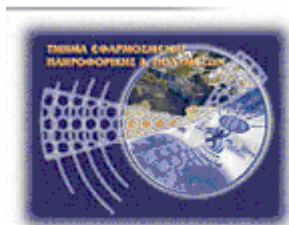




Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

Σχολή Τεχνολογικών Εφαρμογών

Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων



Πτυχιακή Εργασία

Τίτλος: Οπτική Σημασιολογική Ανάλυση Βίντεο Μουσικών Κομματιών με Πιάνο

Πασγιάννης Εμμανουήλ (ΑΜ: 2145)

Επιβλέπων καθηγητής : Τριανταφυλλίδης Γεώργιος

Επιτροπή Αξιολόγησης :

Ημερομηνία παρουσίασης :

Ευχαριστίες

Θα ήθελα πρώτα από όλους να ευχαριστήσω τον κ. Τριανταφυλλίδη για την καθοδήγηση του, χωρίς την οποία θα ήταν αδύνατη η εκπόνηση αυτής της εργασίας. Επίσης θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου για τις πολύτιμες ιδέες τους αλλά και την υποστήριξη που μου πρόσφεραν.

Abstract

The purpose of this thesis is to develop a system that can extract information from an already recorded video which presents a piano performance. The most important of these are: which key is played, when, for how long and from which finger. The combination of the image and the information can significantly help to distant piano learning, which becomes increasingly popular with the development of technology. So far this can be achieved using a video camera or a midi recorder. In the first case the student has only the image from video, while in the second only information about the notes. In this context, the main goal of this paper is to produce all of the information required, for distant piano learning, from the visual analysis of the video feed, without any need of midi support.

Σύνοψη

Σκοπός της Πτυχιακής Εργασίας είναι η ανάπτυξη ενός συστήματος που μπορεί να εξάγει πληροφορίες από ένα ήδη καταγεγραμμένο βίντεο το οποίο παρουσιάζει την εκτέλεση ενός μουσικού κομματιού από έναν πιανίστα. Οι πιο σημαντικές από αυτές είναι: ποιο πλήκτρο πατιέται, πότε, για πόσο χρονικό διάστημα και από ποιο δάχτυλο. Ο συνδυασμός εικόνας και πληροφορίας μπορεί να βοηθήσει σημαντικά στην εκμάθηση πιάνο εξ' αποστάσεως, η οποία γίνεται όλο και πιο δημοφιλής με την εξέλιξη της τεχνολογίας. Μέχρι στιγμής αυτό επιτυγχάνεται με τη χρήση βιντεοκάμερας ή μιας συσκευής εγγραφής midi. Στην πρώτη περίπτωση ο μαθητής έχει μόνο την εικόνα από το βίντεο, ενώ στη δεύτερη μόνο την πληροφορία για τις νότες. Σε γενικές γραμμές, ο κύριος στόχος της πτυχιακής εργασίας είναι η παραγωγή όλων των απαραίτητων πληροφοριών, για την εξ' αποστάσεως εκμάθηση πιάνο, από την οπτική ανάλυση του βίντεο, χωρίς την ανάγκη υποστήριξης από την τεχνολογία midi.

Περιεχόμενα

1	Εισαγωγή.....	1
1.1	Περίληψη.....	1
1.2	Κίνητρο για την Διεξαγωγή της Εργασίας	1
1.3	Σκοπός και Στόχοι Εργασίας.....	2
1.4	Δομή Εργασίας.....	2
2	Ψηφιακή Ανάλυση Εικόνας και Ήχου στη Μουσική.....	4
2.1	Αναγνώριση Μουσικών Οργάνων.....	4
2.2	Αναγνώριση Μουσικών Χειρονομιών.....	5
2.2.1	Μοντελοποίηση σε δυο και τρεις Διαστάσεις	6
2.2.2	Ανάλυση Χειρονομιών	8
3	Ευφυή Συστήματα στη Μουσική	10
3.1	Ελεγκτές	10
3.1.1	Matrix	10
3.1.2	PalmDriver	11
3.2	Εικονικά Όργανα.....	11
3.2.1	Εικονικά Ντραμς	12
3.2.2	Χορευτικά Παπούτσια (Expressive Footwear).....	12
3.2.3	Εικονικά Πιάνο.....	13
3.2.4	Εικονική Μαστρία	14
4	Ανάλυση του Συστήματος.....	16
4.1	Απαιτήσεις Συστήματος	16
4.2	Σύντομη Περιγραφή Αλγορίθμου.....	17
4.3	Προ-επεξεργασία και Αρχικοποιήσεις Τιμών	17
5	Εντοπισμός Χεριών	20
5.1	Ταίριασμα Προτύπων.....	20
5.2	Εντοπισμός στο Πεδίο του Χρόνου.....	20
5.3	Εντοπισμός στο Πεδίο του Χώρου.....	20

5.3.1	Εντοπισμός στον Χρωματικό Χώρο RGB.....	21
5.3.2	Εντοπισμός στον Χρωματικό Χώρο YCbCr	22
5.3.3	Χρωματική κβαντοποίηση.....	25
6	Ανίχνευση Ακροδαχτύλων	29
6.1	Εντοπισμός Περιγράμματος	29
6.2	Εντοπισμός Ακροδαχτύλων.....	29
6.3	Ονοματοθεσία Ακροδαχτύλων	32
7	Εντοπισμός Πατήματος Πλήκτρων	36
7.1	1 ^η Προσέγγιση	36
7.2	2 ^η Προσέγγιση	37
7.3	Σύγκριση Μεθόδων	37
8	Συμπεράσματα.....	40
8.1	Επισκόπηση του Συστήματος.....	40
8.2	Μελλοντική δουλειά.....	41

Εικόνες

Εικόνα 2.1: Μοντέλο φίλτρου-πηγής για διαχωρισμό ήχου.....	4
Εικόνα 2.2: Χωρικά μοντέλα χειρονομιών.....	6
Εικόνα 2.3: Πρότυπα χειρονομιών	7
Εικόνα 2.4: Κατάτμηση και εντοπισμός χεριού	8
Εικόνα 2.5: Σκιαγράφημα χεριού	8
Εικόνα 2.6: Περίγραμμα χεριού	9
Εικόνα 2.7: Θέσεις ακροδάχτυλων.....	9
Εικόνα 3.1: Matrix.....	11
Εικόνα 3.2: PalmDriver	11
Εικόνα 3.3: Εικονικά ντραμς.....	12
Εικόνα 3.4: Χορευτικά παπούτσια (Expressive Footwear).....	13
Εικόνα 3.5: Εικονικό πιάνο	13
Εικόνα 3.6: Ψηφιακή ράβδος μαεστρίας (Digital Baton).....	14
Εικόνα 3.7: Σακάκι του μαέστρου (Conductor's Jacket)	15
Εικόνα 4.1: Θέση βιντεοκάμερας και πηγής φωτός	16
Εικόνα 4.2: Θέση και οπτικό πεδίο βιντεοκάμερας	17
Εικόνα 4.3: Διάγραμμα ροής αλγορίθμου	18
Εικόνα 4.4: Οπτικό πεδίο βιντεοκάμερας.....	19
Εικόνα 5.1: Εντοπισμός χεριών στο πεδίο του χρόνου	20
Εικόνα 5.2: Χρωματικός χώρος RGB	21
Εικόνα 5.3: Κατανομή αποχρώσεων δέρματος στον RGB	22
Εικόνα 5.4: Το επίπεδο CbCr για σταθερό Y=0,5 και οριοθέτηση των δερματικών αποχρώσεων	22
Εικόνα 5.5: Τμηματοποίηση χεριών στον YCbCr.....	23
Εικόνα 5.6: Εντοπισμός χεριού (α) με χρήση φίλτρου (β) χωρίς χρήση φίλτρου	23
Εικόνα 5.7: Εντοπισμός ακροδαχτύλων (α) με χρήση φίλτρου (β) χωρίς χρήση φίλτρου.....	24
Εικόνα 5.8: Διάγραμμα Voronoi	26
Εικόνα 5.9: Τμηματοποίηση χεριών με χρήση κβαντοποίησης	28

Εικόνα 6.1: Περίγραμμα χεριών.....	30
Εικόνα 6.2: Λανθασμένος εντοπισμός ακροδαχτύλων.....	30
Εικόνα 6.3: Περιοχή ανίχνευσης ακροδαχτύλων	31
Εικόνα 6.4: Μη εντοπισμός ακροδαχτύλων που δεν πιέζουν πλήκτρα.....	31
Εικόνα 6.5: Εντοπισμός χεριών και ακροδάχτυλων (α),(γ) με χρήση φίλτρου εξομάλυνσης (β),(δ) χωρίς χρήση φίλτρου εξομάλυνσης.....	31
Εικόνα 6.6: Αφαίρεση λανθασμένα εντοπισμένου ακροδαχτύλου	32
Εικόνα 6.7: Παράδειγμα 1 ^{ου} κανόνα ονοματοθεσίας ακροδαχτύλων.....	33
Εικόνα 6.8: Παράδειγμα 2 ^{ου} κανόνα ονοματοθεσίας ακροδαχτύλων.....	33
Εικόνα 6.9: Διπλότυπη ονομασία ακροδαχτύλων με χρήση του 2 ^{ου} κανόνα (α) Πριν την εφαρμογή του κανόνα (β) μετά την εφαρμογή του κανόνα.....	34
Εικόνα 6.10: Διπλότυπη ονομασία ακροδαχτύλων με χρήση του 3 ^{ου} κανόνα (α) Πριν την εφαρμογή του κανόνα (β) μετά την εφαρμογή του κανόνα (έστω ότι χάνει το $i=2$).....	34
Εικόνα 6.11: Διπλότυπη ονομασία ακροδαχτύλων με χρήση του 4 ^{ου} κανόνα (α) Πριν την εφαρμογή του κανόνα (β) μετά την εφαρμογή του κανόνα (έστω ότι χάνει το $i=2$).....	35
Εικόνα 7.1: Σύστημα απόφασης (α) Η περιοχή του πλήκτρου στην αρχή του βίντεο (β) Η περιοχή του πλήκτρου στο καρέ που εξετάζεται (γ) Οι διαφορές μεταξύ των δυο εικόνων...	36
Εικόνα 7.2: Σφάλμα σταθερού κατωφλιού.....	36
Εικόνα 7.3: Πλήκτρο μακριά από το κέντρο της εικόνας (α) Η αρχική του κατάσταση (β) Η κλίση που αποκτά όταν βρίσκεται υπό πίεση.....	37
Εικόνα 7.4: Πλήκτρο κοντά στο κέντρο της εικόνας (α) Η αρχική του κατάσταση (β) Η αίσθηση γραμμικής προοπτικής που αποκτά όταν βρίσκεται υπό πίεση	37
Εικόνα 7.5: Εσφαλμένα εντοπισμένο πάτημα πλήκτρου	38
Εικόνα 7.6: Εντοπισμός ακροδαχτύλων και πατημάτων πλήκτρων.....	39
Εικόνα 8.1: Οπτική απεικόνιση των αποτελεσμάτων του συστήματος.....	41

Γραφήματα

Γράφημα 5.1: Σύγκριση αποδόσεων με ή χωρίς φίλτρο εξομάλυνσης	24
Γράφημα 5.2: Σύγκριση σφαλμάτων εντοπισμού με ή χωρίς φίλτρο εξομάλυνσης	25
Γράφημα 5.3: Ποσοστά ευστοχίας των μεθόδων YCbCr και κβαντοποίησης.....	27
Γράφημα 5.2: Σύγκριση σφαλμάτων εντοπισμού των μεθόδων YCbCr και κβαντοποίησης .	27
Γράφημα 7.1: Σύγκριση προτεινόμενων μεθόδων	38
Γράφημα 8.1: Διαχείριση χρόνου ανάλυσης ενός καρέ	40

1 Εισαγωγή

Οι πιο διαδεδομένες μέθοδοι για την εκμάθηση πιάνο εξ' αποστάσεως μπορεί να χρησιμοποιούν είτε μια βιντεοκάμερα, είτε μια συσκευή εγγραφής midi. Στην πρώτη περίπτωση το μέσο μάθησης είναι η εικόνα από το βίντεο, ενώ στη δεύτερη η πληροφορία που παράγεται από τη συσκευή. Με τη μέθοδο που προτείνεται σε αυτή την Πτυχιακή μπορούν να συνδυαστούν εικόνα και πληροφορία προς όφελος του μαθητή.

1.1 Περίληψη

Εξαιτίας της ανάπτυξης της τεχνολογίας, η εκμάθηση πιάνο εξ' αποστάσεως γίνεται όλο και πιο δημοφιλής. Σε πολλές περιπτώσεις αυτός είναι ο μόνος τρόπος να μάθει κάποιος μουσική, ιδιαίτερα σε απομακρυσμένες περιοχές. Φυσικά αυτή η μέθοδος έχει κάποια πολύ σημαντικά πλεονεκτήματα. Ένας μαθητής μπορεί να επιλέξει ένα διαδικτυακό πρόγραμμα σπουδών της προτίμησής του, αντί να παρακολουθήσει το πρόγραμμα κάποιου σχολείου το οποίο είναι πολύ πιθανό να επιλέξει με βάση το πόσο κοντά βρίσκεται στο σπίτι του. Η εκμάθηση μουσικής εξ' αποστάσεως μπορεί να γίνει οπουδήποτε, οποιαδήποτε στιγμή, παρέχοντας στον μαθητή ένα ευέλικτο πρόγραμμα με περισσότερο ελεύθερο χρόνο.

Στην εκμάθηση πιάνο εξ' αποστάσεως συνήθως χρησιμοποιούνται βίντεο που παρουσιάζουν την εκτέλεση κομματιών από τον καθηγητή, συνοδευόμενη από ομιλία στην οποία εξηγεί τις κινήσεις του. Ο μαθητής μπορεί να δει τον τρόπο που παίζονται οι νότες αλλά είναι δύσκολο να αντιστοιχίσει ένα πλήκτρο στο βίντεο με το αντίστοιχο στο δικό του πιάνο. Μια άλλη μέθοδος χρησιμοποιεί μια συσκευή εγγραφής midi, στην οποία αποστέλλονται οι νότες μέσα σε ένα αρχείο midi. Η τεχνολογία εγγραφής midi επιτρέπει στον καθηγητή να παίζει ένα κομμάτι στο πιάνο του και ο μαθητής να βλέπει ένα άλλο πιάνο σε κάποιο άλλο μέρος να παίζει μόνο του. Με την χρήση αυτής της μεθόδου ο μαθητής δεν μπορεί να γνωρίζει τον τρόπο με τον οποίο πιέζονται τα πλήκτρα. Η πτυχιακή εργασία στοχεύει στο συνδυασμό των δυο προηγούμενων μεθόδων για την παραγωγή βέλτιστου αποτελέσματος.

Η μέθοδος που περιγράφεται στην πτυχιακή χρησιμοποιεί μια βιντεοκάμερα για να καταγράψει μια μουσική εκτέλεση πιάνο. Η εικόνα από το βίντεο δίνεται σαν είσοδος στο σύστημα για επεξεργασία. Αρχικά γίνεται μια τμηματοποίηση βασισμένη στο χρώμα για να εντοπιστούν οι περιοχές του πιάνο αλλά και του κάθε πλήκτρου ξεχωριστά. Επίσης τμηματοποίηση βασισμένη στο χρώμα γίνεται και για τον εντοπισμό των χεριών. Ο εντοπισμός των ακροδάχτυλων γίνεται βρίσκοντας τα τοπικά ακρότατα στο περίγραμμα του χεριού. Όταν η θέση ενός ακροδαχτύλου ανήκει στην περιοχή ενός πλήκτρου, τότε το συγκεκριμένο πλήκτρο πρέπει να εξεταστεί επιπλέον. Για να ληφθεί η απόφαση αν το πλήκτρο είναι πατημένο ή όχι εξετάζονται τυχόν μεταβολές στο σχήμα του. Η έξοδος του συστήματος είναι ένας πίνακας με πληροφορίες σχετικά με το ποιο πλήκτρο πιάστηκε, ποια χρονική στιγμή, για πόσο διάστημα και από ποιο δάχτυλο.

1.2 Κίνητρο για την Διεξαγωγή της Εργασίας

Είναι αρκετά δύσκολο για κάποιον που δεν έχει προηγούμενη εμπειρία, να μάθει να παίζει πιάνο με μαθήματα εξ' αποστάσεως. Αυτό οφείλεται στη μη αμεσότητα της διδασκαλίας του μαθήματος. Επομένως, όσες περισσότερες είναι οι πληροφορίες που παρέχονται στον μαθητή τόσο καλύτερα μπορεί να αφομοιώσει τη γνώση που προσπαθεί να του μεταφέρει ο καθηγητής. Η παρούσα πτυχιακή στοχεύει στην κατασκευή ενός συστήματος που διευκολύνει τη κατανόηση της γνώσης που παρέχεται στον μαθητή, βοηθώντας τον να βελτιωθεί με τον ταχύτερο δυνατό ρυθμό.

1.3 Σκοπός και Στόχοι Εργασίας

Σκοπός της πτυχιακής είναι η ανάπτυξη ενός συστήματος που μπορεί να εξάγει πληροφορίες από ένα ήδη καταγεγραμμένο βίντεο το οποίο παρουσιάζει την εκτέλεση ενός μουσικού κομματιού από έναν πιανίστα. Οι πληροφορίες αυτές μπορούν να βοηθήσουν σημαντικά στην εκμάθηση πιάνο εξ' αποστάσεως. Πιο συγκεκριμένα οι απαραίτητες πληροφορίες είναι: πότε πατήθηκε το κάθε πλήκτρο, για πόσο χρονικό διάστημα και από ποιο δάχτυλο. Για να επιτευχθεί αυτό, πρώτα από όλα πρέπει να γίνει εντοπισμός των πλήκτρων του πιάνο. Ακολουθεί τμηματοποίηση των χεριών και εντοπισμός των ακροδαχτύλων. Στη συνέχεια πρέπει να ληφθεί η απόφαση αν κάποιο πλήκτρο έχει πατηθεί και να γίνει αντιστοίχηση του με το κατάλληλο δάχτυλο.

Χρησιμοποιώντας τις πληροφορίες που εξάγει το σύστημα είναι δυνατόν να δημιουργηθεί ένα αρχείο midi το οποίο αναπαράγει το κομμάτι. Επίσης, είναι εφικτό να κατασκευασθεί κάποια άλλη εφαρμογή η οποία θα έχει τη δυνατότητα χρησιμοποιώντας την αντιστοιχία πλήκτρου-δαχτύλου, να δημιουργεί μια τρισδιάστατη αναπαράσταση της μουσικής εκτέλεσης του κομματιού. Η τρισδιάστατη αναπαράσταση που δημιουργείται από την εφαρμογή σε συνδυασμό με τον ήχο από το midi αρχείο προσφέρουν μια πλήρη οπτικοακουστική εμπειρία στον μαθητεύομενο, η οποία δεν υστερεί σε τίποτα από την απευθείας διδασκαλία και ανοίγει νέους ορίζοντες στη μουσική διδασκαλία εξ' αποστάσεως. Τελειώνοντας την αναφορά στο σκοπό της εργασίας, αξίζει να σημειωθεί ότι είναι πολύ πιο εύκολο και γρήγορο να σταλεί μέσω internet στον μαθητή ένα πολύ μικρό αρχείο που περιέχει τις προαναφερθείσες πληροφορίες, αντί να σταλεί ένα μεγάλο σε μέγεθος βίντεο.

1.4 Δομή Εργασίας

Η δομή της Πτυχιακής Εργασίας έχει ως εξής:

- Κεφάλαιο 2: Ψηφιακή Ανάλυση Εικόνας και Ήχου στη Μουσική
Γενικά για την ανάλυση ήχου και εικόνας καθώς και για το πώς μπορούν να αποκτήσουν εφαρμογή στη μουσική.
- Κεφάλαιο 3: Ευφυή Συστήματα στη Μουσική
Αναφορά και παρουσίαση μερικών εφαρμογών ευφυών συστημάτων στη μουσική.
- Κεφάλαιο 4: Ανάλυση του Συστήματος
Σύντομη περιγραφή του προτεινόμενου συστήματος και αναφορά στις απαιτήσεις της εφαρμογής.
- Κεφάλαιο 5: Εντοπισμός Χεριών
Αναφορά σε μεθόδους εντοπισμού χεριών και σύγκρισή τους.
- Κεφάλαιο 6: Ανίχνευση Ακροδαχτύλων
Περιγραφή της μεθόδου που προτιμήθηκε για την ανίχνευση των ακροδαχτύλων.

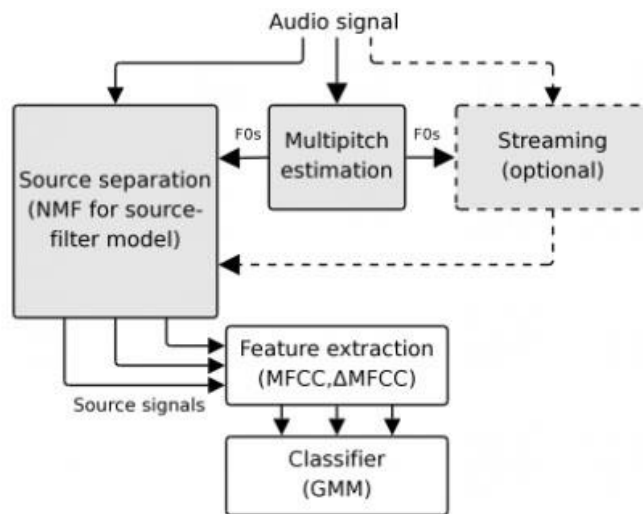
- Κεφάλαιο 7: Σύστημα Λήψης Απόφασης
Σε αυτό το κεφάλαιο προτείνονται κάποιες λύσεις στο πρόβλημα της απόφασης για το πότε ένα πλήκτρο πιέζεται ή όχι.
- Κεφάλαιο 8: Συμπεράσματα
Παρουσίαση των αποτελεσμάτων της Πτυχιακής Εργασίας και των μελλοντικών σχεδίων για την εξέλιξη του συστήματος.

2 Ψηφιακή Ανάλυση Εικόνας και Ήχου στη Μουσική

Η ανάλυση ήχου και εικόνας χρησιμοποιείται ευρέως στη μουσική. Στις μέρες μας εφαρμογές που αποσκοπούν στην αυτόματη αναγνώριση τραγουδιστών και τραγουδιών γίνονται όλο και πιο διαδεδομένες. Τέτοιες εφαρμογές μπορούν να βοηθήσουν στην έρευνα που γίνεται γύρω από την κατανόηση του τρόπου έκφρασης ενός μουσικού κατά τη διάρκεια μιας μουσικής παράστασης. Η έρευνα σχετικά με την ανάλυση εικόνας στη μουσική βρίσκεται ακόμα σε πρώιμο στάδιο αλλά τα πρώτα βήματα φαντάζουν πολλά υποσχόμενα.

2.1 Αναγνώριση Μουσικών Οργάνων

Η αυτόματη αναγνώριση της ηχητικής πηγής παίζει πολύ σημαντικό ρόλο στην ανάπτυξη εφαρμογών αυτόματης καταγραφής και ανάκτησης από μια βάση δεδομένων. Τέτοιες εφαρμογές έχουν τη δυνατότητα να σώσουν τους ανθρώπους από χρονοβόρες αναζητήσεις μέσω τεράστιων ποσοτήτων ψηφιακού ηχητικού υλικού που είναι διαθέσιμο. Για παράδειγμα, θα ήταν πιο χρήσιμο αν μπορούσαν να βρεθούν δείγματα που ακούγονται παρόμοια με το δοσμένο ηχητικό παράδειγμα. Η αυτόματη αναγνώριση μουσικών οργάνων μπορεί να λύσει αυτά τα προβλήματα και μπορεί επιπλέον να παρέχει χρήσιμες πληροφορίες σε άλλες περιοχές της αναγνώρισης της ηχητικής πηγής, όπως είναι η αναγνώριση του ομιλητή. Ωστόσο, η ανάλυση του μουσικού σήματος δεν έχει καταφέρει να προσεγγίσει εμπορικό ενδιαφέρον όπως για παράδειγμα, η αναγνώριση του ομιλητή και του λόγου. Αυτό συμβαίνει επειδή τα θέματα γύρω από την επεξεργασία της φωνής έχουν πιο εύκολη εμπορική εφαρμογή, παρ' όλα αυτά και οι δυο περιοχές θεωρούνται υψηλής πολυπλοκότητας.



Εικόνα 2.1: Μοντέλο φίλτρου-πηγής για διαχωρισμό ήχου

Σε έναν μουσικό ήχο μπορούν να αναγνωριστούν τέσσερα χαρακτηριστικά: τόνος, ένταση, διάρκεια και ποιότητα. Χάρη σε αυτά είναι δυνατόν ένας ακροατής να διακρίνει έναν μουσικό ήχο από έναν άλλο. Ο τόνος είναι καλά ορισμένος και είναι σχεδόν ίσος με τη θεμελιώδη συχνότητα. Η ένταση είναι ανάλογη του τετραγώνου του πλάτους της ακουστικής πίεσης. Η ποιότητα είναι το λιγότερο κατανοητό από τα χαρακτηριστικά. Όταν δυο ήχοι ίδιας έντασης και τόνου είναι ανόμοιοι, έχουν διαφορετική ποιότητα.

Έχουν αναπτυχθεί συστήματα τα οποία προσπαθούν να εξάγουν πληροφορίες από ήχους μουσικών οργάνων και να αναγνωρίσουν τις πηγές τους. Βέβαια απέχουν πολύ από το να εφαρμοστούν στον πραγματικό κόσμο. Η πλειονότητα των ερευνών έως τώρα έχει γίνει

χρησιμοποιώντας απομονωμένες νότες ή αποσπάσματα από σόλο εκτελέσεις. Μια ανασκόπηση των προτεινόμενων προσεγγίσεων στην βασισμένη σε απομονωμένες νότες αναγνώριση υπάρχει στο [1]. Με τα χρόνια αυξάνεται το ερευνητικό ενδιαφέρον για αναγνώριση ήχου, παραγόμενου από πολλά όργανα. Κάποιες τεχνικές αποσπών μουσικά χαρακτηριστικά από το σήμα, αποφεύγοντας τον διαχωρισμό της πηγής [4] [7]. Σε πολυφωνικές μείξεις αποτελούμενες από πολλά όργανα, η παρεμβολή ταυτόχρονων ήχων περιορίζει την απόδοση της αναγνώρισης. Η παρεμβολή μπορεί να μειωθεί χωρίζοντας την μείξη σε σήματα για κάθε ανεξάρτητη πηγή ήχου. Επιπλέον, ο διαχωρισμός των ηχητικών πηγών έχει εφαρμογές στο χειρισμό ήχου. Ένα σύστημα διαχωρισμού βασισμένο σε αυτή την προσέγγιση στοχεύει στη βελτίωση της ακρίβειας της αναγνώρισης ενός τραγουδιστή όταν υπάρχει μουσική υπόκρουση [8]. Αρμονικά μοντέλα εκπαιδευμένα πάνω σε συγκεκριμένα όργανα μπορούν να επιτύχουν διαχωρισμό και αναγνώριση ταυτόχρονα [10].

Έχουν προταθεί πολλοί αλγόριθμοι βασισμένοι στην παραγοντοποίηση του πίνακα του φασματογράμματος της μίξης. Το σήμα αναπαρίσταται σαν ένα άθροισμα συστατικών έχοντας σταθερό φάσμα και μεταβαλλόμενο ως προς τον χρόνο κέρδος. Η αποσύνθεση μπορεί να γίνει χρησιμοποιώντας ανεξάρτητη ανάλυση συστατικών ή μη-αρνητική παραγοντοποίηση πίνακα, το δεύτερο συνήθως οδηγεί σε καλύτερη ποιότητα διαχωρισμού. Το πλεονέκτημα των μεθόδων αυτών είναι η ικανότητα να γνωστοποιούν τα φασματικά χαρακτηριστικά της κάθε πηγής από μια μείξη, ενεργοποιώντας τον διαχωρισμό των πηγών στο πεδίο του χρόνου και στο πεδίο των συχνοτήτων. Συστήματα αναγνώρισης οργάνων βασισμένα στην αποσύνθεση που αποκτήθηκε από την ανεξάρτητη ανάλυση συστατικών, εξάγουν τα χαρακτηριστικά από τα αναμενόμενα φασματικά διανύσματα [11]. Μια ατέλεια της αποσύνθεσης του φασματογράμματος είναι ότι κάθε τόνος από κάθε διαφορετικό όργανο πρέπει να αναπαρασταθεί από μια ξεχωριστή συνάρτηση. Αυτό απαιτεί πληθώρα συναρτήσεων, πράγμα που κάνει τον διαχωρισμό και την ταξινόμηση δύσκολα.

Οι Virtanen και Klauri [12] πρότειναν τη μοντελοποίηση κάθε βασικού φασματικού διανύσματος σαν προϊόν διέγερσης και φιλτραρίσματος. Τα μοντέλα διέγερσης παράγουν χρονικά μεταβαλλόμενο τόνο χρησιμοποιώντας μια ακολουθία χαρακτήρων, τον οποίο μπορούν να μοιραστούν διαφορετικά όργανα. Για τη βελτίωση της απόδοσης αυτής της μεθόδου ο Fitzgerald πρότεινε τη μοντελοποίηση των διεγέρσεων θεμελιωδών συχνοτήτων σαν άθροισμα ημιτόνων στις αρμονικές συχνότητες [13]. Οι Badeau, Epiya και David [14] χρησιμοποίησαν επίσης ένα αρμονικό μοντέλο για τη διέγερση, αλλά μοντελοποίησαν το φίλτρο χρησιμοποιώντας ένα μοντέλο με μεταβλητό μέσο όρο, καταλήγοντας σε μια ομαλή απόκριση συχνότητας. Οι Vincent, Bertin και Badeau [15] μοντελοποίησαν τα βασικά διανύσματα του φάσματος σαν ένα σταθμισμένο άθροισμα των αρμονικά περιορισμένων φασμάτων, έχοντας περιορισμένο εύρος ζώνης. Σε αυτό το μοντέλο τα βάρη της κάθε ζώνης συχνοτήτων παραμετροποιούν το ανώμαλο φασματικό σχήμα του κάθε οργάνου. Οι Heittola, Klauri και Virtanen [16] παρουσίασαν μια προσέγγιση χρησιμοποιώντας ένα μοντέλο φιλτραρίσματος της πηγής όπως φαίνεται στην εικόνα 4.1. Το σήμα αποσυντίθεται σε ένα άθροισμα φασματικών μοντέλων σαν προϊόν διεγέρσεων και φιλτραρισμάτων. Οι διεγέρσεις περιορίζονται στα αρμονικά φάσματα και οι θεμελιώδεις συχνότητες τους υπολογίζονται στη συνέχεια χρησιμοποιώντας έναν εκτιμητή πολλαπλών τόνων. Τα φίλτρα περιορίζονται για να έχουν ομαλές αποκρίσεις συχνότητας μοντελοποιώντας τα σαν ένα άθροισμα στοιχειωδών συναρτήσεων στην κλίμακα Mel-frequency. Η πληροφορία για τον τόνο και τη ποιότητα του χρησιμοποιούνται στην οργάνωση ανεξάρτητων νοτών σε πηγές ήχου.

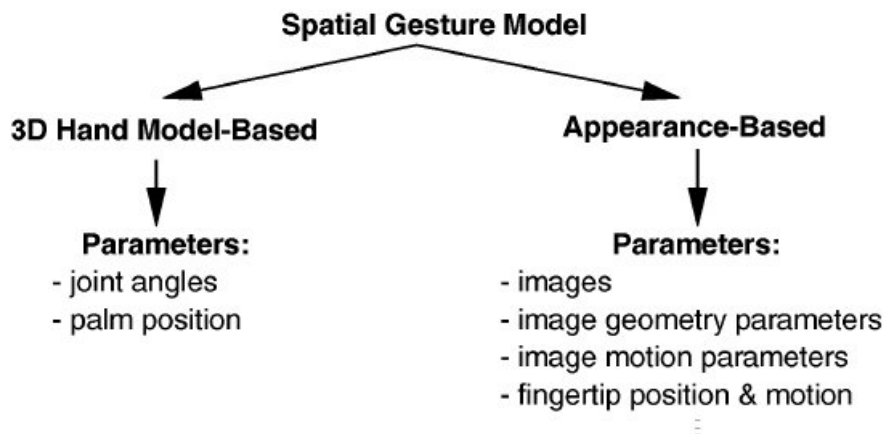
2.2 Αναγνώριση Μουσικών Χειρονομιών

Για την αλληλεπίδραση σε ένα σύστημα χρήστη-υπολογιστή μπορούν να χρησιμοποιηθούν διάφορες συσκευές εισόδου ανάλογα με το πρόβλημα. Όταν το πρόβλημα είναι η αναγνώριση χειρονομιών των χεριών η πιο διαδεδομένη τεχνική χρησιμοποιεί γάντια [30] [31] που συνδέονται με πολλά καλώδια σε έναν υπολογιστή. Τα καλώδια όμως περιορίζουν το εύρος των κινήσεων που μπορεί να πραγματοποιήσει ένα χέρι, καθώς δεν

έχουν πάντα αρκετά μεγάλο μήκος και πολλές φορές μπλέκονται μεταξύ τους ή ακόμα και με το ίδιο το χέρι κάνοντας ακόμα και την πιο απλή κίνηση να πραγματοποιείται με δυσκολία. Για την αποφυγή τέτοιων περιορισμών προτάθηκε μια προσέγγιση βασισμένη στην όραση [28]. Χάρη σε αυτή την προσέγγιση το χέρι έχει πολύ μεγάλη ελευθερία κινήσεων. Παρ' όλα αυτά, η υλοποίηση της είναι αρκετά περίπλοκη γι' αυτό και οι περισσότερες εφαρμογές εστιάζουν σε μια μόνο πτυχή της αναγνώρισης χειρονομιών, για παράδειγμα, εντοπισμός χεριού ή ερμηνεία της στάσης του χεριού. Η έρευνα στην αναγνώριση χειρονομιών μπορεί να χωριστεί σε 2 κατηγορίες: Η πρώτη ασχολείται με τη διατύπωση του προβλήματος της χρήσης συστημάτων που χρησιμοποιούν όραση για την αναγνώριση χειρονομιών. Η Άλλη είναι η κατασκευή ενός μαθηματικού μοντέλου του χεριού το οποίο ενθυλακώνει αρκετές λεπτομέρειες για την ανάλυση της χειρονομίας. Σχετικά με την δεύτερη κατηγορία, η σωστή μοντελοποίηση των χειρονομιών του χεριού επηρεάζει άμεσα την ποιότητα της εφαρμογής. Ένα πρόχειρο μοντέλο πολλές φορές είναι αρκετό για απλές εφαρμογές, αλλά όσο αυξάνονται οι απαιτήσεις χρειάζεται ένα μοντέλο που να υποστηρίζει όσο το δυνατόν περισσότερες χειρονομίες.

2.2.1 Μοντελοποίηση σε δυο και τρεις Διαστάσεις

Το χέρι μπορεί να κινηθεί στον τρισδιάστατο χώρο, ως εκ τούτου η περιγραφή των χειρονομιών συμπεριλαμβάνει επίσης τον χαρακτηρισμό των χωρικών ιδιοτήτων τους. Ο χαρακτηρισμός αυτός εξαρτάται από το είδος της εφαρμογής για την οποία έχει σχεδιαστεί η διεπαφή. Κάποιες εφαρμογές, όπως ο έλεγχος ηλεκτρικών οικιακών συσκευών [33], απαιτούν μόνο ένα απλό μαθηματικό μοντέλο του χεριού, ενώ άλλες εφαρμογές όπως η σχεδίαση χρησιμοποιώντας χειρονομίες απαιτεί πιο εξελιγμένα μοντέλα χεριού [20]. Το ανθρώπινο χέρι αποτελείται από αρθρώσεις, για να μπορέσει δηλαδή ένα μοντέλο να υλοποιήσει όλες τις πιθανές χειρονομίες που μπορεί να πραγματοποιήσει ένα χέρι, αρκεί να είναι γνωστή η θέση όλων των αρθρώσεων αλλά και των ακροδάχτυλων στον τρισδιάστατο χώρο (εικόνα 2.2).



Εικόνα 2.2: Χωρικά μοντέλα χειρονομιών

Αυτός ο ορισμός παρέχει όλες τις πληροφορίες που χρειάζονται για τη σωστή ανάλυση των χειρονομιών. Υπάρχει όμως ένα πρόβλημα. Για να αποκτηθούν αυτές τις πληροφορίες χρησιμοποιώντας τεχνητή όραση είναι εξαιρετικά πολύπλοκο [36]. Υπάρχουν δυο προσεγγίσεις που έχουν ξεπεράσει το πρόβλημα της πολυπλοκότητας του προαναφερθέντος μοντέλου [25] και φαίνονται στην εικόνα 2.1. Η πρώτη μοντελοποιεί τις χειρονομίες χρησιμοποιώντας ένα τρισδιάστατο μοντέλο χεριού. Η δεύτερη βασίζεται στην εμφάνιση,

καθώς συγκρίνει την εμφάνιση του χεριού με την εμφάνιση ήδη γνωστών χειρονομιών όπως αυτές που φαίνονται στην εικόνα 2.3.

Τρισδιάστατο μοντέλο χεριού

Ένα τρισδιάστατο μοντέλο χεριού μπορεί να κατασκευαστεί μελετώντας τη διαμόρφωση των αρθρώσεων του στο χώρο [27]. Στην πράξη σπανίως χρησιμοποιούνται όλες οι αρθρώσεις. Συνήθως χρησιμοποιούνται κάποιες από αυτές αρκεί να είναι δυνατόν να παραχθεί ισοδύναμο αποτέλεσμα. Η απλοποίηση του μοντέλου επιτυγχάνεται χρησιμοποιώντας κάποιες παραδοχές, για παράδειγμα, μπορούν να εισαχθούν ανεξαρτησίες ανάμεσα στις αρθρώσεις ή ακόμα και να επιβληθούν όρια στο εύρος κινήσεων των αρθρώσεων. Τα περισσότερα τρισδιάστατα μοντέλα βασίζονται σε απλοποιημένους σκελετούς του ανθρώπινου χεριού. Αν η κίνηση του βραχίονα είναι περισσότερο σημαντική χρησιμοποιούνται κυλινδρικά μοντέλα βραχίονα [34] [35] [36].

Δισδιάστατο μοντέλο χεριού

Η δεύτερη κατηγορία μοντέλων βασίζεται στην εμφάνιση του χεριού σε μια δισδιάστατη εικόνα. Αυτό σημαίνει ότι οι παράμετροι αυτού του μοντέλου δεν συμπεριλαμβάνουν τις παραμέτρους του προηγούμενου μοντέλου. Για να αναγνωριστεί μια χειρονομία αρκεί να γίνει σύγκριση της εμφάνισης της χειρονομίας με κάθε μια από ένα σετ προκαθορισμένων πρότυπων χειρονομιών όπως αυτό της εικόνας 2.2. Κάποια από τα μοντέλα που ανήκουν σε αυτή την κατηγορία βασίζονται σε ένα δισδιάστατο πρότυπο του χεριού ικανό να παραμορφωθεί [39] [40] [41]. Αυτά τα πρότυπα περιλαμβάνουν σημεία ενός αντικειμένου, τα οποία προσεγγίζουν το περίγραμμα του. Μια άλλη κατηγορία δισδιάστατων μοντέλων χρησιμοποιεί ακολουθίες εικόνων του χεριού σαν πρότυπα χειρονομιών. Κάθε επιτρεπόμενη χειρονομία μοντελοποιείται από μια ακολουθία αντιπροσωπευτικών εικόνων. Επιπλέον κάθε εικόνα μιας χειρονομίας αντιστοιχεί σε μια διαφορετική όψη της χειρονομίας αυτής. Στις περισσότερες των περιπτώσεων χρησιμοποιούνται μια ή δυο όψεις. Οι παράμετροι σε τέτοια μοντέλα μπορούν να είναι είτε αυτούσιες οι εικόνες [42] [43], είτε χαρακτηριστικά που έχουν αντληθεί από αυτές [44].



Εικόνα 2.3: Πρότυπα χειρονομιών

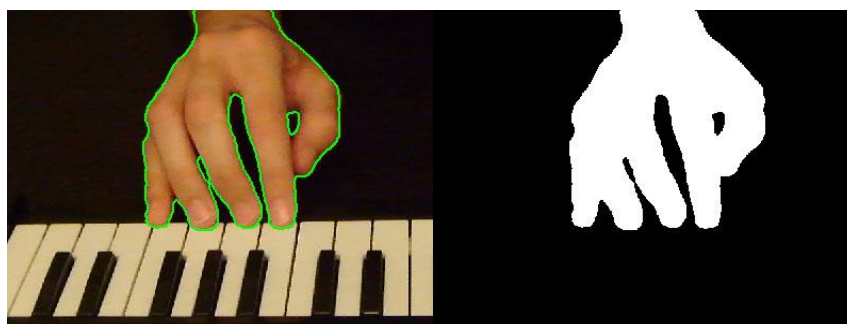
Τέτοια χαρακτηριστικά μπορούν να είναι περιγράμματα, ακμές, ακόμα και ιδιοδιανύσματα. Η τρίτη και τελευταία κατηγορία χρησιμοποιεί τις θέσεις των ακροδάχτυλων σαν παραμέτρους [22]. Αυτή η προσέγγιση είναι βασισμένη στο ότι οι θέσεις των ακροδάχτυλων σε σχέση με την παλάμη, είναι σχεδόν πάντα αρκετή για να διαφοροποιήσει έναν πεπερασμένο αριθμό χειρονομιών.

2.2.2 Ανάλυση Χειρονομιών

Οι παράμετροι των μοντέλων χειρονομιών εξάγονται από την ανάλυση ακολουθίας εικόνων που προέρχονται από βίντεο που έχει τραβηχτεί από μια ή περισσότερες κάμερες. Η φάση ανάλυσης της χειρονομίας αποτελείται από τρία βήματα: κατάτμηση του χεριού, εξαγωγή χαρακτηριστικών και υπολογισμός παραμέτρων.

Κατάτμηση και εντοπισμός χεριού

Η κατάτμηση είναι το πρώτο στάδιο ανάλυσης της χειρονομίας και περιλαμβάνει την εξαγωγή του χεριού από την υπόλοιπη εικόνα όπως φαίνεται στην εικόνα 2.3. Αυτή είναι μια πολύπλοκη διαδικασία λόγω του υπολογιστικού εύρους ζώνης που χρησιμοποιείται στην επεξεργασία εικόνας. Για να ελαττωθεί ο φόρτος εργασίας χρησιμοποιούνται τρία είδη περιορισμών: στο φόντο, στον χρήστη και στην απεικόνιση. Περιορισμοί στο φόντο είναι η χρήση ομοιόμορφου και σκουρόχρωμου φόντου, κάτι που απλοποιεί κατά πολύ τη διαδικασία της κατάτμησης. Οι περιορισμοί στον χρήστη και στην απεικόνιση απλοποιούν τον εντοπισμό του χεριού. Με την επιβολή των περιορισμών αυτών η κατάτμηση γίνεται εφαρμόζοντας ένα κατώφλι χρώματος ή εντάσεων απευθείας στην εικόνα. Άλλες τεχνικές αναλύουν την κίνηση της σκηνής. Οι κινήσεις στην εικόνα παράγονται κυρίως από τις κινήσεις των χεριών και έτσι μπορεί να γίνει κατάτμηση του χεριού από τα υπόλοιπα στατικά αντικείμενα, με την προϋπόθεση ότι το φόντο και η κάμερα παραμένουν σταθερά [32] [33].



Εικόνα 2.4: Κατάτμηση και εντοπισμός χεριού

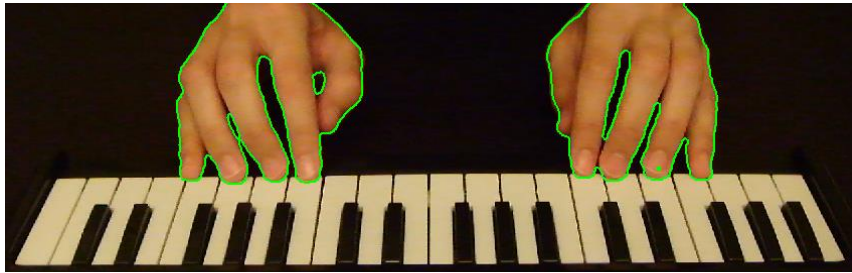
Εξαγωγή χαρακτηριστικών

Σκιαγραφήματα – Από τα πιο απλά χαρακτηριστικά. Χρησιμοποιείται ευρέως [39]. Παράγονται εύκολα από εικόνες χεριών σε φόντο με περιορισμούς (εικόνα 2.4). Χρησιμοποιούνται στην ανάλυση τρισδιάστατων αλλά και δισδιάστατων μοντέλων χεριών.



Εικόνα 2.5: Σκιαγράφημα χεριού

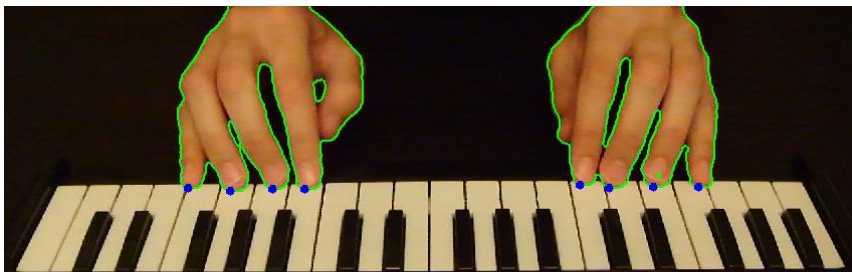
Περιγράμματα – Παράγονται εντοπίζοντας τις άκρες των αντικειμένων που περιέχονται σε μια εικόνα (εικόνα 2.5). Η τεχνική αυτή μπορεί να χρησιμοποιηθεί σε σκιαγραφήματα ή εικόνες αποχρώσεων του γκρι. Χρησιμοποιούνται στην ανάλυση δισδιάστατων και τρισδιάστατων μοντέλων χεριού.



Εικόνα 2.6: Περίγραμμα χεριού

Θέσεις ακροδάχτυλων – Χρησιμοποιούνται για την απόκτηση παραμέτρων διαστάσεων και τρισδιάστατων μοντέλων χεριών. Υπάρχουν 5 προσεγγίσεις στην ανίχνευση των θέσεων:

- Τα ακροδάχτυλα διακρίνονται χάρη σε σηματοδοτημένα γάντια στα χέρια του χρήστη [37] [29].
- Τεχνικές ταιριάσματος προτύπων μπορούν να χρησιμοποιηθούν για να ταιριάξουν εικόνες με πρότυπα ακροδάχτυλων [44] και δαχτύλων [38].
- Κάποιοι αλγόριθμοι εξαγωγής βασίζονται σε χαρακτηριστικές ιδιότητες των ακροδάχτυλων στην εικόνα. Για παράδειγμα η κυρτότητα ενός δαχτύλου (εικόνα 2.6) ακολουθεί ένα χαρακτηριστικό πρότυπο.
- Ευρετικοί αλγόριθμοι χρησιμοποιούνται.
- Έμμεσες προσεγγίσεις όπως ανάλυση εικόνας με τη χρήση ειδικά ρυθμισμένων πυρήνων Gabor [22].



Εικόνα 2.7: Θέσεις ακροδάχτυλων

Υπολογισμός παραμέτρων

Σε αυτό το βήμα τα χαρακτηριστικά που εξάχθηκαν από τις εικόνες, αντιστοιχίζονται στο στατικό μοντέλο χεριού. Ο τύπος του υπολογισμού εξαρτάται από τα χαρακτηριστικά και τη φύση των παραμέτρων του μοντέλου. Οι παράμετροι των περισσότερων τρισδιάστατων μοντέλων χειρονομιών υπολογίζονται χρησιμοποιώντας διαδοχικές προσεγγιστικές μεθόδους. Η βασική ιδέα είναι η συνεχής αλλαγή των παραμέτρων του μοντέλου έως ότου τα χαρακτηριστικά που εξάχθηκαν από το μοντέλο να ταιριάξουν με αυτά που συλλέχθηκαν από τις εικόνες. Η διαδικασία του ταιριάσματος συνήθως ξεκινάει από την παλάμη και καταλήγει στο ταιρίασμα των δαχτύλων. Οι αρχικές παράμετροι του μοντέλου συνήθως επιλέγονται έτσι ώστε να ταιριάξουν σε μια γενική θέση του χεριού.

3 Ευφυή Συστήματα στη Μουσική

Η μουσική είναι μια εξαιρετικά περίπλοκη τέχνη. Απαιτεί ανεπτυγμένες μεθόδους μηχανικής για να καταγραφεί και να παρουσιαστεί με αρκετές λεπτομέρειες. Η κατανόηση του πως εκφράζεται ένας μουσικός στη τέχνη του είναι ένα πολύ ενδιαφέρον και καρποφόρο αντικείμενο έρευνας, το οποίο παραμένει πλούσιο σε δυνατότητες καθώς έχει γίνει πολύ λίγη έρευνα πάνω σε αυτό. Πιθανές κατευθύνσεις έρευνας για την ενσωμάτωση της επιστήμης των υπολογιστών και της μουσικής φαίνονται παρακάτω:

- Με τη συνεχή ανάπτυξη των μεθόδων ανάλυσης μουσικών εκτελέσεων, μπορούν να αναπτυχθούν εργαλεία ανάλυσης.
- Ενσωματώνοντας με προσοχή ενισχυμένο δειγματοληπτημένο ήχο σε ζωντανές εκτελέσεις, η συνολική παλέτα ήχου αυξάνεται.
- Ειδικά οπτικά εφέ μπορούν να προστεθούν σε μουσικές εκτελέσεις, όπου η συμπεριφορά των οπτικών στοιχείων οδηγείται από τη μορφή της μουσικής εκτέλεσης.
- Αντί να αλλάξουν οι μουσικοί τις τεχνικές τους, τα υπολογιστικά συστήματα μπορούν να προγραμματιστούν για να ακολουθούν τις φυσικές τους κινήσεις.

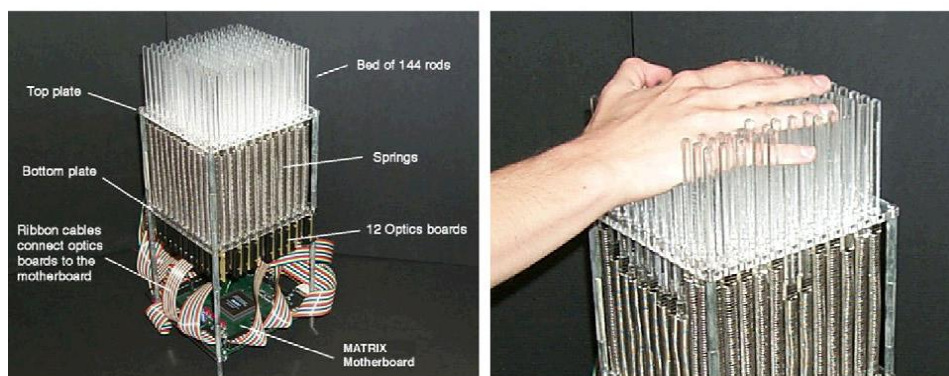
Σε αυτό το κεφάλαιο αναλύονται μερικές κατηγορίες διεπαφών χρήστη που προωθούν τη διαδραστική μουσική δημιουργικότητα. Έχουν γίνει πολλές προσπάθειες, ακαδημαϊκές και εμπορικές, ανάπτυξης και παρουσίασης νέων μουσικών εκτελέσεων που επισημαίνουν μια δυναμική αλληλεπίδραση μεταξύ παραδοσιακών και νέων τεχνολογιών [6]. Αυτές οι υλοποιήσεις χρησιμοποιούν τελευταίας τεχνολογίας αισθητήρες, λογισμικό αναγνώρισης χειρονομιών, μοντελοποίηση των ανθρώπινων εκφράσεων και μπορούν να βοηθήσουν στην επέκταση των δυνατοτήτων ενός μουσικού.

3.1 Ελεγκτές

Στις μουσικές εφαρμογές, ένας ελεγκτής είναι μια συσκευή ικανή να εισάγει πολυδιάστατες τιμές. Για παράδειγμα, ένα αναλογικό χειριστήριο εισάγει μια δισδιάστατη τιμή. Τέτοιες συσκευές σχεδιάστηκαν έτσι ώστε να ταιριάζουν στην ανθρώπινη ανατομία για να διευκολύνουν την μουσική έκφραση, όπως ένα πετάλι ποδιού. Ένα τέτοιο πετάλι είναι παρόμοιο με αυτό ενός πιάνο, πιέζοντας το ο ήχος που βγάζουν οι νότες διατηρείται περισσότερη ώρα. Στην ηλεκτρονική μουσική ένα πετάλι μπορεί να χρησιμοποιηθεί για να αλλάξει την ποσότητα της παραμόρφωσης που δημιουργείται από έναν ενισχυτή, καθώς και τα 2 χέρια του μουσικού είναι απασχολημένα. Κάποιες πρόσφατες έρευνες σε τέτοιου είδους ελεγκτές φαίνεται στη συνέχεια.

3.1.1 Matrix

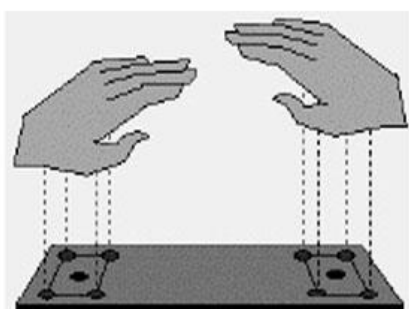
Το Matrix [23] (Multipurpose Array of Tactile Rods for Interactive expression) είναι μια νέα μουσική διεπαφή και φαίνεται στην εικόνα 3.1. Δίνει στο χρήστη τη δυνατότητα να ελέγξει μουσική με τα χέρια του και μπορεί να χρησιμοποιηθεί σε συνδυασμό με κάποιο μουσικό όργανο και ένα μικρόφωνο ή σαν μια αυτοδύναμη συσκευή εισόδου χειρονομιών. Η επιφάνεια του λειτουργεί σαν μια διεπαφή πραγματικού χρόνου που μπορεί να χειριστεί τις παραμέτρους ενός αλγορίθμου ειδικών εφέ ανάλογα με τις χειρονομίες του χρήστη.



Εικόνα 3.1: Matrix

3.1.2 PalmDriver

Το PalmDriver φαίνεται στην εικόνα 3.2. Είναι μια ηλεκτρονική συσκευή βασισμένη στην υπέρυθη τεχνολογία (IR) και αποτελείται από έως οκτώ σύνολα τεσσάρων αισθητήρων [26]. Επίσης στη συσκευή είναι ενσωματωμένος ένας μικροεπεξεργαστής για προεπεξεργασία των δεδομένων των αισθητήρων. Μπορεί να λειτουργήσει σαν αυτοδύναμο σύστημα καθώς είναι εξοπλισμένο με θύρα εξόδου MIDI. Η απόσταση σε διαφορετικές περιοχές του χεριού υπολογίζεται από την ποσότητα του ανακλώμενου φωτός που συνέλλεξαν οι αισθητήρες. Αυτές οι μετρήσεις είναι αρκετά ακριβείς λαμβάνοντας υπ' όψιν τις ανωμαλίες στο σχήμα και το χρώμα της ανθρώπινης παλάμης. Είναι δυνατόν να ανακατασκευαστούν οι θέσεις και οι κινήσεις των χεριών από τα δεδομένα που παράγονται από το PalmDriver. Η συσκευή αυτή είναι σταθερή και έχει πολύ καλή ανταπόκριση, κατά συνέπεια ο ήχος που παράγεται από τον υπολογιστή δίνει την αίσθηση στον μουσικό ότι 'αγγίζει τον ήχο'.



Εικόνα 3.2: PalmDriver

3.2 Εικονικά Όργανα

Ένα εικονικό όργανο είναι μια εφαρμογή λογισμικού που αλληλεπιδρά με αντικείμενα του πραγματικού κόσμου με τη βοήθεια αισθητήρων. Επιτρέπει στον χρήστη την παραγωγή μουσικής από τον υπολογιστή του. Τα εικονικά όργανα προσομοιώνουν είτε ένα αληθινό φυσικό όργανο, είτε νέα φανταστικά όργανα.

3.2.1 Εικονικά Ντραμς

Μια εφαρμογή εικονικών ντραμς [21] αποσκοπεί στη προσομοίωση ενός εικονικού περιβάλλοντος από ντραμς, στο οποίο κάποιος μπορεί να παίξει ακριβώς με τον ίδιο τρόπο όπως με ένα αληθινό σετ. Για να επιτευχθεί αυτό είναι πολύ σημαντικό να γίνει σωστή οπτικοποίηση. Η οπτικοποίηση συμπεριλαμβάνει τη χρήση τρισδιάστατων γραφικών για τη δημιουργία ενός ρεαλιστικού μοντέλου ντραμς, όπως αυτό στην εικόνα 3.3. Για να είναι ρεαλιστικό το μοντέλο, απαιτούνται ντραμς με οπτικά εφέ που αναπαριστούν την κίνηση ενός πραγματικού μοντέλου όταν ο χρήστης παίζει. Ένα τέτοιου είδους εφέ μπορεί να είναι, ένας κυματισμός που ξεκινάει από το κέντρο ενός τύμπανου και κινείται προς τα έξω. Σε συνδυασμό με τα οπτικά εφέ είναι σημαντικό να οπτικοποιηθεί το τυμπανόξυλο. Η θέση και ο προσανατολισμός των τυμπανόξυλων αναπαρίστανται γραφικά παρέχοντας στο χρήστη οπτική ανάδραση της κίνησης των πραγματικών τυμπανόξυλων που κρατάει στα χέρια του. Εξίσου σημαντικό είναι η παραγωγή ρεαλιστικού ήχου από τα ντραμς. Για την επίτευξη αυτού, κάθε κύμβαλο και τύμπανο έχει τον δικό του κατάλληλο ήχο. Επιπλέον, χρησιμοποιείται χωρικός ήχος, ο οποίος δίνει μια υπόδειξη για την θέση της πηγής ενός ήχου στον χώρο. Ενισχύεται έτσι ο ρεαλισμός των εικονικών ντραμς. Η αυξομείωση της έντασης ενός ήχου υποδεικνύει πόσο δυνατά χτυπήθηκε ένα τύμπανο ή κύμβαλο. Το τελευταίο σημαντικό κομμάτι των εικονικών ντραμς είναι η αλληλεπίδραση. Ο τρόπος που παίζονται τα εικονικά ντραμς πρέπει να είναι πανομοιότυπος με αυτόν στα πραγματικά ντραμς, ή όσο το δυνατόν πιο κοντινός. Αληθινά τυμπανόξυλα χωρίς καλώδια είναι απαραίτητα. Τα οπτικά και ακουστικά εφέ ενεργοποιούνται όταν κάποιο τυμπανόξυλο κινηθεί σε θέση που είναι τοποθετημένο ένα τύμπανο ή κύμβαλο.



Εικόνα 3.3: Εικονικά ντραμς

3.2.2 Χορευτικά Παπούτσια (Expressive Footwear)

Ένα ζευγάρι αθλητικά παπούτσια (εικόνα 3.4) στα οποία έχουν τοποθετηθεί αισθητήρες ικανοί να μετρήσουν πίεση, περιστροφή, θέση και επιτάχυνση [24]. Οι ενδείξεις των αισθητήρων μετατρέπονται σε ψηφιακά σήματα πάνω στο παπούτσι και στη συνέχεια εκπέμπονται σε κάποιο κοντινό δέκτη. Υπάρχουν 2 κατηγορίες εφαρμογών στις οποίες μπορούν να χρησιμοποιηθούν τα παπούτσια. Ανεπτυγμένες εκδόσεις του συστήματος μπορούν να χρησιμοποιηθούν από χορογράφους και χορευτές, προσθέτοντας μουσική πραγματικού χρόνου σε χορευτικές επιδείξεις. Είναι επίσης πολύ χρήσιμα σε προβλήματα ανάλυσης της κίνησης, όπως την αναγνώριση ενός συγκεκριμένου είδους χορού. Αρκεί ανάλυση μερικών δευτερολέπτων δεδομένων πραγματικού χρόνου για να αναγνωριστεί το είδος.



Εικόνα 3.4: Χορευτικά παπούτσια (Expressive Footwear)

3.2.3 Εικονικά Πιάνο

Ένα εικονικό πιάνο, όπως αυτό της εικόνας 3.5, αποτελείται από ένα σύστημα ψηφιακής ανάλυσης εικόνας πραγματικού χρόνου. Ο πιανίστας αλληλεπιδρά με το σύστημα χρησιμοποιώντας μόνο τα χέρια του [19] χωρίς να απαιτείται κάποιος επιπλέον εξοπλισμός. Κάθεται ως συνήθως σε μια καρέκλα πιάνου και δεν έχει μπροστά του τίποτα παρά μόνο μια κάμερα μερικά μέτρα μακριά η οποία κοιτάζει τα χέρια του. Μια φανταστική γραμμή σχεδιάζεται στο σύστημα στο ύψος που θα ήταν τα πλήκτρα αν υπήρχε πιάνο. Όταν ένα δάχτυλο ή χέρι διασχίσει αυτή τη γραμμή με φορά προς τα κάτω καταγράφεται σαν πάτημα κάποιου πλήκτρου. Το σημείο στο οποίο διασχίστηκε η γραμμή ορίζει το πλήκτρο. Λόγω του ότι το πιάνο δεν είναι ορατό, είναι δύσκολο για τον πιανίστα να πατήσει το σωστό πλήκτρο με ακρίβεια.



Εικόνα 3.5: Εικονικό πιάνο

3.2.4 Εικονική Μαστρία

Λέγεται πως η επίδοση του μαέστρου καθορίζει την επιτυχία ή την αποτυχία μιας ορχήστρας. Η τέχνη της καθοδήγησης μιας ορχήστρας συμπεριλαμβάνει μια μεγάλη ποσότητα γνώσης που αναμένεται να κατανοηθεί με τη βοήθεια της τεχνολογίας. Από τους ελεγκτές έως τα εικονικά όργανα και την καθοδήγηση μιας ορχήστρας, οι αντιστοιχίες σημάτων σε επιθυμητές αντιδράσεις γίνονται όλο και λιγότερο καθορισμένες. Όσο λιγότερο καθορισμένες είναι οι αντιστοιχίες τόσο πιο περίπλοκοι γίνονται οι κανόνες της αλληλεπίδρασης. Ο συνδυασμός μουσικών με ηλεκτρονικά μέσα σε ζωντανές εκτελέσεις για τη δημιουργία συνεκτικού ήχου παραμένει μείζων πρόβλημα [9]. Ο μαέστρος μπορεί να ελέγχει τα ηλεκτρονικά μέσα όπως ακριβώς θα έκανε αν είχε απέναντι του μουσικούς. Τεχνικές αναγνώρισης προτύπων χρησιμοποιούνται για να αναγνωριστούν οι κινήσεις των χεριών του μαέστρου. Οι αναγνωρισμένες κινήσεις πρέπει να αντιστοιχηθούν με κινήσεις χρησιμοποιώντας γνώσεις της τεχνικής της μαστρίας και γνώση του κομματιού που καθοδηγείται. Τα περισσότερα συστήματα μπορούν να ακολουθήσουν μόνο το ρυθμό του κομματιού. Ο κύριος σκοπός της έρευνας στην ακολουθία ενός μαέστρου, είναι η ερμηνεία του ρυθμού καθώς και άλλων μουσικών εκφράσεων αλλά και η μουσική αντίδραση σε πραγματικό χρόνο.

3.2.4.1 Ψηφιακή Ράβδος Μαστρίας (Digital Baton)

Η ψηφιακή ράβδος [5] [2] είναι ένα νέο ηλεκτρονικό όργανο το οποίο έχει σχεδιαστεί για την καθοδήγηση της παραγωγής μουσικής από υπολογιστή. Το κίνητρο για τη σχεδίαση του ήταν η δημιουργία ενός ελεγκτή χειρονομιών που αναπαράγει με όσο το δυνατόν μεγαλύτερη ακρίβεια την αίσθηση μιας παραδοσιακής ράβδου μαστρίας ενώ διατηρεί το μέγιστο αριθμό παραμέτρων ελέγχου για το χρήστη. Περιέχει διάφορα συστήματα αισθητήρων τα οποία καταγράφουν πολλές από τις λεπτομέρειες της κίνησης του χεριού. Το λογισμικό της χρησιμοποιεί τις λειτουργίες των αισθητήρων σε μουσικά συστήματα εκτέλεσης πραγματικού χρόνου. Ο εξοπλισμός όπως φαίνεται στην εικόνα 3.6, αποτελείται από μια ράβδο, έναν εξωτερικό αισθητήρα υπερύθρων, μια μονάδα εντοπισμού και ένα μικροϋπολογιστή. Στη ράβδο υπάρχουν ένα LED υπερύθρων για τον εντοπισμό της θέσης, αισθητήρες πίεσης για την ανίχνευση της πίεσης στα δάχτυλα και τη παλάμη καθώς και τρία επιταχυνσιόμετρα για την ανίχνευση του ρυθμού. Ο υπέρυθρος αισθητήρας και η ράβδος στέλνουν μέσω ενός καλωδίου διαφορετικές ροές δεδομένων στη μονάδα εντοπισμού, η οποία μετατρέπει και στέλνει τα σήματα στον υπολογιστή.



Εικόνα 3.6: Ψηφιακή ράβδος μαστρίας (Digital Baton)

Εκτός από την κανονική της χρήση σαν μουσικός ελεγκτής χρησιμοποιείται επίσης σε εντελώς νέες εφαρμογές. Για παράδειγμα, από τη στιγμή που μετράει τις κινήσεις των δαχτύλων και της παλάμης, μπορεί να χρησιμοποιηθεί όπως ένα μουσικό όργανο

αποτελούμενο από πλήκτρα. Η έρευνα που έγινε από τους Marrin Nakra και Paradiso [5] έφερε στην επιφάνεια δυο σημαντικά προβλήματα. Το πρώτο είναι η εργονομία της συσκευής: Η ψηφιακή ράβδος σχεδιάστηκε για να μοιάζει σε μια παραδοσιακή ράβδο μαεστρίας με μεγαλύτερη λαβή λόγω των αισθητήρων που τοποθετήθηκαν στη βάση της, αλλά ήταν υπερβολικά μεγάλη και βαριά για να χρησιμοποιηθεί άνετα από έναν μαέστρο. Επίσης το καλώδιο μπορούσε εύκολα να μπερδέψει κάπου, πράγμα μη ιδανικό για παραστάσεις. Το δεύτερο πρόβλημα ήταν ότι η αντιστοίχιση χειρονομιών σε εκφράσεις ήταν βασισμένη εξ' ολοκλήρου σε υπολογισμούς και δοκιμές. Αυτά τα ζητήματα οδήγησαν την έρευνα σε μια δεύτερη εργασία πάνω σε αυτόν τον τομέα, η οποία ήταν το σακάκι του μαέστρου (The Conductor's Jacket).

3.2.4.2 Σακάκι του Μαέστρου (Conductor's Jacket)

Οι Marrin και Picard [3] παρουσίασαν τη σχεδίαση μιας συσκευής που έχει τη δυνατότητα να συλλέγει και να αναλύει δεδομένα που παράγονται από τις κινήσεις ενός μαέστρου. Πρόκειται για ένα συνηθισμένο ένδυμα στο οποίο έχουν τοποθετηθεί διάφοροι αισθητήρες, όπως φαίνεται στην εικόνα 3.7, με τέτοιο τρόπο ώστε να επιτρέπει φυσιολογική δραστηριότητα και να μην παρεμποδίζει τις κινήσεις αυτού που το φοράει. Καταγράφει πληροφορίες ικανές να περιγράψουν τις κινήσεις του μαέστρου με σκοπό την καλύτερη κατανόηση της έκφρασης του καθώς καθοδηγεί την ορχήστρα. Η ιδέα της χρήσης εξοπλισμού που μπορεί να φορεθεί στέφθηκε με απόλυτη επιτυχία, επειδή δεν παρέμβαινε στις κινήσεις και εκφράσεις του ατόμου. Με τη βοήθεια αυτού του εξοπλισμού έχουν γίνει πολλά πειράματα συλλογής δεδομένων και έχει γίνει μια συνεχής πλατφόρμα για έρευνα, εκπαίδευση και παραστάσεις.



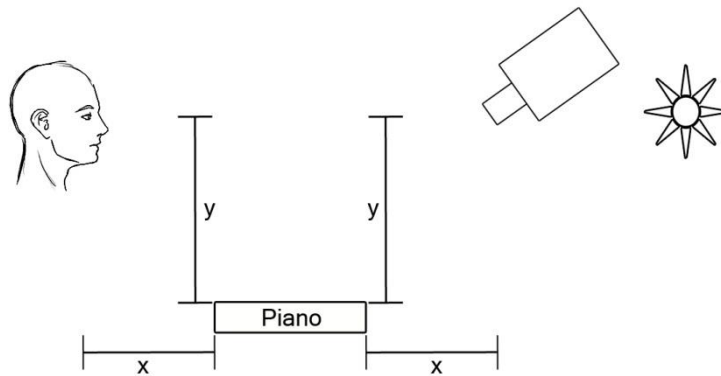
Εικόνα 3.7: Σακάκι του μαέστρου (Conductor's Jacket)

4 Ανάλυση του Συστήματος

Η μέθοδος που περιγράφεται στην πτυχιακή εργασία χρησιμοποιεί μια βιντεοκάμερα για να καταγράψει την εκτέλεση ενός μουσικού κομματιού από έναν πιανίστα. Η εικόνα από το βίντεο μπορεί να χρησιμοποιηθεί για την εξαγωγή χρήσιμων πληροφοριών, όπως είναι η μουσική σύνθεση του κομματιού, οι νότες δηλαδή που απαρτίζουν το κομμάτι, η χρονική στιγμή που χρησιμοποιήθηκε η κάθε νότα, αλλά και η διάρκεια της. Η επιπλέον πληροφορία που εισάγει το σύστημα που προτείνεται στην πτυχιακή εργασία είναι η αντιστοιχία της κάθε νότας με το δάχτυλο που την έπαιξε.

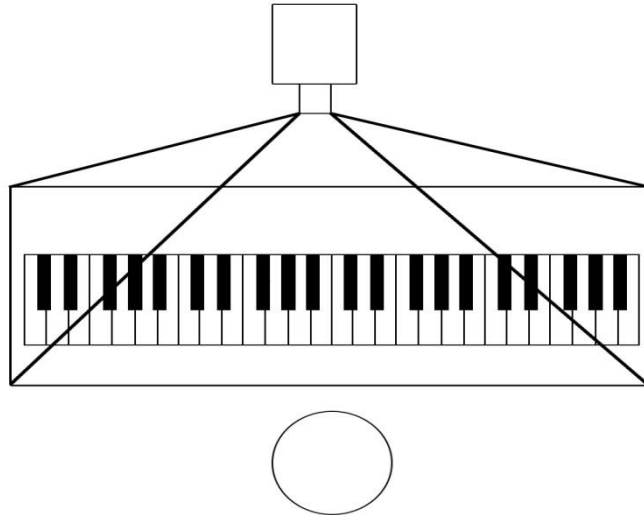
4.1 Απαιτήσεις Συστήματος

Για το βίντεο της εφαρμογής χρησιμοποιήθηκε η Fujifilm FinePix Real 3D W3, η οποία για την εγγραφή βίντεο παρέχει εικόνες ανάλυσης 1280x720, με βάθος χρώματος 24bit και ρυθμό καταγραφής 25 καρέ ανά δευτερόλεπτο. Ο αλγόριθμος του συστήματος αναπτύχθηκε με χρήση του λογισμικού Matlab και δοκιμάστηκε σε έναν ηλεκτρονικό υπολογιστή με επεξεργαστή Intel i5 με συχνότητα ρολογιού στα 3,4GHz. Ο μέσος χρόνος ανάλυσης ενός καρέ στο συγκεκριμένο επεξεργαστή είναι στα 102ms, χρόνος ικανοποιητικός για το είδος της εφαρμογής.



Εικόνα 4.1: Θέση βιντεοκάμερας και πηγής φωτός

Η συσκευή καταγραφής βίντεο πρέπει να τοποθετηθεί στην πίσω πλευρά του πιάνο, περίπου στην ίδια οριζόντια και κατακόρυφη απόσταση που έχει το πιάνο από τα μάτια του πιανίστα, όπως φαίνεται στην εικόνα 4.1. Ο λόγος που προτιμήθηκε η συγκεκριμένη θέση είναι ότι υπάρχει μονίμως οπτική επαφή με όλα τα πλήκτρα και τα δάχτυλα σπανίως αλληλεπικαλύπτονται. Ο πιανίστας επιβάλλεται να μην φοράει αξεσουάρ στα χέρια του, τα οποία θα εμποδίσουν τον σωστό εντοπισμό τους, όπως για παράδειγμα δαχτυλίδια, βραχιόλια, ακόμα και μια μακρυμάνικη μπλούζα θα δημιουργούσε προβλήματα. Αυτοί οι περιορισμοί μπορούν εύκολα να αποτραπούν αλλά αυτό είναι πέρα από τα πλαίσια της Πτυχιακής Εργασίας. Ιδιαίτερη προσοχή πρέπει να δοθεί στο φωτισμό, καθώς το σύστημα δεν αποδίδει εξίσου αποτελεσματικά σε καταστάσεις έντονου ή περιορισμένου φωτισμού. Για καλύτερα αποτελέσματα η πηγή φωτός πρέπει να βρίσκεται όσο το δυνατόν πιο κοντά στην κάμερα, κατά προτίμηση ακριβώς κάτω από αυτήν. Όσον αφορά το φόντο, πρέπει να αποτελείται από ουδέτερα χρώματα τα οποία δεν θυμίζουν τις αποχρώσεις του δέρματος. Στο οπτικό πεδίο της κάμερας πρέπει να είναι ορατά όλα και ολόκληρα τα πλήκτρα που σκοπεύει να χρησιμοποιήσει ο πιανίστας, αλλά και περίπου 20 εκατοστά της περιοχής μπροστά από τα πλήκτρα του πιάνο, το σημείο δηλαδή στο οποίο βρίσκονται τα χέρια του πιανίστα κατά την εκτέλεση του κομματιού (εικόνα 4.2).



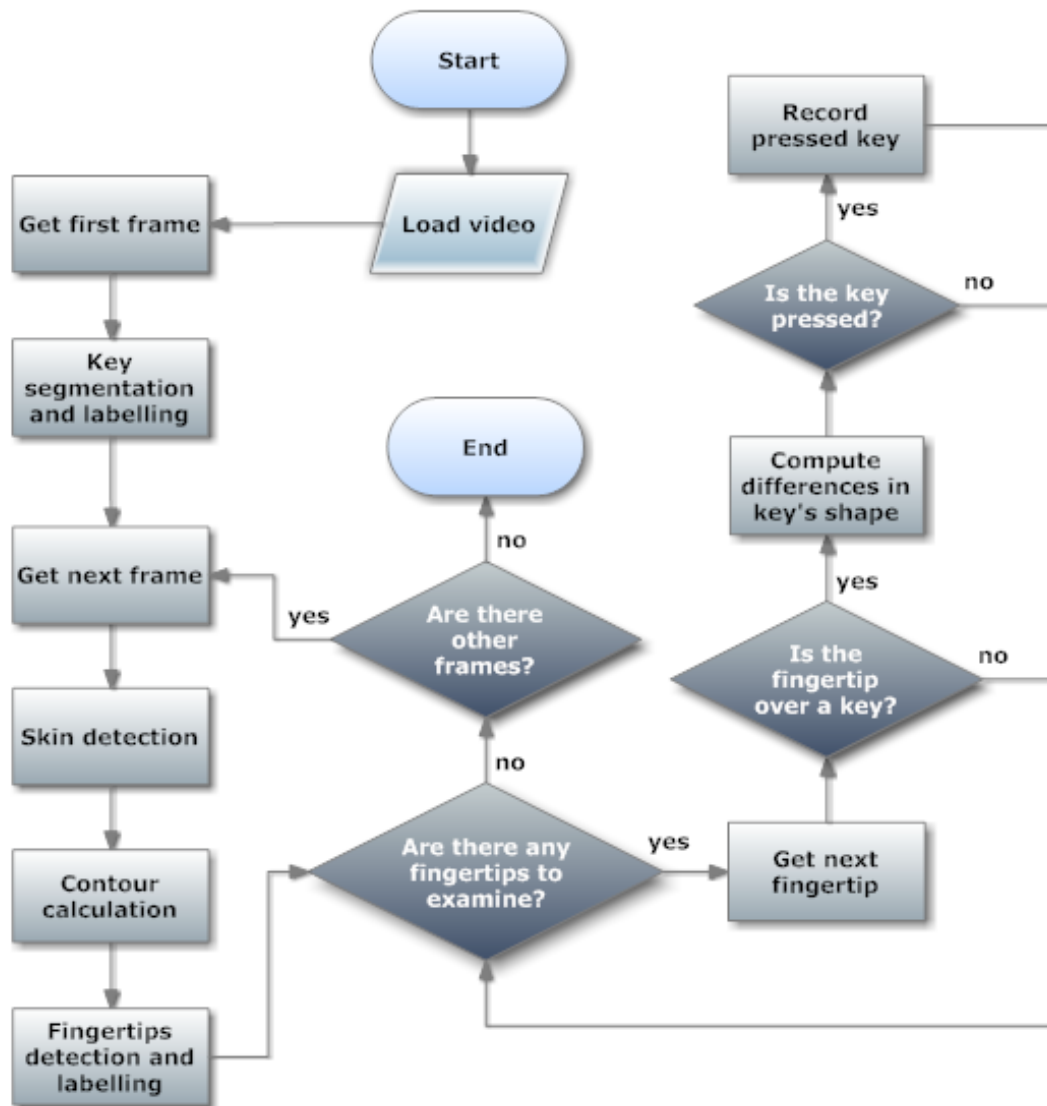
Εικόνα 4.2: Θέση και οπτικό πεδίο βιντεοκάμερας

4.2 Σύντομη Περιγραφή Αλγορίθμου

Ο αλγόριθμος της εφαρμογής χωρίζεται σε τέσσερα βασικά στάδια. Το πρώτο στάδιο ασχολείται με την εξέταση της καταλληλότητας του βίντεο και την απόδοση τιμών στις μεταβλητές που εξαρτώνται από τις παραμέτρους λήψης του βίντεο. Το δεύτερο στάδιο έχει να κάνει με τον εντοπισμό των χεριών, ενώ το τρίτο με την εύρεση της θέσης των ακροδαχτύλων. Το τέταρτο και τελευταίο στάδιο είναι αυτό που ελέγχει την κατάσταση του κάθε πλήκτρου και αποφασίζει ποιες νότες αναπαράγονται. Στην εικόνα 4.3 φαίνεται το διάγραμμα ροής του αλγορίθμου.

4.3 Προ-επεξεργασία και Αρχικοποιήσεις Τιμών

Για την σωστή λειτουργία του αλγορίθμου είναι βασικό να γίνουν κάποιες αρχικοποιήσεις. Οι αρχικοποιήσεις αυτές γίνονται με χρήση του πρώτου καρέ του βίντεο, στο οποίο φαίνεται το πιάνο χωρίς να είναι εμφανή τα χέρια του πιανίστα, όπως στην εικόνα 4.4. Αρχικά πρέπει να εντοπιστούν και να ονομαστούν τα πλήκτρα που είναι εμφανή στην εικόνα. Η τμηματοποίηση των πλήκτρων γίνεται με βάση τις αποχρώσεις τους. Για την επίτευξη αυτού δεν μπορεί να χρησιμοποιηθεί σταθερό κατώφλι στον αλγόριθμο διεκπεραίωσης του συγκεκριμένου βήματος, καθώς είναι δύσκολο δυο βιντεοσκοπήσεις που πραγματοποιήθηκαν σε διαφορετικό χρόνο, τόπο και με διαφορετικές συνθήκες φωτισμού να έχουν ίδια κατανομή των τιμών φωτεινότητας στο ιστόγραμμα τους. Για την υποβοήθηση της συνάρτησης υπολογισμού της τιμής του κατωφλίου απαιτείται από τον χρήστη να πληκτρολογήσει, όταν του ζητηθεί, τον αριθμό των πλήκτρων που υπάρχουν στην εικόνα. Με την προϋπόθεση ότι έχουν τηρηθεί οι συνθήκες καταγραφής του βίντεο που περιγράφηκαν στην ενότητα 4.1, μπορεί να υπολογισθεί η ζητούμενη τιμή. Ο κανόνας απόφασης για το αν ένα εικονοστοιχείο ανήκει σε κάποιο πλήκτρο περιγράφεται αναλυτικά στην επόμενη παράγραφο.



Εικόνα 4.3: Διάγραμμα ροής αλγορίθμου

Δουλεύοντας στον χρωματικό χώρο RGB και δίνοντας τιμή 100 στο κατώφλι, ψάχνουμε για εικονοστοιχεία που η κόκκινη συνιστώσα τους έχει τιμή μεγαλύτερη του κατωφλίου. Αφού αγνοήσουμε τα πολύ μικρά αντικείμενα (όσα αποτελούνται από λιγότερα του 0,1% των εικονοστοιχείων της εικόνας), αριθμούμε αυτά που εντοπίστηκαν. Αν ο αριθμός των αντικειμένων που εντοπίστηκαν με την εφαρμογή του κανόνα συμβαδίζει με τον αριθμό που έδωσε ο χρήστης, τότε κάθε αντικείμενο αντιστοιχεί σε ένα πλήκτρο. Στην περίπτωση που τα αντικείμενα στην εικόνα είναι λιγότερα από τα πλήκτρα, αυξάνουμε το κατώφλι κατά 10. Αν τα αντικείμενα είναι περισσότερα από τα πλήκτρα, το κατώφλι μειώνεται κατά 1. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να εντοπιστεί ο σωστός αριθμός αντικειμένων. Αν το κατώφλι ξεπεράσει την τιμή 255 σημαίνει πως το βίντεο δεν είναι κατάλληλο για την εφαρμογή και πρέπει να γίνει νέα εγγραφή βίντεο δίνοντας μεγαλύτερη βάση στις απαιτήσεις φωτισμού. Τέλος, δίνεται μια μοναδική τιμή σε κάθε πλήκτρο, η οποία χρησιμεύει στον διαχωρισμό των πλήκτρων.



Εικόνα 4.4: Οπτικό πεδίο βιντεοκάμερας

5 Εντοπισμός Χεριών

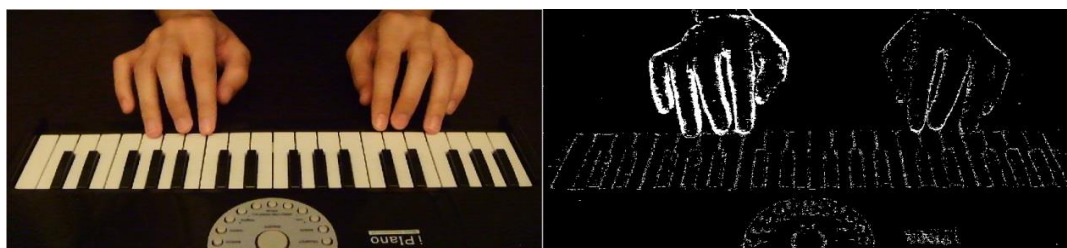
Μέχρι σήμερα είναι γνωστές πολλές τεχνικές εντοπισμού μελών του ανθρώπινου σώματος. Μπορούν να χωριστούν σε τρεις κύριες κατηγορίες: Ταίριασμα προτύπων, Εντοπισμός στο πεδίο του χρόνου και εντοπισμός στο πεδίο του χώρου. Η τεχνική που προτιμήθηκε σε αυτή την πτυχιακή ανήκει στην τρίτη κατηγορία και προτιμήθηκε λόγω της μεγάλης ακρίβειας που έχει στον εντοπισμό των χεριών.

5.1 Ταίριασμα Προτύπων

Η τεχνική αυτή χρησιμοποιείται για τον εντοπισμό μερών μιας εικόνας που μοιάζουν σε μια εικόνα πρότυπο. Υπάρχουν δυο προσεγγίσεις. Στην πρώτη η σύγκριση των εικόνων μπορεί να γίνει χρησιμοποιώντας διάφορα χαρακτηριστικά τους όπως για παράδειγμα κάποιες ακμές ή γωνίες. Είναι περισσότερο αποδοτική υπολογιστικά από την δεύτερη μέθοδο η οποία αναζητά οπτικές ομοιότητες ανάμεσα στις δυο εικόνες. Υπάρχει όμως ένα βασικό ελάττωμα. Είναι πιθανό ένα αντικείμενο σε μια εικόνα να επισκιάζεται ή ακόμα και να καλύπτεται μερικώς από κάποιο άλλο αντικείμενο. Φυσικά και δεν είναι δυνατόν να καλυφθούν όλες αυτές οι περιπτώσεις χρησιμοποιώντας πρότυπα εικόνων.

5.2 Εντοπισμός στο Πεδίο του Χρόνου

Έχοντας σαν προϋπόθεση ότι δεν υπάρχουν άλλα κινούμενα αντικείμενα, είναι δυνατόν να γίνει εντοπισμός των χεριών με βάση τις μεταβολές εικονοστοιχείων που εντοπίζονται στο βίντεο [17]. Μεταβολή ορίζεται ως η αλλαγή στην τιμή της φωτεινότητας ενός εικονοστοιχείου, πέρα από κάποια ορισμένη τιμή κατωφλιού, σε σχέση με την προηγούμενη του κατάσταση. Η τεχνική αυτή είναι αρκετά γρήγορη και αποτελεσματική εφόσον τα χέρια του πιανίστα κινούνται ασταμάτητα και με γρήγορο ρυθμό. Κάτι τέτοιο όμως είναι πρακτικά αδύνατο. Επίσης είναι αναγκαίο να υπάρξει πρόβλεψη για την κίνηση των πλήκτρων όταν κάποιο από αυτά πατηθεί. Το πρόβλημα είναι ότι αν ένα πλήκτρο πατηθεί για ένα σχετικά μεγάλο χρονικό διάστημα, δεν θα φαίνεται κίνηση του δαχτύλου και επομένως δε θα μπορεί να εντοπιστεί το πάτημα παρά μόνο τη χρονική στιγμή εκκίνησης και τερματισμού του πατήματος. Φυσικά και θα μπορούσε να γίνεται έλεγχος με πιο αργό ρυθμό έτσι ώστε αργές μεταβολές στην πραγματικότητα, να φαίνονται γρήγορες στο σύστημα, αλλά το στιγμιαίο πάτημα ενός πλήκτρου είναι πολύ πιθανό να μην μπορέσει να εντοπιστεί. Αυτό φαίνεται καλύτερα στην εικόνα 5.1 όπου το δεξί χέρι κινείται με ταχύτερο ρυθμό σε σχέση με το αριστερό και όπως είναι εμφανές, ο εντοπισμός του είναι πιο διακριτός.



Εικόνα 5.1: Εντοπισμός χεριών στο πεδίο του χρόνου

5.3 Εντοπισμός στο Πεδίο του Χώρου

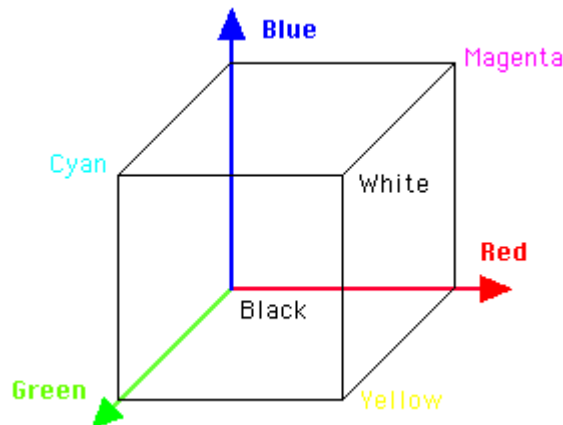
Οι τεχνικές αυτής της κατηγορίας βασίζονται στην ανάλυση των αποχρώσεων και των φωτεινοτήτων της εικόνας. Κάθε εικονοστοιχείο ελέγχεται ανεξάρτητα από τα υπόλοιπα. Αρκεί λοιπόν να οριοθετηθεί μια περιοχή στον χρωματικό χώρο που χρησιμοποιείται, η οποία

περιέχει όσο το δυνατόν περισσότερες από τις αποχρώσεις του αντικειμένου, χωρίς παράλληλα να περιέχονται άλλες αποχρώσεις σε αυτήν. Η οριοθέτηση της περιοχής είναι ένα περίπλοκο πρόβλημα για τον χρωματικό χώρο RGB, καθώς όπως φαίνεται και στην εικόνα 5.3 οι αποχρώσεις του δέρματος δεν είναι ομοιόμορφα κατανεμημένες. Η μετατροπή της εικόνας από RGB σε κάποιον άλλο χρωματικό χώρο όπως ο YCbCr αν και είναι αρκετά χρονοβόρα φαίνεται να είναι μονόδρομος για τμηματοποιήσεις εικόνων με βάση το χρώμα.

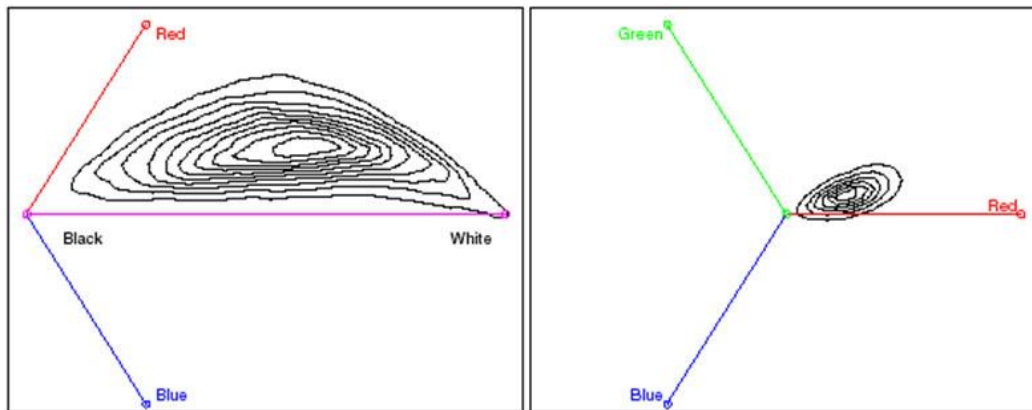
Σε αυτή την πτυχιακή εργασία έχουμε ένα πολύ περιορισμένο εύρος του φάσματος χρωμάτων που χρησιμοποιείται. Πιο συγκεκριμένα, έχουμε το χρώμα του δέρματος για τα χέρια του πιανίστα, το λευκό χρώμα για τα πλήκτρα του πιάνο, το μαύρο χρώμα για το υπόλοιπο πιάνο, καθώς και το χρώμα του φόντου το οποίο να μην είναι άγνωστο αλλά λόγω της γωνίας θέασης της κάμερας μπορούμε να είμαστε σχεδόν βέβαιοι πως πρόκειται για κάποιο ουδέτερο χρώμα χωρίς μεγάλες μεταβολές σε αποχρώσεις και φωτεινότητα, πράγμα που σημαίνει πως μπορεί να εντοπιστεί πολύ εύκολα. Για όλους τους παραπάνω λόγους προτιμήθηκε να χρησιμοποιηθεί η τεχνική της χρωματικής κβαντοποίησης, χάρη στην απλότητα της οποίας μπορούμε να επιτύχουμε πολύ καλά αποτελέσματα εντοπισμού σε πολύ μικρό χρονικό διάστημα.

5.3.1 Εντοπισμός στον Χρωματικό Χώρο RGB

Ο χρωματικός χώρος RGB είναι ο πιο διαδεδομένος στις ψηφιακές εικόνες. Κωδικοποιεί κάθε χρώμα σαν ένα συνδυασμό τριών βασικών χρωμάτων: κόκκινο πράσινο και μπλε. Οπτικοποιείται σαν ένας κύβος όπου οι συνιστώσες των κόκκινου, πράσινου και μπλε είναι οι 3 κάθετοι μεταξύ τους άξονες του κύβου (εικόνα 5.2). Λόγω της μεγάλης συσχέτισης των τριών καναλιών χρώματος αλλά και της μίξης των δεδομένων χρώματος και φωτεινότητας [45] δεν είναι η καλύτερη επιλογή για εφαρμογές που απαιτούν εντοπισμό δερματικών αποχρώσεων. Οι αποχρώσεις στον κύβο να μην είναι ομοιόμορφα κατανεμημένες αλλά όπως φαίνεται και στην εικόνα 5.3 είναι πολύ δύσκολο να οριστεί μια περιοχή που να περιέχει μόνο δερματικές αποχρώσεις.



Εικόνα 5.2: Χρωματικός χώρος RGB



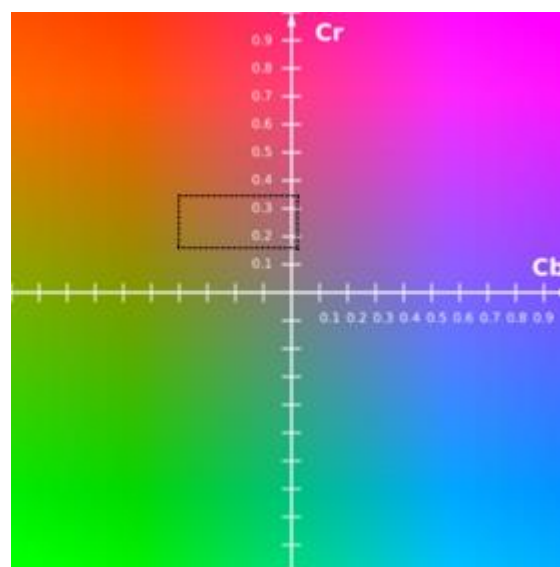
Εικόνα 5.3: Κατανομή αποχρώσεων δέρματος στον RGB

5.3.2 Εντοπισμός στον Χρωματικό Χώρο YCbCr

Σε αντίθεση με τον RGB, στον YCbCr οι δερματικές αποχρώσεις βρίσκονται σε μια ευδιάκριτη τετραγωνική περιοχή [4]. Η συνιστώσα Y αντιπροσωπεύει την ένταση του φωτός και είναι ανεξάρτητη από την απόχρωση η οποία, όπως βλέπουμε στην εικόνα 5.4, βρίσκεται αποθηκευμένη στις συνιστώσες Cb και Cr. Μια πολύ καλή επιλογή για τον εντοπισμό των χρωματικών αποχρώσεων του δέρματος μπορεί να οριστεί από τις ευθείες:

$$cr = 0,17 \quad cr = 0,35 \quad cb = -0,4 \quad cb = -0,01$$

Όπως φαίνεται και στην εικόνα 5.5 ο εντοπισμός γίνεται με μεγάλη ακρίβεια. Φυσικά είναι δυνατόν η νέα εικόνα να φιλτραριστεί με ένα φίλτρο μεσαίας τιμής για να μειωθούν τα σφάλματα στο ελάχιστο (εικόνα 5.6), αλλά κάτι τέτοιο είναι χρονοβόρο. Στην εικόνα 5.7 φαίνονται τα αποτελέσματα του εντοπισμού των ακροδαχτύλων με και χωρίς την εφαρμογή του φίλτρου εξομάλυνσης.



Εικόνα 5.4: Το επίπεδο CbCr για σταθερό Y=0,5 και οριοθέτηση των δερματικών αποχρώσεων

Παρότι τα αποτελέσματα είναι αρκετά ικανοποιητικά, δεν υπάρχει η δυνατότητα ανάγνωσης του αρχείου βίντεο απευθείας στον YCbCr, απαιτείται δηλαδή η μετατροπή από RGB για κάθε frame ξεχωριστά. Μια πολύ καλή μέθοδος για αυτή τη μετατροπή περιγράφεται αναλυτικά στο βιβλίο Video Demystified [18] και έχει τη δυνατότητα να μετατρέψει μια εικόνα αρκετά γρήγορα και με μεγάλη ακρίβεια στην απόδοση των χρωμάτων. Πιο συγκεκριμένα, οι συνιστώσες του χρωματικού χώρου μπορούν να υπολογισθούν ως εξής:

$$Y = (0.257 \times R) + (0.504 \times G) + (0.098 \times B) + 16$$

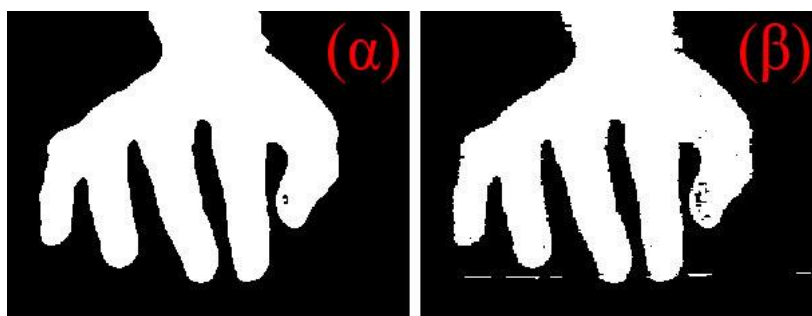
$$Cb = - (0.148 \times R) - (0.291 \times G) + (0.439 \times B) + 128$$

$$Cr = (0.439 \times R) - (0.368 \times G) - (0.071 \times B) + 128$$

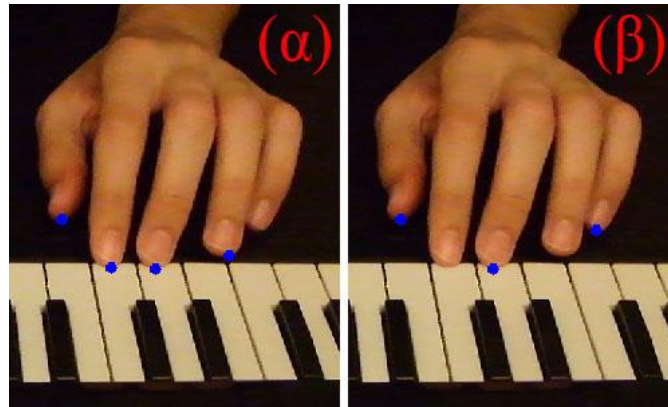


Εικόνα 5.5: Τμηματοποίηση χεριών στον YCbCr

Με χρήση αυτής της τεχνικής ο μέσος χρόνος μετατροπής ενός καρέ ανάλυσης 1280x720 για το μηχάνημα που χρησιμοποιήθηκε στις δοκιμές είναι 69ms. Δεδομένου ότι η πληροφορία που είναι αποθηκευμένη στην Y συνιστώσα είναι περιττή για τον εντοπισμό των επιθυμητών αποχρώσεων, ο υπολογισμός της τιμής της μπορεί να παρακαμφθεί με αποτέλεσμα ο χρόνος να μειωθεί στα 46ms. Με την εφαρμογή φίλτρου εξομάλυνσης ο χρόνος αυξάνεται στα 133ms και 110ms αντίστοιχα. Αφού ολοκληρωθεί επιτυχώς η μετατροπή ενός καρέ, ακολουθεί η εφαρμογή του κανόνα για τον εντοπισμό των χεριών. Με την προϋπόθεση ότι κάθε χρωματική συνιστώσα δέχεται τιμές από 0 έως 255, ένα εικονοστοιχείο ανήκει στην περιοχή του χεριού αν έχει τιμή Cr μεταξύ 150 και 173 και τιμή Cb μεταξύ 76 και 126.

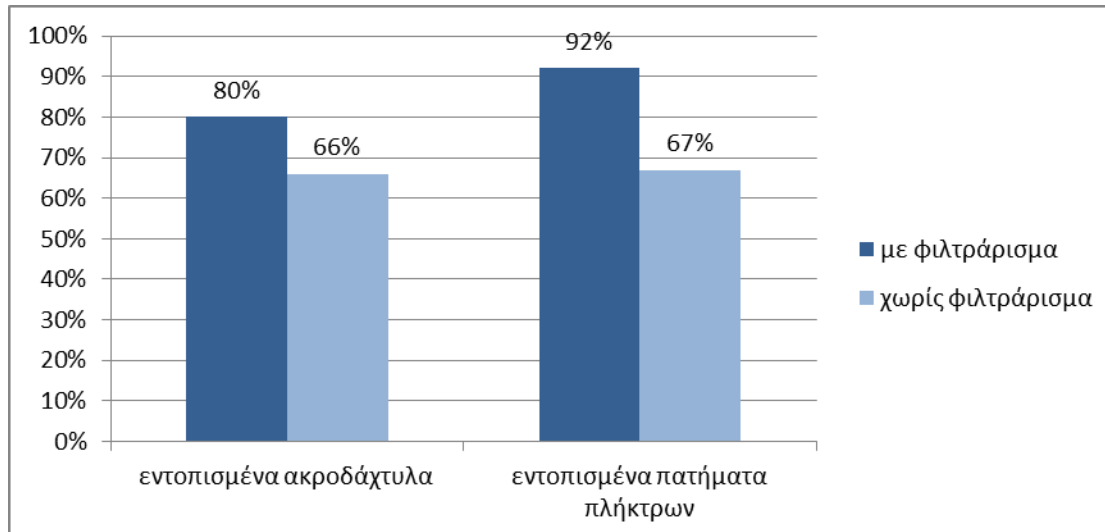


Εικόνα 5.6: Εντοπισμός χεριού (α) με χρήση φίλτρου (β) χωρίς χρήση φίλτρου

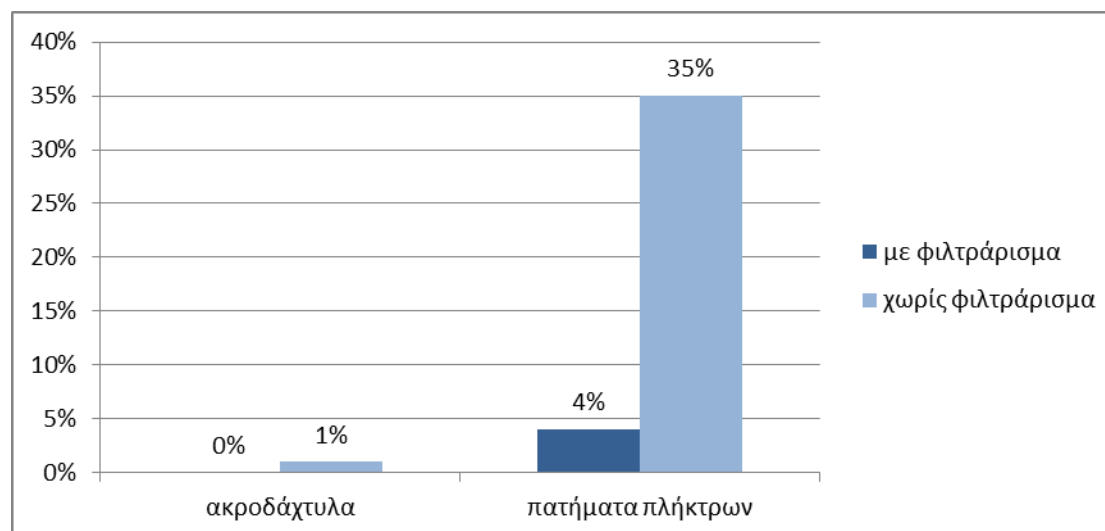


Εικόνα 5.7: Εντοπισμός ακροδαχτύλων (α) με χρήση φίλτρου (β) χωρίς χρήση φίλτρου

Όπως φαίνεται και στην εικόνα 5.7 ο εντοπισμός των ακροδαχτύλων είναι πιο αποτελεσματικός όταν χρησιμοποιείται φίλτρο εξομάλυνσης στο περίγραμμα του χεριού. Είναι όμως η διαφορά στην απόδοση ικανή να καλύψει τα 64ms που απαιτούνται για την εφαρμογή του φίλτρου εξομάλυνσης σε ένα frame; Σε τυχαίο δείγμα 100 καρέ από το βίντεο, εντοπίστηκαν οπτικά 1000 ακροδάχτυλα και 120 πατήματα πλήκτρων. Χωρίς την χρήση του φίλτρου, ο αλγόριθμος εντόπισε 664 ακροδάχτυλα και 80 πατήματα πλήκτρων, ενώ εντόπισε λανθασμένα ένα δάχτυλο και 28 πατήματα πλήκτρων. Με χρήση του φίλτρου εντοπίστηκαν 796 ακροδάχτυλα και 110 πατήματα πλήκτρων, τα λανθασμένα πατήματα πλήκτρων που εντοπίστηκαν ήταν μόλις 4 ενώ δεν υπήρξε κανένας λάθος εντοπισμός δαχτύλου. Τα αποτελέσματα των δοκιμών φαίνονται καλύτερα στα γραφήματα 5.1 και 5.2.



Γράφημα 5.1: Σύγκριση αποδόσεων με ή χωρίς φίλτρο εξομάλυνσης



Γράφημα 5.2: Σύγκριση σφαλμάτων εντοπισμού με ή χωρίς φίλτρο εξομάλυνσης

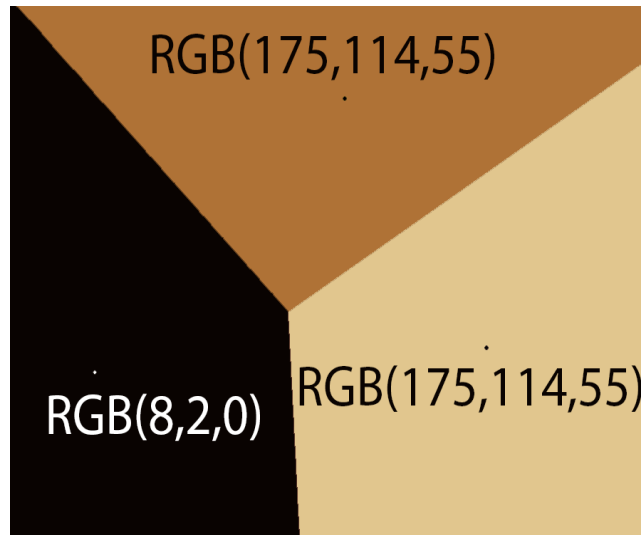
Το γράφημα 5.1 παρουσιάζει αρκετό ενδιαφέρον το στατιστικό του δείχνει σε τι ποσοστό εντοπίστηκαν με επιτυχία τα πατήματα των πλήκτρων. Όπως είναι εμφανές η εφαρμογή του φίλτρου βελτιώνει σημαντικά την ευστοχία του συστήματος από 67% σε 92%. Ενώ παράλληλα μειώνεται αρκετά ο αριθμός των εσφαλμένα εντοπισμένων πατημάτων από 14 σε 2. Συνολικά η ανάλυση ενός καρέ του βίντεο απαιτεί χωρίς φιλτράρισμα περίπου 129ms, ενώ για το ίδιο καρέ αν χρησιμοποιηθεί φιλτράρισμα ο χρόνος αυξάνεται στα 193ms. Συνδυάζοντας όλα τα παραπάνω και δίνοντας αυξημένη βαρύτητα στην απόδοση, η εφαρμογή φιλτραρίσματος θεωρείται απαραίτητη για την σωστή λειτουργία της εφαρμογής, εφόσον ότι θα χρησιμοποιηθεί αυτή η τεχνική τμηματοποίησης.

5.3.3 Χρωματική κβαντοποίηση

Η χρωματική κβαντοποίηση είναι μια τεχνική απωλεστικής συμπίεσης, κατά την οποία μειώνονται οι αποχρώσεις μιας εικόνας μέχρι έναν ορισμένο αριθμό. Η διαδικασία αυτή χωρίζεται σε 4 φάσεις [47]:

- 1) Δειγματοληψία της εικόνας για την εξαγωγή στατιστικών των χρωμάτων.
- 2) Επιλογή ενός χρωματικού χάρτη βασισμένου σε αυτά τα στατιστικά.
- 3) Αντικατάσταση κάθε χρώματος της εικόνας με το αντίστοιχο του στον χάρτη.
- 4) Κβαντοποίηση και σχεδιασμός της νέας εικόνας.

Για να επιτευχθεί η χρωματική κβαντοποίηση συνήθως χρησιμοποιείται ένας αλγόριθμος μέτρησης της απόστασης μεταξύ δυο χρωματικών αποχρώσεων, η οποία απόσταση καθορίζεται από την ευθεία γραμμή μεταξύ των σημείων των δυο χρωμάτων στον τρισδιάστατο χρωματικό χώρο. Με άλλα λόγια αν έχουμε δυο χρώματα $(r1, g1, b1)$ και $(r2, g2, b2)$ μπορούμε να υπολογίσουμε την ευκλείδεια απόσταση τους ως εξής: $\sqrt{(r1 - r2)^2 + (g1 - g2)^2 + (b1 - b2)^2}$. Με αυτή τη τεχνική ο χρωματικός κύβος αποσυντίθεται σε ένα διάγραμμα Voronoi όπως αυτό της εικόνας 5.8, στο οποίο όλα τα χρώματα που ανήκουν στο ίδιο κελί οποία αντιστοιχίζονται σε ένα κοινό χρώμα.



Εικόνα 5.8: Διάγραμμα Voronoi

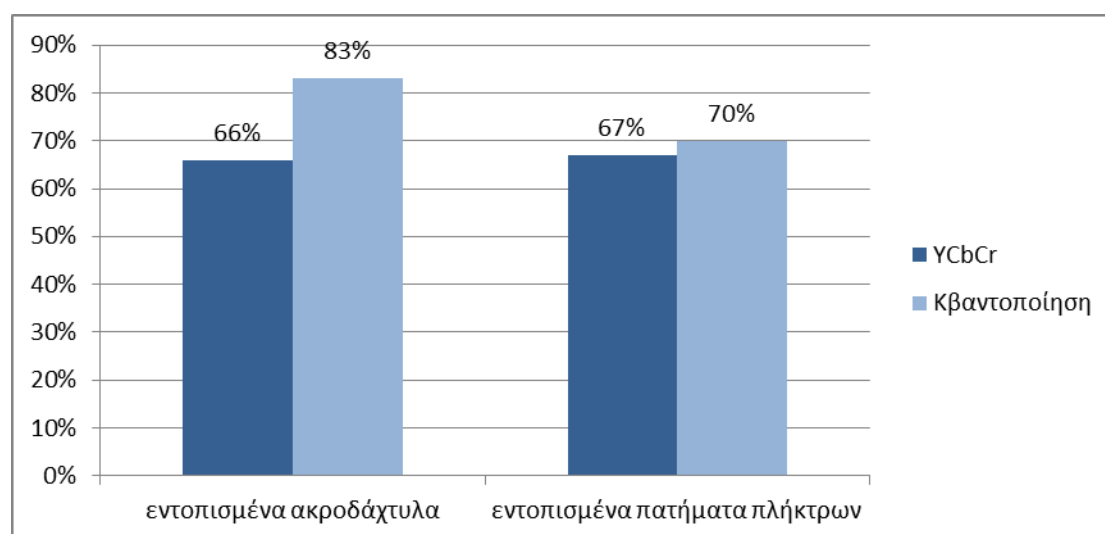
Στα πλαίσια της πτυχιακής εργασίας, δημιουργείται ένα διάγραμμα Voronoi αποτελούμενο από 3 κελιά για κάθε καρέ του βίντεο. Φυσικά θα μπορούσε να είχε δημιουργηθεί ένα καθολικό διάγραμμα αλλά το σύστημα δεν θα είχε την ίδια ευελιξία σε καταστάσεις μη ελεγχόμενης μεταβλητής φωτεινότητας. Για λόγους απλότητας προτιμήθηκε το φόντο να έχει κοινό χρώμα με το πιάνο με σκοπό να ανήκουν στο ίδιο κελί. Το ένα από τα άλλα δυο κελιά περιέχει όλες τις δερματικές αποχρώσεις, ενώ το τελευταίο περιέχει τα λευκά χρώματα που αντιστοιχούν στα πλήκτρα του πιάνο. Κατά προσέγγιση το αντιπροσωπευτικό χρώμα του κάθε κελιού για το βίντεο των δοκιμών φαίνεται παρακάτω:

Φόντο – Πιάνο: RGB(8,2,0)

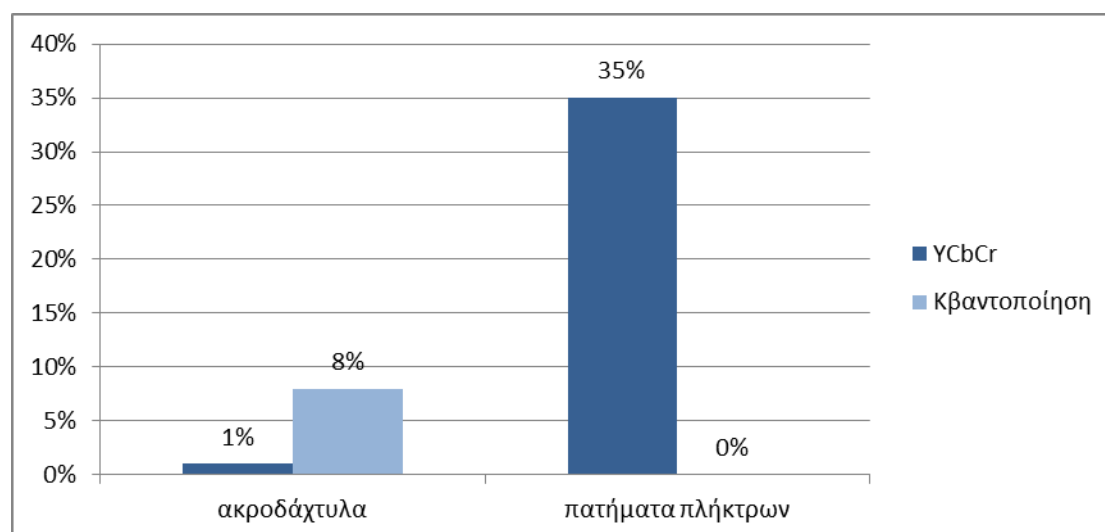
Χρωματικές αποχρώσεις: RGB(175,114,55)

Πλήκτρα πιάνο: RGB(225,199,142)

Περνώντας από την θεωρία στην πράξη, τα αποτελέσματα φαίνονται να είναι εντυπωσιακά. Πιο συγκεκριμένα, για την τμηματοποίηση μιας εικόνας με χρήση της νέας μεθόδου απαιτούνται μόλις 6ms. Σε σύγκριση με τα 46ms που απαιτούσε η τμηματοποίηση στον YCbCr η διαφορά μοιάζει χαοτική. Αξίζει να σημειωθεί ότι οι μετρήσεις έγιναν χωρίς το φίλτρο εξομάλυνσης καθώς έχουν γίνει κάποιες τροποποιήσεις στον εντοπισμό των ακροδαχτύλων και πλέον η εφαρμογή του φίλτρου δεν είναι αναγκαία. Στα παρακάτω γραφήματα μπορούμε να δούμε τα νέα ποσοστά ευστοχίας του συστήματος σε σύγκριση με αυτά της προηγούμενης μεθόδου.



Γράφημα 5.3: Ποσοστά ευστοχίας των μεθόδων YCbCr και κβαντοποίησης

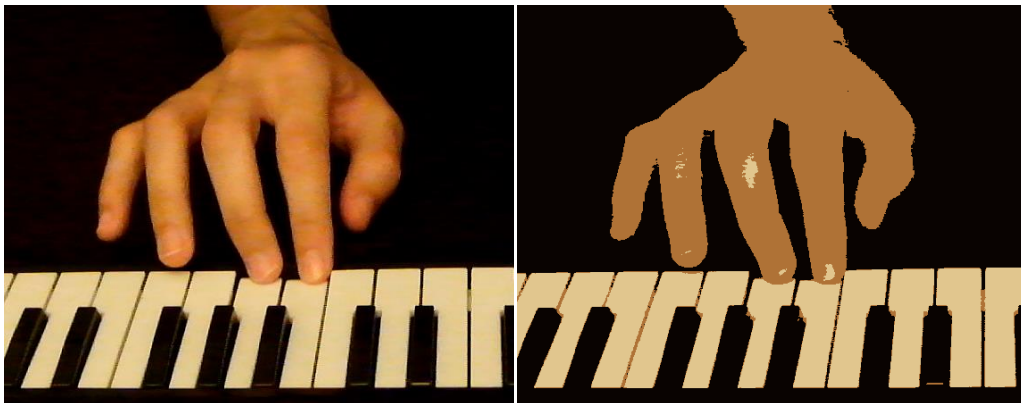


Γράφημα 5.2: Σύγκριση σφαλμάτων εντοπισμού των μεθόδων YCbCr και κβαντοποίησης

Από τα παραπάνω γραφήματα μπορούμε να εξάγουμε 2 συμπεράσματα. Το πρώτο είναι ότι ο εντοπισμός των χεριών δεν είναι τόσο ακριβής όσο πριν και αυτό φαίνεται από τα πολλά σφάλματα στον εντοπισμό των ακροδαχτύλων. Φυσικά υπάρχει αισθητή βελτίωση στον εντοπισμό των πλήκτρων που πατήθηκαν, καθώς έχει μηδενιστεί το σφάλμα. Τα ποσοστά του YCbCr όμως βελτιώνονται αρκετά όταν χρησιμοποιηθεί φιλτράρισμα, πράγμα που δεν συμβαίνει με την κβαντοποίηση, και εδώ τίθεται το ερώτημα: Γιατί προτιμήσαμε την τεχνική της κβαντοποίησης; Η απάντηση είναι απλή και βρίσκεται στους περιορισμούς της κάθε μεθόδου.

Όπως είδαμε στην εικόνα 5.4, επιλέξαμε μια παραλληλόγραμμη περιοχή στον YCbCr η οποία αντιπροσωπεύει τις δερματικές αποχρώσεις. Η περιοχή αυτή ορίζεται από τον χρήστη και επομένως είναι διαφορετική για κάθε βίντεο. Δεν υπάρχει κάποιος γενικός κανόνας που να προσαρμόζει τα όρια της περιοχής ανάλογα με τις απαιτήσεις του κάθε βίντεο. Αν για παράδειγμα προσπαθήσουμε να χρησιμοποιήσουμε την περιοχή που ορίστηκε στο βίντεο της δοκιμής, σε κάποιο άλλο δοκιμαστικό βίντεο τότε είναι πολύ πιθανόν να οδηγηθούμε σε σφάλμα. Αυτό συμβαίνει διότι είναι πολύ δύσκολο να αναπαράγουμε τις συνθήκες φωτισμού που έχουν καταγραφεί στο πρώτο βίντεο. Αν προσθέσουμε και το ενδεχόμενο να χρησιμοποιηθεί διαφορετική κάμερα, τότε η πιθανότητα να μπορέσουμε να εξάγουμε αποτελέσματα χωρίς να επαναπροσδιορίσουμε την περιοχή των δερματικών αποχρώσεων είναι σχεδόν μηδενική.

Αντιθέτως στην τεχνική της κβαντοποίησης, όπως είπαμε και προηγουμένως, δημιουργείται για κάθε καρέ του βίντεο ένα διάγραμμα Voronoi το οποίο δεν συνδέεται με αυτά των προηγούμενων καρέ. Αν συνυπολογίσουμε πως δεν είναι απαραίτητη η παραμετροποίηση του αλγορίθμου, μπορούμε πολύ εύκολα να αντιληφθούμε το πόσο πιο αποτελεσματική είναι αυτή η μέθοδος σε καταστάσεις μη ιδανικού φωτισμού. Ο χρήστης μπορεί πολύ εύκολα να τραβήξει ένα βίντεο χωρίς να ανησυχεί ιδιαίτερα για τον φωτισμό και στη συνέχεια να εξάγει τα αποτελέσματα που επιθυμεί χωρίς καμία αλληλεπίδραση με το σύστημα. Τα αποτελέσματα της εφαρμογής αυτής της τεχνικής φαίνονται στην εικόνα 5.9.



Εικόνα 5.9: Τμηματοποίηση χεριών με χρήση κβαντοποίησης

6 Ανίχνευση Ακροδαχτύλων

Υπό την προϋπόθεση ότι έχει γίνει τμηματοποίηση με μεγάλη ακρίβεια του επιθυμητού αντικειμένου, είναι δυνατόν να εξετασθεί το περίγραμμα με σκοπό τον εντοπισμό των ακροδαχτύλων. Σύμφωνα με το κεφάλαιο 5 η τμηματοποίηση που χρησιμοποιήθηκε σε αυτή την Πτυχιακή Εργασία είναι ικανή να εντοπίσει με μεγάλη αποτελεσματικότητα τα χρώματα του ανθρωπίνου δέρματος. Ως εκ τούτου, το περίγραμμα κάθε αντικειμένου είναι αρκετά ξεκάθαρο και επομένως κατάλληλο για χρήση.

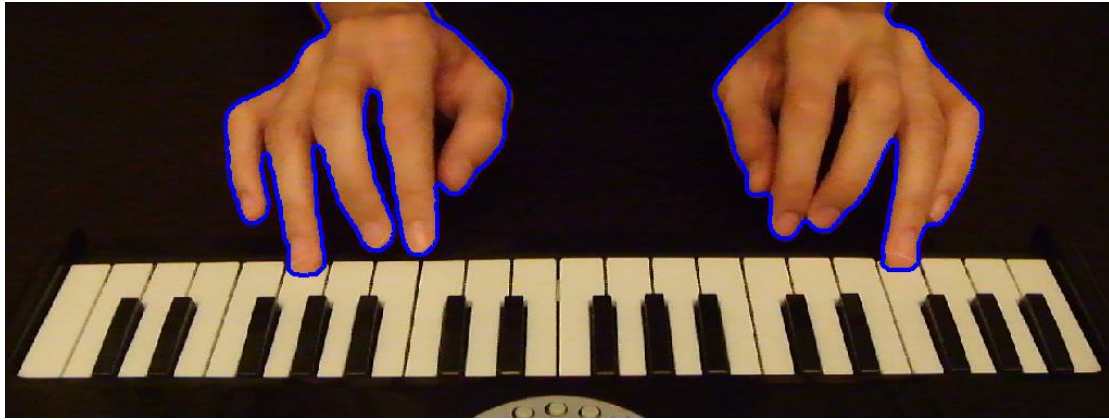
6.1 Εντοπισμός Περιγράμματος

Προτού όμως πραγματοποιηθεί η ανάλυση, είναι απαραίτητο να εντοπισθεί το περίγραμμα. Δεδομένου ότι έχει γίνει ήδη τμηματοποίηση της περιοχής του χεριού, μπορεί να εφαρμοσθεί κάποιος αλγόριθμος ιχνηλασίας γραμμών. Ο αλγόριθμος που επιλέχθηκε ονομάζεται Square Tracing [46], είναι ίσως ο πιο απλός της κατηγορίας και προτιμήθηκε λόγω του μικρού του υπολογιστικού κόστους, χωρίς όμως να υστερεί ποιοτικά στην συγκεκριμένη εφαρμογή από άλλους περισσότερο περίπλοκους. Αν θεωρηθούν γνωστά τα σημεία εκκίνησης και τερματισμού, η ακολουθία βημάτων έχει ως εξής:

Έστω ότι τοποθετείται ένα έντομο στο αρχικό εικονοστοιχείο έχοντας κατεύθυνση προς τα κάτω. Αν βρεθεί σε εικονοστοιχείο που ανήκει σε αντικείμενο στρίβει δεξιά κατά 90 μοίρες, ειδάλλως στρίβει αριστερά, επίσης κατά 90 μοίρες. Αφού αλλάξει κατεύθυνση κινείται στο επόμενο εικονοστοιχείο και επαναλαμβάνει την ίδια διαδικασία. Τα εικονοστοιχεία αντικειμένου που συνάντησε μέχρι να μεταβεί στο σημείο τερματισμού απαρτίζουν το ζητούμενο περίγραμμα. Τα σημεία εκκίνησης και τερματισμού εντοπίζονται αυτόματα, αρκεί να εξετασθεί η πρώτη γραμμή εικονοστοιχείων της εικόνας για τυχόν συνδεδεμένες περιοχές που ανήκουν σε κάποιο από τα χέρια του πιανίστα. Προφανώς το πρώτο εικονοστοιχείο της πρώτης από αριστερά προς δεξιά περιοχής, αντιστοιχεί στο σημείο εκκίνησης για το δεξί χέρι ενώ το τελευταίο εικονοστοιχείο της ίδιας περιοχής στο σημείο τερματισμού. Το ίδιο ισχύει και για την δεύτερη περιοχή που ανήκει στο αριστερό χέρι. Για την περίπτωση που εντοπισθεί μόνο ένα χέρι, θεωρείται πως είναι το δεξί λόγω του ότι υπάρχουν περισσότερες πιθανότητες ο πιανίστας να είναι δεξιόχειρας.

6.2 Εντοπισμός Ακροδαχτύλων

Ο εντοπισμός των τοπικών ελαχίστων της γραφικής αναπαράστασης του περιγράμματος των χεριών στο δισδιάστατο επίπεδο (εικόνα 6.1) είναι μια πολύ καλή επιλογή για τον εντοπισμό των ακροδαχτύλων, καθώς τα περισσότερα από αυτά ανήκουν σε ακροδάχτυλα. Με τη βοήθεια ενός κατωφλιού που καθορίζει το ελάχιστο μήκος καμπύλης μετρούμενο σε αριθμό εικονοστοιχείων ανάμεσα σε δυο ακροδάχτυλα, τα αποτελέσματα είναι αρκετά ικανοποιητικά. Είναι πιθανό φυσικά, όπως φαίνεται και στην εικόνα 6.2 να συνυπάρξουν δυο τοπικά ελάχιστα σε ένα δάχτυλο και ως εκ τούτου να εντοπισθούν περισσότερα από πέντε ακροδάχτυλα σε ένα χέρι. Σε αυτήν τη περίπτωση τα λανθασμένα ακροδάχτυλα εντοπίζονται αργότερα και δεν δημιουργείται εγγραφή για αυτά στον πίνακα αποτελεσμάτων. Δεδομένου ότι δυο τοπικά ελάχιστα που βρίσκονται πολύ κοντά δεν μπορούν να ανήκουν σε διαφορετικά ακροδάχτυλα, το σφάλμα μπορεί να ελαχιστοποιηθεί εάν εξετάζεται η μονοτονία ανά 15 εικονοστοιχεία, για τιμή κατωφλιού 15. Θα εξετασθούν δηλαδή κατά σειρά το 16^ο εικονοστοιχείο, το 32^ο, το 47^ο και ούτω καθεξής.

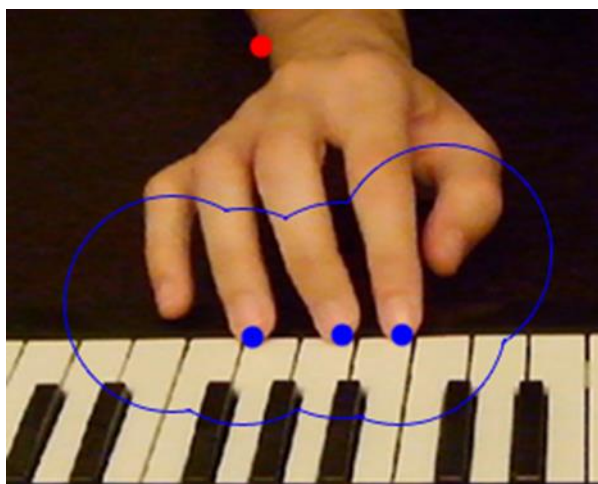


Εικόνα 6.1: Περίγραμμα χεριών



Εικόνα 6.2: Λανθασμένος εντοπισμός ακροδαχτύλων

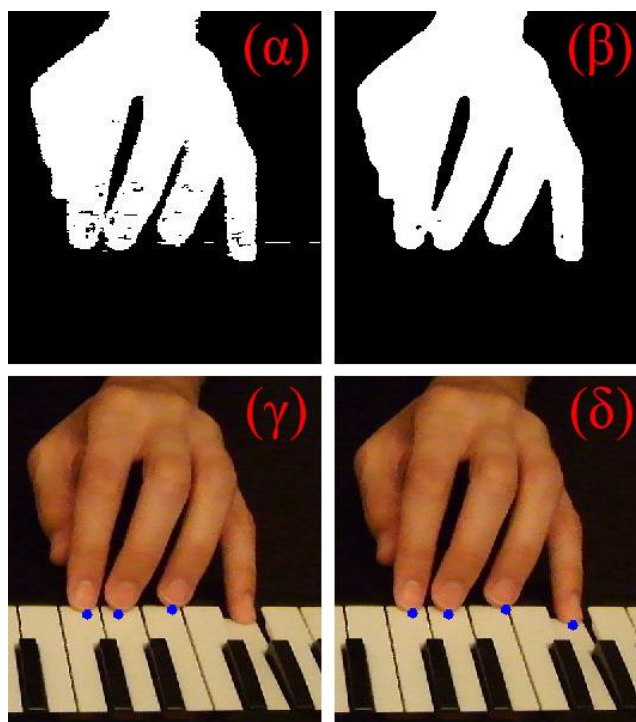
Ο σωστός εντοπισμός ενός ακροδαχτύλου έχει νόημα, εφόσον το δάχτυλο αυτό πιέζει κάποιο πλήκτρο, καθώς αν δεν το πιέζει είτε εντοπιστεί είτε όχι στον πίνακα αποτελεσμάτων θα καταχωρηθούν οι ίδιες τιμές. Ο εντοπισμός που έγινε στην εικόνα 6.4 είναι ικανοποιητικός καθώς παρέχει στον χρήστη όλες τις πληροφορίες που χρειάζεται. Δεδομένου ότι το πλήκτρο βρίσκεται υπό πίεση, όλα τα τοπικά ελάχιστα στο περίγραμμα του χεριού που βρίσκονται στην περιοχή του συγκεκριμένου πλήκτρου, θεωρούνται ότι ανήκουν στο ίδιο δάχτυλο οπότε καταγράφεται μόνο ένα πάτημα από ένα και μόνο δάχτυλο. Η περίπτωση δηλαδή που εντοπίζεται λανθασμένα ένα ακροδάχτυλο δεν δημιουργεί κάποιο πρόβλημα στη σωστή λειτουργία του συστήματος. Για να μειωθεί το ποσοστό των εσφαλμένα εντοπισμένων ακροδαχτύλων, εξετάζεται και η απόσταση του κάθε ακροδάχτυλου από όλα τα ακροδάχτυλα που εντοπίστηκαν στο προηγούμενο καρέ (εικόνα 6.3). Αν η απόσταση από ένα τουλάχιστον ακροδάχτυλο του προηγούμενου καρέ είναι μικρότερη του σταθερού κατωφλιού που έχει δηλωθεί εξ' αρχής, τότε ο εντοπισμός του ακροδάχτυλου είναι πιθανότατα σωστός. Αν σε καμία περίπτωση η απόσταση δεν είναι μικρότερη του κατωφλιού τότε δεν υπάρχει ακροδάχτυλο στο εν λόγω σημείο. Η απόσταση μετριέται με χρήση της συνάρτησης Manhattan distance. Το κατώφλι μέγιστης απόστασης έχει ρυθμιστεί στα 150 εικονοστοιχεία αλλά φυσικά εξαρτάται από την ανάλυση του βίντεο.



Εικόνα 6.3: Περιοχή ανίχνευσης ακροδαχτύλων



Εικόνα 6.4: Μη εντοπισμός ακροδαχτύλων που δεν πιέζουν πλήκτρα



Εικόνα 6.5: Εντοπισμός χεριών και ακροδάχτυλων (α),(γ) με χρήση φίλτρου εξομάλυνσης (β),(δ) χωρίς χρήση φίλτρου εξομάλυνσης

Αντιθέτως, ο μη εντοπισμός ενός ακροδαχτύλου, ιδιαίτερα αν αυτό πιέζει κάποιο πλήκτρο οδηγεί στον μη εντοπισμό του πατήματος του πλήκτρου και επομένως σε λάθος καταχώρηση στον πίνακα αποτελεσμάτων. Ο μη εντοπισμός μπορεί να οφείλεται είτε στην πολύ μικρή απόσταση μεταξύ δυο ακροδαχτύλων, μικρότερη του κατωφλιού που χρησιμοποιείται, είτε στην χρήση μη ομαλού περιγράμματος (εικόνα 6.5). Η πρώτη περίπτωση δεν μας απασχολεί δεδομένου ότι δυο δάχτυλα ενωμένα μεταξύ τους δεν μπορούν να πιέσουν δυο πλήκτρα ακόμα και αν αυτά βρίσκονται το ένα δίπλα στο άλλο. Αυτό φυσικά και δεν είναι αδύνατο, αλλά δε θα το έκανε ποτέ κανένας επαγγελματίας πιανίστας.

6.3 Ονοματοθεσία Ακροδαχτύλων

Αφού εντοπιστούν τα ακροδάχτυλα είναι απαραίτητο να γίνει γνωστή η ονομασία του καθενός. Για το λόγο αυτό ακολουθήθηκε η εξής διαδικασία: Όταν εντοπιστούν για πρώτη φορά 5 ακροδάχτυλα σε ένα χέρι, ονομάζονται όλα ένα-ένα με τη σειρά. Στα επόμενα καρέ εφόσον εντοπισθούν 5 ακροδάχτυλα θα ονομαστούν πάλι όλα ένα-ένα με την σειρά. Είναι πιθανό φυσικά, σε κάποιο καρέ να εντοπισθούν λιγότερα ή περισσότερα από 5 ακροδάχτυλα. Στην περίπτωση που τα ακροδάχτυλα είναι περισσότερα, υπολογίζεται η απόσταση του καθενός από το πλησιέστερο ακροδάχτυλο, με βάση την τελευταία γνωστή θέση του καθενός. Αφαιρούνται με τη σειρά τα ακροδάχτυλα ξεκινώντας από αυτό με τη μεγαλύτερη απόσταση (εικόνα 6.6) μέχρι να μείνουν 5 και να γίνει η γνωστή διαδικασία.

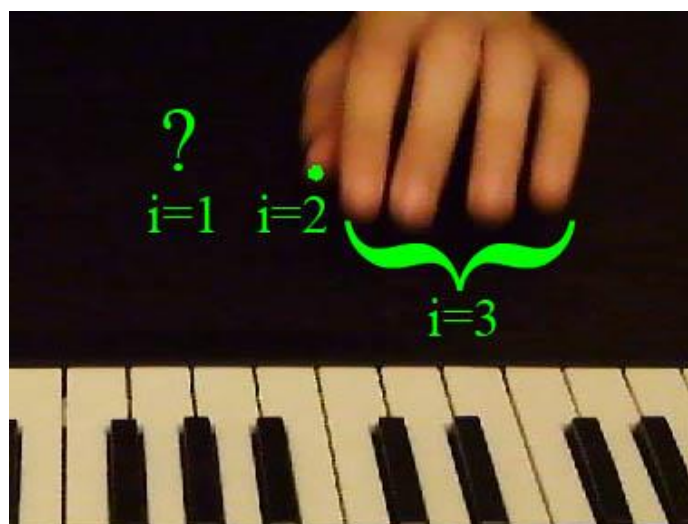


Εικόνα 6.6: Αφαίρεση λανθασμένα εντοπισμένου ακροδαχτύλου

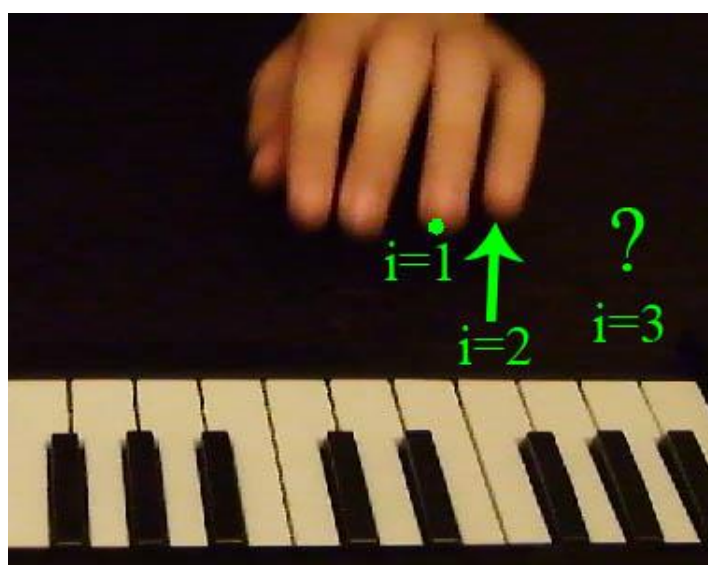
Αν όμως εντοπισθούν λιγότερα από 5 ακροδάχτυλα, το καθένα αντιστοιχίζεται στο πλησιέστερο ακροδάχτυλο που έχει εντοπιστεί στο παρελθόν, αφού πρώτα κάνουμε το εξής: Αριθμούμε με τη σειρά τα ακροδάχτυλα που εντοπίστηκαν σε αυτό το καρέ και αναφερόμαστε σε αυτά χρησιμοποιώντας τη μεταβλητή i . Προφανώς το i παίρνει τιμές από 1 έως τον αριθμό των εντοπισμένων ακροδαχτύλων τον οποίο θα ονομάσουμε S . Αριθμούμε επίσης και τα 5 ακροδάχτυλα για τα οποία ξέρουμε την τελευταία θέση που εντοπίστηκαν και αναφερόμαστε σε αυτά με χρήση της μεταβλητής j . Η τιμή του j καθορίζει το όνομα του ακροδαχτύλου i που δείχνει σε αυτό. Αρκεί να ισχύουν:

$$1^{\text{o}} \text{ κανόνας: } \boxed{i - j \leq 0}$$

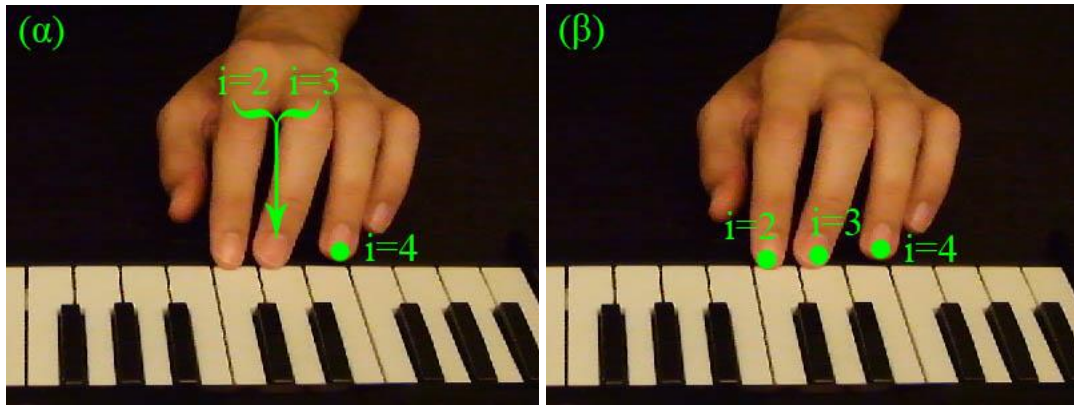
$$2^{\text{o}} \text{ κανόνας: } \boxed{i + 5 - j \geq S}$$

Εικόνα 6.7: Παράδειγμα 1^{ου} κανόνα ονοματοθεσίας ακροδαχτύλων

Αν για παράδειγμα έχουμε εντοπίσει 3 ακροδάχτυλα ($S=3$), το 2^ο ($i=2$) από αυτά δεν μπορεί να αντιστοιχεί στο 1^ο ($j=1$) από τα ήδη γνωστά ακροδάχτυλα γιατί $i - j = 1 > 0$ και επομένως δεν τηρείται ο πρώτος κανόνας. Η εξήγηση αυτού φαίνεται στην εικόνα 6.7 όπου στον αντίχειρα ($j=1$) δεν μπορεί να αντιστοιχεί το δάχτυλο $i=2$ γιατί το $i=1$ θα πρέπει να βρίσκεται αριστερά από αυτό και δεν υπάρχει άλλο δάχτυλο αριστερά του αντίχειρα.

Εικόνα 6.8: Παράδειγμα 2^{ου} κανόνα ονοματοθεσίας ακροδαχτύλων

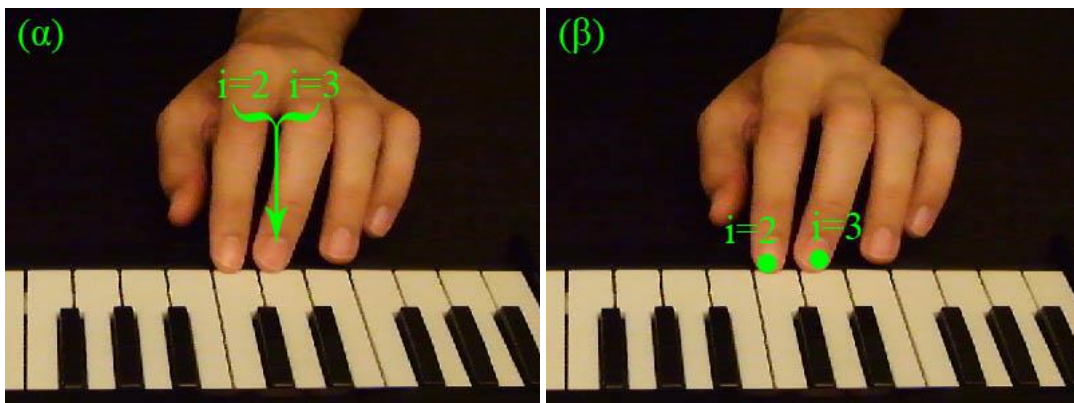
Έχοντας εντοπίσει ξανά 3 ακροδάχτυλα, σύμφωνα με τον 2^ο κανόνα δεν είναι δυνατόν το ακροδάχτυλο $i=1$ να αντιστοιχεί στο ακροδάχτυλο $j=4$ καθώς $1 + 5 - 4 \geq 3$ που προφανώς δεν ισχύει. Πρακτικά αυτό σημαίνει πως το δάχτυλο $i=1$ πρέπει να έχει δεξιά του, όπως φαίνεται στην εικόνα 6.8, τα ακροδάχτυλα $i=2$ και $i=3$ και επομένως δεν μπορεί να αντιστοιχεί στον παράμεσο.



Εικόνα 6.9: Διπλότυπη ονομασία ακροδαχτύλων με χρήση του 2^{ου} κανόνα (α) Πριν την εφαρμογή του κανόνα (β) μετά την εφαρμογή του κανόνα

Με την εφαρμογή των προηγούμενων κανόνων να μην γίνεται μια καλή προσέγγιση των θέσεων των ακροδαχτύλων αλλά πολλές φορές δυο ακροδάχτυλα μπορεί να πάρουν κοινή ονομασία. Για να διορθωθεί αυτό πρέπει να γίνει ένας επιπλέον έλεγχος. Έστω λοιπόν δυο ακροδάχτυλα $i=2$ και $i=3$ (όπου $i=2$ βρίσκεται πάντα αριστερά σε σχέση με το $i=3$) που διεκδικούν την ίδια θέση $j=3$. Αν υπάρχει i που αντιστοιχεί στη θέση $j=2$ και κανένα ακροδάχτυλο δε διεκδικεί τη θέση $j=4$, τότε το ακροδάχτυλο $i=3$ πρέπει να δείξει στη θέση $j=4$. Φυσικά ισχύει και το αντίθετο. Για το ίδιο παράδειγμα αν υπάρχει i που αντιστοιχεί στη θέση $j=4$ και κανένα ακροδάχτυλο δε διεκδικεί τη θέση $j=2$, τότε το ακροδάχτυλο $i=2$ πρέπει να δείξει στη θέση $j=2$ (εικόνα 6.9).

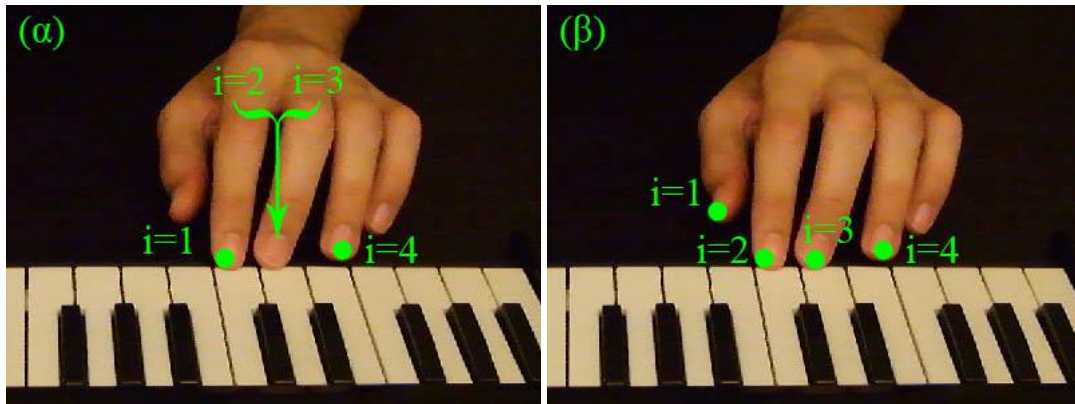
Στην περίπτωση που 2 ακροδάχτυλα έστω $i=2$ και $i=3$ διεκδικούν την ίδια θέση έστω $j=3$ και δεν υπάρχουν ακροδάχτυλα που διεκδικούν τις θέσεις $j=2$ και $j=4$, τότε αυτό που βρίσκεται πιο κοντά στη θέση $j=3$ κερδίζει και το άλλο θα πάρει κάποια από τις θέσεις $j=2$ ή $j=4$ ανάλογα με το ποια βρίσκεται πιο κοντά σε αυτό (εικόνα 6.10). Επειδή όμως το $i=2$ βρίσκεται πάντα αριστερά του $i=3$, αν χάσει το $i=2$ θα δείξει στην θέση $j=2$, ενώ αν χάσει το $i=3$ θα δείξει στη θέση $j=4$.



Εικόνα 6.10: Διπλότυπη ονομασία ακροδαχτύλων με χρήση του 3^{ου} κανόνα (α) Πριν την εφαρμογή του κανόνα (β) μετά την εφαρμογή του κανόνα (έστω ότι χάνει το $i=2$)

Υπάρχει όμως μια ακόμα περίπτωση όπου δυο ακροδάχτυλα διεκδικούν την ίδια θέση. Η περίπτωση που και η θέση $j-1$ αλλά και η $j+1$ είναι κατειλημμένες. Τότε το ακροδάχτυλο που βρίσκεται πιο κοντά στη θέση j κερδίζει και το άλλο θα δείξει στην θέση $j-1$ ή $j+1$ ανάλογα με τη θέση του, αλλά και αυτό που έδειχνε σε αυτή τη θέση θα πρέπει να

δείξει κάπου αλλού. Αν για παράδειγμα έχουμε τα ακροδάχτυλα $i=1$ να δείχνει στη θέση $j=2$, $i=2$ και $i=3$ να δείχνουν στη θέση $j=3$ και $i=4$ να δείχνει στη θέση $j=4$ και κερδίσει το ακροδάχτυλο $i=3$ τότε το $i=2$ θα δείξει στη θέση $j=2$ και το $i=1$ στη θέση $j=1$ (εικόνα 6.11).



Εικόνα 6.11: Διπλότυπη ονομασία ακροδαχτύλων με χρήση του 4^{ου} κανόνα (α) Πριν την εφαρμογή του κανόνα (β) μετά την εφαρμογή του κανόνα (έστω ότι χάνει το $i=2$)

Γενικά για οποιαδήποτε i_2 και i_3 που δείχνουν στην ίδια θέση j , με i_2 να βρίσκεται πάντα αριστερά του i_3 , ισχύουν οι εξής κανόνες:

1^{ος} κανόνας:

$$\forall \exists i \rightarrow j - 1 \wedge \exists i \rightarrow j + 1 \text{ τότε } i_3 \rightarrow j + 1$$

2^{ος} κανόνας:

$$\forall \exists i \rightarrow j + 1 \wedge \exists i \rightarrow j - 1 \text{ τότε } i_2 \rightarrow j - 1$$

3^{ος} κανόνας:

$$\forall \exists i \rightarrow j - 1 \wedge i \rightarrow j + 1 \text{ τότε} \\ \forall \exists d(i_2, j) > d(i_3, j) \text{ τότε } i_2 \rightarrow j - 1 \\ \text{αλλιώς } i_3 \rightarrow j + 1$$

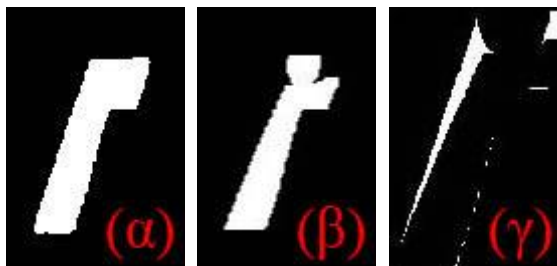
4^{ος} κανόνας:

$$\forall \exists i_1 \rightarrow j - 1 \wedge i_4 \rightarrow j + 1 \text{ τότε} \\ \forall \exists d(i_2, j) > d(i_3, j) \text{ τότε } i_2 \rightarrow j - 1 \wedge i_1 \rightarrow j - 2 \\ \text{αλλιώς } i_3 \rightarrow j + 1 \wedge i_4 \rightarrow j + 2$$

7 Εντοπισμός Πατήματος Πλήκτρων

Στο τελικό στάδιο του αλγορίθμου, τα πλήκτρα εξετάζονται για το αν βρίσκονται υπό πίεση. Βεβαίως, δεν υπάρχει λόγος να εξεταστούν όλα τα πλήκτρα, αλλά μόνο όσα έχει εισχωρήσει κάποιο δάχτυλο στην περιοχή τους.

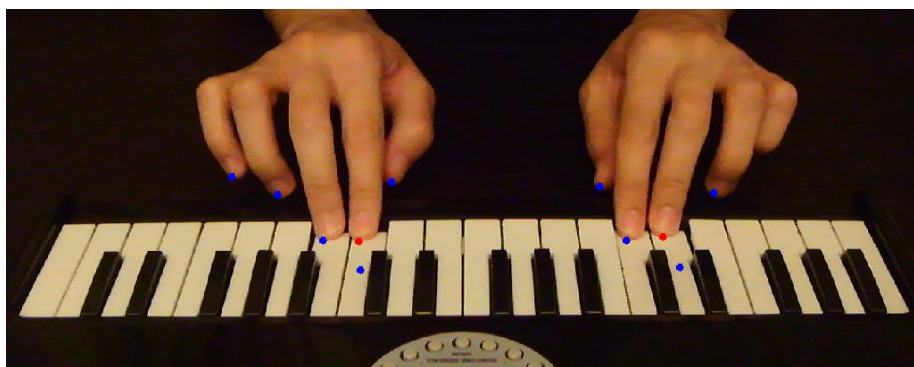
7.1 1^η Προσέγγιση



Εικόνα 7.1: Σύστημα απόφασης (α) Η περιοχή του πλήκτρου στην αρχή του βίντεο (β) Η περιοχή του πλήκτρου στο καρτέ που εξετάζεται (γ) Οι διαφορές μεταξύ των δυο εικόνων

Προκειμένου να ληφθεί η απόφαση για το αν πιέζεται ένα πλήκτρο, μια πρώτη προσέγγιση είναι να εξετασθεί η τωρινή του κατάσταση (εικόνα 7.1(β)) σε σύγκριση με την αρχική (εικόνα 7.1(α)). Με χρήση του ίδιου κανόνα που χρησιμοποιήθηκε στον εντοπισμό των πλήκτρων στην αρχή της επεξεργασίας, αυτή τη φορά με χρήση της ήδη γνωστής τιμής κατώφλιού που υπολογίστηκε νωρίτερα, είναι δυνατόν να εντοπισθούν ξανά τα πλήκτρα σε κάθε καρτέ. Βέβαια για να γίνει σωστή τμηματοποίηση υπήρχε ο περιορισμός να μην υπάρχει κάποιο χέρι στην εικόνα, οπότε από τη νέα δυαδική εικόνα που παράγεται από την τμηματοποίηση πρέπει να αφαιρεθεί η δυαδική εικόνα που δημιουργήθηκε από την τμηματοποίηση των χεριών σε αυτό το καρτέ.

Για κάθε πλήκτρο που υπάρχει ανάγκη να ελεγχθεί η κατάσταση του, εντοπίζονται όλα τα εικονοστοιχεία του πλήκτρου που έχουν υποστεί μεταβολή φωτεινότητας μεταξύ των δυο δυαδικών εικόνων όπως για παράδειγμα στην εικόνα 7.1(γ). Στη συνέχεια εξετάζεται αν ο αριθμός των εικονοστοιχείων που διαφέρουν είναι αρκετά μεγάλος και εφόσον το αποτέλεσμα ξεπερνάει κάποιο κατώφλι καταγράφεται το πάτημα του πλήκτρου. Τα αποτελέσματα είναι εύστοχα σε ικανοποιητικό ποσοστό, αλλά υπάρχουν περιθώρια βελτίωσης. Το κατώφλι που χρησιμοποιήθηκε ήταν σταθερό και ίδιο για όλα τα πλήκτρα, πράγμα που όπως φάνηκε αργότερα δεν είναι αποδοτικό (εικόνα 7.2).



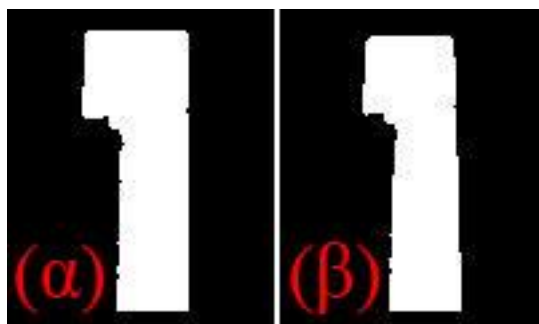
Εικόνα 7.2: Σφάλμα σταθερού κατώφλιού

7.2 2^η Προσέγγιση

Παρατηρώντας καλύτερα τα πλήκτρα μπορεί να γίνει μια πολύ σημαντική παρατήρηση · Όταν αλλάζει η κατάσταση ενός πλήκτρου που βρίσκεται μακριά από το κέντρο της εικόνας, η αποτύπωση του στο επίπεδο φαίνεται να έχει μια κλίση σε σχέση με την αρχική του κατάσταση (εικόνα 7.3). Όσο μεγαλύτερη είναι η απόσταση από το κέντρο της εικόνας, τόσο μεγαλύτερη είναι και η κλίση. Η μετατόπιση των γωνιών στην κορυφή του αντικειμένου δείχνει την κλίση και εφόσον δεν είναι μηδενική μπορεί να ειπωθεί με σιγουριά ότι το πλήκτρο βρίσκεται υπό πίεση. Όσον αφορά τα πλήκτρα που βρίσκονται κοντά στο κέντρο της εικόνας, όταν πιέζονται αποκτούν μια αίσθηση γραμμικής προοπτικής όπως στην εικόνα 7.4. Με πιο απλά λόγια, οι κάθετες στο επίπεδο ευθείες του παραλληλογράμμου που σχηματίζει το αντικείμενο τείνουν να συγκλίνουν όσο πλησιάζουν το πάνω μέρος της εικόνας. Ως εκ τούτου, η απόσταση των γωνιών της κορυφής του αντικειμένου είναι μικρότερη σε σχέση με αυτή της αρχικής του κατάστασης. Χρησιμοποιώντας αισθητά λιγότερους υπολογισμούς σε σχέση με πριν, μειώνεται σημαντικά ο χρόνος ανάλυσης από τα 40ms σε μόλις 3ms. Επίσης, λόγω της καλύτερα εμπεριστατωμένης άποψης της νέας βελτιωμένης προσέγγισης, πέρα από την ταχύτητα βελτιώνεται και η απόδοση ελαττώνοντας αισθητά το ποσοστό σφάλματος.



Εικόνα 7.3: Πλήκτρο μακριά από το κέντρο της εικόνας (α) Η αρχική του κατάσταση (β) Η κλίση που αποκτά όταν βρίσκεται υπό πίεση

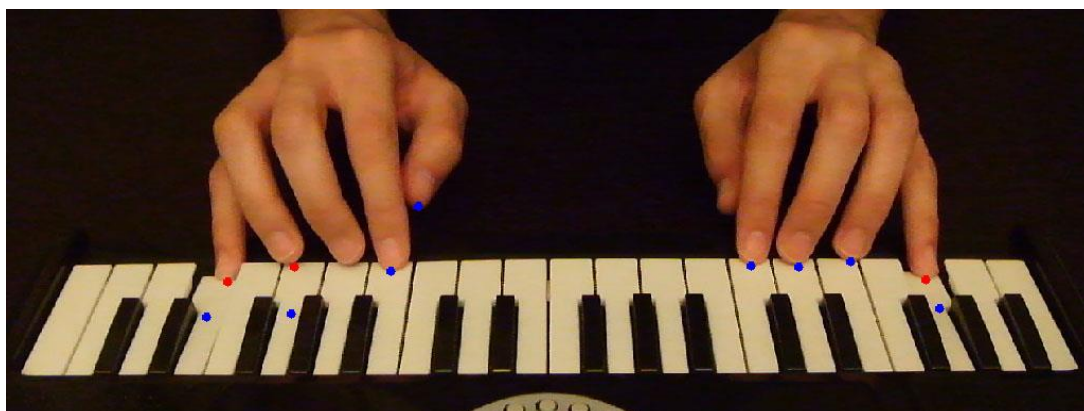


Εικόνα 7.4: Πλήκτρο κοντά στο κέντρο της εικόνας (α) Η αρχική του κατάσταση (β) Η αίσθηση γραμμικής προοπτικής που αποκτά όταν βρίσκεται υπό πίεση

7.3 Σύγκριση Μεθόδων

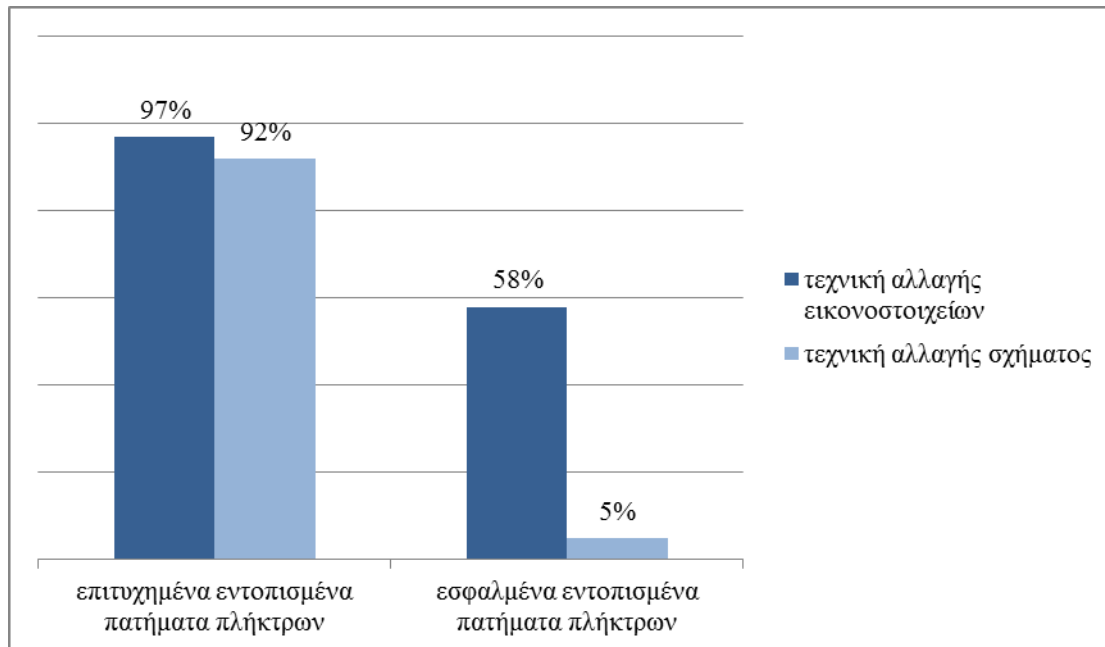
Σύμφωνα με το γράφημα 7.1, τα εσφαλμένα εντοπισμένα πατήματα πλήκτρων χρησιμοποιώντας τη νέα μέθοδο είναι 14 φορές λιγότερα σε σχέση με πριν. Στην εικόνα 7.5 φαίνεται ότι ένα δάχτυλο έχει σημειωθεί πως πιέζει το πλήκτρο πάνω στο οποίο βρίσκεται ενώ στην πραγματικότητα δεν συμβαίνει κάτι τέτοιο. Με τον νέο σύστημα τέτοια σφάλματα

τυχαίνουν σε ποσοστό 4%. Σε σχέση με το 46% της προηγούμενης τεχνικής έχουμε ένα αρκετά μεγάλο χάσμα.



Εικόνα 7.5: Εσφαλμένα εντοπισμένο πάτημα πλήκτρου

Τα αποτελέσματα των δοκιμών που παρουσιάστηκαν στο κεφάλαιο 5 έγιναν με χρήση της 2^{ης} τεχνικής και επομένως η απόδοση της είναι ήδη γνωστή. Στα 100 καρέ που χρησιμοποιήθηκαν στις δοκιμές είχε ευστοχία σε ποσοστό 92% ενώ εντόπισε λανθασμένα 4 πατήματα πλήκτρων. Η 1^η τεχνική υποβλήθηκε στην ίδια διαδικασία δοκιμών και είχε ευστοχία 97% αλλά 56 εσφαλμένα πατήματα. Δεδομένης και της μεγαλύτερης ταχύτητας της 2^{ης} τεχνικής, που είναι 3,5 φορές πιο γρήγορη από την 1^η, η βέλτιστη λύση είναι ξεκάθαρη.



Γράφημα 7.1: Σύγκριση προτεινόμενων μεθόδων

Συνδυάζοντας τα δεδομένα του γραφήματος 7.1 με την εικόνα 7.2 γεννάται το εξής ερώτημα: «Η 1^η τεχνική έχει μεγαλύτερο ποσοστό επιτυχίας, μήπως αν αντί για σταθερό κατώφλι θα μπορούσε να δημιουργηθεί κάποια ρουτίνα με σκοπό τον υπολογισμό

μεταβλητού καταφλιού ξεχωριστά για κάθε πλήκτρο; Δεν θα ήταν λιγότερα τα εσφαλμένα εντοπισμένα πατήματα;» Βεβαίως και είναι δυνατόν να γίνει αυτό και είναι πολύ πιθανό να έχει εξίσου καλά, ίσως και καλύτερα αποτελέσματα από την τεχνική που χρησιμοποιείται. Η απάντηση του γιατί δεν έγινε μια τέτοια προσπάθεια βρίσκεται στο γράφημα 7.1 και πιο συγκεκριμένα στον μέσο χρόνο που χρειάζεται η κάθε τεχνική για να αναλύσει ένα καρέ. Το 5% της αυξημένης απόδοσης που θα μπορούσε να έχει το σύστημα είναι πολύ μικρό για να θυσιάσουν 37ms χρόνου επεξεργασίας για κάθε καρέ.



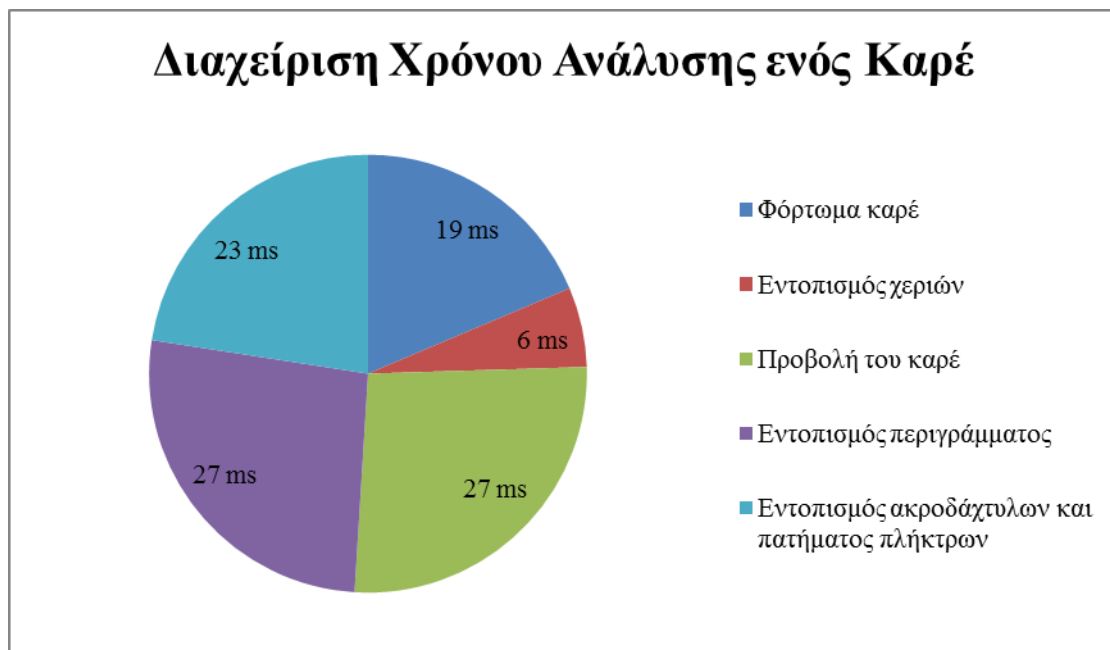
Εικόνα 7.6: Εντοπισμός ακροδαχτύλων και πατημάτων πλήκτρων

8 Συμπεράσματα

Όπως είδαμε και στο γράφημα 7.1 του προηγούμενου κεφαλαίου, το σύστημα είναι ικανό να εντοπίσει το πάτημα ενός πλήκτρου σε ποσοστό 70%. Το νούμερο αυτό είναι αν μη τι άλλο ικανοποιητικό και πολλά υποσχόμενο.

8.1 Επισκόπηση του Συστήματος

Για τη σωστή λειτουργία του συστήματος είναι απαραίτητο να έχουν τηρηθεί όλες οι προϋποθέσεις για την καταγραφή του βίντεο. Αν για παράδειγμα ο πιανίστας φοράει διάφορα αξεσουάρ στα χέρια του, ο εντοπισμός των χεριών θα αποτύχει. Ο λόγος που δεν υπήρξε πρόβλεψη για κάτι τέτοιο είναι καθαρά για να μειωθεί ο χρόνος ανάλυσης ενός καρέ. Η θέση της κάμερας είναι πολύ σημαντική γιατί ο εντοπισμός ενός πατήματος κρίνεται σε λεπτομέρειες και η λανθασμένη τοποθέτηση της κάμερας μπορεί να τον επηρεάσει αρνητικά. Εξίσου σημαντική είναι η θέση της πηγής φωτός καθώς και η ένταση του φωτός για τον ίδιο ακριβώς λόγο.

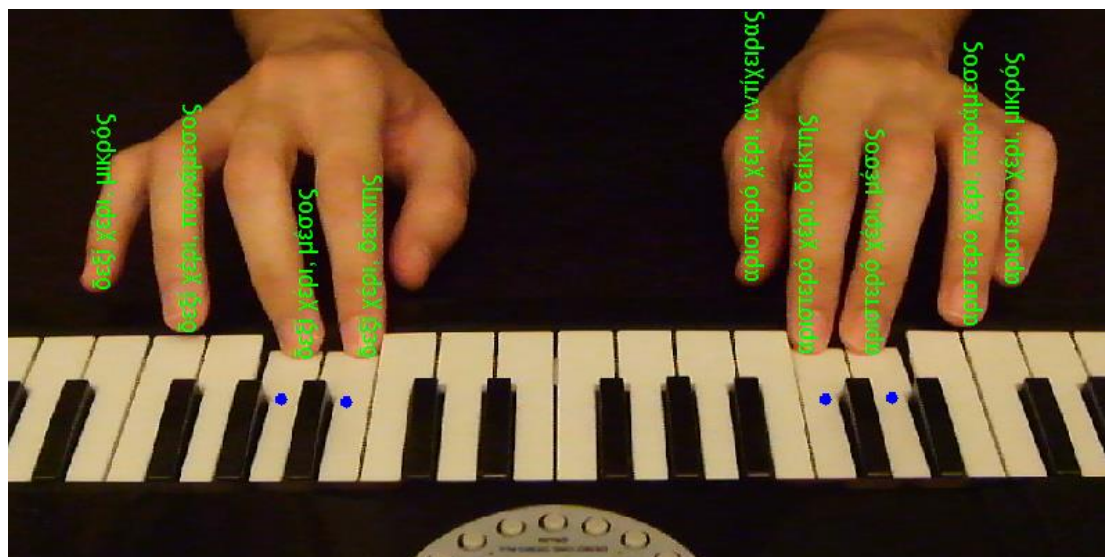


Γράφημα 8.1: Διαχείριση χρόνου ανάλυσης ενός καρέ

Ο εντοπισμός των χεριών έγινε με χρωματική κβαντοποίηση. Έχει πολύ καλά αποτελέσματα ακόμα και σε περιπτώσεις κακού φωτισμού. Η μοναδική απαίτηση της τεχνικής αυτής είναι να είναι γνωστός και σταθερός ο αριθμός των αποχρώσεων μεταξύ των καρέ. Για την συγκεκριμένη εφαρμογή αυτό δεν είναι πρόβλημα, καθώς λόγω της θέσης και της κλίσης της κάμερας μπορούμε να είμαστε βέβαιοι πως οι αποχρώσεις που εμφανίζονται στο οπτικό της πεδίο είναι τρεις με τέσσερις (πιάνο, πλήκτρα, φόντο, χέρια).

Για τον εντοπισμό των ακροδαχτύλων πρέπει πρώτα να γίνει εξαγωγή του περιγράμματος από την εικόνα των χεριών, η οποία επιτυγχάνεται με τη χρήση ενός πολύ απλού αλγορίθμου. Στη συνέχεια σημειώνονται όλα τα τοπικά ελάχιστα του περιγράμματος σαν πιθανά ακροδάχτυλα. Με την εφαρμογή των κανόνων που αναλύθηκαν στο κεφάλαιο 6, κάποια από αυτά θα σημειωθούν σαν ακροδάχτυλα και θα πάρουν το καθένα το όνομα του ακροδάχτυλου στο οποίο αντιστοιχεί.

Αφού εντοπισθούν τα ακροδάχτυλα, το επόμενο βήμα είναι να παρθεί μια απόφαση για το ποια από αυτά πιέζουν κάποιο πλήκτρο. Αυτό επιτυγχάνεται παρατηρώντας τις διαφορές στο σχήμα του πλήκτρου σε σχέση με την αρχική του κατάσταση. Ένα παράδειγμα εφαρμογής του συστήματος φαίνεται στην εικόνα 8.1



Εικόνα 8.1: Οπτική απεικόνιση των αποτελεσμάτων του συστήματος

8.2 Μελλοντική δουλειά

Ο αλγόριθμος ναι μεν παράγει αρκετά καλά αποτελέσματα για τα λευκά πλήκτρα, δεν έχει γίνει όμως πρόβλεψη για τα μαύρα πλήκτρα. Το πάτημα ενός μαύρου πλήκτρου πολλές φορές είναι δύσκολο να εντοπιστεί μόνο με την όραση ακόμα και από τον άνθρωπο, καθώς τυχόν αλλαγές στο σχήμα του πλήκτρου καλύπτονται από το δάχτυλο που το πατάει.

Από τον πίνακα αποτελεσμάτων που παράγεται με την εφαρμογή του αλγορίθμου σε ένα βίντεο μπορεί να δημιουργηθεί ένα midi αρχείο το οποίο θα αναπαράγει τον ήχο από την εκτέλεση του μουσικού κομματιού που καταγράφηκε. Επιπλέον, από τις εγγραφές του ίδιου πίνακα είναι εφικτό να κατασκευαστεί μια τρισδιάστατη απόδοση του βίντεο η οποία σε συνδυασμό με τον ήχο από το midi αρχείο μπορεί να βοηθήσει σημαντικά στην έρευνα που γίνεται γύρω από την κατανόηση του τρόπου έκφρασης ενός μουσικού, αλλά και στην εκπαίδευση μαθητευόμενων μουσικών.

Αναφορές

- [1] P. Herrera-Boyer, A. Klapuri, and M. Davy. Automatic classification of pitched musical instrument sounds. In *Signal Processing Methods for Music Transcription*, pages 163–200. Springer, 2006.
- [2] T. Marrin Nakra, Possibilities for the Digital Baton as a General-Purpose Gestural Interface, *Computer Human Interaction Conference on Human Factors in Computing Systems*, pages 311-312, 1997.
- [3] T. Marrin Nakra and R. Picard, The Conductor's Jacket: A Device for Recording Expressive Musical Gestures, *International Computer Music Conference*, Ann Arbor, MI, pages 215-219, October 1998.
- [4] T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. Okuno. Musical instrument recognizer "instrogram" and its application to music retrieval based on instrumentation similarity. *International Symposium on Multimedia*, pages 265–274, 2006.
- [5] T. Marrin Nakra and J. Paradiso, The Digital Baton: a Versatile Performance Instrument, *International Computer Music Conference*, Thessaloniki, Greece, pages 313-316, 1997.
- [6] T. Marrin Nakra, Inside the Conductors Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture, Ph.D. Thesis, Media Laboratory. Cambridge, MA, Massachusetts Institute of Technology, February 2000.
- [7] S. Essid, G. Richard, and David.B. Instrument recognition in polyphonic music based on automatic taxonomies. *IEEE Transactions on Audio, Speech & Language Processing*, 14(1):68–80, 2006.
- [8] H. Fujihara, T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno. Singer identification based on accompaniment sound reduction and reliable frame selection. In *Proc. ISMIR 2005*, pages 329–336, 2005.
- [9] T. Ilmonen, Tracking Conductor of an Orchestra Using Artificial Neural Networks, Master's thesis, Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology, 1999.
- [10] P. Leveau, E. Vincent, G. Richard, and L. Daudet. Instrument-specific harmonic atoms for mid-level music representation. *IEEE Transactions on Audio, Speech & Language Processing*, 16(1):116–128, 2008.
- [11] P. Jinchaitra. Polyphonic instrument identification using independent subspace analysis. In *Proceeding of the IEEE International Conference on Multimedia and Expo*, pages 1211–1214. IEEE, 2004.

- [12] T. Virtanen and A. Klapuri. Analysis of polyphonic audio using source-filter model and non-negative matrix factorization. In *Advances in Models for Acoustic Processing, Neural Information Processing Systems Workshop*, 2006.
- [13] D. FitzGerald, M. Cranitch, and E. Coyle. Extended nonnegative tensor factorisation models for musical source separation. *Computational Intelligence and Neuroscience*, 2008.
- [14] R. Badeau, V. Emiya, and B. David. Expectationmaximization algorithm for multi-pitch estimation and separation of overlapping harmonic spectra. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009.
- [15] E. Vincent, N. Bertin, and R. Badeau. Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 109–112, 2008.
- [16] T. Heittola, A. Klapuri, and T. Virtanen. Musical instrument recognition in polyphonic audio using source-filter model for sound separation. *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, 2009.
- [17] T. Aach and A. Kaup. Bayesian algorithms for adaptive change detection in image sequences using markov random fields. *Signal Processing: Image Communication*, 7, 1995.
- [18] K. Jack. *Video Demystified, Fourth Edition (Demystifying Technology)*. 2004.
- [19] L. Tarabella, pCM (pure C Music) A Real-Time Music Language, *Proceedings of Generative ART 2002*, Politecnico di Milano, 2002.
- [20] V. Lalioti, A. Malan, J. Pun and J. Wind, Ndebele Painting in VR, *IEEE Computer Graphics and Applications*, volume 20, number 6, pages 54-65, November/December 2000.
- [21] J. Lane, *Using Catadioptrics for Multidimensional Interaction in Computer Graphics*, Thesis for Master of Science, Department of Computer Science, University of Pretoria, South Africa, 2002.
- [22] C. Näolker and H. Ritter, Detection of Fingertips in Human Hand Movement Sequences, *Gesture and Sign Language in Human-Computer Interaction, Proceedings of the International Gesture Workshop 1997*, pages 209-218.
- [23] D. Overholt, *The MATRIX: A Novel Controller*, MIT Media Laboratory, Massachusetts Institute of Technology, 2001.

- [24] J. Paradiso, K. Hsiao, A. Benbasat and Z. Teegarden, Design and Implementation of Expressive Footwear, IBM Systems Journal, volume 39, numbers 3 & 4, pages 511-529, October 2000.
- [25] V. Pavlovic, R. Sharma and T. Huang, Visual Interpretation of Hand Gestures for Human computer Interaction: A Review, IEEE Transaction on Pattern Analysis and Machine Intelligence, pages 677-695, volume 19, number 7, July 1997.
- [26] L. Tarabella, G. Bertini, T. Sabbatini, The Twin Towers: a remote sensing device for controlling live-interactive computer music, Proceedings of Workshop on Mechatronical Computer Systems for Perception and Action, Pisa, 1997.
- [27] J. Davis and M. Sash, Determining 3D Hand Motion, Proceedings of the 28th Asilomar Conference on Signals, Systems and Computer, 1994.
- [28] C. Maggioni, Gesture computer | New Ways of Operating a Computer, Proceedings of International Workshop on Automatic Face and Gesture-Recognition, pages 166-171, June 1995.
- [29] Y. Kuno, M. Sakamoto, K. Sakata, and Y. Shirai, Vision Based Human Computer Interface with User Centered Frame, Proceedings of International Conference on Intelligent Robots and Systems, pages 2923-2929, 1994.
- [30] D. J. Sturman and D. Zeltzer, A Survey of Glove-Based Input, IEEE Computer Graphics and Applications, volume 14, pages 30-39, January 1994.
- [31] D. L. Quam, Gesture Recognition with a Dataglove, Proceedings of the 1990 IEEE National Aerospace and Electronics Conference, volume 2, 1990.
- [32] F. K. H. Quek, Eyes in the Interface, Image and Vision Computing, volume 13, August 1995.
- [33] W. T. Freeman and C. D. Weissman, Television Control by Hand Gestures, Proceedings of International Workshop on Automatic Face and Gesture-Recognition, pages 179-183, June 1995.
- [34] E. Clergue, M. Goldberg, N. Madrane, and B. Merialdo, Automatic Face and Gestural Recognition for Video Indexing, Proceedings of International Workshop on Automatic Face and Gesture-Recognition, pages 110-115, June 1995.
- [35] A. C. Downton and H. Doret, Image Analysis for Model-Based Sign Language Coding, Progress in Image Analysis and Processing, in Proceedings of the 6th International Conference on Image Analysis and Processing, pages 637-644, 1991.

[36] D. M. Gavrila and L. S. Davis, towards 3D Model Based Tracking and Recognition of Human Movement: A Multi View Approach, Proceedings of International Workshop on Automatic Face and Gesture-Recognition, pages 272-277, June 1995.

[37] J. Lee and T. L. Kunii, Model-Based Analysis of Hand Posture, IEEE Computer Graphics and Applications, pages 77-86, September 1995.

[38] J. M. Rehg and T. Kanade, Visual Tracking of Self-Occluding Articulated Objects, Technical Report 224, School of Computer Science, Carnegie Mellon University, Pittsburgh PA, December 1994.

[39] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, Active Shape Models Their Training and Application, Computer Vision and Image Understanding, volume 61, pages 38-59, January 1995.

[40] C. Kervrann and F. Heitz, Learning Structure and Deformation Modes of Non-rigid Objects in Long Image Sequences, Proceedings of International Workshop on Automatic Face and Gesture-Recognition, June 1995.

[41] A. Lanitis, C. J. Taylor, T. F. Cootes and T. Ahmed, Automatic Interpretation of Human Faces and Hand Gestures Using Flexible Models, Proceedings of International Workshop on Automatic Face- and Gesture-Recognition, pages 98-103, June 1995.

[42] T. Darrell and A. Pentland, Space-Time Gestures, Proceedings of Computer Vision and Pattern Recognition Conference, 1993.

[43] T. Darrell and A. Pentland, Attention Driven Expression and Gesture Analysis in an Interactive Environment, Proceedings of International Workshop on Automatic Face and Gesture-Recognition, pages 135-140, June 1995.

[44] J. L. Crowley, F. Berard and J. Coutaz, Finger Tracking as an Input Device for Augmented Reality, Proceedings of International Workshop on Automatic Face and Gesture-Recognition, pages 195-200, June 1995.

[45] V. Vezhnevets, V. Sazonov, A. Andreeva. A Survey on Pixel-Based Skin Color Detection Techniques, In Proceedings of the GraphiCon 2003, pp. 85-92, 2003.

[46] http://www.imageprocessingplace.com/downloads_V3/root_downloads/tutorials/contour_tracing_Abeer_George_Ghuneim/square.html

[47] Heckbert, P. "Color Image Quantization for Frame Buffer Display", Computer Graphics, Vol 16, #3, pp. 297-303, 1982

Παράρτημα Α: Πηγαίος Κώδικας

Το κυρίως μέρος του προγράμματος:

```

close all;
clear;
clc;
movieObj = VideoReader('DSCF0027.AVI');
nFrames = uint32(movieObj.NumberOfFrames);
vidHeight = movieObj.Height;
vidWidth = uint16(movieObj.Width);
numOfKeys = uint8(22);
mid = zeros(nFrames, numOfKeys, 'uint8');
mov = read(movieObj, 100);
[keys, keyStruct, keyDetectThreshold] =
keysDetection(mov, numOfKeys, vidHeight);
keys = uint8(bwlabel(keys, 8));
fingerKey =
struct('x', {0,0,0,0,0;0,0,0,0,0}, 'y', {0,0,0,0,0;0,0,0,0,0}, 'finger', {
'null', 'null', 'null', 'null', 'null'; 'null', 'null', 'null', 'null', 'null'
}, 'key', {'null', 'null', 'null', 'null', 'null'; 'null', 'null', 'null', 'nul
l', 'null'});
finDtect(2) = false;
results = cell(nFrames, numOfKeys);
results(:) = {'null'};
for k = uint32(175) : 1 : nFrames
    mov = read(movieObj, k);
    [handImg, map] = rgb2ind(mov, 3, 'nodither');
    figure, imshow(mov);
    c2 = squareTracing(handImg);
    if size(c2, 2) > 2
        [hands, existH(1), existH(2)] = handPoints(c2);
    else
        existH(1) = false;
        existH(2) = false;
    end
    hold on
    for i = uint8(1) : 1 : uint8(2)
        if existH(i)
            [finDtect(i), fingerKey(i,:), results(k,:)] =
fingertips(hands(i), keys, keyStruct, keyDetectThreshold, mov, i, fingerKey
(i,:), finDtect(i), results(k,:), numOfKeys);
        end
    end
    hold off
end
end

```

Ακολουθεί η συνάρτηση `keysDetection` που κάνει τον εντοπισμό των πλήκτρων του πιάνο

```

function [bwImg, keys, keyDetectThreshold] =
keysDetection(img, numOfKeys, vidHeight)
bwImg(size(img, 1), size(img, 2)) = false;
keyDetectThreshold = uint8(100);
detectedKeys = 0;
while detectedKeys ~= numOfKeys
    for i = 1 : size(img, 1)

```

```

    for j = 1 : size(img,2)
        if (img(i,j,1) > keyDetectThreshold)
            bwImg(i,j) = 255;
        else
            bwImg(i,j) = 0;
        end
    end
end
bwImg = bwareaopen(bwImg,vidHeight);
keys = regionprops(bwImg, 'Centroid', 'BoundingBox');
detectedKeys = size(keys,1);
if detectedKeys < numOfKeys
    keyDetectThreshold = keyDetectThreshold + 10;
end
if detectedKeys > numOfKeys
    keyDetectThreshold = keyDetectThreshold - 1;
end
if keyDetectThreshold > 255
    error('not enough keys were found')
    break
end
end
end

```

Η squareTracing εντοπίζει το περίγραμμα των χειρών

```

function [contours] = squareTracing(img)
contours(2,1)=0;
sizeIndex=1;
curPos=sizeIndex+1;
for j = uint16(1) : 1 : uint16(size(img,2))
    if (img(1,j) == 1 && img(1,j-1) == 0)
        curPoint=[j,1];
        i=15;
        direction = 'right';
        while (curPoint(2)~=0)
            if img(curPoint(2),curPoint(1)) == 1
                if mod(i,15)==0
                    contours(1,curPos) = curPoint(1);
                    contours(2,curPos) = curPoint(2);
                    contours(2,sizeIndex) = contours(2,sizeIndex) +
1;
                    curPos = curPos + 1;
                end
                if strcmp(direction,'up')
                    direction = 'right';
                    curPoint(1) = curPoint(1) + 1;
                else
                    if strcmp(direction,'down')
                        direction = 'left';
                        curPoint(1) = curPoint(1) - 1;
                    else
                        if strcmp(direction,'left')
                            direction = 'up';
                            curPoint(2) = curPoint(2) - 1;
                        else
                            if strcmp(direction,'right')
                                direction = 'down';
                                curPoint(2) = curPoint(2) + 1;
                            end
                        end
                    end
                end
            end
            i=i-1;
            curPoint(2) = curPoint(2) - 1;
        end
    end
end

```



```

end
end
end
end
i = i + 1;
else
if strcmp(direction,'up')
direction = 'left';
curPoint(1) = curPoint(1) - 1;
else
if strcmp(direction,'down')
direction = 'right';
curPoint(1) = curPoint(1) + 1;
else
if strcmp(direction,'left')
direction = 'down';
curPoint(2) = curPoint(2) + 1;
else
if strcmp(direction,'right')
direction = 'up';
curPoint(2) = curPoint(2) - 1;
end
end
end
end
end
end
end
sizeIndex = curPos;
curPos = sizeIndex+1;
end
end

```

Η συνάρτηση `handPoints` ελέγχει τα περιγράμματα, αποφασίζει πόσα χέρια έχουν εντοπιστεί και διαχωρίζει τα περιγράμματα τους

```

function [hands,hands1,hands2] = handPoints(c)
meion=0;
j=0;
hands(1)=struct('positions', zeros(2, c(2,1), 'uint16'));
hands(2)=hands(1);
hands1=false;
hands2=false;
for i=1:size(c,2)
if c(1,i)==0
j=j+1;
contours(j) = struct('positions', zeros(2, c(2,i),
'uint16'));
cl(j+1)=c(2,i);
meion=meion+cl(j);
else
contours(j).positions(1,i-j-meion)=c(1,i);
contours(j).positions(2,i-j-meion)=c(2,i);
end
end
end
j=1;
for i=1:size(contours,2)
if size(contours(i).positions,2)>100
hands(j)=contours(i);

```



```

                                j=j+1;
                                prevKey =
keys (hands.positions (2,i),hands.positions (1,i));
                                end
                                end
                                else
                                pos (j).x=hands.positions (1,i);
                                pos (j).y=hands.positions (2,i);
                                j=j+1;
                                prevKey =
keys (hands.positions (2,i),hands.positions (1,i));
                                end
                                break
                                end
                                end
                                end
                                end
                                i=i+1;
end
if j > 6 && finDtect == true
    dis = 20000;
    for m = 1 : j - 2
        if abs (int16 (pos (m).x)-int16 (pos (m+1).x)) +
abs (int16 (pos (m).y)-int16 (pos (m+1).y)) < dis
            dis = abs (int16 (pos (m).x)-int16 (pos (m+1).x)) +
abs (int16 (pos (m).y)-int16 (pos (m+1).y));
            temp = m;
        end
    end
    for m = temp : j - 2
        pos (m) = pos (m+1);
    end
    em = pos;
    clear pos;
    for i = 1 : 5
        pos (i) = em (i);
    end
end
if j == 6 && finDtect == false
    if dir == 1
        pos (1).finger='δεξί χέρι, μικρός';
        pos (2).finger='δεξί χέρι, παράμεσος';
        pos (3).finger='δεξί χέρι, μέσος';
        pos (4).finger='δεξί χέρι, δείκτης';
        pos (5).finger='δεξί χέρι, αντίχειρας';
    else
        pos (1).finger='αριστερό χέρι, αντίχειρας';
        pos (2).finger='αριστερό χέρι, δείκτης';
        pos (3).finger='αριστερό χέρι, μέσος';
        pos (4).finger='αριστερό χέρι, παράμεσος';
        pos (5).finger='αριστερό χέρι, μικρός';
    end
    fingerKey = pos;
    finDtect = true;
else
    if finDtect == true
        for i = 1 : 1 : size (pos,2)
            dis = 20000;
            for j = 1 : 1 : 5

```

```

        if i-j <= 0 && i+5-j >= size(pos,2) &&
abs(int16(pos(i).x)-int16(fingerKey(j).x)) + abs(int16(pos(i).y)-
int16(fingerKey(j).y)) < dis
        dis = abs(int16(pos(i).x)-int16(fingerKey(j).x))
+ abs(int16(pos(i).y)-int16(fingerKey(j).y));
        temp = j;
    end
end
posTemp(i) = fingerKey(temp);
posTemp(i).x = pos(i).x;
posTemp(i).y = pos(i).y;
end
for i = 1 : 1 : size(pos,2)
    for k = 1 : 1 : 5
        if strcmp(posTemp(i).finger,fingerKey(k).finger)
            break
        end
    end
    for j = size(pos,2) : -1 : i + 1
        kmeion = false;
        ksun = false;
        if strcmp(posTemp(i).finger,posTemp(j).finger)
            if k > 1 && i > 1
                if strcmp(posTemp(i - 1).finger,fingerKey(k -
1).finger)
                    kmeion = true;
                end
            end
            if k < 5 && j < size(pos,2)
                if strcmp(posTemp(j + 1).finger,fingerKey(k
+ 1).finger)
                    ksun = true;
                end
            end
            if kmeion == false && ksun == true
                posTemp(i).finger = fingerKey(k-1).finger;
            else
                if kmeion == true && ksun == false
                    posTemp(j).finger =
fingerKey(k+1).finger;
                else
                    if kmeion == false && ksun == false
                        if abs(int16(pos(i).x)-
int16(fingerKey(j).x)) > abs(int16(pos(j).x)-int16(fingerKey(j).x))
                            posTemp(i).finger = fingerKey(k-
1).finger;
                        else
                            posTemp(j).finger =
fingerKey(k+1).finger;
                        end
                    else
                        if abs(int16(pos(i).x)-
int16(fingerKey(j).x)) > abs(int16(pos(j).x)-int16(fingerKey(j).x))
                            posTemp(i).finger = fingerKey(k-
1).finger;
                        else
                            posTemp(i-1).finger =
fingerKey(k-2).finger;
                        else
                            posTemp(j).finger =
fingerKey(k+1).finger;
                    end
                end
            end
        end
    end
end

```


Στη συνέχεια η συνάρτηση `keysDiff` μετράει την θέση των πάνω γωνιών του πλήκτρου καθώς και την απόσταση τους, όταν αυτό δεν πατιέται

```
function [minX,maxX] = keysDiff(img,posY,posX)
tempX = posX;
maxX = 0;
minX = 20000;
while img(posY,tempX) ~= 0
    maxX = tempX;
    tempX = tempX + 1;
end
tempX = posX - 1;
while img(posY,tempX) ~= 0
    minX = tempX;
    tempX = tempX - 1;
end
```

Τέλος η συνάρτηση `curKeyDetect` μετράει την θέση των πάνω γωνιών του πλήκτρου καθώς και την απόσταση τους, όταν υπάρχει υποψία ότι πατιέται

```
function [minX,maxX] = curKeyDetect(img,keyDetectThreshold,posY,posX)
tempX = posX;
maxX = 0;
minX = 20000;
while img(posY,tempX,1) > keyDetectThreshold
    maxX = tempX;
    tempX = tempX + 1;
end
tempX = posX - 1;
while img(posY,tempX,1) > keyDetectThreshold
    minX = tempX;
    tempX = tempX - 1;
end
```

Παράρτημα Β: Παρουσίαση Πτυχιακής Εργασίας

*Πτυχιακή Εργασία Τμήματος Εφαρμοσμένης
Πληροφορικής & Πολυμέσων*

**Οπτική Σημασιολογική Ανάλυση
Βίντεο Μουσικών Κομματιών με
Πιάνο**

*Πασγιάννης Εμμανουήλ (ΑΜ: 2145)
Επιβλέπων καθηγητής : Τριανταφυλλίδης Γεώργιος*

Εισαγωγή

- Η ανάλυση ήχου και εικόνας χρησιμοποιείται ευρέως στη μουσική
- Εφαρμογές αναγνώρισης τραγουδιστών και τραγουδιών γίνονται όλο και πιο διαδεδομένες
- Τέτοιες εφαρμογές βοηθούν στην έρευνα γύρω από την κατανόηση του τρόπου έκφρασης ενός μουσικού





Περιορισμοί του Συστήματος

- Αποδίδει καλύτερα με HD video ($\geq 720p$)
- Χρειάζεται απαλό φωτισμό
- Δεν πρέπει να υπάρχουν σκιές πάνω στα πλήκτρα
- Ουδέτερο φόντο
- Ο πιανίστας δεν πρέπει να φοράει αξεσουάρ στα χέρια του



Διάγραμμα Ροής Αλγορίθμου



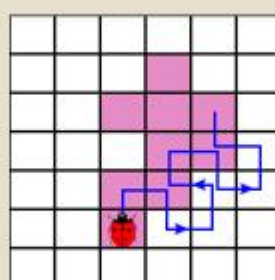
Εντοπισμός Χεριών

- Χρωματική κβαντοποίηση
- Δειγματοληψία και εξαγωγή χαρακτηριστικών των χρωμάτων
- Δημιουργία χρωματικού χάρτη (Διάγραμμα Voronoi)
- Αντικατάσταση κάθε χρώματος με το αντίστοιχο του στον χάρτη



Εντοπισμός Περιγράμματος

- Square Tracing
- Έλεγχος στην γραμμή $y=1$ για εύρεση των σημείων εκκίνησης και τερματισμού
- Το έντομο στρίβει δεξιά αν βρεθεί σε pixel αντικειμένου
- Το έντομο στρίβει αριστερά αν βρεθεί σε pixel φόντου
- Σταματάει όταν φτάσει στο σημείο τερματισμού



Εντοπισμός Ακροδαχτύλων

- Εντοπισμός τοπικών ελαχίστων στο περίγραμμα των χεριών
- Εξετάζεται η απόσταση του κάθε ακροδάχτυλου από όλα τα ακροδάχτυλα που εντοπίστηκαν στο προηγούμενο καρέ
- Αν εντοπιστούν περισσότερα από 5 ακροδάχτυλα, αφαιρείται αυτό με τη μεγαλύτερη απόσταση από το πλησιέστερο του ακροδάχτυλο, με βάση την τελευταία γνωστή του θέση



Ονοματοθεσία Ακροδαχτύλων (1/2)

- Αριθμούμε με τη σειρά τα ακροδάχτυλα με βάση την τελευταία θέση που εντοπίστηκαν και αναφερόμαστε σε αυτά με χρήση της μεταβλητής j
- Αριθμούμε με τη σειρά τα ακροδάχτυλα που εντοπίστηκαν σε αυτό το καρέ και αναφερόμαστε σε αυτά με χρήση της μεταβλητής i
- Η τιμή του j καθορίζει το όνομα του ακροδαχτύλου i που δείχνει σε αυτό



Ονοματοθεσία Ακροδαχτύλων (2/2)

Αρκεί να ισχύουν:

$$i - j \leq 0$$

$$i + 5 - j \geq S$$

Για οποιαδήποτε i_2 και i_3 που δείχνουν στην ίδια θέση j , με i_2 να βρίσκεται πάντα αριστερά του i_3 , ισχύουν οι εξής κανόνες:

Αν $\exists i \rightarrow j - 1 \wedge \nexists i \rightarrow j + 1$ τότε $i_3 \rightarrow j + 1$

Αν $\exists i \rightarrow j + 1 \wedge \nexists i \rightarrow j - 1$ τότε $i_2 \rightarrow j - 1$

Αν $\nexists i \rightarrow j - 1 \wedge i \rightarrow j + 1$ τότε
 Αν $d(i_2, j) > d(i_3, j)$ τότε $i_2 \rightarrow j - 1$
 αλλιώς $i_3 \rightarrow j + 1$

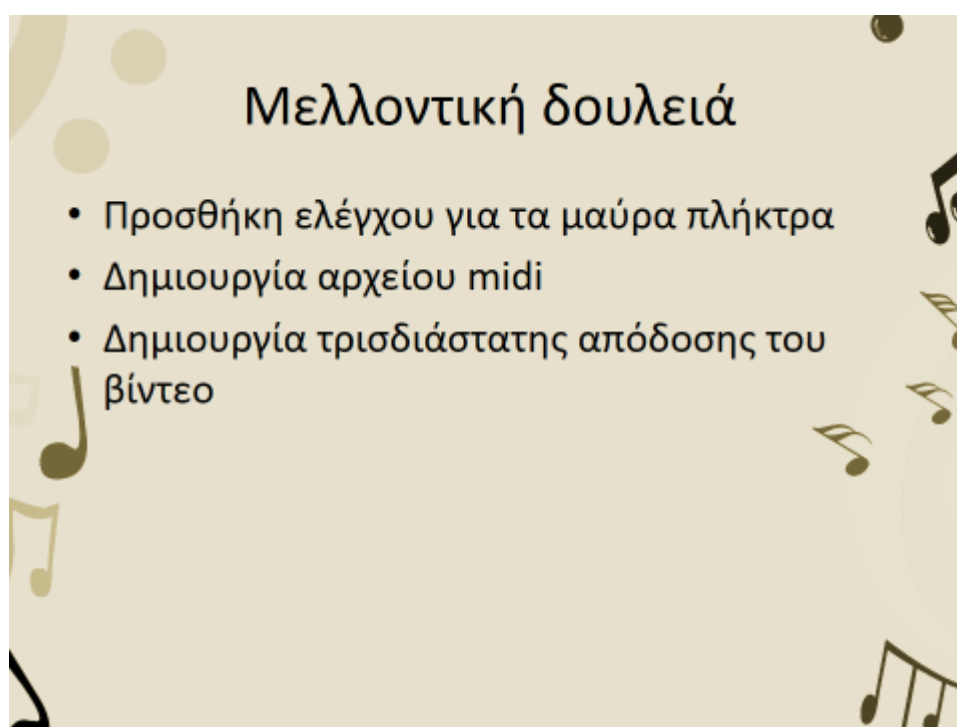
Αν $\exists i_1 \rightarrow j - 1 \wedge i_4 \rightarrow j + 1$ τότε
 Αν $d(i_2, j) > d(i_3, j)$ τότε $i_2 \rightarrow j - 1 \wedge i_1 \rightarrow j - 2$
 αλλιώς $i_3 \rightarrow j + 1 \wedge i_4 \rightarrow j + 2$

Εντοπισμός Πατήματος Πλήκτρων

- Τα πλήκτρα που βρίσκονται κοντά στο κέντρο της εικόνας, όταν πιέζονται αποκτούν μια αίσθηση γραμμικής προοπτικής
- Όταν αλλάζει η κατάσταση ενός πλήκτρου που βρίσκεται μακριά από το κέντρο της εικόνας, η αποτύπωση του στο επίπεδο φαίνεται να έχει μια κλίση σε σχέση με την αρχική του κατάσταση







Παράρτημα Γ: Περίληψη Πτυχιακής Εργασίας

Οπτική Σημασιολογική Ανάλυση Βίντεο Μουσικών Κομματιών με Πιάνο

Πασγιάννης Εμμανουήλ

Τμήμα Εφαρμοσμένης Πληροφορικής και Πολυμέσων, ΑΤΕΙ Κρήτης, Email: manospasj@gmail.com

Σύνοψη — Σκοπός της Πτυχιακής Εργασίας είναι η ανάπτυξη ενός συστήματος που μπορεί να εξαγάγει πληροφορίες από ένα ήδη καταγεγραμμένο βίντεο το οποίο παρουσιάζει την εκτέλεση ενός μουσικού κομματιού από έναν πιανίστα. Οι πιο σημαντικές από αυτές είναι: ποιο πλήκτρο πατιέται, ποια χρονική στιγμή, για πόσο χρονικό διάστημα και από ποιο δάχτυλο. Ο συνδυασμός εικόνας και πληροφορίας μπορεί να βοηθήσει σημαντικά στην εκμάθηση πιάνο εξ' αποστάσεως, η οποία γίνεται όλο και πιο δημοφιλής με την εξέλιξη της τεχνολογίας. Μέχρι στιγμής αυτό επιτυγχάνεται με τη χρήση βιντεοκάμερας ή μιας συσκευής εγγραφής midi. Στην πρώτη περίπτωση ο μαθητής έχει μόνο την εικόνα από το βίντεο, ενώ στη δεύτερη μόνο την πληροφορία για τις νότες. Σε γενικές γραμμές, ο κύριος στόχος της πτυχιακής εργασίας είναι η παραγωγή όλων των απαραίτητων πληροφοριών, για την εξ' αποστάσεως εκμάθηση πιάνο, από την οπτική ανάλυση του βίντεο, χωρίς την ανάγκη υποστήριξης από την τεχνολογία midi.

Λέξεις κλειδιά — Ψηφιακή Επεξεργασία Εικόνας, Τμηματοποίηση Δέρματος, Ανίχνευση Δαχτύλων, Ανάλυση Βίντεο με Πιάνο

Ι. ΕΙΣΑΓΩΓΗ

Εξαιτίας της ανάπτυξης της τεχνολογίας, η εκμάθηση πιάνο εξ' αποστάσεως γίνεται όλο και πιο δημοφιλής. Σε πολλές περιπτώσεις αυτός είναι ο μόνος τρόπος να μάθει κάποιος μουσική, ιδιαίτερα σε απομακρυσμένες περιοχές. Φυσικά αυτή η μέθοδος έχει κάποια πολύ σημαντικά πλεονεκτήματα. Ένας μαθητής μπορεί να επιλέξει ένα διαδικτυακό πρόγραμμα σπουδών της προτίμησής του, αντί να παρακολουθήσει το πρόγραμμα κάποιου σχολείου το οποίο είναι πολύ πιθανό να επιλέξει με βάση το πόσο κοντά βρίσκεται στο σπίτι του. Η εκμάθηση μουσικής εξ' αποστάσεως μπορεί να γίνει οπουδήποτε, οποιαδήποτε στιγμή, παρέχοντας στον μαθητή ένα ευέλικτο πρόγραμμα με περισσότερο ελεύθερο χρόνο.

Στην εκμάθηση πιάνο εξ' αποστάσεως συνήθως χρησιμοποιούνται βίντεο που παρουσιάζουν την εκτέλεση κομματιών από τον

καθηγητή, συνοδευόμενη από ομιλία στην οποία εξηγεί τις κινήσεις του. Ο μαθητής μπορεί να δει τον τρόπο που παίζονται οι νότες αλλά είναι δύσκολο να αντιστοιχίσει ένα πλήκτρο στο βίντεο με το αντίστοιχο στο δικό του πιάνο. Μια άλλη μέθοδος χρησιμοποιεί μια συσκευή εγγραφής midi, στην οποία αποστέλλονται οι νότες μέσα σε ένα αρχείο midi. Η τεχνολογία εγγραφής midi επιτρέπει στον καθηγητή να παίζει ένα κομμάτι στο πιάνο του και ο μαθητής να βλέπει ένα άλλο πιάνο σε κάποιο άλλο μέρος να παίζει μόνο του. Με την χρήση αυτής της μεθόδου ο μαθητής δεν μπορεί να γνωρίζει τον τρόπο με τον οποίο πιέζονται τα πλήκτρα. Η πτυχιακή εργασία στοχεύει στο συνδυασμό των δυο προηγούμενων μεθόδων για την παραγωγή μιας νέας βελτιωμένης μεθόδου.

Η εκτεταμένη χρήση της τεχνολογία midi έχει αποθαρρύνει τις προτάσεις στη βιβλιογραφία για εκμάθηση πιάνο εξ' αποστάσεως που βασίζονται στην εικόνα. Η ανεξάρτητη χρήση της συσκευής εγγραφής midi και της βιντεοκάμερας δημιουργεί το πρόβλημα του συγχρονισμού των δεδομένων. Έχει προταθεί από τον Gorodnichy μια εφαρμογή που ονομάζεται C-MIDI [1] η οποία συνδυάζει την εγγραφή midi με την εγγραφή βίντεο για να καταγράφει το πάτημα ενός πλήκτρου αλλά και ποιο δάχτυλο το πάτησε. Στο [2] ο Takegawa εισήγαγε ένα σύστημα ανίχνευσης-σημαδιών για να επιτύχει ανίχνευση των δαχτύλων σε πραγματικό χρόνο για εφαρμογές με πιάνο.

Η μέθοδος που περιγράφεται στην πτυχιακή χρησιμοποιεί μια βιντεοκάμερα για να καταγράψει μια μουσική εκτέλεση πιάνο. Η εικόνα από το βίντεο δίνεται σαν είσοδος στο σύστημα για επεξεργασία. Αρχικά γίνεται μια τμηματοποίηση βασισμένη στο χρώμα για να εντοπιστούν οι περιοχές του πιάνο αλλά και του κάθε πλήκτρου ξεχωριστά. Για τον εντοπισμό των χεριών χρησιμοποιείται τμηματοποίηση βασισμένη στη χρωματική κβάντοποίηση. Ο εντοπισμός των ακροδάχτυλων γίνεται εντοπίζοντας τα τοπικά ελάχιστα στο περίγραμμα του χεριού. Όταν η θέση ενός ακροδαχτύλου ανήκει στην περιοχή ενός πλήκτρου, τότε το συγκεκριμένο πλήκτρο πρέπει να εξεταστεί επιπλέον. Για να ληφθεί η απόφαση αν το

πλήκτρο είναι πατημένο ή όχι εξετάζονται τυχόν μεταβολές στο σχήμα του. Η έξοδος του συστήματος είναι ένας πίνακας με πληροφορίες σχετικά με το ποιο πλήκτρο πιάστηκε, ποια χρονική στιγμή, για πόσο διάστημα και από ποιο δάχτυλο.

Η συνέχεια της παρουσίασης οργανώνεται ως εξής. Η ενότητα 2 αναλύει τον αλγόριθμο. Η ενότητα 3 αναφέρει τις απαιτήσεις του συστήματος και παρουσιάζει κάποια πειραματικά αποτελέσματα. Τέλος, στην ενότητα 4 επισημαίνονται κάποια συμπεράσματα αλλά και η μελλοντική δουλειά.

II. ΠΕΡΙΓΡΑΦΗ ΑΛΓΟΡΙΘΜΟΥ

Πρώτα από όλα, πρέπει να γίνουν κάποιες αρχικοποιήσεις τιμών. Οι απαραίτητοι υπολογισμοί γίνονται με χρήση του πρώτου καρέ του βίντεο, στο οποίο δεν είναι εμφανή τα χέρια του πιανίστα, όπως στην εικόνα 1.



Εικόνα 1: Οπτικό πεδίο βιντεοκάμερας

Ο χρήστης πρέπει να εισάγει στο σύστημα τον αριθμό των πλήκτρων που εμφανίζονται σε αυτό το καρέ. Αυτό βοηθάει το σύστημα να εντοπίσει τα πλήκτρα με επιτυχία. Αρκεί να εφαρμοσθεί τμηματοποίηση βασισμένη στο χρώμα. Χρησιμοποιείται μεταβλητό κατώφλι, το οποίο αλλάζει τιμή μέχρι ο αριθμός των εντοπισμένων πλήκτρων να συμφωνεί με αυτόν που έδωσε ο χρήστης σαν είσοδο. Ξεκινώντας με τιμή 100, αν ο αριθμός των εντοπισμένων πλήκτρων είναι μικρότερος από τον αριθμό του χρήστη, το κατώφλι αυξάνεται κατά 10 αλλιώς μειώνεται κατά 1. Η διαδικασία αυτή επαναλαμβάνεται έως ότου οι δυο αριθμοί να συμφωνούν. Ακολουθεί η τιτλοφόρηση των πλήκτρων.



Εικόνα 2: Τμηματοποίηση πλήκτρων

Η ακόλουθη διαδικασία εκτελείται σε κάθε καρέ της εικόνας και περιγράφεται καλύτερα στο διάγραμμα της εικόνας 3.

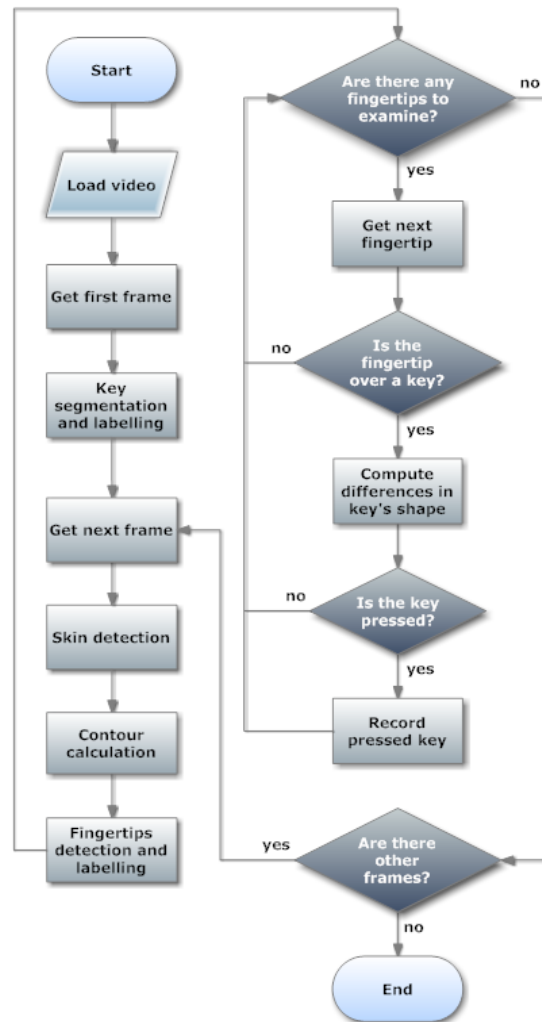
Για την τμηματοποίηση των χεριών

χρησιμοποιήθηκε χρωματική κβαντοποίηση. Η χρωματική κβαντοποίηση είναι μια τεχνική απολεστικής συμπίεσης κατά την οποία μειώνονται οι αποχρώσεις της εικόνας έως την επίτευξη του επιθυμητού αποτελέσματος. Έχει δυο πολύ σημαντικά πλεονεκτήματα:

- Έχει πολύ καλά αποτελέσματα αν είναι γνωστός ο αριθμός των χρωμάτων που υπάρχουν στην οθόνη.
- Παράγει εξίσου καλά αποτελέσματα σε συνθήκες κακού φωτισμού.

Χωρίζεται σε τέσσερις φάσεις [3]:

- 1) Δειγματοληψία και εξαγωγή στατιστικών των χρωμάτων
- 2) Επιλογή χρωματικού χάρτη
- 3) Αντικατάσταση χρωμάτων με το αντίστοιχο τους στον χάρτη
- 4) Σχεδιασμός της νέας εικόνας

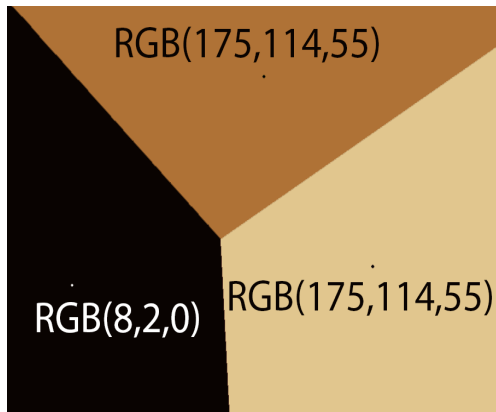


Εικόνα 3: Διάγραμμα ροής αλγορίθμου

Η πρόκληση στην εφαρμογή αυτής της τεχνικής βρίσκεται στα δυο πρώτα βήματα. Ο χρωματικός χώρος χωρίζεται σε κλάσεις με βάση έναν αλγόριθμο μέτρησης της απόστασης μεταξύ των χρωμάτων. Στη συνέχεια εντοπίζονται τα

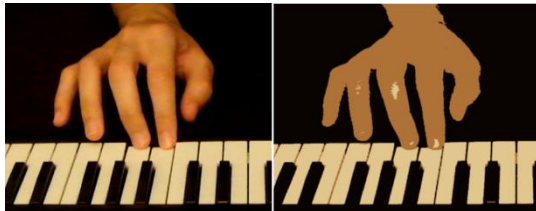
κέντρα της κάθε κλάσης και από αυτά δημιουργείται ένα διάγραμμα Voronoi (Εικόνα 4) κατά το οποίο κάθε χρώμα αντιστοιχίζεται σε μια άλλη απόχρωση που αντιπροσωπεύει την κλάση στην οποία ανήκει η συγκεκριμένη απόχρωση.

Στα πλαίσια αυτής της Πτυχιακής Εργασίας δημιουργείται δια κάθε καρέ του βίντεο ένα διάγραμμα Voronoi αποτελούμενο από 3 κλάσεις (κελιά). Η πρώτη κλάση αντιπροσωπεύει το φόντο και τον σκελετό του πιάνο, η δεύτερη τα πλήκτρα του πιάνο ενώ η τρίτη και τελευταία αντιπροσωπεύει τις δερματικές αποχρώσεις. Το αποτέλεσμα της τμηματοποίησης φαίνεται στην εικόνα 5.



Εικόνα 4: Διάγραμμα Voronoi

Το επόμενο βήμα είναι ο εντοπισμός των ακροδαχτύλων. Χωρίζεται σε δυο μέρη: Στον εντοπισμό του περιγράμματος και στην ανάλυση του.



Εικόνα 5: Εντοπισμός χειρών

Για τον εντοπισμό του περιγράμματος εφαρμόζεται ο αλγόριθμος Square Tracing [4]. Ο αλγόριθμος αυτός απαιτεί τη δήλωση δυο εικονοστοιχείων: αυτό από το οποίο θα ξεκινήσει η διαδικασία και αυτό στο οποίο θα τερματιστεί. Τα εικονοστοιχεία αυτά ανιχνεύονται στην ευθεία $y=1$ ψάχνοντας για ενωμένα εικονοστοιχεία που ανήκουν στην κλάση του χεριού. Σε κάθε ομάδα εικονοστοιχείων αυτής της κλάσης, το πρώτο ορίζεται σαν σημείο εκκίνησης ενώ το τελευταίο σαν σημείο τερματισμού. Έστω ότι ένα έντομο έχει τοποθετηθεί στο σημείο εκκίνησης και είναι στραμμένο προς τα κάτω. Αν το εικονοστοιχείο που βρίσκεται το έντομο ανήκει στην περιοχή του χεριού, τότε αυτό κινείται στο εικονοστοιχείο που βρίσκεται δεξιά του, ειδάλλως κινείται σε αυτό που βρίσκεται στα αριστερά του. Αυτή η διαδικασία επαναλαμβάνεται έως ότου το έντομο να φτάσει

στο σημείο τερματισμού. Τα εικονοστοιχεία που διέσχισε το έντομο και ανήκουν στην περιοχή του χεριού, ανήκουν επίσης και στο περίγραμμα του (εικόνα 6).

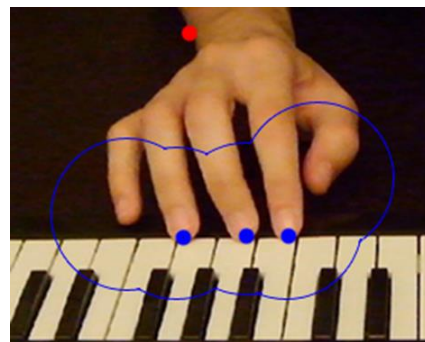


Εικόνα 6: Περίγραμμα χειρών

Ο έλεγχος για τοπικά ελάχιστα στο περίγραμμα παράγει πολύ καλά αποτελέσματα όσον αφορά τον εντοπισμό των δαχτύλων (εικόνα 7). Αν εντοπισθεί ένα τοπικό ελάχιστο, τα επόμενα 15 εικονοστοιχεία του περιγράμματος αγνοούνται λόγω του ότι είναι αδύνατον να υπάρχουν δυο ακροδάχτυλα τόσο κοντά και είναι πολύ πιθανό να πρόκειται για το ίδιο δάχτυλο. Για να μειωθεί το ποσοστό των εσφαλμένα εντοπισμένων ακροδαχτύλων, εξετάζεται και η κατά Manhattan απόσταση του κάθε ακροδάχτυλου από όλα τα ακροδάχτυλα που εντοπίστηκαν στο προηγούμενο καρέ (εικόνα 8). Αν η απόσταση από ένα τουλάχιστον ακροδάχτυλο του προηγούμενου καρέ είναι μικρότερη του σταθερού κατωφλιού που έχει δηλωθεί εξ' αρχής, τότε ο εντοπισμός του ακροδάχτυλου είναι πιθανότατα σωστός. Το σταθερό αυτό κατώφλι έχει οριστεί στα 150 εικονοστοιχεία για το βίντεο των δοκιμών. Αν σε καμία περίπτωση η απόσταση δεν είναι μικρότερη του κατωφλιού τότε δεν υπάρχει ακροδάχτυλο στο εν λόγω σημείο.



Εικόνα 7: Εντοπισμός ακροδαχτύλων



Εικόνα 8: Περιοχή ανίχνευσης ακροδαχτύλων

Ακολουθεί η διαδικασία της ονοματοθεσίας των δαχτύλων. Αν εντοπισθούν 5 ακροδάχτυλα σε ένα χέρι ονομάζονται ένα-ένα με τη σειρά. Αν τα ακροδάχτυλα είναι περισσότερα, υπολογίζεται η απόσταση του καθενός από το πλησιέστερο ακροδάχτυλο με βάση την τελευταία γνωστή θέση του καθενός. Αφαιρούνται με τη σειρά τα ακροδάχτυλα ξεκινώντας από αυτό με τη μεγαλύτερη απόσταση μέχρι να μείνουν 5.



Εικόνα 9: Ονοματοθεσία δαχτύλων

Αν εντοπισθούν λιγότερα από 5 ακροδάχτυλα, το καθένα αντιστοιχίζεται στο πλησιέστερο ακροδάχτυλο που έχει εντοπιστεί στο παρελθόν, αφού πρώτα κάνουμε το εξής: Αριθμούμε με τη σειρά τα ακροδάχτυλα που εντοπίστηκαν σε αυτό το καρέ και αναφερόμαστε σε αυτά χρησιμοποιώντας τη μεταβλητή i . Προφανώς το i παίρνει τιμές από 1 έως τον αριθμό των εντοπισμένων ακροδαχτύλων τον οποίο θα ονομάσουμε S . Αριθμούμε επίσης και τα 5 ακροδάχτυλα για τα οποία ξέρουμε την τελευταία θέση που εντοπίστηκαν και αναφερόμαστε σε αυτά με χρήση της μεταβλητής j . Η τιμή του j καθορίζει το όνομα του ακροδαχτύλου i που δείχνει σε αυτό. Αρκεί να ισχύουν:

$$(i - j \leq 0) \quad (i + 5 - j \geq 5)$$

Πολλές φορές δυο ακροδάχτυλα μπορεί να πάρουν κοινή ονομασία. Για να διορθωθεί αυτό γίνεται ο έλεγχος που ακολουθεί. Για οποιαδήποτε i_2 και i_3 που δείχνουν στην ίδια θέση j , με i_2 να βρίσκεται πάντα αριστερά του i_3 , ισχύουν οι εξής κανόνες:

$$\text{Αν } \exists i \rightarrow j - 1 \wedge \exists i \rightarrow j + 1 \text{ τότε } i_3 \rightarrow j + 1$$

$$\text{Αν } \exists i \rightarrow j + 1 \wedge \exists i \rightarrow j - 1 \text{ τότε } i_2 \rightarrow j - 1$$

$$\begin{aligned} &\text{Αν } \exists i \rightarrow j - 1 \wedge i \rightarrow j + 1 \text{ τότε} \\ &\text{Αν } d(i_2, j) > d(i_3, j) \text{ τότε } i_2 \rightarrow j - 1 \\ &\text{αλλιώς } i_3 \rightarrow j + 1 \end{aligned}$$

$$\begin{aligned} &\text{Αν } \exists i_1 \rightarrow j - 1 \wedge i_4 \rightarrow j + 1 \text{ τότε} \\ &\text{Αν } d(i_2, j) > d(i_3, j) \text{ τότε } i_2 \rightarrow j - 1 \wedge i_1 \rightarrow j - 2 \\ &\text{αλλιώς } i_3 \rightarrow j + 1 \wedge i_4 \rightarrow j + 2 \end{aligned}$$

Μετά την εφαρμογή και των προηγούμενων ελέγχων, το αποτέλεσμα της ονοματοθεσίας είναι όπως αυτό της εικόνας 9 όπου η τιμή του i δείχνει την ονομασία του κάθε δαχτύλου.

Εάν ένα εντοπισμένο ακροδάχτυλο ανήκει στην περιοχή ενός πλήκτρου, υπάρχει πιθανότητα αυτό το δάχτυλο να ασκεί πίεση στο εν λόγω πλήκτρο. Για να παρθεί απόφαση σχετικά με το αν πιέζεται το πλήκτρο ή όχι, το σύστημα αναζητά διαφορές ανάμεσα στο σχήμα του πλήκτρου σε αυτό το καρέ και στο σχήμα του πλήκτρου στο πρώτο καρέ που είναι γνωστό ότι κανένα πλήκτρο δεν πιεζόταν. Σημαντικό ρόλο παίζει η θέση του πλήκτρου στην εικόνα. Έστω ότι πρέπει να ελεγχθεί ένα πλήκτρο που βρίσκεται κοντά στο κέντρο της εικόνας. Εάν πατηθεί το σχήμα του στην εικόνα αποκτά μια αίσθηση γραμμικής προοπτικής όπως στην εικόνα 10. Στην περίπτωση που πατηθεί ένα πλήκτρο το οποίο βρίσκεται μακριά από το κέντρο της εικόνας, το σχήμα του στο επίπεδο φαίνεται να αποκτά μια μικρή κλίση προς το κέντρο της εικόνας (εικόνα 11).

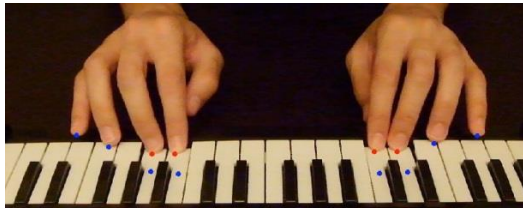


Εικόνα 10: Η γραμμική προοπτική που αποκτά ένα πλήκτρο όταν πατηθεί

Και στις δυο περιπτώσεις τα εικονοστοιχεία του πλήκτρου για τα οποία ισχύει $y \leq \text{finger}(i) \cdot y$ αφαιρούνται από την περιοχή του πλήκτρου και στις δυο εικόνες. Στη συνέχεια, εντοπίζεται η θέση των δυο πάνω γωνιών στο σχήμα του πλήκτρου και στις δυο εικόνες. Εάν έστω μια από αυτές έχει μετατοπισθεί κατά τουλάχιστον 3 εικονοστοιχεία στον άξονα x τότε το πλήκτρο βρίσκεται υπό πίεση. Το αποτέλεσμα φαίνεται στην εικόνα 12.

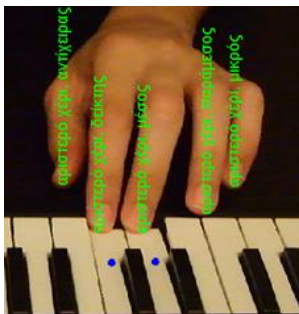


Εικόνα 11: Η κλίση που αποκτά ένα πλήκτρο όταν πατηθεί



Εικόνα 12: Εντοπισμός πατήματος πλήκτρων

Το σύστημα δημιουργεί ένα διδιάστατο άδειο πίνακα συμβολοσειράς με κάθε στήλη να αντιπροσωπεύει ένα πλήκτρο και κάθε γραμμή ένα καρέ. Έστω ότι ένα πλήκτρο πατηθεί σε κάποιο καρέ, η αντίστοιχη τιμή στον πίνακα θα αλλάξει και θα γίνει ίδια με το όνομα του δαχτύλου που πίεσε το πλήκτρο.

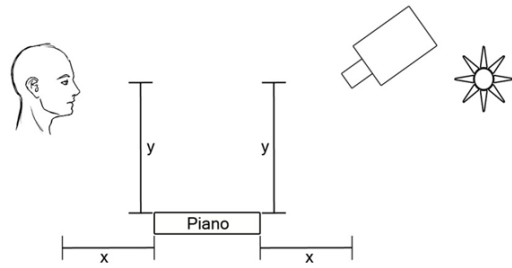


Εικόνα 13: Οπτική απεικόνιση αποτελεσμάτων

III. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

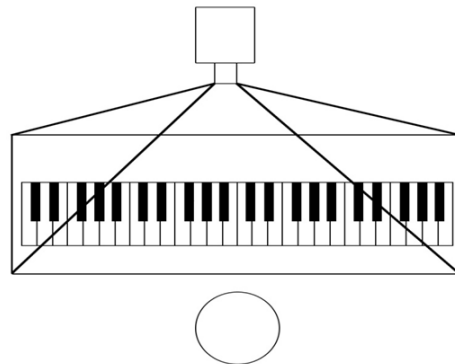
A. Απαιτήσεις του συστήματος

Για το βίντεο των δοκιμών χρησιμοποιήθηκε η Fujifilm FinePix Real 3D W3, η οποία για την εγγραφή βίντεο παρέχει εικόνες ανάλυσης 1280x720, με βάθος χρώματος 24bit και ρυθμό καταγραφής 25 καρέ ανά δευτερόλεπτο. Ο αλγόριθμος του συστήματος αναπτύχθηκε με χρήση του λογισμικού Matlab. Για τις δοκιμές χρησιμοποιήθηκε ένας ηλεκτρονικός υπολογιστής με επεξεργαστή intel core i5 με συχνότητα ρολογιού στα 3,4GHz.



Εικόνα 14: Θέση βιντεοκάμερας και πηγής φωτός

Η συσκευή καταγραφής βίντεο πρέπει να τοποθετηθεί στην πίσω πλευρά του πιάνο, περίπου στην ίδια οριζόντια και κατακόρυφη απόσταση που έχει το πιάνο από τα μάτια του πιανίστα, όπως φαίνεται στην εικόνα 14. Ο λόγος που προτιμήθηκε η συγκεκριμένη θέση είναι ότι υπάρχει μονίμως οπτική επαφή με όλα τα πλήκτρα και τα δάχτυλα σπανίως αλληλεπικαλύπτονται. Ο πιανίστας επιβάλλεται να μην φοράει αξεσουάρ στα χέρια του, τα οποία θα εμποδίσουν τον σωστό εντοπισμό τους, όπως για παράδειγμα δαχτυλίδια, βραχιόλια ή ακόμα και μακρυμάνικη μπλούζα.



Εικόνα 15: Θέση και οπτικό πεδίο βιντεοκάμερας

Ιδιαίτερη προσοχή πρέπει να δοθεί στο φωτισμό, καθώς το σύστημα δεν αποδίδει εξίσου αποτελεσματικά σε καταστάσεις έντονου ή περιορισμένου φωτισμού. Για καλύτερα αποτελέσματα η πηγή φωτός πρέπει να βρίσκεται όσο το δυνατόν πιο κοντά στην κάμερα, κατά προτίμηση ακριβώς κάτω από αυτήν. Όσον αφορά το φόντο, πρέπει να αποτελείται από ουδέτερα χρώματα τα οποία δεν θυμίζουν τις αποχρώσεις του δέρματος αλλά ούτε και αυτές των πλήκτρων. Στο οπτικό πεδίο της κάμερας πρέπει να είναι ορατά όλα και ολόκληρα τα πλήκτρα που σκοπεύει να χρησιμοποιήσει ο πιανίστας, αλλά και περίπου 20 εκατοστά της περιοχής μπροστά από τα πλήκτρα του πιάνο, το σημείο δηλαδή στο οποίο βρίσκονται τα χέρια του πιανίστα κατά την εκτέλεση του κομματιού (εικόνα 15).

B. Αποτελέσματα

Ένα βίντεο κατάλληλο για ανάλυση τηρεί τις παρακάτω προϋποθέσεις:

- Η βιντεοκάμερα πρέπει να βρίσκεται ακριβώς απέναντι από τον πιανίστα.
- Η πηγή φωτός πρέπει να είναι κάτω από την βιντεοκάμερα.
- Να μην υπάρχουν σκιές στην περιοχή των πλήκτρων.
- Ο πιανίστας να μην φοράει αξεσουάρ στα χέρια του.
- Το φόντο να μην αποτελείται από λευκές ή δερματικές αποχρώσεις.

Ο εντοπισμός των χεριών έχει πολύ καλά αποτελέσματα ακόμα και σε περιπτώσεις κακού φωτισμού. Η μοναδική απαίτηση της τεχνικής αυτής είναι να είναι γνωστός και σταθερός ο αριθμός των αποχρώσεων μεταξύ των καρτέ. Για την συγκεκριμένη εφαρμογή αυτό δεν είναι πρόβλημα, καθώς λόγω της θέσης και της κλίσης της κάμερας μπορούμε να είμαστε βέβαιοι πως οι αποχρώσεις που εμφανίζονται στο οπτικό της πεδίο είναι τρεις με τέσσερις (πιάνο, πλήκτρα, φόντο, χέρια).

Ο εντοπισμός των δαχτύλων συνδέεται άμεσα με την τμηματοποίηση των χεριών, καθώς όσο πιο ομαλό είναι το περίγραμμα των χεριών θα είναι και πιο ακριβής ο εντοπισμός των ακροδαχτύλων πάνω σε αυτό. Η τεχνική που χρησιμοποιείται έχει επιτυχία σε ποσοστό που αγγίζει το 83%. Σε αρκετά υψηλά επίπεδα κινείται και το ποσοστό του εντοπισμού των πλήκτρων που πατήθηκαν, το οποίο κυμαίνεται γύρω στο 70%.

Ο μέσος χρόνος ανάλυσης ενός καρτέ για τον επεξεργαστή των δοκιμών είναι 102ms, χρόνος ικανοποιητικός για τις απαιτήσεις της εφαρμογής.

IV. ΣΥΜΠΕΡΑΣΜΑΤΑ

Το παρών έγγραφο προτείνει ένα σύστημα βασισμένο στην ανάλυση βίντεο το οποίο μπορεί να βοηθήσει σημαντικά στην εκμάθηση πιάνο εξ' αποστάσεως. Στο μέλλον, το σύστημα μπορεί να βελτιωθεί εξετάζοντας επιπλέον και τα μαύρα πλήκτρα πέρα από τα λευκά. Το πάτημα ενός μαύρου πλήκτρου είναι πολλές φορές δύσκολο να εντοπιστεί ακόμα και από τον άνθρωπο, καθώς τυχόν αλλαγές στην όψη ενός πλήκτρου καλύπτονται από το δάχτυλο που το πιέζει.

Από τον πίνακα αποτελεσμάτων που παράγεται με την εφαρμογή του αλγορίθμου σε ένα βίντεο μπορεί να δημιουργηθεί ένα midi αρχείο το οποίο θα αναπαράγει τον ήχο από την εκτέλεση του μουσικού κομματιού που καταγράφηκε. Επιπλέον, από τις εγγραφές του ίδιου πίνακα είναι εφικτό να κατασκευαστεί μια τρισδιάστατη απόδοση του βίντεο η οποία σε συνδυασμό με τον ήχο από το midi αρχείο μπορεί να βοηθήσει σημαντικά στην έρευνα που γίνεται γύρω από την κατανόηση του τρόπου έκφρασης ενός μουσικού, αλλά και στην εκπαίδευση μαθητευόμενων μουσικών.

ΑΝΑΦΟΡΕΣ

[1] D. Gorodnichy and A. Yogeswaran, "Piano player finger detection using the crevice detection operator," Conference on Computer Vision and Pattern Recognition, New York, NY, USA. IEEE Computer Society 2006.

[2] Yoshinari Takegawa, Tsutomu Terada, and Shojiro Nishio, "Design and Implementation of a Real-Time Fingering Detection System for Piano Performance," in Proc International Computer Music Conference (ICMC 2006), pp.67-74, Nov 2006.

[3] Heckbert, P. "Color Image Quantization for Frame Buffer Display", Computer Graphics, Vol 16, #3, pp. 297-303, 1982

[4] Square Tracing Algorithm http://www.imageprocessingplace.com/downloads_V3/root_downloads/tutorials/contour_tracing_Abeer_George_Ghuneim/square.html