ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΤΟΠΟΓΡΑΦΙΑΣ & ΓΕΩΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΜΣ: ΓΕΩΧΩΡΙΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΕΘΟΔΟΙ ΑΝΙΧΝΕΥΣΗΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗΣ ΠΡΟΣΩΠΩΝ

ΠΟΛΛΟ ΡΕΝΤΙ

ΑΜ: 1615

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

ΕΠΙΒΛΕΠΟΝΤΕΣ:

ΓΡΑΜΜΑΤΙΚΟΠΟΥΛΟΣ Λ.

ΠΕΤΣΑ Ε.

ΚΑΛΗΣΠΕΡΑΚΗΣ Η.

ΑΘΗΝΑ, ΣΕΠΤΕΜΒΡΙΟΣ 2020

Στο θείο μου Σωκράτη

Στη θεία μου Κυράτσα

Στον Σολ

Frustra fit per plura quod potest fieri per pauciora

William of Ockham

«Είναι μάταιο να κάνεις κάτι με πολλά όταν αυτό μπορεί να γίνει με λίγα»

Ευχαριστίες

Με την ευκαιρία της ολοκλήρωσης της μεταπτυχιακής μου εργασίας θα ήθελα να ευχαριστήσω τους ανθρώπους που συνέβαλαν στο ταξίδι μου στις γεωχωρικές τεχνολογίες. Θα ήθελα πρωτίστως να ευχαριστήσω τους καθηγητές μου κ. Λάζαρο Γραμματικόπουλο και κ. Έλλη Πέτσα για την υπομονή, επιμονή και όλη τη βοήθεια που μου προσέφεραν καθ' όλη την διάρκεια των μεταπτυχιακών μου σποδών. Δεν θα μπορούσα να ξεχάσω τον συμφοιτητή και φίλο Παναγιώτη Νάκη, ο οποίος με την στήριξη και την συμπαράστασή του συνέβαλε σημαντικά στην ολοκλήρωση των μεταπτυχιακών μου σπουδών. Βέβαια, το μεγαλύτερο ευχαριστώ το οφείλω στην οικογένειά μου για όλα όσα μου έχει προσφέρει όλα αυτά τα χρόνια.

Περίληψη

Η ανίχνευση αντικειμένων αποτελεί μια διαδικασία που σχετίζεται με την όραση υπολογιστών και την επεξεργασία εικόνας. Η όραση υπολογιστών και η επεξεργασία εικόνας, ως μια υποκατηγορία της επεξεργασίας ψηφιακού σήματος, υποδηλώνει έναν αριθμό μεθόδων για την ενίσχυση, την τροποποίηση και την ταξινόμηση των εικόνων με διάφορους τρόπους. Τα τελευταία χρόνια, η ζήτηση για την επεξεργασία εικόνας αυξάνεται εκθετικά λόγω της μεγάλης ποικιλίας εφαρμογών σε διάφορους τομείς. Μια από τις πιο σημαντικές εφαρμογές της είναι στον τομέα της ανίχνευσης προσώπων, η οποία με την σειρά της αποτελεί ένα ισχυρό εργαλείο για την αναγνώριση, την παρακολούθηση προσώπου και άλλες σημαντικές εφαρμογές της όρασης υπολογιστών. Το πρόσωπο είναι το πιο σημαντικό μέρος του ανθρώπινου σώματος για την αναγνώριση ενός ατόμου. Ειδικά τα τελευταία χρόνια, είναι το πιο «πολύ-φωτογραφημένο» αντικείμενο στον κόσμο.

Οι μέθοδοι ανίχνευσης και αναγνώρισης προσώπων ως μια υποκατηγορία της ανίχνευσης αντικειμένων μπορούν να ταξινομηθούν σε δυο κατηγορίες. Πρόκειται για τις προσεγγίσεις με βάση τα χαρακτηριστικά (Feature Based Approach) και με βάση την εμφάνιση ή την εικόνα (Image Based Approach). Οι μέθοδοι της πρώτης κατηγορίας έχουν στόχο την εξαγωγή χαρακτηριστικών μιας δοθείσας εικόνας και στην συνέχεια την αντιστοίχισή της με μοντέλα που δημιουργούνται με βάση την γνώση του ερευνητή για τα χαρακτηριστικά του προσώπου. Αντιθέτως, η προσέγγιση με βάση την εμφάνιση προσπαθεί να πραγματοποιήσει την καλύτερη αντιστοίχιση, ή αλλιώς ταξινόμηση, μεταξύ των εικόνων εκπαίδευσης (training images) με τις αντίστοιχες κατηγορίες προσώπων . Η τελευταία κατηγορία αναφέρεται σε μεθόδους μηχανικής μάθησης καθώς ένα σύστημα εκπαιδεύεται μέσω παραδειγμάτων για να επιτευχθεί η τελική ταξινόμηση. Η ανίχνευση προσώπου θεωρείται ως ένα πρόβλημα ταξινόμησης, καθώς στόχος του συστήματος είναι να αναγνωρίσει πρότυπα σε μια εικόνα ώστε να τα ταξινομήσει στην αντίστοιχη κατηγορία. Το ανθρώπινο πρόσωπο όμως είναι ένα δυναμικό αντικείμενο με υψηλό βαθμό μεταβλητότητας στην εμφάνισή του, γεγονός που καθιστά την ανίχνευση και αναγνώριση προσώπου ένα δύσκολο πρόβλημα για την επιστήμη της όρασης υπολογιστών.

Στο πλαίσιο αυτής της διπλωματικής εργασίας υλοποιείται ένα σύστημα αναγνώρισης προσώπων με δυο διαφορετικές μεθόδους. Πρόκειται για προσεγγίσεις βάσει της εμφάνισης που εμπίπτουν στις τεχνικές της μηχανικής μάθησης. Η ανίχνευση προσώπου γίνεται σύμφωνα με τον αλγόριθμο των Viola & Jones, ο οποίος αποδεικνύεται εξαιρετικός ανιχνευτής προσώπου. Οι μέθοδοι αναγνώρισης EigenFace & Bag of Visual Features υλοποιούνται με την προβολή των εικόνων προς αναγνώριση στον χώρο των χαρακτηριστικών και βαρών με σκοπό την τελική ταξινόμηση και την μείωση του υπολογιστικού κόστους. Η υλοποίηση των συναρτήσεων έγινε στο περιβάλλον του MATLAB.

Λέξεις Κλειδιά: Ανίχνευση, Εντοπισμός και Αναγνώριση Προσώπου, Αλγόριθμος Viola & Jones, Eigen Faces, Principal Component Analysis, Bag of Words, Bag of Visual Features

University of West Attica
School of Engineering
Department of Surveying & Geoinformatics
Master of Science in Geospatial Technologies

Master Thesis

Face Detection and Recognition Methods

Pollo Redi

September 2020

Abstract

Object detection is a process related to computer vision and image processing. Computer vision and image processing, as a subcategory of digital signal processing, suggest a number of methods for enhancing, modifying, and classifying images in a variety of ways. In recent years, the demand for image processing is growing exponentially due to the wide variety of applications in various fields. One of its most important applications take place in the field of face detection, which in turn is a powerful tool for identification, face tracking and other important applications of computer vision. The face is the most important part of the human body for identifying a person. Especially in recent years, it is the most photographed object in this world. Methods of detecting and identifying persons as a subcategory of object detection can be classified into two categories. These are feature-based and image-based approaches. The methods of the first category aim at extracting the characteristics of a given image and then matching it with models created based on the researcher's knowledge of facial features and types. On the contrary, the appearance-based approach strives to achieve the best matching classification between training images with the corresponding categories. The last category lies in machine learning methods as a system is trained through examples to achieve the final classification. Face detection is considered a classification problem as the system aims to identify patterns in an image in order to classify them into the appropriate categories. The human face is a dynamic object and has a high degree of variability in its appearance, which makes face detection and recognition a difficult problem for the science of computer vision. In the context of this dissertation, a system of face recognition is implemented with two different approaches. These are appearance-based approaches that fall into the techniques of machine learning. The face detection is done according to the Viola & Jones algorithm which proves to be an excellent face detector. The Eigen Face & Bag of Visual Features identification methods are implemented by displaying the identification images in the area of features and weights in order to finalize the classification and reduce the computational cost. These methods were implemented in MATLAB.

Keywords: Face Detection, Face Recognition, Viola & Jones Algorithm, Eigen Faces, Principal Component Analysis, Bag of Words, Bag of Visual Features

Πίνακας περιεχομένων

Εισαγωγή

Η ανίχνευση και αναγνώριση προσώπων αποτελεί ένα από τα πιο γρήγορα εξελισσόμενα πεδία τεχνολογιών της όρασης υπολογιστών και της μηχανικής μάθησης καθώς περιλαμβάνει την αναγνώριση της προσωπικής ταυτότητας από ένα σύστημα βάσει των γεωμετρικών ή στατιστικών χαρακτηριστικών που προέρχονται από εικόνες προσώπου. Η όραση υπολογιστών ασχολείται με την απόκτηση, την ανάλυση και την κατανόηση εικόνων ή βίντεο και έχει ως σκοπό να δώσει στα υπολογιστικά συστήματα μια εποπτεία και κατανόηση του πραγματικού κόσμου. Σε ένα υπολογιστικό σύστημα, μια εικόνα διαθέτει μια ψηφιακή αναπαράσταση. Στην πιο απλή και συνηθισμένη της μορφή, μια δισδιάστατη εικόνα αναπαρίσταται με ένα ψηφιακό σήμα δυο διαστάσεων [i,j]. Η τιμή του σήματος σε κάθε σημείο, γνωστό ως εικονοστοιχείο (pixel), του επιπέδου αφορά την τιμή της έντασης του χρώματος της εικόνας στη θέση αυτή. Επομένως, τα εικονοστοιχεία αποτελούν την είσοδο σε αλγορίθμους που «μιμούνται» την ανθρώπινη κατανόηση και αντίληψη για τα πρόσωπα, με σκοπό την ταξινόμησή τους ώστε να επιτευχθεί πρώτα η ανίχνευση, ο εντοπισμός και εν τέλει η αναγνώριση προσώπου.

Η αναγνώριση προσώπου είναι μία από τις λίγες βιομετρικές μεθόδους που διαθέτουν τα πλεονεκτήματα τόσο της υψηλής ακρίβειας όσο και της χαμηλής εισβολής στον προσωπικό χώρο του ανθρώπου. Έχει την ακρίβεια μιας φυσιολογικής προσέγγισης χωρίς όμως να είναι ενοχλητική. Για το λόγο αυτό, από τις αρχές της δεκαετίας του '70 [1], η αναγνώριση προσώπου έχει τραβήξει την προσοχή ερευνητών σε τομείς από την ασφάλεια, την ψυχολογία, την επεξεργασία εικόνων και στην όραση υπολογιστών. Γενικότερα έχουν προταθεί πολυάριθμοι αλγόριθμοι για την ανίχνευση και αναγνώριση προσώπου οι οποίοι μπορούν να ταξινομηθούν σε δύο κύριες κατηγορίες βάσει της προσέγγισής τους. Στην πρώτη κατηγορία είναι εκείνοι που επιτυγχάνουν τον στόχο με την εξαγωγή χαρακτηριστικών χαμηλού επιπέδου ή προτύπων και μοτίβων, και η ακρίβειά τους βασίζεται κυρίως στην «γνώση» του σχεδιαστή του συστήματος. Σε αυτές τις μεθόδους εισάγονται μοντέλα γνώσης και ο τελικός στόχος επιτυγχάνεται με σύγκριση των μοντέλων με την εισαγόμενη προς ανίχνευση ή αναγνώριση εικόνα. Αυτές οι τεχνικές, αν και μπορούν να επιτύχουν αρκετά ικανοποιητικά αποτελέσματα, παρουσιάζουν μεγάλη ευαισθησία στις συνθήκες λήψης των εικόνων ή του σύνθετου φόντου τους. Στην δεύτερη κατηγορία ανήκουν πιο σύνθετοι και πολύπλοκοι αλγόριθμοι, οι οποίοι μπορούν να ανιχνεύσουν και να αναγνωρίσουν ένα πρόσωπο μέσω μιας διαδικασίας που λέγεται εκπαίδευσή. Η προσέγγιση αυτή εμπίπτει σε τεχνικές μηχανικής μάθησης όπου το σύστημα τροφοδοτείται αρχικά από ένα σύνολο εικόνων εκπαίδευσης(training set). Ένα τέτοιο σύστημα «μαθαίνει» τα χαρακτηριστικά του προσώπου και στην συνέχεια καλείται να ταξινομήσει ανάλογα με τα μοτίβα που δημιουργούνται. Ένα τέτοιο σύστημα μηχανικής μάθησης θα μπορούσε να ταξινομηθεί σε τρεις κατηγορίες:

- Εκμάθηση χαρακτηριστικών τύπου HAAR, SIFT, SURF, HOG κ.α.
- Εκμάθηση χαρακτηριστικών εικόνων (Eigen Faces)
- Αυτόματη εκμάθηση χαρακτηριστικών (Neural Networks)

Η σχέση της ανίχνευσης και αναγνώρισης προσώπου είναι πολύ στενή καθώς η ανίχνευση αποτελεί μέρος της διαδικασίας αναγνώρισης. Τα περισσότερα συστήματα αναγνώρισης προσώπων είναι σχεδιασμένα με τρόπο ώστε να μπορούν να επιτυγχάνουν και ανίχνευση. Ο διαχωρισμός των τεχνικών αυτών σχετίζεται συνήθως με τις εικόνες εκπαίδευσης. Για την ανίχνευση προσώπων οι εικόνες εκπαίδευσης διαχωρίζονται σε «θετικά» (εικόνες με πρόσωπο) και «αρνητικά» (εικόνες χωρίς πρόσωπα), ενώ για την αναγνώριση οι εικόνες διαχωρίζονται στις κατηγορίες-κλάσεις των προσώπων προς ανίχνευση, δηλαδή το όνομα του ατόμου προς αναγνώριση. Αυτές οι τεχνικές με βάση την

εμφάνιση πετυχαίνουν αναγνώριση σε πραγματικό χρόνο (real time recognition) και σε πολύ υψηλές ακρίβειες που φτάνουν έως και το 99%.

Αν και οι αλγόριθμοι αναγνώρισης προσώπων είναι γενικά αρκετά επιτυχημένοι, εξακολουθούν να υπάρχουν πολλές ανοιχτές προκλήσεις. Η γήρανση του ατόμου, η στάση του, οι εκφράσεις του προσώπου, η ομοιότητα των ανθρώπων όπως στα αδέλφια, τα περιβαλλοντικά χαρακτηριστικά, όπως ο φωτισμός και το φόντο, η χαμηλή ανάλυση των εικόνων, η κλίμακα (αφορά την απόσταση του ατόμου από την κάμερα), οι μερικές αποκρύψεις (occlusions) είναι μόνο μερικές από τις προκλήσεις που αντιμετωπίζει ένα σύστημα αναγνώρισης προσώπου [2].

Η διαδεδομένη χρήση της αναγνώρισης προσώπου αντανακλά εν μέρει τον αριθμό των δραστηριοτήτων που ήδη απαιτούν ταυτοποίηση ενός ατόμου. Η σάρωση προσώπου έχει αντικαταστήσει ή αυξήσει τους ελέγχους ταυτοτήτων στα ξενοδοχεία, τις πτήσεις και τα τρένα καθώς και στις τράπεζες και τα νοσοκομεία. Η ανίχνευση και αναγνώριση προσώπων χρησιμοποιείται όμως και για να επεκτείνει την παρακολούθηση στο «όνομα της ασφάλειας». Πολλές χώρες, όπως η Kiva, παρουσιάζουν στατιστικά για την μείωση της εγκληματικότητας καθώς το δίκτυο καμερών που έχει απλωθεί στα αστικά κέντρα οδήγησε στην αναγνώριση εγκληματιών με αποτέλεσμα την άμεση επέμβαση της αστυνομίας. Από την άλλη μεριά, η εγκατάσταση ενός συστήματος παρακολούθησης προσώπου σε μια τάξη στο πανεπιστήμιο Φαρμακευτικής της Kίνας στο πλαίσιο του «έξυπνου πανεπιστημίου» πυροδότησε μια εκτεταμένη κριτική στο Διαδίκτυο. Η σχολή εγκατέστησε σαρωτές προσώπου για την παρακολούθηση της συμπεριφοράς των φοιτητών στην τάξη, δηλαδή πχ. αν ασχολούνται με τα κινητά τους ή αν μιλάνε με τους συμφοιτητές την ώρα του μαθήματος. Η Kiva, ως πρωτοπόρα δύναμη στην αναγνώριση προσώπων, έχει εισαγάγει και πιλοτικά προγράμματα (point system) σε πόλεις της για την παρακολούθηση ανθρώπων όπου οι άνθρωποι θα επιβραβεύονται ή θα έχουν κυρώσεις ανάλογα με την κοινωνική τους συμπεριφορά .Όλες αυτές οι εφαρμογές φαντάζουν όσο εκπληκτικές για τις δυνατότητές τους τόσο και απειλητικές για τον τρόπο με τον οποίο θα εφαρμοσθούν και για την κατάχρηση των προσωπικών δεδομένων και ελευθεριών στο μέλλον.

Με στόχο την κατανόηση και εμβάθυνση των μεθόδων αναγνώρισης προσώπων με μηχανική μάθηση, στο πλαίσιο της διπλωματικής εργασίας, γίνεται η απόπειρα υλοποίησης ενός αλγορίθμου ανίχνευσης και αναγνώρισης προσώπων.

Δομή κειμένου

Στο πρώτο κεφάλαιο γίνεται μια αναφορά στη μηχανική μάθηση και τις βασικότερες μεθόδους ταξινόμησης που χρησιμοποιήθηκαν για την υλοποίηση τους αλγορίθμου αναγνώρισης προσώπου. Στο δεύτερο κεφάλαιο υπάρχει εκτενής αναφορά στις κυριότερες μεθόδους ανίχνευσης προσώπων. Το κεφάλαιο αυτό αποτελείται από δυο κύριες ενότητες που αναλύουν την ανίχνευση προσώπου βάσει της γνώσης και βάσει της εμφάνισης-εικόνας. Το τρίτο κεφάλαιο περιλαμβάνει το θεωρητικό υπόβαθρο των μεθόδων αναγνώρισης που χρησιμοποιήθηκαν, ενώ στο τέταρτο κεφάλαιο αναλύεται η υλοποίηση του αλγορίθμου. Στο πέμπτο κεφάλαιο υπάρχει η παρουσίαση των αποτελεσμάτων και, τέλος, στο έκτο κεφάλαιο συνοψίζονται τα συμπεράσματα που προέκυψαν από την εφαρμογή του αλγορίθμου. Η εργασία ολοκληρώνεται με την βιβλιογραφία που χρησιμοποιήθηκε.

.

Κεφάλαιο 1. Μηχανική Μάθηση

Από επιστημονική και φιλοσοφική άποψη, η μηχανική μάθηση είναι ενδιαφέρουσα, διότι η ανάπτυξη της κατανόησής μας για αυτήν συνεπάγεται με την ανάπτυξη της κατανόησης των αρχών που διέπουν την ανθρώπινη νοημοσύνη. Η μηχανική μάθηση μας δίνει τη δυνατότητα να αντιμετωπίσουμε εργασίες που είναι υπερβολικά δύσκολο να επιλυθούν με συγκεκριμένα προγράμματα γραμμένα και σχεδιασμένα από ανθρώπους. Σε αυτόν τον σχετικά επίσημο ορισμό της λέξης «εργασία», η ίδια διαδικασία της μάθησης δεν είναι εργασία. Η μάθηση είναι το μέσο για την επίτευξη της ικανότητας εκτέλεσης της εργασίας. Για παράδειγμα, εάν θέλουμε ένα ρομπότ να μπορεί να κινείται, τότε η κίνηση είναι η εργασία. Θα μπορούσαμε να προγραμματίσουμε το ρομπότ να μάθει να κινείται, ή θα μπορούσαμε να προσπαθήσουμε να γράψουμε άμεσα ένα πρόγραμμα που καθορίζει πώς να κινείται χειροκίνητα.

Ένας αλγόριθμος μηχανικής μάθησης είναι ένας αλγόριθμος που μπορεί να μάθει από τα δεδομένα. Ο Mitchell το 1997 [3] παρέχει έναν συνοπτικό ορισμό: " Ένα πρόγραμμα ηλεκτρονικού υπολογιστή λέγεται ότι μαθαίνει από την εμπειρία **E** σε σχέση με κάποια τάξη εργασιών **T** και μέτρηση απόδοσης **P**, αν η απόδοσή του στις εργασίες T, όπως μετρείται με P, βελτιώνεται με την εμπειρία **E** ".

Σύμφωνα με τους Goodfellow et al. [4] οι εργασίες **T** μηχανικής μάθησης περιγράφονται συνήθως ως προς τον τρόπο με τον οποίο το σύστημα μηχανικής μάθησης πρέπει να επεξεργάζεται ένα παράδειγμα από ένα σύνολο δεδομένων. Ένα σύνολο δεδομένων είναι μια συλλογή πολλών παραδειγμάτων. Μερικές φορές τα παραδείγματα καλούνται και σημεία δεδομένων. Ένα από τα παλαιότερα σύνολα δεδομένων που μελετήθηκαν από τους στατιστικολόγους και τους ερευνητές της μηχανικής μάθησης είναι το σύνολο δεδομένων Iris [5], όπου στο σύνολο δεδομένων αντιπροσωπεύονται τρία διαφορετικά είδη. Πρόκειται για μια συλλογή μετρήσεων διαφορετικών τμημάτων 150 φυτών ίριδας. Κάθε μεμονωμένο φυτό αντιστοιχεί σε ένα παράδειγμα και χαρακτηριστικά σε κάθε παράδειγμα είναι οι μετρήσεις κάθε μέρους του φυτού: το μήκος του σέπαλου, το πλάτος του σέπαλου, το μήκος του πέταλου και το πλάτος του πέταλου. Το σύνολο δεδομένων καταγράφει επίσης σε ποιο είδος ανήκε κάθε φυτό. Επομένως ένα παράδειγμα καλείται μια συλλογή χαρακτηριστικών που έχουν μετρηθεί ποσοτικά σε κάποιο αντικείμενο και εισάγεται στο σύστημα μηχανικής μάθησης προς επεξεργασία. Συνήθως το παράδειγμα αντιπροσωπεύεται από ένα διάνυσμα $x \in R^n$, όπου κάθε καταχώριση x_i του διανύσματος είναι ένα άλλο χαρακτηριστικό. Στο σύνολο δεδομένων IRIS τα χαρακτηριστικά είναι το πλάτος και το μήκος σέπαλου και πέταλου. Αντίστοιχα, σε μια εικόνα τα χαρακτηριστικά της στην απλουστευμένη τους μορφή είναι συνήθως οι τιμές των εικονοστοιχείων.

Για να αξιολογήσουμε τις ικανότητες ενός αλγορίθμου μηχανικής μάθησης, πρέπει να σχεδιάσουμε ένα ποσοτικό μέτρο της απόδοσής του **P**. Συνήθως αυτό το μέτρο απόδοσης είναι προσαρμοσμένο για την εργασία που εκτελείται από το σύστημα. Για εργασίες όπως η ταξινόμηση, η ταξινόμηση με ελλιπείς εισόδους και η μεταγραφή, συχνά υπολογίζεται η ακρίβεια του μοντέλου. Η ακρίβεια είναι απλώς η αναλογία των παραδειγμάτων για τα οποία το μοντέλο παράγει το σωστό αποτέλεσμα. Μπορούμε επίσης να λάβουμε ισοδύναμες πληροφορίες μετρώντας το ποσοστό σφάλματος, την αναλογία των παραδειγμάτων για τα οποία το μοντέλο παράγει ένα λανθασμένο αποτέλεσμα. Συνήθως άξιο ενδιαφέροντος είναι το πόσο καλά ο αλγόριθμος μηχανικής μάθησης αποδίδει σε δεδομένα που δεν έχει δει ποτέ, αφού αυτό καθορίζει το πόσο καλά θα λειτουργήσει όταν αναπτυχθεί στον πραγματικό κόσμο. Επομένως, αξιολογεί κανείς αυτά τα μέτρα απόδοσης χρησιμοποιώντας ένα δοκιμαστικό σύνολο δεδομένων που είναι ξεχωριστό από τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση του συστήματος μηχανικής μάθησης. Η επιλογή του μέτρου απόδοσης μπορεί να μοιάζει απλή και αντικειμενική, αλλά συχνά είναι δύσκολο να επιλεγεί ένα μέτρο απόδοσης που αντιστοιχεί βέλτιστα στην επιθυμητή συμπεριφορά του συστήματος. Σε ορισμένες περιπτώσεις, αυτό οφείλεται στο γεγονός ότι είναι δύσκολο να αποφασιστεί τι πρέπει να μετρηθεί.

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να κατηγοριοποιηθούν ευρέως ως μη εποπτευόμενοι ή εποπτευόμενοι με βάση το είδος της εμπειρίας **E** που τους επιτρέπεται να έχουν κατά τη διάρκεια της μαθησιακής διαδικασίας. Οι μη εποπτευόμενοι αλγόριθμοι μάθησης αντιμετωπίζουν ένα σύνολο δεδομένων που περιέχει πολλές δυνατότητες και, στη συνέχεια, μαθαίνουν χρήσιμες ιδιότητες της δομής αυτού του συνόλου δεδομένων. Στο πλαίσιο της βαθιάς μάθησης, είναι συνήθως θεμιτό να «μαθευτεί» ολόκληρη η κατανομή πιθανότητας που δημιούργησε ένα σύνολο δεδομένων, είτε ρητά όπως στην εκτίμηση πυκνότητας είτε σιωπηρά για εργασίες όπως η σύνθεση. Ορισμένοι άλλοι μη εποπτευόμενοι αλγόριθμοι μάθησης εκτελούν διαφορετικούς ρόλους, όπως ομαδοποίηση, ο οποίος συνίσταται στην διαίρεση του συνόλου δεδομένων σε ομάδες παρόμοιων παραδειγμάτων. Οι εποπτευόμενοι αλγόριθμοι μάθησης αντιμετωπίζουν ένα σύνολο δεδομένων που περιέχει χαρακτηριστικά, αλλά κάθε παράδειγμα σχετίζεται επίσης με μία ετικέτα ή έναν στόχο. Για παράδειγμα, το σύνολο δεδομένων Iris σχολιάζεται με το είδος κάθε φυτού ίριδας. Ένας εποπτευόμενος αλγόριθμος μάθησης μπορεί να μελετήσει το σύνολο δεδομένων Iris και να μάθει να ταξινομεί τα φυτά ίριδας σε τρία διαφορετικά είδη με βάση τις μετρήσεις τους. Παραδοσιακά, οι άνθρωποι αναφέρονται σε προβλήματα παλινδρόμησης, ταξινόμησης και δομημένων αποτελεσμάτων ως εποπτευόμενη μάθηση. Αν και η μη εποπτευόμενη και η εποπτευόμενη μάθηση δεν είναι εντελώς τυπικές ή διακριτές έννοιες, βοηθούν στην κατηγοριοποίηση ορισμένων διεργασιών που γίνονται με τους αλγόριθμους μηχανικής μάθησης. Ορισμένοι αλγόριθμοι μηχανικής μάθησης δεν αντιμετωπίζουν μόνο ένα σταθερό σύνολο δεδομένων. Για παράδειγμα, οι αλγόριθμοι ενισχυτικής μάθησης αλληλοεπιδρούν με ένα περιβάλλον, οπότε υπάρχει ένας βρόχος ανατροφοδότησης μεταξύ του συστήματος μάθησης και των εμπειριών του.

1.1 Εργασίες μηχανικής μάθησης

Ταξινόμηση: Η ταξινόμηση αποτελεί μια από τις σημαντικότερες εργασίες της μηχανικής μάθησης. Σε αυτού του τύπου την εργασία, τα δεδομένα εισόδου χωρίζονται σε δύο ή περισσότερες κλάσεις, και το σύστημα πρέπει να κατασκευάσει ένα μοντέλο το οποίο θα ταξινομεί τα δεδομένα σε μία ή περισσότερες κλάσεις(ταξινόμηση multi-label). Αυτό συνήθως εμπίπτει στην εποπτευόμενη μάθηση. Ένα παράδειγμα μιας εργασίας ταξινόμησης είναι η αναγνώριση αντικειμένου, όπου η είσοδος είναι μια εικόνα και έξοδος είναι ένας αριθμητικός κωδικός που προσδιορίζει το αντικείμενο στην εικόνα. Η πιο απλουστευμένη μορφή ταξινόμησης για την αναγνώριση προτύπων είναι ο υπολογισμός της ελάχιστης διαφοράς των εικόνων. Για κάθε κλάση εικόνων υπολογίζεται η μέση τιμή των παραδειγμάτων (εικόνων εκπαίδευσης) και στην συνέχεια υπολογίζεται η διαφορά τους από την εικόνα εισόδου. Έτσι η εικόνα εισόδου θα μπορούσε να ταξινομηθεί στην κλάση με την ελάχιστη διαφορά. Προφανώς αυτή η τεχνική, εφόσον γίνεται στο επίπεδο των εικονοστοιχείων, παρουσιάζει πολύ μεγάλο υπολογιστικό κόστος, χαμηλή απόδοση και μεγάλη ευαισθησία σε μεταθέσεις, κλίμακα, φωτισμό κ.ά. του προς αναγνώριση αντικειμένου στην εικόνα. Για την επίλυση των προβλημάτων αυτών έχει προταθεί πληθώρα αλγορίθμων που στοχεύουν στον αυτόματο ή μη υπολογισμό χαρακτηριστικών σημείων στην εικόνα και, στην συνέχεια, την αντιστοίχισή τους σε κάθε κλάση ώστε να επιτευχθεί η ταξινόμηση για κάθε νέα εικόνα εισόδου. Η σύγχρονη αναγνώριση αντικειμένων εμπίπτει σε τεχνικές βαθιάς μάθησης [6, 7] και υλοποιείται με την χρήση νευρωνικών δικτύων.

Ταξινόμηση με ελλιπείς εισόδους: Η ταξινόμηση γίνεται πιο δύσκολη εάν στο σύστημα δεν είναι εγγυημένο ότι κάθε μέτρηση/χαρακτηριστικό θα παρέχεται πάντα στο διάνυσμα εισόδου του. Για να λυθεί αυτή η εργασία ταξινόμησης, ο αλγόριθμος μάθησης πρέπει να καθορίσει μόνο μία συνάρτηση που θα συνδέεται από ένα διάνυσμα εισόδου σε ένα κατηγορικό αποτέλεσμα. Όταν όμως κάποιες από τις εισόδους μπορεί να λείπουν, αντί να παρέχει μια μοναδική συνάρτηση ο αλγόριθμος μάθησης πρέπει να μάθει ένα σύνολο από συναρτήσεις. Σε κάθε συνάρτηση αντιστοιχεί το να ταξινομήσει το διάνυσμα x με διαφορετικό υποσύνολο των εισόδων που της λείπουν. Ένας τρόπος για να προσδιορι-

στεί αποτελεσματικά ένα τόσο μεγάλο σύνολο συναρτήσεων είναι να βρεθεί μια κατανομή πιθανότητας για όλες τις σχετικές μεταβλητές, και έπειτα να λυθεί η εργασία ταξινόμησης με την απόρριψη των ελλειπουσών μεταβλητών.

Παλινδρόμηση: Σε μία τέτοια εργασία, ζητείται η πρόβλεψη μίας αριθμητικής τιμής αφού δοθεί κάποια είσοδος. Έτσι, για να λυθεί, ο αλγόριθμος μάθησης ζητείται να παράγει μια συνάρτηση f:Rn => R. Αν και αυτή η εργασία είναι παρόμοια με την εργασία ταξινόμησης, η διαφορά έγκειται στη μορφή της εξόδου καθώς τα αποτελέσματα είναι συνεχή και όχι διακριτά. Ένα παράδειγμα εργασίας παλινδρόμησης είναι η πρόβλεψη του αναμενόμενου ποσού αξίωσης που θα υποβάλει ο ασφαλισμένος ή η πρόβλεψη μελλοντικών τιμών ασφάλειας, η πρόβλεψη των αξιών ακινήτων κ.ά. Αυτά τα είδη προβλέψεων χρησιμοποιούνται επίσης για αλγοριθμικές συναλλαγές.

Μεταγραφή: Σε μία τέτοια εργασία, το σύστημα μηχανικής μάθησης ζητείται να παρατηρήσει μία σχετικά αδόμητη αναπαράσταση κάποιου είδους δεδομένων και να μεταγράψει την πληροφορία σε διακριτή μορφή κειμένου. Για παράδειγμα, σε οπτική αναγνώριση χαρακτήρων OCR στο σύστημα εισάγεται μια εικόνα κειμένου και καλείται να επιστραφεί αυτό το κείμενο με τη μορφή ακολουθίας χαρακτήρων. Ένα άλλο παράδειγμα είναι η αναγνώριση ομιλίας, όπου εισάγεται μια ηχητική κυματομορφή και εκπέμπεται μια ακολουθία χαρακτήρων ή κωδικών ID λέξεων που περιγράφουν τις λέξεις που εκφωνήθηκαν στην ηχογράφηση. Η βαθιά μάθηση είναι ένα κρίσιμο στοιχείο των σύγχρονων συστημάτων αναγνώρισης ομιλίας που χρησιμοποιούνται σε μεγάλες εταιρείες, συμπεριλαμβανομένων των Microsoft, IBM και Google [8].

Μηχανική Μετάφραση: Σε μια εργασία μηχανικής μετάφρασης, η είσοδος αποτελείται από μια ακολουθία συμβόλων σε κάποια γλώσσα και το πρόγραμμα υπολογιστή πρέπει να το μετατρέψει σε ακολουθία συμβόλων σε άλλη γλώσσα. Αυτό εφαρμόζεται συνήθως σε φυσικές γλώσσες, όπως η μετάφραση ενός κειμένου από Αγγλικά σε Ελληνικά.

Δομημένο αποτέλεσμα: Οι δομημένες εργασίες αποτελέσματος περιλαμβάνουν κάθε εργασία όπου το αποτέλεσμα είναι ένα διάνυσμα (ή άλλη δομή δεδομένων που περιέχει πολλές τιμές) με σημαντικές σχέσεις μεταξύ των διάφορων στοιχείων. Ένα παράδειγμα είναι η ανάλυση-αντιστοίχιση μιας φράσης φυσικής γλώσσας σε ένα δέντρο που περιγράφει τη γραμματική του δομή επισημαίνοντας τους κόμβους των δέντρων ως ρήματα, ουσιαστικά, επιρρήματα κ.ο.κ. Αυτές οι εργασίες ονομάζονται δομημένες εργασίες αποτελέσματος επειδή το πρόγραμμα πρέπει να παράγει πολλές τιμές που είναι στενά αλληλένδετες. Για παράδειγμα, οι λέξεις που παράγονται από ένα πρόγραμμα λεζάντας εικόνας πρέπει να αποτελούν μια έγκυρη πρόταση.

1.2 Ταξινομητές - Μοντέλα μηχανικής μάθησης

1.2.1 Νευρωνικά δίκτυα (Neural networks – NN)

Τα νευρωνικά δίκτυα (NN) αποτελούν μια πανίσχυρη τεχνική μηχανικής μάθησης και τα τελευταία χρόνια χρησιμοποιούνται ευρέως για την επίλυση προβλημάτων τεχνητής νοημοσύνης (Artificial Intelligence – AI). Τα NN εντάσσονται πλέον στις πιο σύγχρονες τεχνικές ταξινόμησης αντικειμένων καθώς χρησιμοποιούνται για την πρόβλεψη όχι μόνο για τα γνωστά δεδομένα αλλά και για άγνωστα δεδομένα, και χρησιμοποιούνται σε πολλούς τομείς όπως η ανίχνευση και αναγνώριση προσώπου, η ερμηνεία οπτικών σκηνών, η αναγνώριση ομιλίας, η αναγνώριση δακτυλικών αποτυπωμάτων, η αναγνώριση ίριδας κ.λπ.

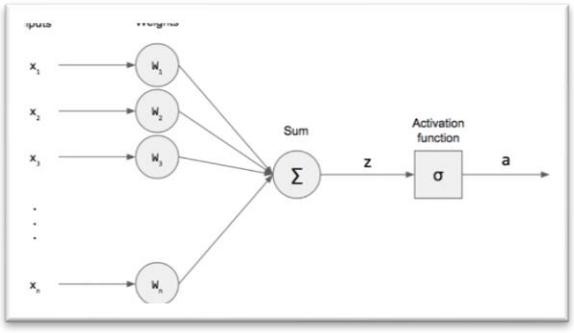
Η μελέτη των τεχνητών νευρικών δικτύων (ΤΝΔ) έλκει την έμπνευσή της εν μέρει από την παρατήρηση ότι τα βιολογικά συστήματα μάθησης είναι κατασκευασμένα από πολύπλοκους ιστούς διασυνδεδεμένων νευρώνων. Προσεγγιστικά μιλώντας, θα έλεγε κανείς ότι τα τεχνητά νευρικά δίκτυα κατασκευάζονται από ένα πυκνά διασυνδεδεμένο σύνολο απλών μονάδων, όπου κάθε μονάδα παίρνει έναν αριθμό από πραγματικές εισόδους

(πιθανώς τα αποτελέσματα άλλων μονάδων) και παράγει ένα πραγματικό αποτέλεσμα (που μπορεί να γίνει η είσοδος πολλών άλλων μονάδων).

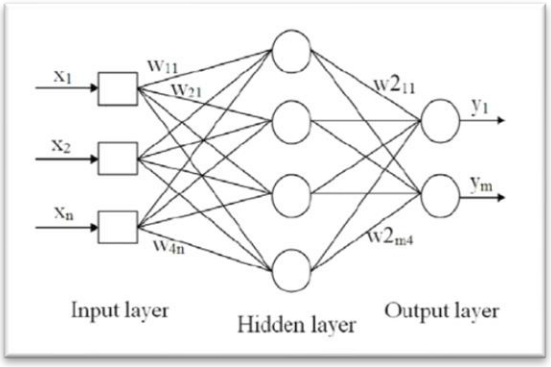
Η βασική δομή των ΤΝΔ είναι ένα δίκτυο μικρών μονάδων επεξεργασίας ή κόμβων, τα οποία συνδέονται μεταξύ τους με σταθμισμένες συνδέσεις. Στην αρχή τα ΤΝΔ παρουσιάστηκαν ως ένα μαθηματικό μοντέλο το οποίο προσομοιώνει την πολύπλοκη λειτουργία εγκεφαλικών νευρώνων του ανθρώπινου εγκεφάλου. Ένα τεχνητό νευρωνικό δίκτυο αποτελείται από ένα δίκτυο τεχνητών νευρώνων, επίσης γνωστών ως "κόμβων" (nodes), που επικοινωνούν και αλληλεπιδρούν, συνδεόμενοι μεταξύ τους με τις λεγόμενες συνάψεις (synapses). Η κάθε σύναψη έχει διαφορετικό βαθμό αλληλεπίδρασης, ο οποίος καθορίζεται από τα συναπτικά βάρη (synaptic weights) , τα οποία μεταβάλλονται ανάλογα κατά την εκπαίδευση. Έτσι ένας κόμβος αποτελείται από m εισόδους $[x_1, x_2, \dots x_m]^T$ και m συναπτικά βάρη $[w_1, w_2, \dots w_m]$ και μια πόλωση b (bias):

$$u = \sum_{i=1}^{m} w_i \, x_i + b$$

Μπορούμε να πούμε πως στα NN υπάρχει μια επικοινωνία και προσαρμοστικότητα των βαρών ώστε να επιτευχθεί η εκπαίδευση, καθώς τα βάρη μεταβάλλονται καθ' όλη την διάρκεια της εκπαίδευσης αποδυναμώνοντας ή ενδυναμώνοντας την ισχύ του δεσμού. Το δίκτυο ενεργοποιείται παρέχοντας μία είσοδο σε μερικούς ή και όλους τους κόμβους, και στην συνέχεια αυτή η ενεργοποίηση εξαπλώνεται σε όλο το δίκτυο κατά μήκος των σταθμισμένων συνδέσεων. Συνεπώς τα βάρη αποτελούν μια κωδικοποίηση της εμπειρικής γνώσης η οποία έρχεται μέσω της εκπαίδευσης.



Εικόνα 1. Perceptron: Το πρώτο Τεχνητό Νευρωνικό δίκτυο



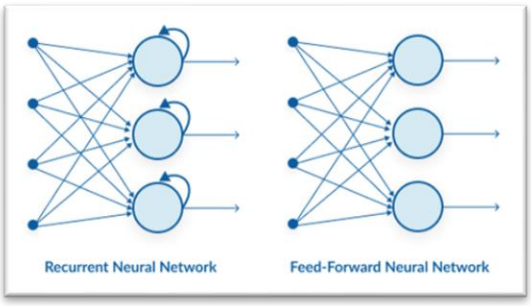
Εικόνα 2. Δίκτυο πολλαπλών επιπέδων (Multi-Layer Perceptron)

Υπάρχουν δύο κυρίως είδη ΤΝΔ, τα κυκλικά και τα μη κυκλικά. Τα κυκλικά συνήθως ονομάζονται feedback, recursive, ή recurrent νευρωνικά δίκτυα, ενώ τα μη κυκλικά feed-forward νευρωνικά δίκτυα (FNN). Το πιο γνωστό και διαδομένο FNN είναι το πολυεπίπεδο perceptron (multi-layer perceptron).

Ένα νευρωνικό δίκτυο εκτελεί μια διαδικασία αναγνώρισης προτύπων (ανίχνευση-αναγνώριση με βάση την εικόνα) αφού πρώτα περάσει από μία διαδικασία εκπαίδευσης. Κατά την διαδικασία της εκπαίδευσης στο δίκτυο παρουσιάζεται, επαναληπτικά, ένα σύνολο προτύπων εισόδου $x_1, x_2, \dots x_m$ μαζί με την κλάση $y_1, y_2, \dots y_m$ στην οποία ανήκει το καθένα, δηλαδή την επιθυμητή έξοδο. Στην συνέχεια, αφού ολοκληρωθεί η διαδικασία της εκπαίδευσης, το δίκτυο μπορεί να ανιχνεύσει και να αναγνωρίσει την κλάση των αντικειμένων στην οποία ανήκουν τα άγνωστα πρότυπα βάσει της πληροφορίας που έχει εξαχθεί κατά την διάρκεια της εκπαίδευσης. Ένα νευρωνικό δίκτυο είναι σε θέση να αναπαραστήσει ένα σύνολο προτύπων σε έναν πολυδιάστατο χώρο αποφάσεων, ο οποίος διαιρείται σε περιοχές ανάλογα με τις κλάσεις του προβλήματος. Δεδομένου ότι η ανίχνευση προσώπου μπορεί να αντιμετωπιστεί ως ένα πρόβλημα αναγνώρισης προτύπων δύο κατηγοριών (πρόσωπο – μη πρόσωπο), έχουν προταθεί διάφορες αρχιτεκτονικές νευρωνικών δικτύων. Το πλεονέκτημα της χρήσης νευρωνικών δικτύων για την ανίχνευση προσώπου είναι η σκοπιμότητα της κατάρτισης ενός συστήματος για την καταγραφή της σύνθετης κατηγορίας υπό όρους πυκνότητας των προτύπων προσώπου. Ωστόσο, ένα μειονέκτημα είναι ότι η αρχιτεκτονική του δικτύου πρέπει να συντονιστεί εκτενώς (αριθμός επιπέδων, αριθμός κόμβων, ρυθμοί εκμάθησης κ.λπ.) για να επιτύχει εξαιρετική απόδοση.

1.2.1.1 Recurrent Neural Network

Τα Επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent neural network – RNN) [9] [4] είναι ένα είδος νευρωνικών δικτύων για την επεξεργασία αλληλουχίας δεδομένων. Όπως και τα CNN, είναι ένα εξειδικευμένο είδος νευρωνικού δικτύου για την επεξεργασία ενός πίνακα τιμών $\boldsymbol{X}$ όπως μία εικόνα. Ένα RNN είναι ένα νευρωνικό δίκτυο ειδικό για την επεξεργασία μιας αλληλουχίας τιμών $\boldsymbol{x^{(1)}, \dots, x^{(\tau)}}$. Τα περισσότερα RNN μπορούν να επεξεργαστούν αλληλουχίες μεταβλητού μήκους. Η μετάβαση από τα δίκτυα πολλαπλών επιπέδων σε RNN στηρίχτηκε στο πλεονέκτημα που υπάρχει ήδη στις αρχικές ιδέες της μηχανικής μάθησης και των στατιστικών μοντέλων: την κοινή χρήση παραμέτρων σε διάφορα μέρη ενός μοντέλου. Η κοινή χρήση παραμέτρων καθιστά δυνατή την επέκταση και την εφαρμογή του μοντέλου σε παραδείγματα διαφορετικού μήκους και στην γενίκευσή τους. Ένα πρόβλημα που ανακύπτει είναι ότι οι κλίσεις που διαδίδονται σε πολλά στάδια τείνουν είτε να εξαφανίζονται είτε να εκρήγνυνται. Ακόμα και εάν ειπωθεί ότι οι παράμετροι είναι τέτοιες ώστε το RNN να είναι σταθερό, η δυσκολία με μακροπρόθεσμες εξαρτήσεις (long-term dependencies) προκύπτει από τα εκθετικά μικρότερα βάρη που δίδονται σε μακροπρόθεσμες αλληλεπιδράσεις σε σύγκριση με τις βραχυπρόθεσμες. Αυτό το πρόβλημα, που αφορά συγκεκριμένα τα RNN, ονομάζεται και «vanishing and exploding gradient problem» [10, 11].

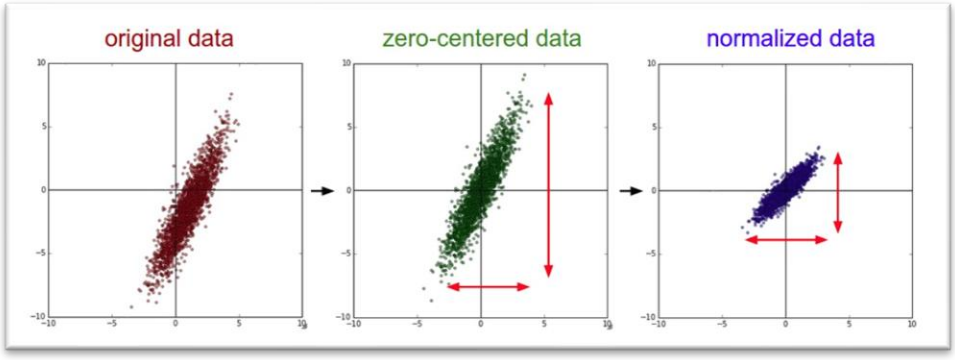


Εικόνα 3 Δομή RNN

1.2.1.2 Convolutional Neural Networks – CNN

1.2.1.2.1 Συγκέντρωση δεδομένων και αρχιτεκτονική δικτύου

Τα συνελικτικά νευρωνικά δίκτυα, γνωστά ως convolutional neural networks – CNN, προτάθηκαν από τους LeCun & Bengio [12] και είναι να εξειδικευμένο είδος νευρωνικού δικτύου για την επεξεργασία δεδομένων που έχει μια γνωστή τοπολογία τύπου πλέγματος. Τα παραδείγματα περιλαμβάνουν κυρίως δεδομένα εικόνων ή χρονοσειρών, τα οποία μπορούν να θεωρηθούν ως ένα δισδιάστατο πλέγμα εικονοστοιχείων. Τα CNN αποτελούν το θεμέλιο των περισσότερων τεχνολογιών όρασης υπολογιστών για την ανίχνευση και αναγνώριση αντικειμένων καθώς, σε αντίθεση με τις παραδοσιακές αρχιτεκτονικές πολλαπλών επιπέδων perceptron, χρησιμοποιούν δύο λειτουργίες, που ονομάζονται συνέλιξη (convolution) και υποδειγματοληψία (pooling), για να μειώσουν τις διαστασεις μιας εικόνας στα βασικά χαρακτηριστικά της και χρησιμοποιούν αυτά τα χαρακτηριστικά για να κατανοήσουν και, τέλος, να ταξινομήσουν μια εικόνα. Το όνομα «convolutional neural network» υποδηλώνει ότι το δίκτυο χρησιμοποιεί συνάψεις (βάρη) με την μορφή φίλτρων, τα οποία εφαρμόζονται με την διαδικασία της συνέλιξης στην εικόνα εισόδου και τα παράγωγά της. Αξίζει να σημειωθεί ότι στις τεχνικές μηχανικής μάθησης τα δεδομένα υπόκεινται συνήθως σε μια προεπεξεργασία, η οποία αφορά κυρίως την κανονικοποίηση των χαρακτηριστικών. Όταν πρόκειται για εικόνα, ως χαρακτηριστικό θεωρείται το κάθε της εικονοστοιχείο. Ειδικότερα[1] κατά την διαδικασία αυτή σημαντικό είναι τα δεδομένα να κεντραριστούν (center the data) με την αφαίρεση της «μέσης» εικόνας από την κάθε εικόνα. Αυτό έχει ως αποτέλεσμα οι τιμές των εικονοστοιχείων να περιγράφονται πλέον από τιμές σε μια κλίμακα από -127 έως 127. Στην συνέχεια αυτές οι τιμές μετασχηματίζονται σε μια κλίμακα από [-1 έως 1].

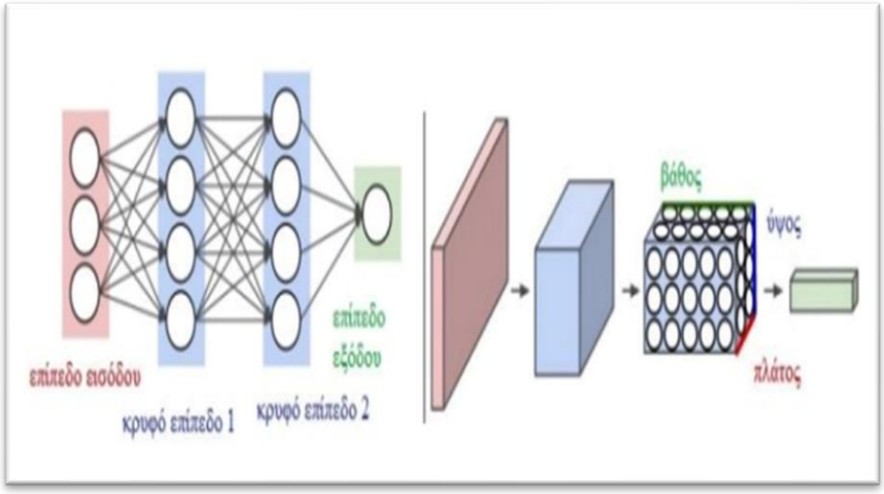


Εικόνα 4 Κανονικοποίηση (normalization) δεδομένων εισόδου (εικόνων)

Η συνέλιξη είναι ένα εξειδικευμένο είδος γραμμικής πράξης. Συνεπώς, τα CNN είναι απλά νευρωνικά δίκτυα που χρησιμοποιούν συνέλιξη σε τουλάχιστον ένα από τα επίπεδά τους, καθώς μπορεί και να ερμηνευθεί ως ο υπολογισμός του εσωτερικού γινομένου μεταξύ των τιμών των φίλτρων και των δεδομένων (όγκου) εισόδου. Η συνέλιξη γίνεται με την εφαρμογή ενός συνόλου εκπαιδευόμενων φίλτρων, ή αλλιώς πυρήνων (kernel), για τον εντοπισμό χαρακτηριστικών σημείων και προτύπων σε μια εικόνα εισόδου. Τα φίλτρα που χρησιμοποιούνται στα πρώτα επίπεδα του δικτύου ανιχνεύουν τα γενικώς πιο σημαντικά χαρακτηριστικά ενός αντικειμένου. Για παράδειγμα, σε ένα πρόσωπο θα ανιχνευτούν αρχικά χαρακτηριστικά όπως τα μάτια, η μύτη κ.λπ. Η εφαρμογή αυτών των φίλτρων παράγει ως έξοδο τους χάρτες χαρακτηριστικών και, σε συνδυασμό με κάποιες συναρτήσεις ενεργοποίησης, τους λεγόμενους χάρτες ενεργοποίησης. Σε ένα CNN εφαρμόζονται πολλά διαφορετικά φίλτρα που εξάγουν διάφορα χαρακτηριστικά μιας

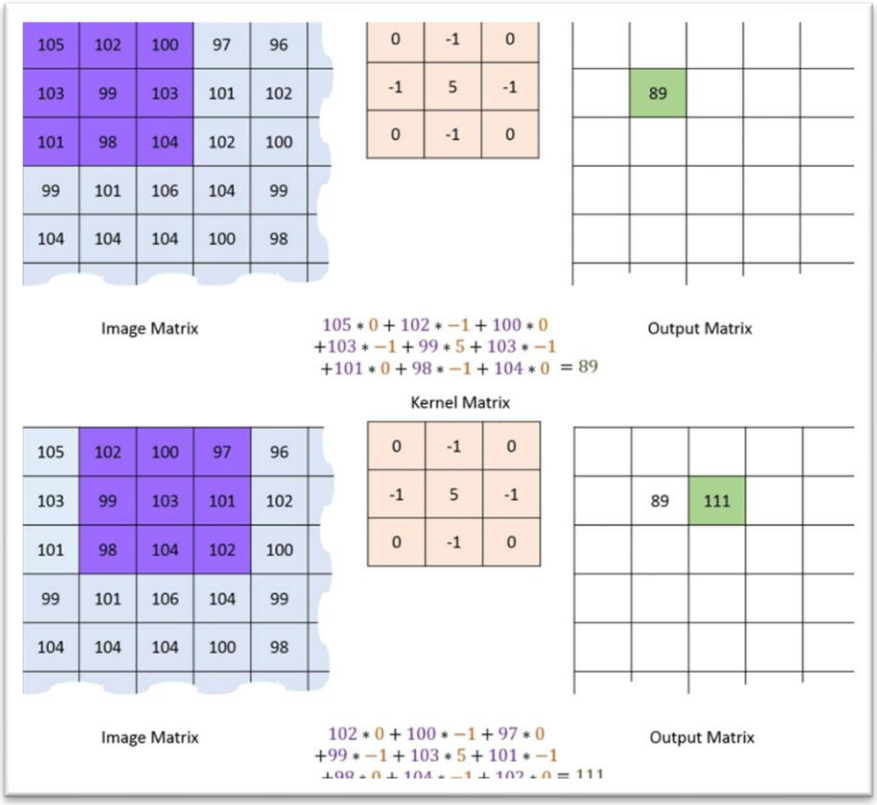
[1] CS231n Convolutional Neural Network for visual recognition

εικόνας. Καθώς προχωράμε στο δίκτυο, τα χαρακτηριστικά που εξάγονται από μια εικόνα γίνονται όλο και πιο συγκεκριμένα καθώς η συνέλιξη είναι μια πολύ αποτελεσματική μέθοδος για την εξομάλυνση εικόνας, την ανίχνευση ακμών και γενικότερα για την εξαγωγή χρήσιμων και σημαντικών πληροφοριών μιας εικόνας.

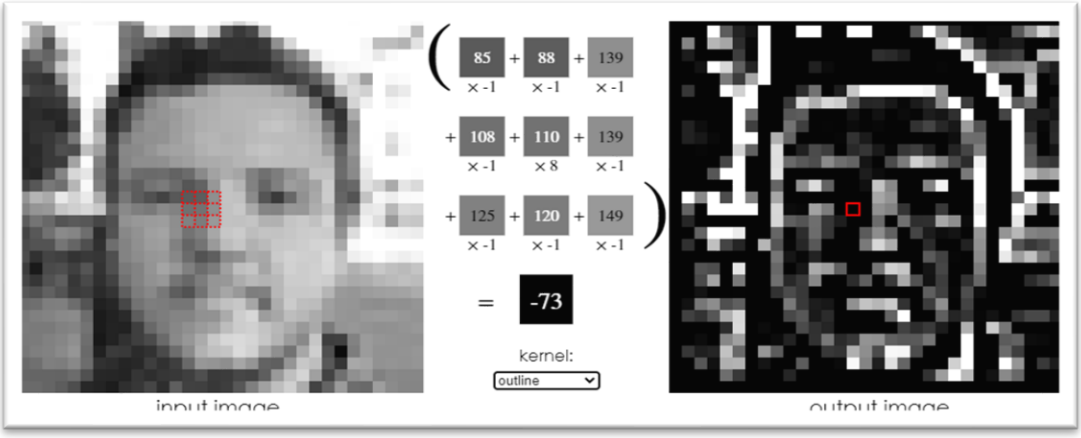


Εικόνα 5. Διαφορά αρχιτεκτονικής δικτύων TNN(αριστερά) και CNN (δεξιά)

.



Εικόνα 6. Παράδειγμα Συνέλιξης σε μια εικόνα (δισδιάστατο πλέγμα εικονοστοιχείων)

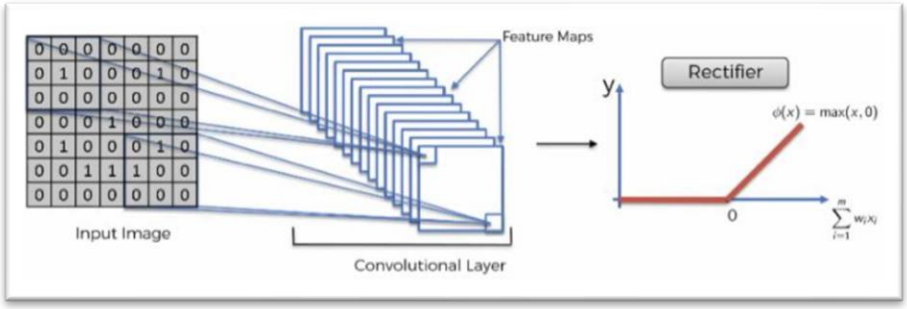


Εικόνα 7. Εφαρμογή φίλτρου σε μια εικόνα προσώπου για την εξαγωγή σημαντικών χαρακτηριστικών του

Στο επίπεδο (layer) της συνέλιξης δημιουργείται ένας πίνακας ίσος ή μικρότερος σε διαστασεις [w,h] από εκείνον της εικόνας εισόδου. Αυτό εξαρτάται από τις υπερ-παραμέτρους που ορίζονται πριν από την εκπαίδευση. Οι υπερ-παράμετροι αφορούν τον αριθμό των φίλτρων, το μέγεθός τους που θα εφαρμοσθεί στην εικόνα για τις συνελίξεις, το βήμα μετακίνησης (stride), το οποίο καθορίζει ανά πόσα στοιχεία του όγκου εισόδου θα μετακινηθεί το φίλτρο, και το μέγεθος της επέκτασης με μηδενικά (padding). Στην συνέχεια εφαρμόζεται η συνάρτηση ενεργοποίησης $\varphi(u)$. Αυτή ορίζει την έξοδο ενός νευρώνα βάσει της τιμής ενεργοποίησης u. Η συνάρτηση ενεργοποίησης είναι ένας μετασχηματισμός μη γραμμικός που εφαρμόζεται στα δεδομένα εισόδου σε ένα επίπεδο του δικτύου. Η μετασχηματισμένη έξοδος στέλνεται στο επόμενο επίπεδο, για το οποίο και αποτελεί σήμα εισόδου. Η σημασία των συναρτήσεων ενεργοποίησης είναι πολύ μεγάλη καθώς λόγω της μη γραμμικής σχέσης επιτρέπουν στο δίκτυο να εκπαιδεύεται «αυτόματα» με τον αλγόριθμο *backpropagation*, ο οποίος με μια επαναληπτική μέθοδο μπορεί να προσαρμόζει τα βάρη με σκοπό την επίτευξη υψηλών ποσοστών ανίχνευσης και αναγνώρισης.

Η πιο συνηθισμένη συνάρτηση ενεργοποίησης είναι η «Ανορθωμένη» Γραμμική ή Συνάρτηση Ράμπας (Rectified Linear Unit - ReLu). Η συνάρτηση αυτή επιτρέπει ταχύτερη και αποτελεσματικότερη εκπαίδευση καθώς μηδενίζει τις αρνητικές τιμές και διαχειρίζεται μόνο τις θετικές τιμές που προκύπτουν. Οι συναρτήσεις αυτές λέγονται συναρτήσεις ενεργοποίησης διότι επιτρέπουν μόνο στα «ενεργοποιημένα χαρακτηριστικά» να προχωρήσουν στο επόμενο επίπεδο.

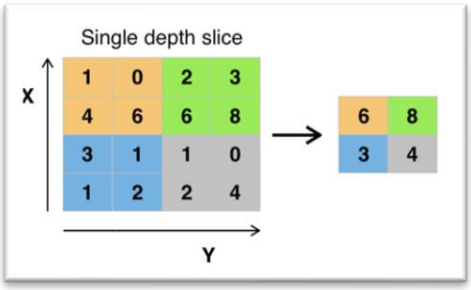
$$\varphi(\upsilon) = \max(0, \upsilon)$$



Εικόνα 8. Επίπεδα συνέλιξης και συνάρτηση ενεργοποίησης ReLU

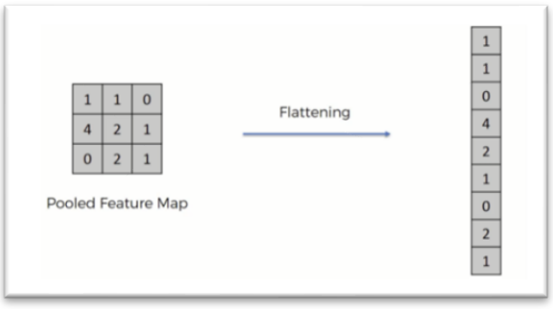
Στο επόμενο επίπεδο γίνεται υποδειγματοληψία (pooling). Σε αυτό το στάδιο γίνεται η μείωση των χωρικών διαστάσεων των δεδομένων εισόδου, με αποτέλεσμα να απλοποιεί τα δεδομένα εξόδου εκτελώντας μη γραμμική δειγματοληψία και μειώνοντας τον αριθμό των παραμέτρων που πρέπει να μάθει το δίκτυο. Με αυτόν τον τρόπο μειώνονται και οι

πιθανότητες υπερ-εκπαίδευσης (overfitting), δηλαδή της δημιουργίας ενός μοντέλου το οποίο θα είναι εκπαιδευμένο αποκλειστικά στα παραδείγματα εισόδου και θα αδυνατεί να ανταποκριθεί στην αναγνώριση άλλων δεδομένων εισόδου τα οποία θα παρουσιάζουν διαφορές. Για παράδειγμα, ένα μοντέλο το οποίο θα μπορεί με μεγάλη επιτυχία να ανιχνεύει-αναγνωρίζει μόνο πρόσωπα που έχουν εισαχθεί στο σύνολο εκπαίδευσης και θα αδυνατεί να αναγνωρίζει νέα πρόσωπα. Η πιο συνήθης μορφή υποδειγματοληψίας είναι η επιλογή της μεγαλύτερης τιμής σε ένα παράθυρο. Στα CNN συνήθως τα δεδομένα εισόδου υπο-τετραπλασιάζονται με την εφαρμογή ενός φίλτρου διαστάσεων 2x2, όπως παρατηρείται στην παρακάτω εικόνα.



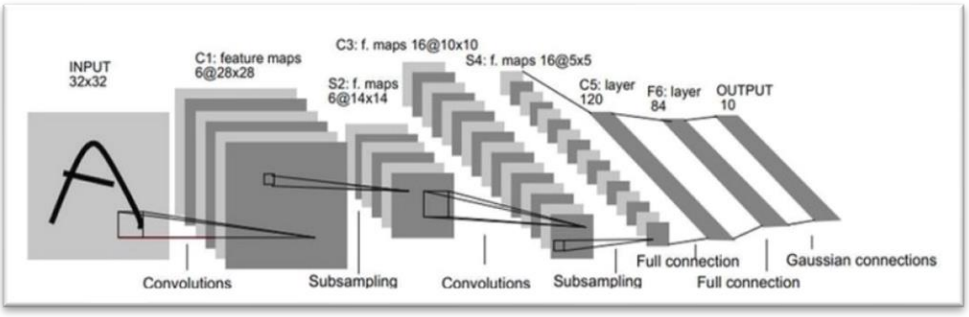
Εικόνα 9. Εφαρμογή Max Pooling

Μετά την εκμάθηση χαρακτηριστικών σε πολλά επίπεδα, η αρχιτεκτονική ενός CNN εμβαθύνει στην ταξινόμηση. Αφού ολοκληρωθούν οι παραπάνω διεργασίες δημιουργείται ένας συγκεντρωτικός χάρτης χαρακτηριστικών. Πλέον τα CNN λειτουργούν ως ένα απλό νευρωνικό δίκτυο. Για αυτό τον λόγο γίνεται η «γραμμικοποίηση» (flattening) του συγκεντρωτικού χάρτη χαρακτηριστικών σε ένα μονοδιάστατο διάνυσμα. Έτσι όλα τα χαρακτηριστικά που προέκυψαν παρουσιάζονται με την μορφή ενός ενιαίου μεγάλου (μονοδιάστατου) διανύσματος δημιουργώντας ένα πλήρως συνδεδεμένο επίπεδο.



Εικόνα 10. «Γραμμικοποίηση» (flattening) του συγκεντρωτικού χάρτη χαρακτηριστικών σε διάνυσμα

Το τελευταίο επίπεδο είναι ένα πλήρως συνδεδεμένο επίπεδο που εξάγει ένα διάνυσμα διαστάσεων K, όπου το K είναι ο αριθμός των κλάσεων που θα μπορεί να προβλέψει το δίκτυο. Κάθε στοιχείο του διανύσματος υποδεικνύει, έτσι, την πιθανότητα να ανήκει η εικόνα εισόδου σε μια κλάση. Με λίγα λόγια, ένα πλήρως συνδεδεμένο επίπεδο καθορίζει τη σχέση μεταξύ της θέσης των χαρακτηριστικών στην εικόνα και μια κλάση. Επειδή ο πίνακας εισόδου είναι το αποτέλεσμα του προηγούμενου επιπέδου, αντιστοιχεί σε κάθε χάρτη χαρακτηριστικών μια δεδομένη λειτουργία. Έτσι οι υψηλές τιμές του χάρτη υποδεικνύουν τη θέση ενός συγκεκριμένου χαρακτηριστικού στην εικόνα με αποτέλεσμα να θεωρείται ένα σημαντικό στοιχείο της εικόνες, οπότε, λαμβάνει και ένα αντίστοιχο βάρος [13]. Είναι προφανές ότι ένα τέτοιο δίκτυο χρησιμοποιείται γενικά για ανίχνευση αντικειμένων πολλών διαφορετικών κλάσεων, συνεπώς μπορεί να χρησιμοποιηθεί και για την αναγνώριση προσώπου. Το τελικό επίπεδο της αρχιτεκτονικής CNN χρησιμοποιεί ένα επίπεδο ταξινόμησης, όπως το Softmax, το οποίο παρέχει την έξοδο ως πιθανότητα για την τελική ταξινόμηση.



Εικόνα 11. Αρχιτεκτονική ενός δικτύου CNN LeNet5

Η συνάρτηση Softmax αποτελεί μια γενίκευση της λογιστικής συνάρτησης σε πολλαπλές διαστάσεις. Χρησιμοποιείται σχεδόν πάντοτε στο τελευταίο επίπεδο των νευρωνικών δικτύων ως η τελευταία λειτουργία ενεργοποίησής τους για την ομαλοποίηση της εξόδου ενός δικτύου σε μια κατανομή πιθανότητας σε σχέση με τις προβλεπόμενες τάξεις εξόδου. Η συνάρτηση Softmax, ως μια μορφή λογικής παλινδρόμησης, λαμβάνει ως είσοδο ένα διάνυσμα z με K στοιχεία. Σημειώνεται ότι τα K στοιχεία περιγράφουν τις κλάσεις προς ταξινόμηση. Στόχος λοιπόν της συνάρτησης Softmax είναι η ομαλοποίηση του διανύσματος σε κατανομή πιθανότητας που αποτελείται από K πιθανότητες ανάλογες με τα εκθετικά των αριθμών εισαγωγής. Έτσι, το τελικό διάνυσμα θα εμπεριέχει τιμές από 0 έως 1 και η κάθε τιμή θα περιγράφει την πιθανότητα (%) της αναγνώρισης του αντικειμένου. Η μαθηματική περιγραφή της συνάρτησης Softmax είναι η εξής:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

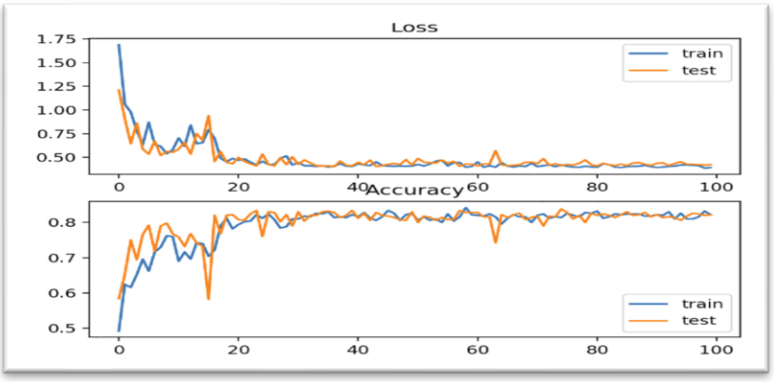
1.2.1.2.2 Εκπαίδευση Μοντέλου

Στην προηγούμενη ενότητα έγινε η περιγραφή της αρχιτεκτονικής ενός συνελικτικού νευρωνικού δικτύου CNN. Η επιτομή όμως της επιτυχίας ενός τέτοιου μοντέλου είναι η εκπαίδευσή του. Κατά την εκπαίδευση το σύστημα μαθαίνει να κατανοεί και να επεξεργάζεται αντίστοιχα τα δεδομένα εισόδου για την δημιουργία μιας αλληλουχίας κανόνων με σκοπό την επίτευξη υψηλών ποσοστών θετικής αναγνώρισης στο τελευταίο βήμα της ταξινόμησης. Η εκπαίδευση ενός τέτοιου συστήματος γίνεται με μια επαναληπτική μέθοδο. Κατά την πρώτη επανάληψη συνέλιξης τα φίλτρα έχουν τυχαίες τιμές (βάρη), οπότε η ταξινόμηση που θα προκύψει προφανώς θα έχει πολλές αστοχίες. Ένα παράδειγμα αστοχίας είναι όταν σε μια εικόνα αναγνωρίζεται πρόσωπο με πιθανότητα 80% ενώ στην πραγματικότητα στην εικόνα δεν υπάρχει κάποιο πρόσωπο. Έτσι το σφάλμα αυτό υπολογίζεται με την εισαγωγή συναρτήσεων κόστους ή απώλειας[2]. Με την χρήση τέτοιων συναρτήσεων το σύστημα μαθαίνει από τα λάθη του. Η διαδικασία της μάθησης των βαθιών νευρωνικών δικτύων[3] γίνεται με αλγορίθμους βελτιστοποίησης Stochastic Gradient Descent. Μια απλούστευση της διαδικασίας αυτής μπορεί να συνοψισθεί σε τρία βήματα:

- Η εκπαίδευση ξεκινά με την θέσπιση τυχαίων βαρών.
- Στην πρώτη επανάληψη υπολογίζεται το σφάλμα μέσω της συνάρτησης απώλειας. Στην συνέχεια τα βάρη μετασχηματίζονται με σκοπό την ελαχιστοποίηση του σφάλματος.
- Η παραπάνω διαδικασία επαναλαμβάνεται κυκλικά έως ότου το σφάλμα ελαχιστοποιηθεί

[2] Στην βιβλιογραφία αναφέρονται ως Loss Functions ή Cost Functions.

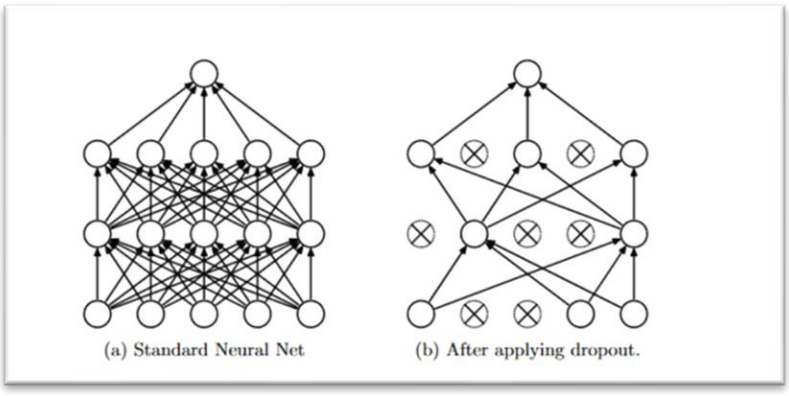
[3] Ο όρος αυτός αναφέρεται στην παρουσία πολλών κρυφών επιπέδων σε ένα νευρωνικό δίκτυο



Εικόνα 12. Διάγραμμα απώλειας και ακρίβειας

Στην παραπάνω εικόνα παρουσιάζονται τα διαγράμματα της συνάρτησης απώλειας και της ακρίβειας του μοντέλου κατά την εκπαίδευση ενός δικτύου CNN[4]. Είναι λογικό ότι με την μείωση του σφάλματος (απώλειας) το σύστημα αυξάνει συνάμα την ακρίβεια του. Είναι αξιοσημείωτο ότι με την μέθοδο SGD το ελάχιστο σφάλμα εντοπίζεται κατά τις πρώτες επαναλήψεις, γεγονός που προσφέρει ταχύτητα στην εκπαίδευση χιλιάδων εικόνων.

Ένα από τα σημαντικότερα προβλήματα που προκύπτουν κατά την διάρκεια της εκπαίδευσης δεδομένων είναι μεγάλη διαστασιμότητα (curse of dimensionality) και η υπερ-εκπαίδευση του συστήματος. Η επίλυση αυτού του ζητήματος γίνεται με την χρήση της τεχνικής απόσυρσης (dropout). Η τεχνική αυτή χρησιμοποιείται για την αποτροπή της υπερ-εκπαίδευσης (overfitting) και παρέχει έναν αποτελεσματικό συνδυασμό εκθετικά πολλών διαφορετικών αρχιτεκτονικών νευρωνικών δικτύων. Ο όρος «dropout» αναφέρεται στην απόσυρση ενός πλήθους νευρώνων (κρυφών και ορατών) σε ένα νευρωνικό δίκτυο. Με την εγκατάλειψη ενός νευρώνα εννοείται η προσωρινή κατάργησή του από το δίκτυο, μαζί με όλες τις εισερχόμενες και εξερχόμενες συνδέσεις του, όπως φαίνεται και στην παρακάτω εικόνα. Η τυχαιότητα περιγράφεται μέσω μίας τυχαίας μεταβλητής η οποία ακολουθεί κατανομή Bernoulli. Στην απλούστερη περίπτωση, κάθε νευρώνας διατηρείται με μια σταθερή πιθανότητα p ανεξάρτητη από άλλες μονάδες, όπου το p μπορεί να επιλεγεί με χρήση ενός συνόλου δεδομένων δοκιμής ή μπορεί απλώς να οριστεί στο 0.5, το οποίο φαίνεται να είναι σχεδόν βέλτιστο για ένα ευρύ φάσμα δικτύων και εργασιών. Ωστόσο, για τις μονάδες εισόδου η βέλτιστη πιθανότητα διατήρησης είναι συνήθως πιο κοντά στο 1 παρά στο 0.5.

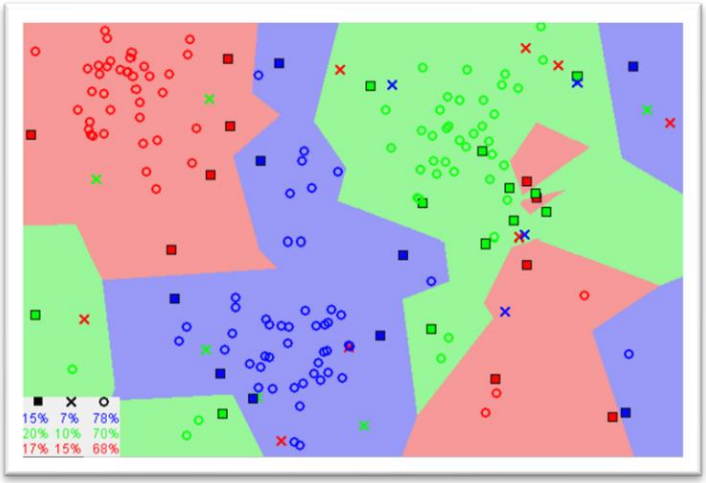


Εικόνα 13. Μοντέλο Dropout νευρωνικού δικτύου. Αριστερά: ένα τυπικό νευρωνικό δίκτυο με δύο κρυφά επίπεδα. Δεξιά: παράδειγμα αραιωμένου δικτύου που παράχθηκε με την εφαρμογή dropout στο δίκτυο στα αριστερά.

[4] Για λεπτομερέστερη περιγραφή των CNN προτείνεται η επίσκεψη στην ιστοσελίδα https://cs231n.github.io/

1.2.2 Αλγόριθμος Κ εγγύτερων γειτόνων (K nearest neighbors)

Σε μια αδημοσίευτη έκθεση της Ιατρικής Αεροπορίας των ΗΠΑ για την Πολεμική Αεροπορία το 1951, οι Fix & Hodges παρουσίασαν μια μη παραμετρική μέθοδο για την ταξινόμηση μοτίβων που έκτοτε έγινε γνωστή ως κανόνας Κ-εγγύτερων γειτόνων [14]. Η ταξινόμηση Κ-εγγύτερων γειτόνων αναπτύχθηκε για τις ανάγκες της διακριτικής ανάλυσης όταν οι αξιόπιστες παραμετρικές εκτιμήσεις της πυκνότητας πιθανότητας είναι άγνωστες ή δύσκολο να προσδιοριστούν. Η ταξινόμηση K-NN είναι μια από τις πιο θεμελιώδεις και απλές μεθόδους ταξινόμησης, και συνήθως χρησιμοποιείται ως μια από τις πρώτες επιλογές για μια εργασία ταξινόμησης όταν υπάρχει περιορισμένη ή καθόλου προηγούμενη γνώση σχετικά με τη κατανομή των δεδομένων. Ο αλγόριθμος KNN υποθέτει ότι παρόμοια πράγματα βρίσκονται πολύ κοντά στον χώρο προβολής τους.

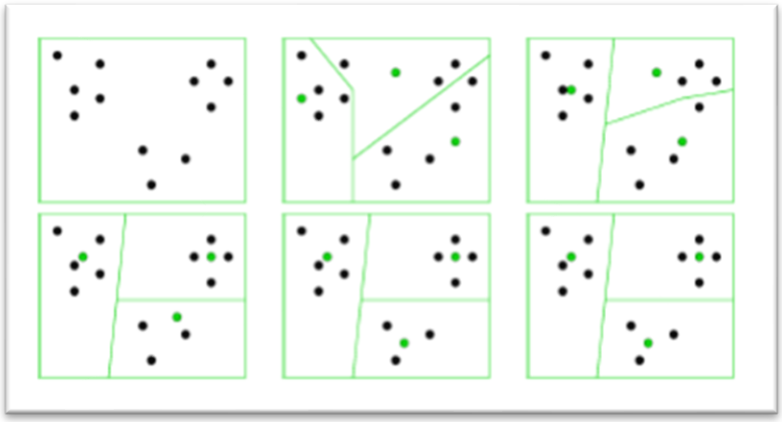


Εικόνα 14. Ταξινόμηση δεδομένων βάσει της ομοιότητάς τους

1.2.3 K-means ομαδοποίηση (clustering)

Ο αλγόριθμος K-means προτάθηκε από τον MacQueen το 1967 [15] και αποτελεί έναν αλγόριθμο μη εποπτευόμενης μάθησης καθώς το μοντέλο μπορεί να εντοπίσει αυτόματα ποια είναι εκείνα τα χαρακτηριστικά που διαφοροποιούν κάθε κλάση. Αποτελεί έναν από τους απλούστερους αλγορίθμους μη εποπτευόμενης μάθησης που επιλύουν ένα πρόβλημα ομαδοποίησης. Η διαδικασία ακολουθεί έναν απλό και εύκολο τρόπο για την ταξινόμηση ενός συνόλου δεδομένων μέσω ενός συγκεκριμένου αριθμού k συστάδων ο οποίος ορίζεται εκ των προτέρων. Η κύρια ιδέα του αλγορίθμου είναι να οριστούν k κεντροειδή, ένα για κάθε συστάδα που θα δημιουργηθεί. Μπορούμε να πούμε πως στόχος του κεντροειδούς είναι να αποτελεί το κέντρο βάρους της συστάδας. Αρχικά αυτά τα k κεντροειδή τοποθετούνται τυχαία και συνήθως όσο το δυνατόν πιο μακριά το ένα από το άλλο. Με υπολογισμό των αποστάσεων του κάθε χαρακτηριστικού στα δεδομένα εισόδου από το κεντροειδές δημιουργείται μια συστάδα χαρακτηριστικών που βρίσκεται πλησιέστερα στο κεντροειδές. Στην συνέχεια το κεντροειδές τοποθετείται στο κέντρο βάρους των συστάδων που δημιουργούνται. Η διαδικασία αυτή επαναλαμβάνεται έως ότου τα κεντροειδή να σταματήσουν να μεταβάλλονται. Στην παρακάτω εξίσωση, το m_i εκφράζει το κεντροειδές της συστάδας C_i, ενώ το $d(x, m_i)$ εκφράζει την Ευκλείδεια απόσταση μεταξύ των χαρακτηριστικών και του κεντροειδούς. Στόχος του αλγορίθμου είναι η ελαχιστοποίηση της απόστασης Ε:

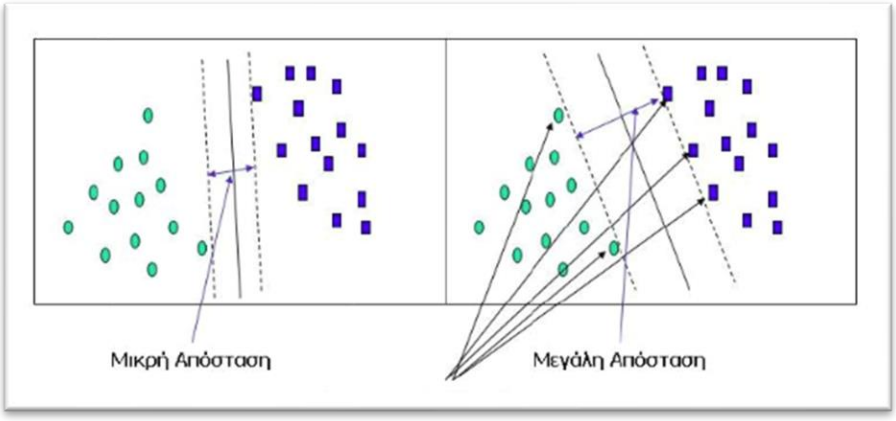
$$E = \sum_{i=1}^{c} \sum_{x \in C} d(x, m_i)$$



Εικόνα 15. Παράδειγμα ομαδοποίησης με 4 επαναλήψεις

1.2.4 Μηχανές Διανυσμάτων Υποστήριξης

Οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines – SVM) [16] [17] ανήκουν σε μια ομάδα αλγορίθμων εποπτευόμενης μάθησης και χρησιμοποιούνται ευρέως για εφαρμογές ταξινόμησης αλλά και παλινδρόμησης. Στόχος των SVM είναι ο υπολογισμός ενός βέλτιστου διαχωριστικού υπερεπιπέδου (separating hyperplane) ανάμεσα στα δεδομένα με μεγιστοποίηση της απόστασής του από τις κοντινότερες παρατηρήσεις του συνόλου εκπαίδευσης, η εύρεση δηλαδή του λεγόμενου υπερεπιπέδου μεγίστου περιθωρίου (max margin hyperplane). Όπως περιγράφεται στην παρακάτω εικόνα, στις δύο διαστάσεις το υπερεπίπεδο αυτό εμφανίζεται ως μια ευθεία, στις τρεις διαστάσεις ως ένα επίπεδο, ενώ σε περισσότερες διαστάσεις ως υπερεπίπεδο. Επομένως ένα μοντέλο SVM είναι μια αναπαράσταση του training set ως σημείων στο χώρο, τα οποία χαρτογραφούνται ώστε να χωρίζονται από κατά το δυνατόν σαφέστερο και ευρύτερο περιθώριο (margin).



Εικόνα 16. Ταξινόμηση βάσει μηχανών διανυσμάτων υποστήριξης

Κεφάλαιο 2. Ανίχνευση Προσώπων

Η ανίχνευση προσώπου είναι η διαδικασία της ταξινόμησης/διάκρισης προσώπων από μη πρόσωπα που υπάρχουν σε μια εικόνα εισαγωγής και μπορεί να θεωρηθεί ως μια ειδική περίπτωση ανίχνευσης κατηγορίας αντικειμένων η οποία ασχολείται με την εύρεση της θέσης και του μεγέθους διαφορετικών αντικειμένων που εμπίπτουν σε μια δεδομένη κατηγορία. Η ανίχνευση προσώπου αποτελεί ένα απαραίτητο και πολλές φορές πρώτο βήμα σε αλγορίθμους αναγνώρισης προσώπων καθώς σκοπός της είναι ο εντοπισμός και ο διαχωρισμός της περιοχής του προσώπου από το υπόβαθρο. Η βιβλιογραφία αναφέρει πληθώρα μεθόδων και τεχνικών οι οποίες κατατάσσονται σε δύο κύριες κατηγορίες βάσει της προσέγγισής τους στην ανίχνευση των χαρακτηριστικών του προσώπου. Πρόκειται για τις προσεγγίσεις βάσει χαρακτηριστικών (Feature Based Approach) και βάσει της εμφάνισης ή εικόνας (Image Based Approach).

Οι μέθοδοι της πρώτης κατηγορίας έχουν στόχο την εξαγωγή χαρακτηριστικών μιας δοθείσας εικόνας και στην συνέχεια την αντιστοίχισή τους με μοντέλα που δημιουργούνται με βάση την γνώση για τα χαρακτηριστικά του προσώπου. Αντίθετα η προσέγγιση με βάση την εμφάνιση προσπαθεί να πραγματοποιήσει την καλύτερη αντιστοίχιση-ταξινόμηση μεταξύ των εικόνων εκπαίδευσης (training Images) με τις αντίστοιχες κατηγορίες. Η τελευταία προσέγγιση έγκειται σε μεθόδους μηχανικής μάθησης καθώς ένα σύστημα εκπαιδεύεται μέσω παραδειγμάτων για να επιτευχθεί η τελική ταξινόμηση. Η ανίχνευση προσώπου θεωρείται ως ένα πρόβλημα ταξινόμησης καθώς στόχος του συστήματος είναι να αναγνωρίσει πρότυπα σε μια εικόνα ώστε να τα ταξινομήσει στην αντίστοιχη κατηγορία. Το ανθρώπινο πρόσωπο είναι ένα δυναμικό αντικείμενο και έχει υψηλό βαθμό μεταβλητότητας στην εμφάνισή του, γεγονός που καθιστά την ανίχνευση προσώπου ένα γενικά δύσκολο πρόβλημα για το επιστημονικό πεδίο της όρασης υπολογιστών.

Η ανίχνευση προσώπου αποτελεί το πρώτο στάδιο κάθε αυτόματου συστήματος αναγνώρισης προσώπου αφού για να αναγνωριστεί ένα πρόσωπο σε μια εικόνα θα πρέπει πρώτα να εντοπιστεί η θέση του. Συχνά η ανίχνευση προσώπου αναφέρεται ως διαδικασία η οποία ολοκληρώνεται σε δυο φάσεις. Αρχικά η εικόνα εξετάζεται προκειμένου να βρεθούν περιοχές που ανιχνεύονται ως πρόσωπα (detection) και, αφού εκτιμηθεί κατά προσέγγιση η θέση και το μέγεθος ενός προσώπου, ακολουθεί μια διαδικασία εντοπισμού (localization) η οποία παρέχει ακριβέστερη εκτίμηση της πραγματικής θέσης και κλίμακας του προσώπου. Δηλαδή, ενώ ο στόχος της ανίχνευσης είναι να βρεθούν κατά προσέγγιση όλα τα πρόσωπα σε εικόνες με πολλά πρόσωπα και με σύνθετο υπόβαθρο, η διαδικασία του εντοπισμού (localization) δίνει έμφαση στη χωρική ακρίβεια η οποία επιτυγχάνεται συνήθως με την ανίχνευση συγκεκριμένων χαρακτηριστικών του προσώπου. Όπως προαναφέρθηκε, οι μέθοδοι ανίχνευσης προσώπου μπορούν να οργανωθούν αποτελεσματικά σε δύο ευρείες κατηγορίες που ξεχωρίζουν από την διαφορετική προσέγγιση στη χρήση της γνώσης του προσώπου.

Οι τεχνικές εντοπισμού βάσει χαρακτηριστικών εφαρμόζονται μετά από την ανίχνευση χαρακτηριστικών προσώπου σε μια εικόνα και λειτουργούν ως τεχνικές επαλήθευσης για την μείωση της ψευδούς ανίχνευσης. Αυτές οι μέθοδοι κάνουν σαφή χρήση της γνώσης του προσώπου και ακολουθούν την κλασική μεθοδολογία εντοπισμού στην οποία τα χαρακτηριστικά χαμηλού επιπέδου (ακμές, χρώμα, ένταση εικόνας) προέρχονται από ανάλυση βασισμένη σε προϋπάρχουσα γνώση-μοντέλα χρησιμοποιώντας ορισμένους προκαθορισμένους κανόνες προσδιορισμού προσώπου. Αυτοί οι κανόνες αφορούν τα χαρακτηριστικά (συστατικά του προσώπου) και τη σχέση μεταξύ τους. Βάσει αυτών των κανόνων, πρώτα αναζητούνται τα στοιχεία προσώπου στη δεδομένη εικόνα εισαγωγής και στη συνέχεια προσδιορίζονται τα υποψήφια πρόσωπα. Αφού εφαρμοστούν οι κανόνες ανίχνευσης και εντοπιστεί το υποψήφιο πρόσωπο, τότε γίνεται επαλήθευση για την απόρριψη ψευδούς εντοπισμού.

Αυτή η επαλήθευση βασίζεται σε κανόνες που εξάγονται με βάση την ανθρώπινη γνώση σχετικά με την τυπική γεωμετρία του προσώπου και τις θέσεις των χαρακτηριστικών του αλλά και το χρώμα του. Δεδομένου ότι το ανθρώπινο πρόσωπο παρουσιάζει μία εγγενή συμμετρία και τα κύρια χαρακτηριστικά του εμφανίζονται με μία φυσική αλληλουχία από πάνω προς τα κάτω και από αριστερά προς τα δεξιά, είναι δυνατό να ορίσει κανείς κανόνες που περιγράφουν το σχήμα, το μέγεθος, την υφή και άλλα παρόμοια γνωρίσματα χαρακτηριστικών του προσώπου αλλά και να περιγράφει τις μεταξύ τους χωρικές σχέσεις. Οι σχέσεις αυτές αναφέρονται κυρίως στις σχετικές θέσεις, αποστάσεις και την μεταβλητότητα της έντασης των χαρακτηριστικών. Για παράδειγμα, μια τυπική εικόνα προσώπου θα έχει δύο μάτια συμμετρικά μεταξύ τους, ένα στόμα και μια μύτη. Ένα άλλο παράδειγμα είναι ότι το κεντρικό τμήμα του προσώπου θα έχει τιμές έντασης που είναι ομοιόμορφες και λαμβάνεται η μέση διαφορά μεταξύ των τιμών έντασης του άνω μερους και του κεντρικού τμήματος.

Συνήθως υιοθετείται μία ιεραρχική προσέγγιση που εξετάζει την εικόνα του προσώπου σε διαφορετικά επίπεδα ανάλυσης. Στα υψηλότερα επίπεδα οι υποψήφιες περιοχές ανιχνεύονται με βάση μία χονδρική περιγραφή της γεωμετρίας του προσώπου. Σε χαμηλότερα επίπεδα, εξάγονται χαρακτηριστικά του προσώπου όπως εκείνα που αναφέρθηκαν παραπάνω, και μία περιοχή χαρακτηρίζεται ως "πρόσωπο" ή "μη-πρόσωπο" βάσει προκαθορισμένων κανόνων για τα χαρακτηριστικά του προσώπου και τις σχετικές τους θέσεις.

Οι τεχνικές ανίχνευσης βάσει εμφάνισης, αντί για προκαθορισμένους κανόνες, χρησιμοποιούν προκαθορισμένα πρότυπα-μοτίβα προσώπου που ταιριάζουν με τα τμήματα στην εικόνα εισόδου και, μέσω μιας ταξινόμησης ή σύγκρισης, καθορίζεται η παρουσία του προσώπου. Αυτή η προσέγγιση χρησιμοποιεί ένα μεγάλο πλήθος δεδομένων εκπαίδευσης διαφορετικών ατόμων υπό διάφορες συνθήκες (έκφρασης, έντασης ή πόζας κ.α.) ώστε να εκπαιδευτούν ταξινομιτές (classifiers). Η ανίχνευση γίνεται με μετασχηματισμό-mapping και ταξινόμηση. Σε αντίθεση με την πρώτη κατηγορία, σε αυτές τις τεχνικές δεν χρησιμοποιούνται τα χαρακτηριστικά και οι σχέσεις που προκύπτουν από την γνώση του προσώπου, όμως ενσωματώνουν έμμεσα την γνώση του προσώπου μέσω της εκπαίδευσης και ταξινόμησης. Η προσέγγιση βάσει εμφάνισης θεωρεί την ανίχνευση προσώπου κυρίως ως πρόβλημα αναγνώρισης μοτίβου. Τα μοτίβα προσώπου αναγνωρίζονται μέσω μιας διαδικασίας εκπαίδευσης που ταξινομεί τα συλλεγμένα μεγάλα παραδείγματα σε κατηγορίες προσώπου και μη προσώπου. Πρόκειται για μια μαθησιακή προσέγγιση (learning-based) όπου η μάθηση χρησιμοποιείται για την αναγνώριση ενός μοτίβου προσώπου από παραδείγματα. Λέγοντας παραδείγματα εννοούμε τα δεδομένα εκπαίδευσης (training sets). Τα δεδομένα εκπαίδευσης περιέχουν θετικά και αρνητικά παραδείγματα, δηλαδή οι εικόνες που περιέχουν πρόσωπο είναι θετικά παραδείγματα και αυτές που δεν περιέχουν πρόσωπα είναι αρνητικά. Αυτή η προσέγγιση απαιτεί μεγάλο αριθμό παραδειγμάτων και χρόνο για την εκπαίδευση των ταξινομητών, όμως τέτοιοι αλγόριθμοι είναι ισχυροί, γρήγοροι και αποτελεσματικοί καθώς μπορούν να ανιχνεύουν πρόσωπα σε διαφορετική στάση και προσανατολισμό εξαλείφοντας το σφάλμα μοντελοποίησης λόγω ανακριβούς «γνώσης» του προσώπου. Εφόσον γίνει η εκπαίδευση των ταξινομητών, αυτοί μπορούν να εφαρμοσθούν σε εφαρμογές πραγματικού χρόνου (real-time applications).

2.1 Προκλήσεις της ανίχνευσης προσώπου

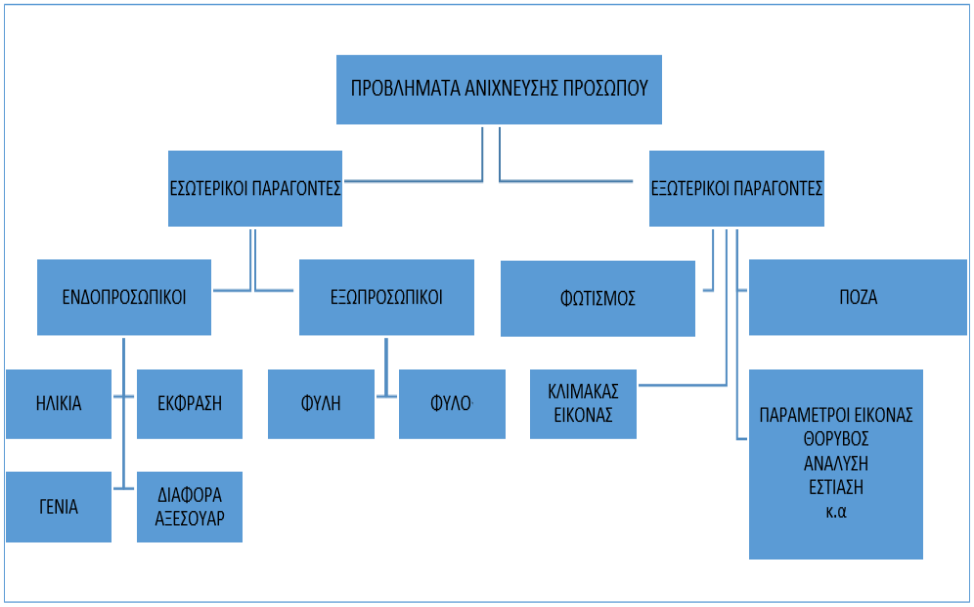
Η ανίχνευση προσώπου σε έναν δισδιάστατο χώρο όπως είναι μια εικόνα αντιμετωπίζει πολλές προκλήσεις. Τα πρόσωπα των ανθρώπων παρουσιάζουν, όπως προαναφέρθηκε, σημαντικές διαφορές και μεγάλη διακύμανση εμφάνισης που μπορεί να οφείλονται στα χαρακτηριστικά των ανθρώπων τα οποία κατατάσσονται κυρίως σε δύο κύριες κατηγορίες, τους εσωτερικούς και τους εξωτερικούς παράγοντες.

Οι εσωτερικοί παράγοντες σχετίζονται με τα χαρακτηριστικά των προσώπων όπως είναι τα μάτια ή μύτη, τα χείλη κ.ά. Οι παράγοντες αυτοί μπορούν να ταξινομηθούν σε δυο βασικές κατηγορίες, τους ενδοπροσωπικούς παράγοντες και τους εξωπροσωπικούς.

Στους ενδοπροσωπικούς παράγοντες κατατάσσονται εκείνοι οι παράγοντες που επηρεάζουν την διακύμανση της εμφάνισης του προσώπου του ιδίου του ατόμου, όπως είναι η ηλικία του ή έκφρασή του (θυμωμένη ή χαρούμενη), τα γένια ή και τα διάφορα αξεσουάρ που μπορεί να υπάρχουν σε ένα πρόσωπο, όπως γυαλιά, καπέλο κ.ά. Στους εξωπροσωπικούς παράγοντες εντάσσονται παράγοντες που αφορούν γενικότερες αλλαγές στην διακύμανση της εμφάνισης των ανθρώπων τα οποία κυρίως προκύπτουν από την φυλή (γεωμετρία προσώπου, χρώμα, σχήμα ματιών κ.ά.) και το φύλο.

Οι εξωτερικοί παράγοντες σχετίζονται με τις συνθήκες λήψης της εικόνας όπως αυτές επηρεάζουν την εμφάνιση του ανθρώπου. Σε αυτούς του παράγοντες περιλαμβάνονται:

- Η κλίμακα της εικόνας (Image Scale). Σε μια εικόνα μπορεί να βρίσκονται ένα ή περισσότερα πρόσωπα σε διαφορετικό βάθος, με αποτέλεσμα αυτά να έχουν διαφορετικό ύψος και πλάτος.
- Η στάση του προσώπου (Pose) όπως προβάλλεται στην εικόνα μπορεί επίσης να δημιουργήσει σύγχυση στους αλγόριθμους διότι ενδέχεται να αποκρύπτονται σημαντικά στοιχεία όπως μάτια, στόμα κ.ά καθώς να αλλάζει και η γεωμετρία του προσώπου.
- Ο θόρυβος της εικόνας (Image Noise). Ο θόρυβος της εικόνας μπορεί να προκληθεί είτε από το περιβάλλον είτε από τα χαρακτηριστικά της μηχανής λήψης.
- Η στροφή (Rotation): Τα πρόσωπα σε μια εικόνα μπορεί να έχουν διαφορετικές στροφές
- Ο φωτισμός(Illumination Variation): Σε μια εικόνα διάφορα πρόσωπα μπορεί να έχουν διαφορετικό φωτισμό.
- Εμπόδια (Occlusion): Σε μια φωτογραφία τα πρόσωπα μπορεί εν μέρει να αποκρύπτονται.
- Χαμηλή ποιότητα εικόνας, χαμηλή ανάλυση(Low Image quality – resolution).



Εικόνα 17. Προκλήσεις στην ανίχνευση προσώπου

2.2 Ανίχνευση προσώπου με βάση τα χαρακτηριστικά

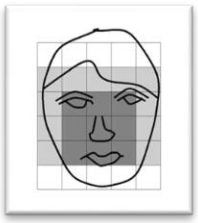
Σε αυτήν την προσέγγιση, οι μέθοδοι ανίχνευσης προσώπων αναπτύσσονται με βάση τους κανόνες που προέρχονται από τη γνώση του ερευνητή για τα ανθρώπινα πρόσωπα. Ακόμη και αν θεωρείται εύκολο να επινοήσουμε απλούς κανόνες για να περιγράψουμε τα χαρακτηριστικά ενός προσώπου και τις σχέσεις τους, αποδεικνύεται ότι πρόκειται για μια πολύ δύσκολη εργασία λόγω της ποικιλομορφίας του προσώπου και των

διαφορετικών χαρακτηριστικών που έχουν οι άνθρωποι. Όμως, όπως λέχθηκε, μπορούμε να πούμε πως ένα πρόσωπο εμφανίζεται κατά κανόνα σε μια εικόνα με δύο μάτια που είναι συμμετρικά μεταξύ τους, μια μύτη και ένα στόμα. Οι σχέσεις μεταξύ των χαρακτηριστικών μπορούν να αναπαρασταθούν από τις σχετικές αποστάσεις και τις θέσεις τους. Αρχικά εξάγονται τα χαρακτηριστικά του προσώπου σε μια εικόνα εισαγωγής και τα υποψήφια πρόσωπα αναγνωρίζονται βάσει των κωδικοποιημένων κανόνων. Στη συνέχεια εφαρμόζεται μια διαδικασία επαλήθευσης για τη μείωση ψευδών εντοπισμών. Το πρόβλημα όμως σε αυτήν την προσέγγιση είναι η δυσκολία στη μετάφραση της ανθρώπινης γνώσης σε καλά καθορισμένους κανόνες καθώς το αν οι κανόνες είναι λεπτομερείς (δηλ. αυστηροί), οπότε ενδέχεται να μην εντοπίζονται πρόσωπα που δεν πληρούν όλους τους κανόνες. Εάν οι κανόνες είναι πολύ γενικοί, ενδέχεται να δώσουν πολλά ψευδώς θετικά πρόσωπα. Επιπλέον, είναι δύσκολο να επεκταθεί αυτή η προσέγγιση για τον εντοπισμό προσώπων σε διαφορετικές πόζες, καθώς δεν είναι απλό να απαριθμηθούν όλες οι πιθανές περιπτώσεις. Από την άλλη πλευρά, αυτή η προσέγγιση εμφανίζει ικανοποιητικά αποτελέσματα στην ανίχνευση μετωπικά απεικονιζόμενων προσώπων.

Οι Yang & Huang χρησιμοποίησαν μια ιεραρχική μέθοδο με βάση τη γνώση για την ανίχνευση προσώπων [18]. Το σύστημά τους αποτελείται από τρία επίπεδα κανόνων. Στο υψηλότερο επίπεδο, τα υποψήφια πρόσωπα ανιχνεύονται με την σάρωση ενός παραθύρου διαφορετικών μεγεθών καθώς εφαρμόζονται ένα σύνολο κανόνων σε κάθε θέση. Οι κανόνες σε αυτό το επίπεδο είναι γενικοί περιγραφείς προσώπου, όπως είναι το σχήμα του, ενώ στα ανώτερα επίπεδα οι κανόνες σχετίζονται με την περιγραφή των χαρακτηριστικών του προσώπου, όπως μάτια, μύτη κ.ά. Έτσι δημιουργείται μια ιεραρχία εικόνων και κανόνων.

Εικόνα 18. Εικόνες που προκύπτουν από την σάρωση παραθύρων (4x4,8x8,16x16). Το κάθε εικονοστοιχείο της νέας εικόνας που δημιουργείται περιγράφει τον μέσο όρο του παραθύρου σάρωσης.

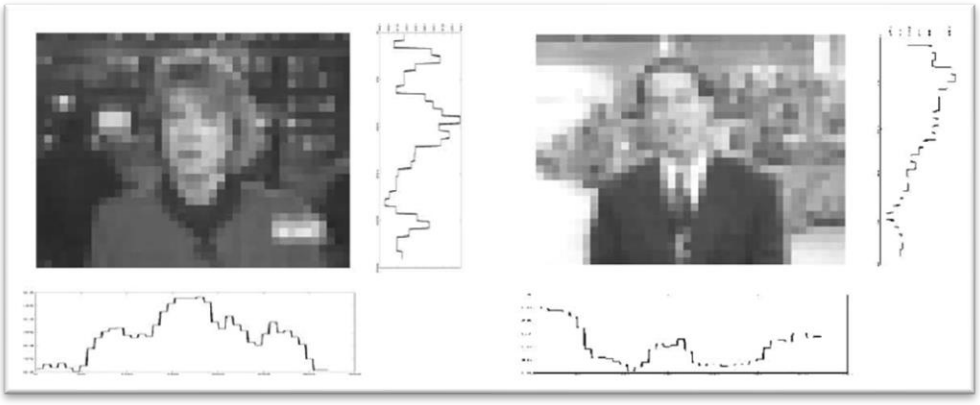
Στη συνέχεια εφαρμόζονται κωδικοποιημένοι κανόνες για τον εντοπισμό των υποψήφιων προσώπων στην εικόνα με την χαμηλότερη ανάλυση, το οποίο περιλαμβάνει το κεντρικό τμήμα του προσώπου.



Εικόνα 19. Ένα τυπικό πρόσωπο με προσέγγιση βάσει της γνώσης. Οι κανόνες κωδικοποιούνται βάσει της γνώσης για τα χαρακτηριστικά του προσώπου (ένταση, διασπορά, διαφορά) των περιοχών του προσώπου

Στην παραπάνω εικόνα παρατηρούμε πως το κέντρο του προσώπου και οι περιοχές γύρω από αυτό παρουσιάζουν αντίστοιχα μια ομοιομορφία στην ένταση, με αποτέλεσμα η διαφορά μεταξύ των περιοχών αυτών να είναι αξιοσημείωτη. Η ανάλυση αυτή συμβαίνει στο πρώτο επίπεδο, δηλαδή από την εικόνα με την χαμηλότερη ανάλυση προκύπτουν οι υποψήφιες περιοχές προσώπου. Στην συνέχεια στο δεύτερο επίπεδο εφαρμόζεται επεξεργασία τοπικού ιστογράμματος στις περιοχές που έχουν προκύψει από το

προηγούμενο επίπεδο μαζί με την ανίχνευση ακμών στα υποψήφια πρόσωπα. Τέλος, στο τρίτο επίπεδο τα εναπομένοντα υποψήφια πρόσωπα εξετάζονται με ένα άλλο σύνολο κωδικοποιημένων κανόνων που σχετίζονται με τα χαρακτηριστικά του προσώπου.



Εικόνα 20 Επεξεργασία τοπικού ιστογράμματος. Παρατηρείται πως στης περιοχές του προσώπου παρουσιάζεται τοπικό μέγιστο.

Μια άλλη προσέγγιση ανίχνευσης προσώπου με βάση τα χαρακτηριστικά στηρίζεται στην παρατήρηση ότι οι άνθρωποι μπορούν εύκολα να εντοπίσουν πρόσωπα και αντικείμενα σε διαφορετικές πόζες και συνθήκες φωτισμού και, επομένως, πρέπει να υπάρχουν ιδιότητες ή χαρακτηριστικά που είναι αμετάβλητα έναντι αυτών των μεταβολών. Έτσι, σε αντίθεση με την προηγούμενη προσέγγιση η οποία βασίζεται στην δημιουργία γενικών κανόνων προσώπου που προκύπτουν από την αντίληψη του ερευνητή για το πρόσωπο, αυτή η προσέγγιση αναζητεί την εύρεση αμετάβλητων χαρακτηριστικών του προσώπου (facial features). Οι μέθοδοι που βασίζονται σε αυτή την προσέγγιση εξασφαλίζουν πρώτα την ανίχνευση χαρακτηριστικών του προσώπου όπως είναι τα μάτια, η μύτη, το στόμα και τα μαλλιά με τον εντοπισμό ακμών[5]. Σημειώνεται ότι η ανίχνευση ακμών αποτελεί από την αρχή της επιστήμης της όρασης υπολογιστών μια σημαντική διαδικασία καθώς έχει πολλές εφαρμογές στην ανίχνευση και τον εντοπισμό προσώπων. Είναι λογικό πως μια τέτοια προσέγγιση, σε συνδυασμό με άλλες προσεγγίσεις (όπως χρώματος), μπορεί να επιτύχει σημαντικά αποτελέσματα. Τα αντικείμενα όπως και τα πρόσωπα διακρίνονται από τα μοναδικά χαρακτηριστικά τους. Τα μάτια, τα αυτιά, τα φρύδια, η μύτη, τα χείλη, οι τρίχες, το μέτωπο, το χρώμα του δέρματος κ.λπ. αποτελούν τα κύρια χαρακτηριστικά των προσώπων. Έτσι αντί να χρησιμοποιηθεί ένα μόνο πρότυπο για την αναπαράσταση ολόκληρου του προσώπου, αυτή η μεθοδολογία χρησιμοποιεί πρότυπα για αναζήτηση κάθε μεμονωμένου χαρακτηριστικού του προσώπου και μια περιοχή στην εικόνα δηλώνεται ως πρόσωπο εάν εντοπιστεί ένας κατάλληλος συνδυασμός χαρακτηριστικών προσώπου. Αυτή η προσέγγιση ανίχνευσης μεμονωμένων χαρακτηριστικών προσώπου (facial features) αντί για ολόκληρο το πρόσωπο καθιστά δυνατή την ανίχνευση όχι μόνο προσώπων σε μετωπικές λήψεις αλλά και προσώπων με ελαφρές στροφές. Αλλά για την ανίχνευση εσωτερικών χαρακτηριστικών του προσώπου, τα πρόσωπα πρέπει να έχουν μεγαλύτερο μέγεθος στην εικόνα. Επομένως, χαμηλής ανάλυσης εικόνες, πχ. 50x50, είναι ακατάλληλες για ανίχνευση προσώπων [19].

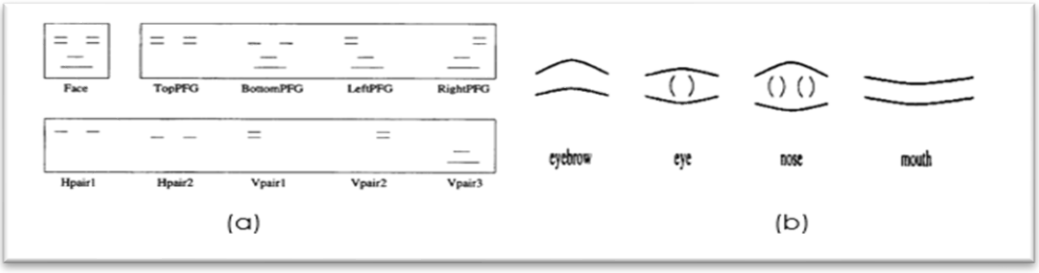
Ο Sirohey πρότεινε μια μεθοδολογία εντοπισμού για την κατάτμηση ενός προσώπου από ένα γεμάτο και σύνθετο φόντο [20]. Σύμφωνα με αυτή την μέθοδο χρησιμοποιείται ένας χάρτης ακμών (ανιχνευτής Canny [21]) και ένα σύνολο κωδικοποιημένων κανόνων για να αφαιρεθούν περιοχές μη ενδιαφέροντος, και στην συνέχεια οι ακμές ομαδοποιούνται έτσι ώστε να διατηρούνται μόνο αυτές εντός του περιγράμματος του προσώπου.

[5] Οι ακμές θεωρούνται στην επιστήμη της υπολογιστικής όρασης ως το πιο παλαιό χαρακτηριστικό. Ο τελεστής ή φίλτρο Sobel (Sobel operator filter) [15], [16], το φίλτρο Maar-Hilreth [17] και ο αλγόριθμος Canny Edge Detector [17] [18] είναι μερικά από τα πιο συνήθη φίλτρα ανίχνευσης ακμών που συναντά κανείς στην βιβλιογραφία.

Στη συνέχεια, προσαρμόζεται μια έλλειψη στο όριο μεταξύ της περιοχής του κεφαλιού και του φόντου. Αυτός ο αλγόριθμος επιτυγχάνει 80% ακρίβεια σε μια βάση δεδομένων 48 εικόνων με «ακατάστατο» φόντο.

Μια άλλη διαφορετική προσέγγιση για την εξαγωγή αντίστοιχων χαρακτηριστικών εφαρμόστηκε από τον Graf [22]. Η μέθοδος αυτή χρησιμοποιεί ένα φίλτρο διέλευσης ζώνης (band pass) και διάφορες μορφολογικές πράξεις σε εικόνες κλίμακας grayscale για την ανάδειξη περιοχών με μεγάλη ένταση (intensity) οι οποίες έχουν συγκεκριμένο σχήμα (πχ. μάτια). Τότε, με εφαρμογή ενός ιστογράμματος μπορούν να εξαχθούν διακεκριμένα ακρότατα. Με βάση αυτά τα ακρότατα και το πλάτος τους χρησιμοποιούνται κατάλληλα κατώφλια και εξάγονται δυαδικές εικόνες με τα χαρακτηριστικά του προσώπου. Τέλος, οι συνδυασμοί των χαρακτηριστικών, μαζί με ένα σύνολο ταξινομητών, καθορίζουν την παρουσία ή όχι του προσώπου.

Μια άλλη αξιοσημείωτη εργασία ανίχνευσης προσώπου βάσει χαρακτηριστικών προσώπου προτείνεται από τους Yow & Cipolla [23]. Παρουσιάζουν έναν αρκετά γενικό αλγόριθμο ο οποίος ανιχνεύει σημεία του προσώπου εφαρμόζοντας χωρικά φίλτρα. Τα σημεία που προκύπτουν ομαδοποιούνται με την χρήση περιορισμών (γεωμετρικών και επιπέδων του γκρι). Αρχικά μοντελοποιείται το πρόσωπο ως ένα επίπεδο με 6 προσανατολισμένα χαρακτηριστικά προσώπου, πιο συγκεκριμένα τα φρύδια, τα μάτια, την μύτη και το στόμα (Εικόνα 21a). Σημειώνουν πως σε περιοχές στα μάγουλα, δηλαδή κάτω από τα μάτια και δεξιά και αριστερά από την μύτη, δεν θα πρέπει να ανιχνευθούν ακμές (edge & feature free). Η χρήση ενός τέτοιου μοντέλου με προσανατολισμένα χαρακτηριστικά έχει το πλεονέκτημα πως εκτείνεται σε μεγάλη περιοχή του προσώπου κάνοντας έτσι την ανίχνευση πιο αξιόπιστη και ισχυρή. Λόγω πιθανών εμποδίων που πιθανώς να αποκρύπτουν το πρόσωπο, αναλύουν το παραπάνω μοντέλο σε συστατικά τα οποία αποτελούνται από 4 χαρακτηριστικά τα οποία λέγονται Partial Face Groups (PGG). Για μια πιο αξιόπιστη ανίχνευση χαρακτηριστικών, θα πρέπει να χρησιμοποιηθούν μόνο χαρακτηριστικά που είναι αμετάβλητα σε αλλαγές κλίμακας και έντασης. Έχει παρατηρηθεί ωστόσο πως σε εικόνες χαμηλής ανάλυσης τα έξι παραπάνω χαρακτηριστικά εμφανίζονται ως σκούρες κηλίδες σε ένα πιο φωτεινό φόντο (του προσώπου). Έτσι, δεδομένου ότι οι ακμές είναι αναλλοίωτες σε μεγάλο βαθμό, τα παραπάνω έξι χαρακτηριστικά του προσώπου μοντελοποιούνται ως ζεύγη προσανατολισμένων ακμών.

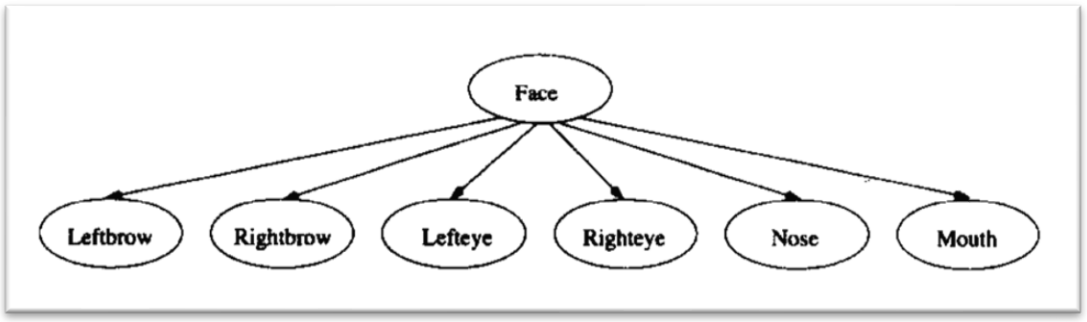


Εικόνα 21. Μοντελοποίηση Προσώπου

Όπως περιγράφεται στην επόμενη εικόνα, στο πρώτο στάδιο γίνεται η εξαγωγή σημείων ενδιαφέροντος. Τα σημεία ενδιαφέροντος εξάγονται από τα τοπικά μέγιστα που προκύπτουν με εφαρμογή φίλτρο δευτέρας παραγώγου Gauss (second derivative Gaussian filter) στην αρχική εικόνα. Στην συνέχεια, αφού εξαχθούν οι ακμές, εξετάζονται οι περιοχές γύρω από τα σημεία ενδιαφέροντος και ομαδοποιούνται σε περιοχές βάσει του προσανατολισμού και της ομοιότητας ως προς τον προσανατολισμό και την εγγύτητα με τα παραπάνω μοντέλα.

Εικόνα 22. Διαδικασία ανίχνευσης χαρακτηριστικών του προσώπου

Αφού βρεθούν οι υποψήφιες περιοχές, γίνονται υπολογισμοί όπως για το μήκος των ακμών, την διαφορά της έντασης κ.ά, και κωδικοποιούνται ως ένα διάνυσμα χαρακτηριστικών (feature vector). Στην συνέχεια, τα διανύσματα αυτά συγκρίνονται με πίνακες που έχουν προκύψει από την εκπαίδευση του μοντέλου και, τέλος, οι περιοχές επαληθεύονται αν η απόσταση Mahalanobis μεταξύ του διανύσματος και του πίνακα των χαρακτηριστικών είναι μικρότερη από ένα οριζόμενο κατώφλι. Τα εξαγόμενα χαρακτηριστικά έχουν πλέον χαρακτηριστεί (ως μάτια, μύτη, στόμα ή φρύδια) από την προηγούμενη διαδικασία, και στην συνέχεια ομαδοποιούνται με βάση το μοντέλο γνώσης το οποίο περιγράφει τις μεταξύ τους σχέσεις και θέσεις. Τέλος, το κάθε χαρακτηριστικό και ομάδα χαρακτηριστικών εισάγονται σε ένα δίκτυο Bayes. Ως δίκτυα Bayes ορίζονται πιθανοτικά γραφικά μοντέλα που αναπαριστούν ένα σύνολο τυχαίων μεταβλητών και τις αλληλεξαρτήσεις τους μέσα από έναν κατευθυνόμενο ακυκλικό γράφο, και χρησιμοποιούνται για την εξαγωγή συμπεράσματος σε συνθήκες αβεβαιότητας [24]. Τα δίκτυα Bayes δεν υποθέτουν την ανεξαρτησία μεταξύ των χαρακτηριστικών αλλά κωδικοποιούν τις μεταξύ τους εξαρτήσεις [25].



Εικόνα 23 Δίκτυο Bayes

Αυτή η μέθοδος που προτείνεται από τον Cipolla & Yow έχει σημαντική επιτυχία στην ανίχνευση προσώπων με διαφορετικό προσανατολισμό και σε διαφορετικές πόζες. Το συνολικό ποσοστό ανίχνευσης σε ένα σύνολο 110 εικόνων με πρόσωπα υπό διαφορετική κλίμακα, οπτική γωνία και διαφορετικούς προσανατολισμούς αγγίζει το 85%, ωστόσο η εσφαλμένη ανίχνευση ανέρχεται στο 28%. Προκύπτει πως η μέθοδος αυτή είναι αποτελεσματικότερη για πρόσωπα μεγαλύτερα από 60x60 pixel.

Η υφή αποτελεί σημαντική πληροφορία για την αναγνώριση και ταξινόμηση των αντικειμένων επομένως και των προσώπων [26]. Η βασική προϋπόθεση για να θεωρηθεί μια περιοχή ως υφή είναι ότι πρέπει να υπάρχει ένας σημαντικός αριθμός χαρακτηριστικών με χωρικές διακυμάνσεις έντασης, τα οποία να είναι ορατά σε κάποιο βαθμό και να είναι πυκνά και ομοιόμορφα στο σύνολο τους [27]. Οι Dai & Nakano [28] προτείνουν χρήση του πίνακα SGLD (Space Gray-Level Dependence matrix) για την ανίχνευση προσώπων Ασιατών. Ο πίνακας SGLD [29] χρησιμοποιείται για τον εντοπισμό της υφής σε εικόνες της κλιμακας του γκρι και εντοπίζει την επαναλαμβανόμενη παρουσία της ίδιας έντασης στην εικόνα. Με βάση τις παραμέτρους του πίνακα SGLD μπορεί να εξαχθεί ένα μοντέλο υφής προσώπου που βασίζεται σε ένα σύνολο ανισοτήτων. Στην συνέχεια ενσωματώνουν σε αυτό το μοντέλο και την πληροφορία του χρώματος ενισχύντας, όμως,

το πορτοκαλί χρώμα καθώς με αυτόν τον τρόπο καταφέρνουν να ενισχύσουν την περιοχή του δέρματος των Ασιατών. Τέλος, δημιουργώντας ένα σχήμα σάρωσης της εικόνας και ελέγχοντας την μέση τιμή του παραθύρου σάρωσης κάνουν σύγκριση με τις ανισότητες που προέκυψαν από το μοντέλο και επιτυγχάνεται η ανίχνευση του προσώπου. Ένα πλεονέκτημα αυτής της προσέγγισης είναι ότι μπορεί να ανιχνεύσει πρόσωπα που δεν είναι όρθια ή έχουν χαρακτηριστικά όπως γένια και γυαλιά. Τα αναφερόμενα ποσοστα ανίχνευσης φαίνονται ενδεικτικά στον επόμενο πίνακα.

test samples		rate determined as faces
case	number	
verti. facial img.	35	100%
incli. facial img.	10	100%
2/3 facial img.	35	80%
1/2 facial img.	40	40%
1/3 facial img.	40	6%
nonfacial img.	150	4%

Εικόνα 24 Ποσοστά επιτυχίας ανίχνευσης προσώπου βάσει της υφής

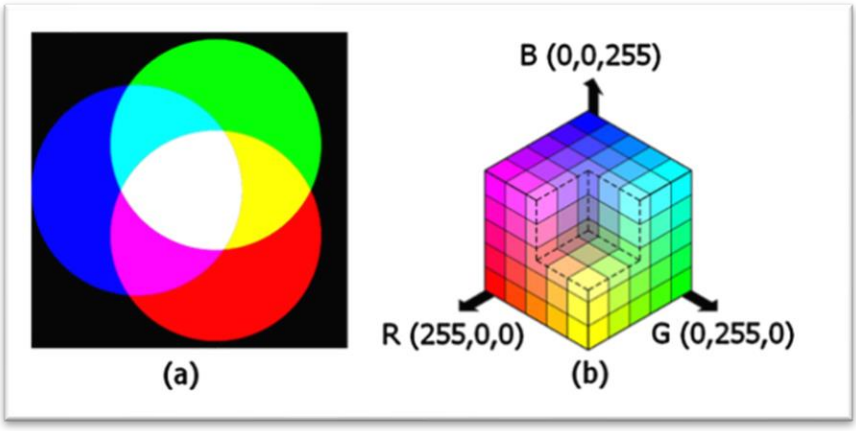
Το χρώμα του δέρματος θεωρείται ως ένα από τα πιο σημαντικά χαρακτηριστικά των ανθρώπινων προσώπων και αποτελεί σημαντικό χαρακτηριστικό για την ανίχνευση και παρακολούθηση-κίνηση ενός προσώπου. Η επεξεργασία χρωμάτων για την αναγνώριση χαρακτηριστικών του προσώπου παρουσιάζει αρκετά πλεονεκτήματα καθώς είναι πολύ ταχύτερη στην εξαγωγή χαρακτηριστικών σε σύγκριση με άλλες μεθόδους, και σε κατάλληλες συνθήκες φωτισμού τα χαρακτηριστικά βάσει του χρώματος είναι ανεξάρτητα του προσανατολισμού του προσώπου. Με την πληροφορία του χρώματος σε μια εικόνα μπορούν να εξαχθούν υποψήφιες περιοχές για την ανίχνευση προσώπου. Με αυτόν τον τρόπο μειώνεται σημαντικά ο χώρος αναζήτησης στην εικόνα με αποτέλεσμα την ταχύτερη διαδικασία ανίχνευσης. Παρ' όλα αυτά, η παρακολούθηση ανθρώπινων προσώπων που χρησιμοποιούν το χρώμα ως χαρακτηριστικό παρουσιάζει πολλές δυσκολίες καθώς η χρωματική αναπαράσταση ενός προσώπου που λαμβάνεται από μια κάμερα επηρεάζεται από πολλούς παράγοντες, όπως το φως περιβάλλοντος και η κίνηση, ενώ παρουσιάζονται αστοχίες σε περίπτωση όπου το υπόβαθρο-φόντο παρουσιάζει παρόμοιες τιμές χρώματος με εκείνο του προσώπου. Επίσης, αν και διαφορετικοί άνθρωποι έχουν διαφορετικό χρώμα δέρματος, αρκετές μελέτες έχουν δείξει ότι η κύρια διαφορά μεταξύ τους έγκειται σε μεγάλο βαθμό στην έντασή τους παρά στην απόχρωσή τους [30]. Για την επισήμανση των εικονοστοιχείων ως δέρματος προσώπου έχουν χρησιμοποιηθεί αρκετά χρωματικά μοντέλα. Το πιο συνηθισμένο χρωματικό μοντέλο για τις εικόνες είναι το μοντέλο RGB, στο οποίο τα χρώματα παρουσιάζονται ως συνδυασμοί του κόκκινου, του πράσινου και του μπλε χρώματος.

Το 1996 οι Yang & Waibel [31] ενέφεραν πως η κύρια διακύμανση στην χρωματική εμφάνιση του δέρματος οφείλεται σε μεγάλο βαθμό στην αλλαγή της φωτεινότητας και προτείνουν την χρήση κανονικοποιημένων χρωμάτων RGB για το φιλτράρισμα της φωτεινότητας. Η κανονικοποίηση γίνεται μέσω της σχέσης:

$$r = R/(R + B + G)$$

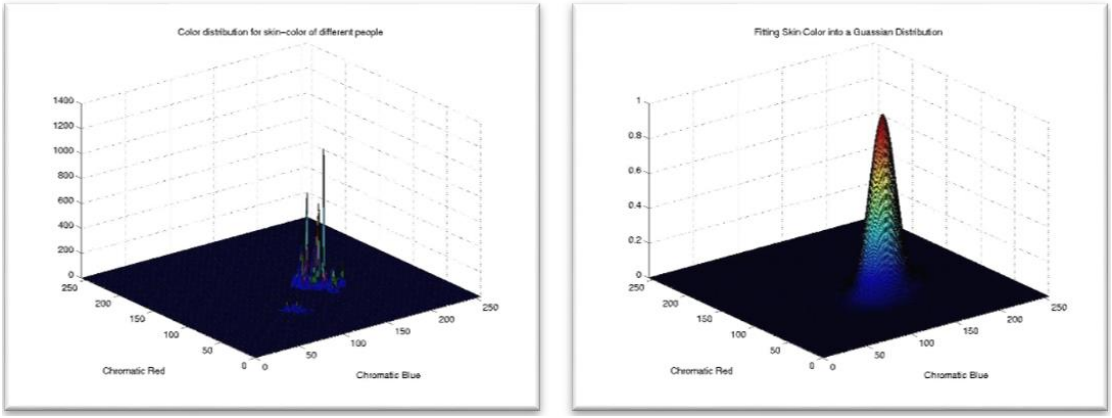
$$g = G/(R + B + G)$$

$$b = B/(R + B + G)$$



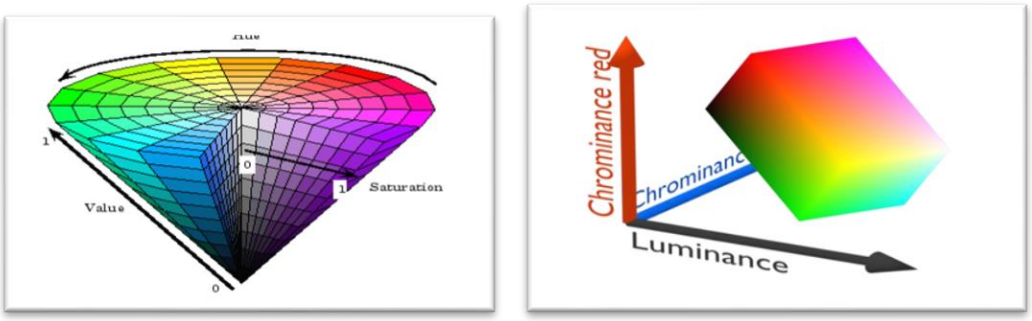
Εικόνα 25 (α) Μίξη χρωμάτων σε μορφή RGB και (β) το χρωματικό μοντέλο RGB χαρτογραφημένο ως κύβος

Σύμφωνα με τις παραπάνω σχέσεις προκύπτει πως το άθροισμα των κανονικοποιημένων τιμών είναι 1, με αποτέλεσμα μια από τις τρεις συνιστώσες να εκφραστεί σε σχέση με τις άλλες δυο πχ. b = 1 – r - g. Έτσι μπορούμε να πούμε πως έχουμε μείωση των διαστάσεων του χρωματικού χώρου. Χρησιμοποιώντας τις κανονικοποιημένες τιμές r & g σε μια Γκαουσιανή κατανομή προκύπτει πως τα χρώματα του προσώπου παρουσιάζουν μια συστάδα (cluster) [31]. Αντίστοιχα, οι φοιτητές του πανεπιστημίου Stanford Chang & Robles παρουσιάζουν την Γκαουσιανή κατανομή. Έτσι ένα pixel μπορεί να ταξινομηθεί να ανήκει σε χρώμα δέρματος αν h(r,g) ≥ τ, όπου τ είναι ένα κατώφλι που προκύπτει εμπειρικά από ανάλυση ιστογραμμάτων από ένα δείγμα εικόνων προσώπου.



Εικόνα 26 Κατανομή του χρώματος ανθρώπων διαφορετικών εθνικοτήτων (αριστερά) και Gaussian κατανομή χρώματος ανθρώπων (δεξιά)

Ένα άλλο χρωματικό μοντέλο που χρησιμοποιείται για την ανίχνευση του χρώματος δέρματος, επομένως και του προσώπου, είναι το μοντέλο HSV. Το μοντέλο αυτό εκφράζει το χρώμα (Hue) με τιμές που διακυμαίνονται από 0-1, τον κορεσμό (Saturation) με τιμές από 0-1 και την ένταση (Value) με τιμές από 0-1 .



Εικόνα 27 Το χρωματικό μοντέλο HSV & Το χρωματικό μοντέλο YCbCr

Το YCbCr είναι ένα κωδικοποιημένο μη γραμμικό σήμα RGB, που χρησιμοποιείται συνήθως από ευρωπαϊκά τηλεοπτικά στούντιο και για εργασίες συμπίεσης εικόνας. Το μοντέλο αυτό αντιπροσωπεύεται από luminance (που αντιστοιχεί στη φωτεινότητα που υπολογίζεται από το μη γραμμικό χώρο RGB) κατασκευασμένο ως σταθμισμένο άθροισμα τιμών RGB [32]. Η απλότητα του μετασχηματισμού και ο ρητός διαχωρισμός των στοιχείων φωτεινότητας και χρώματος καθιστούν το χρωματικό χώρο YCbCr ένα ευρέως χρησιμοποιούμενο χρωματικό χώρο στον ψηφιακό τομέα βίντεο. Σε αυτήν τη μορφή, οι πληροφορίες φωτεινότητας αποθηκεύονται ως ένα μόνο συστατικό (Y) και οι πληροφορίες χρωματισμού αποθηκεύονται ως δύο συστατικά διαφοράς χρώματος (Cb και Cr).

$$Y = 0.299R + 0.287G + 0.11B$$

$$Cr = R - Y$$

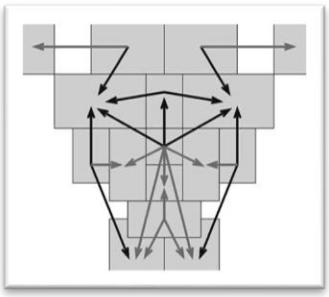
$$Cb = B - Y$$

Η βιβλιογραφία αναφέρει πληθώρα εργασιών που χρησιμοποιούν την προσέγγιση με βάση το χρώμα. Οι αλγόριθμοι συνήθως στοχεύουν στον εντοπισμό κατάλληλου κατωφλίου για την κατάτμηση της εικόνας, και στην συνέχεια ανιχνεύουν τις περιοχές όπου εντοπίζεται το χρώμα του δέρματος. Οι αλγόριθμοι που κάνουν συνδυασμούς των χρωματικών μοντέλων RGB-HSV-YCbCr εμφανίζουν μεγάλη επιτυχία στην ανίχνευση δέρματος. Αυτοί οι αλγόριθμοι είναι σε θέση να επεξεργάζονται εικόνες διαφορετικών συνθηκών φωτισμού και δίνουν αρκετά ικανοποιητικά αποτελέσματα όσον αφορά την ακρίβεια και την ταχύτητα για την ανίχνευση προσώπων, χεριών καθώς και χειρονομιών.

Εικόνα 28. Εικόνες RGB HSV YCbCr

Οι Sakai et al. [33] έκαναν μια πρώιμη προσπάθεια ανίχνευσης μετωπικά απεικονιζόμενων προσώπων σε εικόνες. Αυτή η εργασία βασίστηκε στην ανάλυση γραμμικών «σχεδίων» των προσώπων από εικόνες, με στόχο τον εντοπισμό χαρακτηριστικών του προσώπου, με χρήση δηλαδή υπο-προτύπων (subtemplates) για τη μύτη, τα μάτια, το στόμα και το περίγραμμα του προσώπου ώστε να μοντελοποιηθεί ένα πρόσωπο. Κάθε υποπρότυπο ορίζεται με την έννοια των τμημάτων γραμμής. Οι γραμμές στην εικόνα εισό-

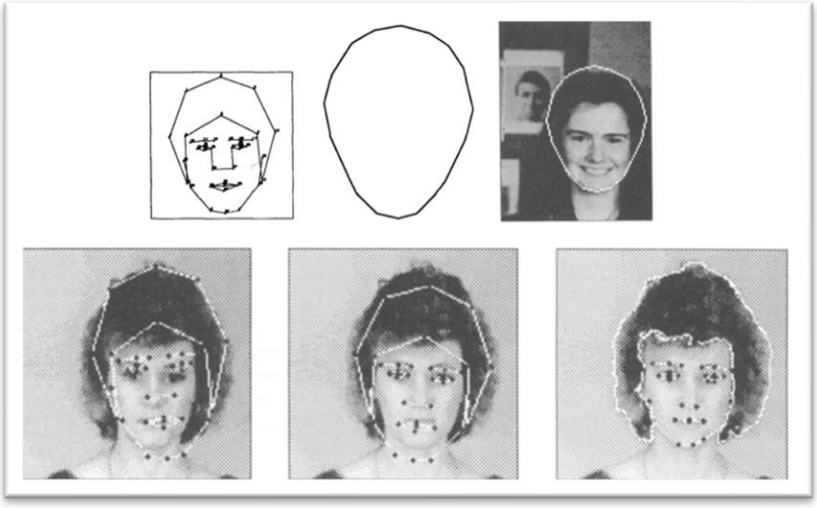
δου εξάγονται με βάση τη μεγαλύτερη αλλαγή κλίσης (gradient) και έπειτα αντιστοιχίζονται με τα υπο-πρότυπα. Για την ανίχνευση υποψήφιων θέσεων προσώπων υπολογίζονται οι συσχετίσεις μεταξύ υποεικόνων (subimages) και προτύπων περιγράμματος. Στην συνέχεια, γίνεται η αντιστοίχιση με τα υπόλοιπα υποπρότυπα για να επαληθευτεί η ανίχνευση. Με άλλα λόγια, η πρώτη φάση καθορίζει την εστίαση της προσοχής ή την περιοχή ενδιαφέροντος και η δεύτερη φάση εξετάζει τις λεπτομέρειες για να προσδιορίσει την ύπαρξη ενός προσώπου. Ο Sinha χρησιμοποίησε ένα μικρό σύνολο αναλλοίωτων χωρικών εικόνων για να περιγράψει το χώρο των μοτίβων προσώπου [34], [35]. Η βασική του ιδέα για το σχεδιασμό αναλλοίωτων χωρικών εικόνων (templates) βασίζεται στο ότι, ενώ οι παραλλαγές στον φωτισμό αλλάζουν συγκεκριμένα την φωτεινότητα διαφορετικών τμημάτων προσώπων (όπως μάτια, μάγουλα και μέτωπο), η σχετική φωτεινότητα αυτών των μερών παραμένει σε μεγάλο βαθμό αμετάβλητη. Ο καθορισμός ζευγών αναλογιών (pairwise ratios) της φωτεινότητας μερικών τέτοιων περιοχών και η διατήρηση μόνο των "κατευθύνσεων" αυτών των αναλογιών παρέχει μια ισχυρή αναλλοίωτη χωρική σχέση σε αυτά τα τμήματα.



Εικόνα 29. Ένα πρότυπο αναλογίας 14x16 pixel για εντοπισμό προσώπου με βάση τη μέθοδο Sinha. Το πρότυπο αποτελείται από 16 περιοχές (τα γκρι τμήματα) και 23 σχέσεις (εμφανίζονται με βέλη) [36]

Στις αρχές της δεκαετίας '90 προτάθηκαν αρκετές μέθοδοι που συνδυάζουν διάφορα χαρακτηριστικά του προσώπου για την ανίχνευση ή εντοπισμό του. Τα περισσότερα από αυτά χρησιμοποιούν χαρακτηριστικά όπως το μέγεθος, το χρώμα του δέρματος και το σχήμα για να βρουν υποψήφια πρόσωπα και, επακόλουθα, να επαληθεύσουν αυτά τα υποψήφια χρησιμοποιώντας λεπτομερή, τοπικά χαρακτηριστικά όπως η μύτη, τα φρύδια και τα μαλλιά.

Οι Craw et al. [37] παρουσίασαν μία μέθοδο εντοπισμού με βάση ένα πρότυπο σχήματος ενός προσώπου σε μετωπική όψη. Αρχικά χρησιμοποιείται ένα φίλτρο Sobel για την εξαγωγή ακμών. Αυτές οι ακμές ομαδοποιούνται για να αναζητήσουν το πρότυπο ενός προσώπου με βάση διάφορους περιορισμούς. Αφού εντοπιστεί το περίγραμμα της κεφαλής, η ίδια διαδικασία επαναλαμβάνεται σε διαφορετικές κλίμακες για τον εντοπισμό χαρακτηριστικών όπως τα μάτια, τα φρύδια και τα χείλη. Αργότερα, οι Craw et al. [38] περιέγραψαν μία μέθοδο εντοπισμού που χρησιμοποιεί ένα σύνολο 40 προτύπων για την αναζήτηση χαρακτηριστικών προσώπου και μια στρατηγική ελέγχου (control strategy) για να καθοδηγήσει και να αξιολογήσει τα αποτελέσματα από τους ανιχνευτές χαρακτηριστικών που βασίζονται σε πρότυπα.



Εικόνα 30. Ανίχνευση προσώπου με αντιστοίχιση προτύπων

Οι Sobottka & Pitas [39] έχουν προτείνει μια μέθοδο για τον εντοπισμό προσώπου (face localization) και την εξαγωγή χαρακτηριστικών προσώπου χρησιμοποιώντας το χρώμα και το σχήμα. Αρχικά, πραγματοποιείται η κατάτμηση χρώματος στο χώρο HSV που αποσκοπεί στον εντοπισμό περιοχών που μοιάζουν με το δέρμα. Στην συνέχεια, με κανόνες «συνδεσμολογίας» προσδιορίζονται από την περιοχή που αναπτύσσεται σε μια χονδροειδή ανάλυση. Για κάθε συνδεδεμένο μέρος, υπολογίζεται η καλύτερη έλλειψη με χρήση γεωμετρικών ροπών (geometric moments). Τα συνδεδεμένα μέρη που είναι καλώς προσεγγισμένα από την έλλειψη επιλέγονται ως υποψήφια πρόσωπα. Έπειτα τα υποψήφια πρόσωπα επαληθεύονται με την αναζήτηση χαρακτηριστικών προσώπου εντός των συνδεδεμένων μερών. Χαρακτηριστικά, όπως τα μάτια και τα στόματα, εξάγονται με βάση την παρατήρηση ότι είναι πιο σκούρα από το υπόλοιπο πρόσωπο. Σε αντίθεση με μεθόδους που βασίζονται μόνο σε εικονοστοιχεία, οι Yang & Ahuja [40] προτείνουν μία μέθοδο ανίχνευσης με βάση τη δομή, το χρώμα και την γεωμετρία. Αρχικά, εκτελείται κατάτηση πολλαπλών κλιμάκων [41] για την εξαγωγή ομοιογενών περιοχών σε μια εικόνα. Χρησιμοποιώντας ένα Gaussian μοντέλο χρώματος δέρματος, εξάγουν περιοχές του τόνου του δέρματος και ομαδοποιούνται σε ελλείψεις. Η ανίχνευση προσώπου γίνεται εφόσον υπάρχουν χαρακτηριστικά προσώπου, όπως μάτια και στόμα, σε αυτές τις ελλειπτικές περιοχές . Πειραματικά αποτελέσματα δείχνουν ότι αυτή η μέθοδος είναι σε θέση να ανιχνεύσει πρόσωπα σε διαφορετικούς προσανατολισμούς με ιδιαίτερα χαρακτηριστικά προσώπου, όπως γυαλιά και γενειάδα.

Με την ανάπτυξη της τεχνολογίας των υπολογιστών και νέων δυναμικών αλγορίθμων που βασίζονται στις αρχές της μηχανικής μάθησης, οι τεχνικές και προσεγγίσεις με βάση την γνώση σταμάτησαν να εξελίσσονται καθώς η ταχύτητα και δυναμικότητα αυτών των τεχνικών υστερούσε σημαντικών των νέων μεθόδων. Η εξέλιξη της μηχανικής μάθησης και της όρασης των υπολογιστών άνοιξαν τον δρόμο στην τεχνολογία ανίχνευσης αντικειμένων σε πραγματικό χρόνο με πολύ υψηλά ποσοστά επιτυχίας.

2.3 Προσέγγιση με βάση την εμφάνιση - εικόνα

Οι μέθοδοι ανίχνευσης προσώπου με βάση την εικόνα εμπίπτουν σε προσεγγίσεις που βασίζονται στη μηχανική μάθηση είτε σε προσεγγίσεις που βασίζονται στη βαθιά μάθηση. Οι προσεγγίσεις βάσει της μηχανικής μάθησης ορίζουν εξαρχής τα χαρακτηριστικά (features) και στην συνέχεια χρησιμοποιούν τεχνικές ταξινόμησης. Από την άλλη πλευρά οι προσεγγίσεις βαθιάς μάθησης είναι σε θέση να εκπαιδεύσουν συστήματα ανίχνευσης και αναγνώρισης χωρίς την εκ των προτέρων εισαγωγή χαρακτηριστικών καθώς τα χαρακτηριστικά προκύπτουν κατά την διάρκεια της εκπαίδευσης. Οι προσεγγίσεις αυτές βασίζονται κυρίως σε συνελικτικά νευρωνικά δίκτυα (CNN).

Ο ορισμός του προβλήματος της ανίχνευσης προσώπων είναι ο προσδιορισμός της θέσης των αντικειμένων σε μια δεδομένη εικόνα εισαγωγής και η ταξινόμησή της σε ποια κλάση ανήκει (αν είναι πρόσωπο ή όχι). Έτσι η διαδικασία της ανίχνευσης, βασιζόμενη στις αρχές της μηχανικής μάθησης, μπορεί να διαιρεθεί σε τρία στάδια: ανίχνευση των πιθανών παραθύρων προσώπου, εξαγωγή χαρακτηριστικών προσώπου και, τέλος, ταξινόμηση. Η επιλογή του παραθύρου είναι μια δύσκολη διαδικασία καθώς τα πρόσωπα μπορεί να βρίσκονται σε οποιαδήποτε θέση και κλίμακα όπως και μπορεί να παρουσιάζουν μεταβλητότητα λόγω των εσωτερικών και εξωτερικών παραγόντων. Για την ανίχνευσή τους μια εύλογη επιλογή είναι να σαρωθεί η εικόνα από παράθυρα ανίχνευσης πολλαπλών κλιμάκων (multi-scale sliding windows), όμως είναι προφανές πως μια τέτοια προσέγγιση θα έχει πολύ σημαντικό υπολογιστικό κόστος. Για την ανίχνευση προσώπου είναι απαραίτητη η εξαγωγή οπτικών χαρακτηριστικών, τα οποία μπορούν να παρέχουν μια σημασιολογική και πληροφοριακή βάση αναγνώρισης. Τα βασικότερα χαρακτηριστικά που χρησιμοποιούνται σε αυτό το στάδιο είναι χαρακτηριστικά τύπου Haar, SURF και HoG. Τα χαρακτηριστικά αυτά αποτελούν μια εξαιρετική επιλογή για εντοπισμό αντικειμένων, μολονότι οι εξωτερικοί παράγοντες όπως αναφέρθηκαν παραπάνω αποτελούν μια πρόκληση. Η τελική απόφαση για την παρουσία προσώπου γίνεται μέσων ενός ταξινομητή. Ο ταξινομητής κρίνεται απαραίτητος για να διακρίνει ένα αντικείμενο-στόχο. Οι πιο συνηθισμένοι ταξινομητές είναι οι Μηχανές Διανυσμάτων Υποστήριξης(SVM) και οι ταξινομητές Softmax και Cascade.

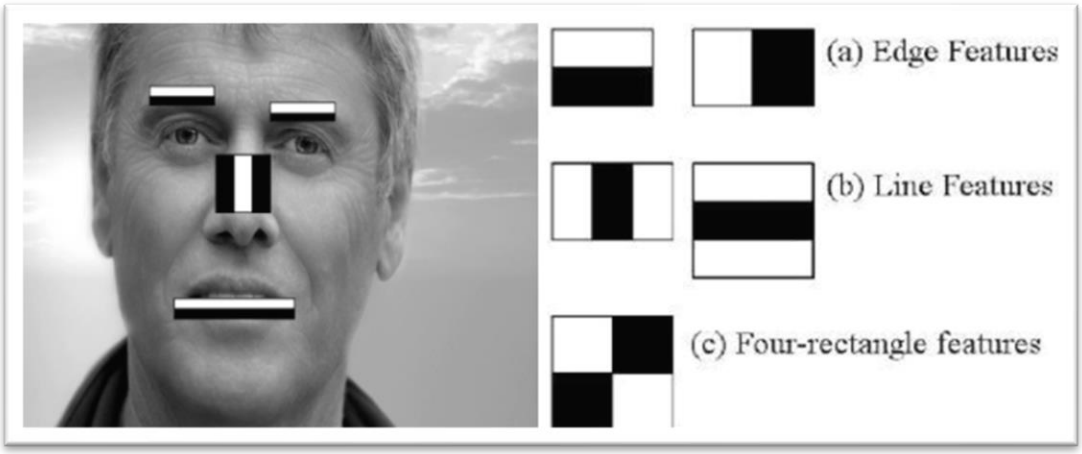
2.3.1 Viola & Jones Algorithm

Στην προηγούμενη ενότητα παρουσιάστηκε ένα σύνολο μεθόδων ανίχνευσης προσώπου, οι οποίες βασίζονται στην γνώση του ερευνητή για το πρόσωπο, και με την επεξεργασία των εικονοστοιχείων σε επίπεδο έντασης, χρώματος και ακμών επιτυγχάνεται η ανίχνευση προσώπου. Στα τέλη της δεκαετίας του '90 οι Papageorgiou et. al [42] παρουσίασαν μια εναλλακτική μέθοδο για την ανίχνευση αντικειμένων. Για πρώτη φορά εισήγαγαν την χρήση ενός συνόλου χαρακτηριστικών τύπου Haar για ανίχνευση αντικειμένων και προσώπων. Η προσέγγιση αυτή αντιμετωπίζει τον εντοπισμό προσώπου σαν ένα πρόβλημα αναγνώρισης προτύπων. Έτσι ελαχιστοποιείται η πιθανότητα σφάλματος μοντελοποίησης λόγω ελλιπούς ή λανθασμένης γνώσης προσώπου. Με αφορμή αυτή την εργασία, λίγα χρόνια αργότερα οι φοιτητές του Cambridge Paul Viola & Michael Jones παρουσιάζουν τον αλγόριθμό τους, ο οποίος αποτελεί τομή στο πεδίο της ανίχνευσης προσώπων και δημιούργησε την υποδομή για την γενικότερη ανίχνευση αντικειμένων. Η συγκεκριμένη μέθοδος μπορεί να εκπαιδευτεί για την ανίχνευση διάφορων κατηγοριών αντικειμένων, ωστόσο η ιδέα που προκάλεσε την υλοποίησή της ήταν η ανίχνευση προσώπων. Ο αλγόριθμος αυτός είχε τον μεγαλύτερο αντίκτυπο στον χώρο της ανίχνευσης αντικειμένων καθώς παρήγαγε ικανοποιητικά αποτέλεσμα σε πραγματικό χρόνο. Για την εφαρμογή του αλγορίθμου σε πραγματικό χρόνο, και για την επίλυση του προβλήματος της ταχύτητας επεξεργασίας, προτείνουν την χρήση της «ολοκληρωμένης εικόνας» (integral image) για τον γρήγορο υπολογισμό των χαρακτηριστικών τύπου Haar, την χρήση του αλγορίθμου Adaboost για την επιλογή των χαρακτηριστικών και την εκμάθηση του ταξινομητή, ενώ, τέλος, με την χρήση ενός δένδρου απόφασης, το οποίο ονομάζεται *Cascade* Classifier, επιτυγχάνεται η ανίχνευση προσώπου.

2.3.1.1 Χαρακτηριστικά τύπου HAAR

Τα χαρακτηριστικά τύπου Haar (HAAR Like Features) οφείλουν την ονομασία τους στην ομοιότητα τους με τα κυματίδια Haar (Haar Wavelet) και προκύπτουν από την εφαρμογή του μετασχηματισμού wavelet με την χρήση των συναρτήσεων τύπου Haar στην εικόνα. Ένα χαρακτηριστικό τύπου Haar λαμβάνει υπόψη τις γειτονικές ορθογώνιες περιοχές σε μια συγκεκριμένη θέση σε ένα παράθυρο ανίχνευσης. Αρχικά υπολογίζεται ο μέσος όρος της έντασης-φωτεινότητας των εικονοστοιχείων σε κάθε περιοχή, και στη συνέχεια υπολογίζεται η διαφορά μεταξύ αυτών των αθροισμάτων. Είναι προφανές πως οι τιμές

των εικονοστοιχείων μιας εικόνας επηρεάζονται έντονα από τις αλλαγές του φωτισμού της σκηνής, οι αλλαγές αυτές όμως επηρεάζουν ομοιόμορφα όλα τα εικονοστοιχεία. Παρά ταύτα, όλα τα παραδείγματα υπο-παραθύρων που χρησιμοποιήθηκαν για την εκπαίδευση κανονικοποιούνται για την ελαχιστοποίηση της επίδρασης διαφορετικών συνθηκών φωτισμού. Έτσι, η τιμή μιας συνάρτησης που εξετάζει τη μέση διαφορά ανάμεσα σε δύο ή τρεις περιοχές της ίδιας εικόνας θα παραμένει σε μεγάλο βαθμό ανεπηρέαστη. Σε ένα ανθρώπινο πρόσωπο μπορεί εύκολα να παρατηρήσει κανείς πως η περιοχή της μύτης είναι πιο φωτεινή στο κέντρο της από ό,τι στα άκρα της (αριστερά και δεξιά, αντίστοιχα). Επομένως, η χρήση χαρακτηριστικών τύπου Haar για την ανίχνευση προσώπου μπορεί να θεωρηθεί πολύ αποτελεσματική για τον εντοπισμό τέτοιων χαρακτηριστικών εάν υπολογιστεί η διαφορά της μέσης φωτεινότητας της περιοχής της μύτης σε σχέση με τα μάγουλα. Η ανίχνευση προσώπου βάσει της προσέγγισης Viola & Jones ταξινομεί τις εικόνες (πρόσωπο και μη πρόσωπο) βάσει της τιμής των χαρακτηριστικών Haar καθώς είναι εύκολο να κωδικοποιήσουν (μέσω της εκπαίδευσης) την γνώση του προσώπου, και γενικά η προσέγγιση με βάση τα χαρακτηριστικά είναι πολύ ταχύτερη και πιο αποδοτική από την επεξεργασία και τον υπολογισμό των εικονοστοιχείων.



Εικόνα 31. Χαρακτηριστικά τύπου Haar

Τα χαρακτηριστικά Haar που χρησιμοποιούνται παρουσιάζονται στην παραπάνω εικόνα και αποτελούνται από τρία είδη. Το πρώτο είδος το οποίο έχει διαστάσεις [2wxh] ή [wx2h], όπου w,h>0 χρησιμεύει κυρίως για την ανίχνευση ακμών (Edge Features). Το δεύτερο είδος χαρακτηριστικών τύπο Haar με διαστάσεις [3wxh] ή [wx3h] χρησιμεύει για την ανίχνευση γραμμικών χαρακτηριστικών (οριζόντιων και κατακόρυφων) (Line Features) και, τέλος, το τρίτο είδος τεσσάρων ορθογωνίων έχει διαστάσεις [2w,2h] (Four rectangle Features). Οι συναρτήσεις Haar υπολογίζουν τη διαφορά ανάμεσα στους μέσους όρους των τιμών των εικονοστοιχείων δύο (ή τριών) περιοχών. Οπότε όσο μεγαλύτερη είναι αυτή η διαφορά τόσο πιο ισχυρό είναι και το χαρακτηριστικό τύπου Haar. Στη συνέχεια αυτή η διαφορά χρησιμοποιείται για την κατηγοριοποίηση των υπο-τμημάτων μιας εικόνας. Για την εύρεση ισχυρών χαρακτηριστικών τύπου Haar σε μια εικόνα, η εικόνα χρειάζεται να σαρωθεί από όλα τα παραπάνω χαρακτηριστικά τύπου Haar (σε όλες τις επιτρεπτές διαστάσεις του) σε όλο το μήκος και πλάτος της με βήμα ενός εικονοστοιχείου. Η προσέγγιση αυτή χρησιμοποιείται για να υπάρχει περίσσεια πληροφορίας και δίνει την δυνατότητα να βρεθούν χαρακτηριστικά προσώπου ακόμα και αν υπάρχει διαφορά κλίμακας. Είναι προφανές πως η διαδικασία αυτή παράγει ένα πολύ μεγάλο πλήθος χαρακτηριστικών. Για μια εικόνα με διαστάσεις [W,H] οι συντελεστές μεγέθυνσης των χαρακτηριστικών είναι X=[W/w], Y=[H/h]. Το πλήθος των χαρακτηριστικών για τον κάθε τύπο προκύπτει από την σχέση:

$$Haar\ Features = XY \cdot \left(W + 1 - w\,\frac{X+1}{2}\right) \cdot \left(H + 1 - h\,\frac{Y+1}{2}\right)$$

Έτσι, σε μια εικόνα με διαστάσεις [24,24] παράγονται παραπάνω από 160.000 χαρακτηριστικά όπως φαίνεται και στον επόμενο πίνακα.

HAAR χαρακτηριστικά	**w**	**h**	**X**	**Y**	**ΠΛΗΘΟΣ**
	2	1	12	24	43.200
	1	2	24	24	43.200
	1	3	24	8	27.600
	3	1	8	24	27.600
	2	2	12	12	20.736
					162.336

Πίνακας 1. Πλήθος Χαρακτηριστικών Haar σε μια εικόνα 24x24

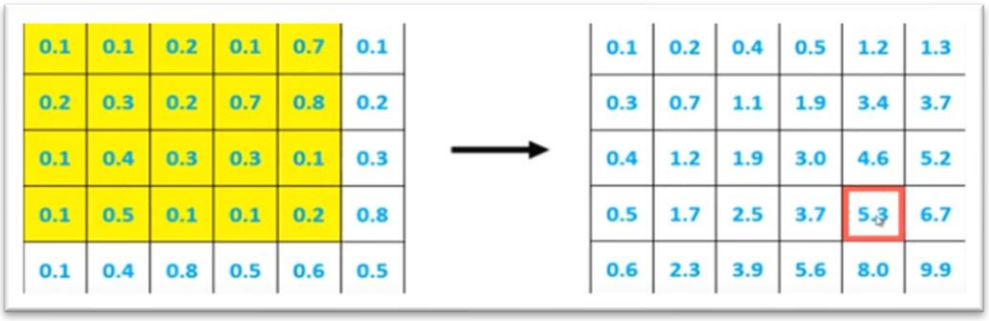
Για την εκπαίδευση χιλιάδων εικόνων είναι προφανές πως το υπολογιστικό κόστος θα ήταν ασύμφορο αν ο υπολογισμός τους γινόταν σε επίπεδο εικονοστοιχείων. Για την επίλυση αυτού του ζητήματος οι Viola & Jones εισάγουν στην μέθοδό τους την χρήση της ολοκληρωμένης εικόνας.

2.3.1.2 Ολοκληρωμένη Εικόνα

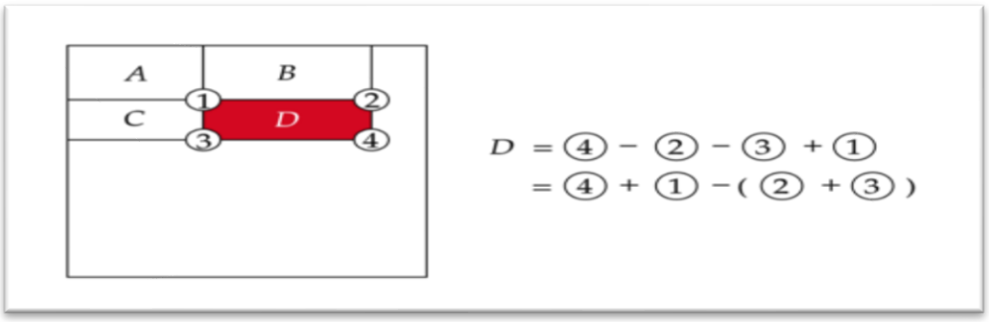
Δεδομένου ότι τα χαρακτηριστικά τύπου Haar είναι ορθογώνια ίσων διαστάσεων, η χρήση της ολοκληρωμένης εικόνας αποδεικνύεται πολύ αποδοτική για τον υπολογισμό των αθροισμάτων των ορθογώνιων υποσυνόλων των χαρακτηριστικών. Το ολοκλήρωμα εικόνας αναφέρεται στην βιβλιογραφία ως πίνακας προστιθέμενου εμβαδού (summed area table) και αποτελεί μια ενδιάμεση κατάσταση εικόνας καθώς κατασκευάζεται από την πρόσθεση των τιμών των εικονοστοιχείων ανά πλάτος και μήκος της εικόνας. Η τιμή στο σημείο ii(x,y) της ολοκληρωμένης εικόνας προκύπτει από την σχέση:

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x', y')$$

όπου ii(x,y) είναι οι τιμές της ολοκληρωμένης εικόνας στην θέση x,y και i(x',y') είναι η τιμή της αρχικής εικόνας στην θέση (x',y'). Ο υπολογισμός φαίνεται στην παρακάτω εικόνα. Το πλεονέκτημα της χρήσης αυτής της μεθόδου είναι πως αυτή η εικόνα χρειάζεται να υπολογιστεί μόνο μια φορά. Στην συνέχεια ο υπολογισμός των χαρακτηριστικών τύπου Haar θα γίνεται μόνο με την πρόσθεση και αφαίρεση αυτών των τιμών.



Εικόνα 32. Υπολογισμός της ολοκληρωμένης εικόνας



Εικόνα 33. Απεικόνιση υπολογισμού των ορθογωνίων περιοχών των χαρακτηριστικών Haar

Ο υπολογισμός κάθε ορθογωνίου του χαρακτηριστικού τύπου Haar υπολογίζεται ως:

$$\sum_{(x,y)\in \mathrm{ABCD}} i(x,y) = \mathrm{i(D)} + \mathrm{ii(A)} - \mathrm{ii(B)} - \mathrm{ii(C)}$$

Αξίζει να σημειωθεί πως οι περισσότερες προσεγγίσεις για την ανίχνευση χαρακτηριστικών σε μια εικόνα, πχ. SIFT [43], για να επιλύσουν το πρόβλημα της κλίμακας κατασκευάζουν μια πυραμίδα εικόνων DoG, στις οποίες ο ανιχνευτής εφαρμόζεται διατηρώντας τις διαστάσεις του. Αυτές οι προσεγγίσεις έχουν το μειονέκτημα πως εκτός από το κόστος της ανίχνευσης στις εικόνες της πυραμίδας προστίθεται και το κόστος της κατασκευής τους. Αντιθέτως, στην προσέγγισή τους οι Viola & Jones, επωφελούμενοι της ευκολίας μεταβολής των διαστάσεων του ανιχνευτή χαρακτηριστικών HAAR, εφαρμόζουν στην ίδια εικόνα ανιχνευτές διαφορετικών διαστάσεων. Άλλο ένα πλεονέκτημα είναι πως με την χρήση της ολοκληρωμένης εικόνας ο υπολογισμός αυτών των χαρακτηριστικών δεν επηρεάζεται από το μέγεθος του ανιχνευτή με αποτέλεσμα να έχει το ίδιο υπολογιστικό κόστος για κάθε κλίμακα παράθυρου ανίχνευσης.

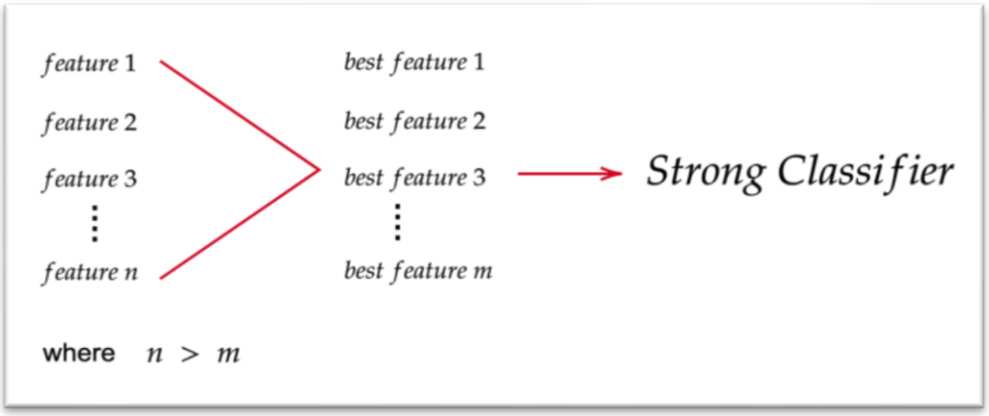
2.3.1.3 Αλγόριθμος Adaboost

Η παραπάνω διαδικασία εντοπισμού χαρακτηριστικών τύπου Haar εξάγει μια πληθώρα χαρακτηριστικών η οποία ξεπερνά πχ. τα 160.000 χαρακτηριστικά σε μια εικόνα. Είναι εύκολο να καταλάβει κανείς πως για την εξαγωγή χαρακτηριστικών από πολλές εικόνες αυτός ο αριθμός μπορεί εύκολα να υπερ-πολλαπλασιαστεί όπως και ότι ένα μεγάλο μέρος των χαρακτηριστικών θα ανιχνευθούν σε περιοχές όπου δεν θα υπάρχει πρόσωπο. Έτσι δημιουργείται η ανάγκη να ανιχνευθούν τα καλύτερα χαρακτηριστικά τα οποία θα περιγράφουν αποκλειστικά ένα πρόσωπο. Είναι λοιπόν απαραίτητο η επιλογή ενός υποσυνόλου χαρακτηριστικών τα οποία θα μπορούν να παράγουν πληροφορία για την ανίχνευση και εντοπισμό ενός αντικειμένου-προσώπου. Στο σύστημά τους οι Viola & Jones χρησιμοποιούν μια παραλλαγή του αλγορίθμου AdaBoost τόσο για την επιλογή των χαρακτηριστικών όσο και για την εκπαίδευση του ταξινομητή [36].

Στην αρχική του μορφή, ο αλγόριθμος εκμάθησης AdaBoost χρησιμοποιείται για την ενίσχυση της απόδοσης ταξινόμησης ενός απλού αλγορίθμου εκμάθησης. Η διαδικασία αυτή επιτυγχάνεται από τον συνδυασμό μιας συλλογής από συναρτήσεις ασθενούς ταξινόμησης (weak classification functions) για να σχηματιστεί ένας ισχυρότερος ταξινομητής (Strong Classifier).

Ένας αλγόριθμος απλής μάθησης καλείται αδύναμος εκπαιδευόμενος (weak learner) και ένα τέτοιο παράδειγμα αποτελεί ο αλγόριθμος εκμάθησης perceptron. Είναι ένας τύπος γραμμικού ταξινομητή, δηλαδή ένας αλγόριθμος ταξινόμησης που κάνει τις προβλέψεις του βασισμένος σε μια λειτουργία γραμμικής πρόβλεψης που συνδυάζει ένα σύνολο βαρών με το διάνυσμα χαρακτηριστικών. Αξίζει να σημειωθεί πως ο εκπαιδευόμενος (learner) καλείται ασθενής-αδύναμος επειδή δεν επιτυγχάνει την καλύτερη λειτουργία ταξινόμησης, να ταξινομήσει δηλαδή με επιτυχία τα δεδομένα εκπαίδευσης. Προκειμένου να ενισχυθεί ο αδύναμος «μαθητής», καλείται να λύσει μια σειρά μαθησιακών προβλημάτων. Μετά τον πρώτο γύρο της μάθησης, τα παραδείγματα επαναβαθμίζονται προκειμένου να τονιστούν εκείνα που ταξινομήθηκαν λανθασμένα από τον προηγούμενο ασθενή ταξινομητή. Έτσι ο τελικός ισχυρός ταξινομητής έχει τη μορφή ενός perceptron, ενός σταθμισμένου συνδυασμού ασθενών ταξινομητών ακολουθούμενου από ένα κατώφλι.

Βασισμένος στην λειτουργία του αλγορίθμου Perceptron, ο αλγόριθμος Adaboost καλείται να επιλέξει ένα υποσύνολο ασθενών ταξινομητών (weak classifiers) με στόχο την κατασκευή ενός τελικού ισχυρού ταξινομητή (strong classifier). Ο ισχυρός ταξινομητής κατασκευάζεται μέσω της εκπαίδευσης, μιας επαναληπτικής διαδικασίας στην οποία επιτυγχάνεται η ελαχιστοποίηση του σφάλματος ταξινόμησης, από ένα μεγάλο σύνολο εικόνων με μέγεθος που αντιστοιχεί στο παράθυρο ανίχνευσης. Το σύνολο πρέπει να περιέχει θετικά παραδείγματα για το επιθυμητό αντικείμενο ανίχνευσης (πχ. μόνο μπροστινή όψη προσώπων) και αυστηρά αρνητικά παραδείγματα (μη πρόσωπα).



Εικόνα 34. Δημιουργία ισχυρού ταξινομητή

Όλα τα παραδείγματα λαμβάνουν ένα συγκριμένο βάρος που είναι ίδιο για όλα. Το κάθε χαρακτηριστικό που έχει εξαχθεί θεωρείται ως ένας ασθενής ανιχνευτής. Έτσι ο ασθενής ταξινομητής $h_j(x)$, όπου x είναι το πλαίσιο του παραδείγματος που εξετάζεται, αποτελείται από ένα συγκεκριμένο και μοναδικό χαρακτηριστικό τύπου Haar $f_j(x)$ και θα εξετάζει το κάθε παράδειγμα εικόνας αν είναι ή δεν είναι πρόσωπο. Η εξέταση γίνεται μέσω ενός κατωφλίου θ_j και της ισοτιμίας p_j η οποία υποδεικνύει την κατεύθυνση της ανισότητας.

$$h_j(x) = 1 \; \alpha\nu \; p_j f_j(x) < p_j \theta_j \text{ για πρόσωπα}$$

$$h_j(x) = -1 \; \alpha\nu \; p_j f_j(x) \geq p_j \theta_j \text{ για μη πρόσωπα}$$

Αναλυτικότερα, η διαδικασία που ακολουθείται για την επιλογή του ισχυρού ταξινομητή έχει ως εξής:

- *Επιλογή των παραδειγμάτων εκπαίδευσης (x_i, y_i), i=1-N με y_i=1 για θετικά παραδείγματα και y_i=0 για αρνητικά παραδείγματα*

- *Γίνεται αρχικοποίηση των βαρών με βάρος:*

$$w_{1,j} = \frac{1}{2m}, \frac{1}{2l}$$

όπου: m αριθμός θετικών παραδειγμάτων

n αριθμός αρνητικών παραδειγμάτων

- *Για t = 1 έως T (αριθμός επαναλήψεων) έτσι ώστε $w_{t,\lambda}$ να είναι μια κατανομή πιθανοτήτων*

1. *Κανονικοποίηση βαρών:*

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$

2. *Για κάθε χαρακτηριστικό j, εκπαιδεύεται ταξινομητής h_j ο οποίος χρησιμοποιεί ένα μοναδικό χαρακτηριστικό. Το σφάλμα αξιολογείται ως προς τα βάρη:*

$$\varepsilon_j = \sum w_i \left| h_j(x_i) - y_i \right|$$

3. *Επιλέγεται ταξινομητής h_j με το μικρότερο σφάλμα ε_{min}*
4. *Γίνεται ενημέρωση των βαρών*

$$w_{t+1,i} = w_t, \beta_t^{1-e_i}$$

όπου $e_i = 0$ για κάθε σωστή ταξινόμηση και $e_i = 1$ για κάθε εσφαλμένη ταξινόμηση

$\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

- *Δημιουργία του τελικού ισχυρού ταξινομητή $h_{(x)}$*

$$h(x) = \begin{cases} 0, & \sum_{t=1}^{T} \alpha_t h_t < \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 1, & \sum_{t=1}^{T} \alpha_t h_t \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \end{cases}$$

'οπου $a = log(\frac{1}{\beta_t})$

2.3.1.4 Ταξινομητής Classifier

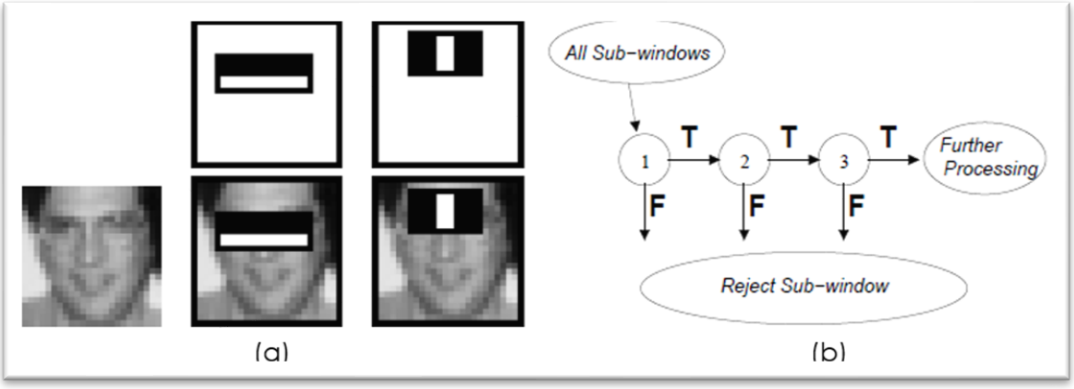
Οι Viola & Jones παρουσιάζουν έναν αλγόριθμο για την κατασκευή ενός καταρράκτη ταξινομητών που επιτυγχάνει αυξημένη απόδοση ανίχνευσης μειώνοντας ριζικά τον χρό-

νο υπολογισμού. Η βασική ιδέα αυτής της προσέγγισης είναι ότι μπορούν να κατασκευαστούν μικρότεροι, και επομένως πιο αποτελεσματικοί, ενισχυμένοι ταξινομητές οι οποίοι θα απορρίπτουν πολλά από τα αρνητικά (μη πρόσωπα) παράθυρα ανίχνευσης, ενώ θα ανιχνεύουν σχεδόν όλες τις θετικές παρουσίες (δηλαδή το κατώφλι ενός ενισχυμένου ταξινομητή μπορεί να ρυθμιστεί έτσι ώστε ο ψευδώς αρνητικός ρυθμός να πλησιάζει το μηδέν). Αρχικά οι απλούστεροι ταξινομητές θα χρησιμοποιούνται για την απόρριψη της πλειονότητας των δευτερευόντων παραθύρων ανίχνευσης, και στην συνέχεια πιο σύνθετοι ταξινομητές καλούνται να «επιβεβαιώσουν» αν στο παράθυρο όντως ανιχνεύεται το αντικείμενο.

Η μορφή της διαδικασίας ανίχνευσης είναι αυτή ενός εκφυλισμένου δέντρου αποφάσεων. το οποίο στην ορολογία της ανίχνευσης αντικειμένων λέγεται «καταρράκτης» (cascade). Ονομάζεται καταρράκτης λόγω του σχεδιασμού του. Στην ουσία η διαδικασία αυτή αποτελείται από στάδια τα οποία συνδέονται σε περίπτωση θετικής εκτίμησης για την ανίχνευση αντικειμένου. Ουσιαστικά ένα θετικό αποτέλεσμα από τον πρώτο ταξινομητή ενεργοποιεί την αξιολόγηση ενός δεύτερου ταξινομητή που έχει επίσης προσαρμοστεί ώστε να επιτυγχάνει πολύ υψηλά ποσοστά ανίχνευσης. Ένα θετικό αποτέλεσμα από τον δεύτερο ταξινομητή ενεργοποιεί έναν τρίτο ταξινομητή και ούτω καθεξής. Ένα αρνητικό αποτέλεσμα σε οποιοδήποτε σημείο οδηγεί στην άμεση απόρριψη του υποπαραθύρου. Τα στάδια του καταρράκτη κατασκευάζονται με την εκπαίδευση ταξινομητών χρησιμοποιώντας τον αλγόριθμο AdaBoost και στη συνέχεια προσαρμόζουν ένα κατώφλι για να ελαχιστοποιηθούν τα ψευδώς αρνητικά υπο-παράθυρα. Αξίζει να σημειωθεί πως το προεπιλεγμένο κατώφλι του AdaBoost έχει σχεδιαστεί για να αποδίδει χαμηλό ποσοστό σφάλματος στα δεδομένα εκπαίδευσης, καθώς ένα χαμηλό κατώφλι θα μπορούσε να παρέχει υψηλότερα ποσοστά ανίχνευσης αντικειμένων αλλά ταυτοχρόνως να παραχθεί και μεγάλος αριθμός ψευδώς θετικών παραθύρων.

Για παράδειγμα, ένας αποτελεσματικός ταξινομητής πρώτου σταδίου μπορεί να κατασκευαστεί από έναν ισχυρό ταξινομητή δύο χαρακτηριστικών και, σε συνδυασμό με την μείωση του κατωφλίου, επιτυγχάνεται και η μείωση των ψευδώς αρνητικών παραθύρων. Στην παρακάτω εικόνα παρατηρούμε τα δύο πρώτα χαρακτηριστικά ως ασθενείς ταξινομητές προσώπου (α). Το πρώτο χαρακτηριστικό μετρά τη διαφορά έντασης μεταξύ της περιοχής των ματιών και μιας περιοχής στα άνω μάγουλα. Το χαρακτηριστικό αξιοποιεί την παρατήρηση ότι η περιοχή των ματιών είναι συχνά πιο σκούρα από τα μάγουλα. Το δεύτερο χαρακτηριστικό συγκρίνει τις εντάσεις στις περιοχές των ματιών με την ένταση κατά μήκος της γέφυρας της μύτης. Ο συνδυασμός αυτών των χαρακτηριστικών αποτελεί τον πρώτο ισχυρό ταξινομητή ο οποίος εισάγεται στο πρώτο στάδιο του καταρράκτη ταξινομητών. Αξίζει να αναφερθεί πως με την χρήση αυτού του ισχυρού ταξινομητή στο πρώτο στάδιο επιτυγχάνεται η ανίχνευση προσώπου με ποσοστό100% και ποσοστό 40 % ψευδώς θετικών παραθύρων προσώπου.

Σημαντικό πλεονέκτημα του ταξινομητή *cascade* είναι η μείωση των ψευδώς θετικών. Με την χρήση ενός ταξινομητή πολλών επιπέδων, οι false-positive εικόνες όμως δεν αποτελούν ουσιαστικό πρόβλημα καθώς αναμένεται να απορριφθούν σε επόμενο στάδιο ώστε να μειωθούν σε πολύ μεγάλο βαθμό στα τελικά στάδια.

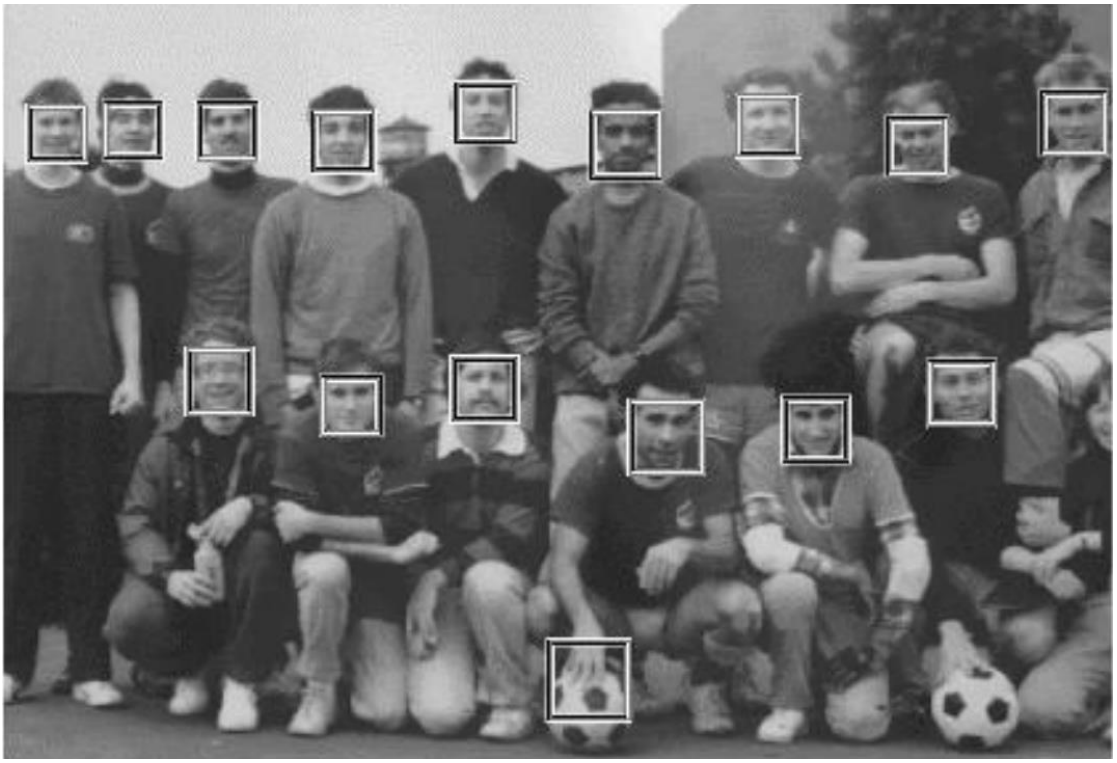


Εικόνα 35. Δυο ισχυροί ταξινομητές (a) συνδυάζονται και εισάγονται στο πρώτο στάδιο του καταρράκτη ανίχνευσης. Σχηματική απεικόνιση ενός καταρράκτη ανίχνευσης (b)

Η δομή του καταρράκτη ταξινομητών στηρίζεται στο γεγονός ότι σε οποιαδήποτε εικόνα, η συντριπτική πλειοψηφία των υπο-παραθύρων είναι αρνητική. Ως εκ τούτου, ο καταρράκτης προσπαθεί να απορρίψει όσο το δυνατόν περισσότερα αρνητικά υπο-παράθυρα στα πρώτα στάδια. Ο καταρράκτης ταξινομητών μοιάζει με ένα δέντρο αποφάσεων όπου οι ταξινομητές του επόμενου σταδίου εκπαιδεύονται χρησιμοποιώντας αυτά τα παραδείγματα που περνούν από όλα τα προηγούμενα στάδια. Αυτό έχει ως αποτέλεσα ο δεύτερος ταξινομητής να αντιμετωπίζει μεγαλύτερη δυσκολία από τον πρώτο. Έτσι σε κάθε στάδιο του καταρράκτη μειώνεται ο ρυθμός ανίχνευσης όπως και ο ρυθμός των ψευδώς θετικών παραθύρων.

Για την εκπαίδευση του καταρράκτη ταξινομητών λαμβάνονται υπόψιν δυο σημαντικοί περιορισμοί. Στις περισσότερες περιπτώσεις οι ταξινομητές με περισσότερα χαρακτηριστικά επιτυγχάνουν υψηλότερα ποσοστά ανίχνευσης και χαμηλότερα ψευδώς θετικά ποσοστά. Την ίδια στιγμή όμως οι ταξινομητές με περισσότερα χαρακτηριστικά απαιτούν μεγαλύτερη υπολογιστική ισχύ. Λαμβάνοντας κανείς αυτά υπόψιν, ορίζει ένα πλαίσιο βελτιστοποίησης και μιας ισορροπίας ως προς τον αριθμό των σταδίων ταξινόμησης, τον αριθμό των χαρακτηριστικών σε κάθε στάδιο και, τέλος, τον ορισμό ενός κατωφλίου για κάθε στάδιο, προκειμένου να ελαχιστοποιηθεί ο αναμενόμενος αριθμός αξιολογούμενων χαρακτηριστικών. Έτσι σε κάθε στάδιο του καταρράκτη μειώνεται το ποσοστό των ψευδώς θετικών όπως το ποσοστό της ανίχνευσης ορίζοντας μια τιμή σαν ελάχιστο ποσοστό απόρριψης και μια αντίστοιχη τιμή σαν μέγιστο ποσοστό ορθής ανίχνευσης. Στην συνέχεια κάθε στάδιο εκπαιδεύεται με την προσθήκη χαρακτηριστικών έως ότου επιτευχθούν οι παραπάνω περιορισμοί. Τέλος όλα τα στάδια προστίθενται έως ότου επιτευχθεί ο συνολικός στόχος για ψευδώς θετικό και ποσοστό ανίχνευσης.

Οι Viola & Jones εκπαίδευσαν έναν ταξινομητή cascade με 38 στάδια για την ανίχνευση προσώπων. Για την εκπαίδευση χρησιμοποίησαν ένα σύνολο 4916 εικόνων προσώπων χειροκίνητα επισημασμένων σε ανάλυση $24x24$. Τα υποπαράθυρα μη προσώπων που χρησιμοποιήθηκαν για την εκπαίδευση προήλθαν από 9544 εικόνες, οι οποίες αυστηρά δεν περιείχαν κανένα πρόσωπο. Η δοκιμή του αλγορίθμου έγινε στη βάση MIT+CMU η οποία περιέχει ένα σύνολο με 130 εικόνες και 507 πρόσωπα ,και κατάφεραν να επιτύχουν ακρίβεια 93.9%.



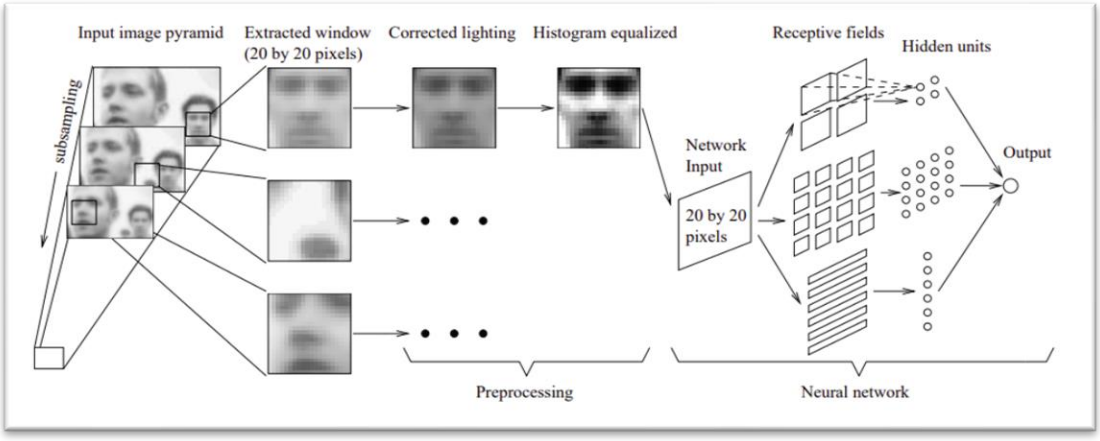
Εικόνα 36. Ανίχνευση προσώπου με τον αλγόριθμο Viola & Jones

2.3.2 Neural Networks

Όπως προαναφέρθηκε, τα νευρωνικά δίκτυα αποτελούν την επιτομή των τεχνικών που χρησιμοποιούνται τα τελευταία χρόνια για την ανίχνευση αντικειμένων, συνεπώς παρουσιάζουν εξαιρετικά αποτελέσματα και για την ανίχνευση και αναγνώριση προσώπων. Τα νευρωνικά δίκτυα χρησιμοποιήθηκαν για πρώτη φορά αποκλειστικά για την ανίχνευση προσώπων το 1992 από τον Argui [44]. Πρόκειται για ένα πολυεπίπεδο δίκτυο όπου το πρώτο επίπεδο αποτελείται από δύο παράλληλα υπο-δίκτυα. Οι τιμές έντασης των εικονοστοιχείων της εικόνας και οι τιμές έντασής τους αφού έχει γίνει χρήση ενός φίλτρου Sobel 3x3 εισάγονται ως δεδομένα εισόδου. Οι είσοδοι του δεύτερου επιπέδου αποτελούνται από τις εξόδους των υπο-δικτύων και τις τιμές των χαρακτηριστικών που έχουν εξαχθεί, όπως είναι η απόκλιση των τιμών έντασης των εικονοστοιχείων της δοθείσας εικόνας και ο λόγος των λευκών εικονοστοιχείων προς το συνολικό πλήθος των εικονοστοιχείων που έχουν κωδικοποιηθεί με το δυαδικό σύστημα στο παράθυρο σάρωσης. Οπότε η τιμή εξόδου από το δεύτερο επίπεδο δηλώνει την παρουσία προσώπου στην περιοχή ενδιαφέροντος. Από πειράματα αποδείχθηκε πως η συγκεκριμένη μέθοδος επιτυγχάνει την ανίχνευση προσώπων, με την προϋπόθεση ότι τα πρόσωπα στις εικόνες εισόδου έχουν το ίδιο μέγεθος.

Το 1996 ο Rowley [45] παρουσιάζει μια μέθοδο επαναστατική για την τότε εποχή, καθώς πετυχαίνει αρκετά υψηλά ποσοστά επιτυχίας ανίχνευσης τα οποία φτάνουν το 79,6% με συγχρόνως πολύ χαμηλά ποσοστά ψευδώς θετικών ανιχνεύσεων. Το πρώτο συστατικό του συστήματός ανίχνευσης είναι ένα νευρωνικό δίκτυο που λαμβάνει ως είσοδο μια περιοχή με διαστάσεις 20x20 της εικόνας και παράγει δυαδική έξοδο που κυμαίνεται από [1,-1] υποδηλώνοντας την παρουσία ή την απουσία ενός προσώπου, αντίστοιχα. Για την ανίχνευση προσώπων σε διαφορετικές θέσεις στην εικόνα εισόδου, το δίκτυο εφαρμόζεται σε κάθε θέση της εικόνας. Για την ανίχνευση προσώπων μεγαλύτερων από το μεγεθος του παραθύρου, η εικόνα εισόδου κλιμακώνεται επανειλημμένα σε μέγεθος (με

υποδειγματοληψία), δημιουργώντας μια πυραμίδα[6] εικόνων ώστε ο ανιχνευτής να εφαρμόζεται σε κάθε μέγεθος. Μετά την εξαγωγή του τμήματος 20x20 από μια συγκεκριμένη θέση και κλίμακα της πυραμίδας εικόνων εισόδου, το τμήμα υποβάλλεται σε επεξεργασία, με προσαρμογή ιστογράμματος, για την διόρθωση φωτισμού. Το προεπεξεργασμένο τμήμα εισάγεται στην συνέχεια στο νευρωνικό δίκτυο το οποίο αποτελείται από ένα κρυφό επίπεδο που περιλαμβάνει 26 μονάδες, όπου οι 4 μονάδες εξετάζουν υποπεριοχές διαστάσεων 10x10, οι 16 μονάδες εξετάζουν περιοχές διαστάσεων 5x5 και οι υπόλοιπες 6 ερευνούν περιοχές διαστάσεων 20x5 μέσα στις οποίες περιέχονται επικαλυπτόμενες οριζόντιες λωρίδες. Κάθε μία από αυτές τις περιοχές έχει πλήρεις συνδέσεις με μία κρυφή μονάδα αν και αυτές οι μονάδες μπορεί να αναπαραχθούν (replicated). Για την βελτίωση του συστήματος ανίχνευσης πρόσθεσε ένα πολυεπίπεδο δίκτυο perceptron (multilayer perceptron) με ένα κρυφό επίπεδο 26 μονάδων εξόδου για την επίτευξη ανίχνευσης σε πρόσωπα με στροφή.

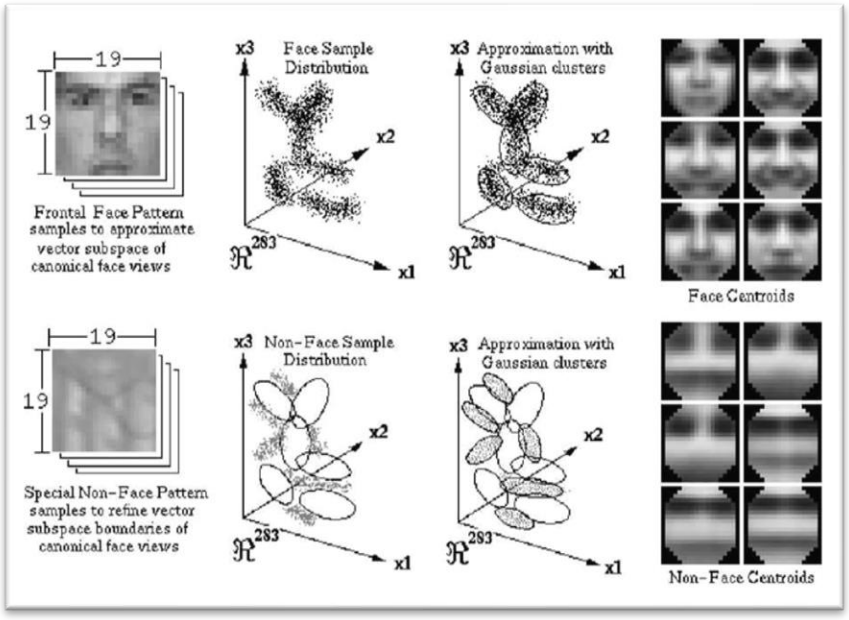


Εικόνα 37. Ο βασικός αλγόριθμος για την ανίχνευση προσώπων κατά Rowley

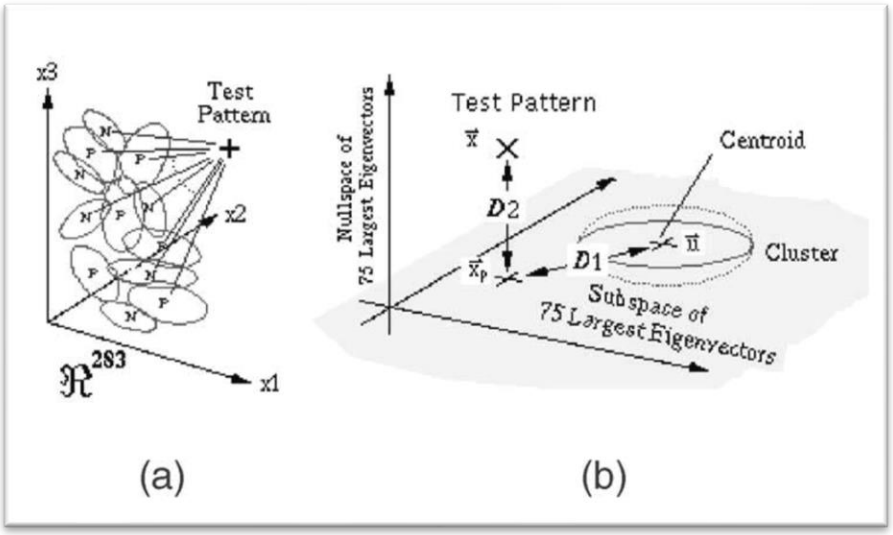
2.3.3 Ανίχνευση βάσει κατανομής

Οι Sung & Poggio [46] ανέπτυξαν ένα σύστημα βασισμένο στη κατανομή για την ανίχνευση προσώπων, το οποίο έδειξε πώς οι κατανομές μοτίβων εικόνας από μια κατηγορία αντικειμένων μπορούν να εκπαιδευτούν από θετικά και αρνητικά παραδείγματα (δηλαδή εικόνες) αυτής της κλάσης. Το σύστημά τους αποτελείται από δύο συστατικά, τα μοντέλα βάσει της κατανομής για μοτίβα προσώπων και μη προσώπων, και έναν ταξινομητή Perceptron πολλαπλών επιπέδων. Για κάθε παράδειγμα προσώπου και μη προσώπου γίνεται πρώτα κανονικοποίηση (normalization) και η μετατροπή σε μια εικόνα διαστάσεων 19x19. Η εικόνα/μοτίβο πλέον αναπαρίσταται ως ένα μονοδιάστατο διάνυσμα 361 θέσεων. Στη συνέχεια, τα μοτίβα ομαδοποιούνται σε έξι όψεις και έξι συστάδες «non-faces» με βάση έναν τροποποιημένο αλγόριθμο k-means, όπως φαίνεται και στην παρακάτω εικόνα. Κάθε συστάδα που δημιουργείται αναπαρίσταται ως πολυδιάστατη συνάρτηση Gauss και συνοδεύεται με την μέση εικόνα και έναν πίνακα συμμεταβλητότητας.

[6] Η εικόνες υποκλιμακώνονται με συντελεστή1.2 για κάθε βήμα στην πυραμίδα



Εικόνα 38. Συστάδες προσώπων και μη προσώπων



Εικόνα 39. Μέτρα απόστασης για ταξινόμηση

Στην παραπάνω εικόνα παρουσιάζονται τα μέτρα απόστασης που χρησιμοποίησαν οι Sung & Poggio. Για την τελική ταξινόμηση υπολογίζονται δύο μετρήσεις απόστασης μεταξύ του μοτίβου εικόνας εισόδου και των συστάδων που έχουν προκύψει από τα παραδείγματα εκπαίδευσης. Έτσι με δεδομένο ένα μοτίβο δοκιμής (test pattern), υπολογίζεται η απόσταση μεταξύ αυτού του μοτίβου εικόνας και κάθε συστάδας (Εικόνα 33α). Σε αυτό το στάδιο γίνεται η σύγκριση ενός συνόλου 12 αποστάσεων μεταξύ του μοτίβου δοκιμής και των 12 κεντροειδών των συστάδων του μοντέλου. Η κάθε απόσταση μεταξύ του προτύπου δοκιμής και του μοντέλου γίνεται με τον υπολογισμό των δυο αποστάσεων στον χώρο των ιδιοτιμών. Από τον υπολογισμό των ελάχιστων αποστάσεων προκύπτει και η τελική ταξινόμηση.

2.3.4 Μηχανές Διανυσμάτων Υποστήριξης

Οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines – SVMs) χρησιμοποιήθηκαν πρώτα για την ανίχνευση προσώπου από τους Osuna et al. [47]. Τα SVMs μπορούν να θεωρηθούν ως ένα νέο πρότυπο για την εκπαίδευση πολυωνυμικών συναρτήσεων, νευρωνικών δικτύων ή συναρτήσεις ακτινικού τύπου (ΣΑΤ) ταξινομητών. Ενώ οι περισσότερες μέθοδοι για την εκπαίδευση ενός ταξινομητή (δηλ. Bayesian, νευρωνικά δίκτυα,και ΣΑΤ) βασίζονται στην ελαχιστοποίηση του σφάλματος εκπαίδευσης, γνωστό ως εμπειρικό κίνδυνο (empirical risk), τα SVMs λειτουργούν με μία άλλη αρχή επαγωγής, που ονομάζεται ελαχιστοποίηση των δομικών κινδύνων (structural risk minimization), πράγμα οποίο στοχεύει στην ελαχιστοποίηση ενός ανώτερου ορίου στο αναμενόμενο σφάλμα γενίκευσης. Ένας ταξινομητής SVM είναι ένας γραμμικός ταξινομητής όπου το διαχωριστικό υπερεπίπεδο (hyperplane) επιλέγεται για να ελαχιστοποιηθεί το αναμενόμενο σφάλμα ταξινόμησης των μη γνωστών προτύπων δοκιμής. Αυτό το βέλτιστο hyperplane ορίζεται από σταθμισμένο συνδυασμό ενός μικρού υποσυνόλου των διανυσμάτων εκπαίδευσης που ονομάζονται διανύσματα υποστήριξης (support vectors). Η εκτίμηση του βέλτιστου hyperplane ισοδυναμεί με την επίλυση ενός γραμμικού περιορισμένου τετραγωνικού προβλήματος προγραμματισμού. Ωστόσο, ο υπολογισμός απαιτεί χρόνο και μνήμη. Στο [47] οι Osuna et al. ανέπτυξαν μια αποτελεσματική μέθοδο για να εκπαιδεύσουν ένα SVM για προβλήματα μεγάλης κλίμακας και την εφαρμόζουν στην ανίχνευση προσώπων. Με βάση δύο σύνολα δεδομένων δοκιμής 10.000.000 προτύπων δοκιμής με 19x19 εικονοστοιχεία, το σύστημά τους έχει ελαφρώς χαμηλότερα ποσοστά σφάλματος και εκτελείται περίπου 30 φορές πιο γρήγορα από το σύστημα των Sung & Poggio [46].

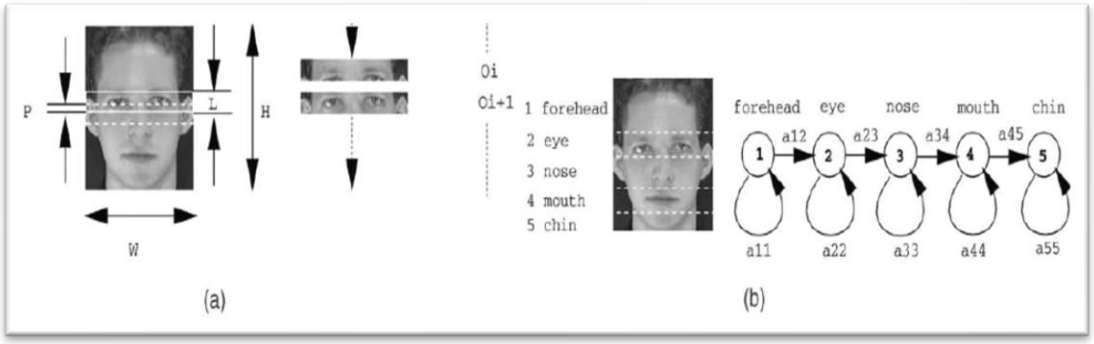
2.3.5 Άλλες Μέθοδοι

Οι Schneiderman & Kanade [48] περιέγραψαν έναν ταξινομητή «naive Bayes» για την εκτίμηση της κοινής πιθανότητας τοπικής εμφάνισης και θέσης των προτύπων προσώπου (υποπεριοχές του προσώπου) σε πολλαπλές αναλύσεις. Τονίζουν την τοπική εμφάνιση, επειδή ορισμένα τοπικά πρότυπα ενός αντικειμένου είναι «πιο μοναδικά» από άλλα, για παράδειγμα τα πρότυπα έντασης γύρω από τα μάτια είναι πιο διακριτά από τα πρότυπα που βρίσκονται γύρω από τα μάγουλα. Υπάρχουν δύο λόγοι για τη χρήση ενός ταξινομητή naive Bayes (δηλαδή δεν υπάρχει στατιστική εξάρτηση μεταξύ των υποπεριοχών). Πρώτον, δίνει καλύτερη εκτίμηση των συναρτήσεων πυκνότητας υπό όρους αυτών των υποπεριοχών. Δεύτερον, ένας ταξινομητής naive Bayes παρέχει μία functional form της μεταγενέστερης πιθανότητας για να συλλάβει τις κοινές στατιστικές της τοπικής εμφάνισης και θέσης στο αντικείμενο. Σε κάθε κλίμακα, μια εικόνα προσώπου αποσυντίθεται σε τέσσερις ορθογώνιες υποπεριοχές. Αυτές οι υποπεριοχές προβάλλονται έπειτα σε ένα χώρο χαμηλότερης διάστασης με χρήση PCA και διακριτοποιούνται σε ένα πεπερασμένο σύνολο προτύπων, και τα στατιστικά κάθε υποπεριοχής που έχει προβληθεί υπολογίζονται από τα δείγματα που έχουν προβληθεί για να κωδικοποιήσουν την τοπική εμφάνιση. Με αυτήν την διατύπωση, η μέθοδός τους αποφασίζει ότι ένα πρόσωπο είναι παρόν όταν ο λόγος πιθανότητας είναι μεγαλύτερος από το λόγο των προηγούμενων πιθανοτήτων.

Μια άλλη αξιοσημείωτη μέθοδος ανίχνευσης προσώπων αποτελεί η χρήση Hidden Markov Model. Η υποκείμενη υπόθεση του Hidden Markov Model (HMM) είναι ότι τα πρότυπα μπορούν να χαρακτηριστούν σε μια τυχαία παραμετρική διαδικασία και ότι οι παράμετροι αυτής της διαδικασίας μπορούν να εκτιμηθούν με έναν ακριβή, καλά καθορισμένο τρόπο.

Κατά την ανάπτυξη ενός HMM για ένα πρόβλημα αναγνώρισης προτύπων πρέπει να αποφασιστεί μια σειρά από κρυφές καταστάσεις (hidden states) για να σχηματίσουν ένα μοντέλο. Στη συνέχεια, μπορεί κανείς να εκπαιδεύσει το HMM για να μάθει τη μεταβατική πιθανότητα μεταξύ των καταστάσεων από τα παραδείγματα όπου κάθε παράδειγμα παρουσιάζεται ως ακολουθία παρατηρήσεων. Ο στόχος της εκπαίδευσης ενός HMM

είναι η μεγιστοποίηση της πιθανότητας παρατήρησης των δεδομένων εκπαίδευσης με προσαρμογή των παραμέτρων σε ένα μοντέλο HMM με την τυπική μέθοδο κατάτμησης Viterbi και τους αλγορίθμους Baum-Welch [49]. Αφού εκπαιδευτεί το HMM, η πιθανότητα εξόδου μιας παρατήρησης καθορίζει την κλάση στην οποία ανήκει. Οι μέθοδοι που βασίζονται σε HMM συνήθως αντιμετωπίζουν ένα μοτίβο προσώπου ως μια αλληλουχία παρατηρήσεων, όπου το κάθε διάνυσμα είναι μια σειρά εικονοστοιχείων όπως φαίνεται στην παρακάτω Εικόνα 40a. Κατά τη διάρκεια της εκπαίδευση και των δοκιμών, μια εικόνα σαρώνεται συνήθως από πάνω προς τα κάτω και η παρατήρηση λαμβάνεται ως ένα μπλοκ στοιχείων.



Εικόνα 40 Hidden Markov Model για τον εντοπισμό προσώπου.

Τα μοτίβα του προσώπου περιγράφονται από τα όρια μεταξύ λωρίδων εικονοστοιχείων και αναπαρίστανται από πιθανοτικές μεταβάσεις μεταξύ καταστάσεων, όπως φαίνεται στην Εικόνα 40b. Τα δεδομένα εικόνας σε μια περιοχή μοντελοποιούνται από μια πολυπαραμετρική κατανομή Gauss. Έτσι μια ακολουθία παρατήρησης αποτελείται από όλες τις τιμές έντασης από κάθε μπλοκ, και οι εξαγόμενες καταστάσεις αντιστοιχούν στις τάξεις στις οποίες ανήκουν οι παρατηρήσεις. Με το τέλος της εκπαίδευσης η πιθανότητα εξόδου μιας παρατήρησης καθορίζει και την κατηγορία στην οποία ανήκει. Ο Samaria [50] απέδειξε στην εργασία του πως οι καταστάσεις που μαθαίνονται σε ένα HMM αντιστοιχούν στις περιοχές προσώπου. Με λίγα λόγια, μια κατάσταση του μοντέλου είναι υπεύθυνη για τον χαρακτηρισμό των διανυσμάτων παρατήρησης των ανθρώπινων μετώπων και μια άλλη κατάσταση είναι υπεύθυνη για τον χαρακτηρισμό των διανυσμάτων παρατήρησης των ανθρώπινων ματιών κ.ο.κ.

Κεφάλαιο 3. Αναγνώριση Προσώπου

Τα τελευταία χρόνια, η αναγνώριση προσώπου έχει γίνει μια από τις πιο απαιτητικές εργασίες στον τομέα της αναγνώρισης προτύπων αποτελώντας σημείο αναφοράς για μια πλειάδα εφαρμογών, όπως την ταξινόμηση εικόνων, την παρακολούθηση βίντεο, την ανάκτηση ταυτότητας από μια βάση δεδομένων και γενικά σε βιομετρικά συστήματα. Κάθε βιομετρικό σύστημα έχει τέσσερα κύρια χαρακτηριστικά:

- Ανίχνευση προσώπου,
- Προεπεξεργασία
- Εξαγωγή χαρακτηριστικών
- Αναγνώριση προσώπου.

Εικόνα 41. Αρχιτεκτονική Αναγνώρισης Προσώπου

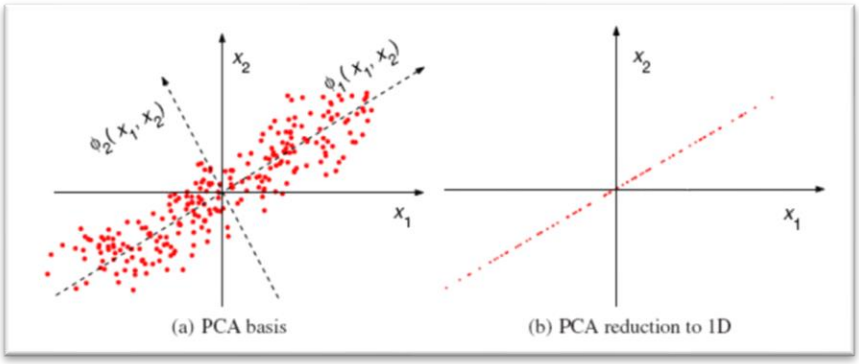
Όπως φαίνεται στην παραπάνω εικόνα η πρώτη εργασία ενός συστήματος αναγνώρισης προσώπου είναι η λήψη εικόνας (μέσω βίντεο, κάμερας ή από τη βάση δεδομένων) και αυτή η εικόνα δίνεται στο περαιτέρω βήμα του συστήματος. Στο πρώτο στάδιο γίνεται η ανίχνευση του προσώπου. Η κύρια λειτουργία σε αυτό το στάδιο είναι να ανιχνευθεί το πρόσωπο από τη λήψη της εικόνας ή την επιλεγμένη εικόνα από τη βάση δεδομένων. Αυτή η διαδικασία ανίχνευσης προσώπου επαληθεύει, στην πραγματικότητα, εάν στην συγκεκριμένη εικόνα υπάρχει πρόσωπο ή όχι. Μετά την ανίχνευσή του, το πρόσωπο θα δοθεί στο περαιτέρω στάδιο της προ-επεξεργασίας. Σε αυτό το στάδιο συνήθως γίνεται ο «καθαρισμός» της εικόνας, δηλαδή ο θόρυβος, η θολότητα, οι ποικίλες συνθήκες φωτισμού, η επίπτωση της σκιάς απομακρύνονται με την χρήση τεχνικών προεπεξεργασίας. Η συνήθης εργασία για την αναγνώριση αντικειμένων απαιτεί την διαδικασία της εξαγωγής χαρακτηριστικών HOG, SURF, Haar κ.ά. τα οποία στην συνέχεια ταξινομούνται με διάφορες μεθόδους ταξινόμησης, όπως τα SVM, ώστε να επιτευχθεί η αναγνώριση. Άλλες τεχνικές επιτυγχάνουν την εξαγωγή χαρακτηριστικών με την εξαγωγή ιδιοπροσώπων (eigen faces) και, σε συνδυασμό με αλγορίθμους μείωσης της διαστασιμότητας, επιτυγχάνουν σημαντικές επιτυχίες αναγνώρισης προσώπων. Οι πιο διαδεδομένες μέθοδοι για την αναγνώριση αντικειμένων και προσώπων χρησιμοποιούν βαθιά νευρωνικά δίκτυα καθώς παρέχουν υψηλή ταχύτητα αναγνώρισης, συνεπώς μπορούν να χρησιμοποιηθούν με επιτυχία σε συστήματα αναγνώρισης σε πραγματικό χρόνο. Στην εργασία αυτή αναλύονται δυο μέθοδοι μηχανικής μάθησης: Αναγνώριση Προσώπου με Ιδιοπρόσωπα (Eigen Faces) και Αναγνώριση Προσώπου με Bag of Visual Features.

3.1 Αναγνώριση προσώπου με Ιδιοπρόσωπα (Eigen Faces)

Η αναγνώριση προσώπου είναι ένα πρόβλημα αναγνώρισης προτύπων υψηλής διαστασιμότητας. Ακόμα και οι εικόνες προσώπου χαμηλής ανάλυσης δημιουργούν τεράστιους χώρους διαστάσεων (10.000 διαστάσεις στην περίπτωση μιας εικόνας προσώπου 100x100 εικονοστοιχείων). Εκτός από τα προβλήματα μεγάλης υπολογιστικής πολυπλοκότητας και αποθήκευσης μνήμης, το πρόβλημα της διαστασιμότητας καθιστά πολύ δύσκολη την λήψη στατιστικών μοντέλων του χώρου της εικόνας με την χρήση καθορισμένων παραμετρικών μοντέλων. Ωστόσο, η διαστασιμότητα (dimensionality) του χώρου του προσώπου είναι πολύ χαμηλότερη από τη διασπασιμότητα του χώρου της

εικόνας, καθώς τα πρόσωπα είναι παρόμοια στην εμφάνιση και περιέχουν πολλές και αρκετά σημαντικές στατιστικές κανονικότητες (statistical regularities).

Η Ανάλυση Κύριων Συνιστωσών (PCA) είναι ένα τυπικό εργαλείο στη σύγχρονη ανάλυση δεδομένων επειδή είναι μια απλή, μη παραμετρική μέθοδος για την εξαγωγή σχετικών πληροφοριών από ένα σύνολο σύνθετων δεδομένων μεγάλων διαστάσεων [51]. Είναι μια από τις πλέον επιτυχημένες τεχνικές που έχουν χρησιμοποιηθεί στην αναγνώριση και συμπίεση εικόνας, και αποτελεί μια από τις καλύτερες μεθόδους μείωσης της διαστασιμότητας. Η PCA είναι μια στατιστική μέθοδος υπό τον ευρύ όρο της ανάλυσης παραγόντων (factor analysis). Ο σκοπός της είναι να μειώσει τη μεγάλη διάσταση του χώρου δεδομένων (παρατηρούμενες μεταβλητές) στη μικρότερη εγγενή διάσταση του χώρου χαρακτηριστικών (ανεξάρτητες μεταβλητές), οι οποίες απαιτούνται για την περιγραφή των δεδομένων. Αυτό όμως συμβαίνει όταν υπάρχει μια σχετικά ισχυρή συσχέτιση μεταξύ των παρατηρούμενων μεταβλητών. Η ιδέα βασίζεται στο γεγονός ότι ένα σύνολο δεδομένων μεγάλων διαστάσεων, όπως είναι μια εικόνα, συχνά περιγράφεται από συσχετιζόμενες μεταβλητές, και έτσι η αποκλειστική πληροφορία του προσώπου μπορεί να περιγραφεί από ένα μικρότερο σύνολο από σημαντικές πληροφορίες μικρότερης διάστασης. Η PCA βρίσκει τις διευθύνσεις με τη μεγαλύτερη διακύμανση στα δεδομένα οι οποίες καλούνται κύριες συνιστώσες. Οι κύριες συνιστώσες στην ουσία είναι ένας τρόπος αναπαράστασης/προβολής των πιο σημαντικών πληροφοριών των δεδομένων. Μερικές από της εργασίες που μπορούν να επιλυθούν με την χρήση της PCA είναι *η εξαγωγή χαρακτηριστικών, η πρόβλεψη, η αφαίρεση πλεονασμού, η συμπίεση δεδομένων* κ.λπ.



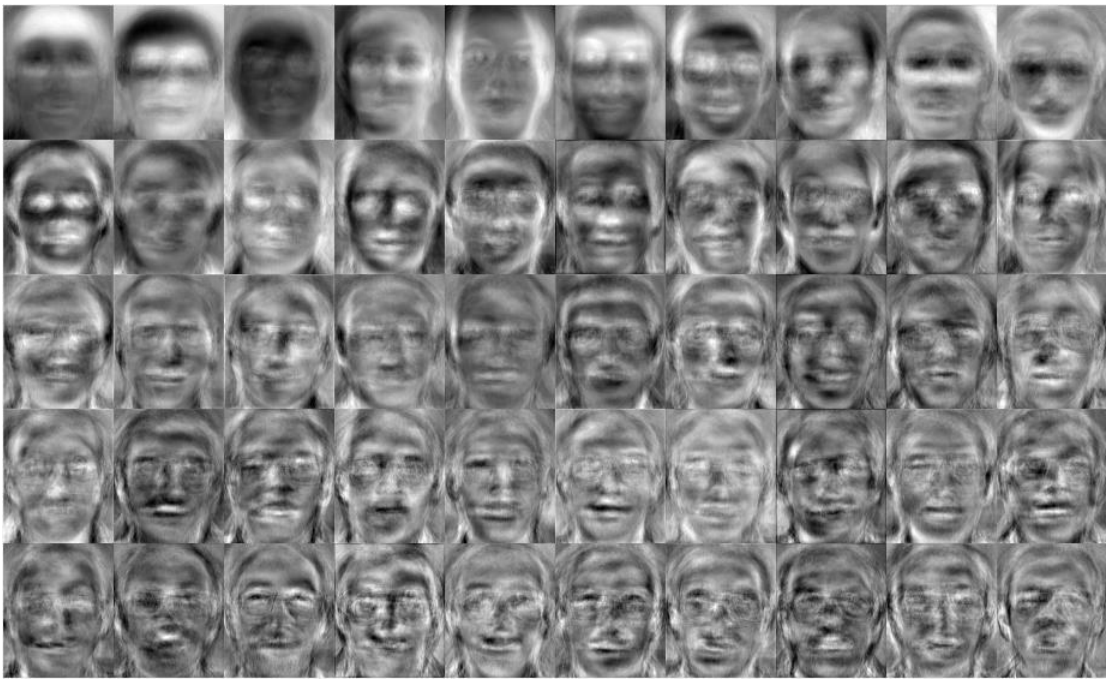
Εικόνα 42. Προβολή δεδομένων στον χώρο των κύριων συνιστωσών – Μείωση 2D διάστασης

Η κύρια ιδέα της χρήσης της PCA για την αναγνώριση προσώπου είναι η έκφραση των διανυσμάτων 1D εικονοστοιχείων, τα οποία κατασκευάζονται στην ουσία από τις 2D εικόνες προσώπου, στις βασικές συνιστώσες του χώρου χαρακτηριστικών. Αυτό μπορεί να ονομαστεί ως η προβολή του ιδιοχώρου (eigenspace projection). Σημειώνεται ότι ένας ιδιοχώρος (eigenspace) ορίζεται ως η συλλογή των ιδιοδιανυσμάτων (eigen vectors) που σχετίζονται με ιδιοτιμές (eigen values) για τον γραμμικό μετασχηματισμό που εφαρμόζεται στα ιδιοδιανύσματα. Ο γραμμικός μετασχηματισμός είναι ένας τετραγωνικός πίνακας και στην γραμμική άλγεβρα ένα ιδιοδιάνυσμα (eigenvector) ενός γραμμικού μετασχηματισμού είναι ένα μη μηδενικό διάνυσμα το οποίο αλλάζει ως προς τον συντελεστή κλίμακας (ιδιοτιμή) όταν του εφαρμόζεται ο γραμμικός μετασχηματισμός. Η αναγνώριση προσώπου με βάση τον ιδιοχώρο (eigenspace) αποτελεί μια από τις πιο επιτυχημένες μεθοδολογίες για την ανίχνευση και αναγνώριση προσώπων σε ψηφιακές εικόνες.

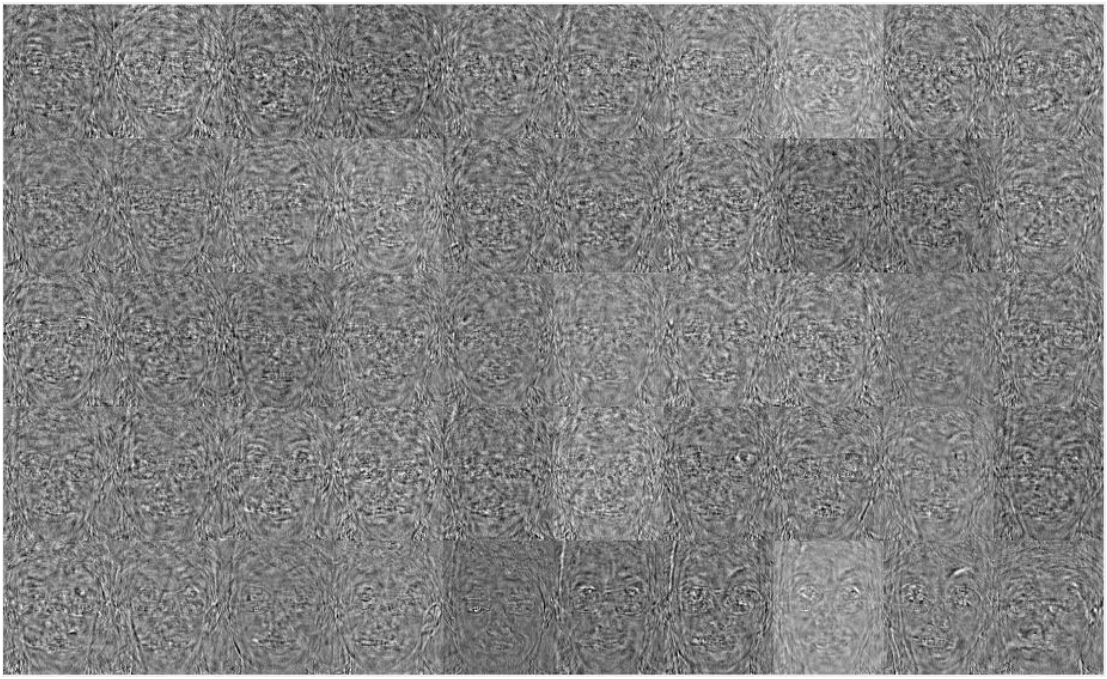
Ένα πρώτο παράδειγμα της χρήσης ιδιοδιανυσμάτων (eigenvectors) στην αναγνώριση προσώπου έγινε από τον Kohonen το 1989 [52], στο οποίο ένα απλό νευρωνικό δίκτυο εκτελεί αναγνώριση προσώπου για ευθυγραμμισμένες και ομαλοποιημένες εικόνες προσώπου. Στην εργασία τους, ένα απλό νευρωνικό δίκτυο υπολόγισε μια περιγραφή προσώπου προσεγγίζοντας τα ιδιοδιανύσματα (eigen vectors) του πίνακα αυτοσυσχέτισης

(autocorrelation) της εικόνας. Αυτά τα ιδιοδιανύσματα έγιναν ύστερα γνωστά ως «ιδιοπρόσωπα» (eigenfaces). Το σύστημα του Kohonen δεν είχε ωστόσο πρακτική επιτυχία, λόγω της ανάγκης για ακριβή ευθυγράμμιση και ομαλοποίηση. Τα επόμενα χρόνια πολλοί ερευνητές δοκίμασαν σχήματα αναγνώρισης προσώπου με βάση τις ακμές, τις αποστάσεις μεταξύ χαρακτηριστικών και άλλες προσεγγίσεις νευρωνικών δικτύων. Ενώ πολλές ήταν επιτυχημένες σε μικρές βάσεις δεδομένων ευθυγραμμισμένων εικόνων, κανένας δεν αντιμετώπισε με επιτυχία το πιο ρεαλιστικό πρόβλημα των μεγάλων βάσεων δεδομένων όπου η θέση και η κλίμακα του προσώπου είναι άγνωστη.

Λίγα χρόνια αργότερα οι Kirby & Sirovich [53] απέδειξαν ότι οι εικόνες των προσώπων μπορούν να κωδικοποιηθούν γραμμικά με χρήση μικρού αριθμού βασικών (basis) εικόνων. Η εργασία τους βασίστηκε στον μετασχηματισμό Karhunen & Loeve (KLT), ο οποίος αναφέρεται στην βιβλιογραφία και ως Hotelling Transform ή Ανάλυση κύριων συνιστωσών – PCA [53]. Ο στόχος της προσέγγισής τους είναι ή απεικόνιση μιας εικόνας ενός προσώπου σε ένα βέλτιστο σύστημα συντεταγμένων, στον χώρο του προσώπου. Η διαδικασία αυτή προσφέρει την μείωση της διαστασιμότητας, καθώς αφαιρεί πολλές μη χρήσιμες πληροφορίες μετασχηματίζοντας τη δομή του προσώπου σε ένα ορθοκανονικό σύνολο αξόνων στις διευθύνσεις όπου τα νέα δεδομένα-χαρακτηριστικά παρουσιάζουν μέγιστη συμμεταβλητότητα (eigenfaces). Δεδομένου ενός συνόλου εικόνων εκπαίδευσης με διαστάσεις [m,n] που αναπαρίστανται ως ένα διάνυσμα μεγέθους $m * n$, τα βασικά διανύσματα (basis vectors) που εκτείνονται σε έναν βέλτιστο υποχώρο προσδιορίζονται έτσι ώστε να ελαχιστοποιηθεί το μέσο τετραγωνικό σφάλμα (mean square error) μεταξύ της προβολής των εικόνων εκπαίδευσης σε αυτόν τον υποχώρο και των αρχικών εικόνων. Το σύνολο των βέλτιστων βασικών διανυσμάτων (basis vectors) καλείται ιδιοπρόσωπο (eigenface) επειδή είναι απλώς τα ιδιοδιανύσματα (eigen vectors) που έχουν προκύψει από τον πίνακα συμμεταβλητότητας ο οποίος προκύπτει από τις διανυσματικές εικόνες προσώπου στο σύνολο των δεδομένων εκπαίδευσης.

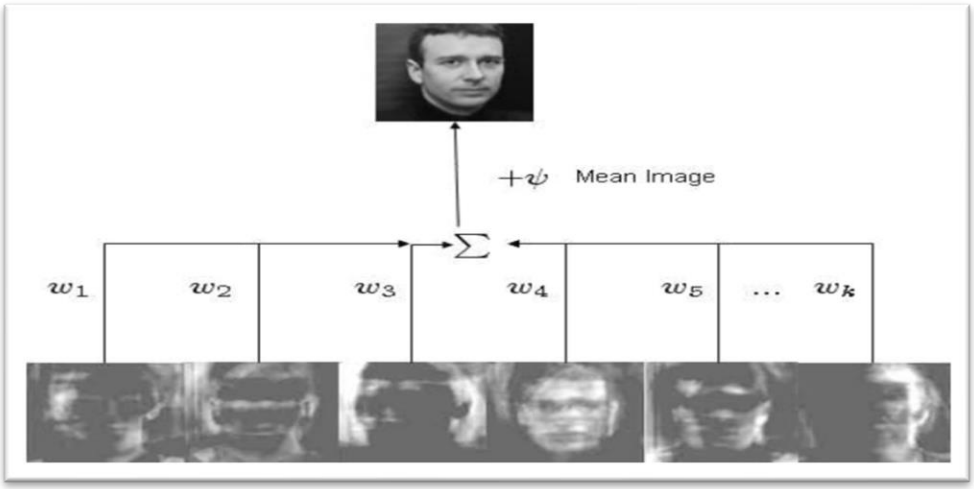


Εικόνα 43. Τα πρώτα 50 ιδιοπρόσωπα



Εικόνα 44. Τα τελευταία 50 ιδιοπρόσωπα

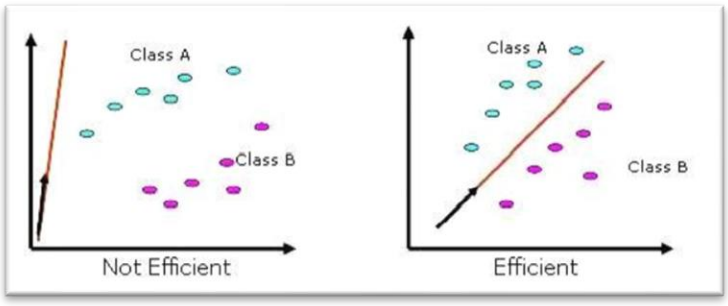
Όλες οι εικόνες προσώπων του συνόλου εκπαίδευσης προβάλλονται πάνω στον υποχώρο προσώπου (face space) προκειμένου να βρεθεί ένα σύνολο βαρών που περιγράφει τη συμμετοχή κάθε διανύσματος στον υποχώρο προσώπου. Στην παραπάνω εικόνα παρατηρείται πως τα πρώτα ιδιοπρόσπα που εκφράζουν την μέγιστη συμμεταβλητότητα παρέχουν σημαντικές πληροφορίες ενός προσώπου, ενώ τα τελευταία ιδιοπρόσωπα παρουσιάζονται με την μορφή θορύβου γεγονός που υποδηλώνει μικρότερη συμμεταβλητότητα. Έτσι μπορούμε να πούμε πως τα τελευταία ιδιοδιανύσματα δεν παρέχουν σημαντική πληροφορία και μπορούν να παραληφθούν από την διαδικασία. Ο συνδυασμός αυτών των ιδιοπροσώπων με την χρήση κατάλληλων βαρών μπορούν να επιτύχουν την ανακατασκευή της αρχικής εικόνας, όπως φαίνεται και στην αμέσως επόμενη εικόνα. Με αυτόν τον τρόπο ο Kirby & Sirovich [53] απέδειξαν πως η ανάλυση κύριων συνιστωσών θα μπορούσε να χρησιμοποιηθεί σε μια συλλογή εικόνων προσώπου για να σχηματίσει ένα σύνολο *βασικών χαρακτηριστικών*, τα ιδιοπρόσωπα, τα οποία μπορούν να συνδυαστούν γραμμικά μαζί με το μέσο πρόσωπο για να ανακατασκευάσουν εν τέλει τις εικόνες στο αρχικό σετ εκπαίδευσης.



Εικόνα 45. Ανακατασκευή προσώπου από ιδιοπρόσωπα

Εικόνα 46. Ανακατασκευή προσώπου με 1-50-100-150-200-250-300-400-405 ιδιοπρόσωπα

Οι Turk & Pentland [54], βασιζόμενοι στην προηγούμενη εργασία των Kirby & Sirovich [53], εφαρμόζουν την ανάλυση κύριων συνιστωσών σε ένα σύνολο δεδομένων εκπαίδευσης από εικόνες προσώπων για την δημιουργία των ιδιοπροσώπων προκειμένου να επιτύχουν αναγνώριση προσώπων. Η προσέγγιση αφορά την αναγνώριση μετωπικά απεικονιζόμενων προσώπων, δηλαδή σε όρθια και εμπρόσθια θέση. Τα ιδιοδιανύσματα αποτελούν στην ουσία την αποδοτικότερη αναπαράσταση δεδομένων. Αυτό το γεγονός είναι πολύ χρήσιμο για την μείωση του υπολογιστικού κόστους. Όπως προαναφέρθηκε, ο χώρος του προσώπου σε μια εικόνα είναι μικρότερος από τον χώρο της εικόνας καθώς τα πρόσωπα εμφανίζουν συγκεκριμένα χαρακτηριστικά διαθέτοντας στατιστικές κανονικότητες. Έτσι ο χώρος του προσώπου μπορεί να παρουσιαστεί με την μορφή βασικότερων ιδιοδιανυσμάτων με χρήση μόνο των βασικών χαρακτηριστικών των προσώπων και παραλείποντας εκείνα που εκδηλώνονται σαν θόρυβος. Η επόμενη εικόνα παρουσιάζει ένα πρόβλημα ταξινόμησης δυο κλάσεων. Δεξιά η συνιστώσα δεν συνεισφέρει στην ταξινόμηση των δυο κλάσεων, ενώ αριστερά η συνιστώσα που αντιπροσωπεύει την διακύμανση των δεδομένων επιτυγχάνει την ταξινόμηση των δυο κλάσεων Α και Β.



Εικόνα 47. Υπολογισμός Βασικών συνιστωσών για την βέλτιστη και αποδοτικότερη ταξινόμηση

Σε μια διαδικασία αναγνώρισης προσώπων το σύνολο των ιδιοπροσώπων που προκύπτει είναι πάντοτε ίσο με τον αριθμό Μ των εικόνων που εισάγονται στο σύστημα κατά την διαδικασία της εκπαίδευσης. Γνωρίζοντας πως η αναπαράσταση ενός προσώπου είναι δυνατή με την χρήση κάποιων βασικών ιδιοπροσώπων, τότε μπορεί κανείς να υπολογίσει τα πρόσωπα επιλέγοντας τα Κ (Κ<Μ) βασικότερα ιδιοδιανύσματα (ιδιοπρόσωπα). Ο αριθμός Κ των σημαντικότερων ιδιοπροσώπων προκύπτει εμπειρικά και βάσει των δεδομένων που εισάγονται στη βάση προσώπων. Σημειώνεται πως για την δημιουργία τέτοιων δεδομένων είναι σημαντικό όλες οι εικόνες να είναι κεντραρισμένες και θα πρέπει να έχουν τις ίδιες διαστάσεις [m x n].

Εικόνα 48. Παράδειγμα κεντραρισμένων εικόνων προσώπου

3.1.1 Διαδικασία Εκπαίδευσης

Οι Turk & Pentland [54] παρουσιάζουν την μέθοδό τους ως εξής:

1. Εισαγωγή των M εικόνων $[I_1, I_2, I_3 \ldots\ldots I_M]$ εκπαίδευσης
2. Αναπαράσταση κάθε εικόνας I_ι ως ένα μονοδιάστατο διάνυσμα Γ_i διαστάσεων [m*n x 1]
3. Υπολογισμός της μέσης εικόνας (μέσο διάνυσμα προσώπου

$$\Psi = \frac{1}{M}\sum_{\iota=1}^{M}\Gamma_\iota$$

4. Υπολογισμός της διαφοράς Φ_i των εικόνων εκπαίδευσης Γ_i με την μέση εικόνα Ψ. Η διαφορά αυτή έχει σκοπό την αφαίρεση των κοινών χαρακτηριστικών προσώπου για την εξαγωγή των ιδιαίτερων χαρακτηριστικών

$$\Phi_{i=}\Gamma_i - \Psi$$

5. Υπολογισμός του πίνακα συμμεταβλητότητας (Covariance Matrix)

$$C = AA^T$$

όπου A=$[\Phi_1, \Phi_2, \Phi_3, \ldots. \Phi_M]$. Ο υπολογισμός του πίνακα συμμεταβλητότητας δημιουργεί έναν πολύ μεγάλο σε διαστάσεις πίνακα [m*n x M]. Για εικόνες προσώπου με διαστάσεις 112x92 θα δημιουργηθεί ένας πίνακας 10.304*10.304. Δεδομένου του υπολογιστικού κόστους, για τον υπολογισμό του πίνακα συμμεταβλητότητας υπολογίζεται ο πίνακας A^TA ο οποίος θα έχει διαστάσεις [M x M].

6. Υπολογισμός των ιδιοτιμών και ιδιοδιανυσμάτων του πίνακα A^TA. Το διάνυσμα των ιδιοτιμών λ_i που προκύπτουν έχουν διάσταση [M x 1] και το ιδιοδιάνυσμα v_i [M*M]. Οι ιδιοτιμές ταξινομούνται από το μεγαλύτερο στο μικρότερο για την εύρεση των ισχυρότερων ιδιοδιανυσμάτων v_i. Τελικά τα ιδιοπρόσωπα θα έχουν διάσταση [m*n x M] και μπορούν να υπολογιστούν από την σχέση:

$$u_i = Av_i$$

Όπως απέδειξαν οι Kirby & Sirovich [53] κάθε πρόσωπο Φ_i αποτελεί έναν γραμμικό συνδυασμό ιδιοπροσώπων με την χρήση κατάλληλων βαρών. Έτσι, η «εκπαίδευση» του συστήματος γίνεται με τον υπολογισμό των βαρών w_j των ιδιοπροσώπων:

$$w_j = u_j^T\Phi_i$$

Στο σημείο αυτό γίνεται η ταξινόμηση των k κυριότερων ιδιοδιανυσμάτων αναλόγως των ιδιοτιμών τους. Έτσι προκύπτουν οι k βασικότερες συνιστώσες (eigenvectors), στις οποίες τα ιδιοδιανύσματά τους αντιστοιχούν στις k μεγαλύτερες ιδιοτιμές. Τα ιδιοπρόσωπα με μικρές ιδιοτιμές v_i μπορούν να παραλειφθούν, καθώς περιγράφουν ένα μικρό μόνο

μέρος των ισχυρών χαρακτηριστικών των προσώπων. Ο αριθμός των ισχυρότερων ιδιοδιανυσμάτων προκύπτει εμπειρικά. Κατά την διάρκεια της εκπαίδευσης, αναγνωρίζοντας κανείς τις μεγαλύτερες ιδιοτιμές μπορεί να βγάλει συμπέρασμα για την βέλτιστη επιλογή του αριθμου Κ. Τέλος, η κάθε εικόνα Φ της εκπαίδευσης μπορεί να εκφραστεί στον ιδιοχώρο (eigenspace) από την παρακάτω σχέση:

$$\Omega_i = \begin{bmatrix} w_1^j \\ w_2^j \\ \dots \\ w_K^j \end{bmatrix}$$

3.1.2 Διαδικασία Αναγνώρισης

Για την αναγνώριση ενός προσώπου μιας νέας εικόνας γίνεται η σύγκριση εικόνων προβαλλόμενων στον ιδιοχώρο. Η νέα εικόνα προς αναγνώριση εισάγεται στο σύστημα και αναπαραστάται σε ένα μονοδιάστατο δυάνυσμα. Έπειτα αφαιρείται από αυτή το μέσο πρόσωπο για την απομάκρυση των γενικών χαρακτηριστικών των προσώπων και την ανάδειξη των ιδιαίτερων χαρακτηρηστικών. Η διαδικασία αυτή κανονικοποιεί (normalize) την νέα εικόνα:

$$\Phi_= \Gamma - \Psi$$

Στην συνέχεια υπολογίζονται τα βάρη w_j της εικόνας Φ ώστε τελικά η εικόνα να προβήθεί στον ιδιοχώρο Ω:

$$w_i = u_i^T \Phi_i$$

έτσι ώστε

$$\Omega = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_3 \end{bmatrix}$$

Εφόσον η εικόνα έχει προβληθεί στον ιδιοχώρο, η ταξινόμησή της γίνεται με υπολογισμό της ελάχιστης Ευκλείδιας απόστασης της ζητούμενης εικόνας με την κάθε εικόνα που έχει λάβει μέρος στην διαδικασία της εκπαίδευσης:

$$e_r = min\|\Omega - \Omega_i\|$$

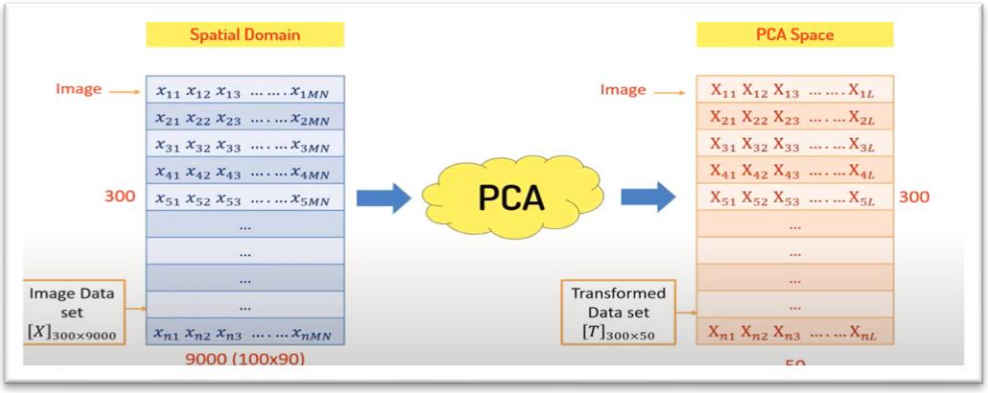
Η τελική αναγνώριση επιτχάνεται με την επιλογή ενός κατωφλίου Τ, το οποίο προκύπτει εμπειρικά από τις παρατηρήσεις:

$$e_r < T_r$$

3.1.3 Πρακτική περιγραφή της Αναγνώρισης προσώπου με τον αλγόριθμο Eigen Faces

Η αναγνώριση προσώπου απαιτεί την εκπαίδευση ενός συστήματος όπου θα προκύψουν οι σημαντικότερες πληροφορίες/χαρακτηριστικά για το κάθε πρόσωπο/παράδειγμα. Σε αυτό το στάδιο τα δεδομένα εκπαίδευσης (training set) περιγράφονται από έναν

ενιαίο πίνακα. Έτσι για ένα σετ 300 εικόνων εκπαίδευσης η κάθε 2D εικόνα [m=100, n=90] μπορεί να αναπαρασταθεί ως ένα διάνυσμα [1,(mxn)]. Με αυτόν τον τρόπο προκύπτει ένας πίνακας χωρικής αναπαράστασης εικόνων 300x9000. Δεδομένου του προβλήματος της διαστασιμότητας, χρησιμοποιείται η μέθοδος ανάλυσης κύριων συνιστωσών για να μετατρέψει ένα σύνολο πιθανώς συσχετισμένων μεταβλητών σε ένα μικρότερο σύνολο ασυσχέτιστων μεταβλητών. Στόχος της διαδικασίας αυτής είναι αρχικά η μείωση της διαστασιμότητας, δηλαδή η εύρεση εκείνων των χαρακτηριστικών τα οποία συνεισφέρουν την μέγιστη πληροφορία, και στην συνέχεια η ταξινόμηση αυτής της πληροφορίας. Ας θεωρήσουμε, χάριν παραδείγματος, πως μπορούμε να έχουμε αρκετά καλά αποτελέσματα αναγνώρισης με την επιλογή των 50 ισχυρότερων ιδιοδιανυσμάτων. Έτσι επιλύεται και το πρόβλημα της διαστασιμότητας, καθώς από τον χώρο της εικόνας με διαστάσεις (300x9000) το σετ μπορεί να αναπαρασταθεί στον χώρο των κύριων συνιστωσών με έναν πίνακα με διαστάσεις (300x50).



Εικόνα 49. Μετατροπή από τον χώρο της εικόνας στον υποχώρο των κύριων συνιστωσών

Πρακτικά ο αλγόριθμος έχεις ως εξής:

1. Οι εικόνες $\Gamma=\{\Gamma_1,\Gamma_2,\ldots\Gamma_n\}$ των *n* 2D παραδειγμάτων μετατρέπονται σε ένα διάνυσμα (1D). Ο πίνακας εικόνων που προκύπτει έχει διαστάσεις 300x9000.
2. Για το σύνολο n των εικόνων των παραδειγμάτων x υπολογίζεται το μέσο πρόσωπο Ψ. Η εικόνα Ψ έχει διαστάσεις 1x9000.
3. Υπολογίζεται η διαφορά Φ όλων των εικόνων με την μέση εικόνα Ψ και παρουσιάζονται σε έναν ενιαίο πίνακα A με διαστάσεις 9000x300
4. Υπολογίζεται ο πίνακας συμμεταβλητότητας (covariance matrix) C. Δεδομένου ότι ο πίνακας $C = AA^T$ θα έχει διαστάσεις 9000x9000, υπολογίζεται ο πίνακας $C = A^TA$ ο οποίος θα έχει διαστάσεις 300x300.
5. Υπολογίζονται τα ιδιοδιανύσματα u_i και οι ιδιοτιμές v_i του πίνακα συμμεταβλητότητας $u_i = Av_i$ Ο πίνακας ιδιοτιμών που προκύπτει έχει διαστάσεις 300x1 και ο πίνακας ιδιοδιανυσμάτων έχει διαστάσεις 300x300.
6. Γίνεται η κανονικοποίηση των ιδιοδιανυσμάτων ώστε $\|v_i\|=1$

$$u_i = \frac{u_i}{\sqrt{v_i}}$$

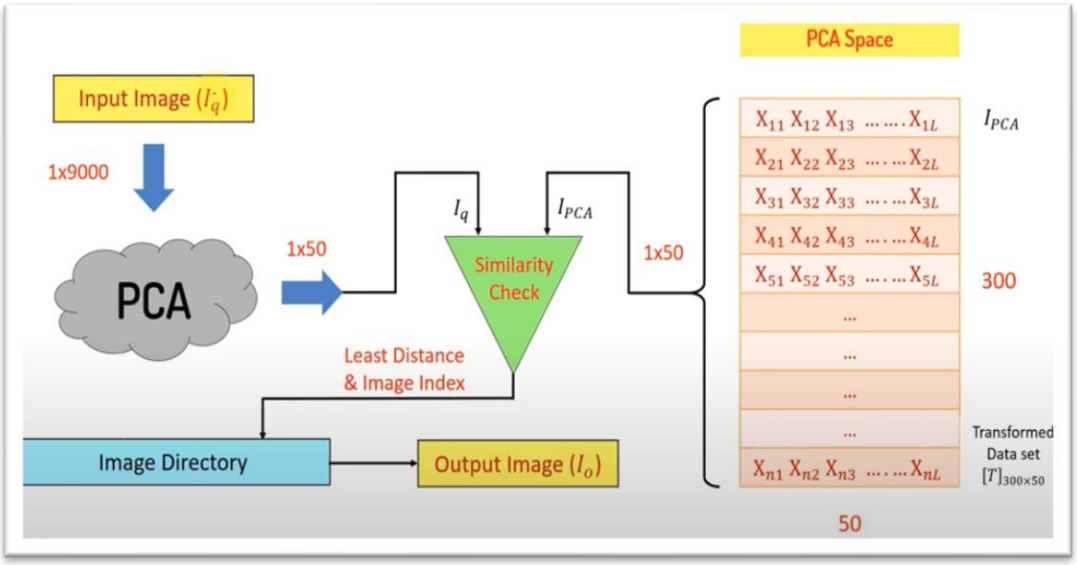
7. Γίνεται η ταξινόμηση των k κυριότερων ιδιοδιανυσμάτων ανάλογα με τις ιδιοτιμές τους. Έτσι προκύπτουν οι k βασικότερες συνιστώσες (eigenvectors), στις οποίες τα ιδιοδιανύσματά τους αντιστοιχούν στις k μεγαλύτερες ιδιοτιμές .
8. Γίνεται ο υπολογισμός των ιδιοπροσώπων. Τα ιδιοπρόσωπα δημιουργούν έναν πίνακα με διαστάσεις 9000 x k.

$$u_j = A * u_i$$

9. Υπολογίζονται τα βάρη. Ο πίνακας που προκύπτει έχει διαστάσεις kx300.

$$w_i = u_i^T \Phi_i$$

Για την αναγνώριση προσώπου η νέα εικόνα Φ που προκύπτει από την διαφορά της εικόνας με την μέση εικόνα προβάλλεται στον ιδιοχώρο. Η εικόνα αυτή μετασχηματίζεται στον χώρο των βαρών με διαστάσεις 1xk και η αναγνώριση γίνεται με τον υπολογισμό της Ευκλείδειας απόστασης του πίνακα βαρών της εισαγόμενης εικόνας με τον πίνακα των βαρών που προέκυψε στο τελευταίο βήμα.



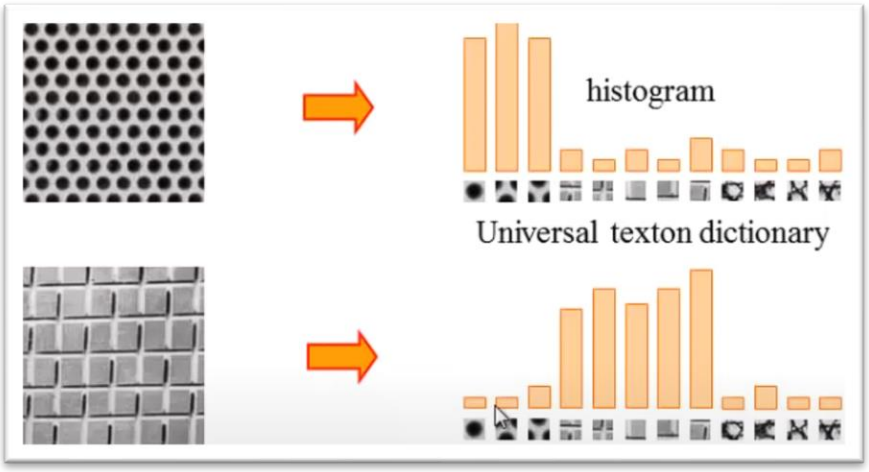
Εικόνα 50. Διαδικασία αναγνώρισης προσώπου

Η μέθοδος αυτή επιτυγχάνει πολύ ικανοποιητικά αποτελέσματα που φτάνουν το 95%.

3.2 Αναγνώριση προσώπου με Bag of Visual Features

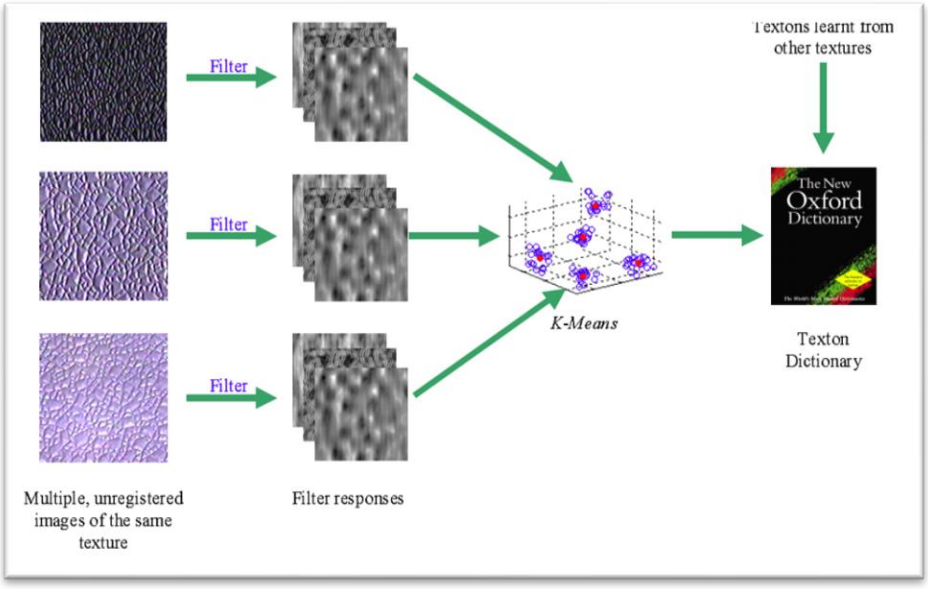
Μια εναλλακτική μέθοδος για την ανίχνευση και αναγνώριση αντικειμένων, συνεπώς και προσώπων, είναι η μέθοδος *Bag of Words*. Πηγή έμπνευσης ήταν η επεξεργασία φυσικών γλωσσών (natural language processing), και η μέθοδος εφαρμόσθηκε αρχικά για την κατηγοριοποίηση κειμένων από τους Joachim et al. [55]. Στόχος της κατηγοριοποίησης κειμένου είναι ο καταμερισμός των εγγράφων σε έναν καθορισμένο αριθμό προκαθορισμένων κατηγοριών. Η προσέγγιση αυτή υλοποιείται με την εκπαίδευση ταξινομητών από παραδείγματα (κειμένων). Οι λέξεις, ως μια σειρά χαρακτήρων (strings of characters), μετασχηματίζονται σε χαρακτηριστικά με συγκεκριμένες τιμές καθώς «ποσοτικοποιούνται» δημιουργώντας ένα οπτικό λεξικό (bag of visual words dictionary) στο οποίο κωδικοποιείται η κάθε λέξη με την επαναληψιμότητά της (συχνότητα εμφάνισης). Ονομάζεται Bag of Words («σάκος» λέξεων), επειδή οι τυχόν πληροφορίες σχετικά με τη σειρά ή τη δομή των λέξεων στο έγγραφο απορρίπτονται και το μοντέλο ασχολείται μόνο με το αν υπάρχουν γνωστές λέξεις στο έγγραφο και όχι για την συγκεκριμένη θέση τους. Τέλος, ο ταξινομητής εκπαιδεύεται με μηχανές διανυσμάτων υποστήριξης για την επίτευξη της κατηγοριοποίησης κειμένου.

Μια σειρά εργασιών που αφορούν την αναπαράσταση υφής στις εικόνες για την σχετική ταξινόμησή τους ([57], [58], [59]) παρουσιάζουν μεγάλη επιτυχία ταξινόμησης. Οι εργασίες αυτές ερευνούν την ταξινόμηση της υφής σε μεμονωμένες εικόνες, οι οποίες προκύπτουν από διαφορετικές λήψεις και συνθήκες φωτισμού. Οι Manik & Zisserman [58] εφαρμόζουν μια στατιστική προσέγγιση για την μοντελοποίηση της υφής χρησιμοποιώντας διάφορα φίλτρα σε εικόνες εκπαίδευσης για να εξάγουν αποκρίσεις φίλτρων (filter responses).

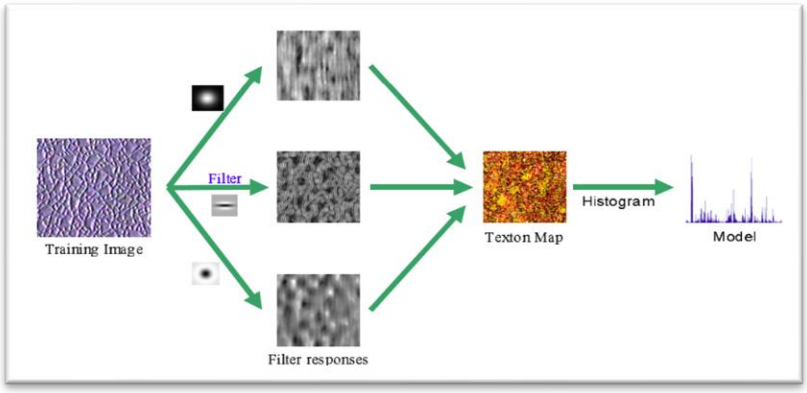


Εικόνα 51. Κωδικοποίηση οπτικών χαρακτηριστικών

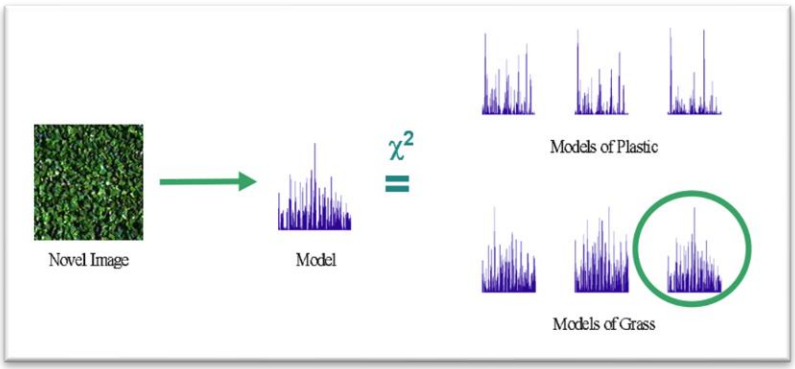
Οι αποκρίσεις στην συνέχεια ομαδοποιούνται με τον υπολογισμό των κέντρων των συστάδων απόκρισης φίλτρου (textons). Τα textons των διαφορετικών κατηγοριών συνδυάζονται εν τέλει για την δημιουργία ενός οπτικού λεξικού texton. Έτσι, η μοντελοποίηση της υφής υλοποιείται από την κοινή κατανομή της πιθανότητας (joint probability distribution), η οποία παρουσιάζεται υπό την μορφή ιστογράμματος συχνότητας των κέντρων των συστάδων απόκρισης φίλτρου (textons). Στο στάδιο της ταξινόμησης ακολουθείται η ίδια διαδικασία για την δημιουργία του αντίστοιχου ιστογράμματος της νέας προς αναγνώριση εικόνας. Το ιστόγραμμα που προκύπτει συγκρίνεται με τα ιστογράμματα του μοντέλου μέσω του αλγορίθμου των εγγύτερων γειτόνων (K-Nearest Neighbour) algorithm), και η τελική ταξινόμηση γίνεται με τον στατιστικό έλεγχο χ^2. Αναλυτικότερα, η διαδικασία αυτή περιγράφεται από τις επόμενες εικόνες.



Εικόνα 52. Δημιουργία λεξικού Texton: Σε εικόνες κάθε κατηγορίας από το σετ εκπαίδευσης εφαρμόζεται συνέλιξη με ένα σετ διάφορων φίλτρων. Το αποτέλεσμα της απόκρισης φίλτρων συγκεντρώνεται και ομαδοποιείται με την χρήση του αλγορίθμου (K-Mean) παράγοντας τα Texton. Τα Texton από τις διαφορετικές κλάσεις δημιουργούν εν τέλει το λεξικό Texton.



Εικόνα 53. Μοντελοποίηση: Για κάθε εικόνα από το σετ εκπαίδευσης δημιουργείται το αντίστοιχο μοντέλο. Στην εικόνα γίνεται συνέλιξη για την παραγωγή αποκρίσεων φίλτρου οι οποίες αντιστοιχούνται με τα Texton που έχουν προκύψει στο προηγούμενο στάδιο. Το ιστόγραμμα των Texton αναπαριστά την συχνότητα εμφάνισης του κάθε Texton στην εικόνα δημιουργώντας το μοντέλο που αντιστοιχεί στην εικόνα εκπαίδευσης.



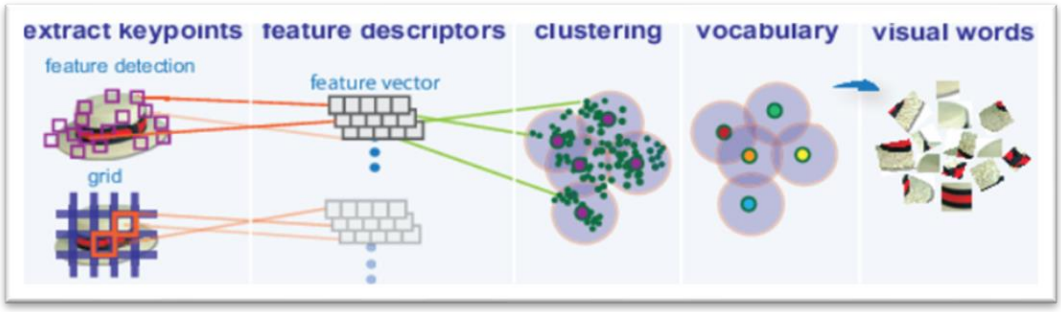
Εικόνα 54. Ταξινόμηση: Για κάθε νέα εικόνα προς ταξινόμηση υπολογίζεται το αντίστοιχο ιστόγραμμα Texton και με την χρήση του αλγορίθμου εγγύτερου γείτονα επιλέγεται το πιο «κοντινό» μοντέλο (υπό την έννοια του ελέγχου χ^2). Έτσι η νέα εικόνα δηλώνεται ότι ανήκει στην κατηγορία υφής του πλησιέστερου μοντέλου.

3.2.1 Αρχιτεκτονική της μεθόδου Bag of Visual Features

Λόγω της μεγάλης επιτυχίας στην κατηγοριοποίηση κειμένων, σε συνδυασμό με την παραπάνω προσέγγιση αναγνώρισης υφής, η αναπαράσταση Bag of Words (BoW) έγινε πολύ δημοφιλής μέθοδος για την αναπαράσταση του περιεχομένου μιας εικόνας και εφαρμόσθηκε από τις αρχές του 2000 με μεγάλη επιτυχία στην ταξινόμηση αντικειμένων όπως και την αναγνώριση προσώπων. Για την ανίχνευση αντικειμένων αυτή η μέθοδος συναντάται στην βιβλιογραφία και ως bag of features ή bag of visual features (BoVF). Η γενική προσέγγιση της μεθόδου BoVF για την ταξινόμηση αντικειμένων περιλαμβάνει μια αλληλουχία διεργασιών, όπου το αποτέλεσμα της προηγούμενης διεργασίας αποτελεί είσοδο για την επόμενη, ενώ όλες οι διεργασίες είναι ανεξάρτητες μεταξύ τους. Σε αντίθεση με τις προσεγγίσεις που έχουν αναφερθεί προηγουμένως, η BoVF δεν χρησιμοποιεί λέξεις ή φίλτρα για την δημιουργία του οπτικού λεξικού αλλά εξάγει τμήματα της εικόνας που περιγράφονται από *περιγραφείς χαρακτηριστικών*. Με δεδομένο εισόδου τις εικόνες εκπαίδευσης, αρχικά υπολογίζονται οι *τοπικοί περιγραφείς* οι οποίοι στην συνέχεια χρησιμοποιούνται από το σύστημα για την εκμάθηση της αναπαράστασης των δεδομένων, δηλαδή την εκμάθηση ενός λεξικού χαρακτηριστικών για την κωδικοποίηση δεδομένων. Στην συνέχεια εφαρμόζεται μια χωρική πυραμιδική συγκέντρωση στα κωδικοποιημένα δεδομένα ώστε να εξαχθεί το τελικό χαρακτηριστικό διάνυσμα που θα εισαχθεί εν τέλει σε έναν ταξινομητή.

Στην αναγνώριση αντικειμένων και προσώπων ειδικότερα, στο πρώτο στάδιο του αλγορίθμου εξάγονται τμήματα (patches) της εικόνας με καθορισμένες διαστάσεις και σταθερή απόσταση (stride) με σκοπό τον υπολογισμό και την εξαγωγή των περιγραφέων τους. Οι τοπικοί αυτοί περιγραφείς αποτελούνται από διανύσματα τα οποία διαχειρίζονται την διατήρηση των χαρακτηριστικών ιδιοτήτων των χαρακτηριστικών. Στην ουσία είναι διανύσματα χαρακτηριστικών που είναι αναλλοίωτα σε μεταβολές κλίμακας της εικόνας, στην ένταση της φωτεινότητας, τις στροφές και τις διακυμάνσεις των τμημάτων της εικόνας. Οι συνηθέστεροι και αποτελεσματικότεροι περιγραφείς χαρακτηριστικών είναι οι Scale-Invariant Feature Transform (SIFT) [43], Histograms of oriented Gradients (HoG) [60] και Speeded-Up Robust Features (SURF) [61]. Με την εξαγωγή τέτοιων περιγραφέων, όπως θα αναλυθούν παρακάτω, επιτυγχάνεται μια πληρέστερη και ισχυρή αναπαράσταση των μοτίβων που υπάρχουν στα τμήματα της εικόνας σε σχέση με την απλούστερη αναπαράστασή τους στον χώρο των εικονοστοιχείων. Τα χαρακτηριστικά που προκύπτουν σε αυτό το στάδιο αναφέρονται ως *χαμηλού επιπέδου*, καθώς υπολογίζονται και εξάγονται από τους τοπικούς περιγραφείς. Στο δεύτερο στάδιο υπολογίζονται τα χαρακτηριστικά μεσαίου επιπέδου (mid-level), καθώς πρόκειται για χαρακτηριστικά που προκύπτουν από την εκμάθηση των αναπαραστάσεων από τα δεδομένα. Σε αυτό το στάδιο γίνεται στην ουσία η δημιουργία του λεξικού μάθησης (learning codebook) ώστε στην συνέχεια, βασισμένη σε αυτό, να ολοκληρωθεί η κωδικοποίηση των χαρακτηριστικών δεδομένων (coding). Η διαδικασία αυτής της κωδικοποίησης δίνει και το όνομα Bag of Visual Features σε αυτή την μέθοδο. Ο όρος bag προκύπτει από την απόρριψη της χωρικής πληροφορίας λόγω της δημιουργίας του λεξικού, ενώ ο όρος visual features ορίζει την οπτικοποίηση των χαρακτηριστικών και την ποσοτικοποίησή τους υπολογισμένη από την επαναληψημότητά τους στην εικόνα. Έτσι θα μπορούσαμε να πούμε πως η μέθοδος αυτή αντιμετωπίζει το πρόσωπο ως μια συλλογή από περιοχές με χαρακτηριστικά στον χώρο της εικόνας, αγνοώντας την δομή τους στον χώρο. *Γενικά* [62], *η διαδικασία κωδικοποίησης των τοπικών περιγραφέων σε μία αρχιτεκτονική Bag-of-Visual-Features υλοποιείται συνήθως μέσω μεθόδων αυστηρού κβαντισμού*[7] *που δημιουργούν το χωρικό ιστόγραμμα χρήσης των στοιχείων του λεξικού (πχ. k-means)* [63]. Στο τρίτο στάδιο γίνεται η χωρική πυραμιδική συγκέντρωση (spatial pyramid pooling) όλων των χαρακτηριστικών χαμηλού επιπέδου που παράχθηκαν στο προηγούμενο στάδιο με στόχο την δημιουργία ενός ενιαίου καθολικού χαρακτηριστικού που θα αντιστοιχεί στην κάθε αντίστοιχη εικόνα του σετ εκπαίδευσης. Σε αυτή την φάση δεν δημιουργείται στην ουσία κάποιο νέο χαρακτηριστικό, με την έννοια κάποιου χαρακτηριστικού υψηλότερης τάξης, απλώς τα χαμηλού επιπέδου χαρακτηριστικά συγκεντρώνονται και ενσωματώνονται μαζί σε ομάδες για την διανυσματική αναπαράσταση των τμημάτων (patches) της εικόνας. Έτσι, το τελικό διάνυσμα χαρακτηριστικών όλων των εικόνων του σετ εκπαίδευσης δημιουργείται από την «ένωση» όλων των διανυσμάτων που προκύπτουν από την διαδικασία συγκέντρωσης σε κάθε τμήμα της εικόνας. Τέλος, ο ταξινομητής εκπαιδεύεται με μηχανές διανυσμάτων υποστήριξης για την επίτευξη του στόχου, την αναγνώριση του προσώπου.

[7] Ο κβαντιστής είναι η διάταξη εκείνη η οποία αντιστοιχίζει τις τιμές των δειγματοληπτούμενων χαρακτηριστικών σε διακριτές κατηγορίες



Εικόνα 55. Διαδικασία δημιουργίας οπτικού λεξικού

Κεφάλαιο 4. Περιγραφή Αλγορίθμου

4.1 Η βάση ORL

Για την εκπαίδευση ενός συστήματος αναγνώρισης προσώπου θα χρησιμοποιηθεί η βάση ORL. Η βάση αυτή δημιουργήθηκε από την εταιρία AT&T και περιλαμβάνει 400 μετωπικές εικόνες προσώπων 40 ατόμων με 10 εικόνες ανά άτομο. Οι εικόνες έχουν διαστάσεις 112x92, είναι κλίμακας του γκρι (grayscale) και έχουν σκοτεινό ομοιογενές φόντο. Οι εικόνες έχουν καταγραφεί σε μεταβαλλόμενο φωτισμό, υπό διαφορετικές γωνίες λήψης (υπάρχει στροφή του προσώπου που φτάνει τις 20°, με διακύμανση της κλίμακας έως 10%) και διαφορετικές εκφράσεις προσώπου. Οι εικόνες της βάσης είναι της μορφής .pgm και έχουν τοποθετηθεί σε φακέλους με όνομα s1 έως s40.

Εικόνα 56. Παράδειγμα εικόνων προσώπου από την βάση ORL

Με αφετηρία αυτό το σύνολο εικόνων, η βάση ORL εμπλουτίζεται με νέες εικόνες που έχουν προκύψει από βίντεο με έγχρωμη εικόνα. Για την δημιουργία του σετ εικόνων προσώπου, το πρόσωπο αρχικά ανιχνεύεται και στην συνέχεια η εικόνα που προκύπτει μετατρέπεται σε εικόνα grayscale και, τέλος, προσαρμόζεται στις διαστάσεις των εικόνων της βάσης ORL. Στο σύνολο πρόκειται για 50 εικόνες 5 ατόμων (10 εικόνες για κάθε άτομο).

4.2 Ανίχνευση Προσώπου

Για να επιτευχθεί η αναγνώριση προσώπου σε μια εικόνα αρχικά χρειάζεται να γίνει η ανίχνευση του προσώπου ή των προσώπων σε μια εικόνα. Η ανίχνευση των προσώπων γίνεται με την χρήση ενός ανιχνευτή cascade (vision.CascadeObjectDetector) στο περιβάλλον matlab. Ο ανιχνευτής αυτός βασίζεται στον αλγόριθμο των Viola & Jones και δίνει την δυνατότητα ανίχνευσης προσώπων, σώματος, ματιών, προφίλ και μύτης:

$$faceDetector = vision.CascadeObjectDetector('MergeThreshold', 10);$$

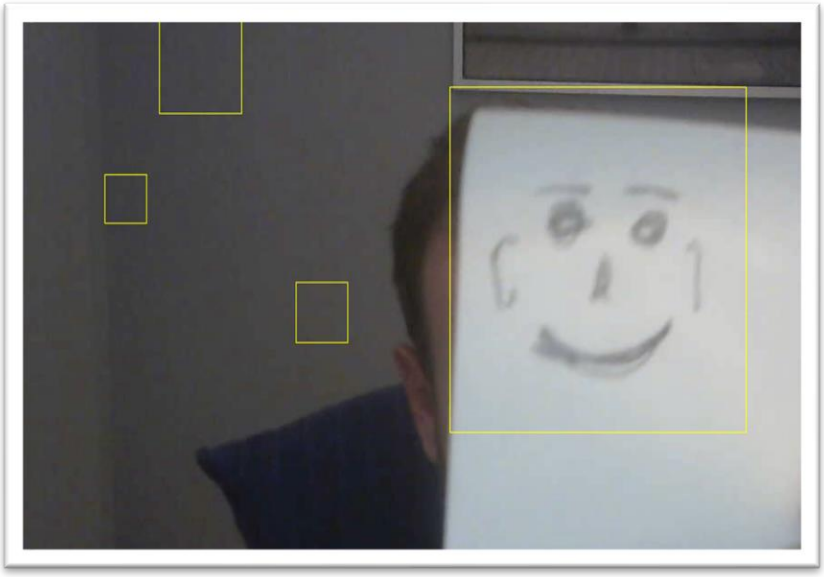
Ο αλγόριθμος CascadeObjectDetector δίνει την δυνατότητα στον χρήστη να χειρίζεται την «ευαισθησία» της ανίχνευσης με τον ορισμό ενός κατωφλίου. Σε δοκιμές που έγιναν με διαφορετικά κατώφλια από 1 έως 10 προκύπτουν κάποιες αξιοσημείωτες παρατηρήσεις. Για κατώφλι ίσο με 1 ο ανιχνευτής είναι πολύ ευαίσθητος καθώς εντοπίζει μεγάλο αριθμό ψευδώς θετικών προσώπων σε μια εικόνα. Ο αλγόριθμος Viola & Jones έχει εκπαιδευτεί με την χρήση των χαρακτηριστικών Haar, τα οποία στην ουσία εντοπίζουν χαρακτηριστικά σημεία από τον υπολογισμό της διαφοράς της έντασης, και στην συνέχεια

ο ταξινομητής εκπαιδεύεται με σκοπό την δημιουργία ενός μοντέλου στο οποίο τα χαρακτηριστικά αυτά παρουσιάζουν μια συγκεκριμένη αλληλουχία, όπως αν εντοπίζονται μάτια, φρύδια, μύτη κλ.π. Το γεγονός αυτό, σε συνδυασμό με ένα πολύ χαμηλό κατώφλι, μπορεί να αυξήσει πολύ σημαντικά την ψευδή ανίχνευση προσώπων.

Μοντέλο ταξινόμησης	**Διαστάσεις εικόνων που χρησιμοποιήθηκαν κατά την εκπαίδευση**
'FrontalFaceCART'(Default)	[20 20]
'FrontalFaceLBP'	[24 24]
'UpperBody'	[18 22]
'EyePairBig' 'EyePairSmall'	[1145] [5 22]
'LeftEye' 'RightEye'	[12 18]
'LeftEyeCART' 'RightEyeCART'	[20 20]
'ProfileFace'	[20 20]
'Mouth'	[15 25]
'Nose'	[15 18]

Εικόνα 57. Μοντέλα ταξινόμησης Cascade object detector στο matlab

Όπως παρατηρείται στην επόμενη εικόνα, για κατώφλι 1 ανιχνεύτηκαν 4 ψευδώς θετικά πρόσωπα. Αξίζει να σημειωθεί ότι η ζωγραφιά ενός προσώπου ανιχνεύεται ως πρόσωπο αφού στην συγκεκριμένη φωτογραφία ο αλγόριθμος εντοπίζει χαρακτηριστικά/διαφορές που θυμίζουν πρόσωπο, για παρόμοια όμως ζωγραφιά χωρίς «μύτη» ή «στόμα», ή για ζωγραφιά προσώπου όπου δεν διατηρούνται οι αναλογίες μήκους των χαρακτηριστικών (πχ. μάτια σε μεγάλη απόσταση μεταξύ τους) δεν ανιχνεύεται κανένα πρόσωπο. Όσο αυξάνεται το κατώφλι ο αριθμός των ψευδώς θετικών προσώπων μειώνεται.



Εικόνα 58. Ανίχνευση ψευδώς θετικών προσώπων για κατώφλι 1

Για κατώφλι 4, που είναι το προκαθορισμένο κατώφλι του CascadeObjectDetector, ο αριθμός των ψευδώς θετικών προσώπων μειώνεται σημαντικά, καθώς πλέον δεν ανιχνεύονται πρόσωπα στο φόντο. Στο παράδειγμα με την ζωγραφιά όμως, ο ταξινομητής συνεχίζει να «μπερδεύεται». Παρατηρήθηκε ότι αν η ζωγραφιά είναι πολύ κοντά στην κάμερα δεν ανιχνεύεται κανένα πρόσωπο, όσο απομακρύνεται όμως γίνονται αρκετές ανιχνεύσεις ψευδώς θετικών προσώπων. Παρατηρήθηκε επίσης πως και φωτισμός δεν επηρεάζει σημαντικά την ανίχνευση προσώπου καθώς για το προκαθορισμένο κατώφλι

4 γίνεται ανίχνευση προσώπου. Στην παρακάτω εικόνα γίνεται ανίχνευση προσώπου και ανίχνευση ψευδώς θετικού προσώπου σε συνθήκες χαμηλού φωτισμού. Στο παραπάνω παράδειγμα δεν προέκυψε κανένα αποτέλεσμα ψευδώς θετικού προσώπου στο φόντο.

Τέλος, για κατώφλι 10 δεν παρατηρήθηκε καμία ανίχνευση ψευδούς προσώπου. Παρατηρήθηκε όμως αδυναμία του ανιχνευτή να ανιχνέυσει ένα πρόσωπο καθ' όλη την διάρκεια ενός βίντεο δεδομένου ότι το μοντέλο είναι πολύ πιο αυστηρό. Το προκαθορισμένο CascadeObjectDetector χρησιμοποιεί το μοντέλο ταξινόμησης FrontalFaceCART. Η εκπαίδευσή του έγινε από μετωπικές εικόνες προσώπων διαστάσεων 20x20 με αποτέλεσμα σε στροφές του προσώπου να παρατηρείται αδυναμία ανίχνευσης.

Εικόνα 59. Ανίχνευση προσώπου σε σκοτεινή εικόνα. Κατώφλι 4

Εικόνα 60. Παράδειγμα Αδυναμίας ανίχνευσης σε πρόσωπα με στροφή

Για την εξερεύνηση των μεθόδων αναγνώρισης και της επιτυχίας που παρέχουν όχι μόνο σε συγκεκριμένες εικόνες από βάσεις δεδομένων προσώπου έγινε εισαγωγή νέων εικόνων προσώπου. Προστέθηκαν πέντε νέα πρόσωπα με 10 εικόνες για το κάθε ένα. Η εξαγωγή των εικόνων προσώπου έγινε με την χρήση του CascadeObjectDetector σε βίντεο έγχρωμων εικόνων. Η λήψη έγινε υπό συνθήκες σταθερού φωτισμού και σε σύνθετο φόντο. Μετά την ανίχνευση των προσώπων οι εικόνες μετατράπηκαν σε grayscale με διαστάσεις 112x92. Τέλος, δημιουργούμε ένα σύνολο εικόνων για την εκπαίδευση του συστήματος (training set) και ένα σύνολο για τον τελικό έλεγχο. Για το σύνολο εκπαίδευ-

σης θα αξιοποιηθούν τα 9/10 των εικόνων για κάθε άτομο, για δε τον τελικό έλεγχο (validation) θα χρησιμοποιηθεί το 1/10. Έτσι, το τελικό σετ εκπαίδευσης θα αποτελείται από 405 εικόνες και το σετ ελέγχου από 45 εικόνες.

Εικόνα 61. Εικόνες προσώπου που εισήχθησαν στη βάση ORL

4.3 Αναγνώρισης προσώπου με Eigen Faces

Η πρώτη μέθοδος αναγνώρισης προσώπου που εφαρμόστηκε είναι με την χρήση των ιδιοπροσώπων [53], [37]. Για την εξαγωγή τους δημιουργείται ο ενιαίος πίνακας. Όλες οι εικόνες μετατρέπονται σε ένα διάνυσμα 1x10.304. Συνεπώς ο ενιαίος πίνακας θα έχει διαστάσεις [405x10304]. Για τον υπολογισμό της μέσης εικόνας και του πίνακα συμμεταβλητότητας, ο πίνακας αυτός θα έχει την τελική μορφή στις διαστάσεις 10304x405.

```
%%
imdata = []
for numberofpersons = 1:45% 45 αριθμός ατόμων προς αναγνώρισης
    folder_index_str = num2str(numberofpersons); % φάκελος ατόμου
    folder_loc = strcat('FaceDatabaseORLnew/s',folder_index_str);
    cd(folder_loc);
    ImageData_Temp = zeros(9, 112*92); %Πίνακας εικόνων
    for Imagenumber = 1:9 % Επιλογή αριθμού προσώπων για το κάθε άτομο
        Image = imread(strcat(num2str(Imagenumber),'.pgm')); %ανάγνωση
κάθε εικόνα
        ImageData_Temp(Imagenumber,:) = reshape(Image, [1, 112*92]);
%μονοδιάστατη
    end
    imdata = [imdata; ImageData_Temp]; % τελικός πίνακας εικόνων
    cd('C:\Users\Apollon\Desktop\Matlab_thesis\');
end
imdata = imdata';
```

Κώδικας 1. Ανάγνωση εικόνων βάσης

Εικόνα 62. Παράδειγμα Υπολογισμού της διαφοράς των εικόνων της βάσης με την μέση εικόνα

Κατόπιν υπολογίζεται η μέση εικόνα και η διαφορά των εικόνων προσώπου από αυτήν.

```
%% Υπολογισμός της διαφοράς (Φ) εικόνας με την μέση εικόνα
mean_img = mean(imdata,2); %Υπολογισμός μέσης τιμής όλων των μονοδιάστατων
εικόνων
for i = 1:size(imdata,2)
    imdata(:,i) = imdata(:,i)-mean_img;
end
```

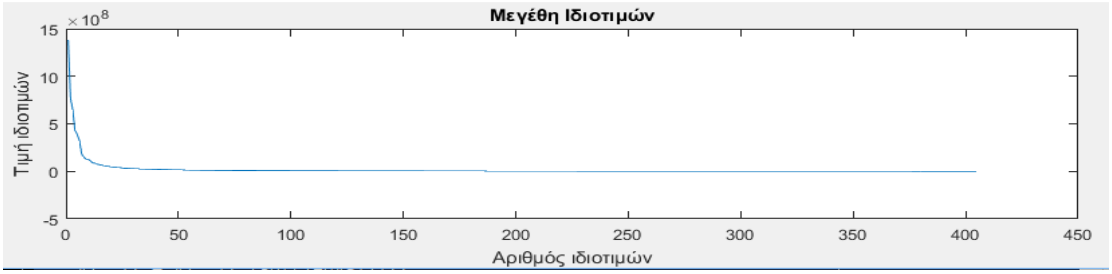
Κώδικας 2. Υπολογισμός της μέσης εικόνας

Στο επόμενο βήμα υπολογίζεται ο πίνακας συμμεταβλητότητας, εξάγονται τα ιδιοδιανύσματα και ταξινομούνται αναλόγως της ιδιοτιμής τους.

```
%% Covariance Matrix
cor_mat = imdata'*imdata;
 %% Eigen vector
[V,D] = eig(cor_mat);
egVal = diag(D);
%% order by largest eigenvalues
egVal = egVal(end:-1:1);
V = V(:,end:-1:1);
```

Κώδικας 3. Υπολογισμός Πίνακα συνδιακύμανσης και ιδιοδιανυσμάτων

Στην επόμενη εικόνα παρουσιάζονται οι μεγαλύτερες ιδιοτιμές. Παρατηρείται ότι ένα σύνολο 405 εικόνων μπορεί να περιγραφεί με την χρήση 20-30 ιδιοδιανυσμάτων.



Εικόνα 63. Μέγεθος ιδιοτιμών για την επιλογή των κυριότερων ιδιοδιανυσμάτων

Έπεται η κανονικοποίηση των ιδιοδιανυσμάτων και η επιλογή των ισχυρότερων.

```
for i = 1:size(V,2)
    V(:,i) = V(:,i)./sqrt(egVal(i));
end
eigFaces = imdata*V;
eigFaces=eigFaces(:,1:K); %επιλογή των K ιδιοδιανυσμάτων
```

Κώδικας 4. Κανονικοποίηση και Επιλογή των ισχυρότερων ιδιοδιανυσμάτων

Η εκπαίδευση του συστήματος αναγνώρισης γίνεται μέσω υπολογισμού των βαρών.

```
k=0;
all_faces=zeros(112*92,45,numberofpictures);
weightsum=zeros(eigenused,numberofpictures*45);
for ni=1:45
    for kimg=1:numberofpictures
        filename=sprintf('FaceDatabaseORLnew/s%i/%i.pgm',ni,kimg);
        images{ni,kimg}=imread(filename);
        one_face=images{ni,kimg};% load

        % Μετασχηματισμός εικόνας σε ένα μονοδιάστατο διάνυσμα
        all_faces(:,ni,kimg)=reshape(one_face,112*92,1);

        % Υπολογισμός διαφοράς εικόνων με τη μέση εικόνα
        one_face=double(all_faces(:,ni,kimg))-mean_img;

        % Υπολογισμός Βαρών
        k=k+1;
        weights=eigFaces'*one_face;
        weightsum(:,k)=weights;
    end
end
```

Κώδικας 5. Εκπαίδευση Συστήματος Αναγνώρισης

Η αναγνώριση του προσώπου για μια νέα εικόνα γίνεται σε πέντε βήματα:

1. Μετασχηματισμός εικόνας σε ένα διάνυσμα (1D)
2. Υπολογισμός της διαφοράς με την μέση εικόνα
3. Υπολογισμός βαρών της νέας εικόνας
4. Υπολογισμός της ελάχιστης Ευκλείδειας απόστασης
5. Έλεγχος κατωφλίου

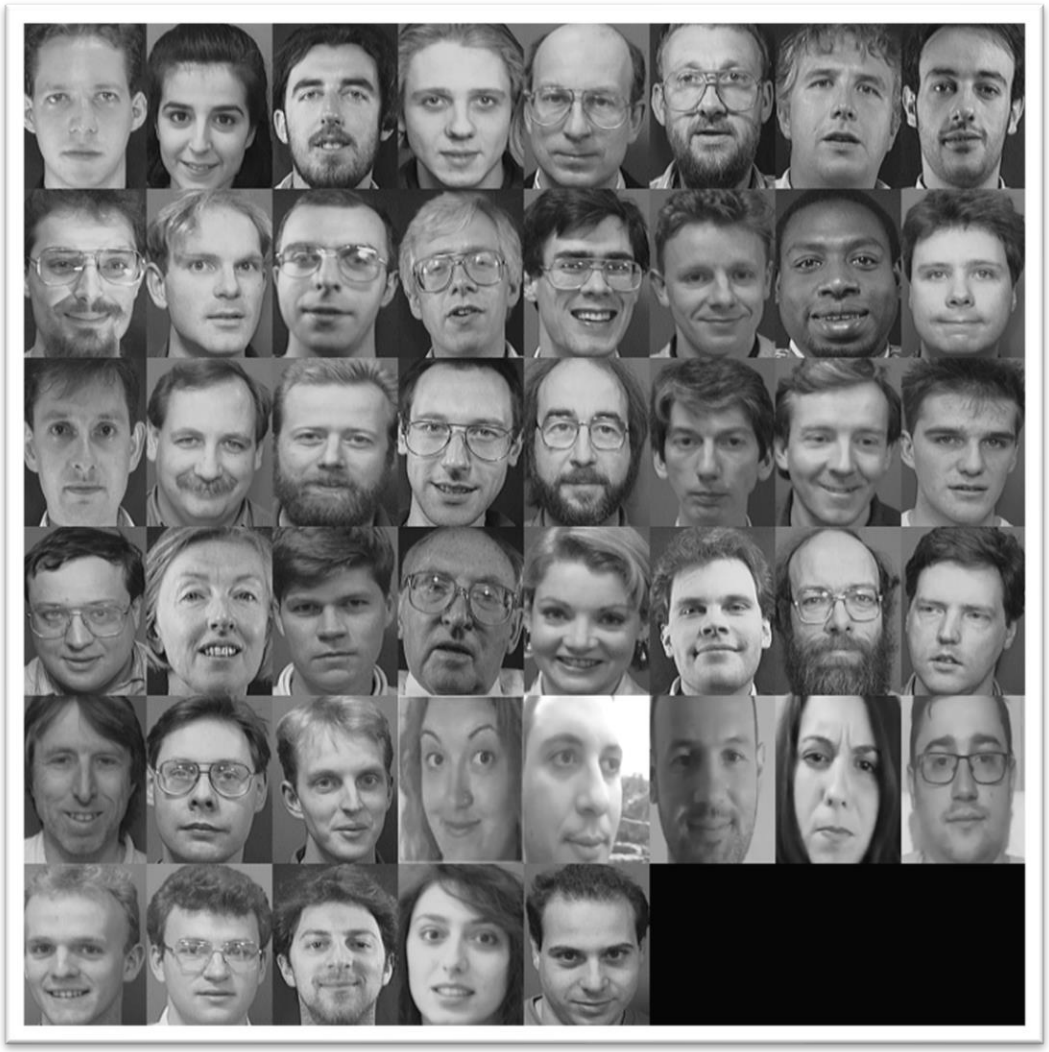
```
for test_category=1:45
    l=0;

test_filename=sprintf('FaceDatabaseORLnew/s%i/%i.pgm',test_category,10);
    test_img=imread(test_filename);
    orgImg=test_img;
    test_img=reshape(test_img,112*92,1);
    test_img=double(test_img)-mean_img;
    % Υπολογισμός βαρών της εικόνας εισαγωγής
    test_wt=eigFaces'*test_img;
    dist=zeros(45,numberofpictures);
    for ni=1:45
        for kimg=1:numberofpictures;
            l=l+1;
            % Υπολογισμός
            dist(ni,kimg)=sum(abs(weightsum(:,l)-test_wt));
        end
    end
    if min_val<T
    [min_val,idx]=min(dist(:));
    [class_pred,col]=ind2sub(size(dist),idx);
    sprintf('Test category and predicted class are %i, %i',test_category,
class_pred);
        if test_category==class_pred
        k=k+1;
    end
    end
end
accuracy=k/45*100
```

Κώδικας 6. Αναγνώριση Προσώπου

4.4 Αναγνώριση προσώπου με Bag Of Visual Features[8]

Για την εκπαίδευση του συστήματος χρησιμοποιείται η προαναφερθείσα βάση δεδομένων. Τα άτομα (κλάσεις) προς αναγνώριση είναι 45 και παρουσιάζονται στην επόμενη εικόνα. Η διαδικασία αναγνώρισης προσώπου με την μέθοδο Bag of Features υλοποιείται σε 4 βήματα.

1. Δημιουργία συνόλου εκπαίδευσης και δοκιμής (training & test set)
2. Δημιουργία «σάκων» χαρακτηριστικών
3. Εκπαίδευση του ταξινομητή Bag of Visual Features
4. Αναγνώριση προσώπου



Εικόνα 64 Τα πρόσωπα των ατόμων προς αναγνώριση

4.4.1 Δημιουργία συνόλων εκπαίδευσης (training & test set)

Η βάση εικόνων οργανώνεται σε δύο σύνολα, το σύνολο εκπαίδευσης (training set) και το σύνολο δοκιμής (test set). Η οργάνωση εικόνων σε αυτές τις κατηγορίες διευκολύνει πρώτα τον χειρισμό μεγάλων συνόλων εικόνων αλλά χρησιμοποιείται και για τον υπολογισμό της επιτυχίας αναγνώρισης. Το σύνολο εκπαίδευσης αποτελείται συνήθως από το 90% των δεδομένων.

[8] https://www.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html

```
Database = imageDatastore('FaceDatabaseORLnew','IncludeSubfolders',true,…
'LabelSource','foldernames');
labels = countEachLabel(Database);
[trainingSet, validationSet] = splitEachLabel(Database, 0.9, 'randomize');
```

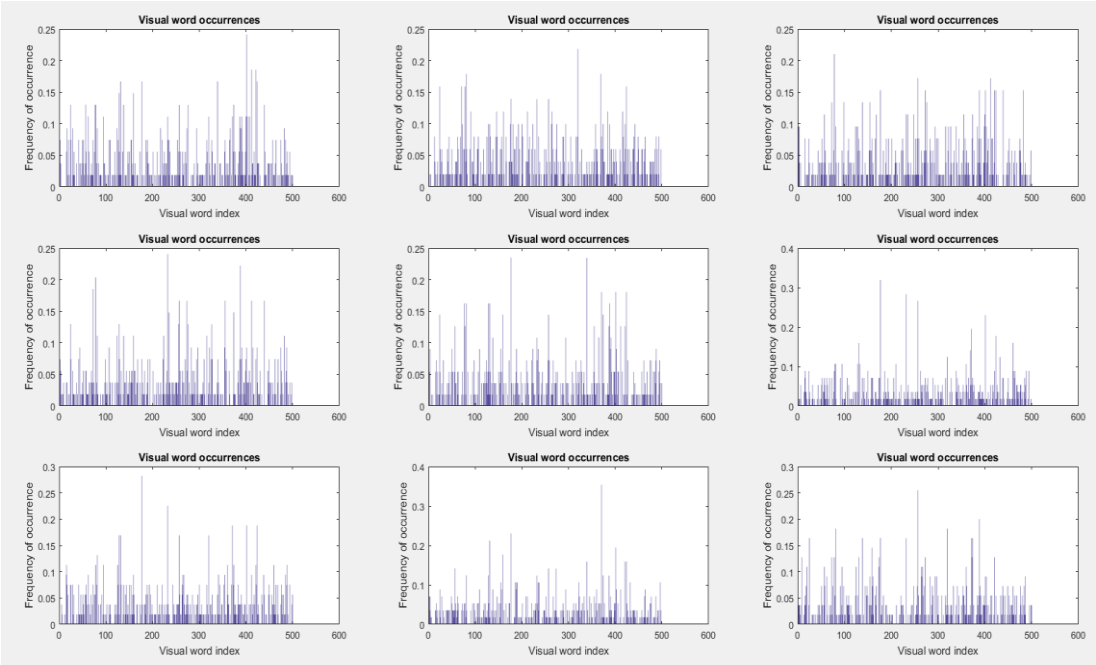
Κώδικας 7. Οργάνωση Δεδομένων σε σύνολο εκπαίδευσης και σύνολο δοκιμής

4.4.2 Δημιουργία οπτικών «σάκων» χαρακτηριστικών[9]

Σε αυτό το στάδιο καθορίζονται τα οπτικά χαρακτηριστικά με την χρήση του αλγορίθμου k-Means. Ο αλγόριθμος αυτός ομαδοποιεί τους περιγραφείς με μια επαναληπτική μέθοδο σε 500 συστάδες. Η διαδικασία αυτή δημιουργεί συστάδες που είναι συμπαγείς και διαχωρίζονται με παρόμοια χαρακτηριστικά. Το κάθε κεντροειδές που προκύπτει αποτελεί και το οπτικό χαρακτηριστικό. Επομένως, αποτέλεσμα αυτής της διαδικασίας είναι η δημιουργία του οπτικού λεξικού. Στην συνέχεια ορίζονται τμήματα (patches) για την εξαγωγή περιγραφέων χαρακτηριστικών. Σημειώνεται ότι με την χρήση αυτής της μεθόδου ενδέχεται να χαθούν σημαντικές πληροφορίες της κλίμακας. Οι περιγραφείς χαρακτηριστικών εντοπίζονται σε τμήματα (patches) διαστάσεων 32x32 με βήμα 8x8. Στο σύνολο του σετ εκπαίδευσης, το οποίο αποτελείται από 405 εικόνες, εξάγονται 272160 χαρακτηριστικά. Από αυτό το σύνολο διατηρείται το 80% των ισχυρότερων χαρακτηριστικών. Με την μέθοδο ομαδοποίησης K-mean δημιουργείται το τελικό οπτικό λεξικό. Ο αριθμός K ορίζεται ως 500 με αποτέλεσμα την δημιουργία 500 συστάδων οπτικών χαρακτηριστικών. Ακολουθούν τα αντίστοιχα ιστογράμματα οπτικών χαρακτηριστικών που παρατηρούνται σε κάθε εικόνα εκπαίδευση για τα 5 νέα άτομα που εισήχθησαν στη βάση ORL.

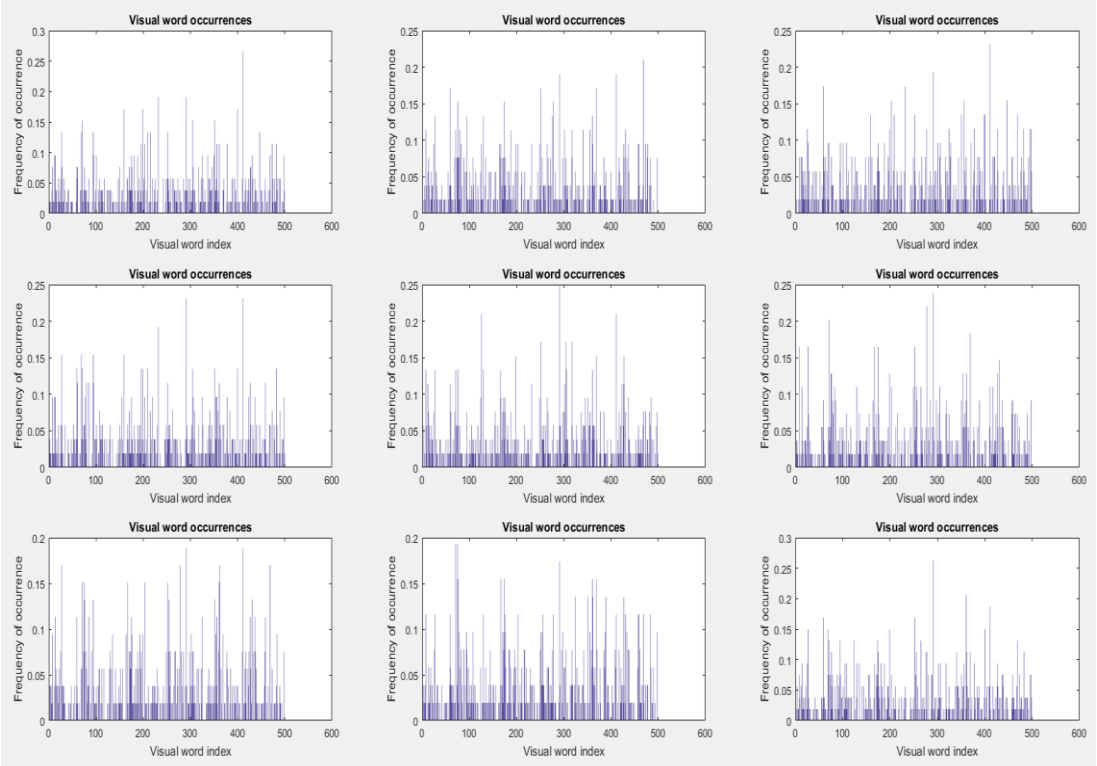
```
bag = bagOfFeatures(trainingSet); %δημιουργία οπτικού λεξικού 500 οπτικών
χαρακτηριστικών
featureVector = encode(bag, img); %Έλεγχος οπτικών χαρακτηριστικών που
εμφανίζονται σε κάθε εικόνα
```

Κώδικας 8. Δημιουργία οπτικού λεξικού

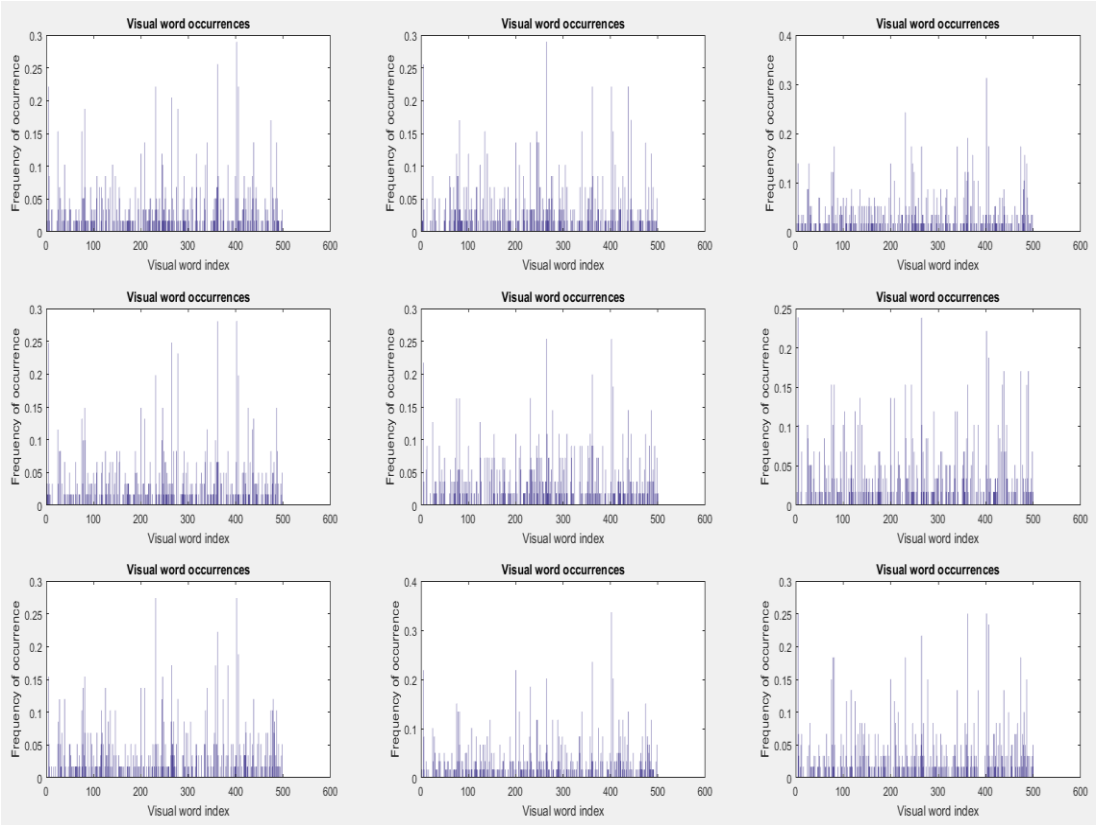


Εικόνα 65. Αναπαράσταση οπτικών Χαρακτηριστικών για την Κλέα

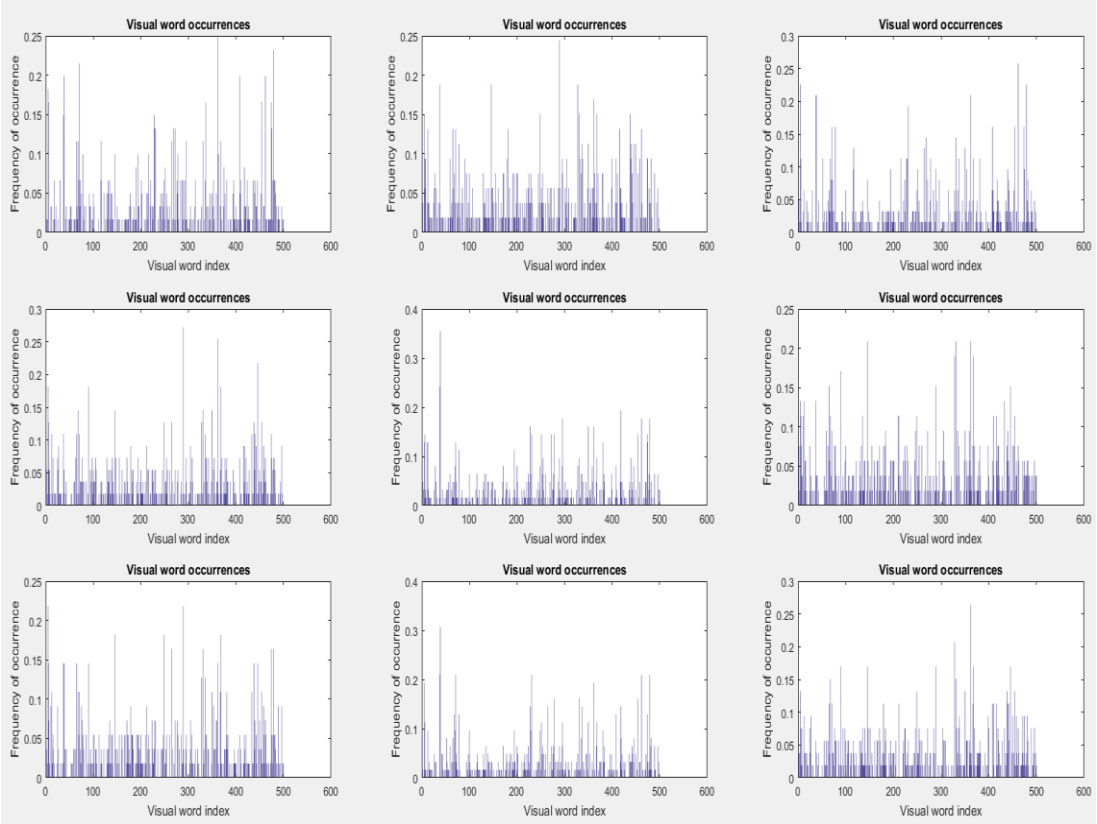
[9] https://www.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html



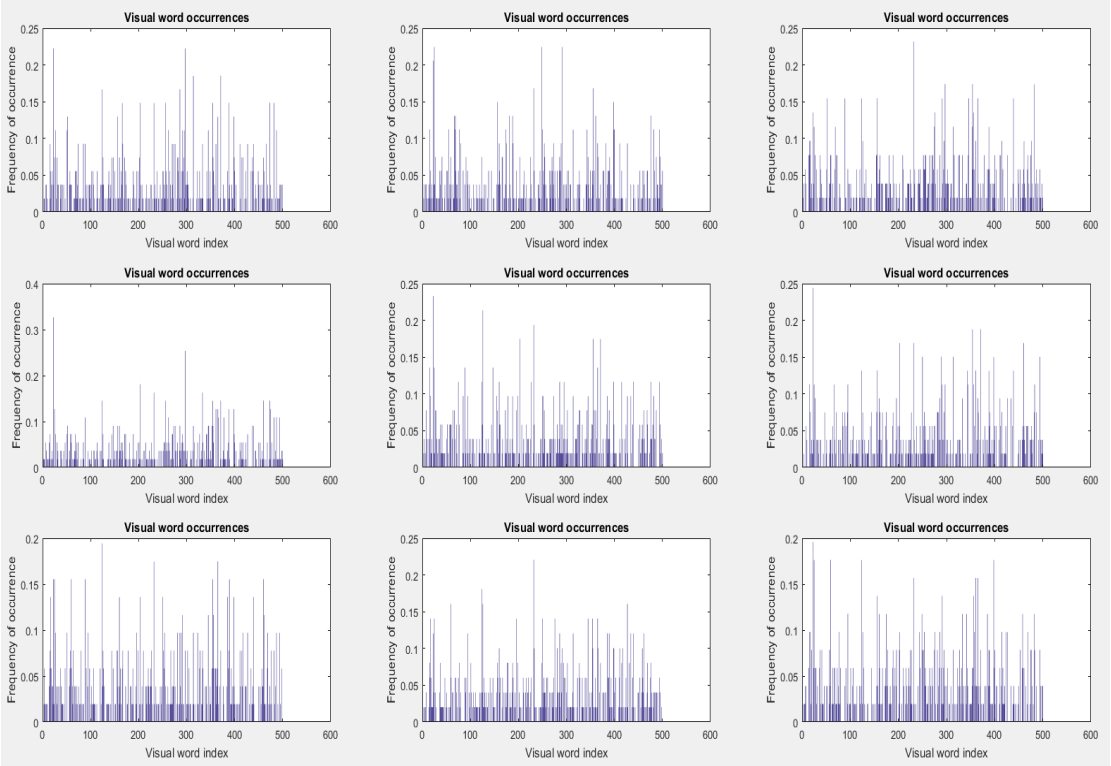
Εικόνα 66. Αναπαράσταση οπτικών Χαρακτηριστικών για τον Παναγιώτη



Εικόνα 67. Αναπαράσταση οπτικών χαρακτηριστικών για τον Redi



Εικόνα 68. Αναπαράσταση οπτικών χαρακτηριστικών για την Αθηνά

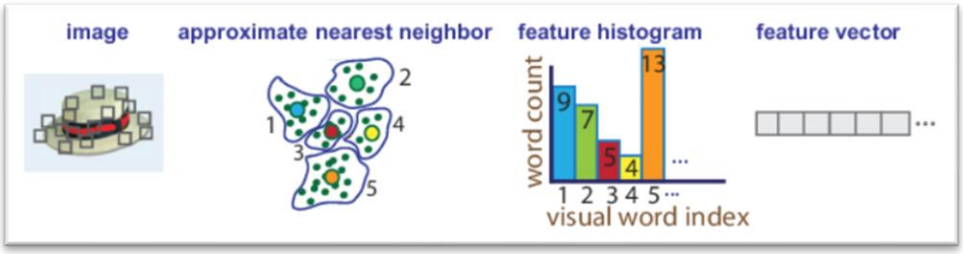


Εικόνα 69. Αναπαράσταση οπτικών χαρακτηριστικών για τον Στέλιο

4.4.3 Εκπαίδευση ταξινομητή Bag of Visual Features[10]

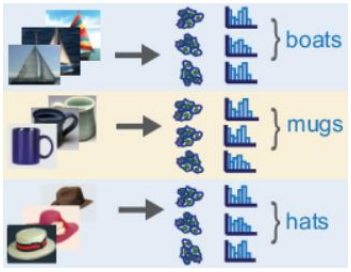
Σε αυτό το στάδιο γίνεται η εκπαίδευση ταξινομητή εικόνων (image classifier). Εκπαιδεύεται λοιπόν ένας ταξινομητής πολλαπλών κλάσεων (multiclass) με την χρήση της ροής εργασίας διόρθωσης σφαλμάτων Error-Correcting Output Codes (ECOC) και ταξινομητών γραμμικών Support Vector Machines. Κατά την εκπαίδευση τα οπτικά χαρακτηριστικά χρησιμοποιούνται για να κωδικοποιήσουν τις εικόνες του σετ εκπαίδευσης σε ιστογράμματα οπτικών χαρακτηριστικών. Έτσι, τα ιστογράμματα αυτά μπορούν να χρησιμοποιηθούν εν τέλει ως θετικά ή αρνητικά παραδείγματα για την διαδικασία της εκπαίδευσης. Η πορεία αυτή μπορεί να κωδικοποιηθεί σε τρία βήματα:

- Πρώτα εξάγονται οι περιγραφείς χαρακτηριστικών και με την χρήση του αλγορίθμου των εγγύτερων γειτόνων δημιουργείται ένα αντίστοιχο ιστόγραμμα. Οι περιγραφείς χαρακτηριστικών συγκρίνονται με τα κεντροειδή των συστάδων ώστε να δημιουργηθεί ένα νέο ιστόγραμμα οπτικών χαρακτηριστικών. Το ιστόγραμμα, τέλος, μετασχηματίζεται σε διάνυσμα οπτικών χαρακτηριστικών.



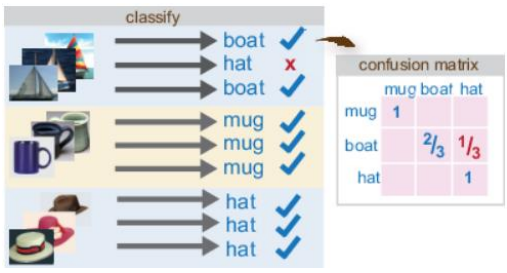
Εικόνα 70. Διαδικασία υπολογισμού οπτικών χαρακτηριστικών για κάθε εικόνα εκπαίδευσης

- Η παραπάνω διαδικασία επαναλαμβάνεται για κάθε εικόνα του σετ εκπαίδευσης



Εικόνα 71. Δημιουργία οπτικών χαρακτηριστικών για κάθε εικόνα.

- Γίνεται η αξιολόγηση της εκπαίδευσης. Για τον σκοπό αυτό δημιουργείται ένας πίνακας σύγχυσης (confusion matrix), ο οποίος παίρνει τιμές από 0 έως 1 που εκφράζουν την ακρίβεια της ταξινόμησης



Εικόνα 72 Αξιολόγηση Εκπαίδευσης

[10] https://www.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html

```
categoryClassifier = trainImageCategoryClassifier(trainingSet, bag);
%εκπαίδευση ταξινομητή
confMatrix = evaluate(categoryClassifier, trainingSet); % Αξιολόγηση
ταξινόμησης του σετ εκπαίδυσης
mean(diag(confMatrix)) % Ακρίβεια ταξινόμησης
confMatrix1 = evaluate(categoryClassifier, validationSet); % Αξιολόγηση
ταξινόμησης του σετ δοκιμής
mean(diag(confMatrix1)) % Ακρίβεια ταξινόμησης
```

Κώδικας 9. Εκπαίδευση ταξινομητή εικόνων (image classifier)

4.4.4 Αναγνώριση προσώπου

Η αναγνώριση προσώπου γίνεται μέσω του ταξινομητή που δημιουργήθηκε κατά την διάρκεια της εκπαίδευσης.

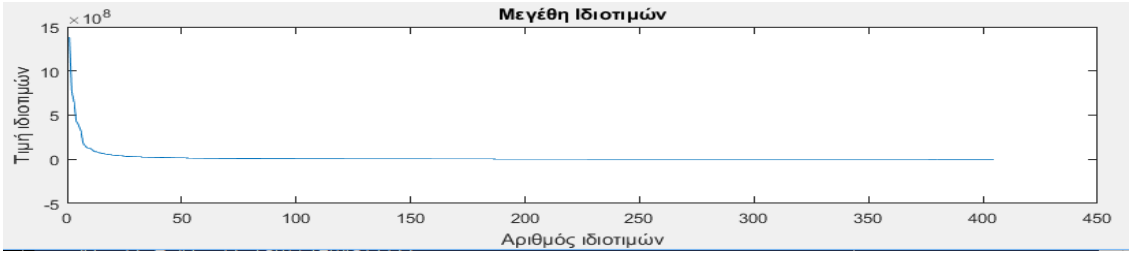
```
[labelIdx, scores] = predict(categoryClassifier, test_img); % Ταξινόμηση
εικόνας στην αντίστοιχη κλάση-κατηγορία
CLASS=categoryClassifier.Labels(labelIdx) % K
```

Κώδικας 10. Αναγνώριση προσώπου

Κεφάλαιο 5. Αποτελέσματα

Προτού παρουσιαστούν και αναλυθούν τα αποτελέσματα της διπλωματικής εργασίας θα πρέπει να δηλωθεί πως η όλη εργασία υλοποιήθηκε με το λογισμικό Matlab και τις εργαλειοθήκες Statistics and Machine Learning Toolbox και Computer Vision Toolbox.

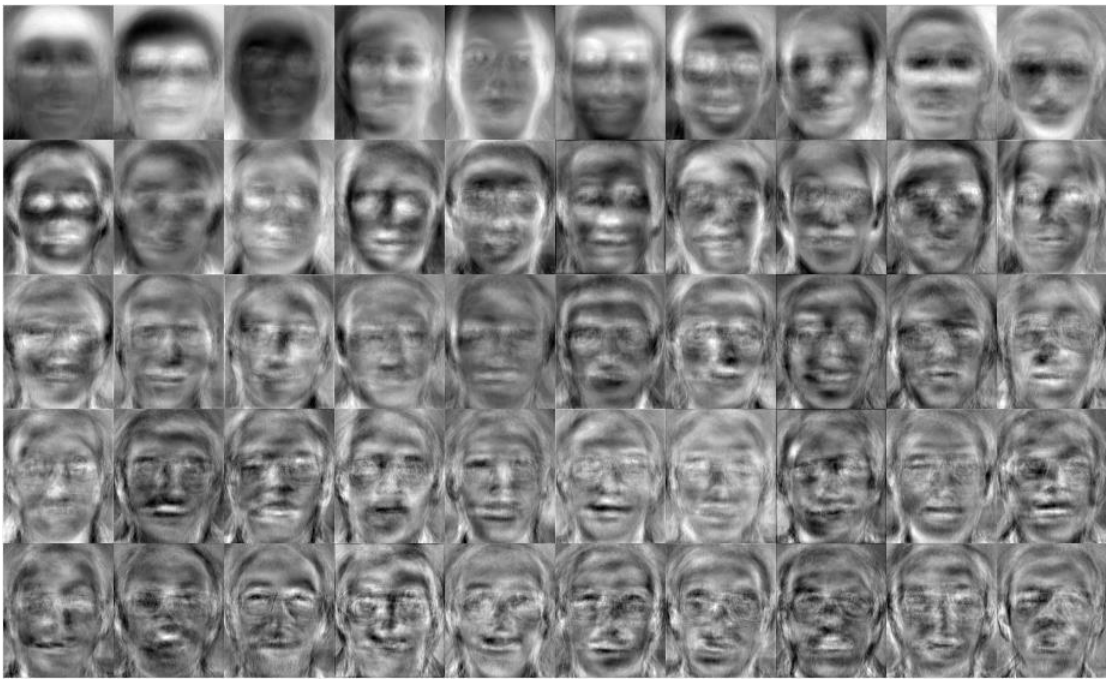
Η πρώτη μέθοδος που χρησιμοποιήθηκε για την αναγνώριση προσώπου είναι ο αλγόριθμος EigenFaces. Αυτός βασίζεται στην εξαγωγή των κυριότερων και ισχυρότερων χαρακτηριστικών που υπάρχουν σε ανθρώπινα πρόσωπα. Η εν λόγω διαδικασία γίνεται με τον υπολογισμό των ισχυρότερων ιδιοτιμών ώστε να επιλεγούν τα αντίστοιχα ιδιοδιανύσματα (eigenfaces). Οπότε ένα από τα κριτήρια που αξίζει ερευνηθούν είναι η επιλογή του κατάλληλου αριθμού των ισχυρότερων ιδιοπροσώπων. Άλλο σημαντικό κριτήριο είναι η επιλογή (σε ποσοστό) των εικόνων εκπαίδευσης και δοκιμής. Η επιλογή αυτού του αριθμού ιδιοπροσώπων και του αριθμού εικόνων εκπαίδευσης ελέγχεται ως προς το ποσοστό ορθότητας (accuracy) του συστήματος. Το ποσοστό αυτό επιτυχίας προκύπτει από την σχέση acc=m/n, όπου m είναι το σύνολο των θετικών αποτελεσμάτων αναγνώρισης και n είναι το σύνολο των προς αναγνώριση εικόνων. Στον παρακάτω πίνακα παρουσιάζεται η επιτυχία του συστήματος σε 40 πειράματα που έγιναν. Αυτά αφορούν συνδυασμούς ιδιοπροσώπων και αριθμών εικόνων εκπαίδευσης για κάθε κλάση. Επομένως όταν στο σύστημα εισάγεται μόνο ένα πρόσωπο για κάθε κλάση, ο συνολικός αριθμός των εικόνων εκπαίδευσης είναι 45, όταν επιλέγονται 5 πρόσωπα ανά κλάση ο συνολικός αριθμός των εικόνων εκπαίδευσης είναι 225 κ.ο.κ. Αντίστοιχα, στις επόμενες εικόνες παρουσιάζονται τα μεγέθη των αντίστοιχων ιδιοτιμών που προκύπτουν από το κάθε σετ εκπαίδευσης. Μια αξιοσημείωτη παρατήρηση που προκύπτει από αυτά τα πειράματα είναι η μείωση του μεγέθους των ιδιοτιμών, η οποία είναι ανάλογη με την μείωση των δεδομένων εκπαίδευσης. Αξίζει να επαναληφθεί πως τα ισχυρότερες ιδιοτιμές στην ουσία εκφράζουν τα ισχυρότερα χαρακτηριστικά των εικόνων.



Εικόνα 73. Μέγεθος ιδιοτιμών – 405 Ιδιοτιμές (9 εικόνες εκπαίδευσης/ατομο)

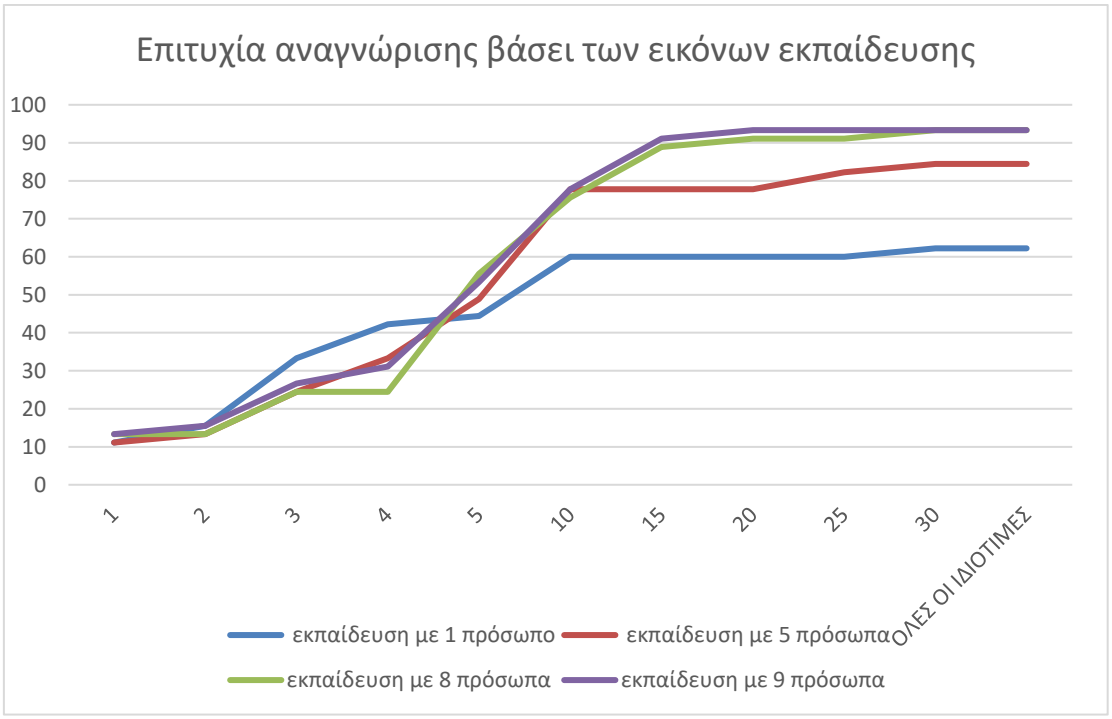
Εικόνα 74. Μέγεθος ιδιοτιμών – 225 Ιδιοτιμές (5 εικόνες εκπαίδευσης/ατομο)

Εικόνα 75. Μέγεθος ιδιοτιμών – 45 Ιδιοτιμές (1 εικόνες εκπαίδευσης/ατομο)



Εικόνα 76. Τα πρώτα 50 ιδιοπρόσωπα

Στην ανωτέρω εικόνα απεικονίζονται τα ισχυρότερα ιδιοπρόσωπα. Παρατηρείται ότι στα ισχυρότερα από αυτά (πρώτη σειρά) τα χαρακτηριστικά ενός προσώπου, όπως τα μάτια, τα μαλλιά, η μύτη και το στόμα, είναι πολύ έντονα και εμφανή. Όσο τα μεγέθη των ιδιοτιμών μειώνονται παρατηρείται πως τα ιδιοπρόσωπα απεικονίζονται περισσότερο με την μορφή θορύβου, χάνοντας έτσι την πληροφορία των χαρακτηριστικών. Παρατηρείται πως το συγκεκριμένο σύστημα μπορεί να επιτύχει την μέγιστη ακρίβεια με την χρήση 50 ιδιοπροσώπων. Παρατηρείται επίσης πως ο αριθμός των εικόνων εκπαίδευσης επηρεάζει σημαντικά την ακρίβεια του συστήματος. Η μέγιστη ακρίβεια του συστήματος επιτυγχάνεται με την εισαγωγή του 80% των εικόνων εκπαίδευσης.



Εικόνα 77. Επιτυχία αναγνώρισης βάσει των εικόνων εκπαίδευσης και των αριθμό των ιδιοπροσώπων

Αριθμός ιδιοπροσώπων	Επιτυχία Αναγνώρισης % εκπαίδευση με 1 πρόσωπο	Επιτυχία Αναγνώρισης % εκπαίδευση με 5 πρόσωπα	Επιτυχία Αναγνώρισης % εκπαίδευση με 8 πρόσωπα	Επιτυχία Αναγνώρισης % εκπαίδευση με 9 πρόσωπα
1	11,11	11,11	13.33	13,33
2	15,55	13,33	13.33	15.55
3	33,33	24,44	24.44	26.66
4	42,22	33,33	24.44	31.11
5	44,44	48,88	55.55	53.33
10	60	77,77	75.55	77.77
15	60	77,77	88.88	91.11
20	60	77,77	91.11	93.33
25	60	82,22	91.11	93.33
50	62,22	84,44	93.33	93.33
ΟΛΕΣ ΟΙ ΙΔΙΟΤΙΜΕΣ	**62,22**	**84,44**	**93.33**	**93.33**

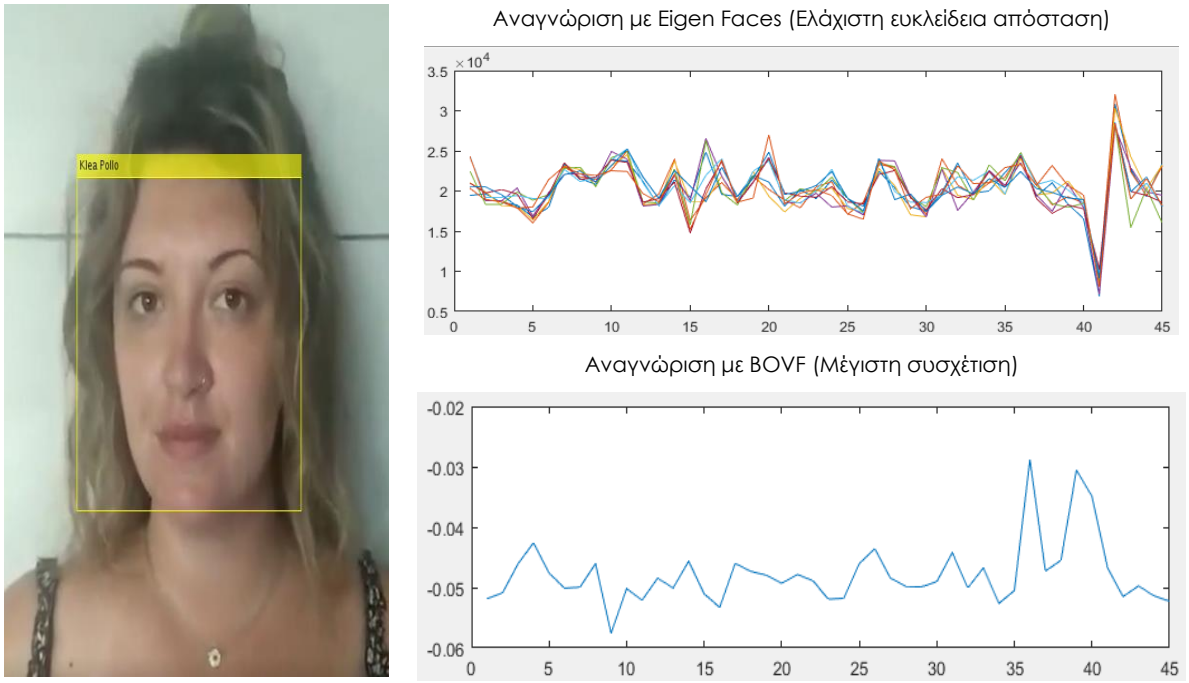
Πίνακας 2. Επιτυχία Αναγνώρισης προσώπων με την μέθοδο EigenFaces

Η δεύτερη μέθοδος αναγνώρισης που υλοποιήθηκε είναι εκείνη των Bag of Visual Features. Η μέθοδος αυτή εντοπίζει χαρακτηριστικά τμήματα μιας εικόνας και δημιουργεί ένα οπτικό λεξικό στο οποίο κωδικοποιούνται και ποσοτικοποιούνται τα χαρακτηριστικά αυτά τμήματα. Επομένως, ένα σημαντικό κριτήριο για την επιτυχία ενός τέτοιου συστήματος αναγνώρισης είναι ο έλεγχος της σημασίας των δεδομένων εκπαίδευσης και του αριθμού των εικόνων που εισάγονται στο σύστημα εκπαίδευσης. Ο Πίνακας 3 παρουσιάζει την επιτυχία αναγνώρισης του συστήματος με σετ εκπαίδευσης που αποτελείται από το 10-90% των δεδομένων. Παρατηρείται πως η απόδοση του συστήματος είναι αρκετά καλή ακόμη και με την χρήση του 20% των δεδομένων εκπαίδευσης. Αυτό, μπορεί να ερμηνευτεί από το είδος των δεδομένων καθώς πρόκειται για εικόνες που παρουσιάζουν χαμηλή μεταβλητότητα εφόσον πρόκειται για μετωπικές εικόνες προσώπων.

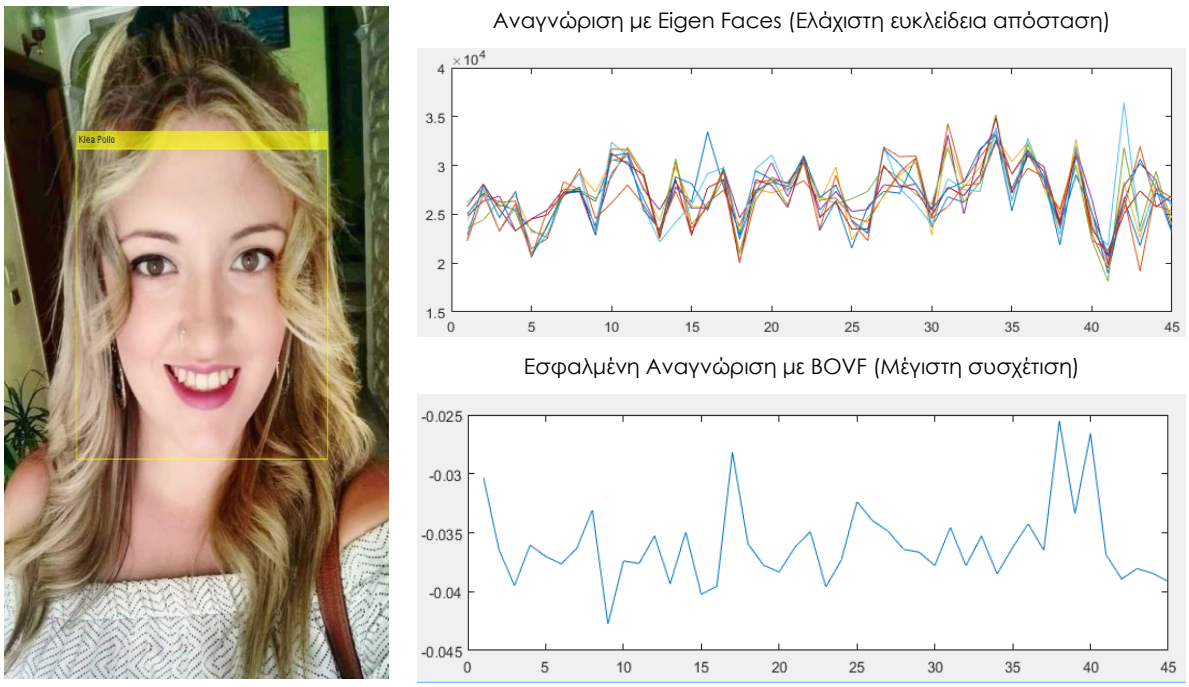
Ποσοστό Δεδομένων εκπαίδευσης	Ποσοστό επιτυχίας
10%	82%
20%	90%
30%	93%
40%	93%
50%	98%
60%	98%
70%	97%
80%	99%
90%	100%

Πίνακας 3. Επιτυχία Αναγνώρισης προσώπων με την μέθοδο Bag Of Visual Features

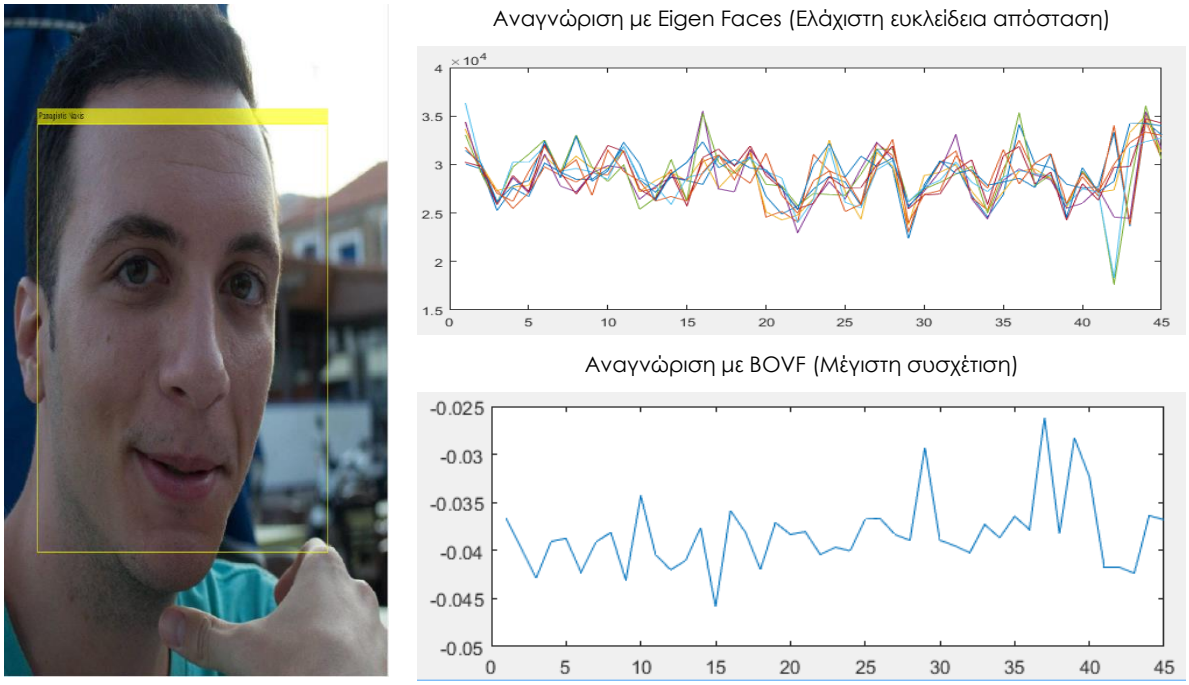
Δεδομένου του υψηλού ποσοστού επιτυχίας που παρουσιάζει το σύστημα αναγνώρισης σε ένα σετ δεδομένων έχει σημασία η επιτυχία αυτή να ελεγχθεί και από άλλες εικόνες των ιδίων ατόμων. Συνεπώς επιλέγονται εικόνες των 5 ατόμων που εμπλούτισαν την βάση ORL. Οι νέες εικόνες ελέγχου ανήκουν στον χρωματικό χώρο RGB, η λήψη τους έγινε υπό διαφορετικές συνθήκες φωτισμού, επιλέχθηκαν τυχαία και οι διαστάσεις τους παρουσιάζουν μεγάλη ποικιλία. Επομένως, για την αναγνώριση του προσώπου στις εικόνες γίνεται αρχικά η ανίχνευση και ο εντοπισμός του προσώπου, η κανονικοποίηση των εικόνων, η μετατροπή τους στον χώρο της κλίμακας grayscale, ενώ και οι διαστάσεις του προσώπου προσαρμόζονται στις διαστάσεις των εικόνων της βάσης (112x92). Παρακάτω ακολουθούν μερικά παραδείγματα.



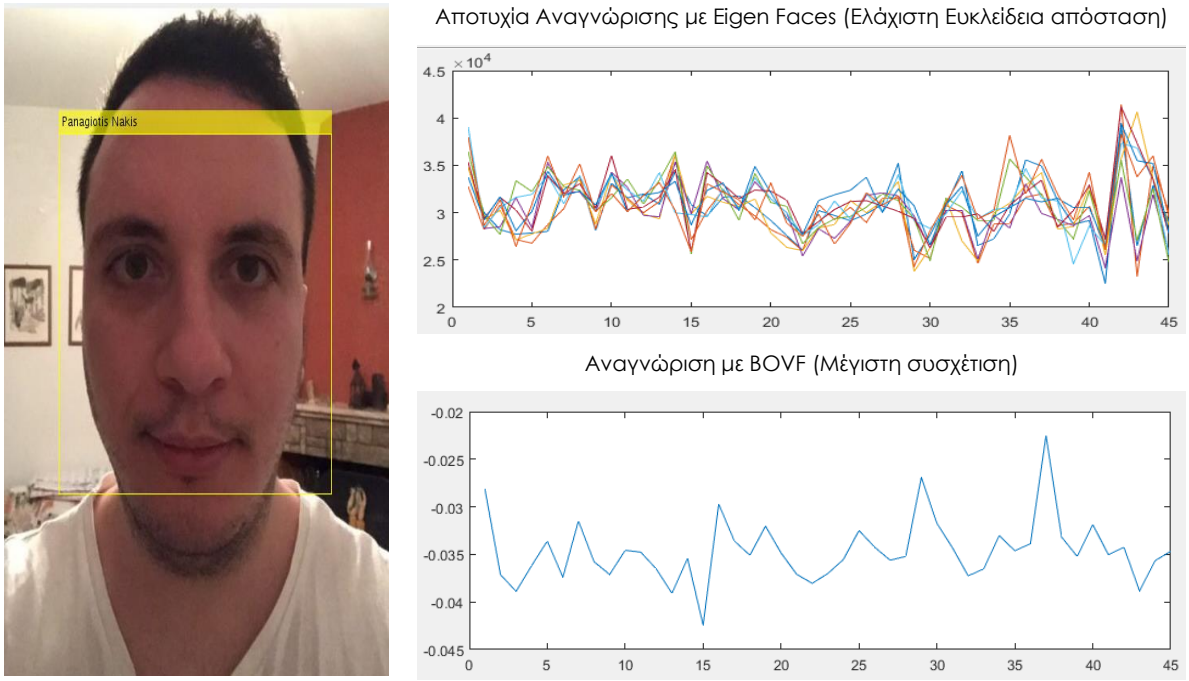
Εικόνα 78. Αναγνώριση προσώπου (Κλέα). Επιτυχής αναγνώριση για εικόνα προσώπου εκτός βάσης



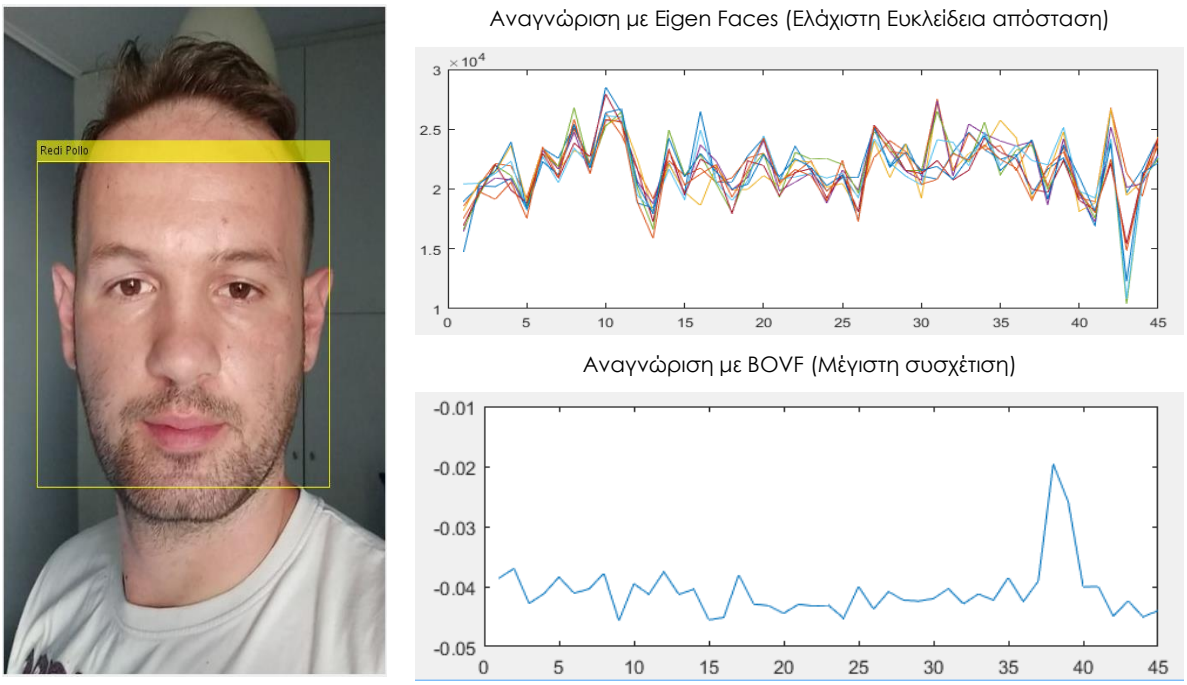
Εικόνα 79. Αναγνώριση προσώπου (Κλέα). Επιτυχής αναγνώριση για εικόνα προσώπου εκτός βάσης με την μέθοδο eigen faces και αποτυχία αναγνώρισης με την μέθοδο Bag of Visiual features



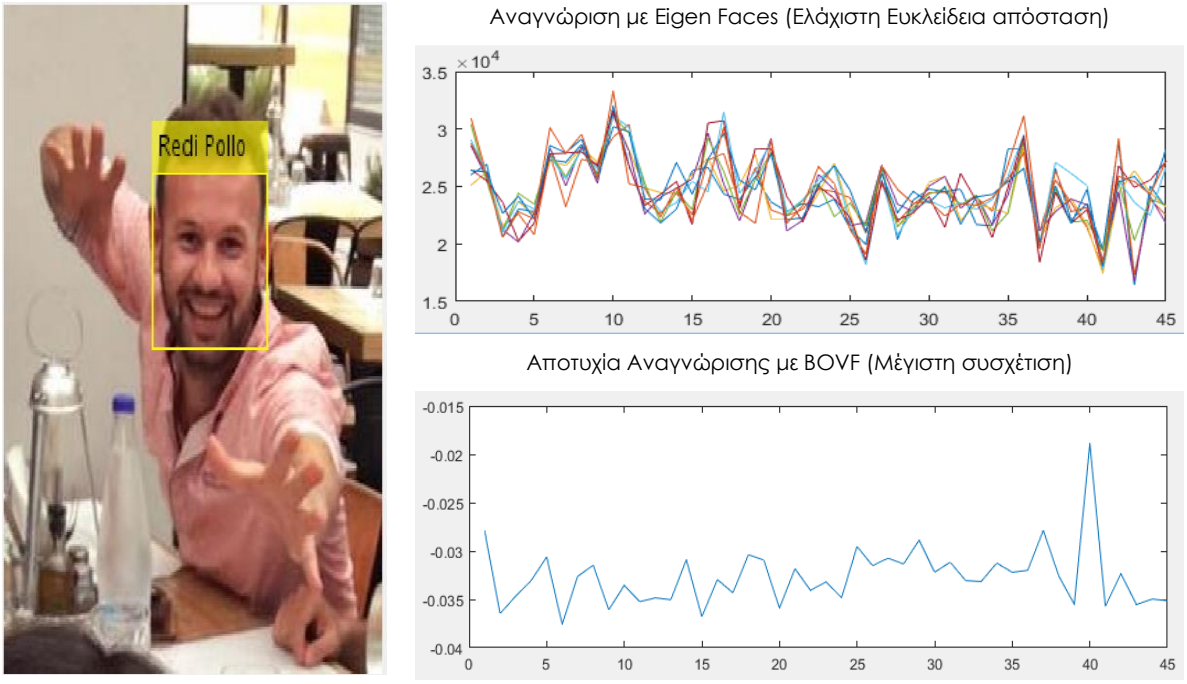
Εικόνα 80. Αναγνώριση προσώπου (Παναγιώτης). Επιτυχής αναγνώριση για εικόνα προσώπου εκτός βάσης



Εικόνα 81 Αναγνώριση προσώπου (Παναγιώτης). Αποτυχία αναγνώρισης για εικόνα προσώπου εκτός βάσης με την μέθοδο eigen faces και επιτυχία αναγνώρισης με την μέθοδο Bag of Visiual Features.

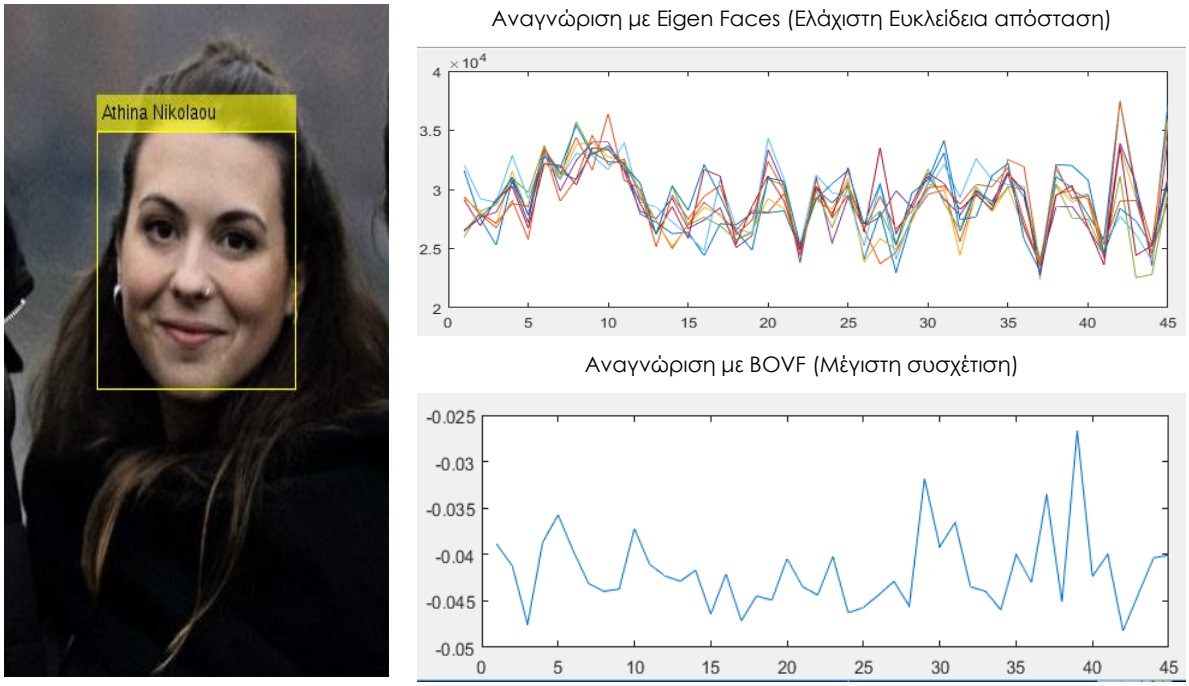


Εικόνα 82. Αναγνώριση προσώπου (Ρέντι). Επιτυχής αναγνώριση για εικόνα προσώπου εκτός βάσης



Εικόνα 83. Αναγνώριση προσώπου (Ρέντι). Επιτυχής αναγνώριση για εικόνα προσώπου εκτός βάσης με την μέθοδο eigen faces και αποτυχία αναγνώρισης με την μέθοδο Bag of Visiual features

Εικόνα 84. Αναγνώριση προσώπου (Αθηνά). Επιτυχής αναγνώριση για εικόνα προσώπου εκτός βάσης



Εικόνα 85 Αναγνώριση προσώπου (Αθηνά). Επιτυχής αναγνώριση για εικόνα προσώπου εκτός βάσης

Κεφάλαιο 6. Συμπεράσματα

Στην ενότητα αυτή παρουσιάζονται οι βασικότερες παρατηρήσεις για τις μεθοδολογίες που ακολουθήθηκαν. Το σύστημα αναγνώρισης που υλοποιήθηκε αποσκοπεί στην αναγνώριση προσώπων σε εικόνες με ένα πρόσωπο. Η διαδικασία αναγνώρισης υλοποιείται σε 5 βασικά βήματα:

1. Ανίχνευση και εντοπισμός προσώπου σε μια εικόνα
2. Επεξεργασία εικόνας προσώπου
3. Εξαγωγή χαρακτηριστικών
4. Ταξινόμηση προσώπου στην αντίστοιχη κλάση
5. Αξιολόγηση αναγνώρισης

Η ανίχνευση προσώπου έγινε με την χρήση του ταξινομητή Cascade, ο οποίος βασίζεται στον αλγόριθμο των Viola & Jones και αποδεικνύεται αρκετά αξιόπιστος για την ανίχνευση προσώπων. Μια σειρά προκλήσεων ως προς την ανίχνευση – όπως η κλίμακα της εικόνας, οι αποκρύψεις, η στάση και οι στροφές του προσώπου – φαίνεται να επηρεάζουν σημαντικά την επιτυχία της. Παρουσιάζεται δηλαδή αδυναμία ανίχνευσης σε εικόνες όπου το πρόσωπο βρίσκεται πολύ κοντά στην κάμερα, σε πρόσωπα που έχουν στροφές μεγαλύτερες από 20^{o} είτε σε δημιουργεί προβλήματα η στάση του προσώπου. Οι αδυναμία αυτή δικαιολογείται από την βάση δεδομένων που χρησιμοποιήθηκε για την εκπαίδευση του ταξινομητή, καθώς οι εικόνες εκπαίδευσης που χρησιμοποιήθηκαν αναφέρονται σε μετωπικές λήψεις προσώπων. Όσον αφορά τον θόρυβο και την ανάλυση (resolution) των υπό εξέταση εικόνων, ο ταξινομητής δεν φαίνεται να αντιμετωπίζει ιδιαίτερο πρόβλημα. Αξίζει να σημειωθεί ότι ο αλγόριθμος αυτός βασίζεται στην εξαγωγή χαρακτηριστικών τύπου Haar, τα οποία παρουσιάζουν την διαφορά της φωτεινότητας στις χαρακτηριστικές περιοχές του προσώπου. Λόγω αυτής της αρχιτεκτονικής παρατηρήθηκε πως ο αλγόριθμος είναι ευαίσθητος στη αναγνώριση ψευδώς θετικών προσώπων σε εικόνες που παρουσιάζουν αντίστοιχες διαφορές. Η ευαισθησία του αλγορίθμου ρυθμίζεται μέσω ενός κατωφλίου (0-10). Για την επιλογή του κατωφλίου δεν υπάρχει συγκεκριμένος κανόνας καθώς βασίζεται κυρίως στα δεδομένα που εισάγονται.

Εφόσον γίνει η ανίχνευση προσώπου ξεκινάει η διαδικασία της αναγνώρισης. Η υλοποίηση της μεθόδου αναγνώρισης με eigenfaces αρχίζει με την αφαίρεση της μέσης εικόνας προσώπων του συνόλου εκπαίδευσης από την νέα εικόνα εισόδου. Η αφαίρεση αυτή στην ουσία «ενδυναμώνει» την εικόνα εισόδου καθώς τα χαρακτηριστικά όπως ακμές και γωνίες γίνονται πιο έντονα. Στην συνέχεια η εικόνα προβάλλεται στον ιδιοχώρο με σκοπό τον υπολογισμό αντίστοιχων βαρών των ιδιοπροσώπων. Αυτή η τεχνική αποδεικνύεται πολύ αποτελεσματική για την μείωση της διαστασιμότητας. Η μείωση των διαστάσεων της εικόνας από M [m x n] σε k [1x k] όπου k<<M δεν επηρεάζει την απόδοση του συστήματος σε ποσοστά επιτυχίας και συνάμα προσφέρει μεγάλη ταχύτητα. Σε πείραμα που υλοποιήθηκε για τον υπολογισμό των Ευκλείδειων αποστάσεων εικόνων στον χώρο των ιδιοπροσώπων και στον χώρο των βαρών, η τεχνική αυτή φάνηκε να αποδίδει κατά 10 φορές ταχύτερα. Η μέθοδος eigenfaces παρουσιάζει όμως και ορισμένα μειονεκτήματα που αφορούν κυρίως την επιλογή των εικόνων εκπαίδευσης. Για την ορθή λειτουργία της μεθόδου τα πρόσωπα χρειάζεται να είναι κεντραρισμένα (centering of data). Αυτό προϋποθέτει την επιλογή δεδομένων σε παρόμοιες στροφές και πόζες. Το γεγονός αυτό επιφέρει αδυναμία στο σύστημα για αναγνώριση προσώπων σε διαφορετικές πόζες και στροφές από τα πρόσωπα που λαμβάνουν μέρος στην εκπαίδευση. Παρά ταύτα, με δεδομένο πως οι περισσότερες εικόνες που λαμβάνονται τις τελευταίες δεκαετίες έχουν συγκεκριμένες πόζες και στροφές (selfies), η μέθοδος αυτή θεωρείται αρκετά απλή και αξιόπιστη για την επίλυση ταξινόμησης/αναγνώρισης προσώπων.

Η δεύτερη μέθοδος αναγνώρισης είναι η Bag of Visual Words. Η μέθοδος αυτή επιτυγχάνει υψηλά ποσοστά επιτυχίας στις εικόνες δοκιμής (test data), και η επιτυχία αυτή επιβεβαιώνεται εν μέρει από άλλες εικόνες προσώπων που δεν συμμετείχαν στην εκπαίδευση

και την αξιολόγηση του συστήματος. Η μέθοδος αυτή στην ουσία κατατέμνει τις εικόνες σε τμήματα (patches) όπου υπολογίζονται οι περιγραφείς χαρακτηριστικών. Οι περιγραφείς αυτοί ομαδοποιούνται δημιουργώντας έτσι τα οπτικά χαρακτηριστικά. Λόγω της ομαδοποίησής τους μέσω του αλγορίθμου (k-means) αυτά τα χαρακτηριστικά χάνουν την χωρική τους πληροφορία, γεγονός που προσδίδει σε αυτή την μέθοδο την ιδιότητα να εξετάζεται κατά βάση η ύπαρξη και η συχνότητα αυτών των χαρακτηριστικών στον χώρο μιας εξεταζόμενης εικόνας. Για την κατανόηση της λογικής αυτής της μεθόδου, θα μπορούσαμε να παρομοιάσουμε τα οπτικά χαρακτηριστικάμε παζλ ίσων διαστάσεων. Έτσι σε μια νέα εικόνα σκοπός είναι να υπολογιστούν ποια και πόσες φορές οι ψηφίδες παζλ μπορούν να αντιστοιχηθούν στην νέα εικόνα. Συνεπώς αυτή η μέθοδος δεν προϋποθέτει την ανίχνευση του προσώπου, με την έννοια της θέσης του στην εικόνα, αρκεί το πρόσωπο αυτό να είναι στην ίδια περίπου κλίμακα με τις εικόνες εκπαίδευσης. Επομένως μπορούμε να συμπεράνουμε πως ένα από τα σημαντικότερα μειονεκτήματα αυτής της μεθόδου είναι η απουσία της χωρικής πληροφορίας των χαρακτηριστικών, η οποία είναι πολύ σημαντική για την αναπαράσταση της εικόνας. Επίσης, η διαφορά κλίμακας εικόνων εκπαίδευσης και προς αναγνώριση εικόνων μπορεί να δημιουργήσει αδυναμία ανίχνευσης καθώς η ευαισθησία της κλίμακας βασίζεται μόνο στις δυνατότητες των περιγραφέων χαρακτηριστικών (πχ. SURF).

Τέλος, η αναγνώριση προσώπου, πέραν όλων των γενικών προκλήσεων που συναντώνται στην ανίχνευση/αναγνώριση αντικειμένων, αποδεικνύεται ένα πολύ δυσκολότερο πρόβλημα λόγω της αλλαγής των ανθρώπινων προσώπων με την πάροδο του χρόνου. Ο χρόνος αποδεικνύεται ένας πολύ σημαντικός παράγοντας που δυσκολεύει την δημιουργία καθολικών βάσεων δεδομένων προσώπων, καθώς ένα πρόσωπο μπορεί να παρουσιάζει μεγάλες διακυμάνσεις λόγω της ηλικίας (βρέφος-ανήλικος-ενήλικος-ηλικιωμένος) όπως και της αλλαγής βασικών χαρακτηριστικών, όπως τα μαλλιά (μήκος, χρώμα, είδος κουρέματος), η γενειάδα, η χρήση αξεσουάρ και make up κ.λπ.

Βιβλιογραφία

[1] M. Kelly, *Visual Identification of People by Computer*, 1970. Stanford AI Project, Stanford, CA, Technical Report.

[2] Ekenel H.K., Stiefelhagen R., 2009. Why is facial occlusion a challenging problem? *International Conference on Biometrics*, Springer, pp. 299-308.

[3] Mitchell T.M., 1997. *Machine Learning*, McGraw-Hill, Inc.

[4] Goodfellow I., Bengio Y., Courville A., 2016. *Deep Learning*, MIT Press.

[5] Fisher R.A., 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* vol. 7, pp. 179–188.

[6] Krizhevsky A., Sutskever I., Hinton G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*.

[7] Ioffe S., Szegedy C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. *Proc. 32nd International Conference on Machine Learning, PMLR* 37, pp. 448-456.

[8] Hinton G., Deng L., Yu D., Dahl G.E., Mohamed A.-R., Jaitly N., Senior A., Vanhoucke V., Nguyen P., Sainath T. N. et al., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine,* 29, pp. 82–97.

[9] Rumelhart D.E., Hinton G.E., Williams R.J., 1986. Learning representations by back-propagating errors. *Nature,* 323, pp. 533–536.

[10] Bengio Y., Simard P., Frasconi P., 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks,* 5, pp. 157–166.

[11] Pascanu R., Mikolov T., Bengio Y., 2013. On the difficulty of training recurrent neural networks. *International Conference on Machine Learning*.

[12] Lecun Y., Bengio Y., 1995. Convolutional networks for images, speech, and time-series. In: Arbib M.A. (Ed.), *The Handbook of Brain Theory and Neural Networks*, MIT Press.

[13] Λουρίδας Χ., 2020. *Ενσωμάτωση Drone σε Φυσικό Περιβάλλον και Υποβοήθηση στην Ανίχνευση και Προστασία Δασών από Πυρκαγιές*. Διπλωματική Εργασία, Πανεπιστήμιο Πατρών.

[14] Silverman B.W., Jones M.C., 1989. An important contribution to nonparametric discriminant analysis and density estimation: commentary on Fix and Hodges (1951). International Statistical Review, 57(3), pp. 233-238.

[15] MacQueen J., 1967. Some methods for classification and analysis of multivariate observations. *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281-297

[16] Boser B.E., Guyon I.M., Vapnik V.N. 1992. A training algorithm for optimal margin classifiers. Proc. 5th Annual Workshop on Computational Learning Theory, pp. 144–152.

[17] Cortes C., Vapnik V., 1995. Support-vector networks. *Machine Learning*, 20, pp. 273–297.

[18] Yang G., Huang T., 1994. Human face detection in complex background. *Pattern Recognition,* 27, pp. 53-63.

[19] Kadan A.B., 2014. «A survey on face detection methods,». *International Conference on Advanced Trends in Engineering and Technology*.

[20] Sirohey S.A., 1993. *Human Face Segmentation and Identification*. Technical Report, CS-TR-3176, Computer Vision Laboratory, University of Maryland.

[21] Canny J., 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), pp. 679-698,

[22] Graf H., Chen T., Petajan E., Cosatto E., 1995. Locating Faces and Facial Parts. *Proc. 1st International Workshop Automatic Face and Gesture Recognition*, pp. 41–46.

[23] Yow K.C., Cipolla R., 1997. Feature based human face detection. *Image and Vision Computing*, 15(9), pp. 713-735

[24] Ζάρδα Ι., 2014. *Δίκτυα Bayes: Ένα Εργαλείο Απόφασης για την Αξιολόγηση της Πιστοληπτικής Ικανότητας*. Μεταπτυχιακή Διπλωματική Εργασία, ΕΜΠ

[25] Rusell S., Norvig P., 1994. *Artificial Intelligence: A Modern Approach*. Prentice Hall.

[26] Hjelmås E., Low B.K, 2001. Face detection: a survey. *Computer Vision and Image Understanding*, 83(3), pp. 236-274.

[27] McGunnigle G., Chantier M., 1999. Rotation invariant classification of rough surfaces. *IET Proc. Vision, Image and Signal Processing*, 146(6), pp. 345-352.

[28] Dai Y., Nakano Y., 1996. Face-texture model based on SGLD and its application in face detection in a color scene. *Pattern Recognition*, 29(6), pp. 1007-1017.

[29] Haralick R., Shanmugam K., Dinstein I., 1973. Texture features for image classification. *Studies in Media and Communication*, 3(6), pp. 610-621.

[30] Chaves-González J.M., Vega-Rodríguez M.A., J Gómez-Pulido A., Sánchez-Pérez J.M., 2010. Detecting skin in face recognition systems: A colour spaces study. *Digital Signal Processing*, 20(3), pp. 806-823.

[31] Yang J., Waibel A., 1996. A real-time face tracker. *Proc. 3rd IEEE Workshop on Applications of Computer Vision*, pp. 142-147.

[32] Kumar A., Malhotra S., 2015. Real-time human skin color detection algorithm using skin color map. *2nd International Conference on Computing for Sustainable Global Development*, pp. 2002-2006.

[33] Sakai T., Nagao M., Fujibayashi S., 1969. Line extraction and pattern detection in a photograph. *Pattern recognition*, 1(3), pp. 233–248.

[34] Sinha P., 1994. Object recognition via image invariants: a case study. *Investigative Ophthalmology and Visual Science,* 35, pp. 1735--1740.

[35] Sinha P., 1995. Processing and Recognizing 3D Forms. Ph.D. Thesis, MIT.

[36] Scassellati B., 1998. Eye finding via face detection for a foevated, active vision system. Proc. 15th National/10th Conference on Artificial intelligence/Innovative Applications of Artificial Intelligence, pp. 969–976.

[37] Craw I., Ellis H., Lishman J.R., 1987. Automatic extraction of face-features. *Pattern Recognition Letters, 5*, pp. 183–187.

[38] Craw I., Tock D., Bennett A., 1992. Finding face features. Proc. *European Conference on Computer Vision*, pp. 92-96.

[39] Sobottka K., Pitas I., 1996. Face localization and facial feature extraction based on shape and color information. *Proc. 3rd IEEE International Conference on Image Processing*, pp. 483-486.

[40] Yang M.-H., Ahuja N., 1998. Detecting human faces in color images. *Proc. International Conference on Image Processing*, vol. 1, pp.127-130.

[41] Ahuja N., 1996. A transform for multiscale image segmentation by integrated edge and region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18, pp. 1211–1235.

[42] Papageorgiou C.P., Oren M., Poggio T., 1998. A general framework for object detection. IEEE 6th International Conference on Computer Vision, pp. 555-562.

[43] Lowe D.G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision,* 60(2), pp. 91–110.

[44] Agui T., Kokubo Y., Nagashashi H., Nagao T., 1992. Extraction of face recognition from monochromatic photographs using neural networks. *Proc. 2nd International Conference on Automation, Robotics and Computer Vision*.

[45] Rowley H.A., Baluja S., Kanade T., 1996. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20(1), pp. 23-38.

[46] Sung K.-K., Poggio T., 1998. Example-based learning for view-based human face detection. *IEEE Transactions on pattern analysis and machine intelligence,* 20, pp. 39–51.

[47] Osuna E., Freund R., Girosit F., 1997. Training support vector machines: an application to face detection. *Proc. IEEE Conference on Computer Vision and PatternRecognition*, pp. 130-136.

[48] Schneiderman H., Kanade T., 1998. Probabilistic modeling of local appearance and spatial relationships for object recognition. *Proc IEEE Conference on Computer Vision and Pattern Recognition*, pp. 45-51.

[49] Rabiner L., 1993. *Fundamentals of Speech Recognition*. Pearson Education (US), Prentice Hall.

[50] Samaria F., 1994. *Face Recognition Using Hidden Markov Models*. Ph.D. Thesis, University of Cambridge.

[51] Shlens J., 2014. *A Tutorial on Principal Component Analysis*. 2014arXiv1404.1100S

[52] Kohonen T., 1989. *Self-Organization and Associative Memory*. Springer.

[53] Kirby M., Sirovich L., 1990. Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 12, pp. 103–108.

[54] Turk M., Pentland A., 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience,* 3, pp. 71–86.

[55] Joachims. T., 1998. *Text Categorization with Suport Vector Machines: Learning with Many Rrelevant Features*. Technical Report 23, Universität Dortmund.

[56] Julesz B., 1981. Textons, the elements of texture perception, and their interactions. *Nature*, 290, pp. 91–97.

[57] Cula O.G., Dana K.J., Murphy F.P., Rao B.K. 2005. Skin texture modeling. *International Journal of Computer Vision,* 62, pp. 97–119.

[58] Varma M., Zisserman A., 2005. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62, pp. 61–81.

[59] Lazebnik S., Schmid C., Ponce J., 2005. A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), pp. 1265-1278.

[60] Dalal N., Triggs B., 2005. Histograms of oriented gradients for human detection. Proc. *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* vol. 1, pp. 886-893.

[61] Bay H., Ess A., Tuytelaars T., van Gool L.., 2008. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), pp. 346-359.

[62] Τσουρούνης Δ., 2017. *Βαθιά Αραιή Κωδικοποίηση*. Μεταπτυχιακή Διπλωματική Εργασία, Πανεπιστήμιο Πατρών.

[63] Wagstaff K., Cardie C.T, Rogers S., Schrödl S., 2001. Constrained k-means clustering with background knowledge. *Proc.18th International Conference on Machine Learning*, pp. 577–584.