) \quad (61)$$

Για πολλά άλλα μοντέλα, τέτοιες αναλυτικές εκτιμήσεις δεν είναι εφικτές. Εντούτοις, είναι δυνατό να κατασκευαστούν σχεδόν βέλτιστες προσεγγίσεις της βέλτιστης importance density με τη χρησιμοποίηση τεχνικών τοπικής γραμμικοποίησης. Τέτοιες γραμμικοποιήσεις χρησιμοποιούν μια importance density που είναι Gaussian προσέγγιση της $p(x_k | x_{k-1}, z_k)$. Μια άλλη προσέγγιση είναι να εκτιμηθεί μια Gaussian προσέγγιση της $p(x_k | x_{k-1}, z_k)$ χρησιμοποιώντας τον unscented μετασχηματισμό.

Τέλος, είναι συχνά βολικό να επιλεγεί η importance density ως η prior:

$$q(x_k | x_{k-1}^i, z_k) = p(x_k | x_{k-1}^i) \quad (62)$$

Αντικατάσταση της (62) στη (48) δίνει

$$w_k^i \propto w_{k-1}^i p(z_k | x_k^i) \quad (63)$$

Αυτό θα ήταν η πιο συνήθης επιλογή importance density δεδομένου ότι είναι δεισθητική και απλή στην υλοποίηση. Εντούτοις, υπάρχει πληθώρα άλλων densities που μπορούν να χρησιμοποιηθούν και αυτή η επιλογή είναι το κρίσιμο σημείο στο σχεδιασμό του particle filtering.

Resampling: Η δεύτερη μέθοδος με την οποία μπορούν να μειωθούν τα αποτελέσματα του εκφυλισμού έγκειται στη χρήση της επαναδειγματοληψίας (resampling) όποτε παρατηρείται σημαντικός εκφυλισμός (δηλαδή όταν ο αριθμός N_{eff} είναι μικρότερος κάποιου κατώτατου ορίου). Η βασική ιδέα του resampling είναι να αποβάλλονται δείγματα που έχουν μικρά βάρη και δίνεται περισσότερη σημασία στα δείγματα που έχουν μεγάλα βάρη. Η διαδικασία του resampling περιλαμβάνει την παραγωγή ενός νέου συνόλου $\{x_k^{*i}\}_{i=1}^{N_s}$ δειγματοληψιώντας (με αντικατάσταση) N_s φορές από μια κατά προσέγγιση διακριτή αναπαράσταση της $p(x_k | z_{1:k})$ που δίνεται από την ακόλουθη σχέση:

$$p(x_k | z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i) \quad (64)$$

έτσι ώστε $\Pr(x_k^{*i} = x_k^j) = w_k^j$. Το προκύπτον δείγμα είναι στην πραγματικότητα ένα δείγμα i.i.d. από τη διακριτή πυκνότητα (64), επομένως τα βάρη επαναρυθμίζονται στην τιμή $w_k^i = 1/N_s$. Η διαδικασία του resampling είναι δυνατό να υλοποιηθεί σε $O(N_s)$ βήματα.

Σημειώνεται ότι μπορούν να εφαρμοστούν και άλλοι αποδοτικοί (ως προς τη μειωμένη διακύμανση MC) αλγόριθμοι, όπως stratified sampling και residual sampling, ως εναλλακτικές λύσεις σε αυτό τον αλγόριθμο. Συνήθως προτιμάται ο αλγόριθμος systematic resampling (αφού είναι εύκολος στην υλοποίηση, χρειάζεται $O(N_s)$ βήματα και ελαχιστοποιεί τη διακύμανση MC), ο οποίος περιγράφεται στον αλγόριθμο 2, όπου $U[a, b]$ είναι η ομοιόμορφη κατανομή στο διάστημα $[a, b]$. Για κάθε δείγμα που επαναδειγματοληφτείται, αυτός ο αλγόριθμος αποθηκεύει επίσης το δείκτη του δείγματος από το οποίο προήλθε, το οποίο συμβολίζεται με i^j . Αυτό το στοιχείο αποδεικνύεται χρήσιμο σε μια εναλλακτική μέθοδο particle filtering που ονομάζεται auxiliary sampling importance resampling filter.

Η γενική περιγραφή του particle filtering (με resampling) δίνεται στον αλγόριθμο 3.

Αλγόριθμος 2: Αλγόριθμος Resampling

- $$\left[\left\{ x_k^{j*}, w_k^j, i^j \right\}_{j=1}^{N_s} \right] = \text{RESAMPLE} \left[\left\{ x_k^i, w_k^i \right\}_{i=1}^{N_s} \right]$$
- Αρχικοποίηση της CDF: $c_1 = 0$
 - FOR $i = 1 : N_s$
 - Κατασκευή της CDF: $c_i = c_{i-1} + w_k^i$
 - END FOR
 - Αρχή στη βάση της CDF: $i = 1$
 - Επιλογή σημείου εκκίνησης: $u_1 \sim U[0, N_s^{-1}]$
 - FOR $j = 1 : N_s$
 - Κίνηση κατά μήκος της CDF: $u_j = u_1 + N_s^{-1}(j-1)$
 - WHILE $u_j > c_i$
 - * $i = i + 1$
 - END WHILE
 - Ανάθεση δείγματος: $x_k^{j*} = x_k^i$
 - Ανάθεση βάρους: $w_k^j = N_s^{-1}$
 - Ανάθεση δείκτη: $i^j = i$
 - END FOR

Εικόνα 5.3: Αλγόριθμος 2

Αλγόριθμος 3: Γενικό Particle Filter

- $$\left[\left\{ x_k^i, w_k^i \right\}_{i=1}^{N_s} \right] = \text{PF} \left[\left\{ x_{k-1}^i, w_{k-1}^i \right\}_{i=1}^{N_s}, z_k \right]$$
- FOR $i = 1 : N_s$
 - Λήψη $x_k^i \sim q(x_k | x_{k-1}^i, z_k)$
 - Ανάθεση βάρους w_k^i στο particle, σύμφωνα με τη (48)
 - END FOR
 - Υπολογισμός συνολικού βάρους: $t = \text{SUM} \left[\left\{ w_k^i \right\}_{i=1}^{N_s} \right]$
 - FOR $i = 1 : N_s$
 - Κανονικοποίηση βαρών: $w_k^i = t^{-1} w_k^i$
 - END FOR
 - Υπολογισμός \hat{N}_{eff} , από την (51)
 - IF $\hat{N}_{eff} < N_T$
 - Resample, σύμφωνα με τον Αλγόριθμο 2:
 - * $\left[\left\{ x_k^i, w_k^i, - \right\}_{i=1}^{N_s} \right] = \text{RESAMPLE} \left[\left\{ x_k^i, w_k^i \right\}_{i=1}^{N_s} \right]$
 - END IF

Εικόνα 5.4: Αλγόριθμος 3

Παρόλο που η διαδικασία του resampling μειώνει τα αποτελέσματα του εκφυλισμού, δημιουργεί άλλα πρακτικά προβλήματα. Πρώτον, περιορίζει τη δυνατότητα παραλληλοποίησης αφού πρέπει να συνδυάζονται όλα τα δείγματα. Δεύτερον, τα δείγματα που έχουν μεγάλα βάρη στατιστικά επιλέγονται πολλές φορές. Αυτό οδηγεί σε απώλεια ποικιλομορφίας μεταξύ των δειγμάτων, αφού το προκύπτον δείγμα περιέχει πολλά επαναλαμβανόμενα σημεία. Αυτό το πρόβλημα, το οποίο είναι γνωστό ως εξασθένηση δειγμάτων (sample impoverishment), είναι ιδιαίτερα σοβαρό στην περίπτωση μικρού θορύβου διαδικασίας. Στην πραγματικότητα, για την περίπτωση του πολύ μικρού θορύβου διαδικασίας, όλα τα δείγματα θα εκφυλιστούν σε ένα μόνο σημείο μέσα σε μερικές επαναλήψεις. Τρίτον, δεδομένου ότι η ποικιλομορφία στην πορεία των δειγμάτων μειώνεται, οποιεσδήποτε ομαλές εκτιμήσεις βασισμένες στις πορείες των δειγμάτων θα εκφυλιστούν (degenerate). Υπάρχουν διάφοροι τρόποι αντιμετώπισης αυτής της επίδρασης. Σύμφωνα με μια προσέγγιση οι καταστάσεις των δειγμάτων προκαθορίζονται από το μπροστινό (forward) φίλτρο και έπειτα λαμβάνονται οι ομαλές εκτιμήσεις υπολογίζοντας εκ νέου τα βάρη των δειγμάτων μέσω αναδρομής από το τελευταίο στο πρώτο βήμα. Μια άλλη προσέγγιση είναι να χρησιμοποιηθεί Markov chain Monte Carlo (MCMC) δειγματοληψία (βλ. [Khan et al., 2005]).

Έχουν προταθεί πρόσφατα μερικές συστηματικές τεχνικές για τη λύση του προβλήματος της εξασθένησης δειγμάτων. Μια τέτοια τεχνική είναι ο αλγόριθμος resample-move. Αν και αυτή η τεχνική μοιάζει εννοιολογικά με τις τεχνολογίες των importance sampling-resampling και MCMC sampling, αποφεύγει την εξασθένηση δειγμάτων. Αυτό το πετυχαίνει κατά τρόπο που εξασφαλίζει ότι τα δείγματα προσεγγίζουν ασυμπτωτικά τα δείγματα της posterior. Μια εναλλακτική λύση στο ίδιο πρόβλημα είναι η regularization (συστηματοποίηση). Αυτή η προσέγγιση συχνά βελτιώνει την απόδοση, παρά τη λιγότερο αυστηρή παραγωγή και η χρήση της είναι αρκετά διαδεδομένη.

Τεχνικές για την παράκαμψη της χρήσης μιας σχεδόν βέλτιστης (suboptimal) importance density: Συχνά δεν είναι διαθέσιμη μια καλή importance density. Παραδειγματος χάριν, εάν χρησιμοποιείται η prior $p(x_k | x_{k-1})$ ως importance density και είναι πολύ ευρύτερη κατανομή από την πιθανότητα $p(z_k | x_k)$, τότε μόνο μερικά δείγματα θα έχουν μεγάλο βάρος. Υπάρχουν διάφορες μέθοδοι για την προώθηση των δειγμάτων στη σωστή θέση. Τόσο η χρήση bridging densities όσο και η χρήση προοδευτικής διόρθωσης εισάγουν τις ενδιάμεσες κατανομές μεταξύ της prior και της πιθανότητας. Έπειτα αναθέτονται στα δείγματα νέα βάρη, σύμφωνα με αυτές τις ενδιάμεσες κατανομές και επαναδειγματοληπτούνται. Αυτή η διαδικασία "συγκεντρώνει" τα δείγματα στο σωστό μέρος του χώρου κατάστασης.

Μια άλλη προσέγγιση, γνωστή ως partitioned sampling είναι χρήσιμη αν η πιθανότητα είναι πολύ οξυμένη αλλά μπορεί να παραγοντοποιηθεί σε έναν αριθμό ευρύτερων κατανομών. Τυπικά αυτό συμβαίνει επειδή κάθε μια από τις χωρισμένες κατανομές είναι συνάρτηση κάποιων (όχι όλων) καταστάσεων. Με τη μεταχείριση κάθε μιας από αυτές τις χωρισμένες κατανομές διαδοχικά και επαναδειγματοληπτώντας βάσει κάθε τέτοιας χωρισμένης κατανομής, τα δείγματα οδηγούνται πάλι προς την οξυμένη πιθανότητα.

5.4.3. Άλλες μέθοδοι particle filtering

Ο αλγόριθμος sequential importance sampling που παρουσιάστηκε προηγουμένως αποτελεί τη βάση για τα περισσότερα είδη particle filtering που έχουν αναπτυχθεί. Οι διάφορες παραλλαγές του particle filtering που υπάρχουν στη βιβλιογραφία μπορούν να θεωρηθούν ως ειδικές περιπτώσεις αυτού του γενικού αλγορίθμου SIS. Αυτές οι ειδικές περιπτώσεις μπορούν να παραχθούν από τον αλγόριθμο SIS επιλέγοντας κατάλληλα την importance sampling density ή/και τροποποιώντας τη διαδικασία του resampling.

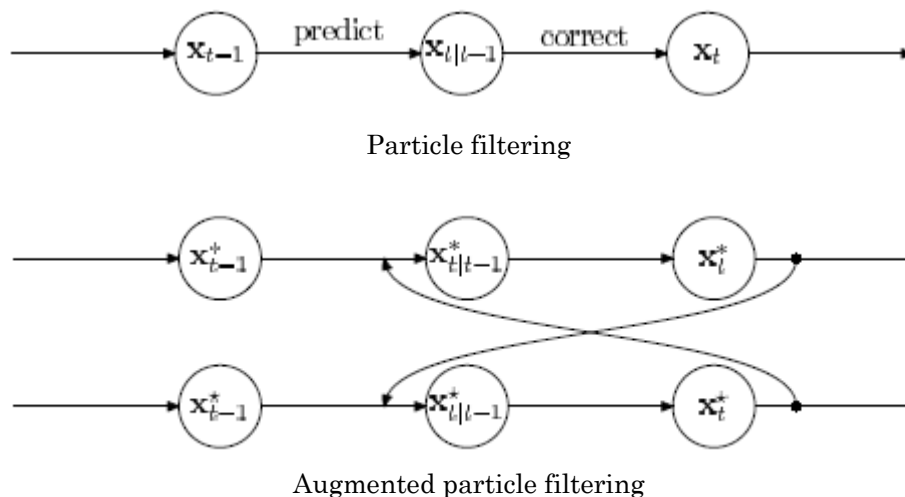
Τρεις παραλλαγές του κλασικού particle filtering που προτείνονται στο [Arulampalam et al., 2002] είναι:

- ✓ φίλτρο sampling importance resampling (SIR)
- ✓ φίλτρο auxiliary sampling importance resampling (ASIR)
- ✓ regularized particle filter (RPF)

Ο αλγόριθμος SIR χρησιμοποιεί ως importance density $q(x_k | x_{k-1}^i, z_{1:k})$ την prior density $p(x_k | x_{k-1}^i)$, ενώ εκτελείται resampling σε κάθε χρονική στιγμή. Ο ASIR αλγόριθμος εισάγει τη χρήση της importance density $q(x_k, i | z_{1:k})$, η οποία δειγματοληπτεί το ζεύγος $\{x_k^i, i^i\}_{j=1}^{M_s}$, όπου ο i^j είναι ο δείκτης του particle τη στιγμή $k-1$, όπως αυτός υπολογίζεται στον αλγόριθμο του systematic resampling. Τέλος, ο αλγόριθμος RPF είναι παρόμοιος με τον SIR, με τη διαφορά ότι στο resampling χρησιμοποιεί μια συνεχή προσέγγιση της posterior density $p(x_k | z_{1:k})$ (ο SIR χρησιμοποιεί την διακριτή προσέγγιση της σχέσης (64)).

Στο [Zhang et al., 2004] προτείνεται ένα νέο πλαίσιο particle filtering που, με την ενσωμάτωση ενός ειδικού έμμεσου μοντέλου μέτρησης, μπορεί να βελτιώσει αισθητά την ικανότητα αναπαράστασης του συνόλου των δειγμάτων, παράγοντας μια ακριβή εκτίμηση της posterior κατανομής για την παρακολούθηση. Ειδικότερα, προτείνεται μια add-on resampling τεχνική για να ενσωματώσει την έμμεση μέτρηση. Κατ' αυτό τον τρόπο, μπορεί να αντιμετωπιστεί η ανάγκη χρήσης μεγάλου αριθμού δειγμάτων που απαιτούνται στο συμβατικό particle filtering.

Στο [Shen et al., 2005] παρουσιάζεται μια άλλη παραλλαγή του βασικού αλγορίθμου particle filtering, που ονομάζεται augmented particle filtering και υιοθετεί δειγματοληψία δύο βημάτων. Συγκεκριμένα, χρησιμοποιούνται δύο επαυξημένα (augmented) φίλτρα, όπως φαίνεται και στο σχήμα που ακολουθεί. Φυσικά μπορεί η μέθοδος να επεκταθεί στην περίπτωση περισσότερων από δύο φίλτρα. Η βασική ιδέα είναι να επεκτείνουμε ένα particle filter (ή ένα φίλτρο Kalman) στο πρώτο στάδιο ώστε να μάθει μια κατανομή δειγματοληψίας. Στο δεύτερο στάδιο, η εκτίμηση που παράγεται στο πρώτο στάδιο παρέχεται ως είσοδος στο δεύτερο particle filter για την κατασκευή ενός βέλτιστου proposal. Η διαδικασία μπορεί να επαναληφθεί για να παραγάγει καλύτερα αποτελέσματα.



Εικόνα 5.5: Κλασικό particle filtering και Augmented particle filtering

Η μέθοδος active particle filtering που περιγράφεται στο [Zeng, Ma, 2002] αποτελεί μια από τις προσπάθειες για μείωση του υπολογιστικού κόστους λόγω ανάγκης χρήσης μεγάλου αριθμού δειγμάτων σε υψηλής διάστασης χώρους κατάστασης. Στην πραγματικότητα, η παρακολούθηση ενός αντικειμένου σε σύνθετο περιβάλλον μπορεί να αντιμετωπιστεί ως εύρεση του ολικού μέγιστου της posterior density. Όταν η prior density είναι ανακριβής, η πιθανότητα θα μπορούσε να είναι κυρίαρχη posterior density. Τα δειγματοληπτημένα particles στο κλασικό particle filtering σταθμίζονται άμεσα από στοιχεία παρατήρησης. Στη μέθοδο active particle filtering αντίθετα, τα particles οδηγούνται στα τοπικά μέγιστα της πιθανότητας πριν σταθμιστούν. Τα particles της μεθόδου αυτής αποκαλούνται ενεργά (active) γιατί έχουν πιθανότητες να κινηθούν προς μεγαλύτερα βάρη. Η αποδοτικότητα κάθε particle βελτιώνεται και ο απαιτούμενος αριθμός τους για την εύρεση του ολικού μέγιστου της posterior density μειώνεται σημαντικά. Συγκεκριμένα, ο αριθμός των particles του active particle filtering βασίζεται περισσότερο στο βαθμό συνθετότητας του περιβάλλοντος και στο fitting range κάθε particle παρά στο μέγεθος του χώρου διάταξης (configuration space) του μοντέλου και εκεί ακριβώς βρίσκεται και η διαφορά του αλγορίθμου αυτού σε σχέση με άλλες μεθόδους particle filtering.

Στο [Pantrigo et al., 2005] παρουσιάζεται ένας αλγόριθμος, που ονομάζεται αλγόριθμος Local Search Particle Filter (LSPF) και εφαρμόζεται σε προβλήματα εκτίμησης σε ακολουθιακές διαδικασίες που μπορούν να εκφραστούν χρησιμοποιώντας το μοντέλο χώρου κατάστασης. Ο στόχος αυτού του αλγορίθμου είναι να βελτιώσει την αποτελεσματικότητα του βασικού αλγορίθμου particle filtering χρησιμοποιώντας μια διαδικασία τοπικής αναζήτησης. Αυτή η πρόταση είναι ιδιαίτερα κατάλληλη για εφαρμογές που απαιτούν εκτιμήσεις υψηλής ποιότητας. Ο αλγόριθμος LSPF συνδυάζει την τοπική αναζήτηση (local search (LS)) και το particle filtering (PF) σε δύο διαφορετικά στάδια:

- ✓ Στο στάδιο του particle filter, ένα σύνολο particles (λύση) διαδίδεται στο χρόνο και ανανεώνεται με μετρήσεις για να προκύψει ένα νέο σύνολο. Αυτό το στάδιο επικεντρώνεται στη χρονική εξέλιξη των καλύτερων λύσεων που βρέθηκαν στις προηγούμενες χρονικές στιγμές.

- ✓ Στο στάδιο της τοπικής αναζήτησης, επιλέγονται οι καλύτερες λύσεις από το σύνολο των particles και εφαρμόζεται μια διαδικασία τοπικής αναζήτησης στη γειτονιά τους. Αυτό το στάδιο στοχεύει στη βελτίωση της εκτίμησης του PF.

Τέλος, στο [Marron et al., 2005] παρουσιάζεται η μέθοδος του extended particle filtering (XPF), που τροποποιεί δύο στοιχεία του βασικού αλγορίθμου:

- ✓ Καταρχήν, στο κλασικό particle filtering δεν είναι δυνατό να εκτιμηθεί η εμφάνιση νέων αντικειμένων στο περιβάλλον επειδή όλες οι μετρήσεις που σχετίζονται με τα νέα αντικείμενα θα απορριπτονταν κατά το resampling αφού δε θα υπήρχαν particles παρόμοια με αυτές τις μετρήσεις. Για να λυθεί αυτό το πρόβλημα, εισήχθη ένα στάδιο εκ νέου αρχικοποίησης στο βασικό αλγόριθμο του PF για την εισαγωγή νέων particles απευθείας από τις μετρήσεις στο σύνολο δειγμάτων. Το στάδιο του resampling θα πρέπει επίσης να τροποποιηθεί για να επιτρέψει την ενσωμάτωση των νέων particles στο σύνολο δειγμάτων.
- ✓ Δεύτερον, για την παρακολούθηση πολλαπλών αντικειμένων, πρέπει να τροποποιηθεί το στάδιο του importance sampling επειδή η πιθανότητα πρέπει να υπολογιστεί ανάλογα με την ομοιότητα μεταξύ κάθε particle και των μετρήσεων από το αντίστοιχο αντικείμενο. Διαφορετικά, τα particles που αντιστοιχούν στο αντικείμενο μπορεί επίσης να απορριφθούν όπως αναφέρθηκε προηγουμένως. Για την επίλυση αυτού του προβλήματος, πρέπει να παρεμβληθεί μια τροποποίηση στη συνάρτηση για να υπολογίζονται τα βάρη των particles ανάλογα με τις αντίστοιχες μετρήσεις.

Σημειώνουμε ότι η παρακολούθηση πολλών αντικειμένων καθώς και μεταβλητού αριθμού αντικειμένων με PF μελετάται στο [Hue et al., 2002].

5.4.4. Επίδραση του αριθμού των particles

Έχει αποδειχθεί ότι ο αριθμός των particles είναι πολύ σημαντικός για την απόδοση του particle filtering. Απαιτούνται πολλά δείγματα για να διατηρηθεί μια ικανοποιητική αναπαράσταση της posterior κατανομής και ο αριθμός τους αυξάνεται εκθετικά με τη διάσταση του χώρου διατάξεων (configuration space) του μοντέλου [Zhang et al., 2004]. Σε διαφορετική περίπτωση είναι εύκολο το σύστημα να χάσει την τροχιά του στόχου και είναι δύσκολο να την ανακτήσει εξαιτίας της μείωσης των δειγμάτων στο χώρο κατάστασης. Επιπλέον απαιτείται ακριβής αρχικοποίηση του μοντέλου.

Η απόδοση της παρακολούθησης, εκτός από τον αριθμό των particles, εξαρτάται και από το ρυθμό λήψης νέων δειγμάτων, ο οποίος με τη σειρά του εξαρτάται από τα χαρακτηριστικά του υπό παρακολούθηση αντικειμένου. Ο αριθμός των particles και ο ρυθμός επεξεργασίας επηρεάζουν άμεσα την πολυπλοκότητα του υλικού (hardware) του υλοποιημένου φίλτρου. Για αποδοτικές υλοποιήσεις υλικού σε εφαρμογές παρακολούθησης είναι επιτακτική η ανάγκη χρήση όσο το δυνατόν λιγότερων particles κάθε στιγμή. Η ελαχιστοποίηση του αριθμού των particles μπορεί να πραγματοποιηθεί μεταβάλλοντας δυναμικά τον αριθμό των particles χρησιμοποιώντας τις επεξεργαζόμενες παρατηρήσεις.

Μερικές υπάρχουσες τεχνικές αλλαγής του αριθμού των particles κατά το φιλτράρισμα, όπως παρουσιάζονται στο [Bolic et al., 2002], περιλαμβάνουν τη

δειγματοληψία με την απόσταση Kullback - Leibler (Kullback - Leibler distance KLD sampling) και την προσαρμογή βάσει της πιθανότητας (likelihood based adaptation). Αυτές οι μέθοδοι βασίζονται στην ιδέα της διάδοσης μικρού αριθμού particles όταν η συνάρτηση πυκνότητας πιθανότητας είναι συγκεντρωμένη σε μια μικρή περιοχή (για παράδειγμα υπό την παρουσία μικρών αβεβαιοτήτων) του χώρου κατάστασης και τη διάδοση μεγάλου αριθμού particles σε περιπτώσεις μεγάλων αβεβαιοτήτων. Για υλοποιήσεις υλικού, η μέθοδος KLD είναι υπολογιστικά απαιτητική. Στη μέθοδο προσαρμογής βάσει πιθανοτήτων, παράγονται νέα particles μέχρι το άθροισμα των μη-κανονικοποιημένων πιθανοτήτων να ξεπεράσει ένα προκαθορισμένο κατώφλι. Με τον περιορισμό του ελάχιστου και μέγιστου αριθμού των particles, η μέθοδος αυτή είναι πολύ πιο απλή στην υλοποίηση. Ένα πρόβλημα που παρουσιάζεται είναι ότι κάποια βάρη μπορεί να είναι τόσο μεγάλα ώστε να αποτρέπουν τη δημιουργία νέων particles, γεγονός που οδηγεί σε ανακριβείς εκτιμήσεις. Ένα άλλο πρόβλημα της προσαρμογής βάσει πιθανότητας είναι η εισαγωγή εξαρτήσεων δεδομένων στα στάδια της δειγματοληψία και importance steps καθιστώντας την παράλληλη υλοποίηση υλικού δυσκολότερη. Ένα άλλο σημαντικό ζήτημα είναι το resampling των particles του φίλτρου. Υπάρχουν διάφορες μέθοδοι για να εκτιμηθεί αν το resampling είναι απαραίτητο. Συνήθως χρησιμοποιείται η διασπορά των βαρών, μέθοδος που εισάγει όμως επιπρόσθετες υπολογιστικές απαιτήσεις. Στο [Bolic et al., 2002] προτείνεται μια μέθοδος που κρίνει αν είναι απαραίτητο το resampling βάσει του τρέχοντος αριθμού των particles.

5.5. Εφαρμογή particle filtering στην παρακολούθηση χειρονομιών

5.5.1. Εισαγωγή

Ο αλγόριθμος του particle filtering επιτρέπει, όπως έχει ήδη αναφερθεί, τη μοντελοποίηση μιας στοχαστικής διαδικασίας με μια αυθαίρετη συνάρτηση πυκνότητας πιθανότητας (probability density function (pdf)), προσεγγίζοντας την αριθμητικά με ένα σύνολο σημείων που ονομάζονται particles στο χώρο κατάστασης (state-space). Ένας βασικός στόχος του particle filtering είναι η ακριβής παρακολούθηση (tracking) μιας μεταβλητής όπως αυτή εξελίσσεται στο χρόνο. Ένα παράδειγμα είναι η παρακολούθηση αντικειμένων σε ακολουθίες εικόνων (video) για τον προσδιορισμό των κινηματικών τους χαρακτηριστικών.

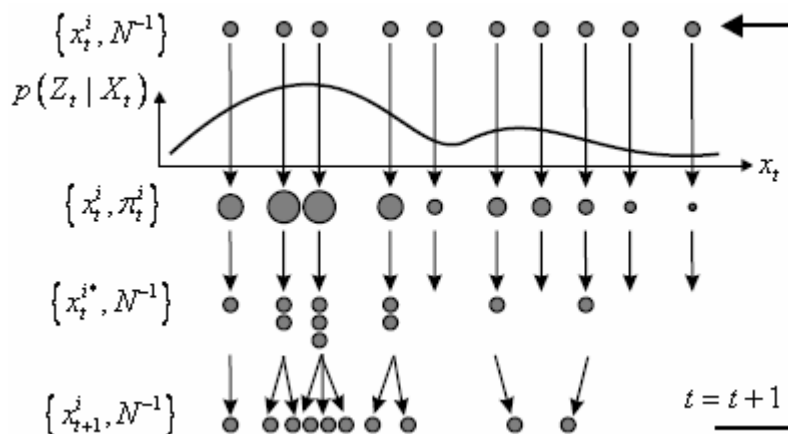
5.5.2. Αλγόριθμος particle filtering για παρακολούθηση κίνησης

Τα βήματα του αλγορίθμου για παρακολούθηση ενός στόχου είναι:

1. Αρχικοποίηση: Δημιουργείται ένα σύνολο N δειγμάτων (particles) από το αρχικό διάνυσμα κατάστασης x_0 . Τα βάρη των δειγμάτων είναι σταθερά και ίσα με $w_i = 1/N$. Κάθε δείγμα εκπροσωπεί μια υποθετική κατάσταση του αντικειμένου με αντίστοιχη διακριτή πιθανότητα δειγματοληψίας w_i . Η

κατάσταση του παρακολουθούμενου αντικειμένου περιλαμβάνει χαρακτηριστικά όπως η θέση του και ο προσανατολισμός του.

2. Πρόβλεψη-Εξέλιξη: Προβλέπεται η κατάσταση x_k^i κάθε δειγματος με ένα μοντέλο εξέλιξης. Για παρακολούθηση αντικειμένων μπορούμε να χρησιμοποιήσουμε ένα απλό μοντέλο (σταθερή θέση, σταθερή ταχύτητα, σταθερή επιτάχυνση κλπ) ή ένα μοντέλο μάθησης. Κάθε particle i , κινείται σύμφωνα με το μοντέλο κατάστασης και η εξέλιξή του σχετίζεται με την τιμή του βάρους του όπως αυτό καθορίζεται από τη συνάρτηση πιθανότητας $p(x_k | x_{k-1}^i)$.
3. Αναπροσαρμογή βαρών: Ανανεώνεται το βάρος κάθε δειγματος, ανάλογα με τις μετρήσεις. Η ανανέωση γίνεται πολλαπλασιάζοντας το βάρος w_k^i κάθε δειγματος με την πιθανότητα παρατήρησης του z_k δεδομένης της κατάστασης x_k^i . Το νέο βάρος είναι δηλαδή $w_k^i = p(z_k | x_k^i) w_k^i$. Το διάνυσμα μετρήσεων z_k κάθε χρονική στιγμή k προκύπτει από το σύστημα. Στη συνέχεια τα βάρη κανονικοποιούνται έτσι ώστε $\sum_i w_k^i = 1$.
4. Εκτίμηση: Η τιμή της ποσότητας N_{eff} , που ορίζεται στην ενότητα 5.4.2, δίνει μια εκτίμηση της κατάστασης του συστήματος.
5. Resampling: Επιλέγεται, εφόσον χρειάζεται βάσει της εκτίμησης, ένα νέο σύνολο N δειγμάτων, λαμβάνοντας υπόψη ότι όσο μεγαλύτερο το βάρος ενός δειγματος τόσο περισσότερες πιθανότητες έχει να επιλεγεί. Ένα δείγμα με μικρό βάρος απομακρύνεται. Τέλος, τα βάρη των νέων δειγμάτων τίθενται ίσα με $w_i = 1/N$. Χωρίς resampling η διασπορά των βαρών αυξάνει στοχαστικά στο χρόνο και έτσι οδηγούμαστε σε χαμηλή αποτελεσματικότητα. Στη συνέχεια επαναλαμβάνεται η διαδικασία από το βήμα 2.



Εικόνα 5.6: Αλγόριθμος particle filtering

5.5.3. Παρακολούθηση κίνησης χεριού με particle filtering

Η τεχνική του particle filtering έχει χρησιμοποιηθεί σε αρκετά συστήματα παρακολούθησης της κίνησης των χεριών, κυρίως για εφαρμογές αλληλεπίδρασης ανθρώπου-υπολογιστή και είναι η τεχνική που χρησιμοποιήθηκε για την παρακολούθηση του χεριού στο σύστημα που υλοποιήθηκε στην παρούσα εργασία. Μερικά παραδείγματα παρουσιάζονται στα [Bray et al., 2004], [Brethes et al., 2004], [Bretzner et al., 2002], [Chang et al., 2005], [Chateau et al., 2005], [Laptev, Lindeberg, 2001], [Lee et al., 2002], [Liang et al., 2005], [MacCormick, Isard, 2000], [Pantrigo et al., 2005], [Shan, Wei, Qiu, Tan, 2004], [Shan, Wei, Tan, Ojardias, 2004], [Vacavant, Chateau, 2005].

Η παρακολούθηση του χεριού αποτελεί ένα πρόβλημα πολλών βαθμών ελευθερίας, λόγω της κατασκευής του χεριού. Γενικά, υπάρχουν δύο προσεγγίσεις που προτείνονται για την παρακολούθηση πολλών βαθμών ελευθερίας (βλ. [Chang et al., 2005]). Η μια προσέγγιση είναι βασισμένη στην εμφάνιση (appearance based) και εκτιμά την κίνηση άμεσα από τις εικόνες αντιστοιχώντας τα χαρακτηριστικά γνωρίσματα της εικόνας στο χώρο κατάστασης του αντικειμένου. Η δεύτερη προσέγγιση βασίζεται στη χρήση μοντέλου (model based), και υπολογίζει την κίνηση προβάλλοντας ένα τρισδιάστατο μοντέλο στην εικόνα και συγκρίνοντας έπειτα τις προβολές με τις παρατηρήσεις. Ένα πλεονέκτημα των πρώτων συστημάτων είναι ότι οι παρατηρήσεις μπορούν να επεξεργαστούν από αυθαίρετες οπτικές γωνίες. Εντούτοις, οι μεγάλες και υψηλής ανάλυσης εικόνες αναφοράς πρέπει να συλλεχθούν εκ των προτέρων για να λάβουμε μια ακριβή εκτίμηση. Επίσης, η χρήση σε μεγάλο σύνολο εικόνων είναι υπολογιστικά απαιτητική. Η δεύτερη προσέγγιση μπορεί να παρέχει μια ακριβή εκτίμηση όταν το τρισδιάστατο μοντέλο αρχικοποιείται καλά, αλλά η αναζήτηση σε ένα χώρο μεγάλης διάστασης είναι αρκετά σύνθετη.

Στη βασισμένη σε μοντέλο προσέγγιση η κίνηση ανακτάται από την τρισδιάστατη διάταξη με τη μέγιστη ομοιότητα. Αυτό το πρόβλημα έχει διατυπωθεί ως πρόβλημα βελτιστοποίησης, και μπορεί επίσης να αντιμετωπιστεί σε ένα πιθανολογικό πλαίσιο σαν εκτίμηση κατάστασης ενός δυναμικού συστήματος. Δεδομένου ότι οι λύσεις κλειστής μορφής ενός ιδιαίτερα μη γραμμικού δυναμικού συστήματος είναι intractable, εισήχθησαν οι ακολουθιακές μέθοδοι Monte Carlo (sequential Monte Carlo methods) όπως το particle filtering για να λύσουν αυτό το πρόβλημα.

Εντούτοις, οι περισσότερες από τις βασισμένες στο particle filtering μεθόδους παρακολούθησης χεριού πολλών dof χρησιμοποιούν μόνο τις οπτικές πληροφορίες από τις προηγούμενες χρονικές στιγμές. Αν και η εφαρμογή των εκτιμητών κατάστασης σε ένα δυναμικό σύστημα έχει αποδειχθεί αποτελεσματική για την οπτική παρακολούθηση, έχει ορισμένους περιορισμούς. Κατ' αρχάς, χρησιμοποιούνται μόνο οι αρχικές καταστάσεις, οπότε η διαδικασία παρακολούθησης μπορεί να παγιδευτεί σε τοπικά ελάχιστα. Δεύτερον, οι υπάρχουσες μέθοδοι εκτίμησης κατάστασης δυσκολεύονται να εφαρμόσουν τις γνωστές πληροφορίες εμφάνισης ενός αντικειμένου για να βελτιώσουν την απόδοση της παρακολούθησης, ακόμα και όταν τέτοιες πληροφορίες είναι εύκολο να αποκτηθούν. Για να ξεπεραστούν αυτές οι δυσκολίες, προτείνεται στο [Chang et al., 2005] η εκτίμηση κατάστασης ενός δυναμικού συστήματος υπό την υπόθεση ότι υπάρχουν μερικοί γνωστοί ελκυστές (attractors), εκτός από την αρχική κατάσταση, στο χώρο κατάστασης. Ως ελκυστές αναφέρονται τα διανύσματα του χώρου

κατάστασης των οποίων οι παρατηρήσεις είναι γνωστές. Για ένα πρόβλημα οπτικής παρακολούθησης, οι ελκυστές είναι εικόνες αναφοράς των αντικειμένων με γνωστή κατάσταση κίνησης και χρησιμεύουν ως εκ των προτέρων γνώση για να καθοδηγήσουν την παρακολούθηση σε μεγάλης διάστασης χώρους.

Όπως έχει αναφερθεί και σε προηγούμενα κεφάλαια, το χρώμα του δέρματος είναι το διακριτικό στοιχείο του χεριού, όχι μόνο γιατί αποτελεί χαρακτηριστικό του γνώρισμα, αλλά και γιατί είναι αμετάβλητο στην κλίμακα και την περιστροφή. Επειδή λοιπόν τα περιγράμματα και οι μορφές του χεριού ποικίλλουν πολύ κατά τη διάρκεια της φυσικής κίνησης του χεριού, στις περισσότερες εφαρμογές επιλέγεται και στη μέθοδο του particle filtering το χρώμα του δέρματος ως χαρακτηριστικό γνώρισμα του χεριού. Σε κάποιες εφαρμογές, κυρίως όταν το χέρι είναι το μοναδικό κινούμενο αντικείμενο στην εικόνα, χρησιμοποιείται και το στοιχείο της κίνησης ως δεύτερο χαρακτηριστικό γνώρισμα.

Σε εφαρμογές όπου αναγνωρίζονται συγκεκριμένες διατάξεις του χεριού, όπως στο [Laptev, Lindeberg, 2001], θεωρούμε ότι η μεταβλητή κατάστασης X του συστήματος περιλαμβάνει χαρακτηριστικά όπως είναι η θέση $(x; y)$, το μέγεθος s , ο προσανατολισμός a και η μορφή λ του μοντέλου του χεριού, δηλαδή $X = (x, y, s, a, \lambda)$. Στο [Laptev, Lindeberg, 2001] εκτός του X χρησιμοποιείται και το διάνυσμα Y της χρονικής παραγώγου των τεσσάρων πρώτων μεταβλητών, δηλαδή $Y = (\dot{x}, \dot{y}, \dot{s}, \dot{a})$. Όταν αρχίζει η παρακολούθηση του στόχου, όλα τα δείγματα αρχικά κατανέμονται ομοιόμορφα στους χώρους των παραμέτρων X και Y . Μετά από κάθε βήμα του particle filtering, εκτιμάται η καλύτερη υπόθεση για ένα χέρι επιλέγοντας αρχικά την πιο πιθανή διάταξη του χεριού και έπειτα υπολογίζοντας το μέσο όρο της $p(X_t, Y_t | \tilde{I}_t)$ για εκείνη τη διάταξη, όπου \tilde{I}_t είναι οι παρατηρήσεις μέχρι τη στιγμή t . Η διάταξη i επιλέγεται εάν $w_i = \max_j (w_j)$, όπου w_j είναι το άθροισμα των βαρών όλων των δειγμάτων με κατάσταση j . Κατόπιν, οι συνεχείς παράμετροι εκτιμώνται με τον υπολογισμό ενός σταθμισμένου μέσου όρου όλων των δειγμάτων με κατάσταση i .

Για την υλοποίηση του συστήματος παρακολούθησης και αναγνώρισης χειρονομιών της παρούσας εργασίας εφαρμόστηκε η μέθοδος του particle filtering στο στάδιο της παρακολούθησης του κινούμενου χεριού. Το διάνυσμα μετρήσεων αποτελείται από δύο στοιχεία: το πρώτο είναι οι πιθανότητες των pixels να ανήκουν σε περιοχή του χεριού και το δεύτερο είναι μια απλοϊκή πληροφορία εκτίμησης κίνησης. Περισσότερες λεπτομέρειες δίνονται στο κεφάλαιο 7, όπου περιγράφεται αναλυτικά το σύστημα.

6

Σύμπτωση Προτύπων (Template Matching)

6.1. Εισαγωγή

Η σύμπτωση προτύπων (template matching) αποτελεί την τρίτη βασική τεχνική που χρησιμοποιεί το σύστημα ανίχνευσης, παρακολούθησης και αναγνώρισης χειρονομιών της παρούσας εργασίας και με αυτή ασχολείται το κεφάλαιο αυτό. Πρόκειται για την πιο απλοϊκή μορφή αναγνώρισης προτύπων (pattern recognition). Στην τεχνική αυτή αποθηκεύεται ένα ίχνος ή μάσκα ή πρωτότυπο (template) για κάθε διαφορετικό πρότυπο. Το πρότυπο εισόδου συγκρίνεται με καθένα από τα πρωτότυπα και η ταξινόμηση βασίζεται σε ένα προαποφασισμένο κριτήριο ομοιότητας. Η απόφαση λαμβάνεται για αντιστοίχιση της εισόδου σε εκείνο το πρωτότυπο για το οποίο η ομοιότητα είναι μεγαλύτερη. Η τεχνική αυτή χρησιμοποιήθηκε παλιότερα (1970) για αναγνώριση τυπωμένων κειμένων με τυποποιημένους χαρακτήρες. Πολλές φορές είναι δύσκολη η επιλογή "καλών πρωτοτύπων", όπως και "καλών κριτηρίων ομοιότητας". Η τεχνική αυτή παράγει τα καλύτερα αποτελέσματα όταν οι τάξεις που αντιπροσωπεύουν τα πρωτότυπα είναι ισχυρά διαφοροποιημένες και υπάρχει μεγάλη συνέπεια στα πρωτότυπα.

Η σύμπτωση προτύπων έχει πολυάριθμες σημαντικές εφαρμογές σε πεδία όπως η επεξεργασία εικόνας, η ανάκτηση πληροφοριών βάσει περιεχομένου από βάσεις δεδομένων εικόνων, η επεξεργασία δακτυλικών αποτυπωμάτων, η οπτική αναγνώριση χαρακτήρων, τα γεωγραφικά συστήματα πληροφοριών, η επεξεργασία αεροφωτογραφιών και αστρονομικών εικόνων, και η έρευνα για γνωστές δομές (όπως οι πρωτεΐνες) σε τρισδιάστατα μοντέλα βιολογικών ιών.

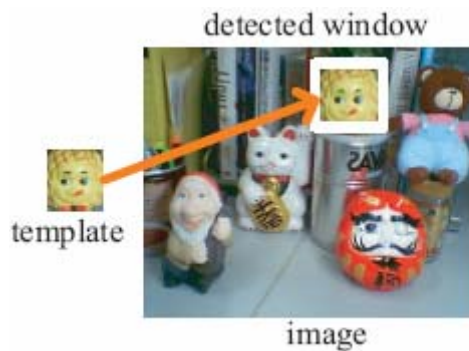
6.2. Σύμπτωση προτύπων και επεξεργασία εικόνων

6.2.1. Το πρόβλημα της σύμπτωσης προτύπων

Η αναζήτηση και ο εντοπισμός σχημάτων σε εικόνες και βίντεο είναι σημαντικό τμήμα πολλών συστημάτων όρασης υπολογιστών. Η σύμπτωση προτύπων είναι η πιο διαδεδομένη προσέγγιση στο πρόβλημα αυτό και χρησιμοποιείται από την

πρόωρη έρευνα στην αναγνώριση προτύπων. Η σύμπτωση προτύπων είναι λοιπόν ένα κλασικό πρόβλημα στην ανάλυση εικόνων: δεδομένης μιας εικόνας αναφοράς ενός αντικειμένου, εξετάζεται αν το αντικείμενο αυτό υπάρχει στην υπό ανάλυση εικόνα, και αν υπάρχει εντοπίζεται η θέση του. Η σύμπτωση προτύπων έχει χρησιμοποιηθεί ευρέως στην επεξεργασία εικόνων, στην οπτική παρακολούθηση αντικειμένων, στην αναγνώριση προτύπων, στην όραση υπολογιστών και στη συμπίεση εικόνων και video.

Στην παρακάτω εικόνα φαίνεται ένα παράδειγμα εφαρμογής της σύμπτωσης προτύπων, όπου σε μια εικόνα αναζητείται η πιο όμοια με ένα προκαθορισμένο πρωτότυπο περιοχή:



Εικόνα 6.1: Παράδειγμα εφαρμογής της σύμπτωσης προτύπων (από το [Kawanishi et al., 2004])

Ουσιαστικά, το πρόβλημα της σύμπτωσης προτύπων είναι ένα πρόβλημα εύρεσης του κοντινότερου γείτονα, το οποίο μπορεί να επιλυθεί από τον απλό αλγόριθμο πλήρους αναζήτησης (full-search algorithm).

6.2.2. Τεχνικές σύμπτωσης προτύπων

Η κλασική διαδικασία της σύμπτωσης προτύπων περιλαμβάνει τον υπολογισμό της συσχέτισης (cross-correlation) του πρωτοτύπου (template) με την εικόνα και τον καθορισμό της απόκλισης βάσει ενός μέτρου της μεταξύ τους ομοιότητας. Σε γενικευμένους αλγορίθμους υπολογίζεται η συσχέτιση μεταξύ κάθε θέσης της εικόνας και κάθε περιστροφής και γενικότερα ενός αφινικού μετασχηματισμού του πρωτοτύπου, ώστε να συμπεριληφθεί η περίπτωση που το πρωτότυπο υπάρχει περιστραμμένο στην υπό εξέταση εικόνα και σε διαφορετική κλίμακα. Στη συνέχεια παρουσιάζονται αλγόριθμοι και των δύο κατηγοριών.

Το μειονέκτημα της σύμπτωσης προτύπων είναι το υψηλό υπολογιστικό κόστος. Στην αναζήτηση ενός αντικειμένου, πολλές μικρές περιοχές μιας εικόνας (παράθυρα) που έχουν το ίδιο μέγεθος με το πρωτότυπο συγκρίνονται με το πρωτότυπο.

Ο βασικός αλγόριθμος σύμπτωσης προτύπων συνίσταται στην ολισθήση του πρωτοτύπου στην περιοχή αναζήτησης και στον υπολογισμό σε κάθε θέση της τιμής μιας συνάρτησης παραμόρφωσης ή συσχέτισης που υπολογίζει το βαθμό ανομοιότητας ή ομοιότητας μεταξύ του πρωτοτύπου και της εικόνας. Κατόπιν, η θέση ελάχιστης παραμόρφωσης ή μέγιστης συσχέτισης, λαμβάνεται ως ένδειξη της

εμφάνισης του πρωτοτύπου στην υπό εξέταση εικόνα, με ένα κατώτατο όριο στο μέτρο ομοιότητας/ανομοιότητας, που ρυθμίζεται σύμφωνα με τις απαιτήσεις της εφαρμογής, ώστε να απορρίπτονται τα μη βέλτιστα ταιριάσματα. Τα χαρακτηριστικά μέτρα παραμόρφωσης που χρησιμοποιούνται στους αλγορίθμους σύμπτωσης προτύπων είναι το άθροισμα των απόλυτων διαφορών (sum of absolute differences - SAD) και το άθροισμα των τετραγώνων των διαφορών (sum of square differences - SSD), ενώ η ομαλοποιημένη συσχέτιση (normalized cross-correlation - NCC) είναι το δημοφιλέστερο μέτρο συσχέτισης.

Θεωρούμε T το πρωτότυπο, $W^{(u,v)}$ ένα παράθυρο της εικόνας στη θέση (u,v) και $d(T, W^{(u,v)})$ τη διαφορά μεταξύ των T και $W^{(u,v)}$. Το άθροισμα των απολύτων διαφορών (sum of absolute distances SAD) ορίζεται στο [Kawanishi et al., 2004] ως:

$$SAD : d(T, W^{(u,v)}) = \sum_{p=0}^{N-1} |T^p - W^p|$$

όπου T^p και W^p είναι οι χρωματικές τιμές του p -pixel του T και του $W^{(u,v)}$ αντίστοιχα και N είναι ο αριθμός των pixels του T και του $W^{(u,v)}$. Το άθροισμα των τετραγώνων των διαφορών είναι επίσης διαθέσιμο για χρήση. Αν το μέγεθος του πρωτοτύπου είναι $(N_1 \times N_2)$ και της εικόνας $(M_1 \times M_2)$, τότε ο αριθμός των παραθύρων είναι $(M_1 - N_1) \times (M_2 - N_2)$. Οι συνήθεις μέθοδοι σύμπτωσης προτύπων υπολογίζουν τη διαφορά $d(T, W^{(u,v)})$ για κάθε παράθυρο ξεχωριστά και βρίσκουν το ελάχιστο.

Δεδομένης τώρα μιας $M_1 \times M_2$ εικόνας I και μιας $N_1 \times N_2$ μικρότερης εικόνας πρωτοτύπου T , το σύννηθες μέτρο παραμόρφωσης τετραγωνικού σφάλματος του ταιριάσματος είναι, όπως αναφέρεται στο [Yang et al., 2004]:

$$D(u, v) = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} (T(i, j) - I(u+i, v+j))^2$$

όπου $(u, v) \in S = \{(u, v) \mid 0 \leq u \leq M_1 - N_1 + 1, 0 \leq v \leq M_2 - N_2 + 1\}$ και S είναι το σύνολο όλων των πιθανών θέσεων ταιριάσματος. Στόχος της σύμπτωσης προτύπων είναι προφανώς η εύρεση της θέσης καλύτερου ταιριάσματος:

$$(\bar{u}, \bar{v}) = \arg \min_{(u,v) \in S} D(u, v)$$

Όσον αφορά τη δεύτερη κατηγορία μέτρων σφάλματος όπου χρησιμοποιείται η συσχέτιση, αν υποθέσουμε ότι το t είναι ένα πρωτότυπο που ανιχνεύεται σε μια εικόνα f τότε η σύμπτωση προτύπων με ομαλοποιημένη συσχέτιση υπολογίζει την ακόλουθη ποσότητα σε κάθε σημείο (u, v) της εικόνας, όπως αναφέρεται στο [Schweitzer et al., 2002]:

$$h(u, v) = \frac{\sum_{x,y} f(u+x, v+y)t(x, y)}{\sqrt{\sum_{x,y} f^2(u+x, v+y)}}$$

όπου τα αθροίσματα αφορούν όλες τις συντεταγμένες του πρωτοτύπου. Μια μεγάλη τιμή του $h(u, v)$ δείχνει ένα πιθανό ταίριασμα στη θέση (u, v) . Μπορεί να αποδειχθεί ότι ένα ταίριασμα που μεγιστοποιεί το $h(u, v)$ είναι ίδιο με το πρωτότυπο t όσον αφορά την κλίμακα.

Η ομαλοποιημένη συσχέτιση (Normalized Cross-Correlation NCC) και η μηδενικής μέσης τιμής ομαλοποιημένη συσχέτιση (Zero mean Normalized Cross-Correlation ZNCC) χρησιμοποιούνται ευρέως σαν συναρτήσεις ομοιότητας στη σύμπτωση προτύπων καθώς επίσης και στην ανάλυση κίνησης, την τρισδιάστατη όραση υπολογιστών, τις βιομηχανικές επιθεωρήσεις και πολλές άλλες εφαρμογές, δεδομένου ότι η διαδικασία κανονικοποίησης που ενσωματώνεται στις NCC και ZNCC επιτρέπει το χειρισμό των γραμμικών παραλλαγών της φωτεινότητας. Επιπλέον, χάρη στην αφαίρεση της μέσης έντασης, η συνάρτηση ZNCC είναι ακόμη πιο σθεναρή λύση από την NCC δεδομένου ότι μπορεί να χειριστεί και τις ομοιόμορφες παραλλαγές φωτεινότητας. Αφού οι NCC και ZNCC είναι υπολογιστικά δαπανηρές, έχουν προταθεί διάφορες μη εξαντλητικές τεχνικές που στοχεύουν στη μείωση του υπολογιστικού κόστους. Ωστόσο, οι μη εξαντλητικοί αλγόριθμοι δεν ερευνούν ολόκληρο το διάστημα αναζήτησης και ως εκ τούτου μπορούν να παγιδευτούν σε τοπικά μέγιστα, παράγοντας κατά συνέπεια μια μη βέλτιστη λύση. Στο [Di Stefano et al., 2004] προτείνεται ένας αλγόριθμος που βρίσκει ακριβώς την ίδια βέλτιστη λύση με μια διαδικασία σύμπτωσης προτύπων βασισμένη στη ZNCC αλλά με σημαντικά μειωμένο υπολογιστικό κόστος. Ο προτεινόμενος αλγόριθμος επεκτείνει την έννοια της περιορισμένης μερικής συσχέτισης (Bounded Partial Correlation BPC), η οποία προηγουμένως επινόησε για τη βασισμένη στην NCC διαδικασία σύμπτωσης προτύπων, στη συνάρτηση ZNCC.

Δεδομένου ότι ειδικά στις μεγάλου μεγέθους εικόνες και πρωτότυπα η διαδικασία μπορεί να είναι υπολογιστικά πολύ δαπανηρή, έχουν επινοηθεί πολυάριθμες τεχνικές που στοχεύουν στην επιτάχυνση της βασικής προσέγγισης, κυρίως εστιάζοντας στην εκτίμηση κίνησης και χρησιμοποιώντας τις ομοιότητες μεταξύ των διαδοχικών εικόνων (frames). Μεταξύ των γενικών τεχνικών (δηλαδή αυτών που είναι εφαρμόσιμες τόσο με τα μέτρα παραμόρφωσης όσο και με τα μέτρα συσχέτισης), οι σημαντικότερες είναι:

- ✓ η χρήση των σχεδίων πολλαπλής ανάλυσης (δηλαδή εντοπίζοντας ένα πρωτότυπο χαμηλής ανάλυσης στην εικόνα χαμηλής ανάλυσης και βελτιώνοντας έπειτα την αναζήτηση στα επίπεδα υψηλότερης ανάλυσης)
- ✓ υπο-δειγματοληψία της εικόνας και του πρωτοτύπου
- ✓ ταίριασμα δύο σταδίων (δηλαδή ταιριάζοντας ένα subtemplate πρώτα, και έπειτα ολόκληρο το πρωτότυπο μόνο στις βέλτιστες θέσεις)

Στο [Kawanishi et al., 2004] παρουσιάζεται ένα παράδειγμα γρήγορου αλγορίθμου συσχέτισης που ανήκει στην τελευταία κατηγορία. Ο αλγόριθμος ανίχνευσης διαδοχικής ομοιότητας (Sequential Similarity Detection Algorithm - SSDA) μειώνει το κόστος της σύγκρισης με το να τερματίζει τη σύγκριση εάν η συσσωρευμένη απόσταση υπερβαίνει ένα κατώτατο όριο. Η σύμπτωση προτύπων

δύο σταδίων μειώνει το υπολογιστικό κόστος με τη χρησιμοποίηση ενός τμήματος του πρωτοτύπου (subtemplate) και την εφαρμογή του υπολοίπου του πρωτοτύπου μόνο όταν η απόσταση μεταξύ του subtemplate και μιας μικρής περιοχής του παραθύρου (subwindow) δεν υπερβαίνει ένα κατώτατο όριο. Αυτές οι προσεγγίσεις απαιτούν ενός είδους σύγκριση σε κάθε παράθυρο της εικόνας. Έχει αναπτυχθεί μια αποδοτική στρατηγική αναζήτησης που υιοθετεί την περικοπή ανώτερου ορίου ομοιότητας, η οποία καλείται ενεργός αναζήτηση (Active Search). Η ενεργός αναζήτηση περικόπτει το διάστημα αναζήτησης παραβλέποντας παράθυρα των οποίων το ανώτερο όριο ομοιότητας είναι κατώτερο από το κατώφλι. Εντούτοις, επειδή χρησιμοποιείται μόνο ένα ιστόγραμμα χρώματος σαν χαρακτηριστικό γνώρισμα, οι γεωμετρικές πληροφορίες και οι αποστάσεις μεταξύ των χρωμάτων χάνονται. Επομένως, η ενεργός αναζήτηση παρουσιάζει μια δυσκολία στον ακριβή εντοπισμό της θέσης του αντικειμένου ή στο διαχωρισμό των αντικειμένων με παρόμοια ιστογράμματα χρώματος. Το χαρακτηριστικό γνώρισμα των ιστογραμμάτων χρώματος μπορεί να χρησιμοποιηθεί μόνο περιορισμένα στη σύμπτωση προτύπων.

Εντούτοις, οι τεχνικές των τριών παραπάνω κατηγοριών υπονοούν μια μη λεπτομερή διαδικασία αναζήτησης δεδομένου ότι δε συγκρίνουν την εικόνα πλήρους ανάλυσης με το πρωτότυπο πλήρους ανάλυσης σε κάθε θέση αναζήτησης. Κατά συνέπεια, δεν υπάρχει καμία εγγύηση ότι ο αλγόριθμος βρίσκει το ολικό ελάχιστο παραμόρφωσης (ή το μέγιστο συσχέτισης).

Από την άλλη, στην περίπτωση των μέτρων παραμόρφωσης, δύο ενδιαφέρουσες τεχνικές, που παρουσιάζονται στο [Di Stefano, Mattoccia, 2003] και ονομάζονται αλγόριθμος διαδοχικού αποκλεισμού (successive elimination algorithm SEA) και αποκλεισμός μερικής παραμόρφωσης (partial distortion elimination PDE) επιτρέπουν την αξιολογή επιτάχυνση του υπολογισμού που απαιτείται για μια διαδικασία εξαντλητικής αναζήτησης. Εκτός αυτού, και οι δύο μπορούν να θεωρηθούν ως βασικές τεχνικές, που ενσωματώνονται στις προηγούμενες αναφερθείσες μη λεπτομερείς μεθόδους. Η SEA στηρίζεται στη γρήγορη εκτίμηση ενός ελάχιστου ορίου για το μέτρο παραμόρφωσης: εάν η συνάρτηση ορίου υπερβαίνει το τρέχον ελάχιστο, η θέση αυτή μπορεί να παραβλεφθεί χωρίς να υπολογιστεί η αντίστοιχη παραμόρφωση. Η PDE συνίσταται στον τερματισμό της εκτίμησης του μέτρου παραμόρφωσης εάν αυτό υπερβαίνει το τρέχον ελάχιστο.

Στο [Di Stefano, Mattoccia, 2003] επεκτείνονται οι προσεγγίσεις SEA και PDE για την περίπτωση της συσχέτισης. Αυτό γίνεται καθορίζοντας πρώτα ένα άνω όριο για τη συνάρτηση NCC και ενσωματώνοντας στη συνέχεια την εκτίμηση του ορίου αυτού σε ένα σχέδιο μερικής συσχέτισης.

Η ευρεία χρήση της σύμπτωσης προτύπων στις εφαρμογές της επεξεργασίας εικόνων οφείλεται κυρίως στην ευκολία υλοποίησής της μαζί με τους πολλούς ταχείς αλγορίθμους, όπως αυτοί που αναφέρθηκαν, που μπορούν να χρησιμοποιηθούν για να επιταχύνουν τη διαδικασία ταιριάσματος για τις διάφορες εφαρμογές. Σε ένα μη ομαλό περιβάλλον όμως μπορούν να εμφανιστούν προβλήματα όπως θόρυβος ή μερικός αποκλεισμός του αντικειμένου ενδιαφέροντος από άλλα κατά τη διάρκεια της διαδικασίας σύμπτωσης προτύπων. Σε τέτοιες περιπτώσεις, τα κριτήρια SAD και SSD δεν είναι πλέον κατάλληλα γιατί δε λαμβάνουν υπόψη παραμέτρους όπως το θόρυβο ή το μερικό αποκλεισμό κατά τον υπολογισμό των μέτρων σφάλματος. Μια πιθανή λύση για αυτή την αδυναμία είναι η χρήση ενός σθεναρού κριτηρίου αντί των SAD ή SSD. Η τεχνική των M-

εκτιμητών, που παρουσιάζεται στο [Chen et al., 2003], είναι μια από τις δημοφιλέστερες μεθόδους για τη λύση του προβλήματος της σθεναρής εκτίμησης παραμέτρων και έχει εφαρμοστεί σε πολλές περιπτώσεις. Η βασική ιδέα της τεχνικής των M-εκτιμητών είναι ο περιορισμός της επιρροής των παραμέτρων που αναφέρθηκαν στο σφάλμα ταιριάσματος.

Σε όσα έχουν αναφερθεί μέχρι τώρα δεν έχει εξεταστεί η περίπτωση που το πρωτότυπο εμφανίζεται περιστραμμένο στην υπό εξέταση εικόνα. Στη συνέχεια εξετάζεται λοιπόν το πρόβλημα της εύρεσης της ακριβούς και της κατά προσέγγιση εμφάνισης ενός πρωτοτύπου σε μια εικόνα, όταν επιτρέπονται όλες οι πιθανές περιστροφές του πρωτοτύπου. Αυτό το πρόβλημα ονομάζεται αμετάβλητη της περιστροφής σύμπτωση προτύπων.

Το πρώτο πρόβλημα για σύμπτωση προτύπων ανεξάρτητη της περιστροφής είναι ο καθορισμός της έννοιας του ταιριάσματος. Αν η εικόνα και το πρωτότυπο θεωρηθούν πλέγματα, τότε ο ορισμός της έννοιας του ταιριάσματος δεν είναι τετριμμένος όταν το πρωτότυπο περιστραφεί, αφού από τη στιγμή που κάθε κελί του πλέγματος του πρωτοτύπου επικαλύπτεται με περισσότερα από ένα κελιά του πλέγματος της εικόνας και αντίστροφα δεν είναι ξεκάθαρο τι θα πρέπει να ταιριάζει με τι. Για το σκοπό αυτό έχει προταθεί ένα απλό μοντέλο, το μοντέλο ακριβούς ταιριάσματος (exact matching model), σύμφωνα με το οποίο:

- ✓ Το γεωμετρικό κέντρο του πρωτοτύπου πρέπει να ταυτίζεται με το γεωμετρικό κέντρο ενός κελιού της εικόνας (υπόθεση κέντρου-προς-κέντρο, center-to-center assumption)
- ✓ Τα κελιά της εικόνας που εμπλέκονται στο ταιρίασμα είναι αυτά των οποίων τα γεωμετρικά κέντρα επικαλύπτονται από το πρωτότυπο
- ✓ Κάθε κελί της εικόνας που εμπλέκεται σε ταιρίασμα θα πρέπει να ταιριάζει με το κελί του πρωτοτύπου που επικαλύπτει το κέντρο του

Μια επέκταση του μοντέλου αυτού, που ονομάζεται MinMax model και είναι καταλληλότερο για γκριζες εικόνες, αναφέρει ότι η τιμή κάθε κελιού της εικόνας που εμπλέκεται σε ταιρίασμα πρέπει να κυμαίνεται μεταξύ της ελάχιστης και μέγιστης τιμής των εννέα γειτονικών κελιών του αντίστοιχου κελιού του πρωτοτύπου.

Όταν τα αντικείμενα είναι περιστραμμένα στην εικόνα, οι μέθοδοι που έχουν αναφερθεί δεν μπορούν να χρησιμοποιηθούν και πρέπει να χρησιμοποιηθεί ένα σύνολο πρωτοτύπων σε διαφορετικούς προσανατολισμούς. Αυτή η διαδικασία δεν είναι φυσικά πρακτική για επεξεργασία σε πραγματικό χρόνο όταν η γωνία περιστροφής είναι αυθαίρετη ή δεν υπόκειται σε κάποιο περιορισμό. Σε τέτοιες εφαρμογές, μπορούν να χρησιμοποιηθούν μέθοδοι ανεξάρτητες από οποιαδήποτε περιστροφή, που βασίζονται στις ροπές (moments). Οι μέθοδοι αυτές εντούτοις, απαιτούν πολλούς υπολογισμούς και είναι ευαίσθητες στο θόρυβο. Στο [Choi, Kim, 2002] προτείνεται μια μέθοδος δύο σταδίων, ανεξάρτητη της περιστροφής. Στο πρώτο στάδιο επιλέγονται τα πιθανά ταιριάσματα χρησιμοποιώντας ένα χαρακτηριστικό γνώρισμα με χαμηλό υπολογιστικό κόστος. Αυτό το χαρακτηριστικό γνώρισμα ονομάζεται διανυσματικό άθροισμα των κυκλικών προβολών της υπο-εικόνας. Για να επιταχυνθεί η διαδικασία επιλογής πιθανών ταιριασμάτων, υιοθετήθηκε η υλοποίηση της συνέλιξης στο πεδίο συχνότητας. Στο δεύτερο στάδιο, η ανεξάρτητη περιστροφής σύμπτωση προτύπων εκτελείται μόνο στα πιθανά ταιριάσματα χρησιμοποιώντας τις ροπές Zernike (Zernike moments).

6.2.3. Σύμπτωση προτύπων και αναγνώριση χειρονομιών

Η σύμπτωση προτύπων χρησιμοποιείται, όπως προαναφέρθηκε, στην επεξεργασία εικόνων κυρίως για ανίχνευση αντικειμένων σε εικόνες. Είναι λοιπόν αναμενόμενο η τεχνική αυτή να βρίσκει εφαρμογή στην αναγνώριση ανθρώπινων χειρονομιών, και ιδιαίτερα στην αναγνώριση στατικών χειρομορφών. Επιπλέον, είναι αρκετά απλή στην εφαρμογή της. Έτσι, προτιμήθηκε για το στάδιο της αναγνώρισης των χειρομορφών στο σύστημα που υλοποιήθηκε για τις ανάγκες της παρούσας εργασίας.

Για την εφαρμογή της σύμπτωσης προτύπων είναι απαραίτητη η εκ των προτέρων αποθήκευση των πρωτοτύπων των χειρομορφών που είναι επιθυμητό να αναγνωριστούν. Στη συνέχεια, δεδομένης μιας εικόνας εκτελείται η σύμπτωση προτύπων για αυτά τα πρωτότυπα. Φυσικά, όταν τα πρωτότυπα είναι πολλά η διαδικασία απαιτεί πολύ υπολογιστικό χρόνο. Για την επιτάχυνσή της είναι απαραίτητο να εφαρμοστεί κάποιος από τους διαθέσιμους αλγορίθμους επιτάχυνσης. Εναλλακτικά, είναι χρήσιμο να ληφθεί υπόψη ότι δεν αναζητείται μια τυχαία περιοχή μέσα στη δεδομένη εικόνα, αλλά μια περιοχή που περιλαμβάνει ένα (ή και περισσότερα) χέρια. Όπως έχει αναφερθεί σε προηγούμενο κεφάλαιο, η ανίχνευση περιοχών ανθρώπινου δέρματος, όπως τα χέρια και το κεφάλι, είναι μια ευρέως διαδεδομένη εφαρμογή κατάτμησης εικόνων και μάλιστα δεν απαιτεί πολύ υπολογιστικό χρόνο για την εκτέλεσή της. Έτσι, για την επιτάχυνση της διαδικασίας, μπορεί να προηγηθεί το στάδιο της ανίχνευσης των περιοχών δέρματος και στη συνέχεια να εκτελεστεί η σύμπτωση προτύπων μεταξύ των πρωτοτύπων και των ανιχνευμένων περιοχών που περιέχουν τα χέρια. Με τον τρόπο αυτό, η αναζήτηση των πρωτοτύπων περιορίζεται στις περιοχές που ανιχνεύονται με τη διαδικασία της κατάτμησης και δεν γίνεται σε ολόκληρη την εικόνα.

Η αναγνώριση των χειρομορφών με χρήση της σύμπτωσης προτύπων πραγματοποιείται ακολουθώντας τη διαδικασία που περιγράφεται στη συνέχεια. Αφού ανιχνευτεί η περιοχή του χεριού, υπολογίζεται το κέντρο της, ο προσανατολισμός της, καθώς και οι διαστάσεις του χεριού (ύψος και πλάτος). Έπειτα, μεταβάλλεται το μέγεθος του πρωτοτύπου της χειρομορφής για την οποία εκτελείται η διαδικασία σύμφωνα με το ύψος και το πλάτος της περιοχής του χεριού, όπως αυτά έχουν υπολογιστεί. Το νέο πρωτότυπο περιστρέφεται κατάλληλα, έτσι ώστε να ευθυγραμμίζεται στην ίδια κατεύθυνση με την περιοχή δέρματος. Αφού υπολογιστεί και το κέντρο του πρωτοτύπου τα δύο κέντρα, της περιοχής του χεριού και του πρωτοτύπου, ταυτίζονται. Τέλος, υπολογίζεται η συσχέτιση μεταξύ της περιοχής του χεριού και του πρωτοτύπου. Η έξοδος του συστήματος είναι η χειρομορφή εκείνη για την οποία η συσχέτιση είναι μέγιστη. Η διαδικασία αυτή χρησιμοποιήθηκε στην αναγνώριση της διάταξης του χεριού στο σύστημά μας. Περισσότερες λεπτομέρειες για τον τρόπο υπολογισμού των διαφόρων μεγεθών που χρησιμοποιούνται θα δοθούν στο κεφάλαιο της υλοποίησης του συστήματος.

7

Υλοποίηση Συστήματος Ανίχνευσης, Παρακολούθησης και Αναγνώρισης Ανθρώπινων Χειρονομιών

7.1. Εισαγωγή

Στα προηγούμενα κεφάλαια παρουσιάστηκαν διάφορες μέθοδοι υπολογιστικής όρασης που χρησιμοποιούνται στην παρακολούθηση και αναγνώριση αντικειμένων. Στο κεφάλαιο αυτό περιγράφεται διεξοδικά το σύστημα ανίχνευσης, παρακολούθησης και αναγνώρισης ανθρώπινων χειρονομιών, που δημιουργήθηκε στα πλαίσια της παρούσας εργασίας, και χρησιμοποιεί τις μεθόδους αυτές. Παρουσιάζονται επίσης μερικά ενδεικτικά αποτελέσματα εφαρμογής του αλγορίθμου.

Το σύστημα αυτό υλοποιήθηκε στο Matlab και περιλαμβάνει την αναγνώριση έξι προκαθορισμένων χειρονομιών που εκτελούνται από ένα χέρι. Η παρακολούθηση του χεριού και η αναγνώριση των χειρονομιών δε γίνονται σε πραγματικό χρόνο, ωστόσο υπάρχει δυνατότητα επέκτασης του συστήματος σε λειτουργία πραγματικού χρόνου. Είναι επίσης εφικτή η παρακολούθηση περισσότερων του ενός χεριών, χωρίς όμως να αναγνωρίζονται οι συγκεκριμένες χειρονομίες που εκτελούνται. Υπάρχει όμως δυνατότητα επέκτασης και προς αυτή την κατεύθυνση.

7.2. Περιγραφή συστήματος

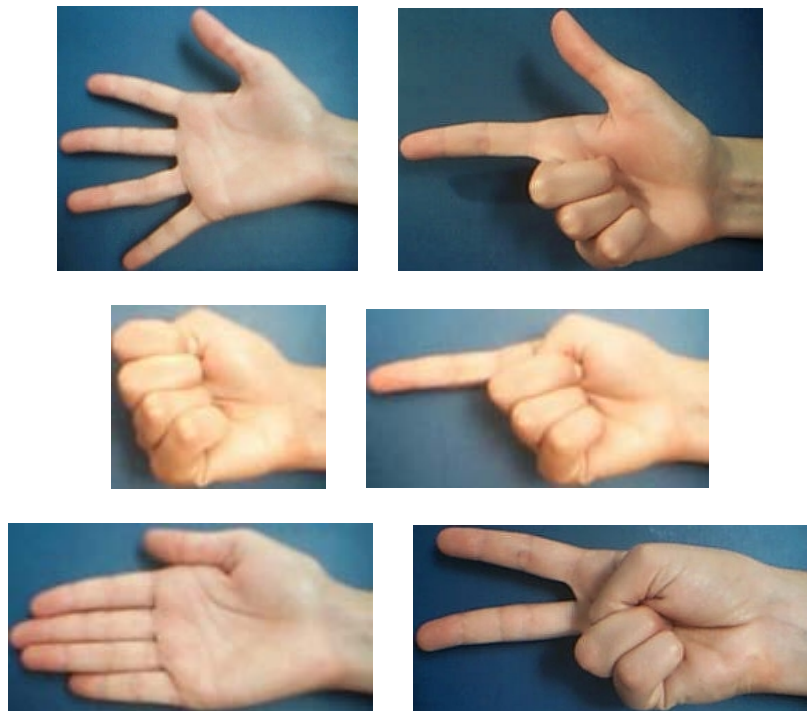
7.2.1. Γενική περιγραφή του συστήματος

Όπως ήδη αναφέρθηκε, το σύστημα που υλοποιήθηκε δεν είναι πραγματικού χρόνου. Αυτό σημαίνει ότι η παρακολούθηση και αναγνώριση των χειρονομιών δε γίνεται την ώρα που διεξάγονται. Αντίθετα, η κίνηση του χεριού καταγράφεται από μια κάμερα (στην καταγραφή των video που χρησιμοποιήθηκαν για τον έλεγχο του αλγορίθμου και την παραγωγή των αποτελεσμάτων που παρουσιάζονται στη

συνέχεια χρησιμοποιήθηκε webcam) και τα αποτελέσματα της αναγνώρισης προκύπτουν από την επεξεργασία του video με χρήση μεθόδων υπολογιστικής όρασης.

Τα βασικά δομικά στοιχεία του συστήματος είναι το μοντέλο χρώματος, που χρησιμοποιείται για την αναγνώριση των περιοχών κάθε frame του video που έχουν χρώμα δέρματος, η μέθοδος particle filtering, που χρησιμοποιείται για την παρακολούθηση του χεριού και η τεχνική της σύμπτωσης προτύπων (template matching), που χρησιμοποιείται για την αναγνώριση της συγκεκριμένης διάταξης του χεριού (χειρονομίας) σε κάθε frame.

Οι χειρομορφές που αναγνωρίζονται από το σύστημα απεικονίζονται στα παρακάτω σχήματα:



Εικόνα 7.1: Χειρομορφές που αναγνωρίζονται από το σύστημα

Σημειώνεται ότι οι εικόνες αυτές χρησιμοποιήθηκαν για την εξαγωγή των πρωτοτύπων των χειρομορφών για την εφαρμογή της τεχνικής της σύμπτωσης προτύπων.

Στη συνέχεια περιγράφονται συνοπτικά τα βήματα που ακολουθούνται για το σχεδιασμό του συστήματος ανίχνευσης, παρακολούθησης και αναγνώρισης χειρονομιών. Αναλυτικότερα τα βήματα περιγράφονται στις επόμενες παραγράφους.

- ✓ Πρώτο βήμα είναι η δημιουργία ενός μοντέλου χρώματος, για την ανίχνευση περιοχών δέρματος και συγκεκριμένα για την εύρεση της πιθανότητας του κάθε pixel να ανήκει σε περιοχή δέρματος. Η πιθανότητα αυτή χρησιμεύει στο επόμενο στάδιο, του particle filtering, ως ένα στοιχείο του διανύσματος μετρήσεων.

- ✓ Επόμενο βήμα είναι η αρχικοποίηση της μεθόδου particle filtering: επιλέγεται ένας αριθμός N particles, τα οποία τοποθετούνται τυχαία πάνω στο επίπεδο της εικόνας, που θεωρείται ως χώρος κατάστασης για την εφαρμογή της μεθόδου, και ανατίθεται στο καθένα βάρος $1/N$.
- ✓ Στη συνέχεια εφαρμόζεται ο αλγόριθμος του particle filtering που παρουσιάστηκε στην παράγραφο 5.5.2. για την παρακολούθηση του χεριού. Ως στοιχεία για το διάνυσμα μετρήσεων χρησιμοποιούνται οι πιθανότητες των pixels να ανήκουν σε περιοχή του χεριού και μια πληροφορία κίνησης για να περιοριστούν οι επιδράσεις της ύπαρξης κάποιου αντικειμένου παρόμοιου χρώματος με αυτό του χεριού στο φόντο.
- ✓ Το τελευταίο βήμα είναι η αναγνώριση της χειρομορφής, με χρήση της τεχνικής της σύμπτωσης προτύπων (template matching). Συγκεκριμένα, συγκρίνεται η εικόνα της περιοχής του χεριού που έχει ανιχνευτεί μέσω του particle filtering με καθένα από τα προκαθορισμένα πρωτότυπα χειρομορφών.

Τα δύο τελευταία βήματα εφαρμόζονται σε κάθε frame του video εισόδου (που έχει καταγραφεί από webcam με ρυθμό 30 frames/sec), ή εναλλακτικά, για εξοικονόμηση υπολογιστικού χρόνου, ανά δεδομένο αριθμό frames.

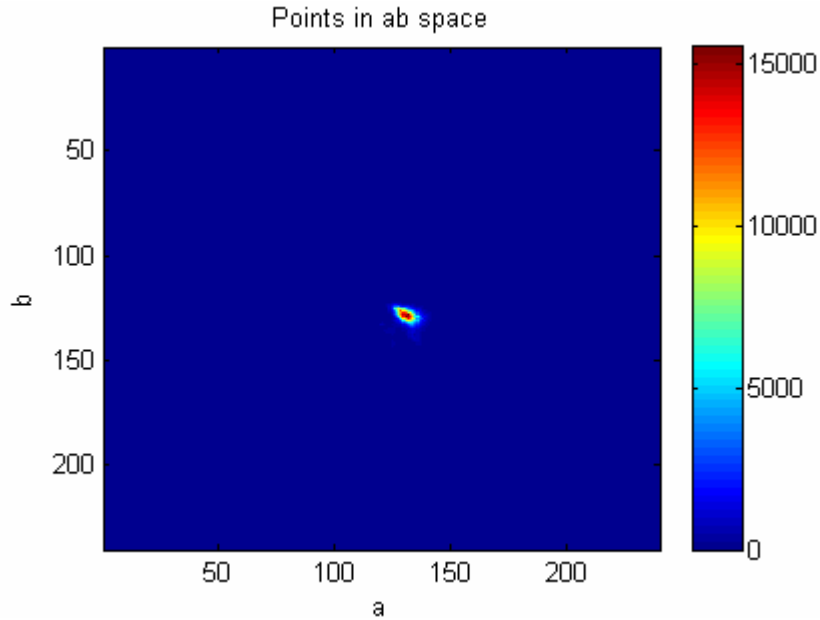
7.2.2. Δημιουργία μοντέλου χρώματος

Όπως αναφέρθηκε και στην ενότητα 4.3.4., πρώτο βήμα για τη δημιουργία ενός μοντέλου χρώματος είναι η συλλογή ενός συνόλου μάθησης, που αποτελείται από δείγματα pixels περιοχών του χρώματος που μας ενδιαφέρει, από διάφορες εικόνες. Οι RGB συνιστώσες του συνόλου αυτού μετατρέπονται σε χρωματικές συνιστώσες, αντιστοιχίζονται δηλαδή στο δισδιάστατο χώρο χρωματικότητας. Τέλος επιλέγεται ένα μοντέλο για την περιγραφή της κατανομής των δειγμάτων στο 2Δ χώρο χρωματικότητας. Το βασικό πλεονέκτημα της χρήσης του χώρου χρωματικότητας είναι ότι περιορίζεται η επίδραση από τις αλλαγές στο φωτισμό, αφού οι συνιστώσες χρωματικότητας είναι ανεξάρτητες της συνιστώσας φωτεινότητας.

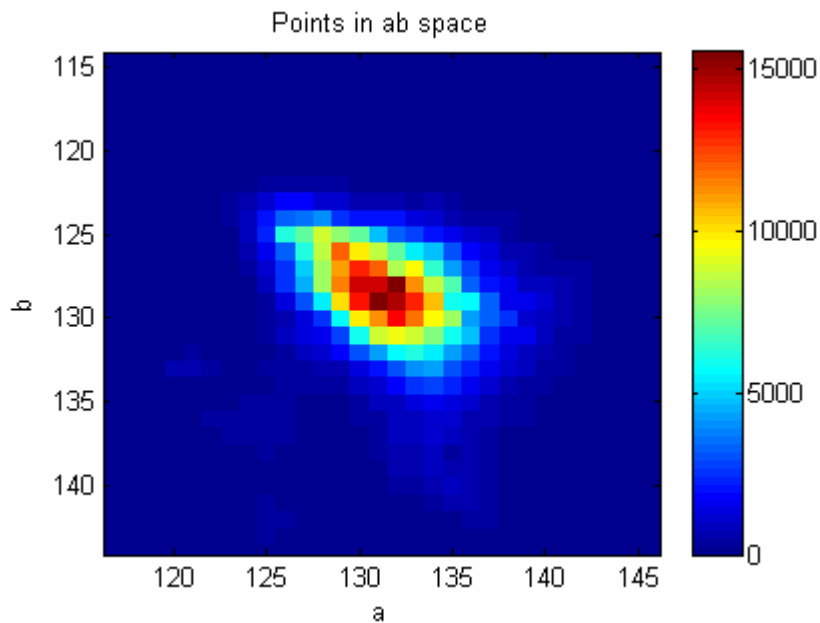
Το χρώμα που μας ενδιαφέρει στη δεδομένη εφαρμογή είναι το χρώμα του δέρματος. Το σύνολο μάθησης αποτελείται λοιπόν από εικόνες που περιέχουν το χρώμα αυτό και συγκεκριμένα από διάφορες εικόνες χεριών υπό διάφορες συνθήκες φωτισμού, ώστε να περιορίζεται ακόμη περισσότερο η επίδραση των αλλαγών στο φωτισμό. Οι διαστάσεις της εικόνας του συνόλου μάθησης που χρησιμοποιήθηκε είναι 970x810 pixels. Για ακόμα καλύτερα αποτελέσματα και γενίκευση του μοντέλου, προτείνεται επιπλέον ο συνδυασμός εικόνων από διαφορετικούς ανθρώπους.

Αφού δημιουργηθεί η εικόνα του συνόλου μάθησης, μετατρέπεται στο χώρο CIE-Lab και τα ζεύγη των τιμών (a, b) κάθε pixel τοποθετούνται στο επίπεδο των χρωματικών συνιστωσών (a, b) , ακολουθώντας τη διαδικασία που περιγράφεται στη συνέχεια. Αρχικά είναι απαραίτητο να μετρηθεί ο αριθμός εμφανίσεων κάθε ζεύγους (a, b) . Για το σκοπό αυτό μετατρέπονται οι τιμές (a, b) από το διάστημα $[-120, 120]$ που βρίσκονται (εξ' ορισμού του χρωματικού χώρου CIE-Lab) στο διάστημα $[1, n]$, όπου n είναι ο αριθμός των τάξεων που θέλουμε να χωρίσουμε το

χώρο. Οι νέες, διακριτοποιημένες, τιμές (a_{ind}, b_{ind}) χρησιμοποιούνται ως δείκτες σε έναν πίνακα μεγέθους $n \times n$, κάθε στοιχείο του οποίου περιλαμβάνει τον αριθμό εμφανίσεων κάθε ζεύγους (a_{ind}, b_{ind}) . Το αποτέλεσμα που προκύπτει είναι ένα "σύννεφο" σημείων, συγκεντρωμένο σε μια σχετικά μικρή περιοχή, όπως φαίνεται και στο παρακάτω σχήμα (όπου $n=240$):

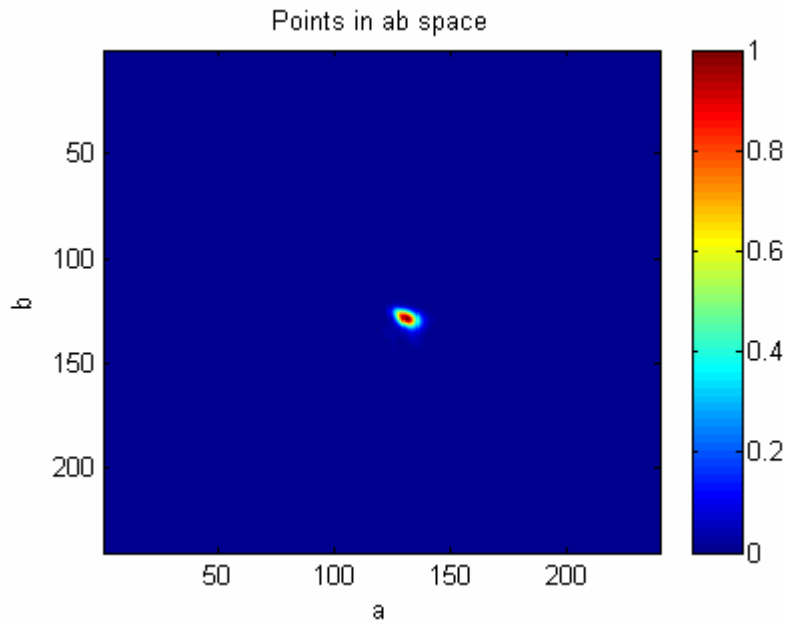


Εικόνα 7.2: Απεικόνιση των ζευγών (a_{ind}, b_{ind}) στο επίπεδο ab

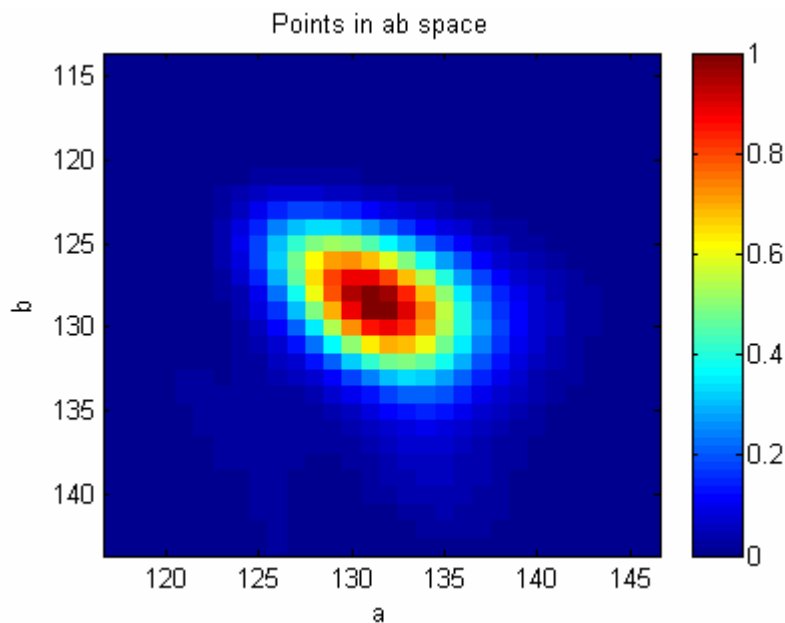


Εικόνα 7.3: Απεικόνιση των ζευγών (a_{ind}, b_{ind}) στο επίπεδο ab (λεπτομέρεια)

Για την ομαλοποίηση της εικόνας αυτής, εφαρμόζουμε συνέλιξη των σημείων αυτών με μια Gaussian και κανονικοποιούμε το αποτέλεσμα στο διάστημα $[0, 1]$. Τελικά προκύπτει το μοντέλο χρώματος που χρησιμοποιείται για την ανάθεση των πιθανοτήτων ύπαρξης χρώματος δέρματος σε μια εικόνα:



Εικόνα 7.4: Το μοντέλο χρώματος μετά τη συνέλιξη με Gaussian. Κάθε σημείο της γραφικής παράστασης δίνει την πιθανότητα το αντίστοιχο ζεύγος τιμών (a,b) να είναι δέρμα.

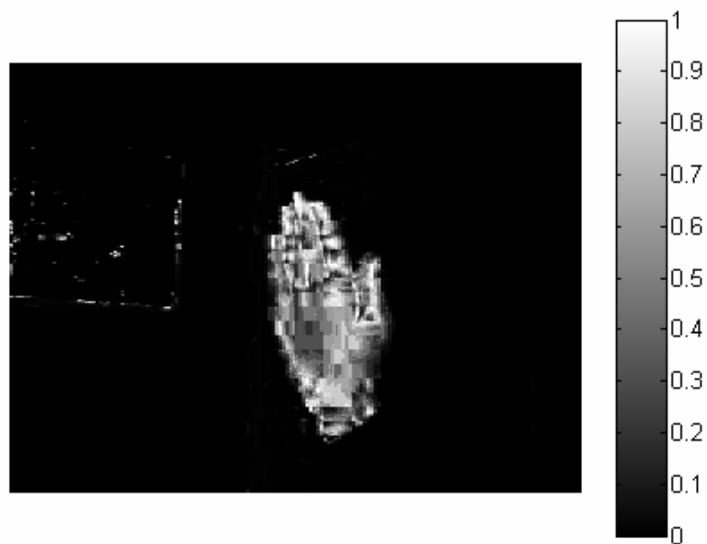


Εικόνα 7.5: Το μοντέλο χρώματος μετά τη συνέλιξη με Gaussian. Κάθε σημείο της γραφικής παράστασης δίνει την πιθανότητα το αντίστοιχο ζεύγος τιμών (a,b) να είναι δέρμα. (λεπτομέρεια)

Η εύρεση της πιθανότητας κάθε pixel μιας τυχαίας εικόνας να ανήκει σε περιοχή δέρματος γίνεται αντιστοιχίζοντας την τιμή (a,b) κάθε pixel της εικόνας στην αντίστοιχη τιμή από 0 έως 1 που δίνει το μοντέλο. Φυσικά, για να γίνει αυτή η αντιστοίχιση είναι απαραίτητη η μετατροπή των τιμών (a,b) από το διάστημα $[-120, 120]$ στο διάστημα $[1, n]$. Στα ακόλουθα σχήματα φαίνεται ένα τυχαίο frame ενός video που καταγράφηκε για τον έλεγχο του συστήματος αναγνώρισης χειρονομιών και η εικόνα πιθανότητας ύπαρξης χρώματος δέρματος στο frame αυτό:



Εικόνα 7.6: Εικόνα εισόδου



Εικόνα 7.7: Εικόνα πιθανοτήτων

7.2.3. Χρήση του **particle filtering** στην παρακολούθηση του χεριού

Η παρακολούθηση του χεριού αποτελεί το βασικότερο κομμάτι του αλγορίθμου, αφού παρέχει πληροφορίες για την κατάσταση του χεριού, όπως για παράδειγμα για τη θέση του χεριού στην εικόνα καθώς και για το σχήμα του χεριού (προσεγγιστικά). Όπως έχει ήδη αναφερθεί, η παρακολούθηση του χεριού πραγματοποιήθηκε με χρήση της μεθόδου **particle filtering**, τα βήματα εφαρμογής της οποίας παρουσιάζονται αναλυτικά στη συνέχεια. Σημειώνεται ότι ως χώρος κατάστασης χρησιμοποιείται το επίπεδο της εικόνας.

Αρχικοποίηση της μεθόδου: Ορίζεται ένας αριθμός δειγμάτων (particles), έστω $N=800$, τα οποία τοποθετούνται σε τυχαίες θέσεις (x, y) στο επίπεδο της εικόνας. Επίσης, καθορίζεται το αρχικό βάρος w_0^i κάθε particle i , ως $w_0^i = 1/N$.

Εξέλιξη particles: Κάθε particle i κινείται προς οποιαδήποτε κατεύθυνση σύμφωνα με μια τυχαία κατανομή δεδομένης μέσης τιμής και διασποράς. Έτσι, η

νέα θέση κάθε particle προκύπτει από την προηγούμενη και μια ποσότητα που εκφράζει τη μεταβολή της θέσης. Η ποσότητα αυτή, που μπορεί να θεωρηθεί ως η "ταχύτητα" κάθε particle είναι ίση με το άθροισμα της προηγούμενης τιμής της ταχύτητας και μιας τυχαίας ποσότητας που ορίζεται από μια ομοιόμορφη κατανομή.

Αναπροσαρμογή βαρών: Ανανεώνεται το βάρος κάθε δείγματος, ανάλογα με τις μετρήσεις. Οι μετρήσεις αποτελούνται από δύο στοιχεία: το πρώτο είναι η πιθανότητα κάθε pixel να είναι pixel περιοχής δέρματος και το δεύτερο είναι ένα στοιχείο κίνησης. Αναλυτικότερα, η πιθανότητα κάθε pixel να είναι pixel περιοχής δέρματος γίνεται με χρήση του μοντέλου χρώματος που παρουσιάστηκε στην προηγούμενη παράγραφο. Η πληροφορία της κίνησης ενσωματώνεται στο σύστημα για να αντιμετωπιστούν οι επιδράσεις της ύπαρξης αντικειμένων στο φόντο με χρώμα παρόμοιο με αυτό του δέρματος. Η πληροφορία αυτή εξάγεται ως εξής: Σε κάθε βήμα υπολογίζεται η διαφορά φωτεινότητας μεταξύ του παρόντος και του προηγούμενου frame. Στην εικόνα που προκύπτει θέτουμε ίσα με 1 τα pixels που η τιμή τους είναι μεγαλύτερη από ένα προκαθορισμένο όριο (εδώ ορίστηκε ίσο με 0,03) και τα υπόλοιπα τίθενται ίσα με 0. Η διαφορά φωτεινότητας των pixels του φόντου είναι σχεδόν μηδέν, οπότε έτσι προκύπτει μια εικόνα για τις κινούμενες περιοχές. Για να απομονώσουμε μόνο κινούμενες τις περιοχές με χρώμα δέρματος μπορούμε να εφαρμόσουμε τη λογική πράξη ΚΑΙ (AND) μεταξύ της εικόνας διαφορών και της εικόνας πιθανοτήτων, όπως αναφέρεται στο [Shan, Wei, Tan, Ojardias, 2004]. Συνδυάζοντας λοιπόν αυτές τις πληροφορίες μπορούμε να εντοπίσουμε το κινούμενο χέρι. Η ανανέωση των βαρών γίνεται πολλαπλασιάζοντας το προηγούμενο βάρος w_{k-1}^i κάθε particle με το σταθμισμένο άθροισμα των δύο αυτών πληροφοριών, δίνοντας όμως μεγαλύτερη βαρύτητα στο στοιχείο της πιθανότητας κάθε pixel να ανήκει σε περιοχή δέρματος. Συγκεκριμένα, το νέο βάρος του particle i , που βρίσκεται στη θέση (x_i, y_i) και αντιστοιχεί σε πιθανότητα χρώματος $p(x_i, y_i)$ και σε διαφορά φωτεινότητας $diff(x_i, y_i)$ είναι, στο frame k :

$$w_k^i = (\lambda_1 \cdot p(x_i, y_i) + \lambda_2 \cdot diff(x_i, y_i)) \cdot w_{k-1}^i$$

όπου στο σύστημά μας χρησιμοποιήθηκαν: $\lambda_1 = 0,8$ και $\lambda_2 = 0,2$. Στη συνέχεια, τα βάρη κανονικοποιούνται, έτσι ώστε $\sum_i w_k^i = 1$.

Εκτίμηση απόδοσης particle filtering: Σε αυτό το βήμα υπολογίζεται το μέγεθος $\hat{N}_{eff} = 1 / \sum_{i=1}^N (w_k^i)^2$ των αποτελεσματικών δειγμάτων, που αποτελεί μια εκτίμηση του εκφυλισμού του αλγορίθμου particle filtering και είναι κριτήριο για την εφαρμογή επαναδειγματοληψίας (resampling).

Resampling: Αν η ποσότητα \hat{N}_{eff} είναι μικρότερη ενός προκαθορισμένου ορίου thresh, δηλαδή αν τα αποτελεσματικά particles είναι λιγότερα από το όριο αυτό, τότε εκτελείται επαναδειγματοληψία. Το όριο που τέθηκε για την εφαρμογή της παρακολούθησης του χεριού είναι το 85% του συνόλου των particles, δηλαδή thresh=680. Η λογική του resampling είναι ότι αποβάλλονται τα particles που έχουν μικρό βάρος και δίνεται περισσότερη προσοχή στα particles με μεγάλο βάρος.

Συγκεκριμένα, στον αλγόριθμο υλοποίησης του resampling, μη αποδεκτά θεωρούνται τα particles των οποίων το βάρος είναι μικρότερο του 30% του μέγιστου υπαρκτού βάρους. Τα particles αυτά τοποθετούνται στις ίδιες θέσεις με τα αποδεκτά particles. Η επιλογή των νέων θέσεων γίνεται τυχαία, δίνονται όμως περισσότερες πιθανότητες στην επιλογή των θέσεων που αντιστοιχούν σε particles με μεγάλα βάρη. Αναλυτικότερα, η επιλογή γίνεται ως εξής: ο χώρος των αποδεκτών particles (που έχουν βάρος μεγαλύτερο του 30% του μέγιστου βάρους) χωρίζεται σε τρεις περιοχές: στην πρώτη περιοχή ανήκουν τα particles με βάρος μεταξύ του 30% και 40% του μέγιστου βάρους, στη δεύτερη περιοχή ανήκουν τα particles με βάρος μεταξύ του 40% και 60% του μέγιστου βάρους, και στην τρίτη περιοχή ανήκουν τα particles με βάρος μεγαλύτερο του 60% του μέγιστου βάρους. Στη συνέχεια ορίζεται ένας τυχαίος αριθμός που ανήκει στο διάστημα $[0,1]$ και η επιλογή της νέας θέσης κάθε μη αποδεκτού particle γίνεται με βάση την τιμή του αριθμού αυτού: αν ο αριθμός είναι μικρότερος του 0,2, τότε επιλέγεται (τυχαία) μια θέση από τις θέσεις των particles που ανήκουν στην πρώτη περιοχή αποδεκτών particles, αν ο αριθμός είναι μεταξύ του 0,2 και του 0,5, τότε επιλέγεται (τυχαία) μια θέση από τις θέσεις των particles που ανήκουν στη δεύτερη περιοχή αποδεκτών particles και αν ο αριθμός είναι μεγαλύτερος του 0,5, τότε επιλέγεται (τυχαία) μια θέση από τις θέσεις των particles που ανήκουν στην τρίτη περιοχή αποδεκτών particles. Σημειώνεται ότι ο αριθμός N των συνολικών particles παραμένει ο ίδιος. Σε κάθε particle του νέου συνόλου αντιστοιχίζεται ένα νέο βάρος, το οποίο είναι $w_k^i = 1/N$.

Τελικά, το αποτέλεσμα του particle filtering είναι μια εικόνα όπου όλα τα particles είναι ομοιόμορφα κατανεμημένα στην περιοχή του χεριού. Τα στάδια της εξέλιξης, της αναπροσαρμογής των βαρών, της εκτίμησης της απόδοσης και του resampling (εφόσον είναι απαραίτητο) εκτελούνται για κάθε frame.

7.2.4. Αναγνώριση χειρομορφών με σύμπτωση προτύπων

Όπως αναφέρθηκε στην προηγούμενη παράγραφο, το αποτέλεσμα της παρακολούθησης του κινούμενου χεριού είναι η εικόνα των particles, που αποτελεί μια προσεγγιστική εικόνα της μορφής του χεριού. Με δεδομένη την εικόνα αυτή το επόμενο στάδιο είναι η αναγνώριση της συγκεκριμένης χειρομορφής για κάθε frame. Η τεχνική που χρησιμοποιείται για την αναγνώριση των χειρομορφών είναι η σύμπτωση προτύπων.

Τα πρωτότυπα που χρησιμοποιούνται για σύγκριση είναι οι αντίστοιχες δυαδικές εικόνες των έξι χειρομορφών που δόθηκαν στην αρχή αυτού του κεφαλαίου. Αυτές οι δυαδικές εικόνες, που έχουν τιμή 1 στην περιοχή του χεριού και 0 στο φόντο, προκύπτουν από τις εικόνες των χειρομορφών (Εικόνα 7.1) με απλή κατάτμηση της κάθε εικόνας με βάση το χρώμα. Για αυτή την κατάτμηση ακολουθείται η παρακάτω διαδικασία. Αρχικά η RGB εικόνα του πρωτοτύπου μετατρέπεται στο χώρο HSV. Στη συνέχεια, όσα pixels της εικόνας ανήκουν στην περιοχή που ορίζεται από προκαθορισμένα ανώτερα και κατώτερα όρια για κάθε συνιστώσα της εικόνας τίθενται ίσα με 1, ενώ τα υπόλοιπα τίθενται στο 0. Για την εξάλειψη περιοχών που βρίσκονται εκτός του χεριού διατηρείται μόνο η μεγαλύτερη περιοχή, που είναι η περιοχή του χεριού.

Τα βασικά βήματα που ακολουθούνται για την αναγνώριση της χειρομορφής με βάση την τεχνική της σύμπτωσης προτύπων είναι:

- ✓ Επέκταση της εικόνας των particles, ώστε να προκύψει μια ακριβέστερη προσέγγιση της περιοχής του χεριού. Η επέκταση της περιοχής του χεριού, όπως αυτή προκύπτει από το particle filtering, γίνεται με βάση την εικόνα πιθανοτήτων των pixels να ανήκουν σε περιοχή δέρματος. Έτσι, ελέγχεται η περιοχή μεγέθους 5x5 γύρω από κάθε particle και σε όλα pixels της περιοχής αντιστοιχεί πιθανότητα μεγαλύτερη ενός προκαθορισμένου ορίου (έστω 0,2) αυτά προστίθενται στην εικόνα των particles, δηλαδή στην εικόνα του χεριού. Τελικά η εικόνα που προκύπτει είναι μια ακριβέστερη προσέγγιση του σχήματος της περιοχής του χεριού.
- ✓ Επόμενο βήμα είναι η εύρεση του κέντρου μάζας (\bar{x}, \bar{y}) της περιοχής αυτής. Οι σχέσεις υπολογισμού των (\bar{x}, \bar{y}) δίνονται στο [Nallaperumal et al., 2006]:

$$\bar{x} = \frac{1}{A} \sum_{i=1}^m \sum_{j=1}^n j \cdot B(i, j)$$

$$\bar{y} = \frac{1}{A} \sum_{i=1}^m \sum_{j=1}^n i \cdot B(i, j)$$

όπου A είναι ο αριθμός των pixels της περιοχής του χεριού (εμβαδό επιφάνειας) και $B(i, j)$ είναι η εικόνα της περιοχής του χεριού, διαστάσεων $m \times n$, όπως αυτή προκύπτει από το προηγούμενο βήμα.

- ✓ Ένα άλλο απαραίτητο στοιχείο της περιοχής του χεριού είναι ο προσανατολισμός. Για την ακρίβεια, υπολογίζεται η γωνία της διεύθυνσης της ευθείας εκείνης για την οποία το άθροισμα των τετραγώνων των αποστάσεων μεταξύ των σημείων της περιοχής και της ευθείας είναι ελάχιστη. Η γωνία αυτή, υπολογίζεται από την παρακάτω σχέση, η οποία δίνεται στο [Nallaperumal et al., 2006]:

$$\theta = \frac{1}{2} \arctan\left(\frac{b}{a-c}\right)$$

όπου

$$a = \sum_{i=1}^m \sum_{j=1}^n (x'_{ij})^2 \cdot B(i, j)$$

$$b = 2 \cdot \sum_{i=1}^m \sum_{j=1}^n x'_{ij} \cdot y'_{ij} \cdot B(i, j)$$

$$c = \sum_{i=1}^m \sum_{j=1}^n (y'_{ij})^2 \cdot B(i, j)$$

και

$$x' = x - \bar{x}$$

$$y' = y - \bar{y}$$

- ✓ Στη συνέχεια υπολογίζονται οι διαστάσεις της περιοχής του χεριού. Για το σκοπό αυτό είναι απαραίτητο να περιστραφεί κατάλληλα η εικόνα ώστε η

περιοχή του χεριού να είναι τοποθετημένη σε κατακόρυφη είτε σε οριζόντια θέση, ώστε να υπολογιστούν οι ακριβείς διαστάσεις της. Οι διαστάσεις αυτές υπολογίζονται με τη βοήθεια τεσσάρων δεικτών, που είναι οι μέγιστες και ελάχιστες τιμές της περιοχής στις δύο διαστάσεις της εικόνας. Έτσι, ανάλογα με τον προσανατολισμό της περιστραμμένης περιοχής, το ύψος και το πλάτος της περιοχής του χεριού υπολογίζονται ως οι διαφορές των μέγιστων και ελάχιστων τιμών στην y και x διάσταση αντίστοιχα, όπου γενικά θεωρούμε ότι το πλάτος είναι η απόσταση από τον καρπό μέχρι τα δάχτυλα.

- ✓ Καθένα από τα έξι πρωτότυπα πρέπει να τοποθετηθεί στις ίδιες συντεταγμένες με την περιοχή του χεριού. Έτσι, κάθε (δυναμική) εικόνα του πρωτοτύπου μετασχηματίζεται στο ύψος και πλάτος της περιοχής του χεριού και περιστρέφεται κατά κατάλληλη γωνία ώστε να ευθυγραμμιστεί με τη διεύθυνση της περιοχής του χεριού. Στη συνέχεια υπολογίζεται το κέντρο μάζας της νέας εικόνας του πρωτοτύπου και ταυτίζεται με το κέντρο μάζας της περιοχής του χεριού. Επόμενο βήμα είναι ο υπολογισμός της συσχέτισης (cross-correlation) μεταξύ της εικόνας του χεριού και της εικόνας του μετασχηματισμένου πρωτοτύπου. Η χειρομορφή που αναγνωρίζεται είναι αυτή για την οποία η τιμή της συσχέτισης είναι μέγιστη.

Το τελικό αποτέλεσμα αυτού του σταδίου είναι το περίγραμμα της χειρομορφής που αναγνωρίστηκε, το οποίο υπερτίθεται στην εικόνα εισόδου.

7.3. Πλεονεκτήματα και περιορισμοί συστήματος

7.3.1. Πλεονεκτήματα συστήματος

Στην παρούσα παράγραφο παρουσιάζονται τα πλεονεκτήματα του συστήματος ανίχνευσης, παρακολούθησης και αναγνώρισης χειρονομιών:

- ✓ Το βασικό πλεονέκτημα της μεθόδου είναι ότι δεν είναι ιδιαίτερα υπολογιστικά δαπανηρή: απαιτούνται περίπου 1,5sec για την επεξεργασία κάθε frame. Σημειώνεται ότι αν παραλείψουμε το στάδιο αναγνώρισης της χειρομορφής και το σύστημα χρησιμοποιηθεί μόνο για παρακολούθηση του χεριού απαιτούνται περίπου 0,4sec/frame. Η λήψη αυτών των αποτελεσμάτων έγινε με χρήση του Matlab, σε φορητό υπολογιστή Toshiba, Intel® Pentium® M 1,73 GHz, 512 MB RAM.
- ✓ Λειτουργεί ικανοποιητικά σε μη ομοιόμορφο φόντο: δεν είναι απαραίτητο το φόντο να είναι μονόχρωμο ή στατικό.
- ✓ Είναι ανεπηρέαστο σε αλλαγές του φωτισμού: δύο λόγοι συνηγορούν για αυτό. Ο πρώτος λόγος είναι ότι το στοιχείο του χρώματος είναι ανεξάρτητο από τη φωτεινότητα αφού χρησιμοποιούνται μόνο οι χρωματικές συνιστώσες. Ο δεύτερος λόγος είναι ότι το χρωματικό μοντέλο που χρησιμοποιήθηκε περιλαμβάνει δείγματα που έχουν ληφθεί σε διάφορες συνθήκες φωτισμού.

7.3.2. Περιορισμοί συστήματος

Οι απαραίτητες προϋποθέσεις για τη σωστή λειτουργία του συστήματος, αλλά και κάποιοι περιορισμοί στη λειτουργία του είναι:

- ✓ Είναι απαραίτητο στο video να μην υπάρχουν άλλα κινούμενα αντικείμενα που έχουν χρώμα δέρματος, αλλά θα πρέπει να υπάρχει μόνο ένα κινούμενο χέρι ώστε να αναγνωριστεί η διάταξή του.
- ✓ Επιπλέον, θα πρέπει ο βραχίονας του χεριού να μην είναι εκτεθειμένος, γιατί θα αναγνωρίζεται και αυτός ως κινούμενο χέρι λόγω του χρώματός του, με αποτέλεσμα η αναγνώριση των χειρομορφών να είναι λανθασμένη. Μια λύση είναι στην καταγραφή του video ο χρήστης να φοράει ρούχα με μακριά μανίκια.
- ✓ Η αναγνώριση είναι επίσης πιθανό να καταλήξει σε λανθασμένη εκτίμηση της χειρομορφής κατά τη διάρκεια μετάβασης από τη μια χειρομορφή στην επόμενη, αφού λόγω της αλλαγής της διάταξης του χεριού, το σχήμα του δεν είναι κανένα από τα προκαθορισμένα πρωτότυπα.
- ✓ Ένας ακόμη περιορισμός έχει σχέση με τον υπολογισμό του προσανατολισμού της περιοχής του χεριού. Όπως αναφέρθηκε υπολογίζεται μόνο η διεύθυνση της περιοχής, χωρίς να καθορίζεται πού είναι ο καρπός και πού τα δάχτυλα. Έτσι, δεν είναι δυνατός ο διαχωρισμός δύο χειρομορφών που διαφέρουν κατά 180° , όπως αυτές που φαίνονται για παράδειγμα στα παρακάτω σχήματα:



Εικόνα 7.8: Χειρομορφές που διαφέρουν κατά 180°

- ✓ Τέλος, σημειώνεται ότι το σύστημα περιλαμβάνει διάφορες καθοριστικές για την απόδοσή του παραμέτρους που δεν ορίζονται αυτόματα, γεγονός που δυσκολεύει τη γενίκευση της χρήσης του χωρίς σφάλματα. Αυτές οι παράμετροι είναι ο αριθμός των τάξεων στις οποίες χωρίζεται ο χρωματικός χώρος για τη δημιουργία του μοντέλου χρώματος, ο αριθμός των particles στο particle filtering, αλλά και τα διάφορα κατώφλια που ορίζονται.

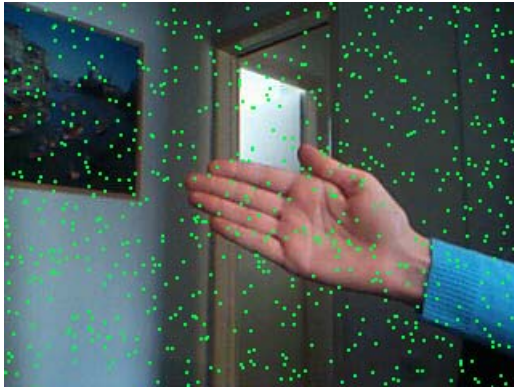
7.4. Αποτελέσματα και παρατηρήσεις

7.4.1. Αποτελέσματα

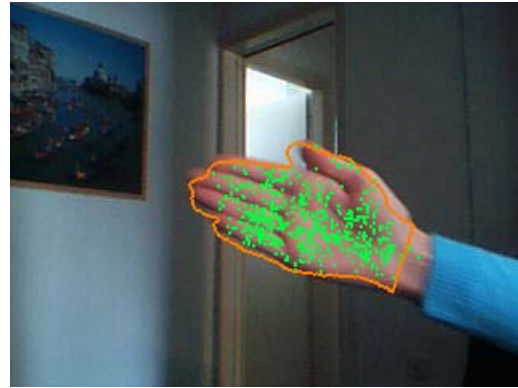
Στην ενότητα αυτή παρουσιάζονται ορισμένα ενδεικτικά αποτελέσματα που προκύπτουν από τον έλεγχο λειτουργίας του συστήματος ανίχνευσης, παρακολούθησης και αναγνώρισης χειρονομιών που περιγράφηκε στις

προηγούμενες παραγράφους. Παρατίθεται επίσης η γραφική παράσταση του απαιτούμενου χρόνου επεξεργασίας για κάθε frame.

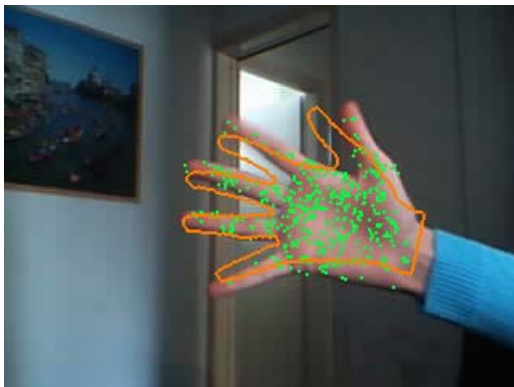
Στο πρώτο video που χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων το φόντο είναι πολύπλοκο και ο φωτισμός μη ελεγχόμενος. Σημειώνεται ότι οι κουκκίδες που φαίνονται στις παρακάτω εικόνες είναι τα particles, ενώ το περίγραμμα της χειρομορφής που απεικονίζεται είναι το αποτέλεσμα της σύμπτωσης προτύπων, δηλαδή το περίγραμμα της χειρομορφής που αναγνωρίζεται ως επικρατέστερη.



7.9.1



7.9.2



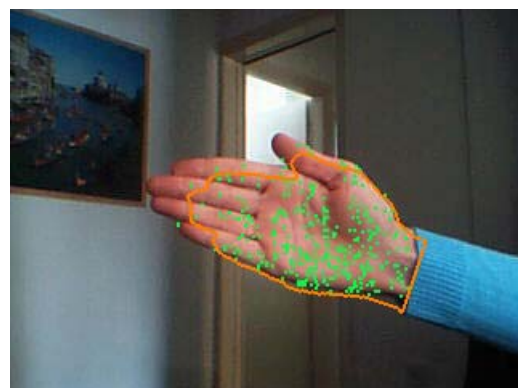
7.9.3



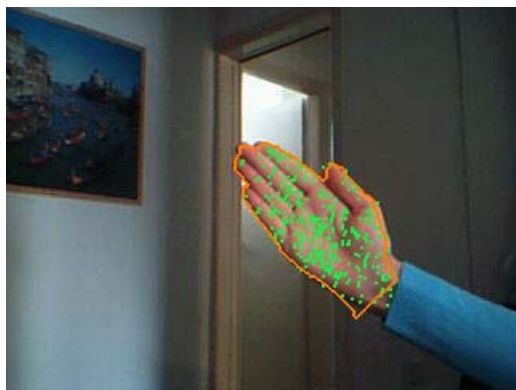
7.9.4



7.9.5



7.9.6



7.9.7



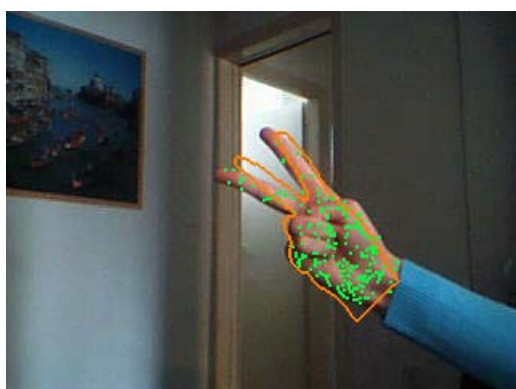
7.9.8



7.9.9



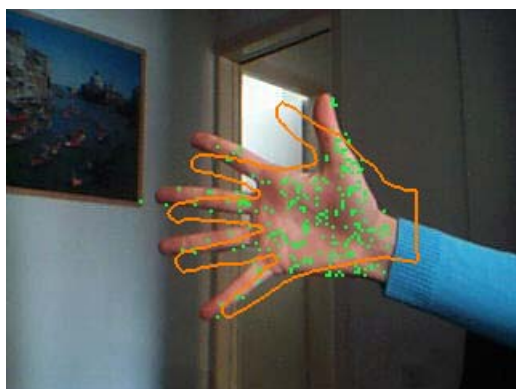
7.9.10



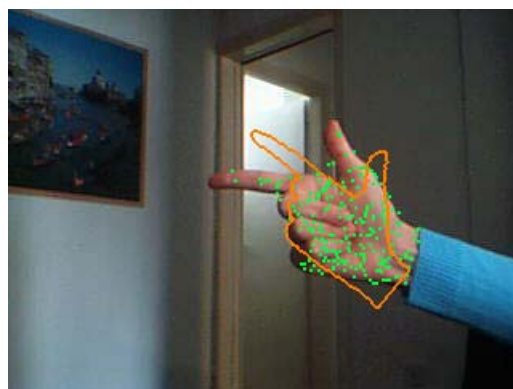
7.9.11



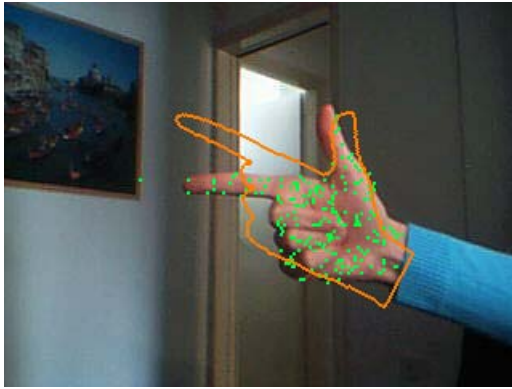
7.9.12



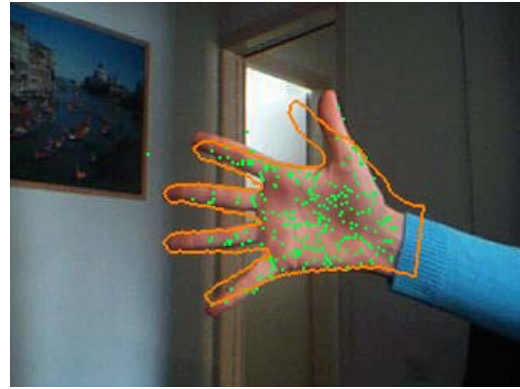
7.9.13



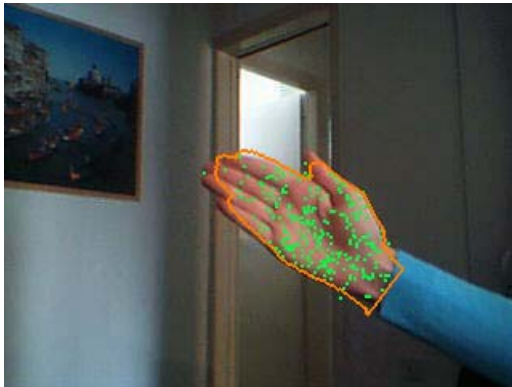
7.9.14



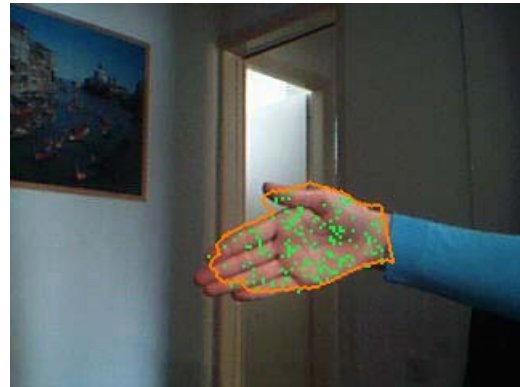
7.9.15



7.9.16

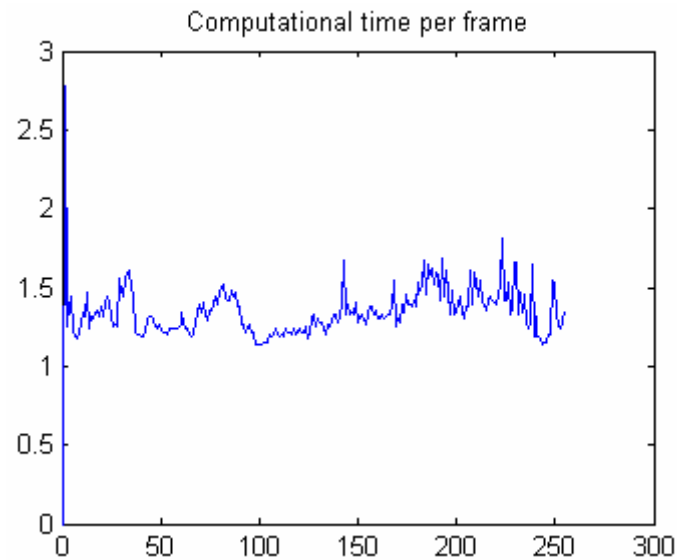


7.9.17



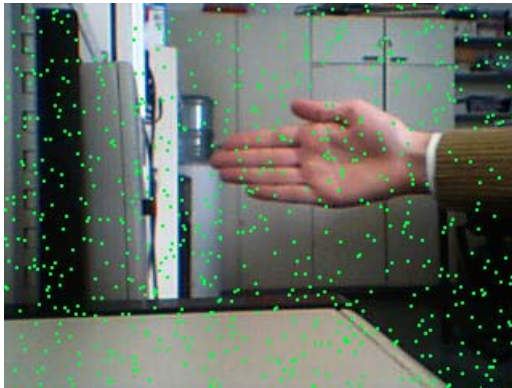
7.9.18

Εικόνα 7.9: Στιγμιότυπα από τα αποτελέσματα λειτουργίας του συστήματος. Στην εικόνα 7.9.1 παρουσιάζεται η αρχικοποίηση του συστήματος, όπου τοποθετούνται τυχαία τα particles στο επίπεδο της εικόνας. Στις υπόλοιπες εικόνες φαίνονται τα αποτελέσματα της παρακολούθησης και αναγνώρισης των χειρονομιών.



Εικόνα 7.10: Απαιτούμενος χρόνος επεξεργασίας για κάθε frame

Στο δεύτερο video που χρησιμοποιήθηκε για την εξαγωγή των αποτελεσμάτων το φόντο και οι συνθήκες φωτισμού είναι διαφορετικά σε σχέση με το πρώτο. Ωστόσο, και σε αυτό το video το φόντο επίσης πολύπλοκο και ο φωτισμός μη ελεγχόμενος.



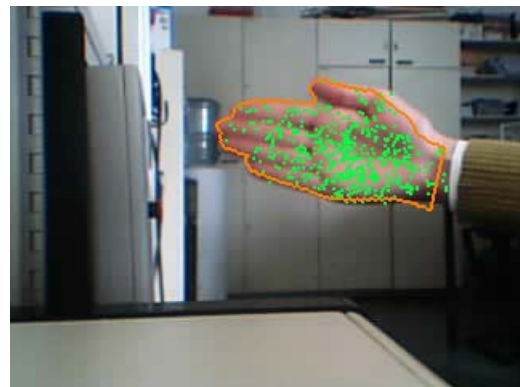
7.11.1



7.11.2



7.11.3



7.11.4



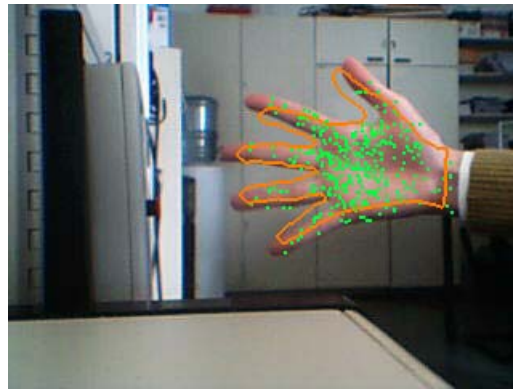
7.11.5



7.11.6



7.11.7



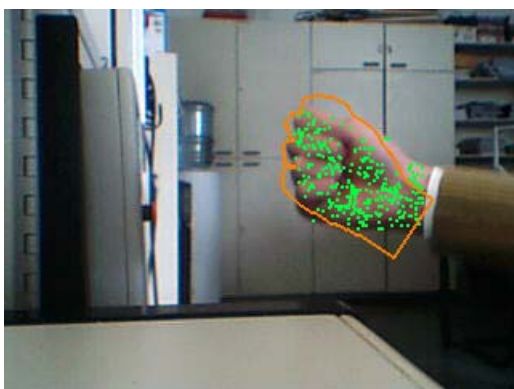
7.11.8



7.11.9



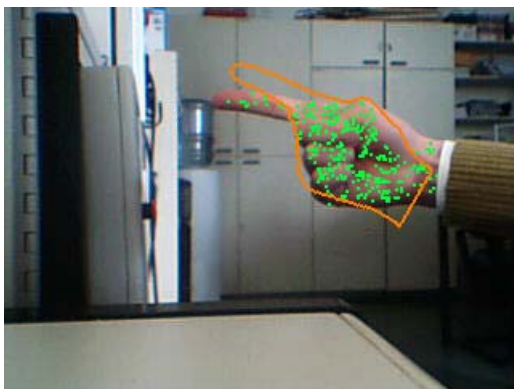
7.11.10



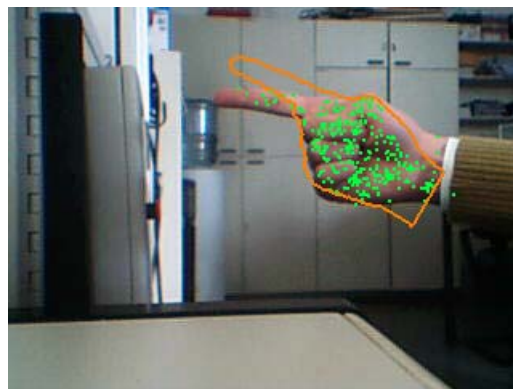
7.11.11



7.11.12



7.11.13



7.11.14



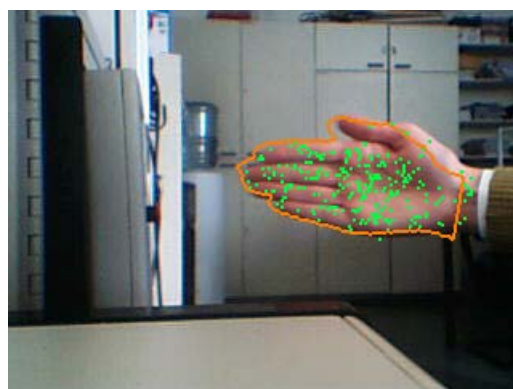
7.11.15



7.11.16



7.11.17



7.11.18



7.11.19



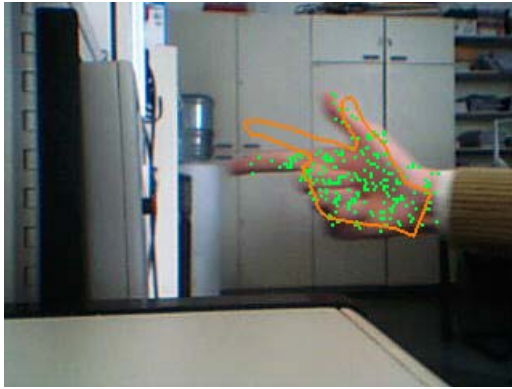
7.11.20



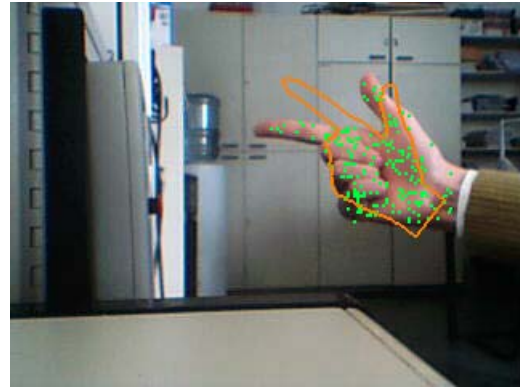
7.11.21



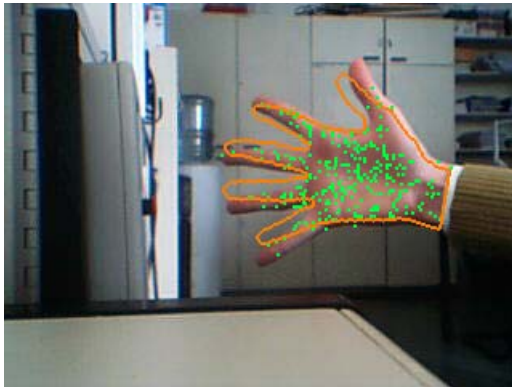
7.11.22



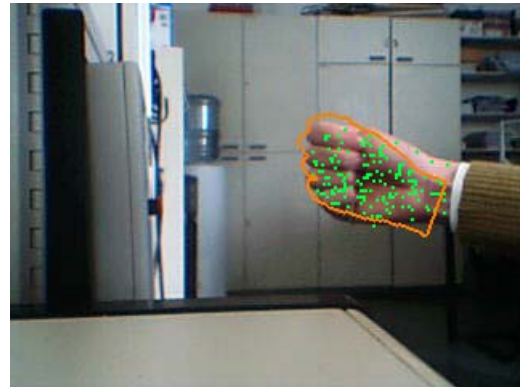
7.11.23



7.11.24

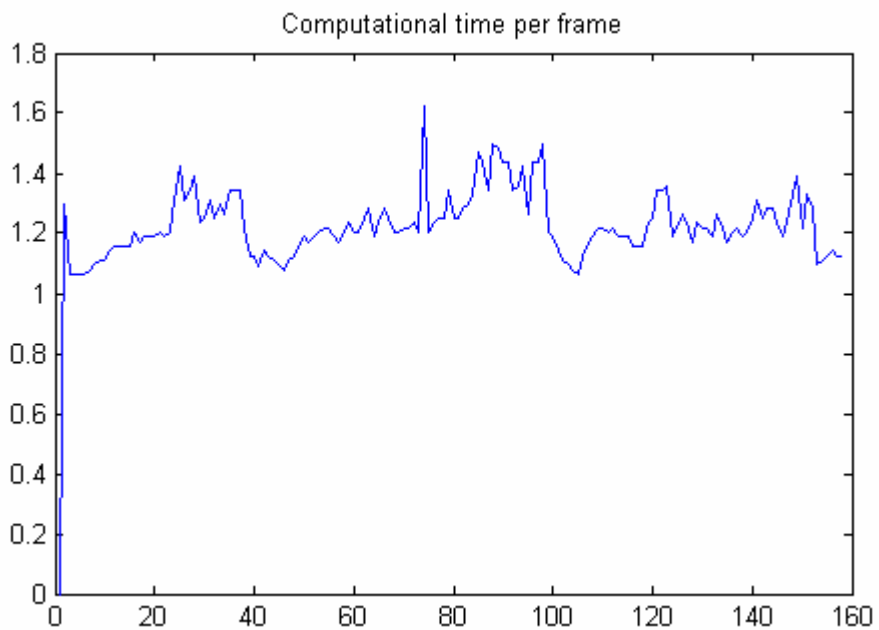


7.11.25



7.11.26

Εικόνα 7.11: Στιγμιότυπα από τον έλεγχο λειτουργίας του συστήματος. Στην εικόνα 7.11.1 παρουσιάζεται η αρχικοποίηση του συστήματος, όπου τοποθετούνται τυχαία τα particles στο επίπεδο της εικόνας. Στις υπόλοιπες εικόνες φαίνονται τα αποτελέσματα της παρακολούθησης και αναγνώρισης των χειρονομιών.



Εικόνα 7.12: Απαιτούμενος χρόνος επεξεργασίας για κάθε frame

Στις επόμενες εικόνες παρουσιάζονται τα αποτελέσματα της παρακολούθησης και αναγνώρισης χειρονομιών για ένα τρίτο video, στο οποίο το χέρι κινείται σε περισσότερους προσανατολισμούς. Και στην περίπτωση αυτή το φόντο είναι πολύπλοκο και ο φωτισμός μη ελεγχόμενος.



7.13.1



7.13.2



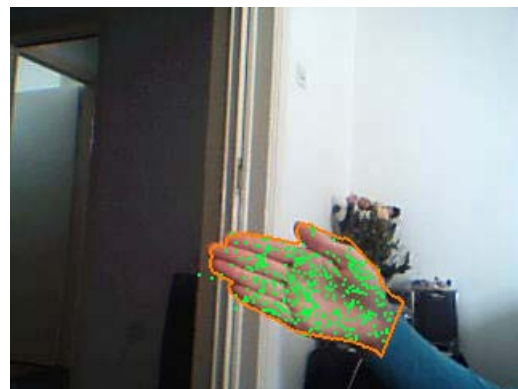
7.13.3



7.13.4



7.13.5



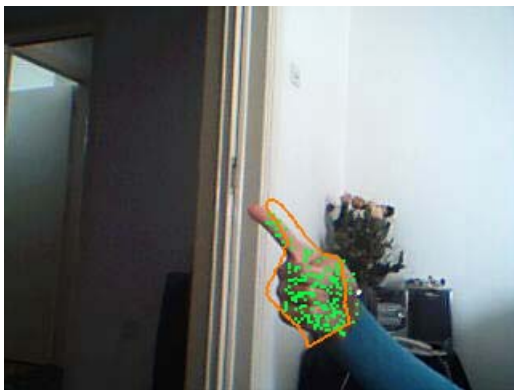
7.13.6



7.13.7



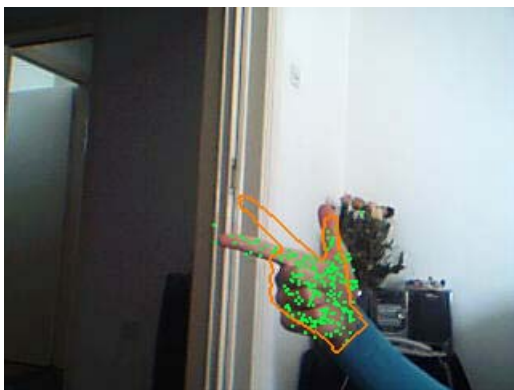
7.13.8



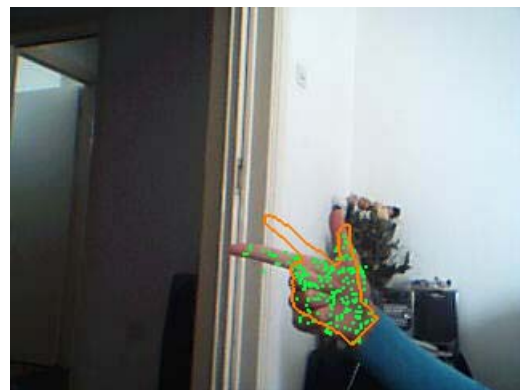
7.13.9



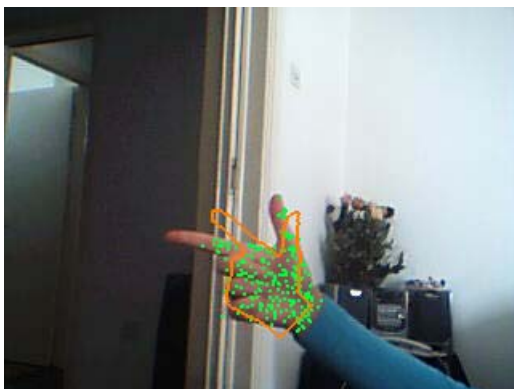
7.13.10



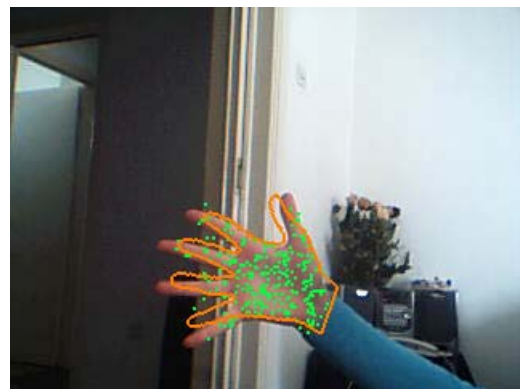
7.13.11



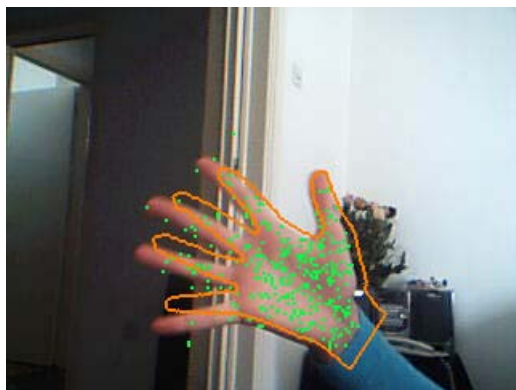
7.13.12



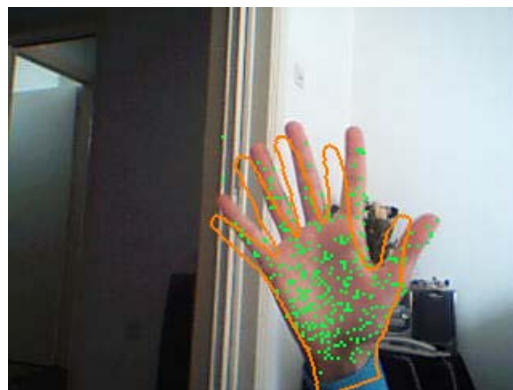
7.13.13



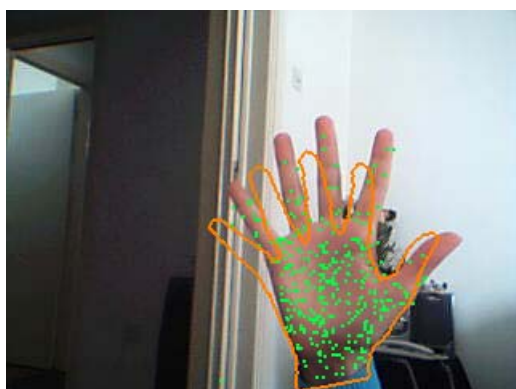
7.13.14



7.13.15



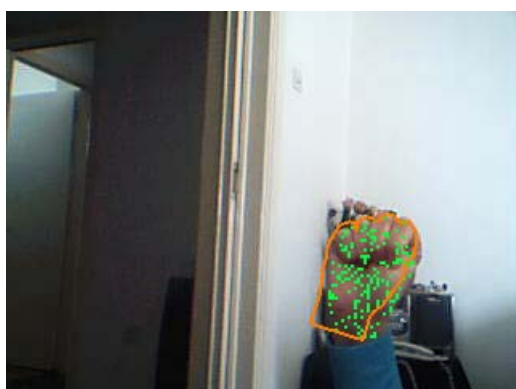
7.13.16



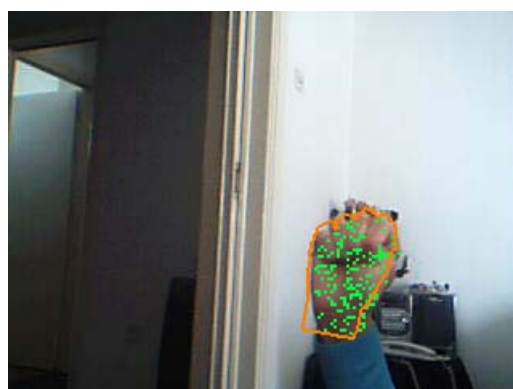
7.13.17



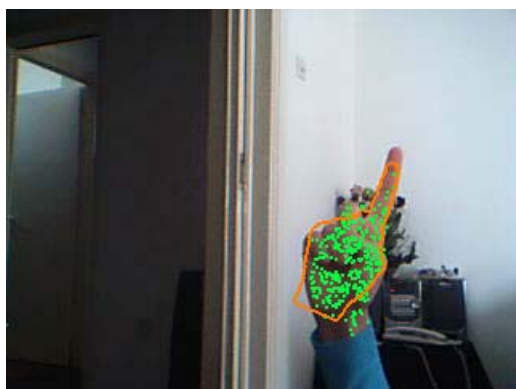
7.13.18



7.13.19



7.13.20



7.13.21



7.13.22



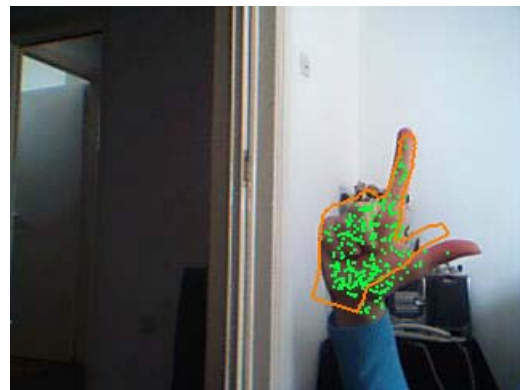
7.13.23



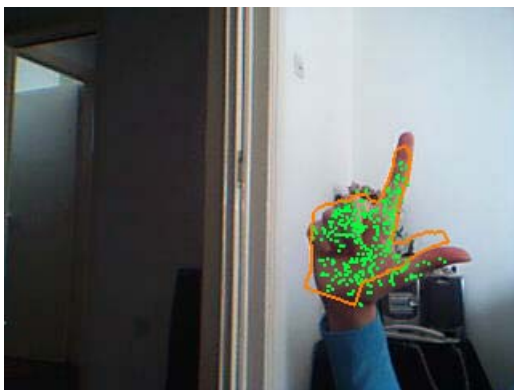
7.13.24



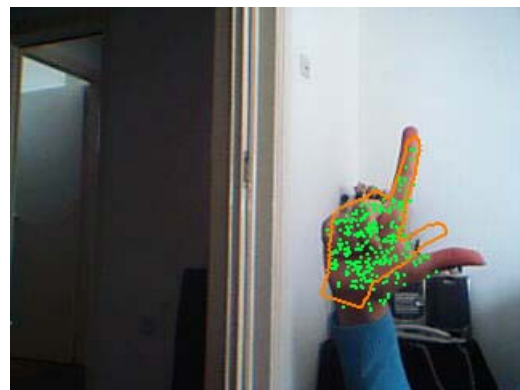
7.13.25



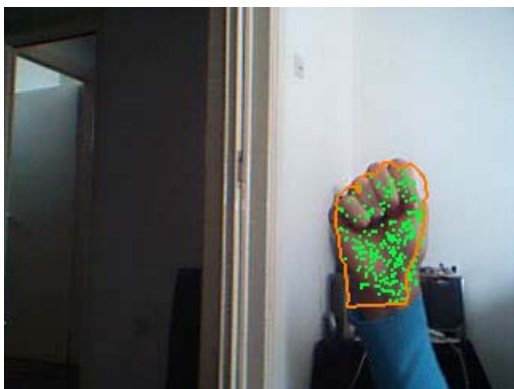
7.13.26



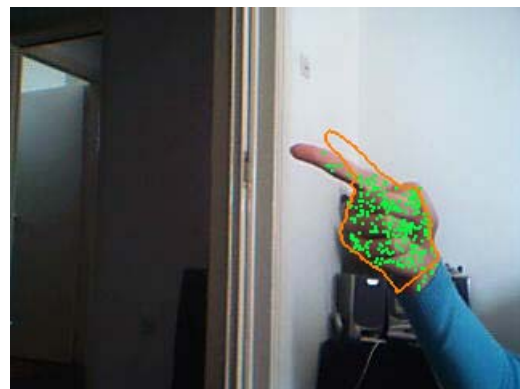
7.13.27



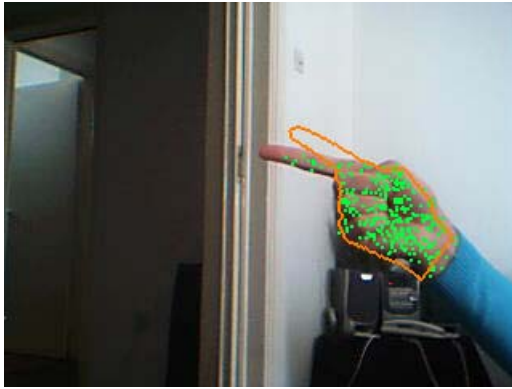
7.13.28



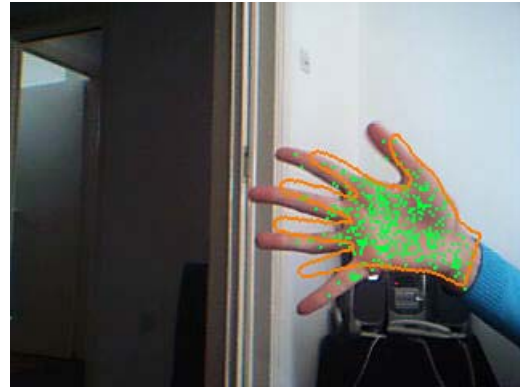
7.13.29



7.13.30



7.13.31



7.13.32

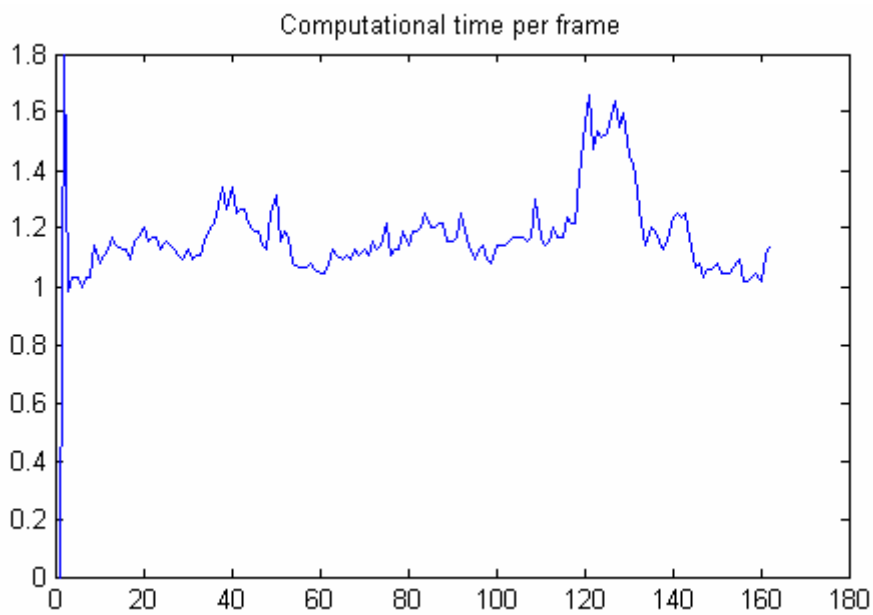


7.13.33



7.13.34

Εικόνα 7.13: Στιγμιότυπα από τον έλεγχο λειτουργίας του συστήματος. Στην εικόνα 7.13.1 παρουσιάζεται η αρχικοποίηση του συστήματος, όπου τοποθετούνται τυχαία τα particles στο επίπεδο της εικόνας. Στις υπόλοιπες εικόνες φαίνονται τα αποτελέσματα της παρακολούθησης και αναγνώρισης των χειρονομιών.



Εικόνα 7.14: Απαιτούμενος χρόνος επεξεργασίας για κάθε frame

Τέλος, παραθέτουμε τα αποτελέσματα της παρακολούθησης δύο χεριών, χωρίς αναγνώριση των χειρομορφών. Σημειώνουμε ότι στο παρακάτω video, εκτός του ότι καταγράφεται η κίνηση δύο χεριών, έχουμε και μερικό αποκλεισμό (occlusion) του ενός χεριού, χωρίς να επηρεάζεται η απόδοση της παρακολούθησης.



7.15.1



7.15.2



7.15.3



7.15.4



7.15.5



7.15.6



7.15.7



7.15.8



7.15.9



7.15.10



7.15.11



7.15.12



7.15.13



7.15.14



7.15.15



7.15.16



7.15.17



7.15.18



7.15.19



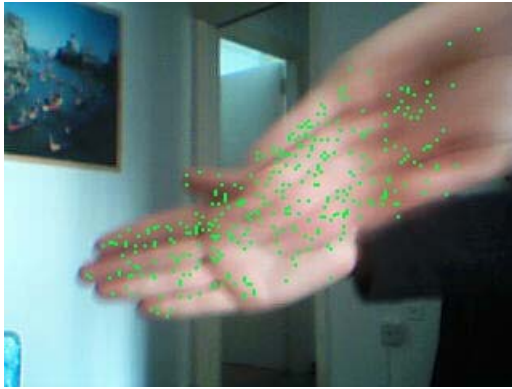
7.15.20



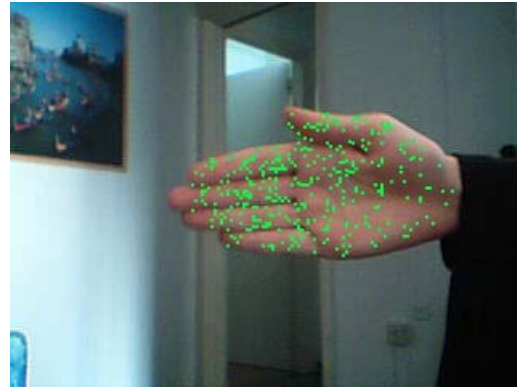
7.15.21



7.15.22

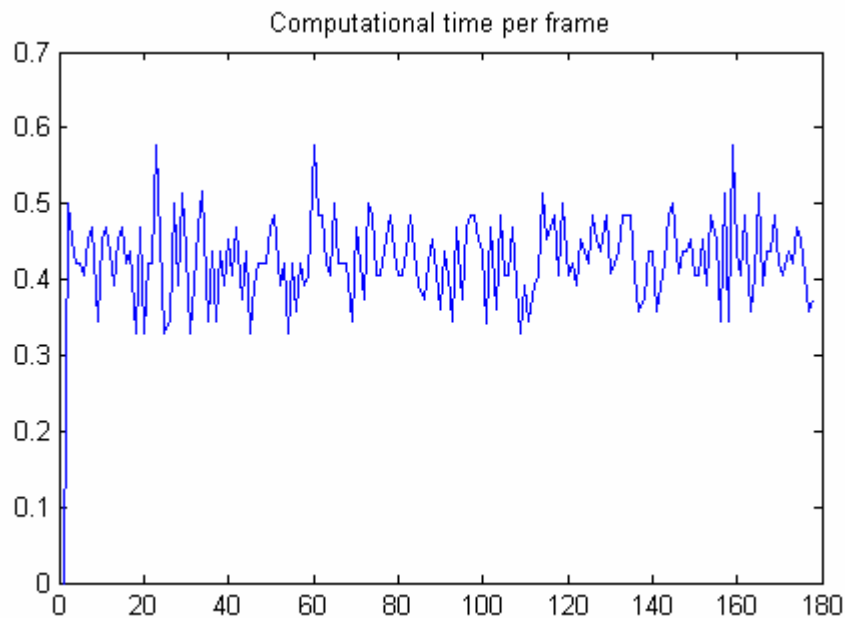


7.15.23



7.15.24

Εικόνα 7.15: Στιγμιότυπα από τον έλεγχο λειτουργίας του συστήματος. Στην εικόνα 7.15.1 παρουσιάζεται η αρχικοποίηση του συστήματος, όπου τοποθετούνται τυχαία τα particles στο επίπεδο της εικόνας. Στις υπόλοιπες εικόνες φαίνονται τα αποτελέσματα της παρακολούθησης του ενός και των δύο χεριών.



Εικόνα 7.16: Απαιτούμενος χρόνος επεξεργασίας για κάθε frame

7.4.2. Παρατηρήσεις

Όπως παρατηρούμε από τα παραπάνω αποτελέσματα, η παρακολούθηση και αναγνώριση των χειρονομιών είναι επιτυχής τόσο σε διάφορα μεγέθη, όσο και σε διάφορους προσανατολισμούς του χεριού. Ο μη ελεγχόμενος φωτισμός δεν επηρεάζει την απόδοση του συστήματος, αφού το μοντέλο χρώματος που χρησιμοποιείται περιέχει δείγματα που έχουν ληφθεί σε διάφορες συνθήκες φωτισμού. Επίσης, το πολύπλοκο φόντο δεν προκαλεί προβλήματα στη λειτουργία του συστήματος. Ωστόσο, υπάρχουν ορισμένες περιπτώσεις όπου τα αποτελέσματα της αναγνώρισης των χειρομορφών δεν είναι τα σωστά. Αυτό συμβαίνει όταν τα particles δεν καλύπτουν ομοιόμορφα την περιοχή του χεριού και όταν εμφανίζονται particles έξω από την περιοχή αυτού. Επίσης, αφού το σύστημα

αναγνωρίζει στατικές χειρομορφές, είναι αναμενόμενο να υπάρχουν προβλήματα στην αναγνώριση κατά τη διάρκεια της μετάβασης του χεριού από τη μια χειρονομία στην άλλη, επειδή τότε το σχήμα του είναι ασαφές. Κατά μέσο όρο το ποσοστό επιτυχούς αναγνώρισης των χειρονομιών προσεγγίζει το 85%, για όλα τα frames που επεξεργάζεται το σύστημα (μαζί με τα frames των μεταβάσεων από τη μια διάταξη του χεριού στην άλλη). Τέλος, όπως φαίνεται από τα αποτελέσματα του τελευταίου video, είναι επιτυχής τόσο η παρακολούθηση του χεριού όταν υπάρχει μερικός αποκλεισμός (occlusion) αυτού, όσο και η παρακολούθηση δύο χεριών.

Η βασικότερη παράμετρος που επηρεάζει άμεσα την απόδοση του συστήματος είναι ο αριθμός των particles που χρησιμοποιούνται για την παρακολούθηση του χεριού. Όταν χρησιμοποιούνται λίγα particles, δεν καλύπτουν ολόκληρη την περιοχή του χεριού, με αποτέλεσμα η προσεγγιστική εικόνα του χεριού που προκύπτει από την παρακολούθηση με το particle filtering να μην επαρκεί για την εξαγωγή της σωστής χειρομορφής από τη σύμπτωση προτύπων. Από την άλλη, η χρήση πολύ μεγάλου αριθμού particles δεν εξασφαλίζει απαραίτητα την ακριβή εκτίμηση της διάταξης του χεριού, αφού λόγω του resampling θα υπάρχουν πολλά particles στις ίδιες θέσεις. Στην υλοποίηση της παρούσας εργασίας χρησιμοποιήθηκαν, όπως αναφέρθηκε 700 particles, αριθμός που δίνει ικανοποιητικά αποτελέσματα.

Άλλες παράμετροι που επηρεάζουν την απόδοση είναι τα διάφορα όρια που θέτουμε, όπως για παράδειγμα στον υπολογισμό του μέτρου του εκφυλισμού του particle filtering, στα όρια που τίθενται για τον χωρισμό των βαρών σε διάφορες περιοχές κατά τη διαδικασία του resampling, και το όριο που θέτουμε για τον υπολογισμό της πληροφορίας κίνησης. Για τη μείωση αυτής της επιρροής θα πρέπει να βρεθεί κάποια μεθοδολογία για τον αυτόματο υπολογισμό αυτών των ορίων ή για την προσαρμογή τους σε κάθε frame.

Φυσικά, θα πρέπει να ληφθεί επίσης υπόψη ότι τα πρωτότυπα των χειρομορφών δεν έχουν εξαχθεί με κάποια ακριβή διαδικασία, αλλά πρόκειται για απλές εικόνες ενός χεριού σε διάφορες διατάξεις. Αυτό σημαίνει ότι τα πρωτότυπα δεν είναι ιδανικά και έτσι η διαδικασία της σύμπτωσης προτύπων είναι περισσότερο επιρρεπής σε σφάλματα.

8

Σύνοψη και Μελλοντικές Επεκτάσεις

Σε αυτό το τελευταίο κεφάλαιο ανακεφαλαιώνονται τα σημαντικότερα σημεία της διπλωματικής εργασίας, τόσο σε θεωρητικό επίπεδο, όσο και σε ότι αφορά την υλοποίηση του συστήματος ανίχνευσης, παρακολούθησης και αναγνώρισης των χειρονομιών. Προτείνονται επίσης ορισμένες χρήσιμες πιθανές επεκτάσεις του συστήματος αυτού, αλλά και γενικότερες κατευθύνσεις για μελλοντική έρευνα πάνω στο πρόβλημα της παρακολούθησης και αναγνώρισης ανθρώπινων χειρονομιών.

8.1. Σύνοψη της εργασίας και συμπεράσματα

Αντικείμενο της εργασίας αποτέλεσε η μελέτη, ο σχεδιασμός και η υλοποίηση ενός συστήματος αναγνώρισης ενός συνόλου προκαθορισμένων στατικών χειρονομιών που εκτελούνται από ένα κινούμενο χέρι, η κίνηση του οποίου καταγράφεται σε ακολουθίες εικόνων (video). Η επεξεργασία του video για την εξαγωγή των επιθυμητών αποτελεσμάτων έγινε χρησιμοποιώντας τις ακόλουθες μεθόδους της Όρασης Υπολογιστών:

- ✓ Χρήση του χρώματος του δέρματος με σκοπό τη δημιουργία ενός χρωματικού μοντέλου για ταξινόμηση των pixels της εικόνας σύμφωνα με το χρώμα τους.
- ✓ Μέθοδος particle filtering για παρακολούθηση του κινούμενου χεριού.
- ✓ Τεχνική σύμπτωσης προτύπων (template matching) για αναγνώριση των χειρομορφών.

Για λόγους πληρότητας αρχικά μελετήθηκαν θεωρητικά τα παραπάνω πεδία. Έτσι, εξηγήθηκαν τα βασικότερα στοιχεία για την αναπαράσταση του χρώματος στα υπολογιστικά συστήματα και περιγράφηκαν οι σημαντικότερες περιοχές χρήσης του χρώματος στην Όραση Υπολογιστών, δίνοντας έμφαση στην ανίχνευση περιοχών δέρματος, όπως τα χέρια, με βάση το χρώμα. Στη συνέχεια, παρουσιάστηκε εκτενώς η μέθοδος του particle filtering για την παρακολούθηση κινούμενων αντικειμένων και δόθηκε ο γενικός αλγόριθμος εφαρμογής της μεθόδου για την παρακολούθηση χειρονομιών. Παρουσιάστηκε τέλος η γενική θεωρία γύρω από τη σύμπτωση προτύπων και δόθηκε έμφαση στη χρήση της για την αναγνώριση χειρονομιών.

Ο σχεδιασμός του συστήματος παρακολούθησης και αναγνώρισης βασίστηκε στις εφαρμογές των μεθόδων αυτών στην παρακολούθηση και αναγνώριση χειρονομιών. Η δομή του συστήματος είναι η ακόλουθη:

- ✓ Πρώτο στάδιο είναι η δημιουργία ενός μοντέλου χρώματος, για την ταξινόμηση των pixels κάθε frame και συγκεκριμένα για την εύρεση της πιθανότητας του κάθε pixel να ανήκει σε περιοχή δέρματος. Η πιθανότητα αυτή χρησιμεύει στο στάδιο του particle filtering.
- ✓ Στη συνέχεια εφαρμόζεται ο βασικός αλγόριθμος του particle filtering για την παρακολούθηση του χεριού. Ως στοιχεία για τη διάδοση των particles χρησιμοποιούνται δύο πληροφορίες: οι πιθανότητες των pixels να ανήκουν σε περιοχή του χεριού και μια πληροφορία για την κίνηση των pixels.
- ✓ Το τελευταίο στάδιο είναι η αναγνώριση της χειρομορφής, με χρήση της τεχνικής της σύμπτωσης προτύπων, όπου συγκρίνεται η προσεγγιστική εικόνα της περιοχής του χεριού που έχει ανιχνευτεί μέσω του particle filtering με καθένα από τα προκαθορισμένα πρωτότυπα χειρομορφών.

Από την παρατήρηση των αποτελεσμάτων λειτουργίας του συστήματος προέκυψαν ορισμένα χρήσιμα συμπεράσματα. Παρατηρήθηκε λοιπόν η ικανότητα του συστήματος να ανταπεξέρχεται τόσο σε μη ομοιόμορφο φόντο, όσο και σε διάφορες συνθήκες φωτισμού. Επίσης, είναι δυνατή η αναγνώριση των χειρομορφών σε διαφορετικούς προσανατολισμούς και διαφορετικές κλίμακες του χεριού. Το σημαντικότερο στοιχείο στη λειτουργία του συστήματος ήταν όμως η ταχύτητα επεξεργασίας και εξαγωγής των αποτελεσμάτων. Σε σχέση λοιπόν με άλλες μεθόδους, η μέθοδος του particle filtering αποδείχτηκε καταλληλότερη για γρήγορη επεξεργασία, οπότε κρίνεται και κατάλληλη για χρήση σε συστήματα παρακολούθησης χειρονομιών σε πραγματικό χρόνο. Ωστόσο, για επέκταση του συστήματος αυτού για λειτουργία πραγματικού χρόνου θα πρέπει να αναζητηθεί μια ταχύτερη μέθοδος για την αναγνώριση των χειρομορφών, αφού η σύμπτωση προτύπων, όπως υλοποιήθηκε, δεν είναι αρκετά γρήγορη υπολογιστικά.

Η παρακολούθηση και αναγνώριση ανθρώπινων χειρονομιών μπορεί να χρησιμοποιηθεί σε ένα ευρύ πεδίο εφαρμογών. Οι εφαρμογές αυτές περιλαμβάνουν τον τηλεχειρισμό ρομποτικών οχημάτων, ηλεκτρονικών υπολογιστών και διάφορων συσκευών που χρησιμοποιούμε καθημερινά, όπως τηλεόραση, κινητό τηλέφωνο, ακόμα και οικιακές συσκευές. Γενικότερα, είναι δυνατή η αντικατάσταση των τηλεχειριστηρίων ορισμένων συσκευών με ένα σύστημα αναγνώρισης χειρονομιών. Μια πολύ σημαντική εφαρμογή είναι επίσης η αναγνώριση της νοηματικής γλώσσας και η μετατροπή της σε κείμενο μέσω ηλεκτρονικού υπολογιστή, αλλά και η εκτίμηση της συναισθηματικής κατάστασης ενός ομιλητή από τις χειρονομίες του. Όταν κατασκευαστούν πολύ αξιόπιστα συστήματα, όπου το περιθώριο λάθους είναι μηδαμινό, το πεδίο εφαρμογών θα διευρυνθεί ακόμα περισσότερο. Έτσι, θα είναι εφικτή η δημιουργία οχημάτων όπου κάποιες λειτουργίες (ακόμα και η οδήγηση) θα εκτελούνται μέσω χειρονομιών, αλλά και η καθοδήγηση διάφορων ιατρικών εργαλείων, όπως η κάμερα του ενδοσκοπίου για διεξαγωγή λαπαροσκοπικών επεμβάσεων και εξετάσεων.

8.2. Μελλοντικές επεκτάσεις

Στην ενότητα αυτή παρουσιάζονται ορισμένες βελτιώσεις και επεκτάσεις που μπορούν να πραγματοποιηθούν στο σύστημα που υλοποιήθηκε για τις ανάγκες της διπλωματικής εργασίας. Οι επεκτάσεις αυτές βασίζονται κυρίως στον περιορισμό, ακόμα και στην εξάλειψη, των αδυναμιών του συστήματος.

- ✓ Σημαντική είναι η βελτίωση του τρόπου υπολογισμού της κατεύθυνσης, αλλά και του μεγέθους της περιοχής του χεριού, για εξαγωγή ακριβέστερων αποτελεσμάτων.
- ✓ Εξίσου σημαντική είναι και η εισαγωγή ενός βελτιωμένου, αλλά και υπολογιστικά γρήγορου αλγορίθμου για την εκτίμηση της κίνησης, ώστε το αποτέλεσμα της παρακολούθησης του χεριού με τη μέθοδο του particle filtering να είναι πιο αξιόπιστο και πιο ακριβές.
- ✓ Φυσικά κρίνεται απαραίτητη η εισαγωγή όσο το δυνατόν περισσότερων χειρομορφών προς αναγνώριση.
- ✓ Επίσης, κρίνεται σκόπιμο να αρθούν οι περιορισμοί σχετικά με την ύπαρξη άλλων περιοχών με χρώμα δέρματος στην εικόνα.
- ✓ Θα πρέπει ακόμη το σύστημα να συμπεριλάβει την αναγνώριση των χειρονομιών που εκτελούνται από περισσότερα από ένα χέρια. Σημειώνεται ότι το σύστημα που υλοποιήθηκε εδώ παρέχει μόνο τη δυνατότητα της παρακολούθησης περισσότερων από ένα χεριών, χωρίς όμως να αναγνωρίζονται οι χειρονομίες.
- ✓ Τέλος, είναι χρήσιμη η εξέλιξη του συστήματος ώστε να αναγνωρίζονται δυναμικά οι χειρονομίες και όχι μόνο η (στατική) διάταξη του χεριού σε κάθε frame.
- ✓ Προφανώς, το ήδη υπάρχον σύστημα αλλά και οι προηγούμενες βελτιώσεις είναι απαραίτητο να τροποποιηθούν ώστε να λειτουργούν σε πραγματικό χρόνο.

Είναι λοιπόν φανερό ότι το σύστημα που μελετήθηκε και υλοποιήθηκε για τις ανάγκες της παρούσας εργασίας, αποτελεί τη βάση για τη δημιουργία συστημάτων ακριβούς και ταχείας παρακολούθησης και αναγνώρισης χειρονομιών.

Βιβλιογραφία

- [1] Γ. Καραγιάννης, Γ. Σταϊνχάουερ: "Μάθηση Μηχανών και Αναγνώριση Προτύπων", Αθήνα, Νοέμβριος 2001
- [2] Π. Μαραγκός, "Ανάλυση Εικόνων και Όραση Υπολογιστών", Αθήνα, Ιούνιος 2005
- [3] A. Almeida, J. Almeida, R. Araujo: "Real-Time Tracking of Moving Objects Using Particle Filters", Proceedings of the IEEE International Symposium on Industrial Electronics, Vol. 4, pp. 1327 - 1332, 2005.
- [4] A. A. Argyros, M. I. A. Lourakis, "Vision-Based Interpretation of Hand Gestures for remote control of a Computer Mouse", Springer Verlag Lecture Notes in Computer Science, Vol. 3979, Computer Vision in Human-Computer Interaction, 2006
- [5] M. S. Arulampalam, S. Maskell, N. Gordon, T. Clapp: "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking", IEEE Transactions on Signal Processing, Volume 50, Issue 2, pp. 174 - 188, Feb. 2002
- [6] O. Bernier, D. Collobert: "Head and hands 3D tracking in real time by the EM algorithm", IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems Proceedings, pp. 75 - 81, 2001.
- [7] H.-J. Boehme, U.-D. Braumann, A. Corradini, H.-M. Gross: "Person Localization and Posture Recognition for Human-Robot Interaction", Springer Verlag, Lecture Notes in Computer Science, Volume 1739, Gesture-Based Communication in Human-Computer Interaction: International Gesture Workshop Proceedings, 1999
- [8] M. Bolic, S. Hong, P. M. Djuric: "Performance and complexity analysis of adaptive particle filtering for tracking applications", Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, Vol. 1, pp. 853 - 857, 2002.
- [9] M. Bray, E. Koller-Meier, L. Van Gool: "Smart particle filtering for 3D hand tracking", Sixth IEEE International Conference on Automatic Face and Gesture Recognition Proceedings, pp. 675 - 680, 2004.
- [10] L. Brethes, P. Menezes, F. Lerasle, J. Hayet: "Face tracking and hand gesture recognition for human-robot interaction", IEEE International Conference on Robotics and Automation Proceedings, Vol.2, pp. 1901 - 1906, 2004

- [11] L. Bretzner, I. Laptev, T. Lindeberg: "Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering", Fifth IEEE International Conference on Automatic Face and Gesture Recognition Proceedings, pp. 405 - 410, 2002
- [12] M. de Bruijne, M. Nielsen: "Image segmentation by shape particle filtering", Proceedings of the 17th International Conference on Pattern Recognition, Vol. 3, pp. 722 - 725, 2004
- [13] J. Cai, A. Goshtasby, C. Yu: "Detecting Human Faces in Color Images", Int'l Workshop on Multimedia Database Management Systems, pp. 124 - 131, Aug. 5-7, 1998
- [14] D. Chai, K. N. Ngan: "Face segmentation using skin-color map in videophone applications", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 9, Issue 4, pp. 551 - 564, June 1999
- [15] Wen-Yan Chang, Chu-Song Chen, Yi-Ping Hung: "Appearance-guided particle filtering for articulated hand tracking", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 235 - 242, 2005.
- [16] T. Chateau, A. Vacavant, J. M. Lavest: "Skin detection and tracking by monocular vision", International Symposium on Signals, Circuits and Systems, Vol. 1, pp. 79 - 82, 2005.
- [17] Jiun-Hung Chen, Chu-Song Chen, Yong-Sheng Chen: "Fast algorithm for robust template matching with M-estimators", IEEE Transactions on Signal Processing, Vol. 51, Issue 1, pp. 230 - 243, Jan. 2003
- [18] M. S. Choi, W. Y. Kim: "A novel two stage template matching method for rotation and illumination invariance", Elsevier Science Publishers, Pattern Recognition Vol. 35, Issue 1, pp. 119 - 129, January 2002.
- [19] X. Deng, J. Xie: "Nonlinear target tracking based on particle filter", Fifth World Congress on Intelligent Control and Automation, Vol. 2, pp. 1618 - 1620, 2004.
- [20] L. Di Stefano, S. Mattoccia: "Fast template matching using bounded partial correlation", Springer Verlag Machine Vision and Applications Journal, Vol. 13, Number 4, February 2003
- [21] L. Di Stefano, S. Mattoccia, F. Tombari: "An Algorithm for Efficient and Exhaustive Template Matching", Springer Verlag Lecture Notes in Computer Science, Vol. 3211, Image Analysis and Recognition, 2004
- [22] A. Doucet, S. Godsill, C. Andrieu: "On sequential Monte Carlo sampling methods for Bayesian filtering", Springer Verlag Statistics and Computing Journal, Vol. 10, Number 3, July 2000

- [23] K. Fredriksson, V. Mäkinen, G. Navarro: "Rotation and Lighting Invariant Template Matching", Springer Verlag, Lecture Notes in Computer Science, Vol. 2976, LATIN 2004: Theoretical Informatics, 2004
- [24] K. Fredriksson, G. Navarro, E. Ukkonen: "Sequential and Indexed Two-Dimensional Combinatorial Template Matching Allowing Rotations", Elsevier Science Publishers, Theoretical Computer Science, Vol. 347, Issues 1-2, pp. 239 - 275, November 2005
- [25] M. Frigola, J. Fernandez, J. Aranda: "Visual human machine interface by gestures", IEEE International Conference on Robotics and Automation Proceedings, Vol.1, pp. 386 - 391, 2003
- [26] S. Funck: "Video-Based Handsign Recognition for Intuitive Human-Computer-Interaction", Springer Verlag, Lecture Notes in Computer Science, Volume 2449, Pattern Recognition: 24th DAGM Symposium Proceedings, 2002
- [27] D. Geer: "Will gesture recognition technology point the way?", IEEE Computer, Vol. 37, Issue 10, pp. 20 - 23, Oct. 2004
- [28] T. Gevers: "Robust segmentation and tracking of colored objects in video", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, Issue 6, pp. 776 - 781, June 2004
- [29] R. C. Gonzalez, R. E. Woods, S. L. Eddins: "Digital Image Processing using MATLAB", Pearson Prentice-Hall, 2004
- [30] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, P.-J. Nordlund: "Particle filters for positioning, navigation, and tracking", IEEE Transactions on Signal Processing, Vol. 50, Issue 2, pp. 425 - 437, Feb. 2002
- [31] N. Habili, Cheng Chew Lim, A. Moini: "Segmentation of the face and hands in sign language video sequences using color and motion cues", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, Issue 8, pp. 1086 - 1097, Aug. 2004
- [32] C. von Hardenberg, F. Bérard, "Bare-Hand Human-Computer Interaction", ACM Press, Proceedings of the ACM Workshop on Perceptive User Interfaces, Orlando, Florida, USA, Nov. 15-16 2001
- [33] Md. Hasanuzzaman, T. Zhang, V. Ampornaramveth, P. Kiatisevi, Y. Shirai, H. Ueno: "Gesture based human-robot interaction using a frame based software platform", IEEE International Conference on Systems, Man and Cybernetics, Vol.3, pp. 2883 - 2888, 2004
- [34] C. Hue, J.-P. Le Cadre, P. Perez: "Sequential Monte Carlo methods for multiple target tracking and data fusion", IEEE Transactions on Signal Processing, Vol. 50, Issue 2, pp. 309 - 325, Feb. 2002

- [35] T. Kawanishi, T. Kurozumi, K. Kashino, S. Takagi: "A fast template matching algorithm with adaptive skipping using inner-subtemplates' distances", Proceedings of the 17th International Conference on Pattern Recognition, Vol. 3, pp. 654 - 657, 2004.
- [36] Zia Khan, T. Balch, F. Dellaert: "MCMC-based particle filtering for tracking a variable number of interacting targets", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, Issue 11, pp. 1805 - 1819, Nov. 2005
- [37] B. Kwolek: "Person following and mobile camera localization using particle filters", Proceedings of the Fourth International Workshop on Robot Motion and Control, pp. 265 - 270, 2004.
- [38] I. Laptev, T. Lindeberg: "Tracking of multi-state hand models using particle filtering and multi-scale image features", Springer Verlag Lecture Notes in Computer Science, Vol. 2106, pp. 63 - 74, Third International Conference on Scale-Space and Morphology in Computer Vision Proceedings, 2001
- [39] C. Laschi, M. Gonzalo-Tasis, J. F. Codes, P. Dario: "Recognizing hand posture by vision: applications in humanoid personal robotics", IEEE International Conference on Robotics and Automation Proceedings, Vol. 2, pp. 1439 - 1444, 2002.
- [40] Mun Wai Lee, I. Cohen, Soon Ki Jung: "Particle filter with analytical inference for human body tracking", Workshop on Motion and Video Computing Proceedings, pp. 159 - 165, 2002
- [41] Wei Liang, Cheng Ge, Yunde Jia: "Hand motion tracking using MDPF method", IEEE International Conference on Systems, Man and Cybernetics, Vol. 3, pp. 2230 - 2235, 2005
- [42] Yan-Qui Liu, Yi Shen, Zhi-Yan Liu: "A new smoothing particle filter for tracking a maneuvering target", International Conference on Machine Learning and Cybernetics, Vol. 2, pp. 1004 - 1008, 2003
- [43] J. MacCormick, M. Isard: "Partitioned sampling, articulated objects, and interface-quality hand tracking", Springer Verlag Lecture Notes in Computer Science, Vol. 2106, Proceedings of the 6th European Conference on Computer Vision-Part II, 2000
- [44] M. Marron, M. A. Sotelo, J. C. Garcia, D. Fernandez, D. Pizarro: "'XPFCP": An Extended Particle Filter for Tracking Multiple and Dynamic Objects in Complex Environments", Proceedings of the IEEE International Symposium on Industrial Electronics, Vol. 4, pp. 1587 - 1592, 2005.
- [45] K. Nallaperumal, R. Subban, K. Krishnaveni, L. Fred, R. K. Selvakumar: "Human Face Detection in Color Images Using Skin Color and Template Matching Models for Multimedia on the Web", IFIP International Conference on Wireless and Optical Communications Networks, pp. 1 - 5, April 2006

- [46] J. J. Pantrigo, A. S. Montemayor, A. Sanchez: "Local search particle filter applied to human-computer interaction", Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, pp. 279 - 284, 2005.
- [47] V. I. Pavlovic, R. Sharma, T. S. Huang: "Visual interpretation of hand gestures for human-computer interaction: A review", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 19, Issue 7, pp. 677 - 695, July 1997
- [48] S. L. Phung, A. Bouzerdoum, Sr.; D. Chai, Sr.: "Skin segmentation using color pixel classification: analysis and comparison", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, Issue 1, pp. 148 - 154, Jan 2005
- [49] E. Sánchez-Nielsen, L. Antón-Canalís, M. Hernández-Tejera: "Hand Gesture Recognition for Human-Machine Interaction", Journal of WSCG, Vol.12, No.1-3, WSCG'2004, February 2-6, 2003, Plzen, Czech Republic. Copyright UNION Agency - Science Press
- [50] H. Schweitzer, J.W. Bell, F. Wu: "Very Fast Template Matching", Springer Verlag, Lecture Notes in Computer Science, Vol. 2353, 7th European Conference on Computer Vision Proceedings, Part IV, 2002
- [51] J. Segen, S. Kumar: "Human-Computer Interaction using Gesture Recognition and 3D Hand Tracking", International Conference on Image Processing Proceedings, Vol.3, pp. 188 - 192, 1998.
- [52] C. Shan, Y. Wei, X. Qiu, T. Tan: "Gesture recognition using temporal template based trajectories", Proceedings of the 17th International Conference on Pattern Recognition, Vol. 3, pp. 954 - 957, 2004.
- [53] C. Shan, Y. Wei, T. Tan, F. Ojardias: "Real time hand tracking by combining particle filtering and mean shift", Sixth IEEE International Conference on Automatic Face and Gesture Recognition Proceedings, pp. 669 - 674, 2004.
- [54] C. Shen, M. J. Brooks, A. van den Hengel: "Augmented particle filtering for efficient visual tracking", IEEE International Conference on Image Processing, Vol. 3, pp. III - 856-9, 2005
- [55] N. Soontranon, S. Aramvith, T. H. Chalidabhongse: "Improved face and hand tracking for sign language recognition", International Conference on Information Technology: Coding and Computing, Vol. 2, pp. 141 - 146, 2005
- [56] M. Störring, H. J. Andersen, E. Granum: "Skin colour detection under changing lighting conditions", 7th Symposium on Intelligent Robotics Systems, Coimbra, Portugal, pp. 187 - 195, 20-23 July 1999
- [57] A. Ude, T. Shibata and C. G. Atkeson: "Real-time visual system for interaction with a humanoid robot", Elsevier Science Publishers, Robotics

and Autonomous Systems, Volume 37, Issues 2-3, pp. 115 - 125, 30 November 2001

- [58] A. Vacavant, T. Chateau: "Real time head and hands tracking by monocular vision", IEEE International Conference on Image Processing, Vol. 2, pp. II - 302-5, 2005.
- [59] T. Wilhelm, H. -J. Böhme and H. -M. Gross: "A multi-modal system for tracking and analyzing faces on a mobile robot", Elsevier Science Publishers, Robotics and Autonomous Systems, Volume 48, Issue 1, pp. 31-40, 31 August 2004
- [60] W. Yang, Lu Huanzhang, S. Guangfu: "A fast search algorithm for template matching based on inequality criterion", 7th International Conference on Signal Processing Proceedings, Vol. 2, pp. 1211 - 1214, 2004.
- [61] Wu Ying, T. S. Huang: "Nonstationary color tracking for vision-based human-computer interaction", IEEE Transactions on Neural Networks, Volume 13, Issue 4, pp. 948 - 960, July 2002
- [62] Z. Zeng, S. Ma: "Head tracking by active particle filtering", Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002. Proceedings, pp. 82 - 87, 2002
- [63] H. Zhang, W. Huang, Z. Huang, B. Zhang: "A particle filtering framework with indirect measurements for visual tracking", 8th Control, Automation, Robotics and Vision Conference, Volume 1, pp. 723 - 728, 2004
- [64] http://en.wikipedia.org/wiki/Human-computer_interaction
- [65] <http://sigchi.org/cdg/cdg2.html>
- [66] <http://users.rsise.anu.edu.au/~luke/cvcourse.htm>