Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

Σχολή Τεχνολογικών Εφαρμογών

Τμήμα Μηχανικών Πληροφορικής

Πτυχιακή Εργασία

Τίτλος: Ταυτοποίηση ατόμου με συνδυασμό βίντεο, εικόνας και ήχου

Ονοματεπώνυμο: Γκίκας Στέφανος (ΑΜ: 3650)

Δημητριάδης Αβταντίλ (ΑΜ: 3360)

Επιβλέπων εκπαιδευτικός: Μαριάς Κωνσταντίνος

Τσικνάκης Μανώλης

Επιτροπή Αξιολόγησης: Μαριάς Κωνσταντίνος

Παπαδάκης Νικόλαος

Τσικνάκης Μανώλης

Ημερομηνία Παρουσίασης: 21-09-2018

Ευχαριστίες

Θα θέλαμε να ευχαριστήσουμε θερμά τους επιβλέποντες καθηγητές μας, τον κύριο Μαριά Κώστα και τον κύριο Τσικνάκη Μανώλη που μας έδωσαν την ευκαιρία να εργαστούμε πάνω στην παρούσα πτυχιακή εργασία όπου με την καθοδήγησή τους φέραμε εις πέρας.

Επίσης, θα θέλαμε να ευχαριστήσουμε τον κύριο Καραμπίδη Κώστα για τις συμβουλές του πάνω σε θέματα που σχετίζονται με τον τομέα της Μηχανικής Μάθησης.

Τον κύριο Ξεζωνάκη Ιωάννη, ο οποίος με τον καλύτερο δυνατό τρόπο μας δίδαξε την έννοια του προγραμματισμού και μας ενέπνευσε στα πρώτα χρόνια της φοιτητικής μας σταδιοδρομίας.

Τις οικογένειές μας για την στήριξη τους κατά την διάρκεια των φοιτητικών μας χρόνων.

Abstract

In the 21st century, security is undoubtedly one of the main issues of employment and research of government agencies, business groups and individuals where they want to protect data integrity, the integrity of people and anything that can be considered as requiring protection. One of the most advanced technologies in the security industry is biometrics.
Now, biometrics is considered to be the most advanced method of identifying and certifying the identity of a person. The reason why it renders it superior to other methods is the use of natural or genetic features of people like fingerprints, eye iris, facial features, DNA, and others. An additional secondary reason for which biometrics is preferred is the convenience with which it can be applied and used in commercial systems such as mobile phones and laptops in relation to methods such as codes or motifs.

This diploma thesis explores two of the biometric identification methods, namely facial recognition and voice recognition. In the part of face detection was used the open source code toolkit OpenFace, in the part of face recognition the method of principal component analysis (PCA) and in the part of voice recognition the Mel-Frequency Ceptrum Coefficients (MFCC).

The paper focuses on the study and comparison of methods based on two unimodal models, face recognition and voice recognition, and a multimodal model combining the two above. It also develops two different algorithmic pattern recognition methods, the similarity classifier and the neural network.

Σύνοψη

Στον 21^{ο} αιώνα όπου διανύουμε, ο τομέας της ασφάλειας είναι αναμφισβήτητα ένα από τα κατεξοχήν θέματα απασχόλησης και έρευνας κρατικών υπηρεσιών, επιχειρηματικών ομίλων αλλά και ιδιωτών όπου θέλουν να προστατεύσουν την ακεραιότητα δεδομένων, την ακεραιότητα φυσικών προσώπων και οτιδήποτε μπορεί να θεωρηθεί ότι χρίζει προστασίας. Μία από τις πλέον αναπτυσσόμενες τεχνολογίες στον κλάδο της ασφάλειας είναι η βιομετρία. Πλέον, η βιομετρία θεωρείται η πιο προηγμένη μέθοδος ταυτοποίησης και πιστοποίησης της ταυτότητας ενός ατόμου. Ο λόγος για τον οποίο την καθιστά υπερέχουσα έναντι άλλων μεθόδων είναι η χρήση φυσικών ή γενετικών χαρακτηριστικών των ίδιων των ατόμων όπως τα δακτυλικά αποτυπώματα, η ίριδα του ματιού, τα χαρακτηριστικά του προσώπου, το DNA κα. Ένας επιπλέον δευτερεύων λόγος, για τον οποίο η βιομετρία προτιμάται είναι η ευκολία με την οποία μπορεί να εφαρμοσθεί και να χρησιμοποιηθεί σε εμπορικά συστήματα όπως κινητά τηλέφωνα, και φορητοί ηλεκτρονικοί υπολογιστές σε σχέση με μεθόδους όπως είναι οι κωδικοί ή τα μοτίβα.

Η παρούσα πτυχιακή εργασία, μελετά δύο από τις βιομετρικές μεθόδους ταυτοποίησης, που είναι η αναγνώριση προσώπου και η αναγνώριση φωνής. Στο κομμάτι του εντοπισμού προσώπου χρησιμοποιήθηκε η ανοιχτού κώδικα εργαλειοθήκη OpenFace, στο κομμάτι της αναγνώρισης προσώπου η μέθοδος της ανάλυσης κυρίων συνιστωσών (PCA) και στο κομμάτι της αναγνώρισης φωνής οι συντελεστές συχνότητας Ceptrum του Mel (Mel-Frequency Ceptrum Coefficients, MFCC).

Η εργασία εστιάζει στην μελέτη και σύγκριση μεθόδων που βασίζονται σε δύο μονοτροπικά μοντέλα ταυτοποίησης (unimodal model), αναγνώρισης προσώπου και αναγνώρισης φωνής και σε ένα πολυτροπικό μοντέλο ταυτοποίησης (multimodal model) όπου συνδυάζει τα δύο παραπάνω. Επίσης γίνεται ανάπτυξη δύο διαφορετικών αλγοριθμικών μεθόδων αναγνώρισης προτύπων, του ταξινομητή ομοιότητας και του νευρωνικού δικτύου.

Πίνακας περιεχομένων

Ευχαριστίες ii

Abstract iii

Σύνοψη iv

Λίστα εικόνων viii

Λίστα Πινάκων x

Λίστα Γραφημάτων x

Εισαγωγή 1

1.1 Περίληψη 3

1.2 Κίνητρο για την Διεξαγωγή της Εργασίας – Στόχοι 4

1.3 Δομή εργασίας 4

2 Μεθοδολογία Υλοποίησης 5

2.1 Μέθοδος Ανάλυσης & Ανάπτυξης 5

3 Επισκόπηση Τεχνικών 8

3.1 Εντοπισμός Προσώπου 8

3.1.1 Μέθοδοι με βάση τα χαρακτηριστικά 10

3.1.1.1 Ανάλυση χαμηλού επιπέδου 11

3.1.1.1.1 Ακμές 11

3.1.1.1.2 Κίνηση 12

3.1.1.1.3 Γενικευμένες μετρήσεις 12

3.1.1.1.4 Επίπεδα γκρι 13

3.1.1.1.5 Χρώμα 14

3.1.1.2 Ανάλυση χαρακτηριστικών 15

3.1.1.2.1 Αναζήτηση χαρακτηριστικών 16

3.1.1.2.2 Ανάλυση συστοιχιών 17
3.1.1.3 Μοντέλα ενεργού σχήματος 17
3.1.1.3.1 Snakes 18
3.1.1.3.2 PDM 19
3.1.2 Μέθοδοι με βάση την εικόνα 19
3.1.2.1 Viola-Jones 20
3.1.2.2 Μέθοδοι γραμμικών υποχωρών 22
3.1.2.3 Νευρωνικά Δίκτυα 22
3.1.2.4 Support Vector Machines 25
3.2 Αναγνώριση Προσώπου 26
3.2.1 Δυσδιάστατες μέθοδοι αναγνώρισης 27
3.2.1.1 Στατιστικές Προσεγγίσεις 27
3.2.1.1.1 Ανάλυση κυρίων συνιστωσών (PCA) 28
3.2.1.1.2 Ανάλυση γραμμικού διαχωρισμού (LDA) 30
3.2.1.1.3 Κυμματίδιο Gabor 31
3.2.1.2 Νευρωνικά δίκτυα 32
3.2.1.2.1 Νευρωνικά δίκτυα με χρήση φίλτρων Gabor 32
3.2.1.2.2 Νευρωνικά δίκτυα με μοντέλα Hidden Markov 33
3.2.1.2.3 Ασαφή νευρωνικά δίκτυα 34
3.2.2 Τρισδιάστατες μέθοδοι αναγνώρισης 35
3.2 Αναγνώριση φωνής 36
3.2.1 Μετασχηματισμοί σήματος 37
3.2.1.1 Cross Correlation 37
3.2.1.2 Discrete Laplacian Transform 37
3.2.1.3 Envelope 38
3.2.1.4 Fast Fourier Transform 38
3.2.1.5 Hilbert Transform 38
3.2.1.6 MFCC 38
3.2.1.7 Wavelets 39
3.3 Συναρτήσεις απόφασης 39
3.3.1 Μέτρα απόστασης 39
3.3.1.1 Ευκλείδεια απόσταση 39
3.3.1.2 Ιπποδάμεια απόσταση 40
3.3.1.3 Hamming απόσταση 40
3.3.1.4 Chebyshev απόσταση 40
3.3.1.5 Mahalanobis απόσταση 41
3.3.2 Μέτρα ομοιότητας 41

3.2.2.1 Εσωτερικό γινόμενο 41
3.2.2.1 Tanimoto 41
3.4 Ένωση βιομετρικών χαρακτηριστικών 42
3.4.1 Προ-ταξινόμηση 43
3.4.2 Μετά-ταξινόμηση 43
4 Σχέδιο Δράσης για την εκπόνηση της πτυχιακής Εργασίας 44
4.1 Βιβλιογραφική ανασκόπηση 44
4.1.1 Ταυτοποίηση μέσω χαρακτηριστικών προσώπου 44
4.1.2 Ταυτοποίηση μέσω χαρακτηριστικών φωνής 45
4.1.3 Ταυτοποίηση μέσω συνδυασμού χαρακτηριστικών 46
5 Κύριο μέρος Πτυχιακής Εργασίας 51
5.1 Η βάση δεδομένων 51
5.2 Πειραματικό μέρος 52
5.2.1 Αναγνώριση προσώπου 52
5.2.1.1 Πειραματικό μέρος 1 52
5.2.1.1 Πειραματικό μέρος 2 54
5.2.1.2 Πειραματικό μέρος 3 58
5.2.1.3 Πειραματικό μέρος 4 61
5.2.1.4 Πειραματικό μέρος 5 63
5.2.2 Αναγνώριση φωνής 64
5.2.2.1 Πειραματικό μέρος 1 64
5.2.3 Multimodal 69
5.2.3.1 Μέθοδος 1 69
5.2.3.1.1 Πειραματικό μέρος 1 70
5.2.3.1.2 Πειραματικό μέρος 2 71
5.2.3.2 Μέθοδος 2 73
5.2.3.2.1 Πειραματικό μέρος 1 74
5.2.3.2.2 Πειραματικό μέρος 2 75
5.2.3.3 Μέθοδος 3 76
5.2.3.3.1 Πειραματικό μέρος 1 76
5.2.3.4 Σύγκριση μεθόδων 77
6.1 Συμπεράσματα 78
6.2 Μελλοντική Εργασία και Επεκτάσεις 78

Βιβλιογραφία 79

Λίστα εικόνων

Εικόνα 1: Είδη βιομετρικών χαρακτηριστικών 2
Εικόνα 2: Γενικό διάγραμμα βιομετρικού συστήματος 6
Εικόνα 3: Βασική μέθοδοι της πτυχιακής εργασίας 6
Εικόνα 4: Εντοπισμός προσώπου μέσω ακμών 11
Εικόνα 5: Το σύστημα εντοπισμού προσώπου του Kalman 12
Εικόνα 6: Εντοπισμός προσώπου από την μελέτη των Reisfeld et al. 13
Εικόνα 7: Το ρομπότ του Wong et al. 14
Εικόνα 8: Εντοπισμός πολλαπλών προσώπων βασισμένο στο RGB μοντέλο 15
Εικόνα 9: Σάρωση (από πάνω προς τα κάτω) για εντοπισμό προσώπου 16
Εικόνα 10: Εφαρμογή ενεργού περιγράμματος Α) αρχική θέση Β) ύστερα από 10 επαναλήψεις Γ) σύγκλιση της αναζήτησης 17
Εικόνα 11: Εφαρμογή των snakes για εύρεση περιγράμματος παλάμης 19
Εικόνα 12: Ορθογώνια χαρακτηριστικά τύπου Harr 21
Εικόνα 13: Χρήση ορθογωνίων χαρακτηριστικών σε εικόνα 21
Εικόνα 14: Εφαρμογή Viola-Jones για εντοπισμό προσώπου και ματιών 21
Εικόνα 15: Η βασική δομή ενός απλού τεχνητού νευρωνικού δικτύου 23
Εικόνα 16: Ο βασικός αλγόριθμος του Rowley για εντοπισμό προσώπων 25
Εικόνα 17: Από την εργασία του Kullback. Παράθυρο 30x30 γύρω από το πρόσωπο και οι αντίστοιχες δυαδικές εικόνες 26
Εικόνα 18: Εφαρμογή PCA 29
Εικόνα 19: Ανακατασκευή εικόνας προσώπου στα αρχικά δεδομένα με χρήση των 10,20 και 30 πιο σημαντικών συστατικών 29
Εικόνα 20: Διαφορά μεταξύ PCA και LDA 31
Εικόνα 21: Αναγνώριση μέσω Gabor σωματιδίων[29] 32
Εικόνα 22: Η αρχιτεκτονική του νευρωνικού δικτύου των Bhuiyan και Liu 33
Εικόνα 23: Πιθανοί παράμετροι ενός κρυμμένου μοντέλου Markov 34
Εικόνα 24: Αρχιτεκτονική του τεχνητού νευρωνικού δικτύου των Bhattachrjee et al. 35
Εικόνα 25: Τριών διαστάσεων καταγραφή προσώπου 36
Εικόνα 26: Παράδειγμα εικόνων εκπαίδευσης για ένα άτομο 53
Εικόνα 27: Παράδειγμα εικόνας ελέγχου για ένα άτομο (video 1) 53
Εικόνα 28: Παράδειγμα εικόνας ελέγχου για ένα άτομο (video 2) 53
Εικόνα 29: Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης προσώπου 54

Εικόνα 30: Εικόνες ύστερα από τον εντοπισμό προσώπου 55
Εικόνα 31: Εικόνες ύστερα από το γέμισμα (padding) 55
Εικόνα 32: Αποτελέσματα πειραματικού μέρους 2 αναγνώρισης προσώπου 56
Εικόνα 33: Χρόνοι εκπαίδευσης και ελέγχου αναγνώρισης προσώπου 57
Εικόνα 34: Αποτελέσματα πειράματος resize 58
Εικόνα 35: Παράδειγμα εικόνων εκπαίδευσης για ένα άτομο 59
Εικόνα 36: Αποτελέσματα πειραματικού μέρους 3 αναγνώρισης προσώπου 59
Εικόνα 37: Παράδειγμα αρχικών εικόνων 60
Εικόνα 38: Παράδειγμα εικόνων ύστερα από εντοπισμό προσώπου 61
Εικόνα 39: Παράδειγμα μετωπικής εικόνας 62
Εικόνα 40: Αποτελέσματα πειραματικού μέρους 4 αναγνώρισης προσώπου (1) 62
Εικόνα 41: Αποτελέσματα πειραματικού μέρους 4 αναγνώρισης προσώπου (2) 63
Εικόνα 42: Αποτελέσματα πειραματικού μέρους 5 αναγνώρισης προσώπου 64
Εικόνα 43: Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης φωνής (1) 66
Εικόνα 44: Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης φωνής (2) 67
Εικόνα 45: Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης φωνής (3) 67
Εικόνα 46: Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης φωνής (4) 68
Εικόνα 47: Χρόνοι εκπαίδευσης και ελέγχου (μετασχηματισμοί σήματος) 68
Εικόνα 48: Multimodal μέθοδος 1 69
Εικόνα 49: Αποτελέσματα πειραματικού μέρους 1 (multimodal, μέθοδος 1) 71
Εικόνα 50: Αποτελέσματα πειραματικού μέρους 2 (multimodal, μέθοδος 1) 72
Εικόνα 51: Χρόνοι εκπαίδευσης και ελέγχου (multimodal, μέθοδος 1) 72
Εικόνα 52: Αποτελέσματα πειραματικού μέρους 1 (multimodal, μέθοδος 2) 74
Εικόνα 53: Αποτελέσματα πειραματικού μέρους 2 (multimodal, μέθοδος 2) 75
Εικόνα 54: Αρχιτεκτονική backpropagation νευρωνικού δικτύου 76
Εικόνα 55: Αποτελέσματα πειραματικού μέρους 1 (multimodal, μέθοδος 3) 76

Λίστα Πινάκων

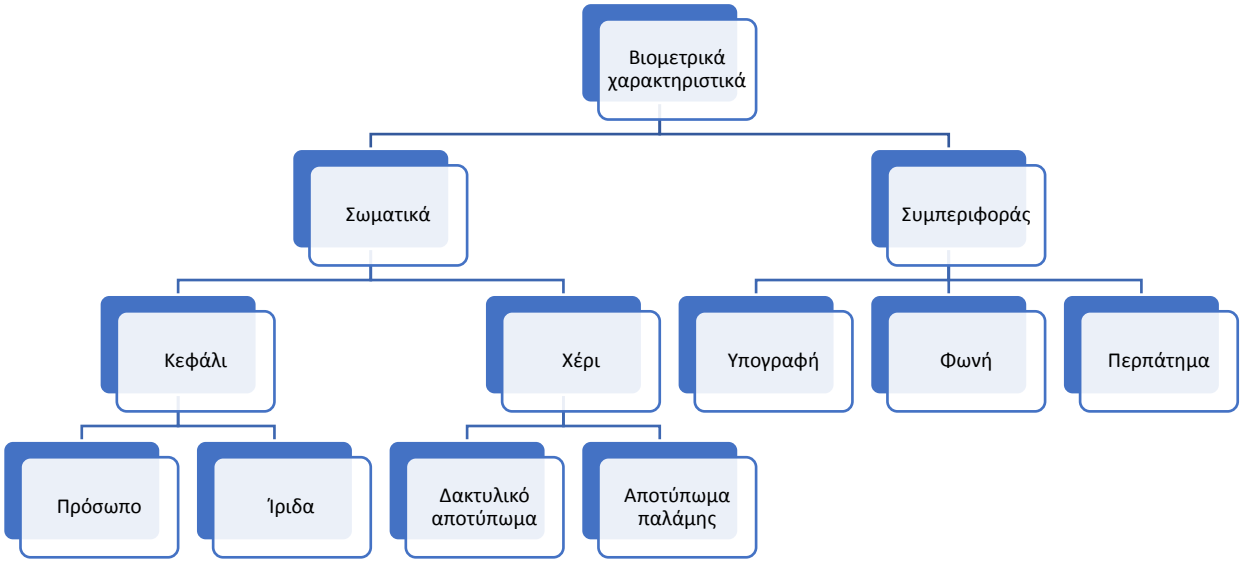
Πίνακας 1: Επισκόπηση ερευνών για ταυτοποίηση μέσω χαρακτηριστικών προσώπου 48
Πίνακας 2: Επισκόπηση ερευνών για ταυτοποίηση μέσω χαρακτηριστικών φωνής 49
Πίνακας 3: Επισκόπηση ερευνών για ταυτοποίηση μέσω συνδυασμού χαρακτηριστικών 50
Πίνακας 4: Επισκόπηση ερευνών για ταυτοποίηση μέσω συνδυασμού χαρακτηριστικών (πρόσωπο-φωνή) 50
Πίνακας 5: Συγκριτικός πίνακας μεθόδων 77

Λίστα Γραφημάτων

Γράφημα 1: Κατηγορίες βιομετρικών χαρακτηριστικών 1
Γράφημα 2: Μέθοδοι εντοπισμού προσώπου 9
Γράφημα 3: Ταξινόμηση νευρωνικών αλγορίθμων [19] 24
Γράφημα 4: Μέθοδοι αναγνώρισης προσώπου 27
Γράφημα 5: Βασική δομή του μοντέλου των Bhattachrjee et al. 34
Γράφημα 6: Προσεγγίσεις ένωσης πληροφορίας [37] 42

Εισαγωγή

Η ταυτοποίηση ατόμων μέσω βιομετρικών χαρακτηριστικών τις δύο τελευταίες δεκαετίες αναπτύσσεται με ραγδαίους ρυθμούς, γεγονός το οποίο οφείλεται πρώτον στην ανάγκη προστασίας είτε ευαίσθητων δεδομένων καθώς πλέον οι περισσότερες πληροφορίες που διακινούνται είναι ψηφιοποιημένες είτε στην προστασία ατόμων-χώρων και δεύτερον στην κατάλληλη τεχνολογική διαθεσιμότητα η οποία επιτρέπει την μελέτη και την πλήρη ανάπτυξη ολοκληρωμένων βιομετρικών συστημάτων ασφαλείας. Οι τεχνολογίες οι οποίες



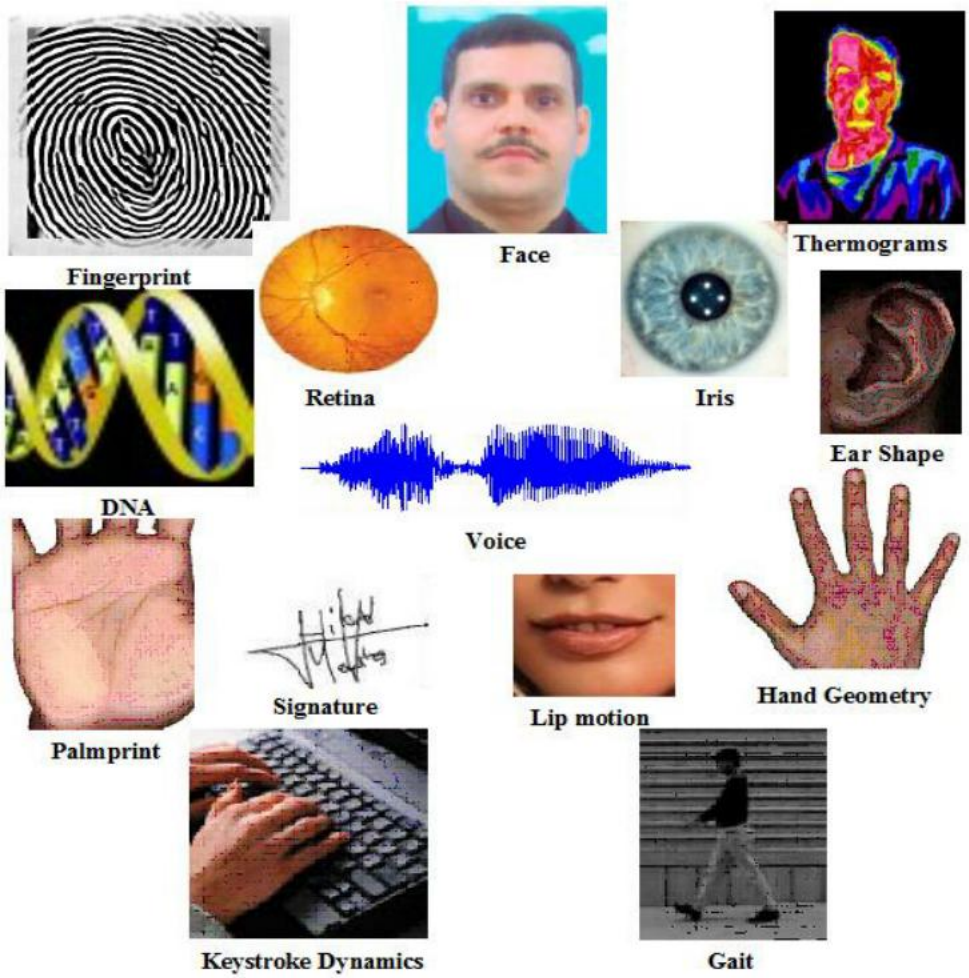
Γράφημα 1: *Κατηγορίες βιομετρικών χαρακτηριστικών*

έχουν αναπτυχθεί ως προς την βιομετρική ταυτοποίηση είναι ποικίλες. Ένας κύριος διαχωρισμός

αυτών είναι στον τύπο των χαρακτηριστικών όπου χρησιμοποιούν. Τεχνολογίες που βασίζονται σε σωματικά χαρακτηριστικά (ίριδα, πρόσωπο, δακτυλικό αποτύπωμα) και σε τεχνολογίες που βασίζονται σε χαρακτηριστικά συμπεριφοράς (περπάτημα, υπογραφή).

Λόγω του πλήθους των πλεονεκτημάτων που κατέχουν οι βιομετρικές μέθοδοι ασφαλείας όπως είναι το υψηλό ποσοστό σωστής ταυτοποίησης το οποίο βέβαια εξαρτάται από την εκάστοτε μέθοδο, έχουν καταστήσει την βιομετρία ως την ανώτερη μορφή ασφάλειας. Σχεδόν όλα τα εθνικά και διεθνή συστήματα ασφάλειας κάνουν χρήση τέτοιων

μεθόδων. Τα γενετικά βιομετρικά στοιχεία χρησιμοποιούνται κατά αποκλειστικότητα για την αναγνώριση ποινικών αδικημάτων καθώς η ύπαρξη εσφαλμένης ταυτοποίησης είναι σχεδόν μηδενική. Βέβαια, με την ανάπτυξη μεθόδων και συστημάτων βιομετρικών χαρακτηριστικών εγείρονται και πολλά δεοντολογικά ερωτήματα και ζητήματα. Μερικά από αυτά είναι, με ποιον τρόπο να δημιουργήσουμε μία βάση βιομετρικών δεδομένων και πώς θα μπορέσουμε να την χρησιμοποιήσουμε με κατάλληλο τρόπο έτσι ώστε να διατηρηθεί η προσωπική ελευθερία του ατόμου. Ο σκοπός είναι μπορέσουμε να επιτύχουμε μία ισορροπία μεταξύ της ασφάλειας δεδομένων-συναλλαγών-ατόμων και της προστασίας της ατομικής ελευθερίας-ιδιωτικότητας. [1]



Εικόνα 1: *Είδη βιομετρικών χαρακτηριστικών*[1]

[1] https://www.researchgate.net/figure/Show-most-of-the-physiological-anatomical-and-behavioral-characteristics-that-are-being_fig1_281830547

1.1 Περίληψη

Σκοπός της παρούσας πτυχιακής εργασίας είναι η μελέτη unimodal και multimodal μεθόδων ταυτοποίησης καθώς και η σύγκριση όλων αυτών (αναγνώριση προσώπου, αναγνώριση φωνής, συνδυασμός προσώπου-φωνής). Επίσης, έγινε εκτενής μελέτη του αλγορίθμου PCA για το πώς αναπτύσσεται (μη κάνοντας χρήση έτοιμων εργαλειοθηκών) καθώς και πειραματική μελέτη ως προς την αποτελεσματικότητά του σε διάφορες συνθήκες περιβάλλοντος. Μέσα από την μελέτη αυτή, προέκυψαν καθοριστικά συμπεράσματα ως προς την χρήση και τα αποτελέσματα του αλγορίθμου PCA σε ένα σύστημα αναγνώρισης προσώπου.

Πιο συγκεκριμένα, έγινε χρήση του dataset ***VidTIMIT Audio-Video Dataset***[2][2], από το οποίο χρησιμοποιήθηκαν τα αρχεία βίντεο, τα αρχεία εικόνων και τα αρχεία ήχων. Επίσης, έγινε χρήση της ανοιχτής κώδικα εργαλειοθήκης **OpenFace**[3][3] με σκοπό να πραγματοποιηθεί ανίχνευση των προσώπων στις εικόνες. Για την αναγνώριση προσώπου αναπτύχθηκε ο αλγόριθμος PCA όπως αναφέρθηκε παραπάνω. Για την αναγνώριση φωνής έγινε χρήση διαφορετικών μετασχηματισμών σήματος και σύγκρισης αυτών, συγκεκριμένα:

- Cross correlation
- Discrete Laplacian Transform
- Envelope
- Fast Fourier Transform
- Hilbert Transform
- MFCC
- Wavelets Decomposition
- Wavelets Transform

Επιπλέον, έγινε ανάπτυξη και συνδυασμός των δύο μεθόδων ταυτοποίησης (πρόσωπο-φωνή) και σύγκρισης αυτών. Για την τελική απόφαση της ταυτοποίησης αναπτύχθηκαν δύο διαφορετικά συστήματα απόφασης, ένας **ταξινομητής ομοιότητας** και ένα **νευρωνικό δίκτυο.**

[2] http://conradsanderson.id.au/vidtimit/ - related_datasets

[3] https://www.cl.cam.ac.uk/~tb346/res/openface.html

1.2 Κίνητρο για την Διεξαγωγή της Εργασίας – Στόχοι

Κίνητρο της παρούσας πτυχιακής εργασίας ήταν η μελέτη και η ανάπτυξη βιομετρικών μεθόδων ταυτοποίησης και εξέταση των διαφόρων παραγόντων που ενδεχομένως να επηρεάζουν τα αποτελέσματα. Στόχος της εργασίας αυτής είναι η εξαγωγή χρήσιμων συμπερασμάτων ως προς την χρήση συγκεκριμένων μεθόδων ταυτοποίησης.

1.3 Δομή εργασίας

Κεφάλαιο πρώτο: Εισαγωγές πληροφορίες για την παρούσα πτυχιακή εργασία

Κεφάλαιο δεύτερο: Μέθοδος που ακολουθήθηκε για έρθει εις πέρας η παρούσα πτυχιακή εργασία

Κεφάλαιο τρίτο: Αναφορά τεχνικών ως προς εντοπισμό προσώπου, αναγνώριση προσώπου, αναγνώριση φωνής, συναρτήσεων απόφασης και ένωσης βιομετρικών χαρακτηριστικών

Κεφάλαιο τέταρτο: Μελέτη των State of the art ερευνητικών εργασιών

Κεφάλαιο πέμπτο: Πειραματικό κομμάτι της παρούσας πτυχιακής εργασίας

Κεφάλαιο έκτο: Συμπεράσματα και μελλοντική εργασία

2 Μεθοδολογία Υλοποίησης

2.1 Μέθοδος Ανάλυσης & Ανάπτυξης

Για να μπορέσει να αντιμετωπιστεί το πρόβλημα της ταυτοποίησης ατόμου, χρειάστηκε να διαχωρισθεί σε επιμέρους στάδια, καθότι οι προσεγγίσεις και οι μέθοδοι αυτών υλοποιούν διαφορετικούς τύπους αλγορίθμων, καθώς επίσης το κάθε επιμέρους στάδιο δέχεται αναγκαίες πληροφορίες από το προηγούμενο για να επιτευχθεί εν τέλη η ταυτοποίηση.

Τα βασικότερα στάδια της μεθοδολογίας είναι:

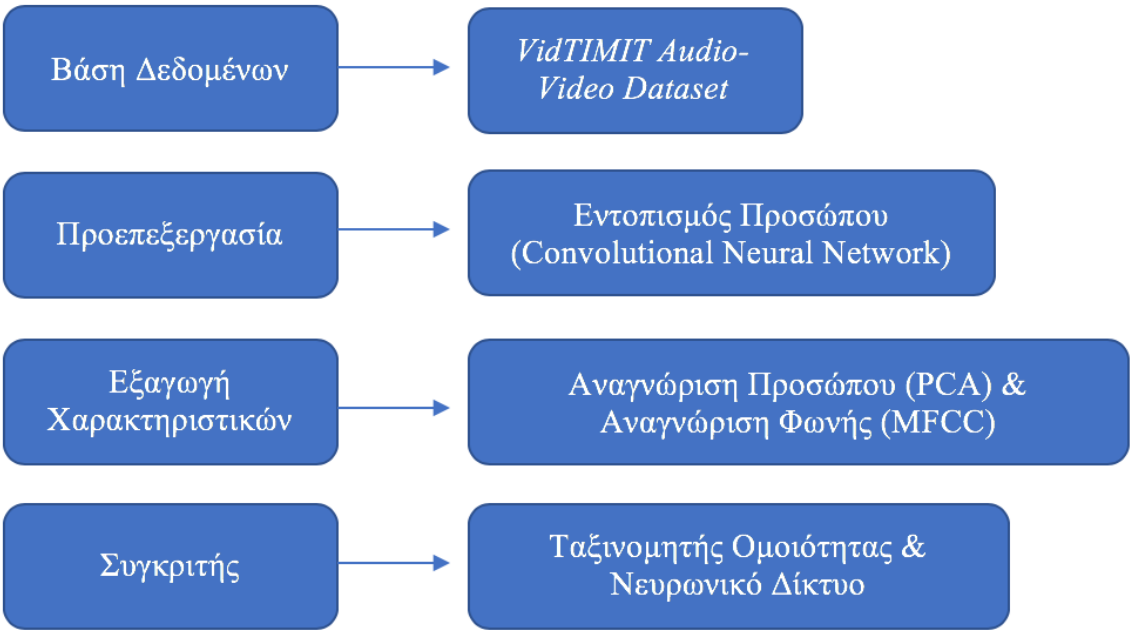
- **Βάση δεδομένων:** το σύνολο των αναγκαίων πληροφοριών (βίντεο, εικόνες, ήχοι) στα οποία θα βασιστεί η όλη μελέτη της παρούσας εργασίας. Είναι τα δεδομένα εκείνα που θα επεξεργαστούμε έτσι ώστε να καταλήξουμε ύστερα σε συμπεράσματα σχετικά με την ταυτοποίηση μέσω unimodal και multimodal συστημάτων. * Κάθε βίντεο, εικόνα ή ήχος το ονομάζουμε αντικείμενο.
- **Προεπεξεργασία:** η απαραίτητη επεξεργασία των δεδομένων της βάσης, έτσι ώστε να είναι δυνατή η χρήση τους (αλλαγή μεγέθους εικόνας, μετατροπή RGB προτύπου σε grayscale , μείωση θορύβου κλπ.)
- **Εξαγωγή χαρακτηριστικών:** η διαδικασία κατά την οποία επιλέγονται στοιχεία από τα προεπεξεργασμένα δεδομένα έτσι ώστε να υπάρξει μείωση του υπολογιστικού κόστους.
- **Δημιουργία προτύπων:** ύστερα από την εξαγωγή χαρακτηριστικών, δημιουργείται διαφορετική αναπαράσταση του κάθε αντικειμένου αποτελούμενο πλέον από χαρακτηριστικά όπου εξήχθησαν στο προηγούμενο στάδιο.
- **Συγκριτής:** είναι το υποσύστημα εκείνο όπου κάνει χρήση αλγορίθμων με βάση των οποίων γίνεται η σύγκριση των προτύπων της βάσης με τα εισερχόμενα πρότυπα και προκύπτει το αποτέλεσμα της ταυτοποίησης.
- **Αποτέλεσμα Ταυτοποίησης:** παρουσίαση του αποτελέσματος της ταυτοποίησης και εκτέλεση της αντίστοιχης ενέργειας από το σύστημα (πχ επιτρέπεται η είσοδος, απαγορεύεται η είσοδος)

Αναλυτικότερα, χρησιμοποιήθηκε μία υπάρχουσα βάση δεδομένων η *VidTIMIT Audio-Video Dataset* από την οποία χρησιμοποιήθηκαν τα δεδομένα της, συγκεκριμένα οι εικόνες και οι ήχοι. Στο στάδιο της προεπεξεργασίας χρειάσθηκε να εφαρμοσθεί τεχνική εντοπισμού

το προσώπου στην εικόνα. Συγκεκριμένα χρησιμοποιήθηκε το toolkit *face_detect* της ανοιχτής κώδικα εργαλειοθήκης ***OpenFace*** στο οποίο έχει υλοποιηθεί και εκπαιδευτεί **Συνελικτικό Νευρωνικό Δίκτυο** (Convolutional Neural Network) το οποίο αποτελεί είδος των Βαθιών Νευρωνικών Δικτύων (Deep Neural Network). Στην συνέχεια, στο στάδιο εξαγωγή χαρακτηριστικών και συγκεκριμένα στη αναγνώριση προσώπου αναπτύχθηκε και εφαρμόσθηκε η μέθοδος **Ανάλυση Κυρίων Συνιστωσών** (PCA) ενώ στην αναγνώριση φωνής εφαρμόσθηκε οι **Συντελεστές Συχνότητας Ceptrum του** Mel (Mel-Frequency Ceptrum Coefficients, MFCC). Τέλος, στο στάδιο του συγκριτή αναπτύχθηκε ένας

Ταξινομητής Ομοιότητας (similarity classifier) καθώς έγινε και προσέγγιση με ένα **Τεχνητό Νευρωνικό Δίκτυο Backpropagation**.

Εικόνα 2: *Γενικό διάγραμμα βιομετρικού συστήματος*



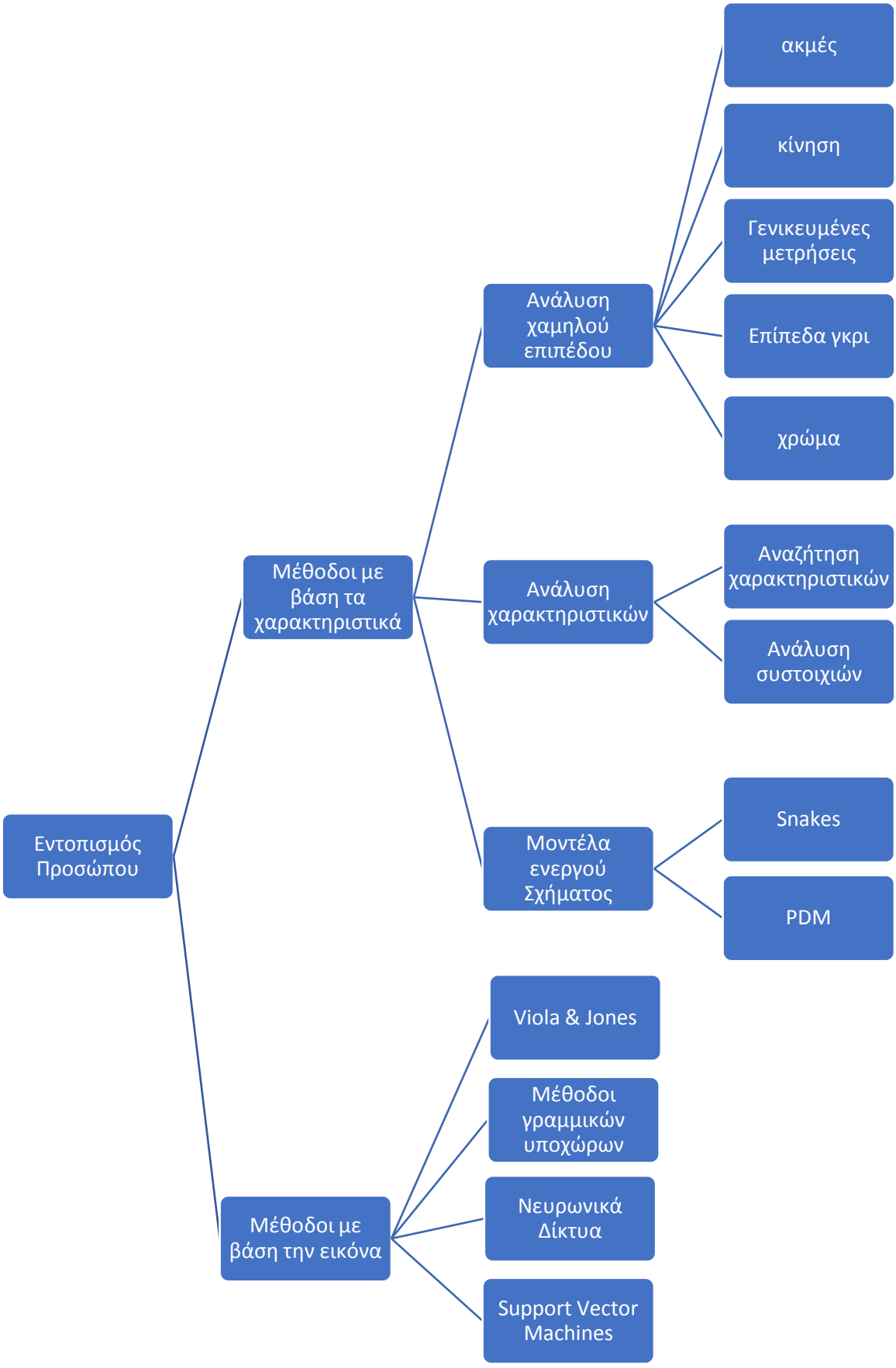
Εικόνα 3: *Βασική μέθοδοι της πτυχιακής εργασίας*

3 Επισκόπηση Τεχνικών

Κατά την διάρκεια της εκπόνησης της παρούσας πτυχιακής εργασίας, μελετήθηκαν ποικίλοι μέθοδοι και αλγοριθμικές προσεγγίσεις ως προς θέματα του εντοπισμού προσώπου, της αναγνώρισης προσώπου , αναγνώρισης φωνής και ταξινόμησης προτύπων. Παρακάτω θα παρουσιαστούν οι πιο αντιπροσωπευτικές από αυτές.

3.1 Εντοπισμός Προσώπου

Η διαδικασία του εντοπισμού προσώπου αποτελεί το πρώτο βασικό στάδιο της για την ανάπτυξη ενός συστήματος ταυτοποίησης μέσω εικόνας προσώπου. Οι τεχνικές εντοπισμού προσώπου διακρίνονται σε δύο βασικές κατηγορίες οι οποίες ξεχωρίζουν ως προς διαφορετικό τρόπο προσέγγισης στην χρήση του προσώπου. Η πρώτη κατηγορία περιγράφεται ως **προσέγγιση με βάση τα χαρακτηριστικά (feature-based)** και κάνει χρήση την υπάρχουσας γνώσης της δομής και της όψης του προσώπου. Η δεύτερη κατηγορία περιγράφεται ως **προσέγγιση με βάση την εικόνα (image-based)** στην οποία δεν γίνεται εξαγωγή χαρακτηριστικών αλλά τα πρόσωπα αναπαρίστανται σε πίνακες εντάσεων. [4]



Γράφημα 2: *Μέθοδοι εντοπισμού προσώπου*

3.1.1 Μέθοδοι με βάση τα χαρακτηριστικά

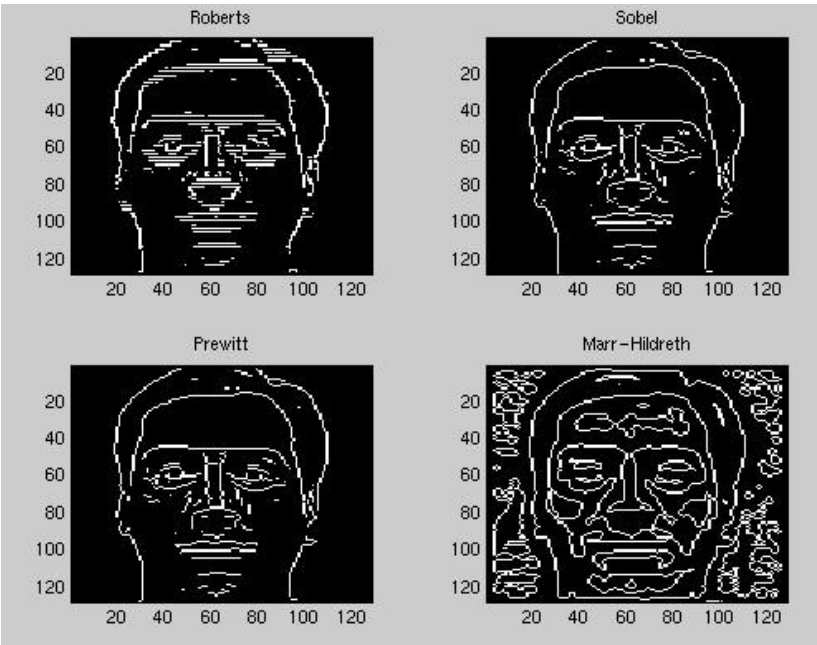
Οι μέθοδοι που βασίζονται στα χαρακτηριστικά, κάνουν γενικότερη χρήση της γνώσης της γεωμετρίας του προσώπου βάση της οποίας μοντέλα τα οποία έχουν την ικανότητα να εντοπίζουν και να ανιχνεύουν την ύπαρξη προσώπων στον χώρο. Τα πρωταρχικά στοιχεία στα οποία στηρίζονται είναι η βασική γνώση του ότι ένα πρόσωπο έχει συγκεκριμένα χαρακτηριστικά, όπως μάτια, στόμα, μύτη καθώς επίσης την θέση στην οποία είναι τοποθετημένα. Οι πρώτες απόπειρες ανάπτυξης τέτοιων μοντέλων ξεκίνησαν την δεκαετία του 60’ αλλά δεν υπάρχουν αναλυτικές δημοσιεύσεις αποτελεσμάτων. Αργότερα, την δεκαετία του 90’ παρουσιάστηκε ένα ολοκληρωμένο σύστημα αναγνώρισης ατόμων, στο οποίο η διαδικασία εντοπισμού του προσώπου γινόταν με χρήση wavelets [5] [6]

Οι μέθοδοι που κάνουν χρήση χαρακτηριστικών διακρίνονται σε τρεις κατηγορίες. Η πρώτη είναι η ανάλυση χαμηλού επιπέδου, η οποία χρησιμοποιεί τις ιδιότητες των pixels όπως είναι το χρώμα και η κλίμακα του γκρι. Αυτό βέβαια, σε πολλές περιπτώσεις λόγω των συνθηκών του περιβάλλοντος μέσα στο οποίο βρίσκεται το πρόσωπο καθιστά την διαδικασία του εντοπισμού δύσκολη καθώς τα χαρακτηριστικά που εξάγονται είναι ασαφή. Η δεύτερη κατηγορία είναι ανάλυση χαρακτηριστικών όπου κάνει χρήση της γεωμετρίας και των χαρακτηριστικών του προσώπου. Αυτή η προσέγγιση συγκριτικά με την προηγούμενη επιφέρει συνήθως καλύτερα αποτελέσματα καθώς μειώνει την ασάφεια των χαρακτηριστικών που προκύπτουν. Τέλος, η τρίτη κατηγορία είναι τα μοντέλα ενεργού περιγράμματος τα περισσότερα από τα οποία βασίζονται στο μοντέλο *snakes* που προτάθηκε την δεκαετία του 80’ από τους Kass, Witkin, Terzopoulos [7] και πλέον τα πιο σύγχρονα μοντέλα διασπαρμένων σημείων που σχετίζονται με την μέση γεωμετρία του σχήματος. [4]

3.1.1.1 Ανάλυση χαμηλού επιπέδου

3.1.1.1.1 Ακμές

Η ανίχνευση ακμών στην είναι μία μέθοδος που χρησιμοποιείται κατά κόρον για την κατάτμηση εικόνων κάνοντας χρήση των μεταβολών στην ένταση. Τα μοντέλα που στηρίζονται στις ακμές διακρίνονται με βάση τα προφίλ της έντασης που τα χαρακτηρίζει.[8] Μία από τις πρώτες προσεγγίσεις για τον εντοπισμό προσώπου ήταν στην εργασία του Sakai et.al. [9] όπου μελετήθηκε η ανάλυση των γραμμών των προσώπων σε φωτογραφίες. Πριν την δεκαετία του 80’ οι περισσότεροι μέθοδοι ανίχνευσης ακμών στηρίζονταν στην χρήση μικρών τελεστών όπως οι μάσκες Sobel. Στην συνέχεια, ο D. Marr και E. Hildreth [10] ανέπτυξαν την μία νέα μέθοδο εντοπισμού ακμών κάνοντας χρήση πιο πολύπλοκων τεχνικών ανάλυσης. Η μέθοδος αυτή, στηρίζεται στον ισχυρισμό ότι οι μεταβολές την έντασης εξαρτώνται από την κλίμακα της εικόνας. [8] Η μέθοδος των Marr και Hildreth χρησιμοποιείται έως και σήμερα για εντοπισμό του προσώπου. Αργότερα, ο Govindaraju κατάφερε να αναπτύξει μία μέθοδο όπου μπορούσε να εντοπίζει τις ακμές του προσώπου σε φυσικό περιβάλλον και όχι σε ελεγχόμενο όπως ίσχυε μέχρι τότε. [11]

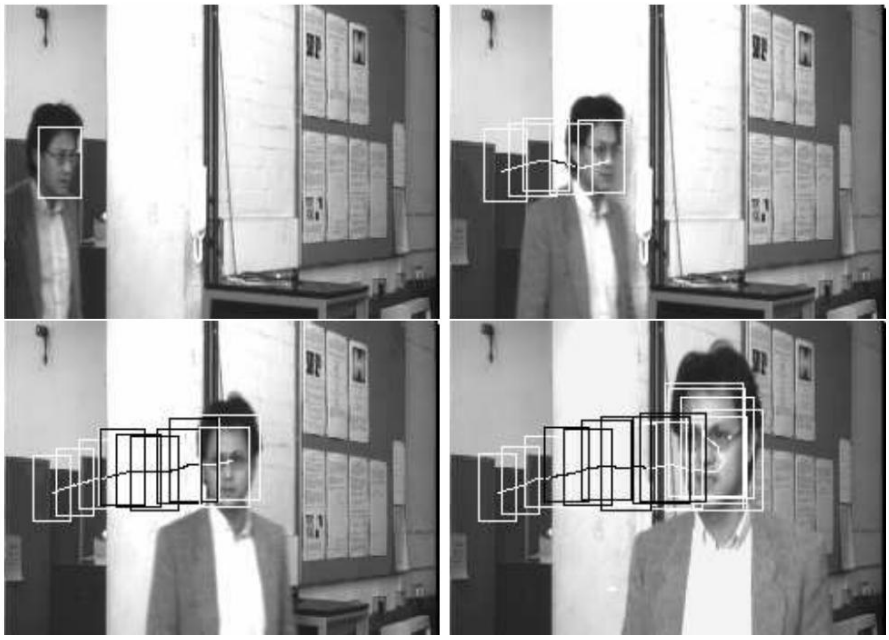


Εικόνα 4: *Εντοπισμός προσώπου μέσω ακμών*[4]

[4] https://www.owlnet.rice.edu/~elec539/Projects97/morphjrks/moredge.html

3.1.1.1.2 Κίνηση

Ένας άλλος τρόπος για τον εντοπισμό του προσώπου είναι η κίνηση η οποία προκύπτει από ακολουθίες εικόνων (βίντεο). Η μέθοδος αυτή στηρίζεται στις μεταβολές που προκύπτουν μεταξύ κοντινών ακολουθιών (frames). Όταν εντοπίζεται έντονη αλλαγή ακμών σε κοντινές περιοχές ενδιαφέροντος είναι ένδειξη κινούμενου αντικειμένου. Οι McKenna, Gong και Liddell [11] ανέπτυξαν ένα σύστημα εντοπισμού προσώπου κάνοντας χρήση χωρο-χρονικών φίλτρων και φιλτραρίσματος Kalman [12].

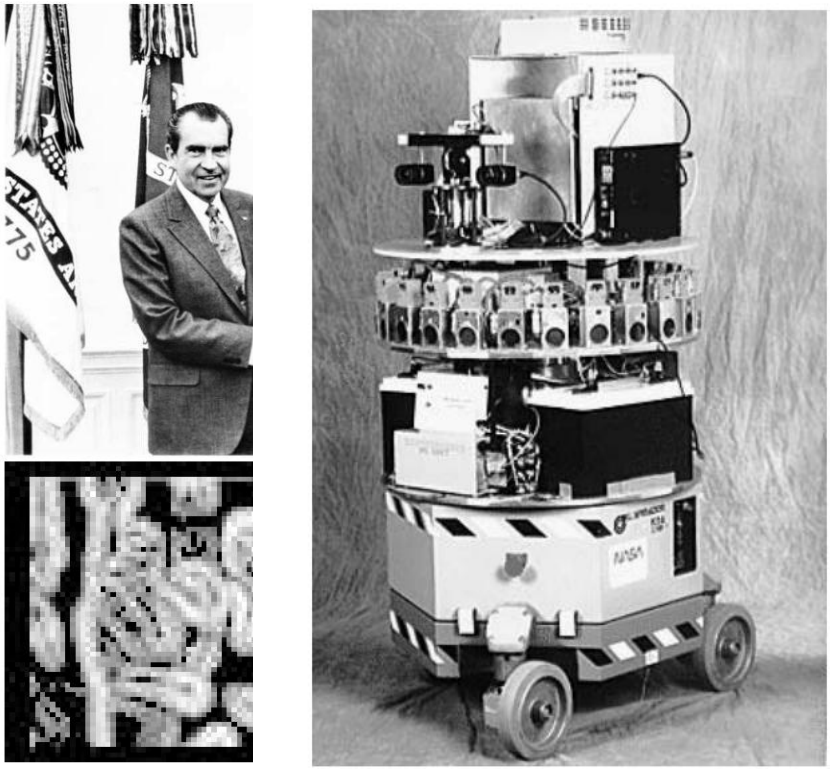


Εικόνα 5: *Το σύστημα εντοπισμού προσώπου του Kalman*[5]

3.1.1.1.3 Γενικευμένες μετρήσεις

Μία άλλη προσέγγιση για τον εντοπισμό προσώπου είναι ο συνδυασμός μετρήσεων κίνησης, ακμών και χρώματος καθώς είναι εκείνα τα χαρακτηριστικά στα οποία βασίζεται η πραγματική ανθρώπινη όραση. Βάση αυτών, έγιναν προσεγγίσεις κάνοντας χρήση επιπλέον της συμμετρίας του προσώπου καθώς αυτή συνέβαλε σημαντικά στον εντοπισμό προσώπου. Βασιζόμενος στο χαρακτηριστικό της συμμετρία αυτής οι Reisfeld, Wolfson και Yeshurun [13] δημιούργησαν ένα σύστημα εντοπισμού πραγματικού χρόνου. [4]

[5] https://pdfs.semanticscholar.org/36d0/9b28ffc70c56dc442e5abacea210916c3579.pdf



Εικόνα 6: *Εντοπισμός προσώπου από την μελέτη των Reisfeld*[6] *et al.*

3.1.1.1.4 Επίπεδα γκρι

Μία επιπλέον μέθοδος είναι ο εντοπισμός προσώπου κάνοντας χρήση της πληροφορίας του γκρι, καθότι παρατηρήθηκε ότι ορισμένες περιοχές του προσώπου όπως είναι το στόμα και τα μάτια παρουσιάζονται πιο σκούρα από ότι τα υπόλοιπες περιοχές του προσώπου. [4] Βάση αυτής της παρατήρησης οι Wong, Kortenkamp και Speich [14] δημιούργησαν ένα ρομπότ που αναγνώριζε ανθρώπους μέσω της αναζήτησης που έκανε για εύρεση ζευγαριού ματιών.

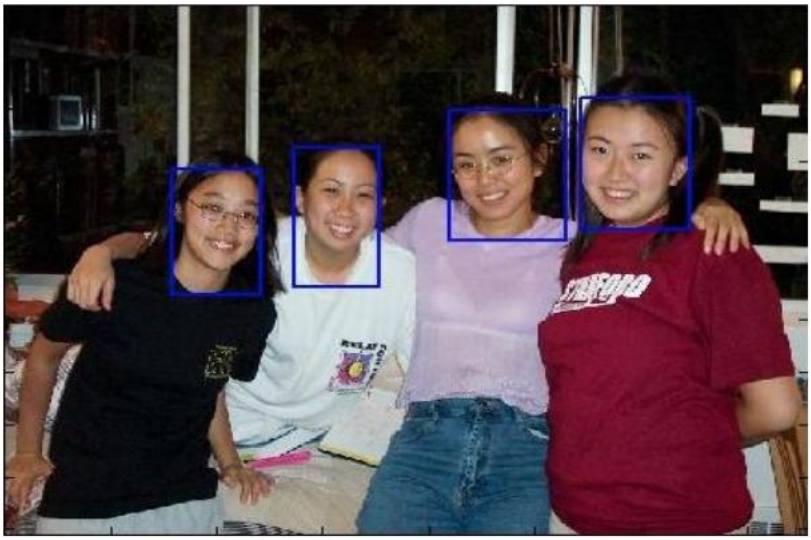
[6] https://pdfs.semanticscholar.org/5597/07ac6ef9cc85db67d4bd5028cdca78c711bd.pdf

***Εικόνα 7**: Το ρομπότ του Wong et al.*[7]

3.1.1.1.5 Χρώμα

Λόγω των δύο παραπάνω διαστάσεων που έχει η πληροφορία του χρώματος σε σχέση με την πληροφορία του γκρι, καθιστά την χρήση του χρώματος μία από τις μεθόδους που βασίζονται αρκετά συστήματα εντοπισμού προσώπου, καθότι προκύπτουν περισσότερα δεδομένα από τα οποία, μπορεί να διακριθεί ένα πρόσωπο από το γενικό περιβάλλον όπου βρίσκεται. Επιπλέον, λόγω συγκεκριμένου εύρους απόχρωσης του προσώπου καθιστά την διαδικασία αυτή ακόμα πιο εύκολη καθότι διαφέρει το χρώμα του προσώπου από το περιβάλλον στο οποίο βρίσκεται. Στην εργασία τους οι Patill [7] ανέπτυξαν ένα σύστημα εντοπισμού προσώπου κάνοντας χρήση του RGB μοντέλου. Βέβαια σε ορισμένες περιπτώσεις όπου υπάρχουν άτομα διαφορετικών φυλών το εκάστοτε σύστημα εντοπισμού ενδεχομένως να αντιμετωπίσει δυσκολίες.

[7] https://pdfs.semanticscholar.org/9be0/36123b41cea87c504799472ca341c5fecf05.pdf



Εικόνα 8: *Εντοπισμός πολλαπλών προσώπων βασισμένο στο RGB μοντέλο*[8]

3.1.1.2 Ανάλυση χαρακτηριστικών

Οι μέθοδοι όπου κάνουν ανάλυση χαμηλού επιπέδου σε ορισμένες περιπτώσεις τα χαρακτηριστικά που εξάγουν είναι ασαφή καθώς οι περιοχές ενδιαφέροντος δηλαδή το πρόσωπο δεν μπορούν να διακριθούν από το περιβάλλον. Χαρακτηριστική περίπτωση είναι οι μέθοδοι που βασίζονται σε μοντέλα χρώματος δέρματος όπου πολλές φορές στο περιβάλλοντα χώρο υπάρχει περιοχή όπου έχει την ίδια απόχρωση με το δέρμα του προσώπου. Για αυτόν τον λόγο, έχουν αναπτυχθεί μέθοδοι που βασίζονται σε υψηλού επιπέδου ανάλυση χαρακτηριστικών. Οι μέθοδοι αυτοί, βασίζονται στην γνώση της γεωμετρίας του προσώπου. Διακρίνονται δύο διαφορετικές κατηγορίες των μεθόδων γεωμετρίας του προσώπου. Η πρώτη αναπτύσσει στρατηγικές διαδοχικής αναζήτησης χαρακτηριστικών όπου όμως σε μεγάλο βαθμό βασίζονται στην θέση των υπολοίπων χαρακτηριστικών του προσώπου. Η δεύτερη κατηγορία βασίζεται στην ομαδοποίηση των χαρακτηριστικών όπου στην συνέχεια δημιουργεί τις λεγόμενες ευέλικτες συστοιχίες. [4]

[8] https://pdfs.semanticscholar.org/87a3/8b79a03457e52a2e7f4fcfbc528632f3cb99.pdf

3.1.1.2.1 Αναζήτηση χαρακτηριστικών

Οι μέθοδοι που βασίζονται στην αναζήτηση χαρακτηριστικών, είναι στις μέρες μας οι πιο διαδεδομένοι, καθώς επιτυγχάνουν υψηλά ποσοστά ακρίβειας στον εντοπισμού του προσώπου. Οι αλγόριθμοι όπου χρησιμοποιούνται κάνουν χρήση των πιο σημαντικών χαρακτηριστικών του προσώπου όπως είναι το ύψος του κεφαλιού, οι αποστάσεις μεταξύ των ματιών , οι αποστάσεις μεταξύ ματιών και μύτης. Το πρώτο σκέλος των αλγορίθμων κάνει αναζήτηση των χαρακτηριστικών αυτών και ύστερα με την χρήση ανθρωπομετρικών μετρήσεων στο πρόσωπο είναι σε θέση να εξάγει και δευτερεύουσας σημασίας χαρακτηριστικά. Οι De Silva, Aizawa και Hatori [15] με την βοήθεια της ακριβής μέτρησης εικονοστοιχείου μπόρεσαν να κάνουν εντοπισμό των ματιών, της μύτης και του στόματος βάση των οποίων ύστερα επετεύχθει ο εντοπισμός του προσώπου. Η βασική λειτουργία του αλγορίθμου είναι η υπόθεση του της κορυφής του κεφαλιού και στην συνέχεια πραγματοποιούνται αναζητήσεις από την κορυφή προς τα κάτω για εύρεση ματιών. Οι εύρεση αυτή, βασίζεται αυξήσεις των εντάσεων των ακμών που προκύπτουν.

Εικόνα 9: *Σάρωση (από πάνω προς τα κάτω) για εντοπισμό προσώπου*[9]

9 https://www.researchgate.net/profile/Liyanage_De_Silva/publication/220241150_Detection_and_Tracking_of_Facial_Features_by_Using_Edge_Pixel_Counting_and_Deformable_Circular_Template_Matching/links/00b4953a30afe7a96b000000/Detection-and-Tracking-of-Facial-Features-by-Using-Edge-Pixel-Counting-and-Deformable-Circular-Template-Matching.pdf?origin=publication_detail

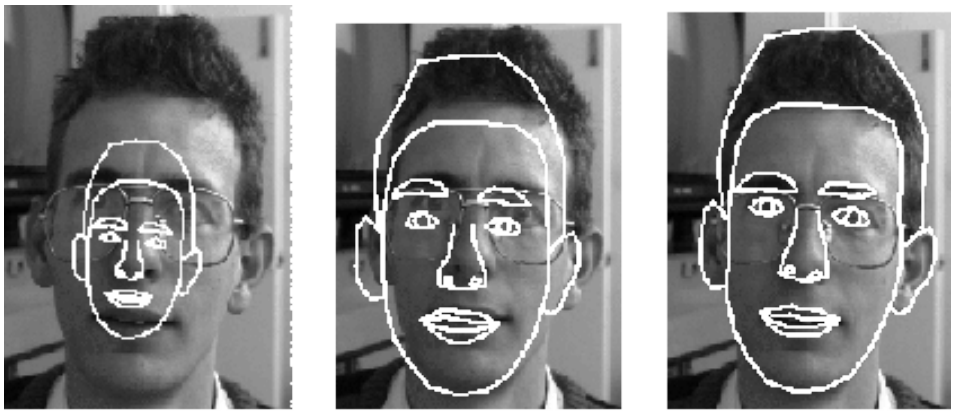
3.1.1.2.2 Ανάλυση συστοιχιών

Οι μέθοδοι όπου κάνουν χρήση της ανάλυσης συστοιχιών, βασίζονται στην κατάλληλη οργάνωση των χαρακτηριστικών που προκύπτουν κατά κύριο λόγο από στατιστικές αναλύσεις, όπως είναι τα μοντέλα πιθανοτήτων και η στατιστική παραμετρική χαρτογράφηση [16]

3.1.1.3 Μοντέλα ενεργού σχήματος

Οι μέθοδοι ενεργού περιγράμματος αναζητούν χαρακτηριστικά υψηλού επιπέδου σε σχέση με τις μεθόδους των μοντέλων που περιεγράφηκαν παραπάνω και βασίζονται σε μοντέλα στατιστικής . Η βασική λειτουργία των μεθόδων αυτών είναι εφαρμογή τους σε ένα χαρακτηριστικό όπου στην συνέχεια διαδοχικά παραμορφώνονται έως ότου πάρουν το σχήμα του ίδιου του χαρακτηριστικού. Οι προσέγγιση αυτή του ενεργού περιγράμματος αναπτύχθηκε την δεκαετία του 90’ από τον Taylor [17] και έχει τα εξής βήματα τα οποία εναλλάσσονται έως να επέλθει το καλύτερο δυνατό αποτέλεσμα:

- Δημιουργεί ένα προτεινόμενο σχήμα γύρω από το σημείο ενδιαφέροντος το οποίο προκύπτει από την παρουσία έντονων ακμών
- Με βάση το αρχικό προτεινόμενο σχήμα, συμπληρώνει το διανομής των σημείων μέχρι να πάρει το σχήμα του χαρακτηριστικού



(A) (B) (Γ)

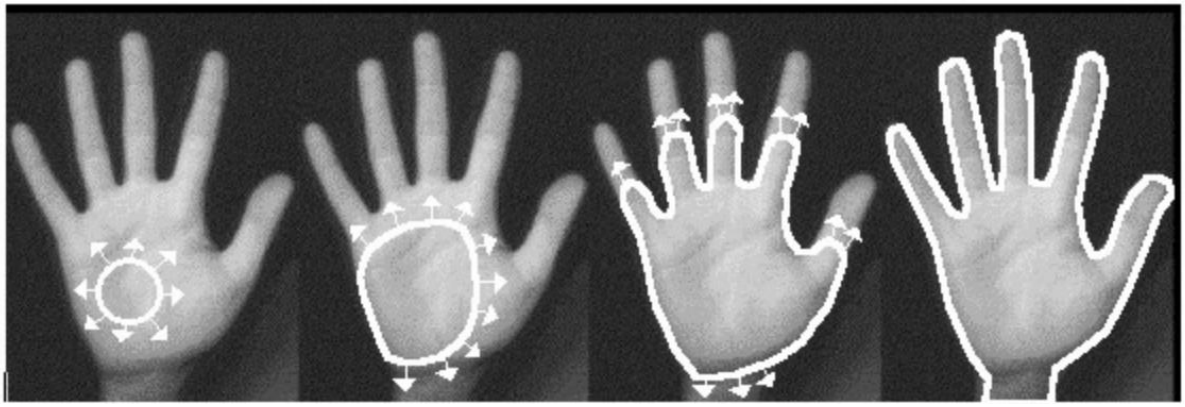
Εικόνα 10: *Εφαρμογή ενεργού περιγράμματος Α) αρχική θέση Β) ύστερα από 10 επαναλήψεις Γ) σύγκλιση της αναζήτησης* [10]

[10] http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/BMVA96Tut/node37.html

3.1.1.3.1 Snakes

Η μέθοδος των snakes (ενεργά περιγράμματα) υλοποιείται κατά κύριο λόγο για να εντοπιστεί το περίγραμμα του κεφαλιού. Η βασική μεθοδολογία των snakes είναι, ο αρχικός ορισμός ενός συνόλου σημείων όπου στην συνέχεια βάση αυτών αναπτύσσονται περιοχές ενώνοντας τα με εικονοστοιχεία της γειτονιάς τους όπου έχουν ορισμένες προκαθορισμένες τιμές, ίδιες ή παρόμοιες με αυτές των σημείων. Ο αρχικός ορισμός των σημείων, συχνά βασίζεται στην φύση του εκάστοτε προβλήματος, αλλά στην περίπτωση όπου η αρχική επιλογή του σημείου δεν είναι δυνατή να γνωρίζεται εκ των προτέρων, τότε η διαδικασία διαφοροποιείται και αντί να υπάρχει αρχικό σημείο στοιχείων, εφαρμόζεται η αναζήτηση ίδιων συνόλων ιδιοτήτων για κάθε εικονοστοιχείο. Από αυτήν την διαδικασία ενδεχομένως να προκύψουν συστάδες εικονοστοιχείων που θα έχουν κοινές ιδιότητες. Ένας παράγοντας που επηρεάζει τα αποτελέσματα των ενεργών περιγραμμάτων είναι ο τύπος των εικόνων. Οι RGB εικόνες παραδείγματος χάρη λόγω περισσοτέρων διαστάσεων και κατ' επέκταση περισσοτέρων πληροφοριών βοηθούν την διαδικασία εύρεσης συστάδων εικονοστοιχείων με παρόμοιες ιδιότητες, ενώ στις μονόχρωμες εικόνες η αναζήτηση συστάδων θα πρέπει να βασιστεί σε ένα περιορισμένο σύνολο περιγραφών για τις κοινές ιδιότητες όπως είναι τα επίπεδα έντασης. [8]

Οι περιγραφές αυτές, σε ορισμένες περιπτώσεις ενδεχομένως να οδηγήσουν σε λαθεμένα αποτελέσματα εάν δεν ληφθούν υπόψιν οι ιδιότητες της συνδεσιμότητας των εικονοστοιχείων. Επίσης, ένα ακόμα πρόβλημα που ίσως παρουσιαστεί στην μέθοδο των ενεργών περιγραμμάτων είναι οι συνθήκες τερματισμού ανάπτυξης των περιγραμμάτων αυτών. Εάν δεν έχουν ορισθεί οι κατάλληλες συνθήκες και τα κατώφλια είναι πιθανό η ανάπτυξη των περιοχών-περιγραμμάτων να συμπεριλάβει και εικονοστοιχεία ή ακόμα και μεγάλες περιοχές εικονοστοιχείων που να μην ανήκουν στην πραγματικότητα στην συστάδα αυτή. [8]



Εικόνα 11: *Εφαρμογή των snakes για εύρεση περιγράμματος παλάμης*[11]

3.1.1.3.2 PDM

Η μέθοδος PDM (point distribution model) ή μοντέλο κατανεμημένων σημείων είναι ένα μοντέλο περιγραφή σχήματος που βασίζεται στην στατιστική. Πιο συγκεκριμένα παρουσιάζει την μέση γεωμετρία του σχήματος και ορισμένους τρόπους γεωμετρικής μεταβολής. Η εφαρμογή του PDM, διαφοροποιείται από της εφαρμογές των υπολοίπων μοντέλων ενεργού σχήματος που παρουσιάστηκαν παραπάνω. Το περίγραμμα που δημιουργείται και στην συνέχεια αναπτύσσεται από την συγκεκριμένη μέθοδο, βασίζεται στην ανάλυση βασικών συνιστωσών των διακυμάνσεων των χαρακτηριστικών έτσι ώστε να χρησιμοποιεί τα χαρακτηριστικά εκείνα τα οποία έχουν την μεγαλύτερη διακύμανση και κατ' επέκταση την περισσότερη πληροφορία.

3.1.2 Μέθοδοι με βάση την εικόνα

Οι μέθοδοι που βασίζονται στην εικόνα, στην ουσία αναπτύσσουν αλγορίθμους που υπάγονται στον κλάδο της αναγνώρισης προτύπων. Δεν επιδιώκουν την εξαγωγή χαρακτηριστικών από την εικόνα, αλλά αντιμετωπίζουν την εικόνα ως σύνολο και πιο συγκεκριμένα την παρουσιάζουν ως δυσδιάστατο πίνακα που περιέχει μέσα τιμές έντασης των εικονοστοιχείων. Με βάση αυτούς του πίνακες πραγματοποιείται εκπαίδευση του συστήματος όπου στην συνέχεια τα πρότυπα (εικόνες) ταξινομούνται σε κλάσεις. Οι κλάσεις και αυτές με την σειρά τους, υπάγονται σε συγκρίσεις μεταξύ τους όπου μέσω των οποίων παίρνεται η απόφαση για την ύπαρξη προσώπου στην εικόνα.[4]

---

[11] http://www.bcamath.org/documentos_public/courses/course_day3.pdf

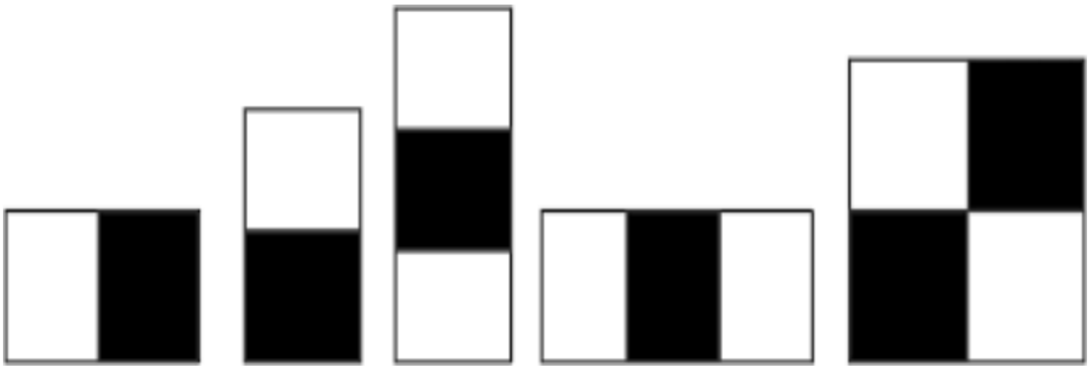
3.1.2.1 Viola-Jones

Ο Viola-Jones, είναι μία μέθοδος που αναπτύχθηκε από τους Viola και Jones [18] ο οποίος επιτυγχάνει εντοπισμό πλήθους αντικειμένων σε πραγματικό χρόνο, αν και χρησιμοποιείται επί το πλείστων για εντοπισμό προσώπων. Ο αλγόριθμος αποτελείται από τέσσερα βασικά στάδια:

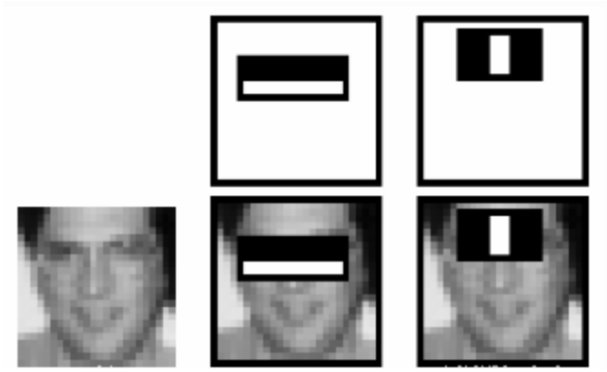
- Επιλογή χαρακτηριστικών Harr
- Δημιουργία ολοκληρωτικής εικόνας
- Εκπαίδευση ταξινομητών με την μέθοδο Adaboost
- Χρήση ταξινομητή Cascade

Η μέθοδος αυτή παρέχει υψηλά ποσοστά εντοπισμού και για τον λόγο αυτό από την αρχή του 2000 έως σήμερα χρησιμοποιείται σε πολλές και εφαρμογές ανίχνευσης προσώπου και όχι μόνο. Ο αλγόριθμος κάνει χρήση ενός ταξινομητή πολλών επιπέδων που έχει την μορφή καταρράκτη (cascade). Επίσης, χρησιμοποιείται ένας επιπλέον αλγόριθμος, ο Adaboost, ο οποίος μαζί με τον Cascade ανιχνευτή επιτυγχάνουν γρήγορη ταχύτητα επεξεργασίας και ανίχνευσης. Η συγκεκριμένη μέθοδος κάνει επιπλέον χρήση μία εικόνας ολοκλήρωσης η οποία συμβάλει στην ταχύτητα υπολογισμού χαρακτηριστικών Harr τα οποία προκύπτουν από τις δοθέντες εικόνες. Τα χαρακτηριστικά τύπου Harr, αποτελούνται από τρία είδη. Το πρώτο είδος είναι τα χαρακτηριστικά εκείνα που προκύπτουν από δύο ορθογώνια (είτε κάθετα είτε οριζόντια) και οι τιμή τους προκύπτει από την διαφορά των αθροισμάτων των εικονοστοιχείων.

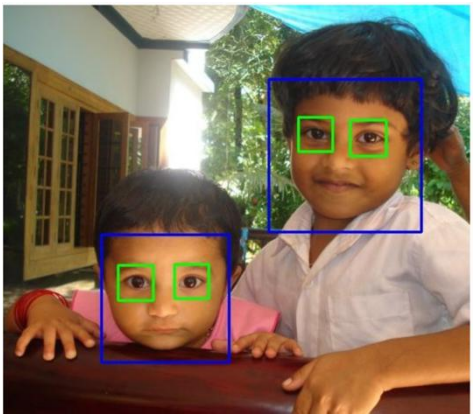
Το δεύτερο είδος, είναι τα χαρακτηριστικά που προκύπτουν από 3 ορθογώνια όπου η τιμή τους προκύπτει από το άθροισμα των δύο ορθογωνίων που βρίσκονται εξωτερικά και στην συνέχεια αφαιρείται από το άθροισμα του ορθογωνίου που βρίσκεται στην μέση. Τέλος, το τρίτο είδος προκύπτει από την χρήση τεσσάρων ορθογωνίων και η τιμή τους ισούται με την διαφορά του αθροίσματος των διαγώνιων ορθογωνίων. [18]



Εικόνα 12: *Ορθογώνια χαρακτηριστικά τύπου Harr*[12]



Εικόνα 13: *Χρήση ορθογωνίων χαρακτηριστικών σε εικόνα* [13]



Εικόνα 14: *Εφαρμογή Viola-Jones για εντοπισμό προσώπου και ματιών*[14]

[12] http://nemertes.lis.upatras.gr/jspui/bitstream/10889/10974/1/ΚΟΤΣΙΑΣ ΔΗΜ. ΣΕΣΕ 206.pdf

[13] https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html

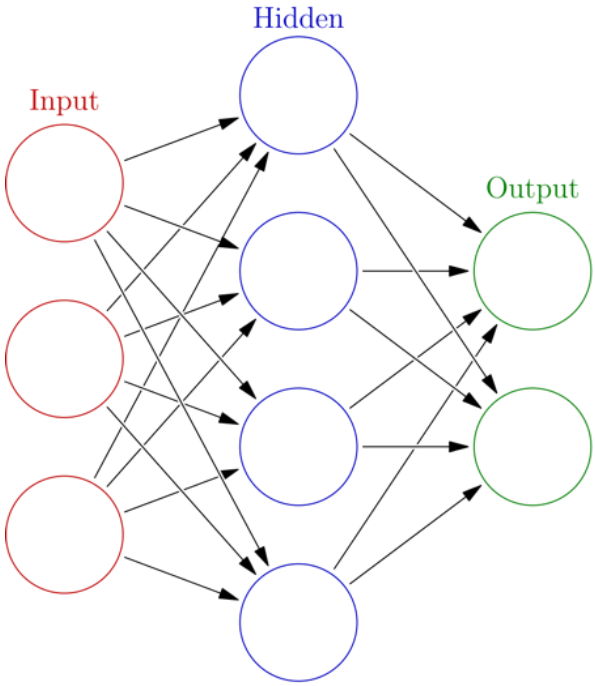
[14] https://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html

3.1.2.2 Μέθοδοι γραμμικών υποχωρών

Ένας επιπλέον τρόπος εντοπισμού προσώπου είναι η εφαρμογή μεθόδων που κάνουν χρήση στατιστικών αναλύσεων και αναπαριστούν τις εικόνες προσώπων σε υποχώρους του συνολικού χώρου εικόνων. Για τον λόγο ότι κατά πλειοψηφία οι μέθοδοι αυτοί όπως ο PCA και LDA χρησιμοποιούνται στον τομέα της αναγνώρισης προσώπου και όχι του εντοπισμού, θα αναλυθούν σε παρακάτω υποκεφάλαιο.

3.1.2.3 Νευρωνικά Δίκτυα

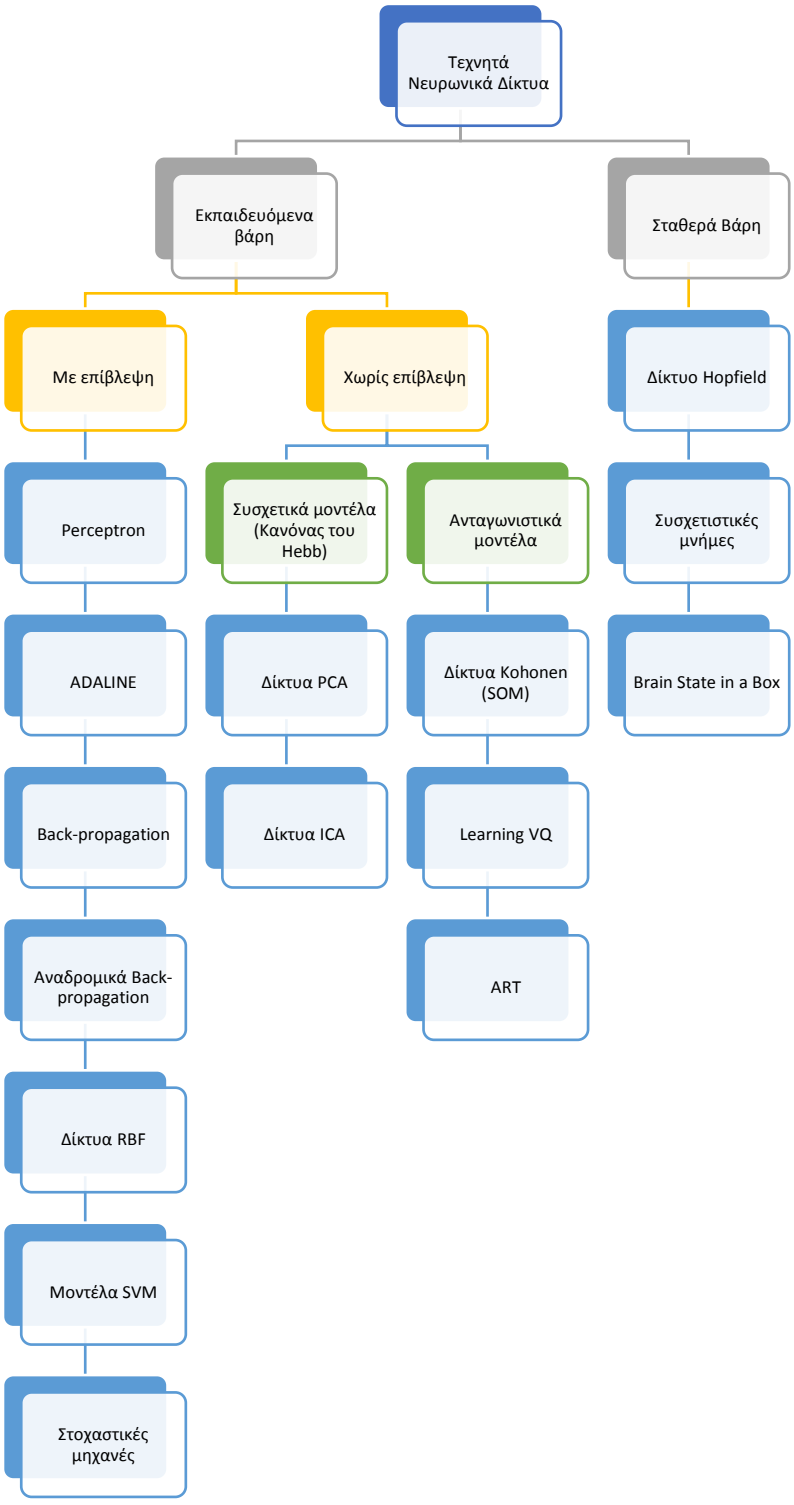
Τα τεχνητά νευρωνικά δίκτυα, αποτελούν μία από τις επικρατέστερες μεθόδους εντοπισμού προσώπου και όχι μόνο. Η εξέλιξή τους τα τελευταία χρόνια είναι ραγδαία λόγω της τεχνολογικής ανάπτυξης καθότι η γενικότερη λειτουργία των νευρωνικών δικτύων χρίζει μεγάλη υπολογιστική δύναμη για να μπορέσουν να εφαρμοσθούν. Βέβαια, και τις προηγούμενες δεκαετίες υπήρχε εκτενής μελέτη και εφαρμογή αυτών, σε απλούστερα θέματα που μπορούσαν να αντιμετωπιστούν από την εκάστοτε υλική υποδομή. Η βασική δομή των νευρωνικών δικτύων παρομοιάζει την δομή των νευρώνων του εγκεφάλου που διαθέτει συνάψεις και με τις οποίες συνδέεται με άλλους νευρώνες. Οι νευρώνες με την σειρά τους δημιουργούν επίπεδα με βάση απλών έως πολύ πολύπλοκων αρχιτεκτονικών.



Εικόνα 15: *Η βασική δομή ενός απλού τεχνητού νευρωνικού δικτύου*[15]

Οι αλγόριθμοι που χρησιμοποιούνται στις μεθόδους νευρωνικών δικτύων είναι ποικίλοι καθότι και η χρήση τους τους έχει μεγάλο εύρος εφαρμογών. Οι σημαντικότεροι αλγόριθμοι είναι:

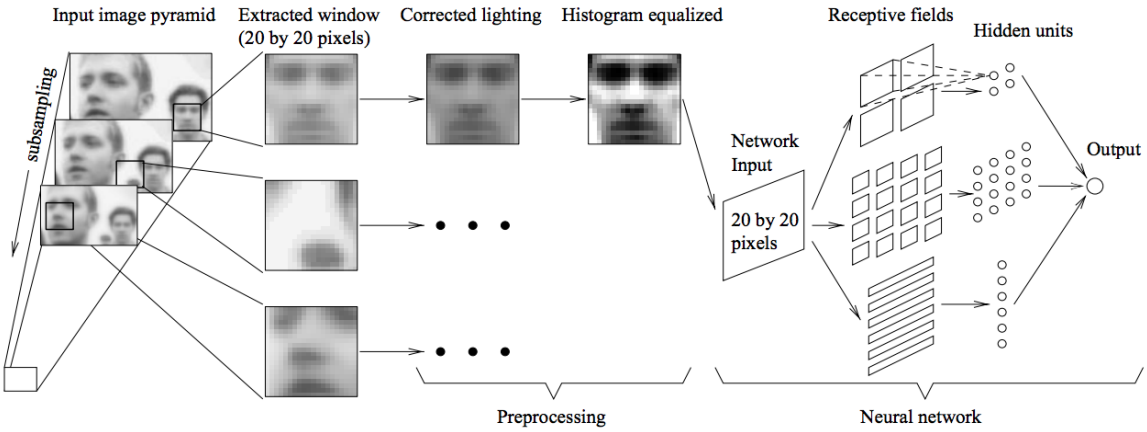
[15] https://en.wikipedia.org/wiki/File:Colored_neural_network.svg



Γράφημα 3: *Ταξινόμηση νευρωνικών αλγορίθμων* [19]

Στον εντοπισμό προσώπων, τα τεχνητά νευρωνικά δίκτυα χρησιμοποιούνται κατά ένα μεγάλο ποσοστό λόγω των επιτυχών αποτελεσμάτων που αποφέρουν. Η μέθοδος με νευρωνικά δίκτυα, είναι αρκετά νεότερη σε σχέση με άλλες μεθόδους. Την δεκαετία του 90’ οι Rowley, Baluja και Kanade [20] έκαναν την πρώτη επιτυχή προσέγγιση εντοπισμού

προσώπου. Το νευρωνικό δίκτυο που δημιούργησαν δέχεται ως είσοδο εικόνες διαστάσεων 20x20 pixels όπου η δομή του δικτύου είναι, ένα hidden layer (κρυφό επίπεδο) όπου αποτελείται από 26 νευρώνες, από τους οποίους, των 4 η περιοχή ενδιαφέροντος τους είναι 10x10 pixel της αρχικής περιοχής, των 16 είναι 5x5 pixels και των υπολοίπων 6 είναι 20x5 pixels. Οι νευρώνες αυτοί, υλοποιούν τεχνικές οριζόντιου σκαναρίσματος, από όπου όμως προκύπτουν προβλήματα επικαλυπτόμενων εντοπισμών , τα οποία αντιμετωπίστηκαν με χρήση δύο επιπλέον τεχνικών οι οποίες είναι το κατώφλι και η απαλοιφή επικαλύψεων. Στην τεχνική του κατωφλίου, εάν ο αριθμός των επιτυχών εντοπισμών ισούται πάνω από το ορίζων κατώφλι τότε σημαίνει ύπαρξη προσώπου στην εικόνα. Στην τεχνική της απαλοιφής επικαλύψεων, εάν μια εικόνα ή περιοχή της εικόνας έχει σημανθεί ως πρόσωπο, τότε οι επικαλυπτόμενοι εντοπισμοί θεωρούνται λανθασμένοι.



Εικόνα 16: *Ο βασικός αλγόριθμος του Rowley για εντοπισμό προσώπων*[16]

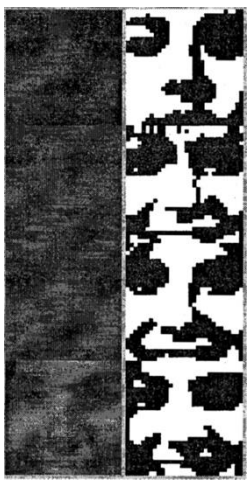
3.1.2.4 Support Vector Machines

Η μέθοδος των Support Vector Machines (μηχανές διανυσμάτων υποστήριξης), αποτελούν είδος αλγορίθμων μηχανικής μάθησης και πιο συγκεκριμένα ανήκουν στην κατηγορία των δικτύων πρόσθιας τροφοδότησης. [21] Οι μηχανές διανυσμάτων υποστήριξης

[16] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.2367&rep=rep1&type=pdf

βασίζονται στην θεωρία στατιστικής εκπαίδευσης και έχουν παρουσιάσει πολύ καλά αποτελέσματα σε εφαρμογές αναγνώρισης και κατηγοριοποίησης αντικειμένων. Οι μηχανές αυτές, λειτουργούν εξαιρετικά καλά με δεδομένα που αποτελούνται πάνω από δύο διαστάσεις όπως είναι οι εικόνες και είναι ο λόγος της υπεροχής τους σε σχέση με άλλες μεθόδους μηχανικής μάθησης. Η βασική ιδέα των μηχανών διανυσμάτων υποστήριξης είναι η χρήση του υπερεπιπέδου μέγιστου περιθωρίου (maximal margin hyperplane) το οποίο είναι υπεύθυνο για τον διαχωρισμό των δεδομένων σε κατηγορίες. [22]

Στην εργασία τους οι Colmenarez και Huang [23], βασίστηκαν στο θεώρημα της απόκλισης Kullback [24] που ένας από τους τρόπους για να επιτευχθεί σύγκριση μεταξύ δύο κατανομών. Επίσης έγινε χρήση ενός ιστογράμματος κατά την διάρκεια της εκπαίδευσης, έτσι ώστε κάθε ζεύγος εικονοστοιχείων να συνδέεται με το ιστόγραμμα αυτό και να δημιουργούνται συναρτήσεις απόφασης ως προς την εύρεση ή μη εύρεση προσώπων. Χρησιμοποιήθηκαν εικόνες προσώπων και μη προσώπων διαστάσεων 30x30 pixels για την διαδικασία της εκπαίδευσης του συστήματος, οι οποίες οδηγούν σε πίνακες αντιστοίχισης των ίδιων των εικόνων με λόγους πιθανότητας. Τέλος, όσα εικονοστοιχεία δεν συνεισφέρουν στην απόκλιση διαγράφονται από τους πίνακες αντιστοίχισης για μείωση υπολογιστικού κόστους και ταχύτητα εντοπισμού προσώπου.

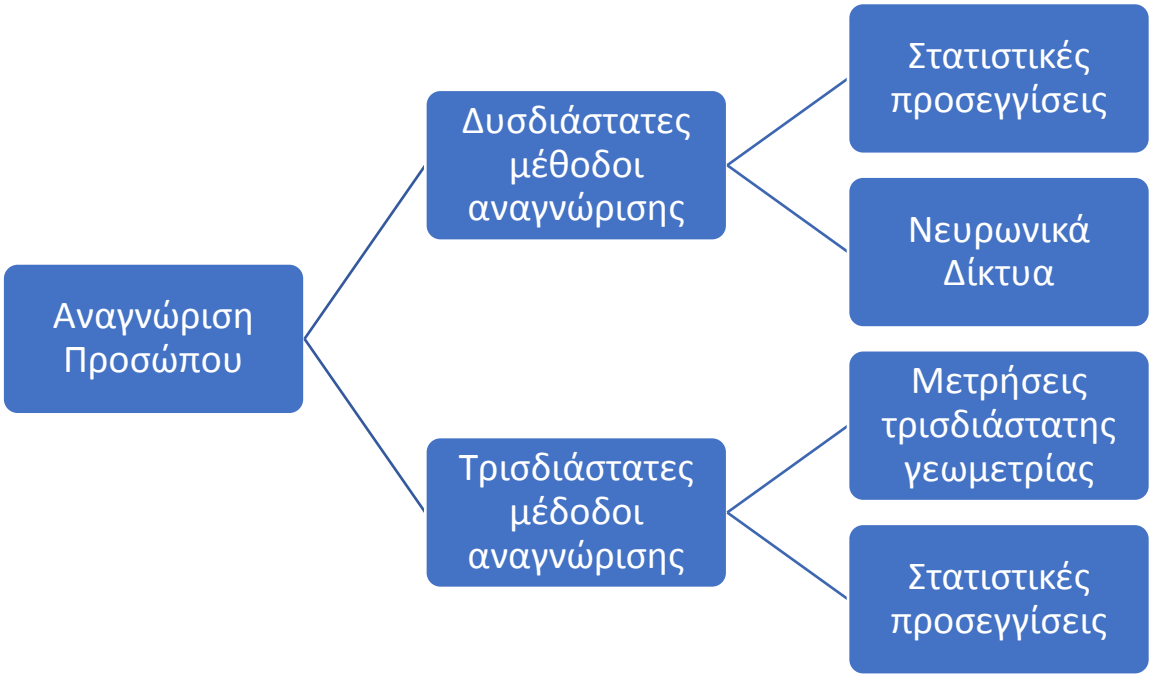


Εικόνα 17: *Από την εργασία του Kullback. Παράθυρο 30x30 γύρω από το πρόσωπο και οι αντίστοιχες δυαδικές εικόνες*

3.2 Αναγνώριση Προσώπου

Το δεύτερο βασικό στάδιο για την ανάπτυξη ενός συστήματος ταυτοποίησης μέσω εικόνας είναι η διαδικασία της αναγνώρισης προσώπου. Η αναγνώριση μπορεί να επιτευχθεί με χρήση εικόνων δύο ή τριών διαστάσεων. Με βάση αυτό το χαρακτηριστικό των

διαστάσεων των εικόνων, αναπτύσσονται αντίστοιχα συστήματα αναγνώρισης τα οποία χρησιμοποιούνται αναλόγως της εφαρμογής και του σκοπού. Οι εικόνες δύο διαστάσεων λόγω λιγότερης πληροφορίας που καταγράφουν, συμβάλουν στην μείωση υπολογιστικού κόστους για το εκάστοτε σύστημα όμως παρουσιάζουν ορισμένα προβλήματα ως προς την ευαισθησία τους σε αλλαγές του φωτός. Οι εικόνες τριών διαστάσεων, έχουν το βασικό πλεονέκτημα ότι δύναται το πρόσωπο να μοντελοποιηθεί λόγω του ότι περιέχουν πληροφορία για το βάθος. [4]



Γράφημα 4: *Μέθοδοι αναγνώρισης προσώπου*

3.2.1 Δυσδιάστατες μέθοδοι αναγνώρισης

Ως προς την δυσδιάστατη αναγνώριση χρησιμοποιούνται οι μέθοδοι των νευρωνικών δικτύων όπως επίσης και στατιστικά μοντέλα. Τα στατιστικά αυτά μοντέλα, αποτελούνται από την μέθοδο της ανάλυσης κυρίων συνιστωσών (PCA), την ανάλυση γραμμικού διαχωρισμού (LDA) και το κυμματίδιο Gabor (Gabor Wavelet) ενώ ως προς τα νευρωνικά δίκτυα γίνεται χρήση αυτών με φίλτρα Gabor και Hidden Markov Models. [4]

3.2.1.1 Στατιστικές Προσεγγίσεις

Οι εικόνες ως προς το περιεχόμενο τους παρουσιάζουν το γεγονός του ότι ενώ παρουσιάζονται ως πολυδιάστατοι πίνακες εικονοστοιχείων, τις περισσότερες φορές το περιεχόμενο αυτό δύναται να παρουσιαστεί σε πίνακες μικρότερων οντότητα. Για να

πραγματοποιηθεί αυτή η παρουσίαση του περιεχομένου σε μικρότερες διαστάσεις οντότητας, γίνεται χρήση στατιστικών μεθόδων και εργαλείων μέσω των οποίων είναι δυνατόν να βρεθούν τα χαρακτηριστικά εκείνα του ευρύτερου χώρου της εικόνας τα οποία μπορούν να χαρακτηρίσουν την συνολική εικόνα και να την κατηγοριοποιήσουν ως προς άλλες.

3.2.1.1.1 Ανάλυση κυρίων συνιστωσών (PCA)

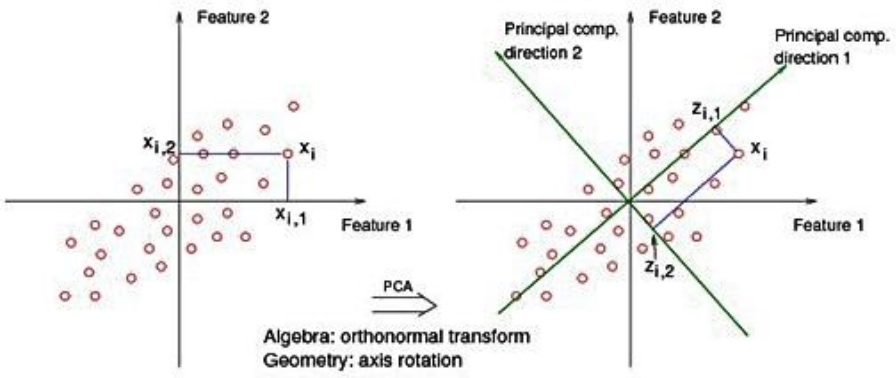
Η μέθοδος της ανάλυσης κυρίων συνιστωσών (PCA), αποτελεί μια γραμμική μέθοδο συμπίεσης των δεδομένων κάνοντας χρήση ενός ορθογώνιου μετασχηματισμού έτσι ώστε να μετατρέψει και να επαναπροσδιορίσει τις συντεταγμένες ενός συνόλου δεδομένων. Οι νέες αυτές συντεταγμένες περιέχουν σημεία τα οποία διατηρούν σε φθίνουσα σειρά την διακύμανση των δεδομένων, δηλαδή την μεταβλητότητα τους και είναι το αποτέλεσμα ενός γραμμικού συνδυασμού που προέρχεται από τις αρχικές μεταβλητές. Λόγω αυτής της φθίνουσας σειράς, το πρώτο (κύριο) συστατικό περιέχει τις περισσότερες πληροφορίες σε σχέση με τα υπόλοιπα. Τα συστατικά που προαναφέρθηκαν ονομάζονται συνιστώσες, οι οποίες συνιστώσες είναι μικρότερες ή ίσες από τον αρχικό αριθμό των μεταβλητών. [25]

Έστω ο X_{nxm} ο πίνακας δεδομένων όπου $x_1,\ldots,x_m$ τα διανύσματα εικόνας και n ο αριθμός των pixel ανά εικόνα. Λύνεται στην συνέχεια το πρόβλημα των ιδιοτιμών

$$C_X = \Phi \Lambda \Phi^T$$

C είναι ο πίνακας συνδιασποράς των δεδομένων

$$C_x = \frac{1}{m}\sum_{i=1}^{m} x_i * x_i^T$$



Εικόνα 18: *Εφαρμογή PCA*[17]

Εικόνα 19: *Ανακατασκευή εικόνας προσώπου στα αρχικά δεδομένα με χρήση των 10,20 και 30 πιο σημαντικών συστατικών*

[17] https://onlinecourses.science.psu.edu/stat857/node/35/

3.2.1.1.2 Ανάλυση γραμμικού διαχωρισμού (LDA)

Η μέθοδος της ανάλυσης γραμμικού διαχωρισμού (LDA), αποτελεί ένα είδος γραμμικού ταξινομητή όπου χρησιμοποιεί υπερεπίπεδα βάση των οποίων διαχωρίζει τα δεδομένα. Κατά κύριο λόγο η μέθοδος αυτή δύναται να μπορεί να ξεχωρίσει δεδομένα που να ανήκουν σε δύο μόνο κλάσεις και χρειάζεται τα δεδομένα των διαφορετικών κλάσεων να είναι μακριά μεταξύ τους ενώ τα δεδομένα των ίδιων κλάσεων να βρίσκονται κοντά. Η ανάλυση γραμμικού διαχωρισμού δεν έχει υψηλές υπολογιστικές απαιτήσεις και για αυτόν τον λόγο προτιμάται να εφαρμόζεται σε συστήματα BCI (σύστημα διάδρασης εγκεφάλου-υπολογιστή), ενώ σε εφαρμογές που υπάρχουν δεδομένα μη γραμμικά έχει σχετικά χαμηλή απόδοση. [26]

Έστω ότι m δείγματα $x_1,...x_m$ όπου ανήκουν σε c κλάσεις και κάθε κλάση έχει m_k στοιχεία. Η συνάρτηση LDA είναι:

$$a_{opt} = argmax \frac{a^T S_b a}{a^T S_t a}$$

$$S_b = \sum_{k=1}^{c} m_k \mu^{(k)} (\mu^{(k)})^T = \sum_{k=1}^{c} (\frac{1}{m_k} (\sum_{i=1}^{m_k} x_i{}^{(k)}))(\frac{1}{l_k} (\sum_{i=1}^{m_k} x_i{}^{(k)}))^T = XW_{mxm} X^T$$

$$S_t = \sum_{i=1}^{m} x_i (x_i)^T = XX^T$$

Όπου W_{mxm} είναι ο διαγώνιος πίνακας

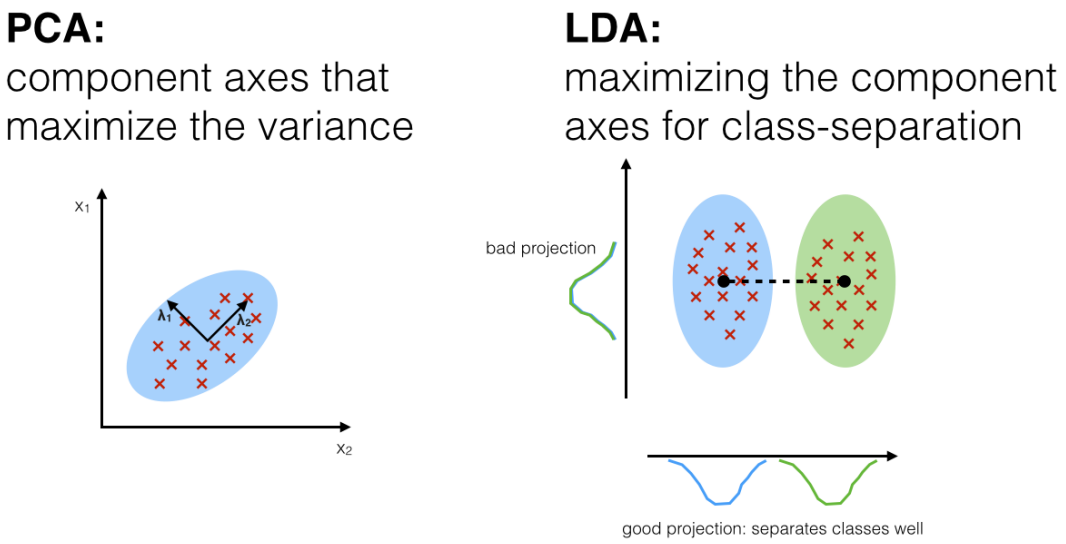
$$W_{mxm} = \begin{bmatrix} W^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W^C \end{bmatrix}$$

Και W^k είναι πίνακας $m_k * m_k$

$$W^k = \begin{bmatrix} \frac{1}{m_k} & \cdots & \frac{1}{m_k} \\ \vdots & \ddots & \vdots \\ \frac{1}{m_k} & \cdots & \frac{1}{m_k} \end{bmatrix}$$

Τέλος το ιδιοπρόβλημα δύναται να γραφεί:

$$S_b a = \lambda S_t a \longrightarrow S_t^{-1} S_b a = \lambda a \longrightarrow X W_{lxl} X^T (X X^T)^{-1} a = \lambda a$$



Εικόνα 20: *Διαφορά μεταξύ PCA και LDA*[18]

3.2.1.1.3 Κυμματίδιο Gabor

Τα κυμματίδια Gabor, είναι εφεύρεση από τον Dennis Gabor [27] όπου χρησιμοποίησε συναρτήσεις που χρησιμεύουν στους μετασχηματισμούς Fourier με σκοπό να εφαρμοσθούν σε θεωρίες πληροφοριών. Μελέτες έδειξαν, ότι δεδομένα που προέρχονται από τον οπτικό φλοιό θηλαστικών υποδεικνύουν πως κύτταρα του οπτικού φλοιού μπορούν ενδεχομένως να θεωρηθούν ως ομάδες σωματιδίων Gabor. Στην μελέτη του ο Lee [28], υλοποιεί συναρτήσεις Gabor και τις χρησιμοποιεί ως χωρικά ζωνοπερατά φίλτρα καταφέρνοντας έτσι να συνδυάζει την ανάλυση πληροφορίας μεταξύ των χωρικών πεδίων δύο διαστάσεων και των πεδίων Fourier. [4] Η συνάρτηση Gabor του Lee:

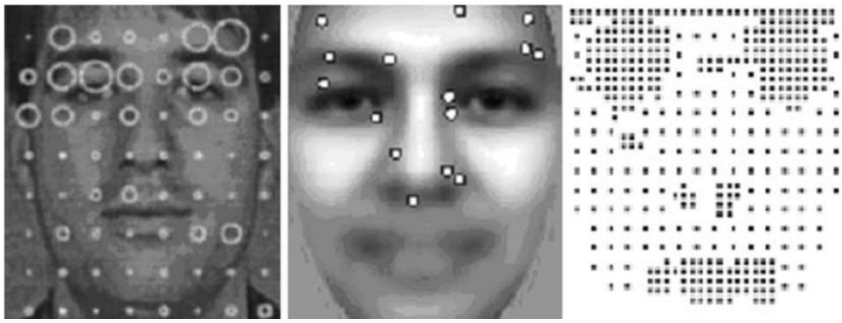
$$\psi_i(\vec{x}) = \frac{\|\vec{k}_l\|}{\sigma^2} e^{\frac{\|\vec{k}_l\|^2 \|\vec{x}\|^2}{2\sigma^2}} \left[e^{j \overrightarrow{k_l x}} - e^{\frac{-\sigma^2}{2}} \right]$$

[18] https://sebastianraschka.com/Articles/2014_python_lda.html

Το ψ_i αποτελεί κύμα που χαρακτηρίζεται από το k_i και μέσω μίας συνάρτησης Gaussian που το περιβάλλει μας δίνει την κεντρική συχνότητα του φίλτρου:

$$\overrightarrow{k_l} = \begin{pmatrix} k_{ix} \\ k_{iy} \end{pmatrix} = \begin{pmatrix} k_\nu cos\theta_a \\ k_\nu sin\theta_\alpha \end{pmatrix};\ k_\nu = 2\pi^{\frac{-\nu+2}{2}};\ \theta_\alpha = \alpha\frac{\pi}{8}$$

Τα φίλτρα Gabor μπορούν να καταγράψουν το φάσμα συχνοτήτων μιας εικόνας, το πλάτος των οποίων χρησιμοποιείται για την αναγνώριση προσώπου. Στην εργασία του οι Shen και Bai πέτυχαν με χρήση σωματιδίων Gabor ποσοστό αναγνώρισης προσώπου της τάξης του 100% χρησιμοποιώντας την βάση δεδομένων ORL [29]



Εικόνα 21: *Αναγνώριση μέσω Gabor σωματιδίων*[29]

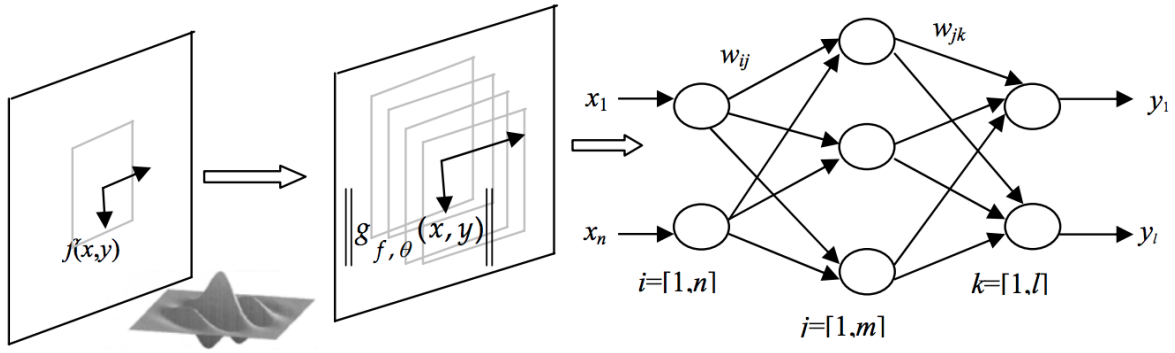
3.2.1.2 Νευρωνικά δίκτυα

Σχετικά με την δομή των τεχνητών νευρωνικών δικτύων έχουμε μιλήσει στην παράγραφο 3.1 περί των μεθόδων εντοπισμού προσώπου. Τα νευρωνικά δίκτυα χρησιμοποιούνται επίσης σε μεγάλο βαθμό και για την αναγνώριση προσώπων. Τα τελευταία χρόνια δε, λόγω της δημιουργίας Deep Learning Networks (βαθιά δίκτυα εκμάθησης) υπάρχει η δυνατότητα για αναγνώριση προσώπων δίχως προεπεξεργασίας των ίδιων των εικόνων. Το δίκτυο από μόνο του δύναται να πραγματοποιήσει feature extraction (εξαγωγή χαρακτηριστικών).

3.2.1.2.1 Νευρωνικά δίκτυα με χρήση φίλτρων Gabor

Μία προσέγγιση τεχνητών νευρωνικών δικτύων είναι αυτή που χρησιμοποιεί επιπλέον τα φίλτρα Gabor που αναφέρθηκαν στο υποκεφάλαιο 3.2.1.1.3. Οι Bhuiyan και Liu [30], δημιούργησαν ένα τέτοιου τύπου νευρωνικό δίκτυο αναγνώρισης προσώπου. Τα βασικά

στάδια του αλγορίθμου τους είναι η κανονικοποίηση των εικόνων ως προς τον θόρυβο και τον φωτισμό, οι εικόνες περνάνε από ένα φίλτρο τύπου Gabor, το οποίο καθώς αυτό αναπαριστά ένα ημιτονοειδές σήμα διαμορφώνεται από μία συνάρτηση τύπου Gaussian. Το δίκτυο, δέχεται ως είσοδο στους νευρώνες εισόδου του τα χαρακτηριστικά Gabor που έχουν προκύψει παραπάνω και ως έξοδο προκύπτουν οι αναγνωρισμένες εικόνες.



Εικόνα 22: *Η αρχιτεκτονική του νευρωνικού δικτύου των Bhuiyan και Liu*[19]

3.2.1.2.2 Νευρωνικά δίκτυα με μοντέλα Hidden Markov

Τα μοντέλα Hidden Markov (HMM), είναι στατιστικά εργαλεία πιθανοτήτων που βασίζονται στο γενικότερο μοντέλου του Markov, που χρησιμοποιείται για να μοντελοποιεί μεταβαλλόμενα συστήματα με χρήση υποθέσεων και χρησιμοποιούνται στην πρόβλεψη πιθανοτήτων. Επίσης χρησιμοποιούνται σε συνδυασμό με τα τεχνητά νευρωνικά δίκτυα για να πραγματοποιούν αναγνώριση προσώπου. Στην εργασία τους οι Bevilacqua, Cariello, Carro, Daleno και Mastronardi [31], υλοποίησαν νευρωνικό δίκτυο με Hidden Markov μοντέλο και πραγματοποίησαν ποσοστά σωστής αναγνώρισης προσώπου της τάξης του 100%. Το δίκτυο αυτό, εκπαιδεύει ψευδοδυσδιάστατα HMM που έχουν τον εξής ορισμό:

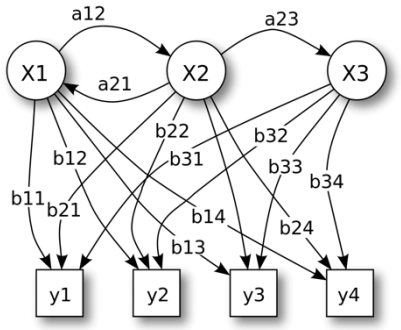
$\lambda = (A, B, \Pi)$

όπου

- $A = [a_{ij}]$ αποτελεί πίνακα πιθανότητας μεταβολής κατάστασης και a_{ij} η πιθανότητα η κατάσταση i να γίνει j
- $B = [b_j(k)]$ αποτελεί πίνακα πιθανότητας μεταβολής κατάστασης και $b_j(k)$ η πιθανότητα να υπάρχει η παρατήρηση k όταν η κατάσταση είναι j

[19] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.280&rep=rep1&type=pdf

- $\Pi = \{\pi_1,\dots,\pi_n\}$ αποτελεί την κατανομή αρχικής κατάστασης όπου π_i η πιθανότητα που σχετίζεται με την κατάσταση i

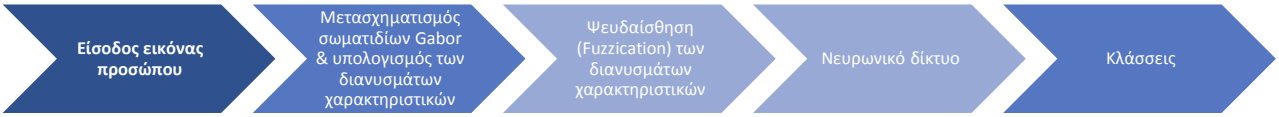


Εικόνα 23: *Πιθανοί παράμετροι ενός κρυμμένου μοντέλου Markov*[20]

X – καταστάσεις, y – πιθανές παρατηρήσεις, a – πιθανότητες μεταβατικής περιόδου, b – πιθανότητες εξόδου

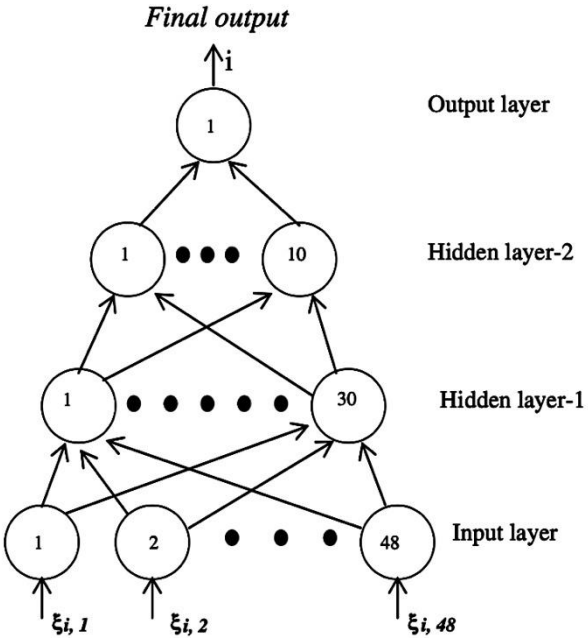
3.2.1.2.3 Ασαφή νευρωνικά δίκτυα

Τα ασαφή νευρωνικά δίκτυα, βασίζονται στα λεγόμενα ασαφή σύνολα τα οποία δημιουργήθηκαν από τον Berkeley Lotfi Zadeh [32] και παρουσιάζουν μία διαφορετική αντίληψη σχετικά με την αλήθεια και το ψεύδος μία πρότασης αφού εισήγαγαν την πλειότιμη (multivalued) λογική στην έννοια της πρότασης. Ως προς τον τομέα της αναγνώρισης προσώπων έχουν γίνει υλοποιήσεις τεχνητών νευρωνικών δικτύων με βάση την ασαφή λογική όπως το σύστημα των Bhattachrjee, Basu, Nasipuri και Kundu [33] κάνοντας χρήση επιπλέον, των κυμματομορφών Gabor που αναφέρθηκαν στην παράγραφο 3.2.1.1.3 μέσω των οποίων αποκτώνται τα διανύσματα χαρακτηριστικών. Η βασική δομή του μοντέλου αυτού είναι η εξής:



Γράφημα 5: *Βασική δομή του μοντέλου των Bhattachrjee et al.*

[20] https://en.wikipedia.org/wiki/File:HiddenMarkovModel.svg

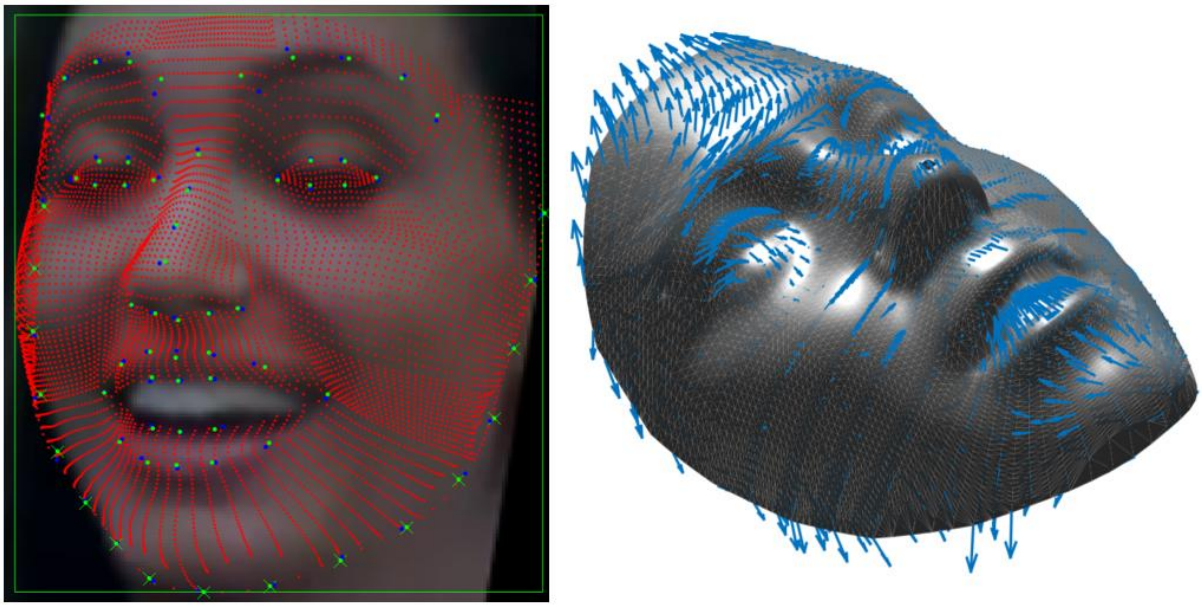


Εικόνα 24: *Αρχιτεκτονική του τεχνητού νευρωνικού δικτύου των Bhattachrjee et al.*[21]

3.2.2 Τρισδιάστατες μέθοδοι αναγνώρισης

Οι τρισδιάστατες μέθοδοι αναγνώρισης υπερτερούν ως προς τις δυσδιάστατες ως προς την δυνατότητα που προσφέρουν να χρησιμοποιούν την επιπλέον πληροφορία που διατίθεται από τις εκάστοτε εικόνες. Στα συστήματα αναγνώρισης προσώπου η επιπλέον πληροφορία αυτή, προκύπτει από την καμπυλότητα του κεφαλιού και των επιμέρους τμημάτων όπως το μέτωπο, το σαγόνι και τα μάγουλα. Μία τέτοια προσέγγιση πέρα την επιπλέον πληροφορίας που προσφέρει έχει την ιδιότητα να μην επηρεάζεται από τις όποιες διακυμάνσεις του φωτός ή της κατεύθυνσης του προσώπου. Το μοναδικό μειονέκτημα τέτοιων μεθόδων αποτελεί το υπολογιστικό κόστος.

[21] https://sci-hub.tw/10.1007/s00500-009-0426-0



Εικόνα 25: *Τριών διαστάσεων καταγραφή προσώπου*[22]

Στην εργασία του ο Gordon [34], υλοποιεί ένα σύστημα κάνοντας χρήση τρισδιάστατων εικόνων, όπου οι καμπυλότητες του προσώπου υπολογίζονται από ένα μεγάλο σύνολο δεδομένων και σε συνδυασμό με την γνώση σχετικά με την δομή του προσώπου πραγματοποιείται εντοπισμός χαρακτηριστικών όπως μάτια, μύτη μέτωπο. Ύστερα, τα καταγεγραμμένα πρόσωπα προβάλλονται πάνω σε ένα κυλινδρικό πλέγμα. Ο όγκος που καταλαμβάνουν τα πρόσωπα στον χώρο χρησιμοποιείται ως μέτρο ομοιότητας. Τα αποτελέσματα του συστήματος αυτού αγγίζουν το 100% σωστής αναγνώρισης.

3.2 Αναγνώριση φωνής

Η αναγνώριση φωνής (voice recognition), πολλές φορές συγχέεται με την αναγνώριση ομιλίας (speech recognition). Η διαφορά αυτών των δύο εννοιών είναι ότι στην αναγνώριση φωνής το εκάστοτε σύστημα αναγνωρίζει το ‘ποιος μιλάει’, ενώ στην αναγνώριση ομιλίας αναγνωρίζει το ‘τι λέγεται’. Επιπλέον, η αναγνώριση φωνής, διακρίνεται σε μεθόδους εξαρτημένες και ανεξάρτητες από κείμενο. Στις εξαρτημένες, ο ομιλητής μπορεί να διακριθεί λέγοντας κάποια συγκεκριμένη φράση ενώ στις ανεξάρτητες ο ομιλητής μπορεί να διακριθεί λέγοντας οποιαδήποτε φράση.

Η βασική προσέγγιση ενός τέτοιου συστήματος, είναι η χρήση μετασχηματισμών σήματος, οι οποίοι αναλόγως τον μετασχηματισμό πραγματοποιούν εξαγωγή

[22] https://www.micc.unifi.it/projects/sparse-3d-face-modeling-for-face-recognition-in-the-wild/

χαρακτηριστικών (feature extraction) από την φωνή και στην συνέχεια γίνεται χρήση ενός συστήματος σύγκρισης, το οποίο συγκρίνει τα εξαγόμενα χαρακτηριστικά και μέσω συναρτήσεων ομοιότητας παίρνεται η απόφαση ταυτοποίησης. Τα τελευταία χρόνια, λόγω της τεχνολογικής προόδου του υλικού υπολογιστών (hardware) χρησιμοποιούνται βαθιά νευρωνικά δίκτυα (deep neural networks) τα οποία με την κατάλληλη αρχιτεκτονική είναι σε θέση να εξάγουν από μόνα τους, ύστερα από της διαδικασία της εκπαίδευσή τους, τα χαρακτηριστικά γνωρίσματα των σημάτων φωνής.

3.2.1 Μετασχηματισμοί σήματος

3.2.1.1 Cross Correlation

Ο μετασχηματισμός αυτός, επιστρέφει την αλληλουχία αυτοσυσχέτισης του σήματος εισόδου. Έστω x το σήμα εισόδου, τότε η έξοδος r ισούται με ένα πίνακα που περιέχει όλες τις αλληλουχίες αυτοσυσχέτισης και διασταυρούμενης συσχέτισης των τιμών του x. [35]

$$R_{ff}(\tau) = (f * g_{-1}(\overline{f}))(\tau) = \int_{-\infty}^{\infty} f(u+\tau)\overline{f}(u)du = \int_{-\infty}^{\infty} f(u)\overline{f}(u-t)du$$

Όπου

f(t) = σήμα εισόδου

R_{ff} = συνεχής αυτοσυσχέτιση του σήματος

$\overline{f}$ = σύνθετο σύζευγμα

3.2.1.2 Discrete Laplacian Transform

Ο διακριτός μετασχηματισμός Λαπλάς, αποτελεί έναν ολοκληρωτικό μετασχηματισμό και απεικονίζει γραμμικά μία συνάρτηση. Ενώ ο μετασχηματισμός Λαπλάς σε ορισμένα σημεία σχετίζεται με τον μετασχηματισμό Φουριέ, διαφέρουν στο ότι ο ένας αναλύει μία συνάρτηση στις ροπές που την αποτελούν, ενώ ο άλλος την αναλύει στο φάσμα των συχνοτήτων που την αποτελούν.

$$\lim_{R\to\infty} \int_0^R f(t)e^{-ts}dt$$

3.2.1.3 Envelope

Η συνάρτηση αυτή, επιστρέφει την άνω και κάτω περιβάλλουσα του σήματος εισόδου. Κάνει χρήση επίσης του διακριτού μετασχηματισμού Φουριέ με την μέθοδο Hilbert για να αναλύσει το σήμα εισόδου.

$$F(x,t) = sin\left[2\pi\left(\frac{x}{\lambda - \Delta\lambda} - (f + \Delta f)t\right)\right] + sin\left[2\pi\left(\frac{x}{\lambda + \Delta\lambda} - (f - \Delta f)t\right)\right]$$

$$\approx 2cos\left[2\pi\left(\frac{x}{\lambda_{mod}} - \Delta f t\right)\right] sin\left[2\pi\left(\frac{x}{\lambda} - ft\right)\right]$$

3.2.1.4 Fast Fourier Transform

Ο γρήγορος μετασχηματισμός Φουριέ, πραγματοποιεί δειγματοληψία στο σήμα εισόδου σε χρονική περίοδο και το διαιρεί σε συνιστώσες συχνότητας.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}$$

3.2.1.5 Hilbert Transform

Ο μετασχηματισμός του Χίλμπερτ, είναι ένας γραμμικός χειριστής που μετατρέπει μια συνάρτηση πραγματικής μεταβλητής τύπου u(t) και παράγει μία άλλη τύπου H(u)(t).

$$H(u)(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{u(\tau)}{t - \tau} dt$$

3.2.1.6 MFCC

Οι συντελεστές συχνότητας Ceptrum του Mel (Mel-Frequency Ceptrum Coefficients), χρησιμοποιούνται σε μεγάλο βαθμό σε εφαρμογές αναγνώρισης φωνής. Οι συντελεστές αυτοί, βασίζονται στο φασματικό περιεχόμενο του σήματος και αποτελούν την

αναπαράσταση του φάσματος ισχύος που προκύπτει από ένα γραμμικό μετασχηματισμό συνημίτονου. Τα βήματα της αλγοριθμικής μεθόδου είναι:

- Υπολογισμός του μετασχηματισμού Fourier
- Εφαρμογή παραθύρου (τριγωνικά αλληλεπικαλυπτόμενα παράθυρα) για την χαρτογράφηση του φάσματος
- Καταγραφή των συχνοτήτων mel
- Εφαρμογή διακριτού μετασχηματισμού συνημίτονου
- Καταγραφή πλάτους του προκύπτοντος φάσματος

3.2.1.7 Wavelets

Τα wavelets, ως μετασχηματισμός είναι μαθηματική συνάρτηση η οποία διαιρεί ένα σήμα συνεχούς χρόνου σε διαφορετικά στοιχεία κλίμακας και μπορούν να χρησιμοποιηθούν σε σήματα συνεχούς χρόνου και όχι σε διακριτού. Υπάρχουν πολλών ειδών wavelets μετασχηματισμών και έχουν χρήση αναλόγως της εφαρμογής.

$$F(a,b) = \left(f, \psi_{a,b}\right) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi \left(\frac{t-b}{a}\right) dt$$

3.3 Συναρτήσεις απόφασης

Οι συναρτήσεις απόφασης είναι το τρίτο στάδιο ενός συστήματος ταυτοποίησης ατόμου, κατά το οποίο παίρνεται η απόφαση σε ποια κατηγορία ανήκει ένα πρότυπο.

3.3.1 Μέτρα απόστασης

Είναι τα μέτρα εκείνα τα οποία κάνουν χρήση των αποστάσεων μεταξύ των προτύπων για να ορίσουν σε ποια κατηγορία ανήκουν αυτά.

3.3.1.1 Ευκλείδεια απόσταση

Η Ευκλείδεια απόσταση αποτελεί μία συνάρτηση που αντιστοιχεί σε δύο διανύσματα. Η συνάρτηση αυτή λειτουργεί κάνοντας χρήση του Πυθαγόρειου θεωρήματος. Η έννοια της Ευκλείδειας απόστασης μεταξύ ενός τυχαίου προτύπου x και ενός προτύπου z ορίζεται ως:

$$D_2(x,z) = \|x-z\|_2 = \sqrt{(x-z)^T(x-z)} = [\sum_{i=1}^{n}(x_i - z_i)^2]^{\frac{1}{2}}$$

3.3.1.2 Ιπποδάμεια απόσταση

Η Ιπποδάμεια απόσταση ή αλλιώς Manhattan ή Cityblock αποτελεί μία συνάρτηση που αντιστοιχεί και αυτή σε δύο διανύσματα. Η χρήση της διαφέρει ως προς Ευκλείδεια απόσταση καθώς σε εφαρμογές όπως είναι οι ιατρικές, η ύψωση των διαφορών των χαρακτηριστικών διανυσμάτων στο τετράγωνο έχει ως αποτέλεσμα να υπερτονίζονται αυτές και να χρειάζονται επιπλέον υπολογισμοί. [36] Ο ορισμός της Ιπποδάμειας είναι ο εξής:

$$D_1(x,z) = \|x - z\|_1 = \sum_{i=1}^{n} |x_i - z_i|$$

3.3.1.3 Hamming απόσταση

Η απόσταση Hamming, αποτελεί μία άλλη προσέγγιση ελάχιστης απόστασης που χρησιμοποιείται σε περιπτώσεις όπου τα χαρακτηριστικά του προτύπου δεν είναι καθαρά αριθμητικά δεδομένα αλλά περιγράφονται ποιοτικά. Ορισμένα ποιοτικά δεδομένα εκφράζονται σε δυαδική μορφή όπως είναι το 'NAI-OXI' με αποτέλεσμα τα διανύσματα χαρακτηριστικών να αποτελούνται από τιμές 1 ή 0. Η απόσταση Hamming κάνει χρήση του δυαδικού τελεστή Exclusive OR ο οποίος συμβολίζεται με $\oplus$ και ορίζεται ως: [36]

$$0 \oplus 0 = 0$$
$$1 \oplus 1 = 0$$
$$0 \oplus 1 = 1$$
$$1 \oplus 0 = 1$$

3.3.1.4 Chebyshev απόσταση

Η απόσταση Chebyshev, είναι μία μέτρηση απόστασης μεταξύ δύο διανυσμάτων όπου η απόσταση αυτών είναι μεγαλύτερη από τις διαφορές τους κατά μήκος τις οποιασδήποτε διάστασης και ορίζεται ως:

$$D_\infty(x,z) = \max_i \{|x_i - z_i|\}$$

3.3.1.5 Mahalanobis απόσταση

Η απόσταση Mahalanobis, είναι μία μέτρηση απόστασης η οποία λαμβάνει υπόψιν στατιστικούς δείκτες όπως είναι η συνδιακύμανση. Ορίζεται ως:

$$D_M(x,z) = (x-z)^T C^{-1}(x-z)$$

Όπου C είναι ο πίνακας συνδιακύμανσης της κατηγορίας και το z το μέση χαρακτηριστικό διάνυσμα της κατηγορίας.

3.3.2 Μέτρα ομοιότητας

Πέραν της ύπαρξης των μέτρων απόστασης τα οποία δείχνουν πόσο διαφορετικά είναι δύο πρότυπα μεταξύ τους υπάρχουν και τα μέτρα ομοιότητας που δείχνουν πόσο όμοια είναι δύο τα δύο πρότυπα αυτά.

3.2.2.1 Εσωτερικό γινόμενο

Το εσωτερικό γινόμενο, είναι από τα πιο γνωστά μέτρα ομοιότητας διανυσμάτων και εξαρτάται από την γωνία που σχηματίζουν τα δύο διανύσματα μεταξύ τους. Το συγκεκριμένο μέτρο ομοιότητας είναι κατάλληλο σε περιπτώσεις όπου τα διανύσματα απέχουν αρκετή απόσταση μεταξύ τους όπως και από την αρχή των αξόνων. Ορίζεται ως:

$$S_i(x,z) = x^T z = \sum_{i=1}^{n} x_i z_i$$

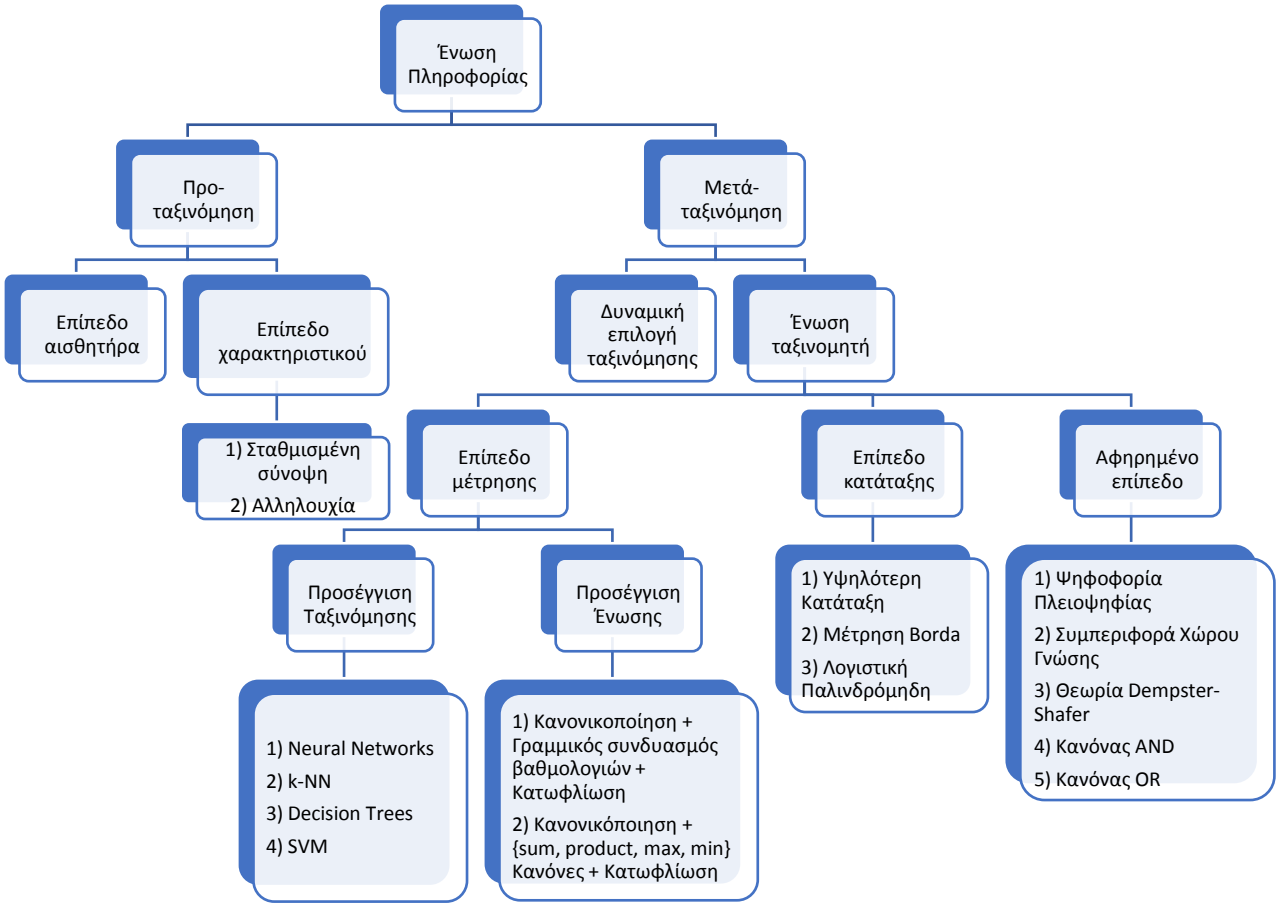
3.2.2.1 Tanimoto

Ένα επιπλέον μέτρο ομοιότητας είναι η μετρική Tanimoto η οποία χρησιμοποιείται και σε πραγματικές και σε διακριτές τιμές. Ορίζεται ως:

$$S_T(x,z) = \frac{x^T z}{x^T x + z^T z - x^T z}$$

3.4 Ένωση βιομετρικών χαρακτηριστικών

Ένα από τα σημαντικότερα τμήματα ενός πολυτροπικού βιομετρικού συστήματος (multimodal biometric system), είναι ο κατάλληλος συνδυασμός των διαφορετικών βιομετρικών χαρακτηριστικών έτσι ώστε να εκμεταλλευτούν με τον καλύτερο δυνατό τρόπο τα πλεονεκτήματα των εκάστοτε τυπικοτήτων (modalities). Υπάρχουν διάφορες προσεγγίσεις που μπορούν να ακολουθηθούν προσφέροντας οφέλη τα οποία εξαρτώνται βέβαια από την εφαρμογή αλλά και από τα ίδια τα βιομετρικά χαρακτηριστικά.



Γράφημα 6: *Προσεγγίσεις ένωσης πληροφορίας* [37]

3.4.1 Προ-ταξινόμηση

Πριν από την ταξινόμηση, η ενσωμάτωση των πληροφοριών μπορεί να πραγματοποιηθεί είτε στο επίπεδο αισθητήρα είτε στο επίπεδο χαρακτηριστικού. Τα δεδομένα σε ακατέργαστη μορφή (raw) συνδυάζονται έτσι ώστε να προκύψει το όποιο αποτέλεσμα. Δεδομένα που συλλέγονται από τους ίδιους αισθητήρες μπορούν να συνδυαστούν όπως είναι οι εικόνες ενός προσώπου, ενώ δεδομένα που συλλέγονται από διαφορετικούς αισθητήρες, ενδεχομένως να μην είναι συμβατά. Στο επίπεδο χαρακτηριστικού, διανύσματα χαρακτηριστικών δημιουργούνται από διαφορετικούς αισθητήρες όπου συνδυάζονται και προκύπτει ένα ενιαίο διάνυσμα χαρακτηριστικών. [37]

3.4.2 Μετά-ταξινόμηση

Οι προσεγγίσεις για ενσωμάτωση πληροφορίας μετά την ταξινόμηση μπορούν σε διακριθούν σε τέσσερις κατηγορίες: δυναμική επιλογή ταξινόμησης, ένωση στο επίπεδο μέτρησης, ένωση στο επίπεδο κατάταξης και ένωση στο αφηρημένο επίπεδο. Ένας δυναμικός ταξινομητής επιλέγει τα αποτελέσματα του ταξινομητή που είναι πιο πιθανόν να δώσει την σωστή απόφαση. Αυτή η προσέγγιση είναι γνωστή ως ‘ο νικητής τα παίρνει όλα’ και η εκάστοτε συσκευή που την υλοποιεί ονομάζεται ως συνδυαστικός διακόπτης. [37]

Η ενσωμάτωση στο αφηρημένο επίπεδο πραγματοποιείται όταν ο κάθε βιομετρικός μετρητής αποφασίζει μεμονωμένα για την καλύτερη αντιστοίχιση βάση της εισόδου που παρουσιάζεται σε αυτόν. Μέθοδοι όπως ψηφοφορία πλειοψηφίας (majority voting), συμπεριφορά χώρου γνώσης (behavior knowledge space), θεωρία Dempster-Shafer και κανόνες AND και OR, εφαρμόζονται σε αυτή την περίπτωση. Ένωση στο επίπεδο κατάταξης, έχουμε όταν η έξοδος κάθε βιομετρικού δείκτη αντιστοιχεί σε ένα υποσύνολο πιθανών αντιστοιχιών που ταξινομούνται με φθίνουσα σειρά. Υπάρχουν τρεις μέθοδοι σε αυτήν την προσέγγιση: η υψηλότερη κατάταξη όπου κάθε πιθανή ταξινόμηση έχει την υψηλότερη κατάταξη όπως υπολογίζεται από διαφορετικά διανύσματα, η μέτρηση Borda όπου χρησιμοποιεί το άθροισμα των τάξεων που έχουν εκχωρηθεί έτσι ώστε να υπολογίσει τις συνδυασμένες τάξεις και η λογιστική παλινδρόμηση όπου είναι μία γενίκευση της μέτρησης Borda και υπολογίζει το σταθμισμένο άθροισμα των μεμονωμένων τάξεων. [37]

Όταν οι βιομετρικοί πίνακες εξάγουν ένα σύνολο πιθανών αποτελεσμάτων μαζί με την ποιότητα της κάθε ταξινόμησης, η ενσωμάτωση μπορεί να γίνει στο επίπεδο μέτρησης. Υπάρχουν δύο διαφορετικές προσεγγίσεις στο επίπεδο μέτρησης: προσέγγιση ταξινόμησης όπου πολυδιάστατα διανύσματα (όσα βιομετρικά χαρακτηριστικά τόσες διαστάσεις)

χαρακτηριστικών εισάγονται σε ταξινομητές όπως νευρωνικά δίκτυα, SVM κλπ. και προσέγγιση ένωσης όπου τα διανύσματα που προκύπτουν από τη κάθε βιομετρική τυπικότητα ενώνονται σε ένα ενιαίο, μέσω κανονικοποίησης και εφαρμογής αλγεβρικών πράξεων μεταξύ των αποτελεσμάτων.

4 Σχέδιο Δράσης για την εκπόνηση της πτυχιακής Εργασίας

4.1 Βιβλιογραφική ανασκόπηση

4.1.1 Ταυτοποίηση μέσω χαρακτηριστικών προσώπου

Μία από τις προσεγγίσεις που ακολουθείται κατά κόρον στην ταυτοποίηση ατόμων είναι μέσω της αναγνώρισης προσώπου καθότι, το πρόσωπο είναι το πρώτο βιομετρικό χαρακτηριστικό το οποίο είναι εμφανές και εύκολα ανιχνεύσιμο. Η αναγνώριση προσώπου είναι δυνατόν να επιτευχθεί με ποικίλους τρόπους καθότι μελετάται αρκετές δεκαετίες από ερευνητές. Το 2014, οι Fakhir et al. [38] εισήγαγαν ένα νέο βιομετρικό χαρακτηριστικό προσώπου που αποτελείται από 3 διαφορετικές αποστάσεις μεταξύ των αυτιών. Το βιομετρικό αυτό, είναι δυνατό να καταγραφεί και να χρησιμοποιηθεί από μετωπικές εικόνες προσώπων μονάχα αλλά μπορεί να επιφέρει υψηλής ακρίβειας ταυτοποίηση. Στην εργασία τους οι Karczmarek et al. [39] βασίστηκαν στην μέθοδο τοπικών περιγραφών (local descriptor) της εικόνας όπως είναι το SIFT (Scale-Invariant Feature Transform) το οποίο βασίζεται στην γειτνίαση των εικονοστοιχείων. Τα εξαγόμενα χαρακτηριστικά από την μέθοδο αυτή, συνδυάζονται στην συνέχεια με την ομαδοποίηση των εικονοστοιχείων αυτών για καλύτερα αποτελέσματα.

Το 2015 οι Schroff et al. [40] προσέγγισαν το πρόβλημα της αναγνώρισης προσώπου δημιουργώντας ένα βαθύ συνελικτικό δίκτυο (deep convolutional network, CNN) 22 επιπέδων εκπαιδεύοντας το με ολόκληρες εικόνες προσώπων. Το αποτέλεσμα εκπαίδευσης που προκύπτει από το νευρωνικό αυτό δίκτυο, είναι διανύσματα τα οποία στην συνέχεια προβάλλονται στον Ευκλείδειο χώρο και υπολογίζονται οι αποστάσεις μεταξύ τους. Η βασική ιδέα είναι ότι εικόνες του ιδίου ατόμου έχουν μικρές αποστάσεις μεταξύ τους και εικόνες διαφορετικών ατόμων έχουν μεγάλες αποστάσεις. Επιπλέον, χρησιμοποιήθηκε η μέθοδος της τριπλής απώλειας (Triple Loss) έτσι ώστε η διαδικασία εκμάθησης να επιφέρει καλύτερα

αποτελέσματα. Παρόμοια προσέγγιση ακολούθησαν και οι Parkhi et al. [41] δημιουργώντας ένα βαθύ συνελικτικό δίκτυο 37 επιπέδων. Η μελέτη τους έδειξε ότι συνελικτικό δίκτυο δεν χρειάζεται να έχει πολύπλοκη δομή για να επιφέρει τα σωστά αποτελέσματα, αλλά την κατάλληλη εκπαίδευση.

Το 2018 οι Xie et al. [42] στην μελέτη τους δημιούργησαν ένα Multicolumn Network (MN) το οποίο είναι ένα τύπος deep learning δικτύου. Το δίκτυο αυτό είναι βασισμένο στο πρότυπο ResNet50 προσθέτοντας δύο επιπλέον μονάδες ελέγχου ποιότητας των εικόνων για την κανονικοποίηση αυτών (χαμηλή ποιότητα, υπερφωτισμός, μη εύρεση προσώπου). Η πρόσθεση των δύο αυτών μονάδων έδειξε ότι υφίσταται βελτίωση στα αποτελέσματα της ταυτοποίησης σε σχέση με τον να μην υπήρχαν.

4.1.2 Ταυτοποίηση μέσω χαρακτηριστικών φωνής

Μία διαφορετική προσέγγιση που ακολουθείται για την ταυτοποίηση ατόμου βασίζεται στην αναγνώριση φωνής του ομιλητή, καθώς είναι δυνατόν να επιτευχθούν υψηλής ακρίβειας αποτελέσματα. Η αναγνώριση φωνής δεν έχει μελετηθεί στον ίδιο βαθμό που έχουν μελετηθεί άλλα βιομετρικές προσεγγίσεις, αλλά τα τελευταία χρόνια γίνεται έρευνα από πολλούς ερευνητές. Ο Zhang [43] στην μελέτη του έκανε χρήση ενός Gaussian mixture model- Universal background model (GMM-UBM), το οποίο αποτελεί στην στατιστική ένα πιθανοτικό μοντέλο για την αντιπροσώπευση της ύπαρξης υποπληθυσμών σε ένα συνολικό πληθυσμό. Με βάση αυτό το μοντέλο εκπαιδεύθηκε το σύστημα από τα κοινά χαρακτηριστικά και τις ιδιότητες των προτύπων. Η μελέτη έδειξε, ότι η χρήση του συγκεκριμένου μοντέλου για να επιφέρει υψηλά αποτελέσματα χρειάζεται μεγάλος όγκος δεδομένων. Επιπλέον, έρευνα έδειξε ότι το σύστημα ανταποκρίνεται καλύτερα όταν γίνεται εφαρμογή του μετασχηματισμού σήματος PNCC (Power Normalize Cepstral Coefficients) στα αρχεία ήχου.

Οι Zhao et al. [44] στην εργασία τους κάνουν μελέτη ως προς την αφαίρεση του θορύβου και των background ήχων στην ομιλία όπως επίσης και ως προς την αναγνώριση του ομιλητή. Για την αναγνώριση του ομιλητή, χρησιμοποιήθηκε και εδώ ένα μοντέλο Gaussian mixture model- Universal background model (GMM-UBM) σε συνδυασμό με τον μετασχηματισμό σήματος GFCC (gammatone frequency cepstral coefficients) και του MFCC (mel-frequency cepstral coefficients). Η μελέτη έδειξε, ότι για να επιτευχθούν ικανοποιητικά αποτελέσματα ταυτοποίησης χρειάζονται πολλά λεπτά ομιλίας που αντιστοιχούν στο κάθε άτομο ως προς αναγνώριση.

Το 2016 οι Matejka et al. [45] κάνουν χρήση ενός Deep Neural Network (DNN) για επιτύχουν αναγνώριση φωνής. Για την διαδικασία εξαγωγής χαρακτηριστικών χρησιμοποιούν συνδυασμό του μετασχηματισμού MFCC (mel-frequency cepstral coefficients) καθώς και των χαρακτηριστικών που προκύπτουν από το Bottleneck (BN) το οποίο είναι ένα συγκεκριμένο επίπεδο του DNN. Ο συνδυασμός αυτός αποφάνθηκε ότι επιφέρει τα καλύτερα δυνατά αποτελέσματα. Υλοποίηση νευρωνικού δικτύου έκαναν και οι Nagraniy et al. [46] καθότι δημιούργησαν ένα CNN το οποίο εκπαιδεύτηκε από ένα μεγάλο όγκο δεδομένων προερχόμενο από την πλατφόρμα YouTube. Στην συνέχεια έγινε σύγκριση μεταξύ των μεθόδων της τεχνολογίας αιχμής για αναγνώριση φωνής όπου προέκυψε ότι η συγκεκριμένη CNN υλοποίηση επιφέρει τα καλύτερα αποτελέσματα.

Οι Mitsianis et al. [47] στην μελέτη τους κάνουν μια διαφορετική προσέγγιση ως προς την αναγνώριση φωνής. Από αρχεία ήχου που διαθέτουν δημιουργούν τα αντίστοιχα φασματογραφήματα αυτών τα οποία είναι εικόνες. Τις εικόνες αυτές ύστερα τις εισάγουν σε ένα CNN δίκτυο όπου προκύπτουν οι κατάλληλοι περιγραφείς που προσδιορίζουν την εκάστοτε εικόνα και πραγματοποιείται με αυτόν τον τρόπο η ταξινόμηση και στην συνέχεια η ταυτοποίηση.

4.1.3 Ταυτοποίηση μέσω συνδυασμού χαρακτηριστικών

Ο συνδυασμός βιομετρικών χαρακτηριστικών σε ένα σύστημα ταυτοποίησης ατόμου τα τελευταία χρόνια βρίσκει χρήση από πληθώρα εφαρμογών, καθότι τα αποτελέσματα δύνανται να είναι καλύτερα σε σχέση με το να χρησιμοποιούνταν ξεχωριστά τα χαρακτηριστικά αυτά καθώς επίσης η υπολογιστική ισχύ είναι σε θέση να επιφέρει την κατάλληλη επεξεργασία και εξαγωγή αποτελεσμάτων. Η επιλογή του συνδυασμού των βιομετρικών χαρακτηριστικών εξαρτάται από την χρήση και τον σκοπό του συστήματος ταυτοποίησης. Ερευνητές έχουν μελετήσει σε βάθος τους συνδυασμούς αυτούς, και έχουν εξάγει πληθώρα συμπερασμάτων.

Οι Pala et al. [48] στην εργασία του πραγματοποίησαν ταυτοποίηση ατόμου μέσω του συνδυασμού ανθρωμετρικών χαρακτηριστικών, όπως είναι οι αποστάσεις κεφαλιού-πατώματος, λαιμού-πατώματος και ώμων-πλάτης. Η καταγραφή των χαρακτηριστικών αυτών έγινε μέσω RGB-D κάμερας που έχει την δυνατότητα λήψης 3 διαστάσεων πληροφορίας. Ένα άλλος γνωστός συνδυασμός βιομετρικών χαρακτηριστικών είναι εκείνος που χρησιμοποιεί πληροφορία δοθείσα από τα χέρια όπως έκαναν στην μελέτη τους οι Angadi et al. [49] κάνοντας χρήση των χαρακτηριστικών της γεωμετρίας των χεριών και της παλάμης με την βοήθεια των Support Vector Machines (μηχανές

διανυσμάτων υποστήριξης) οι οποίες χρησιμοποιήθηκαν για την ταξινόμηση και για την τελική απόφαση ταυτοποίησης. Το 2017 ο Kovoor [50] στην έρευνα του έκανε χρήση τριών βιομετρικών χαρακτηριστικών, του προσώπου, της φωνής και της εικόνας της υπογραφής (μέσω εικόνας). Χρησιμοποίησε για την αναγνώριση του προσώπου τον αλγόριθμο Κύριων Συνιστωσών (Principal Component Analysis, PCA), για την αναγνώριση της φωνής τον μετασχηματισμό MFCC και για την αναγνώριση της υπογραφής τον μετασχηματισμό Gabor Wavelets (GWT). Για την σύγκριση των διανυσμάτων που προέκυψαν έγινε χρήση της Mahalanobits απόστασης για την τελική απόφαση ταυτοποίησης.

Οι Soleymani et al. [51] το 2018, στην εργασία του χρησιμοποίησαν ένα CNN δίκτυο, το οποίο αποτελείται από ξεχωριστά δίκτυα, ένα για κάθε τύπο χαρακτηριστικού (modality). Καθένα από αυτά τα δίκτυα πραγματοποιεί εξαγωγή χαρακτηριστικών (feature extraction) από τα δεδομένα εισόδου. Στην μελέτη τους, παρουσιάζουν διαφορετικές προσεγγίσεις για το κομμάτι της ένωσης (fusion) των χαρακτηριστικών. Προκύπτει το συμπέρασμα ότι η καλύτερη μέθοδος ένωσης είναι η χρήση ενός bilinear CNN model, όπου στην έξοδο του, οι επιμέρους έξοδοι των δικτύων πολλαπλασιάζονται μεταξύ τους για να προκύψουν τα ενοποιημένα δεδομένα. Τα βιομετρικά χαρακτηριστικά που μελετήθηκαν ήταν ο συνδυασμός εικόνων προσώπου, οι συνδυασμός ίριδας δακτυλικών αποτυπωμάτων και ο συνδυασμός προσώπου-ίριδας-δακτυλικού αποτυπώματος. Επίσης, η ίδια ομάδα ερευνητών σε μία άλλη εργασία τους [52] δημιούργησαν το ίδιο μοντέλο, με την διαφορά ότι αντί να πολλαπλασιάζονται οι έξοδοι μεταξύ τους, υπάρχει ένα επιπλέον επίπεδο στο υπάρχων CNN δίκτυο όπου εκεί πραγματοποιείται η ένωση των εξόδων των επιμέρους δικτύων.

Οι Kita et al. [53], πραγματοποίησαν ταυτοποίηση ατόμου μέσω των χαρακτηριστικών του προσώπου και της φωνής κάνοντας χρήση της συσκευής Kinect της Microsoft. Στην αναγνώριση προσώπου χρησιμοποιήθηκε η υλοποιημένη εφαρμογή WebAPI που εντοπίζει και υπολογίζει χαρακτηριστικά του προσώπου όπως είναι τα μάτια, μύτη, στόμα, φρύδια, περίγραμμα προσώπου. Στην αναγνώριση φωνής χρησιμοποιήθηκε το toolkit HTK που κάνει χρήση των συντελεστών συχνότητας cepstral και Δ λογαριθμικών δυνάμεων. Στην συνέχεια αναπτυχθήκαν και συγκρίθηκαν δύο μέθοδοι ταξινόμησης, αυτή των SVM (support vector machines) και αυτή των NN (neural networks) από όπου προέκυψαν καλύτερα αποτελέσματα με χρήση των NN.

Στην εργασία τους οι Ren et al. [54] κάνοντας χρήση των χαρακτηριστικών του προσώπου και της φωνής επέτυχαν υψηλά ποσοστά ταυτοποίησης της τάξης του 91.38 %.

Στην διαδικασία της εξαγωγής χαρακτηριστικών προσώπου χρησιμοποιήθηκε ένα CNN δίκτυο καθώς και ο αλγόριθμος PCA για μείωση των διαστάσεων των χαρακτηριστικών. Στην αναγνώριση της φωνής εφαρμόσθηκε ο μετασχηματισμός MFCC. Τα παραπάνω χαρακτηριστικά στην συνέχεια εισήλθαν σε ένα δίκτυο LSTM (long short term memory) που βασίζεται σε κυψέλες μνήμης για να πραγματοποιήσει ταξινόμηση.

Το 2018 οι Kauffman et al. [55] στην μελέτη τους έκαναν χρήση και αυτοί των δύο βιομετρικών χαρακτηριστικών του προσώπου και της φωνής. Στην αναγνώριση φωνής χρησιμοποιήθηκε για εξαγωγή χαρακτηριστικών ο μετασχηματισμός MFCC και στην συνέχεια ένα Gaussian Mixture Model (GMM) το οποίο είναι ένα μοντέλο πιθανοτήτων. Στον εντοπισμό προσώπου ακολουθήθηκε ο αλγόριθμος Viola-Jones και στην αναγνώριση προσώπου δημιουργήθηκε ένα CNN δίκτυο6 επιπέδων. Η ένωση των δύο βιομετρικών χαρακτηριστικών γίνεται μέσω αθροίσματος των επιμέρους αποτελεσμάτων που προκύπτουν από αυτά. Στο στάδιο της τελικής ταξινόμησης έγινε χρήση ενός συστήματος SVM. Τέλος, οι Santana et al. [56] χρησιμοποιούν Deep Predictive Coding Networks όπου είναι είδος CNN δικτύων για την ταυτοποίηση ατόμου μέσω χαρακτηριστικών προσώπου και φωνής. Η βασική δομή του δικτύου αυτού είναι δύο επιμέρους δίκτυα, ένα για κάθε είδος πληροφορίας.

Στην συνέχεια παρατίθενται πίνακες με όλες τις έρευνες που προαναφέρθηκαν με πληροφορίες ως προς τα σετ δεδομένων που χρησιμοποιήθηκαν καθώς και τα αποτελέσματα τις εκάστοτε έρευνας.

Πίνακας 1: *Επισκόπηση ερευνών για ταυτοποίηση μέσω χαρακτηριστικών προσώπου*

Ταυτοποίηση μέσω χαρακτηριστικών προσώπου				
Μελέτη	**Σετ Δεδομένων**	**Μέθοδος**	**Τύπος Εικόνας**	**Αποτελέσματα**
Fakhir et al. [38]	*Δημιουργία δική τους*	Features Measurement	Frontal	94%
Karczmarek et al. [39]	AT&T database	Local Descriptors (gray scale level)	Frontal	98.45%
Schroff et al. [40]	Labeled Faces in the Wild YouTube Faces DB	Deep Convolutional Network	In the wild	98.87% 95.12%
Parkhi et al. [41]	Labeled Faces in the Wild YouTube Faces DB	Deep Convolutional Network	In the wild	98.95% 97.3%

Xie et al. [42]	VGGFace2	Multicolumn Network	In the wild	false accept rate 0.989

Πίνακας 2: *Επισκόπηση ερευνών για ταυτοποίηση μέσω χαρακτηριστικών φωνής*

Ταυτοποίηση μέσω χαρακτηριστικών φωνής				
Μελέτη	**Σετ Δεδομένων**	**Μέθοδος**	**Εξάρτηση από το κείμενο**	**Αποτελέσματα**
Zhang [43]	*Δημιουργία δική του*	Gaussian mixture model-Universal background model	Όχι	97%
Zhao et al. [44]	NIST Speaker Recognition Evaluation dataset	Gaussian mixture model-Universal background model	Όχι	84,83%
Matejka et al. [45]	PRISM set	Deep Neural Network (Bottleneck)	Όχι	equal error rate 0.009381
Nagraniy et al. [46]	YouTube	Deep Convolutional Network	Όχι	92.1%
Mitsianis et al.	Multi-View Stereo	Deep	NAI	88.88%

[47]	correspondence	Convolutional Network		

Πίνακας 3: *Επισκόπηση ερευνών για ταυτοποίηση μέσω συνδυασμού χαρακτηριστικών*

Ταυτοποίηση μέσω συνδυασμού χαρακτηριστικών				
Μελέτη	**Σετ Δεδομένων**	**Βιομετρικά Χαρακτηριστικά**	**Μέθοδοι**	**Αποτελέσματα**
Pala et al. [48]	KinectREID dataset, RGBD−ID dataset	Ανθρωπομετρικά (αποστάσεις πατώματος-κεφαλιού,λαιμού κλπ)	Ταξινομητής Ευκλείδειας απόστασης	57%, 60%
Angadi et al. [49]	Virtual Multimodal Hand and Palmprint Database	Γεωμετρία χεριών-παλάμης	Support Vector Machines	99.19%
Kovoor [50]	*Δημιουργία δική του*	Πρόσωπο-φωνή-υπογραφή	PCA-MFCC-GWT	Success ratio (mean) 0.81
Soleymani et al. [51]	CMU Multi-Pie database, BIOMDATA multimodal database BioCop multimodal database	Πρόσωπο-ίριδα-δακτυλικά αποτυπώματα	Deep Convolutional Network	97.27% 99.90% 99.30%
Soleymani et al. [52]	BIOMDATA multimodal database BioCop multimodal database	Πρόσωπο-ίριδα-δακτυλικά αποτυπώματα	Deep Convolutional Network	99.91% 99.34%

Πίνακας 4: *Επισκόπηση ερευνών για ταυτοποίηση μέσω συνδυασμού χαρακτηριστικών (πρόσωπο-φωνή)*

Ταυτοποίηση μέσω συνδυασμού χαρακτηριστικών (πρόσωπο φωνή)						
Μελέτη	**Σετ**	**Βιομετρικά**	**Μέθοδοι**	**Τύπος**	**Εξάρτηση**	**Αποτελέσματα**

	Δεδομένων	**Χαρακτηριστικά**		**Εικόνας**	**από το κείμενο**	
Kita et al. [53]	*Δημιουργία δική τους*	Πρόσωπο-φωνή	Neural Networks	Frontal	Ναι	98.7%
Ren et al. [54]	The Bing Bang Theory dataset	Πρόσωπο-φωνή	PCA-MFCC-CNN-LSTM	In the wild	Όχι	91.38%
Kauffman et al. [55]	Digital Democracy Dataset (2014-2016)	Πρόσωπο-φωνή	MFCC-GMM-CNN	In the wild	Όχι	70.18%
Santana et al. [56]	the VIDTIMIT dataset	Πρόσωπο-φωνή	Deep Predictive Coding Networks	Frontal	Όχι	85.7%

5 Κύριο μέρος Πτυχιακής Εργασίας

5.1 Η βάση δεδομένων

Το σετ δεδομένων που χρησιμοποιήθηκε στην παρούσα πτυχιακή εργασία είναι το VidTIMIT Audio-Video dataset. Αποτελείται από βίντεο και αντίστοιχες ηχογραφήσεις43 ατόμων, απαγγέλοντας σύντομες προτάσεις. Μπορεί να είναι χρήσιμο για την έρευνα σε θέματα όπως η αυτόματη αναγνώριση των χειλών, η αναγνώριση προσώπου πολλαπλών προβολών, η πολυτροπική αναγνώριση ομιλίας και η αναγνώριση προσώπου.

Το σετ δεδομένων καταγράφηκε σε 3 συνεδρίες, με μέση καθυστέρηση 7 ημερών μεταξύ της συνεδρίας 1 και 2 και 6 ημερών μεταξύ της συνεδρίας 2 και 3. Οι προτάσεις που λέγονται εξελέγησαν από το τμήμα δοκιμών του σώματος TIMIT. Υπάρχουν 10 προτάσεις ανά άτομο. Οι πρώτες 6 προτάσεις (ταξινομημένες αλφαριθμητικά κατά όνομα αρχείου) αντιστοιχίζονται στη περίοδο 1. Οι επόμενες 2 προτάσεις δίνονται στην περίοδο 2 και με τις υπόλοιπες 2 έως την περίοδο 3. Οι πρώτες 2 προτάσεις για όλα τα άτομα είναι ίδιες, ενώ οι υπόλοιπες οκτώ γενικά διαφέρουν για κάθε άτομο.

Εκτός από τις προτάσεις, κάθε άτομο εκτέλεσε μια σειρά περιστροφής κεφαλής σε κάθε συνεδρία. Η ακολουθία αποτελείται από το άτομο που μετακινεί το κεφάλι του προς τα αριστερά, δεξιά, πίσω στο κέντρο, πάνω, στην συνέχεια κάτω και τελικά επιστρέφει στο κέντρο.

Η εγγραφή έγινε σε περιβάλλον γραφείου χρησιμοποιώντας ψηφιακή βιντεοκάμερα. Το βίντεο κάθε ατόμου αποθηκεύεται ως αριθμημένη σειρά εικόνων JPEG με ανάλυση 512 x 384 pixel. Χρησιμοποιήθηκε ρύθμιση ποιότητας 90% κατά την δημιουργία των εικόνων JPEG. Ο αντίστοιχος ήχος αποθηκεύεται ως μονοφωνικό αρχείο 16 bit 32 kHz WAV.

5.2 Πειραματικό μέρος

5.2.1 Αναγνώριση προσώπου

5.2.1.1 Πειραματικό μέρος 1

Στο πειραματικό μέρος 1, εκπαιδεύσαμε το σύστημα με 5 εικόνες ανά άτομο (σετ εκπαίδευσης). Οι εικόνες αυτές είναι frames του σετ δεδομένων αυτούσια δίχως καμία προεπεξεργασία κλπ. Οι 5 αυτές εικόνες είναι διαφορετικών κλίσεων κεφαλιού μεταξύ τους. Το σύστημα στην συνέχεια ελέγχθηκε με 700 περίπου εικόνες ανά άτομο (σετ ελέγχου) και 37.377 συνολικά. Ελέγχθηκαν εικόνες που προέρχονται από 2 διαφορετικά βίντεο (video 1 & video 2), όπου στο κάθε βίντεο το εκάστοτε άτομο περιστρέφει το κεφάλι του καθώς επίσης οι συνθήκες κατά τις οποίες έχει καταγραφεί είναι διαφορετικές (χρώματα ρούχων, χτένισμα, κούρεμα). Οι διαστάσεις των εικόνων είναι 512 x 384 pixels.

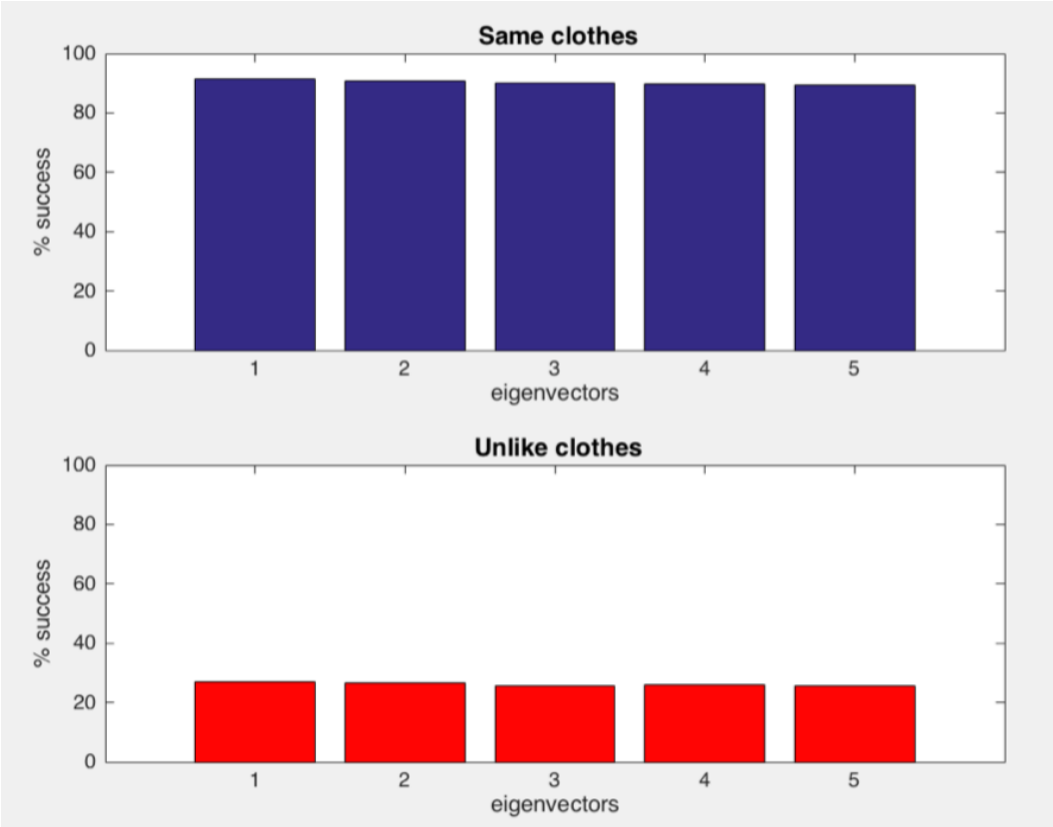
Στο κομμάτι τις αναγνώρισης προσώπου εφαρμόσθηκε ο αλγόριθμος PCA και στο κομμάτι του ελέγχου δημιουργήθηκε και χρησιμοποιήθηκε ένας ταξινομητής ομοιότητας κάνοντας χρήση της νιοστής νόρμας (n-norm). Ο σκοπός του πειραματικού μέρους 1, είναι να βρεθεί ο επαρκής αριθμός ιδιοδιανυσμάτων (eigenvectors) καθώς επίσης να δούμε την συμπεριφορά της ταξινόμησης ως προς εικόνες ελέγχου διαφορετικών συνθηκών.

Εικόνα 26: *Παράδειγμα εικόνων εκπαίδευσης για ένα άτομο*

Εικόνα 27*: Παράδειγμα εικόνας ελέγχου για ένα άτομο (video 1)*

Εικόνα 28: *Παράδειγμα εικόνας ελέγχου για ένα άτομο (video 2)*

Αποτελέσματα



Εικόνα 29: *Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης προσώπου*

Από το διαγράμματα που προέκυψαν βλέπουμε: με μπλε χρώμα τις εικόνες ελέγχου που προέρχονται από το ίδιο βίντεο (video 1) με τις εικόνες εκπαίδευσης με ακρίβεια αποτελεσμάτων περίπου 93% και με κόκκινο χρώμα τις εικόνες που προέρχονται από διαφορετικό βίντεο (video 2) από τις εικόνες εκπαίδευσης με ακρίβεια αποτελεσμάτων περίπου 25%. Καταλήγουμε στο συμπέρασμα ότι το σύστημα επηρεάζεται από την πλεονάζουσα πληροφορία πέραν την πληροφορίας του προσώπου.

5.2.1.1 Πειραματικό μέρος 2

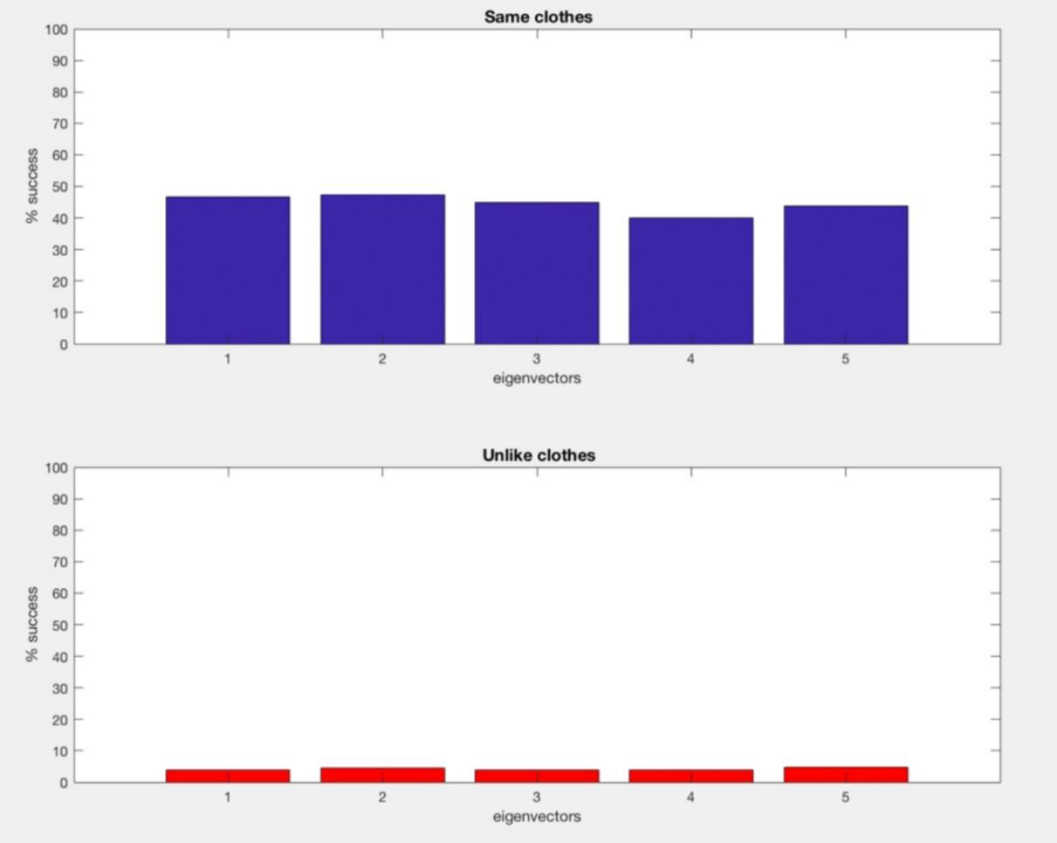
Λόγω της πλεονάζουσας πληροφορίας πέραν του προσώπου κρίθηκε απαραίτητο να γίνει εντοπισμός προσώπου και στην συνέχεια περικοπή έτσι ώστε να παραμείνουν μόνο τα πρόσωπα. Ο εντοπισμός του προσώπου έγινε με χρήση της ανοιχτής κώδικα εργαλειοθήκης OpenFace. Λόγω του ότι μέσω του εντοπισμού και στην συνέχεια της περικοπής, οι εικόνες που προέκυψαν δεν είχαν τα ίδιο μέγεθος (ο αλγόριθμος PCA χρειάζεται ίδιου μεγέθους εικόνες) εφαρμόσθηκε η τεχνική γεμίσματος εικόνας (padding) έτσι ώστε να έχουν όλες οι εικόνες το ίδιο μέγεθος.

Ο αριθμός των εικόνων εκπαίδευσης και ελέγχου είναι ίδιος με το πειραματικό μέρος 1 αλλά οι διαστάσεις των εικόνων είναι πλέον 219 x 211 pixels.

Εικόνα 30: *Εικόνες ύστερα από τον εντοπισμό προσώπου*

Εικόνα 31: *Εικόνες ύστερα από το γέμισμα (padding)*

Αποτελέσματα



Εικόνα 32: *Αποτελέσματα πειραματικού μέρους 2 αναγνώρισης προσώπου*

Από το διαγράμματα που προέκυψαν παρατηρούμε ότι ποσοστά ακρίβειας και στις δύο περιπτώσεις είναι αισθητά μειωμένα. Από 93 % σε 48% και από 25% σε 5%. Βλέπουμε ότι αν και η πληροφορία που δόθηκε ως προς εκπαίδευση και έλεγχο είναι μόνο το πρόσωπο που είναι και το ζητούμενο, τα αποτελέσματα είναι χαμηλά και μη αναμενόμενα.

Λόγω αυτού, αποφασίσθηκε να αυξηθεί ο αριθμός εικόνων εκπαίδευσης από **5** σε **200** εικόνες. Η αύξηση αυτή όμως μας δημιουργεί πρόβλημα ως προς το υπολογιστικό κόστος, καθότι η διαδικασία ελέγχου πλέον, είναι μία χρονοβόρος διαδικασία σε βαθμό τέτοιο που δεν θα μπορούσε να έρθει εις πέρας σε επιτρεπτό χρονικό περιθώριο. Ο παρακάτω πίνακας παρουσιάζει τους χρόνους εκπαίδευσης και ελέγχου. Με 200 εικόνες εκπαίδευσης, ο χρόνος ελέγχου για 1 εικόνα μπορεί να φθάσει έως και 83 λεπτά, κάτι το οποίο είναι απαγορευτικό.

Έτσι λοιπόν, αποφασίσθηκε να γίνει αλλαγή μεγέθους (resize) στις εικόνες και από 219x211

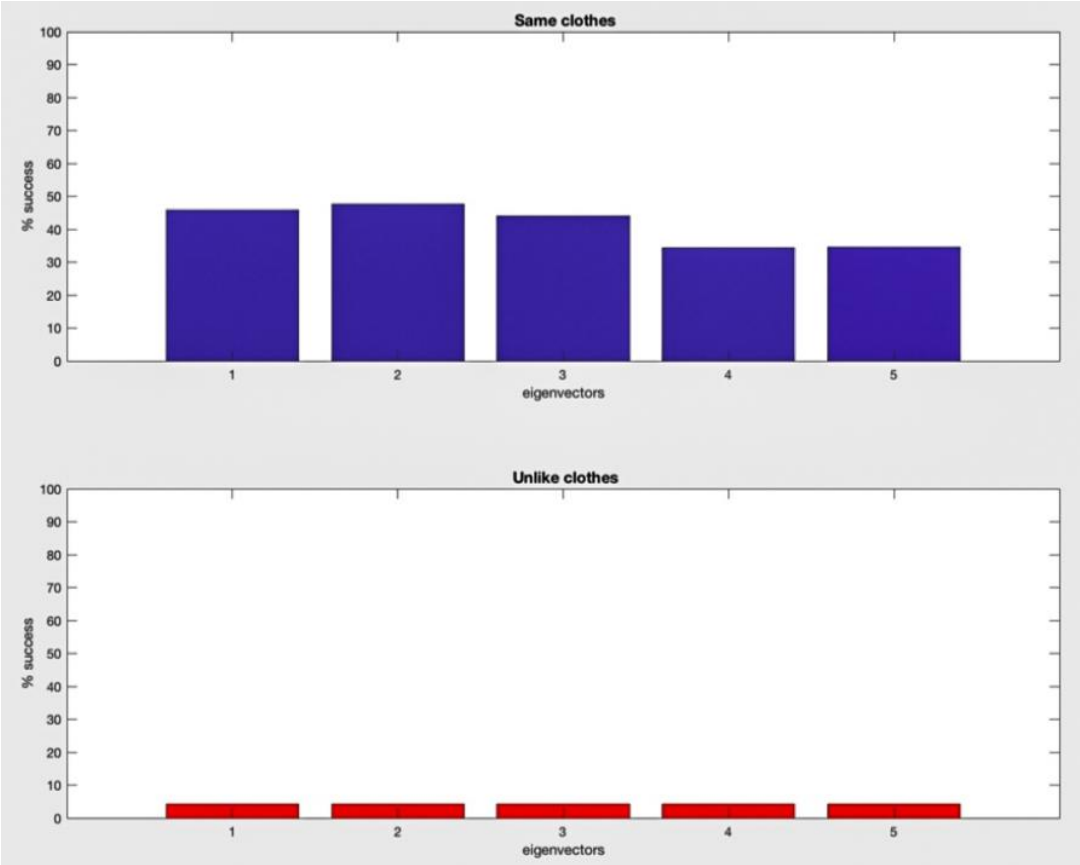
pixels να γίνουν 55x53 pixels. Με αυτή την αλλαγή βλέπουμε ότι ο χρόνος των 83 λεπτών για τον έλεγχο 1 εικόνας μειώθηκε σε 2.83 λεπτά.

5 εικόνες εκπαίδευσης 219x211 (ανά άτομο)					
training time	eigenvector 1	eigenvector 2	eigenvector 3	eigenvector 4	eigenvector 5
	1.1 sec	1.4 sec	1.5 sec	1.6 sec	1.8 sec
testing time (1image)	0.38 sec	0.56 sec	1.09 sec	1.16 sec	1.51 sec
200 εικόνες εκπαίδευσης 219x211 (ανά άτομο)					
training time	eigenvector 1	eigenvector 50	eigenvector 100	eigenvector 150	eigenvector 197
	29.7 sec	32.9 sec	32.2 sec	34.8 sec	48.6 sec
testing time (1image)	17.3 sec	5.25 min	42.0 min	62.0 min	83.0 min

5 εικόνες εκπαίδευσης 55x53 (ανά άτομο)					
training time	eigenvector 1	eigenvector 2	eigenvector 3	eigenvector 4	eigenvector 5
	0.81 sec	0.82 sec	0.92 sec	0.94 sec	0.99 sec
testing time (1image)	0.12 sec	0.14 sec	0.15 sec	0.15 sec	0.16 sec
200 εικόνες εκπαίδευσης 55x53 (ανά άτομο)					
training time	eigenvector 1	eigenvector 50	eigenvector 100	eigenvector 150	eigenvector 197
	14.7 sec	15.1 sec	15.7 sec	20.1 sec	25.3 sec
testing time (1image)	1.20 sec	28.17 sec	1.23 min	2.51 min	2.83 min

Εικόνα 33: *Χρόνοι εκπαίδευσης και ελέγχου αναγνώρισης προσώπου*

Βέβαια, με την αλλαγή των διαστάσεων των εικόνων θέλαμε να βεβαιωθούμε αν τα αποτελέσματα επηρεάζονται. Για τον λόγο αυτό, πραγματοποιήθηκε το προηγούμενο πείραμα αλλά με τις νέες διαστάσεις των εικόνων. Στο παρακάτω γράφημα βλέπουμε ότι τα αποτελέσματα επηρεάστηκαν σε μικρό βαθμό κάτι το οποίο είναι πλήρως ανεκτό αν αναλογιστούμε την μείωση του υπερμεγέθους υπολογιστικού κόστους.



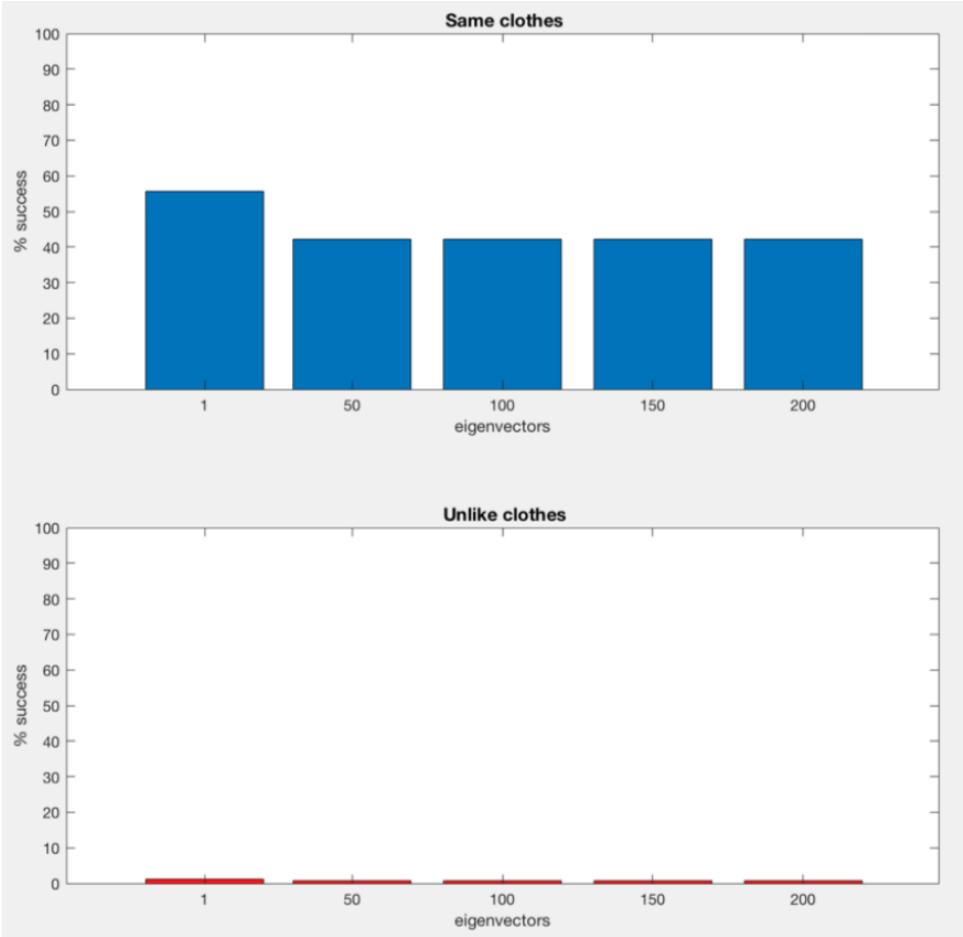
Εικόνα 34: *Αποτελέσματα πειράματος resize*

5.2.1.2 Πειραματικό μέρος 3

Οι εικόνες εκπαίδευσης είναι 200 ανά άτομο με αποτέλεσμα να αυξηθεί η μεταβλητότητα των στάσεων του προσώπου και οι εικόνες ελέγχου είναι περίπου 500 ανά άτομο και 28.992 συνολικά.

Εικόνα 35: *Παράδειγμα εικόνων εκπαίδευσης για ένα άτομο*

Αποτελέσματα

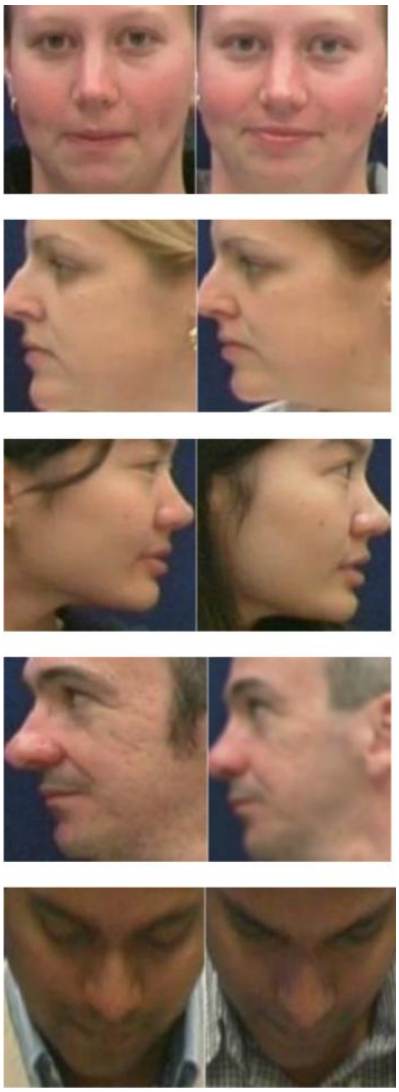


Εικόνα 36: *Αποτελέσματα πειραματικού μέρους 3 αναγνώρισης προσώπου*

Από τα παραπάνω διαγράμματα βλέπουμε ότι στην πρώτη περίπτωση (εικόνες προερχόμενες από το video 1) έχουμε αύξηση ακρίβειας αποτελεσμάτων από 48% σε 55% κάνοντας χρήση του πρώτου ιδιοδιανύσματος αλλά μείωση ακρίβειας αποτελεσμάτων από 5% σε 2% στην δεύτερη περίπτωση κάνοντας χρήση και πάλι του πρώτου ιδιοδιανύσματος.

Λόγου των διαγραμμάτων αυτών, συμπεραίνουμε ότι παρόλο που πραγματοποιήθηκε εντοπισμός προσώπου και περικοπή αυτού, στις εικόνες παραμένουν ορισμένα στοιχεία τα οποία επηρεάζουν και συμβάλλουν καθοριστικά στην λανθασμένη ταυτοποίηση. Τα στοιχεία αυτά είναι το χρώμα των ρούχων, το χτένισμα μαλλιών, το κούρεμα, κλπ. Στις παραπάνω εικόνες βλέπουμε ότι όντως τα στοιχεία αυτά παραμένουν.

Εικόνα 37: *Παράδειγμα αρχικών εικόνων*



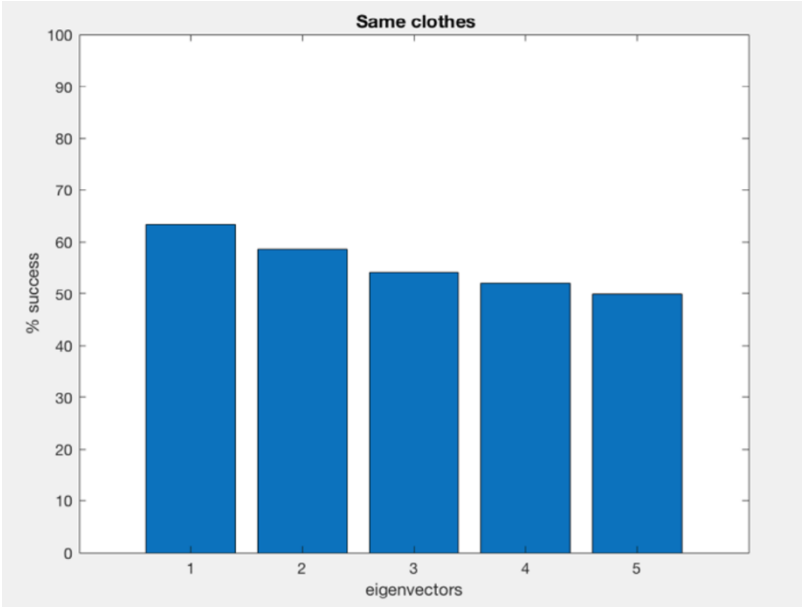
Εικόνα 38: *Παράδειγμα εικόνων ύστερα από εντοπισμό προσώπου*

5.2.1.3 Πειραματικό μέρος 4

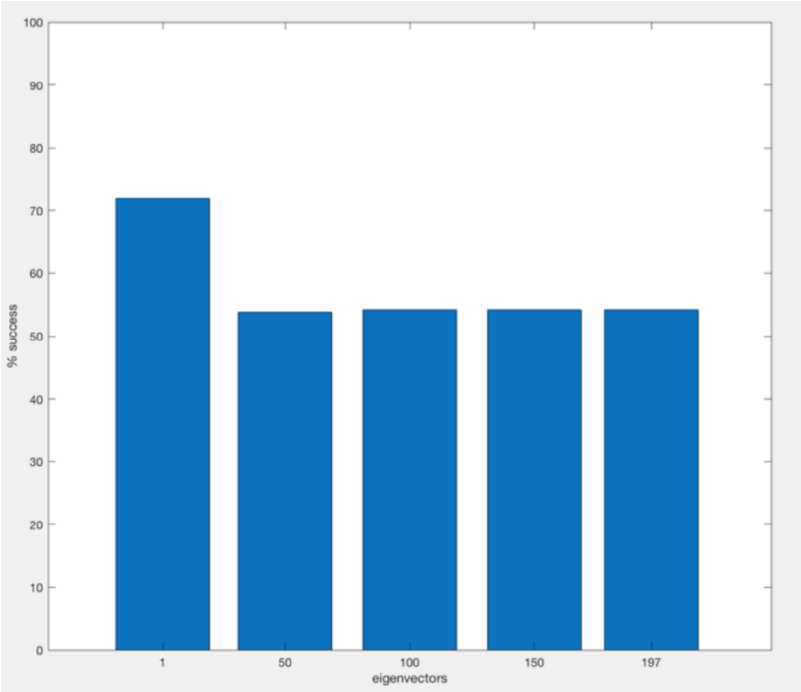
Στο πειραματικό μέρος 4, πραγματοποιήθηκαν δύο πειράματα τα οποία είχαν σκοπό να παρουσιάσουν την συμπεριφορά του συστήματος σε μετωπικές εικόνες στο κομμάτι του ελέγχου. Το πρώτο, αφορά την εκπαίδευση του συστήματος με 5 εικόνες όπως έχει γίνει ήδη αλλά για τον έλεγχο χρησιμοποιήθηκαν εικόνες οι οποίες είναι αποκλειστικά μετωπικές (frontal). Στο δεύτερο μέρος οι εικόνες εκπαίδευσης είναι 200 όπως στο πείραμα 3, αλλά και εδώ χρησιμοποιήθηκαν μετωπικές εικόνες για έλεγχο. Συνολικά έγινε έλεγχος σε 4.946 εικόνες.

Εικόνα 39: *Παράδειγμα μετωπικής εικόνας*

Αποτελέσματα



Εικόνα 40: *Αποτελέσματα πειραματικού μέρους 4 αναγνώρισης προσώπου (1)*



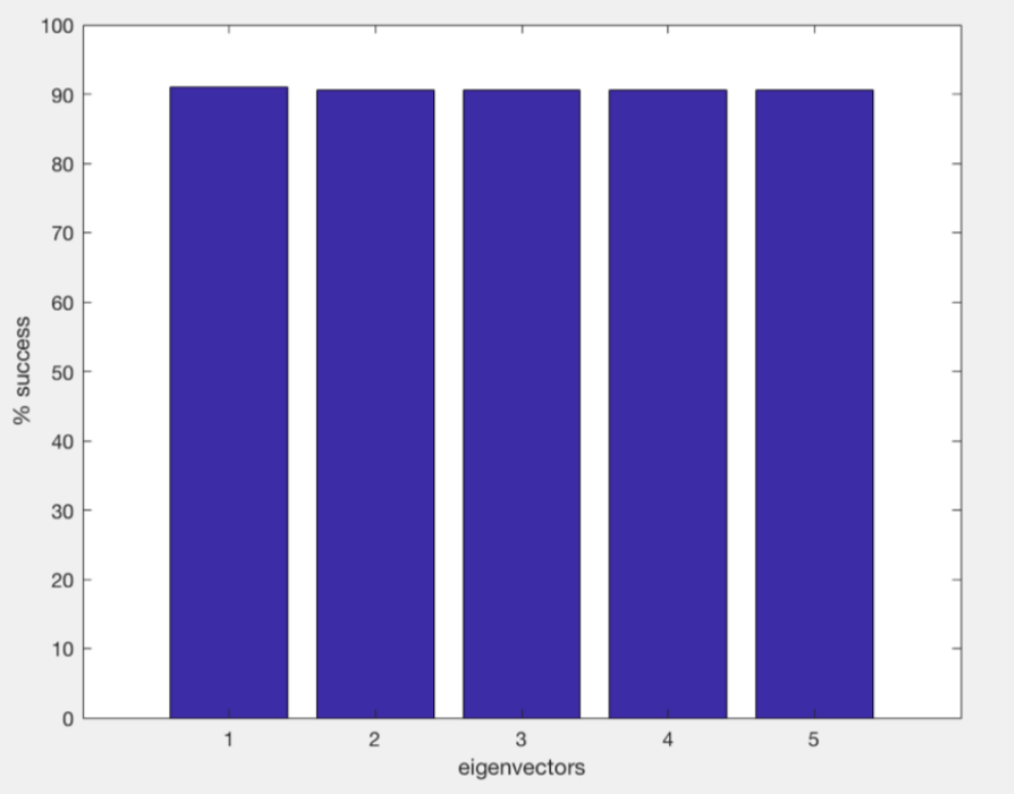
Εικόνα 41: *Αποτελέσματα πειραματικού μέρους 4 αναγνώρισης προσώπου (2)*

Από τα παραπάνω διαγράμματα προκύπτει ότι :

- Κάνοντας χρήση του πρώτου ιδιοδιανύσματος πετυχαίνουμε μεγαλύτερα αποτελέσματα ακρίβειας σε σχέση με τα υπόλοιπα
- Με 5 εικόνες εκπαίδευσης με 1 ιδιοδιάνυσμα έχουμε ακρίβεια της τάξης του 62%
- Με 200 εικόνες εκπαίδευσης με 1 ιδιοδιάνυσμα έχουμε ακρίβεια της τάξης του 72%

5.2.1.4 Πειραματικό μέρος 5

Στο πειραματικό μέρος 5, το σύστημα εκπαιδεύτηκε με 5 μετωπικές εικόνες ανά άτομο και ελέγχθηκαν 50 μετωπικές εικόνες ανά άτομο και 2.150 συνολικά. Ο σκοπός του πειράματος αυτού είναι η συμπεριφορά του συστήματος σε αποκλειστικά μετωπικές εικόνες εκπαίδευσης και ελέγχου.



Εικόνα 42: *Αποτελέσματα πειραματικού μέρους 5 αναγνώρισης προσώπου*

Από το παραπάνω διάγραμμα προκύπτει ότι το σύστημα συμπεριφέρεται κατά μεγάλο ποσοστό ακριβέστερα κάνοντας χρήση μετωπικών εικόνων και μόνο. Με χρήση οποιουδήποτε ιδιοδιανύσματος τα ποσοστά ακρίβειας είναι λίγο μεγαλύτερα του 90%

5.2.2 Αναγνώριση φωνής

5.2.2.1 Πειραματικό μέρος 1

Το πειραματικό μέρος 1 αποτελείται από δύο σκέλη. Το πρώτο αφορά την δημιουργία του σετ ελέγχου και το δεύτερο δοκιμές διαφόρων μετασχηματισμών σήματος στις ηχογραφήσεις των φωνών. Η μελέτη της αναγνώρισης φωνής στην παρούσα πτυχιακή εργασία βασίζεται σε μέθοδο που βιβλιογραφικά αναφέρεται ως ‘εξαρτώμενη από το κείμενο’. Αυτό σημαίνει ότι ο έλεγχος του συστήματος γίνεται με λέξεις ή φράσεις που είναι γνωστές στο σύστημα καθότι έχει εκπαιδευτεί με αυτές. Στην συγκεκριμένη περίπτωση, από σετ δεδομένων VidTIMIT, χρησιμοποιούμε την πρώτη φράση που είναι ίδια για όλα τα

άτομα. Αυτό σημαίνει όμως, ότι δεν υπάρχει άλλο αρχείο με την ίδια φράση για έλεγχο συστήματος.

Για τον λόγο αυτό, δημιουργήσαμε αντίγραφα του αρχείου αυτού (1 για κάθε άτομο) όπου στην συνέχεια τα τροποποιήσαμε ως προς την ένταση (volume), ως προς το μπάσο (bass) και ως προς τον Γκαουσιανό θόρυβο (Gaussian noise). Με τον τρόπο αυτό πετυχαίνουμε ως ένα βαθμό επίσης την δημιουργία πραγματικών συνθηκών γραφείου. Οι τροποποιήσεις που εφαρμόσθηκαν είναι:

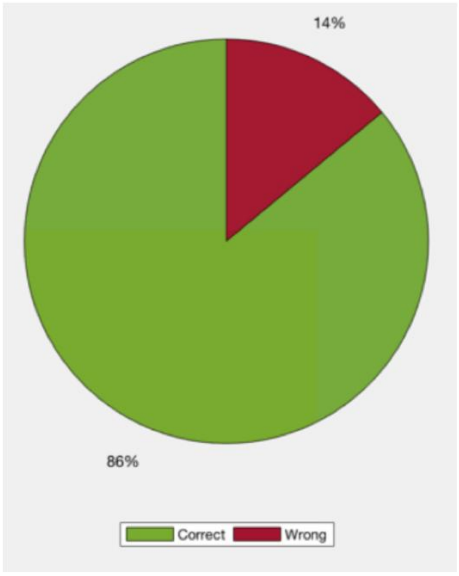
- **Volume:** 15db, -14db, -13db, -12db, -11db, -10db, -9db, -8db, -7db, 6db, -5db, -4db, -3db, -2db, -1db, +1db, +2db, +3db, +4db, +5db, +6db, +7db, +8db, +9db. +10db, +11db, +12db, +13db, +14db, +15db
- **Bass:** 10/Fs, 100/Fs, 1.000/Fs, 10.000/Fs
- **Gaussian Noise:** 15db, -14db, -13db, -12db, -11db, -10db, -9db, -8db, -7db, 6db, -5db, -4db, -3db, -2db, -1db, +1db, +2db, +3db, +4db, +5db, +6db, +7db, +8db, +9db. +10db, +11db, +12db, +13db, +14db, +15db

Με αυτόν τον τρόπο το κάθε άτομο πλέον έχει **1** αρχείο εκπαίδευσης και **67** αρχεία ελέγχου.

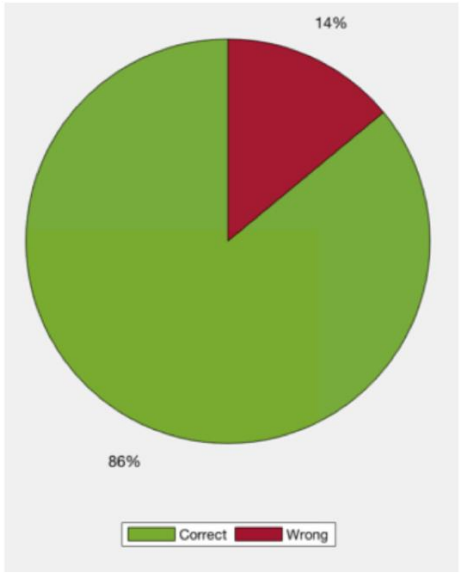
Οι μετασχηματισμοί σήματος[23] που εφαρμόσθηκαν είναι:

- Cross Correlation
- Discrete Laplacian Transform
- Envelope
- Fast Fourier Transform
- Hilbert Transform
- MFCC
- Wavelets Decomposition
- Wavelets Transform

[23] Στο περιβάλλον MATLAB ορισμένοι αλγόριθμοι αναφέρονται ως μετασχηματισμοί σήματος αν και στην πραγματικότητα δεν είναι, όπως ο ‘Envelope’

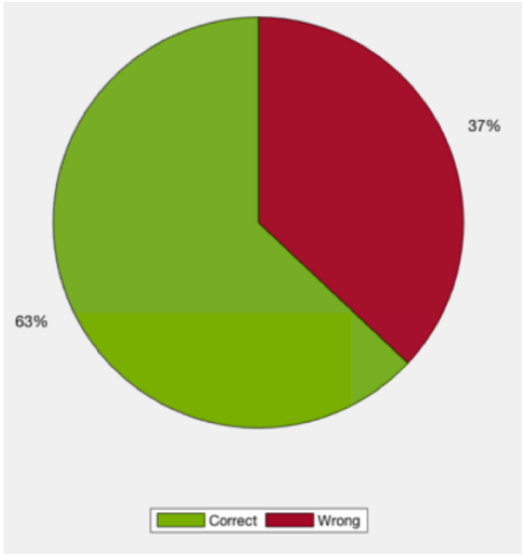


Cross-Correlation

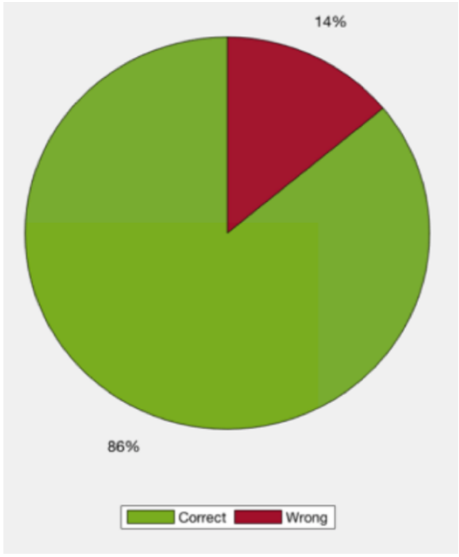


Discrete Laplacian Transform

Εικόνα 43: Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης φωνής (1)

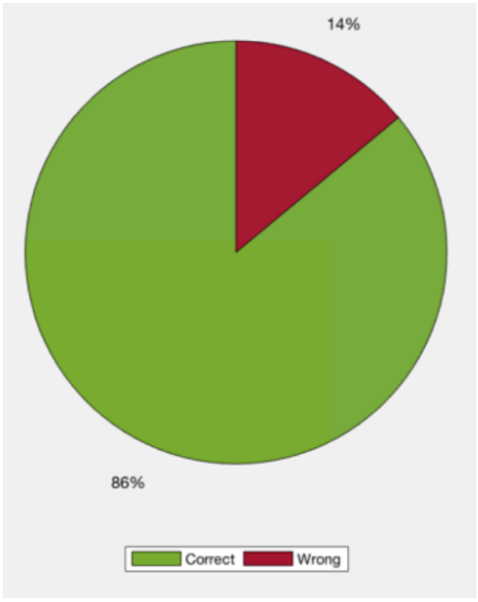


Envelope

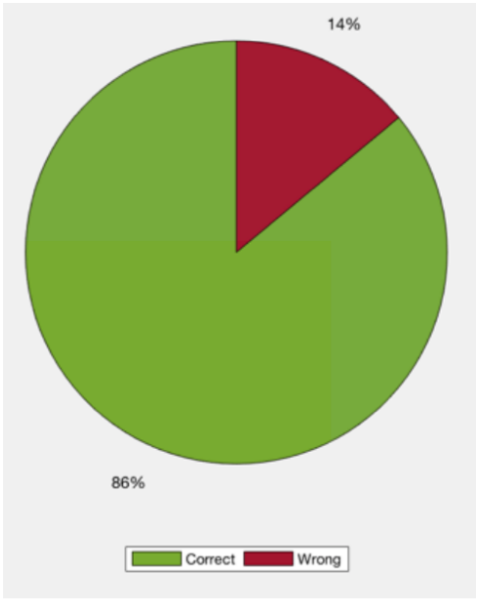


Fast Fourier Transform

Εικόνα 44: *Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης φωνής (2)*

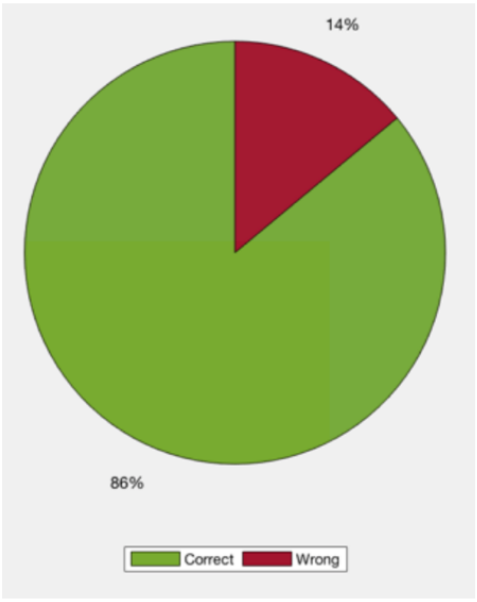


Hilbert Transform

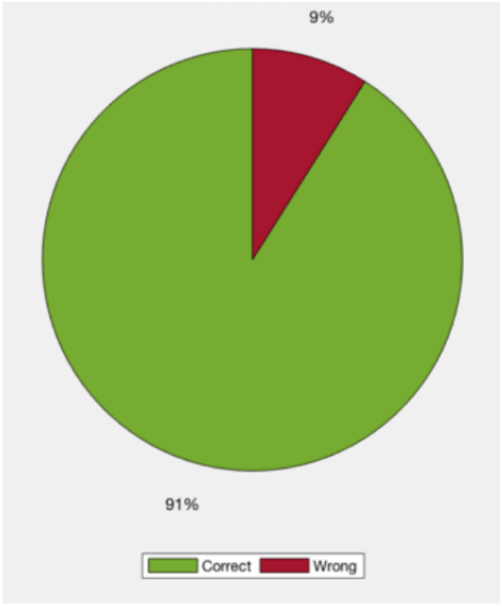


Wavelets Transform

Εικόνα 45: *Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης φωνής (3)*



Wavelets Decomposition



MFCC Transform

Εικόνα 46: *Αποτελέσματα πειραματικού μέρους 1 αναγνώρισης φωνής (4)*

Από τα παραπάνω διαγράμματα προκύπτει ότι μεγαλύτερα ποσοστά ακρίβειας επιτυγχάνονται με την χρήση του μετασχηματισμού MFCC, 91%. Παρακάτω παρουσιάζεται ένας πίνακας με το υπολογιστικό κόστος (χρόνος) για κάθε αρχείο ήχου που εκπαιδεύεται και για κάθε αρχείο ήχου που ελέγχεται από το σύστημα.

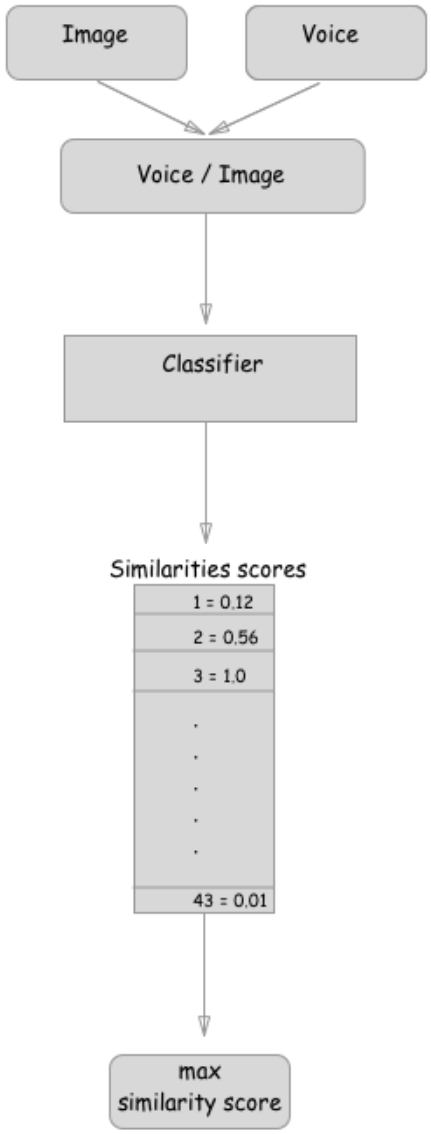
Transform	training time (1 .wav ανά άτομο)	testing time (1 .wav)
Cross Correlation	6.69 sec	0.87 sec
Discrete Laplacian Transform	4.47 sec	0.41 sec
Envelope	4.88 sec	0.41 sec
Fast Fourier Transform	4.45 sec	0.91 sec
Hilbert Transform	6.66 sec	0.90 sec
MFCC	4.20 sec	0.10 sec
Wavelets Decompotition	4.90 sec	0.46 sec
Wavelets Transform	4.74 sec	0.39 sec

Εικόνα 47: *Χρόνοι εκπαίδευσης και ελέγχου (μετασχηματισμοί σήματος)*

5.2.3 Multimodal

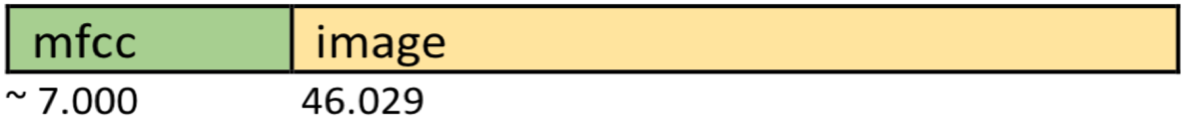
5.2.3.1 Μέθοδος 1

Στην μέθοδο 1, το multimodal σύστημα υλοποιείται ως εξής:



Εικόνα 48: *Multimodal μέθοδος 1*

Γίνεται χρήση των διανυσμάτων (vectors) που προκύπτουν από τον μετασχηματισμό MFCC και των διανυσμάτων που προκύπτουν από τον αλγόριθμο PCA με 5 μεταβλητές εικόνες εκπαίδευσης κάνοντας χρήση το πρώτο ιδιοδιάνυσμα. Το εκάστοτε διάνυσμα φωνής ενώνεται με το εκάστοτε διάνυσμα προσώπου και έτσι πλέον, έχουμε ενιαία διανύσματα εκπαίδευσης και ελέγχου με την παρακάτω μορφή.

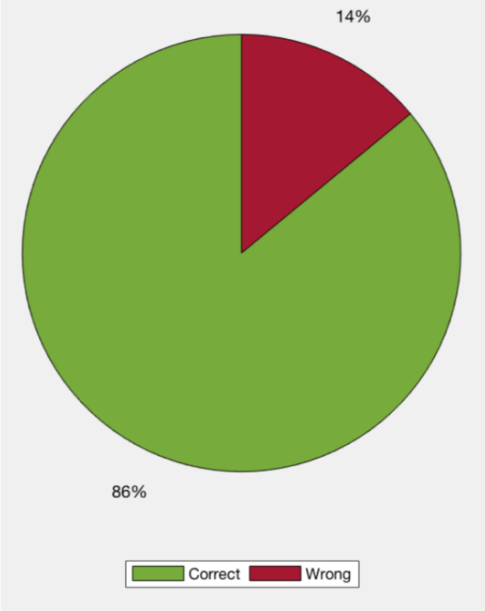


Το διάνυσμα της φωνής έχει μήκος περίπου 7.000 στήλες και το διάνυσμα της εικόνας 46.029 στήλες (~7.000 + 46.029 = ~53.029 στήλες συνολικά). Πλέον το κάθε άτομο έχει τέτοιας μορφής διανύσματα εκπαίδευσης και διανύσματα ελέγχου τα οποία περνούν από τον ταξινομητή ομοιότητας όπου προκύπτουν σκορ ομοιότητας. Από αυτά παίρνεται το μεγαλύτερο και προκύπτει με αυτόν τον τρόπο η εκάστοτε ταξινόμηση.

5.2.3.1.1 Πειραματικό μέρος 1

Στο πειραματικό μέρος 1, το σύστημα εκπαιδεύεται με 5 μετωπικές εικόνες προσώπου και ελέγχεται με μετωπικές εικόνες επίσης όσον αφορά το τμήμα της εικόνας του ενιαίου διανύσματος. Το τμήμα του ήχου είναι το αποτέλεσμα του μετασχηματισμού MFCC όπως έχει προαναφερθεί. Συνολικά γίνεται έλεγχος σε 2.150 x 67 = **144.050** διανύσματα.

Αποτελέσματα



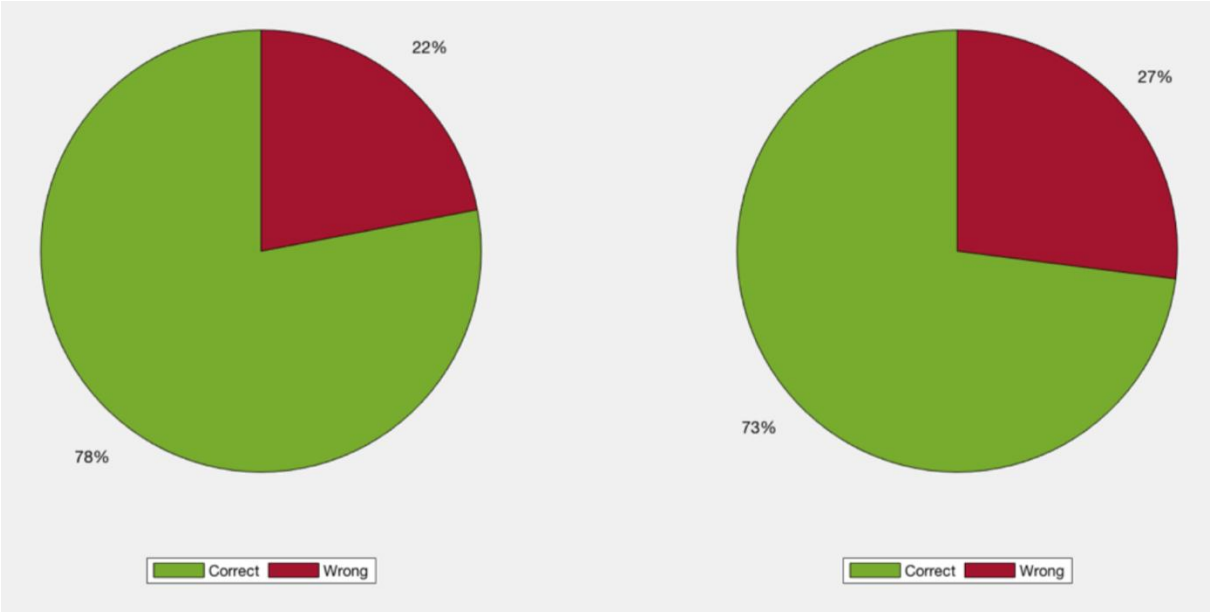
Εικόνα 49: *Αποτελέσματα πειραματικού μέρους 1 (multimodal, μέθοδος 1)*

Από το παραπάνω διάγραμμα προκύπτει ακρίβεια σωστών αποτελεσμάτων 86% και 14% λανθασμένων.

5.2.3.1.2 Πειραματικό μέρος 2

Στο πειραματικό μέρος 2, το σύστημα εκπαιδεύεται με 5 μεταβλητές εικόνες προσώπου και ελέγχεται με μεταβλητές εικόνες επίσης όσον αφορά το τμήμα της εικόνας του ενιαίου διανύσματος. Το τμήμα του ήχου είναι το αποτέλεσμα του μετασχηματισμού MFCC όπως έχει προαναφερθεί. Συνολικά γίνεται έλεγχος σε 37.377 x 67 = **2.504.259** διανύσματα.

Αποτελέσματα



Εικόνα 50: *Αποτελέσματα πειραματικού μέρους 2 (multimodal, μέθοδος 1)*

Στο πρώτο διάγραμμα εμφανίζονται αποτελέσματα ακρίβειας σωστών και λανθασμένων ταυτοποιήσεων 78% και 22% αντίστοιχα, από διανύσματα ελέγχου που προέρχονται από το ίδιο βίντεο (video 1) με τα διανύσματα εκπαίδευσης.

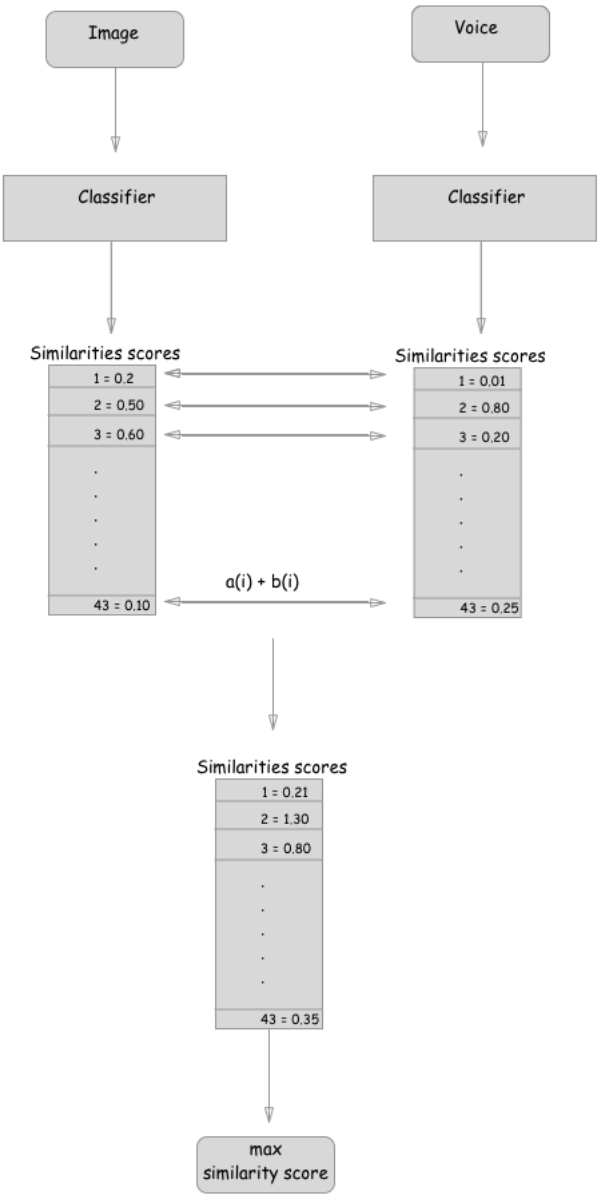
Στο δεύτερο διάγραμμα εμφανίζονται αποτελέσματα ακρίβειας σωστών και λανθασμένων ταυτοποιήσεων 73% και 27% αντίστοιχα, από διανύσματα ελέγχου που προέρχονται από διαφορετικό βίντεο (video 2) με τα διανύσματα εκπαίδευσης. Παρακάτω παρουσιάζεται ένας πίνακας με το υπολογιστικό κόστος (χρόνος) για κάθε διάνυσμα που εκπαιδεύεται και για κάθε διάνυσμα που ελέγχεται από το σύστημα.

	training time	testing time
Frontal	6.20 sec	10.70 sec
Variability	6.50 sec	10.10 sec

Εικόνα 51: *Χρόνοι εκπαίδευσης και ελέγχου (multimodal, μέθοδος 1)*

5.2.3.2 Μέθοδος 2

Στην μέθοδο 1, το multimodal σύστημα υλοποιείται ως εξής:

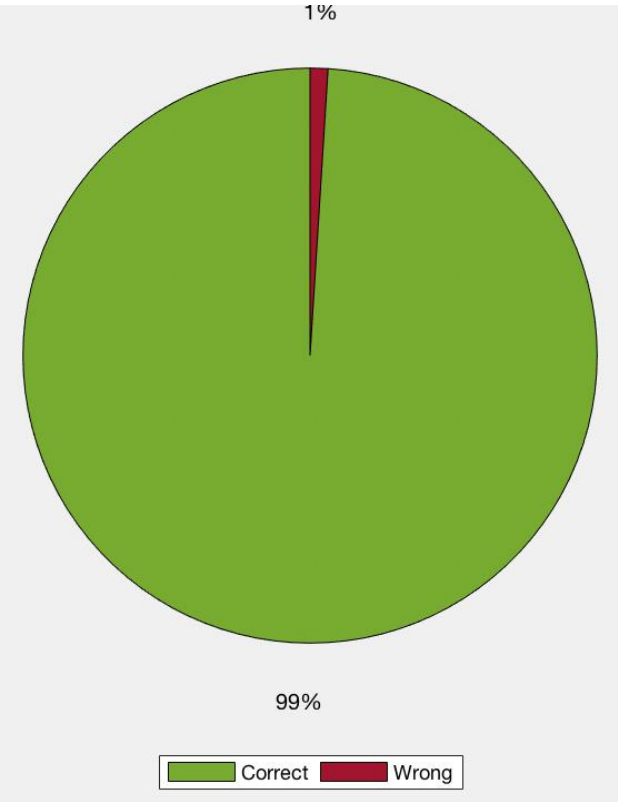


Γίνεται χρήση των διανυσμάτων (vectors) που προκύπτουν από τον μετασχηματισμό MFCC και των διανυσμάτων που προκύπτουν από τον αλγόριθμο PCA κάνοντας χρήση το πρώτο ιδιοδιάνυσμα όπως και στην μέθοδο 1. Η διαφορά είναι ότι η κάθε τυπικότητα (modality) φωνή και εικόνα δηλαδή, έχει την δική της ροή μέσα στο σύστημα. Στην φωνή

υπάρχουν τα αντίστοιχα διανύσματα εκπαίδευσης και ελέγχου τα οποία περνάνε από τον ταξινομητή ομοιότητας και προκύπτουν τα αντίστοιχα σκορ. Με τον ίδιο ακριβώς τρόπο προκύπτουν και τα σκορ ομοιότητας για την εικόνα. Ύστερα, τα σκορ φωνής και εικόνας αθροίζονται και προκύπτουν τα νέα σκορ. Από αυτά παίρνεται το μεγαλύτερο και με αυτόν τον τρόπο πραγματοποιείται η εκάστοτε ταξινόμηση.

5.2.3.2.1 Πειραματικό μέρος 1

Στο πειραματικό μέρος 1, το σύστημα εκπαιδεύεται με 5 μετωπικές εικόνες προσώπου και ελέγχεται με μετωπικές εικόνες επίσης. Οι φωνές είναι το αποτέλεσμα του μετασχηματισμού MFCC όπως έχει προαναφερθεί. Συνολικά γίνεται έλεγχος σε 2.150 x 67 = **144.050** διανύσματα όπως έγινε και στο πειραματικό μέρος 1 της μεθόδου 1.



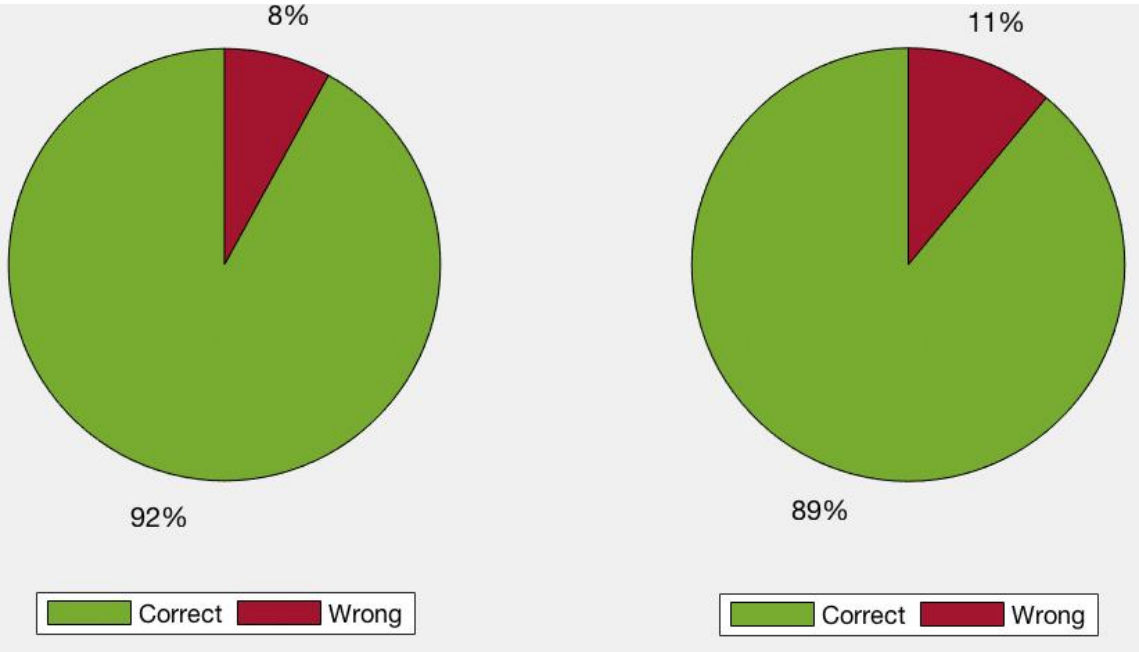
Εικόνα 52: *Αποτελέσματα πειραματικού μέρους 1 (multimodal, μέθοδος 2)*

Από το παραπάνω διάγραμμα προκύπτει ακρίβεια αποτελεσμάτων της τάξης 99%.

5.2.3.2.2 Πειραματικό μέρος 2

Στο πειραματικό μέρος 2, το σύστημα εκπαιδεύεται με 5 μεταβλητές εικόνες προσώπου και ελέγχεται με μεταβλητές εικόνες επίσης. Οι φωνές είναι το αποτέλεσμα του μετασχηματισμού MFCC όπως έχει προαναφερθεί. Συνολικά γίνεται έλεγχος σε 37.377 x 67 = **2.504.259** διανύσματα όπως έγινε και στο πειραματικό μέρος 1 της μεθόδου 1.

Αποτελέσματα



Εικόνα 53: *Αποτελέσματα πειραματικού μέρους 2 (multimodal, μέθοδος 2)*

Στο πρώτο διάγραμμα εμφανίζονται αποτελέσματα ακρίβειας σωστών και λανθασμένων ταυτοποιήσεων 92% και 8% αντίστοιχα, από διανύσματα ελέγχου που προέρχονται από το ίδιο βίντεο (video 1) με τα διανύσματα εκπαίδευσης.

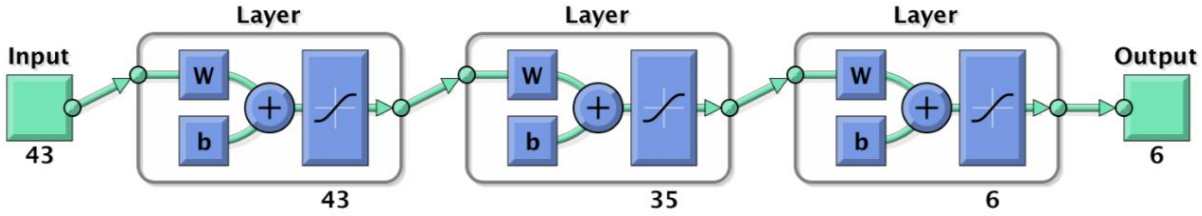
Στο δεύτερο διάγραμμα εμφανίζονται αποτελέσματα ακρίβειας σωστών και λανθασμένων ταυτοποιήσεων 89% και 11% αντίστοιχα, από διανύσματα ελέγχου που προέρχονται από διαφορετικό βίντεο (video 2) με τα διανύσματα εκπαίδευσης.

5.2.3.3 Μέθοδος 3

5.2.3.3.1 Πειραματικό μέρος 1

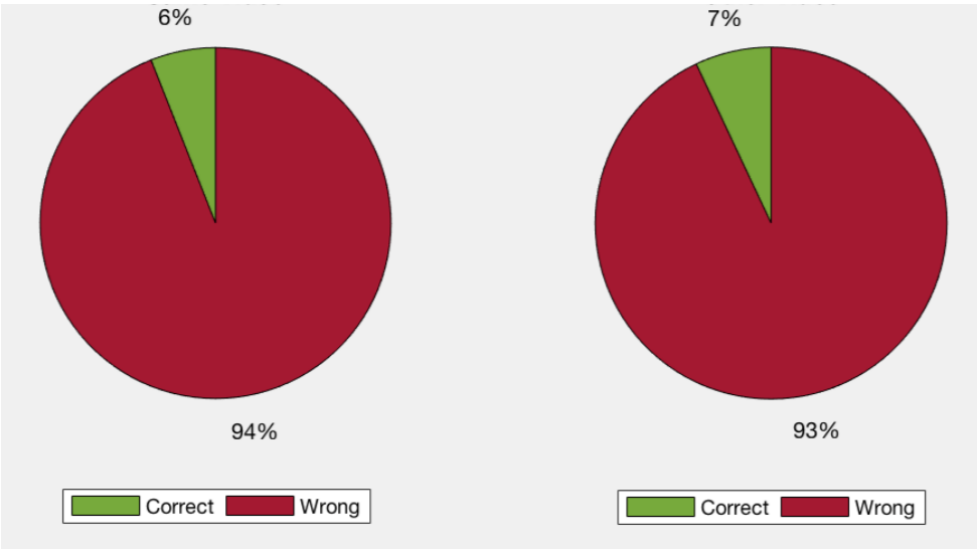
Η μέθοδος 3, είναι η υλοποίηση ενός backpropagation νευρωνικού δικτύου. Η μέθοδος αυτή υλοποιήθηκε καθαρά για λόγους εκπαίδευσης και εξοικείωσης με τα νευρωνικά δίκτυα και όχι για να επιφέρει υψηλά ποσοστά ακρίβειας καθότι γνωρίζαμε εκ των προτέρων ότι η συγκεκριμένη μέθοδος δεν ενδείκνυται για την παρούσα μελέτη και θα επιφέρει χαμηλά ποσοστά ακρίβειας.

Το κάθε άτομο, έχει 68 διανύσματα εκπαίδευσης τα οποία προκύπτουν από τα αποτελέσματα του ταξινομητή ομοιότητας της μεθόδου 1. Το κάθε διάνυσμα αποτελείται από τα σκορ ομοιότητας μεταξύ του ιδίου και των άλλων διανυσμάτων εκπαίδευσης. Παρακάτω παρουσιάζεται η αρχιτεκτονική του νευρωνικού δικτύου.



Εικόνα 54: *Αρχιτεκτονική backpropagation νευρωνικού δικτύου*

Αποτελέσματα



Εικόνα 55: *Αποτελέσματα πειραματικού μέρους 1 (multimodal, μέθοδος 3)*

Στο πρώτο γράφημα έχουμε ακρίβεια σωστών και λανθασμένων ταυτοποιήσεων 6% και 94% αντίστοιχα, από διανύσματα ελέγχου που προέρχονται από το ίδιο βίντεο (video 1) με τα διανύσματα εκπαίδευσης.

Στο δεύτερο γράφημα έχουμε ακρίβεια σωστών και λανθασμένων ταυτοποιήσεων 7% και 93% αντίστοιχα, από διανύσματα ελέγχου που προέρχονται από το ίδιο βίντεο (video 1) με τα διανύσματα εκπαίδευσης.

5.2.3.4 Σύγκριση μεθόδων

Παρακάτω παρουσιάζεται συγκριτικός πίνακας με τα αποτελέσματα των τριών μεθόδων που υλοποιήθηκαν και αναφέρθηκαν παραπάνω.

Πίνακας 5: *Συγκριτικός πίνακας μεθόδων*

	Μέθοδος 1	**Μέθοδος 2**	**Μέθοδος 3**
Μετωπικές εικόνες	86%	99%	-
Μεταβλητές εικόνες	78% 73%	92% 89%	6% 7%

Από τον παραπάνω πίνακα βλέπουμε ότι η μέθοδος 2 επιφέρει τα καλύτερα αποτελέσματα, άνω του 10% σε σχέση με την μέθοδο 1.

6.1 Συμπεράσματα

Ένα από τα βασικά συμπεράσματα της παρούσας πτυχιακής εργασίας αφορούν την μέθοδο PCA. Η συγκεκριμένη μέθοδος προσφέρει μείωση υπολογιστικού κόστους καθώς μειώνει τον αριθμό εικόνων κάτι το οποίο είναι χρήσιμο καθότι εφαρμογές ταυτοποίησης ατόμου ενδεχομένως να υλοποιούνται σε ενσωματωμένα συστήματα περιορισμένων υπολογιστικών δυνατοτήτων. Μέσω των πειραμάτων που πραγματοποιήθηκαν, έγινε εμφανές ότι η μέθοδος PCA, καθιστά δυνατή την ταυτοποίηση προσώπου σε πολύ υψηλά ποσοστά ακρίβειας, όσον αφορά εικόνες εμπρόσθιας όψης. Εάν οι εικόνες εκτός από εμπρόσθιας όψης, είναι προφίλ ή και άλλων ενδιάμεσων γωνιών κλίσης, έγινε επίσης εμφανές ότι η μέθοδος δεν ανταποκρίνεται σε υψηλά ποσοστά, εν αντιθέσει, η ακρίβεια που προσφέρει είναι σε σχεδόν μηδενικά ποσοστά σε ορισμένες περιπτώσεις. Καταλήγουμε ότι ένα σύστημα ταυτοποίησης ατόμων μέσω εικόνων που έχει βασιστεί εξ ολοκλήρου στην μέθοδο PCA, μπορεί να έχει επιφέρει σωστά αποτελέσματα εάν: έχει εκπαιδευτεί και ελέγχεται αποκλειστικά με μετωπικές εικόνες. Σε συνθήκες ‘wild’ ένα τέτοιο σύστημα δεν θα μπορέσει να λειτουργήσει.

Επίσης στην παρούσα πτυχιακή εργασία, μελετήθηκαν μετασχηματισμοί σήματος, οι οποίοι χρησιμοποιούνται για την επίτευξη αναγνώρισης φωνής. Μέσω πειραμάτων, καταλήξαμε ότι ο μετασχηματισμός που προσφέρει τα καλύτερα δυνατά αποτελέσματα είναι ο MFCC μετασχηματισμός. Στην συνέχεια, δείξαμε ότι τα υψηλά αποτελέσματα αναγνώρισης φωνής με τον συγκεκριμένο μετασχηματισμό δεν επηρεάζονται από την παρουσία θορύβου ή ορισμένων αλλοιώσεων την φωνής. Καταλήξαμε, ότι ένα σύστημα ταυτοποίησης ατόμου μέσω φωνής, μπορεί να βασιστεί στον μετασχηματισμό MFCC και στην υλοποίηση του συστήματος που αναπτύξαμε, εφόσον η μέθοδος αναγνώρισης φωνής είναι εξαρτημένη από το κείμενο.

Τέλος, μελετήθηκαν και πραγματοποιήθηκαν πειράματα ως προς ορισμένες από τις προσεγγίσεις ένωσης διαφορετικών modalities σε ένα ενιαίο multimodal σύστημα. Μέσω της μελέτης αυτής, καταλήξαμε ότι η προσέγγιση κατά την οποία τα modalities εισέρχονται στον ταξινομητή ξεχωριστά το ένα από το άλλο και στην συνέχεια προκύπτει ο μέσος όρος ομοιότητας αυτών, επιφέρει τα ακριβέστερα αποτελέσματα

6.2 Μελλοντική Εργασία και Επεκτάσεις

Η μελλοντική εργασία που πιθανός να πραγματοποιηθεί αφορά το ίδιο βιομετρικό σύστημα ταυτοποίησης, με την διαφορά του ότι θα εφαρμοσθεί επιλογή χαρακτηριστικών

(feature selection) στην εικόνα και την φωνή και το αποτέλεσμα αυτής θα εισάγεται στον αντίστοιχο ταξινομητή ομοιότητας.

Βιβλιογραφία

[1] E.-A. Cabanis, J.-Y. Le Gall, R. Ardaillou, and Groupe de travail issu de la Commission I (Biologie), "[Personal identification with biometric and genetic methods].," *Bull. Acad. Natl. Med.*, vol. 191, no. 8, pp. 1779–82, Nov. 2007.

[2] C. Sanderson and B. C. Lovell, "Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference," Springer, Berlin, Heidelberg, 2009, pp. 199–208.

[3] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–10.

[4] Χ. Ι. ΓΕΩΡΓΙΟΥ, "Αναγνώριση ταυτότητας προσώπου από βιντεοσκοπήσεις," Bachelors's thesis, Πανεπιστήμιο Πατρών, 2014.

[5] J. Shi, N. Zhang, and X. Liu, "A novel fractional wavelet transform and its applications," *Sci. China Inf. Sci.*, vol. 55, no. 6, pp. 1270–1279, Jun. 2012.

[6] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching * †," CRC Press, 1999.

[7] M. Kass, A. Witkin, and D. Terzopoulos, Snakes: active contour models, in Proc. of 1st Int Conf. on Computer Vision, London, 1987.

[8] R. C. Gonzalez and R. E. Woods *Digital image processing*, Prentice Hall, 2008

[9] T. Sakai, M. Nagao, and T. Kanade, Computer analysis and classification of photographs of human faces, in Proc. First USA—Japan Computer Conference, 1972, p. 2.7.

[10] D. Marr and E. Hildreth, "Theory of edge detection.," *Proc. R. Soc. London. Ser. B, Biol. Sci.*, vol. 207, no. 1167, pp. 187–217, Feb. 1980.

[11] V. Govindaraju, "Locating human faces in photographs," *Int. J. Comput. Vis.*, vol. 19, no. 2, pp. 129–146, Aug. 1996.

[12] H. Musoff and P. Zarchan, *Fundamentals of Kalman Filtering: A Practical Approach, Second Edition*. Reston ,VA: American Institute of Aeronautics and Astronautics, 2005.

[13] D. Reisfeld, H. Wolfson, and Y. Yeshurun, "Context Free Attentional Operators: the

Generalized Symmetry Transform," 1995.

[14] C. Wong, D. Kortenkamp, and M. Speich, "A mobile robot that recognizes people," *Proc. 7th IEEE Int. Conf. Tools with Artif. Intell.*, 1995.

[15] L. C. De Silva, K. Aizawa, and M. Hatori, "Detection and Tracking of Facial Features by Using a Facial Feature Model and Deformable Circular Template," *Ieice*, vol. E78–D, no. 9, pp. 1195–1207, 1995.

[16] J. Missimer *et al.*, "On two methods of statistical image analysis.," *Hum. Brain Mapp.*, vol. 8, no. 4, pp. 245–58, 1999.

[17] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[18] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," 2001.

[19] Κ. Διαμαντάρας, *ΤΕΧΝΗΤΑ ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ*, Κλειδάριθμος, 2007.

[20] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 1, pp. 23–38, 1998.

[21] S. Haykin, *Neural Networks and Learning Machines Third Edition*, Prentice Hall, 2009.

[22] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Pearson, 2006.

[23] J. Colmenarez and T. S. Huang, "Maximum likelihood face detection," *Autom. Face Gesture Recognition, 1996., Proc. Second Int. Conf.*, pp. 307–311, 1996.

[24] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22. Institute of Mathematical Statistics, pp. 79–86, 1951

[25] Γ. Παπαδουράκης and Κ. Μαριάς, *Η μέθοδος PCA-Ανάλυση Κύριων Συνιστωσών*, [On-line]. Available: https://eclass.teicrete.gr/modules/document/index.php?course=TP223&openDir=/565efd25MNdX/565efe77JftD [November 18, 2018].

[26] Β. Τσιλιγκίρδη, "Σύγχρονες Τεχνικές στις Διεπαφές Ανθρώπινου Εγκεφάλου - Υπολογιστή", Bachelor's thesis, ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ, 2011.

[27] D. Bařina, *GABOR WAVELETS IN IMAGE PROCESSING*, Proceedings of the 17th Conference STUDENT EEICT, Brno, CZ, 2011

[28] T. S. Lee, *Image Representation Using 2D Gabor Wavelets*, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 18, NO. 10, OCTOBER, 1996

[29] L. Shen and L. Bai, “A review on Gabor wavelets for face recognition,” *Pattern Anal. Appl.*, vol. 9, no. 2–3, pp. 273–292, 2006.

[30] A.-A. Bhuiyan and C. H. Liu, “On Face Recognition using Gabor Filters,” 2007.

[31] V. Bevilacqua, L. Cariello, G. Carro, D. Daleno, and G. Mastronardi, “A face recognition system based on Pseudo 2D HMM applied to neural network coefficients,” vol. 12, pp. 615–621, 2008.

[32] L. A. Zadeh, “Fuzzy logic = computing with words,” *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103–111, May 1996.

[33] D. Bhattacharjee, D. K. Basu, M. Nasipuri, and M. Kundu, “Human face recognition using fuzzy multilayer perceptron,” *Soft Comput.*, vol. 14, no. 6, pp. 559–570, 2010.

[34] G. G. Gordon, “Face recognition based on depth maps and surface curvature,” *SPIE 1570, Geom. Methods Comput. Vis.*, vol. 1570, no. 1991, pp. 234–247, 1991.

[35] “Cross-correlation - MATLAB xcorr.” [Online]. Available: https://www.mathworks.com/help/signal/ref/xcorr.html. [Accessed: 07-Sep-2018].

[36] G. Padourakis, *ΒΑΣΙΚΕΣ ΑΡΧΕΣ ΤΗΣ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΤΥΠΩΝ*, [On-line]. Available: https://eclass.teicrete.gr/modules/document/file.php/TP223/%CE%98%CE%B5%CF%89%CF%81%CE%AF%CE%B1%20%28Lectures%29/%CE%91%CF%83%CE%BA%CE%AE%CF%83%CE%B5%CE%B9%CF%82%20%28Assignments%29/pattern1.pdf [November 18, 2018].

[37] A. Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, 2005.

[38] M. M. Fakhir, W. L. Woo, and S. S. Dlay, “Face Recognition Based on Features Measurement Technique,” *2014 Eur. Model. Symp.*, pp. 158–162, 2014.

[39] P. Karczmarek, A. Kiersztyn, W. Pedrycz, and M. Dolecki, “An application of chain code-based local descriptor and its extension to face recognition,” *Pattern Recognit.*, vol. 65, no. December 2016, pp. 26–34, 2017.

[40] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 815--823, 2015.

[41] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition,” *Procedings Br. Mach. Vis. Conf. 2015*, no. Section 3, p. 41.1-41.12, 2015.

[42] W. Xie and A. Zisserman, “Multicolumn Networks for Face Recognition,” pp. 1–12, 2018.

[43] J. Zhang, “Realization and Improvement Algorithm of GMM - UBM Model in Voiceprint Recognition,” *2018 Chinese Control Decis. Conf.*, pp. 2989–2992, 2018.

[44] X. Zhao, Y. Wang, and D. Wang, “Robust speaker identification in noisy and reverberant conditions,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 22, no. 4, pp. 3997–4001, 2014.

[45] J. H. Cernock, “ANALYSIS OF DNN APPROACHES TO SPEAKER IDENTIFICATION Pavel Mat e ˇ ej Novotn ´ Franti s ˇ ek Gr ´ Brno University of Technology , Speech @ FIT and IT4I Center of Excellence , Brno , Czech Republic { matejkap , glembek , inotovny , grezl , burget , iplchot,” pp. 5100–5104, 2016.

[46] A. Nagraniy, J. S. Chungy, and A. Zisserman, “VoxCeleb: A large-scale speaker identification dataset,” *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2017–Augus, pp. 2616–2620, 2017.

[47] E. Mitsianis and T. Giannakopoulos, “Speaker Verification based on extraction of Deep Features,” 2018.

[48] F. Pala, R. Satta, G. Fumera, and F. Roli, “Multimodal Person Reidentification Using RGB-D Cameras,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 4, pp. 788–799, 2016.

[49] S. A. Angadi and S. M. Hatture, “Biometric Person Identification System: A Multimodal Approach Employing Spectral Graph Characteristics of Hand Geometry and Palmprint,” *Int. J. Intell. Syst. Appl.*, vol. 8, no. 3, pp. 48–58, 2016.

[50] B. C. Kovoor, “Performance evaluation of unimodal , bimodal and trimodal biometric identification system,” vol. 7, no. 6, pp. 1–8, 2017.

[51] S. Soleymani, A. Torfi, J. Dawson, and N. M. Nasrabadi, “Generalized Bilinear Deep Convolutional Neural Networks for Multimodal Biometric Identification,” no. ii, 2018.

[52] S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, and N. M. Nasrabadi, “Multi-Level Feature Abstraction from Convolutional Neural Networks for Multimodal Biometric Identification,” no. i, 2018.

[53] E. Kita, Y. Zuo, F. Saito, and X. Feng, “Personal Identification with Face and Voice Features Extracted through Kinect Sensor,” *IEEE Int. Conf. Data Min. Work. ICDMW*, pp. 545–551, 2017.

[54] J. Ren *et al.*, “Look, Listen and Learn - A Multimodal LSTM for Speaker Identification,” pp. 3581–3587, 2016.

[55] D. Kauffman, M. Williams, and C. Washington, “Multimodal Speaker Identification in Legislative Discourse,” 2018.

[56] E. Santana, G. T. Cinar, and J. C. Principe, “Parallel flow in Deep Predictive Coding Networks,” *Proc. Int. Jt. Conf. Neural Networks*, vol. 2015–Septe, 2015.