

# **Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης**

**Σχολή Τεχνολογικών Εφαρμογών**

**Τμήμα Εφαρμοσμένης Πληροφορικής & Πολυμέσων**



## **ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**Τίτλος: Αναγνώριση Εκφράσεων Προσώπου στην  
Μουσική**

**Γιώργος Καββαλάκης (ΑΜ: 2166)**

**Επιβλέπων Καθηγητής: Γιώργος Τριανταφυλλίδης**

**Επιτροπή Αξιολόγησης:**

**Ημερομηνία Παρουσίασης:**

## **Ευχαριστίες**

Για την διεξαγωγή της έρευνας και της μελέτης αυτής χρειάστηκε η ορθή καθοδήγηση και συμπαράσταση κάποιων ανθρώπων και για τους οποίους νοιώθω την ανάγκη να ευχαριστήσω. Ο κύριος παράγοντας για την έκβαση της μελέτης αυτής ήταν ο καθηγητής Γεώργιος Τριανταφυλλίδης. Μέσα από την συνεχή παρακολούθηση της πορείας της μελέτης, το ενδιαφέρον του και την άψογη συνεργασία που υπήρχε για όλο το χρονικό διάστημα που χρειάστηκε ώστε να ολοκληρωθεί η έρευνα, βοηθούσε ανελλιπώς την εξέλιξη της. Επίσης σημαντική ήταν και η συνεισφορά του Τσαμπίκου Κουναλάκη, τον οποίο πρέπει να ευχαριστήσω, καθώς χωρίς κανένα προσωπικό όφελος ήταν πρόθυμος να με κατατοπίσει σε κάποιους τομείς της έρευνας. Πάνω από όλα όμως θα ήθελα να ευχαριστήσω τους γονείς μου που για τα τόσα χρόνια έχουν σταθεί υποστηρικτές μου, τόσο ηθικά όσο και με τα υλικά αγαθά που μου έχουν εξασφαλίσει.

Η παρούσα πτυχιακή αφιερώνεται στους γονείς μου.

## **Abstract**

Human expression recognition is a fundamental problem in computer vision attracting great interest from the research community over the last years. Human expressions involve facial expressions, gestures, voices, etc. In this paper, we will focus on the facial expression recognition which is the task of automatically identifying and classifying expressions in an image or video sequence. We will consider a specific and interesting case of facial expressions recognition: recognition in music video clips. Music is a basic way of expressing human emotions. But bear in mind that there is a rather different way of face expressing emotions in music video clips compared to the same emotions in everyday life.

The proposed scheme employs already known and used methods of classifying emotions such as the Logistic Regression (LogReg), the Classification and Regression Trees (CR-tree), Linear discriminant analysis (LDA [17]), k-Nearest neighbor (k-NN [19]) and Quadratic discriminant analysis (QDA [16]). The paper also introduces the use of SIFT[9] feature extraction algorithm for emotion classification. SIFT is an algorithm in computer vision that detects and describes local features in images. In this context, for any face in an image, interesting points can be extracted to provide a "feature description" of the face. This description, extracted from a training image, can then be used to recognize the facial emotions in the image with a use of a neural network. To perform reliable recognition, it is important that the features extracted from the training image are detectable even under changes in image scale, noise and illumination.

This fact of different expression classification for our case results in the need of creating a new data base, obeying the rules of music video clip's facial expression labeling.

## Σύνοψη

Η αναγνώριση των συναισθημάτων στο ανθρώπινο πρόσωπο είναι ένα θεμελιώδες πρόβλημα στον τομέα των ηλεκτρονικών υπολογιστών το οποίο αποκτά μεγάλο ενδιαφέρον από την ερευνητική κοινότητα τα τελευταία χρόνια. Σε αυτή την εργασία, θα επικεντρωθούμε στην αναγνώριση εκφράσεων προσώπου το οποίο είναι ο κύριος στόχος της, εντοπίζοντας αυτόματα και ταξινομώντας τις εκφράσεις σε μια αλληλουχία εικόνων ή βίντεο. Θα εξετάσουμε μια ειδική και ενδιαφέρουσα περίπτωση της αναγνώρισης των εκφράσεων προσώπου: την αναγνώριση σε μουσικά βίντεο κλιπ. Η μουσική είναι ένα βασικός τρόπος έκφρασης για τα ανθρώπινα συναισθήματα. Αλλά έχουμε κατά νου ότι υπάρχει ένας μάλλον διαφορετικός τρόπος που το πρόσωπο εκφράζει τα συναισθήματα μέσα σε ένα μουσικό βίντεο κλιπ και στην περίπτωση έκφρασης τους στην καθημερινή ζωή.

Θα χρησιμοποιήσουμε τις ήδη γνωστές μεθόδους ταξινόμησης συναισθημάτων, όπως η λογιστική παλινδρόμηση (LogReg), η ταξινόμηση και Δέντρα Παλινδρόμησης (CR-tree), Γραμμική ανάλυση διακρίνουσας (LDA [17]), k- nearest neighbor (k-NN) και την τετραγωνική διακριτική ανάλυση (QDA [16]). Επίσης εισάγει την χρήση του αλγορίθμου SIFT[9] για την εξαγωγή τοπικών χαρακτηριστικών σημείων σε μια εικόνα. Σε αυτό το σημείο, για οποιοδήποτε πρόσωπο μέσα σε μια εικόνα, μπορεί να εξάχθούν ενδιαφέροντα σημεία και να χρησιμοποιηθούν ως χαρακτηριστικά του προσώπου. Το σύνολο αυτών των χαρακτηριστικών μπορούν στην συνέχεια να χρησιμοποιηθούν για την εκπαίδευση ενός νευρωνικού δικτύου και αυτό με την σειρά του για την αναγνώριση ανθρώπινων συναισθημάτων.

Όλα αυτά προϋποθέτουν την δημιουργία μιας νέας βάσης δεδομένων με θέμα τα συναισθήματα του καλλιτέχνη την ώρα που ερμηνεύει ένα μουσικό κομμάτι.

## Περιεχόμενα

Ευχαριστίες.....	2
Abstract .....	3
Σύνοψη.....	4
1. Εισαγωγή.....	10
1.1 Περίληψη .....	10
1.2 Κίνητρο για την διεξαγωγή της εργασίας.....	11
1.3 Σκοπός και στόχοι της εργασίας.....	11
1.4 Δομή εργασίας.....	12
2. Σχέδιο δράσης για την εκπόνηση της εργασίας .....	13
2.1 Βιομετρική .....	13
2.1.1 Απόδοση.....	15
2.2 Εφαρμογές Βιομετρικής.....	15
2.2.1 Τρέχουσες, αναδυόμενες και μελλοντικές εφαρμογές της βιομετρικής.....	15
2.3 Ζητήματα και ανησυχίες.....	16
2.3.1 Μυστικότητα και διάκριση.....	16
2.3.2 Κίνδυνος στους ιδιοκτήτες των εξασφαλισμένων στοιχείων.....	16
2.3.3 Ακυρώσιμη βιομετρική.....	16
2.3.4 Μαλακή βιομετρική.....	17
2.3.5 Οι κυβερνήσεις είναι απίθανο να αποκαλύψουν τις πλήρεις ικανότητες των βιομετρικών επεκτάσεων.....	17
2.4 Βιομετρική ανάλυση ταυτοποίησης ατόμου .....	17
2.4.1 Face Recognition (αναγνώριση προσώπου).....	17
2.4.2 Fingerprint identification (ταυτοποίηση δακτυλικών αποτυπωμάτων).....	18
2.4.3 Σάρωση Αμφιβληστροειδή .....	18
2.4.4 Σάρωση Ιριδας .....	20
2.4.5 Αναγνώριση Υπογραφής.....	20
2.4.6 Ανάλυση Φωνής.....	21
2.4.7 Δυναμική πληκτρολόγηση.....	22
3. Μεθοδολογία Υλοποίησης .....	24
3.1 Ανίχνευση Προσώπου .....	24
3.1.1 Δομή προβλήματος.....	24
3.1.2 Προσεγγίσεις στην ανίχνευση προσώπου.....	25
Ανίχνευσης ανάλογα με το σενάριο.....	25
Χωρισμός μεθόδων ανίχνευσης σε κατηγορίες.....	26
Knowledge-based Methods.....	26
Template Matching Methods.....	27
Appearance-based Methods.....	27
3.1.3 Εφαρμογές.....	31
3.2 Εξαγωγή Χαρακτηριστικών .....	31
3.2.1 Μέθοδοι Επιλογής Χαρακτηριστικών.....	32
3.3 Facial Expression.....	33
3.3.1 Facial Expression Classification .....	34
3.3.2 Πλαίσιο περιγραφής της απεικόνισης των συναισθημάτων στα πρόσωπα.....	34
3.3.3 Απεικονίσεις που σχετίζονται με την συναισθηματική έκφραση.....	35
3.3.4 Νευροφυσιολογία και συναισθήματα.....	36
3.4 ΤΕΧΝΙΚΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΕΚΦΡΑΣΕΩΝ.....	36
Παρακολούθηση της κίνησης του προσώπου.....	37
3.4.1 K Nearest Neighbors.....	37
3.4.2 Linear discriminant analysis.....	38
3.4.3 Square discriminant analysis.....	39
3.4.4 Classification and involution Tree.....	39

3.4.5 Logistic involution.....	39
3.4.6 Πειραματική Μέθοδος- SIFT-based Neural Network.....	40
3.4.6.1 Scale-invariant feature transform (SIFT).....	40
3.4.6.2 Νευρωνικά δίκτυα.....	41
4. Σχεδιασμός Υλοποίησης.....	43
4.1 Εισαγωγή της εικόνας.....	43
4.2 Ανίχνευση Προσώπου.....	45
4.2.1 Δοκιμές Μεθόδων Face Detection.....	45
Face Detection by Mikael Nilsson.....	45
Face Detection System by Omid Sakhi.....	45
Face Detection using Support Vector Machine (SVM).....	47
Face and Eye Detection by Mustafa Ucak.....	47
Fdlbmex – Fast and simple face detection by Wolf Kienzle.....	48
Viola Jones Object Detection.....	48
4.2.2 Υλοποίηση Ανίχνευσης Προσώπου.....	49
Gabor Wavelet Transformation.....	51
Εύρεση περιοχή ματιών (τρόπος λειτουργίας της συνάρτησης extractEyes).....	53
Εύρεση περιοχή στόματος (τρόπος λειτουργίας της συνάρτησης extractMouth).....	56
Εύρεση περιοχή μύτης (τρόπος λειτουργίας της συνάρτησης extractNose).....	58
4.3 Βάση Δεδομένων.....	59
4.4 Ταξινόμηση Συναισθημάτων.....	60
Optical analysis:.....	61
Ταξινόμηση μέσω SIFT-based Neural Network.....	62
Συνδυασμός μεθόδων και τελική απόφαση.....	66
5. Πειραματικά Αποτελέσματα και GUI Matlab.....	68
5.1 Παρουσίαση Γραφικού περιβάλλοντος.....	69
5.2 Ανάλυση Αποτελεσμάτων και Συμπεράσματα.....	72
Ανίχνευση προσώπου.....	77
Δυσκολία εύρεσης συναισθημάτων.....	77
Επιτυχία κάθε αλγόριθμου.....	78
Αποτελέσματα SIFT-based Neural Network.....	79
Καταλληλότητα Βάσης Δεδομένων.....	79
Γενικά συμπεράσματα.....	80
6. Μελλοντική Εργασία και Επεκτάσεις.....	82
6.1 Βελτιώσεις Υπάρχοντος Προγράμματος.....	82
6.1.1 Συμβατότητα Προγράμματος.....	82
6.1.2 Αυτονομία Προγράμματος.....	82
6.1.3 Ρυθμός Δειγματοληψίας.....	83
6.1.4 Βελτίωση Κώδικα.....	83
6.2 Βελτιώσεις Μεθόδων και Τεχνικών.....	84
6.2.1 Βάση δεδομένων.....	84
6.2.2 SIFT-based Neural Network.....	84
6.3 Εφαρμογές Προγράμματος και Τεχνογνωσίας.....	84
BIBΛΙΟΓΡΑΦΙΑ.....	86

## Κατάλογος εικόνων

Εικόνα 1: Συναισθήματα στην καθημερινότητα.....	11
Εικόνα 2: Συναισθήματα και Μουσική.....	11
Εικόνα 3: Σχεδιάγραμμα Βιομετρικών συστημάτων.....	13
Εικόνα 4: Ανάλυση και Αναγνώριση Προσώπου.....	18
Εικόνα 5: Λήψη Δεδομένων από το Δακτυλικό αποτύπωμα.....	18
Εικόνα 6: Η όψη ενός αμφιβληστροειδή.....	19
Εικόνα 7: Τρόπος ανάλυσης ματιού.....	20
Εικόνα 8: Ανάλυση Υπογραφής.....	21
Εικόνα 9: Διάγραμμα υπογραφής.....	21
Εικόνα 10: Ανάλυση Φωνητικού Κύματος.....	22
Εικόνα 11: Λογισμικό με χρήση Template Matching μεθόδου.....	28
Εικόνα 12: Χρήση Appearance-based Μεθόδου.....	29
Εικόνα 13: Μερικά Eigenfaces.....	29
Εικόνα 14: Προβολή ενός feedforward artificial neural network.....	30
Εικόνα 15: SVM λειτουργία.....	30
Εικόνα 16: Probabilistic parameters of a hidden Markov model.....	31
Εικόνα 17: Εφαρμογές ανίχνευσης προσώπου.....	32
Εικόνα 18: Μέθοδοι Επιλογής Χαρακτηριστικών.....	35
Εικόνα 19: Example of k-NN classification.....	40
Εικόνα 20: Διανύσματα του αλγόριθμου SIFT.....	42
Εικόνα 21: Εύρεση κοινών σημείων μέσω SIFT.....	42
Εικόνα 22: Νευρωνικό Δίκτυο.....	44
Εικόνα 23: Σχέδιο κώδικα.....	45
Εικόνα 24: Παράδειγμα Καρέ no1.....	46
Εικόνα 25: Παράδειγμα Καρέ no2.....	46
Εικόνα 26: Παράδειγμα Καρέ no3.....	46
Εικόνα 27: Nilsson's Algorithm result 1.....	47
Εικόνα 28: Nilsson's Algorithm result 2.....	47
Εικόνα 29: Εσφαλμένο Αποτέλεσμα.....	49
Εικόνα 30: Ορθό Αποτέλεσμα.....	49
Εικόνα 31: Σημεία Αναζήτησης.....	49
Εικόνα 32: Χειροκίνητο μαρκάρισμα περιοχής ματιού.....	50
Εικόνα 33: Αποτελέσματα Αλγορίθμου Viola Jones.....	51
Εικόνα 34: Ορθή ανίχνευση προσώπου.....	52
Εικόνα 35: Εσφαλμένη ανίχνευση προσώπου.....	53
Εικόνα 36: Ακατάλληλο πρόσωπο για ανάλυση.....	53
Εικόνα 37: Τρόπος λειτουργίας Face Detection.....	55
Εικόνα 38: Διαφορετικά φίλτρα Kernels με διαφορετικά μεγέθη.....	56
Εικόνα 39: Φιλτραρισμένες εικόνες κατώτερου ορίου.....	57
Εικόνα 40: Εικόνα ακρών (edge picture) της περιοχής των ματιών.....	57
Εικόνα 41: Κατάτμηση των ματιών με χρήση Gabor.....	58
Εικόνα 42: Ένωση της edge picture με φίλτρο Gabor.....	58
Εικόνα 43: Ίδανική κατάτμηση στα μάτια.....	58
Εικόνα 44: Εύρεση σημαντικών περιοχών στην αρχική εικόνα.....	59
Εικόνα 45: Περισσότερες από δυο περιοχές βρέθηκαν στην αρχική εικόνα.....	59
Εικόνα 46: Edge picture της αρχικής εικόνας.....	60
Εικόνα 47: Αποτέλεσμα του Gabor Wavelet transformation.....	60
Εικόνα 48: Ενδιάμεσο αποτέλεσμα με την edge picture.....	61
Εικόνα 49: Filled hole, μειωμένη, εικόνα που έχει υποστεί περικοπές.....	61
Εικόνα 50: Αποτέλεσμα της ανακατασκευής.....	61
Εικόνα 51: Αποτέλεσμα της εξαγωγής.....	61

Εικόνα 52: Εισαγωγή γραμμής στην εικόνα για σύνδεση απομονωμένων περιοχών.....	62
Εικόνα 53: Αποτέλεσμα του Gabor Wavelet transformation.....	62
Εικόνα 54: Εύρεση περιοχής της μύτης.....	63
Εικόνα 55: Διαίρεση της μεγάλης περιοχής με τη βοήθεια της αφαίρεσης της edge picture.....	63
Εικόνα 56: Default nose.....	63
Εικόνα 57: Εικόνες προσώπου που αποπροσανατολίζουν τους αλγορίθμους από την ορθή απόφαση .....	64
Εικόνα 58: SIFT keypoints .....	67
Εικόνα 59: Αρχική εικόνα.....	68
Εικόνα 60: Θέσεις των χαρακτηριστικών.....	68
Εικόνα 61: Τελική εικόνα .....	68
Εικόνα 62: Εφαρμογή SIFT.....	69
Εικόνα 63: Δείγμα βάσεις δεδομένων προσώπων από μουσική.....	69
Εικόνα 64: Emotion Recognition Analysis.....	71
Εικόνα 65: Απαιτήσεις Εμφάνισης ανά Βήμα.....	73
Εικόνα 66: Παράθυρο εφαρμογής.....	74
Εικόνα 67: Παράθυρο επιλογής αρχείου.....	75
Εικόνα 68: Άμεση προβολή του πρώτου καρέ του βίντεο κατά την εισαγωγή του.....	76
Εικόνα 69: Λειτουργία προγράμματος.....	77
Εικόνα 70: Δεύτερη ανάλυση.....	79
Εικόνα 71: Τρίτη ανάλυση.....	80
Εικόνα 72: Τέταρτη ανάλυση.....	81
Εικόνα 73: Πέμπτη ανάλυση.....	82
Εικόνα 74: Εσφαλμένη ανίχνευση προσώπου.....	83
Εικόνα 75: Ολική αποτυχία αλγορίθμων.....	84
Εικόνα 76: Επιτυχία αλγορίθμων.....	84
Εικόνα 77: Αποτυχία CRTree .....	84



## Ευρετήριο πινάκων

Πίνακας 1: αλγορίθμων εξαγωγής χαρακτηριστικών.....	33
Πίνακας 2: Μέθοδοι Επιλογής Χαρακτηριστικών.....	35
Πίνακας 3: Συγκεντρωτικός Πίνακας μεθόδων Face Detection.....	52
Πίνακας 4: Πίνακας Featurematrix[nxp] και Labelvektor[nx1].....	65
Πίνακας 5: Scatterplot των τελικά επιλεγμένων χαρακτηριστικών γνωρισμάτων.....	66
Πίνακας 6: Παράδειγμα λήψης απόφασης για ένα Καρέ.....	72
Πίνακας 7: Αποτελέσματα 1.....	78
Πίνακας 8: Αποτελέσματα Δεύτερης Ανάλυσης.....	79
Πίνακας 9: Αποτελέσματα Τρίτης Ανάλυσης.....	80
Πίνακας 10: Αποτελέσματα Τέταρτης Ανάλυσης.....	81
Πίνακας 11: Αποτελέσματα Πέμπτης Ανάλυσης.....	82
Πίνακας 12: Αποτελέσματα Νευρωνικού δικτύου.....	85
Πίνακας 13: Δυνατότητες προγράμματος.....	87

# 1. Εισαγωγή

Περίπου τα τελευταία δέκα χρόνια, η εύρεση και ανάλυση προσώπων μέσα σε μια εικόνα απασχόλησε πολύ τον κόσμο της πληροφορικής και έγινε ένα δημοφιλές πεδίο έρευνας. Λόγο της φύσης του προβλήματος, δεν έχει δείξει μόνο η επιστήμη της πληροφορικής ενδιαφέρον για το θέμα αυτό, αλλά και επιστήμονες από τον τομέα της νευρολογίας και ψυχολογίας έχουν ασχοληθεί και ασχολούνται ακόμα με αυτό το ζήτημα. Η γενική άποψη επί του θέματος είναι ότι η πρόοδος της έρευνας πάνω στον τομέα της όρασης μηχανής θα αποβεί ωφέλιμο στο μέλλον για την νευρολογία και τους ψυχολόγους στην κατανόηση του τρόπου λειτουργίας του ανθρώπινου εγκεφάλου.

Την πρώτη φορά που σημειώθηκε ότι τα συναισθήματα του ανθρώπου συνδέονται άμεσα με τις εκφράσεις που παίρνει το πρόσωπο κατέγραψε ο Δαρβίνος(1872/1998). Ο Δαρβίνος ισχυρίστηκε ότι οι εκφράσεις του προσώπου είναι το υπολειπόμενο των ενεργειών για πληρέστερη απόκριση συμπεριφοράς και εμφανίζονται σε συνδυασμό με άλλες σωματικές αντιδράσεις, όπως είναι φωνητικά επιφωνήματα, χειρονομίες. Αργότερες μελέτες επαλήθευσαν αυτήν την θεωρία, κατοχυρώνοντας σε κάθε ένα από τα συναισθήματα και μια σειρά από αντιδράσεις τόσο του σώματος όσο και του προσώπου του κάθε ανθρώπου που σχεδόν πάντα υπακούν σε ένα σύνολο κανόνων.

Όταν όμως έχουμε να κάνουμε με την έκφραση του ανθρώπινου προσώπου μέσα σε ένα μουσικό βίντεο, την ώρα που ο καλλιτέχνης ερμηνεύει ένα κομμάτι οι κανόνες αυτοί αλλάζουν και γίνεται ακόμα πιο δύσκολο να αναγνωριστεί το συναίσθημα που εκφράζεται καθαρά και μόνο από την εικόνα. Αυτό είναι και το θέμα που θα μας απασχολήσει και στην υπόλοιπη μελέτη.

## 1.1 Περίληψη

Τα τελευταία χρόνια έχει αναπτυχθεί ιδιαίτερη έρευνα στην αναγνώριση συναισθημάτων μέσω των εκφράσεων του προσώπου. Εκτενή έρευνα σχετικά με τις ανθρώπινες εκφράσεις πραγματοποίησαν ο Paul Ekman και οι συνεργάτες του, και απέδειξαν ότι κάποιες από αυτές είναι παγκόσμιες ενώ άλλες που ποικίλουν εξαιτίας πολιτισμικών παραγόντων ή δεν είναι τόσο εύκολα ανιχνεύσιμες. Οι εκφράσεις που χαρακτηρίζονται ως 'παγκόσμιες' ή 'καθολικές' είναι αυτές που αντιπροσωπεύουν τα συναισθήματα χαράς, λύπης, θυμού, φόβου, έκπληξης και απέχθειας και φυσικά η ουδέτερη έκφραση.

Η αυτόματη αναγνώριση εκφράσεων αποτελεί μια πολυσύνθετη διεργασία καθώς οι φυσιολογικές των προσώπων διαφέρουν σημαντικά από άτομο σε άτομο εξαιτίας πολλών παραγόντων, όπως είναι η ηλικία, η εθνικότητα, το φύλο, η τριχοφυΐα του προσώπου, τα διακοσμητικά και καλλυντικά προϊόντα και αντικείμενα που ενδεχομένως κρύβουν σημεία του προσώπου όπως τα γυαλιά μυωπίας ή ακόμα και τα μαλλιά. Επιπλέον τα πρόσωπα παρουσιάζονται ανόμοια των διαφορετικών συνθηκών φωτισμού. Η έρευνα όμως της αναγνώρισης συναισθημάτων είναι σε έξαρση και αναπτύσσονται πολλές διαφορετικές προσεγγίσεις μεταξύ των οποίων κάποιες επιτυγχάνουν ιδιαίτερα ενθαρρυντικά αποτελέσματα.

Όμως ένα ακόμα πρόβλημα είναι ότι το αντικείμενο μελέτης στην παρούσα εργασία είναι η αναγνώριση των εκφράσεων μέσα σε ένα μουσικό βίντεο κλιπ. Όπως προαναφέραμε οι κανόνες που έχουν αναγνωριστεί από την επιστημονική κοινότητα καταργούνται στην μελέτη αυτή, καθώς στις εικόνες που ακολουθούν διαπιστώνουμε ότι δεν έχουν το ίδιο σχήμα τα χαρακτηριστικά γνωρίσματα στην καθημερινή έκφραση του προσώπου με την έκφραση του προσώπου στην μουσική.



Συναισθήματα χαράς και λύπης στην μουσική



Συναισθήματα χαράς και λύπης στην καθημερινότητα



Εικόνα 2: Συναισθήματα και Μουσική



Εικόνα 1: Συναισθήματα στην καθημερινότητα

Λόγο του πειραματικού σταδίου της έρευνας αυτής και κάποιων δοκιμαστικών μεθόδων, θα επικεντρωθούμε περισσότερο σε δύο συναισθήματα που μπορεί να πάρει το ανθρώπινο πρόσωπο, την χαρά και την λύπη.

## 1.2 Κίνητρο για την διεξαγωγή της εργασίας

Το κίνητρο για την διεξαγωγή της εργασίας ήταν η μελέτη των μεθοδολογιών μέσω των οποίων ένας υπολογιστής μπορεί να εντοπίζει ένα πρόσωπο μέσα από μια εικόνα ή μια σειρά εικόνων και η ανάλυση του διεξάγοντας χρήσιμες πληροφορίες. Αρχικά στόχος ήταν η εύρεση προσώπου μέσα σε ένα μουσικό βίντεο κλιπ, η αναγνώριση του και η εύρεση συναισθημάτων. Κατά την υλοποίηση όμως του προγράμματος διαπιστώθηκε ότι πρόκριτε για ένα μεγάλο πρόγραμμα το οποίο θα απαιτούσε μεγάλο χρονικό διάστημα για να πραγματοποιηθεί, οπότε αφαιρέθηκε το κομμάτι της αναγνώρισης προσώπου, και ασχολήθηκα μόνο με το κομμάτι του ανίχνευσης συναισθημάτων του καλλιτέχνη.

Από μόνο του ήταν μια πρόκληση για εμένα καθώς δεν έβρισκα παρόμοιες μελέτες και θεώρησα ότι είναι κάτι το καινοτόμο. Θέλοντας λοιπόν να δοκιμάσω τα όρια μου το τελικό κίνητρο πέρα από τον καθορισμό νέων κανόνων αναγνώρισης στην ταξινόμηση συναισθημάτων επέλεξα να χρησιμοποιήσω και τον αλγόριθμο SIFT ο οποίος μας έχει δείξει πολύ καλά αποτελέσματα στην αναγνώριση προσώπων, αλλά αυτήν την φορά για την αναγνώριση συναισθημάτων σε ένα ανθρώπινο πρόσωπο.

## 1.3 Σκοπός και στόχοι της εργασίας

Στόχος της μελέτης αυτής είναι μετά την ολοκλήρωση της να έχουν επιτευχθεί τα εξής:

1. Μελέτη ήδη υπαρχόντων λογισμικών emotion detection.
2. Δημιουργία ενός λογισμικού το οποίο να είναι λειτουργικό, εύχρηστο και όσο το δυνατόν αξιόπιστο.
3. Καθορισμός νέων κανόνων αναγνώρισης συναισθημάτων προσώπου μέσα στην μουσική.
4. Σύγκριση των τεχνικών για την ταξινόμηση των γνωρισμάτων.
5. Καθορισμός προοπτικής για την ταξινόμηση συναισθημάτων μέσω νευρωνικών δικτύων (εκπαίδευσης μέσω keypoint SIFT).
6. Δυνατότητες βελτίωσης του υπάρχοντα λογισμικού.

## 1.4 Δομή εργασίας

Στο πρώτο μέρος της πτυχιακής εργασίας αναφέρεται η μέχρι τώρα ιστορία της ανάλυσης του προσώπου και της αναγνώρισης συναισθημάτων, καθώς επίσης και την χρησιμότητα τους τόσο για τον επιστημονικό όσο και τον υπόλοιπο κόσμο, μέσα από πιθανές εφαρμογές που θα μπορούσαν να υλοποιηθούν και βελτιώσεις που θα μπορούσαν να κάνουν στην ζωή μας.

Στο δεύτερο μέρος αναλύουμε πλέον ποιες μεθόδους και για ποιους λόγους επιλέχθηκαν ή απορρίφθηκαν εν τέλει, και για την ανίχνευση προσώπου και την αναγνώριση συναισθημάτων. Παραθέτονται η δυσκολίες που πιθανόν θα συναντήσουμε για κάθε μέθοδο γνωρίζοντας τα μειονεκτήματα της κάθε μιας, και ποιες οι απαιτήσεις που πρέπει να πληρούνται για την ορθή χρήση τους.

Στο τέταρτο κομμάτι παρουσιάζεται το GUI guide tool του Matlab που χρησιμοποιήσαμε για την δημιουργία του interface του προγράμματος, τα πλεονεκτήματα που μας προσφέρει και ο τρόπος υλοποίησής του.

Στην συνέχεια αναφερόμαστε στα αποτελέσματα που λαμβάνουμε από τα πρόγραμμα μας. Το πέμπτο μέρος είναι ουσιαστικά αυτό που κρίνει την αξιοπιστία του τελικού προγράμματος και συγκρίνει το αποτέλεσμα κάθε μεθοδολογίας με απώτερο σκοπό την ορθή κριτική της εφαρμογής για μελλοντικές διορθώσεις.

Τέλος το 6 μέρος παρουσιάζει μελλοντικές βελτιώσεις που μπορεί να κάνει ο οποιοδήποτε πάνω στο πρόγραμμα και στις μεθόδους που χρησιμοποιήθηκαν.

## 2. Σχέδιο δράσης για την εκπονή της εργασίας

Το κύριο σημείο που πρέπει να αναλυθεί και να ερευνηθεί για την σωστή οργάνωση της πτυχιακής εργασίας είναι η ήδη γνωστή μελέτη πάνω στην αναγνώριση προσώπου αλλά και την ανίχνευση συναισθημάτων στο ανθρώπινο πρόσωπο. Στον τομέα όμως που υπάγονται αυτά τα ζητήματα που απασχολούν την επιστήμη της πληροφορικής, την βιομετρική, πρέπει να σταθούμε λίγο παραπάνω και να δούμε την ιστορία της και τι έχει επιτευχθεί μέχρι σήμερα χάρις σε αυτήν. Μόνο έτσι θα μπορούσαμε να κατανοήσουμε πλήρως το πρόβλημα που καλούμαστε να αντιμετωπίσουμε και να βρούμε τις καλύτερες δυνατές λύσεις για αυτό.

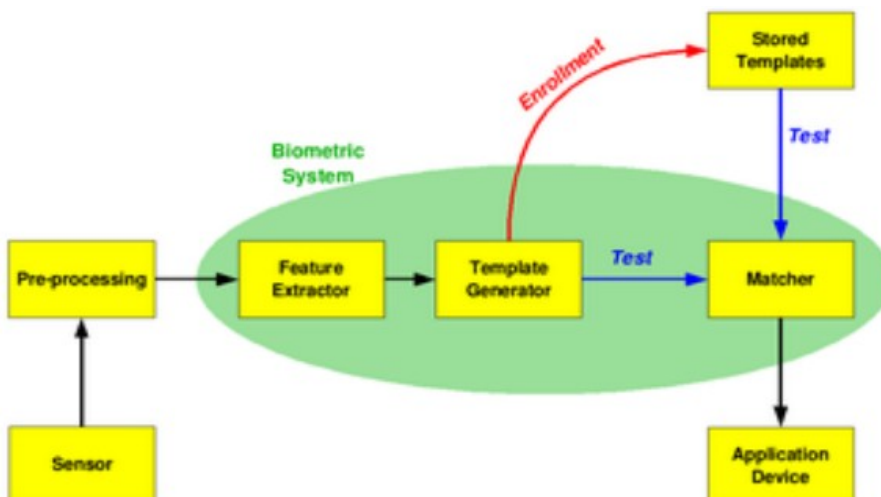
### 2.1 Βιομετρική

Η βιομετρική αποτελείται από τις μεθόδους για μεμονωμένα τους ανθρώπους που βασίζονται σε ένα ή περισσότερα εγγενή φυσικά ή συμπεριφοριστικά γνωρίσματα. Στην πληροφορική, η βιομετρική χρησιμοποιείται ως μορφή διαχείρισης και ελέγχου προσπέλασης πρόσβασης ταυτότητας. Χρησιμοποιείται επίσης για να προσδιορίσει τα άτομα κατά ομάδες που είναι υπό επιτήρηση.

Τα βιομετρικά χαρακτηριστικά μπορούν να διαιρεθούν σε δύο κύριες κατηγορίες:

- Φυσιολογικό: είναι σχετικό με τη μορφή του σώματος. Τα παραδείγματα περιλαμβάνουν, αλλά δεν περιορίζονται στο δακτυλικό αποτύπωμα, αναγνώριση προσώπου, DNA, τυπωμένη ύλη παλαμών, γεωμετρία χεριών, αναγνώριση ίριδας, η οποία έχει αντικαταστήσει κατά ένα μεγάλο μέρος τον αμφιβληστροειδή, και τη μυρωδιά.
- Συμπεριφοριστικό: είναι σχετικό με τη συμπεριφορά ενός προσώπου. Τα παραδείγματα περιλαμβάνουν, αλλά δεν περιορίζονται στο ρυθμό δακτυλογράφησης, του βηματισμού, και της φωνής. Μερικοί ερευνητές έχουν πλάσει το όρο *behaviometrics* για αυτήν την κατηγορία βιομετρικής.

Για να κυριολεκτήσουμε, η φωνή είναι επίσης ένα φυσιολογικό γνώρισμα επειδή κάθε πρόσωπο έχει ένα διαφορετικό φωνητικό κομμάτι, αλλά η αναγνώριση φωνής είναι κυρίως βασισμένη στη μελέτη του τρόπου που ένα πρόσωπο μιλά, συνήθως ταξινομημένος όπως συμπεριφοριστικό.



Εικόνα 3: Σχεδιάγραμμα Βιομετρικών συστημάτων

Είναι δυνατό να καταλάβει εάν ένα ανθρώπινο χαρακτηριστικό μπορεί να χρησιμοποιηθεί για τη

βιομετρική από την άποψη των ακόλουθων παραμέτρων:

- Καθολικότητα - κάθε πρόσωπο πρέπει να έχει το χαρακτηριστικό.
- Μοναδικότητα - είναι πόσο καλά ο βιομετρικός χωρίζει τα άτομα μεταξύ άλλων.
- Μονιμότητα - μέτρα πόσο καλά ένας βιομετρικός αντιστέκεται στη γηράσκουσα και άλλη διαφορά με την πάροδο του χρόνου.
- Collectability - ευκολία της απόκτησης για τη μέτρηση.
- Απόδοση - ακρίβεια, ταχύτητα, και ευρωστία της τεχνολογίας χρησιμοποιούμενη.
- Αποδοχή - βαθμός έγκρισης μιας τεχνολογίας.
- Παράκαμψη - ευκολία - - χρήση ενός υποκατάστατου.

Ένα βιομετρικό σύστημα μπορεί να λειτουργήσει στους ακόλουθους δύο τρόπους:

- Στη λειτουργία ελέγχου του συστήματος εκτελείτε μια one-to-one σύγκριση του συλληφθέντος βιομετρικού με ένα συγκεκριμένο πρότυπο βιομετρικών χαρακτηριστικών που είναι αποθηκευμένα σε μια βιομετρική βάση δεδομένων, προκειμένου να ελέγξει το άτομο εάν είναι το άτομο που ισχυρίζονται ότι είναι. Αυτή η διαδικασία μπορεί να χρησιμοποιεί μια έξυπνη κάρτα, το όνομα χρήστη ή τον αριθμό ταυτότητας (π.χ. PIN) για να δείξει ποιο πρότυπο πρέπει να χρησιμοποιείται για λόγους σύγκρισης.
- Στη λειτουργία αναγνώρισης το σύστημα εκτελεί μια one-to-many σύγκριση σε σχέση με μια βιομετρική βάση δεδομένων στην προσπάθεια να προσδιορίσει την ταυτότητα ενός άγνωστου ατόμου. Το σύστημα θα πετύχει τον προσδιορισμό των διαφορών εάν η σύγκριση των βιομετρικών δειγμάτων σε ένα πρότυπο στη βάση δεδομένων εμπίπτει σε ένα κατώτατο όριο που είχε οριστεί προηγουμένως. Η τελευταία λειτουργία μπορεί να επιτευχθεί μόνο με τη χρήση βιομετρικών στοιχείων από άλλες μεθόδους της προσωπικής αναγνώρισης, αφού οι PINs κωδικούς πρόσβασης είναι αναποτελεσματικοί.

Την πρώτη φορά που ένα άτομο χρησιμοποιεί ένα βιομετρικό σύστημα ονομάζεται εγγραφή. Κατά τη διάρκεια της εγγραφής, βιομετρικές πληροφορίες από ένα άτομο δεσμεύονται και αποθηκεύονται. Σε επόμενες χρήσεις, βιομετρικές πληροφορίες ανιχνεύονται και συγκρίνονται με τις πληροφορίες που αποθηκεύονται κατά τη στιγμή της εγγραφής. Σημειώστε ότι είναι σημαντικό η αποθήκευση και η ανάκτηση των εν λόγω συστημάτων να είναι ασφαλής. Το πρώτο μπλοκ (sensor εικόνα 3) είναι ο σύνδεσμος μεταξύ του πραγματικού κόσμου και του συστήματος. Πρέπει να αποκτήσει όλα τα απαραίτητα στοιχεία. Τις περισσότερες φορές πρόκειται για ένα σύστημα απόκτησης εικόνας, αλλά μπορεί να αλλάξει ανάλογα με τα επιθυμητά χαρακτηριστικά. Το δεύτερο block (εικόνα 3) εκτελεί όλες τις απαραίτητες προ-επεξεργασίες: πρέπει να αφαιρεθούν αντικείμενα από τον αισθητήρα, για την ενίσχυση της εισόδου (π.χ. θόρυβος αφαίρεση φόντου), να χρησιμοποιήσουν κάποιο είδος της κανονικοποίησης, κλπ. Στο τρίτο μπλοκ (εικόνα 3) εξάγονται απαραίτητα χαρακτηριστικά. Αυτό το βήμα είναι ένα σημαντικό βήμα για την εξαγωγή των ορθών χαρακτηριστικών με τον βέλτιστο τρόπο. Ένα διάλυμα αριθμών ή μια εικόνα με συγκεκριμένες ιδιότητες χρησιμοποιείται για να δημιουργήσει ένα πρότυπο. Ένα πρότυπο είναι μια σύνθεση των σχετικών χαρακτηριστικών που εξάγονται από την πηγή. Τα στοιχεία των βιομετρικών μετρήσεων που δεν χρησιμοποιούνται στον αλγόριθμο σύγκρισης απορρίπτονται από το πρότυπο για τη μείωση του μεγέθους του αρχείου και για την προστασία της ταυτότητας του εγγεγραμμένου.

Εάν η εγγραφή εκτελείται, το πρότυπο αποθηκεύεται απλά κάπου (σε μια κάρτα ή μέσα σε μια βάση δεδομένων ή και στις δύο). Εάν μια ταιριάζοντας με φάση εκτελείται, το αποκτηθέν πρότυπο περνούν σε ένα matcher που το συγκρίνει με άλλα υπάρχοντα πρότυπα, που υπολογίζει την απόσταση μεταξύ τους χρησιμοποιώντας οποιοδήποτε αλγόριθμο (π.χ. απόσταση Hamming). Το ταιριάζοντας με πρόγραμμα θα αναλύσει το πρότυπο με την εισαγωγή. Αυτό θα είναι έπειτα παραγωγή για οποιοδήποτε διευκρινισμένο χρήση ή σκοπό (π.χ. είσοδος σε μια περιορισμένη περιοχή).

## 2.1.1 Απόδοση

Τα ακόλουθα χρησιμοποιούνται ως μετρήσεις απόδοσης για τα βιομετρικά συστήματα:

- FAR ή FMR: η πιθανότητα ότι το σύστημα που ταιριάζει εσφαλμένα το πρότυπο εισόδου σε μη ταιριαστά πρότυπα στη βάση δεδομένων. Μετρά το ποσοστό των άκυρων εισροές που έχουν λανθασμένα αποδεκτεί.
- FRR ή FNMR: η πιθανότητα ότι το σύστημα αδυνατεί να ανιχνεύσει έναν ταίρι μεταξύ του προτύπου εισόδου και ένα αντίστοιχο πρότυπο στη βάση δεδομένων. Μετρά το ποσοστό των έγκυρων εισροές που έχουν εσφαλμένα απορριφθεί.
- ROC: Το ROC plot είναι ένας οπτικός χαρακτηρισμός των trade-off μεταξύ της FAR και της FRR. Σε γενικές γραμμές, η αλγόριθμος αντιστοίχισης εκτελεί μια απόφαση με βάση ένα κατώτατο όριο που καθορίζει το πόσο κοντά σε ένα πρότυπο της εισόδου πρέπει να είναι για να μπορεί να θεωρηθεί ένας ταιριαστό. Εφόσον το ανώτατο όριο μειώνεται, θα υπάρχει λιγότερη εσφαλμένα αποτελέσματα, αλλά περισσότερα λάθος. Αντίστοιχα, ένα υψηλότερο κατώτατο όριο θα μειωθεί η FAR, αλλά θα αύξηθεί η FRR. Μια κοινή παραλλαγή είναι το σφάλμα ανίχνευσης trade-off (DET), το οποίο λαμβάνεται με τη χρησιμοποίηση κλιμάκων απόκλισης και για τους δύο άξονες.
- Ίσο ποσοστό σφάλματος ή cross over ποσοστό σφάλματος (EER ή CER): ο ρυθμός με τον οποίο αποδεχόμαστε να απορρίπτουμε τα λάθη είναι ίσος. Η αξία του EER μπορούν εύκολα να προμηθευτούν από την καμπύλη ROC. Το EER είναι ένας γρήγορος τρόπος να συγκριθεί η ακρίβεια των συσκευών με διαφορετικές καμπύλες ROC. Σε γενικές γραμμές, η συσκευή με το χαμηλότερο EER είναι πιο ακριβή.
- Η μη καταχώρηση ποσοστό (FTE ή FER): ο ρυθμός με τον οποίο επιχειρεί να δημιουργήσει ένα πρότυπο από μια είσοδο δεν είναι επιτυχής. Αυτό είναι συνήθως προκαλείται από χαμηλές εισροές ποιότητας.
- FTC: Μέσα σε αυτόματα συστήματα, η πιθανότητα ότι το σύστημα αδυνατεί να ανιχνεύσει μια βιομετρική εισόδου όταν παρουσιάζονται σωστά.
- Χωρητικότητα πρότυπο: ο μέγιστος αριθμός των συνόλων δεδομένων που μπορούν να αποθηκευτούν στο σύστημα.

## 2.2 Εφαρμογές Βιομετρικής

Οι παραδοσιακές μέθοδοι ταυτοποίησης, όπως η πρόσβαση με κωδικό, οι μαγνητικές κάρτες και τα κλειδιά έπαψαν να μας προσφέρουν μεγάλη ασφάλεια και η αξιοπιστία τους μειώθηκε με την ανάπτυξη της τεχνολογίας, καθώς η υποκλοπή τους είναι αρκετά εύκολη, αλλά και πρόκειται για αντικείμενα τα οποία μπορούν να ξεχαστούν ή να χαθούν. Η βιομετρική μας εξασφαλίζει ότι με την χρήση των φυσικών μας χαρακτηριστικών ως 'κλειδί' εισόδου η πιθανότητα λανθασμένης ταυτοποίησης ελαττώνεται αρκετά και καθιστά συστήματα ασφαλείας που κάνουν χρήση βιομετρικών μεθόδων ταυτοποίησης αξιόπιστα.

### 2.2.1 Τρέχουσες, αναδυόμενες και μελλοντικές εφαρμογές της βιομετρικής

Προσκλήσεις προτάσεων για τη βιομετρική επικύρωση για να έχει πρόσβαση σε ορισμένα δημόσια δίκτυα

Ο John Michael (Mike) McConnell, πρώην ναύαρχος στο Ηνωμένο ναυτικό και πρώην διευθυντής της αμερικανικής Εθνικής Υπηρεσίας Πληροφοριών, και ο ανώτερος αντιπρόεδρος Booz Άλλεν Χάμιλτον προώθησαν την ανάπτυξη μιας μελλοντικής ικανότητας να απαιτηθεί η βιομετρική επικύρωση για

πρόσβαση σε ορισμένα δημόσια δίκτυα, στη κεντρική ομιλία τους στη διάσκεψη κοινοπραξίας για την βιομετρική το 2009.

Μια βασική προϋπόθεση στην πρόταση αυτή είναι ότι το πρόσωπο που έχει επικυρωθεί μεμονωμένα χρησιμοποιώντας τη βιομετρική με τον υπολογιστή είναι στην πραγματικότητα επίσης ο πράκτορας που εκτελεί τις ενδεχομένως κακόβουλες ενέργειες από εκείνο τον υπολογιστή. Εντούτοις, εάν ο έλεγχος του υπολογιστή έχει υπονομευθεί, παραδείγματος χάριν στον οποίο ο υπολογιστής είναι μέρος ενός botnet που ελέγχεται από έναν χάκερ, κατόπιν η γνώση της ταυτότητας του χρήστη στο τερματικό δεν βελτιώνει υλικά τις δραστηριότητες επιβολής νόμου ασφάλειας ή ενίσχυσης δικτύων.

## 2.3 Ζητήματα και ανησυχίες

Το πρότυπο κατωτέρω (το REF βελτιώνει το τμήμα) εξετάζεται για τη διαγραφή. Δείτε τα πρότυπα για τη συζήτηση για να βοηθήσετε να επιτύχει μια συναίνεση.

### 2.3.1 Μυστικότητα και διάκριση

Τα στοιχεία που λήφθηκαν κατά τη διάρκεια της βιομετρικής εγγραφής θα μπορούσαν να χρησιμοποιηθούν με τρόπους που το εγγραμμένο άτομο δεν συγκατατίθεται.

### 2.3.2 Κίνδυνος στους ιδιοκτήτες των εξασφαλισμένων στοιχείων

Όταν οι κλέφτες δεν μπορούν να πάρουν την πρόσβαση στις ασφαλείς ιδιοκτησίες, υπάρχει μια πιθανότητα ότι οι κλέφτες θα καταδιώξουν και θα επιτεθούν στον ιδιοκτήτη ακινήτου για να αποκτήσουν πρόσβαση. Εάν το στοιχείο εξασφαλίζεται με μια βιομετρική συσκευή, η ζημία στον ιδιοκτήτη θα μπορούσε να είναι αμετάκλητη, και να κοστίσει ενδεχομένως περισσότερο από την εξασφαλισμένη ιδιοκτησία. Παραδείγματος χάριν, το 2005, οι μαλαισιανοί κλέφτες αυτοκινήτων κόβουν το δάχτυλο ενός ιδιοκτήτη s-class της Mercedes-Benz κατά την προσπάθεια τους να κλέψουν το αυτοκίνητο.

### 2.3.3 Ακυρώσιμη βιομετρική

Ένα πλεονέκτημα των κωδικών πρόσβασης πέρα από τη βιομετρική είναι ότι μπορούν να επανεκδοθούν. Εάν ένα σημείο ή ένας κωδικός πρόσβασης χάνεται ή κλέβεται, μπορεί να ακυρωθεί και να αντικατασταθεί από μια νεώτερη έκδοση. Αυτό δεν είναι φυσικά διαθέσιμο στη βιομετρική. Εάν το πρόσωπο κάποιου συμβιβάζεται από μια βάση δεδομένων, δεν μπορούν να το ακυρώσουν ή να επανεκδώσουν. Η ακυρώσιμη βιομετρική είναι ένας τρόπος με τον οποίο να ενσωματώσει την προστασία και τα χαρακτηριστικά γνωρίσματα αντικατάστασης στη βιομετρική.

Διάφορες μέθοδοι για την ακυρώσιμη βιομετρική έχουν προταθεί. Το πρώτο δακτυλικό αποτύπωμα βάσισε το ακυρώσιμο βιομετρικό σύστημα σχεδιάστηκε και αναπτύχθηκε από Tulyakov. Ουσιαστικά, η ακυρώσιμη βιομετρική εκτελεί μια διαστρέβλωση της βιομετρικών εικόνας ή των χαρακτηριστικών γνωρισμάτων πριν τα αντιστοιχίσουμε. Η μεταβλητότητα στις παραμέτρους διαστρεβλώσεως παρέχει την ακυρώσιμη φύση του σχεδίου. Μερικές από τις προτεινόμενες τεχνικές λειτουργούν χρησιμοποιώντας τις μηχανές αναγνώρισής τους, όπως Teoh, Savvides και άλλες μέθοδοι, όπως Dabbah, αξιοποιούν την πρόοδο της καθιερωμένης βιομετρικής έρευνας για front-end αναγνώρισης τους για να διευθύνουν την αναγνώριση. Αν και αυτό αυξάνει τους περιορισμούς στο σύστημα προστασίας, καθιστά τα cancellable πρότυπα πιο προσιτά για τις διαθέσιμες βιομετρικές τεχνολογίες.



### 2.3.4 Μαλακή βιομετρική

Τα Μαλακά βιομετρικά στοιχεία, λόγω της εσωτερικής φύσης τους, διατηρούν την προστασία της ιδιωτικής ζωής. Επιτρέπουν να περιγράψει ένα θέμα αρχής γενομένης από φυσικές ιδιότητες του. Οι ιδιότητες έχουν χαμηλή διακριτική εξουσία, έτσι δεν είναι σε θέση να αποδώσει ταυτότητα, επιπλέον είναι πλήρως διαθέσιμο σε όλους (π.χ. ύψος, βάρος, φύλο), γεγονός που τις καθιστά ασφαλή προστασία της ιδιωτικής ζωής.

### 2.3.5 Οι κυβερνήσεις είναι απίθανο να αποκαλύψουν τις πλήρεις ικανότητες των βιομετρικών επεκτάσεων

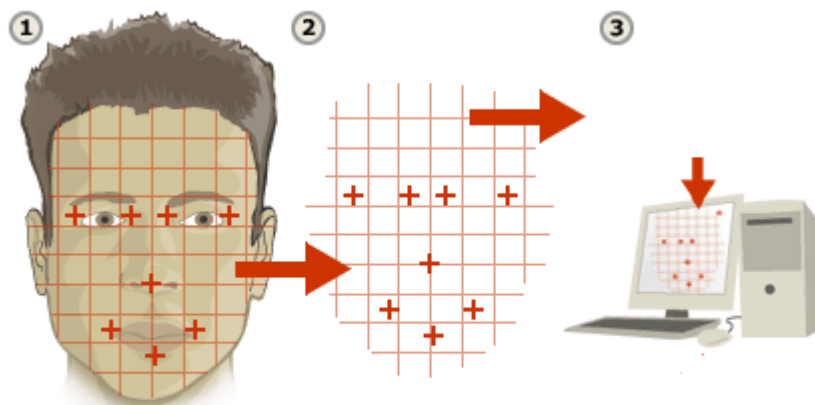
Ορισμένα μέλη της πολιτικής κοινότητας ανησυχούν για το πώς το βιομετρικό στοιχείο χρησιμοποιείται. Δυστυχώς, η πλήρης κοινοποίηση μπορεί να μην είναι προσεχής στην πολιτική κοινότητα.

## 2.4 Βιομετρική ανάλυση ταυτοποίησης ατόμου

Εκ των προτέρων γνωρίζουμε ότι δεν μπορούμε να υποστηρίξουμε ότι υπάρχει μια μέθοδος βιομετρικής ανάλυσης δεδομένων που μας παρακινεί να την καθορίσουμε σαν την πιο αξιόπιστη για την ταυτοποίηση ατόμου. Όμως κάθε βιομετρική μέθοδος ταυτοποίησης προτείνεται για χρήση αναλόγως με τις απαιτήσεις που χρειάζονται και τις συνθήκες που επικρατούν. Φυσικά κάποιες από αυτές είναι περισσότερο δημοφιλείς, άλλες μπορούν να λειτουργήσουν και χωρίς την γνώση του αντικείμενου που εξετάζουν, και άλλες είναι πολύ πιο δύσκολο να λάβουν λανθασμένες αποφάσεις.

### 2.4.1 Face Recognition (αναγνώριση προσώπου)

Από των διάφορων βιομετρικών μεθόδων ταυτοποίησης, η αναγνώριση προσώπου είναι μια από τις πιο ευέλικτες, και λειτουργεί ορθά ακόμα και κάτω από δύσκολες συνθήκες, όπως ότι το πρόσωπο που πρέπει να αναγνωριστεί δεν χρειάζεται να είναι 'στημένο' για μεγάλο χρονικό διάστημα μπροστά σε μια ψηφιακή κάμερα, ούτε να είναι το μοναδικό μέσα στην φωτογραφία που θα ληφθεί.



Εικόνα 4: Ανάλυση και Αναγνώριση Προσώπου

Η αναγνώριση προσώπου λειτουργεί μέσω της συστηματικής ανάλυσης των ιδιαίτερων

χαρακτηριστικών που είναι κοινά σε όλα τα πρόσωπα, η απόσταση των ματιών, το πλάτος της μύτης, η θέση των ζυγωματικών, η γραμμή του σαγονιού, του πηγουνιού και ούτω καθεξής. Αυτές οι αριθμητικές ποσότητες στην συνέχεια συνδυάζονται σε ένα ενιαίο κώδικα που προσδιορίζει μοναδικά κάθε άτομο.

#### 2.4.2 Fingerprint identification (ταυτοποίηση δακτυλικών αποτυπωμάτων)

Τα δακτυλικά αποτυπώματα είναι μη μεταβλητά για ολόκληρη την ζωή μας. Σε πάνω από 140 χρόνια σε όλο τον κόσμο η σύγκριση δακτυλικών αποτυπωμάτων ποτέ δεν έχει βρει ίδια για δυο διαφορετικά άτομα, ακόμα και για πανομοιότυπων διδύμων. Σαρωτές δακτυλικών αποτυπωμάτων πλέον έχουν εγκατασταθεί ακόμα και σε iPAQ, Rocket PCs και PDAs. Μπορεί πλέον να είναι εύκολη η χρήση αυτής της μεθόδου αλλά παρόλα αυτά παραμένουν απαγορευτικές για την χρήση τους για συστήματα ασφαλείας αφού απαιτούν καθαρά χέρια, πράγμα που δεν μπορεί να εξασφαλισθεί σε όλες τις περιπτώσεις.



Εικόνα 5: Λήψη Δεδομένων από το Δακτυλικό αποτύπωμα

Η ταυτοποίηση δακτυλικών αποτυπωμάτων ουσιαστικά συγκρίνει τον τρόπο με τον οποίο καταγράφονται η κορυφογραμμές και τα αυλάκια στις άκρες των δακτύλων, καθώς και μικρά σημεία που παρέχουν όμως σημαντικές πληροφορίες – χαρακτηριστικά όπως μια κορυφογραμμή που χωρίζεται σε δυο – ενός δείγματος εκτύπωσης με μια βάση δεδομένων των εκτυπώσεων σε αρχείο.

Η γεωμετρία του χεριού είναι ένας τομέας της Βιομετρικής που καθορίζει το άτομο από το σχήμα του χεριού του, το μέγεθος της παλάμης καθώς και από τα μήκη και πλάτη των δακτύλων. Οι πρώτες συσκευές καταμέτρησης της γεωμετρίας του χεριού κατασκευάστηκαν στις αρχές της δεκαετίας του 1980 και ήταν το πρώτο βιομετρικό γνώρισμα που έγινε ευρεία η μηχανογραφική του χρήση. Όμως δεν μπόρεσε να κρατηθεί για πολύ στο προσκήνιο αφού δεν ήταν τόσο μοναδικό για κάθε άνθρωπο όπως είναι το δακτυλικό αποτύπωμα ή ίριδα. Δεδομένου αυτού δεν χρησιμοποιείτε σε λογισμικά ταυτοποίησης ατόμου, παρά μόνο με τον συνδυασμό και άλλων μορφών αναγνώρισης όπως οι κάρτες ταυτότητας ή προσωπικών αριθμών, παραμένοντας μη αξιόπιστο σύστημα για την χρήση του σε περιβάλλον με μεγάλο εύρος χρηστών.

#### 2.4.3 Σάρωση Αμφιβληστροειδή

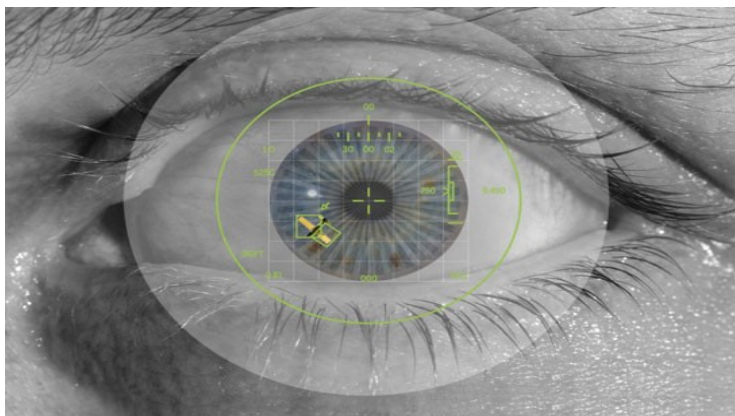
Ο ανθρώπινος αμφιβληστροειδής είναι ένας λεπτός ιστός αποτελούμενος από νευρικά κύτταρα, τα οποία βρίσκονται στο οπίσθιο τμήμα του ματιού. Λόγω της πολυπλοκότητας της δομής των τριχοειδών

αγγείων που τροφοδοτούν τον αμφιβληστροειδή με το αίμα, ο αμφιβληστροειδής παραμένει για το κάθε άτομο μοναδικό. Το δίκτυο των αιμοφόρων αγγείων είναι τόσο περίπλοκο που ακόμα και πανομοιότυπα δίδυμα δεν μοιράζονται παρόμοιο μοτίβο.



Εικόνα 6: Η όψη ενός αμφιβληστροειδή

Παρά το γεγονός ότι ο αμφιβληστροειδής μπορεί να μεταβληθεί σε περιπτώσεις που το άτομο πάσχει από σακχαρώδη διαβήτη, γλαύκωμα ή εκφυλιστικές διαταραχές του αμφιβληστροειδή, παραμένει συνήθως αμετάβλητος από τη γέννηση μέχρι το θάνατο. Λόγω λοιπόν της μοναδικότητάς του, φαίνεται να είναι η πιο ακριβής και αξιόπιστη βιομετρική μέθοδος ταυτοποίησης ατόμου. Μάλιστα υπολογίζεται ότι το ποσοστό λάθους κυμαίνεται στο ένα στο εκατομμύριο. Η ιδέα για την αναγνώριση ατόμου μέσω του αμφιβληστροειδή συνελήφθη πρώτη φορά από τον Dr. Carleton Simon και Dr. Isadore Goldstein το 1953. Η ιδέα ήταν αρκετά εξελιγμένη για την εποχή του, αφού δεν υπήρχε η τεχνολογία για να την αποδείξει. Το 1975 η τεχνολογία είχε εξελιχθεί αρκετά ώστε να μπορέσει να έρθει και πάλι στο προσκήνιο η ιδέα για μια συσκευή σάρωσης του αμφιβληστροειδή, καταφέροντας να υλοποιηθεί το 1978 από τον Robert “Buzz” Hill και την εταιρία του EyeDentify, Inc.



Εικόνα 7: Τρόπος ανάλυσης ματιού

Μια σάρωση αμφιβληστροειδή βασίζεται στο γεγονός ότι τα αιμοφόρα αγγεία μέσα στον αμφιβληστροειδή απορροφούν το φως πιο εύκολα από το περιβάλλοντα ιστό και μέσω μια μικρής δέσμης υπέρυθρου φωτός χαμηλής ενέργειας μπορεί να καταγραφεί το ίχνος του τυποποιημένου μονοπατιού πάνω στον αμφιβληστροειδή. Το μοτίβο των διακυμάνσεων που καταγράφεται μετατρέπεται σε κώδικα και αποθηκεύεται σε μια βάση δεδομένων.

#### 2.4.4 Σάρωση Ίριδας

Η σάρωση της ίριδας του ματιού παρέχει μοναδικά βιομετρικά δεδομένα για κάθε άτομο, οποία είναι πολύ δύσκολο να συναντηθούν σε άλλο άτομο με ακριβώς την ίδια λεπτομέρεια και παραμένει αναλλοίωτη εφόρου ζωής, όπως ακριβώς συμβαίνει και με τον αμφιβληστροειδή. Έχουν αναπτυχθεί τρόποι μέσω των οποίων μπορεί να γίνει η κωδικοποίηση των δεδομένων που λαμβάνονται από την σάρωση της ίριδας κατά τέτοιο τρόπο ώστε η διαχείρισή τους να πραγματοποιείται με ασφάλεια σε μια μορφή “barcode”.

Η σάρωση και η ανάλυση της ίριδας του ματιού γίνεται με την βοήθεια μιας λήψης υψηλής ανάλυσης φωτογραφίας του ματιού. Μέσω χρήσης υπεριώδους φωτισμού μειώνεται η κατοπτρική ανάκλαση από τον κερατοειδή. Η διατήρηση της ίριδας εφόρου ζωής οφείλεται στο γεγονός ότι πρόκειται για ένα εσωτερικό όργανο το οποίο είναι αρκετά προστατευμένο. Ενώ πρόκειται για μια τεχνική η οποία δεν επηρεάζεται από την χρήση φακών επαφής ή γυαλιών μυωπίας, αλλά μόνο τα βλέφαρα και τις βλεφαρίδες, δύο μέρη που μπορούν κάλλιστα να αφαιρέσουν χρήσιμες πληροφορίες καλύπτοντας μέρος του ματιού.

Όπως συμβαίνει με τις περισσότερες μεθόδους βιομετρικής αναγνώρισης, ένα πρόβλημα μένει ακόμα σε μη ικανοποιητικά επίπεδα λύσης και είναι αυτό που παρουσιάζεται με την επαλήθευση του ζωντανού ιστού. Πολλά συστήματα αναγνώρισης ίριδας που είναι διαθέσιμα στο εμπόριο είναι εύκολο να αποπλανηθούν και να δώσουν ψευδές αποτέλεσμα με τον απλό τρόπο της προβολής ενός μη ζωντανού ιστού – για παράδειγμα μέσω μιας φωτογραφίας υψηλής ανάλυσης έναντι ενός πραγματικού προσώπου – και αποτέλεσμα αυτού να κρίνονται μη αξιόπιστες συσκευές για τις εφαρμογές τις οποίες κλήθηκαν να υλοποιήσουν, όπως ο έλεγχος πρόσβασης σε θύρες ασφαλείας. Το πρόβλημα της επαλήθευσης ζωντανού ιστού είναι λιγότερο ανησυχητικό στις εποπτευόμενες εφαρμογές, όπου ένας ανθρώπινος χειριστής εποπτεύει τη διαδικασία.

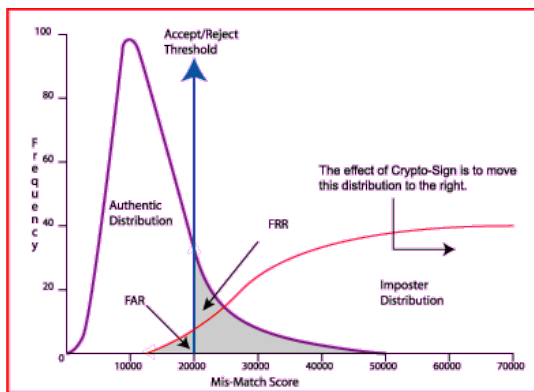
Οι μέθοδοι που έχουν προταθεί για να παρέχουν κάποια προστασία ενάντια στη χρήση των πλαστών ματιών περιλαμβάνουν:

1. Μεταβαλλόμενο περιβαλλοντικό φωτισμό κατά τη διάρκεια του προσδιορισμού της ίριδας, έτσι ώστε η αντανάκλαση της κόρης του οφθαλμού να μπορεί να ελεγχθεί και η εικόνα της ίριδας να καταγράφεται για διαφορετικές διαμέτρους της κόρης του ματιού.
2. Αναλύοντας το χρονικό φάσμα συχνότητας της εικόνας για αιχμές που προκαλούνται από τις απεικονίσεις στους υπολογιστές.
3. Χρησιμοποίηση της φασματικής ανάλυσης της εικόνας έναντι των μονοχρωματικών φωτογραφικών μηχανών για να διακρίνει τον ιστό της ίριδας από άλλο υλικό.
4. Παρατήρηση της χαρακτηριστικής κίνησης ενός βολβού του ματιού.
5. Δοκιμή για αμφιβληστροειδικό retro reflection (επίδραση κόκκινων ματιών)
6. Εξετάζοντας τις αντανάκλασεις από τις τέσσερις οπτικές επιφάνειες του ματιού.
7. Χρησιμοποίηση της τρισδιάστατης απεικόνισης για να ελέγξει τη θέση και τη μορφή της ίριδας σχετικά με άλλα χαρακτηριστικά γνωρίσματα.

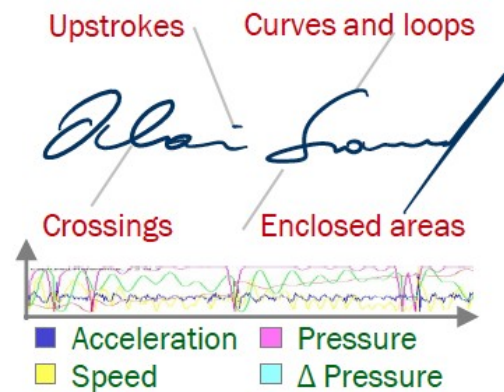
#### 2.4.5 Αναγνώριση Υπογραφής

Η υπογραφή εγγράφων είναι κάτι που οι περισσότεροι ενήλικες είναι εξοικειωμένοι. Στην προσωπική μας ζωή υπογράφουμε τα πάντα, από προσωπικές επιταγές μέχρι και τις κάρτες γενεθλίων. Στον επιχειρηματικό κόσμο που υπογράφουν τα επίσημα έγγραφα, καθιστά την αναγνώριση υπογραφής αρκετά δημοφιλή για μια βιομετρική μέθοδο ταυτοποίησης κυρίως στο ηλεκτρονικό εμπόριο. Η βιομετρική αναγνώριση υπογραφής λειτουργεί σε ένα τρισδιάστατο περιβάλλον, όπου δεν μετρείται μόνο το ύψος και το πλάτος της πένας αλλά και το μέγεθος πίεσης που εφαρμόζεται ώστε να μετρηθεί το βάθος που θα μπορούσε να εμφανιστεί μέσω της διαδρομής της πένας στον αέρα. Αυτό βοηθά να μειωθεί ο κίνδυνος πλαστογραφίας που μπορεί να συμβεί σε σύστημα αναγνώρισης υπογραφής δύο διαστάσεων.

Η δυναμική αναγνώριση υπογραφής χρησιμοποιεί πολλά χαρακτηριστικά στην ανάλυση του γραφικού χαρακτήρα ενός ατόμου. Τα χαρακτηριστικά αυτά διαφέρουν σε χρήση και σημασία από συσκευή σε συσκευή και συλλέγονται μέσω της επαφής ευαίσθητων τεχνολογιών, όπως τα PDA. Τα περισσότερα από τα χαρακτηριστικά που χρησιμοποιούνται είναι δυναμικά και όχι στατικά γεωμετρικά χαρακτηριστικά, αν και ορισμένες συσκευές περιλαμβάνουν τα χαρακτηριστικά αυτά στις αναλύσεις τους. Τα κοινά δυναμικά χαρακτηριστικά περιλαμβάνουν την ταχύτητα, την επιτάχυνση, το χρονοδιάγραμμα, την πίεση και την κατεύθυνση της διαδρομής της υπογραφής όπου και αναλύονται στις τρεις διαστάσεις x,y,z άξονες. Ορισμένοι αλγόριθμοι δυναμικής αναγνώρισης υπογραφής ενσωματώνουν μια μαθησιακή λειτουργία για να λαμβάνονται υπόψη οι φυσικές αλλαγές ή τα ολισθήματα που συμβαίνουν στην υπογραφή ενός ατόμου με τον χρόνο. Τα χαρακτηριστικά που χρησιμοποιούνται για την δυναμική αναγνώριση υπογραφής είναι σχεδόν αδύνατο να αναπαραχθούν. Σε αντίθεση με μια γραφική εικόνα υπογραφής, η οποία μπορεί να αναπαραχθεί από κάποιον πλαστογράφο, τα δυναμικά χαρακτηριστικά είναι πολύπλοκα και μοναδικά για το κάθε άτομο.



Εικόνα 9: Διάγραμμα υπογραφής



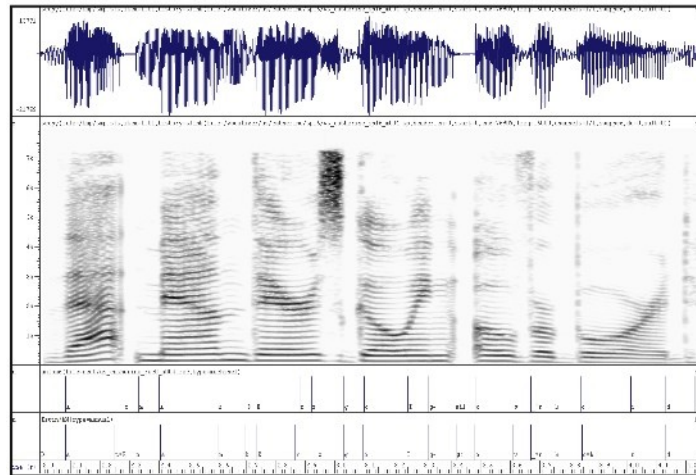
Εικόνα 8: Ανάλυση Υπογραφής

Ένα μειονέκτημα σε αυτή τη μορφή της κρυπτογράφησης είναι ότι οι άνθρωποι δεν υπογράφουν πάντα τα έγγραφα με τον ίδιο ακριβώς τρόπο. Η γωνία με την οποία υπογράφουν μπορεί να είναι διαφορετική λόγω της θέσης καθίσματος ή από τη θέση των χεριών στην επιφάνεια γραφής. Ως εκ τούτου ακόμα κι αν είναι τρισδιάστατο το σύστημα αναγνώρισης υπογραφής του προσθέτει την ικανότητα να εντοπίζει πλαστογραφίες, δεν είναι τόσο ακριβής όσο άλλες μορφές βιομετρικής ταυτοποίησης.

## 2.4.6 Ανάλυση Φωνής

Η ανάλυση της φωνής στην βιομετρική έχει απώτερο σκοπό την αυτόματη αναγνώριση του ομιλητή χρησιμοποιώντας τις πληροφορίες που λαμβάνονται από τις κυματομορφές. Η διαδικασία αναγνώρισης φωνής στηρίζεται στα χαρακτηριστικά γνωρίσματα που επηρεάζονται και από τη φυσική δομή του φωνητικού αποτυπώματος ενός ατόμου και από τα συμπεριφοριστικά χαρακτηριστικά του ατόμου.

Για να επιτευχθεί όμως η αναγνώριση του ομιλητή μέσω της φωνής υπάρχουν δύο μέθοδοι υλοποίησης, εξαρτώμενη του κειμένου και η ανεξάρτητη του κειμένου. Η πρώτη θέτει κάποιους περιορισμούς καθώς το άτομο που θέλουμε να αναγνωρίσουμε πρέπει να διατυπώνει μια σταθερή φράση, όπου και είναι προγραμματισμένη στο σύστημα. Αντίθετα σε ένα ανεξάρτητο του κειμένου σύστημα δεν έχει καμία γνώση της φωνής του ομιλητή και είναι πιο εύκαμπτο στις καταστάσεις όπου το άτομο που υποβάλλεται στην δειγματοληψία είναι απληροφόρητο της συλλογής του δείγματος ή απρόθυμο να συνεργαστεί. Αυτός είναι και ο λόγος που ένα τέτοιο σύστημα παρουσιάζει μεγαλύτερη πρόκληση.



Εικόνα 10: Ανάλυση Φωνητικού Κύματος

Εφαρμογές που μπορείς να λάβει μέρος αυτή η διαδικασία είναι η φωνητική κλήση, τραπεζικές συναλλαγές μέσω τηλεφωνικού δικτύου, υπηρεσίες πρόσβασης σε βάσεις δεδομένων, υπηρεσίες πληροφοριών και κρατήσεις, φωνητικό ταχυδρομείο, τον έλεγχο ασφαλείας των εμπιστευτικών πληροφοριών, καθώς και απομακρυσμένη πρόσβαση σε υπολογιστές. Μια άλλη σημαντική εφαρμογή της τεχνολογίας αναγνώρισης του ομιλητή είναι η χρήση της στην εγκληματολογία.

#### 2.4.7 Δυναμική πληκτρολόγηση

Η δυναμική πληκτρολόγηση χρησιμοποιεί τον τρόπο και τον ρυθμό με τον οποίο ένα άτομο πληκτρολογεί χαρακτήρες σε ένα πληκτρολόγιο για να μπορέσει να αναγνωρίσει την ταυτότητα του. Ο ρυθμός πληκτρολόγησης κάθε χρήστη καταγράφεται και χρησιμοποιείται για να αναπτυχθεί ένα μοναδικό βιομετρικό πρότυπο για τον καθένα ξεχωριστά, ώστε να χρησιμοποιηθούν μελλοντικά για σύγκριση και επικύρωση. Οι ακατέργαστες μετρήσεις που είναι διαθέσιμες από κάθε πληκτρολόγιο μπορούν να καταγραφούν για να καθορίσουν τον χρόνο που ένα πλήκτρο πιέζεται και τον χρόνο πτήσης. Το κατεγγραμμένο στοιχείο συγχρονισμού πληκτρολόγησης υποβάλλεται σε επεξεργασία μέσω ενός μοναδικού νευρωνικού αλγορίθμου, ο οποίος καθορίζει ένα αρχικό σχέδιο για τη μελλοντική σύγκριση.

Τα δεδομένα που απαιτούνται για την ανάλυση δυναμικής πληκτρολόγησης επιτυγχάνονται με την καταγραφή της πληκτρολόγησης. Κανονικά, το μόνο που διατηρείται κατά τη σύνδεση μιας συνόδου πληκτρολόγησης είναι η ακολουθία των χαρακτήρων που αντιστοιχεί στη σειρά με την οποία τα πλήκτρα είχαν πιεστεί και πληροφορίες χρονισμού απορρίπτονται. Κατά την ανάγνωση ενός e-mail, ο δέκτης δεν μπορεί να καταλάβει βασικές πληροφορίες που χαρακτηρίζουν τον τρόπο γραφής ενός ατόμου. Τέτοιες πληροφορίες που αποτελούν χαρακτηριστικά τα οποία εκμεταλλεύεται η αναγνώριση δυναμικής πληκτρολόγησης είναι:

- Πόσο γρήγορη ήταν η πληκτρολόγηση του μηνύματος
- Για τα κεφαλαία γράμματα ποιο shift χρησιμοποιήθηκε ή το Caps Lock
- Τα γράμματα δακτυλογραφήθηκαν με τον ίδιο ρυθμό ή υπήρχε μεγάλη παύση ανάμεσα σε λέξεις
- Και την συχνότητα λαθών που κάνει κατά την πληκτρολόγηση

Η δυναμική πληκτρολόγησης είναι μέρος μιας μεγαλύτερης κατηγορίας βιομετρικής, γνωστής ως συμπεριφοριστική βιομετρική στατιστικής φύσης. Λέγεται πως η συμπεριφοριστική βιομετρική δεν είναι τόσο αξιόπιστη όσο η φυσική βιομετρική που χρησιμοποιείται για την επικύρωση όπως τα δακτυλικά αποτυπώματα ή αμφιβληστροειδικές ανιχνεύσεις ή DNA. Οι παραδοσιακές συγκριτικές μετρήσεις επιδόσεων του ψεύτικου ποσοστού αποδοχής (FAR) και τα ψεύτικα ποσοστά απόρριψης (FRR) δεν έχουν πλέον τις γραμμικές σχέσεις. Το όφελος στη δυναμική πληκτρολόγησης (καθώς επίσης και σε άλλη συμπεριφοριστική βιομετρική) είναι ότι FRR/FAR και μπορεί να ρυθμιστεί με την αλλαγή του κατώτατου ορίου αποδοχής στο μεμονωμένο επίπεδο. Αυτό επιτρέπει το ρητά καθορισμένο μεμονωμένο κίνδυνο κάτι που οι μετριάσμενες βιομετρικές τεχνολογίες δεν θα μπορούσαν ποτέ να επιτύχουν. Ένα άλλο όφελος της δυναμικής πληκτρολόγησης: είναι ότι μπορούν να λαμβάνονται συνεχώς- οι όχι ακριβώς στο ξεκίνημα χρόνοι-και μπορούν να είναι επαρκώς ακριβείς για να προκαλέσουν έναν συναγερό σε ένα άλλο σύστημα ή άτομο για να έρθουν να ελέγξουν την κατάσταση.



## 3. Μεθοδολογία Υλοποίησης

### 3.1 Ανίχνευση Προσώπου

Σήμερα κάποιες εφαρμογές της αναγνώρισης προσώπου δεν απαιτούν και την ανίχνευση του. Σε ορισμένες περιπτώσεις, οι εικόνες προσώπων που αποθηκεύονται στην βάση δεδομένων έχουν ήδη τροποποιηθεί ώστε να λάβουμε το τελικό αποτέλεσμα. Υπάρχει μια τυποποιημένη μορφή εισόδου της εικόνας και έτσι δεν υπάρχει η ανάγκη για την πραγματοποίηση της ανίχνευσης. Ένα παράδειγμα αυτού θα μπορούσε να αποτελεί η βάση δεδομένων ποινικού μητρώου. Εάν υπάρχει νέα καταχώριση και η αστυνομία έχει τη φωτογραφία διαβατηρίου, η ανίχνευση του προσώπου δεν είναι απαραίτητη. Ωστόσο, η συμβατική εικόνα εισόδου των συστημάτων υπολογιστικής όρασης δεν είναι τόσο κατάλληλη. Μπορούν να περιέχουν πολλά αντικείμενα ή πρόσωπα. Σε αυτές τις περιπτώσεις η ανίχνευση προσώπου είναι υποχρεωτική. Είναι επίσης αναπόφευκτη αν θέλουμε να αναπτύξουμε ένα αυτοματοποιημένο σύστημα εντοπισμού προσώπου. Για παράδειγμα, τα συστήματα παρακολούθησης βίντεο προσπαθούν να περιλαμβάνουν ανίχνευση προσώπου, την παρακολούθηση τους και την αναγνώριση αυτών. Έτσι, είναι λογικό να υποθέσουμε ότι η ανίχνευση προσώπου είναι ένα μέρος του προβλήματος της αναγνώρισης προσώπων.

Η ανίχνευση προσώπων καλείται να αντιμετωπίσει διάφορες προκλήσεις, οι περισσότερες εκ των οποίων παρουσιάζονται σε εικόνες που έχουν ληφθεί σε ανεξέλεγκτο περιβάλλον, όπως τα συστήματα παρακολούθησης βίντεο. Οι προκλήσεις αυτές μπορούν να αποδοθούν σε ορισμένους παράγοντες:

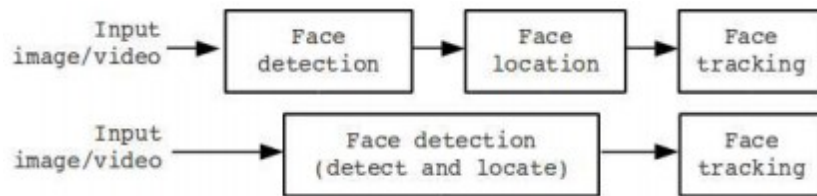
- **Pose Variation.** Το ιδανικό σενάριο για την ανίχνευση προσώπου θα είναι το αντικείμενο προς αναζήτηση να έχει μετωπική θέση στην εικόνα. Αλλά, όπως αναφέρθηκε, αυτό είναι πολύ απίθανο σε γενικές γραμμές όταν το περιβάλλον δεν είναι ελεγχόμενο. Επιπλέον, η απόδοση αλγορίθμων ανίχνευσης προσώπων αποτυγχάνουν σε μεγάλο βαθμό όταν υπάρχει μεγάλη γωνία λήψης του αντικειμένου. Είναι ένα μείζον θέμα της έρευνας, Η μεταβολή της θέσης μπορεί να συμβεί λόγω των κινήσεων του υποκειμένου ή την γωνία της κάμερας.
- **Feature Occlusion.** Η παρουσία στοιχείων όπως γενειάδες, γυαλιά ή καπέλα παρουσιάζει υψηλή μεταβλητότητα στην αποτελεσματικότητα των αλγορίθμων. Τα πρόσωπα μπορούν επίσης να καλύπτονται εν μέρει από αντικείμενα ή άλλα πρόσωπα και να μην ανιχνευθούν από τους αλγόριθμους.
- **Facial Expression.** Τα χαρακτηριστικά του προσώπου μπορεί να διαφέρουν σε μεγάλο βαθμό κατά την διάρκεια έκφρασης συναισθημάτων με αποτέλεσμα να μην μπορούν να γίνουν τόσο αντιληπτά.
- **Imaging Conditions.** Διαφορετικές φωτογραφικές μηχανές και περιβαλλοντικές συνθήκες μπορούν να επηρεάσουν την ποιότητα της εικόνας, με αποτέλεσμα να επηρεαστεί και η εμφάνιση ενός προσώπου.

Υπάρχουν κάποια προβλήματα που σχετίζονται στενά με την ανίχνευση προσώπου, εκτός από της εξαγωγή των χαρακτηριστικών και την ταξινόμηση του προσώπου. Για παράδειγμα, η θέση του προσώπου είναι μια απλοποιημένη προσέγγιση της ανίχνευσης προσώπων. Στόχος της είναι να προσδιορίσει την θέση ενός προσώπου μέσα σε μια εικόνα, όπου υπάρχει μόνο ένα πρόσωπο. Μπορούμε να διαφοροποιήσουμε την ανίχνευση προσώπου με την εύρεση τοποθεσίας προσώπου δεδομένου ότι το τελευταίο είναι μια απλοποιημένη μορφή της πρώτης. Πολλά συστήματα ανίχνευσης προσώπων έχουν σκοπό όχι μόνο την ανίχνευση του προσώπου, αλλά την πραγματοποίηση του σε πραγματικό χρόνο.

#### 3.1.1 Δομή προβλήματος

Η ανίχνευση προσώπων είναι μια έννοια που περιλαμβάνει πολλά επιμέρους προβλήματα. Μερικά συστήματα αναγνωρίζουν και εντοπίζουν τα πρόσωπα την ίδια στιγμή, ενώ άλλα πρώτα πραγματοποιούν την ρουτίνα της ανίχνευσης και στην συνέχεια εντοπίζουν την τοποθεσία του προσώπου, εάν φυσικά η πρώτη διεργασία τελειώσει με θετικά αποτελέσματα. Στην συνέχεια υπάρχει η πιθανότητα να απαιτηθούν κάποιοι αλγόριθμοι εντοπισμού (tracking algorithms).





Οι αλγόριθμοι Face detection συνήθως έχουν κοινά βήματα. Πρώτον, πραγματοποιείτε μείωση ορισμένων δεδομένων διαστάσεων, προκειμένου να επιτευχθεί μια αποδεκτή χρονική απόκριση, αλλά και προ επεξεργασία της εικόνας πριν την εισαγωγή της στον αλγόριθμο είναι ένα σύννητες φαινόμενο ώστε να πληρεί τις προϋποθέσεις του αλγόριθμου. Στην συνέχεια κάποιοι αλγόριθμοι αναλύουν την εικόνα όπως είναι κάποιοι άλλοι προσπαθούν να εξάγουν ορισμένες σχετικές περιοχές του προσώπου. Η επόμενη φάση συνήθως περιλαμβάνει την εξαγωγή χαρακτηριστικών του προσώπου ή των μετρήσεων. Αυτές στην συνέχεια θα σταθμίζονται, αξιολογούνται ή θα συγκρίνονται και αποφασίζεται εάν υπάρχει πρόσωπο εν τέλη ή όχι. Τέλος ορισμένοι αλγόριθμοι έχουν μια μαθησιακή ρουτίνα και περιλαμβάνουν τα νέα δεδομένα με τα μοντέλα τους.

Η λειτουργία του Face detection, ως εκ τούτου, είναι ένα πρόβλημα 2 τάξεων, όπου πρέπει να αποφασίσουμε αν υπάρχει ένα πρόσωπο ή όχι μέσα σε μια εικόνα. Αυτή η προσέγγιση μπορεί να θεωρηθεί ως μια απλοποιημένη αντιμετώπιση του προβλήματος της αναγνώρισης προσώπου. Και αυτό παρατηρείτε εάν συγκρίνουμε της μεθόδους που χρησιμοποιούμε για την αναγνώριση προσώπων και την ανίχνευση προσώπων, δίνοντας μας το τελικό συμπέρασμα ότι οι τεχνικές και των δύο προβλημάτων είναι αρκετές φορές παρόμοιες μεταξύ τους.

### 3.1.2 Προσεγγίσεις στην ανίχνευση προσώπου

Δεν είναι αρκετά εύκολο να ταξινομήσουμε τις μεθόδους ανίχνευσης προσώπου, καθώς δεν υπάρχουν παγκοσμίως αποδεκτά κριτήρια ομαδοποίησης τους. Σε αυτό το κομμάτι θα παρουσιάσουμε δύο κριτήρια ταξινόμησης. Ένας από αυτούς διαφοροποιεί τα μεταξύ τους διακριτά σενάρια. Ανάλογα με αυτά τα σενάρια διαφορετικές προσεγγίσεις μπορεί να απαιτηθούν. Τα υπόλοιπα κριτήρια διαχωρίζουν τους αλγόριθμους ανίχνευσης προσώπου σε τέσσερις κατηγορίες.

#### *Ανίχνευσης ανάλογα με το σενάριο*

- Όταν το περιβάλλον το οποίο λαμβάνεται η εικόνα ή το βίντεο πρόκειται για ελεγχόμενο περιβάλλον τότε έχουμε να αντιμετωπίσουμε την πιο απλοϊκή περίπτωση. Οι φωτογραφίες λαμβάνονται υπό ελεγχόμενο φωτισμό, υπόβαθρο κλπ. Μια απλή τεχνική ανίχνευσης ακμών μπορεί να χρησιμοποιηθεί για την ανίχνευση προσώπων.

- Έχοντας στην κατοχή μας εγχρωμες εικόνας για ανάλυση και επεξεργασία, αρκετά συχνά χρησιμοποιούμε το τυπικό χρώμα του δέρματος ώστε να εντοπίσουμε το πρόσωπο. Όμως αυτή η τεχνική μπορεί να φανεί αδύναμη, καθώς εάν αλλάξει ο φωτισμός το χρώμα του δέρματος θα αλλοιωθεί, με αποτέλεσμα την λανθασμένη λειτουργία της αναγνώρισης προσώπου. Επιπλέον, το ανθρώπινο δέρμα έχει ένα μεγάλο εύρος χρωματισμού, από λευκό έως σχεδόν μαύρο. Όμως, αρκετές μελέτες έχουν αποδείξει ότι η κύρια διαφορά κυμαίνεται στην μεταξύ τους ένταση, έτσι ο χρωματισμός είναι ένα αρκετά καλό χαρακτηριστικό. Δεν είναι εύκολο να δημιουργηθεί μια σταθερή απεικόνιση των χρωμάτων για το ανθρώπινο δέρμα. Ωστόσο υπάρχουν προσπάθειες για τη συγκρότηση ισχυρού αλγόριθμου ανίχνευσης προσώπου με βάση το χρώμα δέρματος.

- Εικόνες σε κίνηση. Η εύρεση προσώπων σε επεξεργασία βίντεο σε πραγματικό χρόνο μας δίνει την δυνατότητα να χρησιμοποιήσουμε την κίνηση του προσώπου για να μπορέσουμε να το εντοπίσουμε.

Σήμερα , τα περισσότερα εμπορικά συστήματα πρέπει να εντοπίζουν πρόσωπα μέσα σε βίντεο.

Υπάρχει μια διαρκής πρόκληση για την βελτίωση των αποτελεσμάτων στον εντοπισμό προσώπων με την καλύτερη δυνατή απόδοση. Μια άλλη προσέγγιση που βασίζεται στην κίνηση των ματιών είναι η ανίχνευση αναλαμπής, η οποία έχει πολλές χρήσεις εκτός από την ανίχνευση των προσώπων.

### *Χωρισμός μεθόδων ανίχνευσης σε κατηγορίες*

Οι Yan, Kriegman και Ahuja παρουσίασαν μια ταξινόμηση που έγινε κοινός αποδεκτή για τους αλγόριθμους ανίχνευσης προσώπων. Οι μέθοδοι χωρίζονται σε τέσσερις κατηγορίες. Οι κατηγορίες αυτές μπορεί να επικαλύπτονται, έτσι ώστε ένας αλγόριθμος να μπορεί να ανήκει σε δύο ή περισσότερες κατηγορίες. Η κατάταξη αυτή μπορεί να γίνει ως εξής:

- Μέθοδοι που βασίζονται στην γνώση (Knowledge-based). Πρόκειται για μεθόδους που κωδικοποιούν της γνώση μας για τα ανθρώπινα πρόσωπα και την καταγράφουν σε βάσεις δεδομένων.
- Μέθοδοι που βασίζονται στην αναζήτηση και την ανάλυση χαρακτηριστικών και αμετάβλητων σημείων του ανθρώπινου προσώπου. (Feature-invariant methods)
- Μέθοδοι που βασίζονται στην σύγκριση και στο ταίριασμα των προτύπων (Template Matching Methods) . Οι αλγόριθμοι συγκρίνουν τις εικόνες εισόδου με αποθηκευμένα πρότυπα προσώπου ή χαρακτηριστικών.
- Μέθοδοι που βασίζονται στην εμφάνιση (Appearance-based methods). Μια μέθοδος που ταιριάζει το πρότυπο με το μοντέλο μέσα από μια βάση δεδομένων που αντλήθηκε από ένα σύνολο εικόνων κατάρτισης.

Αναλυτικότερα για αυτές τις κατηγορίες:

### **Knowledge-based Methods.**

Υπάρχουν σε μεθόδους που είναι βασισμένες σε κανόνες. Προσπαθούν να συλλάβουν την γνώση μας για τα πρόσωπα και να τα μεταφράσουν σε έναν γενικό κανόνα. Είναι εύκολο να μαντέψει κανείς μερικούς απλούς κανόνες. Για παράδειγμα, ένα πρόσωπο έχει συνήθως δύο μάτια συμμετρικά μεταξύ τους και η περιοχή τους έχει πιο σκούρο χρώμα από αυτή που βρίσκονται τα μάγουλα. Χαρακτηριστικά του προσώπου θα μπορούσε να είναι η απόσταση ανάμεσα στα μάτια ή η διαφορά έντασης του χρώματος ανάμεσα στην περιοχή των ματιών και της κατώτερης περιοχής του προσώπου. Το μεγαλύτερο ελάττωμα με αυτές τις μεθόδους είναι η δυσκολία δημιουργίας ενός κατάλληλου γενικού κανόνα. Τα αποτελέσματα θα μπορούσαν να είναι αρκετά απογοητευτικά εάν οι κανόνες ήταν πολύ γενική και αόριστοι. Από τη άλλη πλευρά, θα μπορούσαν να υπάρξουν και πολλά ψευδώς αρνητικά αποτελέσματα, εάν οι κανόνες ήταν υπερβολικά λεπτομερείς και αυστηροί. Μια λύση είναι να οικοδομήσουμε ιεραρχικά μεθόδους γνώσης για να ξεπεραστούν αυτά τα προβλήματα. Ωστόσο, η προσέγγιση αυτή και μόνο είναι πολύ περιορισμένη. Δεν είναι εφικτός να ανιχνεύσει πολλά πρόσωπα μέσα σε μια σύνθετη εικόνα.

Άλλες έρευνες έχουν προσπαθήσει να βρουν κάποια αμετάβλητα χαρακτηριστικά για την ανίχνευση προσώπου. Η ιδέα είναι να ξεπεραστούν τα όρια της ενστικτώδης γνώσης μας περί των προσώπων. Ένας πρώιμος αλγόριθμος αναπτύχθηκε από τους Han, Liao και Yu Chen το 1997. Η μέθοδος αυτή χωρίζεται σε διάφορα στάδια. Πρώτα από όλα, προσπαθεί να βρει τα μάτια, έτσι ώστε να αφαιρέσει τα ανεπιθύμητα pixel από την εικόνα. Μετά από αυτήν την διαδικασία κατάτμησης, θεωρούν κάθε τμήμα που βρέθηκε ως μάτι, υποψήφιο ενός από των δύο ματιών. Στην συνέχεια μέσω ενός συνόλου κανόνων αναζητάτε το ταίρι του ματιού ώστε να δημιουργηθεί το ζευγάρι ματιών. Από την στιγμή που επιλέγονται τα μάτια, ο αλγόριθμος υπολογίζει την περιοχή του προσώπου σαν ένα ορθογώνιο. Οι τέσσερις κορυφές αυτού του ορθογωνίου υπολογίζονται μέσα από μια σειρά συναρτήσεων. Οι περιοχές των προσώπων επιβεβαιώνονται μέσω της χρήσης ενός back-propagation νευρωνικού δικτύου. Τέλος η εφαρμογή μιας συνάρτησης καθορίζει την

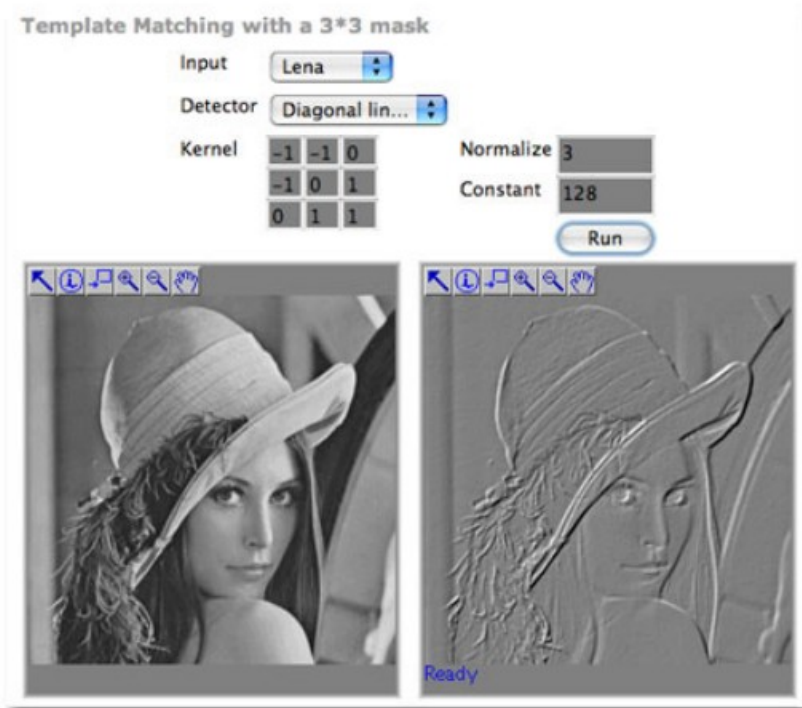
τελική απόφαση. Το ποσοστό επιτυχίας είναι 94%, ακόμα και σε φωτογραφίες με πολλά πρόσωπα. η μέθοδος αυτοί κρίνονται αποτελεσματική κατά κύριο λόγο για απλές εισόδους. Όμως τι μπορεί να συμβεί όταν ένα πρόσωπο στην φωτογραφία φοράει γυαλιά;

Υπάρχουν άλλα χαρακτηριστικά που μπορεί να αντιμετωπίσουν παρόμοιο πρόβλημα. Για παράδειγμα, υπάρχουν αλγόριθμοι που ανιχνεύουν το πρόσωπο, όπως η υφή ή το χρώμα του ανθρώπινου δέρματος. Είναι πολύ σημαντικό να επιλέξουμε το καλύτερο μοντέλο χρωμάτων για την ανίχνευση προσώπων. Σε πρόσφατες έρευνες χρησιμοποιήθηκαν περισσότερα από ένα μοντέλα χρωμάτων. Για παράδειγμα, το RGB και το HSV μπορούν να χρησιμοποιηθούν μαζί με επιτυχία.

## Template Matching Methods

Μέθοδοι που ταιριάζουν πρότυπα προσπαθούν να ορίσουν ένα πρόσωπο ως μια συνάρτηση. Προσπαθούν να βρουν ένα τυποποιημένο πρότυπο όλων των προσώπων. Διαφορετικά χαρακτηριστικά μπορούν να οριστούν αυτοτελώς. Για παράδειγμα, ένα πρόσωπο μπορεί να διαιρεθεί σε μάτια, περίγραμμα προσώπου, μύτη και στόμα. Επίσης, ένα μοντέλο προσώπου μπορεί να οικοδομηθεί από τα άκρα. Αλλά σε αυτές τις μεθόδους υπάρχει ο περιορισμός της μετωπικής και καθαρής λήψης των προσώπων. Ένα πρόσωπο μπορεί επίσης να εκπροσωπείται ως σιλουέτα. Άλλα πρότυπα χρησιμοποιούν τη σχέση μεταξύ των περιφερειών προσώπου ως προς την φωτεινότητα και το σκοτάδι. Η προσέγγιση αυτή είναι απλή στην εφαρμογή της, αλλά είναι ανεπαρκής για την ορθή ανίχνευση προσώπου. Δεν μπορεί να επιτύχει καλά αποτελέσματα με μεταβλητές πόζες, μεγέθη και σχήματα που μπορεί να έχει ένα πρόσωπο.

Εικόνα 11: Λογισμικό με χρήση Template Matching μεθόδου



## Appearance-based Methods

Τα πρότυπα στις μεθόδους που έχουν ως βάση την εμφάνιση κάνουν την εκμάθησή τους μέσω των παραδειγμάτων στις εικόνες. Σε γενικές γραμμές, αυτές οι μέθοδοι στηρίζονται σε τεχνικές από στατιστική ανάλυση και μηχανική μάθηση για να βρουν τα σχετικά χαρακτηριστικά σε μια εικόνα προσώπου. Μερικές από τις μεθόδους αυτές λειτουργούν σε πιθανοτικά δίκτυα (probabilistic networks). Μια εικόνα ή ένα

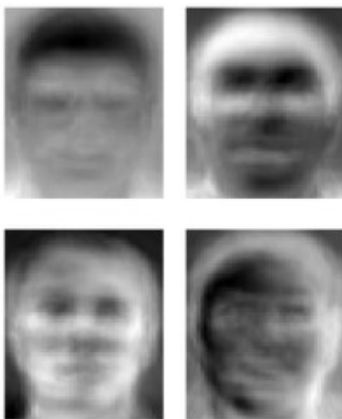
χαρακτηριστικό είναι μια τυχαία μεταβλητή με κάποια πιθανότητα να ανήκει σε ένα πρόσωπο ή όχι. Μια άλλη προσέγγιση είναι να ορίσουμε μια διακρίνουσα συνάρτηση μεταξύ προσώπου και μη-προσώπου. Αυτές οι μέθοδοι είναι επίσης που χρησιμοποιούνται στην εξαγωγή χαρακτηριστικών για την αναγνώριση προσώπου. Παρόλα αυτά, αυτές είναι οι πιο σημαντικές μέθοδοι ή εργαλεία:



Εικόνα 12: Χρήση Appearance-

based Μεθόδων

- Eigenface – based. Τα eigenfaces[7] είναι ένα σύνολο ιδιοδιανυσμάτων που χρησιμοποιείται στο πρόβλημα της όρασης υπολογιστή για την αναγνώριση προσώπου. Η προσέγγιση σύμφωνα με eigenfaces για αναγνώριση αναπτύχθηκε από τους Sirovich και Kirby (1987) και χρησιμοποιήθηκε από τους Matthew Turk και Alex Pentland για την ταξινόμηση των προσώπων. Αυτά τα ιδιοδιανύσματα που προκύπτουν από τη μήτρα συν διακύμανσης της κατανομής πιθανοτήτων των μεγάλων διαστάσεων στον διανυσματικό χώρο των πιθανών όψεων της ανθρώπινης ύπαρξης.

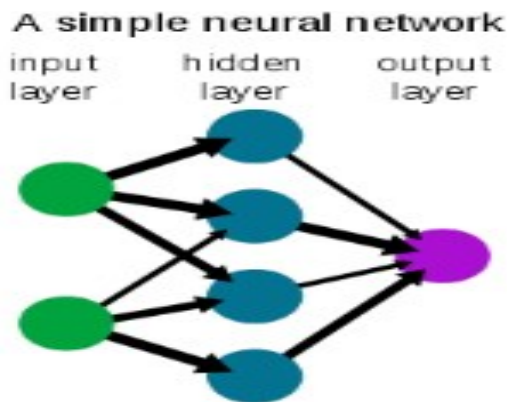


Εικόνα 13: Μερικά Eigenfaces

- Distribution – based Αυτά τα συστήματα προτάθηκαν για πρώτη φορά για το αντικείμενο και ανίχνευση μοτίβο από τον Sung. Η ιδέα είναι η συλλογή ενός αρκετά μεγάλου αριθμού απόψεων του δείγματος για την κατηγοριοποίηση του πρότυπου που θέλουμε να ανιχνευθεί, καλύπτοντας όλες τις πιθανές μεταβολές της εικόνας που μπορεί να υπάρξουν. Θα πρέπει να αντιπροσωπεύει την τάξη μοτίβο ως διανομή από όλες τις επιτρεπόμενες εμφανίσεις της εικόνας. Το σύστημα ταιριάζει με την υποψήφια εικόνα κατά τη διανομή με βάση το κανονικό μοντέλο πρόσωπο. Τέλος, υπάρχει ένας εκπαιδευμένος ταξινομητής που εντοπίζει σωστές εμφανίσεις της τάξης μοτίβο στόχου από μοτίβα εικόνα φόντου, βασίζόμενος σε ένα σύνολο μετρήσεων απόστασης μεταξύ του τρόπου διεξαγωγής των εισροών και η κατανομή με βάση της εκπροσώπησης τάξη στον επιλεγμένο χώρο των χαρακτηριστικών. Αλγόριθμοι όπως ο PCA ή Fisher's Discriminant μπορούν να χρησιμοποιηθούν για τον καθορισμό των υποχώρων του προσώπου που

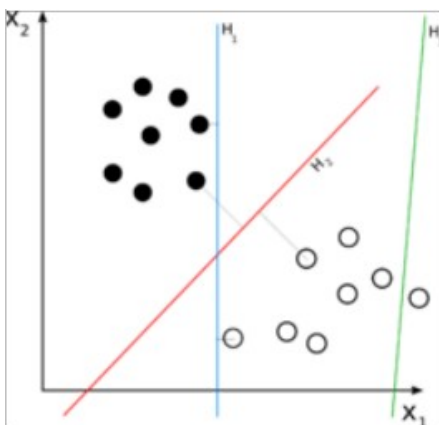
εκπροσωπεί τα πρότυπα.

- Neural Networks. Ο όρος νευρωνικό δίκτυο χρησιμοποιούταν παραδοσιακά για να αναφερθεί σε ένα δίκτυο ή κύκλωμα των βιολογικών νευρών. Η σύγχρονη χρήση του όρου αναφέρεται στα τεχνητά νευρωνικά δίκτυα τα οποία αποτελούνται από νευρώνες ή κόμβους.



Εικόνα 14: Προβολή ενός *feedforward artificial neural network*

- Support Vector Machines. Μια μηχανή διανυσματικής υποστήριξης είναι μια έννοια στην επιστήμη των υπολογιστών για ένα σύνολο που αφορούν την εποπτεία των μεθόδων μάθησης, που αναλύει τα δεδομένα και αναγνωρίζει τα πρότυπα, που χρησιμοποιούνται για την ταξινόμηση και την ανάλυση παλινδρόμησης. Το πρότυπο SVM παίρνει ένα σύνολο δεδομένων εισόδου και προβλέπει, για κάθε δεδομένο εισόδου το οποίο από τις δύο δυνατές τάξεις της εισόδου είναι μέλος της, γεγονός που καθιστά το SVM μη πιθανοτικό δυαδικό ταξινομητή.



Εικόνα 15: SVM λειτουργία

*H3 (green) doesn't separate the two classes. H1 (blue) does, with a small margin and H2 (red) with the maximum*

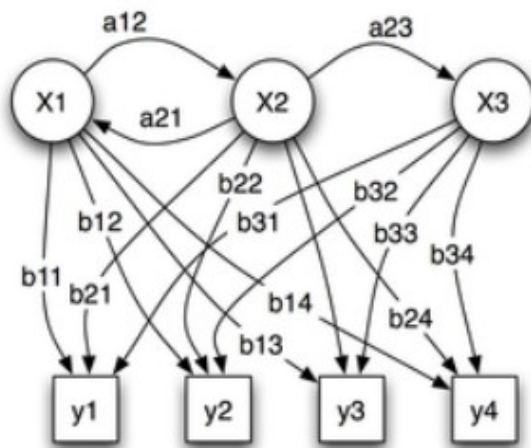
- Sparse Network of Winnows άρχισαν να χρησιμοποιείται για την ανίχνευση από τον Yang et al. Ορίζεται ένα αραιό δίκτυο από δύο μονάδες της γραμμικής ή κόμβους-στόχους, ένας εκπρόσωπος των μοτίβων προσώπου και το άλλο για τα πρότυπα nonface. Το SNoWs μαθαίνει σταδιακά τον χώρο των χαρακτηριστικών. Νέα σήμανση περίπτωσης χρησίμευσε ως θετικό παράδειγμα για ένα στόχο και ως αρνητικό παράδειγμα για τους υπόλοιπους στόχους.



- **Naive Bayes Classifiers** Με απλά λόγια, ένας αφελής ταξινομητής Bayes υποθέτει ότι η παρουσία (ή απουσία) ένας ιδιαίτερου χαρακτηριστικού γνωρίσματος μιας κατηγορίας δεν έχει σχέση με την παρουσία (ή απουσία) οποιοδήποτε άλλου χαρακτηριστικού. Για παράδειγμα, ένα φρούτο μπορεί να θεωρηθεί ότι είναι ένα μήλο και αν είναι κόκκινο, στρογγυλό, και περίπου 4" σε διάμετρο. Ακόμα κι αν αυτά τα χαρακτηριστικά εξαρτώνται μεταξύ τους ή από την ύπαρξη των άλλων χαρακτηριστικών, ένας αφελής ταξινομητής Bayes θεωρεί όλες αυτές οι ιδιότητες συμβάλλουν ανεξάρτητα στην πιθανότητα ότι αυτά τα φρούτα είναι μια μήλα. Ανάλογα με την ακριβή φύση του μοντέλου πιθανότητας, οι αφελής ταξινομητές Bayes μπορούν να εκπαιδευτούν πολύ αποτελεσματικά σε ελεγχόμενο περιβάλλον μάθησης. Σε πολλές πρακτικές εφαρμογές, η εκτίμηση των παραμέτρων για τον αφελή μοντέλο Bayes χρησιμοποιεί τη μέθοδο της μέγιστης πιθανότητας. Με άλλα λόγια, μπορεί κανείς να συνεργαστεί με τον αφελή μοντέλο Bayes χωρίς πίστη στη Bayesian probability ή τη χρήση οποιωνδήποτε Bayesian μεθόδων.

- **Hidden Markov Model** Σε ένα κανονικό μοντέλο Markov, state έχει άμεση ορατότητα προς τον παρατηρητή, και ως εκ τούτου οι πιθανότητες μετάβασης μεταξύ κάθε state είναι οι μοναδικές παράμετροι. Κάθε state έχει μια κατανομή πιθανότητας για την πιθανή μάρκες εξόδου. Ως εκ τούτου η ακολουθία των μαρκών που προκύπτουν από τα HMM δίνει κάποιες πληροφορίες για τη σειρά των state. Σημειώστε ότι «κρύβεται» το αντικείμενο αναφέρεται στην ακολουθία των states μέσω της οποίας το μοντέλο περνά, όχι με τις παραμέτρους του μοντέλου. Ακόμα και αν είναι οι παράμετροι μοντέλου είναι γνωστές, το μοντέλο εξακολουθεί να είναι «κρυμμένο».

Τα κρυφά μοντέλα Markov είναι ιδιαίτερα γνωστά για την εφαρμογή τους σε χρονική αναγνώριση προτύπων, όπως η ομιλία, χειρόγραφου, την αναγνώριση χειρονομίας και της βιοπληροφορικής.



Εικόνα 16: Probabilistic parameters of a hidden Markov model

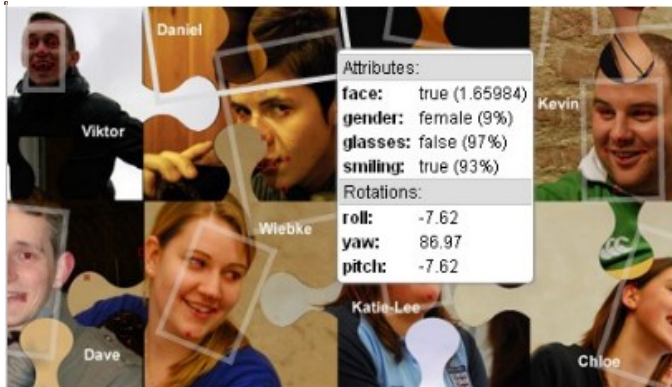
(example)  $x$  - states  $y$  - possible observations  
 $a$  - state transition probabilities  
 $b$  - output probabilities

- **Information-theoretical Approach Markov Random Fields (MRF)**, μπορεί να χρησιμοποιηθεί ως μοντέλο ευρύτερου πλαισίου ενός σχεδίου προσώπου και συσχετίζον χαρακτηριστικά. Η διαδικασία Markov μεγιστοποιεί τη διάκριση μεταξύ τάξεις (μια εικόνα έχει ένα πρόσωπο ή όχι) τη χρήση του Kullback-Leibler απόκλιση. Ως εκ τούτου, αυτή η μέθοδος μπορεί να εφαρμοστεί στην ανίχνευση προσώπου.

- **Inductive Learning** Αυτή η προσέγγιση έχει χρησιμοποιηθεί για την ανίχνευση προσώπων. Έχουν Αλγόριθμοι όπως C4.5 Quinlan ή του Μίτσελ FIND-S, έχουν χρησιμοποιηθεί για το σκοπό αυτό.

### 3.1.3 Εφαρμογές

Η ανίχνευση προσώπου χρησιμοποιείται στη βιομετρική, συχνά ως μέρος (ή μαζί με) ενός συστήματος αναγνώρισης του προσώπου. Χρησιμοποιείται επίσης στην τηλεοπτική επιτήρηση, την ανθρώπου-υπολογιστή διεπαφή και τη διαχείριση βάσεων δεδομένων εικόνας. Μερικές πρόσφατες ψηφιακές κάμερες χρησιμοποιούν την ανίχνευση προσώπου για το autofocus. Η ανίχνευση προσώπου είναι επίσης χρήσιμη για τις περιοχές ενδιαφέροντος για ένα φωτογραφικό slide show.



Εικόνα 17: Εφαρμογές ανίχνευσης προσώπου

Η ανίχνευση προσώπου κερδίζει το συμφέρον των εμπόρων. Μια webcam μπορεί να ενσωματωθεί σε μια τηλεόραση και να ανιχνεύσει οποιοδήποτε πρόσωπο που περπατά κοντά. Το σύστημα υπολογίζει έπειτα τη φυλή, το φύλο και την ηλικία του προσώπου. Μόλις συλλεχθούν οι πληροφορίες, μια σειρά διαφημίσεων μπορεί να παιχτεί που είναι συγκεκριμένη προς την συγκεκριμένη φυλή, φύλο και ηλικία.

Η ανίχνευση προσώπου ερευνάται επίσης στον τομέα της διατήρησης της ενέργειας. Οι τηλεοράσεις και οι υπολογιστές μπορούν να σώσουν την ενέργεια με τη μείωση της φωτεινότητας. Οι άνθρωποι τείνουν να προσέξουν τη TV κάνοντας άλλους στόχους και μην 100% στην οθόνη. Η φωτεινότητα της TV μένει στο ίδιο επίπεδο εκτός αν ο χρήστης το χαμηλώνει με το χέρι. Το σύστημα μπορεί να αναγνωρίσει την κατεύθυνση προσώπου του χρήστη της TV. Όταν ο χρήστης δεν εξετάζει την οθόνη, τη φωτεινότητα της θα μπορούσε να χαμηλώνει. Όταν το πρόσωπο επιστρέφει στην οθόνη, η φωτεινότητα αυξάνεται.

### 3.2 Εξαγωγή Χαρακτηριστικών

Υπάρχουν πολλοί αλγόριθμοι εξαγωγής χαρακτηριστικών. Τα περισσότερα από αυτά χρησιμοποιούνται σε άλλους τομείς εκτός από αυτόν της αναγνώρισης προσώπου. Ερευνητές στην αναγνώριση προσώπου έχουν χρησιμοποιήσει, τροποποιήσει και προσαρμόσει πολλούς αλγόριθμους και μεθόδους για τον σκοπό τους. Παρακάτω παρουσιάζεται ένας πίνακας με μια λίστα μερικών αλγορίθμων εξαγωγής χαρακτηριστικών.

Method	Notes
Principal Component Analysis (PCA)	Eigenvector-based, linear map
Kernel PCA	Eigenvector-based, non-linear map, uses kernel methods
Weighted PCA	PCA using weighted coefficients
Linear Discriminant Analysis (LDA)	Eigenvector-based, supervised linear map
Kernel LDA	LDA-based, uses kernel methods
Semi-supervised Discriminant Analysis (SDA)	Semi-supervised adaptation of LDA
Independent Component Analysis (ICA)	Linear map, separates non-Gaussian distributed features
Neural Network based methods	Diverse neural networks using PCA, etc.
Multidimensional Scaling (MDS)	Nonlinear map, sample size limited, noise sensitive.
Self-organizing map (SOM)	Nonlinear, based on a grid of neurons in the feature space
Active Shape Models (ASM)	Statistical method, searches boundaries
Active Appearance Models (AAM)	Evolution of ASM, uses shape and texture
Gabor wavelet transforms	Biologically motivated, linear filter
Discrete Cosine Transform (DCT)	Linear function, Fourier-related transform, usually used 2D-DCT
MMSD, SMSD	Methods using maximum scatter difference criterion.

Πίνακας 1: αλγορίθμων εξαγωγής χαρακτηριστικών

### 3.2.1 Μέθοδοι Επιλογής Χαρακτηριστικών

Οι αλγόριθμοι επιλογής χαρακτηριστικών στοχεύουν στην επιλογή ενός υποσυνόλου των χαρακτηριστικών που προκαλούν το μικρότερο σφάλμα κατά την ταξινόμηση. Η σημασία αυτού του σφάλματος είναι αυτή που θα καθορίσει την μέθοδο ταξινόμησης που θα χρησιμοποιηθεί. Η πιο απλή προσέγγιση σε αυτό το πρόβλημα θα ήταν να εξετάσουμε κάθε δυνατό υποσύνολο και να επιλέξουμε αυτό που ικανοποιεί το κριτήριο της λειτουργίας. Ωστόσο, αυτό μπορεί να γίνει απρόσιτη διεργασία από την άποψη του υπολογιστικού χρόνου που χρειάζεται. Μερικές αποτελεσματικές προσεγγίσεις για αυτό το πρόβλημα είναι βασισμένες σε αλγόριθμους branch and bound.



Method	Definition	Comments
Exhaustive search	Evaluate all possible subsets of features.	Optimal, but too complex.
Branch and bound	Use branch and bound algorithm.	Can be optimal. Complexity of max $O(2^n)$ .
Best individual features	Evaluate and select features individually.	Not very effective. Simple algorithm.
Sequential Forward Selection (SFS)	Evaluate growing feature sets (starts with best feature).	Retained features can't be discarded. Faster than SBS.
Sequential Backward Selection (SBS)	Evaluate shrinking feature sets (starts with all the features).	Deleted features can't be reevaluated.
"Plus $l$ -take away $r$ " selection	First do SFS then SBS.	Must choose $l$ and $r$ values.
Sequential Forward Floating Search (SFFS) and Sequential Backward Floating Search (SBFS)	Like "Plus $l$ -take away $r$ ", but $l$ and $r$ values automatic pick and dynamic update.	Close to optimal. Affordable computational cost.

Πίνακας 2: Μέθοδοι Επιλογής Χαρακτηριστικών

Πρόσφατα περισσότεροι αλγόριθμοι επιλογής χαρακτηριστικών έχουν προταθεί. Η επιλογή χαρακτηριστικών είναι ένα NP-hard πρόβλημα και για αυτόν τον λόγο οι ερευνητές καταβάλουν προσπάθειες για την δημιουργία ενός ικανοποιητικού αλγόριθμου. Η ιδέα είναι να δημιουργηθεί ένας αλγόριθμος που να μπορεί να επιλέξει το πιο ικανοποιητικό υποσύνολο χαρακτηριστικών, μειώνοντας της διαστάσεις και την πολυπλοκότητα. Μερικές προσεγγίσεις χρησιμοποιούν συντελεστές ομοιότητας ή ικανοποιητικό ποσοστό ως κριτήριο και αλγορίθμους κβαντικής γενετικής.



Εικόνα 18: Μέθοδοι Επιλογής Χαρακτηριστικών

### 3.3 Facial Expression

Οι εκφράσεις του προσώπου είναι μια ορατή εκδήλωση της συναισθηματικής κατάστασης, της πρόθεσης, της προσωπικότητας και της ψυχοπαθολογίας του κάθε ατόμου. Διαδραματίζει επικοινωνιακό

ρόλο στις διαπροσωπικές σχέσεις ανάμεσα στους ανθρώπους. Οι εκφράσεις προσώπου και οι χειρονομίες μπορούν να μεταφέρουν με μη λεκτικό τρόπο συναισθήματα επικοινωνίας από άτομο σε άτομο. Αυτά τα συναισθήματα επίσης συμπληρώνουν την ομιλία του ανθρώπου βοηθώντας τον ακροατή να αποσπάσει την προβλεπόμενη έννοια του ομιλητή.

### 3.3.1 Facial Expression Classification

Η έρευνα στην ανάλυση εκφράσεων έχει έλθει ξανά στο προσκήνιο τα τελευταία χρόνια. Ο Shlosberg σε μια από τις πρώτες προσεγγίσεις χρησιμοποίησε τρεις άξονες για την περιγραφή των συναισθημάτων: αποδοχής - απόρριψης A-R (attention - rejection), ευαρέσκειας - δυσαρέσκειας P-U (pleasantness/unpleasantness), και βαθμού ενεργοποίησης (level of activation). Για παράδειγμα συναισθήματα όπως η περιφρόνηση και η απέχθεια χαρακτηρίζονται από υψηλή τιμή απόρριψης ενώ συναισθήματα όπως η οργή χαρακτηρίζονται από υψηλή τιμή απαρέσκειας. Οι Ekman και Friesen θεωρούν έξι πρωτεύουσες εκφράσεις οι οποίες συνδέονται με αντίστοιχα συναισθήματα και οι οποίες μπορούν να αναγνωριστούν από τις μορφοποιήσεις του προσώπου. Οι εκφράσεις αυτές είναι χαρά, λύπη, έκπληξη, απέχθεια, οργή και φόβος. Πέρα από τις πρωτεύουσες εκφράσεις κατέγραψαν πολλές άλλες οι οποίες όμως δεν είναι τόσο εύκολα ανιχνεύσιμες. Στο πεδίο των γραφικών για υπολογιστές και ιδιαίτερα στο animation έχουν εμφανιστεί πολλές εργασίες οι οποίες μοντελοποιούν τις εκφράσεις με βάση την κίνηση των μυών του προσώπου. Πρωτοποριακή στον τομέα αυτό θεωρείται η εργασία των Terzopoulos και Waters οι οποίοι μεταξύ άλλων πρότειναν και ένα μοντέλο για τους ιστούς και το δέρμα του προσώπου. Πέρα από την μοντελοποίηση χρησιμοποίησαν την αντίστροφη διαδικασία για την ανάλυση των εκφράσεων (analysis by synthesis). Στην ανάλυση εκφράσεων τα χαρακτηριστικά του προσώπου είναι ιδιαίτερα σημαντικά και η αποτελεσματική ανίχνευση τους καθίσταται επιτακτική. Για το σκοπό αυτό οι εικόνες των προσώπων πρέπει να αρκετά υψηλής ανάλυσης και τα χαρακτηριστικά ευδιάκριτα.

### 3.3.2 Πλαίσιο περιγραφής της απεικόνισης των συναισθημάτων στα πρόσωπα

Η παραδοσιακή τακτική αναγνώρισης των συναισθημάτων μέσα από την έκφρασή τους στο πρόσωπο είναι όμοια με αυτήν της αναγνώρισης τους μέσω της ομιλίας: Τα δεδομένα εισόδου είναι στατικά και επιδεικνύουν απλά την κορυφή του συναισθήματος – ‘apex’. Στην περίπτωση της οπτικής πληροφορίας η απεικόνιση είναι απλά μια φωτογραφία στην οποία το υποκείμενο εικονίζεται στο υψηλότερο εκφραστικό επίπεδο. Οι σύγχρονες τεχνικές πάντως τείνουν να εμπλέκουν πληροφορία δυναμικής υφής για την κατανόηση των συναισθημάτων μέσα από τις εκφράσεις του προσώπου. Η αλήθεια είναι ότι η αναγνώριση συναισθημάτων από την οπτική πληροφορία έχει ένα ισχυρότερο υπόβαθρο, προερχόμενο από μελέτες στο τομέα της Νευροφυσιολογίας, από ότι η αναγνώριση τους από ηχητικά δεδομένα. Αυτό οφείλεται σε μεγάλο βαθμό στο γεγονός ότι πολλές μελέτες σχετικές με την αναγνώριση προσώπων διερεύνησαν και τον τρόπο με τον οποίο οι άνθρωποι ερμηνεύουν τα συναισθήματα των συνανθρώπων τους με βάση τις αντιλαμβανόμενες εκφράσεις. Από την άλλη πλευρά η συντριπτική πλειοψηφία των μελετών σχετικά με την κατανόηση συναισθημάτων μέσω των εκφράσεων επικεντρώθηκαν απλά στην ταξινόμηση τους σε έξι πρωτεύουσες κατηγορίες. Η τάση αυτή θα πρέπει να αποδοθεί στον Ekman και τους συνεργάτες του οι οποίοι αποφάνθηκαν ότι πρακτικά μόνο τα έξι πρωτεύοντα συναισθήματα –ή καλύτερα κατηγορίες συναισθημάτων- μπορούν να απεικονιστούν μέσα από συγκεκριμένες μορφοποιήσεις του προσώπου. Το παράδοξο είναι ότι οι Ekman και Friesen είναι από τους ελάχιστους ερευνητές οι οποίοι εξέτασαν τον τρόπο μορφοποίησης και σύνθεσης και μη πρωτεύουσών εκφράσεων –με αλλαγή για παράδειγμα του αριθμού των περιοχών του προσώπου που εμπλέκονται στις εκφράσεις, της χρονικής εξέλιξης τους και της έντασης της δράσης των μυών.

Στην ίδια μελέτη εξετάστηκε και πώς η ανάμιξη συναισθημάτων μπορεί να οδηγήσει σε ουδετεροποίηση των εκφράσεων και φυσικά απώλεια της αντίληψής τους. Για παράδειγμα η ανάμιξη της οργής με τη λύπη μπορεί είτε να οδηγήσει σε μια μη αναγνωρίσιμη έκφραση του προσώπου είτε στην επικάλυψη της λιγότερη ισχυρής εκ των δύο από την άλλη. Σχετικά πρόσφατα ξεκίνησε κάποια πρώιμη έρευνα σχετικά με τη σύνθεση ενδιάμεσων εκφράσεων με προβολή των παραμέτρων μορφοποίησης των

εκφράσεων –ουσιαστικά δράση μυών- στο χώρο ενεργοποίησης–επαλήθευσης. Η επέκταση πάντως των τεχνικών αυτών και στην ανάλυση μη πρωτεύουσών εκφράσεων συναντά σημαντικές δυσκολίες με βασικότερη τη δυσκολία αντίληψης της δράσης των μυών χωρίς τη χρήση ηλεκτροδυναμικών διατάξεων. Βέβαια παρά την επικέντρωση στην ανάλυση των εκφράσεων που σχετίζονται με τα πρωτεύοντα συναισθήματα, η οποία δείχνει πρακτικά προσέγγιση της αναγνώρισης συναισθημάτων μέσα από ένα στενό πρίσμα, ουδείς μπορεί να αμφισβητήσει την αξία της υλοποίησης ενός τέτοιου συστήματος σε περιβάλλοντα επικοινωνίας ανθρώπου μηχανής. Στη πράξη ακόμα και ένας υπολογιστής που αναγνωρίζει τις πρωτεύουσες εκφράσεις είναι σημαντικά πιο έξυπνος από αυτόν που δεν το καταφέρνει.

### 3.3.3 Απεικονίσεις που σχετίζονται με την συναισθηματική έκφραση

Τα χαρακτηριστικά του προσώπου μπορούν να θεωρηθούν είτε στατικά –όπως για παράδειγμα το χρώμα του δέρματος- είτε αργά μεταβαλλόμενα –όπως υφή η οποία μεταβάλλεται με την ανάπτυξη ρυτίδων- είτε κινούμενα –όπως οι βλεφαρίδες, τα φρύδια κοκ. Η ανίχνευση της θέσης των χαρακτηριστικών αυτών από χρονικά σταθερές απεικονίσεις –φωτογραφίες- είναι ο στόχος των στατικών προσεγγίσεων της ανάλυσης εκφράσεων. Παρόλα αυτά υπάρχει ισχυρή ένδειξη ότι η αναγνώριση εκφράσεων από τον άνθρωπο στηρίζεται περισσότερο σε πληροφορία δυναμικής υφής παρά σε στατικές απεικονίσεις. Ο Bassili πραγματοποίησε μια μελέτη στην οποία μια ομάδα ανθρώπων κλήθηκε να αναγνωρίσει εκφράσεις σε ακολουθίες βίντεο στις οποίες υπήρχαν φωτεινές κουκκίδες μόνο στις θέσεις των χαρακτηριστικών του προσώπου και οι υπόλοιπες περιοχές ήταν σκοτεινές. Από την συγκεκριμένη μελέτη προέκυψε ότι αναγνώριση πάνω από το επίπεδο τυχαιότητας ήταν εφικτή για όλες τις εκφράσεις, όταν χρησιμοποιούνταν ακολουθίες βίντεο ως πηγή πληροφορίας, ενώ με βάση τις στατικές εικόνες μόνο οι εκφράσεις “χαρά” και “λύπη” αναγνωρίστηκαν σε ποσοστό υψηλότερο από το επίπεδο τυχαιότητας. Δυστυχώς το συμπέρασμα του Bassili έχει ουσιαστική αξία μόνο όσον αφορά την αντίληψη των εκφράσεων από τον άνθρωπο γιατί και η αναγνώριση εκφράσεων από τον υπολογιστή με βάση ακολουθίες βίντεο αντιμετωπίζει πολλά προβλήματα. Η ειδοποιός διαφορά μεταξύ ανθρώπου και υπολογιστή είναι η ακρίβεια εντοπισμού των προσώπων και των χαρακτηριστικών τους στο χώρο. Το ανθρώπινο οπτικό σύστημα είναι εξαιρετικά αποτελεσματικό στον τομέα αυτό. Αντίθετα στους υπολογιστές τα σφάλματα εντοπισμού του προσώπου και των χαρακτηριστικών του, λειτουργούν προσθετικά και σε πολλές περιπτώσεις καλύπτουν την ουσιαστική πληροφορία κίνησης που διατίθεται από τις ακολουθίες.

Το τελικό συμπέρασμα είναι ότι οι στατικές προσεγγίσεις οι οποίες είναι λιγότερο επιρρεπείς στον εντοπισμό των χαρακτηριστικών του προσώπου διατηρούν την αξία τους όσον αφορά την αναγνώριση των εκφράσεων. Από την άλλη πλευρά τα στάδια προεπεξεργασίας τα οποία αφορούν στον εντοπισμό του προσώπου, των βασικών χαρακτηριστικών του – όπως τα μάτια, μύτη, το στόμα κοκ- και σημείων στη περιοχή των χαρακτηριστικών αυτών είναι εξαιρετικά σημαντικά σε όλες τις περιπτώσεις. Οι δυναμικές προσεγγίσεις αντιμετωπίζουν επιπλέον και την πρόκληση της παρακολούθησης της κίνησης των χαρακτηριστικών με τεχνικές εκτίμησης κίνησης και μοντελοποίησης των μυών και των δράσεων τους. Τη βάση για τα περισσότερα από τα συστήματα αναγνώρισης εκφράσεων αποτέλεσε η εργασία των Ekman και Friesen, οι οποίοι δημιούργησαν ένα σύστημα κωδικοποίησης το οποίο περιγράφει όλες τις οπτικά διαχωρίσιμες κινήσεις του προσώπου το οποίο ονόμασαν FACS - Facial Action Coding System. Το FACS είναι ένα σύστημα ανατομικής περιγραφής το οποίο στηρίζεται στον ορισμό των «μονάδων δράσης» –AU action units. Κάθε AU αντιστοιχεί στην ταυτόχρονη δράση μιας ομάδας μυών οι οποίοι διαμορφώνουν 40 μια συγκεκριμένη δράση στο πρόσωπο. Δεδομένου ότι αρκετοί μύες συμμετέχουν σε περισσότερες από μία AU δεν υπάρχει σαφής αντιστοιχία μυών και AU. Ένα σύνολο από 46 AU καλύπτει πλήρως τον έλεγχο των εκφράσεων ενώ άλλες 12 είναι υπεύθυνες για την θέση και κίνηση της ίριδας των ματιών. Το μοντέλο FACS χρησιμοποιήθηκε αποδοτικά για την σύνθεση εκφράσεων ενώ η χρήση του για ανάλυση εκφράσεων εξακολουθεί να ερευνάται. Ο Ekman και οι συνεργάτες του δημιούργησαν επίσης και ένα λεξικό το EMFACS στο οποίο δηλώνονται οι AU οι οποίες περιγράφουν τις πρωτεύουσες εκφράσεις. Στη συνέχεια δημιούργησαν τη βάση FACSAID η οποία χρησιμοποιείται για τον υπολογισμό των συναισθηματικών εκφράσεων με βάση τις μετρήσεις των παραμέτρων του FACS. Το μοντέλο FACS ενέπνευσε και τη δημιουργία των παραμέτρων περιγραφής προσώπου και απόδοσης κίνησης προσώπου στο πλαίσιο του προτύπου ISO MPEG-4. Το σύνολο παραμέτρων για τον ορισμό προσώπου (Facial Definition Parameter set, FDP) και το σύνολο παραμέτρων απόδοσης κίνησης προσώπου (Facial Animation Parameter set, FAP)

έχουν σχεδιαστεί για να επιτρέπουν τον ορισμό του σχήματος και της υψής ενός προσώπου καθώς και την αναπαράσταση εκφράσεων, συναισθημάτων καθώς και προφοράς ομιλίας. Το σύνολο FDP αποτελείται από ένα τρισδιάστατο πλέγμα, ένα σύνολο από σημεία στο τρισδιάστατο χώρο και ένα σύνολο από εναλλακτικά χαρακτηριστικά όπως μαλλιά, γυαλιά, ρυτίδες κ.ο.κ., τα οποία επιτρέπουν τον ορισμό του ακριβούς σχήματος προσώπου καθώς και της υψής του στην φάση του setup. Εάν αυτά τα στοιχεία χρησιμοποιηθούν στην φάση της αρχικής ρύθμισης ενός συνθετικού μοντέλου, είναι πιθανόν να παράγουμε με ακόμα μεγαλύτερη ακρίβεια τις κινήσεις του συγκεκριμένου προσώπου. Το σύνολο παραμέτρων απόδοσης κίνησης προσώπου (FAPs) είναι βασισμένο στην μελέτη των ελάχιστων δράσεων του προσώπου και είναι στενά συνδεδεμένο με τις δράσεις των μυών.

Οι δράσεις, όπως για παράδειγμα η σύμπτυξη των φρυδιών και το άνοιγμα του στόματος, επιτρέπουν την αναπαράσταση των πιο φυσικών μορφοποιήσεων του προσώπου. Όλες οι παράμετροι που αφορούν την απόδοση κίνησης προσώπου εκφράζονται σύμφωνα με τις μονάδες παραμέτρων αναπαράστασης προσώπου (Facial Animation Parameters Units). Οι μονάδες αυτές έχουν σχεδιαστεί έτσι ώστε να επιτρέπουν την απεικόνιση των παραμέτρων απόδοσης κίνησης σε κάθε μοντέλο προσώπου με έναν συνεχή τρόπο παράγοντας λογικά αποτελέσματα σε ότι αφορά τις εκφράσεις και την ανθρώπινη προφορά. Ανταποκρίνονται στον τεμαχισμό των αποστάσεων ανάμεσα σε κάποια χαρακτηριστικά κλειδιά του ανθρώπινου προσώπου. Το μέγεθος του τεμαχισμού που εφαρμόζεται επιλέγεται ανάλογα με την επιθυμητή ακρίβεια.

### 3.3.4 Νευροφυσιολογία και συναισθήματα

Όπως και στις περιπτώσεις αναγνώρισης προσώπου, υποστηρίζεται ότι υπάρχουν συγκεκριμένες περιοχές του εγκεφάλου οι οποίες είναι υπεύθυνες για την αντίληψη εκφράσεων από εικόνες. Συγκεκριμένα η απώλεια του αμυγδάλου (amygdala) προκαλεί την αδυναμία αναγνώρισης των εκφράσεων «οργή» και «φόβος» σε φωτογραφίες προσώπων. Επίσης εμφανίζεται αυξημένη δραστηριότητα του αμυγδάλου κατά την παρακολούθηση παραλλαγών της έκφρασης «απέχθεια» ακόμα και αν το πρόσωπο στο οποίο απεικονίζεται η έκφραση είναι άγνωστο. Πάντως δεν υπάρχει σαφής ένδειξη για την περιοχή του εγκεφάλου που είναι υπεύθυνη για την κατανόηση ευχάριστων συναισθημάτων όπως η «χαρά», η «ακανοποίηση» κ.ο.κ.

Το πιο αξιοσημείωτο συμπέρασμα πάντως είναι ότι η επεξεργασία της πληροφορίας που σχετίζεται με την αναγνώριση συναισθημάτων πραγματοποιείται σε διαφορετικές περιοχές του εγκεφάλου από ότι η αναγνώριση προσώπων και επιπλέον είναι πιθανόν να εμπλέκει και διαφορετικά είδη διεγέρσεων.

## 3.4 ΤΕΧΝΙΚΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΕΚΦΡΑΣΕΩΝ

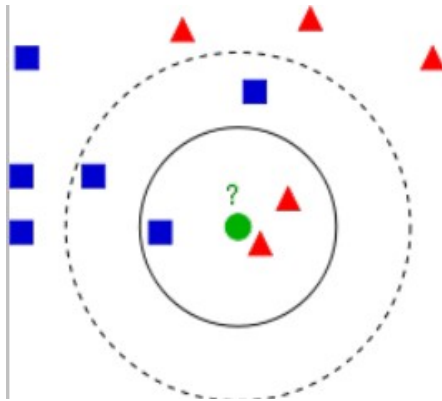
Στην παράγραφο αυτή παρουσιάζονται μερικές από τις τεχνικές οι οποίες έχουν εμφανιστεί στη βιβλιογραφία και πραγματεύονται τους τρόπους με τους οποίους ο υπολογιστής μπορεί να ανακτήσει πληροφορίες σχετικά με τη συναισθηματική κατάσταση κάποιου ατόμου μέσα από τις εκφράσεις του προσώπου του. Οι προσεγγίσεις του ανωτέρω ζητήματος διακρίνονται σε δύο κατηγορίες: (α) στατικές, στις οποίες η αναγνώριση της έκφρασης βασίζεται σε μια απλή φωτογραφία στην οποία εικονίζεται η έκφραση στην κορύφωση της, και (β) δυναμικές, στις οποίες η αναγνώριση πραγματοποιείται με τη χρήση μιας ακολουθίας από καρέ τα οποία επιδεικνύουν την χρονική εξέλιξη της έκφρασης. Η διάρκεια μιας ακολουθίας που απεικονίζει κάποια έκφραση κυμαίνεται από 0.5 έως 4 δευτερόλεπτα. Μια ενδιάμεση κατηγορία αποτελούν οι τεχνικές οι οποίες χρησιμοποιούν δύο καρέ, ένα με το πρόσωπο σε ουδέτερη κατάσταση και ένα με το πρόσωπο στη κορύφωση της έκφρασης. Παρόλο που τεχνικές αυτές πλησιάζουν την περισσότερο τη δυναμική προσέγγιση θα μπορούσε καταχρηστικά κάποιος να τις χαρακτηρίσει ημιστατικές.

## Παρακολούθηση της κίνησης του προσώπου

Όπως αναφέρθηκε στη προηγούμενη παράγραφο οι άνθρωποι οφείλουν ένα μεγάλο ποσοστό της ικανότητας τους να αναγνωρίζουν εκφράσεις στον εξαιρετικά ακριβή εντοπισμό του προσώπου και των χαρακτηριστικών του. Σε κάθε αλγόριθμο αναγνώρισης εκφράσεων ο εντοπισμός του προσώπου και η παρακολούθηση της κίνησης του είναι ζητήματα πολύ σημαντικά. Στο δεύτερο μέρος της διατριβής προτείνονται διάφορα σχήματα εντοπισμού προσώπων. Στη παράγραφο αυτή συνοψίζονται μερικές από τις τεχνικές που χρησιμοποιούνται για την παρακολούθηση της κίνησης του προσώπου. Στις εργασίες το πρόσωπο ανιχνεύεται με βάση την χρωματική κατανομή του δέρματος, σε κάθε καρέ ξεχωριστά, και στη συνέχεια υπολογίζεται η συνολική μετατόπιση του ανάμεσα στα δύο καρέ. Η προσέγγιση αυτή είναι αξιόπιστη όταν εφαρμόζεται σε ακολουθίες βίντεο με ελεγχόμενες συνθήκες φωτισμού και όταν το πρόσωπο κοιτάζει ευθεία στη κάμερα –εφαρμογές τηλεδιάσκεψης. Σε κάθε άλλη περίπτωση η ακρίβεια υπολογισμού της κίνησης είναι χαμηλή και επηρεάζεται από την κλίση και στροφή του προσώπου. Ως τελικό αποτέλεσμα το σφάλμα, όσον αφορά στην εκτίμηση της κίνησης του προσώπου, είναι μεγαλύτερο από την ίδια την κίνηση των χαρακτηριστικών που διαμορφώνουν την έκφραση. Οι τεχνικές παρακολούθησης της κίνησης του προσώπου με χρήση προτύπων, active contours, γράφων, wavelets και RBF συναρτήσεων, είναι σε γενικές γραμμές αναισθητες ως προς affine διακυμάνσεις και μετασχηματισμούς αλλά έχουν υψηλή υπολογιστική πολυπλοκότητα η οποία τις καθιστά ακατάλληλες σε εφαρμογές πραγματικού χρόνου. Εκτός από τις τεχνικές παρακολούθησης της κίνησης του προσώπου σημαντικές είναι και οι τεχνικές οι οποίες προσπαθούν να εκτιμήσουν την θέση του στο τρισδιάστατο χώρο -pose estimation- και να αντισταθμίσουν τις αλλαγές που εμφανίζονται στα χαρακτηριστικά του προσώπου εξαιτίας της οριοθέτησης του.

### 3.4.1 K Nearest Neighbors

Στην αναγνώριση προτύπων, ο k-NN [19] αλγόριθμος είναι μια μέθοδος για την ταξινόμηση των αντικειμένων με βάση το πιο κοντινό παράδειγμα την κατάρτιση στο χώρο των χαρακτηριστικών. Ο k-NN είναι ένας τύπος instance-based learning, ή lazy learning, όπου η λειτουργία είναι μόνο κατά προσέγγιση σε τοπικό επίπεδο και όλοι οι υπολογισμοί αναβάλλονται μέχρι την ταξινόμηση. Ο αλγόριθμος αυτός είναι από τους πιο απλούς όλων των αλγορίθμων μηχανικής μάθησης: ένα αντικείμενο που έχει χαρακτηριστεί από την πλειοψηφία των γειτόνων της, με το αντικείμενο που υπάγονται στην πιο κοινή κλάση μεταξύ των k κοντινότερους γείτονές της (k είναι ένας θετικός ακέραιος, συνήθως μικρό). Αν  $k = 1$ , τότε το αντικείμενο απλά υπάγεται στην κλάση του πλησιέστερου γείτονά της. Η ίδια μέθοδος μπορεί να χρησιμοποιηθεί για την οπισθοδρόμηση, με την απλή ανάθεση της αξίας του ακινήτου για το αντικείμενο που είναι ο μέσος όρος των τιμών των k πλησιέστερων γειτόνων της. Μπορεί να είναι χρήσιμο για τη στάθμιση των εισφορών των γειτόνων, έτσι ώστε οι πιο κοντά τους γείτονες να συμβάλουν περισσότερο στο μέσο όρο από τα πιο μακρινά. (Ένα κοινό σύστημα στάθμισης είναι να δοθεί σε κάθε γείτονα βάρους  $1 / d$ , όπου d είναι η απόσταση από το γείτονα. Το σύστημα αυτό είναι μια γενίκευση της γραμμικής παρεμβολής.) Οι γείτονες που λαμβάνεται από ένα σύνολο αντικειμένων για τα οποία η σωστή ταξινόμηση (ή, στην περίπτωση της παλινδρόμησης, η αξία του ακινήτου) είναι γνωστή. Αυτό μπορεί να θεωρηθεί ως η εκπαίδευση που για τον αλγόριθμο, αν και καμία ρητή βήμα εκπαίδευση είναι απαραίτητη. Ο k-nearest neighbors αλγόριθμος είναι ευαίσθητος στην τοπική δομή των δεδομένων. Είναι επίσης δυνατό να υπολογιστεί το όριο ίδια την απόφαση ρητώς, και μάλιστα με αποτελεσματικό τρόπο, έτσι ώστε η υπολογιστική πολυπλοκότητα είναι συνάρτηση της πολυπλοκότητας όριου.



Εικόνα 19: Example of  $k$ -NN classification

Τα παραδείγματα εκπαίδευσης είναι διανύσματα σε ένα πολυδιάστατο χώρο των χαρακτηριστικών, το καθένα με μια ετικέτα κλάσης. Η φάση της κατάρτισης του αλγορίθμου αποτελείται μόνο από την αποθήκευση του διανύσματα χαρακτηριστικών γνωρισμάτων και τις ετικέτες κατηγορίας των δειγμάτων εκπαίδευσης. Στη φάση της ταξινόμησης,  $k$  είναι ένας ορίζεται από το χρήστη σταθερή, και ένα χωρίς τίτλο διάνυσμα (ένα ερώτημα ή το σημείο δοκιμής) έχει χαρακτηριστεί από την ανάθεση της ετικέτας η οποία είναι πιο συχνή μεταξύ των δειγμάτων εκπαίδευσης  $k$  πλησιέστερο προς το σημείο αυτό το ερώτημα. Συνήθως Ευκλείδεια απόσταση χρησιμοποιείται ως η απόσταση μετρικών? Ωστόσο αυτό ισχύει μόνο για συνεχείς μεταβλητές. Σε περιπτώσεις όπως η ταξινόμηση κειμένου, μια άλλη μετρική όπως η επικάλυψη μετρικό (ή απόσταση Hamming) μπορεί να χρησιμοποιηθεί. Συχνά, η ακρίβεια ταξινόμησης του " $k$ -NN μπορεί να βελτιωθεί σημαντικά αν η μετρική απόσταση μαθαίνεται με εξειδικευμένους αλγόριθμους, όπως μεγάλο περιθώριο πλησιέστερο γείτονα ή Γειτονίας ανάλυση στοιχείων.

Ένα μειονέκτημα για τα βασικά "πλειοψηφία"ταξινόμηση είναι ότι οι τάξεις με την πιο συχνή παραδείγματα τείνουν να κυριαρχήσουν την πρόβλεψη του νέου φορέα, δεδομένου ότι τείνουν να καταλήξουμε σε γείτονες  $k$  πλησιέστερο όταν οι γείτονες υπολογίζονται λόγω του μεγάλου αριθμού τους. [παραπομπή που απαιτείται]Ένας τρόπος για να ξεπεραστεί αυτό το πρόβλημα είναι με το βάρος της ταξινόμησης, λαμβάνοντας υπόψη την απόσταση από το σημείο δοκιμής σε κάθε μία από  $k$  κοντινότερους γείτονες της. KNN είναι μια ειδική περίπτωση μιας μεταβλητής εύρους ζώνης, η πυκνότητα του πυρήνα "μπαλόνι"εκτιμητή με ένα ενιαίο πυρήνα.

### 3.4.2 Linear discriminant analysis

Γραμμική ανάλυση διακρίνουσας (LDA [17]) και η σχετική γραμμική διακρίνουσα του Fisher είναι μέθοδοι που χρησιμοποιούνται στις στατιστικές, αναγνώρισης προτύπων και μηχανικής μάθησης για να βρείτε ένα γραμμικό συνδυασμό των χαρακτηριστικών που χαρακτηρίζουν ή ξεχωριστά δύο ή περισσότερες κατηγορίες αντικειμένων ή γεγονότων. Το αποτέλεσμα του συνδυασμού μπορεί να χρησιμοποιηθεί ως ένα γραμμικό ταξινομητή, ή, συχνότερα, για τη μείωση των διαστάσεων, πριν αργότερα ταξινόμηση. LDA είναι στενά συνδεδεμένη με ANOVA (analysis of variance) και της ανάλυσης παλινδρόμησης, η οποία επίσης επιχειρούν να εκφράσουν μία εξαρτημένη μεταβλητή ως γραμμικός συνδυασμός των άλλων χαρακτηριστικών ή μετρήσεων. Στις άλλες δύο μεθόδους, ωστόσο, η εξαρτημένη μεταβλητή είναι μια αριθμητική ποσότητα, ενώ για LDA είναι μια κατηγορηματική μεταβλητής (δηλαδή η ετικέτα κλάσης). Λογιστική παλινδρόμηση και παλινδρόμησης probit μοιάζουν περισσότερο με LDA, όπως εξηγούν, επίσης, κατηγορηματικά μεταβλητή. Οι άλλες μέθοδοι είναι προτιμότερη σε εφαρμογές όπου δεν είναι λογικό να υποθέσουμε ότι οι ανεξάρτητες μεταβλητές κανονική κατανομή, η οποία αποτελεί θεμελιώδη παραδοχή της μεθόδου LDA. LDA είναι επίσης στενά με τα κύρια ανάλυση στοιχείων (PCA) και η ανάλυση παράγοντα ότι και οι δύο ψάχνουν γραμμικού συνδυασμού των μεταβλητών που εξηγούν καλύτερα τα δεδομένα. LDA επιχειρεί ρητά να μοντελοποιηθεί η διαφορά ανάμεσα στις τάξεις των δεδομένων. Από την άλλη πλευρά δεν λαμβάνει υπόψη τυχόν διαφορά στην κατηγορία, και ανάλυση των παραγόντων χτίζει τους συνδυασμούς λειτουργία με βάση τις διαφορές παρά τις ομοιότητες. Διακριτική ανάλυση είναι επίσης διαφορετική από την ανάλυση των παραγόντων σε ότι δεν είναι μια τεχνική αλληλεξάρτηση: μια διάκριση μεταξύ των ανεξάρτητων μεταβλητών και εξαρτημένη μεταβλητή (που ονομάζεται επίσης μεταβλητό κριτήριο) πρέπει

να γίνει. LDA λειτουργεί όταν οι μετρήσεις που έγιναν σε ανεξάρτητες μεταβλητές για κάθε παρατήρηση είναι συνεχείς ποσότητες.

### 3.4.3 Square discriminant analysis

Η Square discriminant analysis [16] είναι μια από τις πιο συχνά χρησιμοποιούμενες μη γραμμική τεχνικές για την ταξινόμηση μοτίβου. Στο πλαίσιο QDA, η τάξη δεσμευμένη κατανομή θεωρείται ότι είναι Gaussian, ωστόσο, με ένα επίδομα για τους διάφορους πίνακες συνδιακύμανσης. Σε αυτές τις περιπτώσεις, ένα πιο σύνθετο τετραγωνικό όριο μπορεί να δημιουργηθεί. Είναι επομένως λογικό να πιστέψουμε ότι QDA ταιριάζει καλύτερα την πραγματική δομή των δεδομένων. Ωστόσο, λόγω του γεγονότος ότι η πιο ελεύθεροι παράμετροι για να εκτιμηθούν (Γ πίνακες συνδιακύμανσης, όπου C δηλώνει τον αριθμό των τάξεων), σε σύγκριση με εκείνους σε μια LDA-based λύση (1 μήτρα συνδιακύμανσης), ο QDA είναι πιο ευαίσθητος στις έτσι που ονομάζεται μικρό μέγεθος του δείγματος (SSS) πρόβλημα όταν ο αριθμός των δειγμάτων εκπαίδευσης είναι μικρότερη ή ανάλογη με την διάσταση του χώρου δείγμα.

### 3.4.4 Classification and involution Tree

Το CR-Treedecision δέντρο είναι μια δυαδική αναδρομική διαδικασία διαχωρισμού που μπορεί να επεξεργαστεί συνεχή και εικονικών χαρακτηριστικών, όπως στόχους και προγνωστικούς παράγοντες. Τα δεδομένα διακινούνται σε ακατέργαστη μορφή τους. Αρχίζοντας από την ρίζα, τα δεδομένα χωρίζονται σε δύο παιδιά, και κάθε ένα από τα παιδιά είναι με τη σειρά του χωρίζεται σε εγγόνια. Τα δέντρα που καλλιεργούνται σε ένα μέγιστο μέγεθος χωρίς τη χρήση ενός κανόνα διακοπής. Ουσιαστικά το δέντρο είναι μια αναπτυσσόμενη διαδικασία που σταματά όταν δεν απαιτείται περαιτέρω χωρισμούς λόγω έλλειψης στοιχείων. Το μέγιστο μέγεθος του δέντρου στη συνέχεια επανέρχεται στην ρίζα (ουσιαστικά split by split) μέσω της καινοτόμου μέθοδο του κλαδέματος κόστους-πολυπλοκότητας. Η επόμενη διάσπαση να κλαδεύονται είναι αυτός που συμβάλλουν τουλάχιστον στη συνολική απόδοση του δέντρου σε δεδομένα εκπαίδευσης (και περισσότερες από μία διάσπαση μπορεί να αφαιρεθεί σε μια στιγμή). Το CR-Tree μηχανισμός έχει ως στόχο να παράγει ούτε ένα δέντρο, αλλά μια ακολουθία ένθετων κλαδεύονται δέντρα, καθένα από τα οποία είναι υποψήφια να είναι το βέλτιστοδέντρο.

Στην εξόρυξη δεδομένων, τα δέντρα μπορεί να περιγραφεί και ως το συνδυασμό των μαθηματικών και υπολογιστικών τεχνικών για να βοηθήσει στην περιγραφή, την κατηγοριοποίηση και τη γενίκευση ενός συγκεκριμένου συνόλου δεδομένων. Δεδομένα έρχεται στα αρχεία της μορφής:

$$(\mathbf{x}, Y) = (x_1, x_2, x_3, \dots, x_k, Y)$$

Η εξαρτημένη μεταβλητή, Y, είναι η μεταβλητή-στόχο που προσπαθούμε να κατανοήσουμε, να ταξινομήσουμε ή να γενικεύσουμε. Το διάνυσμα x αποτελείται από τις μεταβλητές εισόδου,  $x_1, x_2, x_3$ , κλπ., που χρησιμοποιούνται για την εργασία αυτή.

### 3.4.5 Logistic involution

Η ανάλυση της Logistic Involution προσφέρει δυνατότητα εξέτασης της επίδρασης των διαφόρων επιχειρημάτων ή 'παραγόντων κινδύνου'. Η ιδέα της λογιστικής παλινδρόμησης βασίζεται στην αντίληψη ότι η πιθανότητα ενός γεγονότος μέσω ενός μοντέλου εμπλοκής μπορεί να περιγραφτεί. Η επιρροή από τα επιχειρήματα μπορεί να μοντελοποιηθεί άμεσα. Ένα ακόμα πλεονέκτημα είναι ότι αυτές οι μεταβλητές μπορούν συνήθως να μεταφέρονται στην αρχική τους μορφή με το μοντέλο. Άλλωστε, μόνο οι συντελεστές εμπλοκής πρέπει να γίνονται κατά εκτίμηση, η οποία μειώνει τον αριθμό των αναγκαίων στατιστικών δοκιμών.



### 3.4.6 Πειραματική Μέθοδος- SIFT-based Neural Network

Το κύριο κομμάτι που ασχοληθήκαμε σε αυτή την πτυχιακή εργασία, είναι η ανάπτυξη, πάντα όμως σε αρχική μορφή, μιας τεχνικής αναγνώρισης συναισθημάτων με την χρήση μιας σχετικά καινούργιας μεθόδου, την Scale-invariant feature transform (SIFT) και μέσω αυτής, την εκπαίδευση ενός νευρωνικού δικτύου με απώτερο σκοπό την ταξινόμηση των συναισθημάτων ενός ανθρώπινου προσώπου. Για να καταλάβουμε όμως λίγο θεωρητικά το πως αυτό είναι εφικτό ας δούμε ξεχωριστά τι κάνει ο αλγόριθμος SIFT και πως λειτουργεί ένα νευρωνικό δίκτυο.

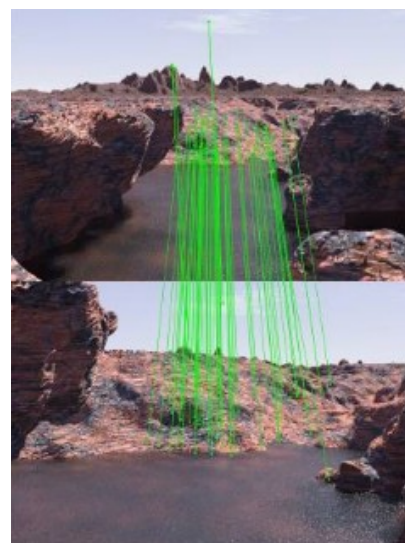
#### 3.4.6.1 Scale-invariant feature transform (SIFT)

Ο αλγόριθμος SIFT [9](Lowe, 2004), αποτελείται από έναν ανιχνευτή χαρακτηριστικών γνωρισμάτων και από έναν περιγραφέα (descriptor). Αρχικά, ο ανιχνευτής γνωρισμάτων εντοπίζει σημεία που φέρουν διακριτή πληροφορία, ανεξάρτητη από πιθανές μεταβολές της εικόνας (π.χ. περιστροφή, αλλαγή φωτισμού κ.ά.). Με τη χρήση μιας γκαουσιανής κατανομής, η εικόνα μετασχηματίζεται από το χρωματικό χώρο στο χώρο κλιμάκωσης που προτάθηκε από τον Lindeberg (1994). Στη συνέχεια, η συνάρτηση διαφοράς του Gauss (DoG) συνδυάζεται με μια παρεμβολή στο χώρο κλιμάκωσης, παράγουν τα σημεία ενδιαφέροντος. Ο περιγραφέας είναι άμεσα συνδεδεμένος με τη διαδικασία ανάθεσης ενός προσανατολισμού, σε κάθε σημείο ενδιαφέροντος, όπως αυτό έχει προκύψει από τον ανιχνευτή γνωρισμάτων. Συγκεκριμένα, ο περιγραφέας είναι ένα διάνυσμα, το μέτρο και η διεύθυνση του οποίου, υπολογίζονται από πληροφορίες της γειτονιάς του εκάστοτε σημείου ενδιαφέροντος.

Τα καλύτερα αποτελέσματα προκύπτουν όταν λαμβάνεται πληροφορία από μια γειτονιά εικονοστοιχείων 4x4. Η διεύθυνση του διανύσματος τυποποιείται έτσι ώστε να λαμβάνει προσανατολισμούς ανά 45°. Με αυτό το τρόπο, τοπικές πληροφορίες ενός αντικειμένου προσδιορίζονται με έναν και μοναδικό τρόπο, γεγονός που εξασφαλίζει τη μοναδικότητα κάθε χαρακτηριστικού σημείου ενός αντικειμένου. Η αποδοτικότητα του περιγραφέα ελέγχεται κατά τη διαδικασία της σύγκρισης της αρχικής εικόνας με τη τελική. Η τελευταία μπορεί να προκύψει είτε από περιστροφή είτε από μετατόπιση της αρχικής σε οποιοδήποτε άξονα. Κατά τη διάρκεια της παραπάνω διαδικασίας, τα διανύσματα των περιγραφέων των δύο εικόνων ελέγχονται για τυχόν ομοιότητες. Το γεγονός ότι η ευρωστία που παρουσιάζεται είναι αξιοσημείωτη (Mikolajczyk, 2005), καθιστά τον SIFT έναν από τους καλύτερους αλγόριθμους ανίχνευσης χαρακτηριστικών γνωρισμάτων που έχουν προταθεί στη διεθνή βιβλιογραφία.



Εικόνα 20: Διανύσματα του αλγόριθμου SIFT



Εικόνα 21: Εύρεση κοινών σημείων μέσω SIFT



### 3.4.6.2 Νευρωνικά δίκτυα

Το **νευρωνικό δίκτυο** είναι ένα δίκτυο από απλούς υπολογιστικούς κόμβους (νευρώνες, νευρώνια), διασυνδεδεμένους μεταξύ τους. Είναι εμπνευσμένο από το Κεντρικό Νευρικό Σύστημα (ΚΝΣ), το οποίο προσπαθεί να προσομοιώσει.

Οι *νευρώνες* είναι τα δομικά στοιχεία του δικτύου. Κάθε τέτοιος κόμβος δέχεται ένα σύνολο αριθμητικών εισόδων από διαφορετικές πηγές (είτε από άλλους νευρώνες, είτε από το περιβάλλον), επιτελεί έναν υπολογισμό με βάση αυτές τις εισόδους και παράγει μία έξοδο. Η εν λόγω έξοδος είτε κατευθύνεται στο περιβάλλον, είτε τροφοδοτείται ως είσοδος σε άλλους νευρώνες του δικτύου. Υπάρχουν τρεις τύποι νευρώνων: οι *νευρώνες εισόδου*, οι *νευρώνες εξόδου* και οι *υπολογιστικοί νευρώνες* ή *κρυμμένοι νευρώνες*. Οι νευρώνες εισόδου δεν επιτελούν κανέναν υπολογισμό, μεσολαβούν απλώς ανάμεσα στις περιβαλλοντικές εισόδους του δικτύου και στους υπολογιστικούς νευρώνες. Οι νευρώνες εξόδου διοχετεύουν στο περιβάλλον τις τελικές αριθμητικές εξόδους του δικτύου. Οι υπολογιστικοί νευρώνες πολλαπλασιάζουν κάθε είσοδό τους με το αντίστοιχο *συναπτικό βάρος* και υπολογίζουν το ολικό άθροισμα των γινομένων. Το άθροισμα αυτό τροφοδοτείται ως όρισμα στη *συνάρτηση ενεργοποίησης*, την οποία υλοποιεί εσωτερικά κάθε κόμβος. Η τιμή που λαμβάνει η συνάρτηση για το εν λόγω όρισμα είναι και η έξοδος του νευρώνα για τις τρέχουσες εισόδους και βάρη.

Εάν  $x_{ki}$  είναι η  $i$ -οστή είσοδος του  $k$  νευρώνα,  $w_{ki}$ : το  $i$ -οστό συναπτικό βάρος του  $k$  νευρώνα και  $\phi(\cdot)$  η συνάρτηση ενεργοποίησης του νευρωνικού δικτύου, τότε η έξοδος  $y_k$  του  $k$  νευρώνα δίνεται από την εξίσωση:

$$y_k = \phi \left( \sum_{i=0}^N x_{ki} w_{ki} \right)$$

Στον  $k$ -οστό νευρώνα υπάρχει ένα συναπτικό βάρος  $w_{k0}$  με ιδιαίτερη σημασία, το οποίο καλείται **πόλωση** ή **κατώφλι** (bias, threshold). Η τιμή της εισόδου του είναι πάντα η μονάδα,  $x_{k0} = 1$ . Εάν το συνολικό άθροισμα από τις υπόλοιπες εισόδους του νευρώνα είναι μεγαλύτερο από την τιμή αυτή, τότε ο νευρώνας ενεργοποιείται. Εάν είναι μικρότερο, τότε ο νευρώνας παραμένει ανενεργός. Η ιδέα προέκυψε από τα βιολογικά νευρικά κύτταρα.

Όπως είναι φανερό, οι αριθμοί οι οποίοι συναποτελούν το διάνυσμα εισόδου (κάθε στοιχείο του διανύσματος τροφοδοτείται κατά τη λειτουργία του δικτύου σε έναν νευρώνα εισόδου), αλλά και οι αριθμοί οι οποίοι συναποτελούν το διάνυσμα εξόδου (κάθε στοιχείο του οποίου εμφανίζεται, μετά το πέρας του ολικού υπολογισμού, σε έναν νευρώνα εξόδου), περιγράφουν χαρακτηριστικά του προς επίλυση προβλήματος. Συνήθως αυτό που μας ενδιαφέρει είναι το δίκτυο να απεικονίζει με ορθό τρόπο διανύσματα εισόδου σε κατάλληλα διανύσματα εξόδου, το πρόβλημα δηλαδή είναι η υλοποίηση μίας συνάρτησης πολλαπλών μεταβλητών, κατά κανόνα περίπλοκης και με άγνωστο ακριβή τύπο. Τέτοιες απεικονίσεις έχουν εφαρμογή σε ποικιλία τομέων της επιστήμης και της τεχνολογίας, αφού λειτουργούν ως αριθμητικά μοντέλα για πολλά διαφορετικά ζητήματα. Το ίδιο δίκτυο μπορεί να υλοποιήσει άπειρες διαφορετικές απεικονίσεις, μία για κάθε διαφορετική επιλογή συνόλου συναπτικών βαρών.

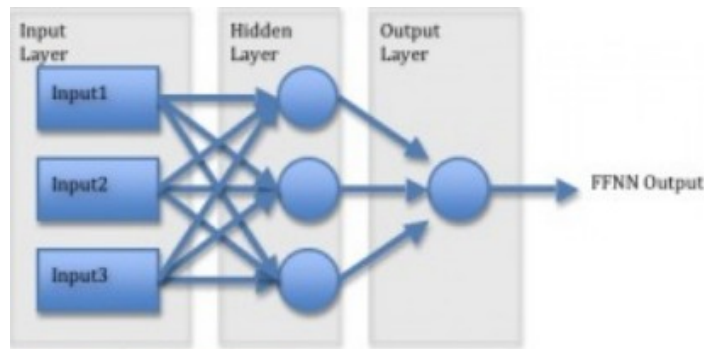


Figure 4 – Feedforward Neural Network

Εικόνα 22: Νευρωνικό Δίκτυο

Το κύριο χαρακτηριστικό των νευρωνικών δικτύων είναι η εγγενής ικανότητα μάθησης. Ως μάθηση μπορεί να οριστεί η σταδιακή βελτίωση της ικανότητας του δικτύου να επιλύει κάποιο πρόβλημα (π.χ. η σταδιακή προσέγγιση μίας συνάρτησης). Η μάθηση επιτυγχάνεται μέσω της **εκπαίδευσης**, μίας επαναληπτικής διαδικασίας σταδιακής προσαρμογής των παραμέτρων του δικτύου (συνήθως των βαρών και της πόλωσής του) σε τιμές κατάλληλες ώστε να επιλύεται με επαρκή επιτυχία το προς εξέταση πρόβλημα. Αφού ένα δίκτυο εκπαιδευτεί, οι παράμετροί του συνήθως «παγώνουν» στις κατάλληλες τιμές και από εκεί κι έπειτα είναι σε λειτουργική κατάσταση. Το ζητούμενο είναι το λειτουργικό δίκτυο να χαρακτηρίζεται από μία ικανότητα **γενίκευσης**: αυτό σημαίνει πως δίνει ορθές εξόδους για εισόδους καινοφανείς και διαφορετικές από αυτές με τις οποίες εκπαιδεύτηκε.

## 4. Σχεδιασμός Υλοποίησης

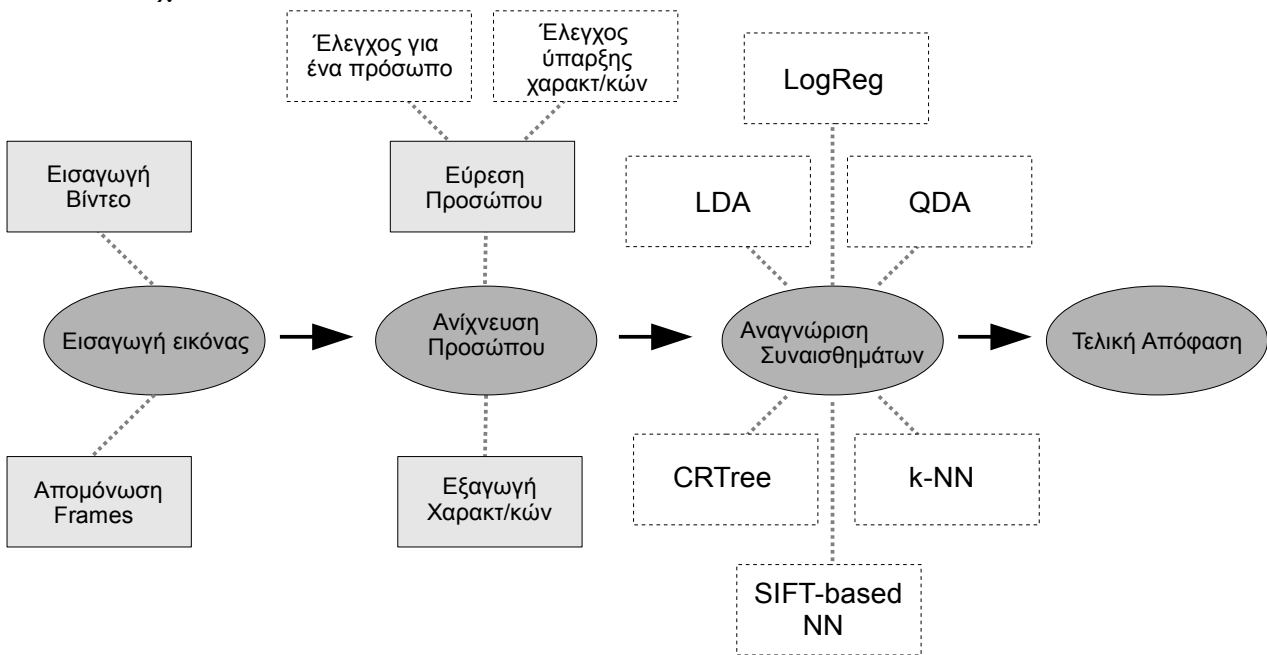
Στο κεφάλαιο αυτό θα δούμε βήμα βήμα την πορεία της υλοποίησης της πτυχιακής εργασίας στο πρακτικό κομμάτι. Θα αναλύσουμε τις μεθόδους, τα προβλήματα που συναντήθηκαν για την κάθε μια, και την απόδοση της. Υπάρχουν πέντε βασικά βήματα που έπρεπε να πραγματοποιηθούν για να μπορέσει να υλοποιηθεί το τελικό πρόγραμμα με επιτυχία.

- Η εισαγωγή των δεδομένων προς επεξεργασία (το βίντεο)
- Η εύρεση προσώπου
- Η ανίχνευση των συναισθημάτων
- Η ανάπτυξη μεθόδου ανίχνευσης συναισθημάτων μέσω SIFT-based Neural Network
- Η λήψη της τελικής απόφασης για το επεξεργαζόμενο βίντεο

Στην συνέχεια αυτού του κεφαλαίου θα αναλυθούν όλα τα παραπάνω.

Στο παρακάτω διάγραμμα παρουσιάζεται το πλάνο και η γενική ιδέα την οποία καλούμαστε να ολοκληρώσουμε με το πρόγραμμα μας.

Εικόνα 23: Σχέδιο κώδικα



Είναι σημαντικό να αναφέρουμε σε αυτό το σημείο ότι το πρόγραμμα έχει δημιουργηθεί σε Matlab 7.11.0 (R2010b), σε Windows 7 x32bit έκδοση. Γνωστοποιούμε από τώρα το περιβάλλον το οποίο χρησιμοποιήθηκε για την υλοποίηση του προγράμματος, καθώς με αυτόν τον τρόπο θέτουμε κάποιους κανόνες για την ορθή λειτουργία του προγράμματος.

### 4.1 Εισαγωγή της εικόνας

Η βασική πηγή από την οποία θα αντλούμε τα δεδομένα μας, τα οποία και θα επεξεργαζόμαστε, είναι μουσικά βίντεο κλιπ. Για την εισαγωγή τους στο Matlab χρησιμοποιούμε την εντολή `mmreader`, η οποία είναι μια εντολή με την οποία το matlab εισάγει στο workspace του αρχεία multimedia. Οι τύποι που μπορούν να εισαχθούν μέσω αυτής της εντολής είναι αρχεία AVI, MPEG-1, Windows Media Video (.wmv, .asf, .asx) και οποιοδήποτε άλλος τύπος υποστηρίζεται από το Microsoft DirectShow. Αφού ανοίξει το παράθυρο επιλογής του βίντεο και ολοκληρωθεί η διαδικασία της εισαγωγής του βίντεο, η εξαγωγή των εικόνων(καρέ) που θα χρησιμοποιηθούν για την ανάλυση του, ξεκινάει έπειτα από την επιλογή του χρήστη να ξεκινήσει η ανάλυση. Συνολικά για το βίντεο θα παρθούν δέκα έγκυρα δείγματα εικόνων προσώπου. Αυτό έχει ως αποτέλεσμα να μην γνωρίζουμε πόσα ακριβώς καρέ θα επεξεργαστούμε και τον λόγο αυτού θα αναλύσουμε και θα εξηγήσουμε στο επόμενο μέρος, αυτό της ανίχνευσης προσώπου.

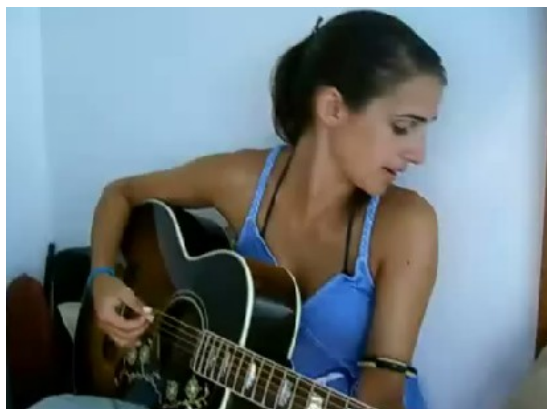
Τα βίντεο που χρησιμοποιήσαμε για να δοκιμάσουμε και να εκπαιδύσουμε το πρόγραμμα μας ήταν κατά κύριο λόγο Covers τραγουδιών με σκηνικά απλά στις περισσότερες περιπτώσεις. Το σκηνικό συνήθως περιείχε ένα πρόσωπο, τον ερμηνευτή, ένα μουσικό όργανο, συνήθως κιθάρα ή πιάνο, και ένα απλό φόντο.



Εικόνα 24: Παράδειγμα Καρέ no1



Εικόνα 25: Παράδειγμα Καρέ no2



Εικόνα 26: Παράδειγμα Καρέ no3

Όπως παρατηρούμε στις εικόνες 24,25 και 26 τα πρόσωπα είναι σχετικά εύκολα ανιχνεύσιμα και το σκηνικό δεν είναι υπερβολικά φορτωμένο. Παρόλα αυτά ο αλγόριθμος μας είναι σχεδιασμένος όπως θα δούμε και στην συνέχεια κατά τέτοιο τρόπο ώστε να αποφεύγει λάθη στον εντοπισμό των προσώπων. Το πιο σημαντικό σημείο που πρέπει να προσέχουμε κατά την εισαγωγή ενός βίντεο στο πρόγραμμα είναι η ανάλυση του, η οποία θα ήταν καλό να είναι καλή, για να πετύχουμε τα καλύτερα δυνατά αποτελέσματα.

## 4.2 Ανίχνευση Προσώπου

Για την ανίχνευση και αποκοπή του προσώπου από την υπόλοιπη εικόνα δοκιμάστηκαν διάφορες μέθοδοι. Από άλλες τα αποτελέσματα ήταν ικανοποιητικά, άλλες ήταν γρήγορες, άλλες συνδύαζαν και τα δύο αυτά πλεονεκτήματα.

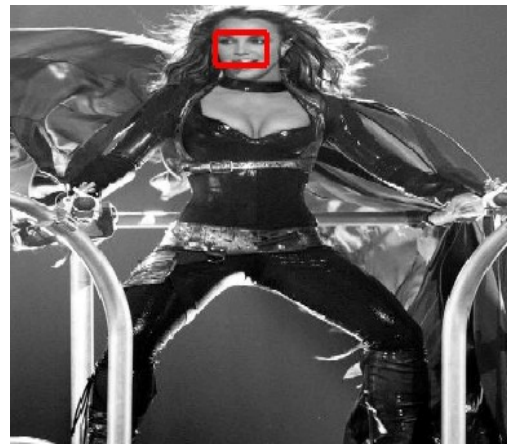
### 4.2.1 Δοκιμές Μεθόδων Face Detection

#### *Face Detection by Mikael Nilsson*

Το συγκεκριμένο σύστημα [23] ανίχνευσης προσώπου βασίζεται σε μια βιβλιοθήκη dll και η μέθοδος που χρησιμοποιείται είναι η Successive Mean Quantization Transform(SMQT) και ο Sparse Network of Winnows(SNoW) ταξινομητής. Γενικά το πρόγραμμα αυτό βασίζεται στο paper: “Face Detection using local SMQT features and split up SNoW classifier”. Πρόκειται για ένα πολύ καλό αλγόριθμο ο οποίος έχει καλά ποσοστά επιτυχίας στην ανίχνευση των προσώπων μέσα σε μια εικόνα. Δοκιμάστηκε σε μια σειρά από εικόνες που είχαν σαν κύριο θέμα την μουσική και ζωντανές εμφανίσεις, με τα αποτελέσματα να είναι άκρος εντυπωσιακά.



Εικόνα 27: Nilsson's Algorithm result 1



Εικόνα 28: Nilsson's Algorithm result

Ενώ βλέπουμε στις εικόνες 27 και 28 ότι έχουμε αρκετά καλά αποτελέσματα, παρατηρείτε ότι το χρονικό διάστημα που απαιτείτε είναι αρκετά μεγάλο για την ανίχνευση ενός και μόνο προσώπου μέσα σε μια εικόνα. Αυτό από μόνο του αποκλείει τον αλγόριθμο αυτό από την χρήση του στο πρόγραμμα μας, καθώς στην καλύτερη των περιπτώσεων θα πρέπει να αναλυθούν δέκα εικόνες για την ανίχνευση προσώπου -και κρίνοντας εκ του αποτελέσματος σε καμία περίπτωση δεν αναλύονται μόνο δέκα εικόνες, καθώς αναζητάμε τις καλύτερες για ανάλυση, πράγμα που θα δούμε στην συνέχεια- και ο χρόνος που απαιτείτε για αυτό είναι αρκετά μεγάλος.

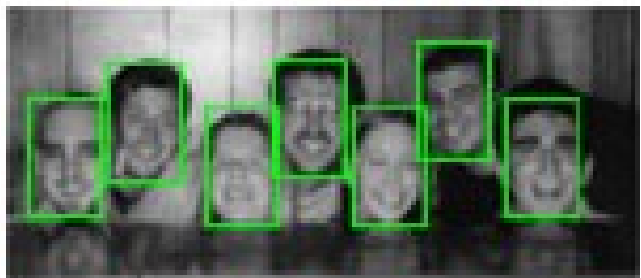
#### *Face Detection System by Omid Sakhi*



Ένα σύστημα Face Detection [20] το οποίο δημιουργήθηκε από τον Omid Sakhi δοκιμάστηκε για τις ανάγκες της ολοκλήρωσης του προγράμματος. Πρόκειται για ένα σύστημα το οποίο χρησιμοποιεί Gabor features και νευρωνικό δίκτυο για την ανίχνευση των προσώπων. Λόγω της μορφής του κώδικα και του τρόπου ανάπτυξης του κρίθηκε ακατάλληλο λόγω και της μη αποτελεσματικότητας του καθώς θα έπρεπε να εκπαιδευτούμε το νευρωνικό δίκτυο από την αρχή, αλλά και του χρόνου που χρειάζεται για την ολοκλήρωση της ανίχνευσης της εικόνας, εάν αυτή είναι μια εικόνα υψηλής ανάλυσης. Παρόλα αυτά αξίζει να επισημάνουμε ότι εάν τα requirements που έχει είναι βολικά με την χρήση που το προσδιορίζουμε, είναι ένας αρκετά καλός αλγόριθμος και έχει αποσπάσει θετικές κριτικές στην επίσημη ιστοσελίδα της MathWorks. Στην παρακάτω εικόνα (29) απεικονίζεται το αποτέλεσμα που λαμβάνουμε όταν δοκιμάσουμε τον αλγόριθμο σε εικόνα τις δικιάς μας θεματολογίας, ενώ στην εικόνα 30 είναι το αποτέλεσμα που παίρνουμε όταν η εικόνα που παρέχουμε στον αλγόριθμο υπακούει στα requirements τα οποία έχουν τεθεί.



Εικόνα 29: Εσφαλμένο Αποτέλεσμα



Εικόνα 30: Ορθό Αποτέλεσμα



Εικόνα 31: Σημεία Αναζήτησης

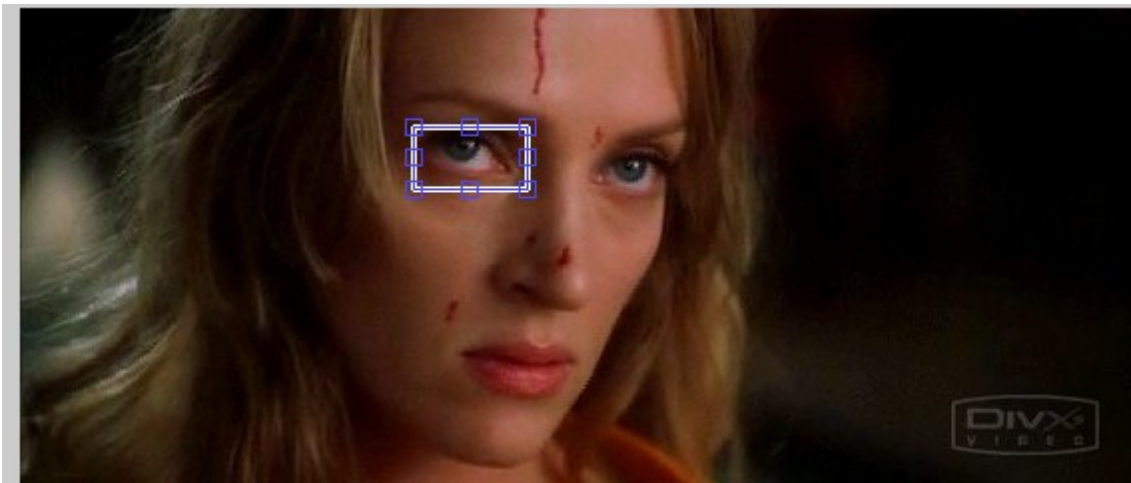
Στην εικόνα 31 φαίνονται τα σημεία τα οποία πρέπει να εξεταστούν ώστε να πάρουμε τα αποτελέσματα μας για την ύπαρξη ή όχι προσώπου μέσα σε αυτό το πλάνο. Από την ποσότητα τους καταλαβαίνουμε ότι είναι μια διαδικασία η οποία δεν πρόκειται να είναι σύντομη.

### ***Face Detection using Support Vector Machine (SVM)***

Το συγκεκριμένο πρόγραμμα [22] είναι ένας κλώνος του προαναφερόμενου (Face detection System by Omid Sakhi). Σε αντίθεση όμως με το προηγούμενο πρόγραμμα, το οποίο έκανε την χρήση νευρωνικού δικτύου, το συγκεκριμένο βασίζεται στο Support Vector Machine (SVM). Παρόλα αυτά δεν παρατηρούμε σημαντικές διαφορές στον τρόπο λειτουργίας του προγράμματος καθώς και διορθώσεις στο χρόνο που απαιτείται για την λήψη των αποτελεσμάτων. Ένα σημαντικό μειονέκτημα αλλά ταυτόχρονα και πλεονεκτήμα που έχουν και οι δύο αλγόριθμοι, είναι ότι μπορούν να εκπαιδευτούν με αντικείμενα που πιθανών να είναι πρόσωπα και αντικείμενα που είναι μη πρόσωπα. Μέσω από αυτήν την διαδικασία θα μπορούσαν να βελτιωθούν τα αποτελέσματα τους αλλά και πάλι παραμένει το πρόβλημα του χρόνου.

### ***Face and Eye Detection by Mustafa Ucak***

Πρόκειται για ένα πρόγραμμα [21] το οποίο εντοπίζει την θέση των ματιών και του προσώπου, με δύο μεγάλα μειονέκτημα να μας κάνουν αρνητικούς εξ αρχής για την μελλοντική του χρήση. Το πρώτο και το κυριότερο είναι ότι μας αναγκάζει να εντοπίσουμε χειροκίνητα τα δυο μάτια στο πρώτο καρέ που μας δίνετε (εικόνα 32). Το δεύτερο είναι ο χρόνος που χρειάζεται για να λάβουμε τα αποτελέσματα.



Εικόνα 32: Χειροκίνητο μαρκάρισμα περιοχής ματιού

## ***Fdlibmex – Fast and simple face detection by Wolf Kienzle***

Ο *fdlibmex* [18] αλγόριθμος αποτελείται από μια βιβλιοθήκη *dll* η οποία περιέχει μεθόδους βασισμένες στον αλγόριθμο του Nilsson, τον οποίο ήδη αναφέραμε. Ουσιαστικά βασίζεται και αυτός σε *Successive Mean Quantization Transform (SMQT)* και ο *Sparse Network of Winnows (SNoW)* ταξινομητή, με την διαφορά ότι ο κώδικας του έχει βελτιστοποιηθεί. Είναι αρκετά αποδοτικός και σαφέστατα γρηγορότερος. Δυστυχώς πρόκειται για έναν αλγόριθμο ο οποίος στηρίζεται πάνω σε μια *dll* βιβλιοθήκη η οποία έχει δημιουργηθεί με *compiler* των 32bit λειτουργίας. Όμως το θετικό του είναι ότι παρέχονται και οι αντίστοιχες βιβλιοθήκες για λειτουργικό των *Mac* και *Linux*.

## ***Viola Jones Object Detection***

Ο αλγόριθμος αυτός [4] είναι δημιούργημα των Paul Viola και Michael Jones, το οποίο και πήρε μορφή το 2001. Είναι ο πρώτος αλγόριθμος ανίχνευσης αντικειμένων που παρείχε ικανοποιητικά αποτελέσματα και αρκετά καλό ποσοστό επιτυχίας σε πραγματικό χρόνο. Αν και μπορούν να εκπαιδευθούν για να ανιχνεύουν μια ποικιλία από κλάσεις αντικειμένων, ο κύριος στόχος τους ήταν η επίλυση του προβλήματος της ανίχνευσης προσώπων.



Εικόνα 33:  
Αποτελέσματα  
Αλγορίθμου Viola Jones

Όνομα Αλγορίθμου	Δημιουργός	Πηγή	Πλεονεκτήματα*	Μειονεκτήματα*
<i>Face detection system</i>	Omid Sakhi	MatlLab Central	Απόδοση	Ταχύτητα - Ανάγκη μικρής ανάλυσης
<i>Face detection using SVM</i>	Omid Sakhi	MatlLab Central	Απόδοση	Ταχύτητα - Ανάγκη μικρής ανάλυσης
<i>Face and Eye Detection</i>	Mustafa Ucak	MatlLab Central	Σχετικά καλή ταχύτητα	Αρχική ένδειξη θέσης ματιού
<i>Face Detection</i>	Nilsson Mikael	MatlLab Central	Απόδοση	Ταχύτητα - Λειτουργία σε 32bit
<i>fdlibmex</i>	Wolf Kienzle	MatlLab Central	Απόδοση και ταχύτητα	Λειτουργία μόνο σε συστήματα 32 bit
<i>Viola Jones Object Detection</i>	Paul Viola- Micheal Jones	MarLab Central	Απόδοση - Ταχύτητα	Χρήση XML files

Πίνακας 3: Συγκεντρωτικός Πίνακας μεθόδων Face Detection

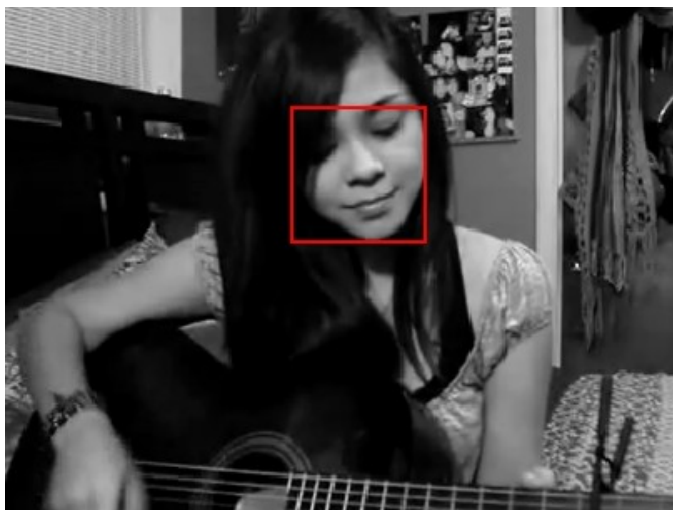


\*Τα πλεονεκτήματα και τα μειονέκτημα είναι με κριτήριο τον σκοπό που προορίζουμε κάθε αλγόριθμο. Αφορούν καθαρά την ωφελιμότητα τους στο συγκεκριμένο πρόγραμμα.

Στον παραπάνω πίνακα γίνεται η συγκέντρωση των αλγορίθμων που δοκιμάστηκαν και ήταν οι δημοφιλέστεροι για να χρησιμοποιηθούν για την υλοποίηση του προγράμματος. Υπήρξαν και δοκιμές άλλων μεθόδων, όπως αυτής της ανίχνευσης προσώπου με βάση το χρώμα δέρματος, οι οποίες λόγω των προδιαγραφών που έπρεπε να τηρούνται σε κάθε εικόνα προσώπου για την ανίχνευση του κρίθηκαν εξ αρχής ακατάλληλες για χρήση, οπότε και παραλείπονται στην παρουσίαση.

## 4.2.2 Υλοποίηση Ανίχνευσης Προσώπου

Για την τελική μέθοδο που επιλέξαμε λάβαμε υπόψιν όλα τα παραπάνω κριτήρια. Ο αλγόριθμος `fdlibmex` κρίθηκε καταλληλότερος για την χρήση που προοριζόταν μέσα στο πρόγραμμα. Έπειτα από ένα μικρό σύνολο δοκιμών παρατηρήθηκε ότι το ποσοστό επιτυχίας του στην ανίχνευση προσώπων σε μη ελεγχόμενο περιβάλλον κυμαίνεται κοντά στο 80%. Είναι ένα αρκετά ικανοποιητικό ποσοστό, εάν λάβουμε υπόψιν και την δυσκολία που υπάρχει στα μουσικά βίντεο κλιπ, καθώς δεν είναι πολλές φορές εφικτό να υπάρχει άριστος φωτισμός, απλοϊκό περιβάλλον και συνεχή λήψη του προσώπου από την ορθή γωνία.



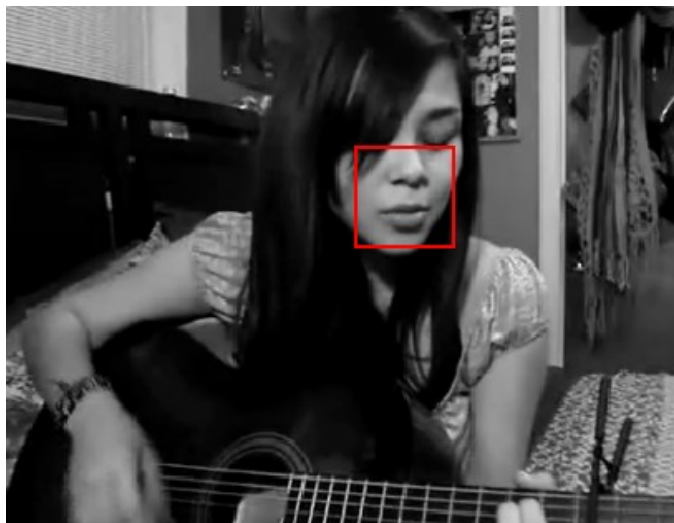
Εικόνα 34: Ορθή ανίχνευση προσώπου

Στην εικόνα 34 ο αλγόριθμος μας ανιχνεύει σωστά το πρόσωπο. Όπως μπορούμε να παρατηρήσουμε το σκηνικό που υπάρχει πίσω από την κοπέλα είναι αρκετά πολύπλοκο, καθώς βρίσκεται μέσα σε ένα δωμάτιο και πολλά είναι τα αντικείμενα που θα μπορούσαν να παραπλανήσουν τον αλγόριθμο μας, πράγμα το οποίο συμβαίνει όπως βλέπουμε στην περίπτωση της εικόνας 35.



Εικόνα 35: Εσφαλμένη ανίχνευση προσώπου

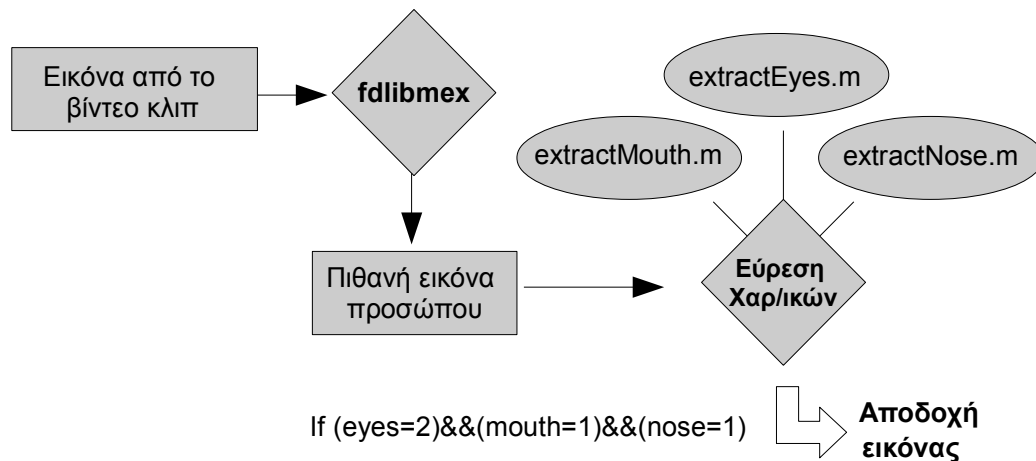
Ένα ακόμα πρόβλημα που καλούμαστε να αντιμετωπίσουμε, πέρα από την εσφαλμένη ανίχνευση προσώπου, είναι και η καταλληλότητα μιας εικόνας προσώπου για την χρήση της για ανάλυση. Όπως παρατηρούμε και στην εικόνα που ακολουθεί (εικόνα 36), υπάρχει μεγάλη πιθανότητα ενώ υπάρχει όντως πρόσωπο στην εικόνα και το σημείο που μας υποδεικνύεται, να μην είναι εμφανή τα χαρακτηριστικά που χρειαζόμαστε για την ανάλυση του προσώπου (μάτια και στόμα).



Εικόνα 36: Ακατάλληλο πρόσωπο για ανάλυση

Για να καταφέρουμε να επιλύσουμε αυτά τα δύο προβλήματα στην αναγνώριση προσώπου – την εσφαλμένη ανίχνευση προσώπων και την απόρριψη μη κατάλληλων προσώπων για ανάλυση – θα χρησιμοποιήσουμε τρεις συναρτήσεις. Και οι τρεις αυτές συναρτήσεις έχουν σκοπό τον εντοπισμό και την εξαγωγή των χαρακτηριστικών του ανθρώπινου προσώπου. Πρόκειται για τις συναρτήσεις `extractMouth.m`, `extractEyes.m` και `extractNose.m`. Σαν όρισμα δέχονται την εικόνα του προσώπου που θέλουμε να εξετάσουμε την καταλληλότητα του και μας επιστρέφει έναν πίνακα με την θέση του αντίστοιχου χαρακτηριστικού.

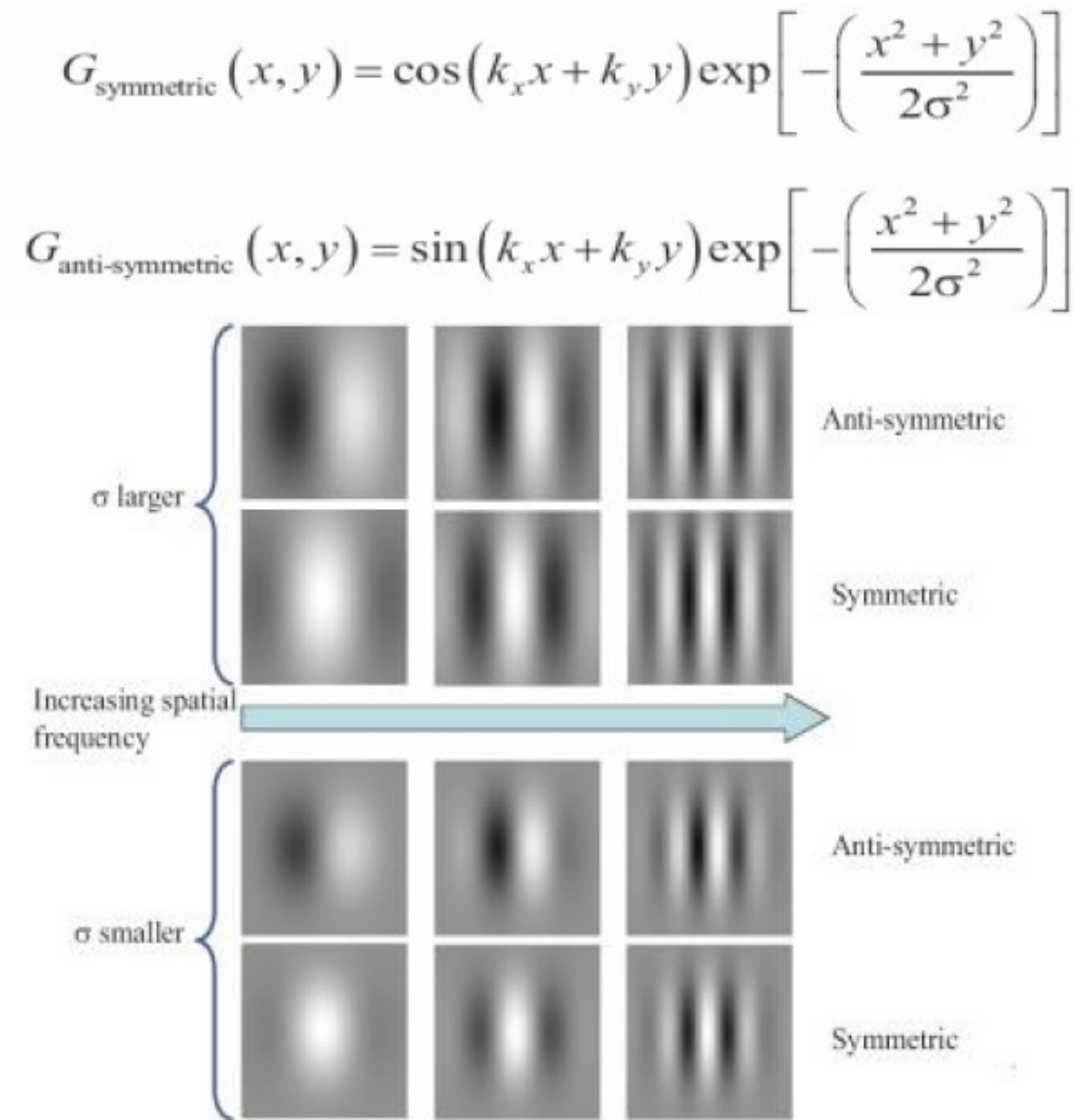
Εικόνα 37: Τρόπος λειτουργίας Face Detection



Όμως πρέπει να δούμε πως καταφέρνεται η εξαγωγή αυτών των χαρακτηριστικών του προσώπου σε κάθε περίπτωση. Οι αλγόριθμοι αυτοί χρησιμοποιήθηκαν στην αναγνώριση προσώπου που σχεδιάστηκε από τον Micheal Nilsson. Ο συνδιασμός του με τον αλγόριθμο μας, μας δίνει το καλύτερο δυνατό αποτέλεσμα. Η χρήση του Gabor wavelet transformation είναι αυτός που μας βοηθά να εντοπίσουμε τα χαρακτηριστικά του προσώπου. Πως όμως λειτουργεί αυτός ο μετασχηματισμός;

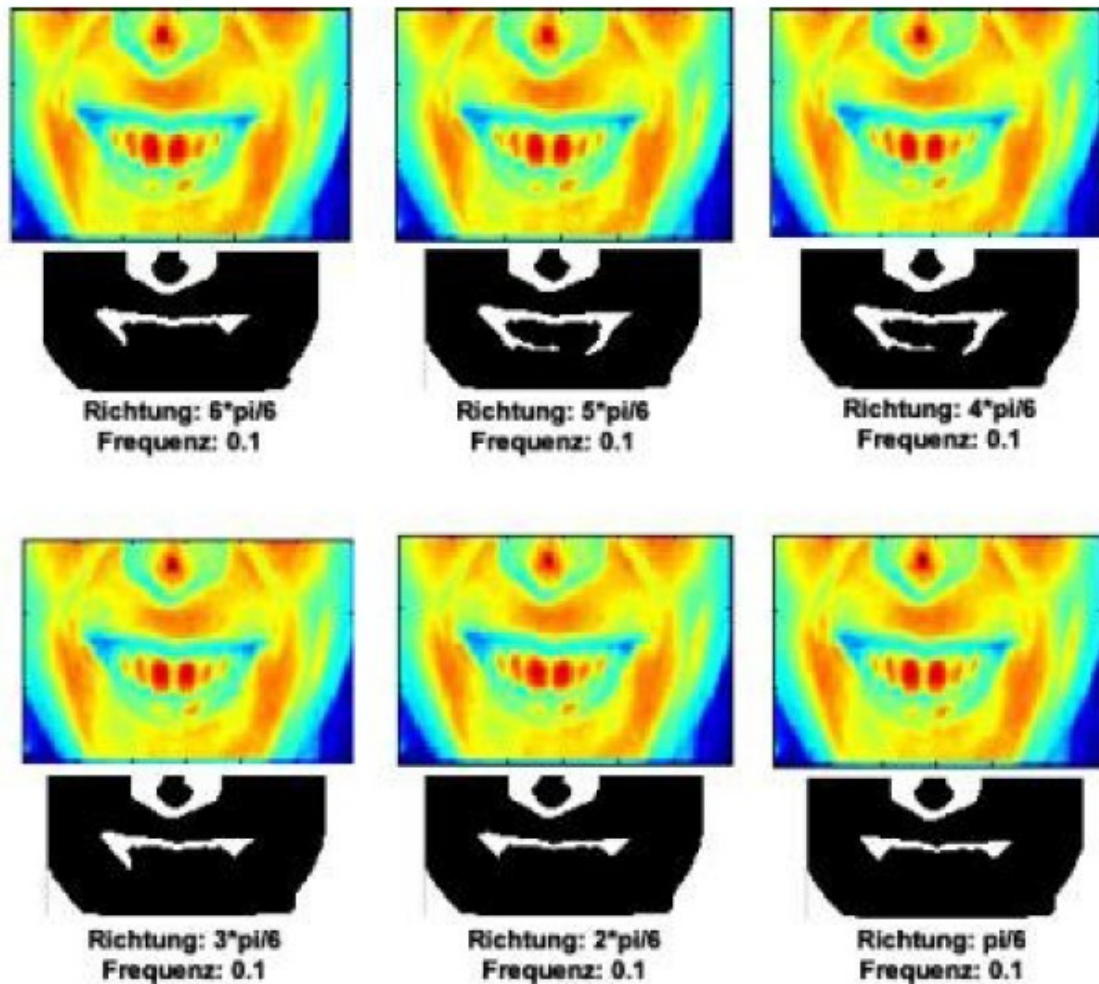
### **Gabor Wavelet Transformation**

Στο πρώτο βήμα της κατάκτησης, γίνεται η αναγνώριση συναισθήματος στα σημαντικά εύρη του προσώπου με τη βοήθεια του Gabor Wavelet μετασχηματισμού. Χρησιμοποιήσαμε τις συναρτήσεις των Ville Kyrki και Joni Kamarainen για την παραγωγή των φίλτρων Gabor (βλ. Αναφορές, Βιβλιοθήκες). Ένα φίλτρο Gabor είναι ένας μετασχηματισμός Fourier που συνδυάζεται με μια Gauss λειτουργία, προκειμένου να φιλτραριστούν οι τοπικές πληροφορίες της εικόνας. Για να λάβουμε μια πολύ καλή περιγραφή της εικόνας, διάφορα φίλτρα με διαφορετικές συχνότητες και κατευθύνσεις μπορούν να εφαρμοστούν. Η μαθηματική μορφή αυτών των φίλτρων είναι η ακολουθία της γραφικής παράστασης που παρουσιάζεται, όπου “kx”, “ky” η χωρική συχνότητα και “σ” το μέγεθος του φίλτρου.



Εικόνα 38: Διαφορετικά φίλτρα *Kernels* με διαφορετικά μεγέθη

Για την κατάτμηση του προσώπου, χρησιμοποιούνται έξι διαφορετικές κατευθύνσεις, με δύο διαφορετικές συχνότητες το περισσότερο. Οι κατευθύνσεις επιλέχτηκαν όπως στο BaGeVen (βλ. Αναφορές, Βιβλιοθήκες) ως εξής: σε  $\pi/6$  βήματα από  $30^\circ$  ως  $180^\circ$ , έτσι έχουμε συνολικά 18 διαφορετικά φίλτρα Gabor. Η συχνότητα γίνεται ανάλογη του εύρους του προσώπου που μπορεί να κατατμηθεί και ποικίλλει (στόμα: 0.1, μύτη: 0.5, μάτια: 0.15), δεδομένου ότι τα διαφορετικά εύρη παρέχουν καλύτερα αποτελέσματα με διαφορετικές συχνότητες. Ο Gabor μετασχηματίζει τις εικόνες που περιέχουν τις σημαντικές περιοχές, όπως το στόμα, τη μύτη και τα μάτια σαν τοπικό ελάχιστο με τη βοήθεια του κατάλληλου κατώτατου ορίου. Οι γραφικές παραστάσεις που ακολουθούν σημειώνουν τις φιλτραρισμένες εικόνες κατώτερου ορίου των αρχικών εικόνων (κατάτμηση του στόματος) και τις εικόνες αποτελέσματος. Το σκούρο και ανοιχτό μπλε αντιπροσωπεύουν τις χαμηλές τιμές, ενώ το κόκκινο τις υψηλές:



Εικόνα 39: Φιλτραρισμένες εικόνες κατώτερου ορίου

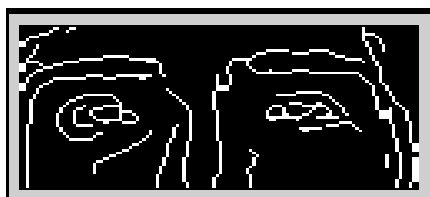
Τελικά οι προκύπτουσες εικόνες είναι ενωμένες και οδηγούν σε μια κατάλληλη αφετηρία για την περαιτέρω επεξεργασία ή/και τη βελτίωση της κατάτμησης της εικόνας.

#### Εύρεση περιοχή ματιών (τρόπος λειτουργίας της συνάρτησης *extractEyes*)

Η εύρεση των ματιών στο πρόσωπο είναι ένα σημαντικό σημείο στην Αναγνώριση Συναισθήματος. Τα μάτια δηλώνουν πολλά για την έκφραση π.χ. τα πολύ ανοικτά μάτια είναι συχνά μια ένδειξη για το γεγονός ότι ένα πρόσωπο φαίνεται έκπληκτο. Η αποπεράτωση είναι πίο σύνθετη και περιγράφεται βαθμιαία εδώ.

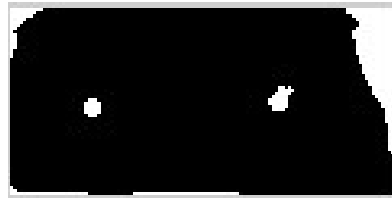
Αναλυτικά τα βήματα:

1. Σε πρώτη φάση παρέχεται μια εικόνα ακρών (edge picture). Αυτή η εικόνα, προβάλλει τα όρια, τα οποία δεν μπορούν να ξεπεραστούν. Το μάτι είναι διαχωρισμένο απο την υπόλοιπη περιοχή και με την αναδημιουργία της άκρης η έκφραση διατηρείται.



Εικόνα 40: Εικόνα ακρών (edge picture) της περιοχής των ματιών

2. Παραγωγή των τμημάτων με τη βοήθεια του Gabor. Όπως ήδη περιγράψαμε, ο μετασχηματισμός χρησιμοποιείται, προκειμένου να βρεθούν οι σημαντικές περιοχές στην εικόνα. Σε αυτό το βήμα εφαρμόζεται στη περιοχή των ματιών. Ως αποτέλεσμα του Gabor Wavelet λαμβάνουμε την ακόλουθη δυαδική εικόνα.



Εικόνα 41: Κατάτμηση των ματιών με χρήση Gabor

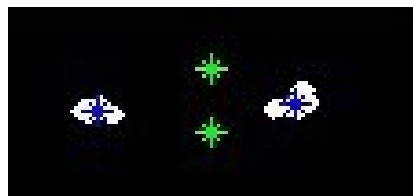
Όπως βλέπουμε από την εικόνα ακρών, κι άλλες περιοχές είναι επίσης έντονα ευδιάκριτες. π.χ. Τα μαλλιά και τα φρύδια βρίσκονται επανειλημμένα ως τμήματα. Προκειμένου να κατασταλούν αυτές οι περιοχές τώρα, η εικόνα επεξεργάζεται.

3. Hole-Filling, Reconstruction and Closing. Αυτό το βήμα για να μπορέσει να εφαρμοστεί, πρέπει τα δυο διαφορετικά αποτελέσματα του πρώτου και δεύτερου βήματος (edge picture και η κατάτμηση της εικόνας με φίλτρο Gabor), να ενωθούν.



Εικόνα 42: Ένωση της edge picture με φίλτρο Gabor

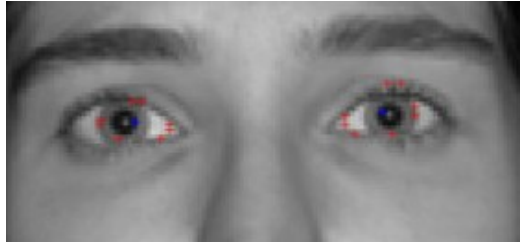
Με βάση αυτήν την εικόνα θα γίνει προσπάθεια για να κλείσουμε τις εσοχές π.χ. Με τη βοήθεια της μεθόδου Closing η οποία θα πραγματοποιηθεί μέσα στα όρια που ορίζει η ενωμένη edge picture. Στο επόμενο βήμα η εικόνα ακρών θα αποκοπεί και θα χρησιμοποιηθεί στην ανακατασκευή της εικόνας. Με βάση αυτό, οι περιοχές που δεν συνδέονται με τα μάτια θα γεμιστούν (filled). Έτσι εξαφανίζουμε περιοχές όπως τα φρύδια και τα μαλλιά. Τέλος, διεκπεραιώνεται εκ νέου ένα Close για να χωρισθούν οι περιοχές που ενδεχομένως συνάπτηκαν. Βάση αυτών των περιοχών θα πραγματοποιηθεί το Postprocessing. Στην ιδανική περίπτωση βρίσκονται δύο περιοχές που αντιπροσωπεύουν τα μάτια. Υπάρχει όμως περίπτωση τα φρύδια ή άλλες μικρότερες και ασήμαντες περιοχές να εμφανιστούν στην εικόνα.



Εικόνα 43: Ιδανική κατάτμηση στα μάτια

Μεταφέροντας στην αρχική εικόνα τα ακόλουθα αποτελέσματα, βρίσκουμε αυτές τις περιοχές όπως φαίνεται στην Εικόνα 44:





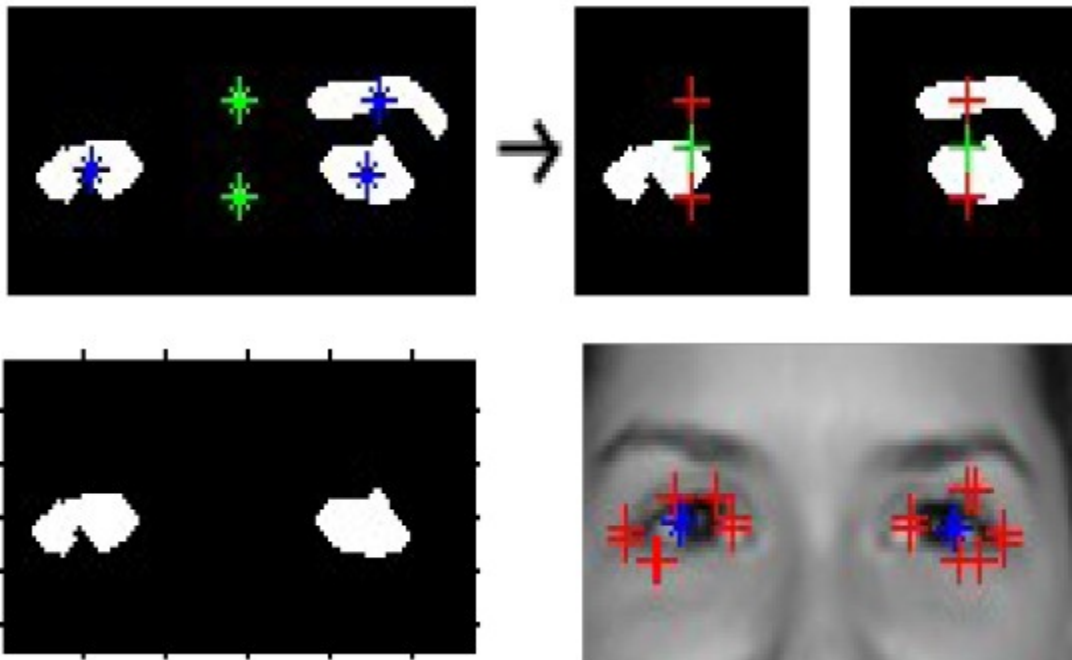
Εικόνα 44: Εύρεση σημαντικών περιοχών στην αρχική εικόνα

Η ακόλουθη προσέγγιση χρησιμοποιείται, προκειμένου να έχουμε επιτυχή εύρεση ματιών:

1. Βάση της επιφάνειας των χαρακτηριστικών αποβάλλονται οι μικρές περιοχές
2. Στην περίπτωση που μόνο ένα μάτι βρεθεί, τότε αυτό καθρεφτίζεται στη δεύτερη πλευρά, δεδομένου ότι αποδεχόμαστε, ότι και το άλλο μάτι βρίσκεται περίπου στην ίδια θέση
3. Υπάρχουν ακόμα περισσότερες από 2 περιοχές διαθέσιμες. Έτσι προκύπτουν τα ακόλουθα:

### Εύρεση ματιών με πολλαπλές περιοχές

Σαν αρχή μπορεί να θεωρηθεί ότι τα μάτια βρίσκονται στο δεξιό και αριστερό μέρος της εικόνας περίπου στο ίδιο ύψος, εκτός κι αν το άτομο κλίνει το κεφάλι του λοξά. Αν βρεθούν περισσότερες από δύο περιοχές, τότε η εικόνα διαιρείται σε αριστερό και δεξιό μέρος και κάθε ένα από αυτά τα δύο μέρη εξετάζονται. Αν ένα από τα δύο μέρη έχει μόνο μια περιοχή, τότε αυτή θεωρείται ότι είναι το μάτι. Έπειτα οι συντεταγμένες αυτών των περιοχών παρέχονται προκειμένου να βρεθεί και στο απέναντι μέρος, στην ίδια θέση, το μάτι. Από τις εναπομείνοντες περιοχές της απέναντης πλευράς, θα προτιμηθεί αυτή που έχει τη μικρότερη ευκλείδεια απόσταση από το κέντρο.



Εικόνα 45: Περισσότερες από δυο περιοχές βρέθηκαν στην αρχική εικόνα.

Απαιτείται περαιτέρω επεξεργασία.

Εάν βρεθούν και στο αριστερό και στο δεξιό μέρος περισσότερες από μια περιοχές, κατόπιν επιλέγεται εκείνη η περιοχή, η οποία εκθέτει τη μικρότερη απόσταση από το κέντρο αυτού του μέρους. Δύο τιμές θα επιστραφούν, “0” εάν καμία περιοχή δεν μπόρεσε να βρεθεί, “2” εάν οι περιοχές βρέθηκαν.

### **Εύρεση περιοχή στόματος (τρόπος λειτουργίας της συνάρτησης *extractMouth*)**

Η πλήρης και σωστή κατάτμηση του στόματος αντιπροσωπεύει το σημαντικότερο μέρος του *segmenting*, δεδομένου ότι από το στόμα έχουμε την καλύτερη αντίληψη όσον αφορά το συναίσθημα. Σαν εικόνα εισαγωγής για την εξαγωγή αυτού του χαρακτηριστικού, λαμβάνεται το χαμηλότερο μέρος της αρχικής εικόνας.

#### **1. Edge picture provide:**

Το πρώτο βήμα είναι η παραγωγή μιας εικόνας ορίων από την αρχική εικόνα εισαγωγής.



*Εικόνα 46: Edge picture της αρχικής εικόνας*

#### **2. Παραγωγή των τμημάτων (segments) με τη βοήθεια του Gabor :**

Η αφετηρία για την κατάτμηση του στόματος αντιπροσωπεύεται από το μετασχηματισμό του Gabor Wavelet.



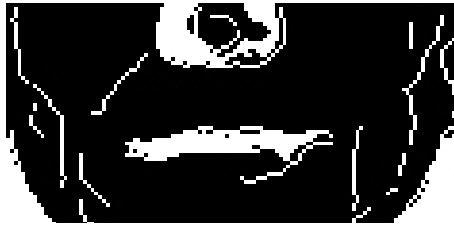
*Εικόνα 47: Αποτέλεσμα του Gabor Wavelet transformation*

Στο αποτέλεσμα του μετασχηματισμού του Gabor επιπλέον εφαρμόζεται η μέθοδος Closing για να κλείσει τις μικρότερες τρύπες στις περιοχές.

#### **3. Fill Holes:**

Μπορεί η περιοχή του στόματος να περιέχει επιπλέον ανοίγματα ή τρύπες (holes), η edge picture που παρήχθηκε στο βήμα 1, τίθεται και κατόπιν χρησιμοποιείται η μέθοδος *get Filling*. Κατά τη διάρκεια αυτής της διαδικασίας γεμίζει ολόκληρη η εικόνα, εάν οι άκρες διαμορφώνουν έναν πλήρη κύκλο. Μερικά pixels γύρω από το όριο αποκόπτονται και κατά συνέπεια η εικόνα γίνεται μικρότερη.





Εικόνα 48: Ενδιάμεσο αποτέλεσμα με την edge picture



Εικόνα 49: Filled hole, μειωμένη, εικόνα που έχει υποστεί περικοπές

#### 4. Ανακατασκευή

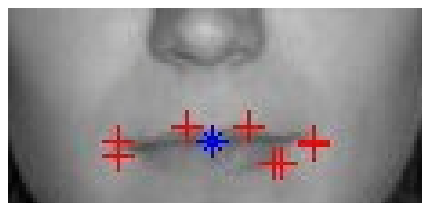
Η διαδικασία για αυτό το βήμα περιλαμβάνει την εξάλειψη των ακρών με σκοπό την αναδημιουργία της εικόνας κι έτσι οι σημαντικές περιοχές διαγράφονται. Περαιτέρω κατά την διάρκεια της εκτέλεσης αποκόπτονται μόνο τα pixel εκείνα που βρίσκονται μακριά από το κέντρο της εικόνας. Κατά συνέπεια με βάση τις πιθανότητες αποτρέπεται η διαγραφή της περιοχής που μας ενδιαφέρει, δηλαδή το στόμα.



Εικόνα 50: Αποτέλεσμα της ανακατασκευής

#### 5. Αποτέλεσμα

Στο τελευταίο βήμα, βρίσκεται και επιλέγεται το κέντρο της εικόνας κι έπειτα επιστρέφεται ως αποτέλεσμα.



Εικόνα 51: Αποτέλεσμα της εξαγωγής

Εντούτοις το στόμα αποτελείται από διάφορα μέρη κι έτσι σαν τιμή επιστροφής, μπορεί να έχουμε μόνο ένα μικρό μέρος, π.χ. τη γωνία ενός στόματος που γελάει. Προκειμένου να συνδεθεί αυτό με τα άλλα μέρη, μια γραμμή τίθεται που περνά από το κέντρο και καταλήγει στην άλλη πλευρά της εικόνας. Τέλος, επαναλαμβάνεται το βήμα 3: “Fill Holes” ώστε το στόμα να αποτελείται τελικά από ένα τμήμα.



Εικόνα 52: Εισαγωγή γραμμής στην εικόνα για σύνδεση απομονωμένων περιοχών

### Εύρεση περιοχή μύτης (τρόπος λειτουργίας της συνάρτησης *extractNose*)

Η κατάτμηση (segmenting) της μύτης διαδραματίζει έναν μικρότερο ρόλο για τον υπολογισμό των χαρακτηριστικών γνωρισμάτων. Χρησιμοποιεί μόνο ως το σημείο αναφοράς, π.χ. για τον υπολογισμό της διαφοράς μεταξύ των γωνιών του στόματος και του κέντρου της στοματικής περιοχής. Στην εικόνα εισαγωγής, όπως με την εύρεση του στόματος (το οποίο βρίσκεται στο χαμηλότερο μέρος της αρχικής εικόνας) υποθέτουμε ότι το σημείο της μύτης πρόκειται να βρεθεί στο ανώτερο σημείο του χαμηλότερου μέρους της αρχικής εικόνας.

#### 1. Edge picture

Παρέχεται η εικόνα ακρών, η οποία αντιπροσωπεύει την εικόνα εισαγωγής και είναι πολύ χρήσιμη για την κατάτμηση, δεδομένου ότι οι περιοχές μπορούν να διαχωριστούν. Για να ενισχύσουμε ή και να κάνουμε παχύτερα τα όρια στην εικόνα ακρών, εφαρμόζεται η μέθοδος Closing γύρω από τις γραμμές.

#### 2. Παραγωγή τμημάτων με τη βοήθεια του Gabor

Στην εικόνα εισαγωγής εφαρμόζεται ο μετασχηματισμός του Gabor Wavelet

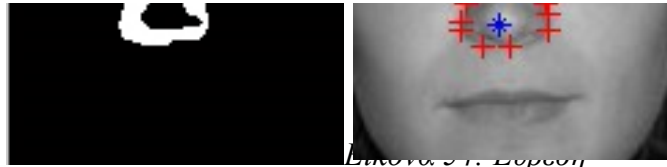


Εικόνα 53:  
Αποτέλεσμα του Gabor Wavelet transformation

Η μέθοδος Closing εφαρμόζεται στην μετασχηματισμένη εικόνα για να καλυφθούν τα μικρότερα ανοίγματα και οι τρύπες.

#### 3. Ανίχνευση της μύτης στην εικόνα

Όπως ήδη αναφέραμε, η μύτη βρίσκεται στο κέντρο του άνω μέρους της κατώτερης περιοχής της εικόνας. Επομένως ο αλγόριθμος ανιχνεύει την εικόνα για αυτή την περιοχή, κι έπειτα εξαπλώνεται ομοιόμορφα δεξιά κι αριστερά. Η εύρεση της περιοχής επιστρέφεται ως αποτέλεσμα.



περιοχής της μύτης

#### 4. Επανάληψη

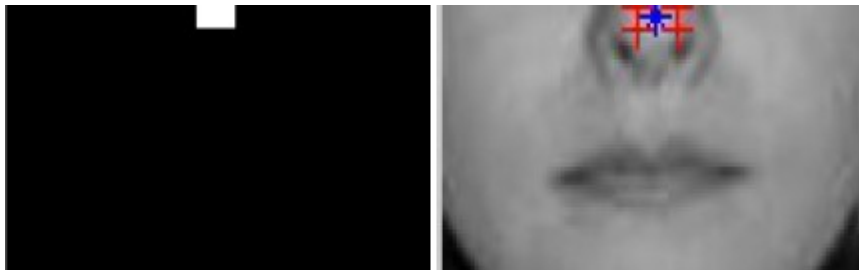
Καμιά φορά η μύτη συνδέεται με άλλες περιοχές και η επιφάνεια της περιοχής που εξετάζεται υπερβαίνει ένα ορισμένο μέγεθος ή το κέντρο της περιοχής βρίσκεται πάρα πολύ μακριά από το κέντρο της εικόνας. Εάν αυτό συμβεί, η εικόνα ακρών που παρέχεται στο βήμα 1 αφαιρείται από την εικόνα και πραγματοποιείται πάλι η ανίχνευση της περιοχής της μύτης.



Εικόνα 55: Διαίρεση της μεγάλης περιοχής με τη βοήθεια της αφαίρεσης της edge picture

#### 5. Default nose

Εάν καμία περιοχή μύτης δεν βρεθεί, επιστρέφεται μια προκαθορισμένη τιμή “default nose”. Ένα μικρό ορθογώνιο εμφανίζεται στο ανώτερο διάστημα.



Εικόνα 56: Default nose

Οι παραπάνω αλγόριθμοι είναι αυτοί που θα επιβεβαιώσουν την ύπαρξη ενός προσώπου μέσα στην εικόνα. Σαν προαπαιτούμενο όμως εμείς, πέρα από την ύπαρξη ενός και μόνο προσώπου μέσα στο καρέ που προορίζεται για εξέταση, είναι και η καταλληλότητα του εμφανιζόμενου αυτού προσώπου για ανάλυση. Με την εφαρμογή των αλγορίθμων αυτών καταφέρνουμε να επιβεβαιώσουμε εάν υπάρχουν τα χαρακτηριστικά που απαιτούμε για την ανάλυση του προσώπου.

### 4.3 Βάση Δεδομένων

Οι αλγόριθμοι που θα χρησιμοποιηθούν προϋποθέτουν την ύπαρξη μιας βάσης δεδομένων που θα καθορίζει την διαφορά ανάμεσα σε ευτυχισμένα πρόσωπα και όχι. Σε αυτό το σημείο πρέπει να αναφέρουμε για ακόμα μια φορά την δυσκολία του στόχου που θέλουμε να επιτευχθεί μέσω αυτής της εργασίας. Υπάρχουν πολλές βάσεις δεδομένων με τις οποίες ένας υπολογιστής μπορεί να εκπαιδευτεί για την ανίχνευση των ανθρώπινων συναισθημάτων που εκφράζονται μέσω του προσώπου. Όμως είναι δυσεύρετες

αυτές που αφορούν τις εκφράσεις του προσώπου ενός ερμηνευτή. Αξίζει να αναφερθεί ότι στην μουσική ένα νέο σύνολο κανόνων τίθεται για την κατηγοριοποίηση των εκφράσεων. Η δυσκολία ακόμα και για τον ανθρώπινο εγκέφαλο να καταλάβει ορθά μέσα και μόνο από μια φωτογραφία ενός μουσικού ερμηνευτή το συναίσθημα του εκείνη την χρονική στιγμή είναι μεγάλη.



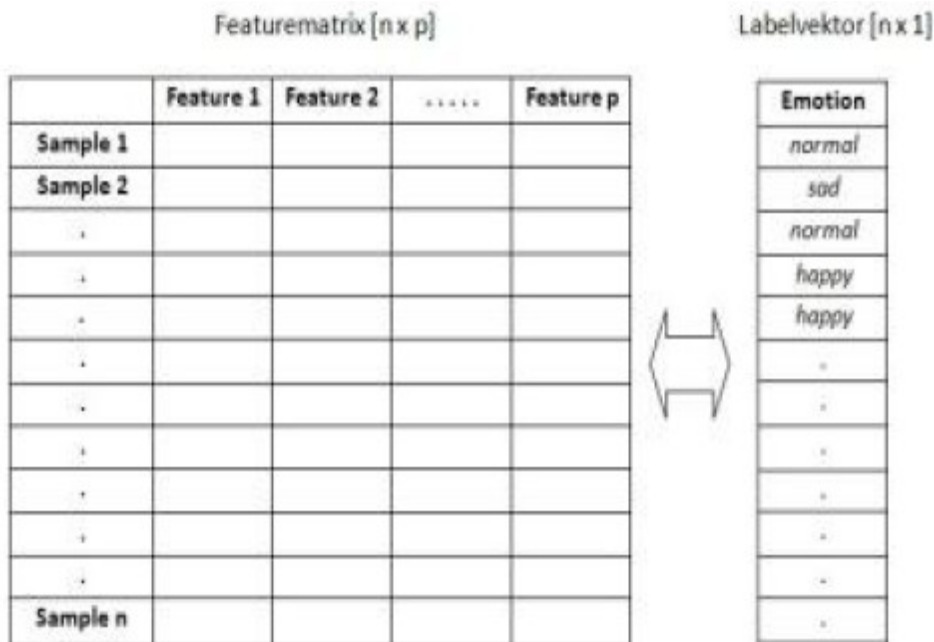
Εικόνα 57: Εικόνες προσώπου που αποπροσανατολίζουν τους αλγορίθμους από την ορθή απόφαση

Στην εικόνα 57 παρουσιάζονται μια σειρά από πρόσωπα που μπορούν να μπερδέψουν έναν αλγόριθμο και σαν αποτέλεσμα αυτού να λάβουμε λάθος αποτέλεσμα. Το γεγονός ότι οι ερμηνευτές όταν τραγουδάνε έχουν ανοικτό το στόμα τους, κάνει αρκετούς αλγόριθμους, οι οποίοι δεν έχουν εκπαιδευθεί σωστά, να βγάζουν το συμπέρασμα ότι πρόκειται για την έκφραση του χαμόγελου. Όμως αυτό δεν αληθεύει στις περισσότερες περιπτώσεις. Τα τρία πρώτα πρόσωπα μπορεί όντως να προέρχονται από μουσικά βίντεο κλιπ τα οποία χαρακτηρίζονται από το συναίσθημα της χαράς, όμως η τρίτη εικόνα του προσώπου αφορά ένα καθαρά μουσικό κομμάτι λύπης. Αυτό που προδίδει την πραγματική συναισθηματική κατάσταση του ερμηνευτή είναι η κατεύθυνση των ματιών του.

Παραθέτοντας αυτές τις εικόνες σαν ένα υποσύνολο παρόμοιων δειγμάτων που λάβαμε έπειτα από μια σειρά από εξέταση προσώπων την ώρα που ερμήνευαν τόσο χαρούμενα όσο και λυπημένα κομμάτια, γίνεται εμφανής η ανάγκη για την δημιουργία μιας νέας βάσης δεδομένων που θα περιέχει πρόσωπα μουσικών καλλιτεχνών την ώρα που ερμηνεύουν διαφορά τραγούδια. Έπειτα από μια αναζήτηση στο διαδίκτυο και σε διάφορες εγκεκριμένες βάσεις δεδομένων, διαπιστώθηκε ότι δεν υπάρχει μια τέτοια βάση δεδομένων, όχι τουλάχιστον την χρονική περίοδο που εξετάστηκε. Για αυτόν τον λόγο με αρκετά προσεκτικές κινήσεις επιλέχθηκαν δέκα ερμηνευτές ως πρότυπα προσώπων – τόσο γυναίκες όσο και άνδρες – και λήφθηκαν από μουσικά βίντεο κλιπ τους που ήταν διαθέσιμα στο διαδίκτυο δέκα στιγμιότυπα με εμφανή τα συναισθήματα τους στο πρόσωπο τους. Σαν αποτέλεσμα είχαμε την δημιουργία μιας βάσης δεδομένων – σχετικά μικρού όγκου αφού περιείχε εκατό εικόνες – αρκετά ικανοποιητικής ώστε να μπορέσουμε να εκπαιδύσουμε με αυτήν το πρόγραμμα μας και τους αλγορίθμους μας για την πιο ορθή ανίχνευση των συναισθημάτων.

#### 4.4 Ταξινόμηση Συναισθημάτων

Αφού λοιπόν λύσαμε το θέμα της βάσης δεδομένων δημιουργήσαμε το training αρχείο δεδομένων, το οποίο λόγω πειραματικού σταδίου περιείχε μόνο τα δύο έντονα συναισθήματα, αυτό της χαράς και αυτό της λύπης.

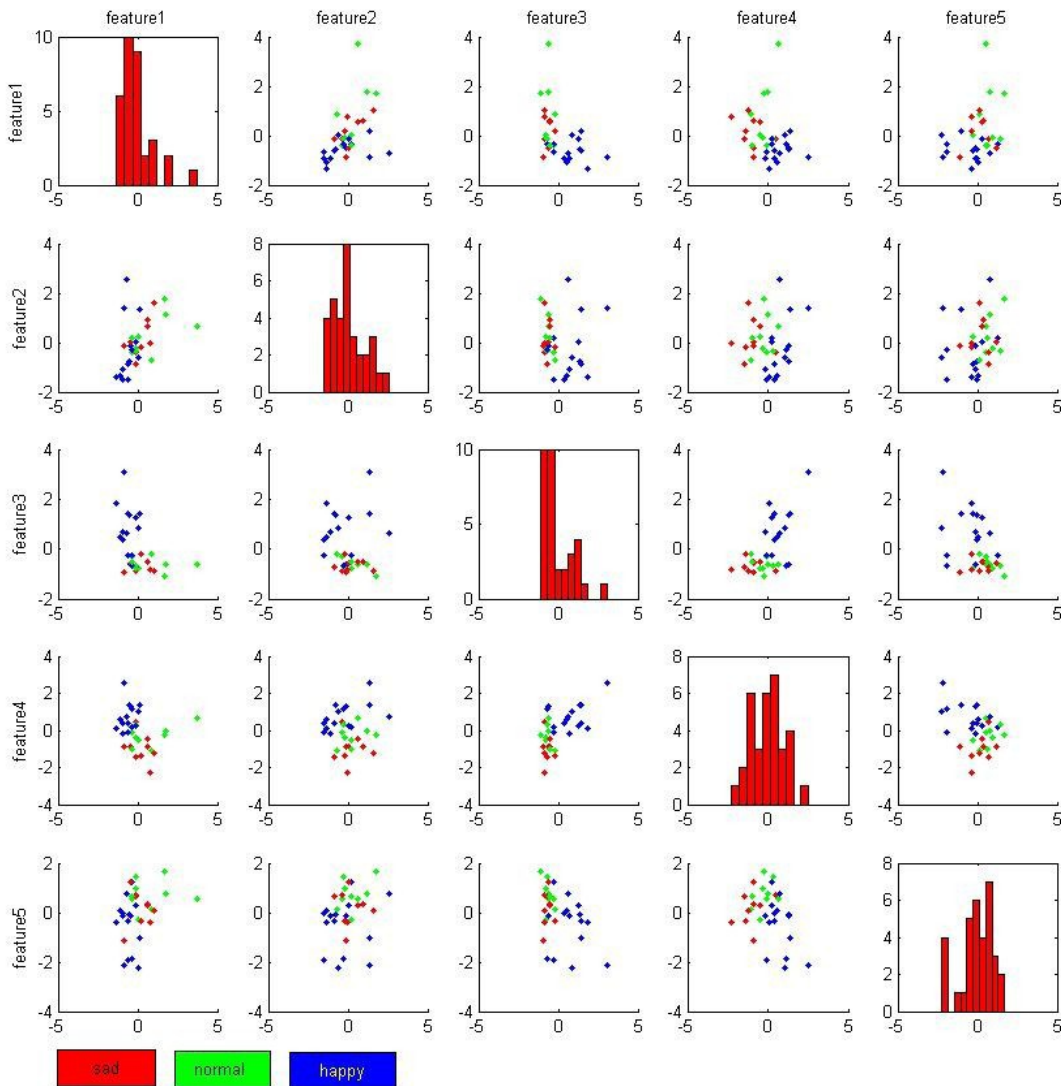


Πίνακας 4: Πίνακας Featurematrix[nxp] και Labelvektor[nx1]

### Optical analysis:

Η οπτική ανάλυση εξετάζει τα διαφορετικά χαρακτηριστικά γνωρίσματα γύρω από την καταλληλότητά τους, για την ταξινόμηση. Ακολουθεί:

- Καλή σύνδεση με την αντίστοιχη κλάση (emotion) καθώς επίσης και ο διαχωρισμός των κλάσεων.
- Αφαίρεση των ισχυρών γραμμικών συνδέσεων μεταξύ των χαρακτηριστικών γύρω από τους πιθανούς πλεονασμούς για να απλοποιήσουμε τη χαρακτηριστική περιοχή. Συγχρόνως επίσης πραγματοποιείται η κλιμάκωση ή η τυποποίηση των χαρακτηριστικών με μέση τιμή 0 και σταθερή απόκλιση 1 γύρω από τα μεμονωμένα χαρακτηριστικά (απαραίτητη μέθοδος για μερικές ταξινομήσεις).



Πίνακας 5: Scatterplot των τελικά επιλεγμένων χαρακτηριστικών γνωρισμάτων

- **Feature 1:** Σχέση μεταξύ του επιπέδου του ματιού και του πλάτους του αριστερού ματιού
- **Feature 2:** Σχέση μεταξύ του επιπέδου του ματιού και του πλάτους του δεξιού ματιού
- **Feature 3:** Σχέση επιφάνειας της έκτασης του στόματος
- **Feature 4:** Μέση τιμή της απόστασης μύτης – κέντρο στόματος και οι αποστάσεις μύτης -γωνίες στόματος
- **Feature 5:** Βαθμός σγέσης του πλάτους του στόματος

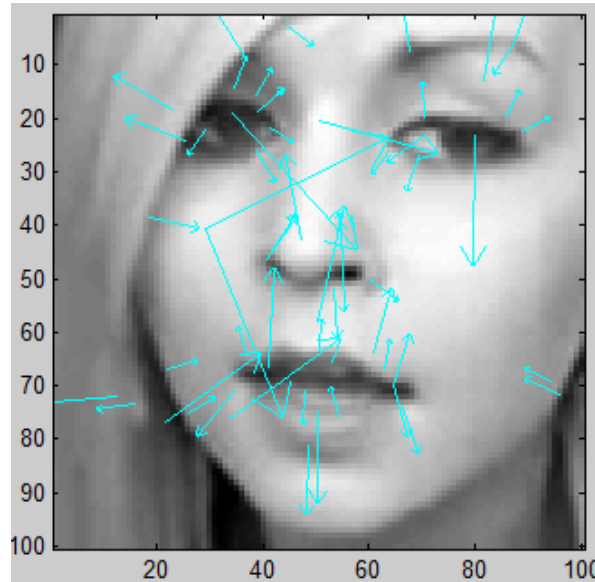
Η κανονική έκφραση (πράσινο) είναι έντονα επικαλυμμένη από την λύπη (κόκκινο) κι έτσι είναι δύσκολο να διαφοροποιηθεί. Αυτή η εντύπωση επιβεβαιώθηκε αργότερα από τα αποτελέσματα.

## Ταξινόμηση μέσω SIFT-based Neural Network

Όπως αναφερθήκαμε στο προηγούμενο κεφάλαιο, για την ταξινόμηση των συναισθημάτων θα χρησιμοποιήσουμε 5 μεθόδους, που η χρήση του ήδη συνιστάται για αυτόν τον σκοπό, και ακόμα μια μεθοδολογία η οποία δεν έχει δοκιμαστεί ακόμα για αυτό τον τομέα της βιομετρικής.

Ας παραλείψουμε εσκεμμένα σε αυτό το κεφάλαιο τις γνωστές μεθόδους, καθώς το μόνο που μπορούμε να δούμε είναι κατά πόσο είναι αποτελεσματικοί ή όχι σε αυτούς τους νέους κανόνες που θέτονται και ας εξετάσουμε το τρόπο λειτουργίας αλλά τα προβλήματα που προέκυψαν από την εφαρμογή του αλγορίθμου SIFT για την αναγνώριση των εκφράσεων του προσώπου.

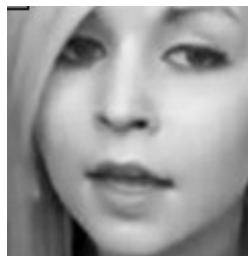
Όπως προαναφέραμε στο προηγούμενο κεφάλαιο (κεφ.3 παράγραφος 3.4.6.1) ο αλγόριθμος SIFT αναζητά και εντοπίζει χαρακτηριστικά γνωρίσματα, τα οποία είναι αμετάβλητα από το μέγεθος και την περιστροφή τους μέσα σε μια εικόνα. Αυτό συνεπάγεται ότι είναι κατάλληλος να βρει κοινά στοιχεία μέσα σε δυο εικόνες. Θα μπορούσε να είναι – και όπως αποδεικνύεται είναι – ένας αρκετά ικανός αλγόριθμος με χρήση στην αναγνώριση προσώπων. Όμως μπορεί να ξεπεράσει κάποια προβλήματα που αντιμετωπίζει και να εντοπίζει τη γενική μορφή ενός χαμόγελου και όχι ακριβώς το ίδιο χαμόγελο; Δίνοντας στον αλγόριθμο SIFT την φωτογραφία προσώπου, όπως μπορούμε να παρατηρήσουμε και στην παρακάτω εικόνα, λαμβάνονται διάφορα keypoints τα οποία δεν είναι όλα το ίδιο ενδιαφέρον για τον σκοπό που χρησιμοποιούμε αυτόν τον αλγόριθμο.



Εικόνα 58: SIFT keypoints

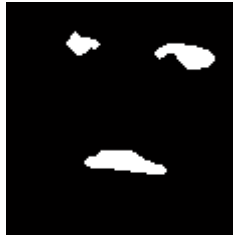
Στην εικόνα αυτή παρατηρούμε ότι ενδιαφέρον keypoints λαμβάνουμε μόνο από τις περιοχές του στόματος και των ματιών. Σαν αποτέλεσμα αυτού είναι τα περιττά χαρακτηριστικά γνωρίσματα να αποπλανούν στην συνέχεια την εκπαίδευση του νευρωνικού δικτύου που θα χρησιμοποιήσουμε για την αναγνώριση συναισθημάτων. Οπότε σκοπός μας είναι να βρούμε έναν τρόπο να απομονώσουμε τις περιοχές αυτές που περιέχουν τα ενδιαφέροντα χαρακτηριστικά γνωρίσματα. Όμως πρέπει να λάβουμε υπόψιν ότι το πρόγραμμα που έχουμε κατασκευάσει μέχρι στιγμής έχει ήδη πολλές διεργασίες να διεξάγει. Οπότε θα ήταν καλό να βρούμε έναν αποτελεσματικό αλλά απλό και γρήγορο τρόπο για να απομονώσουμε τις επιθυμητές περιοχές. Εάν ανατρέξουμε στις προηγούμενες διεργασίες που αφορούν την ανίχνευση προσώπου και την κρίση καταλληλότητας του για ανάλυση παρατηρούμε ότι ήδη έχουμε εντοπίσει το εάν υπάρχουν δύο μάτια, μια μύτη και ένα στόμα. Οπότε το μόνο που μας μένει είναι να 'κόψουμε' αυτές τις περιοχές από την κύρια εικόνα και να δημιουργήσουμε μια νέα βάση δεδομένων που θα περιέχει μόνο τις περιοχές αυτές των προσώπων.

Κάθε συνάρτηση που χρησιμοποιήθηκε για την ανίχνευση των χαρακτηριστικών του προσώπου επιστρέφει έναν πίνακα- εικόνα με binary τιμές, ο οποίος προσδιορίζει την θέση του κάθε χαρακτηριστικού. Αυτό που μας λείπει πλέον είναι να πολλαπλασιάσουμε τους πίνακες αυτούς με τον πίνακα της εικόνας μας.



Εικόνα 59: Αρχική εικόνα

Στην εικόνα 59 έχουμε τη αρχική εικόνα προσώπου, την οποία έχουμε εντοπίσει σε ένα καρέ ενός βίντεο κλιπ. Αυτό συνεπάγεται ότι έχει κριθεί κατάλληλη για ανάλυση αφού εντοπίστηκαν μέσα σε αυτήν τα βασικά χαρακτηριστικά ενός προσώπου. Έχοντας λάβει τις θέσεις των χαρακτηριστικών αυτών δημιουργούμε μια εικόνα σαν αυτή που παρουσιάζεται(εικόνα 60) .



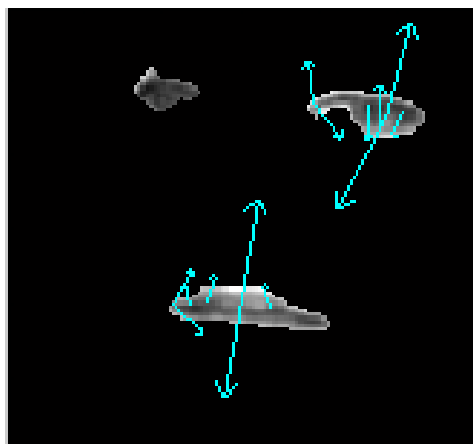
Εικόνα 60:  
Θέσεις των  
χαρακτηριστικών

Μέσω των πολλαπλασιασμού των δύο αυτών εικόνων με την εντολή `BW= uint8(BW1).* uint8(BW)`; καταφέρνουμε να λάβουμε μόνο τις ζητούμενες περιοχές, αποβάλλοντας έτσι τα επιπλέον χαρακτηριστικά τα οποία περιέχουν περιττή πληροφορία για το σύστημα μας. Στην εικόνα 61 παρουσιάζεται το τελικό αποτέλεσμα αυτού του πολλαπλασιασμού, ενώ στην 62 τα αποτελέσματα που λαμβάνουμε όταν στην τελική εικόνα εφαρμόσουμε τον αλγόριθμο SIFT.



Εικόνα 61: Τελική εικόνα

Όπως παρατηρούμε στην τελική εικόνα προς εξέταση, τα βασικά χαρακτηριστικά του προσώπου δεν είναι στο ακέραιο ορθά επιλεγμένα. Όμως αυτό επηρεάζεται κατά πολύ και από την ποιότητα της εικόνας αλλά και από της συνθήκες που επικρατούν μέσα στο ίδιο το βίντεο – όπως είναι το εάν τα μαλλιά του ερμηνευτή κρύβουν εν μέρη κάποια χαρακτηριστικά του, εάν φοράει γυαλιά, ακόμα και ο φωτισμός που επικρατεί – χωρίς όμως να μας αποθαρρύνει για το εάν είναι λειτουργικός ο τρόπος με τον οποίο υλοποιείτε αυτή η ιδέα, βλέποντας και κρίνοντας από την επόμενη εικόνα.





Στην εικόνα 58 είδαμε ότι τα χαρακτηριστικά γνωρίσματα που λαμβάνουμε από τον αλγόριθμο SIFT βασίζονται σε ολόκληρη την εικόνα περιείχε πολλές άχρηστες πληροφορίες. Αντίθετα όμως με την υλοποίηση της προαναφερόμενης ιδέας, βλέπουμε στην εικόνα 62 ότι τα χαρακτηριστικά γνωρίσματα που λαμβάνουμε είναι εμφανέστερα πιο ενδιαφέροντα. Στις περιοχές όπου η πληροφορία έχει αποκλειστεί με την χρήση του μαύρου χρώματος, ο αλγόριθμος SIFT δεν ασχολείται με το να βρει χαρακτηριστικά γνωρίσματα. Έτσι επικεντρώνετε στις περιοχές που του δίνουμε και τα αποτελέσματα, τουλάχιστον θεωρητικά, είναι αυτά που επιθυμούμε ώστε να εκπαιδεύσουμε ένα νευρωνικό δίκτυο για την αναγνώριση συναισθημάτων μέσω των εκφράσεων ενός προσώπου.

Για την δημιουργία του νευρωνικού δικτύου θα χρησιμοποιήσουμε την ίδια βάση δεδομένων, για να εξάγουμε τα χαρακτηριστικά γνωρίσματα, που χρησιμοποιήσαμε και στις προηγούμενες μεθόδους.



Εικόνα 63: Δείγμα βάσεις δεδομένων προσώπων από μουσική

Για κάθε εικόνα γίνεται απομόνωση των σημείων ενδιαφέροντος και στην συνέχεια πραγματοποιείτε η ανίχνευση των χαρακτηριστικών γνωρισμάτων SIFT. Αυτά τα διανύσματα αποθηκεύονται μέσα σε έναν συγκεντρωτικό πίνακα, ο οποίο στο τέλος περιέχει όλα τα διανύσματα. Ένας δεύτερος πίνακας δημιουργείτε ο οποίος χαρακτηρίζει με μηδέν και ένα το σκοπό του κάθε διανύσματος του πρώτου πίνακα. Αυτοί οι δύο πίνακες ουσιαστικά είναι τα στοιχεία τα οποία παρέχουμε στο νευρωνικό δίκτυο μας για την εκπαίδευση του.

```
d=size(descrips);
y=y+d(1,1);

for a=y-d(1,1):y-1
    if i<=45
        motion(a,1)=0;
    else
        motion(a,1)=1;
    end
    for z=1:128

        keypoints(a,z)=descrips(u,z);

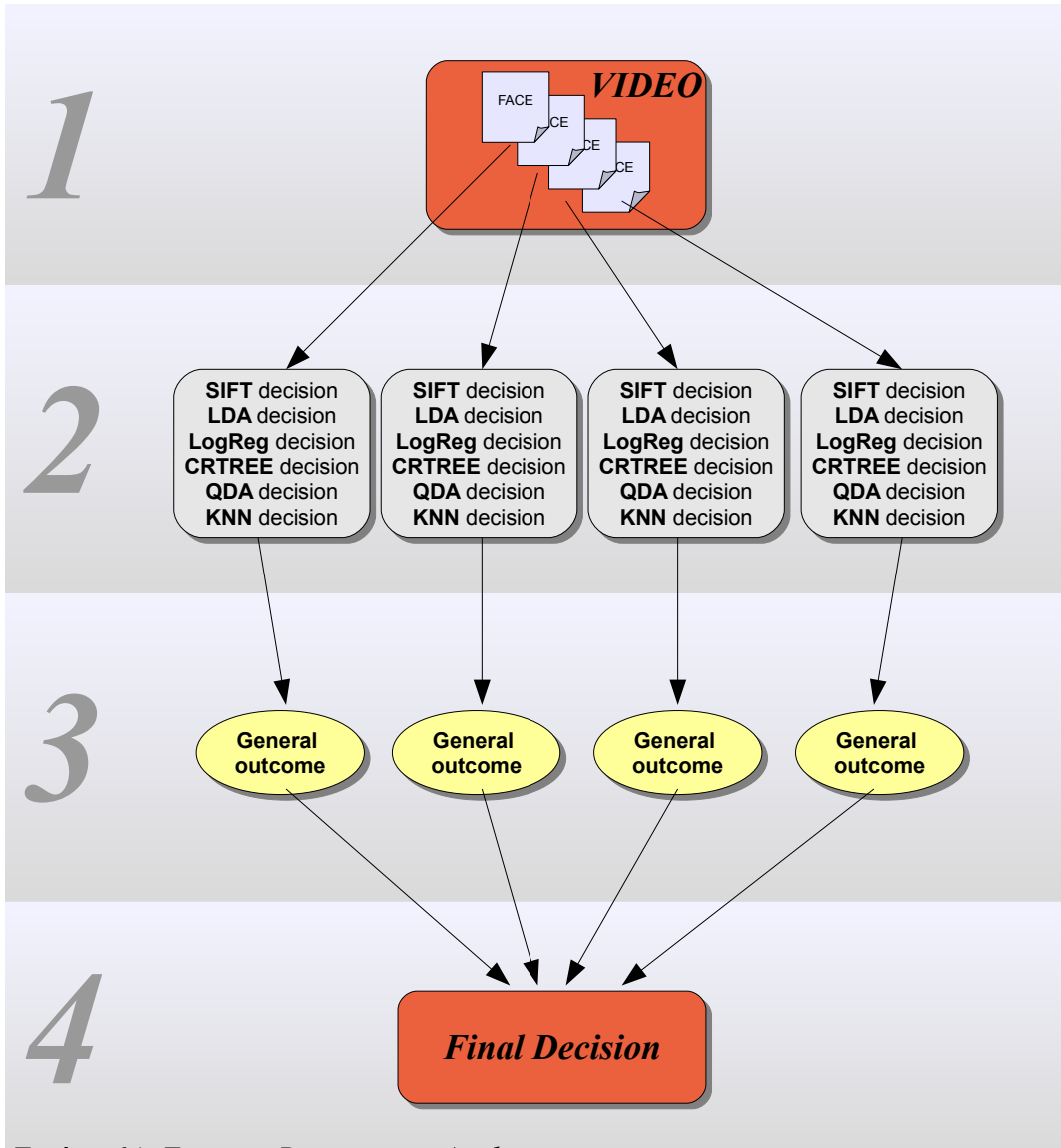
    end
    u=u+1;
end
```

Μέσω του κώδικα που παρουσιάζεται παραπάνω δημιουργούνται οι δύο πίνακες. Οι 45 πρώτες εικόνες είναι αυτές που περιέχουν το συναίσθημα της λύπης και οι υπόλοιπες το συναίσθημα της χαράς. Στο τέλος του υπολογισμού και των 90 εικόνων οι δύο πίνακες αποθηκεύονται σε ένα .mat αρχείο το οποίο θα κλιθεί από το νευρωνικό δίκτυο μόνο για μια φορά για να εκπαιδευτεί.

Για την δημιουργία του νευρωνικού δικτύου μας χρησιμοποιούμε την εντολή nftool μέσω της οποίας τροφοδοτούμε την εκπαίδευση με τα inputs, τα χαρακτηριστικά γνωρίσματα του SIFT, και τα targets, τον χαρακτηρισμό δηλαδή κάθε διανύσματος με μηδέν ή ένα. Μέσω μιας σειράς δοκιμών που γίνονται κατά την εκπαίδευση αρκετές παράμετροι τροποποιούνται ώστε το performance του δικτύου να είναι το βέλτιστο δυνατό.

## Συνδυασμός μεθόδων και τελική απόφαση

Έχοντας θέσει και το νευρωνικό δίκτυο σε λειτουργία, μπορούμε να χρησιμοποιήσουμε 6 μεθόδους για να διασταυρώσουμε το αποτέλεσμα της ανάλυσης συναισθημάτων που θα λάβουμε από κάθε μέθοδο. Η απόφαση για το συναίσθημα κάθε εικόνας θα είναι ένα κριτήριο για την τελική απόφαση.



Εικόνα 64: Emotion Recognition Analysis

Για κάθε καρτέ που λαμβάνουμε από το βίντεο(βλέπε εικόνα 64) θα εξάγεται, έπειτα από την ανάλυση του, μια τελική απόφαση που αφορά το συγκεκριμένο καρτέ και μόνο. Αυτή η διαδικασία θα επαναλαμβάνεται δέκα φορές. Έχοντας πλέον στα χέρια μας ένα δείγμα από δέκα συναισθήματα που βρίσκονται μέσα στο βίντεο κλιπ μας, τότε μπορούμε να δώσουμε ποσοστιαία τα συναισθήματα που επικρατούν σε αυτό.

Για την λήψη της απόφασης για κάθε καρτέ ξεχωριστά, καλούμαστε να λάβουμε υπόψιν τα αποτελέσματα από έξι μεθόδους. Έπειτα από μια σειρά δοκιμών που έχουν διεξαχθεί στην ανίχνευση συναισθημάτων στον τομέα που εξετάζουμε καταλήγουμε στο συμπέρασμα ότι το συναίσθημα της λύπης είναι πιο δύσκολο να εντοπιστεί από τους αλγόριθμους που χρησιμοποιούμε – πράγμα που θα δούμε και στο επόμενο κεφάλαιο – οπότε και του δίνουμε έναν συντελεστή βαρύτητας μεγαλύτερο από αυτό που έχει το συναίσθημα της χαράς. Οπότε για κάθε καρτέ εφαρμόζουμε την εξής εξίσωση:

$$\text{Rate of Happiness} = [(e_1 * W_{r_1}) + (e_2 * W_{r_2}) + \dots + (e_n * W_{r_n})] / n$$

Όπου  $e$  είναι το αποτέλεσμα που λαμβάνουμε από κάθε μέθοδο ξεχωριστά – 1 για το συναίσθημα της χαράς και 0 για το συναίσθημα της λύπης –  $W_r$  είναι η συντελεστής βαρύτητας του κάθε συναισθήματος – 0,6 για το συναίσθημα της χαράς και 1 για το συναίσθημα της λύπης – και  $n$  είναι ο αριθμός των αποτελεσμάτων που λαμβάνουμε υπόψιν για την λήψη της τελικής απόφασης για κάθε καρέ. Ας εξετάσουμε ένα παράδειγμα:

Methods	Results (1 for happy/ 0 for sad)	Weighting Factor( in %)
LDA	1	60
LogReg	1	60
CRtree	0	100
QDA	1	60
kNN	0	100
SIFT	1	60

Πίνακας 6: Παράδειγμα λήψης απόφασης για ένα Καρέ

Στον πίνακα 6 μας παρουσιάζονται τα αποτελέσματα που λαμβάνουμε για μια εικόνα προσώπου που εξετάζεται σε ένα καρέ ενός βίντεο κλιπ. Παρατηρούμε ότι οι αλγόριθμοι LDA, LogReg, QDA και SIFT-based NN μας παρέχουν σαν αποτέλεσμα το συναίσθημα της χαράς για αυτό το πρόσωπο που εξετάζουμε, ενώ οι αλγόριθμοι CRtree και kNN μας δίνουν το συναίσθημα της λύπης. Άρα εφαρμόζοντας την παραπάνω εξίσωση βρίσκουμε

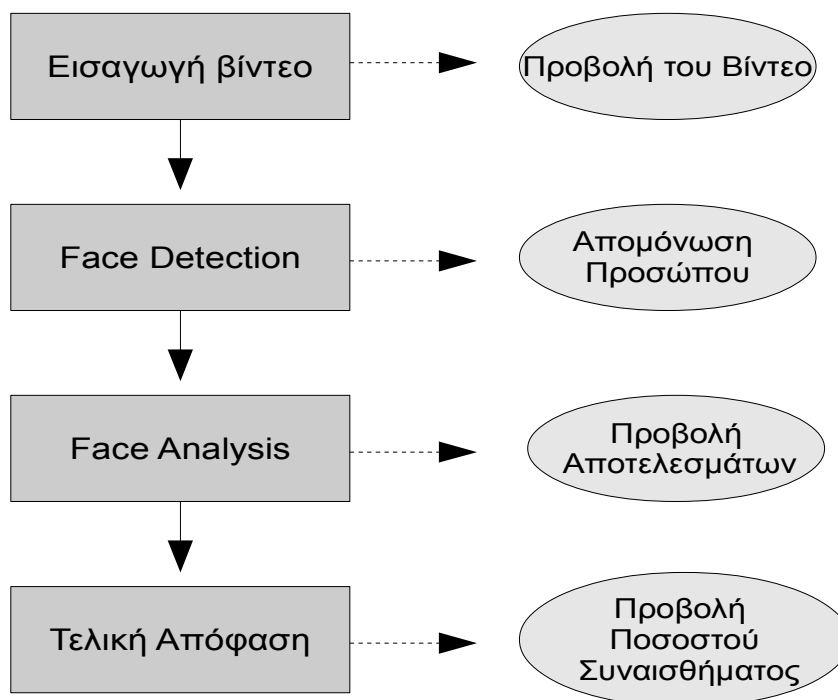
$$\text{Rate of Happiness} = [(1 * 0,6) + (1 * 0,6) + (0 * 100) + (1 * 0,6) + (0 * 100) + (1 * 0,6)] / 6 = 0,4$$

Άρα το συγκεκριμένο πρόσωπο που αναλύσαμε είναι κατά 40% χαρούμενο και στρογγυλοποιώντας το αποτέλεσμα αυτό καταλήγουμε ότι το πρόσωπο φέρει το συναίσθημα της λύπης (και μάλιστα με ποσοστό 60%).

Αφού πραγματοποιηθεί αυτή η διαδικασία για όλες τις εικόνες που εξετάζονται μέσα στο βίντεο μας, τότε λαμβάνουμε έναν πίνακα με 0 και 1 σαν περιεχόμενα του, τα οποία και χαρακτηρίζουν την συναισθηματική κατάσταση του κάθε προσώπου. Βγάζοντας το μέσω όρο των συναισθημάτων που περιγράφονται μέσα σε αυτόν τον πίνακα καταλήγουμε στο τελικό ποσοστό ευτυχίας του τραγουδιού.

## 5. Πειραματικά Αποτελέσματα και GUI Matlab

Σκοπός του προγράμματος μας δεν είναι μόνο να μπορέσουμε να δημιουργήσουμε ένα πρόγραμμα το οποίο θα μπορέσει να ανιχνεύει τα συναισθήματα ενός ερμηνευτή μέσα σε ένα μουσικό βίντεο κλιπ και η εύρεση νέων τεχνικών ανίχνευσης συναισθημάτων προσώπου. Το πρόγραμμα αυτό πρέπει να είναι και αρκετά λειτουργικό και να εύχρηστο για κάθε χρήστη, άσχετα με το επίπεδο γνώσεις του στους ηλεκτρονικούς υπολογιστές. Αυτό απαιτεί την δημιουργία ενός interface μέσω του οποίου ο χρήστης αποβάλλεται από την επαφή του με τον κώδικα, άρα κατά συνέπεια και από την πολυπλοκότητα του προγράμματος. Για να το επιτυγχάνουμε αυτό χρησιμοποιήσαμε ένα εργαλείο που μας παρέχεται στο matlab μέσω του οποίου και μπορούμε να δημιουργήσουμε σχετικά εύκολα αλλά και γρήγορα ένα φιλικό περιβάλλον για τον χρήστη.



Εικόνα 65: Απαιτήσεις Εμφάνισης ανά Βήμα

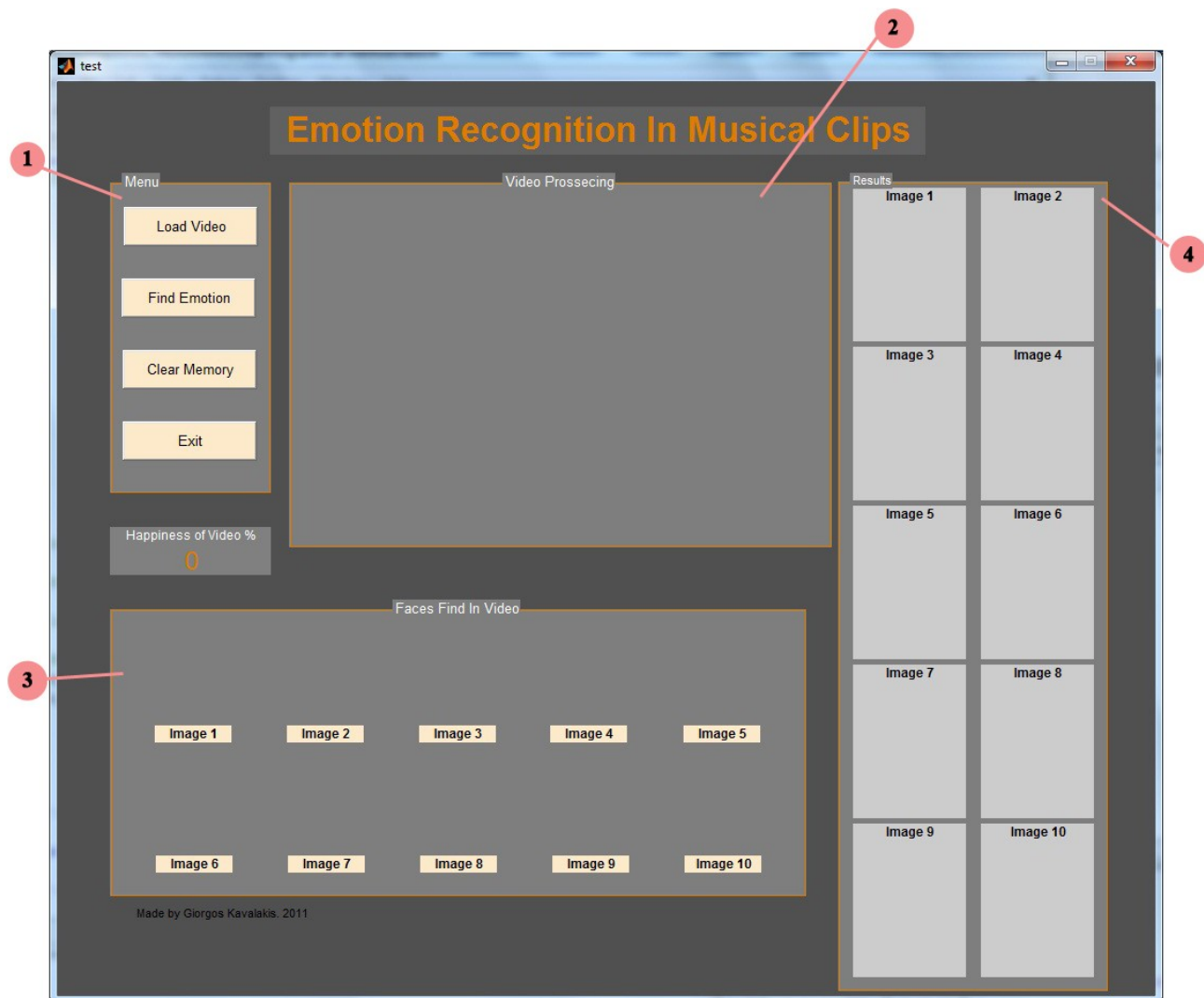
Τέσσερα είναι τα βασικά ζητήματα (εικόνα 65) τα οποία καλούμαστε να επιλύσουμε ώστε να μπορέσουμε να κάνουμε ένα βασικό γραφικό περιβάλλον. Κατά την εισαγωγή του βίντεο να μπορούμε να βλέπουμε μια εικόνα σε κάποιο figure ώστε να είμαστε βέβαιοι ότι έχουμε επιλέξει το σωστό βίντεο προς επεξεργασία και ότι δεν υπάρχει κάποιο σφάλμα κατά την ανάγνωση του από το πρόγραμμα. Επίσης είναι σημαντικό, κατά την ανίχνευση προσώπου, τα πρόσωπα τα οποία εντοπίζονται και κρίνονται κατάλληλα για επεξεργασία να αποκόβονται από το βίντεο και να απομονώνονται. Η εμφάνιση τους σε κάποιο σημείο του προγράμματος είναι χρήσιμη ώστε να μπορούμε να διασταυρώσουμε στο τέλος τα αποτελέσματα.

Είναι παράλογο μέσα σε ένα γραφικό περιβάλλον να μην υπάρχει χώρος για την παρουσίαση των αποτελεσμάτων. Πιο συγκεκριμένα, στην περίπτωση μας, πρέπει να λάβουμε υπόψη ότι έχουμε να εμφανίσουμε δυο διαφορετικά αποτελέσματα, αυτά που αφορούν κάθε πρόσωπο ξεχωριστά και αυτά που αναφέρονται στο σύνολο του βίντεο κλιπ.

## 5.1 Παρουσίαση Γραφικού περιβάλλοντος

Οι βασικές λειτουργίες που πρέπει να έχει ένα πρόγραμμα σαν αυτό είναι η εισαγωγή του βίντεο που θέλουμε να επεξεργαστούμε, η επιλογή εκκίνησης της ανάλυσης του βίντεο και η έξοδος από το πρόγραμμα. Πέρα από αυτές όμως είναι πάντα χρήσιμο να υπάρχει η επιλογή της εκκαθάρισης δεδομένων από το panel μας, για περιπτώσεις που κάτι συμβεί κατά την έκβαση των αποτελεσμάτων μας και σταματήσει το πρόγραμμά μας ή θελήσουμε να επαναλάβουμε μια διαδικασία από την αρχή.

Στην παρακάτω εικόνα θα βλέπουμε την αρχική μορφή που έχει το γραφικό περιβάλλον.

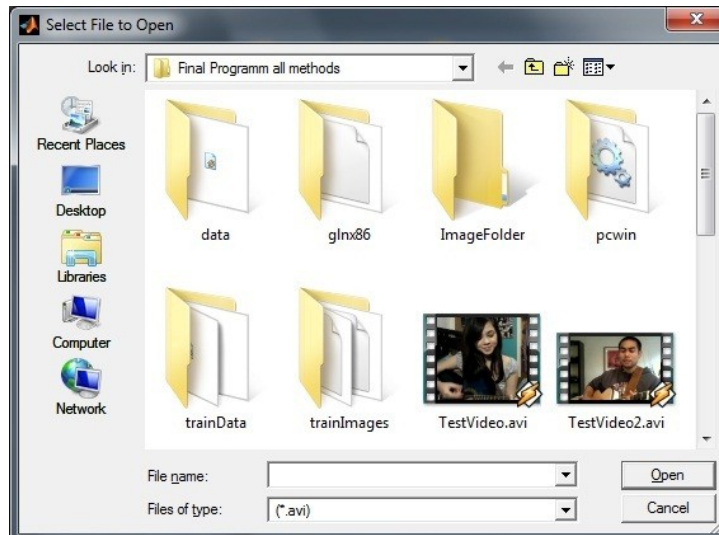


Εικόνα 66: Παράθυρο εφαρμογής

Όπως μπορούμε να παρατηρήσουμε στην εικόνα 66, το αρχικό μας παράθυρο αποτελείται από τέσσερα βασικά panel. Το panel 1 περιέχει το μενού επιλογών, οι οποίες είναι και αυτές που αναφέρθηκαν στην παραπάνω παράγραφο. Στο panel 2 εμφανίζεται το βίντεο το οποίο και επεξεργαζόμαστε. Το panel 3 είναι αυτό το οποίο χρησιμοποιείται για να εμφανίσουμε τα πρόσωπα τα οποία εν τέλει θα αναλύσουμε μέσα

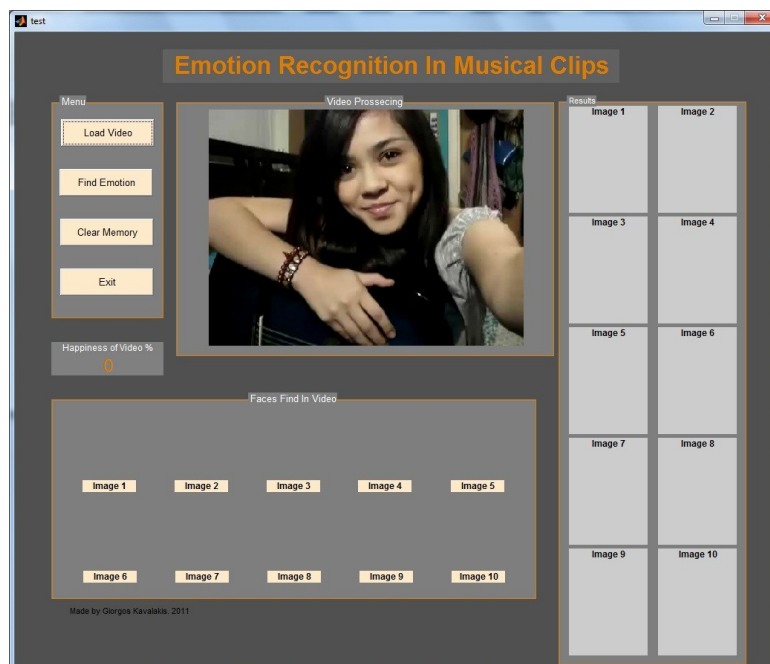
από τα βίντεο κλιπ. Και τέλος στο panel με ονομασία Results (panel 4) γίνεται η εμφάνιση των αποτελεσμάτων που λαμβάνουμε από κάθε μέθοδο ξεχωριστά για κάθε εικόνα.

Ας αναλύσουμε όμως τώρα λίγο παραπάνω πως λειτουργεί κάθε από τα πάνελ μας. Στο μενού μας περιέχονται τέσσερα κουμπιά. Με την πρώτη επιλογή (Load Video) φορτώνουμε στο πρόγραμμα μας το βίντεο το οποίο θέλουμε να επεξεργαστούμε, ανοίγοντας ένα παράθυρο επιλογής αρχείου (εικόνα 67).



Εικόνα 67: Παράθυρο επιλογής αρχείου

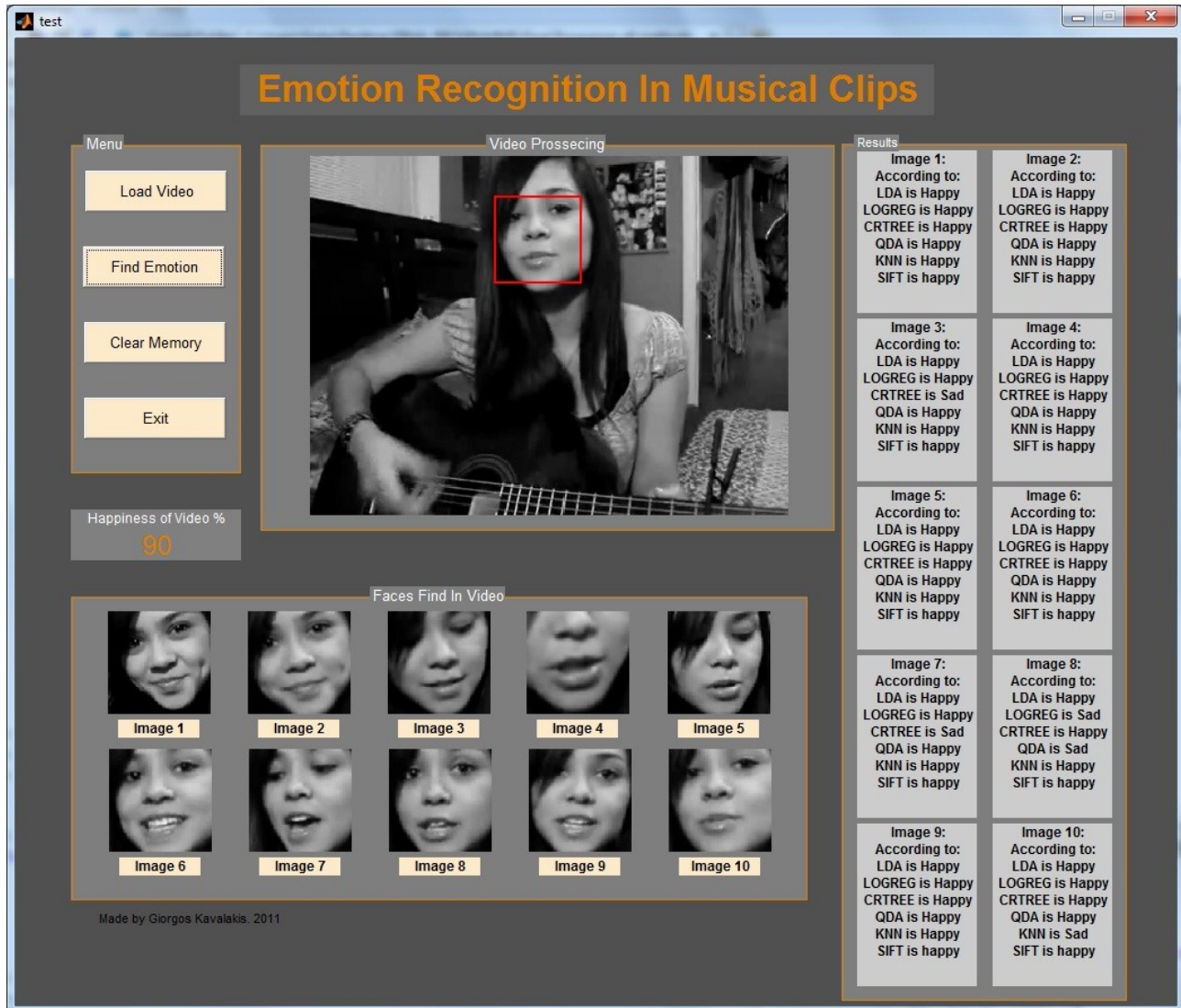
Αφού επιλέξουμε το βίντεο που θέλουμε να επεξεργαστούμε τότε αρχικά μας εμφανίζεται στο πάνελ 2 το βίντεο κλιπ, χωρίς όμως να υπάρχει η αναπαραγωγή του ακόμα, συνεπώς βλέπουμε το πρώτο καρέ του(εικόνα 68).



Εικόνα 68: Άμεση προβολή του πρώτου καρέ του βίντεο κατά την εισαγωγή του



Πλέον το πρόγραμμα είναι έτοιμο για να ξεκινήσει την διαδικασία ανίχνευσης συναισθημάτων μέσα στο μουσικό βίντεο κλιπ. Το μόνο που πρέπει να κάνει ο χρήστης είναι να πατήσει το κουμπί εκκίνησης (Find Emotion). Κατά την εκκίνηση του προγράμματος ξεκινάνε όλες οι διαδικασίες που αναφέραμε στο προηγούμενο κεφάλαιο. Για κάθε πρόσωπο που κρίνεται κατάλληλο για ανάλυση γίνεται η εμφάνιση του στο πάνελ 3 και αμέσως μετά στο πάνελ 4 γίνεται η εμφάνιση των αποτελεσμάτων για το συγκεκριμένο πρόσωπο και μόνο. Αφού βρεθούν και αναλυθούν και τα δέκα δείγματα προσώπων που το πρόγραμμα είναι καθορισμένο να λαμβάνει τότε εμφανίζεται ένα ποσοστό ανάμεσα στο πάνελ 1 και 3, το οποίο είναι και το ποσοστό ευτυχίας του βίντεο κλιπ. Στην παρακάτω εικόνα παρουσιάζεται ένα παράδειγμα ολοκληρωμένης αναγνώρισης συναισθημάτων.



Εικόνα 69: Λειτουργία προγράμματος

## 5.2 Ανάλυση Αποτελεσμάτων και Συμπεράσματα

Έχοντας πλέον ένα αρκετά ικανοποιητικό Interface και φιλικό ως προς τον χρήστη, μπορούμε να ασχοληθούμε περισσότερο με την ποιότητα των αποτελεσμάτων που λαμβάνουμε από την ανάλυση που κάνουμε στο βίντεο. Σαν αρχικό δείγμα θα λάβουμε την εικόνα 69. Στον πίνακα που ακολουθεί είναι συγκεντρωμένα τα αποτελέσματα που έχουμε λάβει από την ανάλυση του βίντεο στην εικόνα 69.

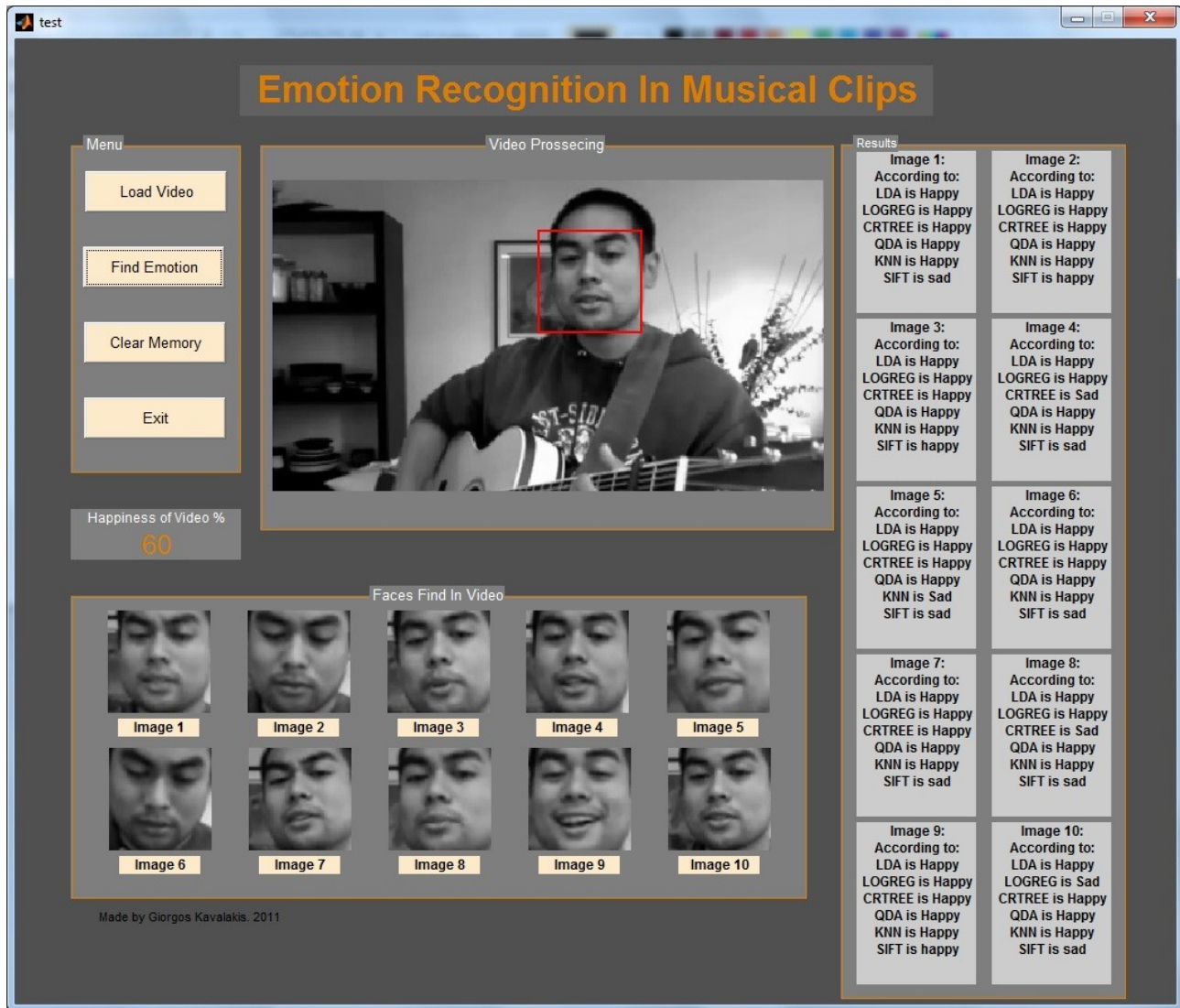
	LDA	LogReg	CRTree	QDA	kNN	SIFT
<b>Face 1</b>	Happy	Happy	Happy	Happy	Happy	Happy
<b>Face 2</b>	Happy	Happy	Happy	Happy	Happy	Happy
<b>Face 3</b>	Happy	Happy	Sad	Happy	Happy	Happy
<b>Face 4</b>	Happy	Happy	Happy	Happy	Happy	Happy
<b>Face 5</b>	Happy	Happy	Happy	Happy	Happy	Happy
<b>Face 6</b>	Happy	Happy	Happy	Happy	Happy	Happy
<b>Face 7</b>	Happy	Happy	Sad	Happy	Happy	Happy
<b>Face 8</b>	Happy	Sad	Happy	Sad	Happy	Happy
<b>Face 9</b>	Happy	Happy	Happy	Happy	Happy	Happy
<b>Face 10</b>	Happy	Happy	Happy	Happy	Sad	Happy

Πίνακας 7: Αποτελέσματα 1

Αρχικά πρέπει να τονίζουμε ότι υπάρχει μια εσφαλμένη αντίληψη καταλληλότητας ενός προσώπου. Στην εικόνα 4 περιέχονται μόνο το στόμα και η μύτη, κρίνοντας το πρόσωπο αυτό ακατάλληλο για ανάλυση, πράγμα το οποίο δεν εντοπίζει σωστά το πρόγραμμα μας σε αυτήν την περίπτωση. Ακόμα μια φορά πρέπει να τονίσουμε ότι τα συναισθήματα που εξετάζουμε είναι αυτά της χαράς και της λύπης, παραμελώντας συναισθήματα όπως του θυμού, της έκπληξης. Το ουδέτερο συναίσθημα δεν συμπεριλαμβάνεται στην γκάμα των συναισθημάτων το οποίο και εξετάζουμε καθώς η μουσική θεωρείται τρόπος έκφρασης του ανθρώπου, και δεν είναι δυνατό ένα κομμάτι να είναι μουσικά ουδέτερο.

Στην συνέχεια παρουσιάζονται ακόμα 4 περιπτώσεις από μουσικά βίντεο κλιπ, τα οποία αναλύθηκαν μέσω του προγράμματος μας.





Εικόνα 70: Δεύτερη ανάλυση

	LDA	LogReg	CRTree	QDA	kNN	SIFT
Face 1	Happy	Happy	Happy	Happy	Happy	Sad
Face 2	Happy	Happy	Happy	Happy	Happy	Happy
Face 3	Happy	Happy	Happy	Happy	Happy	Happy
Face 4	Happy	Happy	Sad	Happy	Happy	Sad
Face 5	Happy	Happy	Happy	Happy	Sad	Sad
Face 6	Happy	Happy	Happy	Happy	Happy	Sad
Face 7	Happy	Happy	Happy	Happy	Happy	Sad
Face 8	Happy	Happy	Sad	Happy	Happy	Sad
Face 9	Happy	Happy	Happy	Happy	Happy	Happy
Face 10	Happy	Sad	Happy	Happy	Happy	Sad

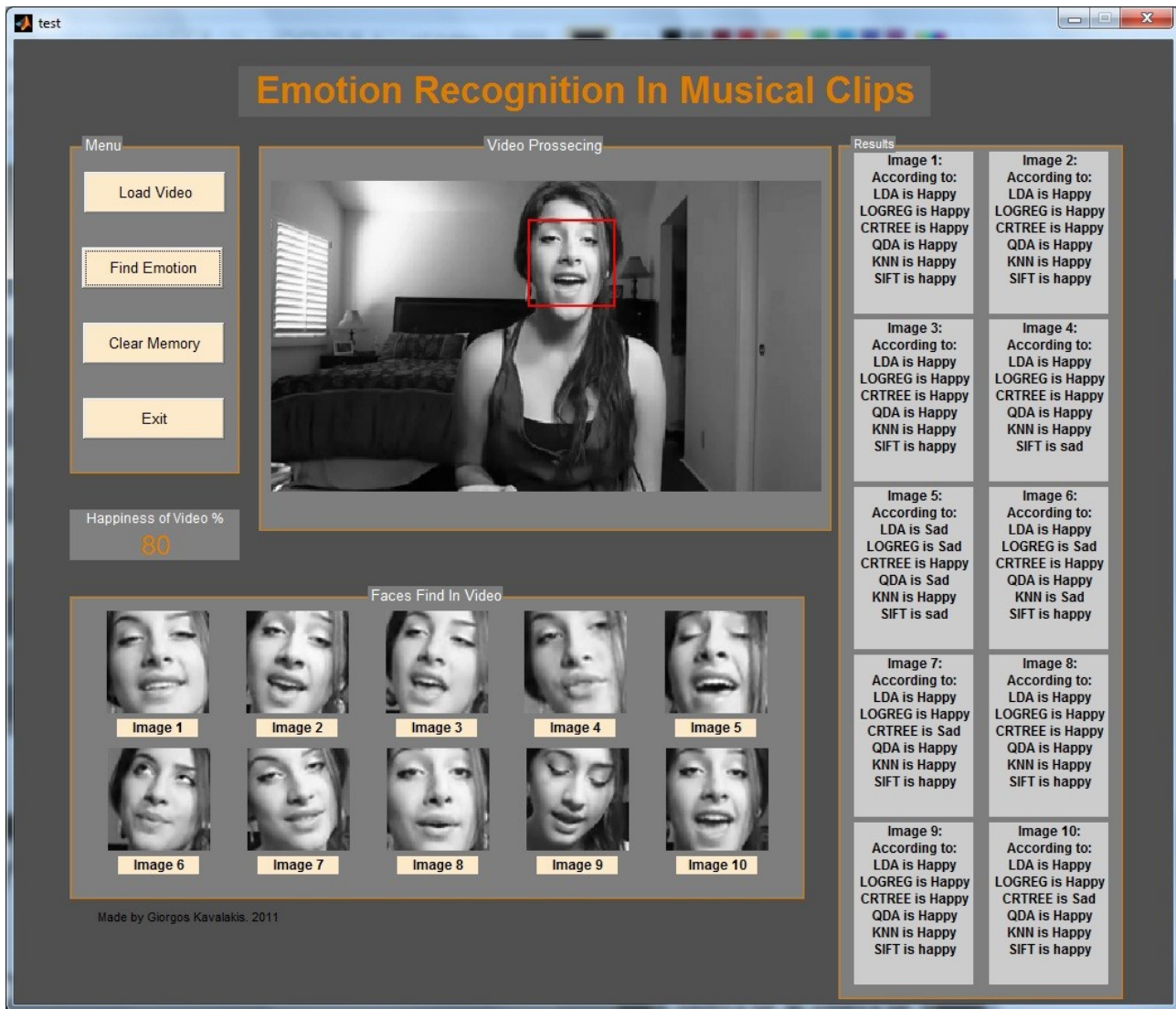
Πίνακας 8: Αποτελέσματα Δεύτερης Ανάλυσης



Εικόνα 71: Τρίτη ανάλυση

	<b>LDA</b>	<b>LogReg</b>	<b>CRTree</b>	<b>QDA</b>	<b>kNN</b>	<b>SIFT</b>
<b>Face 1</b>	Happy	Happy	Happy	Happy	Happy	Happy
<b>Face 2</b>	Sad	Sad	Happy	Sad	Sad	Sad
<b>Face 3</b>	Happy	Happy	Happy	Happy	Happy	Sad
<b>Face 4</b>	Happy	Sad	Sad	Sad	Sad	Sad
<b>Face 5</b>	Happy	Happy	Happy	Happy	Happy	Happy
<b>Face 6</b>	Happy	Sad	Happy	Sad	Sad	Sad
<b>Face 7</b>	Happy	Sad	Sad	Happy	Sad	Sad
<b>Face 8</b>	Sad	Sad	Happy	Sad	Sad	Happy
<b>Face 9</b>	Sad	Sad	Happy	Sad	Sad	Sad
<b>Face 10</b>	Happy	Sad	Happy	Sad	Sad	Happy

Πίνακας 9: Αποτελέσματα Τρίτης Ανάλυσης

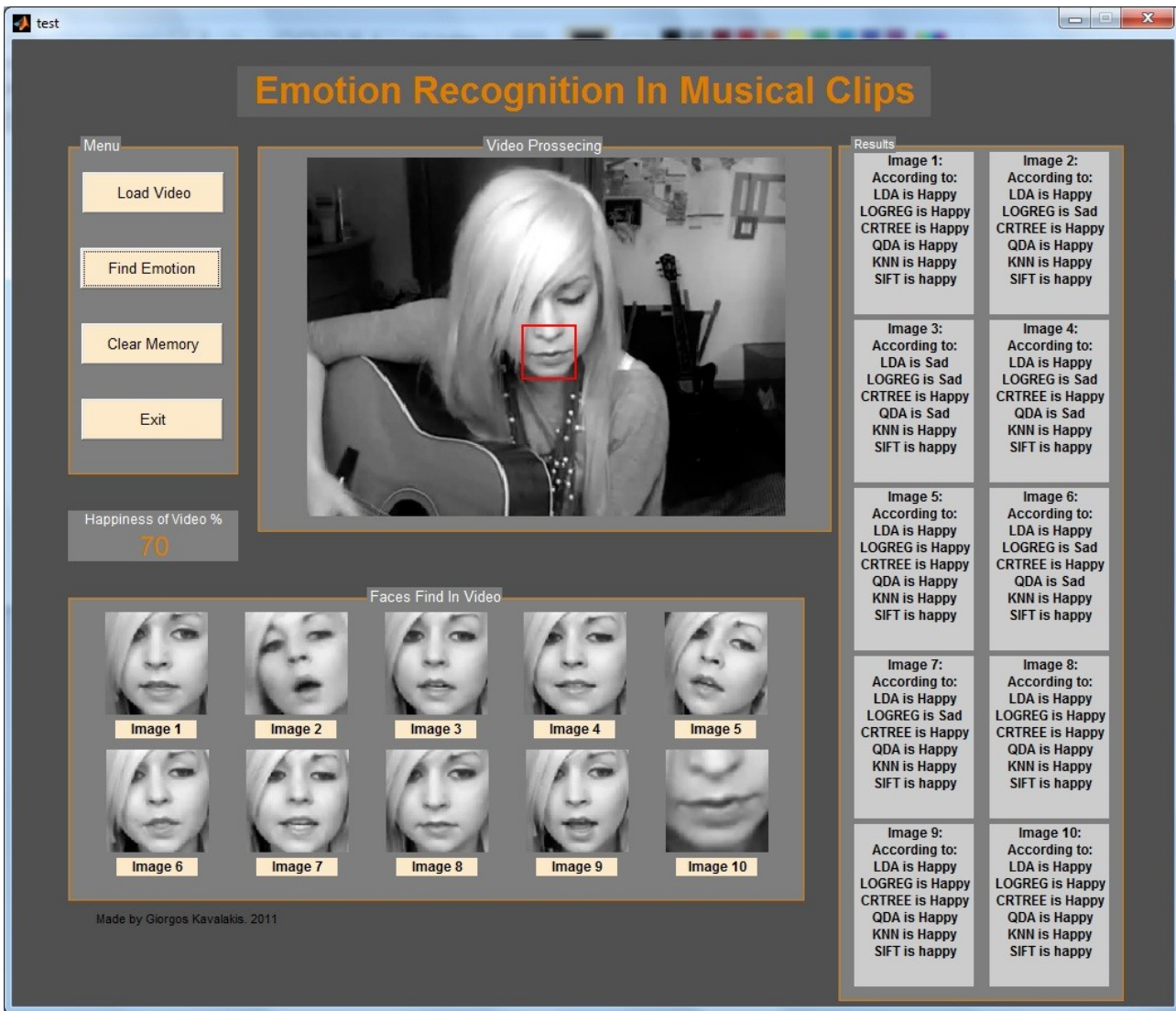


Εικόνα 72: Τέταρτη ανάλυση

	LDA	LogReg	CRTree	QDA	kNN	SIFT
Face 1	Happy	Happy	Happy	Happy	Happy	Happy
Face 2	Happy	Happy	Happy	Happy	Happy	Happy
Face 3	Happy	Happy	Happy	Happy	Happy	Happy
Face 4	Happy	Happy	Happy	Happy	Happy	Sad
Face 5	Happy	Sad	Happy	Sad	Happy	Happy
Face 6	Happy	Sad	Happy	Happy	Sad	Happy
Face 7	Happy	Happy	Sad	Happy	Happy	Happy
Face 8	Happy	Happy	Happy	Happy	Happy	Happy
Face 9	Happy	Happy	Happy	Happy	Happy	Happy
Face 10	Happy	Happy	Sad	Happy	Happy	Happy

Πίνακας 10: Αποτελέσματα Τέταρτης Ανάλυσης





Εικόνα 73: Πέμπτη ανάλυση

	LDA	LogReg	CRTree	QDA	kNN	SIFT
Face 1	Happy	Happy	Happy	Happy	Happy	Happy
Face 2	Happy	Sad	Happy	Happy	Happy	Happy
Face 3	Sad	Sad	Happy	Sad	Happy	Happy
Face 4	Happy	Sad	Happy	Sad	Happy	Happy
Face 5	Happy	Happy	Happy	Happy	Happy	Happy
Face 6	Happy	Sad	Happy	Sad	Happy	Happy
Face 7	Happy	Sad	Happy	Happy	Happy	Happy
Face 8	Happy	Happy	Happy	Happy	Happy	Happy
Face 9	Happy	Happy	Happy	Happy	Happy	Happy
Face 10	Happy	Happy	Happy	Happy	Happy	Happy

Πίνακας 11: Αποτελέσματα Πέμπτης Ανάλυσης

Έχοντας ολοκληρώσει ένα μέρος της παρουσίασης των αποτελεσμάτων που έχουμε λάβει από της

δοκιμές που έχουμε κάνει, θα κρίνουμε καθαρά με αυτά που βλέπουμε, χωρίς να γνωρίζουμε το κομμάτι που ερμηνεύει ο τραγουδιστής. Έτσι θα μπορέσουμε να κρίνουμε εάν ο αλγόριθμος μπορεί να πλησιάσει τον τρόπο 'σκέψης' του δικού μας εγκέφαλου, χωρίς εμείς να έχουμε επηρεαστεί από το κομμάτι που ερμηνεύουν, αλλά κρίνοντας καθαρά από την εικόνα που μας παρουσιάζεται.

## Ανίχνευση προσώπου

Αρχικά πρέπει να δούμε και να κρίνουμε την ποιότητα των δειγμάτων προσώπων που λαμβάνουμε από την ανίχνευση προσώπων. Μέσα σε ένα σύνολο από 50 δείγματα προσώπων που λαμβάνονται και παρουσιάζονται στις παραπάνω σελίδες μόνο 2 είναι αυτά τα οποία δεν είναι κατάλληλα για ανάλυση (εικόνα 69 image 4 – εικόνα 73 image 10). Γεγονός που μεταφράζεται σε αριθμούς σε ένα ποσοστό επιτυχίας 96%. Αυτό οφείλεται όχι μόνο στον βελτιωμένο κώδικα που χρησιμοποιήθηκε για την ανίχνευση προσώπου, αλλά και στην επαλήθευση που κάνουμε με τον έλεγχο ύπαρξης των βασικών χαρακτηριστικών ενός προσώπου.



Εικόνα 74: Εσφαλμένη ανίχνευση προσώπου

Ακόμα και στην εσφαλμένη ανίχνευση προσώπου βλέπουμε ότι έχει εντοπίσει σωστά το πρόσωπο, αλλά δεν έχει οριοθετήσει ορθά την ακριβή τοποθεσία του. Δείγμα το οποίο είναι ενθαρρυντικό, καθώς είναι πολύ δύσκολο να λάβει σαν δείγμα κάτι το οποίο είναι τελείως διαφορετικό από πρόσωπο.

## Δυσκολία εύρεσης συναισθημάτων

Και στις πέντε περιπτώσεις βλέπουμε ότι το συναίσθημα της λύπης είναι πολύ πιο δύσκολα ανιχνεύσιμο από αυτό της χαράς. Πιθανοί λόγοι για αυτό το συμβάν είναι δύο. Ο πρώτος και ο κυριότερος είναι η εγκυρότητα της βάσης δεδομένων, καθώς μια σωστή βάση δεδομένων προϋποθέτει μια μεγάλη μελέτη και προσεκτική επιλογή δειγμάτων. Δυστυχώς δεν είχαμε την πολυτέλεια και την άνεση του χρόνου για κάνουμε μια βάση δεδομένων που θα μπορούσε να κριθεί άψογη. Και για αυτόν τον λόγο δημιουργήσαμε μια βάση η οποία θα μπορούσε να μας εξυπηρετήσει μέχρι έναν βαθμό. Ο δεύτερος λόγος είναι ότι η διαφορές ανάμεσα στις εκφράσεις χαράς και λύπης δεν είναι πολύ μεγάλες πλέον. Και στις δύο περιπτώσεις, το στόμα του ατόμου είναι ανοικτό, γιατί ας μην ξεχνάμε ότι εκείνη την στιγμή τραγουδάει. Οπότε είναι σχετικά εύκολο να αποπλανήσει τους αλγόριθμους μας και να τους οδηγήσει σε λάθος κρίση.

## Επιτυχία κάθε αλγόριθμου

Μπορούμε ρίχνοντας μια ματιά στους πίνακες που προβάλλονται παραπάνω να διαπιστώσουμε ότι δεν έχουν όλοι οι αλγόριθμοι που χρησιμοποιήσαμε την ίδια αποτελεσματικότητα. Σε κάποιες περιπτώσεις – βλέπε εικόνα 71, image5 – κανένας αλγόριθμος δεν μπόρεσε να εντοπίσει το σωστό συναίσθημα.



Εικόνα 75: Ολική αποτυχία αλγορίθμων

Αντίθετα στην ίδια εικόνα αλλά στον 7 δείγμα έχουμε ένα αρκετά ικανοποιητικό ποσοστό επιτυχίας των αλγορίθμων.



Εικόνα 76: Επιτυχία αλγορίθμων

Από τους έξι αλγόριθμους, μόνο δυο ήταν αυτή που δεν είχαν ορθό αποτέλεσμα.



Εικόνα 77: Αποτυχία CRTree

Στην εικόνα 76 παρατηρείτε ακόμα μεγαλύτερο ποσοστό επιτυχίας των αλγορίθμων, με μοναδικό αλγόριθμο που αποτυχαίνει στην ορθή απόφαση να είναι ο CRTree.

Μετά από αρκετές δοκιμές σε βίντεο κλιπ που δοκιμάσαμε την εφαρμογή και των έξι αλγορίθμων για την ανίχνευση συναισθημάτων, δεν μπορούμε με σιγουριά να αποκλείσουμε κάποιες από την λήψη της τελικής απόφασης. Καμία δεν φαίνεται να έχει καλύτερη ή χειρότερη απόδοση από τις υπόλοιπες. Αυτό δεν είναι απαραίτητα κακό, καθώς το γεγονός ότι έχουμε 6 μεθόδους για να λάβουμε αποτελέσματα μόνο για μια εικόνα την φορά μας κάνει να είμαστε πιο ακριβής στην τελική μας απόφαση και πιο σίγουροι. Υπολογιστικά ο χρόνος που δαπανάται για τον υπολογισμό των 6 εξόδων αυτών είναι κατά πολύ μικρότερος από αυτόν που χρειαζόμαστε για την ανίχνευση του προσώπου που τίθεται προς ανάλυση. Αυτό το γεγονός δεν μας αποτρέπει από το να τις χρησιμοποιούμε και τις έξι για τον σκοπό που τις καθορίζουμε, αντίθετα ακόμα και στην περίπτωση που βελτιωθεί η αποτελεσματικότητα τους μέσω της βελτίωσης της βάσης δεδομένων, θα μπορούσαμε να τις χρησιμοποιούμε όλες σαν επαλήθευση η μια της άλλης.

## Αποτελέσματα SIFT-based Neural Network

Τα αποτελέσματα που λαμβάνουμε από το νευρωνικό δίκτυο είναι αρκετά ενθαρρυντικά. Αρχικά έχοντας ένα performance της τιμής 0,44 δεν ήμασταν σίγουροι για το πόσο αποδοτικό θα ήταν το νευρωνικό δίκτυο μας. Φυσικά για ακόμα φορά το εμπόδιο για καλύτερα αποτελέσματα είναι η βάση δεδομένων μας. Παρά το γεγονός αυτό ο αλγόριθμος που βασίζεται στα χαρακτηριστικά διανύσματα που εξάγονται από τον αλγόριθμο SIFT είναι πολύ ικανοποιητικός στην ορθότητα των αποτελεσμάτων του. Μην ξεχνάμε ότι είναι πειραματικό τόσο η μέθοδος όσο και η χρήση που γίνεται. Παρακάτω παρατήθεται ένας πίνακας με τα ορθά και λάθη αποτελέσματα του νευρωνικού δικτύου.

	Ορθά Αποτέλεσμα (στα 10)	Λανθασμένα Αποτέλεσμα (στα 10)	Ποσοστό επιτυχίας
Πρώτη Ανάλυση	9	1	90,00%
Δεύτερη Ανάλυση	8	2	80,00%
Τρίτη Ανάλυση	7	3	70,00%
Τέταρτη Ανάλυση	7	3	70,00%
Πέμπτη Ανάλυση	7	3	70,00%
Συνολικά (μέσος όρος)	8	2	76,00%

Πίνακας 12: Αποτελέσματα Νευρωνικού δικτύου

## Καταλληλότητα Βάσης Δεδομένων

Η μουσική είναι τρόπος έκφρασης. Αυτό σημαίνει ότι είναι δύσκολο να βάλουμε κανόνες στον τρόπο που κάθε καλλιτέχνης εκφράζεται την ώρα που ερμηνεύει ένα μουσικό κομμάτι. Παρόλα αυτά κάποιοι κανόνες είναι ευδιάκριτοι, όπως για παράδειγμα, κατά την διάρκεια ερμηνείας ενός θλιβερού κομματιού συνήθως το βλέμμα του καλλιτέχνη χαμηλώνει ή τα φρύδια του ανασηκώνονται – χαρακτηριστικό του συναισθήματος της έκπληξης στην καθημερινότητα. Για να μπορέσουμε να πραγματοποιήσουμε την εφαρμογή μας μας ζητήθηκε να κατασκευάσουμε μια βάση δεδομένων με πρόσωπα ερμηνευτών, κυρίως από διασκευές κομματιών με όχι πολύπλοκο φόντο, την ώρα που αυτοί ερμηνεύουν διάφορα κομμάτια, τόσο κομμάτια χαράς όσο και λύπης.



Για να καταφέρουμε να πραγματοποιήσουμε αυτό τον στόχο έπρεπε να παρατηρήσουμε κάποιους κανόνες, να μαζέψουμε φωτογραφίες προσώπων και να κρίνουμε εάν είναι χαρούμενα ή όχι, ασχέτως από το τι τραγουδι ερμηνεύουν. Όπως μπορεί να καταλάβει ο καθένας από αυτό η κρίση της κάθε φωτογραφίας ήταν υποκειμενική. Φυσικά δεν λάβαμε υπόψιν μόνο μια γνώμη. Την ίδια βάση δεδομένων έλεγξαν και ανάλυσαν 3 διαφορετικά άτομα για να μπορέσουμε να πάρουμε όσο το δυνατόν καλύτερα αποτελέσματα. Και αυτό γιατί σε πολλές περιπτώσεις κάποιου προσώπου δεν είναι εύκολο ακόμα και για τον ανθρώπινο εγκέφαλο να κρίνει ορθά την συναισθηματική κατάσταση του ατόμου.

Το δεύτερο πρόβλημα που έπρεπε να αντιμετωπίσουμε ήταν αυτό της ποιότητας των εικόνων των προσώπων. Όσο καλύτερη ποιότητα τόσο περισσότερη πληροφορία μπορείς να εξάγεις. Δυστυχώς εδώ σταθήκαμε άτυχοι καθώς ήταν δύσκολο να συλλέξουμε έναν μεγάλο αριθμό από εικόνες ερμηνευτών που να είναι κατάλληλες για τον σκοπό που καθορίζονταν.

Η βάση δεδομένων ήταν το μεγαλύτερο πρόβλημα που αντιμετωπίσαμε στο σύνολο της πτυχιακής εργασίας. Ήταν αρκετά καλή για να μπορέσουμε να έχουμε μια αρχή να δουλέψουμε πάνω στο αντικείμενο αυτό αλλά ήταν και αυτό που, λόγω της χρήσης του κατά την εκπαίδευση των μεθόδων, μας έδωσε όχι τόσο ικανοποιητικά αποτελέσματα για κάθε αλγόριθμο ξεχωριστά. Μπορεί το σύνολο του προγράμματος να προσαρμόστηκε με τα δεδομένα που του δινόταν ώστε τα τελικά αποτελέσματα να είναι αρκετά ικανοποιητικά, όμως με μια ματιά ανά μέθοδο ξεχωριστά μας υπενθυμίζει ότι υπάρχει τρόπος και χώρος βελτίωσης.

## Γενικά συμπεράσματα

Έχοντας ολοκληρώσει την παρουσίαση των αποτελεσμάτων που λάβαμε από την εφαρμογή και τις δοκιμές που κάναμε σε διάφορα μουσικά βίντεο κλιπ καταλήγουμε στο συμπέρασμα ότι η γενική ιδέα της πτυχιακής εργασίας είναι εφικτή, αν και με ένα αρκετό μεγάλο βαθμό δυσκολίας, και υλοποιήσιμη.

Για την ορθή λειτουργία του προγράμματος μας πρέπει να γίνουν κάποιες μετατροπές. Σε γενικές γραμμές όμως για το χρονικό διάστημα και το φόρτο εργασίας που επιβάρυνε μόνο ένα άτομο, η συγκεκριμένη πτυχιακή εργασία έχει καταφέρει να υποδείξει ορθά τρία ζητήματα τα οποία και έχουν λυθεί μέχρι κάποιο βαθμό. Το να αναγνωρίσεις τα συναισθήματα σε έναν καλλιτέχνη πέρα από την πρωτοτυπία του που παρουσιάζει από μόνο του, απαιτεί να καινοτομήσεις σε τουλάχιστον δύο ακόμα πράγματα, στην βάση δεδομένων και την μέθοδο που θα χρησιμοποιήσεις.

Αναπτύχθηκε μια βάση δεδομένων από την αρχή και τέθηκε μια νέα ιδέα και τεχνική για την αναγνώριση συναισθημάτων με βάση τις εκφράσεις προσώπου. Ο αλγόριθμος SIFT-based neural Network μας έδωσε αποτελέσματα αρκετά ικανοποιητικά και πολύ υποσχόμενα. Το πρόγραμμα είναι τελείως φιλικό με τον χρήστη και κάνει την χρήση του ιδανική ακόμα και για αυτούς που δεν έχουν ιδέα από ηλεκτρονικούς υπολογιστές. Για την ακρίβεια ο χρήστης μπορεί να δει να και να επέμβει στο πρόγραμμα, χωρίς να κοιτάζει καν τον κώδικα στις εξής λειτουργίες:

Λειτουργίες	Δυνατότητες
Εισαγωγή βίντεο	✓
Παύση λειτουργίας	✗
Αλλαγή αριθμών δειγμάτων	✗
Προβολή δειγμάτων προσώπου	✓
Προβολή αποτελεσμάτων όλων των μεθόδων	✓
Προβολή ποσοστού	✓

συναισθήματος	
Έξοδος	✓
Εκκαθάριση Μνήμης	✓
Εκκαθάριση Παραθύρου	✓

Πίνακας 13: Δυνατότητες προγράμματος

## 6. Μελλοντική Εργασία και Επεκτάσεις

Παρατηρώντας τις αποδόσεις του προγράμματος μας μπορούμε να πούμε με σιγουριά ότι πρόκειται για ένα επιχείρημα το οποίο με την κατάλληλη προσοχή μπορεί να γίνει ένα ιδανικό εργαλείο για κάθε προσωπικό υπολογιστή ή επιχείρηση. Τα αποτελέσματα ήταν αρκετά ικανοποιητικά και μας παροτρύνουν να πούμε με σιγουριά ότι μπορεί να συνεχιστή περαιτέρω μελέτη πάνω σε αυτόν τον ομολογουμένως καινοτόμο, όσο αφορά τον κλάδο, τομέα. Μια μελέτη με περισσότερη προσοχή στα τεχνικά ζητήματα θα μπορούσε να μας εξασφαλίσει αποτελέσματα τα οποία θα ήταν ικανά για την δημιουργία ενός εμπορικού προγράμματος και την υιοθέτηση της τεχνογνωσίας αυτής από μεγάλους κολοσσούς της ηλεκτρονικής βιομηχανίας της μουσικής

### 6.1 Βελτιώσεις Υπάρχοντος Προγράμματος

Το πρόγραμμα το οποίο παρουσιάστηκε στην συγκεκριμένη πτυχιακή εργασία, παρόλο που δεν υπάρχει κάποιο εμφανές λάθος, μπορεί να δεχτεί κάποιες βελτιώσεις ή τροποποιήσεις ώστε να έχουμε τα καλύτερα δυνατά αποτελέσματα. Είναι μικρά θέματα τα οποία είναι εύκολα να υλοποιηθούν και κάνουν το πρόγραμμα λειτουργίσιμο σε διαφορετικές συνθήκες αλλά και ακριβέστερο ως προς τα αποτελέσματα που μπορούμε να λάβουμε.

#### 6.1.1 Συμβατότητα Προγράμματος

Όπως έχουμε αναφέρει και σε προηγούμενο κεφάλαιο, το πρόγραμμα λόγω της χρήσης κάποιας βιβλιοθήκης dll, θέτει έναν περιορισμό για το λειτουργικό περιβάλλον το οποίο χρειάζεται για να λειτουργήσει σωστά. Το παρόν πρόγραμμα έχει ορθή λειτουργία σε περιβάλλον windows 7 (έκδοση 32 bits) και σε Matlab 7.11.0 (R2010b). Η λειτουργία του σε διαφορετικό λειτουργικό μας παρεμποδίζει από την ολοκλήρωση του face detection που πραγματοποιείται μέσα στο πρόγραμμα. Όμως δεν είναι μόνο αυτός ο λόγος ο οποίος υπάρχει η ασυμβατότητα σε άλλα λειτουργικά. Καλός ή κακός το ίδιο το Matlab πέρα από διαφορές που παρατηρούνται σε εντολές, σε προγράμματα διαφορετικών εκδόσεων, υπάρχουν και αρκετά μεγάλες διαφορές ακόμα και στα αρχεία βίντεο που μπορεί να εισάγει, αναλόγως το λειτουργικό που χρησιμοποιείται. Για παράδειγμα το Matlab στην έκδοση των Linux μπορεί να εισάγει μόνο βίντεο τύπου AVI και αυτό με την προϋπόθεση να μην έχει υποστεί καμία συμπίεση.

Μια βελτίωση που θα μπορούσαμε να κάνουμε μελλοντικά στο πρόγραμμα είναι να μπορεί να εκτελεσθεί ανεξαρτήτως από το λειτουργικό το οποίο εμείς χρησιμοποιούμε. Ακόμα και να χρειαστεί να δημοσιεύσουμε τρεις διαφορετικές εκδόσεις, ώστε να υπάρχει η κατάλληλη για τα τρέχον εμπορικά λειτουργικά τα οποία έχουν την μεγαλύτερη ζήτηση (Windows, Mac OS και Linux), θα είναι αρκετά εξυπηρετικό για τους χρήστες, καθώς δεν αποκλείουμε καμία ομάδα αυτών.

#### 6.1.2 Αυτονομία Προγράμματος

Το πρόγραμμα μας έχει δημιουργηθεί μέσω του Matlab, τουλάχιστον το μεγαλύτερο κομμάτι αυτού. Αυτό δυστυχώς το κάνει να είναι 'εξαρτημένο' από αυτό για να μπορέσει να εκτελεσθεί. Κάθε φορά που ένας χρήστης θέλει να το λειτουργήσει πρέπει να θέσει σε λειτουργία πρώτα το Matlab και στην συνέχεια, μέσω αυτού να μπορέσει να εκτελέσει και το πρόγραμμα. Το Matlab με διάφορες τεχνικές πλέον μας δίνει την δυνατότητα να μετατρέψουμε τα αρχεία κώδικα του σε εκτελέσιμα αρχεία, έτσι ώστε να μπορούν να σταθούν μόνα τους μέσα σε ένα λειτουργικό, χωρίς την ανάγκη ύπαρξης του προγράμματος στο οποίο και

κατασκευάστηκαν.

Αυτό είναι και ένας μελλοντικός μας στόχος ώστε να ευκολύνουμε ακόμα παραπάνω τον κάθε χρήστη, ο οποίος δεν είναι υποχρεωμένος να έχει στην κατοχή του το Matlab, αλλά μόνο το πρόγραμμα Αυτόματης Αναγνώρισης Συναισθημάτων στα μουσικά βίντεο κλιπ. Εάν επιτευχθεί αυτός ο στόχος, το πρόγραμμα πλέον μπορεί να απευθυνθεί σε χρήστες που πραγματικά δεν έχουν μεγάλη εμπειρία και γίνεται ακόμα πιο εύκολο στην χρήση του.

### 6.1.3 Ρυθμός Δειγματοληψίας

Μέσα στο πρόγραμμα κάποια πράγματα που ήταν ζωτικά για την έκβαση των αποτελεσμάτων, ο χρήστης δεν είχε το δικαίωμα μέσω του γραφικού περιβάλλον να τα τροποποιήσει. Δύο από τις παραμέτρους αυτές είναι το διάστημα δειγματοληψίας και ο αριθμός των δειγμάτων που θέλουμε να λάβουμε. Εάν το αποτέλεσμα είναι αμφισβητούμενο, αρκετά μεγάλο ποσοστό σε αυτό έχει και το νούμερο των δειγμάτων που έχουμε λάβει. Για ένα βίντεο διάρκειας τριών λεπτών έχουμε ένα σύνολο από 4500 καρέ(υπολογίζοντας ότι έχουμε ρυθμό μετάδοσης εικόνας στα 25fps). Λαμβάνοντας 10 μόνο δείγματα ανά 2 ή 3 δευτερόλεπτα αφήνουμε ένα πολύ μεγάλο αριθμό του βίντεο κλιπ ανεκμετάλλευτο.

Δίνοντας στον χρήστη αυτήν την επιλογή μπορούμε να αποφύγουμε, ή τουλάχιστον να δώσουμε μικρότερη βαρύτητα, σε καρέ τα οποία ο ερμηνευτής να μην έχει διατυπωμένο το σωστό συναίσθημα του κομματιού. Έτσι με μεγαλύτερο αριθμό δειγμάτων η τελική απόφαση για το σύνολο του μουσικού βίντεο κλιπ σίγουρα θα είναι πιο αντικειμενική.

### 6.1.4 Βελτίωση Κώδικα

Το υπάρχον πρόγραμμα έχει δημιουργηθεί με την χρήση και τον συνδυασμό τμημάτων από διάφορες κώδικες που έχουν σκοπό να καταφέρουν να εκπληρώσουν όλους τους στόχους που είχαμε θέσει από της αρχή αυτής της εργασίας. Πολλά σημεία από αυτούς τους κώδικες έχουν αφαιρεθεί, άλλα έχουν τροποποιηθεί και έχουν συνδυαστεί με νέα κομμάτια κώδικα που είναι κατασκευασμένος καθαρά για την αποτελεσματική εκτέλεση αυτού το προγράμματος. Η ποιότητα του συνολικού κώδικα είναι αυτή που κρίνει το χρόνο που χρειαζόμαστε για να ολοκληρώσουμε μια ανάλυση σε ένα βίντεο κλιπ. Υπό τις συγκεκριμένες ρυθμίσεις ο χρόνος που χρειάζεται για την ολοκλήρωση της ανάλυσης είναι αρκετά ικανοποιητικός. Αλλά τι θα γίνει εάν αυξήσουμε τον αριθμό δειγμάτων που προορίζονται για ανάλυση;

Ο περισσότερος χρόνος που απαιτεί το πρόγραμμα μας είναι για την αναζήτηση κατάλληλων προσώπων για ανάλυση. Όταν αυτό το νούμερο είναι μικρό(10 στην περίπτωση μας) τότε δεν έχουμε πρόβλημα με τον χρόνο που χρειαζόμαστε. Όμως εάν ο αριθμός αυτός αυξηθεί πρέπει να μην αυξηθεί και πολύ ο χρόνος που χρειάζεται το πρόγραμμα για τον εντοπισμό τους. Η βελτίωση του κώδικα θα μας επιτρέψει να έχουμε γρήγορα αποτελέσματα και να κάνουμε ευκολότερη την χρήση του προγράμματος για την ανάλυση μιας μεγάλης βάσης δεδομένων από μουσικά βίντεο κλιπ, μειώνοντας τον χρόνο που θα απαιτούσε με τον τωρινό κώδικα η ολοκλήρωση αυτής της διαδικασίας.

## 6.2 Βελτιώσεις Μεθόδων και Τεχνικών

### 6.2.1 Βάση δεδομένων

Για να μπορέσει κάποιος να φτιάξει μια βάση δεδομένων που θα περιέχει πληροφορίες οι οποίες θα μπορέσουν να χρησιμοποιηθούν για την ολοκλήρωση μιας έρευνας ή την ανάπτυξη ενός προγράμματος, πρώτα πρέπει να κάνει μια μελέτη για τα αντικείμενα που θα τοποθετήσει μέσα στην βάση. Η δημιουργία μιας τέτοιας βάσης είναι αρκετά δύσκολη υπόθεση, καθώς απαιτεί χρόνο, γνώση και διαθεσιμότητα υλικού. Πράγματα που δεν ήταν διαθέσιμα σε μεγάλο βαθμό κατά την δημιουργία της δικιάς μας βάσης που χρησιμοποιήθηκε για την εκπαίδευση των αλγορίθμων μας.

Η ποιότητα των εικόνων που χρησιμοποιήσαμε δεν ήταν η επιθυμητή, και οι κανόνες που λάβαμε υπόψιν δεν ήταν κατοχυρωμένοι. Χρειάζεται περαιτέρω μελέτη για την κατασκευή μια σωστής βάσης δεδομένων ώστε να μπορέσουμε να την χρησιμοποιήσουμε και να γνωρίζουμε ότι δεν επηρεάζει αρνητικά την έκβαση των αποτελεσμάτων μας. Μελλοντικά θα ήταν καλό να ασχοληθούμε με την καταγραφή και ταξινόμηση των εκφράσεων ενός καλλιτέχνη, καθώς είναι αρκετά πολύπλοκο έργο, και ζητά την πλήρες αφοσίωση ενός ή και παραπάνω ατόμου.

### 6.2.2 SIFT-based Neural Network

Το μεγαλύτερο επίτευγμα της παρούσας πτυχιακής εργασίας ήταν η απόδειξη ότι ένας αλγόριθμος όπως είναι ο SIFT μπορεί να χρησιμοποιηθεί για να ταξινομή τα ανθρώπινα συναισθήματα, ασχέτως δυσκολίας. Τα αποτελέσματα που λάβαμε ήταν πολύ ενθαρρυντικά, ειδικά όταν αναλογιστεί κανείς την φύση του προβλήματος και τον πειραματικό του χαρακτήρα. Φυσικά πάντα υπάρχουν περιθώρια βελτίωσης και σε μια τέτοια περίπτωση που λαμβάνουμε θετικά δείγματα λειτουργικότητας για τον συγκεκριμένο αλγόριθμο σε αυτόν τον τομέα, είναι μια από τις πρώτες μας υποχρεώσεις να ασχοληθούμε στο μέλλον με την ορθή ανάπτυξη ενός συστήματος που θα κάνει χρήση αυτής της τεχνογνωσίας.

Το νευρωνικό δίκτυο που βασίστηκε στα διανύσματα SIFT πρέπει να αναλυθεί περισσότερο και να βελτιωθεί ώστε το ποσοστό επιτυχίας του δικτύου να είναι καλύτερο. Μπορούν να δοκιμαστούν και παρεμφερή αλγόριθμοι που έχουν εμπνευστεί ή βασιστεί από αυτό του SIFT, τέτοιο όπως είναι οι SURF και ASIFT.

## 6.3 Εφαρμογές Προγράμματος και Τεχνογνωσίας

Προγράμματα παρεμφερή με αυτά που παρουσιάστηκαν σήμερα μπορούν να χρησιμοποιηθούν για να καλυτερεύσουν κατά πολύ την ζωή των χρηστών των προσωπικών υπολογιστών αλλά και μεγάλες επιχειρήσεις. Η ενθυλάκωση ενός τέτοιου προγράμματος μέσα σε ένα πρόγραμμα αναπαραγωγής βίντεο θα μπορούσε να κάνει αυτοματοποιημένη την ταξινόμηση του μουσικού βίντεο κλιπ που αναπαράγει και όταν του ζητηθεί μέσω του χρήστη στο μέλλον να αναπαράγει μόνο αυτά τα βίντεο κλιπ που αρμόζουν στην συναισθηματική κατάσταση του.

Μεγάλες βάσεις δεδομένων με μουσικά βίντεο κλιπ, όπως είναι το youtube, θα μπορούσαν να χρησιμοποιήσουν τέτοια προγράμματα και να ταξινομήσουν τα κομμάτια τους αναλόγως την

συναισθηματική τους έκφραση και έπειτα από επιλογή του χρήστη να τους προτείνει κομμάτια που να είναι σύμφωνα με την διάθεση τους. Πλέον οι βάσεις δεδομένων θα μπορούσαν να αυτοματοποιήσουν την ταξινόμηση και να την συνδυάσουν με την ήδη υπάρχουσα για να επαληθεύσουν την ορθότητα των αποτελεσμάτων.

Τέλος ο πολλά υποσχόμενος αλγόριθμος SIFT θα μπορούσε χρησιμοποιηθεί σε πολλές άλλες περιπτώσεις ανίχνευσης συναισθημάτων μέσω των εκφράσεων του προσώπου, όπως στην συνομιλία χρηστών μέσω προγραμμάτων συνομιλίας και την αυτόματη αποστολή emoticons με το αντίστοιχο συναίσθημα που εντοπίζεται στον χρήστη, την προσομοίωση του φωτισμού αναλόγως με την διάθεση του και πολλές άλλες εφαρμογές που σκοπό θα είχαν να βελτιώσουν και να ευκολύνουν την καθημερινότητα του κάθε χρήστη που τις εφαρμόζει.

## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] "ΠΡΟΗΓΜΕΝΕΣ ΤΕΧΝΙΚΕΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΙΩΝ ΚΑΙ ΑΝΑΛΥΣΗΣ ΕΚΦΡΑΣΕΩΝ", Νικόλαος Α. Τσαπατσούλης
- [2] "Face Recognition Algorithms", Ion Marqu'és, Retrieved 16 June 2010
- [3] "Facial Recognition Applications". Anometrics. Retrieved 2008-06-04
- [4] "Rapid Object Detection using a Boosted Cascade of Simple Features", Paul Viola, Michael Jones, Retrieved 2001
- [5] "Filipe Tomaz face detection and recognition". W3.ualg.pt. Retrieved 2011-02-15.
- [6] "Elisabeth Zetterholm, Voice Imitation. A Phonetic Study of Perceptual Illusions and Acoustic Success". Phd thesis, Lund University. Retrieved 2003
- [7] "*Eigenface-based facial recognition*", D. Pissarenko, Retrieved 2003
- [8] "Αναγνώριση μορφασμών με δάση τα ενεργά σημεία του προσώπου", Βασίλειου Στασινόπουλου, Retrieved Ιούνιος 2007
- [9] "Distinctive image features from scale-invariant keypoints" David G. Lowe, Retrieved 2004
- [10] "Reading Faces with Conditional Random Fields", Javier Hernandez Rivera and Tomas Simon Kreuz
- [11] "Facial Expression Analysis", David Matsumoto and Paul Ekman
- [12] <http://www.commercialsecuritydevices.com>
- [13] <http://www.technovelgy.com/ct/Technology-Article.asp?ArtNum=16>
- [14] <http://www.face-rec.org/algorithms/>
- [15] <http://www.mathworks.com/products/computer-vision/demos.html>
- [16] Sathyanarayana, Shashi (2010). "Pattern Recognition Primer II". *Wolfram Demonstrations Project*
- [17] [McLachlan](#) (2004) *Discriminant Analysis and Statistical Pattern Recognition* In: *Wiley Interscience*
- [18] face detection: <http://www.mathworks.com/matlabcentral/fileexchange/20976>
- [19] Hall P, Park BU, Samworth RJ (2008). "Choice of neighbor order in nearest-neighbor classification". *Annals of Statistics*
- [20] face detection: <http://www.mathworks.com/matlabcentral/fileexchange/11073-face-detection-system>
- [21] face detection: <http://www.mathworks.com/matlabcentral/fileexchange/13716-face-eye-detection>
- [22] face detection: <http://www.mathworks.com/matlabcentral/fileexchange/29834-face-detection-using-support-vector-machine-svm>
- [23] face detection: <http://www.mathworks.com/matlabcentral/fileexchange/13701>