



Computer vision for the gesture recognition: gesture analysis and stochastic modelling in music interaction

Sotiris Manitsaris

► To cite this version:

Sotiris Manitsaris. Computer vision for the gesture recognition: gesture analysis and stochastic modelling in music interaction. Human-Computer Interaction [cs.HC]. Université de Macédoine, 2010. Greek. NNT: . tel-00551004

HAL Id: tel-00551004

<https://theses.hal.science/tel-00551004>

Submitted on 1 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ



Διδακτορική διατριβή
του ΜΑΝΙΤΣΑΡΗ Σωτηρίου:

**Όραση υπολογιστή για την αναγνώριση
κινήσεων: Ανάλυση και στοχαστική
μοντελοποίηση κινήσεων στη μουσική
αλληλεπίδραση.**

Τριμελής Συμβουλευτική Επιτροπή:

Γ. ΠΕΚΟΣ	Επιβλέπων Καθηγητής
Α. ΚΑΤΟΣ	Καθηγητής
Γ. ΣΤΕΦΑΝΙΔΗΣ	Αναπληρωτής Καθηγητής

Θεσσαλονίκη 2010

Όραση υπολογιστή για την αναγνώριση κινήσεων : Ανάλυση και στοχαστική μοντελοποίηση κινήσεων στη μουσική αλληλεπίδραση.

Περίληψη

Η παρούσα διδακτορική διατριβή παρουσιάζει ένα πρωτότυπο σύστημα όρασης υπολογιστή για την αναγνώριση κινήσεων στη αλληλεπίδραση μεταξύ πιανίστα και μουσικού οργάνου. Η όραση υπολογιστή είναι η μοναδική τεχνολογία που επιτρέπει αναγνώριση κινήσεων, χωρίς περιορισμούς στον πιανίστα και με χαμηλό κόστος. Το σύστημα προτείνει δύο προσεγγίσεις για την αναγνώριση: α) τη στατική προσέγγιση, ή ανάκτηση δακτυλισμών, και β) τη δυναμική προσέγγιση, επέκταση της στατικής προσέγγισης. Η στατική αναγνώριση εφαρμόζεται σε κάθε καρέ του βίντεο. Βασίζεται στην ανάλυση και ερμηνεία χαρακτηριστικών του καρέ, συγκρίνοντάς τα με το αιτιοκρατικό μοντέλο κίνησης. Η δυναμική αναγνώριση εφαρμόζεται σε σύνολο ακολουθιών καρέ του βίντεο. Βασίζεται στην ανάλυση και στοχαστική μοντελοποίηση μέσω Κρυφών Μαρκοβιανών Μοντέλων. Η παρούσα μέθοδος μπορεί να επεκταθεί και σε άλλα πεδία εφαρμογής όπως είναι η παρακολούθηση ανθρώπων με έλλειψη αυτονομίας στο σπίτι, η αξιοποίηση πολιτιστικής κληρονομιάς, η μελέτη της ανθρώπινης συμπεριφοράς και η επικοινωνία ανθρώπου-υπολογιστή.

Λέξεις-Κλειδιά: όραση Η/Υ, μοντελοποίηση, αναγνώριση, κίνηση, σήμα βίντεο, Κρυφά Μαρκοβιανά Μοντέλα, μουσική αλληλεπίδραση.

Vision par ordinateur pour la reconnaissance des gestes : Analyse et modélisation stochastique du geste dans l'interaction musicale.

Résumé

Cette thèse présente un système prototype de vision par ordinateur pour la reconnaissance des gestes dans l'interaction entre le pianiste et l'instrument. La vision par ordinateur est la seule technologie permettant la reconnaissance des gestes, sans interférence entre le pianiste et son instrument, et à un faible coût. Le système propose deux approches pour la reconnaissance : a) l'approche statique, ou reconnaissance des doigtés, et b) l'approche dynamique, extension de l'approche statique. La reconnaissance statique s'applique à chaque image de la vidéo. Elle repose sur l'analyse et l'interprétation des caractéristiques de l'image, en les comparant avec le modèle déterministe du geste. La reconnaissance dynamique s'applique à un ensemble de séquences d'images vidéo. Elle se base sur l'analyse et la modélisation stochastique du geste, à l'aide de Modèles de Markov Cachés. Cette méthode peut être étendue à d'autres champs d'application tels que la surveillance de personnes en perte d'autonomie à domicile, la valorisation du patrimoine culturel, l'étude du comportement humain ou encore l'interaction homme-machine.

Mots-Clés: Vision par ordinateur, modélisation, reconnaissance, geste, signal vidéo, Modèles de Markov Cachés, interaction musicale.

Περιεχόμενα

0 Πρόλογος	12
1 Εισαγωγή	15
1.1 Σύνοψη κεφαλαίου	15
1.2 Το πιάνο	16
1.3 Ερμηνευτικές κινήσεις στο πιάνο	18
1.3.1 Δακτυλικές κινήσεις στο πιάνο	18
1.3.2 Δακτυλισμοί στο πιάνο	18
1.4 Η κίνηση μέσω διαφορετικών προσεγγίσεων	19
1.4.1 Μουσική αλληλεπίδραση	19
1.4.2 Επικοινωνία ανθρώπου υπολογιστή	20
1.4.3 Μουσική μέσω υπολογιστή	21
1.4.4 Μια νέα πρόταση : Όραση υπολογιστή και στοχαστική μοντελοποίηση στη μουσική αλληλεπίδραση	21
1.5 Κίνητρα, σκοπός και στόχοι	23
1.6 Συνεισφορά	24
1.6.1 Πεδία εφαρμογής	25
1.6.1.1 Αξιοποίηση πολιτιστικής κληρονομιάς	25
1.6.1.2 Παρακολούθηση ατόμων με έλλειψη αυτονομίας	25
1.6.1.3 Κατανόηση και βελτίωση αθλητικής συμπεριφοράς και επίδοσης	26
1.6.1.4 Ανάπτυξη μη περιοριστικών λειτουργικών συστημάτων	26
1.7 Διάρθρωση διατριβής	26
2 Ανασκόπηση μεθόδων αναγνώρισης κινήσεων στη μουσική αλληλεπίδραση	29
2.1 Σύνοψη κεφαλαίου	29
2.2 Κίνηση και ανάκτηση πληροφορίας	30
2.3 Υπολογισμός κινήσεων	34
2.3.1 Βιολογική μηχανική	34
2.3.2 Ενεργή όραση	36
2.3.3 Συμπεράσματα	37
2.4 Αναγνώριση κινήσεων	39
2.4.1 Όραση Η/Υ και Κρυφά Μαρκοβιανά Μοντέλα σε διάφορα πεδία εφαρμογής	39
2.4.2 Αισθητήρες	43
2.4.2.1 Δί' επαφής	44
2.4.2.2 Εξ αποστάσεως	46
2.4.3 Συμπεράσματα	49
2.5 Ανάκτηση δακτυλισμών	50
2.5.1 Ιδιαιτερότητες στην τεχνική του πιάνου και της κιθάρας	51
2.5.2 Θεωρία γράφων στην a priori επεξεργασία της παρτιτούρας	52
2.5.2.1 Μέθοδος Radicioni	53
2.5.2.2 Μέθοδος Kasimi	56
2.5.3 Τεχνολογία MIDI στην επεξεργασία σε πραγματικό χρόνο	59
2.5.4 Όραση Η/Υ στην κιθάρα στην επεξεργασία σε πραγματικό χρόνο	61
2.5.5 Ψηφιακή επεξεργασία ηλεκτρικού σήματος στην a posteriori επεξεργασία	64
2.5.6 Συμπεράσματα	65
2.6 Γενικά συμπεράσματα	66

3 Όραση υπολογιστή	69
3.1 Σύνοψη κεφαλαίου	70
3.2 Ανθρώπινη όραση	70
3.3 Κατανόηση ανθρώπινης όρασης	71
3.4 Ανάπτυξη μεθόδου όρασης H/Y	72
3.5 Η πρόταση του David Marr	73
3.5.1 Το μοντέλο David Marr στην αναγνώριση δακτυλικών κινήσεων	74
3.6 Μέθοδοι εντοπισμού χεριού στην εικόνα	75
3.6.1 Εντοπισμός μέσω τμηματοποίησης	75
3.6.1.1 Εντοπισμός μέσω υπολογισμού διαφοράς.....	75
3.6.1.2 Εντοπισμός μέσω ανίχνευσης δέρματος	76
3.6.2 Εντοπισμός μέσω εμφάνισης	80
3.6.3 Συμπεράσματα	81
3.7 Τμηματοποίηση χεριού.....	82
3.7.1 Μαθηματική μορφολογία και ανάκτηση φιγούρας χεριού	83
3.7.1.1 Διαστολή και συστολή	84
3.7.1.2 Άνοιγμα και κλείσιμο	85
3.7.2 Ανάκτηση περιγράμματος χεριού	87
3.7.2.1 Γραμμικό φιλτράρισμα εικόνας.....	88
3.7.2.2 Κλίση μιας εικόνας	89
3.7.2.3 Γεωμετρική ερμηνεία της κλίσης.....	89
3.7.2.4 Μοντέλα περιγράμματος.....	91
3.7.2.5 Κριτήρια απόδοσης.....	93
3.8 Εντοπισμός άκρων δακτύλων.....	95
3.8.1 Αλγόριθμοι προβολής υπογραφών	96
3.8.2 Αλγόριθμοι κυκλικού μετασχηματισμού Hough.....	96
3.8.3 Αλγόριθμοι εξαγωγής γεωμετρικών χαρακτηριστικών	99
3.8.4 Αλγόριθμοι χρωματικών δεικτών.....	100
3.8.5 Συμπεράσματα	100
3.9 Δυναμική αναγνώριση κινήσεων.....	103
3.9.1 Δυναμική Περιτύλιξη του Χρόνου.....	103
3.9.2 Νευρωνικά Δίκτυα	104
3.9.3 Κρύφα Μαρκοβιανά Μοντέλα.....	104
3.10 Γενικά συμπεράσματα	105
 4 Μεθοδολογία για την αναγνώριση κινήσεων	108
4.1 Σύνοψη κεφαλαίου	108
4.2 Ορισμός κίνησης.....	110
4.3 Μοντελοποίηση κινήσεων	112
4.3.1 Χρονική μοντελοποίηση.....	112
4.3.2 Χωρική μοντελοποίηση.....	113
4.4 Δημιουργία δερματικού μοντέλου	114
4.5 Εντοπισμός περιοχής ενδιαφέροντος	116
4.6 Επεξεργασία εικόνας.....	116
4.6.1 Τμηματοποίηση χεριού	117
4.6.2 Εξαγωγή χαρακτηριστικών διανυσμάτων	118
4.7 Αναγνώριση κινήσεων	118
4.8 Συμπεράσματα.....	119
 5 Σύστημα στατικής αναγνώρισης δακτυλικών κινήσεων	122
5.1 Σύνοψη κεφαλαίου	122
5.2 Επιλογή εξοπλισμού	123
5.2.1 Ηλεκτρονικός υπολογιστής.....	123
5.2.2 Πιάνο-συνθεσάιζερ	124

5.2.3 Κάμερα	124
5.2.4 Λογισμικό.....	127
5.3 Προκαταρκτική ανάλυση	128
5.3.1 Τοποθέτηση κάμερας	128
5.3.2 Γωνία λήψης	128
5.3.3 Φωτισμός.....	129
5.3.4 Σύλληψη βίντεο	130
5.4 Ανίχνευση δέρματος.....	130
5.4.1 Χαρακτηριστικά ανθρώπινου δέρματος.....	131
5.4.2 Προσδιορισμός περιοχής ενδιαφέροντος.....	132
5.4.3 Κανονικοποίηση περιοχής ενδιαφέροντος.....	135
5.4.4 Δημιουργία δερματικού μοντέλου	135
5.4.5 Χρήση μοντέλου για ανίχνευση δέρματος σε εικόνα	139
5.5 Τμηματοποίηση χεριού.....	141
5.5.1 Απλοποίηση δυαδικής μάσκας.....	143
5.5.1.1 Μείωση θορύβου με σειριακό εναλλασσόμενο φίλτράρισμα.....	144
5.5.1.2 Ανάκτηση φιγούρας χεριού.....	147
5.5.2 Διαχωρισμός χεριού σε βασικές συνιστώσες.....	151
5.5.2.1 Ανάκτηση περιγράμματος χεριού	152
5.5.2.2 Ανάκτηση άκρων δακτύλων.....	156
5.6 Εντοπισμός δακτύλων	157
5.6.1 Υπολογισμός κεντροειδούς.....	158
5.6.2 Καθορισμός σημείου αναφοράς.....	159
5.6.3 Εντοπισμός θέσης.....	161
5.7 Ανάκτηση δακτυλισμών.....	165
5.8 Συμπεράσματα.....	166

6 Σύστημα δυναμικής αναγνώρισης δακτυλικών κινήσεων με Κρυφά Μαρκοβιανά Μοντέλα	169
6.1 Σύνοψη κεφαλαίου	169
6.2 Επιλογή κινήσεων προς αναγνώριση	171
6.3 Χαρακτηριστικές ιδιότητες και διανύσματα παρατήρησης	181
6.4 Ο ρόλος των Κρυφών Μαρκοβιανών Μοντέλων στο σύστημα αναγνώρισης δακτυλικών κινήσεων.....	186
6.4.1 KMM σε διακριτό ή σε συνεχή χώρο παρατηρήσεων	188
6.4.2 Επιλογή συνεχών KMM.....	191
6.5 Ανάλυση και στοχαστική μοντελοποίηση κινήσεων με Κρυφά Μαρκοβιανά Μοντέλα	192
6.6 Επίλυση των προβλημάτων των Κρυφών Μαρκοβιανών Μοντέλων.....	194
6.6.1 Μοντελοποίηση διανυσμάτων παρατήρησης με Μεικτά Γκαουσιανά Μοντέλα	197
6.6.2 Εύρεση Μεικτού Γκαουσιανού Μοντέλου με τη μέγιστη πιθανοφάνεια	201
6.6.3 Παράμετροι Μεικτού Γκαουσιανού Μοντέλου με τη μέγιστη πιθανοφάνεια.....	203
6.6.4 Αλγόριθμος « forward » του Baum.....	205
6.6.5 Αλγόριθμος του Viterbi.....	207
6.6.6 Επανεκτίμηση των παραμέτρων Κρυφών Μαρκοβιανών Μοντέλων και Μεικτών Γκαουσιανών Μοντέλων κατά Baum-Welch	209
6.7 Παράδειγμα αναγνώρισης μεμονωμένων κινήσεων	213
6.8 Υλοποίηση δυναμικής αναγνώρισης δακτυλικών κινήσεων με το MATLAB... ..	221
6.9 Συμπεράσματα.....	222

7 Λογισμικό PianOrasis και αξιολόγηση.....	225
7.1 Σύνοψη κεφαλαίου	225
7.2 PianOrasis	225
7.2.1 Διεπαφή PianOrasis	226

7.2.2 Στατική αναγνώριση μέσω PianOrasis.....	230
7.2.3 Δυναμική αναγνώριση μέσω PianOrasis	230
7.3 Πρώτη φάση αξιολόγησης	233
7.4 Δεύτερη φάση αξιολόγησης	237
7.5 Τρίτη φάση αξιολόγησης	243
7.6 Τέταρτη φάση αξιολόγησης	245
7.7 Συμπεράσματα.....	247
 8 Γενικά συμπεράσματα.....	250
8.1 Σύνοψη κεφαλαίου	250
8.2 Περίληψη συνεισφοράς.....	250
8.3 Περιορισμοί και επέκταση της μεθόδου	251
8.4 Πρωτοτυπία και ανασκόπηση διατριβής.....	252
8.5 Συμπεράσματα.....	259
 Βιβλιογραφία	261
 Παράρτημα Α : Κώδικας PianOrasis	275
 Παράρτημα Β : « Το τριαντάφυλλο του μικρού πρίγκηπα ».....	297

Κατάλογος σχημάτων

2 Ανασκόπηση μεθόδων αναγνώρισης κινήσεων στη μουσική αλληλεπίδραση

2.2.1 Επιμέρους ερευνητικά πεδία της ανάκτησης πληροφορίας στη MA.....	30
2.2.2 Κατηγορίες ερευνών και περιοχές του σώματος στη MA.....	31
2.2.3 Κατηγορίες ερευνών και περιοχές, θεωρίες και τεχνολογίες στη MA	33
2.3.1 Εξέλιξη βιβλιογραφίας υπολογισμού κινήσεων στη MA	34
2.3.2.1 Διεπαφή Vicon Peak	37
2.3.2.2 Τρισδιάστατο σκελετικό μοντέλο και σημειακοί δείκτες.....	38
2.4.1.1 Μέθοδος αναγνώρισης κινήσεων σε πληκτρολόγιο	42
2.4.2.1 Εξέλιξη συστημάτων αναγνώρισης κινήσεων στη MA	44
2.4.2.1.1 Το Miburi στο Test Studio της Yamaha στο Τόκιο	46
2.4.2.2.1 Εικόνα του Virtual Conductor στο Vienna's House of Music.....	48
2.4.2.2.2 Πλάτος σήματος κινήσεων σε συνάρτηση με το χρόνο	48
2.4.2.2.3 Το Wiimote	49
2.4.2.2.4 Παράδειγμα KMM 5 καταστάσεων για μοντελοποίηση κίνησης στο Wiimote	49
2.5.1 Εξέλιξη συστημάτων ανάκτησης δακτυλισμών στη MA.....	51
2.5.2 Θέση στο τάστο της κιθάρας και χορδή	52
2.5.2.1 Ανάκτηση δακτυλισμών με a priori επεξεργασία παρτιτούρας.....	53
2.5.2.1.1 Παράδειγμα γράφου που παράχθηκε από τις νότες μι-φα-λα.....	56
2.5.2.2.1 Απόσπασμα του Aufschwung και κατακόρυφη ανάλυσή του	57
2.5.2.2.2 Παράδειγμα για οριζόντιο βάρος.....	58
2.5.4.1 Γενική όψη και όψη αριστερού χεριού στην επεξεργασία σε πραγματικό χρόνο	62
2.5.4.2 Βάση κάμερας και όψη δακτύλων	63
2.5.5.1 Θεωρητικό φάσμα και πραγματικό φάσμα στην a posteriori επεξεργασία	65

3 Όραση υπολογιστή

3.7.1.1.1 Διεσταλμένο του αντικειμένου A ως προς το δομικό στοιχείο B.....	85
3.7.1.1.2 Συνεσταλμένο του αντικειμένου A ως προς το δομικό στοιχείο B	85
3.7.1.2.1 Μορφολογικό άνοιγμα και κλείσιμο	86
3.7.1.2.2 Εφαρμογή μορφολογικών μετασχηματισμών σε πίνακα του Henri Metisse « La femme à l'amphore »	87
3.7.1.2.3 Εφαρμογή μορφολογικών μετασχηματισμών	88
3.7.2.3.1 Δύο « λείες » περιοχές που χωρίζονται από περίγραμμα « σκαλοπατιού »	90
3.7.2.4.1 Είδη περιγραμμάτων : « σκαλοπάτι », « ράμπα », « σκεπή », « κορυφή »	92
3.8.1.1 Αλγόριθμος εντοπισμού δακτύλων χρησιμοποιώντας προβολές υπογραφών ...	97
3.8.2.1 Αλγόριθμος εντοπισμού δακτύλων χρησιμοποιώντας κυκλικό μετασχηματισμό « Hough »	98
3.8.3.1 Αλγόριθμος εντοπισμού δακτύλων χρησιμοποιώντας ευκλείδειες αποστάσεις	101
3.8.4.1 Αλγόριθμος εντοπισμού δακτύλων χρησιμοποιώντας χρωματικούς δείκτες ...	102

4 Μεθοδολογία για την αναγνώριση κινήσεων 108

4.3.1 Μεθοδολογία αναγνώρισης δακτυλικών κινήσεων.....	113
--	-----

5 Σύστημα στατικής αναγνώρισης δακτυλικών κινήσεων 122

5.2.1.1 iMac 24	124
5.2.2.1 Συνθεσάιζερ Yamaha CS1X	125

5.2.3.1 Κάμερα της Logitech με ειδική λαστιχωτή βάση	126
5.3.2.1 Λήψη μπροστά από κλαβιέ και λήψη από την κορυφή του κλαβιέ	129
5.4.1 Αλγόριθμος δημιουργίας δερματικού μοντέλου	131
5.4.2.1 Ψηφιακή εικόνα PE^{RGB} που αποτελεί την περιοχή ενδιαφέροντος.....	134
5.4.4.1 Γραφική παράσταση της χρωματικής περιοχής του δέρματος στον rg	137
5.4.4.2 Γραφική αναπαράσταση της χρωματικής περιοχής του δέρματος στον rg	138
5.4.4.3 Απεικόνιση χρωματικών συνιστωσών των εικονοστοιχείων της PE^{rg}	138
5.4.5.1 Ανίχνευση δέρματος στα καρτέ της ακολουθίας του βίντεο	139
5.4.5.2 Ανίχνευση δέρματος υπό διαφορετικές συνθήκες φωτισμού	141
5.5.1 Διαδικασία τμηματοποίησης εικόνας χεριού	142
5.5.2.1 Καρτέ εικόνας βίντεο και δυαδική μάσκα μέσω δερματικού μοντέλου	143
5.5.1.1.1 Εφαρμογή ΣΕΦ στη δυαδική μάσκα	146
5.5.1.2.1 Πρώτη εφαρμογή διαστολής μέσω φίλτρου min	147
5.5.1.2.2 Πρώτη εφαρμογή φίλτρου Gauss	148
5.5.1.2.3 Δεύτερη εφαρμογή διαστολής μέσω φίλτρου min.....	149
5.5.1.2.4 Δεύτερη εφαρμογή φίλτρου Gauss	150
5.5.1.2.5 Εφαρμογή κατωφλίωσης	151
5.5.2.1.1 Κατακόρυφη σάρωση της χ_λ στη θέση μ_0	153
5.5.2.1.2 Ενδεικτική απεικόνιση των τιμών των εικονοστοιχείων.....	154
5.5.2.1.3 Ενδεικτική κλίση της χ_λ στη θέση μ_0	155
5.5.2.1.4 Ανάκτηση περιγράμματος μέσω υπολογισμού της κλίσης της χ_λ	156
5.5.2.2.1 Ανάκτηση άκρων δακτύλων στην α_λ	157
5.6.1.1 Εντοπισμός εικονοστοιχείου που αντιστοιχεί στο κεντροειδές και ανάκτηση άκρων δακτύλων	160
5.6.3.1 Αναπαράσταση διαδικασίας εντοπισμού θέσης δακτύλων	162
5.6.3.2 Αλγόριθμος εντοπισμού θέσης άκρων και ταυτοποίησης δακτύλων	164
5.6.3.3 Προβολή τοπικών μεγίστων σε διάγραμμα εικονοστοιχείου-απόστασης.....	165

6 Σύστημα δυναμικής αναγνώρισης δακτυλικών κινήσεων με Κρυφά Μαρκοβιανά Μοντέλα	169
6.2.1 Ανιούσα κίνηση κλίμακας ντο μείζονα με δακτυλισμούς δεξιού χεριού ...	173
6.2.2 Νότες της ανιούσας κίνησης, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας	174
6.2.3 Κατιούσα κίνηση κλίμακας ντο μείζονα με δακτυλισμούς δεξιού χεριού ..	174
6.2.4 Νότες της κατιούσας κίνησης, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας	175
6.2.5 Ανιούσα κίνηση αρπέζ ντο μείζονα με δακτυλισμούς δεξιού χεριού	176
6.2.6 Νότες της ανιούσας κίνησης αρπέζ, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας	176
6.2.7 Κατιούσα κίνηση αρπέζ ντο μείζονα με δακτυλισμούς δεξιού χεριού	177
6.2.8 Νότες της κατιούσας κίνησης αρπέζ, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας	178
6.2.9 Ανιούσα κίνηση σε τρίτες κλίμακας ντο μείζονα με δακτυλισμούς δεξιού χεριού	178
6.2.10 Νότες της ανιούσας κίνησης αρπέζ, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας	179

6.2.11 Ανιούσα κίνηση σε τρίτες κλίμακας ντο μείζονα με δακτυλισμούς δεξιού χεριού	180
6.2.12 Νότες της κατιούσας κίνησης αρπέζ, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας	181
6.3.1 Χαρακτηριστικά κίνησης και ιδιότητές τους	184
6.4.1.1 Είδη KMM ανάλογα με τους περιορισμούς	188
6.4.1.2 Γραφική αναπαράσταση KMM ανά είδος.....	189
6.4.2.1 Συγκριτικός πίνακας μεταξύ συνεχών και διακριτών KMM	192
6.6.1.1 Παράδειγμα πραγματικής κατανομής και προσέγγισής της από ΜΓΜ	198
6.6.1.2 Παράδειγμα ΜΓΜ και περιγράμματα της κατανομής	199
6.6.1.3 Συνάρτηση πιθανοφάνειας για συνεχή μέτρηση του	200
6.6.1.4 Προσέγγιση συνάρτησης πιθανοφάνειας για συνεχή μέτρηση με ΜΓΜ	202
6.6.5.1 Επιλογή του μονοπατιού εσωτερικά του γράφου Treillis	210
6.7.1 Αλφαβητάρι δακτυλισμών	215
6.7.2 Πίνακας μεταβάσεων με όλες τις δυνατές μεταβάσεις μεταξύ των καταστάσεων	215
6.7.3 Νότες και δακτυλισμοί της κίνησης	216
6.7.4 Αναπαράσταση των διαδικασιών μέσω ενός γράφου Treillis	217
6.7.5 Αλγόριθμος « forward » του Baum	218
6.7.6 Αλγόριθμος του Viterbi	220
6.7.7 Διάγραμμα αναγνωριστή KMM για μεμονωμένες κινήσεις	221

7 Λογισμικό PianOrasis και αξιολόγηση..... 225

7.2.1.1 PianOrasis	227
7.2.1.2 Μενού « File » και « Edit »	228
7.2.1.3 Δήλωση ακολουθίας καρτέ προς αναγνώριση	228
7.2.1.4 Επιλογή λειτουργιών KMM στο PianOrasis.....	230
7.2.3.1 Παράδειγμα οργάνωσης φακέλων εκπαίδευσης κινήσεων των KMM	232
7.3.1 Πρώτο μουσικό απόσπασμα αξιολόγησης.....	233
7.3.2 Έξοδοι ανίχνευσης και ανάκτησης από το PianOrasis	234
7.3.3 Όψη PianOrasis για το καρτέ 112	236
7.3.4 Ανίχνευση δακτύλων χωρίς πρόβλεψη σε 284 καρτέ	237
7.4.1 Δεύτερο μουσικό απόσπασμα από τη σονάτα No.16 του Wolfgang Amadeus Mozar	239
7.4.2 Έξοδοι ανίχνευσης και ανάκτησης από το PianOrasis για τα καρτέ 106 έως 113	240
7.4.3 Έξοδοι ανίχνευσης και ανάκτησης από το PianOrasis για τα καρτέ 180 έως 187	241
7.4.4 Έξοδοι ανίχνευσης και ανάκτησης από το PianOrasis για τα καρτέ 222 έως 229	241
7.4.5 Έξοδοι ανίχνευσης και ανάκτησης από το PianOrasis για τα καρτέ 284 έως 291	242
7.4.6 Ανίχνευση δακτύλων χωρίς πρόβλεψη σε 386 καρτέ	242
7.5.1 Όψη PianOrasis σε λειτουργία εικονικού πιάνου.....	244
7.6.1 Αποτελέσματα δυναμικής αναγνώρισης μέσω PianOrasis	246

0 Πρόλογος και ευχαριστίες

Η χειρονομία είναι μια συνιστώσα της ανθρώπινης έκφρασης. Αναφέρεται σε μια κίνηση του σώματος, ή ενός μέλους αυτού, η οποία ξεδιπλώνεται στο χώρο και το χρόνο. Είναι εύκολο να περιγράψει κάποιος μια χειρονομία ή ακόμα και να σχεδιάσει το σκαρίφημά της. Το σκαρίφημα όμως αυτό, δε θα μπορούσε να εκδηλωθεί χωρίς την παράμετρο του χρόνου.

Η χειρονομία δεν περιορίζεται σε μια κίνηση του σώματος και προφανώς κάθε κίνηση του σώματος δεν αποτελεί χειρονομία. *«Η χειρονομία καθορίζεται ως μια σκόπιμη κίνηση, λιγότερο ή περισσότερο σύνθετη, η οποία έχει ένα στόχο που της αποδίδει μια ιδιαίτερη έννοια, κοινωνική ή ιστορική»* [Cuen 2003].

Ο άνθρωπος δοκιμάζει τις αισθήσεις του καθημερινά, δημιουργώντας έτσι μια εσωτερική αναπαράσταση του σώματός του. Αυτές οι «δοκιμές» του σώματος αποτελούν ένα πολύπλοκο χωρο-χρονικό σύστημα αναφοράς. Ταυτόχρονα, αποτελούν ένα σύνολο δυναμικών πληροφοριών που εξωτερικεύονται μέσω της χαλάρωσης και της εκτόνωσης του νευρομυϊκού συστήματός του. Οι πληροφορίες αυτές επιτρέπουν την ταυτοποίηση ή την *αναγνώριση* βαθύτερων εμπειριών του σώματος, καθώς και των αναφορών του με το φυσικό περιβάλλον.

Ένα εξαιρετικό παράδειγμα, που φωτίζει τις εσωτερικές διεργασίες των χειρονομιών του σώματος, είναι η παρατήρηση ενός μίμου. Σύμφωνα με τον Etienne Decroux (1994), ο μίμος τείνει να αναδεικνύει εμπειρίες ζωής μέσα από μια κίνηση. Η τέχνη αυτή δεν περιορίζεται σε μια απλή μίμηση μορφών ή δομών. Αυτό που βρίσκεται μέσα στη μορφή είναι το σημαντικό. Ο μίμος ενδιαφέρεται για τη συγκίνηση του πνεύματος, αναπαριστώντας κρυφά συναισθήματα, επιδέξια δοσμένα μέσα από μια σωματική κίνηση. Ο πυρήνας αυτών των ευαίσθητων αναπαραστάσεων είναι η συνοχή μεταξύ χειρονομίας και σκέψης. Η συνοχή αυτή αποτελεί την πραγματική έννοια της έκφρασης. «Η μάζα του σώματος» οφείλει να ακολουθεί «τις κινήσεις της ψυχής», με τέτοιο τρόπο ώστε η χειρονομία να αναδεικνύεται

ως μια δυναμική διαδικασία στο χρόνο, και όχι ως μια προκαθορισμένη δομή [Decroux 1994].

«Ακούω κάποιον εκφωνητή στο ραδιόφωνο. Ακούω προσεκτικά τη φωνή του. Μπορώ μέσα από αυτό το άκουσμα να φανταστώ την έκφραση του προσώπου του ή την κίνηση του σώματός του. Κι όμως, δεν τον βλέπω, απλώς τον ακούω από το ραδιόφωνο». Αυτή η απλοϊκή εμπειρία αποδεικνύει την ύπαρξη ενός συνδέσμου μεταξύ *κίνησης* και *ήχου* μέσα σε ένα συγ-κινησιακό περιβάλλον. Για να μπορέσουν οι ζωτικές συγκινήσεις να αποτελέσουν την ψυχολογική βάση της μουσικής εκφραστικότητας, θα πρέπει να μπορούν να *αναγνωρίζονται* ηχητικά/οπτικά από τον ακροατή/θεατή μέσω ενός στιγμιαίου πάθους, το οποίο δίνει ζωή στη *μουσική κίνηση*. Ο Μ. Imberty (1997) απέδειξε ότι η *συγκινησιακή ζωτικότητα* αντιστοιχεί σε *μουσικά γεγονότα*, τα οποία κάθε φορά προκαλούν τον ακροατή να τα ερμηνεύσει.

Μέσα σε ένα πλαίσιο ανταλλαγής γνώσεων και εμπειριών που ισορροπούσαν ανάμεσα στα μαθηματικά, την πληροφορική και τη μουσική, ολοκλήρωση τη διδακτορική μου διατριβή. Αυτά τα τρία χρόνια ήταν ζωτικά για μένα τόσο σε πνευματικό όσο και σε προσωπικό επίπεδο. Αυτό το οφείλω σε ένα σύνολο συγκυριών και πρόσωπων που συγχρονίστηκαν με τις προσδοκίες και τους στόχους μου. Μέσω της ιστορίας «*του μικρού Νικόλα στο διδακτορικό*» επιθυμώ να απαντήσω σε όλους αυτούς που κατά καιρούς με ρωτούν για τη δουλειά της έρευνας. Η ιστορία αυτή δεν αναφέρεται στο αντικείμενο της έρευνας του διδακτορικού μου, ούτε σε πραγματικά γεγονότα. Αναφέρεται στη ζωή του υποψηφίου διδάκτορα γενικότερα.

Σε αυτό το σημείο οφείλω να ευχαριστήσω τα μέλη της τριμελούς και επταμελούς επιτροπής για το χρόνο που αφιέρωσαν να διαβάσουν και να αξιολογήσουν τη δουλειά του διδακτορικού μου. Ειδικότερα, ευχαριστώ τον κύριο Γεώργιο Πέκο για το γεγονός ότι αποδέχτηκε την επίβλεψη του διδακτορικού μου. Τον ευχαριστώ επίσης για την εμπιστοσύνη που μου έδειξε στην επιλογή του θέματος αλλά και στη δρομολόγηση και διεξαγωγή της παρούσας έρευνας. Εργάστηκα σε ένα βαθιά ανθρώπινο περιβάλλον, γεγονός το οποίο μου επέτρεψε την ταχεία ολοκλήρωση της διατριβής μου. Αποτελεί για μένα τιμή το γεγονός ότι συνεργάστηκα μαζί του.

Στη συνέχεια θέλω να ευχαριστήσω τον κύριο Αναστάσιο Κάτο για τη μετάδοση της μεθοδολογικής του εμπειρίας. Η δομή και η διάρθρωση μιας διδακτορικής διατριβής είναι αναπόσπαστο κομμάτι της έρευνας. Η μακροχρόνια εκπαιδευτική και ερευνητική εμπειρία του κυρίου Κάτου αποδείχθηκαν παράγοντες καθοριστικοί

για την οργάνωση και την απόδοση της δουλειάς μου. Η αποτύπωση αυτής του της εμπειρίας θα αποτελέσει για μένα οδηγό και στη μετέπειτα ερευνητική μου πορεία.

Ο κύριος Γεώργιος Στεφανίδης αποτέλεσε για μένα πραγματικό δάσκαλο επί σειρά ετών. Παρακολουθεί τη μορφωτική μου πορεία από τα μαθητικά μου χρόνια μέχρι και σήμερα. Είναι ο θεμελιωτής της μαθηματικής μου κουλτούρας. Μιας κουλτούρας που χαρακτηρίζει την προσέγγιση του διδακτορικού μου από τη σκοπιά της πληροφορικής, αλλά και της μουσικής. Επιθυμώ να τον ευχαριστήσω για το γεγονός ότι μου παρείχε απαντήσεις σε πολλές μαθηματικές μου αναζητήσεις επί σειρά ετών.

Στη συνέχεια επιθυμώ να ευχαριστήσω την κυρία Βίκυ Τσεκουροπούλου, καθηγήτριά μου στις θεωρητικές μουσικές σπουδές στο Δημοτικό Ωδείο Καλαμαριάς, για τη στήριξη που μου παρείχε επιτρέποντάς μου την πρόσβαση σε εμπνεύσεις, ιδέες, κριτικές και συμβουλές επί σειρά ετών. Ευχαριστώ επίσης τον κύριο Κοσμά Δημητρόπουλο για την αμέριστη παροχή της προγραμματιστικής του εμπειρίας στην ανάπτυξη του λογισμικού PianOrasis.

Τέλος, οφείλω να ευχαριστήσω την οικογένειά μου, η οποία με υποστήριξε ηθικά, και όχι μόνο, σε όλη τη διάρκεια των σπουδών μου. Υπήρξε αρωγός και υποστηρικτής των ερευνητικών και μορφωτικών μου αναζητήσεων παρέχοντάς μου με κόπο όλα τα απαραίτητα μέσα για την επίτευξη των στόχων μου.

*Ευχαριστώ πολύ
ΜΑΝΙΤΣΑΡΗΣ Σωτήριος*

1

Εισαγωγή

1.1 Σύνοψη κεφαλαίου**1.2 Το πιάνο****1.3 Ερμηνευτικές κινήσεις στο πιάνο**

1.3.1 Δακτυλικές κινήσεις στο πιάνο

1.3.2 Δακτυλισμοί στο πιάνο

1.4 Η κίνηση μέσω διαφορετικών προσεγγίσεων

1.4.1 Μουσική αλληλεπίδραση

1.4.2 Επικοινωνία ανθρώπου υπολογιστή

1.4.3 Μουσική μέσω υπολογιστή

1.4.4 Μια νέα πρόταση: Όραση υπολογιστή και στοχαστική μοντελοποίηση στη μουσική αλληλεπίδραση

1.5 Κίνητρα, σκοπός και στόχοι**1.6 Συνεισφορά**

1.6.1 Πεδία εφαρμογής

1.6.1.1 Αξιοποίηση πολιτιστικής κληρονομιάς

1.6.1.2 Παρακολούθηση ατόμων με έλλειψη αυτονομίας

1.6.1.3 Κατανόηση και βελτίωση αθλητικής συμπεριφοράς και επίδοσης

1.6.1.4 Ανάπτυξη μη περιοριστικών λειτουργικών συστημάτων

1.7 Διάρθρωση διατριβής**1.1 Σύνοψη κεφαλαίου**

Η παρούσα διατριβή παρουσιάζει μια πρωτότυπη μέθοδο ανάκτησης πληροφορίας σχετικά με τους δακτυλισμούς και την αναγνώριση των κινήσεων των δακτύλων του δεξιού χεριού ενός πιανίστα, χρησιμοποιώντας όραση υπολογιστή και μαθηματική μοντελοποίηση. Το κεφάλαιο αυτό περιγράφει συνοπτικά τα κίνητρα τα οποία οδήγησαν στην επιλογή του πιάνου ως το προς διερεύνηση μουσικό όργανο, της αναγνώρισης κίνησης των δακτύλων του δεξιού χεριού ως περιοχή ενδιαφέροντος, καθώς και τα επιστημονικά πεδία της όρασης υπολογιστών και της μαθηματικής μοντελοποίησης ως μέθοδο απευθείας αναγνώρισης ανθρωπίνων χειρονομιών. Η ενότητα 1.2 περιγράφει σύντομα την ιστορία του πιάνου επεξηγώντας πως η ευρεία χρήση του σε διαφορετικά στυλ μουσικής και πολιτισμού, οδήγησε στην επιλογή του ως όργανο εφαρμογής της διατριβής. Η ενότητα 1.3 ορίζει και επεξηγεί συνοπτικά την έννοια των δακτυλικών κινήσεων σε μουσικό όργανο και ειδικότερα στο πιάνο. Στην ενότητα 1.4 παρουσιάζει

ζονται διαφορετικές ερευνητικές προσεγγίσεις της έννοιας της ανθρώπινης κίνησης στη μουσική αλληλεπίδραση. Η πρώτη προσέγγιση προέρχεται από τη μουσική κοινότητα δια μέσω της Μουσικής Αλληλεπίδρασης (ΜΑ), η δεύτερη από τον τομέα της Επικοινωνίας Ανθρώπου Υπολογιστή (ΕΑΥ), και η τρίτη από τον τομέα της Μουσικής μέσω Υπολογιστή (ΜμΥ). Τέλος, παρουσιάζεται επιγραμματικά, ενώ αναλύεται σε επόμενο κεφάλαιο, μια νέα ερευνητική προσέγγιση χρησιμοποιώντας όραση υπολογιστή και μαθηματική μοντελοποίηση στη μουσική αλληλεπίδραση. Στην ενότητα 1.5 σημειώνεται η ανάγκη για έρευνα στο συγκεκριμένο επιστημονικό πεδίο καθώς και καθορίζονται ο σκοπός και οι στόχοι της παρούσας έρευνας. Στην ενότητα 1.6 περιγράφεται η πιθανή συνεισφορά της διατριβής σε έρευνες διαφόρων επιστημονικών πεδίων, ενώ τέλος, στην ενότητα 1.7 παρουσιάζεται η δομή της διδακτορικής διατριβής.

1.2 Το πιάνο

Ο παρακάτω ορισμός για το πιάνο είναι δημοσιευμένος σε σελίδα της ηλεκτρονικής εγκυκλοπαίδειας Wikipedia:

«Το πιάνο είναι ένα σόλο μουσικό όργανο, που μπορεί να θεωρηθεί έγχορδο ή κρουστό. Παίζεται με πλήκτρα, τα οποία όταν πατηθούν από τον πιανίστα σηκώνουν σφυράκια που χτυπούν τις χορδές του, παράγοντας έτσι ήχους. Η δυνατότητα να δίνει μια διαφορετική νότα από το κάθε δάχτυλο και να κάνει κάθε νότα απαλή ή δυνατή, δίνει στο πιάνο μια εκπληκτική ποικιλία έκφρασης. Το πιάνο μπορεί να αποδώσει μουσική είτε ως σόλο όργανο, είτε μέσα σε μια ορχήστρα. Αν και πολλοί πιστεύουν πώς χρησιμοποιείται κυρίως στην κλασική μουσική, το πιάνο κατέχει έναν ιδιαίτερα σημαντικό ρόλο και στην τζαζ, την μπλουζ και το ροκ εν ρολ, καθώς και στη λαϊκή μουσική όπου είτε κυριαρχεί είτε λειτουργεί ως βοηθητικό για άλλα όργανα.»

Είναι εντυπωσιακό το πλήθος των ανθρώπων που χρησιμοποιούν το πιάνο σε μια μεγάλη ποικιλία μουσικών στυλ και πολιτισμών ως σόλο, συνοδευτικό ή ορχηστρικό όργανο. Ακόμα και μόνο αυτό το στοιχείο θα αρκούσε για να το αναδείξει σε ενδιαφέρουσα ερευνητική περιοχή. Οι μουσικοί που παίζουν πιάνο μπορεί να είναι επαγγελματίες πιανίστες ή και αυτοδίδακτοι δίχως συγκεκριμένη

μουσική κατάρτιση. Το πιάνο συμμετέχει σε διάφορα στυλ μουσικής όπως η κλασσική, το φλαμένκο, μπλουζ, ποπ, τζαζ, ροκ κλπ.

Το πιάνο, με τη μορφή που έχει σήμερα, έχει πίσω του μια ιστορία έρευνας και εξέλιξης περίπου τριακοσίων ετών. Οι πρόγονοί του υπήρξαν το κλαβίχορδο (*Clavichord*) και το τσέμπαλο (*Cembalo*). Το παλαιότερο δείγμα τύπου πιάνου κατασκευάστηκε στη Φλωρεντία από τον Μπαρτολομέο Κριστοφόρι το 1709. Στόχος του ήταν να ικανοποιήσει την επιθυμία των μουσικών της εποχής του να παίζουν σε μια ευρεία δυναμική γκάμα, από πολύ σιγά έως πολύ δυνατά.

«Σήμερα, το σύγχρονο πιάνο έχει σιδερένιο σκελετό και είναι είτε όρθιο είτε με ουρά. Η έκτασή του είναι 7^{1/3} οκτάβες με 88 πλήκτρα. Το πιάνο αποτελείται από ηχείο, σώμα υποστήριξης, πλαίσιο, χορδές, πληκτρολόγιο, πεντάλ και βασικό μηχανισμό. Πάνω στο πλαίσιο ακουμπά το ηχείο κατασκευασμένο από ξύλο ελάτου και μαντέμι. Σε ένα πιάνο υπάρχουν συνήθως 52 λευκά πλήκτρα (από ελεφαντόδοντο ή άσπρο πλαστικό υλικό) και 36 μαύρα πλήκτρα (από έβενο ή μαύρο πλαστικό υλικό). Τα πλήκτρα όταν πιεστούν, σηκώνουν τα ξύλινα σφυράκια που είναι ντυμένα με μαλακό κετσέ (πνιγέας) και τα οποία στηρίζονται στη χορδή. Η χορδή έτσι μπορεί και ταλαντώνεται ελεύθερα παράγοντας ήχο, ώσπου το πλήκτρο, επιστρέφοντας στην αρχική του θέση, ξαναρίχνει τον πνιγέα.» [Γεννάτος 2005]

Το πιάνο θεωρείται σήμερα ο «βασιλιάς των οργάνων», όρος ο οποίος καθιερώθηκε από τον Φρανζ Λιστ. Οφείλεται να αναγνωριστεί η συμβολή όλων εκείνων των συνθετών και εκτελεστών που με το δημιουργικό τους έργο προέβαλαν και καταξίωσαν στη συνείδηση ειδημόνων και μη, αυτό το εξαιρετικών, πραγματικά, δυνατοτήτων όργανο.

Τέλος, από ερευνητική και επιστημονική άποψη, δικαίως θεωρείται ο «βασιλιάς των οργάνων». Η εκπαίδευση κάθε μουσικού στο πιάνο αποτελεί αναπόσπαστο μέρος της γενικότερης μουσικής του κατάρτισης. Σχεδόν σε όλα τα μουσικά ωδεία παγκοσμίως, το στάδιο της εκπαίδευσης του μουσικού στο πιάνο, θεωρείται προϋπόθεση για τη διαμόρφωση της συνολικής μουσικής του παιδείας. Πρόκειται για ένα «ώριμο» και πολύ δημοφιλές όργανο το οποίο αποτελεί βάση για κάθε είδους μουσική έρευνα. Όλα τα παραπάνω χαρακτηριστικά του οδήγησαν στην επιλογή του πιάνου ως όργανο εφαρμογής αυτής της έρευνας.

1.3 Ερμηνευτικές κινήσεις στο πιάνο

Από τη δημιουργία του κόσμου μέχρι και σήμερα, τα χέρια αποτελούσαν για τον άνθρωπο ένα ισχυρό εργαλείο έκφρασης. Πολύ λίγα άλλα στοιχεία της ανθρώπινης φύσης μπόρεσαν διαχρονικά να παίξουν ένα ρόλο τόσο καθοριστικό για τις σχέσεις και την επικοινωνία που ανέπτυξε ο άνθρωπος με το περιβάλλον.

1.3.1 Δακτυλικές κινήσεις στο πιάνο

Η ανάγκη μελέτης των κινήσεων που παράγονται κατά τη διάρκεια ερμηνείας ενός μουσικού έργου έχει τις ρίζες της στον 19^ο αιώνα. Ποικίλες μελέτες, χρησιμοποιώντας τεχνικές της εποχής, αναφέρονται στη βιβλιογραφία [Payeur 2006]. Τα κίνητρα που οδήγησαν στη διεξαγωγή αυτών των μελετών ήταν ποικιλόμορφα. Ξεκινούσαν από τη μελέτη του ήχου και κατέληγαν σε βιοκινητική αξιολόγηση της ερμηνείας του πιανίστα, περνώντας μέσα από την έννοια της εκπαίδευσης στο πιάνο αλλά και της βελτίωσης των μουσικών δεξιοτήτων. Η Marie Jaëll ήταν η πρώτη που ξεκίνησε την έρευνα στον τομέα των κινήσεων του πιανίστα. Ήδη από το 1897 δημοσιεύει μια πρωτότυπη μέθοδο καταγραφής των δακτυλικών κινήσεων. Το 1929 ο Otto Ortman προτείνει την πρώτη μοντελοποίηση των κινήσεων του πιανίστα βασιζόμενος στην ανατομία του χεριού. Λίγα χρόνια μετά, το 1949, ο Schultz επεκτείνει την έρευνα του Ortman εφαρμόζοντας νόμους της μηχανικής στο «πιανιστικό παίξιμο». Οι τρεις τους αποτέλεσαν τους πρώτους επιστήμονες που πραγματοποίησαν αξιόλογες μελέτες στον τομέα της κινητικής του σώματος του πιανίστα και ιδίως στα χέρια.

Η κίνηση των χεριών, και ειδικότερα η κίνηση των δακτύλων, για τη μουσική αποτελεί το αλφαβητάρι της ερμηνευτικής ικανότητας ενός ανθρώπου. Η σχολαστική μελέτη και παρατήρηση των κινήσεων ενός πιανίστα την ώρα που ερμηνεύει ένα μουσικό έργο, αλλά και ο τρόπος που επικοινωνεί και αλληλεπιδρά με αυτό, είναι η ασφαλέστερη μέθοδος για την κατανόηση του ιδίου του έργου αλλά και των συναισθημάτων του πιανίστα.

1.3.2 Δακτυλισμοί στο πιάνο

Σύμφωνα με την ηλεκτρονική εγκυκλοπαίδεια Wikipedia,

«Δακτυλισμός είναι η διαδικασία επιλογής του δακτύλου και της στάσης του χεριού του πιανίστα κατά τη διάρκεια μουσι-

κής ερμηνείας. Για παράδειγμα, παίζοντας πιάνο, οι συγχορδίες και οι μελωδίες της παρτιτούρας μπορούν να αναπαραχθούν με ένα πλήθος δακτυλικών συνδυασμών που πατούν τα πλήκτρα. Σ' αυτό το πλαίσιο λοιπόν, δακτυλισμός είναι η επιλογή του δακτύλου που θα χρησιμοποιηθεί για να πατηθεί ένα συγκεκριμένο πλήκτρο, για κάθε νότα της παρτιτούρας.»

Οι δακτυλισμοί προφανώς αλλάζουν από το ένα μουσικό έργο στο άλλο. Το κριτήριο επιλογής καλών δακτυλισμών σε κάθε περίπτωση, είναι η πραγματοποίηση άνετης και ευέλικτης κίνησης του χεριού, χωρίς συχνή αλλαγή της στάσης του. Ο δακτυλισμός μπορεί να είναι αποτέλεσμα μη δεσμευτικής πρότασης από τη μεριά του ίδιου του συνθέτη, πρωτοβουλίας από τη μεριά του εκδότη ο οποίος εμφανίζει στην παρτιτούρα τους δικούς του δακτυλισμούς ή και απόφαση του πιανίστα μέσα από την προσωπική του εμπειρία.

Ο δακτυλισμός σε μουσικά όργανα που διαθέτουν πληκτρολόγιο, όπως για παράδειγμα το πιάνο ή το συνθεσάιζερ, απεικονίζονται πάνω στην παρτιτούρα με έναν αριθμό από το 1 έως το 5, για το κάθε χέρι. Οι αριθμοί αυτοί αναφέρονται στα ίδια τα δάκτυλα και όχι στη στάση του χεριού πάνω στο όργανο. Είναι κατά συνέπεια προφανές ότι, για μια παρτιτούρα η οποία περιλαμβάνει νότες, υπάρχουν (5+5)^ν πιθανές διατάξεις δακτύλων και για τα δύο χέρια. Πρακτικά, ένας επαγγελματίας μουσικός έχει να επιλέξει μεταξύ ελαχίστων από τις παραπάνω διατάξεις.

Ο «ιδανικός» δακτυλισμός, κατά τον Parncutt, είναι αυτός ο οποίος προσπερνά διαφόρους περιορισμούς που έχουν να κάνουν με παραμέτρους φυσικές (*δομή του κλαβιέ του πιάνου*), ανατομίας (*δομή του χεριού*), κινητικές (*συντονισμός των δακτύλων*), γνωστικές (*απομνημόνευση σύνθετων παρτιτούρων*) και ερμηνευτικές (*συναισθηματική έκφραση*) [Parncutt 2002]. Παρόλο που ο ενδεδειγμένος δακτυλισμός είναι συχνά προφανής ή διαισθητικός για έμπειρους μουσικούς, οι αρχάριοι χρειάζονται εξωτερική βοήθεια εξαιτίας του ότι οι δακτυλισμοί δεν αναφέρονται πάντα στην παρτιτούρα [Gillardino 1975].

1.4 Η κίνηση μέσω διαφορετικών προσεγγίσεων

1.4.1 Μουσική αλληλεπίδραση

Η μουσική ερμηνεία είναι αποτέλεσμα ενός είδους «συμβίωσης» μεταξύ ανθρώπου και μουσικού οργάνου. Ο άνθρωπος με το διπλό του ρόλο ως συνθέτης ή ερμηνευτής και το μουσικό όργανο δημιουργούν

ένα δίκτυο συνεργασίας όπου πολυδιάστατα κανάλια επικοινωνίας συνεργάζονται με σκοπό την πραγματοποίηση σύνθετων μουσικών διεργασιών. Πρόκειται για ένα δίκτυο το οποίο συνδέει τον ερμηνευτή, το μουσικό κείμενο (*παρτιτούρα*) και το μουσικό όργανο.

Στη σχέση του πιανίστα με το μουσικό όργανο παρατηρείται ένα είδος ελκτικής αλληλεπίδρασης η οποία βασίζεται σε μια τριγωνική σχέση που έχει σαν κορυφές τη γνώση (*χρήση μουσικού οργάνου*), την κίνηση (*χέρια*) και την αντίληψη (*όραση και ακοή*). Σε ένα τέτοιο αλληλεπιδραστικό σύστημα, ο πιανίστας και το πιάνο δημιουργούν ένα ενιαίο σύνολο και κατ' επέκταση ένα κανάλι επικοινωνίας μέσω του οποίου ο ερμηνευτής παράγει και ταυτόχρονα μεταδίδει οπτική πληροφορία (*μέσω των κινήσεών του*) αλλά και ηχητική πληροφορία (*μέσω της μουσικής*).

1.4.2 Επικοινωνία ανθρώπου υπολογιστή

Η κλασική αντίληψη της ΕΑΥ στη μουσική μέσω υπολογιστή, ή αλλιώς ψηφιακή μουσική, αφορά κυρίως στην ικανοποίηση των αναγκών του δίδυμου μουσικός - κίνηση.

Η προσέγγιση αυτή δύναται να συνοψιστεί σε δύο στάδια. Το πρώτο στάδιο αναφέρεται στην καταγραφή και ανάλυση των κινήσεων του μουσικού χωρίς να λαμβάνονται καθόλου υπόψη προγραμματιστικά ή τεχνικά εμπόδια στην ανάπτυξη της διεπαφής. Από τη στιγμή που οι κινήσεις έχουν καταγραφεί, επέρχεται το στάδιο της υλοποίησης του συστήματος. Πρόκειται δηλαδή για μια προσέγγιση από «*πάνω προς τα κάτω*» μέσα από το πρίσμα των αναγκών του μουσικού ως χρήστη.

Σε κάθε περίπτωση, συχνά παρατηρείται στο στάδιο της υλοποίησης, οι καταγεγραμμένες ανάγκες του μουσικού να είναι τεχνικά αδύνατο να ικανοποιηθούν. Αυτό διαφαίνεται και στις περιπτώσεις μουσικών αλληλεπιδραστικών συστημάτων όπως αυτό του συστήματος MOODS' για τη διαχείριση μουσικής πληροφορίας σε ορχηστρικά περιβάλλοντα [Bellini 1999]. Στην πράξη τέτοιου είδους συστήματα βασίζονται σε διαδικασίες αλληλεπίδρασης μέσω δικτυακών υπηρεσιών. Δεν απαντούν όμως πραγματικά στις ανάγκες, υστερούν σε εργονομία και απαιτούν ιδιαίτερα ακριβό εξοπλισμό. Για αυτό το λόγο εξάλλου, παραμένουν στη φάση του πρωτότυπου χωρίς τα αποτελέσματά τους να επιβεβαιωθούν μέσα από ευρύτερη χρήση τους.

1.4.3 Μουσική μέσω υπολογιστή

Ο υπολογιστής για την ψηφιακή μουσική είχε πολλές διαφορετικές όψεις μέχρι και σήμερα. Χρησιμοποιούνταν ως μουσικό όργανο, εργαλείο παραγωγής και αναπαραγωγής ήχων, εργαλείο μουσικής σύνθεσης ή ως εργαλείο αυτόματης μουσικής αναπαραγωγής. Σε κάθε μία από τις παραπάνω περιπτώσεις ο υπολογιστής παρεμβάλλεται στη μουσική αλληλεπίδραση ως ένα εναλλακτικό εργαλείο.

Η αναγνώριση μουσικών χαρακτηριστικών (*τονικότητα, διάρκεια, τέμπο κλπ*) μέσω της προσέγγισης της ΜμΥ, μπορεί να επιτευχθεί μόνο στην περίπτωση του ψηφιακού μουσικού οργάνου (*ηλεκτρική κιθάρα, συνθεσάιζερ κλπ*) εφαρμόζοντας κατάλληλους αλγορίθμους επεξεργασίας ηλεκτρικού σήματος. Η απόρριψη των κλασικών μουσικών οργάνων από τη ΜμΥ οδηγεί στην απαξίωσή της από τους μουσικούς και αντιμετωπίζεται ως ξένο σώμα στη διαδικασία μουσικής σύνθεσης και ερμηνείας.

Η αναγνώριση της κίνησης στη μουσική αλληλεπίδραση, υπό το πρίσμα της ΜμΥ, βασίζεται στο ψηφιακό μουσικό όργανο και όχι στις ανάγκες του μουσικού. Συνεπώς, μπορεί να θεωρηθεί ως από «*κάτω προς τα πάνω*» προσέγγιση, βασιζόμενη στην παραμετροποίηση κάποιων συσκευών εισόδου του υπολογιστή.

Ο καθορισμός μιας συγκεκριμένης αισθητικής στην αλληλεπιδραστική μουσική, δεν έγκειται ενδεχομένως σε μια διαφορετική ρύθμιση της κάρτας ήχου του υπολογιστή ή βελτίωση του πρωτοκόλλου MIDI [Verner 1995]. Οφείλει να στοχεύει στη μουσική δημιουργία και τον τρόπο που αυτή ερμηνεύεται. Μέθοδοι πληροφορικής που εφαρμόζονται στη μουσική αλληλεπίδραση, θα ήταν ιδανικό να μην προσεγγίζουν τον υπολογιστή ως ένα μηχάνημα που απλώς επεξεργάζεται άψυχες σειρές δυαδικών ψηφίων, αλλά ως πολυδιάστατο αλληλεπιδραστικό συνεργάτη.

1.4.4 Μια νέα πρόταση: Όραση υπολογιστή και στοχαστική μοντελοποίηση στη μουσική αλληλεπίδραση

Η όραση υπολογιστή, ή αλλιώς και υπολογιστική όραση, έχει ως στόχο να καθορίσει τα βήματα εκείνα τα οποία θα επιτρέψουν στον ηλεκτρονικό υπολογιστή να «κατανοήσει» αυτό που «βλέπει» όταν συνδεθούν σε αυτόν μία ή περισσότερες κάμερες.

Βασικό πλεονέκτημα της όρασης υπολογιστή ως μέθοδος αναγνώρισης ανθρώπινων κινήσεων και χειρονομιών είναι το γεγονός ότι η αλληλεπίδραση μεταξύ ανθρώπου και υπολογιστή επιτυγχάνε-

ται δίχως τα δύο μέλη να έρθουν σε επαφή. Αντίθετα, δημοφιλείς συσκευές αλληλεπίδρασης όπως είναι το πληκτρολόγιο και το ποντίκι προϋποθέτουν την επαφή του ανθρώπου με τη συσκευή ώστε να καταγραφεί η κίνηση που διαγράφεται. Κάτι αντίστοιχο ισχύει και για την πλειονότητα των συσκευών εικονικής πραγματικότητας. Μια τέτοια συσκευή είναι το DataGlove. Παρόλο που μέσω του DataGlove αποκτά κανείς πρόσβαση σε πληροφορίες μεγάλης ακρίβειας, εντούτοις αποτελούν ακριβό εξοπλισμό και δεν προσφέρουν άνεση και ευελιξία ενώ απαιτούν από το μουσικό να προσαρμόσει το προσωπικό του ερμηνευτικό στυλ στα τεχνικά χαρακτηριστικά της συσκευής.

Η εντελώς φυσική συμμετοχή του μουσικού στην αναγνώριση της κίνησης αποτελεί συγκριτικό προτέρημα της όρασης υπολογιστή έναντι των υπολοίπων μεθόδων. Μια ή περισσότερες κάμερες χαμηλού κόστους και ένας ηλεκτρονικός υπολογιστής αποτελούν τον κύριο και απαραίτητο εξοπλισμό για την ανάπτυξη μεθόδων υπολογιστικής όρασης στη μουσική αλληλεπίδραση.

Η κάθε κίνηση, εφόσον προηγουμένως έχει καταγραφεί σε ακολουθία εικόνων, αναλύεται και μοντελοποιείται με τη χρήση στοχαστικών μοντέλων. Τέτοιου είδους μοντέλα είναι τα Κρυφά Μαρκοβιανά Μοντέλα (KMM). Χρησιμοποιούνται ευρύτερα στην αναγνώριση φωνής, κειμένου κλπ, εξαιτίας του γεγονότος ότι διαχειρίζονται επιτυχώς καταστάσεις οι οποίες έχουν στοχαστική διάσταση. Τα KMM υπερτερούν στην αναγνώριση κινήσεων και χειρονομιών έναντι άλλων παράλληλων μεθόδων, όπως της Δυναμικής Περιτύλιξης του Χρόνου (ΔΠΤ), χάριν στο ότι λαμβάνουν υπόψη κινήσεις οι οποίες υπάρχουν ανάμεσα σε γειτονικές χρονικά ακολουθίες εικόνων. Σε πραγματικές συνθήκες που ένας πιανίστας ερμηνεύει κάποιο έργο, μια ακολουθία κινήσεων παρουσιάζει συνήθως μεγαλύτερη συσχέτιση με κοντινότερες ακολουθίες απ' ό,τι με τις πιο απομακρυσμένες χρονικά. Δεν ενδιαφέρει η πρόβλεψη μιας μελλοντικής χειρονομίας αλλά η αναγνώριση διαδοχικών χειρονομιών που καταγράφεται κάθε χρονική στιγμή.

Η χρήση των KMM στην αναγνώριση μουσικών συμβάντων (*πάτημα νότας, παύση, αναγνώριση τονικότητας κλπ*) και ο εντοπισμός της θέσης τους στην παρτιτούρα, μέσω επεξεργασίας στο ηχητικό σήμα, έχει υλοποιηθεί με θετικά αποτελέσματα από το Ινστιτούτο Έρευνας και Συντονισμού Ακουστικής/Μουσικής (IRCAM) του Παρισιού. Η όραση υπολογιστή διαφοροποιείται στο γεγονός ότι βασίζεται στο σήμα του βίντεο και την επεξεργασία του, ενώ υπερτερεί στην ανάκτηση δακτυλισμών και αναγνώριση κινήσεων του μουσικού.

1.5 Κίνητρα, σκοπός και στόχοι

Ο βασικός σκοπός της παρούσας διατριβής είναι η ανάπτυξη μιας μεθόδου για την ανάκτηση δακτυλισμών και την αναγνώριση κινήσεων του δεξιού χεριού ενός πιανίστα, χρησιμοποιώντας όραση υπολογιστή και μαθηματική μοντελοποίηση της κίνησης.

Σημαντικό κίνητρο, αναφορικά με τη γέννηση της ιδέας για την πραγματοποίηση της έρευνας, είναι το χαμηλό κόστος μιας βιντεοκάμερας σε συνδυασμό με την ευκολία καταγραφής ενός βίντεο. Αυτοί οι δύο παράγοντες διαμορφώνουν κατάλληλο πλαίσιο για αυτοματοποιημένη επεξεργασία και ανάλυση του περιεχομένου του βίντεο. Η αναγνώριση ανθρώπων ή ανθρωπίνων συμπεριφορών μέσα στο βίντεο, καθώς και η ανάλυση κάθε μορφής αλληλεπίδρασης που καταγράφεται μέσα του, μπορούν να οδηγήσουν σε βαθύτερη κατανόηση της ανθρώπινης συμπεριφοράς. Η μουσική είναι κατεξοχήν χώρος ελεύθερης έκφρασης, συμπεριφοράς και ερμηνείας. Άρα, μια αστείρευτη πηγή παραγωγής και ανάδειξης συναισθημάτων, τα οποία πραγματώνονται μέσα από την κίνηση του ανθρωπίνου σώματος. Ταυτόχρονα, η συναισθηματική ένδεια, με την οποία μπορούν τα μαθηματικά μοντέλα να περιγράψουν γενικότερα μια κίνηση, συνεισφέρει μοναδικά στην ερμηνεία των ελασσόνων δομικών συστατικών που την απαρτίζουν. Ειδικότερα, για τις δακτυλικές κινήσεις, το βίντεο, ως μια σειρά εικόνων, μπορεί να αναδείξει την επιμέρους διάρθρωση της κίνησης. Έτσι, η επιθυμία που δημιουργήθηκε για βαθύτερη προσέγγιση και μελέτη της χρήσης όρασης H/Y και μαθηματικών μοντέλων στην επίλυση προβλημάτων αναγνώρισης δακτυλικών κινήσεων, με έμφαση στην περίπτωση του πιάνου, οδήγησε στη διεξαγωγή της παρούσας διατριβής.

Για την επίτευξη του σκοπού αυτού καθορίστηκαν στόχοι που πρέπει να ικανοποιούνται σε οποιοδήποτε μέθοδο όρασης υπολογιστή, η οποία εφαρμόζεται στη μουσική αλληλεπίδραση. Οι στόχοι προσδιορίζονται από το ότι η μέθοδος οφείλει να είναι:

- 1) **Ολοκληρωμένη:** ικανή να υπολογίζει όλες τις παραμέτρους που καθορίζουν τις δακτυλικές κινήσεις. *Η όραση υπολογιστή είναι μέθοδος απευθείας λήψης της πληροφορίας της κίνησης.*
- 2) **Οπτικοκεντρική:** να μη χρησιμοποιεί προκαταρκτική μουσικολογική ανάλυση της παρτιτούρας. *Η μέθοδος*

όρασης υπολογιστή χρησιμοποιεί μόνο την εικόνα του μουσικού ως είσοδο και δημιουργεί ως έξοδο τους δακτυλισμούς και το είδος της κίνησης.

- 3) **Μη περιοριστική:** δεν πρέπει να επιβάλλεται στο μουσικό να παίρνει ειδική στάση ή να χρησιμοποιεί με τα χέρια του ειδικό εξοπλισμό. *Η όραση υπολογιστή είναι μια μέθοδος η οποία εστιάζει αποκλειστικά στην εικόνα του μουσικού.*
- 4) **Προσβάσιμη:** να είναι χαμηλού κόστους και κατά συνέπεια προσβάσιμη στον οποιοδήποτε κατέχει ακόμα και μη αναβαθμισμένο εξοπλισμό. *Η όραση υπολογιστή είναι μια μέθοδος η οποία είναι προσβάσιμη στον οποιονδήποτε κατέχει H/Y, κέμερα και πιάνο ή συνθεσάιζερ.*

Οι παραπάνω στόχοι αποτελούν τη βάση για την ανάπτυξη μεθόδου όρασης υπολογιστή με μαθηματική μοντελοποίηση, με έμφαση στη μουσική αλληλεπίδραση. Η λεπτομερής ικανοποίηση του καθενός από τους στόχους αυτούς οδηγεί στην επίτευξη του σκοπού της παρούσας διατριβής.

1.6 Συνεισφορά

Η αναπτυχθείσα μέθοδος όρασης υπολογιστή μπορεί να χρησιμοποιηθεί για εκπαιδευτικούς σκοπούς βοηθώντας αρχάριους στην εκμάθηση του πιάνου και γενικότερα στην επιλογή κατάλληλων δακτυλισμών και την κατανόηση της λειτουργικότητας των κινήσεων του πιανίστα.

Η αναγνώριση κινήσεων του πιανίστα μπορεί να λειτουργήσει ως μέθοδος παραγωγής μελωδιών χωρίς τη χρήση μουσικού οργάνου χάρη στο προηγούμενο της μοντελοποίησης των κινήσεων. Το λογισμικό PianOrasis δύναται να αποτελέσει ένα εικονικό μουσικό όργανο, όπου οι δακτυλικές κινήσεις θα αναγνωρίζονται και θα διασυνδέονται με νότες και ήχους οποιουδήποτε φυσικού μουσικού οργάνου. Κατ' επέκταση, μπορεί να συνδράμει με τη μοντελοποίηση της εικόνας της ερμηνείας του πιανίστα, στην αυτόματη παραγωγή μουσικής και τη γέννηση παρτιτούρων.

Επιπρόσθετα, η επέκταση της παρούσας μεθόδου μπορεί να επιτρέψει τη χρήση του ηλεκτρονικού υπολογιστή ως εργαλείο μουσι-

κής συνοδείας στη ζωντανή ερμηνεία ή να συνδράμει ως αλγόριθμος αρθρωτής κίνησης των δακτύλων σε κάποια εξειδικευμένη μουσική έρευνα στα πλαίσια της ΜΑ. Ο προτεινόμενος αλγόριθμος μπορεί να έχει χρήση, με μικρές τροποποιήσεις, και σε άλλα κρουστά μουσικά όργανα όπου επίσης υφίσταται το πρόβλημα ανάκτησης δακτυλισμών και εντοπισμού θέσης στην παρτιτούρα.

Τέλος, οι έξοδοι του λογισμικού PianOrasis, όπως για παράδειγμα οι κατανομές των θέσεων των δακτύλων ή των αποστάσεων μεταξύ τους, μπορούν να χρησιμοποιηθούν για τη σύγκριση πιανιστικών στυλ και κατανομής δακτύλων στο κλαβιέ. Μέσω του PianOrasis υπάρχουν πλέον τα απαραίτητα δεδομένα για να συγκριθούν πιανιστικά στυλ σύγχρονων μεγάλων ερμηνευτών και κατ' επέκταση να προσεγγίσουμε και να ερμηνεύσουμε ακόμα περισσότερο το ταλέντο των κλασικών συνθέτων.

1.6.1 Πεδία εφαρμογής

Η μέθοδος όρασης Η/Υ στην αναγνώριση κινήσεων έχει πληθώρα πεδίων εφαρμογών, είτε η κίνηση αφορά σε ανθρώπινες χειρονομίες είτε αφορά σε οτιδήποτε παρουσιάζει εξελικτική πορεία. Παρακάτω παρουσιάζονται ενδεικτικά μερικοί πιθανοί τομείς εφαρμογής της μεθόδου.

1.6.1.1 Αξιοποίηση πολιτιστικής κληρονομιάς

Ενδιαφέρον παρουσιάζει ο τομέας της διάσωσης σπάνιων επαγγελμάτων που βασίζονται στη χειροτεχνία. Οι κινήσεις για τέτοιου είδους επαγγέλματα αποτελούν το αλφαβητάρι της δημιουργικής ικανότητας ενός ανθρώπου. Η σχολαστική μελέτη και παρατήρηση των κινήσεων ενός βιοτέχνη, την ώρα που πλάθει τα υλικά του, αλλά και ο τρόπος που επικοινωνεί και αλληλεπιδρά με αυτά, είναι η ασφαλέστερη μέθοδος για την κατανόηση του ιδίου του έργου αλλά και των διεργασιών που ακολουθούνται. Το στοίχημα έγκειται τόσο στην καταγραφή και μοντελοποίηση αυτών των διεργασιών, όσο και στην αναπαράσταση και διάδοση της τεχνογνωσίας.

1.6.1.2 Παρακολούθηση ατόμων με έλλειψη αυτονομίας

Η όραση Η/Υ στην αναγνώριση χειρονομιών μπορεί να προσφέρει υπηρεσίες τόσο σε ανθρώπους με έλλειψη αυτονομίας στο σπίτι όσο και στο κοινωνικό σύστημα υγείας. Η αναγνώριση ανθρώπου πεσμένου στο πάτωμα, αναγνώριση ασθενή εκτός κρεβατιού, αναγνώριση χειρομορφών έκτακτης ανάγκης, ανίχνευση ανθρώπινης παρουσίας

σε σπίτια όπου ζουν μόνοι ηλικιωμένοι, είναι μερικά από τα πιθανά πεδία εφαρμογής.

1.6.1.3 Κατανόηση και βελτίωση αθλητικής συμπεριφοράς και επίδοσης

Ένας από τους τομείς εφαρμογής της ανάλυσης και αναγνώρισης των ανθρώπινων χειρονομιών είναι και η κατανόηση αθλητικών χειρονομιών. Ενδιαφέρον παρουσιάζουν οι λόγοι για τους οποίους μια εξαιρετική επίδοση για έναν αθλητή συναρτάται από την πραγματοποίηση ιδιαίτερων κινήσεων (π.χ. *το άλμα εις ύψος*). Πληθώρα δεδομένων μακροχρόνιων βιομηχανικών ερευνών επιτρέπει τη σύγκρισή τους με δεδομένα σε πραγματικές συνθήκες που θα λαμβάνονται από ένα σύστημα αναγνώρισης κινήσεων.

1.6.1.4 Ανάπτυξη μη περιοριστικών λειτουργικών συστημάτων

Μια εναλλακτική πρόταση απέναντι στις «παραδοσιακές» περιφερειακές συσκευές, όπως για παράδειγμα το πληκτρολόγιο, μπορεί να αναπτυχθεί βασιζόμενη στην προτεινόμενη μέθοδο της παρούσας διατριβής. Ένας H/Y που δε θα απαρτίζεται ούτε από πληκτρολόγιο ούτε από ποντίκι είναι η πρόκληση του μέλλοντος. Υφίσταται υπαρκτά συνάφεια υψηλού βαθμού μεταξύ δακτυλικών κινήσεων ενός πιανίστα και ενός χρήστη H/Y που πληκτρολογεί. Μια κάμερα χαμηλού κόστους, ενσωματωμένη στην οθόνη του H/Y, μπορεί να αποτελέσει το μέσο σύλληψης των κινήσεων του χρήστη, τροφοδοτώντας κατάλληλα το σύστημα αναγνώρισης δακτυλικών κινήσεων. Υπάρχουν όλες οι προϋποθέσεις μια τέτοια μέθοδος να αποτελέσει πραγματικό αντίδοτο στις παθήσεις του καρπού που προκαλούνται από την πολύωρη χρήση του πληκτρολογίου ή του ποντικού.

1.7 Διάρθρωση της διατριβής

Η παρούσα διατριβή αναπτύσσεται σε οκτώ κεφάλαια που αφορούν στο θεωρητικό υπόβαθρο της έρευνας, την ανάπτυξη της μεθόδου, του συστήματος αναγνώρισης δακτυλικών κινήσεων, την αξιολόγηση μεθόδου-συστήματος και τα γενικά συμπεράσματα.

Στο 2^ο κεφάλαιο γίνεται ανασκόπηση των μεθόδων ανάλυσης και μοντελοποίησης κινήσεων. Οι μέθοδοι που παρουσιάζονται χωρίζονται σε τρεις μεγάλες ερευνητικές περιοχές που είναι ο υπολογισμός κινήσεων, η αναγνώριση κινήσεων και η ανάκτηση δακτυλίσμων.

Στο 3^ο κεφάλαιο αναπτύσσεται το θεωρητικό υπόβαθρο της όρασης H/Y. Γίνεται παράθεση των βασικών πτυχών της ανθρώπινης

όρασης. Αναπτύσσονται οι βασικότερες μέθοδοι εντοπισμού χειριού στην εικόνα, τμηματοποίησης και ανίχνευσης ανθρώπινου δέρματος. Περιγράφονται επίσης, οι βασικές λειτουργίες της μαθηματικής μορφολογίας και οι αντίστοιχες τεχνικές μείωσης θορύβου σε δυαδικές εικόνες. Γίνεται σύγκριση των μεθόδων εντοπισμού των άκρων των δακτύλων. Το κεφάλαιο ολοκληρώνεται με την παρουσίαση των μεθόδων δυναμικής αναγνώριση κινήσεων-χειρονομιών.

Στο 4^ο κεφάλαιο παρουσιάζεται μια μεθοδολογία «συρραφής» των εργαλείων όρασης H/Y που παρουσιάστηκαν στο 3^ο κεφάλαιο. Αναπτύσσεται μια γενικευμένη πρόταση ανεξάρτητα με πεδίο εφαρμογής.

Στο 5^ο κεφάλαιο αναπτύσσεται ο εξοπλισμός που χρησιμοποιήθηκε για αναπτυχθεί η παρούσα μέθοδος, καθώς και τα προγραμματιστικά εργαλεία. Στο ίδιο κεφάλαιο αναπτύσσεται το πρωτότυπο σύστημα στατικής αναγνώρισης δακτυλικών κινήσεων.

Στο 6^ο κεφάλαιο αναλύεται ο τρόπος μοντελοποίησης της κίνησης, εκπαίδευσης των KMM και αναγνώρισης των κινήσεων. Το πρωτότυπο σύστημα δυναμικής αναγνώρισης κινήσεων λειτουργεί σε συνέχεια της στατικής αναγνώρισης κινήσεων. Η διαδικασία μοντελοποίησης των συνεχών διανυσμάτων παρατήρησης μέσω MGM.

Στο 7^ο κεφάλαιο παρουσιάζεται η διεπαφή PianOrasis, η οποία συνδέει τα δύο πρωτότυπα συστήματα. Διενεργείται αξιολόγηση του συστήματος αναγνώρισης δακτυλικών κινήσεων σε πραγματικές συνθήκες, από όπου και εξάγονται χρήσιμα συμπεράσματα.

Τέλος, η διδακτορική διατριβή ολοκληρώνεται με το 8^ο κεφάλαιο, όπου διεξάγεται μια ανασκόπηση της εργασίας, παρουσιάζονται οι περιορισμοί του συστήματος καθώς και προτάσεις για περαιτέρω βελτιώσεις και μελλοντική έρευνα.

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Ο επιβλέπων του διδακτορικού

«Για να ξεκινήσεις ένα διδακτορικό, πρέπει να βρεις ένα αφεντικό. Το αφεντικό, είναι ένας κύριος πολύ πολύ σπουδαίος που μου θέτει ένα πρόβλημα και με βοηθάει να το λύσω.»



2 Ανασκόπηση μεθόδων αναγνώρισης κινήσεων στη μουσική αλληλεπίδραση

2.1 Σύνοψη κεφαλαίου

2.2 Κίνηση και ανάκτηση πληροφορίας

2.3 Υπολογισμός κινήσεων

2.3.1 Βιολογική μηχανική

2.3.2 Ενεργή όραση

2.3.3 Συμπεράσματα

2.4 Αναγνώριση κινήσεων

2.4.1 Όραση H/Y και KMM σε διάφορα πεδία εφαρμογής

2.4.2 Αισθητήρες

2.4.2.1 Δι' επαφής

2.4.2.2 Εξ αποστάσεως

2.4.3 Συμπεράσματα

2.5 Ανάκτηση δακτυλισμών

2.5.1 Ιδιαιτερότητες στην τεχνική του πιάνου και της κιθάρας

2.5.2 Θεωρία γράφων στην a priori επεξεργασία της παρτιτούρας

2.5.2.1 Μέθοδος Radicioni

2.5.2.2 Μέθοδος Kasimi

2.5.3 Τεχνολογία MIDI στην επεξεργασία σε πραγματικό χρόνο

2.5.4 Όραση H/Y στην κιθάρα στην επεξεργασία σε πραγματικό χρόνο

2.5.5 Ψηφιακή επεξεργασία ηλεκτρικού σήματος στην a posteriori επεξεργασία

2.5.6 Συμπεράσματα

2.6 Γενικά συμπεράσματα

2.1 Σύνοψη κεφαλαίου

Στο κεφάλαιο αυτό θα γίνει παρουσίαση των μεθόδων ανάλυσης και μοντελοποίησης κινήσεων. Σύμφωνα με τη βιβλιογραφία, οι μέθοδοι αυτές κατηγοριοποιούνται σε τρεις κατηγορίες: α) υπολογισμός κινήσεων (ενότητα 2.3), β) αναγνώριση κινήσεων (ενότητα 2.4) και γ) ανάκτηση δακτυλισμών (ενότητα 2.5).

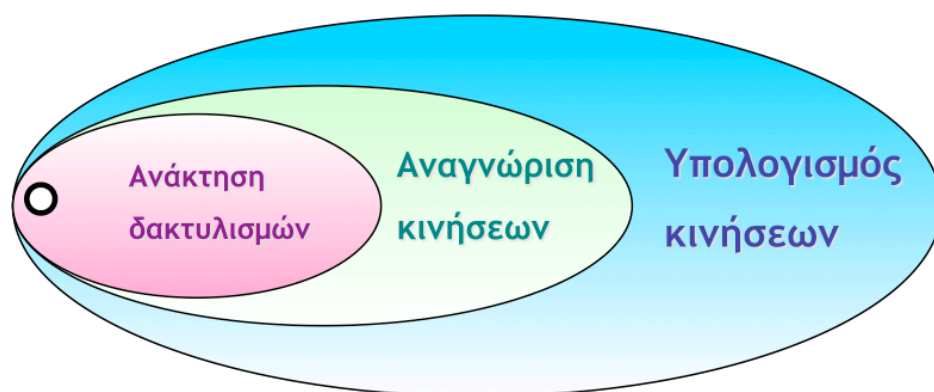
Η πρώτη κατηγορία αφορά κινήσεις που σχετίζονται με τη μετατόπιση επιμέρους μελών του ανθρώπινου σώματος. Αποτελούν μια κατηγορία που έχει ως φάσμα μελέτης σχεδόν όλο το ανθρώπινο σώμα. Η δεύτερη κατηγορία αφορά σε κινήσεις των χεριών του διευθυντή ορχήστρας. Η τρίτη κατηγορία απαρτίζεται από έρευνες που είναι εξειδικευμένες και αφορούν στα δάκτυλα του μουσικού.

Στόχος των ερευνών αυτών είναι συχνά ο καθορισμός προτεινόμενων δακτυλισμών, η ανάκτηση δακτυλισμών σε πραγματικό χρόνο, ή η ανάκτηση δακτυλισμών από προγενέστερη καταγραφή μουσικής ερμηνείας. Στο τέλος κάθε κατηγορίας ανάλυσης και μοντελοποίησης κινήσεων διατυπώνονται συμπεράσματα που αφορούν στη σύγκριση των επιμέρους μεθόδων μεταξύ τους και αναδεικνύονται οι λόγοι που δημιουργούν την ανάγκη για μια νέα εναλλακτική πρόταση.

2.2 Κίνηση και ανάκτηση πληροφορίας

Η κίνηση του σώματος είναι ο πυρήνας παραγωγής της μουσικής. Η μελέτη των κινήσεων του μουσικού και του τρόπου με τον οποίο αλληλεπιδρά με το όργανο αποτελεί την καλύτερη μέθοδο για τη βαθύτερη κατανόηση του έργου αλλά και των ανθρώπινων συναισθημάτων που εκφράζονται μέσω αυτού.

Η πολύ ενδιαφέρουσα βιβλιογραφία στον τομέα της μελέτης των κινήσεων του μουσικού και της πληροφορίας που μπορεί να ανακτηθεί από αυτήν έχει σημαντικά παραδείγματα ερευνών να αναδείξει. Γενικότερα, η βιβλιογραφική έρευνα στον τομέα της ανάκτησης πληροφορίας μέσω της κίνησης στη ΜΑ, αποτελείται από τρεις κατηγορίες: α) υπολογισμός κινήσεων, β) αναγνώριση κινήσεων και γ) ανάκτηση δακτυλισμών (Σχήμα 2.2.1). Η ευρύτερη εκ των τριών, ο υπολογισμός κινήσεων, περιλαμβάνει τις άλλες δύο, ενώ η αναγνώριση κινήσεων περιλαμβάνει την ανάκτηση δακτυλισμών.



Σχήμα 2.2.1 Επιμέρους ερευνητικά πεδία της ανάκτησης πληροφορίας στη ΜΑ

Μεγάλο πλήθος ερευνών που διεξήχθησαν, στόχευαν στην ανάλυση, ή αλλιώς μέτρηση της κίνησης, υπό την έννοια της ασκούμενης ενέργειας ή πράξης. Αυτές οι έρευνες αποτελούν την κατηγορία του υπολογισμού κινήσεων. Εντάσσονται μέσα στο πλαίσιο εκμάθησης και βελτιστοποίησης ερμηνευτικών ικανοτήτων, με σκοπό την απόκτηση αποτελεσματικής κίνησης, η οποία θα είναι συνάμα και ακριβής. Συχνά, είχαν ως στόχο την καταγραφή και επεξεργασία της πληροφορίας που σχετίζονταν με ανεπιθύμητες καταστάσεις οι οποίες θα εξελίσσονταν σε χρόνιο πρόβλημα για το μουσικό. Αφορούσαν κινήσεις οι οποίες σχετίζονταν λιγότερο με την παραγωγή του ήχου και περισσότερο με τη μετατόπιση επιμέρους μελών του ανθρώπινου σώματος. Αποτελούν μια κατηγορία λιγότερο στοχευμένη, με φάσμα μελέτης σχεδόν όλο το ανθρώπινο σώμα (Σχήμα 2.2.2). Το μουσικό όργανο στο οποίο εφαρμόστηκαν κατεξοχήν οι αναπτυχθείσες μέθοδοι υπολογισμού κινήσεων, ήταν το πιάνο.



Σχήμα 2.2.2 Κατηγορίες ερευνών και περιοχές του σώματος στη ΜΑ

Η δεύτερη κατηγορία είναι αυτή της αναγνώρισης κινήσεων. Η συντριπτική πλειοψηφία των καταγεγραμμένων ερευνών αφορούν την αναγνώριση κινήσεων των χεριών του διευθυντή ορχήστρας (μαέστρος) (Σχήμα 2.2.2). Οι μέθοδοι αναγνώρισης κίνησης

ταυτοποιούν την τρέχουσα κίνηση του μαέστρου μέσα από ένα σύνολο κινήσεων οι οποίες δύνανται να αναγνωριστούν από τις μεθόδους. Οι έρευνες που αφορούν στην αναγνώριση κινήσεων στο πιάνο βρίσκονται σε εντελώς αρχαίο επίπεδο και δεν έχουν να επιδείξουν κάποια μέθοδο με σαφήνεια. Αποτελεί συνεπώς για το πιάνο, ένα εντελώς νέο πεδίο έρευνας το οποίο με τη ανάπτυξη μεθόδων όρασης H/Y και μαθηματικής μοντελοποίησης της κίνησης μπορεί να αποφέρει σημαντικά συμπεράσματα στο χώρο.

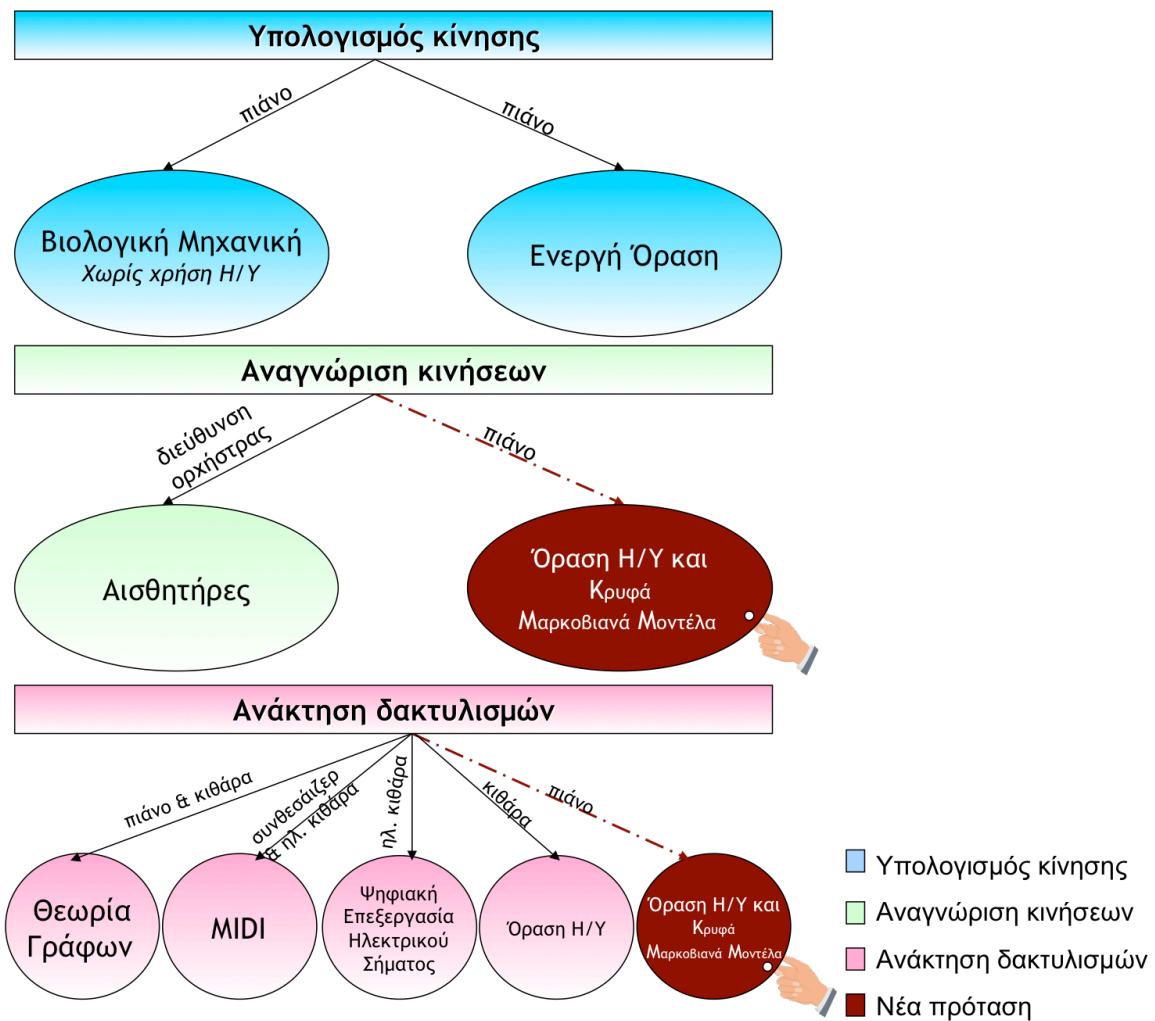
Η τρίτη κατηγορία, η ανάκτηση δακτυλισμών, είναι η λιγότερο γενική αφού εξειδικεύεται στα δάκτυλα του μουσικού. Σε αυτήν εντάσσονται έρευνες οι οποίες αφορούν κυρίως το πιάνο και την κιθάρα. Οι έρευνες αυτές μπορούν να κατηγοριοποιηθούν επιπλέον σε τρεις ενότητες: α) *A priori επεξεργασία της παρτιτούρας*, β) *επεξεργασία, κινήσεων ή συμβάντων σε πραγματικό χρόνο* γ) *A posteriori επεξεργασία του ηλεκτρικού σήματος*. Στόχος αυτών των ερευνών είναι είτε η πρόταση δακτυλισμών (*αντιστοιχία δακτύλου προς νότα*) για μελλοντική χρήση σε κάποιο μουσικό έργο, είτε ο άμεσος εντοπισμός των δακτυλισμών που χρησιμοποιούνται σε πραγματικό χρόνο, είτε τέλος η αναζήτηση και εύρεση των δακτυλισμών που χρησιμοποιήθηκαν σε παρελθόντα χρόνο (Σχήμα 2.2.2).

Μια πιο λεπτομερή προσέγγιση στις θεωρίες πληροφορικής, αλλά και στις τεχνολογίες που αναπτύχθηκαν για τη ΜΑ ή απλώς εφαρμόστηκαν σε αυτήν, μας επιτρέπει να συμπεράνουμε ότι στην ανάκτηση δακτυλισμών έχει πραγματοποιηθεί ο μεγαλύτερος αριθμός ερευνών με ποικιλόμορφη προσέγγιση του αντικειμένου (Σχήμα 2.2.3). Οι έρευνες αυτές εφαρμόστηκαν σε κλασικά ή ηλεκτρονικά όργανα σε πραγματικό χρόνο ή ετεροχρονισμένα.

Στον υπολογισμό κινήσεων, το πλήθος των μελετών είναι πιο περιορισμένο σε σχέση με την ανάκτηση, αλλά παρατηρείται ότι το πιάνο αποτέλεσε βασικό όργανο εφαρμογής (Σχήμα 2.2.3). Από τις πρώτες μεθόδους που αναπτύχθηκαν και εφαρμόστηκαν στη ΜΑ για τη μελέτη της κίνησης ήταν αυτή της βιολογικής μηχανικής. Ο H/Y δεν είχε καν εφευρεθεί τη χρονική περίοδο διεξαγωγής αυτών των ερευνών. Αντίθετα, πιο πρόσφατες μελέτες, προερχόμενες από τον ιατρικό χώρο, αποτέλεσαν την πηγή δημιουργίας για τη μεθοδολογία της ενεργούς όρασης, η οποία μετέπειτα εφαρμόστηκε και στο χώρο της ΜΑ.

Η κατηγορία της αναγνώρισης κινήσεων αποτέλεσε το πεδίο έρευνας με τη μικρότερη εφαρμογή σε μουσικά όργανα, ενώ η συμμετοχή της τεχνολογίας των αισθητήρων ως μέθοδος ήταν αποκλειστική. Το γεγονός αυτό οφείλεται εν πολλοίς στο ότι τεχνολογίες όπως

το ερευνητικό πεδίο της όρασης H/Y είναι νέο σε σχέση με την τεχνολογία των αισθητήρων. Στην περίπτωση αναγνώρισης κινήσεων που δεν έρχονται σε επαφή με μουσικό όργανο (*μαέστρος*), οι αισθητήρες δίνουν αποτελέσματα με ακρίβεια. Στην αναγνώριση κινήσεων όμως κατά την ερμηνεία σε μουσικό όργανο, η χρήση τους είναι απαγορευτική.



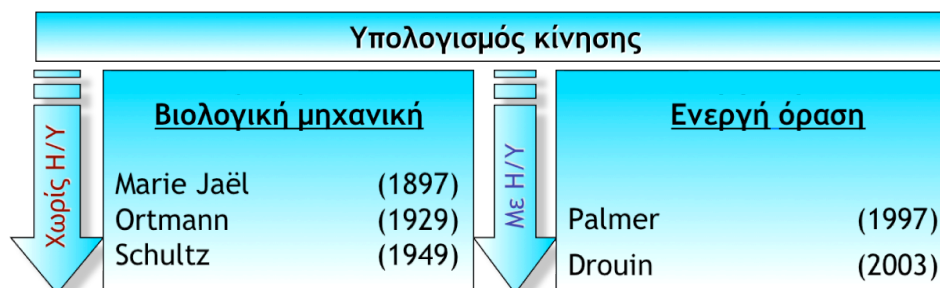
Σχήμα 2.2.3 Κατηγορίες ερευνών, θεωρίες & τεχνολογίες, στη ΜΑ

Τέλος, ένα από τα συμπεράσματα που μπορεί να εξάγει κανείς είναι ότι υπάρχει πρόσφορο πεδίο έρευνας με σκοπό την ανάπτυξη μεθόδου όρασης H/Y και KMM για την αναγνώριση κινήσεων και ανάκτησης δακτυλισμών με έμφαση στο πιάνο.

2.3 Υπολογισμός κινήσεων

Ο τομέας υπολογισμού κινήσεων αποτελεί το παλαιότερο πεδίο έρευνας αναφορικά με τη ΜΑ. Παραδείγματα ερευνών έχουν τις ρίζες τους στις αρχές του 19^{ου} αιώνα και φτάνουν μέχρι σήμερα, με τη χρήση του Η/Υ (Σχήμα 2.3.1).

Δύο είναι επιστημονικοί κλάδοι όπου βασίστηκαν οι μέθοδοι που αναπτύχθηκαν για τον υπολογισμό κινήσεων στη ΜΑ: α) η *βιολογική μηχανική* και β) η *ενεργή όραση*.



Σχήμα 2.3.1 Εξέλιξη βιβλιογραφίας υπολογισμού κινήσεων στη ΜΑ

Μόνο οι μέθοδοι που αναπτύχθηκαν με βάση την ενεργή όραση έκαναν χρήση του Η/Υ. Αντίθετα, οι μέθοδοι της βιολογικής μηχανικής χρησιμοποιούσαν μέσα και τεχνικές της εποχής τους. Είναι αυταπόδεικτο το γεγονός ότι η ανάγκη για μελέτη των αλληλεπιδραστικών κινήσεων του μουσικού με το όργανο είναι διαχρονική.

2.3.1 Βιολογική μηχανική

Η Marie Jaëll το 1897 διεξήγαγε την πρώτη αξιολογη μελέτη των επιδράσεων της επαφής των δακτύλων με τα πλήκτρα του πιάνου, στον παραγόμενο ήχο.

Βασικό πόρισμα της έρευνας της Jaëll, υπήρξε το ότι η επίπτωση που έχει ο τρόπος πατήματος του πλήκτρου στη μουσικότητα, κατά τη δια-

δικασία εκμάθησης του πιάνου, είναι πιο εποικοδομητική στην περίπτωση καλά μελετημένων και «ζυγισμένων» κινήσεων από ότι σε μια ακολουθία πανομοιότυπων κινήσεων. Ως αποτέλεσμα της εν λόγω μελέτης επήλθε η ανάπτυξη μεθόδου για την ανάλυση και κατανόηση δακτυλικών κινήσεων μέσω των αποτυπωμάτων τους [Jaëll 1897]. Μικρά κομματάκια από χαρτόνι τοποθετήθηκαν πάνω στα πλήκτρα του πιάνου. Τα δάκτυλα του πιανίστα μελανώθηκαν πριν τη διεξαγωγή της ερμηνείας ώστε να αποτυπωθούν περισσότερο ή λιγότερο έντονα ίχνη στα πλήκτρα, ανάλογα με την πίεση που ασκήθηκε επάνω τους. Χειροκίνητη ανάλυση των καμπυλών και του πλάτους των αποτυπωμάτων επέτρεψε στην Jaëll να εκτιμήσει τη γωνία και την ένταση με την οποία πατήθηκε το πλήκτρο. Μέσα από τα δύο μεγέθη της γωνίας και της έντασης του πατήματος οδηγήθηκε σε μια αξιολόγηση του παραγόμενου ήχου σε συνάρτηση με τη θέση των δακτύλων και των καρπών του πιανίστα.

Ο Otto Ortmann σύνεδραμε με την έρευνά του το 1929 στη μοντελοποίηση της ανατομίας του χεριού, του καρπού και των δακτύλων του πιανίστα. Προϊόν της έρευνάς του ήταν η διεξαγωγή ποσοτικής ανάλυσης των δυνάμεων που ασκούνται στο κλαβιέ με σκοπό την παραγωγή ήχων [Ortmann 1929]. Οι κλασικοί νόμοι της μηχανικής εφαρμόστηκαν αποβλέποντας στην κατανόηση των κινήσεων του πιανίστα κατά τη διάρκεια της ερμηνείας. Έτσι ο Ortmann οδηγήθηκε στη δημιουργία μηχανικών πρωτοτύπων του ανθρώπινου χεριού και καρπού, καθώς και εργαλείων για τη μέτρηση των μετατοπίσεων των δακτύλων και του «χαμηλώματος» των πλήκτρων. Η καταγραφή των παράλληλων προς το πιάνο κινήσεων επιτεύχθηκε με τη βοήθεια ενός μηχανισμού παντογράφου ο οποίος επισυνάφθηκε με λάστιχο στον καρπό του πιανίστα. Ο παντογράφος αναπαρήγαγε σε κλίμακα τις κινήσεις αυτές πάνω σε χαρτί τοποθετημένο παραπλεύρως του πιάνου ώστε να ελαχιστοποιηθεί η παρεμπόδιση της ερμηνείας του πιανίστα. Η μέθοδος αυτή, όπως είναι φυσικό, απαιτούσε συνεχή παρακολούθηση και ρύθμιση των βοηθητικών συσκευών και παρουσίαζε αδυναμίες όταν ο καρπός του πιανίστα κινούνταν μαζί με το πάτημα των πλήκτρων.

Η ανάπτυξη τέτοιων μεθόδων επέτρεψε να κατανοηθεί καλύτερα ο τρόπος με τον οποίο μπορεί να επιτευχθούν σύνθετες δακτυλικές κινήσεις, αποφεύγοντας καταστάσεις οι οποίες θα προκαλούσαν μεγάλη μυϊκή κούραση. Μερικές δεκαετίες αργότερα, ο Schultz κινούμενος στην ίδια κατεύθυνση, κατέθεσε τη δική του επιστημονική προσέγγιση για το θέμα [Schultz 1949]. Είναι απαραίτητο να σημειωθεί ότι ήδη από την εποχή των Jaëll, Ortmann και Schultz υπήρξαν αντιδράσεις από τη μεριά των μουσικών για τη λειτουργικό-

τητα μεθόδων τεχνικής φύσεως στην κατανόηση φυσικών, κινητικών και ψυχολογικών διαδικασιών στη μουσική αλληλεπίδραση.

2.3.2 Ενεργή όραση

Οι περισσότερες μέθοδοι αναγνώρισης ανθρώπινων κινήσεων οι οποίες αναπτύχθηκαν για να εφαρμοστούν στον τομέα της ιατρικής, του αθλητισμού, της μουσικής κ.ά., κρίνονται αρκετά περιοριστικές αφού απαιτούν εγκατάσταση αισθητήρων με καλωδιώσεις πάνω στον άνθρωπο.

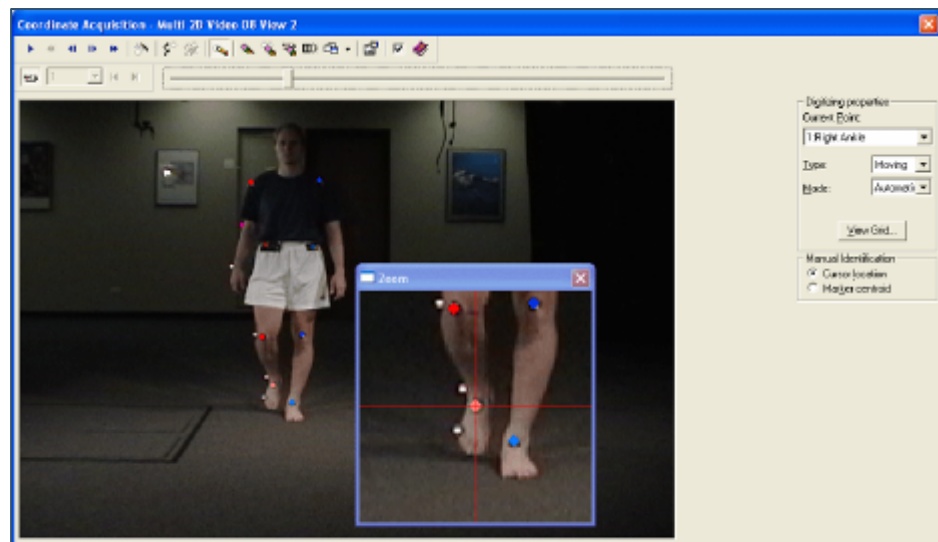
Οι μέθοδοι αναγνώρισης κινήσεων που βασίζονται σε ειδικό εξοπλισμό είναι συνυφασμένες με φυσικούς περιορισμούς στην κίνηση, θεωρούνται ελάχιστα άνετες, απαιτούν μεγάλο χρονικό διάστημα προετοιμασίας προκειμένου να λειτουργήσουν και να αναγνωρίσουν την ανθρώπινη κίνηση και κοστίζουν ακριβά.

Μια μέθοδος η οποία προσπερνάει το πρόβλημα των καλωδιώσεων στο ανθρώπινο σώμα είναι το σύστημα VICON, γνωστό από τον ιατρικό χώρο [Vicon Peak 2005; Jobbagy 2004]. Αποτελεί ένα εξαιρετικό παράδειγμα στον τομέα της ορθοπεδικής έρευνας και ιδιαίτερα χρήσιμο στην αποκατάσταση βαδίσματος. Χρησιμοποιεί την προσέγγιση της *ενεργής όρασης* η οποία έγκειται στην εγκατάσταση μικρών σφαιριδίων αντανάκλασης, σε επιλεγμένα σημεία του σώματος, και στο φιλμάρισμα των μετακινήσεών τους από κάμερες. Οι φακοί των καμερών περιβάλλονται από διόδους εκπομπής φωτός το σήμα των οποίων αντανάκλαται πάνω στα σφαιρίδια (Σχήμα 2.3.2.1). Αυτή η μέθοδος περιορίζει αφενός την αναγκαιότητα απευθείας καλωδίωσης επάνω στον ασθενή και αφετέρου όμως την παρατήρηση των σημείων του σώματος πάνω στα οποία είναι τοποθετημένα τα σφαιρίδια, ενώ απαιτεί χρόνο στην εγκατάσταση ειδικού εξοπλισμού, του οποίου το κόστος δεν είναι αμελητέο.

Ο Palmer (1997) προσπάθησε να μεταφέρει την τεχνική της ενεργής όρασης στο παίξιμο του πιάνου [Palmer 1997; Palmer 2000]. Μια παρόμοια προσπάθεια στον ίδιο επιστημονικό χώρο είναι αυτή του Drouin (2003), ο οποίος υπολόγισε τις κινήσεις ενός πιανίστα που φορούσε ένα μαύρο ένδυμα με πορτοκαλί σφαιρίδια. Η κάμερα κατέγραφε τη μετατόπιση των σφαιριδίων, των οποίων το χρώμα έρχονταν σε μεγάλη αντίθεση με το υπόλοιπο πλάνο [Drouin 2003]. Ως μέθοδος είχε μεν μικρότερο κόστος από αυτήν του Palmer, απαιτούσε δε ειδικό εξοπλισμό όπως εκείνη.

Η χρήση σημειακών δεικτών αντανάκλασης φωτός, όπως είναι τα σφαιρίδια, είναι αρκετά δημοφιλής. Εντάσσεται στη στρατηγική αναζήτησης της θέσης των κινούμενων μελών μέσα από την α-

ντιπαραβολή των προβαλλόμενων σημειακών δεικτών στην εικόνα, με τα σημεία τα οποία καθορίζουν την κίνηση ενός εικονικού σκελετού αν εγκατασταθούν στην ίδια θέση με τους σημειακούς δείκτες [Albrecht 2003]. Ανθρώπινα σκελετικά μοντέλα καθορίστηκαν με στόχο να προσομοιάσουν όλες τις πιθανές στάσεις του ατόμου και να εντοπίσουν αυτήν που αντιστοιχεί καλύτερα στην τρέχουσα κατάσταση που παρατηρείται μέσω των σημειακών δεικτών (Σχήμα 2.3.2.2). Δυστυχώς, το γεγονός ότι οι δείκτες βρίσκονται τοποθετημένοι στα ρούχα του ατόμου, ή σε καλύτερη περίπτωση στο δέρμα του, παράγει συχνά μη ακριβείς μετακινήσεις που δεν αντιστοιχούν απόλυτα στις κινήσεις του μοντέλου. Παρατηρείται συνεπώς απόκλιση μεταξύ της εκτιμώμενης και της πραγματικής κίνησης [Boulic 2000].



Σχήμα 2.3.2.1 Διεπαφή Vicon Peak

2.3.3 Συμπεράσματα

Ο υπολογισμός κινήσεων είναι μια ερευνητική προσέγγιση μακροσκοπικού χαρακτήρα, η οποία έχει να προσφέρει πολλά χρήσιμα συμπεράσματα στην κατανόηση του τρόπου κίνησης των μελών του ανθρώπινου σώματος. Ειδικότερα, επιστήμες όπως η ιατρική μπορούν να αποκομίσουν ακόμα περισσότερα οφέλη στο μέλλον, σε ό,τι αφορά την αποκατάσταση της λειτουργίας μελών του σώματος, π.χ. μετά από ατύχημα. Τα πόδια και το βάδισμα, ή το χέρι και οι λειτουργίες

του, αποτελούν σημαντικά πεδία εφαρμογής μελετών υπολογισμού κινήσεων.



Σχήμα 2.3.2.2 Τρισδιάστατο σκελετικό μοντέλο (γραμμές) και σημειακοί δείκτες (τελείες)

Αντίθετα, στον τομέα της ΜΑ, απαιτείται πρωτίστως έρευνα μικροσκοπικού χαρακτήρα πάνω σε συγκεκριμένα σημεία του σώματος, όπως είναι τα δάκτυλα και τα άκρα τους. Η περαιτέρω κατανόηση του τρόπου κίνησης των χεριών είναι χρήσιμη μεν, αλλά αποδεδειγμένης σημασίας δε. Ο υπολογισμός της κίνησης δύναται να μετρήσει και να καταγράψει τις παραμέτρους της κίνησης του χεριού συνολικά. Ενδεχομένως και να αποδείξει με διαφορετική μέθοδο τη σημασία των καλά «σκεπτόμενων και ζυγισμένων» κινήσεων στη ΜΑ. Παρουσιάζει αδυναμίες όμως στη δημιουργία καλών πρακτικών και τον καθορισμό κριτηρίων που μπορούν συνδράμουν ενεργά και χρηστικά στη ΜΑ. Αδυνατεί να προτείνει λύσεις αναφορικά με τη χρήση δακτυλισμών ή να αναγνωρίσει τις κινήσεις ενός βιρτουόζου του πιάνου.

Τέλος, είναι γεγονός ότι οι μέθοδοι υπολογισμού κινήσεων στη ΜΑ, είτε κάνουν χρήση καλωδιώσεων πάνω στο σώμα είτε χρησιμοποιούν σφαιρίδια για την καταγραφή της μετατόπισης, είναι δύσχρηστες. Οι μέθοδοι αυτές δεν είναι εύκολα προσβάσιμες από το μουσικό επειδή: α) απαιτούν ειδικό εξοπλισμό, β) απαιτούν σημαντική προετοιμασία εγκατάστασης, ενώ, γ) έχουν σημαντικό κόστος. Για όλους τους παραπάνω λόγους, οι μέθοδοι που ήδη αναπτύχθηκαν στον τομέα του υπολογισμού κινήσεων παρέμειναν στα πλαίσια του πρωτότυπου.

2.4 Αναγνώριση κινήσεων

Μέσα από τη βιβλιογραφία στοιχειοθετείται επαρκώς η άποψη ότι η όραση H/Y συνδράμει, με θετικότερα αποτελέσματα, στην καταγραφή εικόνας και τον εντοπισμό της περιοχής ενδιαφέροντος (ΠΕ) σε αυτήν. Κατ' αντιστοιχία, τα KMM, ως μέθοδος διαχείρισης στοχαστικών διαδικασιών και δεδομένου ότι η κίνηση έχει στοχαστικό χαρακτήρα, παράγουν αξιολογότερα συμπεράσματα σε συνδυασμό με την όραση H/Y στην αναγνώριση κινήσεων.

Όπως περιγράφηκε σε προηγούμενη ενότητα, η διεύθυνση ορχήστρας αποτελεί τον κατεξοχήν τομέα εφαρμογής μεθόδων αναγνώρισης κινήσεων. Η αλληλεπίδραση πραγματώνεται μεταξύ μαέστρου και μουσικών. Η κίνηση η οποία γεννιέται από το μαέστρο δεν παράγει άμεσα μουσική αλλά γεννάει οπτική πληροφορία την οποία επεξεργάζονται ανάλογα οι μουσικοί μετατρέποντάς την σε μουσική. Η τεχνολογία των αισθητήρων σ' αυτήν την κατηγορία ερευνών έχει το μεγαλύτερο ιστορικό.

Στις αμέσως παρακάτω ενότητες, γίνεται ανασκόπηση των σημαντικότερων μεθόδων όρασης H/Y και KMM στην αναγνώριση κινήσεων σε διάφορα πεδία εφαρμογής εκτός ΜΑ. Γίνεται επίσης ανασκόπηση των μεθόδων που βασίστηκαν, ως επί το πλείστον, στην τεχνολογία των αισθητήρων για να αναγνωρίσουν κινήσεις μαέστρου στη διεύθυνση ορχήστρας. Τέλος, παρουσιάζονται οι σημαντικότεροι λόγοι για τους οποίους οι αισθητήρες δεν μπορούν να επιφέρουν ικανοποιητικά αποτελέσματα στην αναγνώριση κινήσεων μεταξύ μουσικού και οργάνου.

2.4.1 Όραση H/Y και KMM σε διάφορα πεδία εφαρμογής

Αν και η όραση H/Y και τα KMM παρουσιάζουν σημαντικά αποτελέσματα στην αναγνώριση κινήσεων γενικά, εντούτοις η βιβλιογραφία δεν έχει να αναδείξει μεγάλο αριθμό ερευνών στον τομέα της ΜΑ. Ως εκ τούτου, θα παρουσιαστούν με συνοπτικό τρόπο α) μια μέθοδος αναγνώρισης κινήσεων με όραση H/Y [Coggan 2005], με σκοπό την αντικατάσταση του πληκτρολογίου του H/Y και β) μια μέθοδος αναγνώρισης κινήσεων στο χορό η οποία χρησιμοποιεί KMM [Boukir 2004]. Η μεν πρώτη μέθοδος παρουσιάζει κοινά σημεία με την αναγνώριση κινήσεων στο πιάνο γιατί πρόκειται και στις δύο περιπτώσεις για κινήσεις σε πλήκτρα, ενώ η δεύτερη συσχετίζεται με το πιάνο υπό την έννοια ότι χορευτικές και ερμηνευτικές κινήσεις συνδυάζουν συναισθημα και παραγωγική κίνηση.

Η Melanie Coggan και ο Adrian Ghizaru το 2005, στα πλαίσια ερευνητικού προγράμματος στη όραση H/Y στο πανεπιστήμιο McGill του Καναδά, παρουσίασε μια μέθοδο εξαγωγής και ιχνηλάτησης του ανθρώπινου χεριού. Η μέθοδος αυτή βασίστηκε καταρχήν στην απομόνωση του χεριού από την εικόνα με τη χρήση στατιστικού χρωματικού μοντέλου δέρματος. Στη συνέχεια εφαρμόστηκαν τεχνικές για τον εντοπισμό θέσης των δακτύλων, ενώ τέλος πραγματοποιήθηκε απλοποιημένη αναγνώριση κινήσεων.

Οι Coggan και Ghizaru (2005) έκαναν χρήση τεχνικών βασιζόμενων στην εικόνα. Σε τεχνικές τέτοιου είδους είναι απαραίτητο να καθοριστεί η περιοχή του χεριού στην εικόνα. Αυτό μπορεί να επιτευχθεί με αρκετούς τρόπους. Ένας από αυτούς υλοποιείται κάνοντας χρήση του χαρακτηριστικού χρώματος που έχει το ανθρώπινο δέρμα και το οποίο δεν το συναντούμε εύκολα σε άλλα υλικά. Απαιτεί δημιουργία βιβλιοθήκης φωτογραφιών οι οποίες θα περιέχουν δερματικές περιοχές και μέσω των οποίων θα εντοπίζονται τέτοιες περιοχές σε οποιαδήποτε άλλη εικόνα. Φυσικά υπάρχει ο περιορισμός ότι ως είσοδος στη μέθοδο μπορεί να δοθεί μόνο έγχρωμη εικόνα. Δεδομένης της ευκολίας που μπορούμε να αποκτήσουμε τέτοιες εικόνες από μια ψηφιακή κάμερα, ο περιορισμός δεν είναι και τόσο σημαντικός.

Ο αλγόριθμος αυτός [Coggan 2005] αποτελούνταν από διαδικασίες τμηματοποίησης της εικόνας βασιζόμενη σε τιμές του χώρου RGB, εντοπισμού δακτύλων που εφαρμόζονταν απευθείας στην εικόνα και ανάλυσης χειρονομιών που εξήγαγε την αναγνώριση κάποιων χειρονομιών με βάση το αριθμό των δακτύλων που παρουσιάζονταν σε κάθε εικόνα.

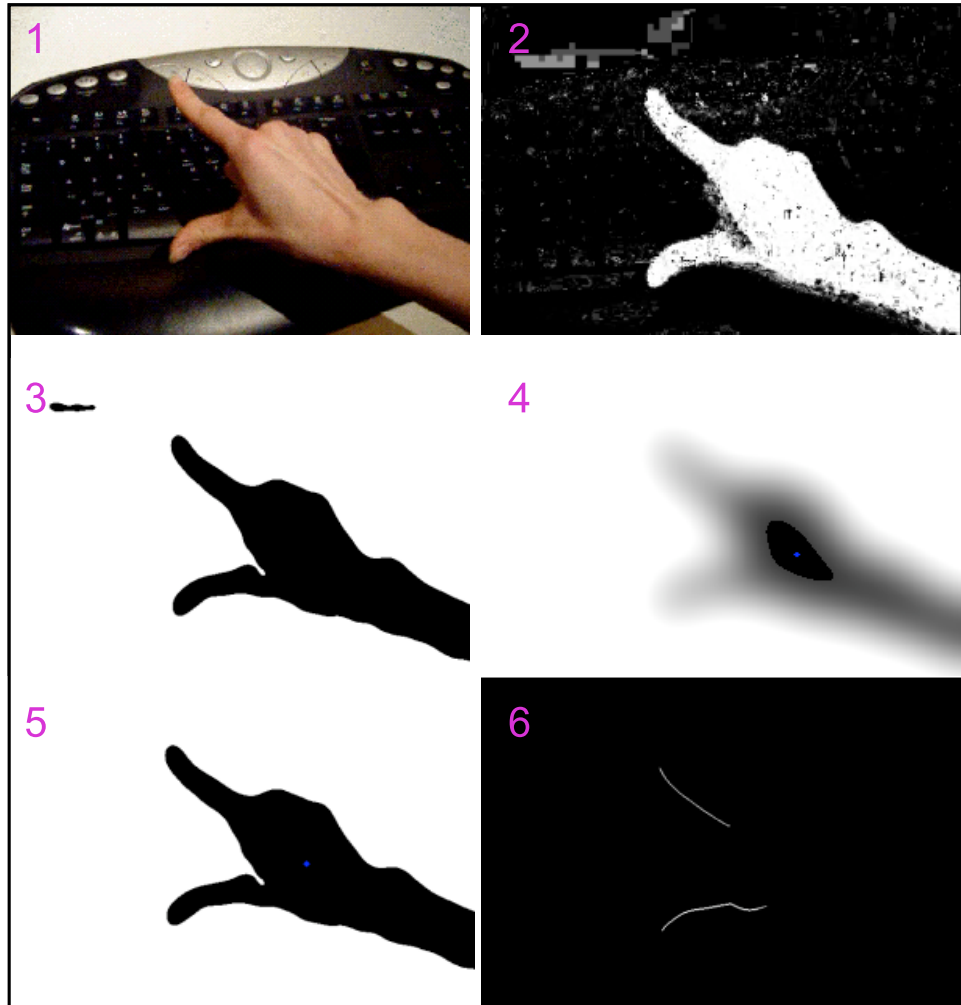
Αποσπάσματα από τα βήματα του αλγορίθμου που ακολουθήθηκε περιγράφονται στο σχήμα 2.4.1.1. Οι Coggan και Ghizaru (2005), με βάση τη βιβλιοθήκη φωτογραφιών που είχαν στη διάθεσή τους, δημιούργησαν κατάλληλο δερματικό μοντέλο, το οποίο χρησιμοποίησαν για να υπολογίσουν την πιθανότητα κάθε εικονοστοιχείο της εικόνας εισαγωγής (Σχήμα 2.4.1.1: 1), να είναι εικονοστοιχείο δέρματος. Λαμβάνοντας υπόψη αυτές τις πιθανότητες, τμηματοποίησαν την εικόνα απομονώνοντας δερματικές περιοχές. Το γεγονός όμως ότι το χρησιμοποιούμενο δερματικό ήταν ατελές οδήγησε στο γεγονός εμφάνισης αντικειμένων του παρασκήνιου ως δερματικές περιοχές (Σχήμα 2.4.1.1: 2). Εφαρμόστηκε επαναληπτικά φίλτρο Gauss μεγάλου βαθμού και κατωφλίωση για καθαρισμό της εικόνας και επίτευξη αυξανόμενου περιμετρικά θαμπώματος του περιγράμματος με σκοπό να οδηγήσουν στον εντοπισμό του κέντρου μάζας

του χεριού (Σχήμα 2.4.1.1: 3, 4 και 5). Με υψηλή κατωφλίωση των καμπυλών του χεριού πέτυχαν την ανάκτηση των σκελετών των δακτύλων (Σχήμα 2.4.1.1: 6) από όπου ήταν εύκολο να υπολογίσουν το πλήθος των εμφανιζόμενων δακτύλων στην εικόνα και άρα να αναγνωρίσουν την κίνηση.

Η αξιολόγηση της εν λόγω έρευνας [Coggan 2005] έδειξε ότι μόνο ένα μικρό τμήμα από το σύνολο των φυσικών ανθρώπινων χειρονομιών, που οφείλει να αναγνωρίζει μια διεπαφή ανθρώπου και Η/Υ, υλοποιήθηκε. Ο βασικότερος περιορισμός της μεθόδου που αναπτύχθηκε ήταν επιρροή των αποτελεσμάτων από τις μεταβολές στις συνθήκες φωτισμού. Υιοθετήθηκε από τους ερευνητές η θέση ότι ο φωτισμός του χώρου οφείλει να είναι τέτοιος ώστε να προσδίδει στο χέρι τη λιγότερη δυνατή σκίαση. Η μέθοδος αυτή παρουσιάζει θετικότερα αποτελέσματα σε ξεκάθαρα καθορισμένες κινήσεις του χεριού.

Η αναγνώριση κινήσεων που εφάρμοσαν οι Coggan και Ghizaru (2005), βασίζεται εξολοκλήρου στις δυνατότητες που προσφέρει η όραση Η/Υ εφαρμόζοντας μια μέθοδο που βασίζεται αποκλειστικά στην εικόνα. Αν και τα αποτελέσματά της είναι θετικά, αφού προσπερνά με επιτυχημένο τρόπο όλες τις απαιτούμενες καλωδιώσεις και το προβλεπόμενο κόστος των αισθητήρων, εντούτοις δε λαμβάνει καθόλου υπόψη τη φύση μιας ανθρώπινης χειρονομίας. Η φύση της χειρονομίας υποδηλώνει ότι μέσα στα πλαίσια της συχνά τυχαίας εναλλαγής χειρονομιών, υπάρχει μια στοχαστικότητα η οποία χαρακτηρίζει ακολουθίες χειρονομιών. Η μέθοδος η οποία διαχειρίζεται καλά τέτοιες καταστάσεις είναι τα KMM.

Οι Samia Boukir και Frédéric Chenevière (2004) ανέπτυξαν ένα ολοκληρωμένο πρωτότυπο για την αναγνώριση χορευτικών κινήσεων βασισμένο σε δισδιάστατες και τρισδιάστατες χωροχρονικές καμπύλες προερχόμενες από σύστημα αισθητήρων για την ανίχνευση της κίνησης. Η μέθοδος αναγνώρισης βασίζεται στα KMM και είναι μια επέκταση μεθόδων ενεργούς όρασης που περιγράφηκαν σε προηγούμενη ενότητα. Αποτελείται από ένα σύνολο KMM, κάθε ένα από τα οποία συνδέεται με την προβολή στο επίπεδο της θέσης ενός σημειακού δείκτη τοποθετημένου στο ένδυμα του χορευτή [Boukir 2004]. Η έρευνά τους βασίστηκε στην αναγνώριση 12 βασικών κινήσεων από 4 χορευτές μπαλέτου.



Σχήμα 2.4.1.1 Μέθοδος αναγνώρισης κινήσεων σε πληκτρολόγιο

1: εικόνα εισαγωγής, 2: τμηματοποίηση εικόνας βασισμένη στο χρώμα του δέρματος, 3: καθαρισμός μάσκας εικόνας, 4: ανάκτηση κέντρου μάζας από θάμπωμα με φίλτρο Gauss, 5: μάσκα χεριού, 6: ανάκτηση σκελετών χεριού από την προηγούμενη εικόνα

Η έρευνα των Boukir και Chenevière (2004) συνέκρινε τα πειραματικά αποτελέσματα των τεχνικών της Δυναμικής Περιτύλιξης του Χρόνου (ΔΠΤ- *Dynamic Time Warping*), των Νευρωνικών Δικτύων Χρονικής Καθυστέρησης (ΝΔΧΚ - *Time Delay Neural Networks*) και των Κρυφών Μαρκοβιανών Μοντέλων (KMM – *Hidden Markov Models*), στον τομέα της αναγνώρισης κινήσεων στο χορό. Απέδειξε πειραματικά ότι τα KMM προσφέρουν καλύτερα αποτελέσματα σχετικά με τα χαρακτηριστικά της ευελιξίας και της ταχύτητας της κίνησης

[Berndt 1994; Rabiner 1989]. Ερμηνεύονται πιο ξεκάθαρα και ευκολότερα από τα ΝΔΧΚ, τα οποία δυσκολεύονται να χειριστούν την παράμετρο του χρόνου και για αυτό τη μετατρέπουν σε χωρική παράμετρο [Weismann 1999]. Τα ΚΜΜ διαχειρίζονται καλύτερα από τα ΔΠΤ αλληλουχίες κινήσεων οι οποίες γειτνιάζουν χρονικά.

Οι έρευνες των Coggan και Ghizaru (2005) και των Boukir και Chenevière (2004) αποτελούν αντιπροσωπευτικά παραδείγματα εφαρμογής μεθόδων όρασης H/Y και ΚΜΜ αντίστοιχα, σε πεδία όπου υφίσταται αλληλεπίδραση, π.χ. άνθρωπος και πληκτρολόγιο. Τα αποτελέσματα των παραπάνω ερευνών είναι σημαντικά για την αναγνώριση κινήσεων στο πληκτρολόγιο και στο χορό, ενώ έχουν αρκετά κοινά χαρακτηριστικά με το πιάνο.

2.4.2 Αισθητήρες

Η τεχνολογία των αισθητήρων στην αναγνώριση κινήσεων είναι μια διαδεδομένη τεχνολογία, η οποία προσφέρει μεγάλη ακρίβεια στην ανάκτηση της πληροφορίας και χρησιμοποιείται σε διάφορες επιστήμες. Πλην όμως, περιορίζει την ελευθερία του ατόμου υποχρεώνοντάς τον να τοποθετήσει επάνω του ειδικό εξοπλισμό, π.χ. γάντια, πάνω στα οποία είναι τοποθετημένοι αισθητήρες. Είναι προφανές ότι στην περίπτωση αλληλεπίδρασης μεταξύ μουσικού και οργάνου αυτό καθίσταται αδύνατο.

Οι τεχνολογίες αισθητήρων, για την αναγνώριση μαέστρων στη διεύθυνση ορχήστρας, διαχωρίζονται σε δύο κατηγορίες. Στην πρώτη κατηγορία ανήκουν οι αισθητήρες που απαιτούν επαφή με το σώμα του ανθρώπου που πραγματοποιεί την κίνηση και στη δεύτερη κατηγορία αυτοί που λειτουργούν εξ αποστάσεως. Έτσι, στην πρώτη κατηγορία συναντούμε ηλεκτρομυογράφους (*Electromyography* - *EMG*), οι οποίοι καταγράφουν την ηλεκτρική δραστηριότητα των μυών, μετρητές απόστασης αλλά και αισθητήρες κίνησης ματιού (*Electrooculography* - *EOG*). Στη δεύτερη κατηγορία συναντούμε υπέρυθρους αισθητήρες και δέκτες ραδιοκυμάτων.

Στις παρακάτω ενότητες περιγράφονται τα σημαντικότερα συστήματα αναγνώρισης κινήσεων στη ΜΑ και συγκεκριμένα στη διεύθυνση ορχήστρας. Παρατηρείται εύκολα, από το σχήμα 2.4.2.1, ότι η πορεία αυτών των συστημάτων έχει τις ρίζες της σχεδόν τρεις δεκαετίες πριν. Το γεγονός αυτό σε συνδυασμό με την εμπλοκή πολλών παγκόσμιων επιστημονικών φορέων αλλά και φορέων της μουσικής βιομηχανίας, αποδεικνύει την αναγκαιότητα έρευνας στον χώρο της ΜΑ και της αναγνώρισης κινήσεων, με τη βεβαιότητα ότι η όραση

ας στα οποία τοποθετήθηκαν αισθητήρες [Bongers 1999]. Επίσης, προσαρμόστηκαν επάνω στο μαέστρο δέκτες υπερήχων, οι οποίοι κατέγραφαν ένταση σήματος παραγόμενη από τα παπούτσια του. Μέσω των μετρήσεων των αποστάσεων χειρών και ύψους από το πάτωμα, υπολογίζονταν συνεχώς συντεταγμένες των χειρών στο χώρο. Παράλληλα, αισθητήρας κίνησης μετρούσε την ταχύτητα των χειρονομιών, ενώ κατάλληλη συσκευή ψηφιοποιούσε τα δεδομένα για τον Η/Υ.

Το «Bodysynth» αναπτύχθηκε από το μηχανικό Ed Severinghaus και τον καλλιτέχνη Chris Van Raalte το 1995. Πρόκειται για ασύρματο ελεγκτή MIDI, ο οποίος ενεργοποιείται από τις κινήσεις των μυών μέσω ηλεκτομυογραφικών αισθητήρων [Seifert 2002]. Μεταδίδει δεδομένα από το σήμα που λαμβάνει από τις κινήσεις των μυών. Τα δεδομένα είχαν τη μορφή εντολών όπως «έναρξη νότας», «portamento» και «συνέχεια νότας», μηνύματα που δέχονταν ο ελεγκτής MIDI και τα μετέτρεπε σε μουσική.

Το σύστημα «Conductor's Jacket» κατασκευάστηκε το 1996 από το MediaLab του πανεπιστημίου MIT [Marrin 1998]. Βασίζονταν στην ταυτοποίηση εικοσιπέντε διαφορετικών χειρονομιών από έξι διευθυντές ορχήστρας. Οι χειρονομίες αυτές μελετήθηκαν και στη συνέχεια σχεδιάστηκε σύστημα αναγνώρισης των χειρονομιών αυτών, ενώ τα αποτελέσματα τα οποία προέκυψαν, χρησιμοποιήθηκαν για τη βελτίωση των ερμηνευτικών δυνατοτήτων των μουσικών. Για την αναγνώριση των χειρονομιών επιλέχτηκε ένα ηλεκτρομυογραφικό σύστημα αποτελούμενο από μια σειρά από αισθητήρες τοποθετημένους σε ειδικό ένδυμα. Οι αισθητήρες λάμβαναν πληροφορίες για τη δραστηριότητα των μυών αυτού που το φοράει. Η επεξεργασία αυτών των πληροφοριών επέτρεπε τελικά την αναγνώριση της χειρονομίας. Η Marrin (1998) συμπέρανε ότι πολλές πληροφορίες της εκφραστικής επικοινωνίας στη ΜΑ μπορούν να καταγραφούν από το σήμα που παράγεται μέσα από τις κινήσεις. Η επεξεργασία αυτών των πληροφοριών απέδειξε την επιρροή των αναπνοών του μαέστρου στην εκφραστικότητα των μουσικών φράσεων από τους μουσικούς, την επιρροή βαθμού άσκησης μυσικής δύναμης στην ακουστότητα του ήχου και τέλος, την αμφιδεξιότητα των χειρών στις μουσικές μεταβολές.

Τέλος, το «Miburi» της Yamaha ήταν μια συσκευή που είχε εμπορικές εφαρμογές και προοπτικές. Αποτελούνταν από κλασικό ένδυμα με αισθητήρες που μετέδιδε σήμα ανάλογα με την κίνηση των αρθρώσεων. Κατάφερε να αποσπάσει το βραβείο G-mark το 1996 (Σχήμα 2.4.2.1.1). Έχει χρησιμοποιηθεί στην πράξη σε πολλές ζωντα-

νές παραστάσεις συμπεριλαμβανομένης και αυτής του Mort Subotnick, με τίτλο "Intimate Immensity" στο Lincoln Center Summer Festival. Το Miburi φορέθηκε από χορευτή, ο οποίος ήλεγχε με αυτό δύο πιάνο Yamaha Disklavier. Πρόκειται δηλαδή για μια συσκευή η οποία ανάλογα με το σήμα που παράγονταν από την κίνηση των αρθρώσεων ενεργοποιούσε κατάλληλα ηλεκτρικά όργανα [Vickery 2001].



Σχήμα 2.4.2.1.1 Το Miburi στο test studio της Yamaha στο Τόκιο

2.4.2.2 Εξ αποστάσεως

Στην κατηγορία της αναγνώρισης κινήσεων με εξ αποστάσεως χρήση αισθητήρων συναντούμε ενδιαφέρουσες έρευνες και συστήματα με σημαντικά έμπρακτα αποτελέσματα.

Μια γενικότερη προσπάθεια στον τομέα της ανάλυσης εκφραστικών χειρονομιών γίνεται και από την επιστημονική ομάδα του

έργου Multisensory Expressive Gesture Applications (*MEGA*). Ο κύριος στόχος της είναι η έρευνα στον τομέα της κίνησης στη ΜΑ με στόχο το σχεδιασμό πρωτότυπων τρόπων ανάδρασης, στρατηγικών αντιστοίχισης χειρονομιών και πολυμεσικών διεπαφών, οι οποίες βασίζονται στην πολυμορφική ανάλυση του εκφραστικού περιεχομένου της μουσικής. Για την επίτευξη των στόχων αυτών έχει δημιουργηθεί το «MEGA System Environment» βασισμένο στη διαδικτυακή πλατφόρμα ανοικτού κώδικα «EyesWeb», που αναπτύχθηκε από το InfoMus Lab του πανεπιστημίου της Genova στην Ιταλία [Camurri 2004]. Έχουν αναπτυχθεί συστήματα ανάκτησης δακτυλισμών μέσω όρασης Η/Υ χρησιμοποιώντας την πλατφόρμα EyesWeb, τα οποία θα αναπτυχθούν σε επόμενη ενότητα.

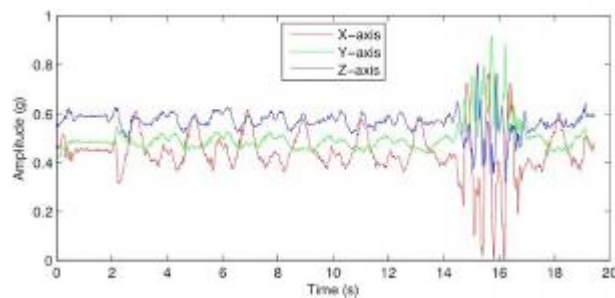
Το «Virtual Conductor» αναπτύχθηκε το 2001 υπό την επίβλεψη του καθηγητή Max Mühlhäuser σε επιστημονική ομάδα η οποία αποτελούνταν από επιστήμονες από τα πανεπιστήμια Darmstadt, Linz, Ulm, Stanford. Πρόκειται για μια συσκευή η οποία σήμερα έχει ιδιαίτερα ψυχαγωγικό χαρακτήρα, αφού επιτρέπει στον οποιονδήποτε χρήστη να ελέγχει με τις χειρονομίες του κρατώντας μία υπέρυθρη μπαγκέτα, μια οπτικοακουστική καταγραφή της φιλαρμονικής της Βιέννης που προβάλλεται σε οθόνη (Σχήμα 2.4.2.2.1). Οι κινήσεις που αναγνωρίζονται επηρεάζουν την αναπαραγωγή του βίντεο αποδίδοντας διαφορετικό τέμπο, ένταση και έμφαση σε ομάδες οργάνων [Hofmann 2001].

Το 2003 αναπτύχθηκε το «You Are the Conductor» από τη συνεργασία του πανεπιστημίου του Stanford, του Immersion Music και του ETH Zurich. Πρόκειται για μια εντελώς παρόμοια κατασκευή με το Virtual Conductor. Είναι ειδικά σχεδιασμένη για παιδιά και φυλάσσεται στο μουσείο της Βοστώνης. Τέλος, μια εντελώς πρόσφατη έρευνα στην αναγνώριση κινήσεων στη ΜΑ είναι αυτή του Grunberg (2008) από το Media Entertainment Technology Lab (METLab) του Πανεπιστημίου Drexel των ΗΠΑ. Ο Grunberg βασίζεται στην τεχνολογία Wii της Nintendo για ασύρματη επικοινωνία με ελεγκτή χειρονομιών. Χρησιμοποιείται το αντίστοιχο χειριστήριο Wiimote της κονσόλας παιχνιδιών για τη μετάδοση των κινήσεων (Σχήμα 2.4.2.2.2), ενώ η μοντελοποίηση και η αναγνώριση των κινήσεων γίνεται με KMM. Το Wiimote διαθέτει μετρητές επιτάχυνσης (Σχήμα 2.4.2.2.2) που επιτρέπουν την καταγραφή και ψηφιοποίηση επιταχυνόμενων κινήσεων. Ο Grunberg περιορίστηκε στις κινήσεις που διαγράφονται πάνω σε ευθεία γραμμή και ημικύκλια, οι οποίες αποτελούν και το αλφαβητάρι κινήσεων των μαέστρων [Grunberg 2008]. Χρησιμοποιήθηκαν περισσότεροι από ένας άνθρωποι για τη

μοντελοποίηση των κινήσεων. Η κίνηση μοντελοποιήθηκε με τη χρήση KMM 10 εσωτερικών καταστάσεων (Σχήμα 2.4.2.2.3). Παρόλο που για ένα άτομο τα αποτελέσματα της μεθόδου κρίνονται ικανοποιητικά, εντούτοις δε λειτουργούν το ίδιο ικανοποιητικά για περισσότερα άτομα.



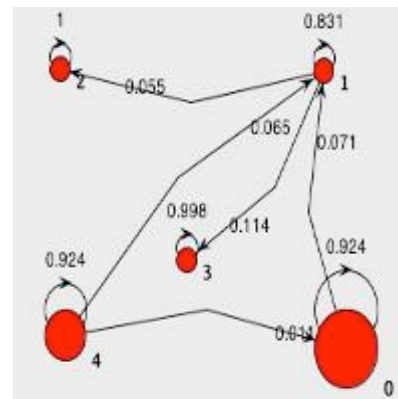
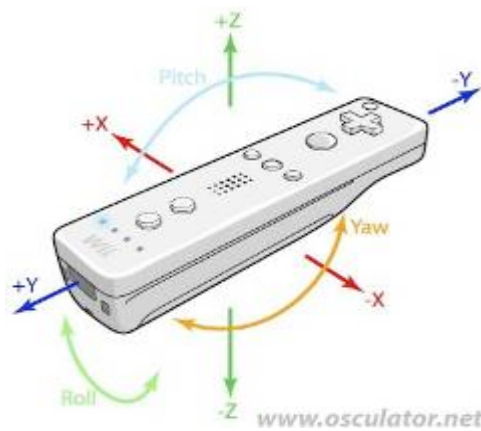
Σχήμα 2.4.2.2.1 Εικόνα του Virtual Conductor στο Vienna's House of Music



Σχήμα 2.4.2.2.2 Πλάτος σήματος κινήσεων σε συνάρτηση με το χρόνο

2.4.3 Συμπεράσματα

Είναι γεγονός ότι οι αισθητήρες κατέχουν το μεγαλύτερο μερίδιο στον τομέα της αναγνώρισης κινήσεων με εφαρμογή στη ΜΑ. Αυτό οφείλεται σε δύο κύριους λόγους. Πρώτον, στο γεγονός ότι η ανάγκη της αναγνώρισης κινήσεων στη ΜΑ, αλλά και η τεχνολογία των αισθητήρων, γεννήθηκαν και αναπτύχθηκαν την ίδια περίπου εποχή. Αντίθετα, εναλλακτικές μέθοδοι βασίζονται σε σχετικά νέους επισημονικούς τομείς, όπως π.χ. η όραση Η/Υ. Δεύτερον, στο γεγονός ότι η τεχνολογία των αισθητήρων προσφέρει μεγάλη ακρίβεια στην ανακτώμενη πληροφορία.



Σχήμα 2.4.2.2.3 Το Wiimote

Σχήμα 2.4.2.2.4 Παράδειγμα KMM 5 καταστάσεων για μοντελοποίηση κίνησης στο Wiimote

Δεν είναι τυχαίο το γεγονός ότι χρησιμοποιήθηκαν στη διεύθυνση ορχήστρας. Είναι ένα πεδίο εφαρμογής όπου δεν απαιτείται επαφή των χεριών του ανθρώπου με κάποιο μουσικό όργανο. Η περίπτωση όμως της εγκατάστασης αισθητήρων ή σφαιριδίων στον πιανίστα την ώρα που ερμηνεύει είναι εντελώς μη λειτουργική και απαξιωμένη από τη μουσική κοινότητα.

Αντίθετα, τεχνολογίες όρασης Η/Υ προσφέρουν πρόσβαση σε πληροφορίες καλής ακρίβειας χωρίς επαφή με τον άνθρωπο, είτε πρόκειται για χρήστη μιας διεπαφής, είτε πρόκειται για μουσικό κλπ. Η χρήση των KMM προσδίδει ακόμα καλύτερη ακρίβεια στην αναγνώριση αφού λαμβάνει υπόψη το στοχαστικό χαρακτήρα των κινήσεων.

2.5 Ανάκτηση δακτυλισμών

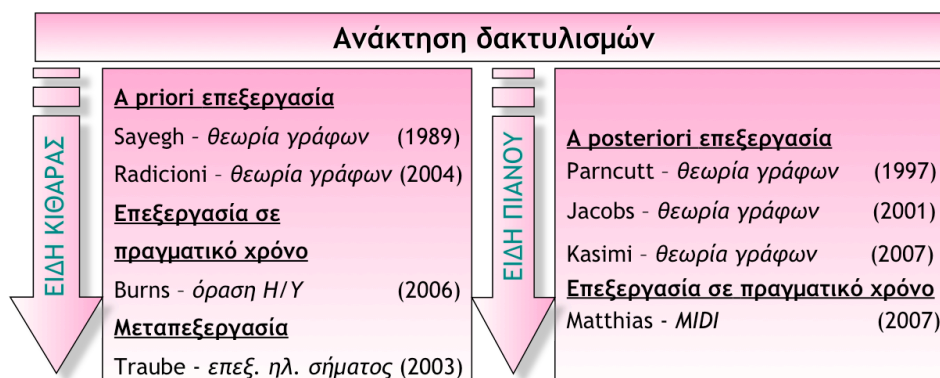
Στα έγχορδα όργανα, όπως είναι η κιθάρα, υπάρχουν πολλοί δυνατοί συνδυασμοί χορδών, δάκτυλων και θέσεων των δακτύλων στο τάστο, μέσω των οποίων μπορεί να παραχθεί ο ίδιος μουσικός τόνος. Παρόμοια είναι η προσέγγιση για τα κρουστά και τα πνευστά όργανα. Δυστυχώς όμως, οι δακτυλισμοί σπάνια αναγράφονται πάνω στην παρτιτούρα.

Οι έρευνες για την ανάκτηση δακτυλισμών, όπως έχει αναφερθεί προηγουμένως, ταξινομήθηκαν σε τρεις επιμέρους κατηγορίες: α) *A priori* επεξεργασία της παρτιτούρας, β) επεξεργασία, κινήσεων ή συμβάντων σε πραγματικό χρόνο γ) *A posteriori* επεξεργασία του ηλεκτρικού σήματος. Στην κατηγορία της προεπεξεργασίας της παρτιτούρας, συναντούμε τη θεωρία γράφων μέσα από σημαντικές έρευνες τόσο στην κιθάρα όσο και στο πιάνο. Η δεύτερη κατηγορία τμηματοποιείται με τη σειρά της σε άλλες δύο υποκατηγορίες, την επεξεργασία συμβάντων και την επεξεργασία κινήσεων σε πραγματικό χρόνο. Στην πρώτη υποκατηγορία, συναντούμε έρευνες οι οποίες στηρίζονται στην τεχνολογία MIDI και απευθύνονται αποκλειστικά σε ηλεκτρονικά όργανα όπως το συνθεσάιζερ και η ηλεκτρική κιθάρα. Στη δεύτερη υποκατηγορία, διακρίνουμε μια πολύ πρόσφατη και ενδιαφέρουσα πρόταση βασισμένη σε όραση H/Y, η οποία απευθύνεται αποκλειστικά στην κιθάρα, ηλεκτρική ή κλασική. Τέλος, στην μεταεπεξεργασία συναντούμε μεθόδους ψηφιακής επεξεργασίας ηλεκτρικού σήματος, οι οποίες απευθύνονται κυρίως στην ηλεκτρική κιθάρα.

Στο σχήμα 2.5.1 αναφέρονται οι σημαντικότερες έρευνες στο πεδίο της ανάκτησης δακτυλισμών χωρισμένες ανά κατηγορίες και όργανα εφαρμογής. Σε ό,τι αφορά το κλασικό πιάνο καταγράφηκαν έρευνες ανάκτησης δακτυλισμών βασιζόμενες αποκλειστικά στη θεωρία γράφων ως προεπεξεργασία της παρτιτούρας. Απουσιάζει από τη βιβλιογραφία οποιαδήποτε ολοκληρωμένη πρόταση για πιάνο σε πραγματικό χρόνο ή με *a posteriori* επεξεργασία. Τέλος, εύκολα συμπεραίνει κανείς ότι η θεωρία γράφων στην ανάκτηση δακτυλισμών, ανεξαρτήτως οργάνου, μονοπώλησε το ενδιαφέρον των επιστημόνων.

2.5.1 Ιδιαιτερότητες στην τεχνική του πιάνου και της κιθάρας

Οι ιδιαιτερότητες, αλλά και τα κοινά σημεία, που έχει η τεχνική του πιάνου, με αυτήν της κιθάρας, αντανακλώνται στις μεθόδους πληροφορικής που αναπτύχθηκαν με σκοπό να εφαρμοστούν σε αυτόν τον τομέα. Η βασική διαφορά στην τεχνική μεταξύ των δύο οργάνων έγκειται στις διαφορετικές λειτουργίες που αναλαμβάνουν να διεκπεραιώσουν τα δύο χέρια.



Σχήμα 2.5.1 Εξέλιξη συστημάτων ανάκτησης δακτυλισμών στη MA

Ιδιαίτερο στοιχείο είναι η στάση των χεριών παίζοντας κιθάρα. Στην κιθάρα, για τους δεξιόχειρες, το αριστερό χέρι επιλέγει τη νότα που θα ακουστεί ενώ το δεξί αναλαμβάνει να την παίξει. Στο πιάνο, και τα δύο χέρια πραγματοποιούν παρόμοιες λειτουργίες. Με άλλα λόγια, στο πιάνο οποιοδήποτε χέρι μπορεί να παίξει μια μελωδία και το άλλο να συνοδεύει, κάτι το οποίο δεν ισχύει στην κιθάρα όπου η μελωδία ακούγεται μόνο από το αριστερό χέρι.

Αυτή η σημαντική διαφορά πρέπει να λαμβάνεται υπόψιν σε κάθε μέθοδο πληροφορικής που αναπτύσσεται απευθυνόμενη στο πιάνο ή την κιθάρα. Ειδικότερα, οφείλει ο ερευνητής να σεβαστεί το γεγονός ότι ο δακτυλισμός στην κιθάρα καθορίζεται από τρεις παραμέτρους, το δάκτυλο, τη χορδή και τη θέση στο τάστο (λαιμός), ενώ ο δακτυλισμός στο πιάνο εξαρτάται από δύο παραμέτρους, το δάκτυλο και το πλήκτρο. Έτσι ορίζεται η τριάδα (δάκτυλο, χορδή, τάστο) (Σχήμα 2.5.2) για κιθάρα και η δυάδα (δάκτυλο, πλήκτρο) για πιάνο. Αυτές οι ιδιαιτερότητες των δύο οργάνων δεν απέκλεισαν την πραγματοποίηση παράλληλων ερευνών για την ανάκτηση δακτυλισμών. Ένα καλό παράδειγμα αποτελούν οι μέθοδοι που βασίζονται στη θε-

ωρία γράφων. Αν ο μελετητής λάβει σοβαρά υπόψη του τις ιδιαιτερότητες και τα κοινά σημεία των δύο οργάνων, τότε εύκολα μελετώντας μια μέθοδο που απευθύνεται στην κιθάρα μπορεί να κατανοήσει και μια άλλη παράλληλη που απευθύνεται στο πιάνο, ή το αντίστροφο.



Σχήμα 2.5.2 Θέση στο τάστο της κιθάρας και χορδή

2.5.2 Θεωρία γράφων στην a priori επεξεργασία της παρτιτούρας

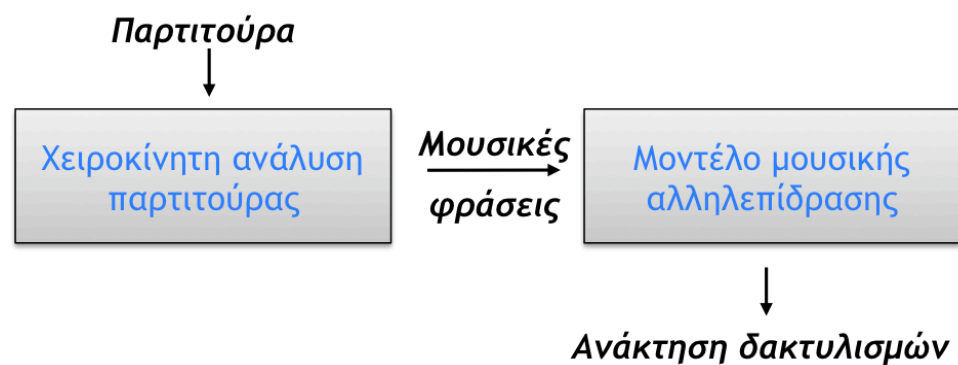
Οι μέθοδοι που βασίζονται στη θεωρία γράφων δέχονται ως είσοδο την παρτιτούρα, την αναλύουν και παράγουν τους ιδανικούς δακτυλισμούς ως έξοδο. Η μέθοδος του Sayegh (1989), μαζί με την επέκτασή της από τον Radicioni (2004), προτείνουν μια υπολογιστικά αποδοτική λύση για έγχορδα όργανα. Η μέθοδος του Parncutt (1997), μαζί με την επέκτασή της από τους Jacobs (2001) και Kasimi (2007), προτείνουν ένα εργονομικό μοντέλο καθορισμού δακτυλισμών για όργανα με πλήκτρα.

Η πλειονότητα των μεθόδων βασίζεται στην αναπαράσταση όλων των πιθανών τριάδων (*κιθάρα*) ή δυάδων (*πιάνο*) χρησιμοποιώντας κόμβους (*nodes*) ενός γράφου (*graph*). Η διαδικασία ξεκινάει αντιστοιχίζοντας ένα βάρος (*weight*) σε κάθε ακμή (*edge*) σύμφωνα με προκαθορισμένο σύνολο κανόνων. Ο γράφος αναζητείται με σκοπό την εύρεση του ιδανικού μονοπατιού που αντιστοιχεί στον υπολογισθέντα ιδανικό δακτυλισμό. Οι δακτυλισμοί των χορδών καθορίζονται με βάση την υπολογιστική θεωρία του Constraint Satisfaction Problem (CSP) [Radicioni 2004a].

Πυρήνας της κατηγορίας των ερευνών που θα μελετηθούν σε αυτήν την ενότητα αποτελούν οι μέθοδοι του Daniele Radicioni από το πανεπιστήμιο του Τορίνο της Ιταλίας (*κιθάρα*) και του Alia Al Kasimi

(πιάνο) και της επιστημονικής ομάδας του Christopher Raphael του Indiana University των ΗΠΑ. Θα αναπτυχθεί αναλυτικά η μέθοδος του Radicioni (2004) και στη συνέχεια θα αναδειχθούν οι διαφοροποιήσεις και οι καινοτομίες της πιο πρόσφατης σχετικής μεθόδου, του Kasimi (2007).

Συνοπτικά, οι μέθοδοι αυτές ξεκινούν με χειροκίνητη τμηματοποίηση της παρτιτούρας (Σχήμα 2.5.2.1), εφαρμόζεται μοντέλο ΜΑ στο κάθε τμήμα (μουσική φράση) και, τέλος, χρησιμοποιώντας αυτό το μοντέλο οι δακτυλισμοί υπολογίζονται και εξάγονται σε οποιοδήποτε άλλο σύστημα.



Σχήμα 2.5.2.1 Ανάκτηση δακτυλισμών με a priori επεξεργασία παρτιτούρας [Radicioni 2004b]

2.5.2.1 Μέθοδος Radicioni

Το πρώτο στάδιο της μεθόδου του Radicioni, αποτελείται από την ανάλυση (τμηματοποίηση) της παρτιτούρας με χειροκίνητη διαδικασία [Radicioni 2004a].

Παρόλο που έχουν καταγραφεί προσπάθειες για αυτόματη τμηματοποίηση παρτιτούρας μέσω της Global Optimization Approach (GOA) [Pardo 2000; Bod 2001], ο Radicioni προτίμησε τη χειροκίνητη τμηματοποίηση για τους παρακάτω λόγους:

1. Η GOA δεν είναι αξιόπιστη λόγω του ότι πρακτικά οι μουσικοί χωρίζουν τα μουσικά έργα σε φράσεις, θέματα, μοτίβα κλπ.

2. Η ερμηνεία ενός μουσικού έργου συνεπάγεται τη διάκριση και την έμφαση σε χαρακτηριστικές δομές της μουσικής, όπως είναι τα δομικά όρια.
3. Εφόσον οι ερμηνευτές στοχεύουν στην ανάδειξη των δομών του έργου μέσω της τμηματοποίησής του, είναι λογικό να επιδιώκουν τη χρήση συγκεκριμένων δακτυλισμών, μέσα στα τμήματα, ώστε να τονίσουν τα δομικά τους όρια.

Η διαφορά μεταξύ GOA και Radicioni έγκειται στο ότι, η GOA εντοπίζει τον ιδανικό δακτυλισμό συνολικά στο κομμάτι, ενώ ο Radicioni τον εντοπίζει πρώτα μέσα στις φράσεις και στη συνέχεια στα όρια μεταξύ των φράσεων.

Η μέθοδος, ως δεύτερο στάδιο, αναλαμβάνει να εφαρμόσει στις μουσικές φράσεις ένα μοντέλο MA, το οποίο θα παράξει τους δακτυλισμούς. Η τεχνική του Radicioni προσανατολίζεται στο σχεδιασμό του μοντέλου με βάση τις παραμέτρους κίνησης και ανατομίας. Το γεγονός αυτό είναι θετικό διότι ικανοποιεί μια βασική μουσική προϋπόθεση: την παραγωγή ομοιογενών ήχων [Sayegh 1989]. Εξασφαλίζει τη μικρότερη δυνατή μετατόπιση του χεριού, για όλες τις κατηγορίες οργάνων, παράγοντας ήχους ομοιόμορφης ποιότητας. Η σωστή θέση του χεριού στο κλαβιέ του πιάνου ή το λαιμό της κιθάρας είναι παράγοντας καθοριστικός, διότι επιτυγχάνεται πρόσβαση σε πλήκτρα ή χορδές αντίστοιχα, μόνο με το άνοιγμα των δακτύλων. Η μοντελοποίηση της MA, ως προς τις παραμέτρους κίνησης και ανατομίας, φαίνεται να είναι γνωστικά και μουσικά αποδεκτή. Εντούτοις, αξίζει να σημειωθεί ότι τα συμπεράσματα του Radicioni στηρίζονται σε μελέτες από μουσικά αποσπάσματα συγκεκριμένου είδους.

Οι Heijink και Meulenbroek (2002) μελέτησαν την πολυπλοκότητα της ερμηνείας στην κλασική κιθάρα. Στη μελέτη αυτή στηρίχθηκε ο Radicioni προκειμένου να μοντελοποιήσει τη MA στην κιθάρα. Όρισε δύο κατηγορίες κίνησης του χεριού: κίνηση **παράλληλα** (οριζόντια) και **εγκάρσια** (κατακόρυφα) του λαιμού της κιθάρας [Radicioni 2004b]. Διατύπωσε την άποψη ότι μεταξύ των δύο κατηγοριών, η κίνηση παράλληλα είναι πιο «επίπονη» για το μουσικό διότι απαιτεί μετατόπιση του χεριού, ενώ για την εγκάρσια κίνηση αρκεί η μετατόπιση των δακτύλων. Η δυσκολία μεταξύ των θέσεων δύο δακτυλισμών i και j υπολογίζεται σύμφωνα με τον παρακάτω τύπο:

$$\Delta\text{υσκολία}(i, j) = \text{παράλληλα} + \text{εγκάρσια} \quad (2.1)$$

όπου το μέγεθος «παράλληλα» υπολογίζεται αποδίδοντας μια τιμή αναφερόμενη στο βάρος δυσκολίας για την κάθε πιθανή θέση. Ο υπολογισμός του μεγέθους «εγκάρσια» βασίζεται στην απόρριψη «μη φυσικών» στάσεων του χεριού και την απόδοση υψηλότερων τιμών βάρους σε θέσεις με μεγάλο κατακόρυφο άνοιγμα των δακτύλων. Ο Radicioni δημιούργησε έναν πίνακα με το μέγιστο επιτρεπτό άνοιγμα των δακτύλων, εκφρασμένο σε τάστα (θέσεις στο τάστο), για κάθε ζεύγος δακτύλων [Radicioni 2005]. Χρησιμοποιώντας τον πίνακα αυτό όρισε βάρη δυσκολιών κατά την εγκάρσια κίνηση σύμφωνα με τους παρακάτω κανόνες:

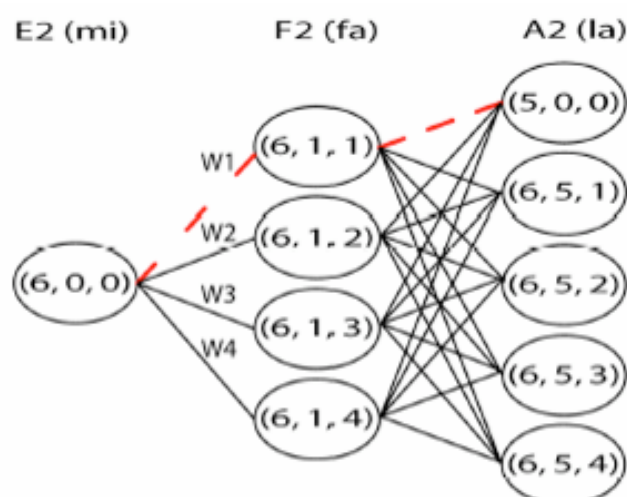
1. Δεν προστίθεται βάρος στη θέση αν το άνοιγμα των δακτύλων, που προκαλεί ο δακτυλισμός, είναι μικρότερο από το μέγιστο επιτρεπτό.
2. Προστίθεται βάρος στη θέση αν το άνοιγμα το δακτύλων είναι μεγαλύτερο από το μέγιστο επιτρεπτό και κατά συνέπεια απαιτείται μετατόπιση του χεριού.

Το πρόβλημα πλέον εστιάζεται στην εύρεση του μονοπατιού ελαχίστου βάρους σε κατευθυνόμενο ακυκλικό γράφο. Η τεχνική αυτή, με τον τεμαχισμό της παρτιτούρας σε μουσικές φράσεις, δημιουργεί και επιλύει γράφους εσωτερικά της κάθε φράσης και στη συνέχεια «λύνει» το δακτυλισμό στα όρια μεταξύ των φράσεων. Στο σχήμα 2.5.2.1.1 απεικονίζεται η στρωματική δομή του γράφου όπου κάθε νότα έχει τουλάχιστον τέσσερις δυνατούς δακτυλισμούς, με εξαίρεση τις νότες που παίζονται σε ανοιχτή χορδή (*ανοιχτή χορδή εννοούμε εκείνη που παίζεται χωρίς να πιέσουμε κάποια νότα επάνω στην κιθάρα*), π.χ. μια πιθανή θέση (*χορδή, τάστο*) στο λαιμό της κιθάρας για κάθε ένα από τα τέσσερα δάκτυλα. Η στρωματική δομή εξασφαλίζει ότι το σύστημα μπορεί να επιλύσει το πρόβλημα του μονοπατιού ελαχίστου βάρους σε γραμμική μορφή $O(n)$, αναφορικά με τον αριθμό νοτών (n) και μια κρυφή σταθερή πολυπλοκότητα που σχετίζεται με τον αριθμό των ακμών ανάμεσα σε κάθε στρώμα (256 στη χειρότερη περίπτωση).

Ο δακτυλισμός μιας συνήχησης απαιτεί μια προσέγγιση διαφοροποιημένη σε σχέση με αυτήν της μεμονωμένης νότας. Ο λόγος είναι ότι η συνήχηση προϋποθέτει περισσότερους από ένα δακτυλισμό τη φορά. Συνεπώς, και ο τρόπος υπολογισμού των δακτυλισμών είναι διαφορετικός. Οι Truchet (2004) και Radicioni (2005) συμφωνούν ότι με τη χρήση του CSP το πρόβλημα υπολογισμού του δακτυ-

λισμού συγχορδίας μπορεί να επιλυθεί. Το CSP αναθέτει τιμές σε μεταβλητές που ικανοποιούν ένα σύνολο από περιορισμούς. Η έννοια της μεταβλητής αντιστοιχεί σε μεμονωμένη νότα, η οποία μαζί με άλλες συνθέτουν τη συνήχηση. Οι τιμές που ανατίθενται είναι όλες οι δυνατές τριάδες (δάκτυλο, χορδή, τάστο), οι οποίες δεν ξεπερνούν τις 16 για κάθε νότα. Οι περιορισμοί κίνησης και ανατομίας, για τον Radicioni, είναι ένα σύνολο κανόνων που λαμβάνονται υπόψη για τον υπολογισμό του «ιδανικού» δακτυλισμού.

Τέλος, είναι πιθανόν να πάρουμε περισσότερες λύσεις που ικανοποιούν το σύνολο των κανόνων. Ο Radicioni πρότεινε μια κατάταξη για την επιλογή της καλύτερης λύσης. Για να επιλύσει το πρόβλημα του δακτυλισμού στην περίπτωση εναλλαγής συνηχήσεων ή μεικτής μετάβασης από τη μελωδία μιας νότας σε συγχορδίες, συμπεριέλαβε τις συγχορδίες στην αναζήτηση του γράφου και την επίλυσή του ως υποπρόβλημα, χρησιμοποιώντας την τεχνική του CSP.



Σχήμα 2.5.2.1.1 Παράδειγμα γράφου που παράχθηκε από τις νότες μι-φα-λα [Radicioni 2005](τα βάρη παραλείπονται από το σχήμα - η διακεκομμένη γραμμή αναπαριστά το μονοπάτι ελαχίστου βάρους).

2.5.2.2 Μέθοδος Kasimi

Η μέθοδος Kasimi (2007) (πιάνο) και Radicioni (2004) (κιθάρα) έχουν παρόμοιο σκεπτικό (Σχήμα 2.5.2.2.1). Ο Kasimi επέκτεινε παλαιότερη μέθοδο του Parncutt (1997). Διαφοροποιείται στο γεγονός ότι μπορεί να δεχτεί ως είσοδο παρτιτούρα πολυφωνικής μουσικής αλλά και στο ότι παράγει αποτελέσματα ευκολότερα στη διαχείρισή τους.

Το βασικότερο χαρακτηριστικό της μεθόδου αυτής είναι η μετατροπή της πολυφωνικής παρτιτούρας σε μονοφωνική, αναλύοντάς την κατακόρυφα. Στηρίζεται στην ιδέα ότι μπορούμε να ξαναγράψουμε την πολυφωνική μουσική ως σύνολο από συγχορδίες και κρατημένες νότες.



Σχήμα 2.5.2.2.1 1: Απόσπασμα του Aufschwung (*Phantasiestück op. 12*) του Schumann, 2: Κατακόρυφη ανάλυση του ίδιου αποσπάσματος [Kasimi 2007]

Σε αντιστοιχία με την εγκάρσια ή παράλληλη κίνηση στη μέθοδο Radicioni, καθώς και τα βάρη που ανατίθενται ανάλογα με το είδος της, στη μέθοδο Kasimi συναντούμε το κατακόρυφο ή οριζόντιο βάρος που προκαλείται από την κίνηση. Το συνολικό βάρος μιας κίνησης προκύπτει ως άθροισμα οριζόντιου και κατακόρυφου βάρους.

Για τον υπολογισμό του κατακόρυφου βάρους ενός κόμβου, τμηματοποιεί τους δακτυλισμούς μέσα στον κόμβο σε ζευγάρια από νότες. Για παράδειγμα, στην τονική της φα μείζονα, για την τριάδα φα-λα-ντο, χρησιμοποιούμε 1°-3°-5° δάκτυλο. Η μέθοδος χωρίζει την τριάδα σε δύο ζευγάρια ΦΑ-ΛΑ (1°-3°) και ΛΑ-ΝΤΟ (3°-5°). Το βάρος για κάθε ζευγάρι εξαρτάται από το ποια δύο δάκτυλα χρησιμοποιήθηκαν και από την απόσταση ανάμεσα στις νότες. Οπότε αποδίδεται βάρος [Kasimi 2007]:

$$Cost_v(n_i) = \sum_{k=1}^{pairs} Cost_v(p_k, d) \quad (2.5.2.2.1)$$

Όπου:

- $Cost_v$, το κατακόρυφο βάρος,

- n_i , ένα σύνολο από νότες,
- p_k , το k -οστό ζευγάρι p των δακτύλων,
- d , η απόσταση των νοτών του ζευγαριού.

Το οριζόντιο βάρος μεταξύ δύο κόμβων υπολογίζεται ως ο μέσος όρος του βάρους μετατόπισης, λαμβάνοντας υπόψη όλους τους συνδυασμούς ζευγαριών για κάθε νότα της πρώτης συγχορδίας με κάθε νότα της επόμενης συγχορδίας [Kasimi 2007]:

$$Cost_h(n_a \rightarrow n_b) = \frac{1}{N} \sum_{i \in I_a, j \in I_b} c_h(i \rightarrow, j, d) \quad (2.5.2.2.2)$$

Όπου:

- $Cost_h$, το οριζόντιο βάρος,
- n_a, n_b δύο σύνολα από νότες,
- N , το πλήθος των ζευγαριών των δακτύλων,
- d , η απόσταση των νοτών του ζευγαριού.

Για το παράδειγμα του σχήματος 2.5.2.2.2, το οριζόντιο βάρος $Cost_h(p, d)$ είναι το άθροισμα των βαρών μετάβασης μεταξύ ΝΤΟ και ΦΑ, ΝΤΟ και ΛΑ, ΜΙ και ΦΑ, ΜΙ και ΛΑ, ΣΟΛ και ΦΑ, ΣΟΛ και ΛΑ, δια 5.



Σχήμα 2.5.2.2.2 Παράδειγμα για οριζόντιο βάρος

Τέλος, τα $Cost_v$ και $Cost_h$ θεωρούνται ίσα στην περίπτωση που π.χ. οι νότες ΝΤΟ-ΜΙ (1° - 3°) παιχτούν ως συγχορδία ($Cost_v$) ή σαν ανιούσα κίνηση σε διάστημα 3^{ns} ($Cost_h$). Υπάρχουν δύο περιπτώσεις όπου παρόλο που ο ισχυρισμός είναι αναληθής, ο Kasimi τον εφαρμόζει. Πρώτον, είναι επιτρεπτή η οριζόντια μετακίνηση του χεριού μεταξύ νοτών, αποδίδοντας αντίστοιχο οριζόντιο βάρος. Δεύτερον, αν η κατεύθυνση της κίνησης είναι κατιούσα, η

διασταύρωση των δακτύλων είναι αναγκαία και η δυσκολία αυξάνει. Ο Kasimi αγνοεί το πρώτο από τα δύο προβλήματα θέτοντας $c_v = c_h$ και στη δεύτερη περίπτωση το c_h καθορίζεται από ξεχωριστή συνάρτηση. Στις διασταυρώσεις που προέρχονται από τον αντίχειρα αντιστοιχίζεται μικρότερο βάρος. Είναι καλό να σημειωθεί ότι η μέθοδος αυτή δε λαμβάνει υπόψη καθόλου τη διαφορά μεταξύ άσπρων και μαύρων πλήκτρων.

2.5.3 Τεχνολογία MIDI στην επεξεργασία σε πραγματικό χρόνο

Η μέθοδος που θα αναλυθεί στην παρούσα ενότητα είναι η μέθοδος που αναπτύχθηκε από την Robine Matthias στο πανεπιστήμιο του Bordeaux 1 της Γαλλίας. Η μέθοδος αυτή στηρίζεται σε δύο βασικούς άξονες: τη θεωρία γράφων και τη τεχνολογία MIDI. Το σκέλος της θεωρίας γράφων, που χρησιμοποιείται από την Matthias (2007), δε διαφοροποιείται ιδιαίτερα από τις μεθόδους που αναλύθηκαν σε προηγούμενες παραγράφους. Για το λόγο αυτό θα αναπτυχθεί η μέθοδος εστιάζοντας στην επεξεργασία σε πραγματικό χρόνο με χρήση MIDI.

Η χρήση της τεχνολογίας MIDI στην ανάκτηση δακτυλισμών χαρακτηρίζεται από ένα βασικό πλεονέκτημα και ένα εξίσου βασικό μειονέκτημα. Το πλεονέκτημα εστιάζεται στο γεγονός ότι είναι μια ευρέως διαδεδομένη τεχνολογία, σχετικά χαμηλού κόστους, η οποία μπορεί να προσφέρει λύσεις σε πραγματικό χρόνο. Το μειονέκτημα είναι ότι η χρήση του MIDI ως μέθοδο ανάκτησης δακτυλισμών αποκλείει αυτόματα το ενδιαφέρον για κλασικά όργανα.

Η Matthias (2007) ισχυρίστηκε ότι η απόδοση βαρών στους επιμέρους κόμβους των γράφων υλοποιείται με εμπειρική ανάθεση από τους Parncutt, Jacobs, Sayegh και Radicioni. Ισχυρίζεται επίσης ότι η πειραματική ανάθεση, με χρήση MIDI σε πραγματικό χρόνο, μπορεί να αποφέρει αποτελέσματα που είναι προσαρμοσμένα στον κάθε ερμηνευτή.

Η μέθοδος συμβολίζει με το γράμμα b τα άσπρα πλήκτρα και με το γράμμα n τα μαύρα πλήκτρα. Χρησιμοποιείται ταυτόχρονα ένας αριθμός, ο οποίος αναπαριστά το πλήθος των συνόρων μεταξύ των πλήκτρων μετάβασης με πρόσημο $+$ εάν πρόκειται για ανοδική μετάβαση και πρόσημο $-$ για καθοδική μετάβαση. Π.χ. η ακολουθία NTO-PE-MI (πρόκειται για τρεις γειτονικές νότες) παράγει ως λίστα μετάβασης ($bb1$, $bb1$). Η μέθοδος βασίζεται πρώτον στο συντελεστή *staccato*, το συντελεστή που καταγράφει το χρόνο μεσολάβησης ανάμεσα στο πάτημα δύο νοτών, και δεύτερον την ταχύτητα πατήματος των πλήκτρων [Matthias 2007].

Ξεκινάει με την κατασκευή ενός γράφου δακτυλισμών, ο οποίος έχει τις μεταβάσεις από τη μια νότα στην άλλη ως ακμές και τους αριθμούς των δακτύλων ως κορυφές. Εφαρμόζει το «φυσιολογικό φιλτράρισμα» σύμφωνα με το οποίο απορρίπτονται μεταβάσεις που θεωρούνται αδύνατες, όπως για παράδειγμα από ΝΤΟ σε ΡΕ (*bb1*) με δάκτυλα $5^{\circ}-1^{\circ}$. Στη συνέχεια προτείνεται ένας τύπος συσχέτισης των τιμών που προκύπτουν από τις τυπικές αποκλίσεις χρονικών παραμέτρων (*staccato*, *ταχύτητα*), μεταξύ:

1. ερμηνείας της κλίμακας του ΝΤΟ σε τέσσερις οκτάβες για τέσσερις ρυθμούς, μόνο με το δεξί χέρι και,
2. το θεωρητικό ρυθμό και τις διαφορές στα πλάτη συχνοτήτων της κλίμακας του ΝΤΟ.

Η Matthias συνεπώς, βασίζεται στη διαφορά που υπάρχει ανάμεσα στην ανάλυση των θεωρητικών παραμέτρων της κλίμακας του ντο, και των παραμέτρων που παρατηρήθηκαν κατά την ερμηνεία.

Η διαφορά της ταχύτητας μεταξύ δύο δακτύλων για τη μετάβαση από μια νότα σε μια άλλη, αποτελεί μια σημαντική παράμετρο. Προτείνεται η μείωση της τιμής της συσχέτισης, πολλαπλασιάζοντάς την με συντελεστή $\alpha < 1$, αν η διαφορά της ταχύτητας μεταξύ των δύο δακτύλων έχει το ίδιο πρόσημο με την αναμενόμενη ταχύτητα από την ανάλυση παρτιτούρας. Στην αντίθετη περίπτωση προτείνεται ο πολλαπλασιασμός με $\beta > 1$. Άρα ο τύπος συσχέτισης C_i , μπορεί να εκφραστεί ως εξής:

$$C_i(d_1, d_2) = u_i(d_1, d_2) \cdot [c_s - cs_i(d_1, d_2)]^2 \quad (2.5.3.1)$$

Όπου:

- t , η μετάβαση από μια νότα σε μια άλλη,
- c_s , ο θεωρητικός συντελεστής *staccato*,
- cs_i , ο συντελεστής *staccato* που καταγράφηκε κατά τη μετάβαση t από το δάκτυλο d_1 στο δάκτυλο d_2 σε πραγματικό χρόνο,
- u_i , η ταχύτητα μετάβασης με:

$$u_i(d_1, d_2) = \begin{cases} a, & \text{αν } du_i(d_1, d_2) \cdot du \geq 0 \\ \beta, & \text{αν } du_i(d_1, d_2) \cdot du < 0 \end{cases} \quad (2.5.3.2)$$

Όπου:

- α, σταθερά με $\alpha < 1$,
- β, σταθερά με $\beta > 1$,
- $du_i(d_1, d_2)$, η διαφορά των καταγεγραμμένων ταχύτητας μεταξύ των δακτύλων d_1 και d_2 ,
- du , η διαφορά που προτείνεται από τις κλίμακες.

Με αυτόν τον τύπο επιτυγχάνεται η απόδοση ενός βάρους σε κάθε κορυφή του γράφου. Έτσι, αποκλείονται μονοπάτια που με τη θεωρία γράφων αναδείχτηκαν ως μονοπάτια ελαχίστου βάρους και στην πραγματικότητα δεν αποτελούν χρηστικούς δακτυλισμούς, ενώ στο τέλος διατηρείται μόνο το καλύτερο μονοπάτι. Η παρούσα μέθοδος εφαρμόζεται για κάθε μετάβαση της παρτιτούρας.

Η ανάκτηση δακτυλισμών με χρήση MIDI σε ερμηνεία σε πραγματικό χρόνο, μπορεί να αποφέρει ασφαλέστερα αποτελέσματα σε σχέση με τα αποτελέσματα που παράγει η θεωρία γράφων στην προεπεξεργασία παρτιτούρας. Αυτό οφείλεται στην ανάδειξη μοναδικού ιδανικού δακτυλισμού. Η μέθοδος μπορεί να λειτουργήσει καλύτερα όταν προσαρμοστεί στον κάθε ερμηνευτή χωριστά.

2.5.4 Όραση H/Y στην κιθάρα στην επεξεργασία σε πραγματικό χρόνο

Στην ενότητα αυτή θα αναλυθεί μια μέθοδος που αναπτύχθηκε από την Anne-Marie Burns το 2006 του πανεπιστημίου McGill στο Montréal του Καναδά. Η μέθοδος αφορά την ανάκτηση δακτυλισμών στην κιθάρα μέσω όρασης H/Y.

Η μέθοδος αυτή εστιάζει στις κινήσεις του κιθαρίστα μέσω της καταγραφής τους από δύο εγκατεστημένες κάμερες. Η πρώτη καταγράφει α) τη γενική όψη (*global view*) και β) την όψη του αριστερού χεριού του κιθαρίστα (*top view*) (Σχήμα 2.5.4.1). Η δεύτερη είναι εγκατεστημένη με ειδική βάση στο λαιμό της κιθάρας, ώστε να μην επηρεάζεται από τις ταλαντώσεις του λαιμού, ενώ λαμβάνει εικόνα από τα δάχτυλα του χεριού (*camera view*) (Σχήμα 2.5.4.2). Διενεργείται απομόνωση της Περιοχής Ενδιαφέροντος (ΠΕ) στην εικόνα, από τα στοιχεία του φόντου. Εφαρμόζεται τμηματοποίηση, υπολογί-

ζοντας τα ελάχιστα της καμπύλης της κίνησης μέσω της διαφοράς σε κάποιες παραμέτρους που αφορούν τα εικονοστοιχεία ανάμεσα στα καρέ. Η επεξεργασία της κίνησης, που διενεργείται από κάθε δάκτυλο χωριστά, υλοποιείται από αλγόριθμο εντοπισμού θέσης [Burns 2006]. Γενικά, η προτεινόμενη μέθοδος βασίζεται σε τρεις άξονες: α) εντοπισμός δακτύλου, β) εντοπισμός τάστου και χορδής, και γ) τμηματοποίηση κίνησης.

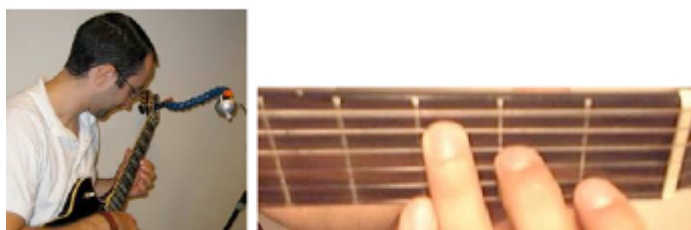
Σε ό,τι αφορά τον εντοπισμό δακτύλου χρησιμοποιείται ο κυκλικός μετασχηματισμός Hough, ο οποίος αναπτύχθηκε από την πλατφόρμα αναγνώρισης κίνησης EyesWeb. Χρησιμοποιήθηκε αυτός ο μετασχηματισμός από την Burns (2006) διότι πρώτον, παρουσιάζει υψηλό βαθμό ακρίβειας της πληροφορίας, δεύτερον, μπορεί να εφαρμοστεί σε σύνθετα περιβάλλοντα όπου υπάρχει μερική καταγραφή του χεριού και τρίτον, μπορεί να δουλέψει σε εικόνες που έχουν υποστεί επεξεργασία ανίχνευσης ακμών.



Σχήμα 2.5.4.1 1: Γενική όψη (*Global view*), 2: Όψη αριστερού χεριού (*Top view*) [Burns 2006]

Ο εντοπισμός τάστου και χορδής αναδεικνύει τον εντοπισμό δακτύλου που υλοποιήθηκε προηγουμένως. Πρόκειται για μια αντιστοιχία του δακτύλου με το χώρο στον οποίο βρίσκεται. Ο χώρος στο πεδίο εφαρμογής της ερμηνείας σε κιθάρα καθορίζεται από τις συντεταγμένες της χορδής και της θέσης στο τάστο. Η απομόνωση της ΠΕ στην εικόνα είναι μια διαδικασία που προφανώς προηγείται. Ο αλγόριθμος ξεκινάει από τον εντοπισμό του λαιμού της κιθάρας, ενώ στη συνέχεια εντοπίζονται οι χορδές και το τάστο μέσω εφαρμογής κατωφλίωσης στην εικόνα που αποτελείται μόνο από τόνους του γκρι. Χρησιμοποιείται φίλτρο Sobel οριζόντια και κατακόρυφα και παράγονται δύο διαφορετικές εικόνες. Έτσι

αναδεικνύονται οι κλίσεις, οριζόντιες και κατακόρυφες, που υπήρχαν στην αρχική εικόνα. Στη συνέχεια, ο γραμμικός μετασχηματισμός Hough εφαρμόζεται πάνω στις δύο εικόνες. Η εφαρμογή αυτή θα επιτρέψει την ανίχνευση γραμμικότητας σε ομάδες εικονοστοιχείων, παράγοντας ένα σύνολο γραμμών πάνω στην εικόνα. Οι γραμμές αυτές ομαδοποιούνται έχοντας ως κριτήριο την εγγύτητά τους, σκοπεύοντας στον καθορισμό της θέσης των έξι χορδών και των θέσεων στο τάστο.



Σχήμα 2.5.4.2 1: Βάση κάμερας, 2: Όψη δακτύλων (*Camera view*) [Burns 2006]

Η Burns (2006), με την τμηματοποίηση της εικόνας, στοχεύει στον εντοπισμό θέσης των δακτύλων και όχι σε μια απλή ιχνηλάτηση των άκρων τους. Στοχεύει δηλαδή στο διαχωρισμό των βοηθητικών κινήσεων, από τις κινήσεις παραγωγής μουσικής. Αυτός ο στόχος ικανοποιείται σχετικά νωρίς, μέσω της γενικής όψης της κάμερας, ώστε να μη χρησιμοποιείται υπολογιστική ισχύς κατά τη διάρκεια της ερμηνείας.

Αυτή η έρευνα αποτέλεσε μια καλή προσπάθεια ανάπτυξης μεθόδου όρασης H/Y για εφαρμογή στη ΜΑ. Σίγουρα αποτελεί ένα σημείο αναφοράς στο χώρο της ανάκτησης δακτυλισμών. Εστιάζει περισσότερο στην κιθάρα και λιγότερο στα έγχορδα όργανα γενικά, ενώ από την άλλη, δεν εισέρχεται καθόλου στο χώρο της αναγνώρισης κινήσεων. Η πληροφορία ανακτάται αποκλειστικά μέσω της ψηφιακής επεξεργασίας εικόνας δίχως τη χρήση στοχαστικών διαδικασιών οι οποίες θα επέτρεπαν την ταυτοποίηση των κινήσεων. Τέλος, η μέθοδος ανάκτησης δακτυλισμών βασίζεται σε συγκεκριμένη πλατφόρμα (EyesWeb), γεγονός που δυσκολεύει τη γενίκευση της μεθόδου.

2.5.5 Ψηφιακή επεξεργασία ηλεκτρικού σήματος στην a posteriori επεξεργασία

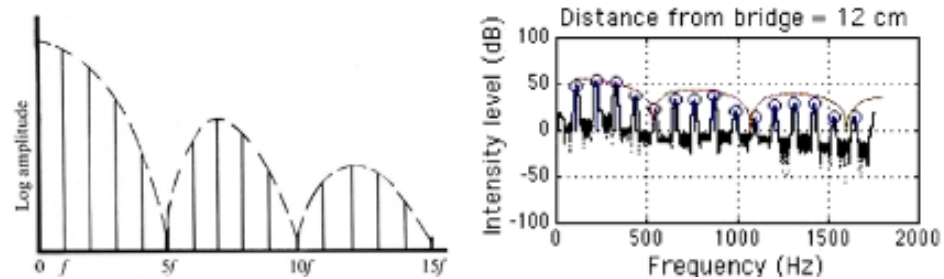
Η τελευταία μέθοδος που θα αναπτυχθεί σε αυτό το κεφάλαιο βασίζεται στην επεξεργασία του ήχου και έχει εφαρμοστεί στην κιθάρα. Ο Traube (2004) και η μέθοδός του στηρίζονται στην παρακολούθηση της χροιάς της κιθάρας. Πιο συγκεκριμένα, υλοποιείται σύγκριση μεταξύ παραμέτρων κάποιου φυσικού μοντέλου και του ηχογραφημένου ήχου της κιθάρας, η οποία οδηγεί στον εντοπισμό κάποιων τεχνικών που αφορούν στην ερμηνεία.

Κύριος στόχος της μεθόδου είναι ο εντοπισμός του σημείου επαφής του δακτύλου με τη χορδή (*plucking point*), αποσκοπώντας στο μεγαλύτερο δυνατό βαθμό ακρίβειας [Traube 2004]. Αφού εντοπιστεί το ακριβές σημείο επαφής πάνω στη χορδή, τότε συγκρίνεται με προκαθορισμένα σημεία επαφής, προφανώς ανάλογα με τη νότα, από όπου εξάγεται τελικά η απόλυτη θέση του δακτυλισμού (*fingering point*).

Το πρώτο βήμα της μεθόδου είναι ο εντοπισμός έναρξης της νότας. Η υλοποίηση του βήματος επιτυγχάνεται με την παρακολούθηση της αύξησης της ενέργειας, αναλύοντας τμήματα της ηχογράφησης με φασματική ανάλυση. Από τη στιγμή που τα σημεία έναρξης νοτών εντοπιστούν, λαμβάνονται δείγματα μεταξύ διαδοχικών ενάρξεων. Από τα δείγματα καθορίζεται η τονικότητα της κάθε νότας με εφαρμογή Fast Fourier Transforms (FFT) σε διάρκεια περίπου ίση με το 1/8 του χρόνου μεταξύ δύο ενάρξεων. Οι τονικότητες καθορίζονται από την εύρεση των μεγίστων σε κάθε φάσμα, χάρη στο ότι μια νότα για την κιθάρα απεικονίζεται από την πρώτη και μεγαλύτερη κορυφή του φάσματος.

Το επόμενο βήμα της μεθόδου είναι η δημιουργία θεωρητικών φασμάτων (*θεωρητικό μοντέλο*) που θα αντιστοιχούν στις ιδανικές θέσεις των χορδών, για κάθε τονικότητα που εντοπίστηκε. Κάθε ένα από τα φάσματα αυτά συγκρίνεται στη συνέχεια με το αντίστοιχο πραγματικό φάσμα κάθε νότας (Σχήμα 2.5.5.1). Το φάσμα που ελαχιστοποιεί το σφάλμα ανάμεσα στα πραγματικά και τα θεωρητικά δεδομένα θα αποτελεί μια πολύ καλή προσέγγιση των πραγματικών νοτών και των *plucking points*. Ο Traube (2004) χρησιμοποιεί μια λογαριθμική έκδοση της συνάρτησης αυτοσυσχέτισης την οποία ονομάζει «λογαριθμική συσχέτιση». Έτσι η λογαριθμική συσχέτιση, προερχόμενη από το θεωρητικό μοντέλο, μαζί με τα πραγματικά δεδομένα παρήγαγαν μια πρώτη προσέγγιση του λόγου (*plucking point/μήκος χορδής*).

Το τρίτο και τελευταίο βήμα στοχεύει στο καθορισμό του *fingering point* και του απόλυτου *plucking point*. Πίνακας με συχνότητες μπορεί να παραχθεί με βάση τις χορδές και τις θέσεις στο τάστο που παρατηρήθηκαν. Με δεδομένο ότι οι τονικότητες της παρτιτούρας είναι γνωστές, οι πιθανές δυνάδες (*χορδή, τάστο*) μπορούν να ανακτηθούν από τον παραπάνω πίνακα. Έτσι, για κάθε (*χορδή, τάστο*) υπολογίζεται ο λόγος (*plucking point/μήκος χορδής*) γνωρίζοντας ότι το μήκος της δονούμενης χορδής είναι αντιστρόφως ανάλογο με τη συχνότητα του ήχου που παράγεται. Ο παραπάνω λόγος (*αντιστοιχούν περισσότεροι από ένας για κάθε νότα*) συγκρίνονται με τον αντίστοιχο λόγο που υπολογίστηκε στο προηγούμενο βήμα. Οι συνδυασμοί (*plucking point, fingering point*), οι οποίοι ελαχιστοποιούν το σφάλμα μεταξύ δύο λόγων, θεωρείται ότι αναπαριστούν τον πραγματικό δακτυλισμό.



Σχήμα 2.5.5.1 1: Θεωρητικό φάσμα, 2: Πραγματικό φάσμα [Traube 2004]

Η μέθοδος υστερεί στην περίπτωση που παιχτούν περισσότερες από μία νότες την ίδια χρονική στιγμή, δεν μπορεί δηλαδή να τις αναγνωρίσει [Traube 2004]. Επίσης, οφείλεται να καταγραφεί ότι δεν υπάρχει μεγάλη ακρίβεια για περιοχές από 3,8 έως 8,3 εκατοστά. Το γεγονός αυτό ενδεχομένως να οφείλεται στην τριβή μεταξύ χορδής και τάστου σε ορισμένα σημεία.

2.5.6 Συμπεράσματα

Ο χώρος της ανάκτησης δακτυλισμών φιλοξενεί τις πολυπληθέστερες έρευνες στο χώρο της ανάκτησης πληροφορίας στη ΜΑ. Θεωρία γράφων, MIDI, ψηφιακή επεξεργασία ηλεκτρικού σήματος και όραση H/Y έχουν να προσφέρουν το δικό τους λιθαράκι στο χώρο.

Όλες αυτές οι αξιόλογες έρευνες υστερούν σε ένα κοινό σημείο, την επέκταση/γενίκευση. Προσφέρουν λύσεις αποκλειστικά για την ανάκτηση δακτυλισμών χωρίς να λαμβάνουν υπόψη τους τη φύση της κίνησης. Αυτό το χαρακτηριστικό, τους στερεί μια πιο ολοκληρωμένη χρήση. Αξιολογούν την κίνηση εκ του αποτελέσματος του οποίου αποφέρει και την επεξεργάζονται αποσπασματικά. Η κίνηση παρόλο που συχνά εμπεριέχει την έννοια του τυχαίου, εντούτοις, τις περισσότερες φορές υπάρχει μια αλληλουχία καταστάσεων που ερμηνεύει το στόχο της κίνησης. Πιο απλά, υπάρχουν καταστάσεις προετοιμασίας, παύσης και παραγωγικές, οι οποίες αποτελούν τα συστατικά της κίνησης. Οι παραπάνω μέθοδοι εστιάζουν αποκλειστικά στην παραγωγική πλευρά της κίνησης, βρίσκονται σε αναμονή έως ότου «εντοπίσουν» το αποτέλεσμα της κίνησης.

2.6 Γενικά συμπεράσματα

Στο κεφάλαιο αυτό πραγματοποιήθηκε μια ανασκόπηση των σημαντικότερων μεθόδων που αφορούν στο χώρο της ανάκτησης πληροφορίας στη ΜΑ. Οι έρευνες αυτές ταξινομήθηκαν σε τρεις μεγάλες κατηγορίες, τον υπολογισμό κινήσεων, την αναγνώριση κινήσεων και την ανάκτηση δακτυλισμών, έχοντας ως κριτήριο το είδος προσέγγισης της κίνησης.

Το πλήθος των ερευνών που διεξήχθησαν στο χώρο της ανάκτησης πληροφορίας βάσει της αναγνώρισης κινήσεων στη ΜΑ αποδεικνύει το διαχρονικό ενδιαφέρον που παρουσιάζει ερευνητικά ο χώρος. Η ποικιλομορφία των μεθόδων πληροφορικής που αναπτύχθηκαν για αυτό το σκοπό επιβεβαιώνουν τον ισχυρισμό ότι ο συνδυασμός των νέων τεχνολογιών με παραδοσιακές επιστήμες μπορεί να αποφέρει θετικότερα αποτελέσματα και να δώσει νέα ώθηση στο χώρο που εφαρμόζεται. Το πιάνο επιβεβαιώνει τη φήμη του ως «βασιλιάς των οργάνων», βάσει του γεγονότος ότι κυριαρχεί ως πεδίο εφαρμογής νέων μεθόδων πληροφορικής.

Κυρίαρχο συμπέρασμα σε ό,τι αφορά τις μεθόδους που αναλύθηκαν είναι ότι αδυνατούν να συνδυάσουν γενίκευση και παραγωγή θετικών αποτελεσμάτων. Ο υπολογισμός κινήσεων, ενώ αποτελεί μια προσέγγιση η οποία διαπραγματεύεται την κίνηση σφαιρικά, εντούτοις δεν αποφέρει απτά αποτελέσματα σχετικά με την αναγνώριση κινήσεων ή μουσικών συμβάντων, τα οποία να μπορούν να χρησιμοποιηθούν στην πράξη για παιδαγωγικούς ή ερμηνευτικούς σκοπούς. Περιορίζεται στο να επιβεβαιώσει επιστημονικά κάποια

συμπεράσματα που τα αντιλαμβάνεται ο καθένας ακόμα και διαισθητικά.

Στη συνέχεια, στο χώρο της αναγνώρισης κινήσεων στη ΜΑ, είναι γεγονός ότι έχουν γίνει σημαντικά βήματα βασιζόμενα στην τεχνολογία των αισθητήρων. Δεν είναι όμως μια απλή, χρηστική, εργονομική και χαμηλού κόστους τεχνολογία, ώστε να μπορέσει να εφαρμοστεί σε μουσικό όργανο. Είναι δύσκολο να έχει ο οποιοσδήποτε πρόσβαση σε τόσο εξειδικευμένο εξοπλισμό αλλά παράλληλα καθιστούν δύσκολη, αν όχι αδύνατη, τη χρήση τους αφού παρεμβάλλονται ανάμεσα στο μουσικό και το όργανο. Από την άλλη μεριά η όραση H/Y, σε συνδυασμό με στατιστικές μεθόδους, έχουν αποδείξει τη λειτουργικότητά τους στο χώρο της αναγνώρισης κινήσεων γενικά. Τέτοια παραδείγματα είναι οι τομείς της ανίχνευσης κίνησης, της βιντεοπαρακολούθησης, της αναγνώρισης προσώπου, της αναγνώρισης κινήσεων στη νοηματική γλώσσα κλπ. Κατά συνέπεια υπόσχονται εξίσου καλά αποτελέσματα και στην αναγνώριση κινήσεων στη ΜΑ. Παράλληλα, η εφαρμογή τους στη ΜΑ, μπορεί να αποφέρει ως λογική συνέπεια και θετικότερα αποτελέσματα και στο χώρο της ανάκτησης δακτυλισμών.

Τέλος, ο χώρος της ανάκτησης δακτυλισμών έχει προσεγγιστεί αποσπασματικά από διάφορους τομείς της πληροφορικής. Παρόλα τα θετικά συμπεράσματα και αποτελέσματα που απέδωσαν αυτές οι μέθοδοι, στερούνται γενίκευσης και αγνοούν τη φυσιολογία της κίνησης, ακόμα και στην περίπτωση που χρησιμοποιήθηκε όραση H/Y.

Ο χώρος της αναγνώρισης κινήσεων στη ΜΑ έχει ανάγκη από γενικές μεθόδους, πρακτικές και με χαμηλό κόστος. Μέθοδοι οι οποίοι στηρίζονται σε πολύ εξειδικευμένο εξοπλισμό, υλικό ή λογισμικό, παραμένουν στη σφαίρα του πρωτότυπου αφού είναι δύσκολη η πρόσβαση σε αυτές από τον καθένα. Χρειάζονται εργαλεία απλά στη χρήση, όπως μια απλή κάμερα για H/Y, λογισμικό το οποίο είναι απλό στη σύλληψή του, μελέτη και αντίληψη της έννοιας «κίνηση» και ταυτόχρονα προσανατολισμός και αξιοποίηση της μουσικής παράδοσης όπως είναι τα κλασσικά μουσικά όργανα.

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Η έρευνα

«Στην αρχή, είναι
τελείως πολύπλοκο,
δεν καταλαβαίνεις
τίποτα.»



THÉOREME 1. - On a l'inégalité

$$\sum_{r \in G} \frac{|K_r|(|K_r| - 1)}{K^2} \log \left(\frac{|K_r| - 1}{K \in A_r} \right) + \frac{K-1}{K^2} \sum_{r \in G} \sum_{s \in K_r} \|\log |s| \alpha_s\|$$

$$\leq \left(1 - \frac{1}{K}\right) \frac{2D}{K} \sum_{i=1}^K h(\alpha_i) + \frac{D}{K} \left(1 + \frac{|G'|}{2D} + \log \frac{K}{2}\right)$$

«Μπορεί να περάσεις ώρες ατελείωτες ψάχνοντας, και να μη βρεις
τίποτα στο τέλος. Σ' αυτήν τη φάση, ο μπαμπάς και η μαμά είναι
χαριτωμένα ανήσυχοι και όταν η μαμά ρωτάει αν ήταν καλή ιδέα
να βάλουμε τον μικρό να κάνει διδακτορικό (αυτός είμαι εγώ), ο
μπαμπάς ανοίγει το στόμα χωρίς να μιλήσει, σηκώνει τους ώμους,
και πάει να διαβάσει την εφημερίδα στο σαλόνι.»



3 Όραση υπολογιστή

3.1 Σύνοψη κεφαλαίου

3.2 Ανθρώπινη όραση

3.3 Κατανόηση ανθρώπινης όρασης

3.4 Ανάπτυξη μεθόδου όρασης Η/Υ

3.5 Η πρόταση του David Marr

3.5.1 Το μοντέλο David Marr στην αναγνώριση δακτυλικών κινήσεων

3.6 Μέθοδοι εντοπισμού χεριού στην εικόνα

3.6.1 Εντοπισμός μέσω τμηματοποίησης

3.6.1.1 Εντοπισμός μέσω υπολογισμού διαφοράς

3.6.1.2 Εντοπισμός μέσω ανίχνευσης δέρματος

3.6.2 Εντοπισμός μέσω εμφάνισης

3.6.3 Συμπεράσματα

3.7 Τμηματοποίηση χεριού

3.7.1 Μαθηματική μορφολογία και ανάκτηση φιγούρας χεριού

3.7.1.1 Διαστολή και συστολή

3.7.1.2 Άνοιγμα και κλείσιμο

3.7.2 Ανάκτηση περιγράμματος χεριού

3.7.2.1 Γραμμικό φιλτράρισμα εικόνας

3.7.2.2 Κλίση μιας εικόνας

3.7.2.3 Γεωμετρική ερμηνεία της κλίσης

3.7.2.4 Μοντέλα περιγράμματος

3.7.2.5 Κριτήρια απόδοσης

3.8 Εντοπισμός άκρων δακτύλων

3.8.1 Αλγόριθμοι προβολής υπογραφών

3.8.2 Αλγόριθμοι κυκλικού μετασχηματισμού Hough

3.8.3 Αλγόριθμοι εξαγωγής γεωμετρικών χαρακτηριστικών

3.8.4 Αλγόριθμοι χρωματικών δεικτών

3.8.5 Συμπεράσματα

3.9 Δυναμική αναγνώριση κινήσεων

3.9.1 Δυναμική Περιτύλιξη του Χρόνου

3.9.2 Νευρωνικά δίκτυα

3.9.3 Κρυφά Μαρκοβιανά Μοντέλα

3.10 Γενικά συμπεράσματα

3.1 Σύνοψη κεφαλαίου

Ο τομέας της όρασης υπολογιστή ασχολείται με την ερμηνεία των αντικειμένων μέσα στην εικόνα της σκηνής. Πρόκειται για μια διαδικασία επεξεργασίας πληροφορίας, η είσοδος της οποίας αποτελείται από μια ή περισσότερες εικόνες.

Η εξέλιξη των τεχνολογικών επιστημών επέτρεψε στην όραση H/Y να στραφεί προς την κατανόηση και την ερμηνεία των αντικειμένων της σκηνής, και ειδικότερα των σκηνών που συμμετέχουν άνθρωποι. Το προσιτό κόστος των καμερών και των καρτών σύλληψης βίντεο επέτρεψε τη σύνδεση μεταξύ των τομέων της «όρασης H/Y» και «επικοινωνίας ανθρώπου-υπολογιστή». Πλέον, οι ερευνητές στον τομέα της επικοινωνίας ανθρώπου-υπολογιστή οραματίζονται νέες διόδους επικοινωνίας, όπου ο χρήστης δε θα περιορίζεται απλώς στο παραδοσιακό δίδυμο ποντίκι-πληκτρολόγιο. Οι διεπαφές γίνονται πλέον αντιληπτικές και «βλέπουν» πλέον τον χρήστη μέσα στο περιβάλλον εργασίας του. Ο όρος «γραφικό περιβάλλον επικοινωνίας» αντικαθίσταται σταδιακά από τον όρο «αντιληπτικό περιβάλλον επικοινωνίας». Από τη μεριά της η όραση H/Y ανακαλύπτει ένα νέο χώρο παρατήρησης, μέσα στον οποίο η δυσκολία είναι η αναγνώριση και η κατανόηση ενός πολύπλοκου αντικειμένου: το χρήστη.

Στο κεφάλαιο αυτό θα γίνει αναλυτική παρουσίαση όλων των μεθόδων και τεχνικών που χρησιμοποιήθηκαν στα πλαίσια ανάπτυξης της μεθόδου της παρούσας διατριβής. Θα γίνει επίσης συνοπτική παρουσίαση και όλων των εναλλακτικών μεθόδων ή τεχνικών που προσφέρονται ως εργαλεία στην όραση H/Y. Το κεφάλαιο ολοκληρώνεται με μια σύντομη περιγραφή μεθόδων δυναμικής αναγνώρισης κινήσεων (*Κρυφά Μαρκοβιανά Μοντέλα κλπ*).

3.2 Ανθρώπινη όραση

Ο κόσμος που μας περιβάλλει έχει μια τρισδιάστατη δομή (3D). Όταν ζητούμε από κάποιον να περιγράψει τι είναι αυτό που βλέπει, δε συναντάει καμία δυσκολία να αναφέρει τα αντικείμενα που τον περιβάλλουν: υπολογιστής, τηλέφωνο, τραπέζι κλπ.

Η πληροφορία που διατίθεται μπροστά στα μάτια του δεν είναι τίποτα άλλο από μια συλλογή από σημεία. Σε κάθε σημείο, ή καλύτερα εικονοστοιχείο για την περίπτωση της εικόνας, περιέχεται μια πληροφορία που δίνει μια ένδειξη σχετικά με την ποσότητα του

φωτός και του χρώματος που προβάλλονται στον αμφιβληστροειδή του ματιού του. Ούτε το τηλέφωνο, ούτε ο υπολογιστής βρίσκονται επάνω στον αμφιβληστροειδή του ματιού. Ο άνθρωπος, οδηγούμενος από την κωδικοποιημένη πληροφορία που περιέχεται στην εικόνα αλλά και από τις γνώσεις που ήδη έχει, καταλήγει στην αντίληψη του χώρου που τον περιβάλλει.

Το τηλέφωνο ή ο υπολογιστής αποτελούν το αποτέλεσμα μιας διαδικασίας ερμηνείας που αποτελεί τμήμα του συστήματος όρασης. Επιπλέον, δεν υπάρχει αντιστοιχία « ένα προς ένα » μεταξύ της πληροφορίας που γίνεται αισθητή (*φως ή χρώμα*) και του τελικού αποτελέσματος (*3Δ αντικείμενα*). Ένα σύστημα όρασης H/Y οφείλει να έχει τροφοδοτηθεί με τις απαιτούμενες γνώσεις ώστε να επιτρέψει μια ξεκάθαρη ερμηνεία της εικόνας.

3.3 Κατανόηση ανθρώπινης όρασης

Ήδη από την αρχαιότητα, πολλές προσπάθειες έχουν γίνει να κατανοηθεί σε βάθος η ανθρώπινη όραση. Πολλοί επιστήμονες και φιλόσοφοι έδειξαν ενδιαφέρον για τη μελέτη της ανατομίας και της λειτουργίας του ανθρώπινου εγκεφάλου. Κατέληξαν σε έναν πολύ σύνθετο μηχανισμό, του οποίου τα μυστικά αδυνατούσαν να εξερευνήσουν πλήρως.

Με τη «γέννηση» των ηλεκτρονικών υπολογιστών, ένας σημαντικός αριθμός μελετών περιστράφηκε γύρω από το πρόβλημα της όρασης, μέσα από μια υπολογιστική σκοπιά: *είναι εφικτό να δημιουργηθεί ένα υπολογιστικό μοντέλο οπτικής αντίληψης;* Ο στόχος, μέσω της απάντησης αυτού του ερωτήματος, δεν ήταν να δοθεί μια πειστική ερμηνεία για το πως λειτουργεί η ανθρώπινη όραση, αλλά η δημιουργία ενός μοντέλου που θα έχει παρόμοιες ιδιότητες με αυτές της ανθρώπινης όρασης. *Το τεχνολογικό αυτό μοντέλο μπορεί να έχει ανάλογη χρησιμότητα με την ανθρώπινη όραση; Μπορεί να αποτελέσει το θεμέλιο μιας νέας τεχνολογίας που θα επιτρέψει στον H/Y να βλέπει;*

Είναι αρκετά νωρίς ώστε να απαντηθούν πλήρως αυτές οι ερωτήσεις και να εξαχθούν ασφαλή συμπεράσματα. Παρά τις φιλότιμες προσπάθειες των επιστημόνων, υπάρχουν λίγες σχετικά πειστικές προτεινόμενες μέθοδοι όρασης H/Y. Πάντως, σε κάθε περίπτωση, θα πρέπει να:

- αναπτύσσεται μια μέθοδος όρασης H/Y που να έχει μακροπρόθεσμους στόχους,

- στοχεύεται η επίλυση συγκεκριμένων επιστημονικών προβλημάτων. Τα επιμέρους εξαγόμενα αποτελέσματα επιτρέπουν γενικότερα την επιβεβαίωση, ή όχι, της ίδιας της μεθόδου.

3.4 Ανάπτυξη μεθόδου όρασης H/Y

Η ανάπτυξη μιας επιστημονικής μεθόδου όρασης H/Y προϋποθέτει τρία στάδια:

1. την ανάπτυξη της μεθόδου και τον προσδιορισμό των βασικών εννοιών στις οποίες στηρίζεται,
2. την απόδοση των εννοιών αυτών με μαθηματική μορφή,
3. την υλοποίηση ενός συνόλου πειραμάτων που θα επιτρέπουν την επαλήθευση της μεθόδου.

Δύο στρατηγικές ανάπτυξης μεθόδων όρασης H/Y χρησιμοποιούνται γενικότερα, α) η « *κάτω προς τα πάνω* » και β) η « *πάνω προς τα κάτω* ». Στην πρώτη περίπτωση, η στρατηγική υποδεικνύει τη δημιουργία μιας αφαιρετικής αναπαράστασης του αντικειμένου (για παράδειγμα ένα σύνολο 3D γεωμετρικών αρχεγόνων) μέσα από τις πληροφορίες που λαμβάνονται από τα αισθητήρια. Στη δεύτερη περίπτωση, η στρατηγική υποδεικνύει την περιγραφή ήδη γνωστών αντικειμένων μέσα από αρχέγονα που εξάγονται από την εικόνα.

Οι *a priori* γνώσεις γύρω από ένα ορατό αντικείμενο μπορεί να είναι τριών ειδών, α) φυσικές, β) γεωμετρικές ή γ) σημασιολογικές. Οι νόμοι της φυσικής καθορίζουν κάποιους περιορισμούς στα σήματα φωτεινότητας που εκπέμπονται από μια πηγή, μεταφέρονται στο χώρο και προβάλλονται στην εικόνα. Οι μορφές των αντικειμένων (πχ επιφάνειες ή ακμές) και η γεωμετρία των μορφών αυτών, καθορίζουν πολύ αυστηρούς περιορισμούς αναφορικά τη δομή που ακολουθεί η διάταξη των μορφών στην εικόνα. Σε ακόμα πιο υψηλό επίπεδο, ένα αντικείμενο μπορεί να περιγραφεί μέσω μιας συνάρτησης που έχει ως πεδίο ορισμού ένα σύνολο συμβόλων. Αυτή η συνάρτηση δεν είναι εύκολο να βρεθεί και να ανακτηθεί μέσω μετρήσεων στην εικόνα. Για παράδειγμα, η λέξη « δάκτυλο » περιγράφει μια συγκεκριμένη κλάση αντικειμένων του φυσικού κόσμου (σημασιολογική προσέγγιση). Παρόλα αυτά, υπάρχει μια μεγάλη ποικιλία δακτύλων

ανάλογα με το χρώμα του δέρματος, το ίδιο το δάκτυλο, το χέρι στο οποίο ανήκει κλπ. Ποια είναι οι κοινές ιδιότητες που χαρακτηρίζουν τα δάκτυλα μέσα σε μια εικόνα;

Το επόμενο στάδιο αναφέρεται στην απόδοση των παραπάνω στρατηγικών με τη μορφή μαθηματικού φορμαλισμού, καθώς και την ανάπτυξη των αντίστοιχων αλγορίθμων. Η απόδοση των αλγορίθμων αυτών πρέπει να ευθυγραμμίζεται με τις ποιοτικές απαιτήσεις ενός συστήματος όρασης H/Y: η οπτική αναγνώριση των αντικειμένων οφείλει να είναι γρήγορη και με καλά αποτελέσματα.

3.5 Η πρόταση του David Marr

Στα τέλη της δεκαετίας του 70, ο David Marr πρότεινε ένα μοντέλο επεξεργασίας και αναπαράστασης οπτικής πληροφορίας [Willshaw 1990]. Οι βασικές άξονες της πρότασής του:

- η εξαγωγή των χαρακτηριστικών ενός αντικειμένου μέσα από εικόνες οδηγεί στην περιγραφή του με χρήση 2Δ ιδιοχαρακτηριστικών. Πρόκειται για το « πρώτο σκαρίφημα » (*première ébauche – primal sketch*), δηλαδή το πρώτο επίπεδο αναπαράστασης.
- Το πρώτο σκαρίφημα αποτελεί την είσοδο ορισμένων διαδικασιών, λιγότερο ή περισσότερο ανεξάρτητων μεταξύ τους. Οι διαδικασίες αυτές καλούνται να υπολογίσουν ένα πλήθος από 3Δ τοπικές ιδιότητες (*τοπική ιδιότητα = η ιδιότητα που σχετίζεται με συγκεκριμένο πλάνο και εντάσσεται σε μια γενικότερη σκηνή*). Πρόκειται δηλαδή για μια αναπαράσταση με σημείο αναφοράς τον παρατηρητή, η οποία ονομάζεται « 2,5Δ σκαρίφημα » (*ébauche 2.5D – 2.5D sketch*). Οι διαδικασίες που υλοποιούνται στα πλαίσια του 2,5Δ σκαριφήματος δύνανται να εφαρμοστούν σε:
 - Ακολουθία εικόνων (*ανάλυση κίνησης*)
 - Ζεύγος εικόνων (*στερεοσκοπία*)
 - Μεμονωμένη εικόνα

Στην περίπτωση της μεμονωμένης εικόνας εφαρμόζονται διαδικασίες που χρησιμοποιούν γνώση α) γεωμετρική (*ανάλυση περιγραμμάτων*), β) γεωμετρική και στατιστική (*ανάλυση υφής*), γ) φωτομετρική (*ανάλυση υφής*) και δ) χρωματομετρική (*ανάλυση αντανάκλασεων*).

- Το 2,5Δ σκαρίφημα αποτελεί το συνδυαστικό κρίκο με τη 3Δ γνώση. Στόχος είναι η περιγραφή της σκηνής με βάση χαρακτηριστικά των αντικειμένων που επιθυμείται να περιγραφούν, ή με χαρακτηριστικά που αφορούν στη διασύνδεση των αντικειμένων αυτών. Πρόκειται δηλαδή, για μια αναπαράσταση με σημείο αναφοράς τη σκηνή, όπου η περιγραφή των αντικειμένων είναι πλέον ανεξάρτητη της θέσης του παρατηρητή.

3.5.1 Το μοντέλο David Marr στην αναγνώριση δακτυλικών κινήσεων

Στην πράξη, η πρόταση του David Marr για την αναγνώριση χειρονομιών μέσω όρασης H/Y μεταφράζεται σε τρία επίπεδα επεξεργασίας: α) εντοπισμός χειριού και τμηματοποίησή του, β) την εξαγωγή χαρακτηριστικών και γ) την αναγνώριση.

Η τμηματοποίηση εικόνων είναι η βάση κάθε συστήματος όρασης H/Y. Η ποικιλομορφία των εικόνων, η εν γένει δυσκολία της τμηματοποίησης της εικόνας αλλά και η εξέλιξη της υπολογιστικής ισχύος των H/Y οδήγησαν στην ανάπτυξη πληθώρας σχετικών αλγορίθμων. Όποια κι αν είναι η προέλευση μιας εικόνας, η ίδια η εικόνα αποτελεί μια αναπαράσταση ενός κόσμου αποτελούμενου από οντότητες: αντικείμενα του παρασκήνιου, επιφάνειες, μέλη ανθρώπινου σώματος κλπ. Ο στόχος κάθε τμηματοποίησης είναι η προετοιμασία του συστήματος για την εξαγωγή ιδιοχαρακτηριστικών που προσδιορίζουν τα αντικείμενα. Η τμηματοποίηση αφορά α) στη μαθηματική μορφολογία, β) στο φιλτράρισμα, γ) στην ανάκτηση περιγράμματος και δ) στην εξαγωγή περιοχών.

Η εξαγωγή των χαρακτηριστικών που αφορούν στα αντικείμενα μιας εικόνας στοχεύει στην περιγραφή της περιοχής ενδιαφέροντος. Δύο βασικές τάσεις εξαγωγής χαρακτηριστικών καταγράφονται γενικότερα. Η πρώτη τάση αφορά στις « *εικόνες της ενέργειας της κίνησης* » (*images d'énergie de mouvement – motion energy image*), όπου η εξαγωγή χαρακτηριστικών είναι μια χωρο-χρονική διαδικασία που εφαρμόζεται πάνω σε ακολουθία εικόνων. Η δεύτερη τάση αφορά στις « *εικόνες του ιστορικού της κίνησης* » (*images de l'historique du mouvement – motion history image*), κάνοντας χρήση κάποιων πρότυπων στατικών εικόνων. Η δυναμική χροιά της κίνησης λαμβάνεται υπόψη μόνο κατά τη διαδικασία αναγνώρισης.

Το στάδιο της αναγνώρισης αφορά στη σύγκριση 2Δ ή 3Δ οπτικών δεικτών, με τους δείκτες των προς αναγνώριση αντικειμένων. Οι μέθοδοι αναγνώρισης συνήθως σχετίζονται με

μεθόδους εντοπισμού αντικειμένων, οι οποίες θα αναπτυχθούν παρακάτω στον παρόν κεφάλαιο.

3.6 Μέθοδοι εντοπισμού χεριού σε εικόνα

Ο εντοπισμός του χεριού στην εικόνα αποτελεί το πρώτο στάδιο της αναγνώρισης χειρονομιών μέσω όρασης Η/Υ. Προηγείται της τμηματοποίησης του χεριού και της εξαγωγής των χαρακτηριστικών. Δύο κατηγορίες αλγορίθμων χρησιμοποιούνται προκειμένου να εντοπιστεί το ανθρώπινο χέρι πάνω σε μια εικόνα: α) ο *εντοπισμός μέσω τμηματοποίησης* και β) ο *εντοπισμός μέσω εμφάνισης*. Στην ενότητα αυτή θα παρουσιαστούν συνοπτικά όλες οι μέθοδοι και οι τεχνικές τους. Θα παρουσιαστεί αναλυτικά η τεχνική εντοπισμού μέσω *ανίχνευσης δέρματος*, η οποία επιλέχτηκε για την ανάπτυξη της μεθόδου και του συστήματος αναγνώρισης δακτυλικών κινήσεων της παρούσας έρευνας.

3.6.1 Εντοπισμός μέσω τμηματοποίησης

Ο εντοπισμός του χεριού μέσω τμηματοποίησης επιτρέπει το διαχωρισμό των εικονοστοιχείων που ανήκουν στο χέρι, από αυτά που δεν ανήκουν. Δύο τεχνικές έχουν αναπτυχθεί. Η πρώτη βασίζεται στην *κίνηση* για να εντοπίσει το χέρι πάνω στην εικόνα, ενώ η δεύτερη βασίζεται στο *ιδιαιτερο χρώμα* του ανθρώπινου δέρματος.

3.6.1.1 Εντοπισμός μέσω υπολογισμού διαφοράς εικόνων

Η τεχνική του εντοπισμού χεριού χρησιμοποιείται ευρέως από τα συστήματα όρασης Η/Υ. Στοχεύει στον εντοπισμό της διαφοράς μεταξύ ενός εικονοστοιχείου σε μια εικόνα και του αντίστοιχού του σε μια άλλη εικόνα. Η διαφορά αυτή για δύο εικονοστοιχεία με συντεταγμένες (x,y) , ορίζεται ως εξής :

$$D(x,y) = |I(x,y) - I'(x,y)|, \forall x,y \quad (3.6.1.1.1)$$

Όπου:

- $D(x,y)$, η διαφορά των δύο εικονοστοιχείων,
- $I(x,y)$, η τιμή του εικονοστοιχείου για την εικόνα I ,
- $I'(x,y)$, η τιμή του εικονοστοιχείου για την εικόνα I' .

Πρόκειται για μια απλή τεχνική που επιτρέπει τη γρήγορη εκτίμηση της θέσης του κινούμενου αντικειμένου. Παρόλα αυτά η συγκεκριμένη τεχνική παρουσιάζει κάποιους περιορισμούς σχετικά με το περιβάλλον που λαμβάνεται το βίντεο [Wellner 1991a; Wellner 1991b; Newman 1992; Chomat 2000]:

1. η κάμερα οφείλει να είναι σταθεροποιημένη, ώστε να μη θεωρηθεί το σύνολο της εικόνας ως κινούμενο αντικείμενο. Ο περιορισμός αυτός δεν αποτελεί εμπόδιο για συστήματα που στοχεύουν σε συγκεκριμένη περιοχή της σκηνής,
2. οι πηγές φωτεινότητας οφείλουν να είναι τοποθετημένες σε συγκεκριμένο σημείο και η τιμή της φωτεινότητας σταθερή. Η αλλαγή στη φωτεινότητα σε κάποιο σημείο, μπορεί να θεωρηθεί ως κίνηση κάποιου αντικειμένου. Η αλλαγή στη φωτεινότητα μπορεί να προκληθεί από το άναμμα μιας λάμπας, το πέρασμα κάποιου ανθρώπου προκαλώντας σκίαση, ή ακόμα και το πέρασμα κάποιου σύννεφου στο παράθυρο,
3. όταν η κίνηση κάποιου αντικειμένου είναι πολύ μικρή, τότε και η περιοχή της εικόνας που ανιχνεύεται κίνηση είναι πολύ μικρή. Αυτό συνεπάγεται την ανίχνευση μόνο ενός μικρού τμήματος από το κινούμενο αντικείμενο,
4. αν ταυτόχρονα δύο αντικείμενα είναι σε κίνηση, απαιτείται επιπρόσθετη διαδικασία διάκρισης των δύο αντικειμένων. Στην περίπτωση που τα δύο αντικείμενα είναι αρκετά απομακρυσμένα, ο υπολογισμός της περιοχής στην οποία κινούνται επαρκεί. Αν τα δύο αντικείμενα είναι πολύ κοντά το ένα στο άλλο λαμβάνονται υπόψη παράμετροι όπως η υφή ή το χρώμα τους.

3.6.1.2 Εντοπισμός μέσω ανίχνευσης δέρματος

Η ανίχνευση δέρματος είναι μια κλασική πλέον μέθοδος στον τομέα του εντοπισμού προσώπων ή χειρών. Βασίζεται στη διάκριση του χρώματος του ανθρώπινου δέρματος από το χρώμα των δεδομένων του παρασκηνίου. Το αντικείμενο εντοπίζεται μέσω της αναζήτησης των εικονοστοιχείων που έχουν χρώμα που πλησιάζει στο χρώμα του αντικειμένου. Αναζητούμε συνεπώς τα εικονοστοιχεία ε , για τα οποία η πιθανότητα του χρώματος $\chi(\varepsilon)$ μεγιστοποιείται, γνωρίζοντας ότι το χρώμα του αντικειμένου έχει μοντελοποιηθεί από το μοντέλο $M_{\chi\theta}$. Η δεσμευμένη πιθανότητα θα σημειώνεται στο εξής ως:

$$P(\chi\rho(\varepsilon)|M_{\chi\rho}) \quad (3.6.1.2.1)$$

Το χρώμα των εικονοστοιχείων μιας εικόνας μοντελοποιείται μέσω μιας τριάδας αριθμών, ενός αριθμού για κάθε κανάλι χρώματος: κόκκινο, πράσινο, μπλε. Για παράδειγμα οι δύο τριάδες <255 0 0> και <100 0 0> αναπαριστούν το κόκκινο χρώμα με διαφορετική φωτεινότητα. Προκειμένου να αποσυμφορηθεί το σύστημα από προβλήματα μεταβολής φωτεινότητας προτείνεται η κανονικοποίηση των τιμών της τριάδας του χρώματος από τη φωτεινότητα. Με τη διαδικασία αυτή διατηρείται η πληροφορία που αναφέρεται στη χρωματικότητα του χρώματος:

$$\varepsilon_{\kappa} = \frac{\varepsilon_{\kappa}}{\varepsilon_{\Phi}} \quad \varepsilon_{\pi} = \frac{\varepsilon_{\pi}}{\varepsilon_{\Phi}} \quad \varepsilon_{\mu} = \frac{\varepsilon_{\mu}}{\varepsilon_{\Phi}} \quad (3.6.1.2.2)$$

Όπου:

- ε_{κ} , ε_{π} , ε_{μ} οι συνιστώσες του εικονοστοιχείου για το κόκκινο, το πράσινο και το μπλε,
- ε_{κ} , ε_{π} , ε_{μ} , οι κανονικοποιημένες συνιστώσες του εικονοστοιχείου ως προς τη φωτεινότητα ε_{Φ} , για την οποία ισχύει: $\varepsilon_{\Phi} = \varepsilon_{\kappa} + \varepsilon_{\pi} + \varepsilon_{\mu}$.

Πρέπει να σημειωθεί ότι οι τρεις κανονικοποιημένες συνιστώσες είναι γραμμικώς εξαρτημένες, δηλαδή ισχύει:

$$\varepsilon_{\kappa} + \varepsilon_{\pi} + \varepsilon_{\mu} = \frac{\varepsilon_{\kappa} + \varepsilon_{\pi} + \varepsilon_{\mu}}{\varepsilon_{\Phi}} = 1 \quad (3.6.1.2.3)$$

Κατά συνέπεια, δύο συνιστώσες αρκούν για να αναπαραστήσουν τη χρωματικότητα ενός εικονοστοιχείου. Στη συνέχεια, η χρωματικότητα του εικονοστοιχείου ε θα συμβολίζεται από το ζεύγος κανονικοποιημένων συνιστωσών χρωματικότητας:

$$\chi\rho(\varepsilon) = \langle \varepsilon_\kappa, \varepsilon_\pi \rangle \quad (3.6.1.2.4)$$

Το μοντέλο της χρωματικότητας $M_{\chi\rho}$ του αντικειμένου (χέρι), το οποίο χρησιμοποιήθηκε στη σχέση 3.5.1.2.1, πρέπει να καθοριστεί επακριβώς. Το $M_{\chi\rho}$ μπορεί να υπολογιστεί μέσω ενός συνόλου εικονοστοιχείων εκπαίδευσης E , προερχόμενα από μια βιβλιοθήκη εικόνων B . Η χρωματικότητα του δέρματος μπορεί να αναπαρασταθεί από δύο μοντέλα: α) ένα ιστόγραμμα ή β) ένα Γκαουσιανό μοντέλο.

1. Ιστόγραμμα χρωματικότητας [Swain 1991]

Οι Swain και Ballard (1991) έδειξαν ότι το ιστόγραμμα της χρωματικότητας είναι ένα καλό μοντέλο για την αναγνώριση χρωματικών περιοχών. Πειραματίστηκαν με διαφορετικούς τύπους ιστογραμμάτων, μεταξύ των οποίων και αυτοί που δημιουργούνται από τις κανονικοποιημένες συνιστώσες του κόκκινου και του πράσινου. Ένα σημείο, στο επίπεδο του ιστογράμματος I_E , με συντεταγμένες (κ, π) , δίνει το πλήθος των εικονοστοιχείων του συνόλου E , τα οποία έχουν συνιστώσα χρωματικότητας για το κόκκινο κ και για το πράσινο π . Αυτό το ιστόγραμμα μπορεί να καθορίσει την πιθανότητα της σχέσης 3.5.1.2.1:

$$P(\chi\rho(\varepsilon) | M_{\chi\rho}) = \frac{1}{\theta_E} I(\chi\rho(\varepsilon)) = \frac{1}{\theta_E} I(\varepsilon_\kappa, \varepsilon_\pi) \quad (3.6.1.2.5)$$

Όπου:

- θ_E , το συνολικό πλήθος των εικονοστοιχείων του συνόλου E .

Ο αλγόριθμος υπολογισμού του ιστογράμματος I_E για το δείγμα των εικονοστοιχείων του συνόλου E δίνεται από τον παρακάτω αλγόριθμο :

Αλγόριθμος 3.6.1.2.1: Αλγόριθμος υπολογισμού του ιστογράμματος χρωματικότητας I_E , με βάση ένα σύνολο εικονοστοιχείων E .

! Αρχικοποίηση:

ΓΙΑ κάθε (κ, π)

$$I(\kappa, \pi) \leftarrow 0$$

ΤΕΛΟΣ_ΕΠΑΝΑΛΗΨΗΣ

! Υπολογισμός του I_E :

ΓΙΑ κάθε εικονοστοιχείο ε του E

$$I(\varepsilon_\kappa, \varepsilon_\pi) \leftarrow I(\varepsilon_\kappa, \varepsilon_\pi) + 1$$

ΤΕΛΟΣ_ΕΠΑΝΑΛΗΨΗΣ

2. Γκαουσσισιανό μοντέλο

Το μοντέλο χρωματικότητας $M_{\chi\theta}$ αναπαρίσταται μέσω μιας Γκαουσσισιανής ΣΠΠ:

$$f(\vec{x}) = \frac{1}{\sqrt{2\pi}|\Sigma|} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})^T} \quad (3.6.1.2.6)$$

Όπου:

- $\vec{x}$, το τυχαία μεταβαλλόμενο διάνυσμα δύο διαστάσεων που αντιπροσωπεύει το ζεύγος της χρωματικότητας,
- $\vec{\mu}$, η μέση τιμή,
- Σ , ο πίνακας συµμεταβλητότητας.

Τα $\vec{x}$, $\vec{\mu}$ και Σ είναι οι παράμετροι του Γκαουσσισιανού μοντέλου. Οι παράμετροι αυτές αντιστοιχούν στις κεντρικές ροπές, τάξης 0 και 1, της κατανομής της χρωματικότητας στο σύνολο E . Οι ροπές αυτές υπολογίζονται με βάση τους παρακάτω τύπους:

$$\vec{\mu} = \begin{bmatrix} \mu_\kappa \\ \mu_\pi \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_\kappa^2 & \sigma_{\kappa\pi} \\ \sigma_{\kappa\pi} & \sigma_\pi^2 \end{bmatrix} \quad (3.6.1.2.7)$$

και

$$\mu_\alpha = \frac{1}{N} \sum_{\varepsilon \in E} \varepsilon_\alpha \quad \sigma_{\alpha\beta} = \frac{1}{N} \left(\sum_{\varepsilon \in E} \varepsilon_\alpha \varepsilon_\beta - \mu_\alpha \mu_\beta \right) \quad (3.6.1.2.8)$$

Όπου:

- Τα α και β παίρνουν τις τιμές των κ και π αντίστοιχα.

Το πρώτο πλεονέκτημα της τεχνικής αυτής, σε σχέση με την τεχνική υπολογισμού διαφοράς εικόνων, είναι ότι σχεδόν απαλείφει τους περιορισμούς σε σχέση με τον περιβάλλοντα χώρο:

1. Η κάμερα μπορεί να βρίσκεται σε κίνηση,
2. Η συνιστώσα χρωματικότητας επηρεάζεται πολύ λιγότερο από μεταβολές στη φωτεινότητα από ότι το χρώμα γενικότερα.

Ως εκ τούτου, αυτή η τεχνική απαιτεί τη δημιουργία ενός μοντέλου με βάση ένα σύνολο εικονοστοιχείων δέρματος που αντλήθηκαν από μια βιβλιοθήκη εικόνων. Συν τοις άλλοις, στην περίπτωση αναγνώρισης και ταυτοποίησης του ματιού για παράδειγμα, προσφέρει τη δυνατότητα εκτίμησης θέσης των ματιών μέσω ανοιγοκλείσματος των βλεφάρων. Πρακτικά, αυτό γίνεται μέσω της ανίχνευσης περιοχής δέρματος κατά το κλείσιμο των βλεφάρων.

Ένα επιπλέον πλεονέκτημα της τεχνικής είναι ότι το μοντέλο μπορεί να εκπαιδευτεί με βάση περισσότερους τύπους δέρματος υπό διαφορετικές συνθήκες φωτεινότητας [Bérard 2000]. Επιτρέπει τέλος, μέσω υπολογισμού της αρνητικής απόχρωσης του δέρματος, να αποκλείονται εντελώς οι μη δερματικές περιοχές του παρασκηνίου [Störing 1999].

3.6.2 Εντοπισμός μέσω εμφάνισης

Ο εντοπισμός μέσω εμφάνισης, ή πιο γενικά η όραση μέσω εμφάνισης [Verdière 1999 ; Chomat 2000 ; Schiele 1997], βασίζεται στον τρόπο που εμφανίζεται το αντικείμενο στο πλάνο. Με τον όρο « εμφάνιση » ενός αντικειμένου θεωρείται το σύνολο των δυνατών αναπαραστάσεων (οποιαδήποτε φωτεινότητα ή γωνία λήψης) στην εικόνα.

Τα προβλήματα που μπορεί να παρουσιάσει η συγκεκριμένη τεχνική χωρίζονται σε δύο κατηγορίες:

1. **Πρόβλημα αλλαγής φωτισμού και κλίμακας αντικειμένου:** Θεωρητικά, τα προς εντόπιση αντικείμενα οφείλουν να βρίσκονται σε κατάλληλες συνθήκες τόσο ως προς α) τη *φωτεινότητα*, όσο και ως προς β) την *κλίμακα μεγέθους* του αντικειμένου στην εικόνα. Ο περιορισμός σ' αυτήν την περίπτωση έγκειται στο γεγονός ότι σε μια πιθανή μεγάλη μεταβολή (πχ λόγω περιστροφής του αντικειμένου) μιας εκ των δύο παραμέτρων, η οποία δεν έχει προκαθοριστεί εκ των προτέρων, παρατηρούνται σημαντικές ατέλειες στον εντοπισμό [Darrell 1992].
2. **Πρόβλημα παραμόρφωσης:** Συχνά, το κινούμενο χέρι στην εικόνα εμφανίζεται παραμορφωμένο διότι εξ' αρχής ανιχνεύτηκε ατελώς η περιοχή του χεριού. Καθώς το χέρι κινείται η παραμόρφωση πιθανόν να επιδεινωθεί. Αυτό το πρόβλημα εμφανίζεται εξαιτίας του γεγονότος ότι η ανίχνευση βασίζεται στη σύγκριση της τρέχουσας εικόνας με μια εικόνα αναφοράς που ονομάζεται *μοτίβο*. Το πρόβλημα αυτό μπορεί να παρακαμφθεί τμηματοποιώντας την εικόνα του χεριού σε επιμέρους συνιστώσες. Μέσω αυτή της διαδικασίας ορίζεται ένα σύνολο μετασχηματισμών των δερματικών εικονοστοιχείων καθώς και μια συνάρτηση σφάλματος ανίχνευσης. Η συνάρτηση σφάλματος επιστρέφει μια αριθμητική τιμή απόκλισης της τρέχουσας εικόνας σε σχέση με το μοτίβο και συνδράμει στη βελτίωση του συνόλου των επιτρεπτών μετασχηματισμών [Black 1996].

3.6.3 Συμπεράσματα

Στην παρούσα ενότητα παρουσιάστηκαν οι μέθοδοι εντοπισμού χεριού μέσω τμηματοποίησης και μέσω εμφάνισης. Περιγράφηκε αναλυτικά η τεχνική εντοπισμού χεριού μέσω ανίχνευσης δέρματος, ως η χρησιμοποιούμενη τεχνική από το σύστημα αναγνώρισης δακτυλικών κινήσεων.

Περιορισμοί σχετικά με τις μεταβολές στο φωτισμό του πλάνου κρίνονται απαγορευτικοί για το σύστημα αναγνώρισης κινήσεων (διαφορά εικόνων – εντοπισμός μέσω εμφάνισης). Ο εντοπισμός του χεριού καθώς αυτό κινείται με μεγάλη ταχύτητα είναι καθοριστική παράμετρος, αφού τα δάκτυλα ενός πιανίστα μπορούν να αναπτύξουν μεγάλη ταχύτητα. Επίσης απαγορευτικοί κρίνονται οι περιορισμοί που αφορούν την ταυτόχρονη κίνηση δύο χεριών στο ίδιο πλάνο, για προφανείς λόγους. Αντίθετα, η αναγνώριση δακτυλικών κινήσεων στη μουσική αλληλεπίδραση, αν και δεν απαιτεί εντοπισμό χεριού μέσω κινούμενης κάμερας, δυνατότητα που προσφέρει η τε-

χνική της ανίχνευσης δέρματος, εντούτοις η απαλλαγή της διαδικασίας από μεταβολές φωτεινότητας είναι σημαντικό πλεονέκτημα της τεχνικής αυτής. Τέλος, προβλήματα παραμόρφωσης και κλίμακας μεγέθους του χεριού στην εικόνα (εντοπισμός μέσω εμφάνισης) υποχρεώνουν τη χρήση επιπρόσθετων αλγορίθμων προκειμένου να επιτευχθεί αποτέλεσμα ισάξιο με την ανίχνευση δέρματος, αυξάνοντας κατά πολύ την πολυπλοκότητα.

Συνεπώς, η ταχύτητα κίνησης των δακτύλων, η ταυτόχρονη κίνηση των δύο χεριών στο ίδιο πλάνο και η σχετική απαλλαγή από διαφοροποιήσεις φωτεινότητας πρέπει να ικανοποιούνται από το σύστημα σε κάθε περίπτωση. Για όλους τους παραπάνω λόγους προτιμήθηκε η μέθοδος του εντοπισμού μέσω τμηματοποίησης και ειδικότερα η τεχνική της ανίχνευσης δέρματος.

3.7 Τμηματοποίηση χεριού

Η τμηματοποίηση της εικόνας είναι ο ακρογωνιαίος λίθος της όρασης H/Y. Η ποικιλομορφία των εικόνων, η εν γένει δυσκολία της τμηματοποίησης, οι διάφορες ερευνητικές σχολές, καθώς και η αύξηση της υπολογιστικής ισχύος των H/Y οδήγησαν στην ανάπτυξη μιας πληθώρας σχετικών αλγορίθμων. Σύμφωνα με τον Παπαμάρκο (2000):

« Τμηματοποίηση εικόνας είναι η διαδικασία με την οποία διαχωρίζεται η εικόνα σε κατάλληλες περιοχές ή αντικείμενα. »

Αποτελεί προϋπόθεση για την περαιτέρω επεξεργασία της εικόνας, ο διαχωρισμός της σε βασικές συνιστώσες. Τα κριτήρια ποιότητας τμηματοποίησης μιας εικόνας προσδιορίζονται σε τέσσερις βασικούς άξονες: α) Οι περιοχές που διαχωρίζονται οφείλουν να είναι ομοιόμορφες και ομογενείς σε σχέση με κάποιο χαρακτηριστικό, όπως για παράδειγμα τα επίπεδα του γκρι, β) το εσωτερικό των περιοχών πρέπει να είναι απλό, χωρίς για παράδειγμα πολλές μικρές οπές, γ) γειτονικές περιοχές πρέπει να έχουν σαφώς διαφορετικές τιμές για το χαρακτηριστικό με βάση το οποίο έγινε διαχωρισμός και δ) τα όρια των περιοχών πρέπει να είναι απλά, όχι απότομα αλλά ακριβή [Papamarkos 2000].

Ένα ισχυρό εργαλείο με βάση το οποίο μπορεί να επιτευχθεί τμηματοποίηση δυαδικής εικόνας κυρίως, είναι η μαθηματική μορφολογία. Στη συνέχεια θα αναπτυχθούν οι μέθοδοι που χρησιμοποι-

ήθηκαν στα πλαίσια της παρούσας έρευνας για τη μείωση θορύβου και την ανάκτηση φιγούρας και περιγράμματος του χεριού.

3.7.1 Μαθηματική μορφολογία και ανάκτηση φιγούρας χεριού

Η μαθηματική μορφολογία είναι ένα μαθηματικό εργαλείο που επιτρέπει κατά πρώτο λόγο την εξερεύνηση της γεωμετρίας των αντικειμένων στην εικόνα. Κατά δεύτερο λόγο, μέσω της ανάπτυξης κατάλληλων τεχνικών, μπορεί να επεκταθεί το πεδίο εφαρμογών της μαθηματικής μορφολογίας, όπως για παράδειγμα στον τομέα της βελτίωσης και αποκατάστασης εικόνας, αλλά και στη μείωση θορύβου.

Συνεπώς, υπό το πρίσμα της αναγνώρισης αντικειμένων, η επεξεργασία της εικόνας στοχεύει στην εξαγωγή γεωμετρικών πληροφοριών (*θέση, περίμετρος, επιφάνεια, προσανατολισμός κλπ*), επιτρέποντας τη διάκριση των αντικειμένων στην εικόνα. Για το λόγο αυτό επιστρατεύονται οι *μορφολογικοί τελεστές*, οι οποίοι εφαρμόζουν μορφολογικούς μετασχηματισμούς. Στην πράξη, τα αντικείμενα μιας σκηνής συνθέτονται με δύο μορφές:

1. ως **ένωση** συνόλων (*επικάλυψη*):

$$X = X_1 \cup X_2 \quad (3.7.1.1)$$

2. ως **τομή** συνόλων (*συρρίκνωση*) :

$$X = X_2 - X_1 = X_1^c \cap X_2 \quad (3.7.1.2)$$

Ο μορφολογικός τελεστής Ψ μπορεί να εφαρμόσει τους παρακάτω μορφολογικούς μετασχηματισμούς πάνω σε μια δυαδική εικόνα:

$$1. \quad \Psi_\delta(X_1 \cup X_2) = \Psi_\delta(X_1) \cup \Psi_\delta(X_2)$$

$$2. \quad \Psi_\varepsilon(X_1 - X_2) = \Psi_\varepsilon(X_1) \cap \Psi_\varepsilon(X_2)$$

Για το υπόλοιπο της διδακτορικής διατριβής, ο πρώτος μετασχηματισμός θα ονομάζεται *μορφολογική διαστολή* (ή απλά *διαστολή*) και ο δεύτερος *μορφολογική συστολή* (ή απλά *συστολή*). Αυτοί οι δύο με-

τασχηματισμοί αποτελούν τη βάση της μαθηματικής μορφολογίας. Ξεκινώντας από αυτή τη βάση θα κατασκευαστούν νέοι μορφολογικοί τελεστές όπως το *μορφολογικό άνοιγμα* και το *μορφολογικό κλείσιμο*.

Στα παρακάτω εδάφια θα οριστεί η έννοια της διαστολής και της συστολής καθώς και του ανοίγματος αλλά και του κλεισίματος.

3.7.1.1 Διαστολή και συστολή

Ορισμός 3.7.1.1.1 (Διαστολή-Σχήμα 3.7.1.1.1) Η διαστολή του A από το B θα σημειώνεται $A \oplus B$ και θα είναι το παρακάτω σύνολο :

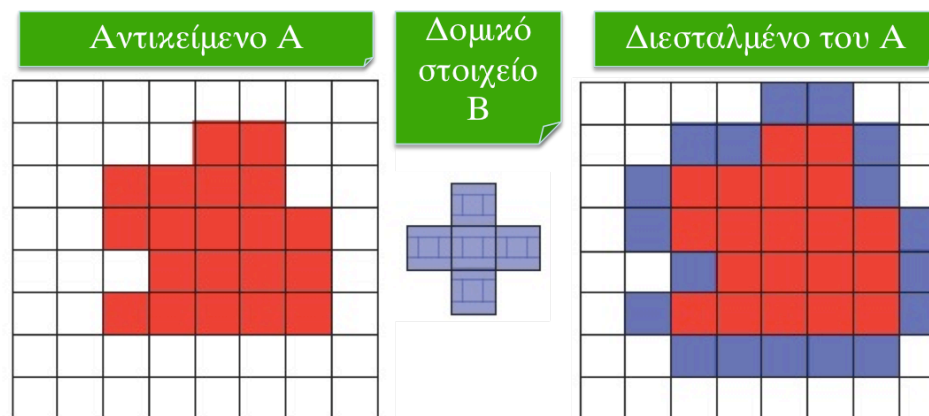
$$A \oplus B = \{x, (\hat{B})_x \cap A \neq \emptyset\} \quad (3.7.1.1.1)$$

Ορισμός 3.7.1.1.2 (Συστολή-Σχήμα 3.7.1.1.2) Η συστολή του A από το B θα σημειώνεται και θα είναι το παρακάτω σύνολο

$$A - B = \{x, (B)_x \subseteq A\} \quad (3.7.1.1.2)$$

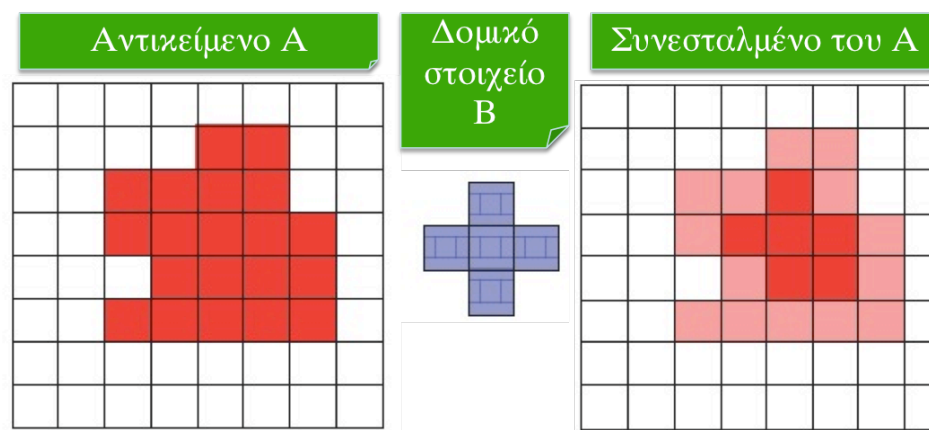
Το B θα είναι ο μορφολογικός τελεστής, ο οποίος στην περίπτωση της μαθηματικής μορφολογίας καλείται και *δομικό στοιχείο*. Συνεπώς, για ένα δυαδικό αντικείμενο A και για ένα δυαδικό συμμετρικό δομικό στοιχείο B , οι μορφολογικοί μετασχηματισμοί διατρέχουν την εικόνα θεωρώντας το B ως μια δυαδική μάσκα :

1. Αν με επίκεντρο το (χ, ψ) του B (αντιστοιχεί στο κεντρικό εικονοστοιχείο του B), το B τέμνει το A , τότε η τιμή για το διεσταλμένο του A ως προς το B στο (χ, ψ) θα είναι ίση με 1, αλλιώς ίση με 0.



Σχήμα 3.7.1.1.1 Διεσταλμένο του αντικειμένου Α ως προς το δομικό στοιχείο Β

2. Αν το B δεν περιέχεται ολόκληρο μέσα στο A , η τιμή του συνεσταλμένου του A ως προς το B στο (x,y) θα είναι ίση με 0, αλλιώς ίση με 1.



Σχήμα 3.7.1.1.2 Συνεσταλμένο του αντικειμένου Α ως προς το δομικό στοιχείο Β

3.7.1.2 Άνοιγμα και κλείσιμο

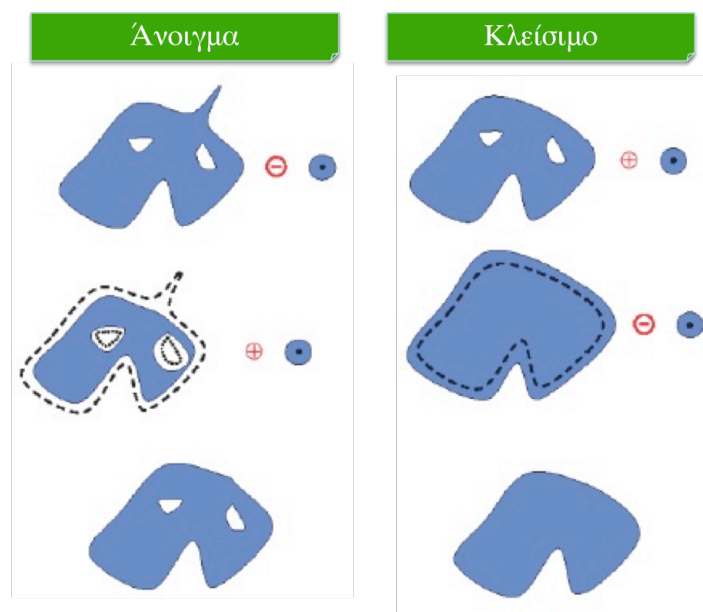
Με βάση τη διαστολή και τη συστολή καθορίζονται δύο νέοι μορφολογικοί μετασχηματισμοί: το *άνοιγμα* και το *κλείσιμο*.

Ορισμός 3.7.1.2.1 (Άνοιγμα-Σχήμα 3.7.1.2.1) Το άνοιγμα του A ως προς το B ορίζεται από τη σχέση:

$$A \circ B = (A - B) \oplus B \quad (3.7.1.2.1)$$

Ορισμός 3.7.1.2.2 (Κλείσιμο-Σχήμα 3.7.1.2.1) Το κλείσιμο του A ως προς το B ορίζεται από τη σχέση:

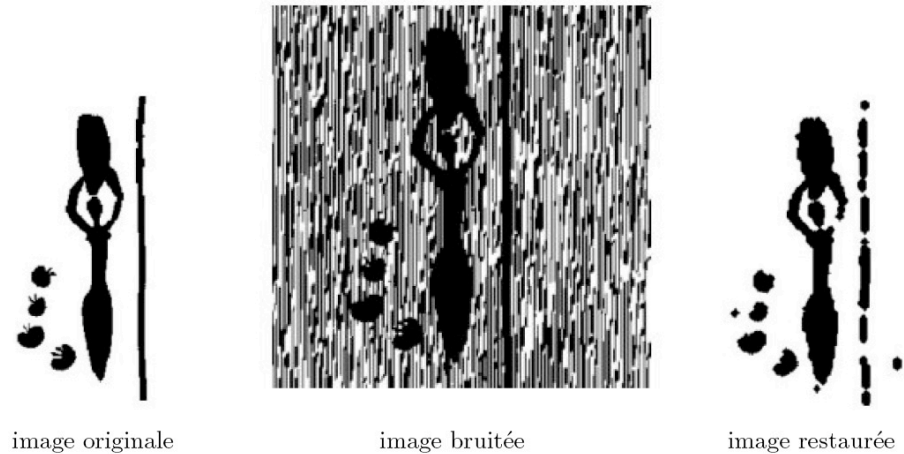
$$A \bullet B = (A \oplus B) - B \quad (3.7.1.2.2)$$



Σχήμα 3.7.1.2.1 Μορφολογικό άνοιγμα και κλείσιμο [Hel-Or 2000]

Γενικά το άνοιγμα εφαρμόζει λείανση στα περιγράμματα μιας εικόνας, σπάει τους ισθμούς μεταξύ των αντικειμένων και μειώνει τα μικρά απομονωμένα αντικείμενα (στο μέγεθος του B). Αντίστοιχα, το κλείσιμο τείνει να λειαίνει τα περιγράμματα της εικόνας, αλλά συνενώνει ταυτόχρονα αντικείμενα που βρίσκονται πολύ κοντά το ένα στο άλλο, καλύπτει μικρές τρύπες (στο μέγεθος του B) και ενώνει τα περιγράμματα. Στο σχήμα 3.7.1.2.1 παρουσιάζεται το αποτέλεσμα μορφολογικού ανοίγματος στην προηγούμενη εικόνα, για την περίπτωση κυκλικού δομικού στοιχείου.

Στο σχήμα 3.7.1.2.2 βλέπουμε μια εφαρμογή των μορφολογικών μετασχηματισμών με σκοπό τη μείωση θορύβου σε εικόνα. Το παράδειγμα βασίζεται σε πίνακα του Henri Metisse (*La femme à l'amphore*, 1952). Ξεκινώντας από τα αριστερά βλέπουμε το πρωτότυπο έργο, στη μέση μια έκδοση αλλοιωμένη από κατακόρυφο θόρυβο και στα δεξιά την επιδιορθωμένη εικόνα μέσω εφαρμογής μορφολογικού ανοίγματος με χρήση προσαρμοσμένου δομικού στοιχείου.



Σχήμα 3.7.1.2.2 Εφαρμογή μορφολογικών μετασχηματισμών σε πίνακα του Henri Metisse « La femme à l'amphore » (1952)

Στο σχήμα 3.7.1.2.3 βλέπουμε μια επιπλέον εφαρμογή των μορφολογικών μετασχηματισμών με σκοπό τη μείωση θορύβου σε εικόνα.

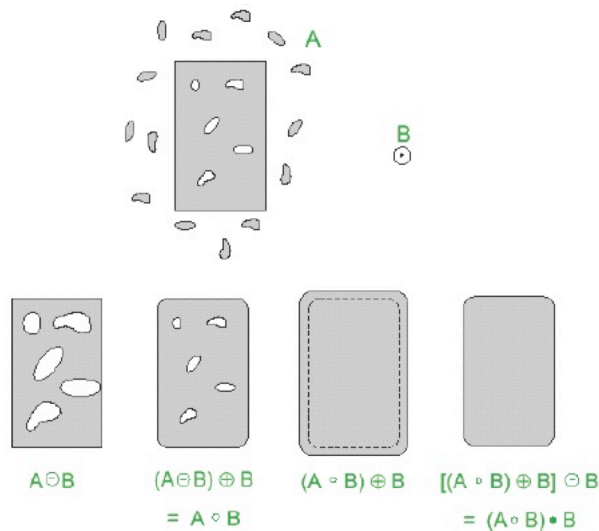
3.7.2 Ανάκτηση περιγράμματος χεριού

Το περίγραμμα ενός αντικειμένου μέσα σε μια δυαδική εικόνα αντιστοιχεί συχνά στα ολικά ακρότατα της κλίσης. Οι δυσκολίες στην ανάκτηση περιγράμματος εντοπίζονται κυρίως στην ύπαρξη σημαντικού θορύβου στην εικόνα. Ο θόρυβος αυτός μπορεί να προέρχεται από την πηγή λήψης, από τη δειγματοληψία ή από ανομοιογενείς επιφάνειες των αντικειμένων.

Προτεραιότητα στην ανάκτηση του περιγράμματος του χεριού στην εικόνα είναι η μείωση του θορύβου. Για να επιτευχθεί η μείωση αυτή θα πρέπει προηγουμένως να καθοριστούν κριτήρια απόδοσης. Στη συνέχεια, επιλέγονται τα κατάλληλα φίλτρα που θα ικα-

νοποιούν τα κριτήρια απόδοσης. Τέλος, η εφαρμογή των φίλτρων θα οδηγήσει στη μείωση του ανεπιθύμητου θορύβου.

Σημαντικό στάδιο στην ανάκτηση περιγράμματος είναι η χρήση των γραμμικών φίλτρων. Τα γραμμικά φίλτρα ενδείκνυνται στην περίπτωση που δε διαθέτουμε εκ των προτέρων πληροφορίες για την εικόνα. Η μέθοδος αυτή καλεί αλγορίθμους χαμηλής πολυπλοκότητας, τα αποτελέσματα των οποίων είναι ικανοποιητικά σε μεγάλο πλήθος εικόνων.



Σχήμα 3.7.1.2.3 Εφαρμογή μορφολογικών μετασχηματισμών [Gonzales 2008]

3.7.2.1 Γραμμικό φιλτράρισμα εικόνας

Φιλτράρισμα εικόνας είναι η συνέλιξη μιας εικόνας $I(x,y)$ με τη συνάρτηση $f(x,y)$. Στην περίπτωση εφαρμογής συνεχούς φίλτρου, η φιλτραρισμένη εικόνα δίνεται από τη σχέση:

$$\begin{aligned}
 I_f(x,y) &= (f * I)(x,y) = \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x',y') I(x-x',y-y') dx' dy' = \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-x',y-y') I(x',y') dx' dy' =
 \end{aligned} \tag{3.7.2.1.1}$$

Στην περίπτωση διακριτού φιλτραρίσματος, τόσο εικόνα I όσο και τη συνάρτηση f κινούνται σε φραγμένο σύνολο. Για την I αυτό θα είναι $[-N/2, +N/2]$ (όπου N η διάσταση της εικόνας) ενώ για την f θα είναι $[-K/2, +K/2]$, με $K \leq N$. Συνεπώς, η συνέλιξη θα γράφεται:

$$\begin{aligned} I_f(x,y) &= (f * I)(x,y) = \\ &= \sum_{i'=-K/2}^{i'=+K/2} \sum_{j'=-K/2}^{j'=-K/2} f(i-i', j-j') I(i', j') \end{aligned} \quad (3.7.2.1.2)$$

3.7.2.2 Κλίση μιας εικόνας

Η κλίση μιας εικόνας υπολογίζεται ως εξής:

$$I_x(x,y) = \frac{\partial I_f(x,y)}{\partial x} \quad (3.7.2.2.1)$$

$$I_y(x,y) = \frac{\partial I_f(x,y)}{\partial y} \quad (3.7.2.2.1)$$

Σε κάθε σημείο της εικόνας (x,y) μπορούμε να υπολογίσουμε το διάνυσμα κλίσης. Το μέτρο και η κατεύθυνση του διανύσματος δίνονται από τις παρακάτω σχέσεις:

$$G = (I_x^2 + I_y^2)^{\frac{1}{2}} \approx \max(I_x, I_y) \approx |I_x| + |I_y| \quad (3.7.2.2.3)$$

$$\phi = \arctan\left(\frac{I_y}{I_x}\right) \quad (3.7.2.2.4)$$

3.7.2.3 Γεωμετρική ερμηνεία της κλίσης

Για να κατανοηθεί καλύτερη η έννοια της κλίσης θεωρούμε μια εικόνα που απαρτίζεται από δύο «λείες» περιοχές, οι οποίες χωρίζονται από ένα περίγραμμα «σκαλοπατιού» (Σχήμα 3.7.2.3.1). Η ερμηνεία της γεωμετρικής διάστασης της κλίσης θα γίνει με τη βοήθεια της συνάρτησης «σκαλοπατιού»:

$$u(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (3.7.2.3.1)$$



Σχήμα 3.7.2.3.1 Δύο «λείες» περιοχές που χωρίζονται από περίγραμμα «σκαλοπατιού»

Επίσης, θα χρησιμοποιηθεί η συνάρτηση «δέλτα» του Dirac:

$$\delta(t) = \begin{cases} 1, & \text{αν } t = 0 \\ 0, & \text{αν } t \neq 0 \end{cases} \quad (3.7.2.3.2)$$

Η συνάρτηση «σκαλοπάτι» μπορεί να επαναδιατυπωθεί με τη βοήθεια της συνάρτησης του Dirac:

$$u(z) = \int_{-\infty}^z \delta(t) dt \quad (3.7.2.3.3)$$

Η πρώτη παράγωγος της $u(t)$ γράφεται:

$$u'(t) = \delta(t) \quad (3.7.2.3.4)$$

Το ευθύγραμμο τμήμα που χωρίζει τις δύο διαφορετικές περιοχές της εικόνας μπορεί να διατυπωθεί σε συνάρτηση της γωνίας θ που σχηματίζει με τον άξονα x . Συνεπώς:

$$-x\eta\mu\theta + y\sigma\upsilon\nu\theta + \rho = 0 \quad (3.7.2.3.5)$$

Κατά συνέπεια η συνάρτηση της εικόνας μπορεί να γραφεί:

$$I(x, y) = B_1 + (B_1 - B_2)u(-x\eta\mu\theta + y\sigma\upsilon\nu\theta + \rho) \quad (3.7.2.3.6)$$

Οι συνιστώσες του διανύσματος κλίσης στο σημείο (x, y) θα είναι:

$$\begin{aligned} I_x &= \sigma\upsilon\nu(\theta + \frac{\pi}{2})(B_2 - B_1)\delta(-x\eta\mu\theta + y\sigma\upsilon\nu\theta + \rho) \\ I_y &= \eta\mu(\theta + \frac{\pi}{2})(B_2 - B_1)\delta(-x\eta\mu\theta + y\sigma\upsilon\nu\theta + \rho) \end{aligned} \quad (3.7.2.3.7)$$

Η κλίση θα έχει την τιμή μηδέν παντού στην εικόνα εκτός από το ευθύγραμμο τμήμα του περιγράμματος «σκαλοπατιού»:

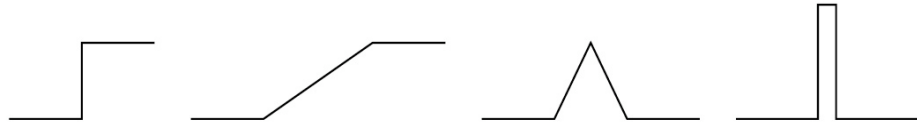
$$\varphi = \theta + \frac{\pi}{2} \quad (3.7.2.3.8)$$

Το μέτρο της κλίσης θα είναι ίσο με την παράγωγο της εικόνας κατά μήκος του ευθύγραμμου τμήματος του περιγράμματος.

3.7.2.4 Μοντέλα περιγράμματος

Βασικό στάδιο στην εξαγωγή χαρακτηριστικών ιδιοτήτων είναι ο καθορισμός του είδους των περιγραμμάτων για τα οποία ενδιαφερόμαστε, καθώς και τα χαρακτηριστικά των θορύβων που συναντούμε. Έτσι, τα πιο συνηθισμένα είδη περιγράμματος, θορύβου αλλά και οι μοντελοποιήσεις αυτών παρατίθενται παρακάτω:

- **Είδη περιγραμμάτων:** Πολλά είδη περιγραμμάτων μπορούν να καταγραφούν μέσα σε εικόνες. Τα περιγράμματα κατηγοριοποιούνται σε τύπου: «σκαλοπατιού», «ράμπας», «σκεπής» και «κορυφής» (Σχήμα 3.7.2.4.1).



Σχήμα 3.7.2.4.1 Είδη περιγραμμάτων : « σκαλοπάτι », « ράμπα », « σκεπή », « κορυφή »

- **Είδη θορύβου:** Ομοίως με τα περιγράμματα, διάφορα είδη θορύβου συναντώνται μέσα στις εικόνες: συνελκτικός, λευκός, παλμικός κλπ.
- **Μοντέλα περιγράμματος:** Τα περιγράμματα είδους «σκαλοπατιού» εμφανίζονται πολύ συχνά μέσα σε εικόνες σε συνδυασμό με λευκό θόρυβο, ο οποίος μοντελοποιείται σχετικά εύκολα.

Έστω A το ύψος του «σκαλοπατιού» και n_0^2 η διασπορά του λευκού θορύβου:

$$I(x) = Au_{-1}(x) + n(x) \quad (3.7.2.4.1)$$

Έστω $u_{-1}(x)$ η παρακάτω συνάρτηση:

$$u_{-1}(x) = \begin{cases} 0, & \forall x \in [-W, 0] \\ 1, & \forall x \in (0, W] \end{cases} \quad (3.7.2.4.2)$$

$I(x)$ είναι ένα σήμα εισόδου, το οποίο αναπαρίσταται ως:

$$n_0^2 = \begin{cases} \int_{-W}^W n^2(x) dx \\ \sum_{-W}^W n^2(x) \end{cases} \quad (3.7.2.4.3)$$

Όπου:

- Στην πρώτη περίπτωση έχουμε τον υπολογισμό του πλάτους θορύβου για συνεχές σήμα εισόδου,
- Στη δεύτερη περίπτωση έχουμε τον υπολογισμό του πλάτους θορύβου για διακριτό σήμα εισόδου,
- n_0 , το πλάτος του θορύβου
- $2W$, η περιοχή διενέργειας ανίχνευσης.

3.7.2.5 Κριτήρια απόδοσης

Γενικά, η εκτίμηση του αποτελέσματος ανάκτησης περιγράμματος είναι μια δύσκολη διαδικασία, διότι ενώ μπορεί να γίνει πολύ εύκολα με το «μάτι», αντίθετα είναι πολύπλοκο να περιγραφεί με μαθηματικούς όρους.

Συνεπώς, είναι απαραίτητο να καθοριστούν κριτήρια απόδοσης για τον ανιχνευτή περιγράμματος. Αυτά τα κριτήρια αναφέρονται:

1. **Ανίχνευση σημείων περιγράμματος:** Ας είναι $S(x)$ ένα ιδανικό περίγραμμα (χωρίς θόρυβο), το οποίο εντοπίζεται στο σημείο $x=0$. Ας είναι $f(x)$ ένα φίλτρο πεπερασμένης κρουστικής απόκρισης (FIR). Θεωρείται ότι το $f(x)$ είναι μηδέν παντού εκτός από το διάστημα $[-W, W]$. Προκειμένου να επιτευχθεί καλή ανίχνευση σημείων περιγράμματος θα πρέπει να μεγιστοποιηθεί ο λόγος σήματος προς θόρυβο ΛΣΘ:

$$\Lambda\S\Theta = \frac{\left| \int_{-W}^W S(-x)f(x)dx \right|}{n_0 \sqrt{\int_{-W}^W f^2(x)dx}} \quad (3.7.2.5.1)$$

Όπου:

- Ο αριθμητής αναπαριστά την απόκριση περιγράμματος για $x=0$,
- Ο παρονομαστής αναπαριστά την απόκριση θορύβου (θεωρείται προσθετικός Γκαουσιανός θόρυβος),
- n_0 , το πλάτος του θορύβου

➤ $2W$, η περιοχή όπου διενεργείται η ανίχνευση.

2. **Εντοπισμός σημείων περιγράμματος:** Στόχος είναι η μεγιστοποίηση της έκφρασης εντοπισμού L :

$$L = \frac{\left| \int_{-W}^W S'(x) f'(-x) dx \right|}{n_0 \sqrt{\int_{-W}^W f'^2(x) dx}} \quad (3.7.2.5.2)$$

Η μεγιστοποίηση της έκφρασης L αντιστοιχεί στην ελαχιστοποίηση του προσδοκώμενου της απόστασης ανάμεσα στο πραγματικό σημείο του περιγράμματος και στο ανιχνευμένο σημείο. Προκειμένου να ικανοποιηθούν τα κριτήρια της καλής ανίχνευσης και εντοπισμού θα πρέπει να μεγιστοποιηθούν οι εκφράσεις L και $\Lambda\Sigma\Theta$.

3. **Μοναδικότητα απόκρισης για κάθε περίγραμμα:** Ο ανιχνευτής δεν πρέπει να εξάγει περισσότερες αποκρίσεις για ένα περίγραμμα. Συνεπώς, αναζητούμε μια έκφραση για την προσδοκώμενη απόσταση ανάμεσα στις κορυφές του θορύβου. Ανάμεσα σε δύο μέγιστα υπάρχουν δύο μηδενισμοί της πρώτης παραγώγου. Έτσι, βρέθηκε ότι η μέση απόσταση μεταξύ μεγίστων είναι δύο φορές η μέση απόσταση μηδενισμού της πρώτης παραγώγου. Θα ισχύει δηλαδή:

$$X_0(f) = \pi \sqrt{\frac{\int_{-W}^W f'^2(x) dx}{\int_{-W}^W f''^2(x) dx}} \quad (3.7.2.5.3)$$

Όπου:

- X_0 , η μέση απόσταση ανάμεσα σε δύο μηδενισμούς της πρώτης παραγώγου.

Θεωρείται το $X_{\max}(f)$ που είναι η μέση απόσταση ανάμεσα στα μέγιστα θορύβου, δηλαδή ισχύει:

$$X_{\max} = 2X_0(f) \quad (3.7.2.5.4)$$

Στόχος είναι να μεγιστοποιηθεί το $X_{\max}(f)$ ώστε να «εξαλειφθούν» τα μέγιστα θορύβου. Σχεδιάζοντας το φίλτρο $f(x)$, ο ερευνητής είναι ελεύθερος να επιλέξει κάποιο k ($0 \leq k \leq 1$) για το οποίο ισχύει $X_{\max}(f) = kW$. Σύμφωνα με τον Rice (1944) N_n θα είναι το προσδοκώμενο πλήθος μεγίστων θορύβου σε μια περιοχή πλάτους $2W$:

$$N_n = \frac{2W}{X_{\max}(f)} = \frac{2}{k} \quad (3.7.2.5.5)$$

Τέλος, αξίζει να σημειωθεί ότι το τρίτο κριτήριο εισήχθηκε λόγω του ότι η ανίχνευση λαμβάνει υπόψη την απόκριση του φίλτρου μόνο στο σημείο του περιγράμματος και όχι στις γειτονικές περιοχές. Είναι το μοναδικό κριτήριο που λαμβάνει υπόψη του τον Γκαουσιανό θόρυβο.

3.8 Εντοπισμός άκρων δακτύλων

Στην ενότητα αυτή θα παρουσιαστούν οι βασικότερες μέθοδοι εντοπισμού άκρων δακτύλων σε εικόνα. Χαρακτηριστικό όλων αυτών των αλγορίθμων είναι ο εντοπισμός της θέσης καθενός δακτύλου χωριστά. Ικανοποιούν δηλαδή την εκ των προτέρων απαίτηση για προσδιορισμό της κίνησης του κάθε δακτύλου χωριστά.

Οι αλγόριθμοι αυτοί αναπτύχθηκαν κυρίως λόγω της ανάγκης καταγραφής μουσικών δακτυλικών κινήσεων. Μπορούν όμως να χρησιμοποιηθούν και σε οποιαδήποτε άλλη περίπτωση απαιτείται εντοπισμός αναγνώρισης δακτυλικών κινήσεων (πχ στην περίπτωση αναγνώρισης λεπτομερών δακτυλικών κινήσεων που προέρχονται από σπάνια επαγγέλματα). Οι αλγόριθμοι αυτοί κατηγοριοποιούνται σε α) προβολή υπογραφής, β) κυκλικού μετασχηματισμού Hough, γ) εξαγωγής γεωμετρικών χαρακτηριστικών ιδιοτήτων και δ) χρωματικών δεικτών. Οι αλγόριθμοι που θα παρουσιαστούν στην παρούσα διδακτορική διατριβή προέρχονται από μακροχρόνια έρευνα των Letessier και Brard (2004).

Τέλος, θα παρουσιαστούν συνοπτικά οι αλγόριθμοι της κατηγορίας α, β και δ, ενώ θα αναλυθεί η κατηγορία γ, ως η βάση ανάπτυξης ενός πιο αποδοτικού αλγορίθμου εντοπισμού δακτύλων. Η

παρουσίαση θα ολοκληρωθεί με μια συνοπτική σύγκριση των αλγορίθμων.

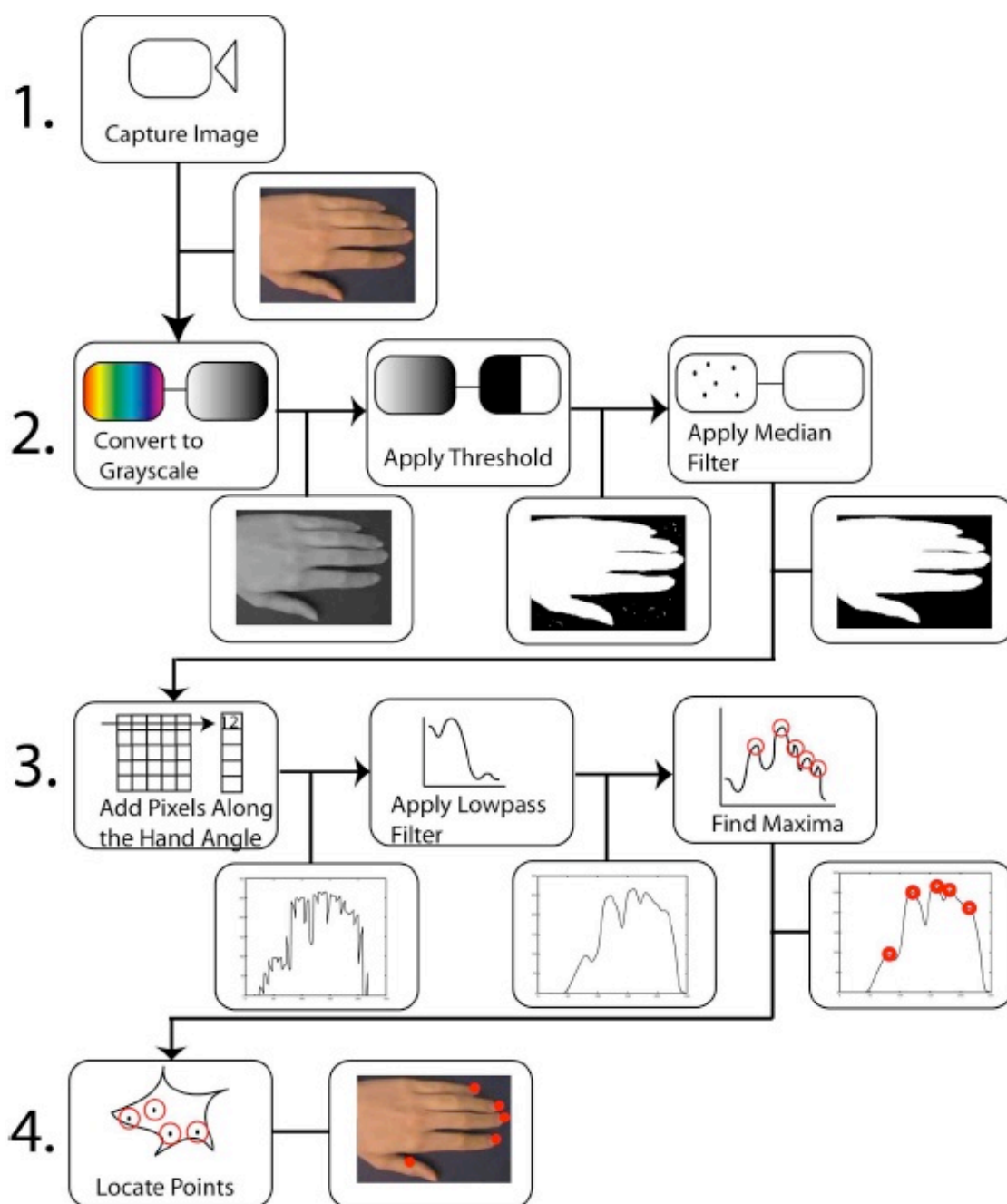
3.8.1 Αλγόριθμοι προβολής υπογραφών

Ένας αλγόριθμος προβολής υπογραφών δρα απευθείας σε εξαγόμενες κατωφλιωμένες δυαδικές εικόνες του χεριού. Ο πυρήνας της διαδικασίας που ακολουθείται από τον αλγόριθμο παρουσιάζεται στην τρίτη γραμμή του σχήματος 3.8.1.1. Ένα χαμηλοπερατό φίλτρο εφαρμόζεται στην υπογραφή ώστε να μειώσει αποκλίσεις σε υψηλές συχνότητες, οι οποίες δημιουργούν σημαντικό πλήθος από τοπικά μέγιστα και πολλαπλασιάζουν τις πιθανές θέσεις δακτύλων. Τα τελικά πέντε σημεία που λαμβάνονται αντιστοιχούν στις θέσεις των πέντε δακτύλων του χεριού.

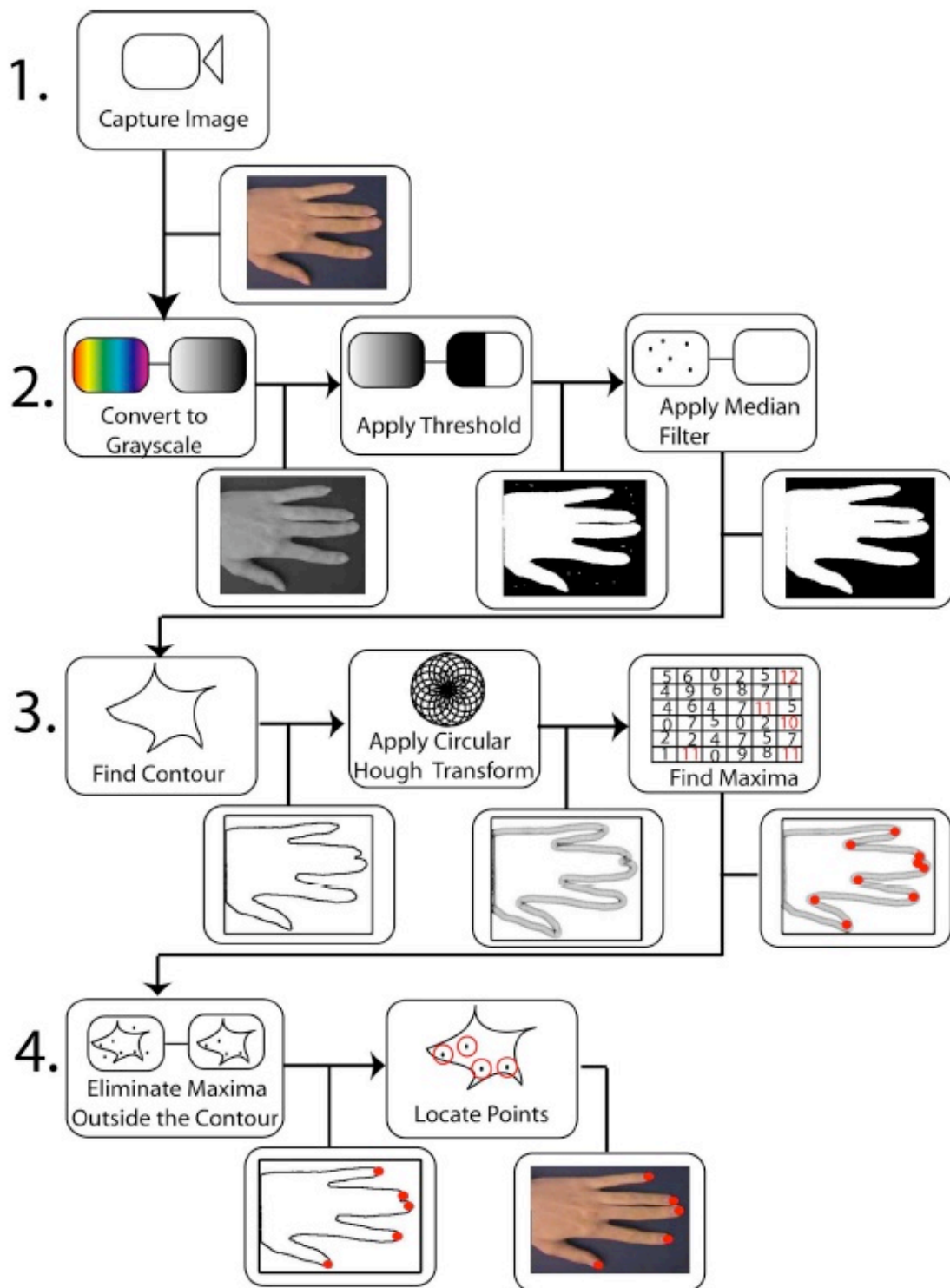
3.8.2 Αλγόριθμοι κυκλικού μετασχηματισμού Hough

Ο κυκλικός μετασχηματισμός του Hough μπορεί να εφαρμοστεί στο περίγραμμα του χεριού, ενώ μπορεί επίσης να εφαρμοστεί σε εικόνα με πολύπλοκο φόντο. Βασίζεται στο σκεπτικό ότι τα άκρα των δακτύλων και οι κοιλάδες των δακτύλων έχουν περίπου κυκλικό σχήμα. Αντίθετα, το περίγραμμα του υπόλοιπου χεριού αποτελείται κυρίως από ευθείες γραμμές.

Οι αντίστοιχοι αλγόριθμοι, αποτυπώνουν ίχνη κύκλων σταθερής ακτίνας πάνω στην εικόνα και οι περιοχές που αντιστοιχούν καλύτερα στους κύκλους αυτούς θεωρείται ότι ανήκουν σε άκρα δακτύλων ή κοιλάδες. Η διαδικασία περιγράφεται στην τρίτη γραμμή του σχήματος 3.8.2.1. Το μειονέκτημα του κυκλικού μετασχηματισμού είναι ότι δεν κάνει απευθείας διαχωρισμό άκρων και κοιλάδων, οπότε απαιτείται περαιτέρω επεξεργασία μέσω φιλτραρίσματος. Ως εκ τούτου, αυξάνεται και η δυσκολία ταξινόμησης και ταυτοποίησης των δακτύλων.



Σχήμα 3.8.1.1 Αλγόριθμος εντοπισμού δακτύλων χρησιμοποιώντας προβολές υπογραφών [Burns 2006]



Σχήμα 3.8.2.1 Αλγόριθμος εντοπισμού δακτύλων χρησιμοποιώντας κυκλικό μετασχηματισμό Hough [Burns 2006]

3.8.3 Αλγόριθμοι εξαγωγής γεωμετρικών χαρακτηριστικών ιδιοτήτων

Η εξαγωγή χαρακτηριστικών ιδιοτήτων παίζει πολύ σημαντικό ρόλο στην αναγνώριση της κίνησης στα συστήματα όρασης H/Y. Ο όρος « χαρακτηριστικές ιδιότητες » συναντάται συχνά στη βιβλιογραφία και ως « χαρακτηριστικά ». Σύμφωνα με τον Παπαμάρκο (1994) οι χαρακτηριστικές ιδιότητες ορίζονται ως εξής :

« Γενικά, ως χαρακτηριστική ιδιότητα μπορεί να θεωρηθεί οποιοδήποτε μετρήσιμο μέγεθος που εξάγεται από μια εικόνα. »

Εφόσον η φιγούρα και το περίγραμμα του χεριού στην εικόνα έχουν ήδη ανακτηθεί, ο προσδιορισμός του χεριού ή των δακτύλων μέσω χαρακτηριστικών ιδιοτήτων, θα επιτρέψει την απεικόνισή τους στον 2Δ χώρο που είναι η εικόνα. Κατ'επέκταση επιτρέπεται πλέον η αναγνώριση θέσης ή στάσης των δακτύλων μέσω της μέτρησης αυτών των μεγεθών. Βασικό κριτήριο επιλογής χαρακτηριστικών ιδιοτήτων, σύμφωνα με τον Παπαμάρκο, είναι ο όσο καλύτερος μονοσήμαντος διαχωρισμός και περιγραφή των προς αναγνώριση αντικειμένων. Αναλυτικότερα, τα κριτήρια που πρέπει να ικανοποιούνται είναι α) η *διακριτικότητα*, η δυνατότητα δηλαδή να διαχωρίζονται τα αντικείμενα μεταξύ τους, β) η *αξιοπιστία*, παρόμοια αντικείμενα πρέπει να περιγράφονται με κοντινές τιμές χαρακτηριστικών ιδιοτήτων, γ) το *μικρό υπολογιστικό κόστος*, το οποίο οφείλει να είναι αρκετά μικρό, ειδικά για εφαρμογές πραγματικού χρόνου, και δ) η *καλή προσαρμογή με τη διαδικασία ταξινόμησης*, οι χαρακτηριστικές ιδιότητες οφείλουν να ταιριάζουν με τη μέθοδο ταξινόμησης που θα ακολουθηθεί στη συνέχεια.

Η δεύτερη κατηγορία αλγορίθμων εντοπισμού άκρων δακτύλων βασίζεται στην εξαγωγή γεωμετρικών χαρακτηριστικών από την εικόνα. Όπως φαίνεται στην τρίτη γραμμή του σχήματος 3.8.3.1, γίνεται χρήση του περιγράμματος του χεριού, πάνω στο οποίο προσδιορίζεται ένα σημείο αναφοράς. Αυτό το σημείο γενικότερα μπορεί αν είναι το κέντρο μάζας του χεριού ή ο καθορισμός ενός σταθερού σημείου πάνω στον καρπό [Yörük 2006]. Υπολογίζονται οι ευκλείδειες αποστάσεις μεταξύ του κέντρου μάζας ή του σημείου αναφοράς και οποιουδήποτε σημείου του περιγράμματος του χεριού. Τα πέντε τοπικά μέγιστα αντιστοιχούν στα πέντε άκρα των

δακτύλων. Τα τοπικά ελάχιστα αντιστοιχούν στις « κοιλάδες » ανάμεσα στα δάκτυλα. Τέλος, εφαρμόζεται και φιλτράρισμα ώστε να μειωθεί το πρόβλημα πολλαπλών πιθανών σημείων άκρων δακτύλων.

3.8.4 Αλγόριθμοι χρωματικών δεικτών

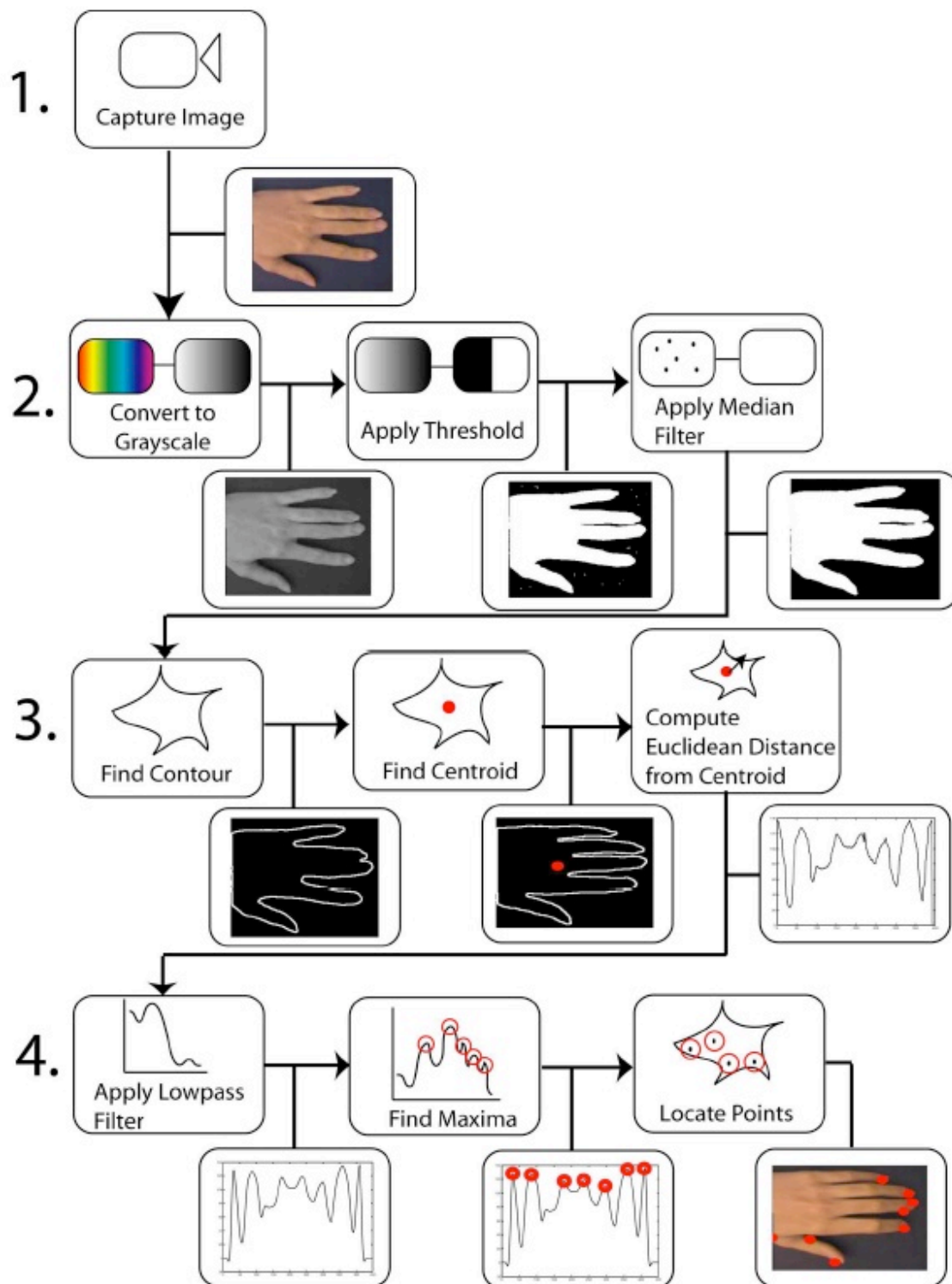
Οι προηγούμενοι τρεις αλγόριθμοι βασίζονται στα χαρακτηριστικά του χεριού ώστε να εντοπίσουν τα άκρα των δακτύλων. Ένας αλγόριθμος χρωματικών δεικτών ανιχνεύει το χρώμα αυτών των δεικτών που είναι τοποθετημένοι πάνω στα άκρα των δακτύλων. Κάθε χρώμα ανιχνεύεται χωριστά χρησιμοποιώντας χρωματική τμηματοποίηση και φιλτράρισμα όπως φαίνεται στη δεύτερη γραμμή του σχήματος 3.8.4.1. Το χρώμα των δεικτών επιτρέπει την ταυτοποίηση και τον εντοπισμό των άκρων των δακτύλων.

3.8.5 Συμπεράσματα

Μετά από αναλυτική σύγκριση που έγινε από την Burns (2006) προέκυψαν χρήσιμα συμπεράσματα ως προς την αποδοτικότητα των παραπάνω αλγορίθμων.

Γενικά, όλοι οι παραπάνω αλγόριθμοι πετυχαίνουν, με διαφορετική ακρίβεια, να εντοπίσουν κάθε δάκτυλο του χεριού. Ο αλγόριθμος προβολής υπογραφών πετυχαίνει μόνο ταυτοποίηση των δακτύλων. Αντίθετα, οι αλγόριθμοι κυκλικού μετασχηματισμού Hough και εξαγωγής γεωμετρικών χαρακτηριστικών ιδιοτήτων μπορούν να εντοπίσουν τόσο τα σημεία ένωσης των δακτύλων όσο και τα σημεία των άκρων των δακτύλων. Αξίζει να σημειωθεί ότι στην περίπτωση που υπάρχει απόκρυψη κάποιου δακτύλου, τα σημεία που εξάγονται ως πιθανά άκρα δεν αντιστοιχούν στα πραγματικά. Στην περίπτωση των γεωμετρικών χαρακτηριστικών γίνεται διάκριση μεταξύ άκρων δακτύλων και κοιλάδων, ενώ αντίθετα ο κυκλικός μετασχηματισμός Hough αδυνατεί να κάνει τέτοιου είδους διάκριση. Ο αλγόριθμος χρωματικών δεικτών μπορεί να ταυτοποιήσει τα δάκτυλα μόνο στην περίπτωση χρήσης διαφορετικού χρώματος ανά δάκτυλο, μειώνοντας σημαντικά τη φιλικότητα προς το χρήστη αφού υποχρεώνεται σε χρήση εξοπλισμού.

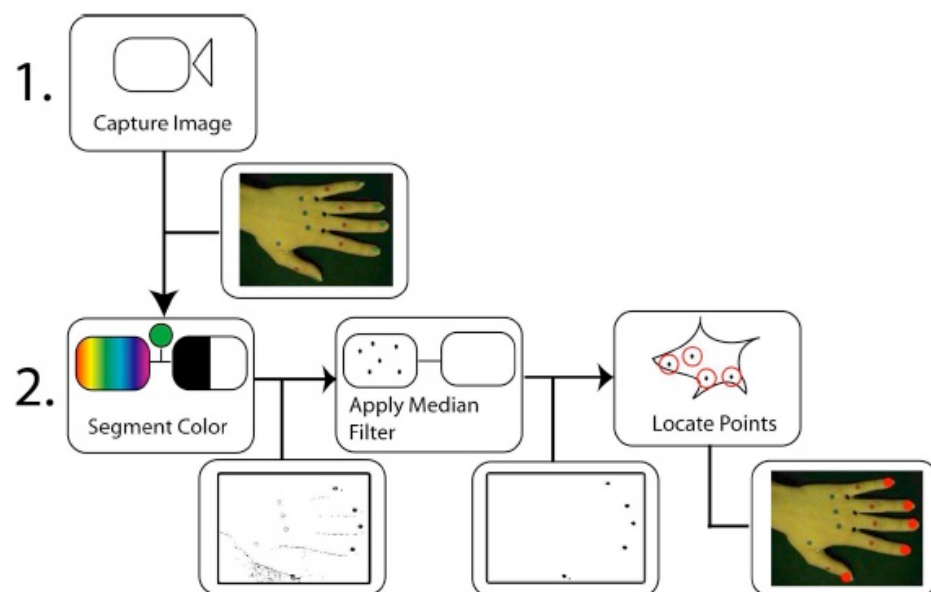
Η προβολή υπογραφών και η εξαγωγή γεωμετρικών χαρακτηριστικών απαιτούν σημαντική τμηματοποίηση πριν την εφαρμογή τους. Ο κυκλικός μετασχηματισμός μπορεί να εφαρμοστεί και σε περιπτώσεις σύνθετου φόντου, προκαλείται όμως σύγχυση στην περίπτωση που στην εικόνα εμφανίζεται επιπλέον κυκλικά αντικείμενα διαφορετικής ακτίνας από αυτή των δακτύλων.



Σχήμα 3.8.3.1 Αλγόριθμος εντοπισμού δακτύλων χρησιμοποιώντας ευκλείδειες αποστάσεις [Burns 2006]

Στα 25 καρέ/δευτερόλεπτο όλοι οι αλγόριθμοι έχουν ικανοποιητικά αποτελέσματα ακόμα και σε πολύ γρήγορη κίνηση των δακτύλων. Παρόλα αυτά, ο αλγόριθμος του κυκλικού μετασχηματισμού είναι πολύ περισσότερο απαιτητικός σε χρήση της ΚΜΕ από ότι οι υπόλοιποι. Αυτός είναι ένας σημαντικός περιορισμός για χρήση σε πραγματικό χρόνο αλλά και όταν παράλληλα λειτουργούν άλλοι αλγόριθμοι αναγνώρισης κινήσεων. Οι αλγόριθμοι κυκλικού μετασχηματισμού και εξαγωγής γεωμετρικών χαρακτηριστικών έχουν παρόμοια καλά αποτελέσματα στην ακρίβεια εντοπισμού της θέσης των δακτύλων.

Τέλος, ο αλγόριθμος προβολής υπογραφών μπορεί να χρησιμοποιηθεί μόνο σε πλήρως ελεγχόμενο περιβάλλον όπου είναι εκ των προτέρων γνωστή η κατεύθυνση των δακτύλων και όπου τα δάκτυλα δεν είναι πολύ κλειστά. Το γεγονός αυτό απορρίπτει τη χρήση της προβολής υπογραφών στην περίπτωση αναγνώρισης δακτυλικών κινήσεων πιανίστα γιατί το χέρι είναι μονίμως σε ημί-κλειστη στάση.



Σχήμα 3.8.4.1 Αλγόριθμος εντοπισμού δακτύλων χρησιμοποιώντας χρωματικούς δείκτες [Burns 2006]

3.9 Δυναμική αναγνώριση κινήσεων

Η αναγνώριση κινήσεων μπορεί να βασίζεται στη σύγκριση καμπυλών, προερχόμενες από μια ακολουθία εισόδου, με τις καμπύλες αναφοράς που υπάρχουν για κάθε προκαθορισμένη κίνηση. Μπορεί ακόμα να βασίζεται στην ανακατασκευή του αντικειμένου προς αναγνώριση με βάση την εμφάνισή του.

Στη συνέχεια θα γίνει μια σύντομη παρουσίαση των δημοφιλέστερων μεθόδων αναγνώρισης κινήσεων : α) της *Δυναμικής Περιτύλιξης του Χρόνου*, β) των *Νευρωνικών Δικτύων Χρονικής Καθυστέρησης*, καθώς και γ) των *Κρυφών Μαρκοβιανών Μοντέλων*.

3.9.1 Δυναμική Περιτύλιξη του Χρόνου

Η μέθοδος Δυναμικής Περιτύλιξης του Χρόνου – *Déformation Temporelle Dynamique* (ΔΠΤ-DTD) υπολογίζει την ομοιότητα μεταξύ δύο ακολουθιών που μπορεί να διαφέρουν μεταξύ τους στο χρόνο ή την ταχύτητα. Για παράδειγμα, εντοπίζονται ομοιότητες μεταξύ δύο βίντεο αυτοκινήτων που κινούνται πάνω σε ένα αυτοκινητόδρομο ακόμα και αν το ένα κινείται αργά, ενώ το άλλο γρήγορα. Η μέθοδος δηλαδή λειτουργεί σε περιπτώσεις μεταβλητού ρυθμού κίνησης.

Η ΔΠΤ χρησιμοποιείται ευρέως στην επεξεργασία βίντεο, ήχου και γραφικών. Γενικά, επιτρέπει την εύρεση της ιδανικής αντιστοιχίας μεταξύ ακολουθιών (πχ *χρονοσειρές*). Οι ακολουθίες « περιτυλίγονται » μη γραμμικά στη διάσταση του χρόνου ώστε να μετρηθεί η ομοιότητα τους ως συνάρτηση μη γραμμικών παραμέτρων.

Η ΔΠΤ βασίζεται στις διαφορές ανάμεσα σε ακολουθίες χρησιμοποιώντας λειτουργίες εισαγωγής, διαγραφής, συμπίεσης-διαστολής και αφαίρεσης υπακολουθιών [Βολουδάκης 2005]. Ο εκ των προτέρων καθορισμός ενός ποσοστού διαφοράς μεταξύ των ακολουθιών πριν και μετά την εκτέλεση των παραπάνω λειτουργιών, συμβάλει στην ταξινόμηση των ακολουθιών.

Η μέθοδος υπόκειται σε ορισμένους σημαντικούς περιορισμούς. Ένα σημαντικό μειονέκτημα της μεθόδου είναι η μονοδιάστατη αντιστοίχιση ως προς την παράμετρο του χρόνου. Επιπλέον, η έννοια της συνέχειας δε λαμβάνεται υπόψη από τη μέθοδο. Αυτό έχει ως συνέπεια να λειτουργεί στη βάση ελιπών δεδομένων. Η απώλεια αυτών των δεδομένων μπορεί να είναι απαγορευτική για τη χρήση της μεθόδου, ειδικά στις περιπτώσεις που

απαιτείται μεγάλη ακρίβεια στην αναγνώριση. Η βελτιστοποίηση της ομοιότητας μεταξύ των ακολουθιών υλοποιείται με τη χρήση δυναμικού προγραμματισμού.

Τέλος, η μέθοδος μπορεί να παράξει καλά αποτελέσματα στην αναγνώριση προκαθορισμένων κινήσεων, όπου το στοχαστικό χαρακτηριστικό των κινήσεων είναι ασθενές. Αγνοεί εντελώς τη σχέση μεταξύ γειτονικών χρονικά ακολουθιών. Γενικά, όμως μια ακολουθία παρουσιάζει μεγαλύτερη συσχέτιση με κοντινότερες ακολουθίες από ότι με τις πιο απομακρυσμένες χρονικά [Βολουδάκης 2005], γεγονός που δεν λαμβάνεται υπόψη.

3.9.2 Νευρωνικά δίκτυα

Οι μέθοδοι νευρωνικών δικτύων που παρουσιάζονται από τη βιβλιογραφία κυρίως αναφέρονται στην ταξινόμηση μοντέλων μέσα από ένα σύνολο δεδομένων.

Το αδύνατο σημείο των νευρωνικών δικτύων είναι η μη συμμετοχή της παραμέτρου του χρόνου. Εναλλακτικά, ο χρόνος προσεγγίζεται ως ένας εξωτερικός μηχανισμός. Τα Νευρωνικά Δίκτυα Χρονικής Καθυστέρησης – Réseaux de Neurones à Lignes de Retard (ΔNT-RNLR) είναι ένα από τα πιο αποτελεσματικά μοντέλα. Εμπνέονται από τα πολυεπίπεδα νευρωνικά δίκτυα. Τα ΝΔΧΚ μετατρέπουν το χρονικό πρόβλημα σε χωρικό πρόβλημα. Έχουν ήδη αποδείξει την αποτελεσματικότητά τους στην αναγνώριση χειρονομιών στη νοηματική γλώσσα.

Τα συστήματα όμως που βασίζονται πάνω στα ΝΔΧΚ έχουν πάντοτε μια πολύ υψηλή πολυπλοκότητα και απαιτούν μια πολύ βαριά επεξεργασία δεδομένων. Ακόμα και στις περιπτώσεις όπου η παράμετρος του χρόνου εισήχθη ως μηχανισμός στο σύστημα, το κόστος υπολογιστηκότητας, η πολυπλοκότητα εκπαίδευσης καθώς και η δυσκολία ερμηνείας των αποτελεσμάτων παραμένουν συχνά σημαντικά εμπόδια στη χρήση των ΝΔΧΚ.

3.9.3 Κρυφά Μαρκοβιανά Μοντέλα

Η μέθοδος των Κρυφών Μαρκοβιανών Μοντέλων – Modèles de Markov Cachés (KMM-MMC) χρησιμοποιούνται ευρέως στην ανάλυση χρονικών σειρών με χωρικές και χρονικές μεταβολές. Έχουν γνωρίσει σημαντική επιτυχία στους τομείς της αναγνώρισης γραφής, φωνής και των χειρονομιών και ειδικότερα στη νοηματική γλώσσα. Αναλυτική παρουσίαση των KMM θα γίνει στο έκτο κεφάλαιο.

Τα ΚΜΜ σε αντίθεση με τη ΔΠΤ και τα ΝΔΧΚ λαμβάνουν υπόψη τους τη χρονική συσχέτιση που υφίσταται ανάμεσα στις ακολουθίες δεδομένων, μέσω της χρήσης μαρκοβιανών αλυσίδων. Τα ΚΜΜ υποθέτουν ότι η παρατηρούμενη ακολουθία (ή αλλιώς ακολουθία διανυσμάτων παρατήρησης) είναι το αποτέλεσμα μιας στοχαστικής κρυφής διαδικασίας που αποτελείται από έναν προκαθορισμένο αριθμό «κρυφών» καταστάσεων [Βολουδάκης 2005]. Τα βασικά συστατικά ενός ΚΜΜ είναι ένα σύνολο από «κρυφές» καταστάσεις, ένα σύνολο καταστάσεων εξόδου, πιθανότητες μετάβασης μεταξύ των «κρυφών» καταστάσεων, πιθανότητες μετάβασης μεταξύ μιας «κρυφής» κατάστασης προς κάποια έξοδο και οι πιθανότητες για την αρχική κατάσταση.

Η επιτυχία των ΚΜΜ στην αναγνώριση φωνής οδήγησε στη χρήση τους και στην αναγνώριση κινήσεων ή χειρονομιών. Η σημαντική διαφορά μεταξύ φωνής και κίνησης είναι ότι η φωνή αποτελείται από σαφείς δομικές μονάδες της γλώσσας (το λεξιλόγιο). Κάτι αντίστοιχο για τις ανθρώπινες κινήσεις δεν υφίσταται. Συνεπώς η όραση H/Y δεν έχει το πλεονέκτημα μετατροπής των εικόνων σε σύμβολα με κάποια αλγοριθμική μέθοδο. Τα ΚΜΜ είναι η πιο δημοφιλής μέθοδος αναγνώρισης κινήσεων λόγω του πλεονεκτήματος που τους παρέχει η στοχαστική τους φύση να διαχειρίζονται την αβεβαιότητα που υπάρχει σε όλες τις κινήσεις.

Μια επέκταση των ΚΜΜ είναι τα Δυναμικά Δίκτυα του Bayes – Réseaux Baysiens Temporels (ΔΔΒ-RBT). Τα ΔΔΒ διαχειρίζονται ικανοποιητικά τρεις ή περισσότερες ξεχωριστές διαδικασίες και αποτελούν μια γενίκευση των ΚΜΜ. Είναι κατευθυντικά γραφικά μοντέλα μιας στοχαστικής διαδικασίας και αναπαριστούν τόσο τις «κρυφές» καταστάσεις όσο και τις καταστάσεις εξόδου των ΚΜΜ, με μεταβλητές κατάστασης, οι οποίες μπορεί να έχουν πολύπλοκες αλληλεξαρτήσεις.

3.10 Γενικά συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκαν τα σημαντικότερα εργαλεία που χρησιμοποιεί η όραση H/Y για την αναγνώριση κινήσεων-χειρονομιών.

Στην αρχή του κεφαλαίου έγινε μια παράθεση των βασικών πτυχών της όρασης, αλλά και το πως κατανοείται αυτή τόσο από τον άνθρωπο όσο και από τον υπολογιστή. Παρουσιάστηκε η πρόταση του David Marr καθώς και η προσαρμογή αυτής στην αναγνώριση δακτυλικών κινήσεων. Βασικοί άξονες του μοντέλου αυτού είναι το 2Δ

σκαρίφημα, το 2,5Δ σκαρίφημα και το 3Δ μοντέλο. Το μοντέλο αυτό καθορίζει τα βήματα που ακολουθούνται από ένα σύστημα ξεκινώντας από τη λήψη της εικόνας και καταλήγοντας στην κατάκτηση της γνώσης που σχετίζεται με την όραση.

Αναπτύχθηκαν οι βασικότερες μέθοδοι εντοπισμού χεριού στην εικόνα. Παρουσιάστηκαν δύο αλγόριθμοι τμηματοποίησης. Στην τμηματοποίηση μέσω χρωματικότητας, λαμβάνεται υπόψη η χαρακτηριστική απόχρωση του ανθρώπινου δέρματος. Η μετατόπιση του χεριού στο χώρο χρησιμοποιείται από τη δεύτερη τεχνική εντοπισμού του χεριού στην εικόνα. Ο εντοπισμός μέσω εμφάνισης βασίζεται στα οπτικά χαρακτηριστικά του προς αναγνώριση αντικειμένου. Παρουσιάστηκαν αναλυτικά τα πλεονεκτήματα και τα μειονεκτήματα της κάθε μεθόδου.

Στη συνέχεια έγινε μια ανασκόπηση των μεθόδων τμηματοποίησης της εικόνας ξεκινώντας από τη λήψη της από την κάμερα και καταλήγοντας στην ανάκτηση του περιγράμματος του χεριού. Περιγράφηκαν οι βασικές πράξεις της μαθηματικής μορφολογίας και οι αντίστοιχες τεχνικές μείωσης θορύβου σε δυαδικές εικόνες. Η κλίση της εικόνας χρησιμοποιείται συχνά ως εργαλείο ανάκτησης περιγράμματος. Παρουσιάστηκαν τα πιο σημαντικά μοντέλα περιγράμματος καθώς και τα κριτήρια απόδοσης ανάκτησης περιγράμματος που ορίζονται από τη βιβλιογραφία.

Στην ενότητα του εντοπισμού των άκρων των δακτύλων, έναν τομέα που αγγίζει εν πολλοίς τη μουσική αλληλεπίδραση, παρουσιάστηκαν οι δημοφιλέστερες μέθοδοι επίσης. Αναπτύχθηκε ο τρόπος λειτουργίας των αλγορίθμων προβολής υπογραφών, κυκλικού μετασχηματισμού Hough, εξαγωγής γεωμετρικών χαρακτηριστικών και χρωματικών δεικτών. Έγινε αναλυτική σύγκριση των μεθόδων, ενώ παρουσιάστηκαν τα δυνατά και τα αδύνατα σημεία της κάθε μιας.

Η δυναμική αναγνώριση κινήσεων-χειρονομιών έχει να επιδείξει μεθόδους όπως η Δυναμική Περιτύλιξη του Χρόνου, τα Νευρωνικά Δίκτυα Χρονικής Καθυστέρησης αλλά και τα Κρυφά Μαρκοβιανά Μοντέλα. Παρουσιάστηκε επιγραμματικά ο τρόπος λειτουργίας της κάθε μεθόδου. Στο κεφάλαιο 6 θα γίνει εκτενέστατη παρουσίαση των Κρυφών Μαρκοβιανών Μοντέλων.

Ενδιαφέρον παρουσιάζει η «συρραφή» όλων των παραπάνω εργαλείων. Στο κεφάλαιο που ακολουθεί θα παρουσιαστούν οι βασικοί άξονες μιας τέτοιας μεθόδου, σε ένα γενικευμένο πλαίσιο δίχως περιορισμούς ανάλογα με το πεδίο εφαρμογής.

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Η ανακάλυψη (1/4)

«Καμιά φορά είναι τέλεια
γιατί ανακαλύπτω
πράγματα που το αφεντικό
μου είχε ζητήσει.
Προφανώς, αυτό μπορεί να
γίνει οποιαδήποτε στιγμή,
και οι γονείς μου δεν είναι
πάντα χαρούμενοι.»



4 Μεθοδολογία για την αναγνώριση κινήσεων

- 4.1 Σύνοψη κεφαλαίου**
- 4.2 Ορισμός κίνησης**
- 4.3 Μοντελοποίηση κινήσεων**
 - 4.3.1 Χρονική μοντελοποίηση
 - 4.3.2 Χωρική μοντελοποίηση
- 4.4 Δημιουργία δερματικού μοντέλου**
- 4.5 Εντοπισμός περιοχής ενδιαφέροντος**
- 4.6 Επεξεργασία εικόνας**
 - 4.6.1 Τμηματοποίηση χεριού
 - 4.6.2 Εξαγωγή χαρακτηριστικών διανυσμάτων
- 4.7 Αναγνώριση κινήσεων**
- 4.8 Συμπεράσματα**

4.1 Σύνοψη κεφαλαίου

Η εξέλιξη των νέων τεχνολογιών και η ευρύτερη χρήση των υπολογιστών σε κάθε πεδίο ανθρώπινης δραστηριότητας απαιτούν νέες μεθόδους αλληλεπίδρασης μεταξύ ανθρώπου και μηχανής. Ο παραδοσιακός συνδυασμός χρήσης πληκτρολογίου και ποντικιού έχει αποδείξει τη χρησιμότητα, τους περιορισμούς και τις αδυναμίες του. Ο άνθρωπος έχει την ανάγκη να αλληλεπιδρά με τη μηχανή χρησιμοποιώντας πιο αποτελεσματικές και εκφραστικές μεθόδους. Οι μέθοδοι αυτές οφείλουν να είναι πιο κοντά στην επικοινωνία ανθρώπου με άνθρωπο (ΕΑΑ) [Picard 1997].

Είναι γεγονός ότι ο άνθρωπος χρησιμοποίησε τα χέρια του, ξεχωριστά ή σε συνδυασμό με άλλα μέρη του σώματός του, για να επικοινωνήσει, να αλληλεπιδράσει με το περιβάλλον του και να χειριστεί αντικείμενα. Κατά συνέπεια, είναι πολύ σημαντικό η τεχνολογία να επιτρέπει την αλληλεπίδραση του ανθρώπου με τις μηχανές χρησιμοποιώντας μεταξύ άλλων και παραδοσιακές μεθόδους αλληλεπίδρασης.

Ο επιστημονικός κλάδος της επικοινωνίας ανθρώπου υπολογιστή (EAY) εφηύρε μια πληθώρα εργαλείων με σκοπό να εκμεταλλευτεί τις ανθρώπινες χειρονομίες. Ένα τέτοιο εργαλείο είναι και η συσκευή Data Glove. Αυτή η συσκευή, όπως και άλλες παρόμοιες, μπορούν να συνδράμουν στη ανάπτυξη χρήσιμων και ενδιαφέροντων εφαρμογών. Πλην όμως αποδεικνύονται συχνά δύσκολες, σύνθετες και ενοχλητικές στη χρήση τους.

Η χρήση της όρασης υπολογιστή είναι μια εναλλακτική πρόταση. Η πρόοδος των τεχνικών υπολογιστικής όρασης σε συνδυασμό με τη διαθεσιμότητα μεγάλης υπολογιστικής ισχύος καθιστούν εφικτή την αναγνώριση χειρονομιών στην EAY για εφαρμογές πραγματικού χρόνου. Ως συνέπεια αυτού, εκτεταμένες έρευνες έγιναν στον τομέα της όρασης υπολογιστή για την αναγνώριση στάσεων του χεριού και στατικών χειρονομιών, αλλά και εντελώς πρόσφατα, για την ερμηνεία της δυναμικής των χειρονομιών [Kohler 2004; Pavlovic 1997]. Τα συστήματα όρασης υπολογιστή είναι πιο φιλικά προς τον χρήστη μιας και χρησιμοποιούν βιντεοκάμερες για την καταγραφή κινήσεων και βασίζονται σε εφαρμογές λογισμικού για να διεξάγουν ανάλυση.

Στο κεφάλαιο αυτό παρουσιάζεται ένα μοντέλο επεξεργασίας κινήσεων χρησιμοποιώντας όραση υπολογιστή. Στην ενότητα 4.2 παρουσιάζεται η δυσκολία να οδηγηθούμε σε μοναδικό ορισμό για την κίνηση υπό την έννοια της χειρονομίας. Προτείνεται ορολογία η οποία θα χρησιμοποιηθεί στη συνέχεια της διατριβής. Οι επόμενες παράγραφοι προσδιορίζουν μια μεθοδολογία για την αναγνώριση δακτυλικών κινήσεων, βασισμένη στα παρακάτω βήματα:

1. Μοντελοποίηση κινήσεων,
2. Δημιουργία δερματικού μοντέλου,
3. Εντοπισμός περιοχής ενδιαφέροντος,
4. Επεξεργασία εικόνας,
5. Αναγνώριση κινήσεων.

Καταρχήν υλοποιείται η αναπαράσταση της κίνησης στο χώρο και το χρόνο όπου και μοντελοποιείται. Στη συνέχεια δημιουργείται το δερματικό μοντέλο με χρήση στατιστικής μέσα από μια βιβλιοθήκη εικόνων. Με βάση αυτό το μοντέλο εντοπίζεται η περιοχή ενδιαφέροντος σε οποιαδήποτε άλλη εικόνα. Το βήμα της επεξεργασίας εικόνας α-

φορά στον αλγόριθμο μετατροπής της εικόνας σε δεδομένα τα οποία αντιστοιχίζονται στο μοντέλο που επιλέχτηκε προηγουμένως. Το βήμα της αναγνώρισης αντιπαραθέτει τα δεδομένα με το μοντέλο με στόχο να τα συνδέσει με μια κίνηση. Τέλος, στα συμπεράσματα, θα επεξηγηθεί ο τρόπος με τον οποίο αυτή η μεθοδολογία μπορεί να χρησιμοποιηθεί στο πρόβλημα ανάκτησης δακτυλισμών και αναγνώρισης κινήσεων δακτύλων του πιανίστα.

4.2 Ορισμός κίνησης

Παρά το γεγονός ότι η ανθρώπινη κίνηση μελετάται από πολλούς επιστημονικούς κλάδους, είναι μάλλον αδύνατο να δοθεί μοναδικός απλός ορισμός γι' αυτήν [Cadoz 2000].

«Η κίνηση είναι μια σωματική μετατόπιση η οποία επιτρέπει τη συμπλήρωση, επεξήγηση και ενίσχυση του λόγου στην ανθρώπινη επικοινωνία. Μπορεί επίσης να χρησιμοποιηθεί ως χειρονομία, προκειμένου να αντικαταστήσει το λόγο. Διακρίνονται διάφορες κατηγορίες σωματικών κινήσεων. Κάποιες κατηγορίες μεταφέρουν μια επιπρόσθετη πληροφορία πάνω στο κυρίως μήνυμα. Άλλες μεταφέρουν ενδείξεις για τις προθέσεις του ομιλητή. Οι κινήσεις-χειρονομίες αποτελούν ένα σημαντικό τομέα της μη-βερμπαλιστικής επικοινωνίας.», σύμφωνα με την ηλεκτρονική εγκυκλοπαίδεια Wikipedia (<http://fr.wikipedia.org>).

Πάντως, μέσα από οποιονδήποτε ορισμό αναδεικνύεται το κοινό σημείο αναφοράς που είναι η ανθρώπινη φυσική συμπεριφορά. Αναδεικνύεται επίσης το γεγονός ότι η κίνηση είναι άρρηκτα συνδεδεμένη με τη χειρονομία, αφού η πρώτη αποτελεί έκφρασης της δεύτερης. Υπάρχουν σημεία στα οποία οι διαφορετικοί ορισμοί της κίνησης δεν τέμνονται. Ένα από αυτά εντοπίζεται στη διάσταση απόψεων για το αν η παραγόμενη κίνηση που περικλείει μια χειρονομία οφείλει να μεταδίδει πληροφορία, ή εναλλακτικά, αρκεί να παράγεται από διαχειριστικό ή εκφραστικό κίνητρο. Στον κλάδο της ΕΑΥ, όπου ο στόχος είναι ο έλεγχος του υπολογιστή με τη χρήση χειρονομιών, είναι επιθυμητό να χρησιμοποιούνται κινήσεις του σώματος και των άκρων, οι οποίες μιμούνται διαχειριστικές και επικοινωνιακές χειρονομίες.

Ο κλάδος της μουσικής τεχνολογίας ορίζει τη χειρονομία ως την φυσική αναπαράσταση της επικοινωνίας ανάμεσα στον μουσικό

και το όργανο. Είναι ο τρόπος με τον οποίο η κίνηση του σώματος του μουσικού, καθώς και η εκφραστικότητά του, μεταδίδονται στο όργανο και μετατρέπονται σε μηχανική ενέργεια. Αποτελεί το βασικό κορμό της παραγωγής ήχου, ενώ παράλληλα πλαισιώνεται από συναίσθημα και συγκίνηση. Μπορεί επομένως η χειρονομία να χωριστεί σε δύο κατηγορίες κατά Delalande: *ενεργός* και *συνοδευτική* χειρονομία (*effective and accompanist gesture*) [Delalande 1988]. Η ενεργός χειρονομία είναι απαραίτητη για την παραγωγή του ήχου. Αποτελεί τη δράση που πραγματοποιείται από τον μουσικό προκειμένου να παράξει μηχανική ενέργεια η οποία θα επιτρέψει στο μουσικό όργανο να «γεννήσει» τον ήχο. Από την άλλη μεριά, η συνοδευτική χειρονομία, η οποία αποκαλείται και ως εκφραστική ή αρθρωτή χειρονομία, χρησιμοποιείται από τον μουσικό για να εκφράσει το συναισθηματικό περιεχόμενο του μουσικού έργου.

Έρευνες απέδειξαν ότι παρότι οι συνοδευτικές χειρονομίες δεν αποτελούν κινητήριο δύναμη για την παραγωγή του ήχου, εντούτοις, περιέχουν ουσιώδη μουσική πληροφορία [Wanderley 2005]. Χαρακτηριστικό παράδειγμα που αποδεικνύει τον ισχυρισμό του Wanderley είναι η διεύθυνση ορχήστρας (*conducting*). Πρόκειται για μία μέθοδο ελέγχου της εκτέλεσης πολλών μουσικών οργάνων δια μέσου συνοδευτικών χειρονομιών του διευθυντή ορχήστρας, χωρίς άμεση επαφή με τα μουσικά όργανα. Στο διαδραστικό περιβάλλον διευθυντή ορχήστρας και μουσικών, η οπτική πληροφορία που εκλαμβάνεται από τους μουσικούς χρησιμοποιείται ως μέσο μεταβίβασης οδηγιών για ενεργές χειρονομίες [Tobey 1996]. Οι χειρονομίες του μαέστρου μεταφέρουν πληροφορίες κυρίως σχετικά με το τέμπο, την εκφραστικότητα και τον συντονισμό των μουσικών. Οι πληροφορίες για το ακριβές περιεχόμενο του προς εκτέλεση μουσικού έργου προέρχονται από τη μουσική σημειογραφία (*παρτιτούρα*). Ως εκ τούτου, μια οποιαδήποτε μέθοδος όρασης υπολογιστή οφείλει να λάβει υπόψη της και τις δύο κατηγορίες χειρονομίας.

Μελετώντας τη χειρονομία από τη σκοπιά της EAY, η χειρονομία ορίζεται ως μια κίνηση η οποία γεννάται από κάποιον άνθρωπο και γίνεται αντιληπτή από παρατηρητή ως ροή από εικόνες. Οι εικόνες αυτές ερμηνεύονται στη συνέχεια από τον παρατηρητή ανάλογα με το εκφραστικό και επικοινωνιακό τους περιεχόμενο. Στην περίπτωση της όρασης υπολογιστή, ο παρατηρητής είναι ένας υπολογιστής ο οποίος καταγράφει τη χειρονομία με μία ή περισσότερες κάμερες. Η ροή εικόνων (*ακολουθία καρτέ του βίντεο*) αναλύεται λαμβάνοντας υπόψη τις χωρικές παραμέτρους του μοντέλου κίνησης, σε καθορισμένο χρονικό διάστημα. Στη συνέχεια, γίνεται σύγκριση

των εξαγομένων παραμέτρων με τις παραμέτρους του μοντέλου ώστε να ταυτοποιηθεί η κίνηση. Ως τελικό στάδιο έχουμε τη μετατροπή των αναγνωρισμένων κινήσεων σε εντολές ή δεδομένα ώστε να ελέγξει ο υπολογιστής κάποια εφαρμογή ή λογισμικό.

4.3 Μοντελοποίηση κινήσεων

Στην ενότητα 4.2 παρουσιάστηκε η αδυναμία να δοθεί ένας ενιαίος ορισμός αναφορικά με τις έννοιες της χειρονομίας και της κίνησης. Κατά συνέπεια η έλλειψη γενικού ορισμού για την κίνηση έχει ως επακόλουθο την απουσία ενός γενικού μοντέλου το οποίο θα επιλύει όλα τα προβλήματα αναγνώρισης κινήσεων στην ΕΑΥ. Παρακάτω θα παρουσιαστεί μια μεθοδολογία για τη μοντελοποίηση κινήσεων των δακτύλων για πιάνο με τη χρήση όρασης H/Y.

Οι εικόνες που ανακτώνται από το βίντεο χρησιμοποιούνται ώστε να εντοπιστεί η θέση του χεριού και η χρονική στιγμή που η κίνηση πραγματοποιήθηκε. Πρόκειται για το χωρικό και το χρονικό μοντέλο αντίστοιχα. Η κίνηση θα τμηματοποιηθεί στη συνέχεια με βάση τα δύο παραπάνω μοντέλα (Σχήμα 4.3.1).

Ακολουθεί το σκέλος της εξαγωγής χαρακτηριστικών από την τμηματοποιημένη εικόνα, το οποίο βασίζεται σε ακολουθία καταστάσεων, οι οποίες αφορούν τη φάση της ηρεμίας, της προετοιμασίας και της δράσης της κίνησης. Ακολουθίες τέτοιων καταστάσεων συνθέτουν μια κίνηση σε συνάρτηση με τα δύο μοντέλα.

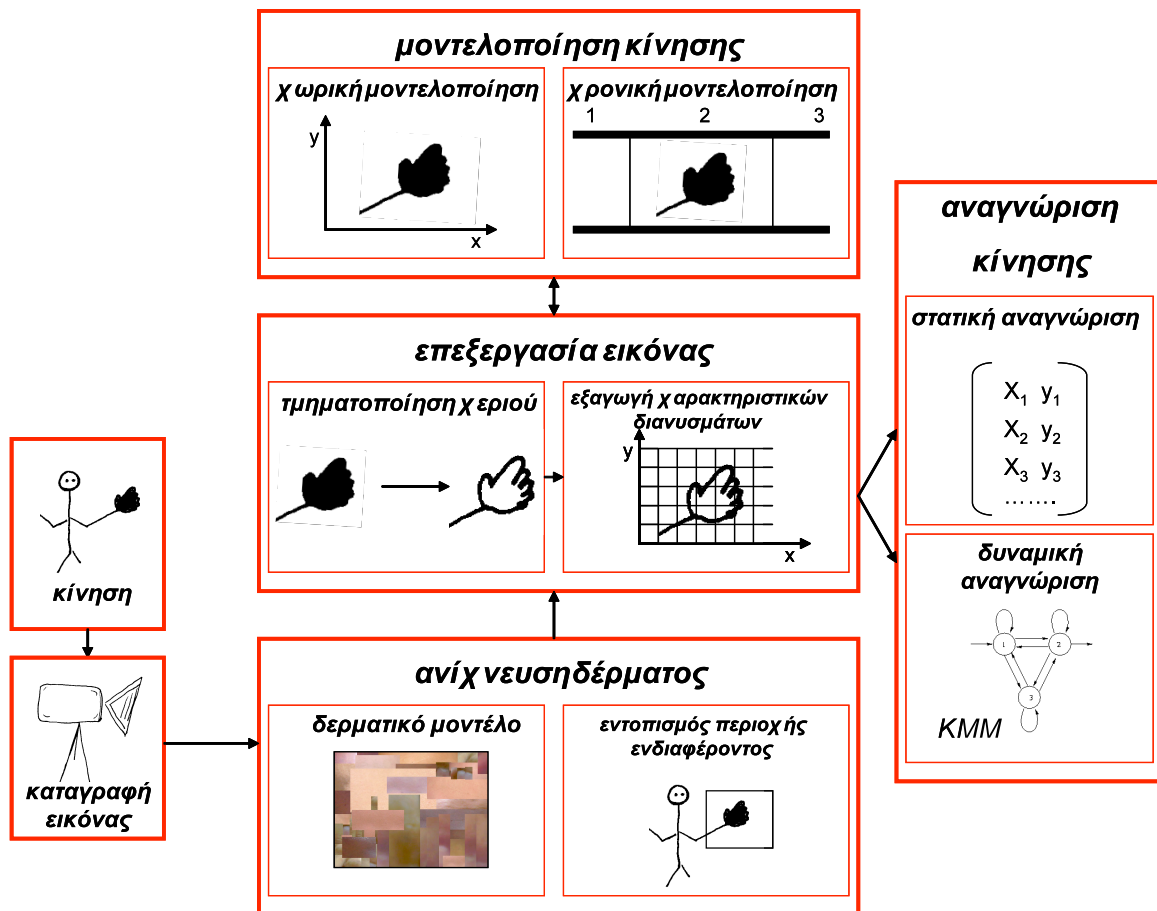
4.3.1 Χρονική μοντελοποίηση

Η κίνηση είναι μια δυναμική ενέργεια που εμπεριέχει χρονικά χαρακτηριστικά με σημαντική πληροφορία. Παρόλα αυτά, σημαντικός όγκος μελετών που διεξήχθησαν στο παρελθόν αναγνώριζαν στατικές χειρονομίες ή στάσεις.

Σημαντική προϋπόθεση για την ανάλυση των δυναμικών συστατικών της κίνησης είναι η κατανόηση της φυσιολογίας της. Οι κινήσεις μπορούν να χωριστούν σε τρεις φάσεις, α) την *ηρεμία*, β) την *προετοιμασία* και γ) τη *δράση*.

Η φάση της ηρεμίας της κίνησης αποτελείται από μια ακολουθία καταστάσεων ηρεμίας, οι οποίες ανιχνεύονται σε γειτονικά καρέ για κάθε δάκτυλο. Η φάση της προετοιμασίας αναφέρεται σε μια ακολουθία αντίστοιχων καταστάσεων όπου το δάκτυλο ετοιμάζεται να ενεργήσει. Τέλος, η φάση της δράσης αποτελεί τον πυρήνα της κίνησης αλλά και τον κυρίως στόχο της αναγνώρισης. Πολλές

φορές είναι δύσκολη η ανίχνευση των επιμέρους φάσεων λόγω της συνθετότητας της κίνησης.



Σχήμα 4.3.1 Μεθοδολογία αναγνώρισης δακτυλικών κινήσεων

4.3.2 Χωρική μοντελοποίηση

Δύο μεγάλες κατηγορίες χωρικών μοντελοποιήσεων της κίνησης καταγράφονται από τους ερευνητές. Αυτές είναι: α) η μοντελοποίηση με βάση την εμφάνιση και β) η μοντελοποίηση με βάση το σκελετόγραμμα.

Στην περίπτωση της αναγνώρισης δακτυλικών κινήσεων, με εφαρμογή στη MA, η μοντελοποίηση με βάση το σκελετόγραμμα προσφέρει αφενός πρόσβαση σε πληροφορία υψηλής ακρίβειας, η οποία μπορεί ενδεχομένως να μη χρειάζεται, αφετέρου παρατηρείται δραματική σπατάλη υπολογιστικής ισχύος [Pavlovic 1997]. Οι δύο

παραπάνω λόγοι καθιστούν απαγορευτική τη χρήση της σε πολλές περιπτώσεις.

Η μοντελοποίηση κινήσεων με βάση την εμφάνιση κάνει χρήση χαρακτηριστικών της εικόνας τα οποία και συγκρίνει με τα αντίστοιχα χαρακτηριστικά προκαθορισμένων δειγμάτων κινήσεων. Τα χαρακτηριστικά αυτά ονομάζονται *χαρακτηριστικά διανύσματα της εικόνας*. Τα χαρακτηριστικά διανύσματα καθορίζονται ανάλογα με το είδος της πληροφορίας που πρέπει να ανακτηθεί. Μπορεί να αφορούν σε συντεταγμένες θέσης του χεριού, των δακτύλων κλπ.

Τα δείγματα των κινήσεων μπορούν να προκαθοριστούν «συμψηφίζοντας» τα χαρακτηριστικά διανύσματα της εικόνας για ένα σύνολο δεδομένων το οποίο αναπαριστά τις προς αναγνώριση κινήσεις, καθώς και τις επιμέρους φάσεις τους. Η προσέγγιση αυτή λειτουργεί θετικά και στην περίπτωση δυναμικής αναγνώρισης κινήσεων όπου συμμετέχει και η τροχιά του κινούμενου αντικειμένου. Ειδικότερα στην περίπτωση των δακτυλικών κινήσεων, καλά και αντιπροσωπευτικά τους μοντέλα μπορούν να παραχθούν αν ληφθούν υπόψη χαρακτηριστικά που αφορούν στα άκρα τους. Τα χαρακτηριστικά των άκρων αποτελούν από μόνα τους μια επαρκή τεχνική περιγραφής ενός συνόλου κινήσεων [Pavlovic 1997].

4.4 Δημιουργία δερματικού μοντέλου

Η ενότητα της αναγνώρισης κινήσεων που αφορά στον εντοπισμό θέσης του κινούμενου αντικειμένου μπορεί να υλοποιηθεί με δύο μεθόδους: α) την *ανίχνευση δέρματος* και β) την *ανίχνευση κίνησης*.

Η μέθοδος της ανίχνευσης κίνησης βασίζεται στην ιδέα ότι το φόντο της εικόνας, οι κάμερες και ο άνθρωπος δεν κινούνται για τη μεγαλύτερη χρονική διάρκεια και ότι το μοναδικό στοιχείο που κινείται είναι το μέλος του ανθρώπινου σώματος του οποίου την κίνηση θέλουμε να αναγνωρίσουμε. Εφαρμόζονται αφαιρετικοί αλγόριθμοι μεταξύ, μιας εικόνας που αποτελεί σημείο αναφοράς και της τρέχουσας εικόνας, ώστε να εντοπιστεί η περιοχή όπου ανιχνεύτηκε κίνηση μέσω των διαφορών του ζεύγους εικόνων. Η μέθοδος αυτή επηρεάζεται από φαινόμενα σκίασης ενώ λειτουργεί καλά σε περιπτώσεις όπου κινείται ένα μόνο αντικείμενο στην εικόνα [Pavlovic 1997]. Κατά συνέπεια είναι προφανές ότι η ανίχνευση κίνησης δεν ενδείκνυται για τον εντοπισμό θέσης πολλών ταυτόχρονα κινούμενων δακτύλων.

Από την άλλη μεριά, έχει αποδειχτεί ότι το χρώμα του δέρματος είναι ένα ισχυρό και χρήσιμο εργαλείο, με εξαιρετικά αποτελέσματα στην αναγνώριση προσώπου, στον εντοπισμό και την ιχνηλά-

τηση μελών του σώματος. Πολλές τεχνικές για τη μοντελοποίηση και αναγνώριση του ανθρώπινου δέρματος έχουν προταθεί τα τελευταία χρόνια. Πολλές έρευνες έχουν διεξαχθεί, συγκρίνοντας τις διαφορετικές τεχνικές οι οποίες έχουν δημοσιευτεί [Zarit 1999; Terrillonet 2000; Brand 2000].

Στόχος είναι η κατασκευή ενός δερματικού μοντέλου, το οποίο θα αποτελεί πρότυπο για την ανίχνευση δερματικών περιοχών σε οποιαδήποτε εικόνα.

Η διαδικασία ξεκινάει από τη συλλογή εικόνων με ανθρώπους αποσκοπώντας στη δημιουργία μιας βιβλιοθήκης εικόνων. Δύο είναι οι καθοριστικοί παράγοντες που αφορούν στη βέλτιστη χρησιμότητα της βιβλιοθήκης:

- Όσο μεγαλύτερος είναι ο αριθμός των εικόνων τόσο καλύτερη θα είναι η ποιότητα του μοντέλου,
- Όσο επιλέγονται καλές και αντιπροσωπευτικές εικόνες (με δερματικές περιοχές), τόσο καλύτερα θα είναι τα εξαγόμενα αποτελέσματα.

Η μείωση των χρωμάτων των εικόνων είναι ένα βήμα προαιρετικό και εξαρτάται από το χώρο εφαρμογής του αλγορίθμου και από το είδος του λογισμικού που θα αναπτυχθεί. Εάν για παράδειγμα πρόκειται για χρήση του αλγορίθμου σε έναν Η/Υ, ο οποίος θα λειτουργεί τοπικά, για εφαρμογή σε σύνολο διακριτών εικόνων, τότε η μείωση του αριθμού χρωμάτων των εικόνων δε θα μειώσει αισθητά την απαιτούμενη υπολογιστική ισχύ. Εάν πρόκειται για εφαρμογή βίντεο που θα λειτουργεί σε πραγματικό χρόνο τότε σίγουρα η μείωση των χρωμάτων θα βελτιώσει την ταχύτητα εκτέλεσης του αλγορίθμου. Για παράδειγμα, μπορεί μια εικόνα στον 3Δ χώρο RGB, από 256x256x256 πλήθος χρωμάτων να μειωθεί σε 32x32x32, ανάλογα με το χώρο εφαρμογής του αλγορίθμου. Η ίδια διαδικασία μπορεί να εφαρμοστεί σε οποιονδήποτε χρωματικό χώρο.

Στη συνέχεια εφαρμόζεται δειγματοληπτική χειροκίνητη ανάκτηση περιοχών δέρματος από κάθε μια εικόνα της βιβλιοθήκης και σταδιακή κατασκευή του δερματικού μοντέλου. Στην πράξη, το μοντέλο θα δημιουργηθεί από μια νέα εικόνα η οποία προφανώς αποτελείται αποκλειστικά από εικονοστοιχεία που περιέχουν περιοχές δέρματος. Καλό είναι αυτές οι περιοχές δέρματος να αποτελούνται από εικονοστοιχεία διαφορετικής φωτεινότητας, γωνίας λήψης κλπ, ώστε να αποκτηθεί όσο το δυνατόν πληρέστερη δειγματοληψία.

Διάφοροι χρωματικοί χώροι μπορούν να χρησιμοποιηθούν για να εντοπιστεί η περιοχή ενδιαφέροντος σε οποιαδήποτε άλλοι εικόνα. Μια πολύ δημοφιλής λύση είναι η χρήση του κανονικοποιημένου RGB, ή αλλιώς *rg*. Ο 3Δ χρωματικός χώρος RGB προβάλλεται στις δύο διαστάσεις κρατώντας μόνο την πληροφορία που αφορά τη χρωματικότητα και αγνοώντας αυτήν της φωτεινότητας. Στη συνέχεια διενεργείται ιστόγραμμα χρώματος. Το μοντέλο που θα χρησιμοποιηθεί είναι το ιστόγραμμα, το οποίο δεν είναι τίποτα άλλο από έναν 2Δ πίνακα.

4.5 Εντοπισμός περιοχής ενδιαφέροντος

Υπάρχουν αρκετές μέθοδοι για τον εντοπισμό της περιοχής ενδιαφέροντος (*ανίχνευση δέρματος*) μέσα σε μια εικόνα. Μια καλή μέθοδος βασίζεται στον υπολογισμό της πιθανότητας κάθε εικονοστοιχείο να έχει χρώμα δέρματος. Η απόφαση για το αν ένα εικονοστοιχείο ανήκει σε δερματική περιοχή εξαρτάται από ένα κατώφλι.

Πρέπει να σημειωθεί ότι για τιμές πιθανοτήτων κοντά στο κατώφλι υπάρχουν διάφορες στρατηγικές που μπορούν να χρησιμοποιηθούν προαιρετικά, οι οποίες αναπτύσσονται αναλυτικά σε συγκριτική μελέτη του Vezhnevets (2003). Για παράδειγμα, για τα εικονοστοιχεία με πιθανότητα κοντά στο κατώφλι, μπορεί να εξεταστεί η αντίστοιχη πιθανότητα των γειτονικών τους, τα οποία συνήθως ανήκουν στο περίγραμμα των δερματικών περιοχών. Η επιλογή εξαρτάται από την ποιότητα της πληροφορίας που απαιτείται για ανάκτηση [Vezhnevets 2003].

4.6 Επεξεργασία εικόνας

Η ενότητα της επεξεργασίας εικόνας αναφέρεται στη διαδικασία τμηματοποίησης της εικόνας και εξαγωγής των χαρακτηριστικών διανυσμάτων της. Εφαρμόζεται σε κάθε καρέ του βίντεο χωριστά, από όπου θα πρέπει να αφαιρεθεί εντελώς το χέρι και τα δάκτυλα από το φόντο, να αφαιρεθεί ο θόρυβος και στη συνέχεια να εξαχθούν τα χαρακτηριστικά διανύσματα του χεριού όπως είναι οι συντεταγμένες θέσης του κάθε δακτύλου, η τροχιά και άλλα.

4.6.1 Τμηματοποίηση χεριού

Από τη στιγμή που η περιοχή ενδιαφέροντος έχει εντοπιστεί πάνω σε κάθε καρτέ του βίντεο, με βάση το δερματικό μοντέλο, το επόμενο βήμα είναι να προετοιμαστεί η εικόνα ώστε να μπορούν να εξαχθούν τα χαρακτηριστικά διανύσματα που αναφέρονται στο χωρικό μοντέλο.

Η τμηματοποίηση του χεριού και των δακτύλων στην εικόνα είναι η πιο σύνθετη και δύσκολη διαδικασία στην αναγνώριση κινήσεων. Η προσέγγιση τμηματοποίησης της εικόνας μπορεί να υλοποιηθεί με τον εντοπισμό θέσης των άκρων των δακτύλων με βάση χρωματικούς δείκτες. Οι δείκτες μπορεί να είναι για παράδειγμα χρωματιστά αυτοκόλλητα, τα οποία τοποθετούνται στην περιοχή των νυχιών με διαφορετικό χρώμα για κάθε δάκτυλο. Έτσι ανιχνεύεται και ταυτοποιείται το κάθε δάκτυλο χωριστά με βάση το χρώμα των δεικτών. Ως τεχνική αποδεικνύεται γρήγορη και αποτελεσματική για αναγνώριση δακτυλικών κινήσεων εκτός ΜΑ, όπως για παράδειγμα κινήσεις πάνω στο πληκτρολόγιο του Η/Υ. Στη ΜΑ όμως, οι χρωματικοί δείκτες περιορίζουν το προσωπικό στυλ μουσικής ερμηνείας του πιανίστα, ενώ απαιτούν προηγουμένως ειδική προετοιμασία.

Μια πιο σύνθετη μέθοδος, η οποία δεν περιορίζει με ειδικό εξοπλισμό το χρήστη, ή τον μουσικό αντίστοιχα, είναι η τεχνική της σύμπτωσης προτύπων όπου το δείγμα μπορεί να είναι μια εικόνα με τα άκρα των δακτύλων ή ένα μοντέλο κίνησης των άκρων. Επίσης μπορούν να χρησιμοποιηθούν χαρακτηριστικές ιδιότητες των δακτύλων, όπως είναι η καμπυλότητα των άκρων, με σκοπό τον εντοπισμό τους στην ακολουθία των καρτέ. Και σε αυτήν την περίπτωση όμως παρουσιάζονται εμπόδια στον εντοπισμό, όπως για παράδειγμα όταν τα άκρα έρχονται πολύ κοντά το ένα στο άλλο η ιδιότητα της καμπυλότητας γίνεται ανεπαρκής. Ως αντίδοτο σε αυτό το εμπόδιο επιλέγεται είτε η χρήση περισσότερων καμερών, είτε η εφαρμογή αλγορίθμων εκτίμησης θέσης για τα δάκτυλα που βρίσκονται σε εγγύτητα.

Η πιο γενική μέθοδος για ανίχνευση κινούμενων αντικειμένων είναι η χρήση μασκών. Η μέθοδος αυτή, με τη βοήθεια του δερματικού μοντέλου, μπορεί να αποφέρει θετικά αποτελέσματα και στην αναγνώριση κινούμενων δακτύλων χωρίς τη χρήση ειδικού εξοπλισμού ή χρωματικών δεικτών. Με την εφαρμογή του δερματικού μοντέλου εντοπίζεται η περιοχή του χεριού στην εικόνα εξάγοντας ως αποτέλεσμα μια δυαδική μάσκα η οποία θα αποδίδει: α) με *άσπρο χρώμα* τα εικονοστοιχεία που θεωρήθηκαν δέρμα, β) με *μαύρο χρώμα* όλα τα υπόλοιπα εικονοστοιχεία. Εφόσον όμως το μοντέλο αυτό εί-

ναι ατελές, συναντάται θόρυβος στην εικόνα, ο οποίος προκύπτει εξαιτίας του ότι κάποια εικονοστοιχεία θεωρήθηκαν δερματικές περι-οχές, ενώ δεν ήταν. Για το λόγο αυτό, προτείνεται η χρήση φίλτρων για τη μείωση του θορύβου. Η κατηγορία των μορφολογικών φίλτρων και φίλτρων μη γραμμικής τάξης, μπορεί να επιδράσει διπλά πάνω στην εικόνα: α) μειώνοντας, έως και αφαιρώντας εντελώς, τον περιττό θόρυβο, β) διαχωρίζοντας ακόμα καλύτερα τα δάκτυλα, αυξάνοντας την απόσταση μεταξύ τους. Στη συνέχεια, η εφαρμογή χαμηλοπερατού φίλτρου μπορεί να οδηγήσει σε ελαφρό θάμπωμα της εικόνας και άρα σε πιο ομαλό περίγραμμα του χεριού. Τέλος, η σταδιακή εφαρμογή κατωφλίωσης μπορεί να οδηγήσει στην εξαγωγή της σιλουέτας του χεριού, από την οποία με κατάλληλη επεξεργασία, μέσα από φίλτρα ανίχνευσης ακμών στην εικόνα, μπορεί να ανακτηθεί το περίγραμμα του χεριού.

4.6.2 Εξαγωγή χαρακτηριστικών διανυσμάτων

Η εξαγωγή των χαρακτηριστικών διανυσμάτων αποτελεί το τελευταίο βήμα της ενότητας της επεξεργασίας εικόνας. Είναι το βήμα όπου το κάθε καρέ του βίντεο μετατρέπεται σε μορφή αντίστοιχη του μοντέλου ώστε να μπορούν να συγκριθούν.

Ο στόχος του βήματος είναι η επιλογή των κατάλληλων χαρακτηριστικών μέσα από ένα σύνολο πληροφοριών που αφορούν τα εικονοστοιχεία και τα οποία μπορούν να ανακτηθούν μέσα από δυαδικές εικόνες, εικόνες περιγράμματος ή ακμών. Στην περίπτωση της μοντελοποίησης με βάση την εμφάνιση, τα χαρακτηριστικά διανύσματα μπορεί να είναι οι συντεταγμένες κάποιων σημείων, για παράδειγμα των άκρων των δακτύλων, ή γωνίες που σχηματίζονται από ευθύγραμμα τμήματα. Τα χαρακτηριστικά διανύσματα αυτά θα συγκριθούν με το μοντέλο για να εντοπιστεί σε ποια προκαθορισμένη κίνηση ταιριάζει η ακολουθία των καρέ. Έτσι, υπολογίζονται κατά προσέγγιση και τυχόν διανύσματα που θεωρούνται ότι απουσιάζουν, όπως για παράδειγμα στην περίπτωση που κάποιο δάκτυλο κρύβεται από κάποιο άλλο.

4.7 Αναγνώριση κινήσεων

Η αναγνώριση της κίνησης είναι το τελευταίο στάδιο ενός ολοκληρωμένου συστήματος αναγνώρισης κινήσεων, όπου τα δεδομένα των χαρακτηριστικών διανυσμάτων αναλύονται και εξαγάγουν την κίνηση.

Η ταυτοποίηση του δακτύλου που χρησιμοποιήθηκε για την κίνηση μπορεί να υλοποιηθεί μέσα από τη σύγκριση των χαρακτηρι-

στικών διανυσμάτων με το χωρικό μοντέλο. Ειδικότερα, στην περίπτωση της ανάκτησης δακτυλισμών στο πιάνο, στόχος είναι η εύρεση του δακτύλου που χρησιμοποιήθηκε για να πατηθεί το εκάστοτε πλήκτρο. Μπορεί να συγκριθεί η συντεταγμένη θέσης του δακτύλου (x,y) , μέσω του χαρακτηριστικού διανύσματος, με το χωρικό μοντέλο ώστε να ανακτηθεί η φάση που βρίσκεται η κίνηση του κάθε δακτύλου. Η περίπτωση αυτή ανήκει στην κατηγορία της *στατική αναγνώριση κινήσεων* όπου κάθε κίνηση θεωρείται ανεξάρτητη από τις υπόλοιπες.

Στη *δυναμική αναγνώριση κινήσεων*, η μοντελοποίηση των χαρακτηριστικών διανυσμάτων της κίνησης, μπορεί να γίνει με χρήση Κρυφών Μαρκοβιανών Μοντέλων (KMM). Η κάθε κίνηση αποτελεί και ένα KMM. Μέσα από τις ακολουθίες των καρέ του βίντεο εξάγονται ακολουθίες παρατήρησης, οι οποίες εξετάζεται αν αποτελούν κάποια κίνηση που το σύστημα μπορεί να αναγνωρίσει. Το κριτήριο είναι η πιθανότητα που εμφανίζει κάθε KMM να έχει παράξει την ακολουθία παρατήρησης. Η ακολουθία παρατήρησης θεωρείται ότι ανήκει στην κίνηση της οποίας το KMM έχει τη μεγαλύτερη πιθανότητα. Στην περίπτωση που κάποια ακολουθία παρατήρησης παρουσιάζει μικρές πιθανότητες για όλα τα μοντέλα, τότε η κίνηση δεν ανήκει στο σύνολο των κινήσεων που αναγνωρίζει το συγκεκριμένο σύστημα.

4.8 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε μια μεθοδολογία για την επίλυση γενικότερων προβλημάτων αναγνώρισης κινήσεων, συνοδευόμενη από την αντίστοιχη ορολογία της. Στα επόμενα κεφάλαια, η μεθοδολογία αυτή θα εφαρμοστεί και θα αξιολογηθεί στην αναγνώριση κινήσεων και την ανάκτηση δακτυλισμών στη ΜΑ.

Το πρόβλημα της αναγνώρισης κινήσεων των χεριών του πιανίστα ακολουθεί τα ίδια βήματα με γενικότερα προβλήματα αναγνώρισης δακτυλικών κινήσεων. Μοντέλα για τη χρονική τμηματοποίηση της κίνησης σε επιμέρους φάσεις (*ηρεμία, προετοιμασία, δράση*), αλλά και τη χωρική τμηματοποίηση της κίνησης (*δάκτυλο, πλήκτρο*) απαιτούνται και στην αναγνώριση κινήσεων στη ΜΑ. Υλοποιείται ανίχνευση δέρματος και εντοπισμός περιοχής ενδιαφέροντος αλλά και επεξεργασία του κάθε καρέ του βίντεο με τμηματοποίηση χεριού και εξαγωγή χαρακτηριστικών διανυσμάτων. Με βάση τα χαρακτηριστικά διανύσματα, και μόνον αυτά, πραγματοποιείται στατική αναγνώριση κινήσεων (*ανάκτηση δακτυλισμών και ταυτοποίηση*

πλήκτρου). Για δυναμική αναγνώριση κινήσεων χρησιμοποιούνται KMM και λαμβάνεται υπόψη η παράμετρος της αλληλουχίας μεταξύ των κινήσεων. Έτσι, η κίνηση (*συνδυασμός χρήσης δακτύλων*) αναγνωρίζεται και ταυτοποιείται μέσα στις ακολουθίες καρέ του βίντεο.

Στο επόμενο κεφάλαιο αναπτύσσεται πλήρως το σύστημα αναγνώρισης δακτυλικών κινήσεων στη μουσική αλληλεπίδραση. Περιγράφεται βήμα-βήμα η δημιουργία του καθώς και τα θεωρητικά εργαλεία στα οποία στηρίζεται το σύστημα. Το κεφάλαιο αυτό αναφέρεται στη στατική αναγνώριση κινήσεων ή αλλιώς ανάκτηση δακτυλισμών.

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Η ανακάλυψη (2/4)

« Διερωτώνται μήπως
τρελαίνομαι σιγά σιγά,
αλλά η μαμά ξέρει ότι δεν
αρέσει στον μπαμπά να το
ακούει. »



5 Σύστημα στατικής αναγνώρισης δακτυλικών κινήσεων

5.1 Σύνοψη κεφαλαίου

5.2 Επιλογή εξοπλισμού

- 5.2.1 Ηλεκτρονικός υπολογιστής
- 5.2.2 Πιάνο-συνθεσάιζερ
- 5.2.3 Κάμερα
- 5.2.4 Λογισμικό

5.3 Προκαταρκτική ανάλυση

- 5.3.1 Τοποθέτηση κάμερας
- 5.3.2 Γωνία λήψης
- 5.3.3 Φωτισμός
- 5.3.4 Σύλληψη βίντεο

5.4 Ανίχνευση δέρματος

- 5.4.1 Χαρακτηριστικά ανθρώπινου δέρματος
- 5.4.2 Προσδιορισμός περιοχής ενδιαφέροντος
- 5.4.3 Κανονικοποίηση περιοχής ενδιαφέροντος
- 5.4.4 Δημιουργία δερματικού μοντέλου
- 5.4.5 Χρήση μοντέλου για ανίχνευση δέρματος σε εικόνα

5.5 Τμηματοποίηση χεριού

- 5.5.1 Απλοποίηση δυαδικής μάσκας
 - 5.5.1.1 Μείωση θορύβου με σειριακό εναλλασσόμενο φιλτράρισμα
 - 5.5.1.2 Ανάκτηση φιγούρας χεριού
- 5.5.2 Διαχωρισμός χεριού σε βασικές συνιστώσες
 - 5.5.2.1 Ανάκτηση περιγράμματος χεριού
 - 5.5.2.2 Ανάκτηση άκρων δακτύλων

5.6 Εντοπισμός δακτύλων

- 5.6.1 Υπολογισμός κεντροειδούς
- 5.6.2 Καθορισμός σημείου αναφοράς
- 5.6.3 Εντοπισμός θέσης

5.7 Ανάκτηση δακτυλισμών

5.8 Συμπεράσματα

5.1 Σύνοψη κεφαλαίου

Το πέμπτο κεφάλαιο διαπραγματεύεται την προσαρμογή της μεθόδου όρασης H/Y, με τους βασικούς άξονες που παρουσιάστηκαν στο τέταρτο κεφάλαιο, στο πεδίο της μουσικής αλληλεπίδρασης. Η προσαρμογή της μεθόδου αυτής οδήγησε παράλληλα στην ανάπτυξη ενός

συστήματος που ικανοποιεί το σκοπό και τους επιμέρους στόχους που τέθηκαν στην ενότητα 1.5.

Στο κεφάλαιο αυτό θα παρουσιαστεί το σκέλος της στατικής αναγνώρισης του συστήματος. Στο έκτο κεφάλαιο θα παρουσιαστεί αντίστοιχα το σκέλος της δυναμικής αναγνώρισης. Η μεν στατική αναγνώριση στοχεύει στην ανάκτηση δακτυλισμών για κάθε καρτέ και ανά δάκτυλο χωριστά. Η δε δυναμική αναγνώριση, ως επέκταση της στατικής, αποσκοπεί στην αναγνώριση « πιανιστικών » κινήσεων μέσα από μια ακολουθία καρτέ του βίντεο. Στην πρώτη περίπτωση δεν υπάρχει συσχέτιση μεταξύ της ανακτώμενης πληροφορίας μεταξύ των καρτέ, ενώ στη δεύτερη περίπτωση υπάρχει. Αξίζει να σημειωθεί ότι οι άξονες του συστήματος που διαπραγματεύονται την ανίχνευση δέρματος, την τμηματοποίηση του χεριού και τον εντοπισμό των δακτύλων είναι κοινοί για τη στατική και τη δυναμική αναγνώριση κινήσεων.

Στις ενότητες που ακολουθούν παρέχονται στον αναγνώστη όλες οι απαραίτητες πληροφορίες, θεμελιώσεις και δικαιολογήσεις, για τις μεθόδους και τους αλγορίθμους που χρησιμοποιήθηκαν για την ανάπτυξη του εν λόγω συστήματος.

5.2 Επιλογή εξοπλισμού

Η επεξεργασία βίντεο είναι μια διαδικασία που απαιτεί εκ των πραγμάτων ικανοποιητικό εξοπλισμό υλικού. Όταν χρησιμοποιείται ρυθμός λήψης βίντεο από 19 καρτέ το δευτερόλεπτο και πάνω είναι αυταπόδεικτο ότι χρειάζεται υπολογιστική ισχύς. Μια ακόμα παράμετρος που δυσκολεύει το έργο του επεξεργαστή είναι το γεγονός της ταχύτητας των δακτύλων την ώρα που ο ερμηνευτής παίζει πιάνο. Οι υψηλές αυτές ταχύτητες δε δίνουν τη δυνατότητα μεγάλης μείωσης του ρυθμού λήψης διότι αυτό θα προκαλούσε σημαντική απώλεια δεδομένων.

Στα εδάφια που ακολουθούν γίνεται ανάλυση του υλικού και του λογισμικού εξοπλισμού που χρησιμοποιήθηκε για την ανάπτυξη του συστήματος στα πλαίσια της διδακτορικής διατριβής.

5.2.1 Ηλεκτρονικός υπολογιστής

Ο ηλεκτρονικός υπολογιστής που χρησιμοποιήθηκε για την αποθήκευση του βίντεο, την επεξεργασία των δεδομένων και το προγραμματισμό του συστήματος είναι ο iMac 24''. Κριτήριο για την επιλογή του ήταν η ταχύτητα του επεξεργαστή ώστε να αξιολογηθεί το σύ-

στημα και σε εφαρμογή σε πραγματικό χρόνο. Πρόκειται για υπολογιστή με διπύρηνο επεξεργαστή στα 2.66 Ghz, με 4 GB Ram και 640 GB χωρητικότητας στο σκληρό δίσκο. Διαθέτει κάρτα γραφικών της Nvidia ειδικά σχεδιασμένη για επεξεργασία τρισδιάστατων γραφικών (Σχήμα 5.2.1.1).



Σχήμα 5.2.1.1 iMac 24

5.2.2 Πιάνο-συνθεσάιζερ

Το συνθεσάιζερ CS1x της Yamaha επιλέχτηκε για τη ανάπτυξη του συστήματος αναγνώρισης δακτυλικών κινήσεων του πιανίστα. Η επιλογή του καθορίστηκε με βάση τη διαθεσιμότητα του εργαστηρίου όπου αναπτύχθηκε το σύστημα. Προαπαιτούμενο του συστήματος είναι η εφαρμογή του σε οποιοδήποτε όργανο διαθέτει οκτάβες πλήκτρων, κλασικό ή μη, πιάνο ή συνθεσάιζερ. Κατά συνέπεια, τα εξαγόμενα αποτελέσματα της μεθόδου δεν επηρεάζονται από την εφαρμογή σε μοντέρνο ή κλασικό όργανο (Σχήμα 5.2.2.1).

5.2.3 Κάμερα

Για τη διεξαγωγή του πειράματος και την ανάπτυξη του συστήματος επιλέχτηκε η κάμερα Logitech QuickCam Communicate Deluxe webcam. Η επιλογή της συγκεκριμένης κάμερας βασίστηκε στους τρεις παρακάτω λόγους:

1. Χαμηλό κόστος,
2. Ικανοποιητικά τεχνικά χαρακτηριστικά,
3. Εύχρηστη βάση,
4. Συμβατότητα με το μαθηματικό πολυεργαλείο MATLAB.



Σχήμα 5.2.2.1 Συνθεσάιζερ Yamaha CS1x

Όπως έχει ήδη προαναφερθεί στο κεφάλαιο 2, οι μέθοδοι όρασης H/Y για την αναγνώριση κινήσεων απαιτούν αξιόπιστη εναλλακτική λύση, αναφορικά με την τεχνολογία αισθητήρων, και λόγω του χαμηλού κόστους ανάπτυξης τους. Συνεπώς, η επιλογή της κάμερας που χρησιμοποιήθηκε για την ανάπτυξη του συστήματος αποτελεί μια οικονομική λύση.

Τα τεχνικά χαρακτηριστικά της κάμερας κρίθηκαν απόλυτα ικανοποιητικά. Έχει δυνατότητα λήψης εικόνας μέχρι 5 megapixels, λήψη βίντεο σε ανάλυση μέχρι 1280x1024 και ρυθμό λήψης μέχρι 30 καρέ το δευτερόλεπτο, ενώ συνδέεται μέσω θύρας USB 2.0. Τέλος, ιδιαίτερα σημαντικό χαρακτηριστικό είναι το καλώδιο μήκους 1,8 μέτρων που διαθέτει και το οποίο προσφέρει τη δυνατότητα τοποθέτη-

σής της σε οποιοδήποτε σημείο του κλαβιέ του πιάνου. (Σχήμα 5.2.3.1)



Σχήμα 5.2.3.1 Κάμερα της Logitech με ειδική λαστιχωτή βάση

Η κάμερα διαθέτει ειδική λαστιχωτή βάση που επιτρέπει την ποικιλόμορφη τοποθέτησή της. Αποφευχθεί η κατασκευή εξωτερικής βάσης για την τοποθέτηση της κάμερας πάνω στο πιάνο, αφού η προεγκατεστημένη βάση της, τη σταθεροποιούσε παρά τους κραδασμούς από το πάτημα των πλήκτρων.

Το προγραμματιστικό σκέλος του συστήματος υλοποιήθηκε με το μαθηματικό πολυεργαλείο MATLAB. Συνεπώς, ήταν αναγκαία η εξασφάλιση συμβατότητας μεταξύ της κάμερας και του λογισμικού. Για το λόγο αυτό ελέγχθηκε η επικοινωνία τους μέσω του δωρεάν λογισμικού AMCap v 9.20. Πρόκειται για μια μικρή εφαρμογή λήψης βίντεο και συμβατής προ-απεικόνισης με το Microsoft DirectShow (παλαιότερο *ActiveMovie*). Βασίζεται στον κώδικα AMCap που προσφέρεται με το Microsoft DirectX 9 SDK. Η συνεργασία της κάμερας, αλλά και οποιασδήποτε άλλης κάμερας, με το συγκεκριμένο λογισμικό εξασφαλίζει τη συμβατότητα μεταξύ κάμερας και MATLAB.

5.2.4 Λογισμικό

Για τις προγραμματιστικές ανάγκες της ανάπτυξης του συστήματος χρησιμοποιήθηκε το πολύ γνωστό μαθηματικό πολυεργαλείο MATLAB της εταιρείας Mathworks (<http://www.mathworks.com>).

Το MATLAB είναι ταυτόχρονα γλώσσα προγραμματισμού και περιβάλλον ανάπτυξης εφαρμογών. Απευθύνεται σε επιστήμονες που επιθυμούν να αναπτύξουν εφαρμογές σχετικά με επιστημονικούς υπολογισμούς, αυτοματισμούς, ψηφιακή επεξεργασία σήματος και επικοινωνίες, επεξεργασία εικόνας, ανάλυση δεδομένων, βιολογικής πληροφορική και χρηματοοικονομικά. Ένα από τα κλειδιά της επιτυχίας του MATLAB είναι το γεγονός δέχεται την εγκατάσταση μιας πληθώρας εργαλειοθηκών (*Toolboxes*), τα οποία αποτελούνται από μια ομάδα μαθηματικών κυρίως συναρτήσεων για μια τεράστια γκάμα επιστημονικών εφαρμογών.

Αναφορικά με την ανάπτυξη του συστήματος αναγνώρισης δακτυλικών κινήσεων, εκτός από το βασικό προγραμματιστικό περιβάλλον του MATLAB (έκδοση 7.6 για *Macintosh*), χρησιμοποιήθηκαν εργαλειοθήκες. Το *image acquisition toolbox (IMAQ)* είναι μια εργαλειοθήκη που επιτρέπει την απευθείας καταγραφή εικόνων και βίντεο στον Η/Υ, με τη βοήθεια κατάλληλου υλικού. Χρησιμοποιήθηκε η έκδοση 3.3 της εργαλειοθήκης για τη καταγραφή του σήματος βίντεο μέσω της κάμερας. Το IMAQ ανέλαβε την αποθήκευση του βίντεο στο σκληρό δίσκο, το χωρισμό του σε καρέ και την αποθήκευση των καρέ με τη μορφή εικόνας.

Η εργαλειοθήκη *DIPimage* χρησιμοποιήθηκε για την ψηφιακή επεξεργασία και ανάλυση της εικόνας του συστήματος. Αναπτύχθηκε από το *Quantitative Imaging Group* του τμήματος εφαρμοσμένων επιστημών του πανεπιστημίου Delft της Ολλανδίας και προσφέρεται δωρεάν για ερευνητικούς σκοπούς [Luengo 2000]. Το πλεονέκτημα της συγκεκριμένης εργαλειοθήκης είναι ότι προσφέρει μεγάλο πλήθος από μαθηματικές συναρτήσεις που αφορούν στην επεξεργασία σήματος. Οι συναρτήσεις αυτές είναι διαθέσιμες μέσω γραφικού περιβάλλοντος. Επιπλέον προσφέρει τη δυνατότητα αυτόματης εμφάνισης σε διαφορετικό παράθυρο όλων των επιμέρους σταδίων ανάλυσης της εικόνας.

Για τον υπολογισμό του κέντρου μάζας μέσω του περιγράμματος του χεριού σε κάθε εικόνα χρησιμοποιήθηκε η δωρεάν εργαλειοθήκη *Centroid Calculation Function* του Fahd A. ABASSI. Στη βιβλιοθήκη μπορεί να έχει πρόσβαση κάποιος μέσω της σελίδας του MATLAB. Η εργαλειοθήκη αυτή δέχεται ως είσοδο μια εικόνα η ο-

ποία περιέχει μόνο ένα αντικείμενο και εξάγει τις συντεταγμένες x και y του κέντρου μάζας.

5.3 Προκαταρκτική ανάλυση

5.3.1 Τοποθέτηση κάμερας

Παρατηρήθηκε ότι το φάσμα λήψης του φακού της κάμερας επιτρέπει την τοποθέτηση επάνω στο κλαβιέ του πιάνου με τρόπο ώστε να καταγράφεται κίνηση κατά μήκος μιας ή δύο οκτάβων. Υπάρχει η δυνατότητα να ανοίξει το φάσμα λήψης, ώστε να καταγράφονται περισσότερες οκτάβες, μέσω της αύξησης της απόστασης ανάμεσα στην κάμερα και το κλαβιέ. Αυτή η παρέμβαση απαιτεί μεγαλύτερη υπολογιστική ισχύ για την εκτέλεση του αλγορίθμου καθώς και απαραίτητη κοπή της περιοχής ενδιαφέροντος αφού θα καταγράφονται και πολλά περισσότερα αντικείμενα παρασκήνιου, τα οποία πιθανόν να έχουμε χρώμα παρόμοιο με το ανθρώπινο δέρμα.

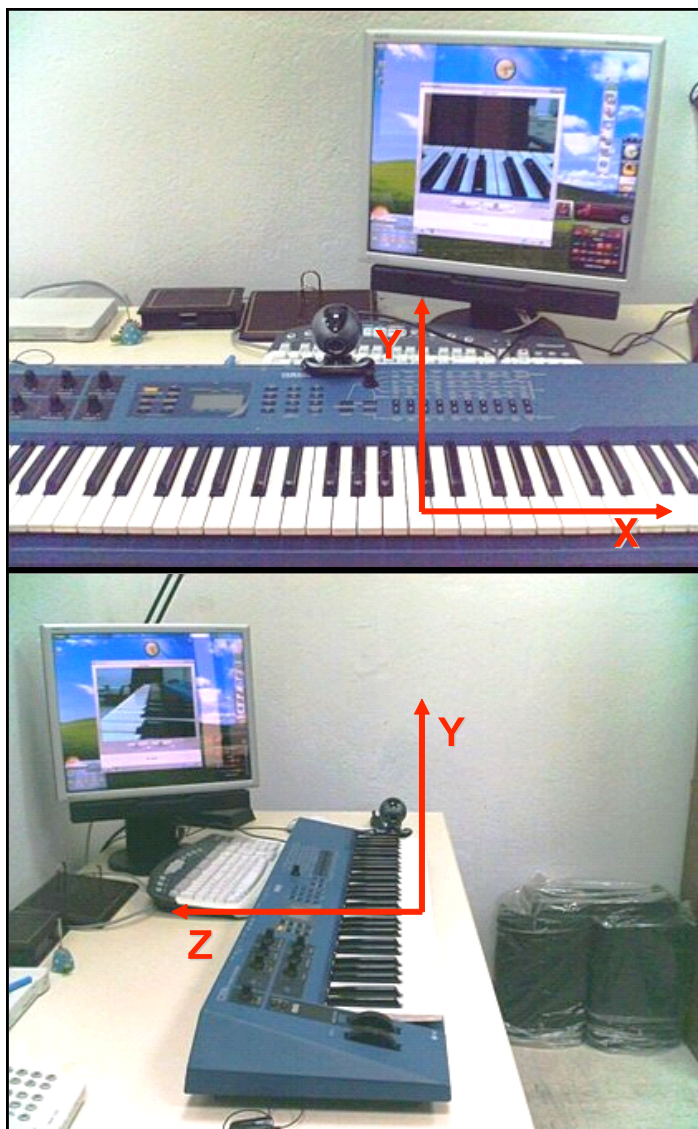
5.3.2 Γωνία λήψης

Τα δάκτυλα του πιανίστα κινούνται στο χώρο πάνω στου τρεις άξονες X , Y , Z . Δύο γωνίες λήψης μελετήθηκαν για την πλήρη καταγραφή των κινήσεων. Η πρώτη συνδέεται με την τοποθέτηση της κάμερας μπροστά από το κλαβιέ και η δεύτερη στην κορυφή του κλαβιέ (Σχήμα 5.3.2.1).

Η πρώτη γωνία εξασφαλίζει την πλήρη καταγραφή της κίνησης των δακτύλων, κοιτώντας μπροστά από το κλαβιέ, στον κατακόρυφο (Y) και τον οριζόντιο (X) άξονα. Η δεύτερη γωνία λειτουργεί επικουρικά για την αναγνώριση πατήματος μαύρων πλήκτρων. Καταγράφει την κίνηση, εκτός από τον άξονα Y , και στον Z . Σε κάθε περίπτωση, η κάθε γωνία λήψης μπορεί να προσφέρει καταγραφή κίνησης σε δύο μόνο άξονες.

Για τις ανάγκες ανάπτυξης του συστήματος χρησιμοποιήθηκε η πρώτη γωνία λήψης για αναγνώριση κινήσεων που αφορά μόνο στα άσπρα πλήκτρα. Έτσι, αναγνωρίζονται κινήσεις που διαγράφονται στους άξονες:

- X , για εντοπισμό της περιοχής των πλήκτρων,
- Y , για εντοπισμό της προετοιμασίας και του πατήματος του πλήκτρου.



Σχήμα 5.3.2.1 1: Λήψη μπροστά από το κλαβιέ, 2: Λήψη από την κορυφή του κλαβιέ

5.3.3 Φωτισμός

Ο φωτισμός παίζει σημαντικό ρόλο στην αναγνώριση κινήσεων μέσω ανίχνευσης δέρματος. Κακός φωτισμός ή φωτισμός που προκαλεί σκίαση αντικειμένων παρασκηνίου πάνω στο δέρμα μπορεί να έχει ως αποτέλεσμα την μη ανίχνευση δερματικών περιοχών στην εικόνα, ή, τη ανίχνευση μη δερματικών περιοχών ως δερματικές.

Για την αποφυγή τέτοιου είδους προβλημάτων επιλέχτηκε λευκό φως. Επίσης, παρακάτω θα παρουσιαστεί η τεχνική με την οποία η συνιστώσα της φωτεινότητας του χρώματος δε συμμετέχει στην ανίχνευση δέρματος.

5.3.4 Σύλληψη βίντεο

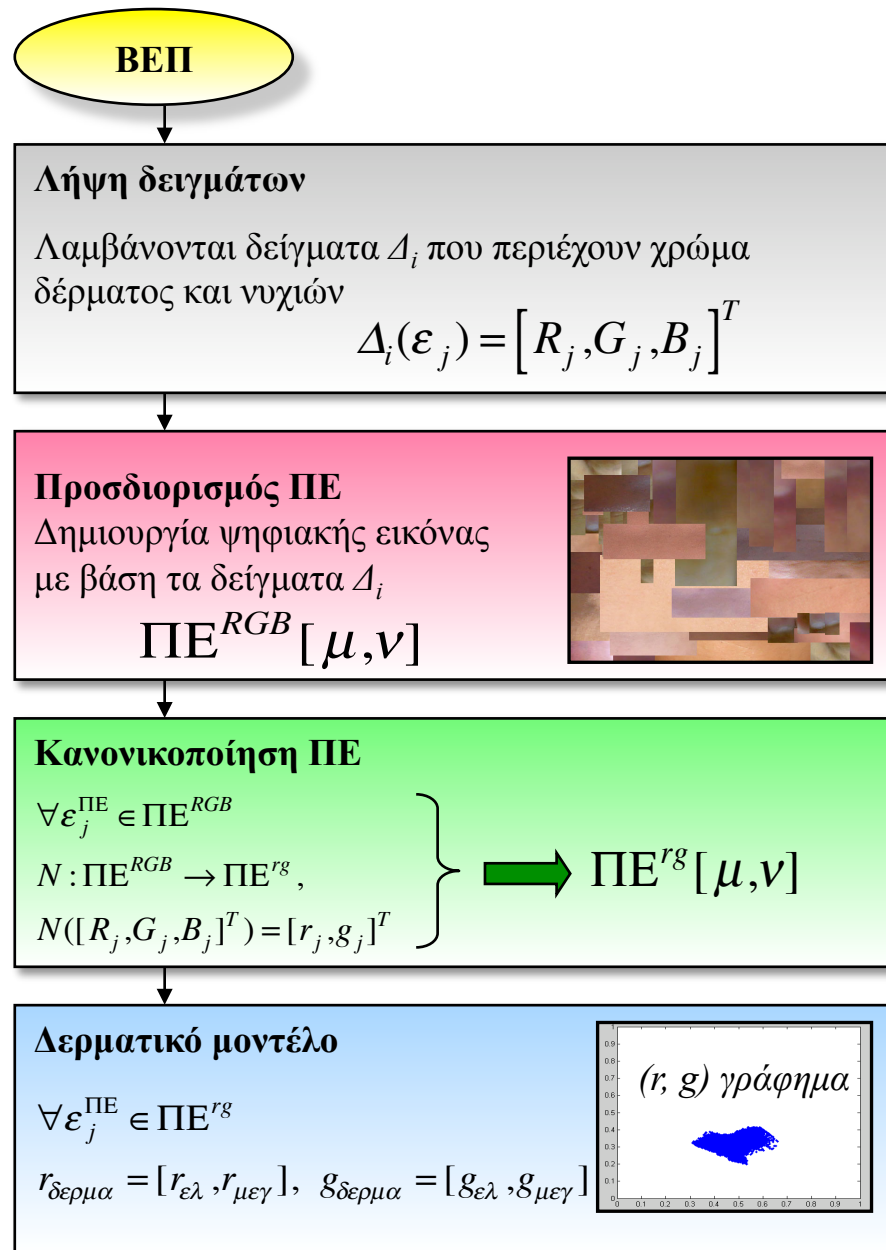
Δοκιμές για τη σύλληψη του βίντεο έγιναν σε χαμηλή (320x240), μέτρια (640x480) και υψηλή ανάλυση (1280x1024). Για την επεξεργασία βίντεο υψηλής ανάλυσης απαιτείται μεγαλύτερη υπολογιστική ισχύς. Λαμβάνοντας υπόψη ότι ο αλγόριθμος αναγνώρισης κινήσεων εφαρμόζεται σε κάθε καρέ του βίντεο χωριστά, η υψηλή ανάλυση δεν ενδείκνυται για την περίπτωση εφαρμογών σε πραγματικό χρόνο.

Η ανάλυση 640x480 προτιμήθηκε λόγω του ότι σε πρώτη φάση η δοκιμή του συστήματος δε θα υλοποιούνταν σε πραγματικό χρόνο, αλλά και του ότι ο αλγόριθμος ανίχνευσης δέρματος λειτουργεί καλά και σε χαμηλή ανάλυση. Τέλος, ο ρυθμός καταγραφής πλαισίων ορίστηκε στα 20 fps ώστε να αποκτηθεί λεπτομερής αποτύπωση της κίνησης των δακτύλων.

5.4 Ανίχνευση δέρματος

Η διαδικασία της ανίχνευσης δέρματος στοχεύει στον εντοπισμό δερματικών περιοχών σε κάθε καρέ του βίντεο εισόδου. Διαχωρίζει τα αντικείμενα του παρασκήνιου, τα οποία δε συμμετέχουν στην αναγνώριση κινήσεων, από εκείνα του προσκήνιου, τα οποία είναι δερματικές περιοχές.

Η διαδικασία στηρίζεται στο στατιστικό δερματικό μοντέλο βάσει του οποίου διεξάγεται ανίχνευση δέρματος. Έχοντας ως πηγή εικονοστοιχείων τη βιβλιοθήκη εικόνων του πιανίστα (BEP) προσδιορίζεται η περιοχή ενδιαφέροντος (ΠΕ) που είναι μια ψηφιακή εικόνα που περιλαμβάνει μόνο χρώμα δέρματος και νυχιών. Στη συνέχεια, η περιοχή ενδιαφέροντος κανονικοποιείται, μέσα από την προβολή του τρισδιάστατου χρωματικού χώρου RGB στον δισδιάστατο rg, ενώ διενεργείται και γραφική απεικόνιση της εικόνας στον χώρο rg. Μέσα από τη γραφική απεικόνιση υπολογίζεται η ελάχιστη και η μέγιστη τιμή στους δύο άξονες r και g. Τα δύο ζευγάρια τιμών $r_{ελ}$, $r_{μεγ}$ και $g_{ελ}$, $g_{μεγ}$ αποτελούν το στατιστικό δερματικό μοντέλο, το οποίο θα χρησιμοποιηθεί για την ανίχνευση δέρματος σε οποιαδήποτε καρέ του βίντεο (Σχήμα 5.4.1).



Σχήμα 5.4.1 Αλγόριθμος δημιουργίας δερματικού μοντέλου

5.4.1 Χαρακτηριστικά ανθρώπινου δέρματος

Το χρώμα του ανθρώπινου δέρματος έχει κάποια ποιοτικά χαρακτηριστικά που το κάνουν να διαφέρει από τα υπόλοιπα χρώματα. Χα-

ρακτηριστικό τα ανθρώπινου δέρματος είναι η χρωματική του κατανομή. Η κατανομή του χρώματος αυτού διαφέρει σχεδόν απόλυτα από τις χρωματικές κατανομές των αντικειμένων του φόντου στην εικόνα. Η ανίχνευση περιοχών ανθρώπινου δέρματος είναι μια τεχνική που εφαρμόζεται ευρέως σε περιπτώσεις αναγνώρισης κινήσεων ή χειρονομιών, παρακολούθησης ή αναγνώρισης προσώπων, μέσα από εικόνα ή βίντεο. Το σύστημα που αναπτύχθηκε κάνει χρήση τεχνικών ανίχνευσης δέρματος.

Οι λόγοι που υποστηρίζουν τη χρήση της συγκεκριμένης τεχνικής προσδιορίζονται σε τρία σημεία. Καταρχήν, έρευνες που διεξήχθησαν στο πρόσφατο παρελθόν αποδεικνύουν ότι οι διαφορές που παρατηρούνται στο χρώμα του δέρματος μεταξύ των διαφορετικών φυλών, οφείλονται στη διαφορά που υπάρχει στη φωτεινότητα του χρώματος (*συνιστώσα φωτεινότητας*) και όχι στο χρώμα (*συνιστώσα χρωματικότητας*) [Yang 1997]. Επιπρόσθετα, η χρήση μόνο των χρωματικών συνιστωσών του χρώματος του δέρματος καθιστά την ανίχνευσή του, ανεπηρέαστη από μεταβολές στο φωτισμό. Τέλος, η υπολογιστική ισχύς που απαιτείται για την ανίχνευση δέρματος με τη χρήση μόνο της χρωματικής συνιστώσας είναι πολύ λιγότερη σε σχέση με τη χρήση και της συνιστώσας φωτεινότητας.

Αξίζει να σημειωθεί ότι οι τεχνικές ανίχνευσης δέρματος υπόκεινται σε κάποιους περιορισμούς που πρέπει να ληφθούν υπόψη κατά την ανάπτυξη των αντίστοιχων αλγορίθμων. Η βελτιστοποίηση των αποτελεσμάτων των αλγορίθμων αυτών ανίχνευσης δέρματος επιτυγχάνεται όταν υπάρχει αντίθεση μεταξύ των αντικειμένων του παρασκήνιου και του ανθρώπινου δέρματος. Συγκεκριμένα για το σύστημα αναγνώρισης δακτυλικών κινήσεων, το χρώμα των πλήκτρων και του δέρματος είναι γνωστό εκ των προτέρων. Αν λοιπόν προτιμηθεί κάποιο σκουρόχρωμο ένδυμα, τότε τα αποτελέσματα της ανίχνευσης, όπως θα αποδειχθεί σε επόμενο εδάφιο, δεν επηρεάζονται από το χρώμα των αντικειμένων του παρασκήνιου. Ο ισχυρισμός αυτός ενισχύεται και από το γεγονός ότι τα μοναδικά κινούμενα μέρη στην εικόνα είναι το χέρι και τα δάκτυλα. Συνεπώς περιοχές που ενδεχομένως θεωρηθούν δερματικές χωρίς να είναι, δεν επηρεάζουν τα αποτελέσματα του συστήματος λόγω του ότι δεν κινούνται.

5.4.2 Προσδιορισμός περιοχής ενδιαφέροντος

Η ανίχνευση δέρματος μπορεί να επιτευχθεί δημιουργώντας ένα όσο το δυνατό καλύτερο στατιστικό δερματικό μοντέλο. Για να λειτουργήσει αποτελεσματικά όμως αυτό το μοντέλο θα πρέπει προηγουμένως να έχει προσδιοριστεί επαρκώς η περιοχή ενδιαφέροντος.

Αν θεωρήσουμε ότι γίνεται μια προβολή βίντεο που εστιάζει στις δακτυλικές κινήσεις ενός πιανίστα κατά την ερμηνεία του. Οι περιοχές που καταγράφονται από το βίντεο μπορούν να ταξινομηθούν στις περιοχές που ανήκουν στα δάκτυλα του πιανίστα και σε αυτές που δεν ανήκουν. Εφόσον το σύστημα στοχεύει στην ανίχνευση δέρματος, με έμφαση σε περιοχές δακτύλων, το χρώμα των νυχιών θα πρέπει να συμμετέχει στη διαδικασία ανίχνευσης.

Μέσα στο πλαίσιο αυτό, προτείνονται τα παρακάτω βήματα για τον προσδιορισμό της περιοχής ενδιαφέροντος:

1. Από κάθε εικόνα που ανήκει στη βιβλιοθήκη εικόνων του πιανίστα (σύνολο φωτογραφιών με διαφορετικό φωτισμό), απομονώνονται χειροκίνητα δείγματα δερματικών περιοχών που περιλαμβάνουν ταυτόχρονα και περιοχές νυχιών. Τα δείγματα αποτελούνται από εικονοστοιχεία. Κάθε ένα από αυτά τα εικονοστοιχεία απεικονίζεται με ένα διάνυσμα τριών συνιστωσών (R,G,B) ή αλλιώς από ένα σημείο στο χρωματικό χώρο λόγω της τριχρωμίας. Συνεπώς, το κάθε δείγμα που λαμβάνεται, αποτελεί έναν τρισδιάστατο πίνακα όπως φαίνεται παρακάτω:

$$\Delta_i(\varepsilon_j) = [R_j, G_j, B_j]^T \quad (5.4.2.1)$$

Όπου:

- Δ_i , το δείγμα,
- ε_j , το εικονοστοιχείο j του δείγματος Δ_i ,
- R_j , η συνιστώσα του κόκκινου χρώματος για το ε_j ,
- G_j , η συνιστώσα του πράσινου χρώματος για το ε_j ,
- B_j , η συνιστώσα του μπλε χρώματος για το ε_j .

2. Οι περιοχές αυτές αντιγράφονται σε νέα εικόνα. Άρα, η περιοχή ενδιαφέροντος αποτελεί μια ψηφιακή εικόνα:

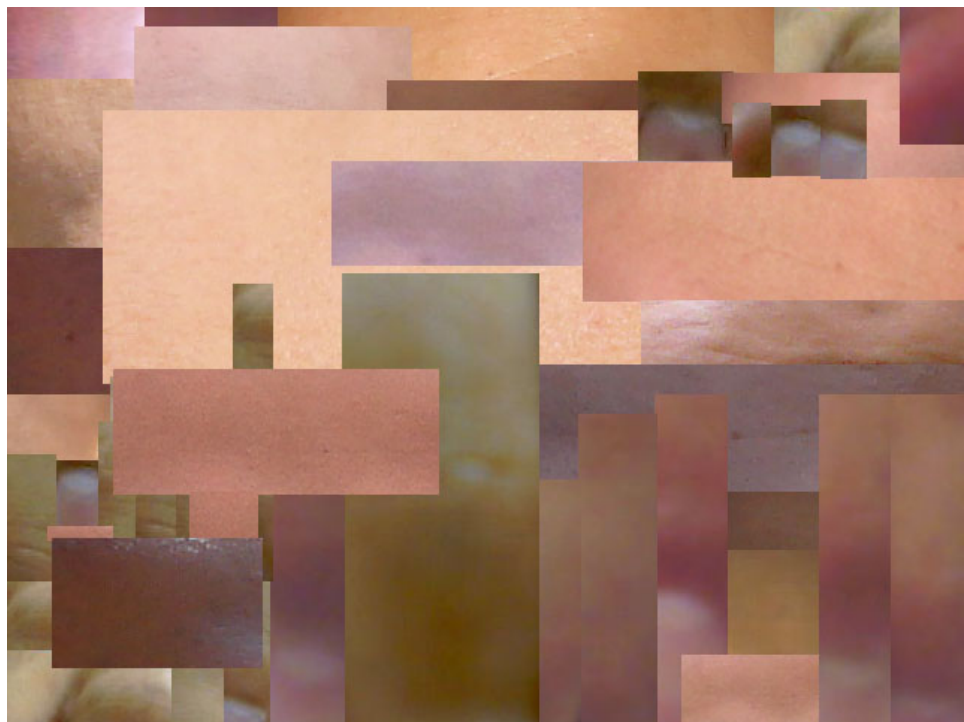
$$PE^{RGB}[\mu, \nu] \quad (5.4.2.2)$$

Όπου:

- PE^{RGB} , η περιοχή ενδιαφέροντος στο χρωματικό χώρο RGB,
- M , το πλήθος των εικονοστοιχείων οριζοντίως,
- N , το πλήθος των εικονοστοιχείων καθέτως,
- $\mu=\{0,1,2,\dots,M-1\}$, η θέση οριζοντίως στην εικόνα,
- $\nu=\{0,1,2,\dots,N-1\}$, η θέση καθέτως στην εικόνα.

Τα σημεία τομής των γραμμών με τις στήλες είναι τα εικονοστοιχεία της περιοχής ενδιαφέροντος.

Η PE^{RGB} φαίνεται στο σχήμα 5.4.2.1. Για την περίπτωση του συστήματος αναγνώρισης δακτυλικών κινήσεων η PE^{RGB} έχει $M=640$ και $N=480$. Παρόλο που η περιοχή ενδιαφέροντος αφορά δερματικές περιοχές μόνο του πιανίστα, παρατηρούμε τις διαφορετικές αποχρώσεις που έχει το δέρμα του ανάλογα με τη φωτογραφία που λήφθηκε το δείγμα (Σχήμα 5.4.2.1). Επιβεβαιώνεται δηλαδή για άλλη μια φορά ότι ο χώρος RGB επηρεάζεται από τις μεταβολές της φωτεινότητας.



Σχήμα 5.4.2.1 Ψηφιακή εικόνα PE^{RGB} που αποτελεί την περιοχή ενδιαφέροντος

5.4.3 Κανονικοποίηση περιοχής ενδιαφέροντος

Χρωματικός χώρος είναι ένα εύρος χρωμάτων στο ορατό φάσμα. Ο χρωματικός χώρος RGB αναπαριστά ένα μεγάλο ποσοστό του ορατού φάσματος αναμειγνύοντας φως κόκκινου R, πράσινου G και μπλε B χρώματος σε διάφορες αναλογίες. Ο RGB χρησιμοποιείται ευρέως στην ψηφιακή επεξεργασία εικόνας και κατ' επέκταση στην επεξεργασία ψηφιακού βίντεο.

Όπως ήδη αναφέρθηκε στο εδάφιο 3.6.1.2, οι τιμές που έχει ένα εικονοστοιχείο, στο χώρο RGB, ποικίλουν ανάλογα με τη φωτεινότητα. Αλλαγές στο φωτισμό προκαλούν μεταβολή της θέσης στο χώρο χρώματος. Η ανίχνευση δέρματος στηρίζεται στην εύρεση περιοχών στην εικόνα που έχουν το χαρακτηριστικό χρώμα του δέρματος. Συνεπώς, αν η αποτύπωση του χρώματος του δέρματος στην εικόνα μεταβάλλεται ανάλογα με το φωτισμό, εύκολα προκαλούνται αστοχίες στον εντοπισμό.

Η τεχνική που υιοθετήθηκε για να επιλυθεί το παραπάνω πρόβλημα είναι η κανονικοποίηση του RGB ως προς τη φωτεινότητα ή την εκπεμπόμενη ακτινοβολία. Ο κανονικοποιημένος RGB, ή απλώς rg, είναι ανεξάρτητος από τις μεταβολές των τιμών της φωτεινότητας. Οι συνιστώσες για το κόκκινο, το πράσινο και το μπλε του κανονικοποιημένου χώρου RGB προκύπτουν χρησιμοποιώντας τις σχέσεις 3.6.1.2.1 έως 3.6.1.2.8. Συνοψίζοντας τη μέθοδο του εδαφίου 3.6.1.2, κατάλληλη συνάρτηση N απεικονίζει τα εικονοστοιχεία της περιοχής ενδιαφέροντος από το χρωματικό χώρο RGB στον rg, ανεξαρτητοποιημένα από μεταβολές στη φωτεινότητα. Η ψηφιακή εικόνα PE^{RGB} δεν είναι κάτι διαφορετικό από έναν πίνακα που περιέχει τριάδες τιμών για κάθε μια συνιστώσα R, G και B. Η συνάρτηση N παράγει ένα νέο πίνακα PE^{rg} που περιέχει ζευγάρια τιμών για κάθε μια συνιστώσα r και g. Όλα τα ζευγάρια τιμών στον νέο πίνακα, προφανώς θα ανήκουν στο διάστημα $[0,1]$.

5.4.4 Δημιουργία δερματικού μοντέλου

Μέσω της κανονικοποίησης των εικονοστοιχείων του χώρου RGB που ανήκουν στην περιοχή ενδιαφέροντος, το σύστημα απέκτησε την ψηφιακή εικόνα PE^{rg} , η οποία αποτελείται από ζευγάρια τιμών για τις συνιστώσες r και g.

Για το σύστημα αναγνώρισης δακτυλικών κινήσεων στη ΜΑ, οι δύο χρωματικές συνιστώσες r και g χρησιμοποιήθηκαν προκειμένου να επιτευχθεί ανίχνευση δερματικών περιοχών σε οποιοδήποτε

καρέ του βίντεο. Προϋπόθεση για την αποτελεσματική ανίχνευση δέρματος είναι οι χρωματικές συνιστώσες να καταλαμβάνουν μια πολύ μικρή περιοχή χρώματος [Wang 1997], όπως φαίνεται στο σχήμα 5.4.4.1. Συνεπώς οι περιοχές, ενός οποιουδήποτε καρέ, των οποίων οι χρωματικές συνιστώσες των εικονοστοιχείων ανήκουν σε αυτήν τη μικρή περιοχή (*σημειώνεται με μπλε χρώμα*), θεωρούνται ως περιοχές ανθρώπινου δέρματος. Κατά συνέπεια, εικονοστοιχεία με χρωματικές συνιστώσες που βρίσκονται εκτός αυτής της μικρής περιοχής, θεωρούνται ως περιοχές μη ανθρώπινου δέρματος. Η εξήγηση αυτού του φαινομένου βασίζεται στο ότι οι χρωματικές συνιστώσες, που αφορούν στο χρώμα του δέρματος, βρίσκονται σε μια μικρή περιοχή του διαστήματος χρώματος.

Το ερώτημα που τίθεται σε αυτό το σημείο, είναι το πως θα περιγραφεί με μαθηματικές σχέσεις αυτή η δερματική περιοχή. Το γεωμετρικό σχήμα που απεικονίζεται στο σχήμα 6-, εξαρτάται από τις εικόνες από τις οποίες λάβαμε δείγματα προκειμένου να δημιουργήσουμε την περιοχή ενδιαφέροντος. Όσο πιο καλά αντιπροσωπευτικές ήταν οι εικόνες που επιλέξαμε τόσο πιο συμπαγές και ομοιόμορφο είναι το γεωμετρικό σχήμα, μειώνοντας τις περιπτώσεις σφάλματος στη μετέπειτα ανίχνευση δέρματος. Παρά την καλή ποιότητα των εικόνων που μπορεί να επιλεγούν, το σχήμα που προκύπτει είναι σχεδόν αδύνατο να αποτελεί κλασικό γεωμετρικό σχήμα (*τετράγωνο, παραλληλόγραμμο, έλλειψη κλπ*). Αν όμως επιτευχθεί ο στόχος το σχήμα να είναι σχετικά συμπαγές, δίχως διάσπαρτα διακριτά σημεία, τότε, η δερματική περιοχή φράσσεται υπολογίζοντας ακρότατα για τις δύο συνιστώσες r και g (Σχήμα 5.4.5.1). Έτσι, μπορούν να προσδιοριστούν τα ζευγάρια τιμών $r_{ελ}$, $r_{μεγ}$ και $g_{ελ}$, $g_{μεγ}$ από τα οποία καθορίζονται τα διαστήματα $r_{δέρμα} = [r_{ελ}, r_{μεγ}]$ και $g_{δέρμα} = [g_{ελ}, g_{μεγ}]$, μέσα στα οποία ανήκουν οι χρωματικές συνιστώσες του ανθρώπινου δέρματος. Έτσι, αν ληφθεί ένα οποιοδήποτε εικονοστοιχείο του οποίου η κάθε μια χρωματική συνιστώσα ανήκει στο αντίστοιχο διάστημα τιμών, τότε θεωρείται ότι το εικονοστοιχείο αυτό έχει χρώμα δέρματος.

Το ερώτημα που γεννάται στην παρούσα φάση είναι τι συμπεράσμα βγαίνει στην περίπτωση που οι χρωματικές συνιστώσες κάποιου εικονοστοιχείου ανήκουν στο παραλληλόγραμμο που σχηματίζουν τα μέγιστα και τα ελάχιστα των δύο αξόνων (*κόκκινη περιοχή*), αλλά δεν ανήκουν στη δερματική περιοχή (*μπλε περιοχή*) (Σχήμα 5.4.4.2). Εφόσον η αναπαράσταση των συνιστωσών του εικονοστοιχείου δεν ανήκει στην μπλε περιοχή τότε σίγουρα δεν πρόκειται για εικονοστοιχείο με χρώμα δέρματος. Από την άλλη μεριά, εφόσον ανήκει στα δύο διαστήματα ($r_{δέρμα}$ και $g_{δέρμα}$) θα θεωρείται ως εικονο-

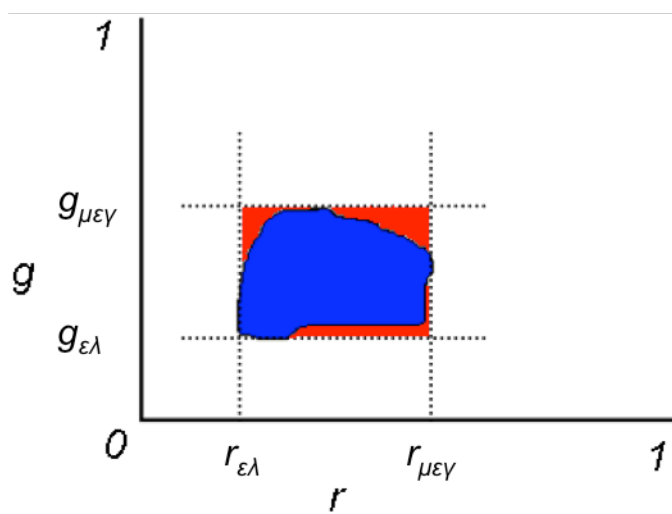
στοιχείο δέρματος. Τι γίνεται λοιπόν όταν ένα εικονοστοιχείο δεν έχει χρώμα δέρματος αλλά θεωρείται ότι έχει; Η περίπτωση αυτή εμπεριέχεται στις περιπτώσεις σφάλματος του δερματικού μοντέλου και οφείλεται στο γεγονός ότι το μοντέλο είναι ατελές. Σε αυτό το σημείο επιβεβαιώνεται για μια ακόμα φορά η ανάγκη για συλλογή αντιπροσωπευτικών και καλής ποιότητας εικόνων, προκειμένου να κατασκευαστεί η βιβλιοθήκη εικόνων. Δεν υπάρχει μια μαθηματική συνάρτηση η οποία να περιγράφει την κατανομή των χρωματικών συνιστωσών, που αφορούν στο χρώμα δέρματος, στον rg. Για το λόγο αυτό, είναι πάρα πολύ δύσκολο, έως αδύνατο, να δημιουργηθεί ένα τέλειο μοντέλο που να ανιχνεύει δερματικές περιοχές δίχως περιθώριο σφάλματος.



Σχήμα 5.4.4.1 Γραφική αναπαράσταση της χρωματικής περιοχής του δέρματος στον rg

Τεχνικές ώστε να παραμείνει σχεδόν ανεπηρέαστο το σύστημα από κάποιο σφάλμα στην ανίχνευση δέρματος, καθώς και σύγκριση των αποτελεσμάτων που παράγει το δερματικό μοντέλο του παρόντος συστήματος, με ένα τυχαίο δερματικό μοντέλο, θα προταθούν και θα παρουσιαστούν σε επόμενο εδάφιο.

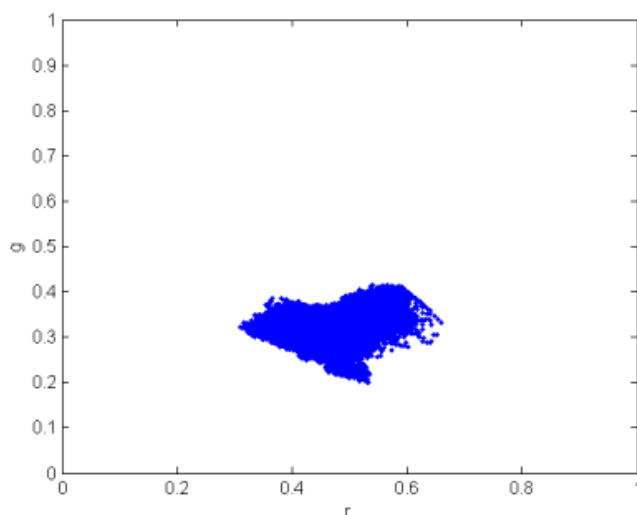
Στο σύστημα αναγνώρισης δακτυλικών κινήσεων, υλοποιήθηκε γραφική απεικόνιση των χρωματικών συνιστωσών των εικονοστοιχείων της περιοχής ενδιαφέροντος. Από τη γραφική απεικόνιση (Σχήμα 5.4.4.3) μπορεί κανείς να εξακριβώσει ότι πρόκειται για μια περιοχή με ελάχιστα διάσπαρτα σημεία. Αυτό είναι θετικό στοιχείο διότι διευκολύνει το φράξιμο των διαστημάτων για τους άξονες r και g . Γίνεται ελαστική χρήση των όρων «ελάχιστο» και «μέγιστο» διαστημάτων, μη λαμβάνοντας υπόψη διάσπαρτα σημεία στο χώρο. Έτσι, με βάση το σχήμα 5.4.4.3, ορίστηκαν τα διαστήματα:



Σχήμα 5.4.4.2 Γραφική αναπαράσταση της χρωματικής περιοχής του δέρματος στον rg

- $r_{\text{δέρμα}}=[0.34, 0.62]$
- $g_{\text{δέρμα}}=[0.25, 0.42]$

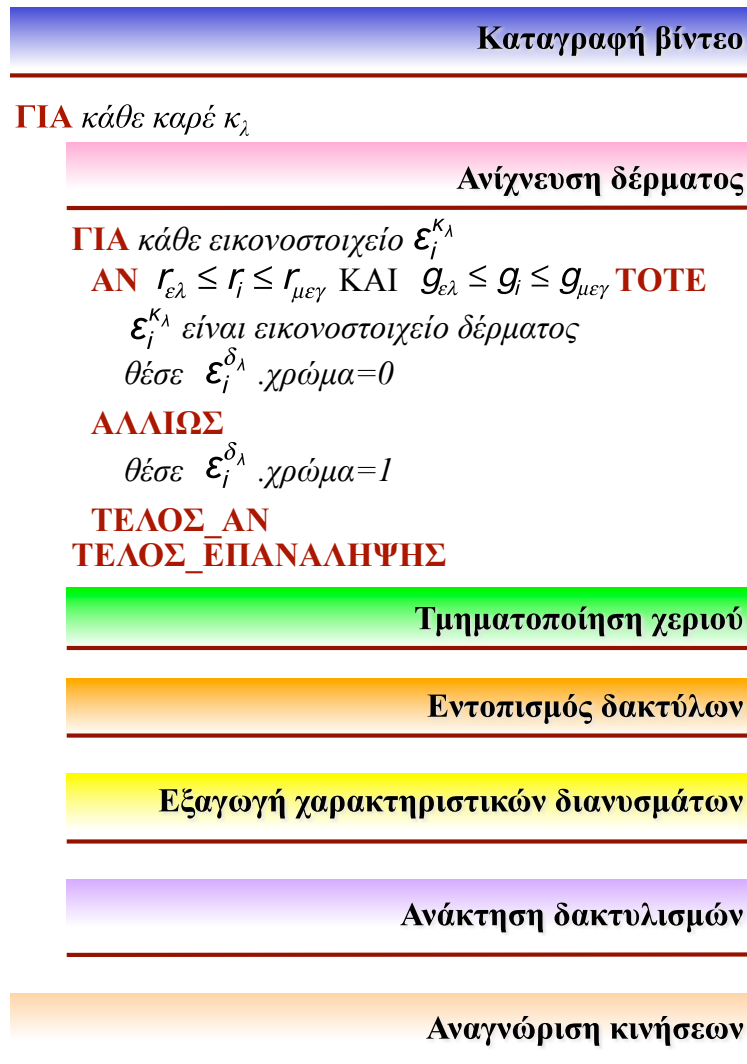
Τα δύο αυτά διαστήματα θα χρησιμοποιηθούν παρακάτω για την ανίχνευση δέρματος στην ακολουθία των καρέ του βίντεο.



Σχήμα 5.4.4.3 Απεικόνιση χρωματικών συνιστωσών των εικονοστοιχείων της PE^g

5.4.5 Χρήση μοντέλου για ανίχνευση δέρματος σε εικόνα

Το σύστημα δέχεται σαν είσοδο μια εικόνα ή μια ακολουθία καρέ του βίντεο στην οποία διενεργεί ανίχνευση δέρματος, βάσει του στατιστικού δερματικού μοντέλου, και εξάγει δυαδικές μάσκες (Σχήμα 5.4.5.1).



Σχήμα 5.4.5.1 Ανίχνευση δέρματος στα καρέ της ακολουθίας του βίντεο

Δυαδική μάσκα είναι μια ψηφιακή εικόνα, τα εικονοστοιχεία της οποίας μπορούν να πάρουν δύο μόνο τιμές χρώματος, το άσπρο

και το μαύρο. Αυτό σημαίνει ότι κάθε εικονοστοιχείο αποθηκεύεται σαν ένα bit (0 ή 1).

Θεωρείται η παρακάτω ακολουθία K :

$$K = \{k_1, \dots, k_\lambda, \dots, k_\Lambda\} \quad (5.4.5.1)$$

Όπου:

- K , η ακολουθία των καρτέ,
- Λ , το πλήθος των καρτέ,
- $\lambda = \{1, 2, \dots, \Lambda\}$, ο δείκτης των καρτέ,
- k_λ , το κάθε καρτέ.

Το κάθε καρτέ k_λ είναι μια ψηφιακή εικόνα η οποία ανήκει στο χρωματικό χώρο RGB. Προκειμένου να πραγματοποιηθεί ανίχνευση δέρματος, θα πρέπει να εφαρμοστεί η διαδικασία κανονικοποίησης του χρωματικού χώρου, όπως ακριβώς περιγράφηκε στο εδάφιο 5.4.3 για την περιοχή ενδιαφέροντος. Έστω λοιπόν, $E_i^{k_\lambda}$ τα εικονοστοιχεία του καρτέ k_λ μετά την κανονικοποίηση, με $i = \{1, 2, \dots, 307200\}$ (χρησιμοποιήθηκε βίντεο ανάλυσης 640x480, άρα 307200 εικονοστοιχεία). Για κάθε ένα από τα 307200 εικονοστοιχεία του κάθε καρτέ του βίντεο γίνεται έλεγχος συνθήκης για το αν τα r_i και g_i του εικονοστοιχείου $E_i^{k_\lambda}$ ανήκουν στα $r_{\text{δέρμα}}$ και $g_{\text{δέρμα}}$ αντίστοιχα. Αναλυτικότερα η διαδικασία περιγράφεται στο σχήμα 5.4.5.1.

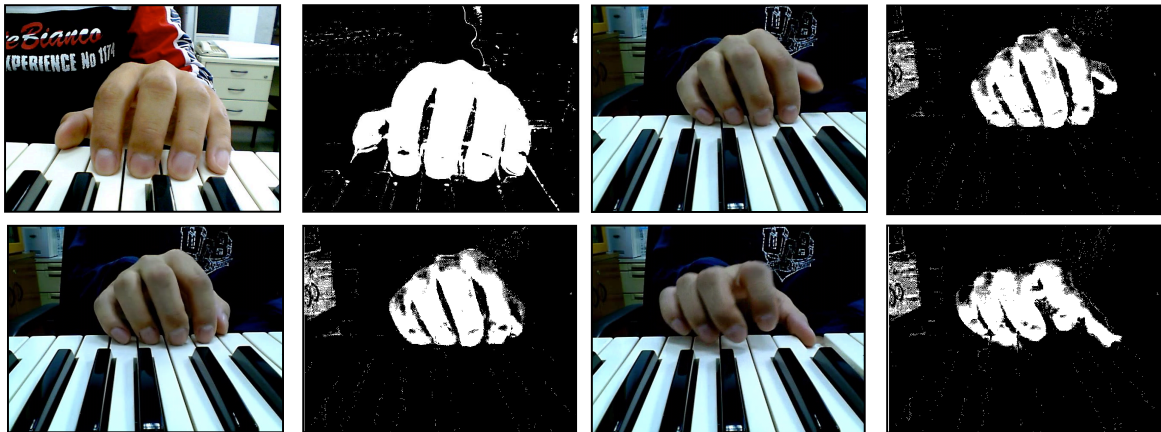
Όπως διατυπώθηκε προηγουμένως, η έξοδος της διαδικασίας ανίχνευσης είναι η δυαδική μάσκα δ_λ , για κάθε καρτέ k_λ του βίντεο. Το χρώμα που αποδίδεται στα αντικείμενα μιας δυαδικής μάσκας είναι το χρώμα του προσκηνίου (0) και το χρώμα του παρασκηνίου (1). Η δυαδική μάσκα δ_λ είναι μια ψηφιακή εικόνα της οποίας τα εικονοστοιχεία $E_i^{\delta_\lambda}$ παίρνουν μόνο δύο τιμές, 0 ή 1. Στις δυαδικές μάσκες που παράχθηκαν από το σύστημα αναγνώρισης δακτυλικών κινήσεων τα αντικείμενα του παρασκηνίου σημειώνονται με μαύρο χρώμα, ενώ αυτά το προσκηνίου σημειώνονται με άσπρο χρώμα.

Στο σχήμα 5.4.5.2 παρατίθενται τα αποτελέσματα ανίχνευσης δέρματος σε τέσσερις διαφορετικές εικόνες του χεριού του πιανίστα. Μελετώντας συνολικά αυτά τα αποτελέσματα, προκύπτει το συμπέρασμα ότι εξάγεται σε πολύ ικανοποιητικό βαθμό η φιγούρα του χεριού, ενώ παράλληλα απομονώνονται αρκετά τα αντικείμενα τα παρασκηνίου.

Ένα δεύτερο συμπέρασμα είναι η επιβεβαίωση ότι το στατιστικό δερματικό μοντέλο που προέκυψε είναι ατελές. Ατελές, διότι δεν έχει απόλυτη επιτυχία στην ανίχνευση δέρματος για λόγους που

αναπτύχθηκαν στην ενότητα 5.4.4. Πάντως, όσο λιγότερες είναι οι ατέλειες του μοντέλου στην ανίχνευση δέρματος, τόσο ευκολότερος είναι ο αποκλεισμός τους σε επόμενο βήμα.

Παρατηρούμε τέλος, ότι η ανίχνευση του δέρματος και η εξαγωγή της φιγούρας του χεριού είναι καλή ακόμα και σε δύσκολες περιπτώσεις, όπως όταν οι αποστάσεις μεταξύ των δακτύλων γίνεται πάρα πολύ μικρές.

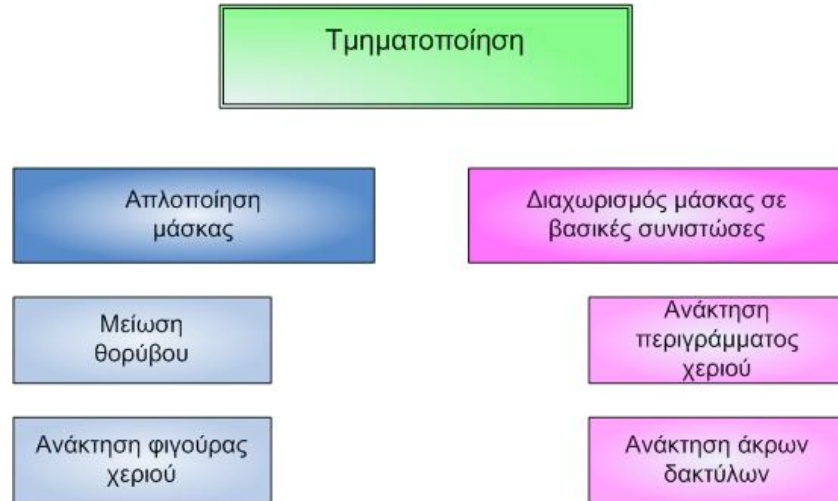


Σχήμα 5.4.5.2 Ανίχνευση δέρματος υπό διαφορετικές συνθήκες φωτισμού

5.5 Τμηματοποίηση χεριού

Η τμηματοποίηση του χεριού είναι η διαδικασία μέσω της οποίας το κάθε καρέ του βίντεο διαχωρίζεται σε περιοχές οι οποίες περιλαμβάνουν καταρχήν το χέρι και στη συνέχεια τα άκρα των δακτύλων του πιανίστα.

Η διαδικασία τμηματοποίησης του χεριού χωρίζεται σε δύο επιμέρους στάδια, την *απλοποίηση της δυναδικής μάσκας* δ_λ και το *διαχωρισμό της σε βασικές συνιστώσες* (Σχήμα 5.5.1). Μέσω της τμηματοποίησης του χεριού, το σύστημα στοχεύει στον εντοπισμό της θέσης των άκρων των δακτύλων.



Σχήμα 5.5.1 Διαδικασία τμηματοποίησης εικόνας χεριού

Η τμηματοποίηση θα πραγματοποιηθεί μέσω διεξοδικής επεξεργασίας της εικόνας με στόχο την απαλοιφή του θορύβου που υπάρχει σε αυτήν και την απλοποίηση της δομής της. Μέριμνα, καθ' όλη τη διαδικασία τμηματοποίησης της δλ, είναι η διατήρηση των κύριων χαρακτηριστικών του χεριού, όπως τα άκρα των δακτύλων και το περίγραμμά του. Κατά τη φάση απλοποίησης της δλ, θα πραγματοποιηθεί α) *μείωση θορύβου*, έχοντας ως προϊόν μια μάσκα με ομαλότερη δομή, η οποία θα είναι πιο εύκολη στη διαχείρισή της και β) *ανάκτηση της φιγούρας του χεριού*, έχοντας ως προϊόν τη σιλουέτα του χεριού που είναι καταλληλότερη για επιπλέον επεξεργασία και διαμέριση σε περιοχές. Κατά τη φάση του διαχωρισμού της μάσκας σε βασικές συνιστώσες, θα γίνει α) *ανάκτηση του περιγράμματος του χεριού*, έχοντας ως αποτέλεσμα μια εικόνα που περιλαμβάνει μόνο τη γραμμή του περιγράμματος και β) *ανάκτηση των άκρων των δακτύλων*, έχοντας σαν αποτέλεσμα τα άκρα των δακτύλων από όπου μπορεί να γίνει ευκολότερα εντοπισμός της θέσης των δακτύλων.

Για την περιγραφή της διαδικασίας τμηματοποίησης του χεριού επιλέχθηκε η εικόνα (1) του σχήματος 5.5.2. Στο σχήμα 5.5.2 εμφανίζεται επίσης η δυαδική μάσκα (2) που παράγει το στατιστικό δερματικό μοντέλο. Μέσω των εικόνων διακρίνεται η ύπαρξη λευκών περιοχών, οι οποίες προφανώς δεν ανήκουν στο χέρι του πιανίστα. Όλη αυτή η πληροφορία είναι σχεδόν αναπόφευκτη και εμφανίζεται συνήθως με την εισαγωγή νέων δομών μικρής κλίμακας και

ακαθόριστης μορφής, μέσα στην εικόνα. Αν δεν αφαιρεθεί, θα λειτουργήσει ανασταλτικά στον εντοπισμό θέσης των δακτύλων.



Σχήμα 5.5.2 1: Καρέ κ_λ , 2: Δυαδική μάσκα δ_λ που παράγεται από το καρέ κ_λ μέσω του στατιστικού δερματικού μοντέλου

5.5.1 Απλοποίηση δυαδικής μάσκας

Η εφαρμογή του στατιστικού δερματικού μοντέλου είναι ατελής. Αυτό έχει ως αποτέλεσμα ορισμένες δερματικές περιοχές να απορρίπτονται ως μη δερματικές, και άλλες μη δερματικές, να θεωρείται ότι έχουν χρώμα δέρματος. Ένας από τους λόγους που δικαιολογεί την ατέλεια αυτή είναι η ιδιαίτερη στάση του χεριού όταν παίζει πιάνο. Η παλάμη του πιανίστα δεν εκτείνεται πλήρως αλλά κατά το ήμισυ, αφού τα δάκτυλα είναι ημίκλειστα.

Ένα άλλο μειονέκτημα, με δεδομένο ότι κοιτάζουμε από μπροστά, είναι ότι η περιοχή ακριβώς πίσω από τα ημίκλειστα δάκτυλα είναι το εσωτερικό της παλάμης, δηλαδή επίσης δερματική περιοχή. Το στοιχείο αυτό δυσκολεύει την εξαγωγή της φιγούρας των άκρων τα δακτύλων και κατ' επέκταση τον εντοπισμό της θέσης τους. Συνεπώς, σε αρκετές περιπτώσεις το χέρι ανιχνεύεται ως μια ενιαία μάζα, ενώ υπάρχει αυξημένη δυσκολία να διακριθούν τα δάκτυλα, ειδικά όταν μειώνεται πολύ η απόσταση μεταξύ τους.

Για όλους τους παραπάνω λόγους καθίσταται δύσκολη η απευθείας ανάκτηση της φιγούρας του χεριού μέσω της δυαδικής μάσκας δ_λ . Έτσι, το σύστημα προγραμματίστηκε να εφαρμόζει μείωση θορύβου πριν την ανάκτηση της φιγούρας του χεριού. Θόρυβος στην

περίπτωση του συστήματος αναγνώρισης κινήσεων ορίζεται ως η «ανεπιθύμητη» πληροφορία.

Σύμφωνα με τη νέα τάση επεξεργασίας εικόνας, το διαχωρισμό (*decomposition*), η δυαδική μάσκα δ_λ χωρίζεται σε βασικές συνιστώσες θ_λ και χ_λ , σύμφωνα με τη σχέση:

$$\delta_\lambda = \chi_\lambda + \theta_\lambda \quad (5.5.1.1)$$

Όπου:

- δ_λ , η δυαδική μάσκα που προκύπτει από το καρέ κ_λ μέσω ανίχνευση δέρματος,
- χ_λ , η δυαδική μάσκα που περιλαμβάνει μόνο τη φιγούρα του χεριού,
- θ_λ , ο θόρυβος που περιλαμβάνεται στη δ_λ .

Κατά συνέπεια, με βάση την παραπάνω σχέση, η συνιστώσα της φιγούρας του χεριού μπορεί να μοντελοποιηθεί ως η συνάρτηση της διαφοράς, δηλαδή:

$$x_\lambda = \delta_\lambda - \theta_\lambda \quad (5.5.1.2)$$

Από την παραπάνω σχέση καθίσταται προφανές ότι, η μείωση του θορύβου οδηγεί στην ανάκτηση της φιγούρας του χεριού.

Το σύστημα αναγνώρισης δακτυλικών κινήσεων εφαρμόζει στη δυαδική μάσκα δ_λ ένα μοντέλο αποθορυβοποίησης, το οποίο αφήνει ανεπηρέαστη τη δομή της μάσκας. Το μοντέλο αποθορυβοποίησης που εφαρμόζεται είναι το σειριακό εναλλασσόμενο φιλτράρισμα με ανακατασκευή. Το φιλτράρισμα αυτό περιλαμβάνει μια σειρά από επαναλήψεις λειτουργιών μαθηματικής μορφολογίας.

5.5.1.1 Μείωση θορύβου με σειριακό εναλλασσόμενο φιλτράρισμα

Στο σχήμα 5.4.5.2, παρατηρείται θόρυβος σε κάθε ένα παράδειγμα δυαδικής μάσκας. Ο θόρυβος προσδιορίζεται σαν ένα σύνολο λευκών περιοχών ή στιγμάτων που δεν ανήκουν στο χέρι του πιανίστα και αποτελεί εμπόδιο στην εξαγωγή των δακτύλων και τον εντοπισμό της θέσης τους

Οι λειτουργίες της μαθηματικής μορφολογίας σε δυαδικές μάσκες είναι η πλέον ενδεδειγμένη τακτική για τη μείωση θορύβου. Οι μορφολογικές λειτουργίες επηρεάζουν το σχήμα των περιοχών. Ολόκληρη η επεξεργασία γίνεται τοπικά. Δηλαδή περιοχές ή μορφές

αντικειμένων επηρεάζονται με τοπικό τρόπο. Στο σύστημα αναγνώρισης κινήσεων, για την απλοποίηση της εικόνας, χρησιμοποιήθηκαν οι παρακάτω μορφολογικές λειτουργίες:

- Μεγέθυνση – Διαστολή αντικειμένων (*Dilation*),
- Σμίκρυνση – Συστολή αντικειμένων (*Erosion*),
- Άνοιγμα – Ομαλοποιεί την κλειστή γραμμή (*σύνορο*) ενός αντικειμένου, σπάει στενούς ισθμούς και εξαλείφει λεπτές προεξοχές (*Opening*),
- Κλείσιμο – Συγχωνεύει στενά σπασίματα και μακριές λεπτές κοιλότητες, εξαλείφει μικρές τρύπες και γεμίζει κενά (*Closing*).

Η εφαρμογή των μορφολογικών λειτουργιών πάνω σε μια δυαδική μάσκα πραγματοποιείται μέσω κατάλληλων φίλτρων. Για τη μείωση θορύβου χρησιμοποιήθηκε η τεχνική του σειριακού εναλλασσόμενου φιλτραρίσματος (ΣΕΦ) [Serra 1998; Sternberg 1986]. Πρόκειται για τη σειριακή εναλλακτική εφαρμογή φίλτρων γενικευμένου ανοίγματος και κλεισίματος πάνω στη δυαδική μάσκα δ_λ , όπως φαίνεται από την παρακάτω εξίσωση:

$$\Psi_{\Sigma\Phi} = \phi_v \left(\gamma_v \left(\cdot \phi_2 \left(\gamma_2 \left(\phi_1 \left(\gamma_1 \left(\delta_\lambda \right) \right) \right) \right) \right) \right) \right) \quad (5.5.1.1.1)$$

Όπου:

- $\Psi_{\Sigma\Phi}$, το αποτέλεσμα της εφαρμογής του σειριακού εναλλασσόμενου φιλτραρίσματος,
- v , η τάξη του φίλτρου,
- γ_v , γενικευμένο άνοιγμα τάξης v ,
- ϕ_v , γενικευμένο κλείσιμο τάξης v ,
- δ_λ , η δυαδική μάσκα.

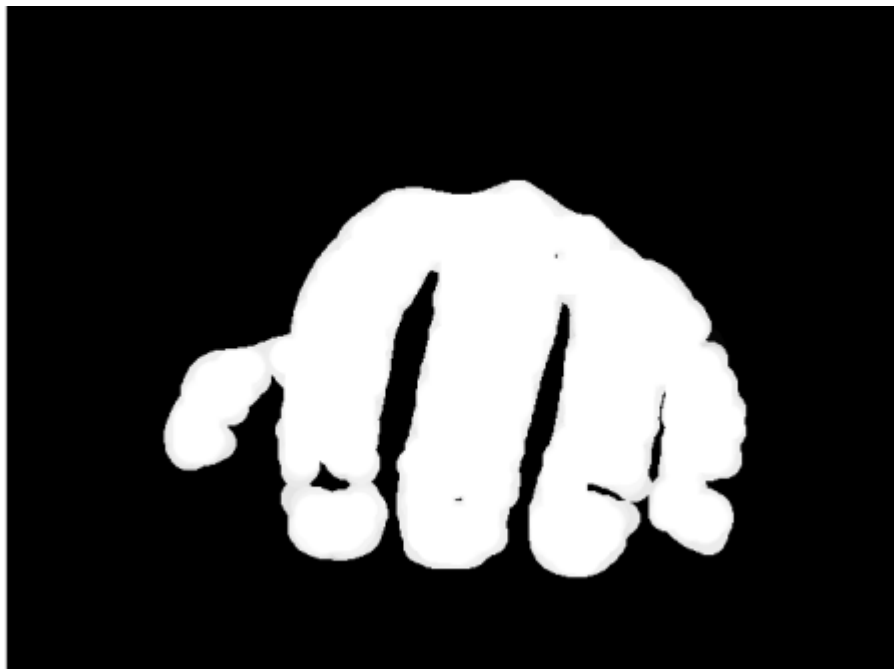
Κατ' ουσία, το ΣΕΦ θα εφαρμόσει ζεύγη φίλτρων ανοίγματος-κλεισίματος, σταδιακά αυξανόμενης τάξης. Η μέγιστη τάξη φιλτραρίσματος που θα χρησιμοποιηθεί εξαρτάται άμεσα από την ποιότητα του βίντεο στο οποίο θα εφαρμοστεί.

Στη δ_λ χρησιμοποιήθηκε φιλτράρισμα τάξης $v=10$ και το αποτέλεσμα φαίνεται στο σχήμα 5.5.1.1.1. Ο βασικός στόχος ο οποίος

υλοποιήθηκε μέσω του ΣΕΦ είναι το μέγλωμα της μαύρης περιοχής. Το στατιστικό δερματικό μοντέλο, όντας ατελές, θεωρεί ως δέρμα περιοχές του παρασκηνίου. Αυτές οι περιοχές πρέπει να περιορίζονται.

Το φίλτρο open ελλειπτικής μορφής που χρησιμοποιείται από το ΣΕΦ, επιτυγχάνει μορφολογικό άνοιγμα στις δυαδικές μάσκες. Διενεργείται συστολή των αντικειμένων της εικόνας και ακολουθεί διαστολή του αποτελέσματος. Αξίζει να σημειωθεί το γεγονός ότι, η διαστολή και η συστολή δεν είναι συμπληρωματικές λειτουργίες, δηλαδή η μία δεν αναιρεί την άλλη. Αντίστοιχα, το φίλτρο close που χρησιμοποιείται από το ΣΕΦ, διενεργεί μορφολογικό κλείσιμο στις εικόνες. Εφαρμόζεται διαστολή των αντικειμένων της εικόνας ακολουθούμενη από συστολή του αποτελέσματος.

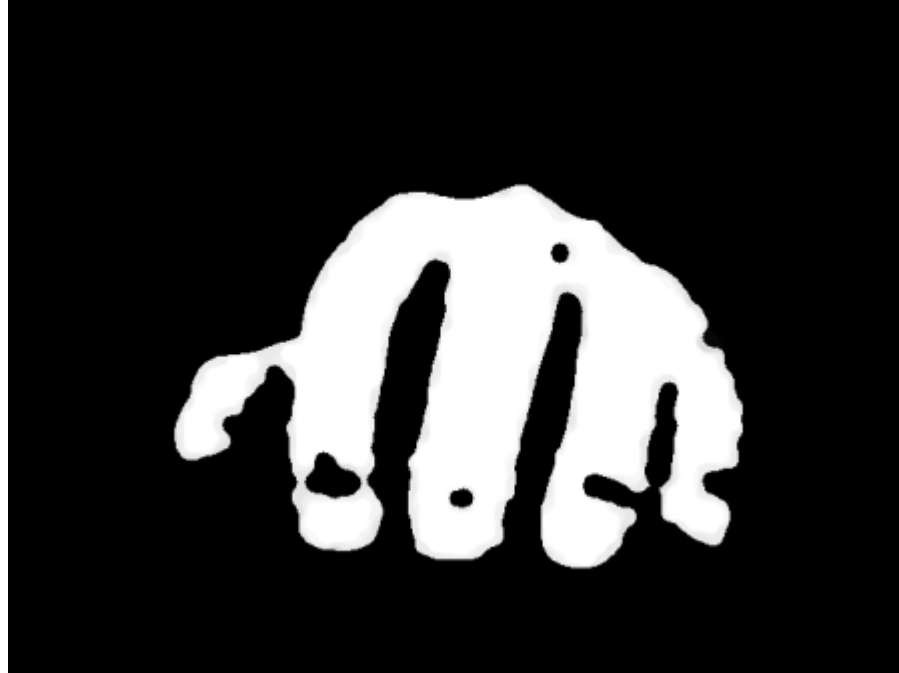
Συγκρίνοντας τα σχήματα 5.5.2 και 5.5.1.1.1 διαπιστώνουμε ότι με την εφαρμογή του ΣΕΦ γίνεται καλύτερος διαχωρισμός των δακτύλων και απαλείφονται εντελώς οι κακώς εννοούμενες δερματικές περιοχές. Το αποτέλεσμα είναι εντυπωσιακό ιδίως, όταν τα δάκτυλα στη δυαδική μάσκα δεν ξεχωρίζουν μεταξύ τους. Παράλληλα, παρατηρούμε ότι οι μικρές μαύρες περιοχές που υπήρχαν μέσα στην περιοχή του χεριού, μεγάλωσαν λόγω του μορφολογικού ανοίγματος. Το γεγονός αυτό δεν είναι ανησυχητικό διότι σε επόμενο στάδιο απαλείφεται αυτή η ατέλεια.



Σχήμα 5.5.1.1.1 Εφαρμογή ΣΕΦ στη δλ

5.5.1.2 Ανάκτηση φιγούρας χεριού

Με την εφαρμογή του ΣΕΦ πάνω στη δ_λ , επιτεύχθηκε μείωση θορύβου αλλά και, όπως προκύπτει από το σχήμα 5.5.1.2.1, ανάκτηση της φιγούρας του χεριού. Η φιγούρα του χεριού επιδέχεται βελτιώσεως και οι τεχνικές που θα εφαρμοστούν αναλύονται παρακάτω.



Σχήμα 5.5.1.2.1 Πρώτη εφαρμογή διαστολής στη χ_λ μέσω του φίλτρου min

Το γεγονός ότι η εφαρμογή μια τεχνικής μείωσης θορύβου απέδωσε τόσο, ώστε να ανακτηθεί και το χέρι παράλληλα, αποδεικνύει τη σωστή λειτουργία του στατιστικού δερματικού μοντέλου. Διότι, όταν το δερματικό μοντέλο λειτουργεί καλά, το αποτέλεσμα που παράγεται είναι το αντικείμενο του προσκηνίου (χέρι) μαζί με κάποιο θόρυβο εύκολα αφαιρέσιμο.

Με βάση τον παραπάνω συλλογισμό, προκύπτει το συμπέρασμα ότι για το σύστημα αναγνώρισης δακτυλικών κινήσεων ισχύει:

$$\chi_\lambda = \Psi_{\Sigma\Phi} \quad (5.5.1.2.1)$$

Με εναλλαγή φίλτρων min και Gauss επιτυγχάνεται άνοιγμα (Σχήμα 5.5.1.2.1) των δακτύλων και θάμπωμα της χ_λ . Το φίλτρο min,

είναι ένα μη γραμμικό φίλτρο τάξης το οποίο κάνει χρήση μασκών προσομοιώνοντας διαδικασίες διαστολής και διάβρωσης. Μέσα από παρατήρηση του σχήματος 5.5.1.2.1, προκύπτει ακόμα μεγαλύτερη επέκταση της μαύρης περιοχής και συρρίκνωση της άσπρης. Πρακτικά, με την εφαρμογή αυτού του φίλτρου εξασφαλίζεται ένας καλός διαχωρισμός δακτύλων ο οποίος θα είναι χρήσιμος για τη συνέχεια, ειδικά για τις περιπτώσεις που σε κάποιο καρέ τα δάκτυλα είναι σχεδόν κολλημένα το ένα με το άλλο.

Στη συνέχεια, εφαρμόζεται φίλτρο Gauss (Σχήμα 5.5.1.2.2) με στόχο να αποκτήσει το χέρι πιο ομοιόμορφο περίγραμμα χωρίς οπές. Είναι ένα χαμηλοπερατό φίλτρο το οποίο προσδίδει θάμπωμα στην εικόνα. Η τεχνική του συγκεκριμένου φίλτρου λειτουργεί καλά για φίλτρα τάξης μεγαλύτερης του 10, λόγω της υπολογιστικής πολυπλοκότητας. Εναλλακτικά στην εφαρμογή ενός φίλτρου Gauss ανώτερης τάξης, μπορούν να χρησιμοποιηθούν επαναληπτικά φίλτρα μικρότερης τάξης. Το αποτέλεσμα ενός τέτοιου φιλτραρίσματος είναι ένα αρκετά ομαλό περίγραμμα του χεριού χωρίς απότομες γωνίες, εσοχές ή εξοχές. Παρατηρείται επίσης, εκ του σχήματος 5.5.1.2.2, ότι κάποιες μαύρες τρύπες που υπήρχαν στο εσωτερικό του χεριού, ιδίως στα άκρα των δακτύλων, μειώνονται αισθητά. Τέλος, οι αποστάσεις μεταξύ των δακτύλων μειώνονται χωρίς τα δάκτυλα να κολλούν, χάρι σε προηγούμενη εφαρμογή διαστολής.



Σχήμα 5.5.1.2.2 Πρώτη εφαρμογή του φίλτρου Gauss στη χ

Το γεγονός ότι τα δάκτυλα ήρθαν πολύ κοντά μετά την ομαλοποίηση, υποχρεώνει την επανάληψη της λειτουργίας της διαστολής. Συνεπώς, νέο φίλτρο min εφαρμόστηκε στη χ_λ , έχοντας ως αποτέλεσμα καλύτερο διαχωρισμό των δακτύλων και κυρίως των άκρων τους (Σχήμα 5.5.1.2.3).

Νέο φίλτρο Gauss εφαρμόζεται για να αποκατασταθεί το ομαλό περίγραμμα του χεριού και να εξαλειφθούν οι απότομες γωνίες. Το αποτέλεσμα αυτής της εφαρμογής φαίνεται στο σχήμα 5.5.1.2.4.



Σχήμα 5.5.1.2.3 Δεύτερη εφαρμογή διαστολής στη χ_λ μέσω του φίλτρου min

Η κατωφλίωση είναι το τελευταίο στάδιο της διαδικασίας ανάκτησης της φιγούρας του χεριού. Η εικόνα του χεριού έχει υποστεί μια διαδικασία φιλτραρίσματος, μέσω της οποίας έχει φτάσει σε κατάλληλη μορφή ώστε με την κατωφλίωση να μπορέσει να ανακτηθεί πλήρως.

Η βασική ιδέα της κατωφλίωσης στηρίζεται στο γεγονός ότι κάθε αντικείμενο της εικόνας με σχετικά μεγάλο μέγεθος δημιουργεί μια κατανομή εικονοστοιχείων στο ιστόγραμμα της εικόνας γύρω από τη μέση τιμή φωτεινότητας του (μ_0). Πρέπει να σημειωθούν δύο σημαντικά στοιχεία: α) με την εφαρμογή του φίλτρου Gauss πάνω στη

χ_λ , η εικόνα παύει να είναι δυαδική και εισέρχεται ξανά η συνιστώσα της φωτεινότητας και β) στη χ_λ όντως το μοναδικό αντικείμενο, με μέγεθος αρκετά μεγάλο, είναι το χέρι του πιανίστα, όποτε μπορεί να εφαρμοστεί η κατωφλίωση.



Σχήμα 5.5.1.2.4 Δεύτερη εφαρμογή του φίλτρου Gauss στη χ_λ

Συνεπώς, αν σ_0 είναι η τυπική απόκλιση της φωτεινότητας από τη μέση τιμή φωτεινότητας μ_0 του αντικειμένου, τότε η περιοχή που καλύπτει το αντικείμενο θα προσεγγίζεται από την περιοχή με τιμή $= 1$ της δυαδικής εικόνας $\chi_\lambda^k[\mu, \nu]$ ($\mu=0,1,\dots,M-1$, με $M=640$ και $\nu=0,1,\dots,N-1$, με $N=480$).

Το αποτέλεσμα της κατωφλίωσης είναι η δυαδική εικόνα χ_λ^k , στην οποία απεικονίζεται με κόκκινο χρώμα η φιγούρα του χεριού και με μαύρο το παρασκήνιο. Όπως παρατηρείται (Σχήμα 5.5.1.2.5), το ομαλό περίγραμμα του χεριού, δίχως γωνίες και οπές, έχει διατηρηθεί μετά την εφαρμογή της κατωφλίωσης. Αξίζει να σημειωθεί ότι το κόκκινο χρώμα στο χέρι είναι μια ιδιαιτερότητα της εργαλειοθήκης DIPimage του MATLAB. Με τη σύγκριση των δύο τελευταίων εικόνων

Όπως αναφέρθηκε στην ενότητα 5.5, στόχος της τμηματοποίησης είναι ο εντοπισμός θέσης των άκρων των δακτύλων. Ο εντο-

πισμός θέσης των άκρων μπορεί να επιτευχθεί μέσω της ανάκτησης του περιγράμματος του χεριού. Το περίγραμμα του χεριού μπορεί να ανακτηθεί με τη χρήση κατάλληλων αλγόριθμων ανίχνευσης ακμών.



Σχήμα 5.5.1.2.5 Κατωφλίωση στη χ_λ (το κόκκινο χρώμα είναι ιδιαιτερότητα του DIPimage)

5.5.2 Διαχωρισμός χεριού σε βασικές συνιστώσες

Στόχος της διαδικασίας τμηματοποίησης του χεριού είναι ο εντοπισμός των άκρων των δακτύλων. Σύμφωνα πάντα με την τάση διαχωρισμού της εικόνας σε βασικές συνιστώσες, που παρουσιάστηκε επιγραμματικά στην ενότητα 5.5, παρουσιάζεται παρακάτω η ακολουθούμενη από το σύστημα μέθοδος για τον εντοπισμό των άκρων των δακτύλων.

Θεωρώντας ως είσοδο την εικόνα της φιγούρας του χεριού χ_λ , παράγονται ως έξοδοι οι ανακτημένες εικόνες α) του περιγράμματος του χεριού (π_λ) και β) των άκρων των δακτύλων (α_λ).

5.5.2.1 Ανάκτηση περιγράμματος χεριού

Η ανάκτηση του περιγράμματος του χεριού από το σύστημα αναγνώρισης δακτυλικών κινήσεων υλοποιείται με ανίχνευση ακμών. Όπως έχει ήδη αναφερθεί στο εδάφιο 3.7.2, οι ακμές οριοθετούν τα αντικείμενα σε μια εικόνα δίνοντας πληροφορίες κίνησης και συνδράμοντας στην εξαγωγή χαρακτηριστικών διανυσμάτων. Η λειτουργία τους βασίζεται στον εντοπισμό των τοπικών μεταβολών σε μια εικόνα γκρι αποχρώσεων.

Το σύστημα κάνει χρήση της έννοιας της κλίσης της εικόνας προκειμένου να ανιχνεύσει ακμές και κατά συνέπεια να ανακτήσει το περίγραμμα π_λ του χεριού. Είναι γνωστό από τα μαθηματικά ότι, η κλίση μιας συνεχούς συνάρτησης $f(x,y): \mathbb{R} \rightarrow \mathbb{R}^2$ είναι ένα διάνυσμα στο επίπεδο xy που ορίζεται σύμφωνα με τη σχέση:

$$\nabla f(x,y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \quad (5.5.2.1.1)$$

Το μέτρο του διανύσματος κλίσης σε ένα σημείο (x,y) μας δείχνει το πόσο απότομη είναι η ακμή και δίνεται από τη σχέση:

$$|\nabla f(x,y)| = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} \quad (5.5.2.1.2)$$

Η κλίση στο σημείο (x,y) είναι μέγιστη αν κινηθούμε κατά τη διεύθυνση του xy επιπέδου που προσδιορίζει η γωνία ϕ σύμφωνα με τη σχέση:

$$\phi = \arctan \left(\frac{\partial f / \partial y}{\partial f / \partial x} \right) \quad (5.5.2.1.3)$$

Τι γίνεται όμως στην περίπτωση της ψηφιακής εικόνας χ_λ , η οποία δεν αποτελεί συνεχή συνάρτηση; Στην περίπτωση αυτή, αντί των μερικών παραγώγων, το σύστημα υπολογίζει τις διαφορές της τιμής ενός εικονοστοιχείου με τα γειτονικά του, οριζοντίως και καθέτως.

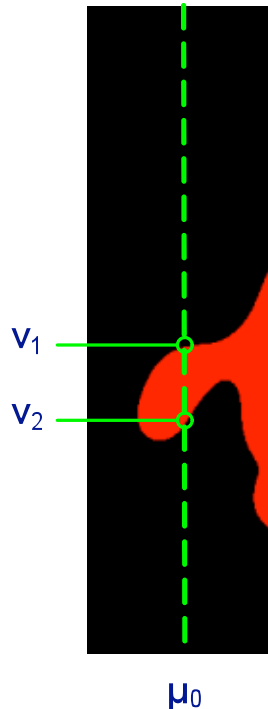
Αν εφαρμοστεί λοιπόν η προσέγγιση αυτή σε τυχαίο εικονοστοιχείο (μ, ν) με το προηγούμενό του στην ψηφιακή εικόνα, τότε η κλίση σ' αυτό θα καθορίζεται από το διάνυσμα:

$$\nabla \chi_\lambda[\mu, \nu] = (\nabla_\mu \chi_\lambda[\mu, \nu], \nabla_\nu \chi_\lambda[\mu, \nu]), \text{ με}$$

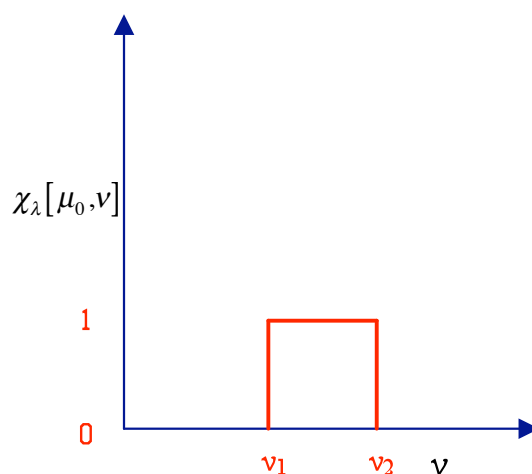
$$\nabla_\mu \chi_\lambda[\mu, \nu] = \chi_\lambda[\mu, \nu] - \chi_\lambda[\mu - 1, \nu] \text{ και}$$

$$\nabla_\nu \chi_\lambda[\mu, \nu] = \chi_\lambda[\mu, \nu] - \chi_\lambda[\mu, \nu - 1]. \quad (5.5.2.1.4)$$

Σαρώνοντας την εικόνα χ_λ κατά τον κατακόρυφο άξονα ν (Σχήμα 5.5.2.1.1) εξάγεται ένας παλμός της μορφής του σχήματος 6-. Η εξαγωγή του παλμού δικαιολογείται από το γεγονός ότι η χ_λ είναι δυαδική εικόνα και από το ότι αποτελείται μόνο από το χέρι και το παρασκήνιο. Συνεπώς, καθώς σαρώνεται η περιοχή του παρασκηνίου, τα εικονοστοιχεία έχουν την τιμή 0. Στο σημείο ν_1 (Σχήμα 5.5.2.1.1 – 5.5.2.1.2) η τιμή των εικονοστοιχείων αλλάζει καθώς ξεκινάει η περιοχή του χεριού, για να ξαναμηδενιστεί στο σημείο ν_2 και μέχρι το τέλος.



Σχήμα 5.5.2.1.1 Κατακόρυφη σάρωση της χ_λ στη θέση μ_0



Σχήμα 5.5.2.1.2 Ενδεικτική απεικόνιση των τιμών των εικονοστοιχείων της χ_λ στη θέση μ_0

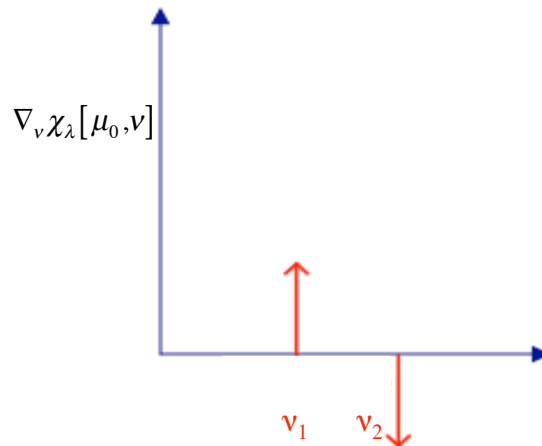
Ας εφαρμόσουμε την προσέγγιση αυτή στο εικονοστοιχείο που βρίσκεται στη θέση (μ_0, ν_1) με το προηγούμενό του καθώς και στο εικονοστοιχείο (μ_0, ν_2) με το προηγούμενό του. Η κλίση στο πρώτο εικονοστοιχείο θα καθορίζεται από το διάνυσμα:

$$\begin{aligned}\nabla \chi_\lambda[\mu_0, \nu_1] &= (\nabla_{\mu_0} \chi_\lambda[\mu_0, \nu_1], \nabla_{\nu_1} \chi_\lambda[\mu_0, \nu_1]) \Rightarrow \\ \nabla_{\nu_1} \chi_\lambda[\mu_0, \nu_1] &= \chi_\lambda[\mu_0, \nu_1] - \chi_\lambda[\mu_0, \nu_1 - 1] \Rightarrow \\ \nabla_{\nu_1} \chi_\lambda[\mu_0, \nu_1] &= 1 - 0 = 1\end{aligned}\quad (5.5.2.1.5)$$

Η κλίση στο δεύτερο εικονοστοιχείο θα καθορίζεται από το διάνυσμα:

$$\begin{aligned}\nabla \chi_\lambda[\mu_0, \nu_2] &= (\nabla_{\mu_0} \chi_\lambda[\mu_0, \nu_2], \nabla_{\nu_2} \chi_\lambda[\mu_0, \nu_2]) \Rightarrow \\ \nabla_{\nu_2} \chi_\lambda[\mu_0, \nu_2] &= \chi_\lambda[\mu_0, \nu_2] - \chi_\lambda[\mu_0, \nu_2 - 1] \Rightarrow \\ \nabla_{\nu_2} \chi_\lambda[\mu_0, \nu_2] &= 0 - 1 = -1\end{aligned}\quad (5.5.2.1.6)$$

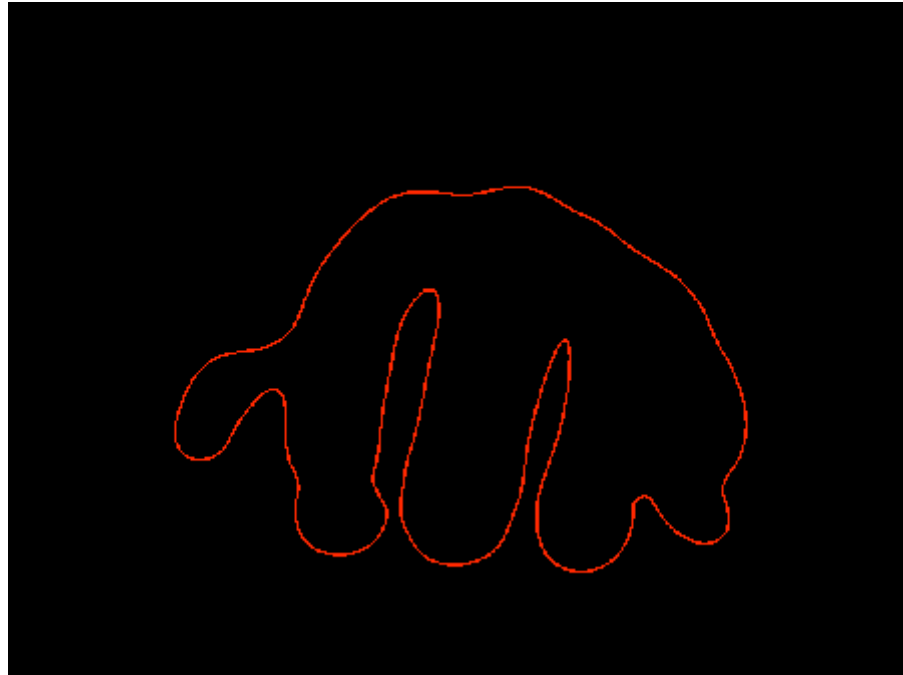
Σε όλες τις ενδιάμεσες θέσεις μεταξύ των v_1 και v_2 , η κλίση παραμένει ίση με το 0. Η κλίση της χ_λ ως προς τον κατακόρυφο άξονα απεικονίζεται στο σχήμα 5.5.2.1.3. Το συμπέρασμα είναι ότι χρησιμοποιώντας την κλίση ενδεικτικά για τη θέση μ_0 στον οριζόντιο άξονα, ανακτήθηκαν τα δύο αντίστοιχα σημεία του κατακόρυφου άξονα τα οποία ανήκουν στο περίγραμμα του χεριού. Άρα, τα εικονοστοιχεία $\chi_\lambda[\mu_0, v_1]$ και $\chi_\lambda[\mu_0, v_2]$ ανήκουν στο περίγραμμα του χεριού.



Σχήμα 5.5.2.1.3 Ενδεικτική κλίση της χ_λ στη θέση μ_0

Το σύστημα σαρώνει την εικόνα χ_λ κατά τον οριζόντιο και τον κατακόρυφο άξονα και ανιχνεύει σημεία τα οποία έχουν κλίση 1 ή -1 και τα συγκαταλέγει στο περίγραμμα του χεριού. Το σύνολο των σημείων αυτών παράγουν τη δυαδική εικόνα του περιγράμματος π_λ (Σχήμα 5.5.2.1.4).

Η ανάκτηση του περιγράμματος δοκιμάστηκε και από τον αλγόριθμο του Canny, ο οποίος λειτούργησε με μεγάλη επιτυχία καταναλώνοντας όμως μεγάλη υπολογιστική ισχύ χωρίς να κρίνεται απαραίτητο με δεδομένο ότι πρόκειται για ανίχνευση ακμών σε δυαδική εικόνα.



Σχήμα 5.5.2.1.4 Ανάκτηση περιγράμματος μέσω υπολογισμού της κλίσης της χ_λ

5.5.2.2 Ανάκτηση άκρων δακτύλων

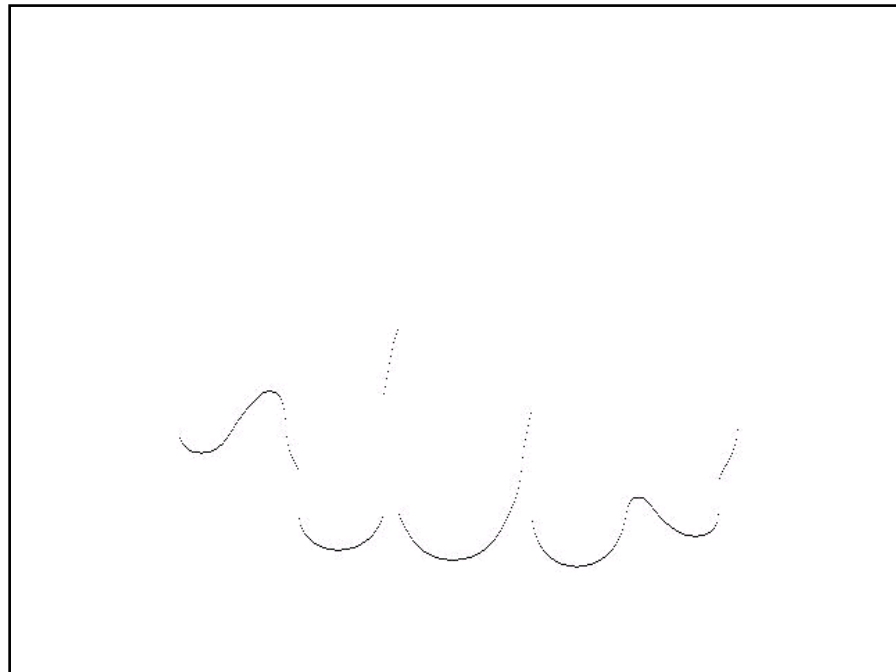
Το τελευταίο βήμα της διαδικασίας τμηματοποίησης του χεριού είναι η ανάκτηση των άκρων των δακτύλων. Η αναγκαιότητα αυτού του βήματος έγκειται στο γεγονός ότι, έχοντας ως στόχο τον εντοπισμό της θέσης των άκρων των δακτύλων, είναι πολύ μεγάλο το πλήθος των εικονοστοιχείων της π_λ που δε συμμετέχουν στην διαδικασία αυτή, αλλά μπορεί να προκαλέσουν σφάλμα στον εντοπισμό της θέσης των άκρων, αυξάνοντας παράλληλα την απαιτούμενη υπολογιστική ισχύ (Σχήμα 5.5.2.2.1).

Στοχεύοντας στη μείωση της απαιτούμενης υπολογιστικής ισχύος αλλά και την πιθανότητας σφάλματος στον εντοπισμό της θέσης των άκρων των δακτύλων, εφαρμόστηκε κατακόρυφη σάρωση της εικόνας π_λ . Έτσι, για κάθε θέση $\mu=\{0,1,2,\dots,M-1\}$, διατηρούνται μόνο τα εικονοστοιχεία τα οποία παρουσιάζουν μέγιστο στο v , παράγοντας τη δυαδική μάσκα α_λ . Διατηρούνται μόνο αυτά, διότι παρουσιάζουν αυξημένη πιθανότητα να ανήκουν στα άκρα των δακτύλων, έναντι των υπολοίπων. Σε αυτό το σημείο είναι σκόπιμο να σημειωθεί ότι για το MATLAB, οι κατακόρυφοι άξονες στην εικόνα μηδενίζο-

νται στην πάνω αριστερή γωνία της. Συνεπώς, δικαιολογείται με αυτόν τον τρόπο ο υπολογισμός των μεγίστων στον κατακόρυφο άξονα.

5.6 Εντοπισμός δακτύλων

Αλγόριθμοι εντοπισμού θέσης δακτύλων παρουσιάστηκαν στην ενότητα 3.8. Οι δημοφιλέστεροι αλγόριθμοι που θα μπορούσαν να εφαρμοστούν κάνουν χρήση: α) προβολής υπογραφών, β) εξαγωγής γεωμετρικών ιδιοτήτων, γ) κυκλικού μετασχηματισμού Hough, δ) χρωματικών δεικτών. Από αυτούς, μόνο δύο θα μπορούσαν να εφαρμοστούν από το σύστημα αναγνώρισης δακτυλικών κινήσεων στον πιάνο.



Σχήμα 5.5.2.2.1 Ανάκτηση άκρων δακτύλων στην a_k

Ο αλγόριθμος ο οποίος χρησιμοποιεί χρωματικούς δείκτες στα δάκτυλα απορρίπτεται λόγω του ότι αντίκειται στον αρχικό στόχο που τέθηκε στην ενότητα 1.5, ότι δηλαδή η μέθοδος πρέπει να είναι *μη περιοριστική*. Ο αλγόριθμος του κυκλικού μετασχηματισμού Hough απορρίπτεται διότι α) χρειάζεται επιπλέον φιλτράρισμα ώστε να απορριφθούν σφάλματα στον εντοπισμό ημικυκλικών περιοχών που δεν ανήκουν στα άκρα των δακτύλων και β) το εξαγόμενο απο-

τέλεσμα αποτελείται από ζεύγη πιθανών θέσεων για άκρα και κοιλάδες τα οποία δεν μπορούν να ταξινομηθούν και να διαχωριστούν εύκολα και γρήγορα. Ομοίως με τον πρώτο λόγο απόρριψης του μετασχηματισμού Hough και για τον αλγόριθμο προβολών υπογραφής. Κατά συνέπεια, σύμφωνα με έρευνα που διεξήχθη από την Α. Μ. Burns (2006), για τον εντοπισμό θέσης δακτύλων του κιθαρίστα, οι αλγόριθμοι, που κάνουν χρήση γεωμετρικών ιδιοτήτων της εικόνας, πλεονεκτούν έναντι των υπολοίπων στη συντριπτική πλειοψηφία των δοκιμών που πραγματοποιήθηκαν. Για αυτούς ακριβώς του λόγους επιλέχτηκε ο δεύτερος αλγόριθμος.

Ο αλγόριθμος εντοπισμού δακτύλων που χρησιμοποιείται από το σύστημα, εκμεταλλεύεται τις γεωμετρικές ιδιότητες που έχουν οι εικόνες του περιγράμματος π_λ και των άκρων των δακτύλων α_λ . Χρησιμοποιείται ένα σημείο αναφοράς το οποίο υπολογίζεται μέσω του κεντροειδούς της εικόνας ϕ_λ , το οποίο είναι ένα σταθερό σημείο της εικόνας. Αποστάσεις μεταξύ του σημείου αναφοράς και των εικονοστοιχείων της εικόνας α_λ υπολογίζονται, έχοντας ως εξαγόμενα αποτελέσματα τα πέντε μέγιστα των αποστάσεων που αντιστοιχούν στα πέντε άκρα των δακτύλων.

Παρόμοιοι αλγόριθμοι εντοπισμού θέσης δακτύλων παρουσιάστηκαν στην ενότητα 3.8. Οι δημοφιλέστεροι αλγόριθμοι που θα μπορούσαν να εφαρμοστούν κάνουν χρήση: α) προβολών υπογραφής, β) γεωμετρικών ιδιοτήτων, γ) κυκλικού μετασχηματισμού Hough, δ) έγχρωμων δεικτών. Σύμφωνα με έρευνα που διεξήχθη από την Α. Μ. Burns (2006), για τον εντοπισμό θέσης δακτύλων του κιθαρίστα, οι αλγόριθμοι, που κάνουν χρήση γεωμετρικών ιδιοτήτων της εικόνας, πλεονεκτούν έναντι των υπολοίπων στη συντριπτική πλειοψηφία των δοκιμών που πραγματοποιήθηκαν.

Στην ενότητα αυτή θα παρουσιαστεί η μεθοδολογία που ακολουθείται από το σύστημα αναγνώρισης δακτυλικών κινήσεων για τον εντοπισμό της θέσης των δακτύλων. Το σύστημα χρησιμοποιεί τις δυαδικές εικόνες ϕ_λ και α_λ , οι οποίες ανακτήθηκαν σε προηγούμενη ενότητα, καθώς και τις συναρτήσεις Centroid Calculation Function και findpeaks του MATLAB.

5.6.1 Υπολογισμός κεντροειδούς

Σύμφωνα με την ηλεκτρονική εγκυκλοπαίδεια Wikipedia

« Το κέντρο μάζας ενός αντικειμένου, είναι το σημείο, γύρω από το οποίο, το γκρι του αντικειμένου κατανέμεται ισόποσα. Λέγεται ότι το σώμα δύναται να ισορροπήσει με μοναδι-

κή στήριξη που τοποθετείται στο κέντρο μάζας. Ειδικότερα, για τις δυαδικές εικόνες το κέντρο μάζας ονομάζεται και κεντροειδές του αντικειμένου».

Στην πραγματικότητα, το κεντροειδές είναι το μεσαίο εικονοστοιχείο από ένα σύνολο εικονοστοιχείων. Συνήθως όμως, προς διευκόλυνση της περιγραφής των διαδικασιών, το κεντροειδές θεωρείται ότι είναι ένα ζευγάρι τιμών, των συντεταγμένων του οριζόντιου και του κατακόρυφου άξονα. Η παραδοχή αυτή θα διατηρηθεί και για την περιγραφή του τρόπου εντοπισμού του κεντροειδούς στην εικόνα π_λ από το σύστημα αναγνώρισης δακτυλικών κινήσεων.

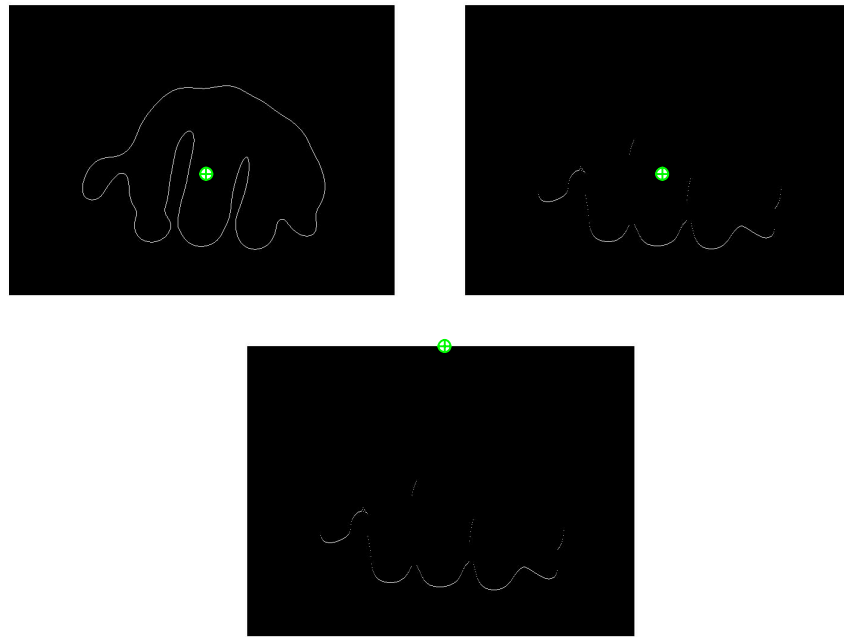
Ο τρόπος υπολογισμού του κεντροειδούς σε εικόνα, για το σύστημα αναγνώρισης δακτυλικών κινήσεων, γίνεται σε κάθε ένα καρέ χωριστά. Το κεντροειδές υπολογίζεται με βάση την εικόνα π_λ , η οποία εισάγεται στη συνάρτηση `ait_centroid` ως όρισμα και εξάγονται οι συντεταγμένες x_k και y_k με τη μορφή διανύσματος:

$$[x_k, y_k] = \text{ait_centroid}(\text{εικόνα}) \quad (5.6.1.1)$$

Το διάνυσμα θέσης που επιστρέφει η συνάρτηση απαρτίζεται από ζεύγος πραγματικών αριθμών. Επειδή στόχος είναι ο εντοπισμός του εικονοστοιχείου του κεντροειδούς, γίνεται αντιστοίχιση της θέσης (x_k, y_k) στο εικονοστοιχείο $\pi_\lambda[\mu_k, \nu_k]$. Έτσι, ο εντοπισμός της θέσης του κεντροειδούς (x_k, y_k) γίνεται αρχικά στην εικόνα π_λ και στη συνέχεια απεικονίζεται και στην α_λ μέσω του εικονοστοιχείου $\alpha_\lambda[\mu_k, \nu_k]$. (Σχήμα 5.6.1.1). Στο σχήμα 5.6.1.1 το κεντροειδές εμφανίζεται με τον πράσινο στόχο.

5.6.2 Καθορισμός σημείου αναφοράς

Αρκετές επιστημονικές έρευνες κάνουν χρήση του κεντροειδούς για τον εντοπισμό θέσης των άκρων των δακτύλων μέσω υπολογισμού αποστάσεων. Στις έρευνες αυτές η παλάμη του χεριού είναι σε πλήρη έκταση με αποτέλεσμα το κεντροειδές του χεριού να βρίσκεται εκτός δακτύλων.



Σχήμα 5.6.1.1 Εντοπισμός εικονοστοιχείου που αντιστοιχεί στο κεντροειδές για τις εικόνες α_λ και α_λ , ανάκτηση άκρων δακτύλων στην α_λ

Στο σύστημα αναγνώρισης δακτυλικών κινήσεων δοκιμάστηκε η παραπάνω μεθοδολογία, δίνοντας αρνητικά αποτελέσματα. Ο λόγος που δεν απέδωσε η μεθοδολογία αυτή οφείλεται στην ειδική στάση του χεριού του πιανίστα, όταν παίζει πιάνο. Το κεντροειδές, λόγω της ημίκλειστης στάσης του χεριού, τείνει να βρίσκεται εσωτερικά κάποιου δακτύλου. Εάν ο εντοπισμός των άκρων των δακτύλων βασιστεί στον υπολογισμό μεγίστων στις αποστάσεις κεντροειδούς και εικονοστοιχείων της α_λ , τότε παράγονται ως έξοδοι πιθανές θέσεις δακτύλων που δεν αντιστοιχούν στις πραγματικές. Μπορεί δηλαδή με άλλα λόγια να θεωρηθεί ως άκρο δακτύλου κάποιο σημείο εσωτερικά, που να ανήκει, για παράδειγμα, στην εσωτερική κοιλάδα μεταξύ δύο δακτύλων.

Τα σφάλματα αυτά εξαλείφονται αν τεθεί η συντεταγμένη του κατακόρυφου άξονα του κεντροειδούς ίση με μηδέν (Σχήμα 5.6.1.1), δηλαδή αν τεθεί ως σημείο αναφοράς του υπολογισμού των άκρων των δακτύλων το εικονοστοιχείο $\alpha_\lambda[\mu_\kappa, 0]$ με θέση $(x_\kappa, 0)$. Το συμπέρασμα αυτό δικαιολογείται από το γεγονός ότι, όσο μεγαλύτερη είναι η απόσταση του σημείου αναφοράς από τα υποψήφια εικο-

νοστοιχεία άκρων δακτύλων, τόσο περισσότερα τα υπολογισθέντα τοπικά μέγιστα θα αντιστοιχούν στα άκρα αυτά.

5.6.3 Εντοπισμός θέσης

Ο εντοπισμός της θέσης των δακτύλων υλοποιείται μέσω του υπολογισμού ευκλείδειων αποστάσεων μεταξύ του σημείου αναφοράς $(x_k, 0)$ και των θέσεων των εικονοστοιχείων της α_λ που ανήκουν στο περίγραμμα των δακτύλων (Σχήμα 5.6.3.1).

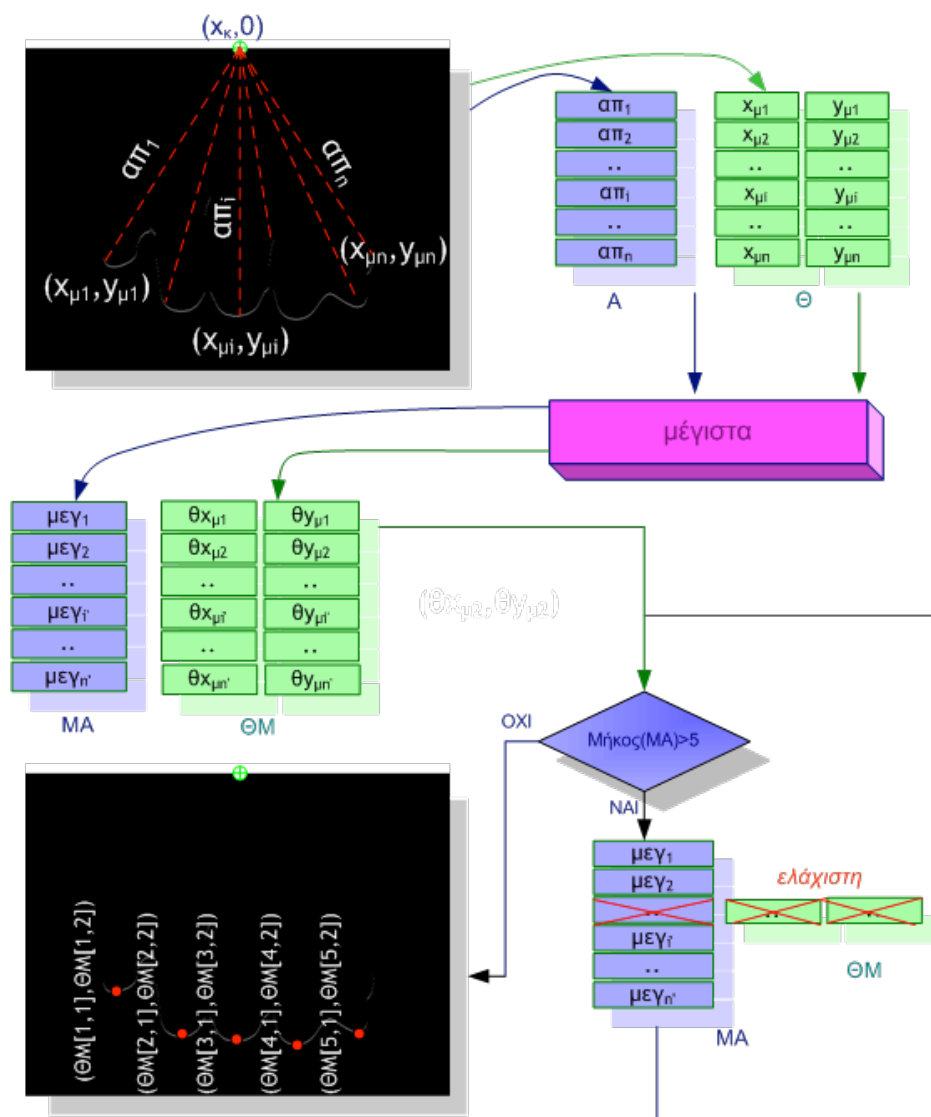
Για τα εικονοστοιχεία του περιγράμματος των δακτύλων, έστω n στο πλήθος, θα ισχύει $\alpha_\lambda[\mu, n] \neq 0$. Για κάθε ένα από τα n εικονοστοιχεία καταχωρούνται α) σε μονοδιάστατο πίνακα A , οι αποστάσεις τους από το σημείο αναφοράς και β) σε δισδιάστατο πίνακα Θ , οι θέσεις τους. Για παράδειγμα, για το εικονοστοιχείο $i = \{1, 2, \dots, n\}$ υπολογίζεται η απόστασή του α_{pi} από το $(x_k, 0)$ και εκχωρείται στον πίνακα ως εξής: $A[i] \leftarrow \alpha_{pi}$. Ομοίως, εκχωρείται η θέση του (x_{mi}, y_{mi}) στον πίνακα ως εξής: $\Theta[i, 1] \leftarrow x_{mi}$ και $\Theta[i, 2] \leftarrow y_{mi}$. Για τον υπολογισμό των ευκλείδειων αποστάσεων χρησιμοποιούνται τα τετράγωνα των συντεταγμένων θέσης του κάθε εικονοστοιχείου. Δηλαδή η απόσταση μεταξύ του εικονοστοιχείου i και του σημείου αναφοράς δίνεται από τη σχέση:

$$\alpha_{pi} = \sqrt{(x_k - x_{mi})^2 + (0 - y_{mi})^2} \quad (5.6.3.1)$$

Όπου:

- α_{pi} , η απόσταση μεταξύ του εικονοστοιχείου i και του σημείου αναφοράς,
- $(x_k, 0)$, οι συντεταγμένες θέσης του σημείου αναφοράς,
- (x_{mi}, y_{mi}) , οι συντεταγμένες θέσης του εικονοστοιχείου i .

Ο λόγος για τον οποίο δε χρησιμοποιούνται οι συντεταγμένες θέσης του εικονοστοιχείου, αλλά τα τετράγωνα τους, είναι ότι έτσι μπορεί να επιτευχθεί μετέπειτα, καλύτερος διαχωρισμός των αποστάσεων και στη συνέχεια να υπολογιστούν ευκολότερα τα τοπικά μέγιστα. Δηλαδή, το τετράγωνο των συντεταγμένων επιτρέπει την καλύτερη διάκριση των θέσεων μέσω των αποστάσεων.



Σχήμα 5.6.3.1 Αναπαράσταση διαδικασίας εντοπισμού θέσης δακτύλων

Έχοντας λοιπόν διαθέσιμους τους πίνακες A και Θ , καλείται η συνάρτηση `findpeaks` του MATLAB ώστε να υπολογίσει τις κατάλληλες αποστάσεις (Σχήματα 5.6.3.1-5.6.3.2). Στη συνάρτηση `findpeaks` θα δοθούν ως ορίσματα οι πίνακες A και Θ και θα εξαχθούν δύο νέοι πίνακες MA και ΘM , οι οποίοι θα περιέχουν τα μέγιστα των αποστάσεων και τις αντίστοιχες θέσεις στις οποίες παρατηρήθηκαν τα μέγιστα. Ειδικότερα, η σύνταξη της συνάρτησης είναι η εξής:

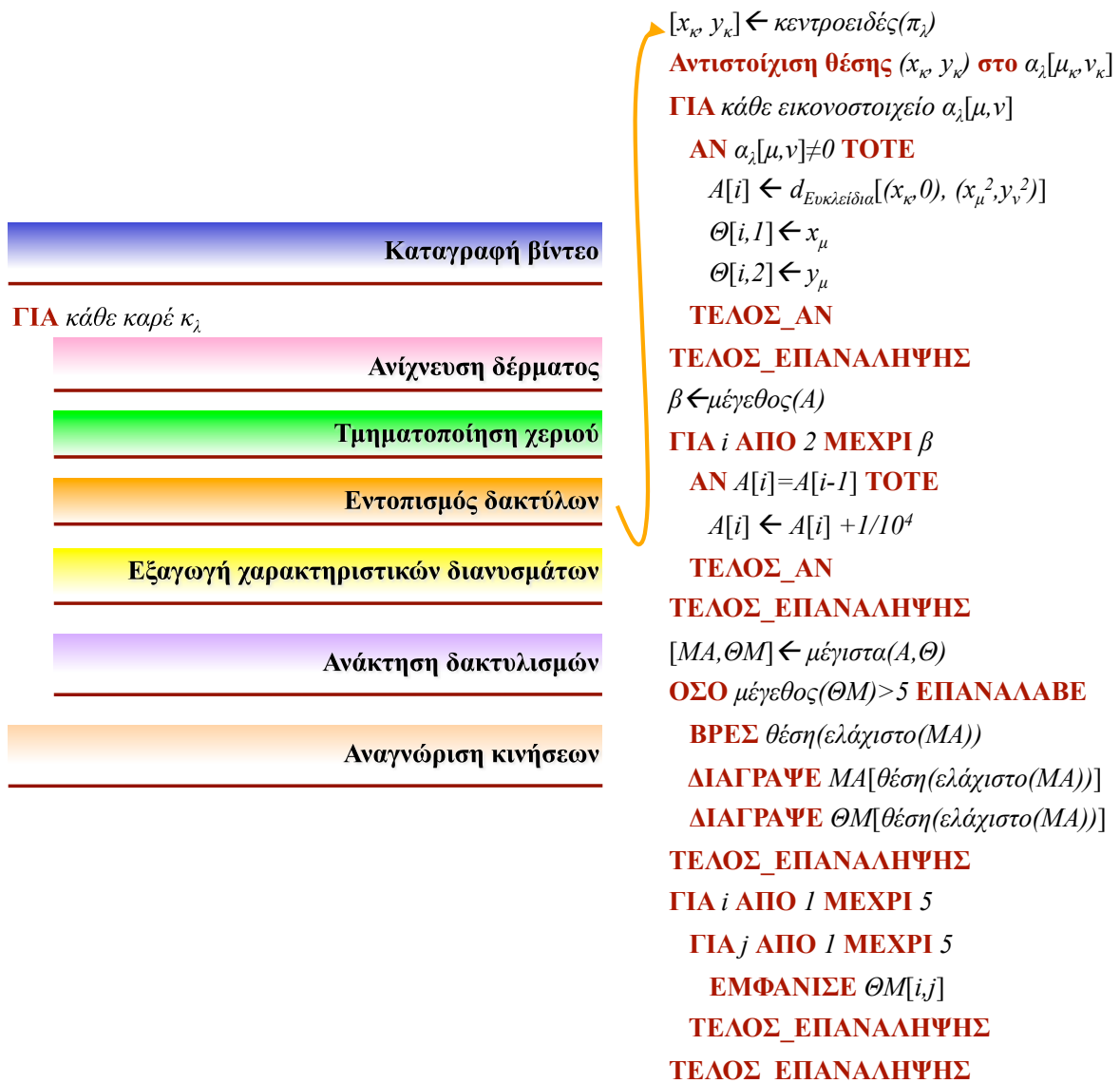
$$[MA, \Theta M] = \text{findpeaks}(A, \Theta) \quad (5.6.3.2)$$

Η συνάρτηση λειτουργεί με τον περιορισμό ότι α) υπάρχουν τουλάχιστον τρεις αποστάσεις καταχωρημένες στον πίνακα A, αλλά και β) δεν παρατηρούνται ίσες γειτονικές τιμές στον πίνακα. Αναφορικά με τον πρώτο περιορισμό είναι σχεδόν απίθανο να κατασκευαστεί πίνακας A με τρεις μόνο τιμές, διότι αυτό θα σήμαινε ότι δεν ανιχνεύτηκε καθόλου δέρμα στο καρέ του βίντεο. Ο δεύτερος περιορισμός εντούτοις, προκάλεσε σφάλματα στον υπολογισμό μεγίστων σε ελάχιστα από τα καρέ που δοκιμάστηκαν και για αυτό εισήχθηκε μια δομή επανάληψης που ελέγχει για ίσες γειτονικές τιμές στον πίνακα. Στην περίπτωση που η τιμή μια θέσης του πίνακα A είναι ίση με την προηγούμενή της, τότε προστίθεται μια ελάχιστη ποσότητα σε αυτήν ώστε να διαφέρουν κατά λίγο:

$$A[i] \leftarrow A[i] + 1/10^4 \quad (5.6.3.3)$$

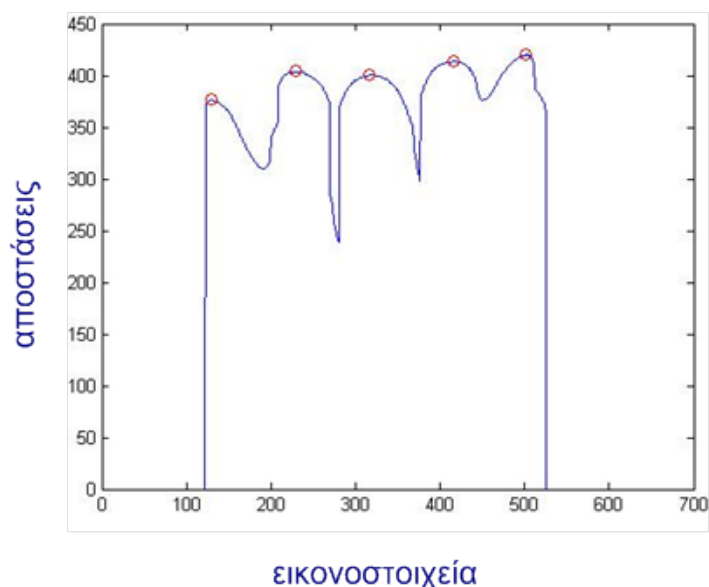
Έτσι, εξασφαλίζουμε ότι οι γειτονικές θέσεις του πίνακα δε θα έχουν ίσες τιμές, δίχως αλλοίωση του αποτελέσματος. Το πλήθος των τοπικών μεγίστων που θα επιστραφούν εξαρτάται από το πόσο «τοπικά» θέλουμε να τα αναζητήσουμε, δηλαδή από το πλήθος των συνεχόμενων τιμών μέσα από τις οποίες θα αναδειχθούν τα μέγιστα. Γενικά, εφόσον ο υπολογισμός των αποστάσεων γίνεται με βάση τα εικονοστοιχεία της εικόνας α_λ , τα οποία αναφέρονται στο μεγαλύτερο μέρος τους σε περιοχές των άκρων των δακτύλων, το πλήθος των επιστρεφόμενων μεγίστων πολύ σπάνια ξεπερνάει το πέντε.

Τι γίνεται όμως στην περίπτωση που επιστραφούν περισσότερες ή λιγότερες πιθανές θέσεις δακτύλων; Στην περίπτωση όπου επιστραφούν περισσότερες από πέντε πιθανές θέσεις δακτύλων, τότε ενεργοποιείται μια δομή επανάληψης η οποία εντοπίζει την ελάχιστη τιμή του πίνακα MA, καθώς και τη θέση στην οποία βρίσκεται. Στη συνέχεια τη διαγράφει από τον πίνακα καθώς και τις τιμές στην αντίστοιχη γραμμή του πίνακα ΘM. Η διαδικασία επαναλαμβάνεται όσο οι επιστρεφόμενες τιμές μεγίστων είναι περισσότερες από πέντε.



Σχήμα 5.6.3.2 Αλγόριθμος εντοπισμού θέσης άκρων και ταυτοποίησης δακτύλων

Μετά τη διαδικασία διαγραφής των ελαχίστων αποστάσεων, και των αντίστοιχων θέσεων των εικονοστοιχείων που τις παρήγαγαν, από τους πίνακες MA και ΘM , το σύστημα έχει αναγνωρίσει τις θέσεις των πέντε δακτύλων. Οι θέσεις αυτές βρίσκονται καταχωρημένες στον πίνακα ΘM και απομένει μόνο η εμφάνισή τους επάνω στην εικόνα α_λ . Το σύστημα απεικονίζει γραφικά τις θέσεις των δακτύλων μέσα από τη γραφική παράσταση των αποστάσεων σε συνάρτηση με τα εικονοστοιχεία (Σχήμα 5.6.3.3). Αυτή η γραφική παράσταση υλοποιείται από το σύστημα για το κάθε καρέ.



Σχήμα 5.6.3.3 Προβολή τοπικών μεγίστων σε διάγραμμα εικονοστοιχείου-απόστασης

5.7 Ανάκτηση δακτυλισμών

Η ανάκτηση δακτυλισμών από το σύστημα στατικής αναγνώρισης δακτυλικών κινήσεων βασίζεται στη σύγκριση θέσεων των άκρων των δακτύλων. Η αναγνώριση διενεργείται για κάθε ένα δάκτυλο χωριστά.

Εφόσον έχει δημιουργηθεί ο πίνακας που έχει αποθηκευμένες τις θέσεις των άκρων των δακτύλων πάνω στην εικόνα μπορεί να ακολουθήσει η ανάκτηση των δακτυλισμών. Ως σημείο αναφοράς καθορίζεται ένα καρέ του βίντεο όπου τα δάκτυλα βρίσκονται σε ηρεμία. Στην αρχή τέθηκε ένα κατώφλι για κάθε δάκτυλο χωριστά, το οποίο καθόριζε την αναγνώριση ή όχι του δακτυλισμού. Αν δηλαδή η τεταγμένη ενός οποιουδήποτε δακτύλου ξεπερνούσε αυτό το κατώφλι, τότε θεωρούνταν ότι υπήρχε δακτυλισμός ή ότι πατήθηκε κάποιο πλήκτρο. Η τεχνική αυτή, αν και λειτουργούσε σωστά στην πλειοψηφία των καρέ, εντούτοις έδινε κάποια σφάλματα σε συγκεκριμένες περιπτώσεις. Μια τέτοια περίπτωση ήταν η μετακίνηση των δακτύλων πάνω στα άσπρα πλήκτρα. Αυτή η μετακίνηση προφανώς

διαφοροποιούσε τις τεταγμένες των δακτύλων με αποτέλεσμα συχνά ανακτώνταν εσφαλμένος δακτυλισμός. Έτσι, επιλέχτηκε το κατώφλι να συγκρίνεται όχι απευθείας με την τεταγμένη του δακτύλου αλλά με τη διαφορά τεταγμένης δακτύλου και τεταγμένης κεντροειδούς. Με αυτόν τον τρόπο δόθηκε λύση στο ζήτημα ενός σημαντικού ποσοστού ψεύτικων δακτυλισμών.

Εμπράκτως, ο τρόπος λειτουργίας του συστήματος θα περιγραφεί στο έβδομο κεφάλαιο όπου και θα αξιολογηθεί επίσης και η αποδοτικότητα του συστήματος μέσω του λογισμικού PianOrasis.

5.8 Συμπεράσματα

Στο πέμπτο κεφάλαιο περιγράφηκαν αναλυτικά όλα τα δομικά στοιχεία και η μεθοδολογική προσέγγιση του συστήματος στατικής αναγνώρισης δακτυλικών κινήσεων.

Το παρόν κεφάλαιο ξεκίνησε με την παρουσίαση του εξοπλισμού, υλικού και λογισμικού, που χρησιμοποιήθηκε για την ανάπτυξη του συστήματος. Έγινε προσπάθεια και επιτεύχθηκε ο εξοπλισμός να είναι όσο το δυνατόν χαμηλού κόστους και κατά συνέπεια άμεσα προσβάσιμος σε έναν μέσο χρήστη. Δεν χρησιμοποιήθηκε εξειδικευμένο υλικό για την επεξεργασία του σήματος του βίντεο (πχ *Graphical Processing Unit*). Μια οικονομική κάμερα, USB ή όχι, και ένας ηλεκτρονικός υπολογιστής της τελευταίας τετραετίας μπορεί να καλύψει τις απαιτήσεις του συστήματος.

Στη συνέχεια αναλύθηκε όλες οι επιμέρους μέθοδοι και τεχνικές όρασης H/Y που χρησιμοποιήθηκαν για την ανάπτυξη του συστήματος. Βασικοί άξονες λειτουργίας του συστήματος είναι: α) η ανίχνευση δέρματος, β) η τμηματοποίηση του χεριού, γ) ο εντοπισμός των δακτύλων, δ) η εξαγωγή των διανυσμάτων παρατήρησης και ε) η ανάκτηση δακτυλισμών. Αρχικά για τον πρώτο άξονα αναπτύχθηκε στατιστικό μοντέλο ειδικά για το παρόν σύστημα, το οποίο στοχεύει στον εντοπισμό δερματικών περιοχών σε κάθε καρέ του βίντεο εισόδου. Ως προς το δεύτερο άξονα, το σύστημα δέχεται ως είσοδο τη δυαδική εικόνα που προέκυψε από πριν. Η τμηματοποίηση επιτυγχάνεται μέσω απλοποίηση της δυαδικής εικόνας, διαχωρισμό της σε βασικές συνιστώσες. Ως προς τον τρίτο άξονα, ο εντοπισμός θέσης των δακτύλων, γίνεται μέσω υπολογισμού ευκλείδειων αποστάσεων μεταξύ ενός σημείου αναφοράς και των θέσεων των εικονοστοιχείων που ανήκουν στο περίγραμμα των δακτύλων. Η ταυτοποίησή τους γίνεται με τη χρήση κατάλληλων ταξινομητών. Ως προς τον τέταρτο άξονα, τα διανύσματα παρατήρησης που καταγράφονται για τη στα-

τική αναγνώριση είναι οι διαφορές της τεταγμένης του κάθε δακτύλου με την τεταγμένη του κεντροειδούς, ενώ τέλος ως προς τον πέμφο άξονα, η ανάκτηση δακτυλισμών επιτυγχάνεται με καθορισμό ενός κατωφλιού για το πάτημα, ή όχι, κάποιου πλήκτρου.

Η στατική αναγνώριση κινήσεων, ή αλλιώς ανάκτηση δακτυλισμών με όρους της ΜΑ, διενεργείται σε κάθε καρτέ και για κάθε δάκτυλο χωριστά. Δεν υπάρχει οποιαδήποτε διασύνδεση των καρτέ μεταξύ τους. Η περίπτωση της αναγνώρισης κινήσεων μέσα από ακολουθίες καρτέ θα παρουσιαστεί αναλυτικά στο έκτο κεφάλαιο και αφορά στη δυναμική αναγνώριση κινήσεων.

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Η ανακάλυψη (3/4)

« Εμένα μου φαίνεται ωραίο όλο αυτό. Εξάλλου, όταν το αφεντικό μου βρίσκει ένα θεώρημα, είναι πολύ περήφανος και οι φίλοι του (που είναι επίσης κύριοι πολύ πολύ σπουδαίοι) είναι πολύ χαρούμενοι για αυτόν. Αλλά αυτό οι γονείς μου δεν το ξέρουν..»



6 Σύστημα δυναμικής αναγνώρισης δακτυλικών κινήσεων με Κρυφά Μαρκοβιανά Μοντέλα

- 6.1 Σύνοψη κεφαλαίου**
- 6.2 Επιλογή κινήσεων προς αναγνώριση**
- 6.3 Χαρακτηριστικές ιδιότητες και διανύσματα παρατήρησης**
- 6.4 Ρόλος των Κρυφών Μαρκοβιανών Μοντέλων στο σύστημα αναγνώρισης δακτυλικών κινήσεων**
 - 6.4.1 Κρυφά Μαρκοβιανά Μοντέλα σε διακριτό ή συνεχή χώρο παρατηρήσεων
 - 6.4.2 Επιλογή συνεχών Κρυφών Μαρκοβιανών Μοντέλων
- 6.5 Ανάλυση και στοχαστική μοντελοποίηση κινήσεων με Κρυφά Μαρκοβιανά Μοντέλα**
- 6.6 Επίλυση των προβλημάτων των Κρυφά Μαρκοβιανών Μοντέλων**
 - 6.6.1 Μοντελοποίηση διανυσμάτων παρατήρησης με Μεικτά Γκαουσιανά Μοντέλα
 - 6.6.2 Εύρεση Μεικτού Γκαουσιανού Μοντέλου με τη μέγιστη πιθανοφάνεια
 - 6.6.3 Παράμετροι Μεικτού Γκαουσιανού Μοντέλου με τη μέγιστη πιθανοφάνεια
 - 6.6.4 Αλγόριθμος « forward » του Baum
 - 6.6.5 Αλγόριθμος του Viterbi
 - 6.6.6 Επανεκτίμηση παραμέτρων Κρυφών Μαρκοβιανών Μοντέλων και Μεικτών Γκαουσιανών Μοντέλων κατά Baum-Welch
- 6.7 Παράδειγμα αναγνώρισης μεμονωμένων κινήσεων**
- 6.8 Υλοποίηση δυναμικής αναγνώρισης δακτυλικών κινήσεων με το MATLAB**
- 6.9 Συμπεράσματα**

6.1 Σύνοψη κεφαλαίου

Στο κεφάλαιο αυτό αναλύεται ο τρόπος που μοντελοποιείται η κίνηση των δακτύλων του πιανίστα, που εκπαιδεύονται τα μοντέλα και ταξινομούνται οι κινήσεις από το σύστημα. Η δυναμική αναγνώριση κινήσεων με χρήση των KMM λειτουργεί σε συνέχεια της στατικής αναγνώρισης κινήσεων, από το στάδιο του εντοπισμού των δακτύλων και μετά.

Για κάθε μουσική κίνηση (*χειρονομία*) υπάρχει και η αποτύπωσή της στο ψηφιακό σήμα μέσω του οποίου καταγράφεται. Με άλλα λόγια, υπάρχει ένα σύνολο χαρακτηριστικών της κίνησης που μετατρέπονται σε χαρακτηριστικά του ψηφιακού σήματος. Η κατάλληλη μοντελοποίηση και ταξινόμηση των κινήσεων αυτών μπορεί να οδηγήσει, με τη χρήση των KMM, στην απαλλαγή από τη χρονική εξάρτησή των κινήσεων, εκμεταλλευόμενοι το δυναμικό τους χαρακτηριστήρα.

Παράλληλες έρευνες, στην αναγνώριση φωνής ή μουσικών οργάνων μέσα από ηχητικό σήμα, κάνουν επίσης χρήση των KMM λόγω του στοχαστικού χαρακτήρα φωνής, ήχου και ανθρώπινης χειρονομίας. Η στοχαστικότητα έγκειται στο γεγονός ότι δεν μπορεί να βρεθεί μια συνάρτηση η οποία να περιγράφει τη φωνή, τον ήχο μουσικών οργάνων ή τις χειρονομίες.

Τα KMM παρεμβάλλονται στη γενικότερη διαδικασία που ακολουθείται από το σύστημα αναγνώρισης δακτυλικών κινήσεων, σε δύο βασικά σημεία:

1. Παραγωγή διανυσμάτων παρατήρησης με κατάλληλες χαρακτηριστικές ιδιότητες, λαμβάνοντας υπόψη συγκεκριμένα χαρακτηριστικά της δακτυλικής κίνησης που ανακτώνται από τα βίντεο εκπαίδευσης των KMM. Τα διανύσματα θα χρησιμοποιηθούν ως είσοδος των KMM, μοντελοποιώντας έτσι την κάθε κίνηση,
2. Ταξινόμηση των δακτυλικών κινήσεων μέσω υπολογισμού της πιθανότητας το διάνυσμα εισόδου να έχει παραχθεί από καθένα από τα KMM. Στη συνέχεια επιλέγεται το KMM το οποίο παρουσίασε τη μεγαλύτερη πιθανότητα να έχει δημιουργήσει το διάνυσμα εισόδου.

Η διαδικασία ταξινόμησης των δακτυλικών κινήσεων πραγματοποιείται με τη χρήση συνεχών KMM. Τα συνεχή KMM είναι συνεχείς κατανομές εξόδου. Η έξοδος για κάθε εσωτερική κατάσταση εξαρτάται από τη ΣΠΠ που ορίστηκε για την κάθε κατάσταση. Η τεχνική των ΜΓΜ μοντελοποιεί τα διανύσματα παρατήρησης πάνω σε συνεχή χώρο, με ένα μείγμα από Γκαουσιανές κατανομές. Η υλοποίηση έγινε με χρήση της εργαλειοθήκης « Kevin Murphy » του MATLAB.

6.2 Επιλογή κινήσεων προς αναγνώριση

Στα έγχορδα όργανα, υφίσταται άμεση επαφή των δακτύλων του μουσικού πάνω στις χορδές. Στα πνευστά όργανα, τα χείλη του μουσικού ελέγχουν και διαμορφώνουν τον επιθυμητό ήχο. Οι ερμηνευτικές κινήσεις του πιανίστα και η παραγωγή του ήχου όμως, επιτυγχάνεται από έναν ηλεκτροφόρο διαμεσολαβητικό μηχανισμό. Αυτή η αποξενωτική παρεμβολή του μηχανισμού καθιστά άκρως περίπλοκο τον έλεγχο στην παραγωγή του επιθυμητού ήχου. Απαιτείται μια προσεκτική και μακρά θητεία προκειμένου να αποκτηθεί η κατάλληλη τεχνική κινήσεων που θα αποφέρει ήχο υψηλής ακρίβειας.

Ως εκ τούτου, η συστηματική εξάσκηση των δακτύλων που βασίζεται σε μια σειρά τεχνικών κινήσεων, μπορεί να οδηγήσει σε άριστες ερμηνευτικές κινήσεις και μουσική ωρίμανση. Η δυναμική αναγνώριση του συστήματος δε θα μπορούσε παρά να εκπαιδευτεί σε κατάλληλες τεχνικές κινήσεις, οι οποίες είναι βασικές για τη σταδιακή δόμηση των ερμηνευτικών κινήσεων του πιανίστα. Οι τεχνικές κινήσεις που αναγνωρίζονται από το σύστημα έχουν ως σημείο αναφοράς τις μουσικές κλίμακες, οι οποίες ορίζονται ως εξής [Πολίτης 2007]:

«Μουσική κλίμακα (ή αλλιώς σκάλα) καλείται μια σειρά από οκτώ συνεχείς φθόγγους (νότες), που είτε ανέρχονται, σχηματίζοντας μια ανιούσα κλίμακα, είτε κατέρχονται, σχηματίζοντας μια κατιούσα κλίμακα. Κάθε φθόγγος της κλίμακας λέγεται βαθμίδα της κλίμακας. Η πρώτη βαθμίδα είναι η βάση της κλίμακας και από αυτήν τη νότα παίρνει την ονομασία της η κλίμακα.»

Οι κινήσεις που επιλέχτηκαν για να αναγνωρίζει το σύστημα ορίζουν ένα σύνολο κινήσεων, το οποίο στο εξής θα ονομάζεται λεξι-κό κινήσεων:

$$\Lambda K = \left\{ \kappa_{iV_i} \right\}_{i \in [1,6]} \quad (6.2.1)$$

Όπου:

- ΛΚ, το λεξικό κινήσεων,
- $κιν_i$, η εκάστοτε κίνηση

Οι κινήσεις που αναγνωρίζονται από το σύστημα είναι οι παρακάτω:

1. ανιούσα κίνηση κλίμακας,
2. κατιούσα κίνηση κλίμακας,
3. ανιούσα κίνηση αρπέζ,
4. κατιούσα κίνηση αρπέζ,
5. ανιούσα κίνησης κλίμακας σε τρίτες,
6. κατιούσα κίνησης κλίμακας σε τρίτες.

Η προκαταρκτική ανάλυση όλων των παραπάνω κινήσεων αποτελεί προϋπόθεση για την επιλογή χαρακτηριστικών διανυσμάτων. Η ανάλυση των κινήσεων θα βασιστεί σε δυο σημαντικές μουσικές έννοιες που είναι α) η αρμονία και β) η μελωδία.

Η αρμονία ορίζεται ως εξής [Πολίτης 2007]:

«Αρμονία είναι το ταυτόχρονο άκουσμα, η συνήχηση δύο ή περισσότερων φθόγγων (νοτών) διαφορετικού ύψους, τοποθετημένων όμως επ'αλλήλων, σύμφωνα με τους καθιερωμένους μουσικούς κανόνες. Η αρμονία είναι η βάση της πολυφωνίας, που χαρακτηρίζει ολόκληρο το οικοδόμημα της Δυτικής μουσικής τους τελευταίους αιώνες, σε αντιδιαστολή με την ομοφωνία της Ανατολικής ή Βυζαντινής μουσικής.»

Η μελωδία ορίζεται ως εξής [Πολίτης 2007]:

«Μελωδία είναι η διαφορά οξύτητας των διαφόρων φθόγγων (νοτών) που συνιστούν ένα μουσικό κομμάτι. Η σειρά των φθόγγων δεν είναι τυχαία. Τουναντίον, καθορίζεται με βάση ορισμένους κανόνες ακουστικής αισθητικής και μουσικής εκφράσεως, ώστε οι μουσικές εκφράσεις να δημιουργούν ένα ευάρεστο ακροαματικό αίσθημα. Η αισθητική κυρίως άποψη της μελωδίας είναι αυτή που χαρακτηρίζει τις μουσικές εποχές και διαφοροποιεί το μπαρόκ του 17^{ου} αιώνα από το ρομαντισμό του 19^{ου}.»

Ας θεωρηθούν οι παραπάνω κινήσεις στην κλίμακα ΝΤΟ μείζονα για το δεξί χέρι. Η συγκεκριμένη κλίμακα για τις εν λόγω κι-

νήσεις δεν κάνει καθόλου χρήση των μαύρων πλήκτρων. Συνεπώς, το γεγονός αυτό διευκολύνει το συγκεκριμένο παράδειγμα. Στο σχήμα 6.2.1 απεικονίζονται οι νότες και δακτυλισμοί που απαρτίζουν την ανιούσα κίνηση της κλίμακας ΝΤΟ, για το δεξί χέρι και για μία οκτάβα. Οφείλεται να σημειωθεί ότι για μουσικές κινήσεις που αφορούν σε κλίμακες ή αρπές έχουν οριστεί ζεύγη δακτύλου-νότας από τη μουσική βιβλιογραφία τα οποία είναι δεσμευτικά για τον ερμηνευτή [Hanon 1923]. Τα ζευγάρια αυτά είναι δεσμευτικά κατ' επέκταση και για τη σύστημα αναγνώρισης δακτυλικών κινήσεων.

Η προσεκτική μελέτη του σχήματος 6.2.1 μας οδηγεί σε συγκεκριμένα συμπεράσματα για την ανιούσα κλίμακα. Τα συμπεράσματα αυτά είναι τα εξής:

1. Η αρμονική ανάλυση υποδεικνύει α) το *πάτημα μιας μόνο νότας* και συνεπώς, β) τη *χρήση ενός μόνο δακτύλου σε κάθε χρονική στιγμή*.



Σχήμα 6.2.1 Ανιούσα κίνηση της κλίμακας ΝΤΟ μείζονα με δακτυλισμούς δεξιού χεριού.

2. Η μελωδική ανάλυση υποδεικνύει ότι οι νότες που χρησιμοποιούνται στην κίνηση είναι *συνεχόμενες και αντιστοιχούν σε διαδοχικά πλήκτρα*. Αν θεωρηθεί ότι τα δάκτυλα πατούν το κάθε πλήκτρο σχεδόν στο μέσον του πλάτους του, τότε η *απόσταση μεταξύ των δακτύλων που πατούν τα πλήκτρα της κίνησης είναι σταθερή*. Στο σχήμα 6.2.2 απεικονίζονται οι νότες της κίνησης, τα πλήκτρα και η σειρά (μπλε αριθμοί) με την οποία θα πατηθούν. Επιβεβαιώνεται η διαδοχικότητα των πλήκτρων καθώς και το γεγονός ότι η κίνηση ξεκινάει από το χαμηλό ΝΤΟ και καταλήγει στο ψηλό ΝΤΟ της επόμενης οκτάβας.
3. Τέλος, τα ζεύγη δακτύλου-νότας που αναγράφονται στο σχήμα 6.2.1 είναι συγκεκριμένα και δεσμευτικά για τον πιανίστα.

Και τα τρία παραπάνω χαρακτηριστικά της εν λόγω κίνησης είναι ιδιαίτερα και τη διακρίνουν από τις υπόλοιπες προς αναγνώριση κινήσεις.



Σχήμα 6.2.2 Νότες της ανιούσας κίνησης, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας.

Η αντίθετη κίνησης της ανιούσας κλίμακας είναι η κατιούσα κλίμακας (Σχήμα 6.2.3). Τα χαρακτηριστικά της κίνησης αυτής είναι παρόμοια με τα χαρακτηριστικά της ανιούσας κλίμακας:

1. Η αρμονική ανάλυση υποδεικνύει α) το πάτημα μιας μόνο νότας σε κάθε χρόνο του μέτρου και συνεπώς β) τη χρήση ενός μόνο δακτύλου σε κάθε χρονική στιγμή.



Σχήμα 6.2.3 Κατιούσα κίνηση της κλίμακας ΝΤΟ μείζονα με δακτυλισμούς δεξιού χεριού.

2. Η μελωδική ανάλυση υποδεικνύει ότι οι νότες που χρησιμοποιούνται στην κίνηση είναι *συνεχόμενες και αντιστοιχούν σε διαδοχι-*

κά πλήκτρα. Αν θεωρηθεί ότι τα δάκτυλα πατούν το κάθε πλήκτρο σχεδόν στο μέσον του πλάτους του, τότε η απόσταση μεταξύ των δακτύλων που πατούν τα πλήκτρα της κίνησης είναι σταθερή. Στο σχήμα 6.2.4 απεικονίζονται οι νότες της κίνησης, τα πλήκτρα και η σειρά με την οποία θα πατηθούν. Επιβεβαιώνεται η διαδοχικότητα των πλήκτρων αλλά και η αντίθεση της κίνησης σε σχέση με την ανιούσα κίνησης, λόγω του ότι η κίνηση στην κατιούσα ξεκινάει από το ψηλό ΝΤΟ και καταλήγει στη χαμηλό ΝΤΟ της προηγούμενης οκτάβας, με έξι ενδιάμεσα πατήματα.

3. Τέλος, τα ζεύγη δακτύλου-νότας, που αναγράφονται στο σχήμα 6.2.3, είναι συγκεκριμένα και δεσμευτικά για τον πιανίστα.

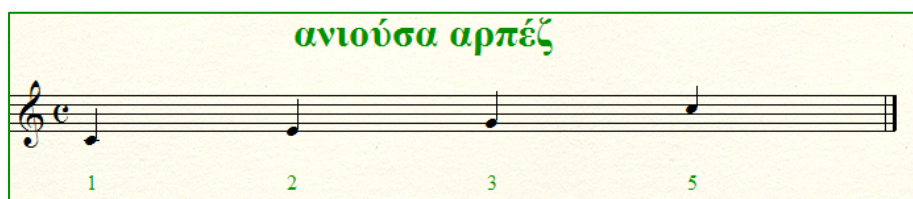


Σχήμα 6.2.4 Νότες της κατιούσας κίνησης, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας.

Η κατάλληλη επεξεργασία των παραπάνω χαρακτηριστικών, παράλληλα με τη στοχευμένη επιλογή χαρακτηριστικών διανυσμάτων που θα διακρίνουν τη μία κίνηση από την άλλη, θα διευκολύνουν το έργο της αναγνώρισης.

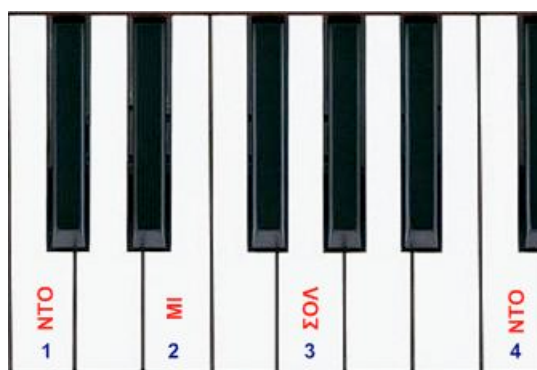
Οι κινήσεις αρπές (ή αλλιώς αρπισμοί) στη μουσική αποτελούν αναπόσπαστο παρτενέρ της κάθε κλίμακας (Σχήμα 6.2.5). Με άλλα λόγια, για κάθε μια κλίμακα υπάρχει αντίστοιχα και ένας αρπισμός που διατηρεί τον οπλισμό (υφέσεις ή διέσεις) της κλίμακας. Τέτοιου είδους κινήσεις είναι πολύ βασικές στην εκμάθηση του πιάνου, στην ανάλυση μουσικών κειμένων κ.α. Η ανιούσα κίνηση αρπές έχει ιδιαίτερα χαρακτηριστικά, κάποια από τα οποία τη διαφοροποιούν σε σχέση με όλες τις υπόλοιπες κινήσεις. Τα χαρακτηριστικά αυτά προσδιορίζονται παρακάτω:

1. Η αρμονική ανάλυση υποδεικνύει α) το πάτημα μιας μόνο νότας και συνεπώς, β) τη χρήση ενός μόνο δακτύλου σε κάθε χρονική στιγμή.



Σχήμα 6.2.5 Ανιούσα κίνηση αρπάζ ΝΤΟ μείζονα με δακτυλισμούς δεξιού χεριού

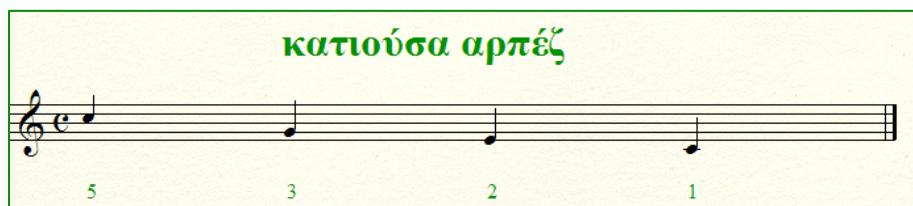
2. Η μελωδική ανάλυση υποδεικνύει ότι οι νότες που χρησιμοποιούνται στην κίνηση δεν είναι συνεχόμενες και συνεπώς δεν αντιστοιχούν σε διαδοχικά πλήκτρα. Ως εκ τούτου, οι αποστάσεις μεταξύ θέσεων πλήκτρων αλλά και δακτύλων που τα πατούν δεν είναι σταθερές, πλην όμως είναι συγκεκριμένες. Στο σχήμα 6.2.5 απεικονίζονται οι νότες της κίνησης, τα πλήκτρα και η σειρά (μπλε αριθμοί) με την οποία θα πατηθούν. Η κίνηση ολοκληρώνεται σε τέσσερα πατήματα πλήκτρων ξεκινώντας από το χαμηλό ΝΤΟ και καταλήγοντας στο ψηλό ΝΤΟ της επόμενης οκτάβας, με δύο ενδιάμεσα πατήματα.
3. Τέλος, τα ζεύγη δακτύλου-νότας που αναγράφονται στο σχήμα 6.2.5 είναι συγκεκριμένα και δεσμευτικά για τον πιανίστα.



Σχήμα 6.2.6 Νότες της ανιούσας κίνησης αρπάζ, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας.

Αντίστοιχα, οι νότες και οι δακτυλισμοί της κατιούσας κίνησης αρπές αναγράφονται στο σχήμα 6.2.7. Τα χαρακτηριστικά της κίνησης αυτής προσδιορίζονται παρακάτω:

1. Η αρμονική ανάλυση υποδεικνύει α) το πάτημα μιας μόνο νότας και συνεπώς, β) τη χρήση ενός μόνο δακτύλου σε κάθε χρονική στιγμή.



Σχήμα 6.2.7 Κατιούσα κίνηση αρπές ΝΤΟ μείζονα με δακτυλισμούς δεξιού χεριού

2. Η μελωδική ανάλυση υποδεικνύει ότι οι νότες που χρησιμοποιούνται στην κίνηση *δεν είναι συνεχόμενες* και συνεπώς *δεν αντιστοιχούν σε διαδοχικά πλήκτρα*. Ως εκ τούτου, οι αποστάσεις μεταξύ θέσεων πλήκτρων αλλά και δακτύλων που τα πατούν δεν είναι σταθερές, πλην όμως είναι συγκεκριμένες. Στο σχήμα 6.2.7 απεικονίζονται οι νότες της κίνησης, τα πλήκτρα και η σειρά (μπλε αριθμοί) με την οποία θα πατηθούν. Η κίνηση ολοκληρώνεται σε τέσσερα πατήματα πλήκτρων ξεκινώντας από το ψηλό ΝΤΟ και καταλήγοντας στο χαμηλό ΝΤΟ της προηγούμενης οκτάβας, με δύο ενδιάμεσα πατήματα.
3. Τέλος, τα ζεύγη δακτύλου-νότας που αναγράφονται στο σχήμα 6.2.7 είναι συγκεκριμένα και δεσμευτικά για τον πιανίστα.

Η «κλίμακα σε τρίτες» αποτελεί αναπόσπαστο μέρος της δακτυλικής εξάσκησης του πιανίστα και προϋπόθεση για την ερμηνεία δύσκολων τεχνικά μουσικών έργων (Σχήμα 6.2.9 και Σχήμα 6.2.10). Συχνά, σε μουσικά έργα μεγάλων συνθετών, όπως για παράδειγμα του Μπετόβεν, συναντούμε τέτοιες κινήσεις. Η δυσκολία της ανιούσας κίνησης κλίμακας σε τρίτες έγκειται στη χρήση ζεύγους δακτυλι-

σμών που απαιτούν ομοιόμορφο και ενωμένο πάτημα πλήκτρων [Hanon 1923]. Τα χαρακτηριστικά αυτά προσδιορίζονται παρακάτω:



Σχήμα 6.2.8 Νότες της κατιούσας κίνησης αρπάζ, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας.

1. Η αρμονική ανάλυση υποδεικνύει α) τη *συνήχηση νοτών* και συνεπώς β) τη χρήση *δύο δακτύλων σε κάθε χρονική στιγμή* (Σχήμα 6.2.9).



Σχήμα 6.2.9 Ανιούσα κίνηση σε τρίτες της κλίμακας ΝΤΟ μείζονα με δακτυλισμούς δεξιού χεριού

2. Η μελωδική ανάλυση υποδεικνύει ότι οι *συνηχήσεις* που *απαρτίζουν* την κίνηση *αποτελούνται* από *ζεύγη διαδοχικών νοτών*.

Δηλαδή, οι χαμηλές νότες των συνηχήσεων, αντίστοιχα και οι ψηλές, είναι διαδοχικές μεταξύ τους. Κατά συνέπεια αντιστοιχούν σε διαδοχικά ζεύγη πλήκτρων. Ως εκ τούτου, οι αποστάσεις μεταξύ θέσεων πλήκτρων, αλλά και δακτύλων που τα πατούν, είναι σταθερές και συγκεκριμένες. Στο σχήμα 6.2.10 απεικονίζονται οι νότες της κίνησης, τα πλήκτρα και η σειρά (μπλε αριθμοί) με την οποία θα πατηθούν. Εφόσον πρόκειται για οκτώ συνηχήσεις ο κάθε αριθμός προτεραιότητας στο σχήμα 6.2.10 εμφανίζεται εις διπλούν, μια για κάθε νότα της συνήχησης. Η κίνηση ολοκληρώνεται σε οκτώ διπλά πατήματα πλήκτρων ξεκινώντας από το χαμηλό ΝΤΟ, σε συνήχηση με ΜΙ, και καταλήγοντας στο ψηλό ΝΤΟ της επόμενης οκτάβας, σε συνήχηση με ΜΙ, με έξι ενδιάμεσα διπλά πατήματα.

3. Τέλος, τα ζεύγη δακτύλου-νότας που αναγράφονται στο σχήμα 7-6 είναι συγκεκριμένα και δεσμευτικά για τον πιανίστα.



Σχήμα 6.2.10 Νότες της ανιούσας κίνησης αρπάζ, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας.

Η αντίστροφη κίνηση της ανιούσα κίνησης κλίμακας σε τρίτες είναι η κατιούσα κίνηση (Σχήμα 6.2.11). Τα χαρακτηριστικά της προσδιορίζονται παρακάτω:

1. Η αρμονική ανάλυση υποδεικνύει α) τη συνήχηση νοτών και συνεπώς β) τη χρήση δύο δακτύλων σε κάθε χρονική στιγμή.



Σχήμα 6.2.11 Ανιούσα κίνηση σε τρίτες της κλίμακας ΝΤΟ μείζονα με δακτυλισμούς δεξιού χεριού

2. Η μελωδική ανάλυση υποδεικνύει ότι οι συνηχήσεις που απαρτίζουν την κίνηση αποτελούνται από ζεύγη διαδοχικών νοτών. Δηλαδή, οι χαμηλές νότες των συνηχήσεων, αντίστοιχα και οι ψηλές, είναι διαδοχικές μεταξύ τους. Κατά συνέπεια αντιστοιχούν σε διαδοχικά ζεύγη πλήκτρων. Ως εκ τούτου, οι αποστάσεις μεταξύ θέσεων πλήκτρων, αλλά και δακτύλων που τα πατούν, είναι σταθερές και συγκεκριμένες. Στο σχήμα 6.2.12 απεικονίζονται οι νότες της κίνησης, τα πλήκτρα και η σειρά (μπλε αριθμοί) με την οποία θα πατηθούν. Εφόσον πρόκειται για οκτώ συνηχήσεις ο κάθε αριθμός προτεραιότητας στο σχήμα 6.2.12 εμφανίζεται εις διπλούν, μια για κάθε νότα της συνήχησης. Η κίνηση ολοκληρώνεται σε οκτώ διπλά πατήματα πλήκτρων ξεκινώντας από το ψηλό ΝΤΟ, σε συνήχηση με ΜΙ, και καταλήγοντας στο χαμηλό ΝΤΟ της προηγούμενης οκτάβας, σε συνήχηση με ΜΙ, με έξι ενδιαμέσα διπλά πατήματα.
3. Τέλος, τα ζεύγη δακτύλου-νότας που αναγράφονται στο σχήμα 6.2.11 είναι συγκεκριμένα και δεσμευτικά για τον πιανίστα.

Εποπεύοντας και τις τέσσερις παραπάνω κινήσεις, στο συγκεκριμένο παράδειγμα, παρατηρείται ότι κάποιες από αυτές παρουσιάζουν ιδιότητες που μπορεί να είναι παραπλήσιες με τις ιδιότητες κάποιων άλλων κινήσεων. Το πλήθος όμως των παραπλήσιων ιδιοτήτων είναι περιορισμένο συγκριτικά με το σύνολο, γεγονός το οποίο καθιστά σαφή τη διάκριση μεταξύ τους. Σε περιπτώσεις που οι παραπλήσιες ιδιότητες μεταξύ των κινήσεων δεν είναι περιορισμένες

στο πλήθος, πρέπει να γίνεται πολύ προσεκτική ανάλυση των κινήσεων ώστε να προσδιοριστούν ιδιότητες που τις ξεχωρίζουν, διαφορετικά υπάρχει αυξημένη πιθανότητα μειωμένης ακρίβειας στην αναγνώριση. Η δυνατότητα διαχωρισμού των παραπάνω κινήσεων ανάλογα με την κλίμακα είναι μια σχετικά απλή διαδικασία, η οποία βασίζεται στην αναγνώριση της αρχικής νότας από την οποία ξεκινάει η κίνηση.



Σχήμα 6.2.12 Νότες της κατιούσας κίνησης αρπάζ, θέσεις των δακτύλων στο κλαβιέ και πλήκτρα που θα χρησιμοποιηθούν με σειρά προτεραιότητας.

Τέλος, οι κινήσεις που αφορούν σε κλίμακες έχουν το χαρακτηριστικό ότι μπορούν να εφαρμοστούν με αρκετά τυποποιημένη μορφή αλλά και παρόμοιο τρόπο ακόμα και από διαφορετικούς χρήστες. Έτσι, η εκπαίδευση των ΚΜΜ, ώστε να αναγνωρίζουν τέτοιους είδους κινήσεις, δεν απαιτεί μεγάλο πλήθος δεδομένων εκπαίδευσης.

6.3 Χαρακτηριστικές ιδιότητες και διανύσματα παρατήρησης

Οι ιδιότητες που κάνουν ευδιάκριτες τις κινήσεις μεταξύ τους, αναφέρονται σε χαρακτηριστικά τροχιάς. Οι τροχιές αυτές αναφέρονται στα δάκτυλα. Τα χαρακτηριστικά διανύσματα που θα προκύψουν σχετίζονται με τις συντεταγμένες του κάθε δακτύλου, χωριστά για κάθε καρέ. Στόχος είναι να καθορίζεται πλήρως η θέση του κάθε δακτύλου, σε κάθε καρέ.

Τα KMM εκπαιδεύονται από ένα σύνολο βίντεο, το καθένα από τα οποία αντιστοιχεί σε κάποια κίνηση. Από τα βίντεο εκπαίδευσης αυτά εξάγονται τα χαρακτηριστικά διανύσματα, τα οποία με τη σειρά τους αντιστοιχούν σε κάποια κίνηση. Τα χαρακτηριστικά διανύσματα χρησιμοποιούνται στη συνέχεια, από τα KMM, για να αναγνωρίσουν ακολουθίες καρτέ, σε οποιοδήποτε βίντεο, οι οποίες «ταιριάζουν» με κίνηση για την οποία έχουν εκπαιδευτεί.

Αξίζει να σημειωθεί ότι οι συντεταγμένες των δακτύλων εξαρτώνται προφανώς από τη θέση του χεριού στην εικόνα. Ειδικότερα, οι θέσεις των δακτύλων εξαρτώνται από τη θέση του χεριού στο κάθε καρτέ. Κατά συνέπεια, η τοποθέτηση των δακτύλων πάνω στα πλήκτρα, πιο συγκεκριμένα το πόσο «μέσα ή έξω» στο πλήκτρο είναι το δάκτυλο, επηρεάζει τα αποτελέσματα της αναγνώρισης. Η θέση του δακτύλου, για το σύστημα αλλά και για την εκπαίδευση των KMM, ανάγεται σε θέση ενός σημείου, όπως περιγράφηκε στην ενότητα 6.2. Το πάτημα ενός πλήκτρου κατ' επέκταση εξαρτάται από την τεταγμένη του σημείου αυτού. Κάτι αντίστοιχο δεν ισχύει προφανώς και για το πλήκτρο, το οποίο δεν είναι σημείο, αλλά μια επιφάνεια.

Συνεπώς, η χρήση μόνο συντεταγμένων θέσης δε θα παρείχε αποτελέσματα ακριβείας στα KMM, διότι η ελάχιστη μετατόπιση της θέσης κάποιου δακτύλου κατά την ερμηνεία, θα συνιστούσε απόκλιση από την αντίστοιχη θέση, στα βίντεο εκπαίδευσης των μοντέλων. Έτσι, αποφασίστηκε η χρήση αποστάσεων για τον κατακόρυφο άξονα. Το σημείο αναφοράς από το οποίο υπολογίζονται οι αποστάσεις του κάθε δακτύλου είναι το κεντροειδές. Εξασφαλίζεται συνεπώς, η ανεξαρτησία των χαρακτηριστικών της κίνησης που αφορούν στον κατακόρυφο άξονα, μέσω χρήσης σχετικών θέσεων των δακτύλων πάνω στα πλήκτρα.

Στον οριζόντιο άξονα διατηρούνται οι τεταγμένες του κάθε δακτύλου. Η επιλογή αυτή στηρίχτηκε στο γεγονός της σταθερής θέσης της κάμερας, της ομοιομορφίας των πλήκτρων αλλά και της ενδεχόμενης ανάγκης της προς αναγνώριση κίνησης σε συγκεκριμένη κλίμακα.

Η καταγραφή των εν λόγω χαρακτηριστικών πραγματοποιείται για κάθε καρτέ, καθενός από τα βίντεο εκπαίδευσης των KMM, παράγοντας κάθε φορά ένα χαρακτηριστικό διάνυσμα. Για κάθε βίντεο εκπαίδευσης, το οποίο αντιστοιχεί σε μια συγκεκριμένη κίνηση, παράγεται μια ακολουθία διανυσμάτων παρατήρησης της μορφής:

$$Y_{0:k} = \{y_1, y_2, \dots, y_k\} \quad (6.3.1)$$

Όπου:

- $Y_{0:k}$, η ακολουθία των διανυσμάτων παρατήρησης,
- k , το πλήθος των διανυσμάτων, με $k \in \mathbb{N}$,
- y_k , το διάνυσμα παρατήρησης, με $y_k \in \mathbb{R}^{14}$.

Οι χαρακτηριστικές ιδιότητες του διανύσματος y_k που καταγράφονται από το σύστημα για κάθε καρέ του βίντεο είναι της μορφής:

$$y_k = [y_1^k, \dots, y_i^k, \dots, y_{14}^k] \quad (6.3.2)$$

Όπου:

- y_k , το διάνυσμα παρατήρησης, με $k \in \mathbb{N}$,
- y_i^k , η χαρακτηριστική ιδιότητα (Σχήμα 6.3.1) του διανύσματος, με $i \in [0,14] \subset \mathbb{N}$:

Χαρακτηριστικές ιδιότητες	Διάνυσμα παρατήρησης y_k	Περιγραφή ιδιοτήτων
y_1^k	x1	Τεταγμένη 1 ^{ου} δακτύλου
y_2^k	yκ-y1	Κατακόρυφη απόσταση κεντροειδούς και 1 ^{ου} δακτύλου
y_3^k	x2	Τεταγμένη 2 ^{ου} δακτύλου
y_4^k	yκ-y2	Κατακόρυφη απόσταση κεντροειδούς και 2 ^{ου} δακτύλου
y_5^k	x3	Τεταγμένη 3 ^{ου} δακτύλου
y_6^k	yκ-y3	Κατακόρυφη απόσταση κεντροειδούς και 3 ^{ου} δακτύλου

y_7^k	x4	Τεταγμένη 4 ^{ου} δακτύλου
y_8^k	yκ-y4	Κατακόρυφη απόσταση κε- ντροειδούς και 4 ^{ου} δακτύλου
y_9^k	x5	Τεταγμένη 5 ^{ου} δακτύλου
y_{10}^k	yκ-y5	Κατακόρυφη απόσταση κε- ντροειδούς και 5 ^{ου} δακτύλου
y_{11}^k	x2-x1	Οριζόντια απόσταση 1 ^{ου} και 2 ^{ου} δακτύλου
y_{12}^k	x3-x2	Οριζόντια απόσταση 2 ^{ου} και 3 ^{ου} δακτύλου
y_{13}^k	x4-x3	Οριζόντια απόσταση 3 ^{ου} και 4 ^{ου} δακτύλου
y_{14}^k	x5-x4	Οριζόντια απόσταση 4 ^{ου} και 5 ^{ου} δακτύλου

Σχήμα 6.3.1 Χαρακτηριστικά κίνησης και ιδιότητές τους

Η καταγραφή των χαρακτηριστικών διανυσμάτων για το κάθε βίντεο εκπαίδευσης παράγει ένα δισδιάστατο πίνακα $E_{14 \times k}$, όπου το πλήθος των γραμμών είναι 14 και το πλήθος των στηλών είναι ίσο με το μήκος του βίντεο εκπαίδευσης.

Τα παραπάνω χαρακτηριστικά διανύσματα, τα οποία χρησιμοποιούνται για την εκπαίδευση των KMM, είναι επαρκή ώστε να επιτευχθεί σωστός διαχωρισμός των κινήσεων μεταξύ τους. Για κάθε μια προς αναγνώριση κίνηση, υπάρχει επαρκές πλήθος χαρακτηριστικών ιδιοτήτων που την κάνουν ευδιάκριτη. Αναλυτικότερα, τα δάκτυλα που πατούν τα πλήκτρα στην ανιούσα ή κατιούσα κίνηση αρπές έχουν μεγαλύτερη απόσταση μεταξύ τους, σε σχέση με την ανιούσα ή κατιούσα κίνηση σε κλίμακα. Συνεπώς οι χαρακτηριστικές ιδιότητες x2-x1, x3-x2, x4-x3, x5-x4 κατανέμονται διαφορετικά στις κινήσεις αρπές σε σχέση με τις υπόλοιπες κινήσεις. Αντίστοιχα, οι ανιούσες από τις κατιούσες κινήσεις διακρίνονται μεταξύ τους από το γεγονός ότι στη μεν ανιούσα οι τιμές των x1, x2, x3, x4, x5 αυξάνονται, ενώ στην κατιούσα μειώνονται. Οι κινήσεις κλίμακας σε τρίτες έχουν την ιδιαιτερότητα των συνηχίσεων, οι οποίες μεταφράζονται

σε αύξηση της κατακόρυφης απόστασης δακτύλου και κεντροειδούς, y_k-y_1 , y_k-y_2 , y_k-y_3 , y_k-y_4 , y_k-y_5 , για ζεύγος δακτύλων ταυτόχρονα.

Είναι χρήσιμο να εξεταστεί η λειτουργικότητα των χαρακτηριστικών διανυσμάτων σε περιπτώσεις όπως α) η αλλαγή μεγέθους, β) η περιστροφή και γ) η μετακίνηση των δακτύλων. Τα αποτελέσματα αναγνώρισης κινήσεων ως προς τις παραπάνω περιπτώσεις μελετώνται σε κάθε σύστημα αναγνώρισης κινήσεων.

Μεγέθυνση ή σμίκρυνση της εικόνας, για το σύστημα αναγνώρισης δακτυλικών κινήσεων, δεν μπορεί να υπάρξει αφού, η θέση της κάμερας, το πλάνο λήψης και το πλάτος των πλήκτρων είναι σταθερά. Τα τρία παραπάνω στοιχεία είναι δεσμευτικά για την τοποθέτηση των δακτύλων στα πλήκτρα του πιάνου και αυτομάτως καθίσταται ευκολότερο το έργο της αναγνώρισης. Η θέση των δακτύλων στον οριζόντιο άξονα, κατά την ερμηνεία ενός πιανίστα, δεν μπορεί να έχει παρά αμελητέες αποκλίσεις σε σχέση με τα βίντεο εκπαίδευσης. Συνεπώς και οι αποστάσεις των δακτύλων που πατούν τα πλήκτρα στον οριζόντιο άξονα, θα έχουν αμελητέες αποκλίσεις λόγω του σταθερού πλάτους των πλήκτρων. Για τον κατακόρυφο άξονα ισχύει κάτι παρόμοιο. Το γεγονός ότι τα δάκτυλα του χεριού οφείλουν να μην «πετάνε», αλλά να είναι τοποθετημένα σε επαφή με τα πλήκτρα, δεν επιτρέπει μεγάλες αποκλίσεις μεταξύ κέντρου μάζας και άκρων δακτύλων.

Το σύστημα αναγνώρισης δακτυλικών κινήσεων είναι απαλλαγμένο από την περιστροφή των δακτύλων στην εικόνα, λόγω του ότι βασίζεται σε θέσεις ή σχετικές θέσεις των δακτύλων και όχι σε γεωμετρικά χαρακτηριστικά όπως για παράδειγμα η επιφάνεια. Οριακές καταστάσεις, που μπορούν ενδεχομένως να δυσκολέψουν την αναγνώριση και σχετίζονται ευρύτερα με την περιστροφή, αναπτύχθηκαν στην ενότητα 5.4 που αφορά στην τμηματοποίηση του χεριού. Οι οριακές καταστάσεις αφορούν ειδικότερα στην περίπτωση που κάποιο δάκτυλο κρύβεται. Στη συντριπτική πλειοψηφία των περιπτώσεων το δάκτυλο που κρύβεται είναι ο αντίχειρας, πιθανόν διότι εναλλάσσεται με άλλο δάκτυλο. Παρατηρήθηκε ότι το δάκτυλο που κρύβεται από τα υπόλοιπα δεν είναι το δάκτυλο που πατάει το πλήκτρο. Συνεπώς, η απόκρυψη αυτή δεν προκαλεί σφάλμα στην ανάκτηση δακτυλισμών. Σε κάθε περίπτωση η θέση του θα εντοπιστεί στα αμέσως επόμενα καρέ. Το ζήτημα της απόκρυψης κάποιου δακτύλου λύνεται οριστικά με την εφαρμογή του αλγόριθμου εκτίμησης θέσης που περιγράφηκε στο εδάφιο 5.6.3.

Η μετακίνηση των δακτύλων πάνω στο κλαβιέ του πιάνου είναι μια υποχρεωτική κίνηση για την ερμηνεία ενός μουσικού έργου.

Το σύστημα αναγνώρισης δακτυλικών κινήσεων είναι προσαρμοσμένο σε τέτοιου είδους κινήσεις. Το είδος της κίνησης καθώς και το μέγεθος της μετατόπισης των δακτύλων εξαρτάται πλήρως από το μουσικό κείμενο, την παρτιτούρα. Συνεπώς, δεν μπορούν να υπάρξουν παρά ελάχιστες αποκλίσεις από τον κάθε πιανίστα σε σχέση με την παρτιτούρα.

Τέλος, η μεθοδολογία, για την αναγνώριση πιανιστικών ερμηνευτικών κινήσεων και ανάκτηση δακτυλισμών, είναι ίδια σε κάθε περίπτωση και σύμφωνα με τους στόχους που τέθηκαν στην ενότητα 1.5, είναι *μη προσαρμοστική*. Γενικά, δεν προαπαιτείται σημαντική προσαρμογή του συστήματος. Παρόλα αυτά, στην περίπτωση εφαρμογής της μεθόδου, σε πιάνο ιδιαίτερων προδιαγραφών, ή αλλαγής θέσης της κάμερας, η μόνη απαιτούμενη προσαρμογή αφορά στη λήψη κατάλληλων βίντεο εκπαίδευσης και όχι στην αλλαγή των χαρακτηριστικών ιδιοτήτων ή του μήκους των χαρακτηριστικών διανυσμάτων.

6.4 Ρόλος των Κρυφών Μαρκοβιανών Μοντέλων στο σύστημα αναγνώρισης δακτυλικών κινήσεων

Η δυναμική αναγνώριση των ερμηνευτικών δακτυλικών κινήσεων του πιανίστα προϋποθέτει τη μοντελοποίηση και την ταξινόμηση των κινήσεων αυτών. Η μέθοδος που επιλέχτηκε για να υποστηρίξει τη διαδικασία της αναγνώρισης στο σύστημα είναι τα KMM. Τα KMM, λόγω του ότι είναι διπλά στοχαστικά μοντέλα, κρίνονται ως η αριότερη μέθοδος για τη μοντελοποίηση των στοχαστικών ιδιοτήτων που έχουν οι κινήσεις που περιγράφηκαν στην ενότητα 6.2.

Τα KMM είναι στατιστικά εργαλεία που επιτρέπουν τη μοντελοποίηση σύνθετων στοχαστικών φαινομένων. Αποτελούνται από ένα σύνολο πεπερασμένων καταστάσεων, οι οποίες συνδέονται με μεταβάσεις. Σε κάθε μια από τις καταστάσεις αυτές αντιστοιχίζεται ένα ζεύγος πιθανοτήτων που αφορούν α) την πιθανότητα μετάβασης και β) την κατανομή πιθανότητας εξόδου. Για το σύστημα αναγνώρισης δακτυλικών κινήσεων επιλέχτηκε η μοντελοποίηση με συνεχή KMM. Συνεπώς, η κατανομή πιθανότητας εξόδου είναι μια συνεχής ΣΠΠ. Με δεδομένη την κατάσταση, καθορίζεται μέσω της κατανομής, η πιθανότητα εμφάνισης κάποιου συνεχούς τυχαίου διανύσματος.

Ο λόγος που επιλέχτηκαν τα KMM ως μέθοδος αναγνώρισης δακτυλικών κινήσεων από το σύστημα είναι ότι η μέθοδος της ΔΠΤ απαιτεί το σύνολο του σήματος για να μπορέσει να πραγματοποιήσει την αναγνώριση. Συνεπώς, δε δύναται να χρησιμοποιηθεί η ΔΠΤ στην ανα-

γνώριση κινήσεων σε πραγματικό χρόνο, γεγονός που θα περιορίζει την επέκταση του παρόντος συστήματος. Τα KMM αντίθετα, προσφέρουν δυνατότητα αναγνώρισης κινήσεων βασιζόμενα μόνο στο σήμα που έχει φτάσει μέχρι μια δεδομένη χρονική στιγμή.

Ερμηνεύοντας και αναλύοντας την κίνηση των δακτύλων του πιανίστα καταλήγουμε στο συμπέρασμα ότι περιέχει χωρικές και χρονικές μεταβολές. Το συμπέρασμα αυτό ανάγει την κίνηση σε μια διπλά στοχαστική διαδικασία. Οι μαρκοβιανές διαδικασίες μπορούν να συνεισφέρουν στον τομέα της αναγνώρισης επειδή και αυτές είναι διπλά στοχαστικές διαδικασίες. Οι εσωτερικές καταστάσεις της κίνησης, οι οποίες αποτελούν και τα δομικά στοιχεία της, δεν μπορούν να παρατηρηθούν άμεσα. Τα KMM μπορούν να επιτύχουν τη μοντελοποίηση ακριβώς τέτοιων διαδικασιών απαλλάσσοντας το σύστημα από χρονικές παραμέτρους. Είναι επίσης αυταπόδεικτο ότι, αν ένας πιανίστας εκτελέσει την ίδια δακτυλική κίνηση περισσότερο από μια φορά, τότε οι κινήσεις αυτές δεν μπορούν να είναι ακριβώς ίδιες. Το γεγονός αυτό επιβεβαιώνει την ανάγκη χρησιμοποίησης κατάλληλων στατιστικών μοντέλων, όπως είναι τα KMM.

Το σήμα του βίντεο ή το σήμα της μουσικής είναι ένα συνεχές σήμα. Ιδιαίτερο ενδιαφέρον παρουσιάζει, στα πλαίσια της παρούσας έρευνας, η αναπαράσταση του σήματος βίντεο μέσα από μοντέλα σήματος. Απώτερος σκοπός τη μοντελοποίησης του σήματος είναι η πρακτική και αποτελεσματική επεξεργασία του, που εν προκειμένου είναι η αναγνώριση των δακτυλικών κινήσεων. Το παρόν σύστημα χρησιμοποιεί στατιστική αναπαράσταση σήματος. Έτσι, το σήμα του βίντεο μπορεί να χαρακτηριστεί ως μια παραμετρική τυχαία διαδικασία, οι παράμετροι της οποίας μπορούν να εκτιμηθούν. Αυτό ακριβώς το σημείο αγγίζει συμβολή των KMM.

Μια μαρκοβιανή διαδικασία είναι ένα σύστημα διακριτού χρόνου, το οποίο βρίσκεται κάθε χρονική στιγμή σε μια κατάσταση μεταξύ ενός συνόλου καταστάσεων. Οι μεταβάσεις μεταξύ των καταστάσεων παράγονται μεταξύ δύο συγκεκριμένων διαδοχικών χρονικών στιγμών, ακολουθώντας τους νόμους των πιθανοτήτων. Η πιθανότητα το μοντέλο να βρεθεί σε κάποια κατάσταση δεν εξαρτάται παρά μόνο από την αμέσως προηγούμενη κατάσταση. Ένα KMM αναπαριστά, με τον ίδιο τρόπο με μια μαρκοβιανή αλυσίδα, ένα σύνολο ακολουθιών παρατήρησης όπου η κατάσταση της κάθε παρατήρησης δεν είναι παρατηρούμενη, αλλά συνδεδεμένη με μια ΣΠΠ. Πρόκειται για διπλά στοχαστικές διαδικασίες όπου οι παρατηρήσεις εξαρτώνται τυχαία από τις καταστάσεις, ενώ η επιλογή κατάστασης

για μια οποιαδήποτε χρονική στιγμή εξαρτάται από τις πιθανότητες μετάβασης της προηγούμενης κατάστασης.

6.4.1 Κρυφά Μαρκοβιανά Μοντέλα σε διακριτό ή συνεχή χώρο παρατηρήσεων

Στόχος αυτού του εδαφίου είναι να παρουσιαστούν επιγραμματικά τα διαφορετικά είδη των KMM.

Το είδος ενός KMM μπορεί να καθοριστεί με βάση [Alani 1994a] (Σχήμα 6.4.1.1):

1. Τους περιορισμούς στον πίνακα μετάβασης μεταξύ των καταστάσεων,
2. Το είδος της πυκνότητας πιθανότητας των παρατηρήσεων,
3. Τη διάρκεια παραμονής σε μία κατάσταση.

Περιορισμοί	Είδη KMM			
σχετικά με τον πίνακα μεταβάσεων	εργοδικά KMM	αριστερά προς τα δεξιά KMM		
σχετικά με την πιθανότητα των παρατηρήσεων	διακριτά KMM	συνεχή KMM	ημι-συνεχή KMM	ημι-μαρκοβιανά KMM
σχετικά με τη διάρκεια παραμονής σε μια κατάσταση	KMM διακριτής μεταβλητής διάρκειας	KMM συνεχούς μεταβλητής διάρκειας		

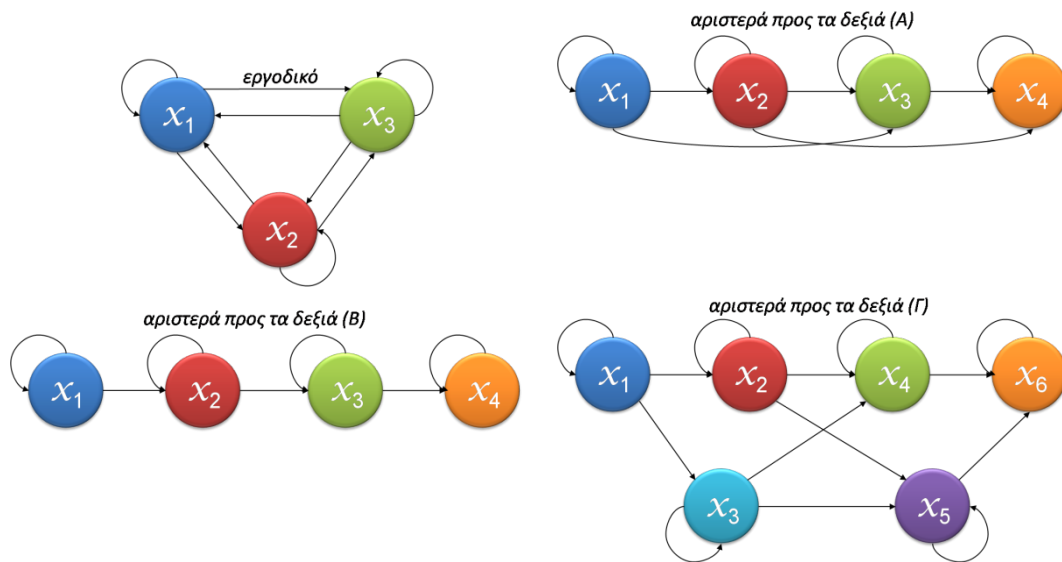
Σχήμα 6.4.1.1 Είδη KMM ανάλογα με τους περιορισμούς

1. Περιορισμοί πάνω στον πίνακα μεταβάσεων

Εργοδικό μοντέλο : Αυτό το μοντέλο δεν έχει περιορισμούς πάνω στον πίνακα μεταβάσεων. Κάθε κατάσταση είναι επισκέψιμη, από οποιαδήποτε άλλη κατάσταση, με απλή μετάβαση (Σχήμα 6.4.1.2).

Αριστερά προς τα δεξιά KMM : Αυτά τα μοντέλα χρησιμοποιούνται για να μοντελοποιήσουν διαδικασίες που χαρακτηρίζονται από μετα-

βλητές ιδιότητες στο χρόνο, όπως για παράδειγμα σήματα ανθρώπινης φωνής. Υπάρχουν περιορισμοί ως προς την επισκεψιμότητα κάποιων καταστάσεων από κάποιες άλλες. Διακρίνονται σε αριστερά προς τα δεξιά KMM με α) περιορισμούς σε σειρά για διπλές μεταβάσεις, β) περιορισμούς σε σειρά για μοναδική μετάβαση, γ) παράλληλους περιορισμούς. Στο σχήμα 6.4.1.2 φαίνονται οι περιορισμοί και οι επιτρεπτές μεταβάσεις μεταξύ των καταστάσεων [Bakis 1976].



Σχήμα 6.4.1.2 Γραφική αναπαράσταση KMM ανά είδος.

Οι επιτρεπτές μεταβάσεις αναπαρίστανται με βέλος. Σε κάθε μετάβαση αντιστοιχεί μια πιθανότητα μετάβασης.

2. Περιορισμοί ως προς τον τύπο της πυκνότητας πιθανότητας των παρατηρήσεων

Ανάλογα με τον τύπο της πυκνότητας της πιθανότητας των παρατηρήσεων, διακριτή ή συνεχή, μπορούν να δημιουργηθούν δύο διαφορετικοί τύποι μοντέλων: α) διακριτά KMM ή β) συνεχή KMM.

Διακριτά KMM : Οι παρατηρήσεις, στην πλειοψηφία τους, είναι πάνω σε συνεχή χώρο αφού προέρχονται από φυσικά συνεχή φαινόμενα. Στην περίπτωση των διακριτών KMM, οι συνεχείς παρατηρήσεις υπόκεινται σε κβαντοποίηση με τη βοήθεια ενός λεξικού. Το μειονέκτημα της διαδικασίας αυτής είναι η σημαντική απώλεια πληροφοριών που συνδέονται με την κβαντοποίηση.

Συνεχή KMM : Προκειμένου να περιοριστούν οι απώλειες σημαντικών πληροφοριών λόγω της κβαντοποίησης, μπορεί να επιλεγεί μια ΣΠΠ για τις συνεχείς παρατηρήσεις. Για να εκτιμηθούν οι παράμετροι της ΣΠΠ, οφείλεται να λαμβάνονται υπόψη οι περιορισμοί που έχει η συνάρτηση στο είδος του μοντέλου. Έχουν καθοριστεί διαδικασίες επανεκτίμησης των παραμέτρων των μοντέλων, ανάλογα με το είδος της ΣΠΠ σε α) ΣΠΠ με ελλειπτική ή λογαριθμική μορφή [Baum 1972; Liporace 1982], β) μείγμα πεπερασμένων πολυδιάστατων Γκαουσιανών κατανομών [Juang 1985a] και γ) αυτοπαλινδρομικές Γκαουσιανές συναρτήσεις [Juang 1985b].

Ημι-συνεχή KMM : Στην περίπτωση των διακριτών KMM, η απώλεια πληροφορίας που συνδέεται με την κβαντοποίηση μπορεί να μειωθεί αν, κατά τη διάρκεια της αποκωδικοποίησης (*αλγόριθμος Viterbi*) χρησιμοποιηθεί συνεχές διάνυσμα παρατηρήσεων αντί για το διακριτό σύμβολο παρατήρησης. Οι Huang και Jack (1988) παρουσίασαν μια προσέγγιση που συνδυάζει διακριτά και συνεχή KMM [Huang 1988]. Το μοντέλο που παράγεται ονομάζεται « Ημι-συνεχές Κρυφό Μαρκοβιανό Μοντέλο ». Το σκεπτικό των μοντέλων αυτών βασίζεται στην αντικατάσταση των πιθανοτήτων των διακριτών συμβόλων παρατηρήσεων, από συνδυασμό διακριτών συμβόλων παρατηρήσεων και συνεχών ΣΠΠ, προερχόμενες από το λεξικό των κβαντοποιημένων διανυσμάτων.

Ημι-μαρκοβιανά KMM : Σε όλα τα παραπάνω είδη KMM θεωρήθηκε δεδομένη η ανεξαρτησία των διανυσμάτων παρατήρησης μιας ακολουθίας. Αυτή η υπόθεση αποτελεί ένα σαφή περιορισμό που δεν επιτρέπει την εφαρμογή των KMM σε ορισμένες δυναμικές διαδικασίες πραγματικού χρόνου. Τα ημι-μαρκοβιανά KMM προτείνουν τη γενίκευση αυτής της υπόθεσης, ορίζοντας μια νέα πιθανότητα γέννησης παρατηρήσεων που λαμβάνει υπόψη τη συσχέτιση μεταξύ των διανυσμάτων που αντιπροσωπεύουν τις συνεχείς παρατηρήσεις [Celeux 1992]

3. Περιορισμοί ως προς τη διάρκεια παραμονής σε μια κατάσταση

Τα KMM χαρακτηρίζονται από απουσία πληροφορίας που να σχετίζεται με την πιθανότητα παραμονής του μοντέλου σε μια κατάσταση. Το στοιχείο αυτό αποτελεί πλεονέκτημα για το σύστημα αναγνώρισης δακτυλικών κινήσεων, διότι δύναται να αναγνωριστεί μια κίνηση α-

νεξάρτητα με το τέμπο που ορίζεται από την παρτιτούρα ή το τέμπο που υιοθετεί ο εκάστοτε πιανίστας. Σε άλλες όμως περιπτώσεις αναγνώρισης, όπως στην αναγνώριση φωνής, είναι χρήσιμο να υπάρχει αυτή η παράμετρος. Για το σκοπό αυτό εισάγονται ως παράμετροι του μοντέλου οι χρόνοι παραμονής σε μια κατάσταση πριν τη μετάβαση του μοντέλου σε μια άλλη κατάσταση. Δύο τέτοιες μέθοδοι έχουν αναπτυχθεί, α) του Ferguson (1980) και β) του Levinson (1986) [Ferguson 1980; Levinson 1986]. Η πρώτη μέθοδος βασίζεται πάνω σε ένα « KMM διακριτής μεταβλητής διάρκειας » ενώ η δεύτερη σε ένα « KMM συνεχόμενα μεταβλητής διάρκειας ».

6.4.2 Επιλογή συνεχών Κρυφών Μαρκοβιανών Μοντέλων

Στόχος του εδαφίου αυτού είναι η επιγραμματική δικαιολόγηση της επιλογής των συνεχών KMM για τη μοντελοποίηση των δακτυλικών κινήσεων του συστήματος.

Στο σχήμα 6.4.2.1 φαίνεται ένας συγκριτικός πίνακας μεταξύ των διακριτών και των συνεχών KMM [Alani 1994b]. Ένα πρώτο συμπέρασμα μελετώντας τον πίνακα είναι ότι τα διακριτά KMM είναι πιο εύχρηστα σε ό,τι αφορά το σχεδιασμό και την υλοποίησή τους. Πλην όμως, υστερούν σε ακρίβεια, ενώ απαιτούν μεγάλο όγκο δεδομένων εκπαίδευσης. Οι παράμετροι που έπαιξαν καθοριστικό ρόλο στην επιλογή των συνεχών KMM, ως μέθοδο μοντελοποίησης των δακτυλικών κινήσεων, είναι α) η ακρίβεια στην ταξινόμηση, κατ'επέκταση και στην αναγνώριση, β) η αποφυγή κβαντοποίησης των δεδομένων, άρα ενδεχόμενης απώλειας σημαντικών πληροφοριών, και τέλος γ) η μη απαίτηση μεγάλου πλήθους ακολουθιών δεδομένων εκπαίδευσης.

Η παράμετρος της ακρίβειας στην ταξινόμηση παίζει σημαντικότερο ρόλο στην ανάπτυξη του συστήματος, δεδομένου ότι τα δάκτυλα ενός πιανίστα αναπτύσσουν μεγάλες ταχύτητες αλλά και ότι πολλές κινήσεις διαφέρουν ελάχιστα μεταξύ τους. Συνεπώς, απαιτείται ακρίβεια και λεπτομέρεια στην ταξινόμηση, ώστε να επιτυγχάνεται αντιστοίχως σωστή αναγνώριση.

Στη συνέχεια, η διαδικασία κβαντοποίησης είναι υποχρεωτική για την περίπτωση των διακριτών KMM αλλά όχι για την περίπτωση των συνεχών. Η διαφορά αυτή αποτελεί ένα συγκριτικό πλεονέκτημα των συνεχών KMM έναντι των διακριτών, αφού δίνει μια σχετική ευελιξία στη χρησιμοποίησή τους, ειδικά στην περίπτωση χρήσης του συστήματος για αναγνώριση σε πραγματικό χρόνο. Επίσης, αποφεύγεται ενδεχόμενη απώλεια σημαντικών πληροφοριών μέ-

σω της κβαντοποίησης, δεδομένα τα οποία συμμετέχουν στην ακριβέστερη ταξινόμηση των κινήσεων.

Τέλος, τα συνεχή KMM δεν απαιτούν μεγάλο πλήθος ακολουθιών δεδομένων εκπαίδευσης, ενώ παράλληλα μπορούν να δεχτούν ακόμα και ελλιπή δεδομένα. Τα δύο αυτά χαρακτηριστικά των συνεχών KMM προσδίδουν μεγάλη ευχρηστία στο σύστημα, αφού κάνουν τη διαδικασία της εκπαίδευσης αρκετά εύελικτη.

Πλήθος παραμέτρων προς εκτίμηση	Μεγάλο	-
	Μικρότερο από τα συνεχή	+
Ακρίβεια στην ταξινόμηση	Υψηλή	+
	Χαμηλότερη από συνεχή	-
Υποθέσεις ως προς τη φύση των παρατηρήσεων	Σημαντικές	-
	Λιγότερο σημαντικές	+
Σχεδιασμός μοντέλου	Χρονοβόρος & δύσκολος	-
	Γρήγορος & εύκολος	+
Πλήθος απαιτούμενων δεδομένων για εκπαίδευση	Μέτριο	+
	Μεγαλύτερο από τα συνεχή	-

Σχήμα 6.4.2.1 Συγκριτικός πίνακας μεταξύ συνεχών (μπλε) και διακριτών (πράσινο) KMM.

6.5 Ανάλυση και στοχαστική μοντελοποίηση κινήσεων με Κρυφά Μαρκοβιανά Μοντέλα

Στο σύστημα αναγνώρισης δακτυλικών κινήσεων ενδιαφερόμαστε για διακριτές τυχαίες καταστάσεις μέσα από ένα πεπερασμένο σύνολο καταστάσεων X (Τα στοιχεία του X χρησιμεύουν στη διάκριση των καταστάσεων της μαρκοβιανής αλυσίδας. Γενικά, οι αλγεβρικές πράξεις πάνω στα στοιχεία του X δεν έχουν κανένα νόημα.). Το X θα θεωρείται, στο εξής, το αλφαβητάρι των δακτυλισμών του εν λόγω συστήματος. Μπορεί να φανταστεί κανείς το X σαν το σύνολο από όλους τους δυνατούς δακτυλισμούς, ή ακόμα ως στάσεις των δακτύλων του πιανίστα στο κλαβιέ του πιάνου. Σε άλλη

περίπτωση το αλφαβητάρι, ενός εντελώς διαφορετικού συστήματος, θα μπορούσε να είναι το $Z=\{1,2,3,4,5,6\}$, το οποίο αντιστοιχεί σε όλες τις δυνατές ρίψεις ενός ζαριού. Παραστατικά, το αλφαβητάρι των δακτυλισμών απαρτίζεται ενδεικτικά από καταστάσεις της μορφής:

$$X=\{ \begin{array}{|c|c|c|c|c|c|} \hline \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \hline \end{array} 1, \dots, \begin{array}{|c|c|c|c|c|c|} \hline \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \hline \end{array} 4, \dots, \begin{array}{|c|c|c|c|c|c|} \hline \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \hline \end{array} 3 \}$$
 (6.5.1)

Πρέπει να σημειωθεί το γεγονός ότι κάθε στοιχείο του συνόλου X δεν αντιστοιχεί σε διακριτή θέση (διακριτές συντεταγμένες) των δακτύλων πάνω στο κλαβιέ, αλλά ανήκουν σε διάστημα τιμών.

Ας θεωρήσουμε τη μαρκοβιανή αλυσίδα $X=(X_k)$ με $k \in \mathbb{N}$, πάνω στον πεπερασμένο χώρο καταστάσεων X , με αρχική κατανομή ρ και πίνακα μεταβάσεων $Q = [Q(x, x')]_{x, x' \in X}$. Η αλυσίδα X δεν είναι απευθείας παρατηρούμενη γι' αυτό και ονομάζεται κρυφή. Διαθέτουμε επίσης, τα διανύσματα παρατήρησης $Y=\{Y_k\}$ με $k \in \mathbb{N}$, με τιμές μέσα από το συνεχή χώρο παρατηρήσεων $\mathcal{Y}$. Τα διανύσματα παρατήρησης Y προέκυψαν από τα καρέ παρατήρησης K (ενότητα 6.3).

Υποθέτουμε ότι τα καρέ παρατήρησης, και κατ' επέκταση τα διανύσματα παρατήρησης, αποκτήθηκαν μέσω μιας *κάμερας χωρίς μνήμη*. Δηλαδή, υποθέτουμε ότι τα διανύσματα παρατήρησης Y_k είναι ανεξάρτητα μεταξύ τους και ότι κάθε διάνυσμα εξαρτάται μόνο από την κατάσταση X_k . Στο εξής, θα σημειώνουμε με $\psi(x, y)$ τη σχετική πιθανοφάνεια.

Τα τρία προβλήματα των KMM που το σύστημα καλείται να επιλύσει είναι τα εξής:

Π1. Αναγνώριση

Έχοντας ως δεδομένες τις παραμέτρους $M = (\rho, Q, \psi)$ του μοντέλου, ποια είναι η πιθανοφάνεια μιας ακολουθίας διανυσμάτων παρατήρησης $Y_{0:k}$; Με άλλα λόγια, ποια είναι η πιθανότερη ακολουθία καταστάσεων $X_{0:k}$ που γέννησε την $Y_{0:k}$;

Αυτό το πρόβλημα επιλύεται από το σύστημα με τη βοήθεια του αλγορίθμου “forward” του Baum [Baum 1972].

Π2. Ερμηνεία

Έχοντας ως δεδομένα α) τις παραμέτρους $M = (\varrho, Q, \psi)$ του μοντέλου και β) μια ακολουθία διανυσμάτων παρατήρησης $Y_{0:k}$, πρέπει να καθοριστεί η πιο πιθανοφανής ακολουθία $X_{0:k}$ ως προς την $Y_{0:k}$.

Αυτό το πρόβλημα επιλύεται από το σύστημα μέσω του αλγορίθμου του Viterbi, με χρήση της μεθόδου εκτίμησης τροχιάς [Rabiner 1989].

Π3. Εκπαίδευση

Έχοντας ως δεδομένη μια ακολουθία διανυσμάτων παρατήρησης $Y_{0:k}$, πρέπει να καθοριστούν οι παράμετροι $M = (\varrho, Q, \psi)$ του μοντέλου που είναι πιθανοφανείς ως προς την $Y_{0:k}$.

Αυτό το πρόβλημα επιλύεται από το σύστημα με τη βοήθεια του αλγορίθμου των Baum-Welch [Dempester 1977], με χρήση της μεθόδους μέγιστης πιθανοφάνειας.

6.6 Επίλυση των προβλημάτων των Κρυφών Μαρκοβιανών Μοντέλων

Προτού αναπτυχθούν οι προτεινόμενες λύσεις των τριών προβλημάτων των KMM, κρίνεται αναγκαίο να παρουσιαστούν τα βασικά εργαλεία που χρησιμοποιούνται από το σύστημα για την επίλυση των προβλημάτων αυτών. Αυτά είναι τα παρακάτω:

1. Μαρκοβιανή αλυσίδα

Θεωρούμε, πάνω στο πεπερασμένο σύνολο X , τη μαρκοβιανή αλυσίδα X :

$$X = (X_k)_{k \in \mathbb{N}} \quad (6.6.1)$$

Όπου:

- X , η μαρκοβιανή αλυσίδα των καταστάσεων,
- X_k , μια τυχαία κατάσταση της μαρκοβιανής αλυσίδας, με $k \in \mathbb{N}$.

2. Πίνακας μεταβάσεων

Θεωρούμε τον πίνακα Q με τις πιθανότητες μετάβασης των καταστάσεων:

$$Q = [Q(x, x')]_{x, x' \in \mathcal{X}} \quad (6.6.2)$$

Όπου:

- Q , ο πίνακας των πιθανοτήτων μετάβασης,
- $Q(x, x')$, η πιθανότητα μετάβασης μεταξύ των καταστάσεων x και x' , με $x, x' \in \mathcal{X}$.

3. Διανύσματα παρατήρησης

Διαθέτουμε ήδη τις παρατηρήσεις Y :

$$Y = \{Y_k\}_{k \in \mathbb{N}} \quad (6.6.3)$$

Όπου:

- Y , το σύνολο των διανυσμάτων παρατήρησης,
- Y_k , ένα τυχαίο διάνυσμα παρατήρησης, με $k \in \mathbb{N}$.

Για κάθε καρτέ παρατήρησης K_k έχει καταγραφεί ένα διάνυσμα παρατήρησης Y_k (ενότητα 6.3), το οποίο χαρακτηρίζεται από 14 ιδιότητες. Οι παρατηρήσεις του K αντιστοιχούν μια προς μια στα διανύσματα παρατήρησης Y , τα οποία κινούνται πάνω στο συνεχή χώρο $\mathcal{Y}$.

Ας θεωρήσουμε, στο σημείο αυτό, ότι $\mathcal{Y} \triangleq \mathbb{R}^{14}$.

Υπόθεση 7.4.1 (πηγή λήψης χωρίς μνήμη) Υποθέτουμε ότι η κάμερα, ως πηγή λήψης των καρτέ και κατά συνέπεια των διανυσμάτων παρατήρησης, δε διαθέτει μνήμη $\forall k \geq 1, y_1, \dots, y_k \in \mathcal{Y}$:

$$\mathcal{P}(Y_{0:k} \in d y_{0:k} | X_{0:k} = x_{0:k}) = \prod_{t=0}^k \mathcal{P}(Y_t \in d y_t | X_t = x_t) \quad (6.6.4)$$

δηλαδή,

$$\kappa\tau\nu(Y_{0:k} | X_{0:k}) = \kappa\tau\nu(Y_0 | X_0 = x_0) \otimes \dots \otimes \kappa\tau\nu(Y_k | X_k = x_k) \quad (6.6.5)$$

όπου, $\kappa\tau\nu$ =κατανομή.

Ορισμός 7.4.2 Ένας μαρκοβιανός πυρήνας πάνω στο $\mathcal{X} \times \mathbb{R}^{14}$ είναι μια οικογένεια συναρτήσεων $\psi = \{\psi(x, \cdot); x \in \mathcal{X}\}$, $\mathbb{R}^{14} \ni y \mapsto \psi(x, y) \in \mathbb{R}$ τέτοια ώστε:

$$1. \psi(x, y) \geq 0 \quad \forall x \in \mathcal{X}, \forall y \in \mathbb{R}^{14},$$

$$2. \int_{\mathbb{R}^{14}} \psi(x, y) d y \geq 0, \forall x \in \mathcal{X}.$$

Υποθέτουμε ότι η δεσμευμένη κατανομή της Y_k , γνωρίζοντας την X_k , δεν εξαρτάται από το k αλλά εξάγει μια πυκνότητα. Θέτουμε λοιπόν:

$$\mathcal{P}(Y_k \in d y_k | X_k = x_k) = \psi(x, y) d y \quad (6.6.6)$$

Προφανώς αποδεικνύεται ότι ο $\psi(x, y)$ είναι ένας γκαουσιανός πυρήνας επίσης. Από την προηγούμενη σχέση εξάγεται το παρακάτω συμπέρασμα:

Πρόταση 7.4.3 Η κατανομή του KMM (X, Y) χαρακτηρίζεται πλήρως από την αρχική κατανομή ρ , τον πίνακα μεταβάσεων Q της μαρκοβιανής αλυσίδας X , καθώς επίσης και από τον μαρκοβιανό πυρήνα ψ . Στην πράξη ισχύει:

$$\mathcal{P}(X_{0:k} = x_{0:k}, Y_{0:k} \in d y_{0:k}) = \psi(x_0) \psi(x_0, y_0) \prod_{t=1}^k [Q(x_{t-1}, x_t) \psi(x_t, y_t)] d y_{0:k} \quad (6.6.7)$$

Από την παραπάνω σχέση εξάγεται ο παρακάτω ορισμός:

Ορισμός 7.4.4 (Κρυφό Μαρκοβιανό Μοντέλο) Ένα KMM (X, Y) σε συνεχή χώρο παρατηρήσεων χαρακτηρίζεται πλήρως από:

1. Την αρχική κατανομή ρ πάνω στη διαδικασία X :

$$\psi(x) \triangleq P(X_0 = x), x \in X \quad (6.6.8)$$

2. Τον πίνακα με τις πιθανότητες μετάβασης Q πάνω στη διαδικασία X :

$$Q(x, x') \triangleq P(X_{k+1} = x' | X_k = x), x, x' \in X \quad (6.6.9)$$

3. Το μαρκοβιανό πυρήνα ψ της διαδικασίας Y γνωρίζοντας τη διαδικασία X :

$$P(Y_k \in dy | X_k = x_k) = \psi(x, y) dy \quad (6.6.10)$$

Αφού ορίστηκαν πλήρως οι παράμετροι $M = (\rho, Q, \psi)$, μπορούν να μοντελοποιηθούν τα διανύσματα παρατήρησης, προτού προταθούν λύσεις για το κάθε πρόβλημα χωριστά.

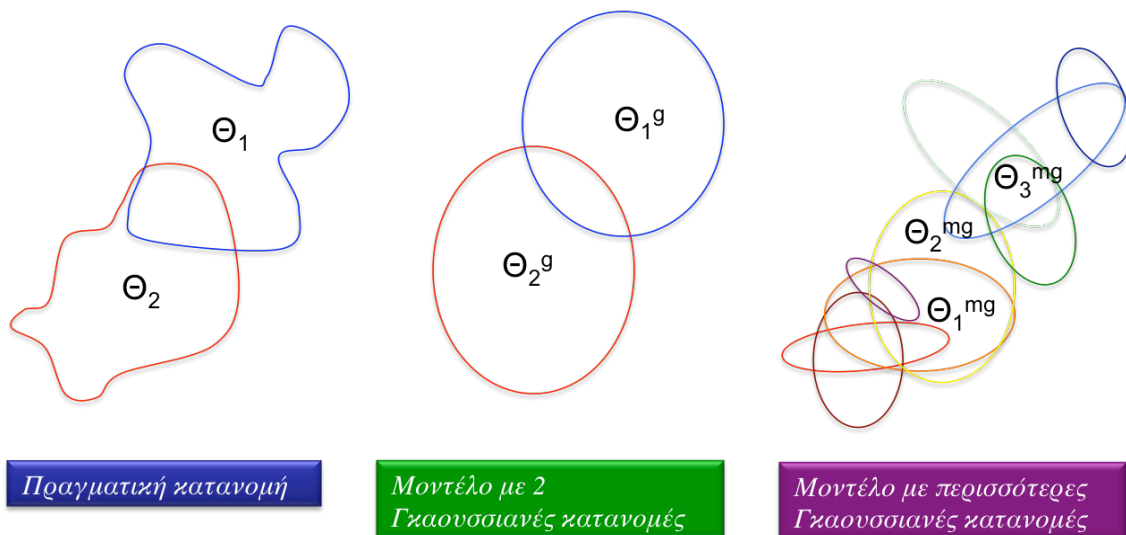
6.6.1 Μοντελοποίηση διανυσμάτων παρατήρησης με Μεικτά Γκαουσιανά Μοντέλα

Τα μοντέλα των μεικτών γκαουσιανών κατανομών, ή αλλιώς Μεικτά Γκαουσιανά Μοντέλα, χρησιμοποιούνται σε πλήθος επιστημονικών πεδίων, όπως είναι για παράδειγμα η επεξεργασία και η αναγνώριση εικόνων ή ανθρώπινης φωνής. Στοχεύουν στη δημιουργία ταξινομητών μέσω της εκπαίδευσης μοντέλων.

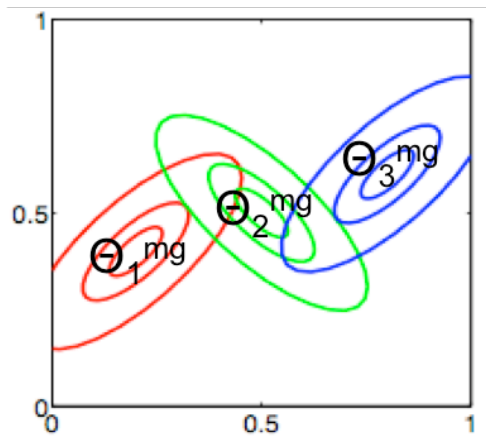
Στην περίπτωση του συστήματος αναγνώρισης δακτυλικών κινήσεων επιθυμείται η πιο λεπτομερής μοντελοποίηση των δεδομένων παρατηρήσεων, λαμβάνοντας υπόψη τη μικρότερη δυνατή απόσταση μεταξύ της μέτρησης $y_{0:k}$ και του κεντροειδούς (Γκαουσιανή κατανομή). Στόχος είναι η προσέγγιση της κατανομής πιθανότητας της ακολουθίας διανυσμάτων παρατήρησης $Y_{0:k}$ σε σχέση με τις πα-

ραμέτρους του ΜΓΜ, δηλαδή ο υπολογισμός της συνάρτησης πιθανοφάνειας.

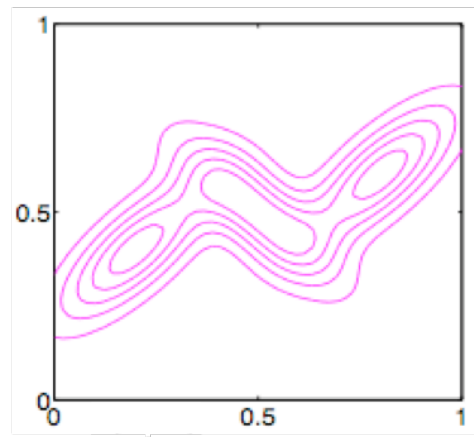
Στα σχήματα 6.6.1.1 και 6.6.1.2 γίνεται πιο κατανοητή η χρησιμότητα μοντελοποίησης των διανυσμάτων παρατήρησης με ΜΓΜ. Ας θεωρήσουμε ότι η πραγματική κατανομή κάποιων συνεχών μετρήσεων μπορεί να αναπαρασταθεί επακριβώς μέσω δύο άγνωστων κατανομών Θ_1 και Θ_2 . Εφόσον οι Θ_1 και Θ_2 είναι άγνωστες, θα πρέπει κάπως να προσεγγιστούν με άλλο τρόπο. Στην περίπτωση αναγνώρισης φωνής, ήχου, ομιλίας ή χειρονομιών, προτείνονται τα ΜΓΜ από τη βιβλιογραφία. Η προσπάθεια να προσεγγιστεί η πραγματική κατανομή των μετρήσεων από δύο Γκαουσιανές κατανομές Θ_1^g και Θ_2^g σαφώς έχει θετικά αποτελέσματα, πλην όμως δεν είναι πλήρως ικανοποιητικά. Αυτό που ενδιαφέρει σε κάθε περίπτωση είναι το περίγραμμα των συνιστάμενων κατανομών. Είναι προφανές ότι το περίγραμμα της συνισταμένης κατανομής στην πρώτη και στη δεύτερη περίπτωση απέχει αρκετά. Χρησιμοποιώντας όμως ένα ΜΓΜ με περισσότερες κατανομές, όπως στην τρίτη περίπτωση, διαπιστώνουμε ότι το περίγραμμα του ΜΓΜ πλησιάζει πολύ ικανοποιητικά την πραγματική κατανομή. Η ομαλότερη προσέγγιση του περιγράμματος της πραγματικής κατανομής από το ΜΓΜ φαίνεται με παραστατικό-τερο τρόπο στο σχήμα 6.6.1.2.



Σχήμα 6.6.1.1 Παράδειγμα πραγματικής κατανομής και προσέγγισής της από ΜΓΜ.



Μοντέλο 3 Γκαουσιανών κατανομών



Περιγράμματα της κατανομής

Σχήμα 6.6.1.2 Παράδειγμα ΜΓΜ και περιγράμματα της κατανομής

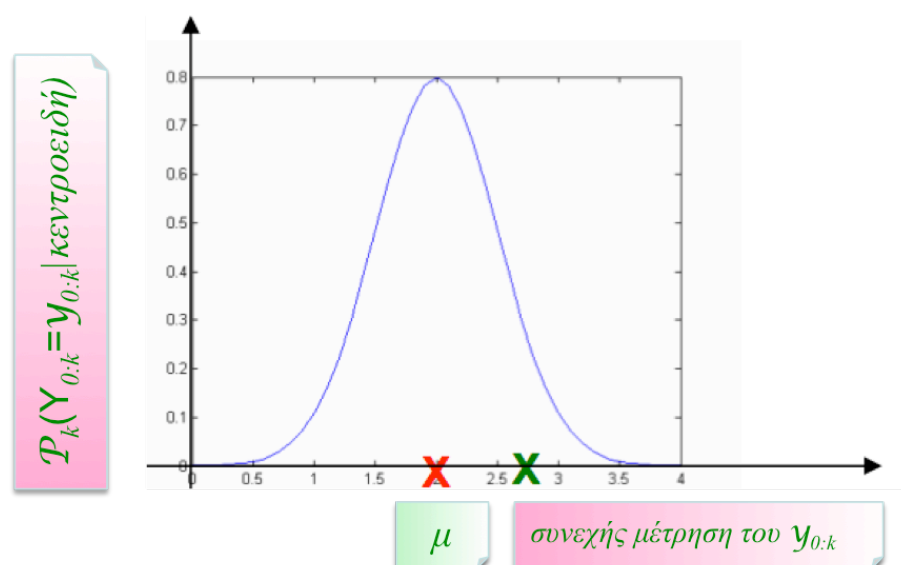
Στο σχήμα 6.6.1.3 απεικονίζεται ένα παράδειγμα συνάρτησης πιθανοφάνειας πάνω σε συνεχείς μετρήσεις του $y_{0:k}$. Η συνάρτηση πιθανοφάνειας που χρησιμοποιήθηκε είναι η :

$$P_k = (Y_{0:k} | \text{κεντροειδή}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_{0:k} - \mu)^2}{2\sigma^2}\right) \quad (6.6.1.1)$$

Όπου:

- $P_k(Y_{0:k} = y_{0:k} | \text{κεντροειδή})$, η συνάρτηση πιθανοφάνειας,
- μ , η μέση τιμή της συνάρτησης πιθανοφάνειας,
- σ , η διασπορά της συνάρτησης πιθανοφάνειας.

Όσο το **X** της μέτρησης $y_{0:k}$ βρίσκεται πιο κοντά στο **X** του κεντροειδούς της μέσης τιμής, τόσο περισσότερο η μέτρηση $y_{0:k}$ είναι πιθανοφανής. Κατ' επέκταση το ίδιο ισχύει και για τη διασπορά. Με άλλα λόγια, αν η απόσταση μεταξύ **X** και **X** είναι ελάχιστη τότε υπάρχει πολύ μεγάλη πιθανοφάνεια, αν η απόσταση είναι πολύ μεγάλη τότε έχουμε ελάχιστη πιθανοφάνεια.



Σχήμα 6.6.1.3 Συνάρτηση πιθανοφάνειας $\mathcal{L}_k$ για συνεχή μέτρηση του $y_{0:k}$.

X ο κεντροειδής της μέσης τιμής,

X μια μέτρηση του $y_{0:k}$.

Στην περίπτωση που είτε δεν είναι γνωστή η συνάρτηση πιθανοφάνειας, είτε είναι γνωστό ότι δεν πρόκειται για γκαουσσισιανή κατανομή, όπως και στην περίπτωση της παρούσας έρευνας, προσεγγίζονται τα διανύσματα παρατήρησης από ένα μείγμα γκαουσσισιανών κατανομών.

Ας θεωρήσουμε δύο γκαουσσισιανές κατανομές με ισοπίθανους κεντροειδείς. Αθροίζοντας τις πιθανοφάνειες των επιμέρους γκαουσσισιανών κατανομών προκύπτει :

$$P(Y_{0:k} = y_{0:k} | \text{MGM}) = \frac{1}{2} P(Y_{0:k} = y_{0:k} | c_1) + \frac{1}{2} P(Y_{0:k} = y_{0:k} | c_2) \quad (6.6.1.2)$$

Όπου:

- $P(Y_{0:k} = y_{0:k} | \text{MGM})$, η συνάρτηση πιθανοφάνειας του MGM,
- $P(Y_{0:k} = y_{0:k} | c_1)$, η πρώτη γκαουσσισιανή κατανομή με κεντροειδή c_1 ,
- $P(Y_{0:k} = y_{0:k} | c_2)$, η δεύτερη γκαουσσισιανή κατανομή με κεντροειδή c_2 .

Στην περίπτωση που οι κεντροειδείς των δύο κατανομών δεν είναι ισοπίθανοι, τότε η σχέση 6.6.1.2 παίρνει τη μορφή :

$$\mathcal{P}(Y_{0:k} = y_{0:k} | \text{MGM}) = P(c_1) \mathcal{P}(Y_{0:k} = y_{0:k} | c_1) + P(c_2) \mathcal{P}(Y_{0:k} = y_{0:k} | c_2) \quad (6.6.1.3)$$

με τη διαφορά ότι κάθε κατανομή συμμετέχει με το δικό της συντελεστή βάρους ($P(c_1)$) για την πρώτη κατανομή και $P(c_1)$ για τη δεύτερη κατανομή, με $P(c_1), P(c_2) \in [0,1]$.

Ας θεωρήσουμε τη γενική περίπτωση όπου διαθέτουμε για το σύστημα αναγνώρισης δακτυλικών κινήσεων k διανύσματα παρατήρησης και θέλουμε να τα μοντελοποιήσουμε με τη χρήση ενός μείγματος L Γκαουσιανών κατανομών, με $k, L \in \mathbb{N}$. Στην περίπτωση αυτή η συνάρτηση πιθανοφάνειας θα δίνεται από τη σχέση :

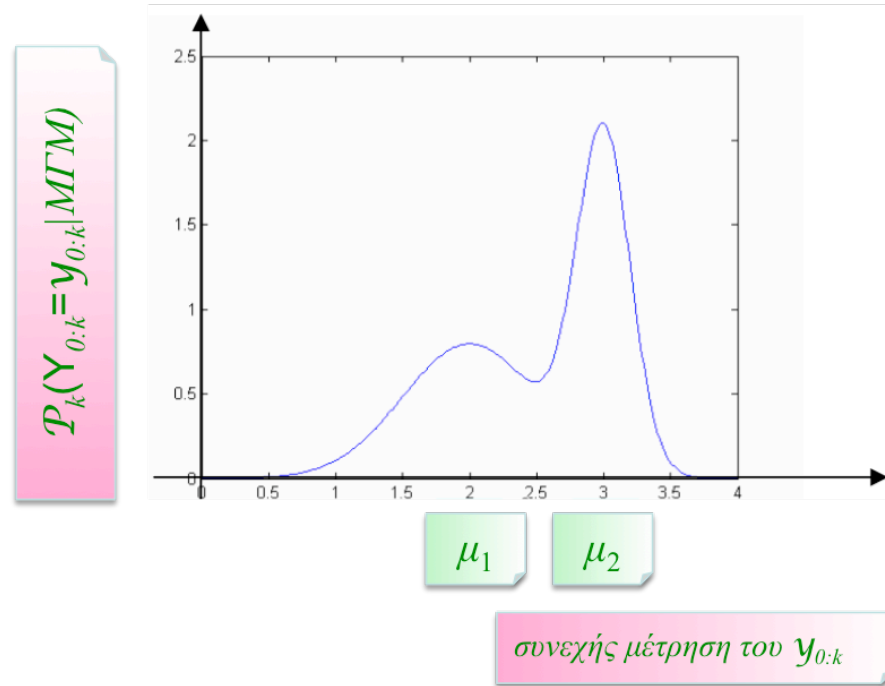
$$\begin{aligned} \mathcal{P}(Y_{0:k} = y_{0:k} | \text{MGM}) &= \sum_l \mathcal{P}(Y_{0:k} = y_{0:k} | c_l) P(c_l) = \\ &= \sum_{l=1}^L P(c_l) \frac{1}{\sigma_l \sqrt{2\pi}} \exp\left(-\frac{(y_{0:k} - \mu_l)^2}{2\sigma_l^2}\right), \text{ με } \sum_{l=1}^L P(c_l) = 1, k, L \in \mathbb{N}. \end{aligned} \quad (6.6.1.4)$$

με τη διαφορά ότι κάθε μια από τις L κατανομές συμμετέχει με το δικό της συντελεστή βάρους $P(c_l)$.

Τα ερωτήματα που τίθενται δικαίως στο σημείο αυτό είναι α) ποιο είναι το καλύτερο μοντέλο και β) πως καθορίζονται οι παράμετροι του καλύτερου μοντέλου;

6.6.2 Εύρεση του Μεικτού Γκαουσιανού Μοντέλου με τη μέγιστη πιθανοφάνεια

Το καλύτερο MGM είναι αυτό για το οποίο τα διανύσματα παρατήρησης παρουσιάζουν τη μέγιστη πιθανοφάνεια. Θα χρησιμοποιηθεί για άλλη μια φορά το κριτήριο μέγιστης πιθανοφάνειας. Το MGM μοντελοποιεί την πηγή που παρήγαγε τις παρατηρήσεις. Στην προκειμένη περίπτωση τις εξόδους της κάμερας που χρησιμοποιήθηκε.



Σχήμα 6.6.1.4 Προσέγγιση συνάρτησης πιθανοφάνειας $\mathcal{L}_k$ για συνεχή μέτρηση του $y_{0:k}$, με ΜΓΜ.

Αρχικά θεωρούμε ότι υπάρχουν L Γκαουσιανές πηγές. Η γέννηση των διανυσμάτων παρατήρησης $Y_{0:k}$ διεξάγεται σε δύο στάδια :

1. επιλέγεται τυχαία μια από τις συνιστώσες του μείγματος, λαμβάνοντας υπόψη το συντελεστή βάρους $P(c_l)$,
2. γεννιούνται τα διανύσματα παρατήρησης $Y_{0:k}$ χρησιμοποιώντας την αντίστοιχη κατανομή πιθανότητας $P(Y_{0:k} = y_{0:k} | c_l)$.

Αντιστρόφως, αν υπάρχει η ακολουθία των διανυσμάτων παρατήρησης $Y_{0:k}$, τότε μπορεί να υπολογιστεί η πηγή που παρουσιάζει τη μεγαλύτερη πιθανότητα να γέννησε την $Y_{0:k}$. Δηλαδή, με βάση την *a posteriori* πιθανότητα για κάθε πηγή παίρνουμε ότι :

$$\begin{aligned}
C &= \max_{c_l} P(c_l | Y_{0:k} = y_{0:k}) = \max_{c_l} \frac{P(Y_{0:k} = y_{0:k} | c_l) P(c_l)}{\sum_{l=1}^L P(Y_{0:k} = y_{0:k} | c_l) P(c_l)} = \\
&= \max_{c_l} \frac{P(Y_{0:k} = y_{0:k} | c_l) P(c_l)}{\sum_{l=1}^L P(Y_{0:k} = y_{0:k} | c_l) P(c_l)} \quad (6.6.2.1)
\end{aligned}$$

Το κριτήριο λοιπόν για τη μεγιστοποίηση της πιθανοφάνειας των k μετρήσεων της ακολουθίας των διανυσμάτων παρατήρησης $Y_{0:k} = y_{0:k}$ είναι η μεγιστοποίηση του :

$$\mathcal{L}_k(\text{MGM}) = \prod_{t=0}^k P(Y_{0:k} = y_{0:k}) \quad (6.6.2.2)$$

ή αλλιώς του :

$$\mathcal{L}_k(\text{MGM}) = \sum_{t=0}^k \log(P(Y_{0:k} = y_{0:k})) \quad (6.6.2.3)$$

6.6.3 Παράμετροι Μεικτού Γκαουσιανού Μοντέλου με τη μέγιστη πιθανοφάνεια

Οι παράμετροι του MGM που αναζητούνται προσδιορίζονται στη μέση τιμή μ_l , τη διασπορά σ_l και το βάρος $P(c_l)$, για κάθε $l \in [1, L]$. Όπως εύκολα διαπιστώνει κανείς, θα πρέπει να υπολογιστούν $3L$ παράμετροι.

Η διαδικασία που ακολουθείται για τον υπολογισμό των παραμέτρων MGM διακρίνεται στην περίπτωση των:

1. **ολοκληρωμένων δεδομένων**, δηλαδή την περίπτωση που για κάθε μέτρηση $y_{0:k}$ της ακολουθίας διανυσμάτων παρατήρησης $Y_{0:k}$, είναι γνωστή την Γκαουσιανή πηγή c_l που την παρήγαγε. Στα πλαίσια της παρούσας έρευνας δε είναι γνωστό εκ των προτέρων ποια Γκαουσιανή πηγή παράγει ποια ακολουθία διανυσμάτων παρατήρησης,

2. ελλιπών δεδομένων, δηλαδή τη γενική περίπτωση που δεν είναι γνωστό ποια Γκαουσιανή πηγή γέννησε κάθε μια από τις ακολουθίες διανυσμάτων παρατήρησης, όπως και στην παρούσα έρευνα.

Στην περίπτωση μη ολοκληρωμένων δεδομένων, που είναι και η περίπτωση του συστήματος αναγνώρισης δακτυλικών κινήσεων, είναι αναγκαίο να οριστεί το σύνολο των παραμέτρων που αναζητούνται :

$$\Theta = \left\{ P(c_l), \mu_l, \sigma_l \right\}_{l \in [1, L]} \quad (6.6.3.1)$$

Όπου:

- Θ , το σύνολο των παραμέτρων του ΜΓΜ,
- c_l , μια τυχαία Γκαουσιανή πηγή, με $l \in [1, L]$,
- $P(c_l)$, το βάρος,
- μ_l , η μέση τιμή της,
- σ_l , η διασπορά.

Θεωρούμε επίσης την ακολουθία των διανυσμάτων παρατήρησης $Y_{0:k}$ καθώς επίσης και την ακολουθία των ελλιπών δεδομένων $Y'_{0:k}$. Προφανώς όλες οι μετρήσεις $y_{0:k}$ της ακολουθίας των διανυσμάτων αποτελούν ελλιπή δεδομένα.

Αν τα ελλιπή δεδομένα ήταν γνωστά, τότε το κριτήριο της πιθανοφάνειας θα γραφόταν :

$$\mathcal{L}_k(\text{ΜΓΜ}) = \sum_{l=0}^k \log \left(P(Y_{0:k} = y_{0:k} | Y'_{0:k} = y'_{0:k}) P(c_{Y_l}) \right) \quad (6.6.3.2)$$

Ο υπολογισμός των παραμέτρων του ΜΓΜ με τη μέγιστη πιθανοφάνεια γίνεται με τη βοήθεια ενός αλγορίθμου Εκτίμησης – Μεγιστοποίησης (EM) [Dempster 1977]. Η μέθοδος στοχεύει στη μεγιστοποίηση της πιθανοφάνειας των διανυσμάτων παρατηρήσεων σε σχέση με τις παραμέτρους $\Theta = \{P(c_l), \mu_l, \sigma_l\}, l \in [1, L]$, του μοντέλου. Με άλλο λόγια, οι παραπάνω παράμετροι οφείλουν να «αντιστοιχούν» όσο το δυνατόν καλύτερα στα διανύσματα παρατήρησης. Η διαδικασία ξεκινάει θέτοντας ένα υποτιθέμενο γνωστό μοντέλο Θ^{mg} , ως

μοντέλο έναρξης. Έτσι, χρησιμοποιήθηκαν οι παρακάτω «επανεκτιμητές»:

$$\bar{P}(c_l) = \frac{1}{k} \sum_{t=0}^k y_{0:t} P(c_l | y_{0:t}, \Theta^{mg}) \quad (6.6.3.3)$$

$$\bar{\mu}_l = \frac{\sum_{t=0}^k y_{0:t} P(c_l | y_{0:t}, \Theta^{mg})}{\sum_{t=0}^k P(c_l | y_{0:t}, \Theta^{mg})} \quad (6.6.3.4)$$

$$\Sigma_l = \frac{\sum_{t=0}^k P(c_l | y_{0:t}, \Theta^{mg}) (y_{0:t} - \bar{\mu}_l)(y_{0:t} - \bar{\mu}_l)^k}{\sum_{t=0}^k P(c_l | y_{0:t}, \Theta^{mg})} \quad (6.6.3.5)$$

Συνεπώς, μια πολυμεταβλητή Γκαουσιανή κατανομή ενός ΜΓΜ θα δίνεται από τη σχέση:

$$\mathcal{P}(Y_{0:k} = y_{0:k} | c_l) = \frac{1}{(2\pi)^{\frac{d}{2}} \det(\Sigma_l)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (y_{0:t} - \bar{\mu}_l)^k \Sigma_l^{-1} (y_{0:t} - \bar{\mu}_l)\right) \quad (6.6.3.6)$$

Όπου:

- $\mathcal{P}(Y_{0:k} = y_{0:k} | c_l)$, η πολυμεταβλητή Γκαουσιανή κατανομή,
- $\bar{\mu}_l$, το διάνυσμα των μέσων τιμών ή αλλιώς κέντρο της συνισταμένης,
- Σ_l , ο πίνακας συμμεταβλητότητας,
- d , η διάσταση του μοντέλου.

6.6.4 Αλγόριθμος « forward » του Baum

Στόχος του αλγορίθμου είναι ο αποτελεσματικός προσδιορισμός της κατανομής πιθανότητας της ακολουθίας διανυσμάτων παρατήρησης $Y_{0:k}$ σε συνάρτηση με τις παραμέτρους του μοντέλου $M = (q, Q, \psi)$, δηλαδή ο προσδιορισμός της συνάρτησης πιθανοφάνειας.

Εισάγουμε την κατανομή των $Y_{0:k}$ και X_k :

$$a_x^k(y_{0:k})d y_{0:k} \triangleq \mathcal{P}(Y_0 \in d y_{0:k}, X_k = x) \quad (6.6.4.1)$$

$\forall k \in \mathbb{N}, y_1, \dots, y_k \in \mathcal{Y}, x \in \mathcal{X}$. Ομοίως εισάγουμε και τη μεταβλητή “forward”:

$$p_k(x) \triangleq a_x^k(Y_{0:k}) \quad (6.6.4.2)$$

$\forall k \in \mathbb{N}, x \in \mathcal{X}$.

Αλγόριθμος “forward” του Baum Η ακολουθία $\{p_k\}, k \in \mathbb{N}$ των μεταβλητών “forward” αποτελεί λύση στην παρακάτω επαναληπτική εξίσωση:

$$p_{k+1}(x) \leftarrow \psi(x, Y_{k+1}) \sum_{x' \in \mathcal{X}} \mathcal{Q}(x, x') p_k(x'), n \geq 0, x \in \mathcal{X} \quad (6.6.4.3)$$

$$p_0(x) \leftarrow \rho(x) \psi(x, Y_0), x \in \mathcal{X} \quad (6.6.4.4)$$

Θέτουμε $\psi(y) \triangleq \text{diag}[(\psi(1, y), \dots, \psi(N, y))]$, οπότε:

$$p_{k+1} \leftarrow p_k \mathcal{Q} \psi(Y_{k+1}), k \geq 0 \quad (6.6.4.5)$$

$$p_0 \leftarrow \rho \psi(Y_0) \quad (6.6.4.6)$$

Η κατανομή των παρατηρήσεων $Y_{0:k}$ εκπέμπει την ακόλουθη πυκνότητα:

$$\mathcal{P}(Y_0 \in d y_{0:k}) = \left[\sum_{x \in \mathcal{X}} a_x^k(Y_{0:k}) \right] d y_{0:k} \quad (6.6.4.7)$$

Ο παραπάνω τύπος αποτελεί τη λύση του προβλήματος Π1 και κατά συνέπεια η συνάρτηση πιθανοφάνειας που συνδέεται με το μοντέλο M δίνεται από τη σχέση:

$$\mathcal{L}_k(M) \triangleq \sum_{x \in \mathcal{X}} p_k(x) \quad (6.6.4.8)$$

6.6.5 Ο αλγόριθμος του Viterbi

Λύση στο πρόβλημα Π2 δίνεται μέσω ενός αλγορίθμου του δυναμικού προγραμματισμού, τον αλγόριθμο του Viterbi [Rabiner 1989]. Ο αλγόριθμος αυτός επιτρέπει τη μεγιστοποίηση της δεσμευμένης κατανομής πιθανότητας της ακολουθίας καταστάσεων $X_{0:k}$. Βασίζεται στη μέθοδο εκτίμησης του *a posteriori* μεγίστου (Maximum *a posteriori* – MAP), η οποία χρησιμοποιείται προκειμένου να εκτιμηθεί ένα πλήθος άγνωστων παραμέτρων, όπως για παράδειγμα οι παράμετροι μιας πυκνότητας πιθανότητας, οι οποίες παράμετροι προφανώς συνδέονται με κάποιο δοσμένο δείγμα (ήχος, βίντεο κλπ) [Gauvain 1994].

Αντίθετα, η μεταβλητή « forward » του Baum, η οποία αναλύθηκε στο εδάφιο 6.6.4, επιτρέπει απλώς τον υπολογισμό της δεσμευμένης κατανομής μιας κατάστασης για δεδομένη χρονική στιγμή, γνωρίζοντας ήδη τις παρατηρήσεις.

Σύμφωνα με τον τύπο του Bayes, η τροχιά που μεγιστοποιεί τη κατανομή της δεσμευμένης πιθανότητας του $X_{0:k}$ γνωρίζοντας τις παρατηρήσεις $Y_{0:k}$ (A), μεγιστοποιεί ταυτόχρονα την από κοινού κατανομή των $X_{0:k}$ και $Y_{0:k}$ (B), δηλαδή :

$$\begin{aligned}
A &= \mathcal{P}(X_{0:k} = x_{0:k} | Y_{0:k} = y_{0:k}) \mathcal{P}(Y_{0:k} \in d y_{0:k}), \\
B &= \mathcal{P}(X_{0:k} = x_{0:k}, Y_{0:k} \in d y_{0:k}), \\
\Gamma &= \rho(x_0) \psi(x_0, y_0) \prod_{t=1}^k [Q(x_{t-1}, x_t) \psi(x_t, y_t)] d y_{0:k}, \\
\text{με } A &= B = \Gamma.
\end{aligned} \tag{6.6.5.1}$$

Κατά συνέπεια, η μεγιστοποίηση της (A) στο $x_{0:k}$ ισοδυναμεί με τη μεγιστοποίηση της (Γ). Για το λόγο αυτό θέτουμε τη μεταβλητή :

$$w_k(x) = \max_{x_{0:k-1}} \rho(x_0) \psi(x_0, y_0) \prod_{t=1}^k [Q(x_{t-1}, x_t) \psi(x_t, y_t)], k \geq 0, x \in \mathcal{X}. \tag{6.6.5.2}$$

Θεώρημα 7.4.1 (Δυναμικός προγραμματισμός) Η ακολουθία $\{w_k\}$, $k \in \mathbb{N}$, είναι η λύση της παρακάτω επαναληπτικής εξίσωσης :

$$\begin{aligned}
w_{k+1}(x) &= \psi(x, Y_{k+1}) \left[\max_{x' \in \mathcal{X}} Q(x, x') w_k(x') \right], k \geq 0, x \in \mathcal{X}, \\
w_0(x) &= \rho(x) \psi(x, Y_0), x \in \mathcal{X}.
\end{aligned} \tag{6.6.5.3}$$

Υποθέτοντας ότι έχουμε επίτευξη μεγίστου σε μοναδικό σημείο, θέτουμε :

$$I_k(x') \triangleq \underset{x \in \mathcal{X}}{\text{Arg max}} [Q(x, x') w_k(x)], k \geq 0, x' \in \mathcal{X}, \tag{6.6.5.4}$$

Αλγόριθμος του Viterbi Το ιδανικό μονοπάτι καταλήγει στην κατάσταση :

$$X_k^{MAP} = \underset{x \in \mathcal{X}}{\text{Arg max}} w_k(x), \tag{6.6.5.5}$$

υποθέτοντας ότι έχουμε επίτευξη μοναδικού μεγίστου τη χρονική στιγμή k .

Αντίστοιχα, στη χρονική στιγμή $k+1$ το ιδανικό μονοπάτι καταλήγει στην κατάσταση :

$$X_k^{MAP} = I_k(X_{k+1}^{MAP}). \quad (6.6.5.6)$$

Στο σχήμα 6.6.5.1 διακρίνουμε τον τρόπο επιλογής του πιθανότερου κόμβου, διαμορφώνοντας το μονοπάτι εσωτερικά ενός γράφου Treillis. Στον κατακόρυφο άξονα διακρίνουμε ένα κόμβο για κάθε κατάσταση της X_k . Στον οριζόντιο άξονα διακρίνουμε την ακολουθία των διανυσμάτων παρατήρησης $Y_{0:k}$. Υπολογίζουμε την πιθανότητα, που προκύπτει από την παράσταση (Γ) της σχέσης 6.6.5.1, για κάθε κατάσταση της X_k , για τη χρονική στιγμή k . Ας υποθέσουμε ότι μεγιστοποιεί την ακολουθία $X_{0:k}$ στο $x_{0:k}$, για X_k^{MAP} , το οποίο αντιστοιχεί στην κατάσταση $x' \in X$. Για να εντοπιστεί η κατάσταση $x \in X$, πιθανότερη διάδοχος της $x' \in X$, για τη χρονική στιγμή $k+1$, θα πρέπει προηγουμένως να βρεθεί το $x_{0:k+1}$ στο οποίο παρουσιάζεται μέγιστο. Δηλαδή, υπολογίζουμε το μέγιστο, μεταξύ των πιθανοτήτων που προκύπτουν από την παράσταση (Γ) της σχέσης 6.6.5.1, για κάθε κατάσταση της X_k . Συνεχίζοντας κατά αυτόν τον τρόπο, και αφού υπολογίσουμε την πιθανότερη κατάσταση για την τελική χρονική στιγμή, ανακτούμε σταδιακά την πιθανότερη ακολουθία καταστάσεων $X_{0:k}$, που γέννησε την ακολουθία των διανυσμάτων παρατήρησης $Y_{0:k}$, μέσω της σχέσης 6.6.5.6.

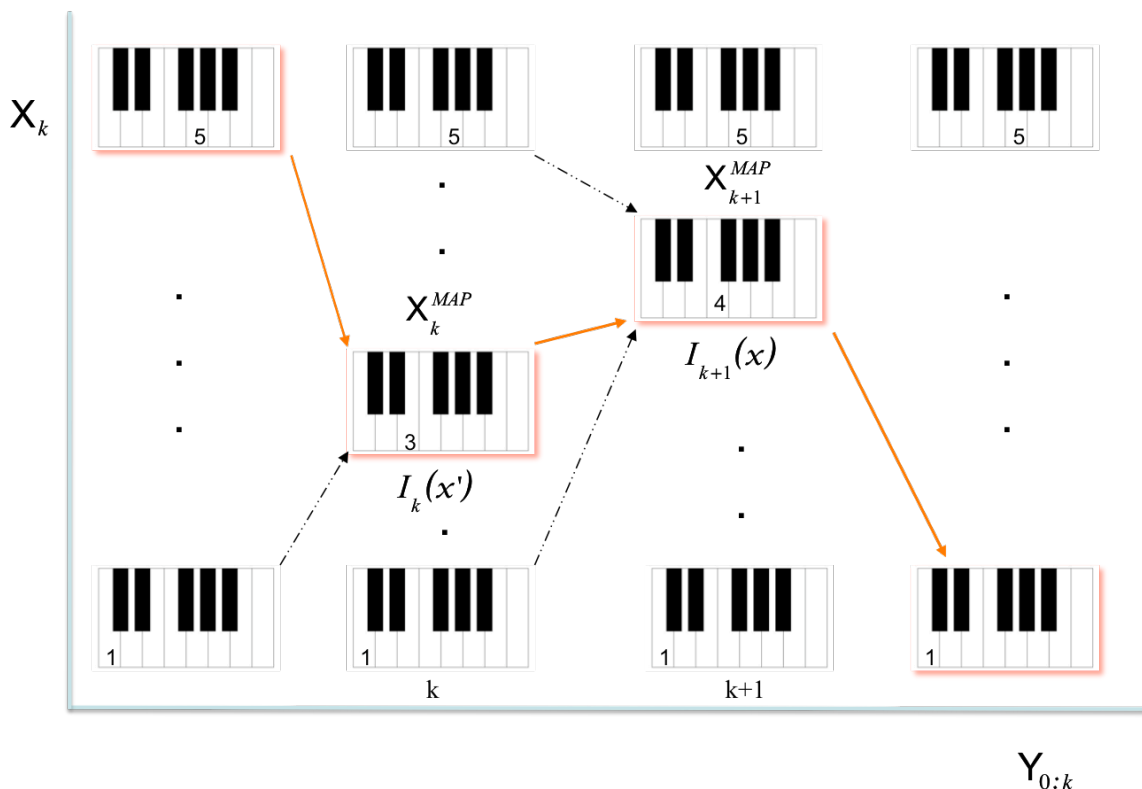
6.6.6 Επανεκτίμηση παραμέτρων Κρυφών Μαρκοβιανών Μοντέλων και Μεικτών Γκαουσιανών Μοντέλων κατά Baum-Welch

Η εκπαίδευση του συστήματος αναγνώρισης δακτυλικών κινήσεων περνάει μέσα από την εφαρμογή της μεθόδου των Baum-Welch, η οποία ανήκει στην κατηγορία αλγορίθμων Εκτίμησης – Μεγιστοποίησης (EM) [Dempster 1977]. Η μέθοδος στοχεύει στη μεγιστοποίηση της πιθανοφάνειας των παρατηρήσεων σε σχέση με τις παραμέτρους $M = (\rho, Q, \psi)$ του και αποτελεί λύση στο πρόβλημα της εκπαίδευσης Π3. Με άλλο λόγια, οι παραπάνω παράμετροι οφείλουν να «αντιστοιχούν» όσο το δυνατόν καλύτερα στα διανύσματα παρατήρησης.

Δεν υπάρχει γενική μέθοδος επίλυσης αυτού του προβλήματος. Το σύστημα προσπαθεί να καθορίσει το M που μεγιστοποιεί τοπικά τη συνάρτηση πιθανοφάνειας:

$$M \rightarrow \mathcal{L}_k(M) \triangleq \mathcal{P}_M(Y_{0:k} = y_{0:k}) \quad (6.6.6.1)$$

όπου $\mathcal{P}$ είναι η πιθανότητα για την οποία το μοντέλο M βρίσκεται εν ισχύ.



Σχήμα 6.6.5.1 Επιλογή του μονοπατιού εσωτερικά του γράφου Treillis μεταξύ των στιγμών k και $k+1$.

Θεωρούμε τις πιθανότητες μετάβασης $x \rightarrow x'$ γνωρίζοντας το μοντέλο M και τα διανύσματα παρατήρησης $Y_{0:k} = y_{0:k}$:

$$\xi'_{x,x'}(y_{0:k}) = \frac{\mathcal{P}_M(X_t = x, X_{t+1} = x', Y_{0:k} = y_{0:k})}{\mathcal{P}_M(Y_{0:k} = y_{0:k})} \quad (6.6.6.2)$$

Δανειζόμαστε την ποσότητα $b_x^t(y_{t+1:k})$ από τον αλγόριθμο « backward » του Baum, για την οποία ισχύει :

$$b_x^t(y_{t+1:k})d y_{t+1:k} \triangleq \mathcal{P}(Y_{t+1:k} \in d y_{t+1:k}, X_t = x), \text{ με } t < k. \quad (6.6.6.3)$$

Δανειζόμαστε επίσης την ποσότητα $a_x^k(y_{0:k})$ από τον αλγόριθμο « forward » του Baum (σχέση 6.6.4.1)

Θέτουμε:

$$\gamma_x^t(y_{0:k}, l) \triangleq \sum_{x' \in \mathcal{X}} \xi_{x, x'}^t(y_{0:k}) \frac{c_{x'l} \mathcal{N}(y_{0:k}, \mu_{x'l}, \Sigma_{x'l})}{\sum_{l=1}^L c_{x'l} \mathcal{N}(y_{0:k}, \mu_{x'l}, \Sigma_{x'l})} \quad (6.6.6.4)$$

Με βάση τις σχέσεις 6.6.6.2, 6.6.6.3 και 6.6.6.4 προκύπτει ότι :

$$\gamma_x^t(y_{0:k}, l) = \frac{a_x^t(y_{0:t})b_x^t(y_{t+1:k})}{\sum_{x' \in \mathcal{X}} a_{x'}^t(y_{0:t})b_{x'}^t(y_{t+1:k})} \frac{P(c_{x'l}) \mathcal{N}(y_{0:k}, \mu_{x'l}, \Sigma_{x'l})}{\sum_{l=1}^L P(c_{x'l}) \mathcal{N}(y_{0:k}, \mu_{x'l}, \Sigma_{x'l})}, \quad (6.6.6.5)$$

που είναι η πιθανότητα το KMM να βρίσκεται στην κατάσταση x τη χρονική στιγμή t και η παρατήρηση $y_{0:k}$ να δημιουργείται με συνεισφορά της l κατανομής του MGM.

Έχοντας ως σημείο εκκίνησης τις παραμέτρους $M = (\rho, Q, \psi)$ του KMM, σε κάθε επανάληψη της διαδικασίας εκπαίδευσης επανεκτιμώνται α) οι παράμετροι κάθε Γκαουσιανής κατανομής και οι συντελεστές του MGM με βάση τις σχέσεις 6.6.3.3 έως 6.6.3.5, αλλά και β) οι παράμετροι του KMM, όπως φαίνεται στις παρακάτω σχέσεις:

$$\bar{P}_x(c_l) = \frac{\sum_{t=0}^k \gamma_x^t(y_{0:k}, l)}{\sum_{t=0}^k \sum_{l=1}^L \gamma_x^t(y_{0:k}, l)}, \quad (6.6.6.6)$$

$$\bar{\mu}_{x,l} = \frac{\sum_{t=0}^k \gamma_x^t(y_{0:k}, l) y_{0:t}}{\sum_{t=0}^k \gamma_x^t(y_{0:k}, l)}, \quad (6.6.6.7)$$

$$\Sigma_{x,l} = \frac{\sum_{t=0}^k \gamma_x^t(y_{0:k}, l) y_{0:t} (y_{0:t} - \bar{\mu}_l)(y_{0:t} - \bar{\mu}_l)^k}{\sum_{t=0}^k \gamma_x^t(y_{0:k}, l)}. \quad (6.6.6.8)$$

Για το KMM χρησιμοποιήθηκαν οι παρακάτω «επανεκτιμητές»:

$$\bar{\rho}(x) = \gamma_x^0(y_{0:k}, l), \quad (6.6.6.9)$$

$$\bar{Q}(x, x') = \frac{\sum_{t=0}^{k-1} \xi_{x,x'}^t(y_{0:k}, l)}{\sum_{t=0}^{k-1} \gamma_x^t(y_{0:k}, l)}, \quad (6.6.6.10)$$

$$\bar{\psi}(x, y) = \frac{\sum_{t=0}^k 1_{y_k=y} \gamma_x^t(y_{0:k}, l)}{\sum_{t=0}^k \gamma_x^t(y_{0:k}, l)}. \quad (6.6.6.11)$$

Με άλλα λόγια ισχύει:

$$\bar{\rho}(x) = A \quad (6.6.6.12)$$

$$\bar{Q}(x, x') = \frac{B}{C} \quad (6.6.6.13)$$

$$\bar{\psi}(x, y) = \frac{D}{C} \quad (6.6.6.14)$$

Όπου:

- A , ο εκτιμητής του πλήθους των καταστάσεων x τη χρονική στιγμή $t=0$,
- B , ο εκτιμητής του πλήθους των μεταβάσεων $x \rightarrow x'$
- C , ο εκτιμητής του πλήθους των μεταβάσεων που ξεκίνησαν από το x ,
- D , ο εκτιμητής του πλήθους των μεταβάσεων που ξεκίνησαν από το x γεννώντας την y .

Με δεδομένες τις παραμέτρους $M=(\rho, \bar{Q}, \bar{\psi})$ και χρησιμοποιώντας σχέσεις 6.6.6.9 έως 6.6.6.11, μπορούν να επανεκτιμηθούν οι παράμετροι του μοντέλου αποκτώντας $\bar{M}=(\bar{\rho}, \bar{Q}, \bar{\psi})$. Οι Baum-Welch απέδειξαν ότι :

1. Είτε το αρχικό μοντέλο M είναι ένα κρίσιμο σημείο της συνάρτησης πιθανοφάνειας :

$$\nabla_M \mathcal{L}_k(M) = \nabla_M \mathcal{P}_M(Y_{0:k} = y_{0:k}) = 0 \quad (6.6.6.15)$$

και άρα οι αρχικές παράμετροι του μοντέλου είναι ίσες με τις επανεκτιμημένες.

2. Είτε το μοντέλο $\bar{M}$ είναι αυστηρά πιο πιθανοφανές υπό την έννοια ότι :

$$\mathcal{L}_k(\bar{M}) = \mathcal{P}_{\bar{M}}(Y_{0:k} = y_{0:k}) > \mathcal{L}_k(M) = \mathcal{P}_M(Y_{0:k} = y_{0:k}). \quad (6.6.6.16)$$

Η επανάληψη της παραπάνω διαδικασίας αποδίδεται από έναν αλγόριθμο EM που επιτρέπει να καθοριστεί το τοπικό μέγιστο της συνάρτησης πιθανοφάνειας.

6.7 Παράδειγμα αναγνώρισης μεμονωμένων δακτυλικών κινήσεων

Στο παράδειγμα αυτό παρουσιάζεται η χρήση του συστήματος ως αναγνωριστής μεμονωμένων δακτυλικών κινήσεων.

Θεωρούνται ως δεδομένα:

A) Το λεξικό 6 κινήσεων που παρουσιάστηκε στην ενότητα 6.2:

$$\Lambda K = \left\{ \kappa \nu_i \right\}_{i \in [1,6]} \quad (6.7.1)$$

Όπου:

- ΛK , το λεξικό κινήσεων,
- $\kappa \nu_i$, η εκάστοτε κίνηση.

B) Τα 6 μοντέλα που αντιστοιχούν ένα προς ένα στις κινήσεις του ΛK :

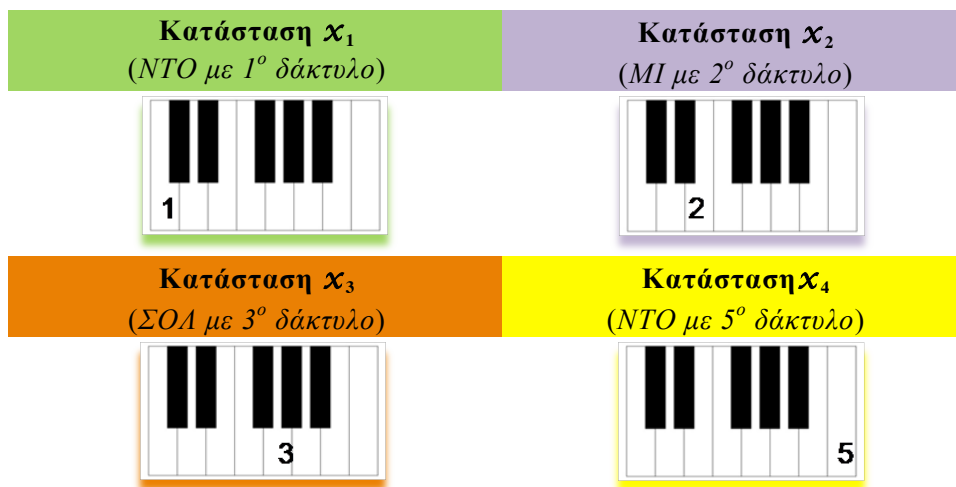
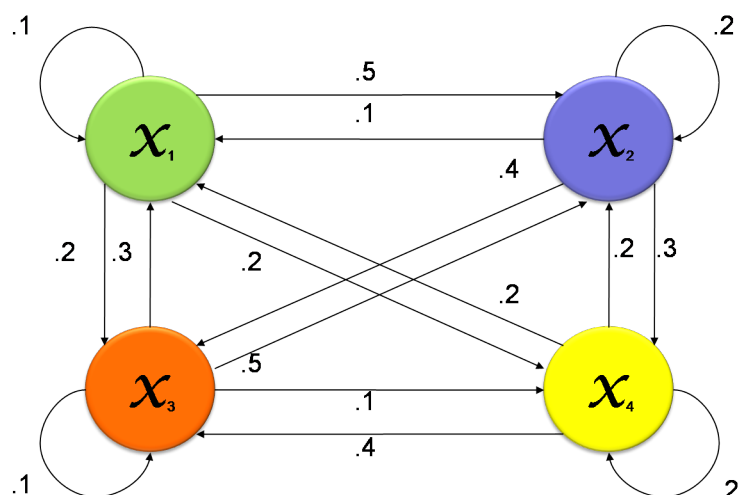
$$MK = \left\{ M_i \right\}_{i \in [1,6]} \quad (6.7.2)$$

Όπου:

- MK , το σύνολο των μοντέλων κινήσεων,
- M_i , οι παράμετροι του εκάστοτε μοντέλου.

Για κάθε μια από τις $\kappa \nu_i$ διατίθεται ένα σύνολο από βίντεο εκπαίδευσης, τα διανύσματα παρατήρησης των οποίων έχουν ιδιότητες που αναπαριστούν πλήρως την κάθε κίνηση.

Ας θεωρηθεί ότι το μοντέλο M_3 , κατ' εξαίρεση για τις ανάγκες του συγκεκριμένου παραδείγματος, είναι εργοδικό (επιτρέπονται μεταβάσεις σε προηγούμενες καταστάσεις). Μπορεί, κατά συνέπεια, να μοντελοποιήσει μια πλήρη κίνηση αρπές στην κλίμακα του ΝΤΟ. Δηλαδή αντιπροσωπεύει τις κινήσεις $\kappa \nu_3$ και $\kappa \nu_4$ με $N=4$ καταστάσεις και αλφαριθμητικό δακτυλισμών X_3 (Σχήμα 6.7.1). Ας θεωρηθεί επίσης, α) ότι το Δt (χρονικό περιθώριο μεταξύ το πάτημα δύο νοτών) καθορίζεται από την παρτιτούρα και β) ότι διατίθεται ο πίνακας μεταβάσεων $Q=[Q(x,x')]$ (Σχήμα 6.7.2).

Αλφαβητάρι δακτυλισμών $\mathcal{X}_3$ Σχήμα 6.7.1 Αλφαβητάρι δακτυλισμών $\mathcal{X}_3$ Σχήμα 6.7.2 Πίνακας μεταβάσεων $Q=[Q(x,x')]$ με όλες τις δυνατές μεταβάσεις μεταξύ των καταστάσεων

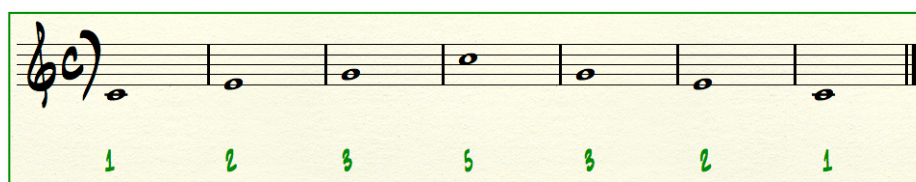
Είναι γνωστό ότι την αρχική χρονική στιγμή ($t=0$) πατήθηκε από τον πιανίστα NTO με 1^ο δάκτυλο, άρα $p(\mathcal{X}_1) = 1$. Έστω επίσης $Y_{0:6}$, η ακολουθία διανυσμάτων παρατήρησης που αφορά σε μια πλήρη κίνηση αρπές και προέκυψε από την ακολουθία των καρέ

παρατήρησης $K_{0:6}$, από $t=0$ μέχρι $t=6$ (συνολικά διάρκεια $T=7$ χρονικών στιγμών). Το ερώτημα που τίθεται είναι το εξής:

Με ποια διαδικασία αναγνωρίζεται από το σύστημα μια πλήρης κίνηση αρπές (Σχήμα 6.7.3 - ανιούσα και κατιούσα κίνηση αρπές χωρίς επανάληψη της NTO με 5°);

Δηλαδή, η παρακάτω κίνηση:

NTO με 1° – MI με 2° – ΣΟΛ με 3° – NTO με 5° – ΣΟΛ με 3° – MI με 2° – NTO με 1° ;



Σχήμα 6.7.3 Νότες και δακτυλισμοί της κίνησης

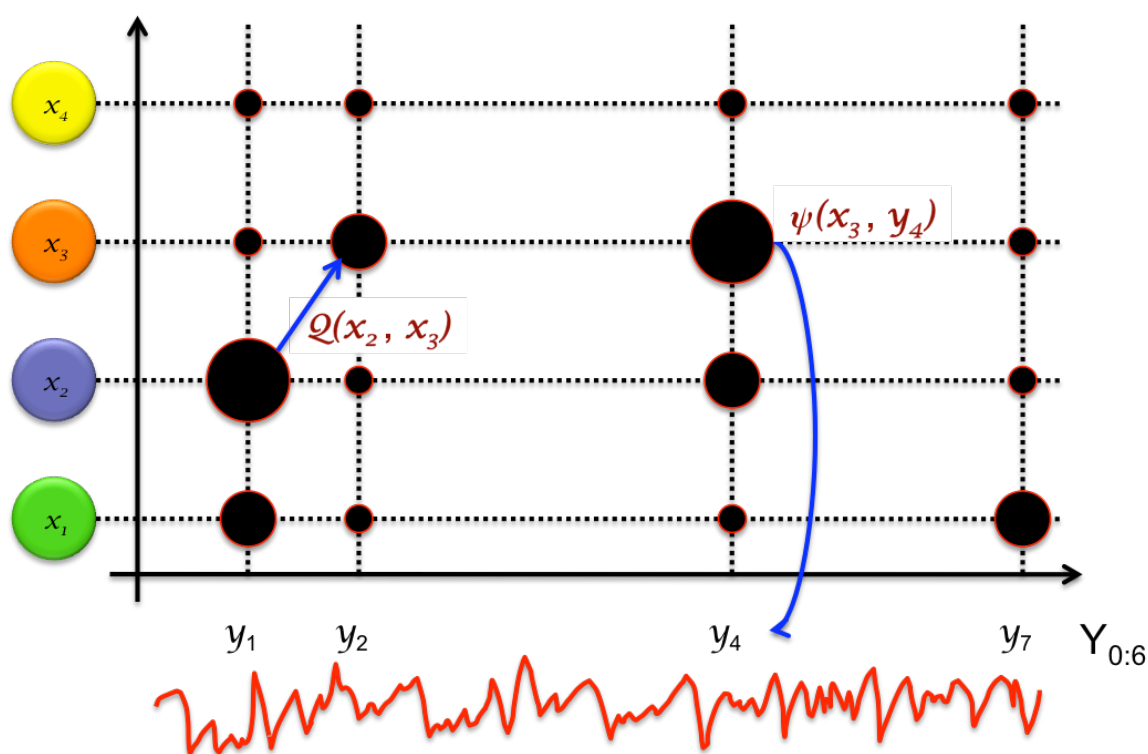
Επίλυση: Αναμένεται το μοντέλο M_3 να δώσει την $X_{0:6}$, ως την πιθανότερη ακολουθία καταστάσεων (*κρυφή ακολουθία*) που γέννησε την $Y_{0:6}$. Αναμένεται επίσης το μοντέλο M_3 να παρουσιάσει μέγιστη πιθανοφάνεια έναντι όλων των υπολοίπων μοντέλων του συστήματος. Αυτό, διότι σύμφωνα με το σενάριο του παραδείγματος α) το M_3 είναι το μοναδικό εργοδικό μοντέλο του συστήματος, μπορεί δηλαδή να συνεισφέρει στην αναγνώριση μιας πλήρους κίνησης αρπές (*ανιούσα και κατιούσα*), σε αντίθεση με το M_4 (*μόνο κατιούσα*) και β) στο παράδειγμα μελετάται το σενάριο αναγνώρισης μεμονωμένων κινήσεων. Διατίθεται λοιπόν η $Y_{0:6}$ και αναμένεται η $X_{0:6}$ από το μοντέλο M_3 :

$$Y_{0:6} = \{y_1, y_2, y_3, y_4, y_5, y_6, y_7\},$$

$$X_{0:6} = \{x_1, x_2, x_3, x_4, x_3, x_2, x_1\}. \quad (6.7.3)$$

Ο τρόπος λειτουργίας των διαδικασιών εσωτερικά ενός KMM αναπαρίσταται μέσω ενός γράφου Treillis (Σχήμα 6.7.4). Στα

αριστερά διακρίνονται όλες οι δυνατές εσωτερικές καταστάσεις του μοντέλου M_3 , ενώ στο κάτω μέρος διακρίνεται η ακολουθία των διανυσμάτων παρατήρησης (*συνεχές σήμα*). Τα μαύρα κυκλάκια (κόμβοι) αντιστοιχούν στην πιθανοφάνεια μια κατάσταση να γεννήσει μια μέτρηση της ακολουθίας των διανυσμάτων παρατήρησης. Για κάθε δυνατή μετάβαση διατίθεται επίσης η αντίστοιχη πιθανότητα μετάβασης. Σχηματικά, όσο μεγαλύτερο είναι το μαύρο κυκλάκι, ο κόμβος δηλαδή που διασταυρώνει κατάσταση με παρατήρηση, τόσο πιο πιθανοφανές είναι μια παρατήρηση να γεννηθήκε από την κατάσταση με την οποία διασταυρώνεται.



Σχήμα 6.7.4 Αναπαράσταση των διαδικασιών μέσα στο M_3 μέσω ενός γράφου Treillis

Στο σημείο αυτό τίθενται τρία ερωτήματα που αναζητούν απάντηση :

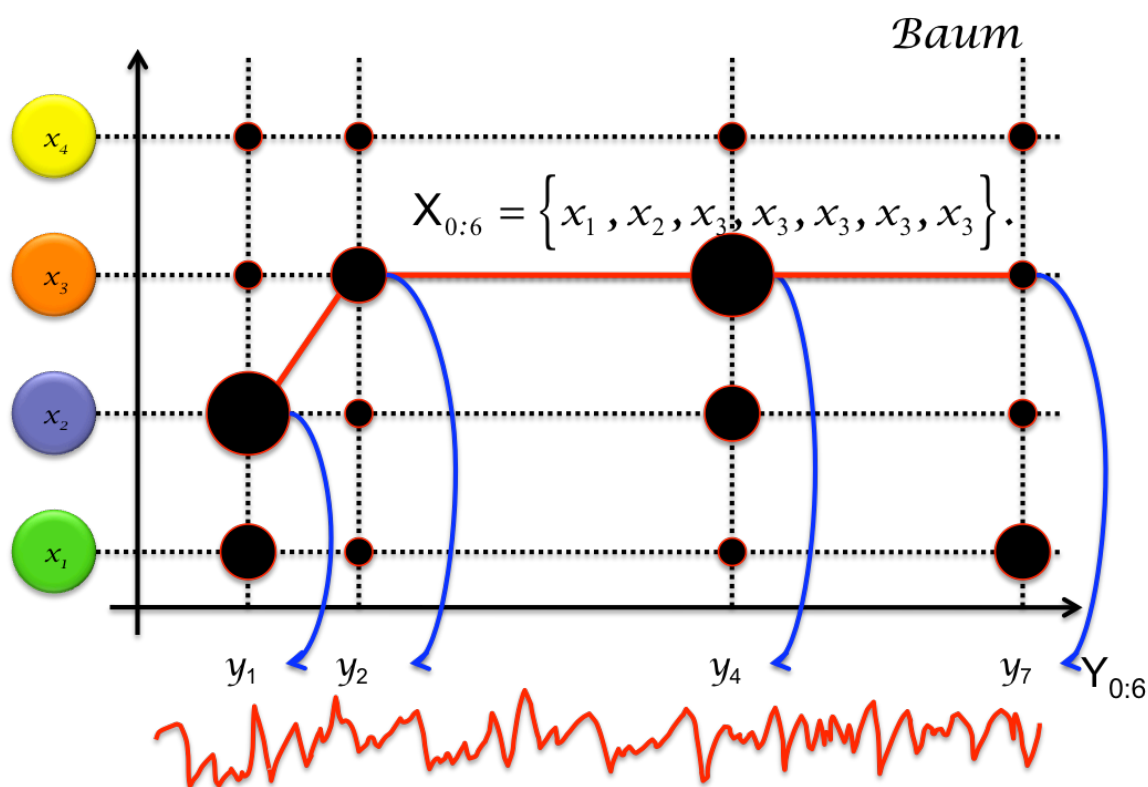
Ερώτημα 1^ο :

- ✓ Γνωρίζουμε το μοντέλο M_3 ,

- ✓ Γνωρίζουμε την ακολουθία των διανυσμάτων παρατήρησης $Y_{0:6}$,
- ? Διερωτόμαστε πως να υπολογίσουμε την $P(Y_{0:6} | M_3)$;

Απάντηση 1^η:

☞ Την απάντηση στο ερώτημα αυτό θα τη δώσει ο αλγόριθμος « forward » του Baum (Σχήμα 6.7.5), διασχίζοντας όλες τις πιθανές ακολουθίες καταστάσεων $X_{0:6}$. Ο αλγόριθμος υπολογίζει την πιθανοφάνεια το μοντέλο να βρίσκεται σε μια κατάσταση π.χ. x_3 για μια δεδομένη χρονική στιγμή $t=4$ έχοντας προηγηθεί παρατηρήσεις για διάρκεια 4 χρονικών στιγμών, ανεξάρτητα με το ποιες καταστάσεις γέννησαν αυτές τις παρατηρήσεις.



Σχήμα 6.7.5 Αλγόριθμος « forward » του Baum

Ερώτημα 2^ο :

- ✓ Γνωρίζουμε το μοντέλο M_3 ;

- ✓ Γνωρίζουμε την ακολουθία των διανυσμάτων παρατήρησης $Y_{0:6}$;
- ? Διερωτόμαστε ποια είναι η καλύτερη (πιθανότερη) $X_{0:6}$ που γέννησε την $Y_{0:6}$;

Απάντηση 1^η:

☞ Την απάντηση σε αυτό ερώτημα, θα τη δώσει ο αλγόριθμος του Viterbi (Σχήμα 6.7.6). Στόχος δεν είναι να βρεθεί η τιμή της μέγιστης πιθανοφάνειας, αλλά το μονοπάτι-ακολουθία $X_{0:6}$ (μονοπάτι του Viterbi) που γέννησε την ακολουθία των διανυσμάτων παρατήρησης $Y_{0:6}$. Το σύστημα ακολουθεί παρόμοια διαδικασία με αυτήν που ακολουθήθηκε για τον υπολογισμό της $P(Y_{0:6} | M_3)$, μέσω του αλγορίθμου « forward ». Υπολογίζεται δηλαδή η πιθανοφάνεια γέννησης της $Y_{0:6}$ μέσω όλων των πιθανών μονοπατιών και επιλέγεται εκείνο που έχει τη μέγιστη πιθανοφάνεια. Είναι γεγονός ότι ο αλγόριθμος του Viterbi και ο αλγόριθμος « forward » του Baum μοιάζουν αρκετά. Η διαφορά τους είναι ότι ο μεν πρώτος υπολογίζει αθροίσματα πιθανοφάνειας που συνδέονται με προγενέστερες καταστάσεις, ενώ ο δεύτερος στοχεύει στη μεγιστοποίηση της πιθανοφανειών αυτών.

Ερώτημα 3^ο :

- ? Πως μπορούν να προσαρμοστούν οι παράμετροι $M_3 = (\varrho_3, Q_3, \psi_3)$ του μοντέλου ώστε να μεγιστοποιηθεί η $P(Y_{0:6} | M_3)$;

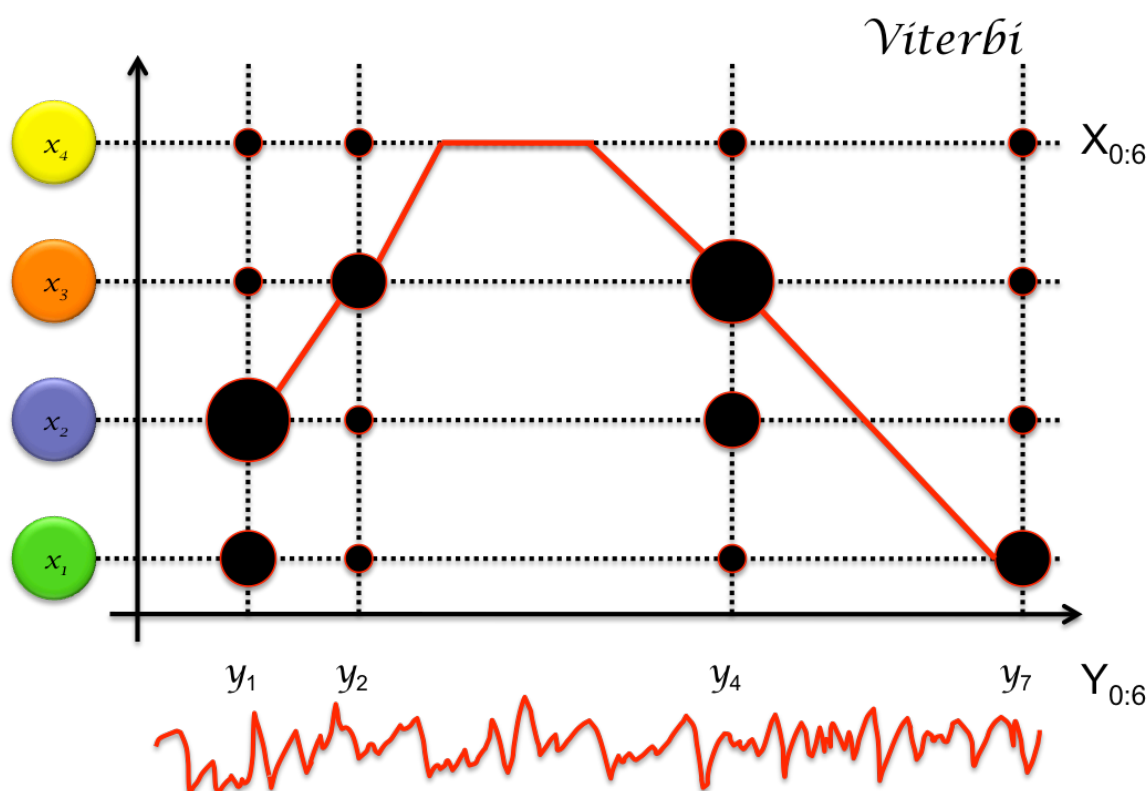
Απάντηση 3^η:

Τα βήματα που ακολουθούνται από το σύστημα είναι τα παρακάτω:

Βήμα 1: Για κάθε μοντέλο εκτιμώνται οι παράμετροί του $M_i = (\varrho_i, Q_i, \psi_i)$, ώστε να βελτιστοποιούν την πιθανοφάνεια του συνόλου των διανυσμάτων παρατήρησης ως προς την $κιν_i$. Αυτό το βήμα αντιστοιχεί στο Π3 των KMM και επιλύεται με τη βοήθεια του αλγορίθμου των Baum-Welch.

Βήμα 2: Για κάθε βίντεο που εισάγουμε στο σύστημα, την κίνηση του οποίου θέλουμε να αναγνωρίσουμε, υλοποιείται η διαδικασία του σχήματος 7-. Πρόκειται αφενός για τη διαδικασία εξαγωγής της ακολουθίας διανυσμάτων παρατήρησης $Y_{0:k}$ μέσα από την ακολουθία

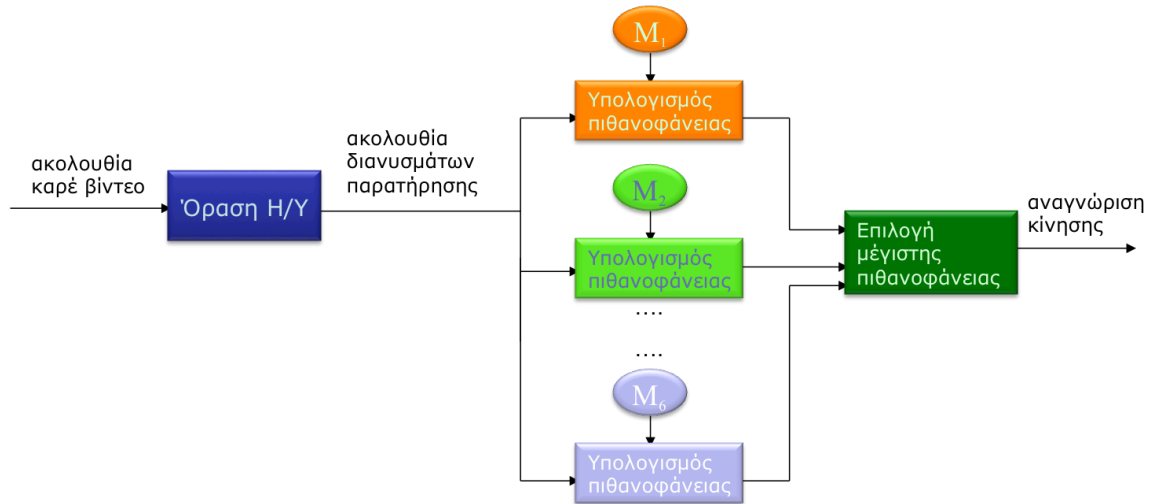
των καρτέ του βίντεο $K_{0:k}$ με χρήση της τεχνολογίας όρασης H/Y, όπως αυτή περιγράφηκε στο κεφάλαιο 6. Αφετέρου, για τη διαδικασία υπολογισμού της πιθανοφάνειας για κάθε μοντέλο M_i , ακολουθούμενη από την επιλογή της κίνησης $κιν_i$, της οποίας το μοντέλο M_i παρουσιάζει μέγιστη πιθανοφάνεια. Με άλλα λόγια, στόχος του βήματος αυτού είναι η βελτιστοποίηση του μοντέλου της κίνησης (*τροποποίηση στο πλήθος των εσωτερικών καταστάσεων, στην επιλογή του λεξικού κινήσεων κλπ*), ώστε να γίνει όσο το δυνατό αρτιότερη η μοντελοποίηση της κίνησης. Πρόκειται για την περίπτωση του Π2 που επιλύεται από το σύστημα με τη βοήθεια του αλγορίθμου του Viterbi.



Σχήμα 6.7.6 Αλγόριθμος του Viterbi

Βήμα 3: Η αναγνώρισης της κίνησης πραγματοποιείται από το σύστημα χρησιμοποιώντας την επίλυση του προβλήματος Π1. Έτσι, αξιολογείται το μοντέλο κάθε κίνησης $κιν_i$, με βάση την ακολουθία διανυσμάτων παρατήρησης $Y_{0:k}$, με σκοπό την επιλογή της κίνησης που

έχει μέγιστη πιθανοφάνεια. Αυτό το βήμα υλοποιείται με τη βοήθεια του αλγορίθμου “forward” του Baum.



Σχήμα 6.7.7 Διάγραμμα αναγνωριστή KMM για μεμονωμένες κινήσεις

6.8 Υλοποίηση δυναμικής αναγνώρισης δακτυλικών κινήσεων με το MATLAB

Όπως ήδη αναφέρθηκε στην εισαγωγή του κεφαλαίου αυτού χρησιμοποιήθηκε η εργαλειοθήκη «Kevin Murphy» για το MATLAB, προκειμένου να υλοποιηθούν τα συνεχή KMM του συστήματος. Βασικό πλεονέκτημα υλοποίησης των συνεχών KMM με τη συγκεκριμένη εργαλειοθήκη, σε σχέση με αντίστοιχη εργαλειοθήκη για διακριτά KMM (π.χ. «Stats 4.1»), είναι η δυνατότητα που παρέχει για εκπαίδευση KMM με περισσότερες από μια χαρακτηριστικές ιδιότητες. Αυτό αποτέλεσε και κριτήριο επιλογής της εργαλειοθήκης «Kevin Murphy».

Η υλοποίηση ξεκινάει με τη δημιουργία του τριδιάστατου πίνακα Data που περιλαμβάνει και τις 14 χαρακτηριστικές ιδιότητες όλων των ακολουθιών καρέ παρατήρησης, για κάθε μια από τις προς αναγνώριση κινήσεις.

Η διεξαγωγή μιας δακτυλικής κίνησης στο πιάνο δεν μπορεί ποτέ να είναι ποτέ απόλυτα όμοια με καμία επανάληψη της ίδιας κίνησης. Πόσο μάλλον στην περίπτωση διαφορετικών ερμηνευτών.

Κατά συνέπεια, οι ακολουθίες των καρτέ παρατήρησης των βίντεο εκπαίδευσης δεν μπορούν να έχουν το ίδιο μήκος, αφού ακόμα και για την ίδια κίνηση, κάθε επανάληψη έχει διαφορετικό χρόνο διεξαγωγής και άρα ο δείκτης k του μήκους της κάθε ακολουθίας δεν είναι σταθερός. Επειδή όμως η εργαλειοθήκη απαιτεί πίνακα σταθερού μήκους προτείνεται η χρήση cell arrays. Κατασκευάζεται, λοιπόν, ένας πίνακας για κάθε προς αναγνώριση κίνηση, ο οποίος αποτελείται από 14 γραμμές και στήλες όσες και το μήκος της ακολουθίας των καρτέ παρατήρησης.

Το πρώτο στάδιο της υλοποίησης με την εργαλειοθήκη « Kevin Murphy » είναι ο ορισμός του:

1. μήκους των ακολουθιών των καρτέ παρατήρησης, δηλαδή των στηλών που θα έχει ο πίνακας Data (αν είναι σταθερό και δε γίνει χρήση cell arrays),
2. πλήθους των χαρακτηριστικών ιδιοτήτων, δηλαδή των γραμμών που θα έχει ο πίνακας Data (αν είναι σταθερό και δε γίνει χρήση cell arrays),
3. πλήθους των ακολουθιών που εισάγονται.

Το δεύτερο στάδιο περιλαμβάνει τις αρχικοποιήσεις των μοντέλων. Εισάγεται ο πίνακας Prior που περιλαμβάνει τις πιθανότητες να βρεθεί σε κάθε μια από τις καταστάσεις το μοντέλο, στην πρώτη επανάληψη του αλγορίθμου. Χάρη στο γεγονός ότι κάθε μια προς αναγνώριση κίνηση έχει προκαθορισμένη έναρξη, μπορούμε να θέσουμε πολύ μεγάλη την πιθανότητα το κάθε μοντέλο να ξεκινάει από την πρώτη κατάσταση. Εισάγεται επίσης ο διδιάστατος πίνακας . Δημιουργήθηκε με τη βοήθεια της συνάρτησης mk_leftright_transmat. Ο πίνακας είναι τύπου « bakis », δηλαδή οι πιθανότητες μετάβασης σε προηγούμενες καταστάσεις είναι 0.

Αναλυτικότερα, ο τρόπος λειτουργίας της ενότητας της δυναμικής αναγνώρισης μέσα στο λογισμικό PianOrasis παρουσιάζεται στο εδάφιο 7.2.3.

6.9 Συμπεράσματα

Το κεφάλαιο 6 αποτελεί την επέκταση προς τη δυναμική αναγνώριση κινήσεων, της προτεινόμενης μεθόδου όρασης H/Y που περιγράφηκε στο κεφάλαιο 5. Στο παρόν κεφάλαιο αναπτύχθηκε ο τρόπος ανάλυ-

σης και στοχαστικής μοντελοποίησης των κινήσεων των δακτύλων του πιανίστα. Αναπτύχθηκε επίσης το μαθηματικό υπόβαθρο εκπαίδευσης των μοντέλων και ταξινόμησης των κινήσεων από το σύστημα.

Παρουσιάστηκαν ενδεικτικές και βασικές κινήσεις πιανιστικής μουσικής αλληλεπίδρασης. Οι κινήσεις αυτές αποτελούν τον πυρήνα της τεχνικής δεξιότητας ενός πιανίστα. Με βάση τις κινήσεις αυτές καθορίστηκαν τα γεωμετρικά χαρακτηριστικά, τα οποία οδηγούν στην δημιουργία διανυσμάτων παρατήρησης για κάθε καρέ της κίνησης. Παρουσιάστηκε αναλυτικά το μαθηματικό υπόβαθρο των Κρυφών Μαρκοβιανών Μοντέλων. Δικαιολογήθηκε η επιλογή των συνεχών KMM.

Ορίστηκε και θεμελιώθηκε μαθηματικά η θεωρία των συνεχών KMM αλλά και ο αλγόριθμος « forward » του Baum, ο αλγόριθμος του Viterbi και ο αλγόριθμος « Baum-Welch ». Παρουσιάστηκε η στατιστική μέθοδος που χρησιμοποιείται από τους αλγορίθμους Εκτίμησης-Μεγιστοποίησης. Αναπτύχθηκε η συνεισφορά και το μαθηματικό υπόβαθρο των Μεικτών Γκαουσιανών Μοντέλων στη μοντελοποίηση συνεχών διανυσμάτων παρατήρησης.

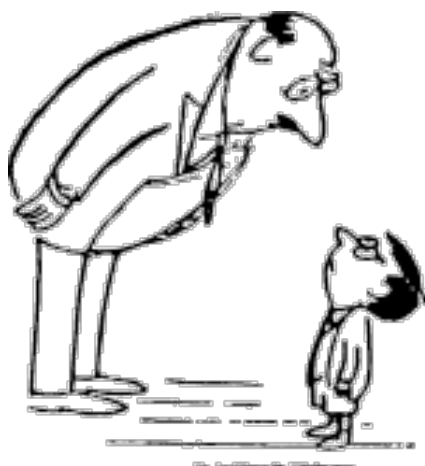
Περιγράφηκε αναλυτικό παράδειγμα χρήσης του συστήματος ως μεμονωμένου αναγνωριστή δακτυλικών κινήσεων. Αποτυπώθηκε με σχετικά σχήματα η χρησιμότητα και η συνεισφορά κάθε πτυχής της προτεινόμενης μεθόδου.

Στο κεφάλαιο που ακολουθεί διενεργείται παρουσίαση του λογισμικού PianOrasis καθώς και του τρόπου λειτουργίας. Τέλος, αξιολογείται το σύστημα σε πραγματικές συνθήκες μουσικής αλληλεπίδρασης.

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Η ανακάλυψη (4/4)

«Καμιά φορά μπερδεύομαι και τα πράγματα δεν πάνε καλά. Και όταν μπερδεύομαι το αφεντικό μου δε γελάει, καθόλου όμως.



‘Κοιτάξτε με στα μάτια Νικόλα’, λέει καθόλου ευχαριστημένος. ‘Αυτό το λέτε έρευνα εσείς;’, με ρωτάει. Και αυτό μοιάζει με ερώτηση, αλλά δεν πρέπει σε καμιά περίπτωση να απαντήσεις. Γιατί αν απαντήσεις, γίνεται κατακόκκινος!...»

7 Λογισμικό PianOrasis και αξιολόγηση

- 7.1 Σύνοψη κεφαλαίου**
- 7.2 PianOrasis**
 - 7.2.1 Διαπεφή PianOrasis
 - 7.2.2 Στατική αναγνώριση μέσω PianOrasis
 - 7.2.3 Δυναμική αναγνώριση μέσω PianOrasis
- 7.3 Πρώτη φάση αξιολόγησης**
- 7.4 Δεύτερη φάση αξιολόγησης**
- 7.5 Τρίτη φάση αξιολόγησης**
- 7.6 Τέταρτη φάση αξιολόγησης**
- 7.7 Συμπεράσματα**

7.1 Σύνοψη κεφαλαίου

Στο κεφάλαιο αυτό παρουσιάζεται το λογισμικό PianOrasis που αναπτύχθηκε για τις ανάγκες της διδακτορικής διατριβής. Αρχικά περιγράφεται και στη συνέχεια αξιολογείται ο τρόπος λειτουργίας του σε πραγματικές συνθήκες.

Το σύστημα αναγνώρισης δακτυλικών κινήσεων αξιολογήθηκε σε δύο διαφορετικές φάσεις. Στη δεύτερη φάση δόθηκε ένα μικρό απόσπασμα από τη « *Sonate pour piano n° 16 en do majeur* » του Wolfgang Amadeus Mozart.

Κατά την αξιολόγηση η κάμερα τοποθετήθηκε πάνω σε ηλεκτρονικό πιάνο. Η παραμετροποίηση για τον καθορισμό της κατάστασης ηρεμίας, για τα πέντε δάκτυλα, έγινε με βάση το αρχικό καρέ κάθε βίντεο. Στη συνέχεια ζητήθηκε στον πιανίστα να παίξει τα επιλεγμένα αποσπάσματα, χρησιμοποιώντας τους δακτυλισμούς που αναγράφονταν στην αντίστοιχη παρτιτούρα.

7.2 PianOrasis

Το PianOrasis είναι ένα λογισμικό που αναπτύχθηκε ειδικά για τις ανάγκες της παρούσας διδακτορικής διατριβής. Δε βασίζεται σε προϋπάρχουσα πλατφόρμα αναγνώρισης κινήσεων. Είναι ειδικά προσαρμοσμένο για τις απαιτήσεις της μουσικής αλληλεπίδρασης με

πίانو. Υλοποιήθηκε εξ ολοκλήρου στο μαθηματικό πολυεργαλείο MATLAB.

Οι δυνατότητες του PianOrasis είναι ακριβώς αυτές που αναπτύσσονται στα κεφάλαια 5 και 6. Περιλαμβάνει δυνατότητες για στατική αναγνώριση (ανάκτηση δακτυλισμών) και δυναμική αναγνώριση μέσω Κρυφών Μαρκοβιανών Μοντέλων.

7.2.1 Διεπαφή PianOrasis

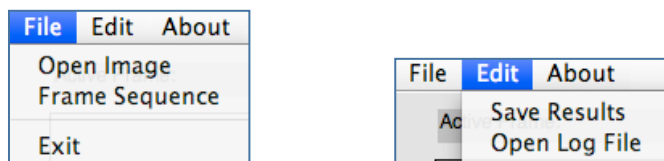
Η διεπαφή του λογισμικού παρουσιάζεται στο σχήμα 7.2.1.1. Χωρίζεται σε δώδεκα επιμέρους περιοχές οι οποίες είναι:

1. **Μενού επιλογών :** Στο μενού « File » συναντάμε όλες τις επιλογές για άνοιγμα εικόνας ή ακολουθίας εικόνων (Σχήμα 7.2.1.2). Ειδικότερα η επιλογή « Open image » επιτρέπει το άνοιγμα μια εικόνας από το λογισμικό. Η δεύτερη επιλογή « Frame Sequence » δίνει τη δυνατότητα αναπαραγωγής ακολουθίας εικόνων (Σχήμα 7.2.1.3). Οι εικόνες αυτές οφείλουν είναι αποθηκευμένες με τη μορφή προθέματος (π.χ. *mozart0000.bmp*, *mozart0001.bmp* κλπ). Δηλώνεται μια εικόνα ως έναρξη της ακολουθίας και στη συνέχεια προσδιορίζεται το μήκος της ακολουθίας. Τέλος υπάρχει η επιλογή εξόδου από το λογισμικό. Στο μενού « Edit » δίνεται η δυνατότητα αποθήκευσης σε « log file » των αποτελεσμάτων αναγνώρισης. Τέλος, το μενού « About » ενημερώνει για την έκδοση του λογισμικού.
2. **Active Frame :** Εμφανίζεται το καρτέ που επεξεργάζεται το σύστημα στη συγκεκριμένη φάση. Πρόκειται για το καρτέ στην αρχική του κατάσταση χωρίς καμία επεξεργασία.
3. **Hand Extraction :** Εμφανίζεται μια δυαδική εικόνα ως το αποτέλεσμα εφαρμογής του δερματικού μοντέλου στην εικόνα του « Active Frame ».
4. **Image Filtering :** Εμφανίζεται μια δυαδική εικόνα ως το αποτέλεσμα φιλτραρίσματος στην έξοδο εφαρμογής του δερματικού μοντέλου. Είναι το αποτέλεσμα ανάκτησης της φιγούρας του χεριού.
5. **Contour :** Εμφανίζεται η δυαδική εικόνα του περιγράμματος του χεριού.



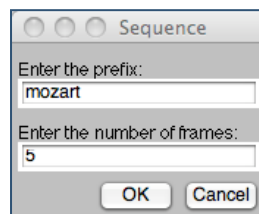
Σχήμα 7.2.1.1 PianOrasis

6. **Possible Edges** : Εμφανίζεται μια δυαδική εικόνα που περιέχει μόνο τα περιγράμματα των πιθανών άκρων των δακτύλων στο καρέ.
7. **Finger Detection** : Εμφανίζονται οι θέσεις των δακτύλων μέσω ανίχνευσης δέρματος και οι προβλεπόμενες θέσεις δακτύλων μέσω ταξινομητών.
- 8.



Σχήμα 7.2.1.2 Μενού « File » και « Edit »

9. **Filtering Parameters** : Είναι η περιοχή που περιλαμβάνει όλες τις επιλογές φιλτραρίσματος για τις περιοχές 4, 5 και 6. Περιέχει δυνατότητες παραμετροποίησης για φίλτρα Open, Min, Gauss καθώς για την ανάκτηση περιγράμματος μέσω της μεθόδου Canny.

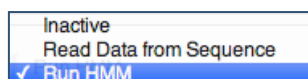


Σχήμα 7.2.1.3 Δήλωση ακολουθίας καρέ προς αναγνώριση

10. **Fingers Position** : Είναι η περιοχή που αναγράφονται οι α) θέσεις ηρεμίας των δακτύλων για ένα συγκεκριμένο βίντεο, β) οι θέσεις των δακτύλων σε κάθε καρέ (εμφανίζονται με κόκκινο) και γ) οι δακτυλισμοί (εμφανίζονται με πράσινο). Τέλος, προσφέρεται η δυνατότητα επαναπροσδιορισμού των θέσεων ηρεμίας των δακτύλων ανάλογα με το βίντεο.
11. **Find Missing Fingers/Show Colors** : Με την επιλογή « Find Missing Fingers » ενεργοποιείται η πρόβλεψη θέσης δακτύλων (εμφα-

νίζεται στην περιοχή 7 με πράσινα τετραγωνάκια). Με την επιλογή « Show Colors » κάθε φορά που ανακτάται ένας δακτυλισμός εμφανίζεται στην περιοχή 9 με πράσινο χρώμα.

12. **Single Frame/Frame Sequence :** Στην περιοχή αυτή διακρίνουμε δύο κουμπιά: α) το « Single Frame », όπου γίνεται ανίχνευση θέσεων δακτύλων και ανάκτηση δακτυλισμών σε μεμονωμένο καρέ και β) το « Frame Sequence », όπου γίνεται ανίχνευση θέσεων δακτύλων και ανάκτηση δακτυλισμών σε ακολουθία καρέ.
13. **Hidden Markov Models :** Στην κορυφή της περιοχής αυτής έχουμε τρεις αρχικές επιλογές: α) « Inactive », επιλέγοντάς την γίνεται στατική αναγνώριση κινήσεων, β) « Read Data from Sequence », επιλέγοντάς την γίνεται αποθήκευση σε κατάλληλο πίνακα των διανυσμάτων παρατήρησης μέσα από μια ακολουθία από καρέ, γ) « Run HMM », επιλέγοντάς την εξάγονται οι πιθανότητες των KMM για το βίντεο του οποίου την κίνηση θέλουμε να αναγνωρίσουμε (Σχήμα 7.2.1.4). Στην περίπτωση που έχουμε επιλέξει μια από τις δύο τελευταίες επιλογές, ενεργοποιούνται τέσσερα κουμπιά στην ίδια περιοχή: α) « Store Sequence Data », πατώντας το αποθηκεύεται ο πίνακας των διανυσμάτων παρατήρησης, β) « HMM Training », πατώντας το γίνεται εκπαίδευση των KMM με βάση τα διανύσματα παρατήρησης που αποθηκεύτηκαν μέσω του κουμπιού « Store Sequence Data » (*ενοποιούνται όλες οι ακολουθίες διανυσμάτων παρατήρησης των βίντεο εκπαίδευσης σε έναν πίνακα*), γ) « Probability HMM », πατώντας εξάγονται οι πιθανότητες του κάθε KMM για την κίνηση του βίντεο που θέλουμε να αναγνωρίσουμε και τέλος δ) « Clear Data », πατώντας το καθαρίζουν οι αποθηκεύσεις των διανυσμάτων παρατήρησης για τα βίντεο εκπαίδευσης. Στο κάτω μέρος της ίδιας περιοχής παρέχονται πληροφορίες σχετικά με το πλήθος των KMM που επεξεργάζεται το λογισμικό στην παρούσα φάση (*Num of HMMs*), το πλήθος των βίντεο εκπαίδευσης που αποθηκεύτηκαν για μία κίνηση (*Num of Videos*), το πλήθος των καρέ για το συγκεκριμένο βίντεο εκπαίδευσης (*Num of Frames*) και η λογαριθμική πιθανότητα του KMM (*logprob*).



Σχήμα 7.2.1.4 Επιλογή λειτουργιών KMM στο PianOasis

7.2.2 Στατική αναγνώριση

Τα βήματα που πρέπει να ακολουθήσουμε για να επιτύχουμε ανάκτηση δακτυλισμών μέσω του PianOasis είναι τα παρακάτω:

1. Ανοίγουμε το πρώτο καρέ του βίντεο
2. Ανοίγουμε το πρώτο καρέ του βίντεο που θέλουμε από File→Open και στη συνέχεια καθορίζουμε τον αριθμό των καρέ που θέλουμε να τρέξουμε File→Frame Sequence (πχ αν βάλουμε 5 θα αποθηκευτούν $5+1=6$ καρέ).
3. Πριν τρέξουμε την ακολουθία πρέπει να έχουμε τσεκάρει τις επιλογές « Find Missing Fingers » και « Show Colors ».
4. Τρέχουμε το βίντεο (*Run/Frame Sequence*).

7.2.3 Δυναμική αναγνώριση

Η δυναμική αναγνώριση δακτυλικών κινήσεων με χρήση του PianOasis ξεκινάει μέσω της εκπαίδευσης των KMM για κάθε κίνηση. Στόχος της εκπαίδευσης είναι να δημιουργηθούν οι πίνακες για κάθε μία κίνηση.

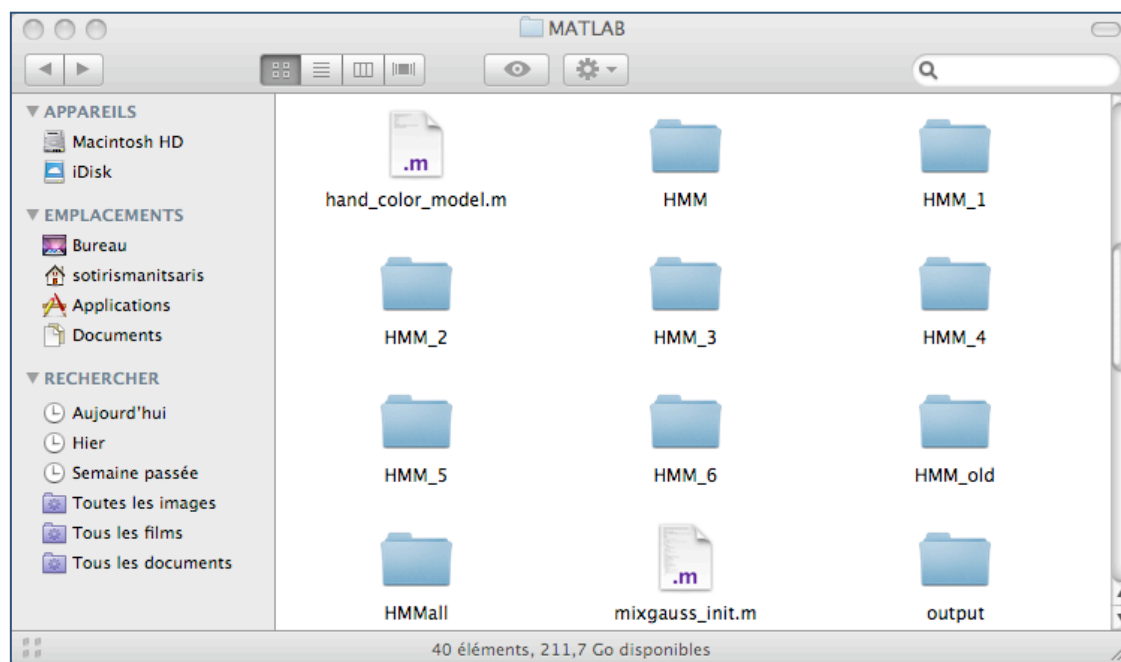
Για την εκπαίδευση των KMM ακολουθείται η εξής διαδικασία:

5. Στην περιοχή « Hidden Markov Model » επιλέγουμε « Read Data from Sequence ».
6. Ανοίγουμε το βίντεο που θέλουμε από File→Open και στη συνέχεια καθορίζουμε τον αριθμό των καρέ που θέλουμε να τρέξου-

με File→Frame Sequence (πχ αν βάλουμε 5 θα αποθηκευτούν $5+1=6$ καρέ).

7. Πριν τρέξουμε την ακολουθία πρέπει να έχουμε τσεκάρει τις επιλογές « Find Missing Fingers » και « Show Colors ».
8. Τρέχουμε το βίντεο (*Run/Frame Sequence*).
9. Μόλις ολοκληρωθεί το βίντεο αποθηκεύουμε τα χαρακτηριστικά διανύσματα για όλα τα καρέ του (*Store Sequence Data*).
10. Επαναλαμβάνοντας τα βήματα 2 έως 5 και προστίθενται στον πίνακα τα χαρακτηριστικά διανύσματα για κάθε βίντεο εκπαίδευσης της συγκεκριμένης κίνησης.
11. Ενεργοποιούμε την εκπαίδευση των HMM (*HMM Training*). Αυτό αποθηκεύει όλους τους εξαγόμενους πίνακες σε ένα αρχείο HMM
12. Αφού έχουμε τα χαρακτηριστικά της κίνησης τα τοποθετούμε σε έναν καινούργιο φάκελο HMM_#κίνηση01#. Έτσι, επιτυγχάνουμε να εκπαιδεύουμε τα KMM μια φορά για κάθε κίνηση.
13. Για να δημιουργήσουμε χαρακτηριστικά άλλης κίνησης επαναλαμβάνουμε τα βήματα 2 έως 7. Καλό είναι να κάνουμε πρώτα ένα Clear Data ή ακόμα και για ασφάλεια να κλείσουμε και να ξανά ανοίξουμε το πρόγραμμα.
14. Τα χαρακτηριστικά της καινούργιας κίνησης τα τοποθετούμε σε ένα άλλο φάκελο με όνομα HMM_#κίνηση02#. Επομένως, στο τέλος της εκπαίδευσης θα έχουμε μία σειρά από N κινήσεις. Τα δεδομένα εκπαίδευσής τους θα βρίσκονται σε φακέλους της μορφής HMM_#κίνησηN# (Σχήμα 7.2.3.1).

Εφόσον ολοκληρωθεί η διαδικασία εκπαίδευσης των KMM, μπορεί να ξεκινήσει η αναγνώριση κινήσεων. Δίνεται ως είσοδος ένα νέο βίντεο και εξάγεται ένα αρχείο που περιέχει τόσες πιθανότητες όσες και τα KMM που ορίστηκαν προηγουμένως. Η μεγαλύτερη πιθανότητα αντιστοιχεί στο KMM της κίνησης που αναγνωρίστηκε.



Σχήμα 7.2.3.1 Παράδειγμα οργάνωσης φακέλων εκπαίδευσης κινήσεων των KMM

Η αναγνώριση επιτυγχάνεται ακολουθώντας τα παρακάτω βήματα:

15. Για να εκτελεστεί η αναγνώριση επιλέγουμε « Run HMM » στην περιοχή « Hidden Markov Model ».
16. Στη συνέχεια διενεργούμε εξαγωγή διανυσμάτων παρατήρησης για την υπό εξέταση ακολουθία επαναλαμβάνοντας τα βήματα 2 έως 4 (δε χρειάζεται να πατήσουμε *Store Sequence Data*).
17. Πατάμε το κουμπί « Propability HMM » έχοντας καθορίσει αρχικά τον αριθμό των μοντέλων που θέλουμε να χρησιμοποιηθούν για την αναγνώριση. Οι πιθανότητες αποθηκεύονται με τη σειρά σε ένα αρχείο probs.txt.

7.3 Πρώτη φάση αξιολόγησης

Το πρώτο μουσικό απόσπασμα που χρησιμοποιήθηκε για την αξιολόγηση του συστήματος παρουσιάζεται στο σχήμα 7.3.1. Αναγράφονται οι δακτυλισμοί που χρησιμοποιήθηκαν από τον πιανίστα.



Σχήμα 7.3.1 Πρώτο μουσικό απόσπασμα αξιολόγησης σε ΝΤΟ μείζονα με τους δακτυλισμούς που χρησιμοποιήθηκαν

Οι έξοδοι του συστήματος για τα καρτέ 109 έως 116 παρατίθενται στο σχήμα 7.3.2. Οι έξοδοι αυτοί αντιστοιχούν στη νότα ΡΕ του δεύτερου μέτρου του αποσπάσματος. Στην ενότητα «Ανίχνευση ανά δάκτυλο» το 0 αντιστοιχεί στην κατάσταση όπου το δάκτυλο δεν ανιχνεύτηκε ως τοπικό μέγιστο κάποιας περιοχής, ενώ το 1 αντιστοιχεί στην κατάσταση όπου το δάκτυλο ανιχνεύτηκε. Στη «Συνολική ανίχνευση» τίθεται η τιμή 0 για κάποιο καρτέ στην περίπτωση που έστω και ένα από τα πέντε δάκτυλα δεν ανιχνεύτηκε ως τοπικό μέγιστο, ενώ τίθεται η τιμή 1 αν και τα πέντε δάκτυλα ανιχνεύτηκαν. Στη συνέχεια, αν για κάποιο καρτέ και κάποιο δάκτυλο στην ενότητα «Ανάκτηση δακτυλισμού» υπάρχει η τιμή 0 σημαίνει ότι για το δάκτυλο αυτό δεν ανακτήθηκε δακτυλισμός. Σε διαφορετική περίπτωση αποδίδεται η τιμή 1. Αν σε κάποιο καρτέ επιστραφούν δύο δακτυλισμοί, και δεν πρόκειται για συγχορδία ή για συνήχηση, τότε στην ενότητα «Πάτημα» αναγράφεται για το αντίστοιχο καρτέ η ένδειξη διπλό ή τριπλό κλπ.

Προσδιορίζεται ότι στην περίπτωση που κάποιο δάκτυλο δεν ανιχνευτεί ως τοπικό μέγιστο, ενεργοποιείται η πρόβλεψη θέσης του δακτύλου με τη χρήση των ταξινομητών. Συνεπώς, σε κάθε περίπτωση υπάρχει μια θέση που επιστρέφεται για κάθε δάκτυλο. Επιβεβαίωση αυτού αποτελεί το σχήμα 7.3.3. Στην όψη του PianOrasis για το καρτέ 112 βλέπουμε ότι ο αντίχειρας δεν ανιχνεύτηκε ως τοπικό μέγιστο (*κόκκινο κυκλάκι στην περιοχή εντοπισμού δακτύλου – finger detection*). Όλα τα υπόλοιπα δάκτυλα εντοπίστηκαν κανονικά. Παρό-

λα αυτά, και εφόσον είναι ενεργοποιημένη η επιλογή «Εύρεση ελλιών δακτύλων - Find Missing Fingers», διενεργείται πρόβλεψη θέσης με βάση την πορεία που είχαν στα προηγούμενα καρέ.

Α/Α καρέ	Ανίχνευση ανά δάκτυλο					Συνολική ανίχνευση	Ανάκτηση δακτυλισμού					Νότα Πάτημα
	1 ^ο	2 ^ο	3 ^ο	4 ^ο	5 ^ο		1 ^ο	2 ^ο	3 ^ο	4 ^ο	5 ^ο	
109	0	1	1	1	1	0	0	1	1	0	0	διπλό διπλό ΡΕ
110	0	1	1	1	1	0	0	1	1	0	0	
111	0	1	1	1	1	0	0	1	0	0	0	
112	0	1	1	1	1	0	0	1	0	0	0	
113	0	1	1	1	1	0	0	1	0	0	0	
114	0	1	1	1	1	0	0	1	0	0	0	
115	0	1	1	1	1	0	0	1	0	0	0	
116	0	1	1	1	1	0	0	1	0	0	0	

Σχήμα 7.3.2 Έξοδοι ανίχνευσης και ανάκτησης από το PianOrasis για τα καρέ 109 έως 116

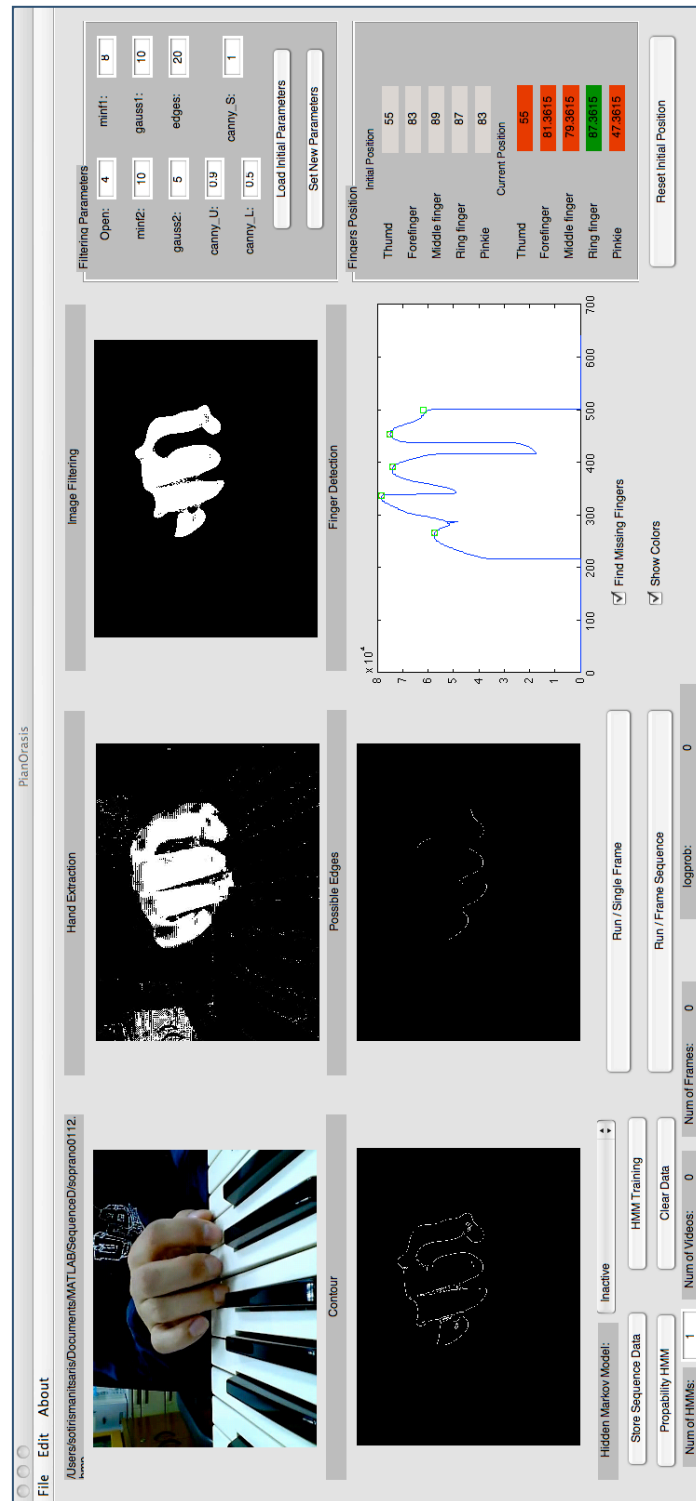
Έτσι, επιστρέφεται η πιθανή θέση του αντίχειρα για το καρέ 112 (πράσινο τετραγωνάκι στην περιοχή εντοπισμού δακτύλου – *finger detection*). Κατά συνέπεια, ακόμα κι αν για κανένα καρέ του σχήματος 7.3.2 δεν ανιχνεύτηκε ο αντίχειρας (τιμή=0 για τον αντίχειρα), με την εφαρμογή πρόβλεψης θέσης για το συγκεκριμένο δάκτυλο, το σύστημα συνεχίζει να λειτουργεί κανονικά. Αξίζει να σημειωθεί, ότι το ποσοστό των δακτύλων που δεν ανιχνεύτηκαν ενώ πατούσαν κάποιο πλήκτρο τείνει στο μηδέν. Όποτε σε κάθε περίπτωση το σύστημα παραμένει ανεπηρέαστο από μια τέτοια περίπτωση.

Παρατηρώντας το σχήμα 7.3.2 και ειδικότερα τα καρέ 109 και 110, διαπιστώνεται ότι ανακτώνται δύο δακτυλισμοί για μια μοναδική νότα. Προφανώς, τα δύο αυτά καρέ αποτελούν προβληματικά σημεία. Το ερώτημα που τίθεται είναι αν πρόκειται για αστοχία του συστήματος ή για σφάλμα του πιανίστα. Λαμβάνοντας υπόψη ότι ο ρυθμός λήψης βίντεο από την κάμερα είναι ίσος με 25 καρέ το δευτερόλεπτο, συμπεραίνεται ότι το 1 καρέ αντιστοιχεί σε 0,04 δευτερόλεπτα. Προφανώς, ακόμα και για δύο καρέ, δηλαδή 0,08 δευτερόλεπτα, η απώλεια της πληροφορίας είναι ελάχιστη, αν αναλογιστεί

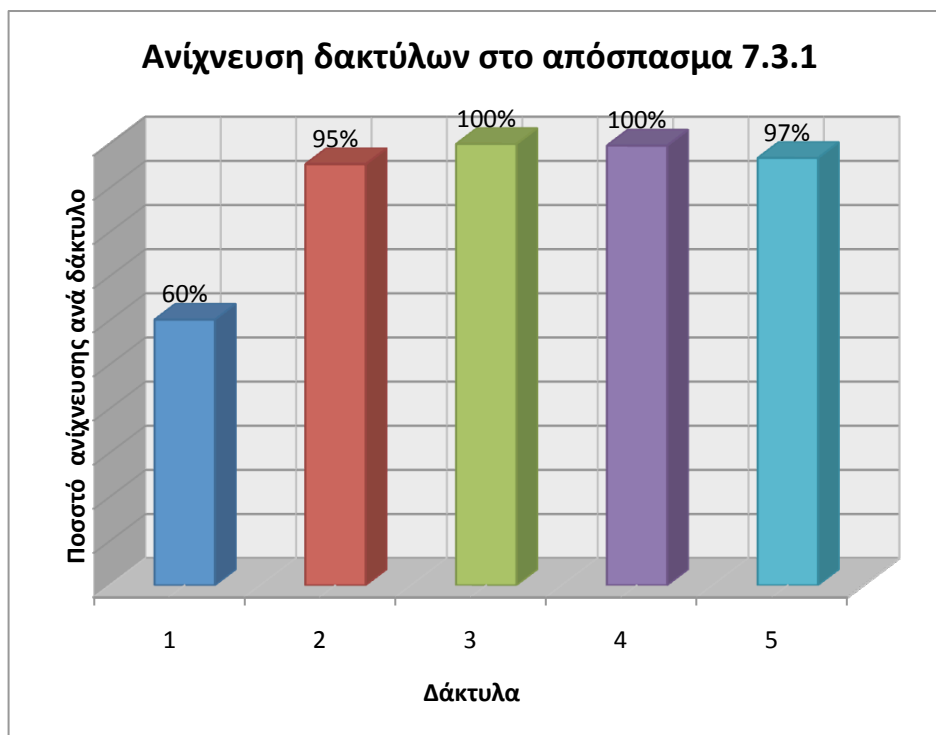
κανείς ότι μια νότα του αποσπάσματος του σχήματος 7.3.1 αντιστοιχεί σε 14,25 καρέ ή σε 0,57 δευτερόλεπτα. Συνεπώς, το σφάλμα των καρέ 109 και 110 αντιστοιχεί σε *ποσοστιαίο σφάλμα ανά νότα* ίσο με $0,08/0,57=0,14=14\%$. Όσο το ποσοστιαίο σφάλμα ανά νότα τείνει στο μηδέν στο 0 τόσο δεν επηρεάζεται πρακτικά η αναγνώριση κινήσεων και είναι ταυτόχρονα πιθανότερο να οφείλεται στον ίδιο τον πιανίστα (*κίνηση staccato ή legato*). Αντίθετα, όσο το ποσοστιαίο σφάλμα ανά νότα μεγαλώνει, τόσο είναι πιθανότερο να οφείλεται σε λάθος παραμετροποίηση του κατωφλιού πατήματος ενός πλήκτρου.

Στο σχήμα 7.3.4 παρουσιάζονται τα ποσοστά ανίχνευσης ανίχνευσης ανά δάκτυλο, σε δείγμα 284 καρέ. Είναι αυταπόδεικτο ότι τα τέσσερα από τα πέντε παρουσιάζουν πολύ υψηλά ποσοστά ανίχνευσης χωρίς να είναι συχνά απαραίτητη η πρόβλεψη θέσης. Το πρώτο δάκτυλο, λόγω της γεωμετρίας του, έχει χαμηλότερο ποσοστό. Δίνεται λύση στο παρόν πρόβλημα μέσω της πρόβλεψης θέσης των δακτύλων στην περίπτωση που δεν ανιχνευτεί η θέση κάποιου δακτύλου. Το δεύτερο και το πέμπτο δάκτυλο έχουν τις αμέσως υψηλότερα ποσοστά ανίχνευσης. Το τρίτο και το τέταρτο δάκτυλο έχουν τα υψηλότερα ποσοστά.

Τέλος, στο συγκεκριμένο βίντεο ανιχνεύτηκαν σωστά οι δακτυλισμοί στο 81% των καρέ. Από τις 16 νότες του αποσπάσματος ανακτήθηκαν οι δακτυλισμοί για τις 15 από αυτές, η αλλιώς για 94%. Στη μια από αυτές, όπου και χρησιμοποιήθηκε τέταρτο δάκτυλο για τη νότα ΦΑ, δεν έγινε ανάκτηση δακτυλισμού για δύο πιθανούς λόγους: α) το κατώφλι πατήματος του τέταρτου δακτύλου δεν είναι αρκετά ευαίσθητο με αποτέλεσμα αν το δάκτυλο δεν πάτησε μέχρι τέρμα το πλήκτρο να μην ανιχνεύτηκε το πάτημα, β) το τέταρτο δάκτυλο κρύφτηκε πίσω από κάποιο μαύρο πλήκτρο με αποτέλεσμα να μην ανιχνεύτηκε σωστή θέση του άκρου του δακτύλου και κατά συνέπεια δεν ανιχνεύτηκε το πάτημα του πλήκτρου. Σε κάθε περίπτωση, μια λύση είναι η διαφορετική παραμετροποίηση του PianOrasis ώστε να διενεργεί πιο ευαίσθητη ανίχνευση.



Σχήμα 7.3.3 Όψη PianOrasis για το καρέ 112



Σχήμα 7.3.4 Ανίχνευση δακτύλων χωρίς πρόβλεψη σε 284 καρτέ

7.4 Δεύτερη φάση αξιολόγησης

Η δεύτερη φάση αξιολόγησης βασίζεται σε ένα απόσπασμα από τη « *Sonate pour piano n° 16 en do majeur* » του Wolfgang Amadeus Mozart (1756-1791). Η σονάτα είναι σε κλίμακα ΝΤΟ μείζονα. Ένα απόσπασμα από το πρώτο μέρος της δόθηκε στο PianOraSis για ανάκτηση δακτυλισμών. Το απόσπασμα αφορά από την αρχή μέχρι την πρώτη νότα του έβδομου μέτρου (πράσινοι δείκτες) για το δεξί χέρι (Σχήμα 7.4.1).

Το βίντεο που χρησιμοποιήθηκε για την αξιολόγηση του συστήματος αποτελείται από 368 καρτέ βίντεο που αντιστοιχούν στο απόσπασμα του σχήματος 7.4.1 από την πρώτη νότα του πρώτου μέτρου μέχρι την πρώτη νότα του έβδομου μέτρου. Στο απόσπασμα διακρίνουμε νότες διαφορετικής αξίας. Ο ρυθμός του κομματιού είναι Allegro (γρήγορο, χαρούμενο). Αντιστοιχεί δηλαδή σε γρήγορα πατήματα πλήκτρων από τον πιανίστα. Συνεπώς, αποτελεί ένα καλό

τεστ για το PianOasis. Χρησιμοποιήθηκε χαμηλός ρυθμός λήψης βίντεο στα 19 καρέ ανά δευτερόλεπτο ώστε να αξιολογηθεί το σύστημα και σε πιο δύσκολες καταστάσεις.

Μια περίπτωση σφάλματος του συστήματος που ξεχώρισε είναι η κατάσταση παύσης όταν αυτή συνοδεύεται από μετακίνηση του χεριού. Τόσο στο δεύτερο όσο και στο τέταρτο μέτρο του αποσπασματος παρατηρείται εμφάνιση αυτού του σφάλματος. Στο σχήμα 7.4.2 παρουσιάζονται οι έξοδοι του PianOasis για τα καρέ 106-113. Τα καρέ αυτά αναφέρονται στην παύση του δεύτερου μέτρου με ταυτόχρονη μετακίνηση του χεριού. Ενώ δεν υπάρχει πραγματικός δακτυλισμός σε αυτά τα καρέ, παρόλα αυτά ανακτώνται δακτυλισμοί από το σύστημα. Αυτό οφείλεται στην αύξηση της απόστασης άκρων δακτύλων και κεντροειδούς. Το πάτημα ενός πλήκτρου ανιχνεύεται αν η παραπάνω διαφορά ξεπεράσει ένα κατώφλι εκ των προτέρων καθορισμένο. Όταν το χέρι μετακινείται ενδέχεται τα δάκτυλα του πιανίστα να « κρέμονται », αυξάνοντας την ποσότητα **τεταγμένη δακτύλου-τεταγμένη κεντροειδούς** και κάνοντάς την να ξεπερνάει το προκαθορισμένο κατώφλι.

Αντίστοιχη είναι η περίπτωση του σχήματος 7.4.3. Παρατηρείται αντίστοιχα ότι για τα καρέ 180-187 επιστρέφονται από το PianOasis δακτυλισμοί ενώ δε θα έπρεπε να επιστραφούν. Μια ενδεχόμενη λύση στην περίπτωση αυτή θα ήταν η αντικατάσταση του κατωφλίου για το πάτημα του πλήκτρου με μια πιο αποδοτική διαδικασία ελέγχου.

Τα καρέ του σχήματος 7.4.4 αναφέρονται στην «ανιούσα κίνηση κλίμακας» του πέμπτου μέτρου. Πρόκειται για γειτονικά πατήματα πλήκτρων που προκαλούνται από γρήγορα passages δεκάτων έκτων σε legato. Στο παρακάτω σχήμα παρατηρούμε ότι ανάμεσα στις νότες PE, MI και FA ανακτώνται συνηχήσεις από νότες.

Ειδικότερα, στα καρέ 224, 225 και 228 επιστρέφονται διπλά πατήματα πλήκτρων. Το καρέ 224 για παράδειγμα αναφέρεται στη χρονική στιγμή όπου το πρώτο δάκτυλο ετοιμάζεται να αφήσει το πλήκτρο, ενώ το δεύτερο δάκτυλο ετοιμάζεται να πατήσει. Κάτι αντίστοιχο συμβαίνει για τα καρέ 225 και 228 (Σχήμα 7.4.4). Ένα τέτοιο φαινόμενο στη μουσική αλληλεπίδραση θεωρείται απόλυτα φυσιολογικό, επειδή ακριβώς οι νότες εδώ εκτελούνται με κάποιο βάρος και όχι με σχετική « τσαχπινάδα ». Σε μια προσέγγιση καθαρά από τη σκοπιά της πληροφορικής, θα μπορούσε να θεωρηθεί το φαινόμενο, ως σφάλμα ή αστοχία του συστήματος. Από τη σκοπιά των μαθηματικών αποδεικνύεται ότι η χρήση ενός απλού κατωφλίου για το πάτημα ή όχι ενός πλήκτρου δεν μπορεί να καλύψει την τεράστια ποικιλία

δακτυλικών κινήσεων ενός πιανίστα. Σε κάθε περίπτωση το σύστημα επιδέχεται παραμετροποίηση της ευαισθησίας αναγνώρισης – ανάκτησης.

Sonate 16 en D^o Majeur
Sonate facile

Allegro W. A. Mozart
K 545

Σχήμα 7.4.1 Δεύτερο μουσικό απόσπασμα από τη σονάτα No.16 του Wolfgang Amadeus Mozart

Στο σχήμα 7.4.5 παρουσιάζονται έξοδοι του PianOrasis για τα καρέ 284 έως και 291. Μέσα από αυτό το σχήμα παρουσιάζεται η συμβολή των ταξινομητών στο σύστημα. Ειδικότερα, στα καρέ 287, 288 και 289 το πρώτο δάκτυλο δεν ανιχνεύεται στην εικόνα (πράσινο). Πιθανόν, να κρύβεται πίσω από κάποιο άλλο δάκτυλο. Παρόλα αυτά, χάρη στην πρόβλεψη θέσης που διενεργείται από τους ταξινομητές, με βάση τα τρία αμέσως προηγούμενα καρέ (284, 285 και 286 – κίτρινο), ανακτάται σωστά ο δακτυλισμός και για τα επόμενα καρέ.

Αξίζει να σημειωθεί ότι είναι σπάνιο, με βάση την αξιολόγηση που έγινε στο PianOasis, δάκτυλο που κρύβεται από άλλα δάκτυλα, να πατάει εκείνη τη χρονική στιγμή και η πληροφορία αυτή να μην καταγραφεί σε κανένα καρτέ.

Α/Α καρτέ	Ανίχνευση ανά δάκτυλο					Συνολική ανίχνευση	Ανάκτηση δακτυλισμού					Νότα Πάτημα
	1°	2°	3°	4°	5°		1°	2°	3°	4°	5°	
106	1	1	1	1	1	1	1	0	0	0	0	Παύση 1 δουλό
107	1	1	1	1	1	1	1	0	1	0	0	
108	1	1	1	1	1	1	0	0	0	0	1	
109	1	1	1	1	1	1	0	0	1	0	1	
110	1	1	1	1	1	1	0	0	1	0	1	
111	1	1	1	1	1	1	0	0	1	0	1	
112	1	1	1	1	1	1	0	0	0	0	0	
113	1	1	1	1	1	1	0	0	0	0	0	δουλό

Σχήμα 7.4.2 Έξοδοι ανίχνευσης και ανάκτησης από το PianOasis για τα καρτέ 106 έως 113

Η μελέτη του σχήματος 7.4.6 δίνει χρήσιμα συμπεράσματα στο βαθμό ανίχνευσης των δακτύλων χωρίς τη χρήση των ταξινομητών. Η σύγκριση του σχημάτων 7.4.6 και 7.3.4 αποδεικνύει το σημαντικό ρόλο του φωτισμού σε ένα σύστημα αναγνώρισης κινήσεων. Η κανονικοποίηση του χρωματικού χώρου (όπως αυτή περιγράφηκε στο εδάφιο 3.6.1.2) μπορεί να κάνει το σύστημα λιγότερο ευαίσθητο σε μεταβολές φωτισμού, πλην όμως δεν μπορεί να καταργήσει τους νόμους της φυσικής: ένας καλός φωτισμός είναι καλύτερος από έναν κακό φωτισμό. Απόδειξη αυτού αποτελεί το ποσοστό ανίχνευσης του πρώτου δακτύλου στην πρώτη και τη δεύτερη φάση αξιολόγησης. Στην πρώτη φάση είχαμε ένα απλό μουσικό απόσπασμα με κακό φωτισμό και ποσοστό ανίχνευσης στο 60%, ενώ στη δεύτερη φάση αξιολόγησης έχουμε ένα πιο σύνθετο μουσικό απόσπασμα με καλό φωτισμό και ποσοστό ανίχνευσης στο 72%. Επίσης, παρατηρείται μια μεγαλύτερη ισορροπία στα ποσοστά ανίχνευσης ανάμεσα στα δάκτυλα, σε σχέση με αυτά του σχήματος 7.3.4.

Α/Α καρτέ	Ανίχνευση ανά δάκτυλο					Συνολική ανίχνευση	Ανάκτηση δακτυλισμού					Νότα Πάτημα			
	1°	2°	3°	4°	5°		1°	2°	3°	4°	5°	Παύση 2	διπλό		
180	1	1	1	1	1	1	0	0	1	0	0				
181	1	1	1	1	1	1	0	0	1	0	0				
182	1	1	1	1	1	1	0	0	1	0	0				
183	1	1	1	1	1	1	0	0	1	1	0				
184	1	1	1	1	1	1	0	0	1	1	0				
185	1	1	1	1	1	1	0	0	1	1	0				
186	1	1	1	1	1	1	0	0	1	1	0				
187	1	1	1	1	1	1	0	0	1	0	0				

Σχήμα 7.4.3 Έξοδοι ανίχνευσης και ανάκτησης από το PianOasis για τα καρτέ 180 έως 187

Α/Α καρτέ	Ανίχνευση ανά δάκτυλο					Συνολική ανίχνευση	Ανάκτηση δακτυλισμού					Νότα Πάτημα	
	1°	2°	3°	4°	5°		1°	2°	3°	4°	5°		
222	0	0	0	0	0	0	1	0	0	0	0	ΡΕ	
223	0	0	0	0	0	0	1	0	0	0	0		
224	0	0	0	0	0	0	1	1	0	0	0		διπλό
225	0	0	0	0	0	0	1	1	0	0	0		διπλό
226	1	1	1	1	1	1	0	1	0	0	0	ΜΙ	
227	1	1	1	1	1	1	0	1	0	0	0		
228	1	1	1	1	1	1	0	1	1	0	0		διπλό
229	1	1	1	1	1	1	0	0	1	0	0	ΦΑ	

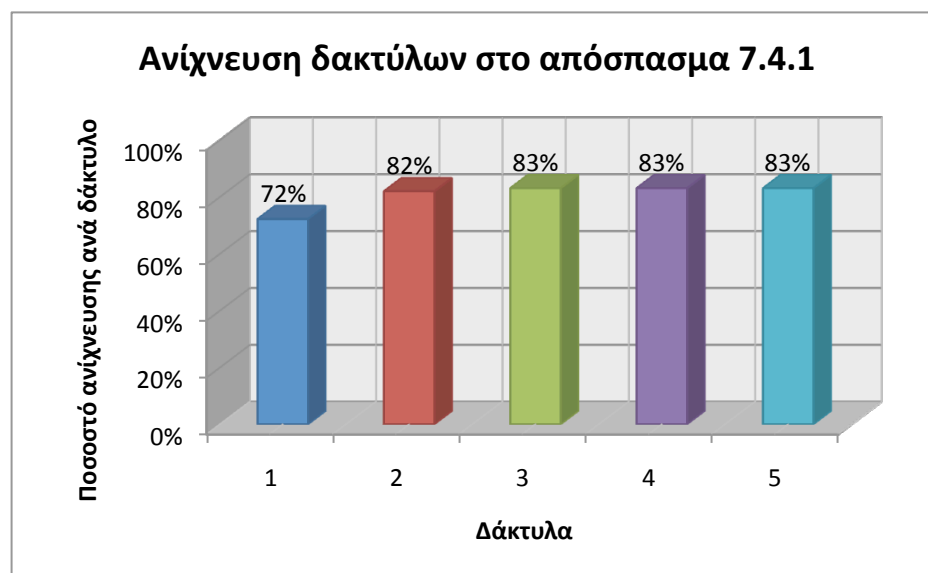
Σχήμα 7.4.4 Έξοδοι ανίχνευσης και ανάκτησης από το PianOasis για τα καρτέ 222 έως 229

Συνολικά, και για τα πέντε δάκτυλα, ο εντοπισμός θέσης των δακτύλων, με ανίχνευση και ταξινομητές, λειτούργησε θετικά για το

83% των καρτέ του βίντεο αυτού, ενώ στο 71% των καρτέ οι δακτυλισμοί ανακτήθηκαν σωστά. Αξίζει να σημειωθεί ότι καρτέ που ανακτήθηκαν διπλοπατήματα πλήκτρων, λόγω της κίνησης legato, ή εσφαλμένα πατήματα λόγω μετακίνησης του χεριού, θεωρήθηκαν για την παρούσα αξιολόγηση ως καρτέ εσφαλμένης ανάκτησης.

Α/Α καρτέ	Ανίχνευση ανά δάκτυλο					Συνολική ανίχνευση	Ανάκτηση δακτυλισμού					Νότα Πάτημα
	1°	2°	3°	4°	5°		1°	2°	3°	4°	5°	
284	1	1	1	1	1	1	1	0	0	0	0	ΣΟΛ
285	1	1	1	1	1	1	1	0	0	0	0	
286	1	1	1	1	1	1	1	0	0	0	0	
287	0	1	1	1	1	0	1	0	0	0	0	
288	0	1	1	1	1	0	1	0	0	0	0	
289	0	1	1	1	1	0	1	0	0	0	0	ΛΑ
290	0	1	1	1	1	0	0	1	0	0	0	
291	0	1	1	1	1	0	0	1	0	0	0	

Σχήμα 7.4.5 Έξοδοι ανίχνευσης και ανάκτησης από το PianOasis για τα καρτέ 284 έως 291



Σχήμα 7.4.6 Ανίχνευση δακτύλων χωρίς πρόβλεψη σε 386 καρτέ

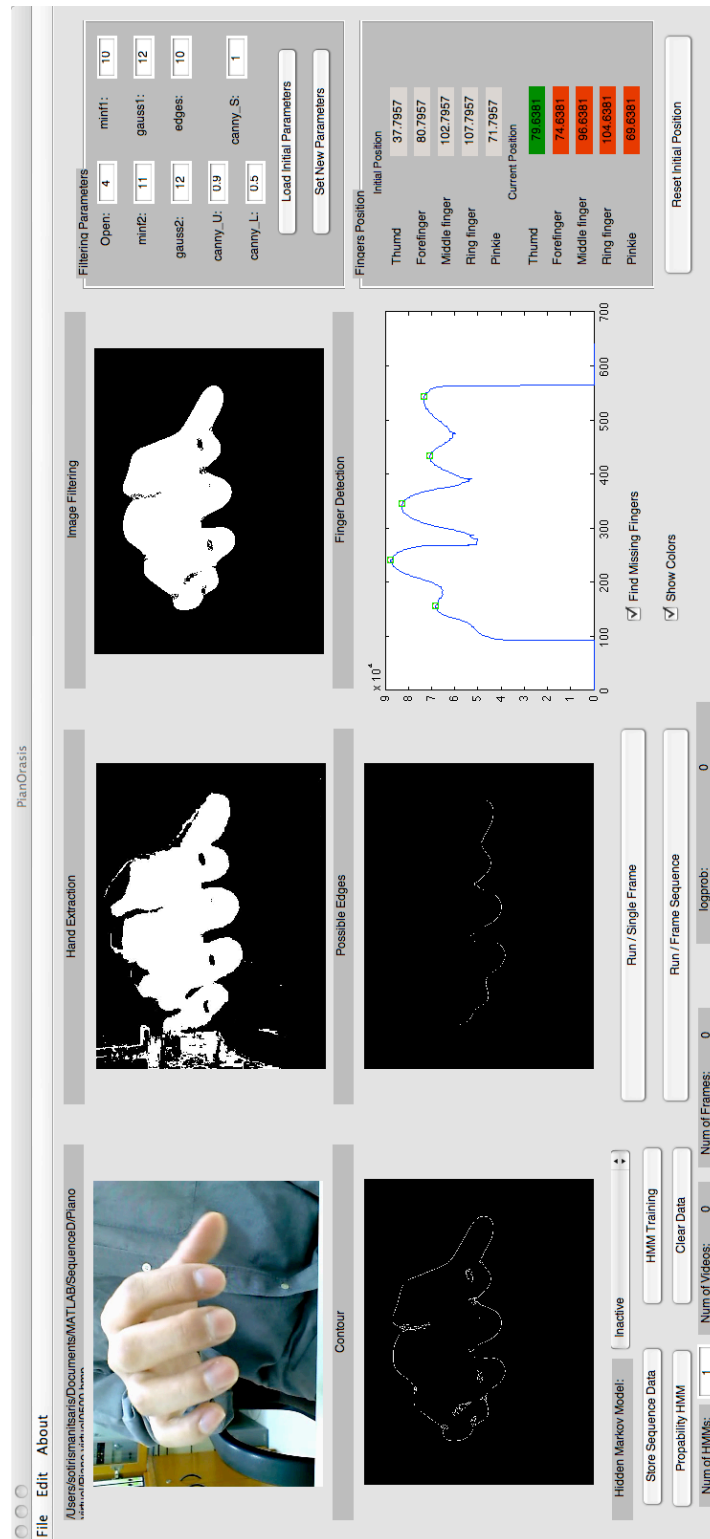
7.5 Τρίτη φάση αξιολόγησης

Κατά την τρίτη φάση αξιολόγησης το σύστημα υποβλήθηκε σε τεστ αναγνώρισης κινήσεων χωρίς τη χρήση μουσικού οργάνου. Ο πιανίστας παρήγαγε ανιούσα κίνηση αρπέζ, ενώ τα δάκτυλά του κινούνταν ελεύθερα στον αέρα.

Αξιολογήθηκαν συνολικά 130 καρτέ βίντεο που αντιστοιχούν στην εν λόγω κίνηση, χωρίς χρήση μουσικού οργάνου. Συνολικά, και για τα πέντε δάκτυλα, ο εντοπισμός θέσης των δακτύλων, με ανίχνευση και ταξινομητές, λειτούργησε θετικά για το 100% των καρτέ του βίντεο αυτού, ενώ στο 97% των καρτέ οι δακτυλισμοί ανακτήθηκαν σωστά.

Ο λόγος που αιτιολογεί τα υψηλά ποσοστά είναι η απουσία πλήκτρων στην εικόνα. Στις προηγούμενες φάσεις αξιολόγησης παρατηρήθηκε ότι τα μαύρα πλήκτρα εμπόδιζαν σε ορισμένες περιπτώσεις την ανίχνευση ολόκληρου του δακτύλου (*κυρίως του 4^{ου} δακτύλου*), με αποτέλεσμα να μην ξεπερνιέται το κατώφλι πατήματος πλήκτρου. Στην παρούσα φάση αξιολόγησης, ο παρών περιορισμός δεν υφίσταται. Τα δάκτυλα κινούνται ελεύθερα στον αέρα χωρίς τον οποιοδήποτε περιορισμό. Στο σχήμα 7.5.1 απεικονίζεται η όψη του PianOrasis για ένα καρτέ που αντιστοιχεί στη νότα ΝΤΟ της κίνησης. Όπως φαίνεται, ανιχνεύεται επιτυχώς ο δακτυλισμός του πρώτου δακτύλου. Συνεπώς, είναι εύκολο να φανταστεί κανείς τις επεκτάσεις του PianOrasis στην κατεύθυνση του εικονικού μουσικού οργάνου.

Μειονέκτημα της χρήσης του PianOrasis ως σύστημα αναγνώρισης δακτυλικών κινήσεων χωρίς μουσικό όργανο είναι η απουσία σημείου αναφοράς για την τοποθέτηση του χεριού στην εικόνα. Όταν υπάρχει μουσικό όργανο, ο πιανίστας τοποθετεί σε συγκεκριμένες θέσεις το χέρι και τα δάκτυλά του. Στην αντίθετη περίπτωση είναι υποχρεώμενος να ανακαλεί εμπειρικά τις θέσεις των πλήκτρων, καθώς και να ικανοποιεί τις αποστάσεις μεταξύ τους.



Σχήμα 7.5.1 Όψη PianOrasis σε λειτουργία εικονικού πιάνου

7.6 Τέταρτη φάση αξιολόγησης

Κατά την τέταρτη φάση αξιολόγησης το σύστημα δυναμικής αναγνώρισης δακτυλικών κινήσεων υποβλήθηκε σε διαδικασία εκπαίδευσης για κάθε μια από τις κινήσεις που παρουσιάστηκαν στην ενότητα 6.2.

Εισήχθησαν στο PianOrasis 20 βίντεο εκπαίδευσης για κάθε μια από τις 6 κινήσεις, δηλαδή συνολικά 120 βίντεο εκπαίδευσης. Το κάθε βίντεο αναλύθηκε καρέ-καρέ από το σύστημα και η επεξεργασία των δεδομένων χρησιμοποιήθηκε για την εκπαίδευση των KMM (ένα για κάθε κίνηση). Χρησιμοποιήθηκαν βίντεο ρυθμού σύλληψης 19 καρέ το δευτερόλεπτο ώστε να μη «βαρύνει» πολύ η διαδικασία εκπαίδευσης. Τα χαρακτηριστικά με τα οποία εκπαιδεύτηκε το σύστημα είναι αυτά που περιγράφηκαν στην ενότητα 6.3. Καμία περαιτέρω επεξεργασία δεν έγινε στα βίντεο εκπαίδευσης χάρη στην επιλογή συνεχών KMM. Οι εσωτερικές καταστάσεις των μοντέλων καθορίστηκαν ανάλογα με το πλήθος των πατημάτων. Συνεπώς, για παράδειγμα για την ανιούσα κίνηση αρπές με 4 πατήματα χρησιμοποιήθηκε μοντέλο 4 εσωτερικών καταστάσεων. Αξίζει να σημειωθεί ότι όσο μεγαλύτερο το πλήθος των εσωτερικών καταστάσεων, τόσο μεγαλύτερο των απαιτούμενων βίντεο εκπαίδευσης για την εκάστοτε κίνηση.

Το σύστημα αξιολογήθηκε ως μεμονωμένος αναγνωριστής δακτυλικών κινήσεων. Αυτό σημαίνει ότι κάθε φορά το σύστημα καλούνταν να αναγνωρίσει μια από τις 6 κινήσεις μέσα από νέο βίντεο που δίνονταν στο PianOrasis. Για κάθε κίνηση χωριστά δόθηκαν 10 βίντεο προς αναγνώριση. Τα αποτελέσματα της αναγνώρισης φαίνονται στο σχήμα 7.6.1.

Μέσα από το παρακάτω σχήμα διαπιστώνεται εύκολα ότι η «ανιούσα κίνηση σε τρίτες» και η «κατιούσα κίνηση σε τρίτες» παρουσίασαν τις μεγαλύτερες δυσκολίες στην αναγνώριση και κατά συνέπεια τα μικρότερα ποσοστά αναγνώρισης. Στα ποσοστά αυτά εντάσσονται και οι περιπτώσεις όπου το σύστημα έδωσε πολύ μικρή πιθανότητα το βίντεο εισόδου να ανήκει σε κάποια από τις 6 κινήσεις. Όταν το βίντεο εισόδου μιας κίνησης δίνει πολύ μικρά ποσοστά τότε κατατάσσεται στις αταξινόμητες κινήσεις. Αυτό οφείλεται στο σύνθετο των κινήσεων αυτών καθώς και στον έντονα στοχαστικό χαρακτήρα που έχουν. Παρατηρείται ακόμα ότι σε δύο περιπτώσεις, για κάθε μια κίνηση, δόθηκε μεγαλύτερη πιθανότητα στην ανιούσα ή κατιούσα κλίμακας αντίστοιχα. Μια αιτιολόγηση του αποτελέσματος

αυτού είναι το κοινό σημείο αφετηρίας και τερματισμού της «ανιούσας κίνησης σε τρίτες» και της «ανιούσας κλίμακας» (καθώς και της κατιούσας). Η μόνη διαφορά στις δύο κινήσεις είναι η συνήχηση νοτών στις κλίμακες σε τρίτες, διαφορά που αποδείχτηκε ανεπαρκής για τον πλήρη διαχωρισμό των κινήσεων. Ένα μεγαλύτερο πλήθος από βίντεο εκπαίδευσης σίγουρα θα βελτίωνε κατά πολύ τα ποσοστά αναγνώρισης αυτών των κινήσεων.

Δακτυλική κίνηση	Ανιούσα κλίμακας	Κατιούσα κλίμακας	Ανιούσα αρπές	Κατιούσα αρπές	Ανιούσα κλίμακας σε τρίτες	Κατιούσα κλίμακας σε τρίτες	Αταξινόμητες κινήσεις
Ανιούσα κλίμακας	7	0	0	0	2	0	1
Κατιούσα κλίμακας	0	8	0	0	0	2	0
Ανιούσα αρπές	0	0	10	0	0	0	0
Κατιούσα αρπές	0	0	0	10	0	0	0
Ανιούσα κλίμακας σε τρίτες	2	0	0	0	6	0	2
Κατιούσα κλίμακας σε τρίτες	0	2	0	0	0	5	3
Αναγνώριση	70%	80%	100%	100%	60%	50%	10%

Σχήμα 7.6.1 Αποτελέσματα δυναμικής αναγνώρισης μέσω PianOraSis

Αντίθετα με τις κινήσεις σε τρίτες, οι κινήσεις αρπές παρουσίασαν απόλυτα ποσοστά επιτυχίας. Τα ποσοστά αυτά δικαιολογούνται από την απλότητα και σαφήνεια των κινήσεων αυτών, τη σχετικά μεγάλη απόσταση μεταξύ των δακτύλων και τα λίγα πατήματα πλήκτρων που αντιστοιχούν σε μικρό πλήθος καρέ. Το μικρό πλήθος καρέ διευκολύνει κατά πολύ ακόμα και την εκπαίδευση του μόντελου της κίνησης. Τα 20 βίντεο εκπαίδευσης που χρησιμοποιήθηκαν για την κάθε μια κίνηση αρπές ενδεχομένως να είναι κατά πολύ περισσότερα από το ελάχιστο απαραίτητο πλήθος. Τέλος, όλες οι παραπάνω παράμετροι ανέδειξαν τις κινήσεις σε αρπές τις πιο εύκολες κινήσεις προς αναγνώριση.

Αναφορικά με την ανιούσα και κατιούσα κλίμακας τα ποσοστά κυμάνθηκαν ανάμεσα σε 70 και 80%. Πρόκειται για κινήσεις με έντονο στοχαστικό χαρακτήρα. Όπως ήδη αναφέρθηκε στην προηγούμενη παράγραφο το PianOrasis αναγνώρισε δύο φορές την κίνηση ως κίνηση σε τρίτες. Η δυσκολία στην ανιούσα κίνηση κλίμακας έγκειται στο υποχρεωτικό πέρασμα του πρώτου δακτύλου πίσω από τα υπόλοιπα. Το γεγονός αυτό προκαλεί εσφαλμένα διανύσματα παρατήρησης για μικρό πλήθος καρτέ και αυξάνει τις απαιτήσεις σε πλήθος βίντεο εκπαίδευσης.

7.7 Συμπεράσματα

Κατά την αξιολόγηση της μεθόδου μέσω του PianOrasis εξετάστηκαν συνολικά 1000 καρτέ βίντεο, από τα οποία παρουσιάστηκαν τα σημαντικότερα συμπεράσματα από τα 782 καρτέ (284 καρτέ στην πρώτη φάση, 368 καρτέ στη δεύτερη φάση και 130 στην τρίτη φάση). Για τη δυναμική αναγνώριση κινήσεων με χρήση KMM χρησιμοποιήθηκαν 120 βίντεο εκπαίδευσης καθώς και 60 βίντεο εξέτασης. Εξετάστηκε η συμπεριφορά του PianOrasis σε διαφορετικές συνθήκες φωτισμού και μουσικών αποσπασμάτων. Ο εντοπισμός θέσης των δακτύλων και η αναγνώριση κινήσεων είχε ποσοστό επιτυχίας 85% κατά μέσο όρο για όλες τις φάσεις αξιολόγησης, ενώ το ποσοστό επιτυχούς αναγνώρισης με χρήση KMM ήταν 71%.

Η εφαρμογή της προτεινόμενης μεθόδου μέσω του λογισμικού PianOrasis ικανοποιεί σε πολύ μεγάλο βαθμό του στόχους που ορίστηκαν στο πρώτο κεφάλαιο:

- Το σύστημα εξάγει τους δακτυλισμούς που επιλέγονται από τον πιανίστα. Αναγνωρίζει κινήσεις μέσα από διαδικασία εκπαίδευσης των KMM, απλά και γρήγορα.
- Το σύστημα δεν απαιτεί καμία προκαταρκτική πληροφορία ή ανάλυση της παρτιτούρας. «Διαβάζει» απευθείας τις επιλογές του μουσικού μέσα από την καταγραφή της μουσικής ερμηνείας.
- Το σύστημα είναι μη διεισδυτικό. Δηλαδή, ο μουσικός δεν είναι υποχρεωμένος να προσαρμόσει το στυλ στις ανάγκες του συστήματος ή να φορέσει ειδικό εξοπλισμό.
- Η μέθοδος βασίζεται σε μια webcamera χαμηλού κόστους, ενώ το PianOrasis απαιτεί ελάχιστες χειροκίνητες ρυθμίσεις, οι ο-

ποίες ενδεχομένως να μπορούν να αυτοματοποιηθούν σε μεταγενέστερη έκδοση. Συνεπώς, το PianOrasis είναι προσβάσιμο από την άποψη κόστους και ευκολίας στη χρήση. Πάντως, απαιτούνται επιπλέον φάσεις αξιολόγησης του συστήματος ώστε να καλυφθεί η περίπτωση διαφορετικού πιανιστικού στυλ.

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Τα σεμινάρια

«Κατά καιρούς, ένας κύριος πολύ πολύ σημαντικός και πάρα πολύ σπουδαίος (αλλά όχι τόσο όσο το αφεντικό μου φυσικά) έρχεται να μας μιλήσει για πράγματα σούπερ πολύπλοκα. Αυτό λέγεται σεμινάριο. Ούτε αυτό είναι για γέλια. Όταν ο κύριος τελειώνει την ομιλία του, το αφεντικό μου του θέτει ερωτήσεις πολύ δύσκολες, και εκείνος δεν ξέρει πάντα να απαντήσει. Και αυτό δεν είναι καθόλου σωστό, γιατί εκείνον δεν τον μαλώνει κανένας!»



8 Επίλογος

- 8.1 Σύνοψη κεφαλαίου
- 8.2 Περίληψη συνεισφοράς
- 8.3 Περιορισμοί και επέκταση της μεθόδου
- 8.4 Πρωτοτυπία και ανασκόπηση διατριβής
- 8.5 Συμπεράσματα

8.1 Σύνοψη κεφαλαίου

Στο κεφάλαιο αυτό θα παρουσιαστεί η περίληψη της συνεισφοράς του διδακτορικού, οι περιορισμοί της μεθόδου και πιθανές μελλοντικές επεκτάσεις της. Θα πραγματοποιηθεί επίσης ανασκόπηση της διατριβής ανακεφαλαιώνοντας τις κεντρικές ιδέες της διδακτορικής διατριβής αναδεικνύοντας την πρωτοτυπία της μεθόδου.

8.2 Περίληψη συνεισφοράς

Έχοντας ως κίνητρο την απουσία ουσιαστικών συνδέσμων μεταξύ όρασης H/Y και μουσικής αλληλεπίδρασης, μελετήθηκε στα πλαίσια της παρούσας έρευνας η σύλληψη και ανάπτυξη μιας μεθόδου στατικής και δυναμικής αναγνώρισης κινήσεων. Δόθηκε έμφαση στον τομέα της μουσικής αλληλεπίδρασης.

Η βασική συνεισφορά της παρούσας έρευνας αγγίζει ταυτόχρονα το χώρο της όρασης H/Y και της μουσικής αλληλεπίδρασης:

- 1) Επιβεβαιώθηκε η τεχνική επάρκεια της όρασης H/Y να απαντήσει σε εξειδικευμένα ζητήματα αναγνώρισης κινήσεων.
- 2) Επιβεβαιώθηκε ότι εναλλακτικές μέθοδοι της τεχνολογίας των αισθητήρων, όπως η όραση H/Y σε συνδυασμό με τα Κρυφά Μαρκοβιανά Μοντέλα, μπορούν να έχουν παρόμοια αποτελέσματα με εξαιρετικά α) *χαμηλότερο κόστος*, β) *μεγαλύτερη φιλικότητα* προς τον άνθρωπο, γ) *εύκολη πρόσβαση* στην αναγνώριση κινήσεων..

- 3) Εφαρμόστηκαν οι αρχικές ιδέες μέσω της σύλληψης και υλοποίησης δύο συστημάτων : α) *ανάκτησης δακτυλίσμων (στατική αναγνώριση κινήσεων)*, β) *δυναμική αναγνώριση κινήσεων* σε μια διεπαφή, το PianOrasis.

8.3 Περιορισμοί και επέκταση της μεθόδου

Η παρούσα μέθοδος υπόκειται σε τεχνικούς περιορισμούς σχετικά με την:

- 1) **Κάμερα:** α) Η αναγνώριση διενεργείται μόνο για τα άσπρα πλήκτρα του πιάνου. β) Η καταγραφή του βίντεο διενεργείται σε δύο οκτάβες στο πιάνο. γ) Η ανάλυση της κάμερας ρυθμίστηκε στα 640x480 εικονοστοιχεία, στα 25 καρέ ανά δευτερόλεπτο (*Για διαφορετική ανάλυση χρειάζεται προσαρμογή των βίντεο εκπαίδευσης*).
- 2) **Μονοδιαδικαστική λήψη δεδομένων από το σύστημα:** Καταγράφεται εικόνα που αφορά μόνο στο δεξί χέρι του πιανίστα. Το σύστημα δε λαμβάνει καμιά πληροφορία σχετικά με τις κινήσεις του αριστερού χεριού.
- 3) **Μερική αναγνώριση:** Δε διεξάγεται αναγνώριση των κινήσεων που αφορούν στα μαύρα πλήκτρα.

Προτάσεις που θα μπορούσαν να οδηγήσουν στην απαλοιφή των περιορισμών και την επέκταση της μεθόδου:

- 1) **Διαφορετικές παραμετροποιήσεις της κάμερας:** α) Πιο ευρυγώνια λήψη. β) Υψηλότερη ανάλυση.
- 2) **Πολυδιαδικαστική λήψη δεδομένων:** α) Χρήση δεύτερης κάμερας. β) Προσθήκη επεξεργασίας ηχητικού σήματος.
- 3) **Δημιουργία ειδικού στηρίγματος για την κάμερα:** Σταθεροποίηση της γωνίας λήψης.
- 4) **Δημιουργία εικονικού μουσικού οργάνου:** Προσαρμογή μεθόδων 3Δ αναπαράστασης και εικονικής πραγματικότητας της κίνησης των χεριών του πιανίστα.

Το πρόβλημα αναγνώρισης κινήσεων και των δύο χεριών ταυτόχρονα για τα άσπρα δάκτυλα μπορεί να επιλυθεί με διπλή εφαρμογή του αλγορίθμου τμηματοποίησης και ανάκτησης άκρων δακτύλων. Η προσαρμογή έγκειται κυρίως στην προσαρμογή των ταξινομητών για τη σωστή ταυτοποίηση των δακτύλων.

Ενδιαφέρον παρουσιάζει επίσης η περίπτωση εκπαίδευσης του συστήματος με βίντεο προερχόμενα από πιανίστες με διαφορετικό ερμηνευτικό στυλ, αλλά και ο εμπλουτισμός του με νέα μοντέλα κινήσεων προς αναγνώριση.

8.4 Πρωτοτυπία και ανασκόπηση διατριβής

Σκοπός της παρούσας διδακτορικής έρευνας ήταν η ανάπτυξη μιας πρωτότυπης μεθόδου όρασης H/Y για την ανάκτηση δακτυλισμών και αναγνώριση κινήσεων του δεξιού χεριού ενός πιανίστα, μέσω ανάλυσης και στοχαστικής μοντελοποίησης της κίνησης.

Στο 1^ο κεφάλαιο τέθηκαν επιμέρους στόχοι για την επίτευξη αυτού του σκοπού. Οι στόχοι αυτοί προσδιόριζαν το πλαίσιο ανάπτυξης της μεθόδου. Η ικανοποίηση των στόχων οδήγησε στην ανάπτυξη της παρούσας μεθόδου που είναι :

- 1) **Ολοκληρωμένη:** ικανή να υπολογίζει όλες τις παραμέτρους που καθορίζουν τις δακτυλικές κινήσεις.
- 2) **Οπτικοκεντρική:** δε χρησιμοποιεί προκαταρκτική μουσικολογική ανάλυση της παρτιτούρας.
- 3) **Μη περιοριστική:** δεν επιβάλλει στο μουσικό να παίρνει ειδική στάση ή να χρησιμοποιεί με τα χέρια του ειδικό εξοπλισμό.
- 4) **Προσβάσιμη:** είναι χαμηλού κόστους και κατά συνέπεια προσβάσιμη στον οποιοδήποτε κατέχει ακόμα και μη αναβαθμισμένο εξοπλισμό.

Στο 2^ο κεφάλαιο έγινε ανασκόπηση όλων των μεθόδων ανάλυσης και μοντελοποίησης κινήσεων, οι οποίες χωρίζονται σε τρεις μεγάλες ερευνητικές περιοχές:

- 1) **Υπολογισμό κινήσεων:** Αφορά κινήσεις που σχετίζονταν λιγότερο με την παραγωγή του ήχου και περισσότερο με τη μετατόπιση επιμέρους μελών του ανθρώπινου σώματος. Αποτελούν μια κατηγορία λιγότερο στοχευμένη, με φάσμα μελέτης σχεδόν όλο το ανθρώπινο σώμα.
- 2) **Αναγνώριση κινήσεων:** Η συντριπτική πλειοψηφία των καταγεγραμμένων ερευνών αφορούν σε κινήσεις των χεριών του διευθυντή ορχήστρας. Οι έρευνες που αφορούν στην αναγνώριση «πιανιστικών κινήσεων», βρίσκονται σε αρχάριο επίπεδο. Συνεπώς, η αναγνώριση δακτυλικών κινήσεων μέσω όρασης Η/Υ και στοχαστικής μοντελοποίησης είναι ένα νέο ερευνητικό πεδίο με πεδίο εφαρμογής τη μουσική αλληλεπίδραση, αλλά και όχι μόνο.
- 3) **Ανάκτηση δακτυλισμών:** Οι έρευνες που ανακτώνται από τη βιβλιογραφία είναι εξειδικευμένες και αφορούν στα δάκτυλα του μουσικού. Στόχος τους μπορεί να είναι α) η πρόταση δακτυλισμών, είτε β) η ανάκτηση δακτυλισμών σε πραγματικό χρόνο, είτε τέλος γ) η ανάκτηση δακτυλισμών από προγενέστερη καταγραφή μουσικής ερμηνείας. Γενικότερα κατηγοριοποιούνται σε τρεις ομάδες: α) την *a priori* επεξεργασία της παρτιτούρας, β) την επεξεργασία, κινήσεων ή συμβάντων, σε πραγματικό χρόνο, και γ) τη *a posteriori* επεξεργασία ηλεκτρικού σήματος.

Οι παραπάνω μέθοδοι δεν ικανοποιούν το σκοπό της παρούσας έρευνας διότι δεν ευθυγραμμίζονται με τους ειδικούς της στόχους. Ειδικότερα, ο τομέας του υπολογισμού κινήσεων είναι μια ερευνητική προσέγγιση μακροσκοπικού χαρακτήρα, η οποία προσφέρει συμπεράσματα που συνδράμουν στην κατανόηση του τρόπου κίνησης των μελών του σώματος. Δεν καλύπτεται η ανάγκη για έρευνα μικροσκοπικού χαρακτήρα εστιάζοντας σε περιοχές όπως τα δάκτυλα και τα άκρα τους.

Η έρευνα στον τομέα της αναγνώρισης κινήσεων στη μουσική αλληλεπίδραση περιορίζεται στο χώρο της τεχνολογίας αισθητήρων. Οι αισθητήρες προσφέρουν μεν μεγάλη ακρίβεια στην ανάκτηση της πληροφορίας, περιορίζουν δε την ελευθερία του μουσικού, τοποθετώντας στο σώμα του ειδικό εξοπλισμό. Είναι προφανές ότι καθί-

σταται αδύνατη η διαμεσολάβηση αισθητήρων μεταξύ μουσικού και οργάνου.

Αναφορικά με την *ανάκτηση δακτυλισμών*, οι *a priori* μέθοδοι είτε δε λαμβάνουν υπόψη όλες τις παραμέτρους που καθορίζουν μια κίνηση στη μουσική αλληλεπίδραση, είτε απαιτούν ανάλυση της μουσικής παρτιτούρας, είτε απευθύνονται μόνο σε ηλεκτρονικά – ψηφιακά μουσικά όργανα. Στις μέθοδοι επεξεργασίας σε πραγματικό χρόνο το πρόβλημα της ανάκτησης δακτυλισμών επιλύεται μερικώς, εξάγοντας ως πληροφορία το ζεύγος (*χορδή, τάστο*). Οι μέθοδοι της *a posteriori* επεξεργασίας βασίζονται στην ηχητική ανάλυση ζωντανής ηχογράφησης και υστερούν σε ένα κοινό σημείο, την επέκταση/γενίκευση. Προσφέρουν λύσεις αποκλειστικά για την ανάκτηση δακτυλισμών χωρίς να λαμβάνουν υπόψη τη φύση της κίνησης.

Με δεδομένο ότι καμία από τις υπάρχουσες μεθόδους δεν ικανοποιούν πλήρως τα κριτήρια και τους στόχους της παρούσας διατριβής, αναπτύχθηκε μια νέα μέθοδος βασισμένη στην όραση H/Y. Στο 3^ο κεφάλαιο αναπτύχθηκαν τα σημαντικά εργαλεία της όρασης H/Y. Έγινε παράθεση των βασικών πτυχών της ανθρώπινης όρασης, αλλά και παρουσίαση του τρόπου που αυτή κατανοείται από τον άνθρωπο. Παρουσιάστηκε η πρόταση του David Marr καθώς και η προσαρμογή της στην αναγνώριση δακτυλικών κινήσεων. Αναπτύχθηκαν οι βασικότερες μέθοδοι εντοπισμού χεριού στην εικόνα. Η *τμηματοποίηση μέσω χρωματικότητας*, βασίζεται στη χαρακτηριστική απόχρωση του ανθρώπινου δέρματος. Η *τμηματοποίηση μέσω διαφοράς εικόνων* βασίζεται στον υπολογισμό της μετατόπισης του χεριού στην εικόνα. Ο *εντοπισμός μέσω εμφάνισης* βασίζεται στα οπτικά χαρακτηριστικά του προς αναγνώριση αντικειμένου.

Στη συνέχεια περιγράφηκαν οι βασικές λειτουργίες της μαθηματικής μορφολογίας και οι αντίστοιχες τεχνικές μείωσης θορύβου σε δυαδικές εικόνες. Η κλίση εικόνας χρησιμοποιείται συχνά ως εργαλείο ανάκτησης περιγράμματος. Παρουσιάστηκαν τα πιο σημαντικά μοντέλα περιγράμματος καθώς και τα κριτήρια απόδοσης ανάκτησης περιγράμματος που ορίζονται από τη βιβλιογραφία.

Σχετικά με τον *εντοπισμό των άκρων των δακτύλων*, αναπτύχθηκε ο τρόπος λειτουργία των αλγορίθμων προβολής υπογραφών, κυκλικού μετασχηματισμού Hough, εξαγωγής γεωμετρικών χαρακτηριστικών και χρωματικών δεικτών. Έγινε αναλυτική σύγκριση των μεθόδων, ενώ παρουσιάστηκαν τα δυνατά και τα αδύνατα σημεία της κάθε μιας.

Το 3^ο κεφάλαιο ολοκληρώθηκε με τη δυναμική αναγνώριση κινήσεων-χειρονομιών. Έγινε περιγραφή των μεθόδων της Δυναμι-

κής Περιτύλιξη του Χρόνου, των Νευρωνικά Δίκτυα Χρονικής Καθυστέρησης αλλά και τα Κρυφών Μαρκοβιανών Μοντέλων. Τα ΚΜΜ είναι η πιο δημοφιλής μέθοδος αναγνώρισης κινήσεων διότι διαχειρίζονται επιτυχώς, μέσω στοχαστικής μοντελοποίησης, την αβεβαιότητα που υπάρχει σε όλες τις κινήσεις.

Η όραση H/Y επιλέχτηκε διότι καλύπτει τις απαιτήσεις της διατριβής, σε αντίθεση με την τεχνολογία των αισθητήρων που συνήθως χρησιμοποιούνται στην Επικοινωνία Ανθρώπου Υπολογιστή και στην αναγνώριση κινήσεων. Το πλεονέκτημα της συγκεκριμένης μεθόδου όρασης H/Y είναι ότι πρόκειται για α) μη διεισδυτική μέθοδο, η οποία β) βασίζεται σε κάμερα χαμηλού κόστους. Συνεπώς, η μέθοδος καθίσταται προσβάσιμη από άποψη κόστους, εργονομίας και χρηστικότητας. Στο 4^ο κεφάλαιο παρουσιάστηκε μια μεθοδολογία «συρραφής» των εργαλείων όρασης H/Y που παρουσιάστηκαν στο 3^ο κεφάλαιο. Η μεθοδολογία διακρίνεται σε βήματα:

- 1) **Μοντελοποίηση κινήσεων:** αναπαρίσταται στο χώρο και στο χρόνο,
- 2) **Δημιουργία δερματικού μοντέλου:** δημιουργείται το δερματικό μοντέλο μέσα από μια βιβλιοθήκη εικόνων, με χρήση στατιστικής,
- 3) **Εντοπισμός περιοχής ενδιαφέροντος:** εντοπίζεται η περιοχή ενδιαφέροντος σε οποιοδήποτε καρέ του βίντεο,
- 4) **Επεξεργασία εικόνας:** αναπτύσσεται αλγόριθμος μετατροπής της εικόνας σε δεδομένα κίνησης,
- 5) **Αναγνώριση κινήσεων:** αντιπαραθέτει για κάθε καρέ χωριστά διακριτές τιμές δεδομένων με το μοντέλο (στατική αναγνώριση) ή μοντελοποιεί στοχαστικά συνεχείς τιμές δεδομένων, από ακολουθίες καρέ, και τις αντιπαραθέτει με κατανομές δεδομένων εκπαίδευσης (δυναμική αναγνώριση).

Πρόκειται για μια γενικευμένη πρόταση ανεξάρτητα με πεδίο εφαρμογής. Εφαρμογή αυτής μπορεί να υπάρξει όπου υφίσταται ανάγκη για καταγραφή, μοντελοποίηση και αναγνώριση δακτυλικών κινήσεων.

Στο 5^ο κεφάλαιο αναπτύχθηκε ο εξοπλισμός που χρησιμοποιήθηκε για αναπτυχθεί η παρούσα μέθοδος. Χρησιμοποιήθηκε α) H/Y iMac με διπύρηνο επεξεργαστή στα 2.66 Ghz, με 4 GB Ram, β) συνθεσάιζερ CS1x της Yamaha, γ) web-κάμερα Logitech QuickCam Communi-cate Deluxe. Προαπαιτούμενο του συστήματος είναι η εφαρμογή του σε οποιοδήποτε όργανο διαθέτει οκτάβες πλήκτρων, κλασικό ή μη, πιάνο ή συνθεσάιζερ. Το σύστημα αναπτύχθηκε στο μαθηματικό πολυεργαλείο MATLAB, με τη συνδρομή της εργαλειοθήκης DIPimage για επεξεργασία εικόνας.

Στο ίδιο κεφάλαιο αναπτύχθηκε το πρωτότυπο σύστημα στατικής αναγνώρισης δακτυλικών κινήσεων. Βασικοί άξονες λειτουργίας του συστήματος είναι:

- 1) **Ανίχνευση δέρματος:** Αναπτύχθηκε στατιστικό μοντέλο ειδικά για το παρών σύστημα. Στοχεύει στον εντοπισμό δερματικών περιοχών σε κάθε καρέ του βίντεο εισόδου. Με βάση μια βιβλιοθήκη εικόνων του πιανίστα προσδιορίζεται η περιοχή ενδιαφέροντος. Πρόκειται για μια ψηφιακή εικόνα που περιέχει μόνο εικονοστοιχεία στο χρώμα του δέρματος και των νυχιών. Μέσω κανονικοποίησης του 3Δ RGB διενεργείται ανίχνευση δέρματος σε οποιοδήποτε καρέ του βίντεο. Εξάγεται μια δυαδική εικόνα, όπου η μια κατάσταση αντιστοιχεί σε δέρμα ή νύχια και η άλλη σε αντικείμενα του παρασκήνι-α.
- 2) **Τμηματοποίηση χεριού:** Η τμηματοποίηση του χεριού δέχεται ως είσοδο τη δυαδική εικόνα που προέκυψε από πριν. Χωρίζεται α) στην *απλοποίηση της δυαδικής εικόνας* και β) στο *διαχωρισμό της σε βασικές συνιστώσες*, με χρήση μαθηματικής μορφολογίας. Στην περίπτωση (α) διενεργείται α) *μείωση θορύβου* και β) *ανάκτηση φιγούρας χεριού*, ενώ στην περίπτωση (β) διενεργείται α) *ανάκτηση περιγράμματος χεριού* και β) *ανάκτηση των άκρων των δακτύλων*.
- 3) **Εντοπισμός δακτύλων:** Ο εντοπισμός θέσης των δακτύλων, γίνεται μέσω υπολογισμού ευκλείδειων αποστάσεων μεταξύ ενός σημείου αναφοράς και των θέσεων των εικονοστοιχείων που ανήκουν στο περίγραμμα των δακτύλων. Η ταυτοποίησή τους γίνεται με τη χρήση κατάλληλων τα-

ξινομητών. Συνεπώς, ο εντοπισμός και η ταυτοποίηση των δακτύλων είναι διαδικασίες ανεξάρτητες από το μέγεθος της παλάμης.

- 4) Α) **Εξαγωγή διανυσμάτων παρατήρησης:** Τα διανύσματα παρατήρησης που καταγράφονται για τη στατική αναγνώριση είναι οι διαφορές της τεταγμένης του κάθε δακτύλου με την τεταγμένη του κεντροειδούς
- 5) Α) **Ανάκτηση δακτυλισμών:** Η ανάκτηση των δακτυλισμών επιτυγχάνεται με καθορισμό ενός κατωφλίου για το πάτημα, ή όχι, κάποιου πλήκτρου. Η ανάκτηση παρουσιάζει πολύ υψηλά ποσοστά επιτυχίας. Ακόμα και σε περίπτωση απόκρυψης κάποιου δακτύλου, ο δακτυλισμός ανακτάται από τα αμέσως επόμενα καρέ, χωρίς καθυστέρηση.

Στο 6ο κεφάλαιο αναλύεται ο τρόπος α) μοντελοποίησης της κίνησης των δακτύλων του πιανίστα, β) εκπαίδευσης των KMM και γ) αναγνώρισης των κινήσεων. Το πρωτότυπο σύστημα δυναμικής αναγνώρισης κινήσεων λειτουργεί σε συνέχεια της στατικής αναγνώρισης κινήσεων, από το στάδιο του εντοπισμού των δακτύλων και μετά. Συνεπώς τα στάδια 4 και 5 του 5^{ου} κεφαλαίου επαναπροσδιορίζονται με όρους δυναμικής αναγνώρισης κινήσεων

- 4) Β) **Εξαγωγή διανυσμάτων παρατήρησης:** Τα διανύσματα παρατήρησης που καταγράφονται για τη δυναμική αναγνώριση είναι: α) οι διαφορές της τεταγμένης του κάθε δακτύλου με την τεταγμένη του κεντροειδούς, β) οι τετμημένες του κάθε δακτύλου, γ) οι διαφορές τετμημένων για γειτονικά δάκτυλα.
- 5) Β) **Αναγνώριση κινήσεων:** Συνεχείς τιμές διανυσμάτων παρατήρησης, από ακολουθίες καρέ, μοντελοποιούνται με MFM, ενώ κάθε κίνηση προς αναγνώριση μοντελοποιείται στοχαστικά με KMM. Αξιολογείται στη συνέχεια το μοντέλο κάθε κίνησης, με βάση τα βίντεο εκπαίδευσης, με σκοπό την επιλογή της κίνησης που έχει μέγιστη πιθανοφάνεια.

Στο 6^ο κεφάλαιο επίσης, ορίστηκε ένα λεξικό κινήσεων προς αναγνώριση, καθώς και ένα αλφαβητάρι δακτυλισμών. Συνδυασμοί δακτυλισμών συνθέτουν μια δακτυλική κίνηση στο πιάνο. Οι κινήσεις αυτές είναι α) ανιούσα κίνηση κλίμακας, β) κατιούσα κίνηση κλίμακας, γ) ανιούσα κίνηση αρπές, δ) κατιούσα κίνηση αρπές, ε) ανιούσα κίνησης κλίμακας σε τρίτες, στ) κατιούσα κίνηση κλίμακας σε τρίτες. Οι κινήσεις αυτές αναλύθηκαν αρμονικά και μελωδικά προκειμένου να προσδιοριστούν οι ελάχιστες δομικές καταστάσεις που συνθέτουν την κάθε κίνηση.

Το σήμα του βίντεο μπορεί να χαρακτηριστεί ως μια παραμετρική τυχαία διαδικασία. Τα KMM ενδείκνυνται για τη μοντελοποίηση τέτοιων διαδικασιών. Παρουσιάστηκε αναλυτικά, τόσο η διαδικασία στοχαστικής μοντελοποίησης των κινήσεων με χρήση συνεχών KMM, όσο και η διαδικασία μοντελοποίησης των συνεχών διανυσμάτων παρατήρησης μέσω MGM. Τα συνεχή KMM επιλέχθηκαν λόγω α) της ακρίβειας στην ταξινόμηση, κατ'επέκταση και στην αναγνώριση, β) της αποφυγής κβαντοποίησης των δεδομένων, άρα ενδεχόμενης απώλειας σημαντικών πληροφοριών, και τέλος γ) της μη απαίτησης μεγάλου πλήθους ακολουθιών δεδομένων εκπαίδευσης. Παράλληλα, αναπτύχθηκαν ο αλγόριθμος «forward» του Baum, ο αλγόριθμος του Viterbi καθώς και ο αλγόριθμος των Baum-Welch.

Τα MGM στοχεύουν στη δημιουργία ταξινομητών μέσω της εκπαίδευσης μοντέλων. Επιλέχθηκαν διότι χρησιμοποιούν μείγματα Γκαουσιανών κατανομών, τα οποία προσεγγίζουν με πολύ μεγάλη ακρίβεια η πραγματική κατανομή των διανυσμάτων παρατήρησης. Γενικότερα, ο συνδυασμός συνεχών KMM και MGM προσφέρει μεγαλύτερη ελευθερία στην εκπαίδευση του συστήματος αφού επιτρέπουν την είσοδο βίντεο εκπαίδευσης με διαφορετικό μήκος (μη κβαντισμένα) ή ακόμα και βίντεο εκπαίδευσης με ελλιπή δεδομένα.

Στο 7^ο κεφάλαιο παρουσιάστηκε η διεπαφή PianOrasis, η οποία συνδέει τα δύο πρωτότυπα συστήματα. Επίσης, παρουσιάστηκαν οι λόγοι για τους οποίους οι αλγόριθμοι ανίχνευσης δέσματος, ανάκτησης, ταυτοποίησης και εντοπισμού θέσης δακτύλων, καθώς και η προσαρμογή της θεωρίας συνεχών KMM στις ανάγκες του παρόντος συστήματος ικανοποιούν τους στόχους που τέθηκαν στο 1^ο κεφάλαιο. Συμπεραίνεται ότι κατάλληλες μέθοδοι όρασης H/Y μπορούν να χρησιμοποιηθούν ακόμα και σε εξειδικευμένα προβλήματα αναγνώρισης κινήσεων. Ειδικότερα, παρουσιάστηκαν οι λόγοι για τους οποίους η μέθοδος όρασης H/Y που εφαρμόστηκε στη μουσική αλληλεπίδραση, μπορεί να χρησιμοποιηθεί για την επίλυση του προ-

βλήματος ανάκτησης δακτυλισμών και αναγνώρισης δακτυλικών κινήσεων του δεξιού χεριού ενός πιανίστα.

Στο 8^ο κεφάλαιο διεξήχθη μια ανασκόπηση της διδακτορικής διατριβής, παρουσιάστηκαν οι περιορισμοί του συστήματος καθώς και προτάσεις για περαιτέρω βελτιώσεις και μελλοντική έρευνα.

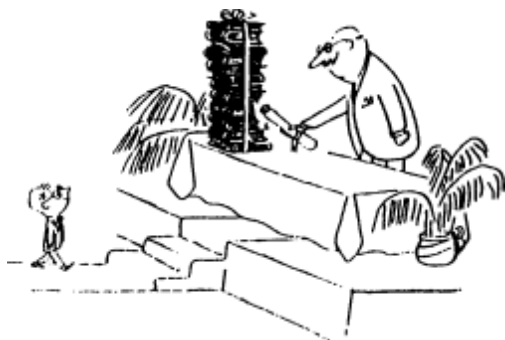
8.5 Συμπεράσματα

Στόχος της όρασης ηλεκτρονικών υπολογιστών είναι ο καθορισμός εκείνων των βημάτων που θα επιτρέψουν στον ηλεκτρονικό υπολογιστή να «κατανοήσει» αυτό που «βλέπει».

Μέσα από την παρούσα διδακτορική διατριβή επιτεύχθηκε ο στόχος της αναγνώρισης ανθρώπινων κινήσεων και χειρονομιών στη μουσική αλληλεπίδραση, δίχως ο άνθρωπος και ο υπολογιστής να έρθουν σε επαφή. Άνθρωπος, μουσικό όργανο και ηλεκτρονικός υπολογιστής συνυπάρχουν αρμονικά χωρίς να επηρεάζεται η φυσική λειτουργία κάποιου μέρους.

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Η υποστήριξη



«Όταν τελειώσω, θα γίνει μια ορχημωσία και θα έρθουν κύριοι πολύ πολύ σπουδαίοι (θα υπάρχουν και άλλα αφεντικά εννοώ) και θα υπάρχει και ένα κύριος πολύ πολύ σπουδαίος που θα μου πει ‘Αγόρι μου, οι

δρόμοι της έρευνας θα είναι λαμπροί για σένα. Είσαι το καμάρι των γονιών σου και η περηφάνια της πατρίδας σου, και όλα αυτά τα χαριτωμένα λόγια. Και μετά θα κάνουμε ένα σούπερ γλέντι με τους φίλους μου. Γιουχουνν!»

«Και όταν ο μπαμπάς μου διαβάσει όλα αυτά στην εφημερίδα, θα είναι πολύ περήφανος και η μαμά μου θα είναι πραγματικά χαρούμενη και θα μου σερβίρει διπλή κρέμα (το αγαπημένο μου επιδόρπιο). Στο τέλος, είναι πραγματικά τέλειο ένα διδακτορικό.»



Βιβλιογραφία

[Βολουδάκης 2005]

Βολουδάκης, Κ. & Μαρμαροκόπος, Κ. (2005). « Μοντελοποίηση και αναγνώριση ανθρώπινων χειρονομιών στην επικοινωνία ανθρώπου – υπολογιστή », Διπλωματική εργασία, Εθνικό Μετσόβειο Πολυτεχνείο, Αθήνα.

[Γεννάτος 2005]

Γεννάτος, Σ. (2005). « Ιστορία του πιάνου », Retrieved January 16, 2009, from: <http://www.xordistis.com/index.files/PianoHistory.pdf/>

[Πολίτης 2007]

Πολίτης, Δ. (2007). « Γλώσσες και διεπαφές στη μουσική πληροφορική », Κλειδάριθμος.

[Alani 1994a]

Alani, T., Guelif, H. (1994). *Modèles de Markov Cachés - Aspects pratiques*. INRIA, France.

[Alani 1994b]

Alani, T. (1994). *Modèles de Markov Cachés - Théorie et techniques de base*. ESIEE, France.

[Albrecht 2003]

Albrecht, I., Haber, J. & Seidel, H. P. (2003). Construction and animation of anatomically based human hand models. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*, San Diego, California, USA.

[Bakis 1976]

Bakis, R. (1976). Continuous speech recognition via centisecond acoustic states. *The Journal of the Acoustical Society of America*, New York, 59(1), 97.

[Baum 1972]

Baum, L. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. In *Proceedings of the Third Symposium on Inequalities*, New York, USA.

[Bellini 1999]

Bellini, P., Fioravanti, F., & Nesi, P. (1999). « Managing music in orchestras », *Computer of the IEEE*, 32(9), 26-34.

[Bérard 2000]

Bérard, F. (2000). « *Vision par Ordinateur pour l'interaction homme-machine fortement couplée* », Thèse de doctorat, Université de Joseph Fourier, Grenoble, France.

[Berndt 1994]

Berndt, D. & Clifford, J. (1994). «Using dynamic time warping to find patterns in time series». In *Proceedings of the Workshop on Knowledge Discovery in Databases*, 77(2), 257-285.

[Black 1996]

Black, A. & Taylor, P. (1996). « Festival Speech Synthesis System: system documentation (1.1.1) ». Rapport Technique HCRC/TR-83, Human Communication Research Center, University of Edinburgh, Royaume Uni.

[Bongers 1999]

Bongers, B. (1999). Exploring Novel Ways of Interaction in Musical Performance. In *Proceedings of the Creativity & Cognition Conference*. Loughborough, United Kingdom.

[Boukir 2004]

Boukir, S. & Chenevière, F. (2004). « Conception d'un système de reconnaissance de gestes dansés ». *Traitement du signal*, 21(3), 195-203.

[Boulic 2000]

Boulic, R., Silaghi, M. & Thalmann, D. (2000). Visualization of Local Movements for Optimal Marker Positioning, *Lecture Notes in Computer Science*, 1899, 133-144.

[Burns 2006]

Burns, A. M. & Wanderley M. (2006). Visual Methods for the Retrieval of Guitarist Fingering. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Paris, France.

[Cadoz 2000]

Cadoz, C., & Wanderley, M. M. (2000). Gesture - music [electronic]. In M. M. Wanderley & M. Battier (Eds.), *Trends in gestural control of music*, 29–65, IRCAM.

[Camurri 2004]

Camurri, A. & Volpe, G. (2004). *Gesture-based Communication in Human-Computer Interaction*. LNAI 2915, Genova, Italy, Springer Verlag.

[Celeux 1992]

Celeux, G., Clairambault, J. (1992). Estimation de chaînes de Markov cachées : méthodes et problèmes. *GDR 134, Traitement du signal et images, Journées thématiques, Approches Markoviennes en signal et images*, 5-19.

[Chomat 2000]

Chomat, O. (2000). « Caractérisation d'éléments d'activités par la statistique conjointe de champs réceptifs », Thèse de doctorat, Institut National Polytechnique de Grenoble.

[Coggan 2005]

Coggan, M. & Ghizaru, A. (2005). *Tracking the human hand*. Retrieved January 16, 2009, from: <http://aghiza.tekahera.com/projects/cs558/>

[Cont 2005]

Cont, A., Schwarz, D. & Schnell, N. (2005). Training IRCAM's score follower, In *Proceedings of the ICASSP 2008 IEEE conference on Acoustics, Speech and Signal Processing*, 3, 253 – 256, Philadelphia, USA.

[Cuen 2003]

Cuen, L., Imberty, M. (2003). La musique comme geste sonore et geste cinétique. In *Abstracts of the 5th Triennial ESCOM Conference*, 187, Hanover, Germany.

[Darrell 1992]

Darrell, T. J. (1992). « Recognition of Space-Time Gestures using a Distributed Representation », Rapport Technique 197, Massachusetts Institute of Technology, Media Laboratory, Perceptual Computing Section, USA.

[Decroux 1994]

Decroux, E. (1994). *Paroles sur le mime*. Librairie Théâtrale : Paris.

[Delalande 1988]

Delalande, F. (1988). La gestique de Gould, éléments pour une sémiologie du geste musical. In G. Guertin (Ed.), *Glenn Gould pluriel* (pp. 85-111). Montréal, Québec, Canada: Louise Courteau.

[Dempster 1977]

Dempster, A. P. (1977). « Maximum Likelihood from incomplete data via the EM algorithm ». *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1-38.

[Drouin 2003]

Drouin, S., Hébert, P. & Parizeau, M. (2003). Simultaneous tracking and estimation of a skeletal model for monitoring human motion. In *Proceedings of Vision Interface Conference*, p. 81-88. Halifax, Canada.

[Exupéry 1943]

Saint Exupéry, A. (1943). *Le Petit Prince*, Folio Junior.

[Ferguson 1980]

Ferguson, J. D. (1980). Variable duration models for speech. In *Proc. Of the Symposium on the Applications of Hidden Markov Models to text and speech*, IDA-CRD, 8-15, Princeton, NJ.

[Gauvain 1994]

Gauvain, J. L., Chini-Hui, L. (1994). « Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains ». *IEEE Transactions on speech and audio proceedings*, 2, 291-228.

[Gilardino 1975]

Gilardino, A. (1975). « Il problema della diteggiatura nelle musiche per chitarra ». *Il Fronimo*, 13, 11-14.

[González 2008]

González, R. C., Woods, R. E. (2008). *Digital Image Processing*. Prentice Hall.

[Grunberg 2008]

Grunberg, D. (2008). *Gesture Recognition for Conducting Computer Music*. Retrieved January 10, 2009, from: <http://schubert.ece.drexel.edu/research/gestureRecognition>

[Hel-Or 2000]

Hel-Or, H. Z. (2000). « Watermarking printed images », Manuscript in préparation.

[Hanon 1923]

Hanon, C. L. (1923). *Le pianiste virtuose*. Bruxelles, Belgique: Editions Schott Frères.

[Heijink 2002]

Heijink, H. & Meulenbroek, R. G. J. (2002). «On the complexity of classical guitar playing: Functional adaptations to task constraints». *Journal of Motor Behaviour*, 34(4), 339-351.

[Hofmann 2001]

Hofmann, P. (2001). *A Vienna Museum For Would-Be Maestros*. Retrieved January 10, 2009, from: <http://query.nytimes.com/gst/fullpage.html?res=950CE0DE113BF931A15754C0A9679C8B63>

[Huang 1988]

Huang, X. D., Jack, M. A. (1988). Semi-Continuous Hidden Markov Models in isolated Word Recognition. In *Proc. IEEE 9th International Conference on Pattern Recognition*, 406-408, Rome, Italy.

[Imberty 1997]

Imberty, M. (1997). Formes de répétition et formes des affects du temps dans l'expression musicale. *Musicae Scientiae*, 1, 33-62.

[Jacobs 2001]

Jacobs, J. P. (2001). Refinements to the ergonomic model for keyboard fingering of Parncutt, Sloboda, Clarke, Raekallio, and Desain. *Music Perception*, 18(4), 505-511.

[Jaëll 1897]

Jaëll, M. (1897). *Le mécanisme du toucher; l'étude du piano par l'analyse expérimentale de la sensibilité tactile*. Armand Colin et Cie, éditeurs.

[Jobbagy 2004]

Jobbagy, A. (2004). « Analysis of finger-tapping movement ». *Journal of Neuroscience Methods*.

[Juang 1985a]

Juang, B. H. (1985). Maximum likelihood estimation for mixture multivariate stochastic observation of Markov chains, *AT&T Tech. Journal*, 64(6), 1235-1249.

[Juang 1985b]

Juang, B. H., Levinson, S. E., Sondhi, M. M. (1985). Mixture autoregressive hidden Markov models for speech signals, *IEEE Trans. Acoust., Speech Signal Processing*, 33(6), 307-309.

[Picard 1997]

Picard, R. W. (1997). *Affective computing*. Massachusetts, United States of America: The MIT Press.

[Kasimi 2007]

Kasimi, A. A., Nichols, E. & Raphael, C. (2007). A Simple Algorithm for Automatic Generation of Polyphonic Piano Fingerings. In *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, Austria.

[Kohler 2004]

Kohler, M. *Vision based hand gesture recognition systems*. Retrieved October 18, 2004, from: <http://ls7-www.cs.unidortmund.de/research/gesture/vbgr-table.html>

[Letessier 2004]

Letessier, J., & Brard, F. (2004). Visual tracking of bare fingers for interactive surfaces. *Seventeenth Annual ACM Symposium on User Interface Software and Technology*, 6(2), 119-122.

[Levinson 1986]

Levinson, S. E. (1986). Continuously Variable Duration Hidden Markov Models for Automatic Speech Recognition, *Computer Speech and Language*, 1, 29-46.

[Liporace 1982]

Liporace, L. R. (1982). Maximum likelihood estimation for multivariate observations of Markov sources, *IEEE Trans. Inform. Theory*, IT-28, 729-734.

[Luengo 2000]

Luengo, C. *DIPimage and DIPlib*. Retrieved March 11, 2009, from: <http://www.diplib.org/home2224>

[Maes 1999]

Maes, A. (1999). « *Le petit Nicolas en thèse* », Retrieved January 7, 2010, from: <http://users.skynet.be/fb819339/nicolas.htm>

[Marrin 1998]

Marrin, T. & Picard, R. (1998). Analysis of Affective Musical Expression with the Conductor's Jacket. In *Proceedings of XII Colloquium for Musical Informatics*. Gorizia, Italy.

[Matthias 2007]

Matthias, R. (2007). Analyse automatique du doigté au piano. In *Proceedings of the Journées d'Informatique Musicale*. Lyon, France.

[Newman 1992]

Newman, W., Wellner, P. (1992). A Desk Supporting Computer-Based Interaction with Paper Document, IN *ACM Conference on Human Factors in Computing Systems*, 587-592.

[Ortmann 1929]

Ortmann, O. (1929). *The Physiological Mechanisms of Piano Technique, an Experimental Study of the Nature of Muscular Action used in Piano Playing, and of the Effects thereof upon the Piano Key and the Piano Tone*. New York: E. P. Dutton & Co, Inc.

[Palmer 1997]

Palmer, C. & Drake, C. (1997). « Monitoring and planning capacities in the acquisition of music performance skills ». *Canadian Journal of Experimental Psychology*, 51, 369-384.

[Palmer 2000]

Palmer, C. & Pfordresher, P. Q. (2000). From my hand to your ear: the faces of meter in performance and perception. In C. Woods, G. Luck, R. Brochard, F. Seddon & J. A. Sloboda (Eds.) *In Proceedings of the 6th International Conference on Music Perception and Cognition*. Keele, UK: Keele University.

[Papamarkos 2000]

Papamarkos, N., Strouthopoulos, C., & Andreadis, I., (2000). "Multithresholding of color and gray level images through a neural network technique", *Image and Vision Computing*, vol. 18, 213-222.

[Parncutt 2002]

Parncutt, R., & Troup, M. (2002). *Science and Psychology of Music Performance: Creative Strategies for Teaching and Learning*, Oxford: Oxford University Press, 285-302.

[Pavlovic 1997]

Pavlovic, V. I., Sharma, R., & Huang, T. S. (1997). Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (7), 677-695.

[Payeur 2006]

Payeur, P., Côte, M. & Comeau, G. (2006). « Les technologies de l'imagerie au service de l'analyse du mouvement en pédagogie du piano », *Revue de recherche en éducation musicale*, 24, 61 – 87.

[Rabiner 1989]

Rabiner, L. R. (1989). «A tutorial on hidden Markov models and selected applications in speech recognition». *Proceedings of the IEEE*, 77(2), 257-285.

[Radicioni 2004a]

Radicioni, D., Anselma, L. & Lombardo, V. (2004a). An Algorithm to compute fingering for string instruments. In *Proceedings of the 2nd national congress of the associazione italiana di scienze cognitive*. Ivrea, Italy.

[Radicioni 2004b]

Radicioni, D., Anselma, L. & Lombardo, V. (2004b). A segmentation-based prototype to compute string instruments fingering. In *Proceedings of the conference on interdisciplinary musicology*. Graz, Austria.

[Radicioni 2005]

Radicioni, D. & Lombardo, V. (2005). Computational model of chord fingering for string instruments. In *Proceedings of the 27th annual conference of the cognitive science society*. Stresa, Italy.

[Ridler 1978]

Ridler, T. & Calvard, S. (1978). «Picture Thresholding Using an Iterative Selection Method». *IEEE Transactions on Systems, Man and Cybernetics*, 8(8), 630-632.

[Sayegh 1989]

Sayegh, S. I. (1989). Fingering for string instruments with the optimum path paradigm. *Computer Music Journal*, 13(3), 76-84.[Schiele 1997]

[Schiele 1997]

Schiele, B. (1997). « *Reconnaissance d'Objets utilisant des Histogrammes Multidimensionnels de Champs Réceptifs* », Thèse de doctorat, Institut National Polytechnique de Grenoble, France.

[Schultz 1949]

Schultz, A. (1949). « *The Riddle of the Pianist's Finger and its Relationship to a Touche-Scheme* », New York: Carl Fischer, USA.

[Seifert 2002]

Seifert, U. (2002). *Bodysynth*. Retrieved January 10, 2009, from: http://www.uni-koeln.de/phil-fak/muwi/hpa7/hpa7_2002/BodySynth_engl.html

[Serra 1998]

Serra, J. (1998). « *Image Analysis and Mathematical Morphology* », *Vol II: Theoretical Advances*, 2, Academic Press.

[Sternberg 1986]

Sternberg, S. (1986). « *Grayscale morphology* ». *Computer Vision, Graphics, and Image Processing*, 35(3), 333-355.

[Stilson 1996]

Stilson, T. (1996). *Biomuse*. Retrieved January 10, 2009, from: <http://ccrma.stanford.edu/CCRMA/Courses/252/sensors/node25.html>

[Störring 1999]

Störring, M., Andersen, H. J. (1999). « *Skin colour detection under changing lighting conditions* ». In *Proceedings of the 7th International Symposium on Intelligent Robotic Systems (SIRS '99)*, 187-195, The University of Coimbra, Portugal.

[Swain 1991]

Swain, M. J. & Ballard, D. H. (1991). « *Color Indexing* ». *International Journal of Computer Vision*, 7(1), 11-32.

[Tobey 1996]

Tobey, F., & Fujinaga, I. (1996). Extraction of conducting gestures in 3d space. In *Proceedings of the International Computer Music Conference*, 305–307, San Francisco, USA.

[Traube 2004]

Traube, C. (2004). *An interdisciplinary study of the timbre of the classical guitar*. Unpublished doctoral dissertation, McGill University.

[Truchet 2004]

Truchet, C. (2004). *Contraintes, recherche locale et composition assistée par ordinateur*. Unpublished doctoral dissertation, Université Paris 7 – Denis Diderot, UFR Informatique.

[Verdière 1999]

Colin de Verdière, V. (1999). « *Représentation et Reconnaissance d'Objets par Champs Réceptifs* ». Thèse de doctorat, Institut National Polytechnique de Grenoble, France.

[Verner 1995]

Verner, J. A. (1995). « MIDI guitar synthesis yesterday, today and tomorrow, an overview of the whole fingerpicking thing », *Recording Magazine*, 8(9), 52-57.

[Vezhnevets 2003]

Vezhnevets, V., Sazonov, V. & Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *Proceedings of Graphicon-2003*, Moscow, Russia.

[Vicon Peak 2005]

Vicon Peak. (2005). *Vicon Motion Capture System*, Lake Forest, CA.

[Vickery 2001]

Vickery, L. (2001). «The Western Edge: some recent electronic music from Western Australia». *Organised Sound*, 6, 69-74.

[Wanderley 2005]

Wanderley, M. M., Vines, B., Middleton, N., McKay, C., & Hatch, W. (2005). « The musical significance of clarinetists' ancillary gestures: An exploration of the field ». *Journal of New Music Research*, 34(1), 97-113.

[Wang 1997]

Wang, H. & Shih-Fu, C. (1997). «A Highly Efficient System for Automatic Face Region Detection in MPEG Video Sequences», *IEEE Trans. on Circuits and Syst. for Video Technology, Special issue on Multimedia Systems and Technologies*, 7(4), 615-628.

[Weismann 1999]

Weismann, J. & Saloman, R. (1999). «Gesture recognition for virtual reality applications using data glove and neural networks». In *Proceedings of the IEEE Conference on Neural Networks*, Washington, USA.

[Wellner 1991a]

Wellner, P. (1991). « The DigitalDesk Calculator: Tactile manipulation on a Desktop », In *ACM Symposium on User Interface Software and Technology*, 27-33.

[Wellner 1991b]

Wellner, P. (1991). « Interacting with Paper on the DigitalDesk », Video.

[Willshaw 1990]

Willshaw, D. J. & Buckingham, J. T. (1990). « An assessment of Marr's theory of the hippocampus as a temporary memory store. », In *Proceedings of the Royal Society of London B*, 329, 205-215.

[Yang 1997]

Yang, J., Lu, W., & Waibel, A. (1997). “Skin-color modeling and adaptation”. *Lecture Notes in Computer Science*, Springer.

[Yörük 2006]

Yörük, E., Dutagãci, H., & Sankur, B. (2006). Hand biometrics. *Image and Vision Computing*, 24(5), 483-497.

Παράρτημα Α : Κώδικας PianOrasis

```
function varargout = PianOrasis(varargin)
% PIANORASIS M-file for PianOrasis.fig
%   PIANORASIS, by itself, creates a new PIANORASIS or raises the
existing
%   singleton*.
%
%   H = PIANORASIS returns the handle to a new PIANORASIS or the han-
dle to
%   the existing singleton*.
%
%   PIANORASIS('CALLBACK',hObject,eventData,handles,...) calls the
local
%   function named CALLBACK in PIANORASIS.M with the given input ar-
guments.
%
%   PIANORASIS('Property','Value',...) creates a new PIANORASIS or
raises the
%   existing singleton*. Starting from the left, property value
pairs are
%   applied to the GUI before PianOrasis_OpeningFcn gets called. An
%   unrecognized property name or invalid value makes property appli-
cation
%   stop. All inputs are passed to PianOrasis_OpeningFcn via varar-
gin.
%
%   *See GUI Options on GUIDE's Tools menu. Choose "GUI allows only
one
%   instance to run (singleton)".
%
% See also: GUIDE, GUIDATA, GUIHANDLES

% Edit the above text to modify the response to help PianOrasis

% Last Modified by GUIDE v2.5 06-Dec-2009 11:01:01

% Begin initialization code - DO NOT EDIT
gui_Singleton = 1;
gui_State = struct('gui_Name',       mfilename, ...
                  'gui_Singleton',   gui_Singleton, ...
                  'gui_OpeningFcn', @PianOrasis_OpeningFcn, ...
                  'gui_OutputFcn',  @PianOrasis_OutputFcn, ...
                  'gui_LayoutFcn',  [] , ...
                  'gui_Callback',    []);
if nargin && ischar(varargin{1})
    gui_State.gui_Callback = str2func(varargin{1});
end

if nargout
    [varargout{1:nargout}] = gui_mainfcn(gui_State, varargin{:});
else
    gui_mainfcn(gui_State, varargin{:});
end
```

```

end
% End initialization code - DO NOT EDIT

% --- Executes just before PianOrasis is made visible.
function PianOrasis_OpeningFcn(hObject, eventdata, handles, varargin)
% This function has no output args, see OutputFcn.
% hObject    handle to figure
% eventdata  reserved - to be defined in a future version of MATLAB
% handles     structure with handles and user data (see GUIDATA)
% varargin   command line arguments to PianOrasis (see VARARGIN)

% Choose default command line output for PianOrasis
handles.output = hObject;

% Update handles structure
guidata(hObject, handles);

% UIWAIT makes PianOrasis wait for user response (see UIRESUME)
% uiwait(handles.figure1);
dip_initialise;

addpath(genpath('C:\Program Files\MATLAB\HMMall'));

global open; global min1; global min2; global gauss1;
global gauss2; global canny_s; global canny_l; global canny_u; global
edges;
open=4;
%min1=8;
min1=9;
gauss1=4;
%gauss1=10;
min2=10;
%gauss2=5;
gauss2=4;
canny_s=1;
canny_l=0.5;
canny_u=0.9;
%edges=20;
edges=4;

global pos1; global pos2; global pos3; global pos4; global pos5;
mean=198;
thres=5;
pos5=248-mean+thres;
pos4=276-mean+thres;
pos3=282-mean+thres;
pos2=280-mean+thres;
pos1=276-mean+thres;

set(handles.text34,'string',num2str(pos1));
set(handles.text33,'string',num2str(pos2));
set(handles.text31,'string',num2str(pos3));
set(handles.text30,'string',num2str(pos4));

```

```

set(handles.text29,'string',num2str(pos5));

global Init_pos;
Init_pos=[pos1 pos2 pos3 pos4 pos5];
global State_vector;
State_vector=[pos1 pos2 pos3 pos4 pos5];

global prefix; global frame_no;
prefix='mozart';
frame_no='0';

global open_file;
open_file=0;

global missing_fingers;
missing_fingers=0;

global show_colors;
show_colors=0;

init_pcks_x=[263 339 396 464 499];
init_e_y=[78961 80089 82369 76176 62001];
global x_coord;
global y_coord;
x_coord=init_pcks_x;
y_coord=init_e_y;
global Centers_of_Classes;
Centers_of_Classes=[init_pcks_x' init_e_y'];
global Samples;
Samples=[init_pcks_x' init_pcks_x' init_pcks_x'];
global Sample_y
Sample_y=[init_e_y' init_e_y' init_e_y'];

%HMM
global data;
data=[];
global data_f;
global num_of_frames;
num_of_frames=1;
global num_of_video;
num_of_video=1;
global prior1;
global transmat1;
global mul;
global Sigma1;
global mixmat1;
global data_HMM;

% --- Outputs from this function are returned to the command line.
function varargout = PianOrasis_OutputFcn(hObject, eventdata, handles)
% varargout cell array for returning output args (see VARARGOUT);
% hObject handle to figure
% eventdata reserved - to be defined in a future version of MATLAB
% handles structure with handles and user data (see GUIDATA)

```

```

% Get default command line output from handles structure
varargout{1} = handles.output;
global data
data=[];

%*****
***
%***** MENU *****

% -----
function File_Callback(hObject, eventdata, handles)
% hObject    handle to File (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% -----
function Edit_Callback(hObject, eventdata, handles)
% hObject    handle to Edit (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% -----
function About_Callback(hObject, eventdata, handles)
% hObject    handle to About (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% -----
function Save_Results_Callback(hObject, eventdata, handles)
% hObject    handle to Save_Results (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
global L; global g; global bwcontour; global Per;
imwrite(L, 'output/hand.jpg');
imwrite(g, 'output/filtering.jpg');
imwrite(bwcontour, 'output/contour.jpg');
imwrite(Per, 'output/per.jpg');

% -----
function Open_Log_File_Callback(hObject, eventdata, handles)
% hObject    handle to Open_Log_File (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
global open_file;
if open_file==0
    open_file=1;
    set(handles.Open_Log_File, 'Checked', 'on');
else
    open_file=0;
    set(handles.Open_Log_File, 'Checked', 'off');

```



```

end

% -----
function Open_Image_Callback(hObject, eventdata, handles)
% hObject      handle to Open_Image (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global file;
global pathname; global filename;
[filename, pathname] = uigetfile({'*.bmp'; '*.jpg'; '*.tif'; '*.png'; '*.*'}, 'Open Image');
if (filename)
    file=[pathname filename];
    set(handles.text1, 'string', file);
    A=imread(file);
    axes(handles.axes1);
    imshow(A);
end
% -----

function Frame_Sequence_Callback(hObject, eventdata, handles)
% hObject      handle to Frame_Sequence (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global prefix; global frame_no;
prompt={'Enter the prefix:', 'Enter the number of frames:',};
name='Sequence';
numlines=1;
defaultanswer={'mozart', '5'};
options.Resize='on';
options.WindowStyle='normal';
options.Interpreter='tex';
answer=inputdlg(prompt, name, numlines, defaultanswer, options);
prefix=answer{1};
frame_no=answer{2};

% -----
function Exit_Callback(hObject, eventdata, handles)
% hObject      handle to Exit (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
h=questdlg('Exit from program');
switch h
    case 'Yes'
        closereq;
end

%***** End of MENU *****

% --- RUN FOR A SINGLE FRAME *****

function pushbutton1_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton1 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB

```

```

% handles      structure with handles and user data (see GUIDATA)
global file;
global open_file;
global filename;
global Init_pos;
global State_vector
global pos1, global pos2; global pos3; global pos4; global pos5;
global missing_fingers;

H=execution_function(file,handles);

if missing_fingers==1
State_vector=[pos1 pos2 pos3 pos4 pos5];
end
Init_pos=round(Init_pos);
H=round(H);
if (open_file==1)
    fout=fopen(['output/' filename '_log.txt'],'w');
    % fprintf(fout,'Initial Positions \r\n');
    % fprintf(fout,'%d %d %d %d %d \r\n',Init_pos(1),Init_pos(2),Init_pos(3),Init_pos(4),Init_pos(5));
    % fprintf(fout,'-----\r\n');
    % fprintf(fout,'%s\r\n',filename);

    str=[];
    if (H(1)>Init_pos(1))
        str=[str 5];
    end
    if (H(2)>Init_pos(2))
        str=[str 4];
    end
    if (H(3)>Init_pos(3))
        str=[str 3];
    end
    if (H(4)>Init_pos(4))
        str=[str 2];
    end
    if (H(5)>Init_pos(5))
        str=[str 1];
    end
    fprintf(fout,'%d ',str);
    fclose(fout);
end

function edit2_Callback(hObject, eventdata, handles)
% hObject      handle to edit2 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit2 as text
%         str2double(get(hObject,'String')) returns contents of edit2 as
a double

```

```

% --- Executes during object creation, after setting all properties.
function edit2_CreateFcn(hObject, eventdata, handles)
% hObject    handle to edit2 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function edit6_Callback(hObject, eventdata, handles)
% hObject    handle to edit6 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit6 as text
%        str2double(get(hObject,'String')) returns contents of edit6 as
a double

% --- Executes during object creation, after setting all properties.
function edit6_CreateFcn(hObject, eventdata, handles)
% hObject    handle to edit6 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function edit7_Callback(hObject, eventdata, handles)
% hObject    handle to edit7 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit7 as text
%        str2double(get(hObject,'String')) returns contents of edit7 as
a double

% --- Executes during object creation, after setting all properties.
function edit7_CreateFcn(hObject, eventdata, handles)
% hObject    handle to edit7 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB

```

```

% handles      empty - handles not created until after all CreateFcns cal-
led

% Hint: edit controls usually have a white background on Windows.
%      See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function edit8_Callback(hObject, eventdata, handles)
% hObject      handle to edit8 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit8 as text
%      str2double(get(hObject,'String')) returns contents of edit8 as
a double

% --- Executes during object creation, after setting all properties.
function edit8_CreateFcn(hObject, eventdata, handles)
% hObject      handle to edit8 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      empty - handles not created until after all CreateFcns cal-
led

% Hint: edit controls usually have a white background on Windows.
%      See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function edit9_Callback(hObject, eventdata, handles)
% hObject      handle to edit9 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit9 as text
%      str2double(get(hObject,'String')) returns contents of edit9 as
a double

% --- Executes during object creation, after setting all properties.
function edit9_CreateFcn(hObject, eventdata, handles)
% hObject      handle to edit9 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      empty - handles not created until after all CreateFcns cal-
led

% Hint: edit controls usually have a white background on Windows.

```

```

%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function edit10_Callback(hObject, eventdata, handles)
% hObject      handle to edit10 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit10 as text
%         str2double(get(hObject,'String')) returns contents of edit10 as
a double

% --- Executes during object creation, after setting all properties.
function edit10_CreateFcn(hObject, eventdata, handles)
% hObject      handle to edit10 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

function edit11_Callback(hObject, eventdata, handles)
% hObject      handle to edit11 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit11 as text
%         str2double(get(hObject,'String')) returns contents of edit11 as
a double

% --- Executes during object creation, after setting all properties.
function edit11_CreateFcn(hObject, eventdata, handles)
% hObject      handle to edit11 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

```

```
end
```

```
function edit14_Callback(hObject, eventdata, handles)
% hObject      handle to edit14 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit14 as text
%        str2double(get(hObject,'String')) returns contents of edit14 as
a double
```

```
% --- Executes during object creation, after setting all properties.
function edit14_CreateFcn(hObject, eventdata, handles)
% hObject      handle to edit14 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
```

```
function edit15_Callback(hObject, eventdata, handles)
% hObject      handle to edit15 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit15 as text
%        str2double(get(hObject,'String')) returns contents of edit15 as
a double
```

```
% --- Executes during object creation, after setting all properties.
function edit15_CreateFcn(hObject, eventdata, handles)
% hObject      handle to edit15 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%       See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end
```

```
% --- Executes on button press in pushbutton2.
```

```

function pushbutton2_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton2 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global open; global min1; global min2; global gauss1;
global gauss2; global canny_s; global canny_l; global canny_u; global
edges;
open=str2num(get(handles.edit2,'string'));
min1=str2num(get(handles.edit6,'string'));
gauss1=str2num(get(handles.edit8,'string'));
min2=str2num(get(handles.edit7,'string'));
gauss2=str2num(get(handles.edit9,'string'));
canny_s=str2num(get(handles.edit15,'string'));
canny_l=str2num(get(handles.edit14,'string'));
canny_u=str2num(get(handles.edit11,'string'));
edges=str2num(get(handles.edit10,'string'));

% --- Executes on button press in pushbutton3.
function pushbutton3_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton3 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global open; global min1; global min2; global gauss1;
global gauss2; global canny_s; global canny_l; global canny_u; global
edges;

set(handles.edit2,'string','4');
set(handles.edit6,'string','8');
set(handles.edit8,'string','10');
set(handles.edit7,'string','10');
set(handles.edit9,'string','5');
set(handles.edit15,'string','1');
set(handles.edit14,'string','0.5');
set(handles.edit11,'string','0.9');
set(handles.edit10,'string','20');

open=str2num(get(handles.edit2,'string'));
min1=str2num(get(handles.edit6,'string'));
gauss1=str2num(get(handles.edit8,'string'));
min2=str2num(get(handles.edit7,'string'));
gauss2=str2num(get(handles.edit9,'string'));
canny_s=str2num(get(handles.edit15,'string'));
canny_l=str2num(get(handles.edit14,'string'));
canny_u=str2num(get(handles.edit11,'string'));
edges=str2num(get(handles.edit10,'string'));

% --- Executes on button press in pushbutton4.
function pushbutton4_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton4 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global Init_pos;
global Centers_of_Classes;

```

```

global Samples;
global Sample_y;
global x_coord;
global y_coord;
global pos1, global pos2; global pos3; global pos4; global pos5;
pos5=str2num(get(handles.text35,'string'));
pos4=str2num(get(handles.text36,'string'));
pos3=str2num(get(handles.text37,'string'));
pos2=str2num(get(handles.text38,'string'));
pos1=str2num(get(handles.text39,'string'));
Init_pos=[pos1 pos2 pos3 pos4 pos5];

set(handles.text34,'string',num2str(pos1));
set(handles.text33,'string',num2str(pos2));
set(handles.text31,'string',num2str(pos3));
set(handles.text30,'string',num2str(pos4));
set(handles.text29,'string',num2str(pos5));

init_pcks_x=x_coord;
init_e_y=y_coord;
Centers_of_Classes=[init_pcks_x' init_e_y']
Samples=[init_pcks_x' init_pcks_x' init_pcks_x'];
Sample_y=[init_e_y' init_e_y' init_e_y'];

% --- RUN FOR A FRAME SEQUENCE
*****
function pushbutton5_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton5 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global prefix; global frame_no; global pathname;
global filename;
global Init_pos;
global open_file;
global State_vector;
global pos1, global pos2; global pos3; global pos4; global pos5;
global file;

y=length(prefix);
fnumber=[filename(y+1) filename(y+2) filename(y+3) filename(y+4)];
start=str2num(fnumber);

%HMM
global num_of_frames;

%val=get(handles.popupmenu1,'value');
%if (val==2)
%    num_of_frames=1;
%end;

```



```

if (open_file==1)
    Init_pos=round(Init_pos);
    fout=fopen(['output/' filename '_plus_' frame_no '_log.txt'],'w');
    % fprintf(fout,'Initial Positions \r\n');
    % fprintf(fout,'%d %d %d %d %d
\r\n',Init_pos(1),Init_pos(2),Init_pos(3),Init_pos(4),Init_pos(5));
    % fprintf(fout,'-----\r\n');
end

for i=start:start+str2num(frame_no);
    f_no=i;

    if (f_no<10)
        f=['000' num2str(f_no)];
    elseif (f_no<100)
        f=['00' num2str(f_no)];
    elseif (f_no<1000)
        f=['0' num2str(f_no)];
    end
    filename2=[pathname prefix f '.bmp'];
    set(handles.text1,'string', filename2);
    I=imread(filename2);
    axes(handles.axes1);
    imshow(I);
    H=execution_function(filename2,handles);
    H=round(H);
    if (open_file==1)
        ff=[prefix f '.bmp'];
        % fprintf(fout,'%s \r\n',ff);
        % fprintf(fout,'%d %d %d %d %d\r\n',H(1),H(2),H(3),H(4),H(5));
        % fprintf(fout,'-----
\r\n');

        str=[];
        finger=0;
        if (length(H)>3)
            if (H(1)>Init_pos(1))
                str=[str 5];
                finger=1;
            end
            if (H(2)>Init_pos(2))
                str=[str 4];
                finger=1;
            end
            if (H(3)>Init_pos(3))
                str=[str 3];
                finger=1;
            end
            if (H(4)>Init_pos(4))
                str=[str 2];
                finger=1;
            end
            if (H(5)>Init_pos(5))
                str=[str 1];
                finger=1;
            end
        end
    end
end

```

```

        if finger==1
            fprintf(fout,'%d ',str);
        end
    end %length

end

end

if (open_file==1)
    fclose(fout);
end
State_vector=[pos1 pos2 pos3 pos4 pos5];
file=filename2;

function [H]=execution_function(file,handles)
global open; global min1; global min2; global gauss1;
global gauss2; global canny_s; global canny_l; global canny_u; global
edges;
global pos1; global pos2; global pos3; global pos4; global pos5;
global L; global g; global bwcontour; global Per;
global State_vector;
global missing_fingers;
global show_colors;
global x_coord;
global y_coord;

%HMM
global num_of_frames;
global data_f;
global prior1;
global transmat1;
global mul;
global Sigma1;
global mixmat1;
global data_HMM;

r_range =[0.34 0.62];
g_range=[0.25 0.42];
[ H L g bwcontour Per y e pcks] = extract_features( fi-
le,r_range,g_range,open,min1,min2,gauss1,gauss2,canny_s,canny_l,canny_u,
edges);
x_coord=pcks;
y_coord=e;

axes(handles.axes2);
imshow(L);
axes(handles.axes3);
g=double(g);
imshow(g);
axes(handles.axes4);
imshow(bwcontour);
axes(handles.axes5);
imshow(Per);

```

```

%axes(handles.axes6);
%plot(y);
%hold on
%plot(pcks,e,'or');
%hold off

if (missing_fingers==1)
    %Estimated_H=compare_function(State_vector,H);
    %State_vector=Estimated_H;
    [Estimated_H,point]=Classifier(pcks,e,H);
    axes(handles.axes6);
    plot(y);
    hold on
    plot(pcks,e,'or');
    hold on
    plot(point(:,1),point(:,2),'gs');
    hold off

    %Hidden Markov Models
    val=get(handles.popupmenu1,'value');
    switch val
        case 1
            %HMM is Inactive
            num_of_frames=1;
        case 2
            set(handles.text44,'string',num2str(num_of_frames));
            featu-
res=[point(1,1);Estimated_H(1);point(2,1);Estimated_H(2);point(3,1);Esti
ma-
ted_H(3);point(4,1);Estimated_H(4);point(5,1);Estimated_H(5);point(2,1)-
point(1,1);point(3,1)-point(2,1);point(4,1)-point(3,1);point(5,1)-
point(4,1)];
            % features=[point(1,1) point(1,2) point(2,1) point(2,2)
point(3,1) point(3,2) point(4,1) point(4,2) point(5,1) point(5,2)
point(2,1)-point(1,1) point(3,1)-point(2,1) point(4,1)-point(3,1)
point(5,1)-point(4,1)];
            data_f(:,num_of_frames)=features;
            num_of_frames=num_of_frames+1;
        case 3
            %HMM run
            set(handles.text44,'string',num2str(num_of_frames));

f_=[point(1,1);Estimated_H(1);point(2,1);Estimated_H(2);point(3,1);Estim
ated_H(3);point(4,1);Estimated_H(4);point(5,1);Estimated_H(5);point(2,1)
-point(1,1);point(3,1)-point(2,1);point(4,1)-point(3,1);point(5,1)-
point(4,1)];
            data_HMM(:,num_of_frames)=f_;
            num_of_frames=num_of_frames+1;

    end

else
    Estimated_H=H;
    num_of_maxima=length(H);
    for i=num_of_maxima+1:5

```

```

        Estimated_H(i)=0;
    end
    axes(handles.axes6);
    plot(y);
    hold on
    plot(pcks,e,'or');
    hold off
end

set(handles.text39,'string',num2str(Estimated_H(1)));
set(handles.text38,'string',num2str(Estimated_H(2)));
set(handles.text37,'string',num2str(Estimated_H(3)));
set(handles.text36,'string',num2str(Estimated_H(4)));
set(handles.text35,'string',num2str(Estimated_H(5)));
if (show_colors==1)

    if (pos1+8>=Estimated_H(1))
        %if (pos1>=Estimated_H(1))
            set(handles.text39,'BackgroundColor',[0.847 0.161 0]);
        else
            set(handles.text39,'BackgroundColor',[0 0.5 0]);
        end
    if (pos2+15>=Estimated_H(2))
        %if (pos2>=Estimated_H(2))
            set(handles.text38,'BackgroundColor',[0.847 0.161 0]);
        else
            set(handles.text38,'BackgroundColor',[0 0.5 0]);
        end
    if (pos3+12>=Estimated_H(3))
        %if (pos3>=Estimated_H(3))
            set(handles.text37,'BackgroundColor',[0.847 0.161 0]);
        else
            set(handles.text37,'BackgroundColor',[0 0.5 0]);
        end
    %if (pos4>=Estimated_H(4))
    if (pos4+12.5>=Estimated_H(4))
        set(handles.text36,'BackgroundColor',[0.847 0.161 0]);
    else
        set(handles.text36,'BackgroundColor',[0 0.5 0]);
    end
    if (pos5+11>=Estimated_H(5))
        %if (pos5>=Estimated_H(5))
            set(handles.text35,'BackgroundColor',[0.847 0.161 0]);
        else
            set(handles.text35,'BackgroundColor',[0 0.5 0]);
        end
    else
        set(handles.text35,'BackgroundColor',[0.847 0.161 0]);
        set(handles.text36,'BackgroundColor',[0.847 0.161 0]);
        set(handles.text37,'BackgroundColor',[0.847 0.161 0]);
        set(handles.text38,'BackgroundColor',[0.847 0.161 0]);
        set(handles.text39,'BackgroundColor',[0.847 0.161 0]);
    end
end
H=Estimated_H;

```

```

function [Estimated_H,point]=Classifier(pcks,e,H)
global Centers_of_Classes;
global Samples;
global Sample_y;
global Init_pos;

point=zeros(5,2);
Estimated_H=zeros(5,1);
div=zeros(5,1);
    num_of_observations=length(pcks);
    if (num_of_observations==5)
        Estimated_H=H;
        point(:,1)=pcks;
        point(:,2)=e;
    else

        for i=1:num_of_observations
            dmin=1000000;
            for j=1:5
                d=abs(Centers_of_Classes(j,1)-pcks(i));
                if (d<dmin)
                    dmin=d;
                    class=j;
                end
            end

            if (point(class,1)==0)
                point(class,1)=pcks(i);
                point(class,2)=e(i);
                Estimated_H(class)=H(i);
            else
                %an yparxei idi kataxorisi tote bale aytin me ti mikro-
                %apostash
                if (dmin<(abs(Centers_of_Classes(class,1)-
point(class,1))));
                    point(class,1)=pcks(i);
                    point(class,2)=e(i);
                    Estimated_H(class)=H(i);
                end
            end
        end

    end

    %gemizw tis theseis toy pinaka point poy exoyn meinei kenes kai kanw
    %update ston Sample
    for k=1:5
        if point(k,1)==0
            point(k,:)=Centers_of_Classes(k,:);
            Estimated_H(k)=Init_pos(k);
        else % oi pinakes poy kratoun tis times gia x kai y apo t ews t-3
            Samples(k,1)=Samples(k,2);
            Samples(k,2)=Samples(k,3);
            Samples(k,3)=point(k,1);
        end
    end
end

```

```

        Sample_y(k,1)=Sample_y(k,2);
        Sample_y(k,2)=Sample_y(k,3);
        Sample_y(k,3)=point(k,2);
    end
end

%ypologizw twra ta nea kentra tvn klasewn kai thn apoklisi
for k=1:5
    Cen=
ters_of_Classes(k,1)=(Samples(k,1)+Samples(k,2)+Samples(k,3))/3;
    Cen=
ters_of_Classes(k,2)=(Sample_y(k,1)+Sample_y(k,2)+Sample_y(k,3))/3;
    % for i=1:3
    % div(k)= (1/3)*(Centers_of_Classes(k,1)-Samples(k,i))^2;
    % end
end

function [Estimated_H]=compare_function(Previous_state,H)
global Init_pos;
count=1;
num_of_maxima=length(H);
Estimated_H = zeros(1,5);
if(num_of_maxima==5)
    Estimated_H =H;
else
    for i=1:5
        %Elegxw an exv antistoixisei ola ta megista
        %an nai tote ta ypoloipa daxyta diathroyn tis
        %prohgoymenes times.
        if count<=num_of_maxima
            if abs(Previous_state(i)-H(count))<10
                Estimated_H(1,i)=H(count);
                count=count+1;
            else
                % Estimated_H(1,i)=Previous_state(i);
                Estimated_H(1,i)=Init_pos(i);
            end
        else
            %Estimated_H(1,i)=Previous_state(i);
            Estimated_H(1,i)=Init_pos(i);
        end
    end
end
end

% --- Executes on button press in checkbox1.
function checkbox1_Callback(hObject, eventdata, handles)
% hObject    handle to checkbox1 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of checkbox1
global missing_fingers

```

```

if get(hObject,'Value')==1
    missing_fingers=1;
else
    missing_fingers=0;
end

% --- Executes on button press in checkbox2.
function checkbox2_Callback(hObject, eventdata, handles)
% hObject      handle to checkbox2 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hint: get(hObject,'Value') returns toggle state of checkbox2
global show_colors
if get(hObject,'Value')==1
    show_colors=1;
else
    show_colors=0;
end

% --- Executes on selection change in popupmenu1.
function popupmenu1_Callback(hObject, eventdata, handles)
% hObject      handle to popupmenu1 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: contents = get(hObject,'String') returns popupmenu1 contents as
cell array
%      contents{get(hObject,'Value')} returns selected item from po-
pupmenu1

% --- Executes during object creation, after setting all properties.
function popupmenu1_CreateFcn(hObject, eventdata, handles)
% hObject      handle to popupmenu1 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      empty - handles not created until after all CreateFcns cal-
led

% Hint: popupmenu controls usually have a white background on Windows.
%      See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

% --- Executes on button press in pushbutton6.
function pushbutton6_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton6 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global data;
global data_f;

```

```

global num_of_video;
if (data_f)
    set(handles.text42,'string',num2str(num_of_video));
    data{num_of_video}(:,:)=data_f;

    %   data(:, :, num_of_video)=data_f;
    %temp{:, :}=data_f;
    %data=cat(3,data,temp);
    num_of_video=num_of_video+1;
    data_f
    data_f=[];
    data
end

% --- Executes on button press in pushbutton7.
function pushbutton7_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton7 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global data;
global data_f;
global num_of_video;
global num_of_frames;
num_of_video=1;
num_of_frames=1;
data=[];
data_f=[];
set(handles.text46,'string','0');
set(handles.text42,'string','0');
set(handles.text44,'string','0');

% --- Executes on button press in pushbutton8.
function pushbutton8_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton8 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global data;
global prior1;
global transmat1;
global mul;
global Sigma1;
global mixmat1;

O = 14;
nex = 10;
T=51;
M = 2;
Q = 4;

%data=randn(14,8,2)

%par=randn(6,23);
%data{1}(:,:)=par;

```



```

%data{2}(:, :)=par;
%data{3}(:, :)=par;
%data{4}(:, :)=par;
%data{5}(:, :)=par;
%data{6}(:, :)=par;
%data{7}(:, :)=par;

%data=data/1000

Q=double(Q);
cov_type = 'full';
%prior0 = [0.6;0.1;0.05;0.05;0.05;0.05;0.05;0.05];
prior0 = [0.6;0.1;0.1;0.2];
%prior0 = [0.7;0.1;0.1];
transmat0 = mk_leftright_transmat(Q,0.3);
[mu0, Sigma0] = mixgauss_init(Q*M, data, cov_type);
mu0 = reshape(mu0, [O Q M]);
Sigma0 = reshape(Sigma0, [O O Q M]);
mixmat0 = mk_stochastic(rand(Q,M));
[LL, prior1, transmat1, mul, Sigmal, mixmat1]=mhmm_em(data, prior0,
transmat0, mu0, Sigma0, mixmat0, 'max_iter', 10, 'thresh', 0.001);
save HMM/prior1.mat prior1
save HMM/transmat1.mat transmat1
save HMM/mul.mat mul
save HMM/Sigmal.mat Sigmal
save HMM/mixmat1.mat mixmat1

% --- Executes on button press in pushbutton9.
function pushbutton9_Callback(hObject, eventdata, handles)
% hObject      handle to pushbutton9 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)
global prior1;
global transmat1;
global mul;
global Sigmal;
global mixmat1;
global data_HMM;
%data_HMM=randn(14,8,1);
data_HMM

hmm_nums=str2num(get(handles.edit16,'string'));

probs=zeros(hmm_nums,1);
for i=1:hmm_nums
    file=['HMM_' num2str(i)];
    prior1=load([file '/prior1.mat']);
    prior1=prior1.prior1;
    transmat1=load([file '/transmat1.mat']);
    transmat1=transmat1.transmat1;
    mul=load([file '/mul.mat']);
    mul=mul.mul;
    Sigmal=load([file '/Sigmal.mat']);
    Sigmal=Sigmal.Sigmal;
    mixmat1=load([file '/mixmat1.mat']);

```

```

mixmat1=mixmat1_.mixmat1;

loglik2 = mhmm_logprob(data_HMM, prior1, transmat1, mul, Sigma1,
mixmat1);
set(handles.text46,'string',num2str(loglik2));
probs(i)=loglik2;

prior1_=[];
transmat1_=[];
mul_=[];
Sigma1_=[];
mixmat1_=[];
prior1=[];
transmat1=[];
mul=[];
Sigma1=[];
mixmat1=[];

end
save probs.txt probs -ascii

function edit16_Callback(hObject, eventdata, handles)
% hObject      handle to edit16 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      structure with handles and user data (see GUIDATA)

% Hints: get(hObject,'String') returns contents of edit16 as text
%          str2double(get(hObject,'String')) returns contents of edit16 as
a double

% --- Executes during object creation, after setting all properties.
function edit16_CreateFcn(hObject, eventdata, handles)
% hObject      handle to edit16 (see GCBO)
% eventdata    reserved - to be defined in a future version of MATLAB
% handles      empty - handles not created until after all CreateFcns called

% Hint: edit controls usually have a white background on Windows.
%          See ISPC and COMPUTER.
if ispc && isequal(get(hObject,'BackgroundColor'),
get(0,'defaultUicontrolBackgroundColor'))
    set(hObject,'BackgroundColor','white');
end

```

Παράρτημα Β

« Το τριαντάφυλλο του μικρού πρίγκηπα »

[Exupéry 1943]

(...) Ο μικρός πρίγκηπας πήγε και ξαναείδε τα τριαντάφυλλα:

- Δεν μοιάζετε καθόλου με το δικό μου τριαντάφυλλο, εσείς δεν είσαστε ακόμα τίποτα, τους είπε. Κανένas δε σας ημέρωσε και εσείς δεν ημερώσατε κανένα. Είσαστε όπως ήταν η αλεπού μου . Ήταν μια αλεπού όμοια μ'εκατό χιλιάδες άλλες . Γίναμε όμως φίλοι, και τώρα είναι μοναδική στον κόσμο.

Και τα τριαντάφυλλα δεν είχαν μούτρα να σταθούν.

- Είσαστε όμορφα, μα είσαστε άδεια, τους είπε ακόμη. Δεν θα μπορούσε κανείς να πεθάνει για το χατήρι σας. Βέβαια, ένας κοινός διαβάτης, βλέποντας το δικό μου τριαντάφυλλο θα νόμιζε πως σας μοιάζει. Μα, από μόνο του αυτό, είναι πιο σημαντικό από όλα εσάς, αφού είναι αυτό που πότισα. Αφού είναι αυτό που έβαλα κάτω από τη γυάλα. Αφού είναι αυτό που προστάτεψα με το παραβάν. Αφού είναι αυτό που του σκότωσα τις κάμπιες (εκτός από δυο ή τρεις που τις άφησα για να γίνουν πεταλούδες). Αφού είναι αυτό που το άκουσα να παραπονιέται ή να κομπάζει ή καμιά φορά και να σωπαίνει. Αφού αυτό είναι το δικό μου τριαντάφυλλο.

Και ξαναγύρισε στην αλεπού.

- Αντίο, της είπε...

- Αντίο, είπε η αλεπού. Να το μυστικό μου. Είναι πολύ απλό:

Μόνο με την καρδιά βλέπεις καλά.

Την ουσία δεν τη βλέπουν τα μάτια.

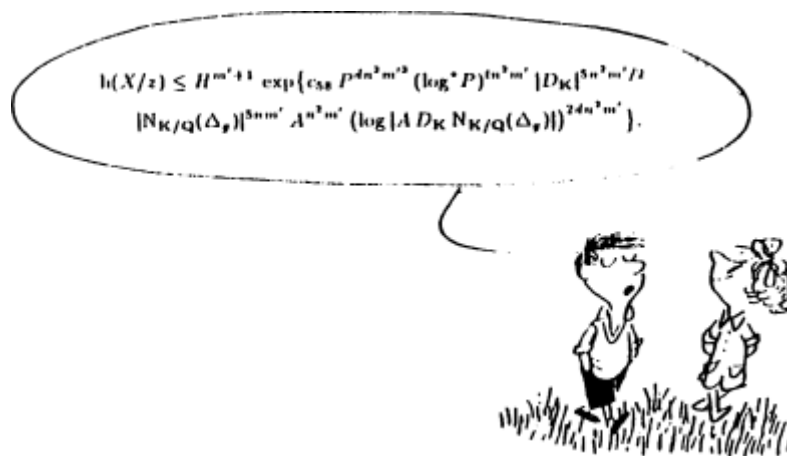
- Την ουσία δεν τη βλέπουν τα μάτια, επανέλαβε ο μικρός πρίγκιπας, για να το θυμάται.
- Ο χρόνος που έχεις χάσει για το τριαντάφυλλό σου είναι που το κάνει τόσο σημαντικό.
- Είναι ο χρόνος που έχω χάσει για το τριαντάφυλλό μου... έκανε ο μικρός πρίγκιπας, για να το θυμάται.
- Οι άνθρωποι έχουν ξεχάσει τούτη την αλήθεια, είπε η αλεπού. Όμως εσύ δεν πρέπει να την ξεχάσεις. Να γίνεις υπεύθυνος για πάντα εκείνου που έχεις εξημερώσει. Είσαι υπεύθυνος για το τριαντάφυλλό σου...
- Είμαι υπεύθυνος για το τριαντάφυλλό μου... ξαναείπε ο μικρός πρίγκιπας, για να το θυμάται.

Ο μικρός πρίγκηπας
του Saint-Exupéry

Ο μικρός Νικόλας στο διδακτορικό [Maes 1999]

Η δόξα

«Επίσης, κάνει ωραία εντύπωση στα κορίτσια όταν ξέρουν ότι έχεις κάνει διδακτορικό στα μαθηματικά και ότι βρήκες πολλά θεωρήματα πραγματικά πολύπλοκα. Πραγματικά όμως. Ακόμα και η μαμά της Μαιρούλας, μου χαμογελάει πλατιά πλέον, ενώ με είχε για αγόρι πολύ ανισόρροπο. »



Σχέδια : J.J. Sempé, **Μαθηματικοί τύποι:** Y. Bugeaud - M. Mignotte - F. Normandin, **Κείμενα :** G. Taviot, **Μορφοποίηση:** G. Taviot, A. Maes, **Μετάφραση στα ελληνικά – προσαρμογή :** Σ. Μανιτσάρης.