



**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

Τμήμα Μηχανικών Βιομηχανικής
Σχεδίασης και Παραγωγής

&

**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΙΓΑΙΟΥ**

Τμήμα Ναυτιλίας και
Επιχειρηματικών Υπηρεσιών



**ΔΙΔΡΥΜΑΤΙΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΝΕΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΣΤΗ ΝΑΥΤΙΛΙΑ ΚΑΙ ΤΙΣ ΜΕΤΑΦΟΡΕΣ»**

«Καταγραφή πραγματικού χρόνου με κάμερα, ατόμων πληρώματος και αναγνώριση αυτών, στο μηχανοστάσιο ενός πλοίου, με χρήση αλγορίθμων μηχανικής όρασης»

“Integrated dead man alarm for vessels’ engine room, with real time crew’s presence supervision and monitoring, using mechanical vision algorithms”

Όνοματεπώνυμο Σπουδαστή:

Ηλίας Γεωργίου

Όνοματεπώνυμο Υπεύθυνου Καθηγητή:

Γρηγόρης Νικολάου

ΔΙΑΤΡΙΒΗ

Φεβρουάριος 2021



**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

Τμήμα Μηχανικών Βιομηχανικής
Σχεδίασης και Παραγωγής

&

**ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΙΓΑΙΟΥ**

Τμήμα Ναυτιλίας και
Επιχειρηματικών Υπηρεσιών



Μέλη Εξεταστικής Επιτροπής

Νικολάου Γρηγόριος

Παπουτσιδάκης Μιχαήλ

Δρόσος Χρήστος



«Καταγραφή πραγματικού χρόνου με κάμερα, ατόμων πληρώματος και αναγνώριση αυτών, στο μηχανοστάσιο ενός πλοίου, με χρήση αλγορίθμων μηχανικής όρασης»

Ηλίας Γεωργίου

Μεταπτυχιακή Διατριβή που υποβάλλεται στο καθηγητικό σώμα για την μερική εκπλήρωση των υποχρεώσεων απόκτησης του μεταπτυχιακού τίτλου του Διϋδρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Νέες Τεχνολογίες στη Ναυτιλία και τις Μεταφορές» του Τμήματος Ναυτιλίας και Επιχειρηματικών Υπηρεσιών του Πανεπιστημίου Αιγαίου και του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής του Πανεπιστημίου Δυτικής Αττικής.

Copyright© Ηλίας Γεωργίου, 2021. Με επιφύλαξη παντός δικαιώματος. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.



ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Γεωργίου Ηλίας του Δημητρίου, με αριθμό μητρώου 20 φοιτητής του Διϊδρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Νέες Τεχνολογίες στη Ναυτιλία και τις Μεταφορές» του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής της Σχολής Μηχανικών Πανεπιστημίου Δυτικής Αττικής, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου».

Ο Δηλών

Ηλίας Γεωργίου

Περιεχόμενα

Κατάλογος Εικόνων.....	- 6 -
Κατάλογος Συντομογραφιών.....	- 7 -
Περίληψη.....	- 8 -
Abstract.....	- 9 -
1. Λίγα λόγια για τη Μηχανική όραση.....	- 10 -
1.1 Πως λειτουργεί η μηχανική όραση.....	- 12 -
1.2 Η εξέλιξη της Μηχανικής Μάθησης.....	- 15 -
2. Ανίχνευση και Αναγνώριση προσώπου.....	- 19 -
2.1 Ανίχνευση προσώπου.....	- 19 -
2.1.1 Ομαδοποίηση και PCA.....	- 21 -
2.1.2 Νευρωνικά Δίκτυα.....	- 21 -
2.1.3 Μηχανές Ανυσματικής Στήριξης.....	- 22 -
2.1.4 Ενίσχυση.....	- 23 -
2.1.5 Δίκτυα μάθησης σε βάθος.....	- 24 -
2.2 Αναγνώριση προσώπου.....	- 25 -
2.2.1 Ενεργή εμφάνιση και μοντέλα 3-Διάστατων σχημάτων.....	- 27 -
2.2.2 Μάθηση σε Βάθος.....	- 31 -
2.2.3 Αναγνώριση αντικειμένων.....	- 33 -
3. Python.....	- 34 -
3.1 Η βιβλιοθήκη OpenCV.....	- 35 -
3.2 Ανίχνευση με Haar-Cascade ταξινομητές.....	- 38 -
3.3 LBPH Ταξινομητής.....	- 41 -
3.3.1 Περιγραφή – Λειτουργία του LBPH.....	- 41 -
3.3.2 Πλεονεκτήματα της LBPH.....	- 44 -
4. Ερμηνεία των υποχρεωτικών κανονισμών στα UMS πλοία.....	- 46 -
4.1 Dead Man Alarm Σύστημα.....	- 53 -
5. Περιγραφή της Εφαρμογής.....	- 57 -
5.1 Κώδικας σε Python.....	- 57 -
5.2 Δημιουργία σετ δεδομένων.....	- 58 -
5.3 Ανίχνευση - Αναγνώριση προσώπου.....	- 59 -

6. Αποτελέσματα – Συμπεράσματα – Μελλοντική έρευνα..... - 62 -
7. Αναφορές - Βιβλιογραφία..... - 64 -

Κατάλογος Εικόνων

- Εικόνα 1: Αναγνώριση αντικειμένων σε τυχαία εικόνα, Πηγή: <https://azati.ai/image-detection-recognition-and-classification-with-machine-learning/> - 11 -
Εικόνα 2: Abraham Lincoln, κλίμακα του γκρι, Πηγή: https://www.researchgate.net/figure/mage-of-Abraham-Lincoln-as-a-matrix-of-pixel-values_fig1_330902210 - 13 -
Εικόνα 3: Δείγμα χώρου μνήμης εικονοστοιχείων - 13 -
Εικόνα 4: Λογική αποθήκευση εικονοστοιχείων στη μνήμη - 14 -
Εικόνα 5: Δείγμα χρωμάτων RGB - 14 -
Εικόνα 6: Ανίχνευση προσώπου και Deep Learning, Πηγή: <https://rapidapi.com/blog/top-facial-recognition-apis/>..... - 16 -
Εικόνα 7: Αποτελέσματα ανίχνευσης προσώπου παραγόμενα από τους Rowley, Baluja, και Kanade (1998) © 1998 IEEE. Μπορείτε να διακρίνετε το ψευδώς ανιχνευμένο (ένα ορθογώνιο πλαίσιο που δεν περιέχει κανένα πρόσωπο) ανάμεσα σε 57 αληθώς ανιχνευμένα αποτελέσματα;..... - 20 -
Εικόνα 8: Ένα νευρωνικό δίκτυο για ανίχνευση προσώπου (Rowley, Baluja, and Kanade 1998) © 1998 IEEE. Επικαλυπτόμενες περιοχές εξάγονται από διαφορετικά επίπεδα μιας πυραμίδας και επεξεργάζονται. Ένα νευρωνικό δίκτυο τριών επιπέδων χρησιμοποιείται έπειτα για να ανιχνεύσει περιοχές που είναι πιθανόν να περιέχουν πρόσωπο..... - 22 -
Εικόνα 9: Οικεία πρόσωπα, εικόνες χαμηλής ανάλυσης - 27 -
Εικόνα 10: Κύριες παραλλαγές στα Μοντέλα Ενεργού Σχήματος (Cootes, Edwards and Taylor 2001). Οι τέσσερις εικόνες δείχνουν την επίδραση των ταυτόχρονων αλλαγών των πρώτων τεσσάρων διακυμάνσεων σε σχήμα και υφή, από το μέσο όρο. Μπορούμε ξεκάθαρα να διακρίνουμε πως το σχήμα του προσώπου και η σκίαση επηρεάζονται ταυτόχρονα. ... - 28 -
Εικόνα 11: Ιχνηλάτηση προσώπου και frontalization: (a) χρησιμοποιώντας 3-Διάστατα Μοντέλα Ενεργής Εμφάνισης, απεικονίζοντας video frames με την εκτιμώμενες παραμέτρους κλίσης και περιστροφής και το 3-Διάστατο παραμορφωμένο πλέγμα (b) χρησιμοποιώντας 6 και μετά 67 σημεία εμπιστοσύνης στο σύστημα DeepFace, χρησιμοποιούμενα για το frontalization του προσώπου (κάτω σειρά). - 30 -
Εικόνα 12: Η αρχιτεκτονική της DeepFace (Taigman, Yang et al. 2014) © 2014 IEEE, ξεκινά με ένα στάδιο frontalization, ακολουθούμενο από μερικά μη-συνεστραμμένα φίλτρα και έπειτα από δυο πλήρως συνδεδεμένα επίπεδα, με softmax K-κλάσεων..... - 31 -
Εικόνα 13: Ανίχνευση ατόμων και αναγνώριση χρησιμοποιώντας ένα συνδυασμένο μοντέλο προσώπου, μαλλιών και στέρνου (Sivic, Zitnick, και Szeliski 2006) © 2006 Springer. (a) Χρησιμοποιώντας μόνο ανίχνευση προσώπου, πολλά κεφάλια δεν ανιχνεύονται. (b) Το συνδυασμένο μοντέλο προσώπου και ρούχων ανιχνεύει επιτυχώς όλα τα μέλη στη φωτογραφία. - 33 -
Εικόνα 14: Λογότυπα της Python και OpenCV - 35 -
Εικόνα 15: Γνωρίσματα γωνιών, γραμμών και ορθογωνίων - 38 -
Εικόνα 16: Απλά γνωρίσματα που χρησιμοποιούνται σε ανιχνευτή προσώπου που βασίζεται στην ενίσχυση (Viola και Jones 2004) © 2004 Springer. Το πρώτο και δεύτερο γνώρισμα

επιλέγεται από τον Adaboost. Το πρώτο γνώρισμα μετράει τις διαφορές στην ένταση των εικονοστοιχείων μεταξύ των ματιών και των μάγουλων, το δεύτερο μεταξύ των ματιών και της γέφυρας της μύτης..... - 39 -

Εικόνα 17: Αναπαράσταση ενός 3x3 Macro-block..... - 42 -

Εικόνα 18: (Αριστερά) Εικόνα προσώπου σε κλίμακα του γκρι, (κέντρο) εικόνα εφαρμόζοντας LBP και (δεξιά) Ιστόγραμμα της LBP εικόνας, Πηγή:
<https://journals.sagepub.com/doi/full/10.1177/56>..... - 44 -

Εικόνα 19: Υπόδειγμα εγκατάστασης Dead Man Alarm, Πηγή:
<http://www.deadmanalarm.com/sm-dma-watch-alarm/>..... - 53 -

Εικόνα 20: Εικόνες σετ Δεδομένων - 62 -

Εικόνα 21: Ανίχνευση και Αναγνώριση Προσώπου με το λογισμικό, σε συνθήκες κανονικού φωτισμού..... - 62 -

Εικόνα 22: Ανίχνευση και αναγνώριση προσώπου με γυαλιά/γκριμάτσα..... - 63 -

Κατάλογος Συντομογραφιών

RGB – Red, Green, Blue

CMYK – Cyan, Magenta, Yellow, Black (Key)

SVM - Support Vector Machines

AAM – Active Appearance Models

MEE – Μοντέλα Ενεργής Εμφάνισης

PCA – Principal Component Analysis

DIFS – Distance In Face Space

DFFS – Distance From Face Space

MLP – Multi-Layer Perceptron

H/Y – Ηλεκτρονικός Υπολογιστής

LFW – Labeled Faces In the Wild

YTF – YouTube Faces

IJB – IARPA Janus Benchmark

CWI – Centrum Wiskunde & Informatica Amsterdam

OpenCV – Open Source Computer Vision Library

IEEE – Institute of Electrical and Electronic Engineers

AI – Artificial Intelligence

TNΔ – Τεχνητά Νευρωνικά Δίκτυα

LBPH – Local Binary Patterns Histogram

UMS – Unmanned Machinery Space

IMO – International Maritime Organization

SOLAS – International Convention for the Safety of Life at Sea

DNV – Det Norske Veritas

IACS – International Association of Classification Societies

H/Z – Ηλεκτροζεύγος

GRT – Gross Register Tonnage

MLC – Maritime Labor Convention

Περίληψη

Στο πλαίσιο αυτής της Διπλωματικής Διατριβής, έγινε προσπάθεια να αναφερθούμε σε κάποια από τα χαρακτηριστικά των αλγορίθμων αναγνώρισης προσώπων, και να υλοποιηθεί ένας τέτοιος αλγόριθμος σε Python shell με τη χρήση της βιβλιοθήκης ανοικτού κώδικα OpenCV. Στόχος της εφαρμογής είναι η μέγιστη αποτελεσματικότητα με το ελάχιστο δυνατό κόστος. Όπως παρουσιάζεται παρακάτω, η αναγνώριση προσώπου με τη χρήση Μηχανικής Όρασης βασίζεται σε δυο διαφορετικές διαδικασίες. Η πρώτη εστιάζεται στο αν υπάρχει κάποιο πρόσωπο στο εξεταζόμενο αρχείο πραγματικού χρόνου (στην προκειμένη περίπτωση καταγραφή βίντεο από μικροκάμερα). Η δεύτερη εστιάζεται στην εκπαίδευση του μοντέλου με συγκεκριμένα δεδομένα αυτόματα, και η σύγκριση των δεδομένων ώστε να κριθεί εάν ταιριάζουν ή όχι με κάποια κριτήρια που ορίζουμε εμείς.

Επιπρόσθετα, και για την υλοποίηση της εφαρμογής “Dead Man Alarm” που χρησιμοποιείται βάσει κανονισμού σε μηχανοστάσια εμπορικών πλοίων, ο αλγόριθμος εξετάζει και αποφασίζει εάν τίθεται λόγος ειδοποίησης των άλλων μελών του πληρώματος σε περίπτωση που κατά την επιθεώρηση μηχανημάτων του πλοίου, εκτός κανονικού ωραρίου, βρεθεί μέλος του πληρώματος ακίνητο.

Έτσι, μετά την επιγραμματική επεξήγηση των επιμέρους τμημάτων του αλγορίθμου, γίνεται αναλυτική παρουσίαση αυτού, καθώς και του χρησιμοποιούμενου εξοπλισμού. Τα αποτελέσματα του αλγορίθμου δείχνουν ότι αποδίδει καλά κάτω από τις δύσκολες συνθήκες που επικρατούν στο μηχανοστάσιο του πλοίου (υψηλές θερμοκρασίες, κραδασμοί ασταθούς συχνότητας και διάρκειας, επικίνδυνα αέρια, κακός φωτισμός) αναγνωρίζοντας και ταυτοποιώντας κάποιο πρόσωπο, και στη συνέχεια εξετάζοντας εάν τίθεται λόγος ανησυχίας λόγω ξαφνικής ακινησίας κάποιου μέλους του πληρώματος.

Abstract

In the following chapters of this Thesis, an attempt was made to explain some of the face recognition algorithms and their features and implement such algorithms in Python shell using the OpenCV open source library. The goal of the application is maximum efficiency with the lowest possible cost. As shown below, face recognition using Mechanical Vision is based on two different processes. The first process focuses on whether there is a person in the real-time file in question (in this case live video stream using a web camera). On top of this, the second process focuses on training the model with specific data obtained from a pre-processed dataset and comparing these data with the live stream, in order to judge if the data fits some of the criteria we set, i.e. if a known person appears.

In addition, for the implementation of the "Dead Man Alarm" application, which is used according to the Class regulations in commercial vessels' machinery space, the algorithm examines and decides whether there is a reason to notify the other crew members in case of, during the inspection of vessel's machinery after normal hours on UMS mode, a crew member becomes alarmingly motionless.

Thus, after the brief explanation of the individual parts of mentioned algorithm, there is a detailed presentation of algorithm. In addition, short mention is given to the used hardware, which was chosen to be simple and cost-effective. The results of the algorithm show that it performs well under the adverse environmental conditions prevailing in the vessel's engine room (high temperatures, vibrations of unstable frequency and duration, dangerous gases, poor lighting) by recognizing and identifying a person. Finally, we examine whether sudden immobility of a crew member, i.e. not acknowledgement of the alarm after preset time, poses a reason of concern.

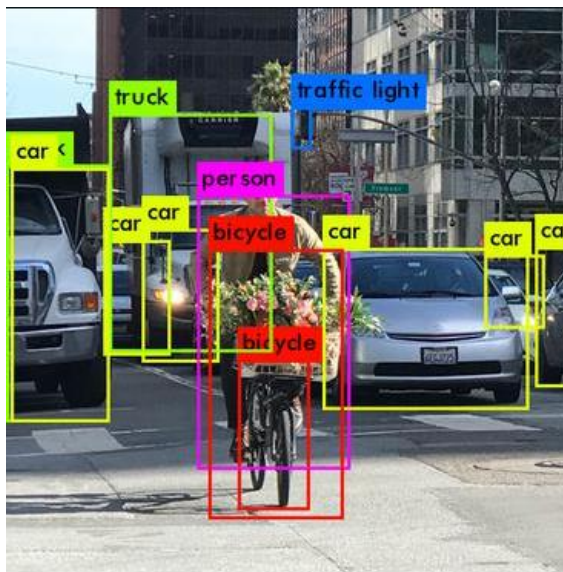
1. Λίγα λόγια για τη Μηχανική Όραση

Ένας από τους πιο ισχυρούς τύπους Τεχνητής Νοημοσύνης (AI), είναι η μηχανική όραση, την οποία σίγουρα πολλοί από εμάς συναντούμε στην καθημερινότητα μας, χωρίς καν να το γνωρίζουμε. Παρακάτω επεξηγείται το πως ακριβώς λειτουργεί (και σίγουρα βελτιώνεται μέρα με τη μέρα).

Η μηχανική όραση είναι το πεδίο της επιστήμης των υπολογιστών που εστιάζει στο να «μιμείται» κάποια τμήματα της ανθρώπινης όρασης ώστε να επιτρέπει σε μια υπολογιστική διάταξη να αναγνωρίζει και να επεξεργάζεται αντικείμενα σε εικόνες και βίντεο, με τον ίδιο τρόπο που το καταφέρνουν και οι άνθρωποι. Μέχρι πολύ πρόσφατα, η Μηχανική Όραση λειτουργούσε με περιορισμένες δυνατότητες.

Με τις πρόσφατες έρευνες και βελτιώσεις στα Νευρωνικά Δίκτυα και την «σε βάθος μάθηση» (Deep Learning), έγιναν άλματα προόδου στο πεδίο της Μηχανικής Όρασης, σε σημείο που πλέον ξεπερνά τον άνθρωπο σε καθήκοντα όπως αναγνώριση και διακρίβωση αντικειμένων.

Ένας από τους σημαντικότερους παράγοντες πίσω από την εξέλιξη της Μηχανικής Όρασης είναι ο όγκος των δεδομένων που δημιουργούμε σήμερα, τα οποία μπορούν να χρησιμοποιηθούν και να εξελίσσουν τη Μηχανική Όραση.



Εικόνα 1: Αναγνώριση αντικειμένων σε τυχαία εικόνα, Πηγή: <https://azati.ai/image-detection-recognition-and-classification-with-machine-learning/>

Επιπρόσθετα του μεγάλου όγκου διαθέσιμων οπτικών δεδομένων (περισσότερες από 3 δισεκατομμύρια εικόνες διαμοιράζονται καθημερινά), η υπολογιστική ισχύς που είναι απαραίτητη για να τα επεξεργαστεί, είναι πλέον διαθέσιμη. Καθώς το πεδίο της Μηχανικής Όρασης έχει αναπτυχθεί με το νέο υλικολογισμικό, έχει βελτιωθεί και το ποσοστό ακριβείας της αναγνώρισης αντικειμένων. Σε λιγότερο από μια δεκαετία, τα σημερινά συστήματα έχουν φτάσει ακόμα και το 99% (από το περίπου 50% που βρισκόταν μέχρι πρότινος) κάνοντάς τα πιο ακριβή και πιο γρήγορα από τον άνθρωπο στην αντίδραση επεξεργασίας οπτικών δεδομένων.

Τα πειράματα στο πεδίο της Μηχανικής Όρασης ξεκίνησαν τη δεκαετία του 1950, η οποία χρησιμοποιήθηκε για εμπορικούς σκοπούς (πιο συγκεκριμένα, να διαχωρίζει δακτυλογραφημένο από χειρόγραφο κείμενο) στη δεκαετία του 1970. Σήμερα, οι εφαρμογές της Μηχανικής Όρασης έχουν αναπτυχθεί εκθετικά.

«Μέχρι το 2022, η αγορά της Μηχανικής Όρασης και του υλικολογισμικού που σχετίζεται, αναμένεται να φτάσει τα \$48.6 δισεκατομμύρια» Πηγή: Tractica report, “Computer Vision Technologies and Markets”, Ιστότοπος: <https://omdia.tech.informa.com/topic-pages/artificial-intelligence>, 2015

1.1 Πως λειτουργεί η μηχανική όραση

Ένα από τα μεγαλύτερα ανοιχτά ερωτήματα στα πεδία της Νευροχειρουργικής και της Μηχανικής Μάθησης είναι: *Πώς ακριβώς λειτουργεί ο ανθρώπινος εγκέφαλος; Πως μπορούμε να πλησιάσουμε ή να μιμηθούμε αυτές τις λειτουργίες με δικούς μας αλγόριθμους;*

Η πραγματικότητα είναι ότι υπάρχουν πολύ λίγες αποδεδειγμένες θεωρίες πάνω στον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου, και παρόλο που τα Νευρωνικά Δίκτυα υποτίθεται πως «μιμούνται τον τρόπο λειτουργίας του εγκεφάλου», κανείς δεν είναι σίγουρος ότι κάτι τέτοιο αληθεύει.

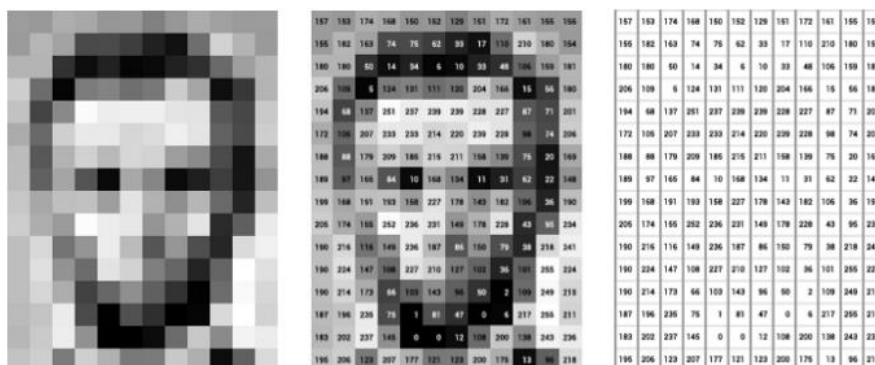
Το ίδιο παράδοξο συμβαίνει και με τη Μηχανική Όραση – από τη στιγμή που δεν έχουμε αποφασίσει πως λειτουργεί και επεξεργάζεται δεδομένα ο εγκέφαλος, δεχόμενος οπτικά δεδομένα από τα μάτια, είναι δύσκολο να αποφανθούμε πόσο καλά οι αλγόριθμοι που χρησιμοποιούνται σήμερα πλησιάζουν τις νοητικές μας διαδικασίες.

Σε ένα συγκεκριμένο επίπεδο, μπορούμε να πούμε ότι η Μηχανική Όραση είναι αναγνώριση μοτίβων – διατάξεων. Έτσι, ένας τρόπος να εκπαιδύσουμε έναν υπολογιστή να καταλαβαίνει οπτικά δεδομένα, είναι να του εισάγουμε εικόνες (χιλιάδες ή ακόμη και εκατομμύρια, αν είναι δυνατόν) οι οποίες έχουν κατηγοριοποιηθεί, και έπειτα να τις υποβάλλουμε σε έλεγχο από εργαλεία λογισμικού που μπορούν να αναγνωρίσουν αυτά τα μοτίβα που σχετίζονται με τις κατηγορίες ή αντικείμενα.

Για παράδειγμα, εάν εισάγουμε σε έναν υπολογιστή εκατομμύρια εικόνες από σκύλους, θα τις υποβάλλει στους αλγορίθμους που αναλύουν χρώματα, σχήματα, αποστάσεις μεταξύ των σχημάτων, όρια κοντινά τοποθετημένων αντικειμένων κλπ. σε σημείο που μπορεί να αναγνωρίσει τι σημαίνει ένα προφίλ «Σκύλος».

Όταν αυτή η διαδικασία έχει ολοκληρωθεί, ο υπολογιστής θα μπορεί (θεωρητικά) να χρησιμοποιήσει αυτή την εμπειρία σε μελλοντικές εικόνες που θα εισάγουμε, και να ξεχωρίσει αυτές που μοιάζουν με σκύλου.

Ας μιλήσουμε όμως πιο τεχνικά. Παρακάτω βλέπουμε μια φωτογραφία του Abraham Lincoln σε κλίμακα του γκρι. Η φωτεινότητα του κάθε εικονοστοιχείου (pixel) εκφράζεται από έναν αριθμό 8-bit, με εύρος κλίμακας από 0 (μαύρο) έως 255 (λευκό):



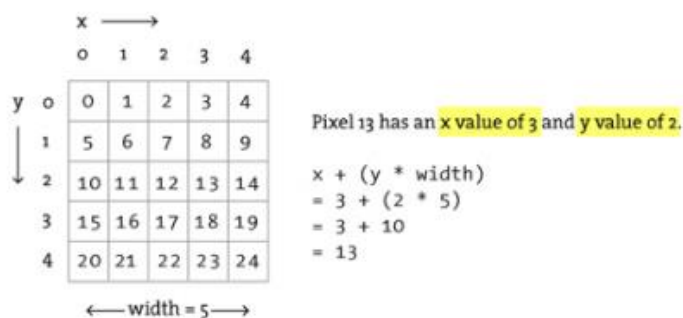
Εικόνα 2: Abraham Lincoln, κλίμακα του γκρι, Πηγή: https://www.researchgate.net/figure/mage-of-Abraham-Lincoln-as-a-matrix-of-pixel-values_fig1_330902210

Γεγονός είναι ότι, οι τιμές των εικονοστοιχείων είναι σχεδόν πάντα αποθηκευμένες, σε επίπεδο hardware, σε έναν μονοδιάστατο πίνακα. Για παράδειγμα, τα δεδομένα της παραπάνω εικόνας αποθηκεύονται με ένα τρόπο παρόμοιο με την παρακάτω μακρά λίστα απροσήμαστων χαρακτήρων:

```
{157, 153, 174, 168, 150, 152, 129, 151, 172, 161, 155, 156,
155, 182, 163, 74, 75, 62, 33, 17, 110, 210, 180, 154,
180, 180, 50, 14, 34, 6, 10, 33, 49, 106, 159, 181,
206, 109, 5, 124, 131, 111, 120, 204, 166, 15, 56, 180,
194, 68, 137, 251, 237, 239, 239, 228, 227, 87, 71, 201,
172, 105, 207, 233, 233, 214, 220, 239, 228, 98, 74, 206,
188, 88, 179, 209, 185, 215, 211, 158, 139, 75, 20, 169,
189, 97, 165, 84, 10, 168, 134, 11, 31, 62, 22, 148,
199, 168, 193, 193, 158, 227, 178, 143, 182, 106, 36, 190,
205, 174, 155, 252, 236, 231, 149, 178, 228, 43, 95, 234,
190, 216, 116, 149, 236, 187, 86, 150, 79, 38, 218, 241,
190, 224, 147, 108, 227, 210, 127, 102, 36, 101, 255, 224,
190, 214, 173, 66, 103, 143, 96, 50, 2, 109, 249, 215,
187, 196, 235, 75, 1, 81, 47, 0, 6, 217, 255, 211,
183, 202, 237, 145, 0, 0, 12, 108, 200, 138, 243, 236,
195, 206, 123, 207, 177, 121, 123, 200, 175, 13, 96, 218};
```

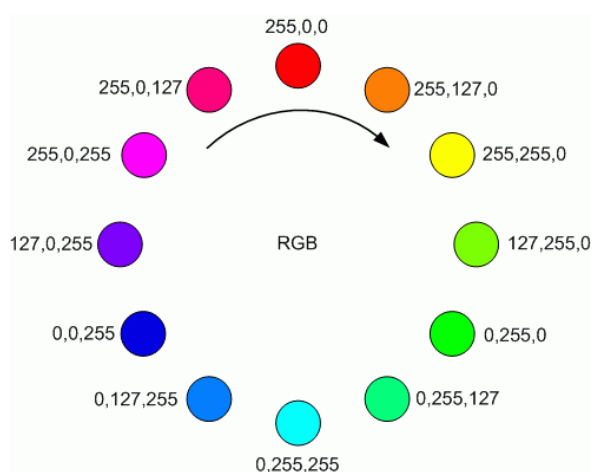
Εικόνα 3: Δείγμα χώρου μνήμης εικονοστοιχείων

Αυτός ο τρόπος αποθήκευσης δεδομένων των εικόνων, μπορεί να λειτουργεί αντίθετα από τις προσδοκίες μας, καθώς σίγουρα τα δεδομένα δείχνουν διδιάστατα όταν εμφανίζονται. Ναι, αυτό ισχύει, καθώς η μνήμη του υπολογιστή αποτελείται απλά από μια γραμμικά αυξανόμενη λίστα χώρων μνήμης.



Εικόνα 4: Λογική αποθήκευσης εικονοστοιχείων στη μνήμη

Τώρα, ας πάμε πίσω στην «Εικόνα 2» και ας φανταστούμε ότι εισάγουμε μια έγχρωμη. Εδώ τα πράγματα αρχίζουν και γίνονται λίγο πιο πολύπλοκα. Οι υπολογιστές συνήθως αναγνωρίζουν το χρώμα σαν σειρά τριών παραμέτρων – κόκκινο, πράσινο, μπλε (RGB) – στο ίδιο εύρος τιμών 0-255. Υπάρχουν και άλλες κωδικοποιήσεις χρωμάτων (CMYK, Pantone κλπ.) στις οποίες δε θα αναφερθούμε σε αυτή την έρευνα. Οι τιμές της κάθε απόχρωσης, είναι συνδυασμός αυτών των τριών διακριτών τιμών των χρωμάτων που προαναφέρθηκαν. Κάθε εικονοστοιχείο (pixel) αντιστοιχείται από τρεις τιμές ώστε να αποθηκευτεί από τον υπολογιστή, επιπρόσθετα της θέσης του. Εάν προσπαθούσαμε να κάνουμε έγχρωμο τον Πρόεδρο Lincoln, θα οδηγούμασταν σε 12x16x3 τιμές, ή 576 αριθμούς.



Εικόνα 5: Δείγμα χρωμάτων RGB

Βλέπουμε ότι απαιτείται αρκετή μνήμη για μια εικόνα μόνο και αρκετά εικονοστοιχεία για να επεξεργαστεί ένας αλγόριθμος. Για να εκπαιδευσουμε ένα μοντέλο με αρκετή αξιοπιστία ωστόσο, ειδικότερα όταν μιλάμε για Μάθηση σε βάθος, θα χρειαζόμασταν δεκάδες χιλιάδες εικόνες, και όσο περισσότερες τόσο το καλύτερο.

1.2 Η εξέλιξη της Μηχανικής Μάθησης

Πριν την έλευση της Μάθησης σε βάθος, οι εργασίες που θα μπορούσε να εκτελέσει η Μηχανική Όραση ήταν πολύ περιορισμένες και απαιτούσαν εκτενή, δυσνόητο κώδικα και αρκετή προσπάθεια από τους προγραμματιστές και τους χειριστές. Για παράδειγμα, αν ήθελε κάποιος να πετύχει αναγνώριση προσώπων, θα έπρεπε να ακολουθήσει τα παρακάτω βήματα:

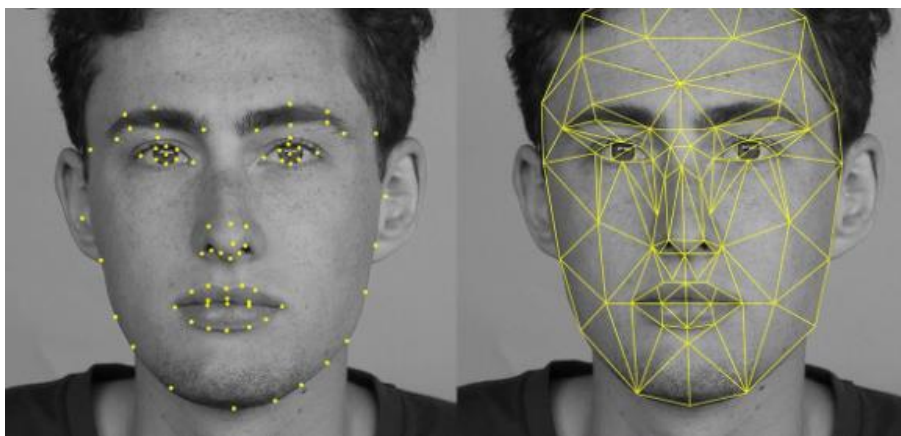
- **Δημιουργία Τράπεζας Δεδομένων:** θα έπρεπε να ληφθούν μεμονωμένες εικόνες από όλα τα πρόσωπα προς αναγνώριση σε συγκεκριμένη μορφή
- **Προσθήκη σχολίων σε εικόνες:** Έπειτα για κάθε εικόνα, θα έπρεπε να εισαχθούν μερικά σημαντικά χαρακτηριστικά-κλειδιά όπως την απόσταση μεταξύ των ματιών, το πλάτος της μύτης, την απόσταση μεταξύ του άνω χείλους και της μύτης, και δεκάδες άλλα γνωρίσματα που ορίζουν τα μοναδικά χαρακτηριστικά του κάθε ατόμου.
- **Λήψη νέων εικόνων:** Τέλος θα χρειαζόταν λήψη νέων εικόνων, φωτογραφίες ή στιγμιότυπα από βίντεο. Έπειτα θα έπρεπε να γίνουν οι απαραίτητες μετρήσεις ξανά, σε σημεία-κλειδιά της εικόνας. Ένας άλλος παράγοντας που πρέπει να ληφθεί υπόψη είναι η γωνία θέασης/λήψης της φωτογραφίας.

Μετά από όλη αυτή τη χειρωνακτική εργασία, η εφαρμογή θα μπορούσε τελικά να συγκρίνει τις μετρήσεις στις νέες εικόνες με τις αποθηκευμένες στη βάση δεδομένων, και να κρίνει αν ταιριάζουν με το προφίλ που ακολουθεί. Στην πραγματικότητα, δεν χρησιμοποιήθηκε σχεδόν καθόλου αυτοματισμός και η περισσότερη δουλειά έγινε χειροκίνητα. Το περιθώριο λάθους παραμένει υψηλό.

Η Μηχανική Μάθηση προσέφερε μια διαφορετική προσέγγιση στην επίλυση των προβλημάτων της μηχανικής όρασης. Με την Μηχανική Μάθηση, οι προγραμματιστές δεν χρειαζόταν πλέον να προγραμματίζουν χειροκίνητα κάθε κανόνα στις εφαρμογές τους. Αντί αυτού, προγραμματίζαν «υπορουτίνες», μικρότερες εφαρμογές που μπορούν να αναγνωρίσουν μοτίβα σε εικόνες. Έπειτα χρησιμοποιούσαν έναν αλγόριθμο στατιστικής μάθησης όπως η

Γραμμική Παλινδρόμηση, δέντρα αποφάσεων ή Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines, SVM) ώστε να αναγνωρίζουν μοτίβα, να κατηγοριοποιούν εικόνες και να αναγνωρίζουν αντικείμενα σε αυτές.

Η Μηχανική Μάθηση βοήθησε στην επίλυση πολλών προβλημάτων που ήταν ιστορικά απαιτητικά για τις κλασικές προσεγγίσεις ή εργαλεία λογισμικού. Για παράδειγμα, χρόνια πριν, μηχανικοί του τομέα κατάφεραν να δημιουργήσουν αλγόριθμο που μπορούσε να προβλέψει τους χρόνους επιβίωσης σε περιπτώσεις καρκίνου του μαστού, καλύτερα από τους ειδήμονες της ιατρικής κοινότητας. Ωστόσο, χρειάστηκαν οι κόποι αρκετών μηχανικών και επιστημόνων και πολύς χρόνος ώστε να αναπτυχθεί αξιόπιστο λογισμικό.



Εικόνα 6: Ανίχνευση προσώπου και Deep Learning, Πηγή: <https://rapidapi.com/blog/top-facial-recognition-apis/>

Η Μάθηση σε βάθος έδωσε τη δυνατότητα μιας εντελώς διαφορετικής προσέγγισης χρησιμοποιώντας Μηχανική Μάθηση. Η Μάθηση εις βάθος βασίζεται σε Νευρωνικά Δίκτυα, μια συνάρτηση γενικής χρήσης που μπορεί να λύσει κάθε πρόβλημα μέσω παραδειγμάτων. Όταν παρέχεις σε ένα Νευρωνικό Δίκτυο πολλά παραδείγματα ενός συγκεκριμένου τύπου δεδομένων, είναι δυνατό να αναγνωρίσει κοινά μοτίβα μεταξύ των παραδειγμάτων και να τα εκφράσει σε μια μαθηματική εξίσωση που θα βοηθήσει στην κατηγοριοποίηση μελλοντικών τμημάτων πληροφοριών.

Για παράδειγμα, για τη δημιουργία μιας εφαρμογής αναγνώρισης προσώπου με Μάθηση εις Βάθος, απαιτεί μόνο τον προγραμματισμό ή την επιλογή ενός ήδη έτοιμου

αλγορίθμου, και την εκπαίδευσή του με παραδείγματα των προσώπων των ατόμων που πρέπει να αναγνωρίσει. Με την παροχή πολλών παραδειγμάτων, το νευρωνικό δίκτυο θα είναι δυνατό να αναγνωρίσει πρόσωπα χωρίς περαιτέρω οδηγίες ή μετρήσεις.

Η Μάθηση σε Βάθος είναι πολύ αποτελεσματική μέθοδος για τη Μηχανική Όραση. Στις περισσότερες περιπτώσεις, η δημιουργία ενός καλού αλγορίθμου Μάθησης σε Βάθος εξαρτάται από την συγκέντρωση μεγάλου αριθμού κατηγοριοποιημένων δεδομένων εκπαίδευσης και τη ρύθμιση των παραμέτρων όπως ο αριθμός και ο τύπος των στρωμάτων των νευρωνικών δικτύων και ο αριθμός των βημάτων εκπαίδευσης. Συγκρινόμενη με παλαιότερες μεθόδους Μηχανικής Μάθησης, η Μάθηση σε Βάθος είναι ευκολότερη και γρηγορότερη.

Στις περισσότερες εφαρμογές Μηχανικής μάθησης σήμερα, όπως η ανίχνευση καρκίνου, τα αυτοκινούμενα οχήματα και η αναγνώριση προσώπου κάνουν χρήση της Μάθησης σε Βάθος. Τα Νευρωνικά Δίκτυα και η Μάθηση σε Βάθος έχουν περάσει από το στάδιο μια ιδέας σε πρακτικές εφαρμογές, χάρη στην μεγάλη διαθεσιμότητα και τις εξελίξεις στο υλικολογισμικό (hardware) και στους πόρους Cloud Computing.

Πόσο χρόνο χρειάζεται η αποκωδικοποίηση μιας εικόνας;

Πλέον, όχι πολύ. Αυτό είναι το κλειδί που η Μηχανική Όραση είναι τόσο συναρπαστική: Ενώ στο παρελθόν ακόμα και Υπερυπολογιστές θα χρειαζόνταν μέρες ή εβδομάδες να ολοκληρώσουν όλους τους υπολογισμούς που απαιτούνται, οι σημερινοί γρήγοροι επεξεργαστές και το υλικολογισμικό, συνεργαζόμενα με το γρήγορο, αξιόπιστο διαδίκτυο και τις Cloud υπηρεσίες, κάνουν όλες τις διεργασίες πολύ γρήγορες. Ένας πολύ σημαντικός παράγοντας είναι και η διάθεση πολλών μεγάλων εταιρειών όπως οι Microsoft, Facebook, Google, IBM να ερευνήσουν πάνω στην Τεχνητή Νοημοσύνη και να μοιραστούν τη δουλειά τους διαθέτοντας κομμάτια ανοικτού κώδικα Μηχανικής Μάθησης.

Αυτό επιτρέπει σε κάποιον να συνεχίσει να χτίζει πάνω στη δουλειά τους αντί να ξεκινήσει από το μηδέν. Ως αποτέλεσμα, η βιομηχανία της Τεχνητής Νοημοσύνης εξελίσσεται, και τα πειράματα που χρειαζόνταν ίσως εβδομάδες παλαιότερα, ίσως χρειάζονται το πολύ 15 λεπτά



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Τμήμα Μηχανικών Βιομηχανικής
Σχεδίασης και Παραγωγής

&

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Τμήμα Ναυτιλίας και
Επιχειρηματικών Υπηρεσιών



σήμερα. Σε πολλές από τις πραγματικές εφαρμογές της μηχανικής όρασης, αυτές οι διαδικασίες εκτελούνται συνεχώς σε μικροδευτερόλεπτα, κάνοντας τους υπολογιστές πλέον να δρουν ακόμα και σε «πραγματικό χρόνο».

2. Ανίχνευση και Αναγνώριση προσώπου

2.1 Ανίχνευση προσώπου

Πριν η ανίχνευση προσώπου να μπορέσει να εφαρμοστεί σε μία γενική εικόνα, οι τοποθεσίες και τα μεγέθη οποιωνδήποτε προσώπων πρέπει πρώτα να βρεθούν (Εικόνα 7). Θεωρητικά, μπορούμε να εφαρμόσουμε έναν αλγόριθμο ανίχνευσης σε κάθε εικονοστοιχείο και κλίμακα (Moghaddam and Pentland 1997) αλλά μια τέτοια διαδικασία θα ήταν πολύ αργή στην πράξη.

Τα τελευταία 40 χρόνια έχει αναπτυχθεί μια τεράστια ποικιλία από γρήγορους αλγόριθμους ανιχνεύσεις προσώπου. Οι Yang, Kriegman, και οι Ahuja (2002) και οι Zhao, Chellappa et al. (2003b) παρέχουν κατανοητές έρευνες για προγενέστερη εργασία σε αυτόν τον τομέα. Σύμφωνα με την ταξινόμηση, η ανίχνευση προσώπου και οι τεχνικές της κατατάσσονται σε: βασισμένες σε γνωρίσματα, βασισμένες σε πρότυπα και βασισμένες στην εμφάνιση. Οι τεχνικές οι οποίες είναι βασισμένες σε γνωρίσματα προσπαθούν να βρουν τη θέση ιδιαίτερων χαρακτηριστικών γνωρισμάτων στην εικόνα όπως τα μάτια, η μύτη και το στόμα. Έπειτα επιβεβαιώνουν εάν αυτά τα χαρακτηριστικά είναι σε ευλογοφανή γεωμετρική διάταξη. Αυτές οι τεχνικές περιέχουν μερικές από τις πιο πρώιμες προσεγγίσεις της αναγνώρισης προσώπου (Fischler και Elschlager 1973; Kanade 1977; Yuille 1991). Επίσης, οι προσεγγίσεις βασίζονται σε αρθρωτούς ιδιοχώρους (Moghaddam και Pentland 1997), τοπικά filter jets (Leung, Burl, και Perona 1995; Penev και Atick 1996; Wiskott, Fellous et al. 1997), υποστηρικτικές μηχανές τομέων – support vector machines (Heisele, Ho et al. 2003; Heisele, Serre, και Poggio 2007) και ενίσχυσης (Schneiderman and Kanade 2004).

Οι προσεγγίσεις οι οποίες είναι βασισμένες σε πρότυπα, όπως τα μοντέλα ενεργής εμφάνισης – active appearance models (AAMs) μπορούν να αντιμετωπίσουν και να καλύψουν μια μεγάλη γκάμα από φιγούρες και εκφράσεις. Τυπικά, απαιτούν καλή αρχικοποίηση με εικόνες κοντά στο πραγματικό πρόσωπο και για αυτό είναι ικανοποιητικά ως γρήγοροι ανιχνευτές προσώπου.

Οι προσεγγίσεις με βάση την εμφάνιση σκανάρουν σε επικαλυπτόμενες ορθογώνιες επιφάνειες της εικόνας ψάχνοντας για πιθανά υποψήφια πρόσωπα, που μπορούν να επεξεργαστούν χρησιμοποιώντας επικάλυψη ακριβότερων αλλά επιλεκτικών αλγορίθμων ανίχνευσης (Sung και Poggio 1998; Rowley, Baluja, και Kanade 1998; Romdhani, Torr et al. 2001; Fleuret και Geman 2001; Viola και Jones 2004). Για να αντιμετωπίσει τις διακυμάνσεις κλίμακας, η εικόνα συνήθως μετατρέπεται σε υπό-οκτάβια πυραμίδα και ένας ξεχωριστός σαρωτής επιδρά σε κάθε επίπεδο. Οι περισσότερες προσεγγίσεις με βάση την εμφάνιση βασίζονται κυρίως στους εκπαιδευτικούς επαληθευτές χρησιμοποιώντας σετ από ετικέτες προσώπων και όχι απλές επιφάνειες.



Εικόνα 7: Αποτελέσματα ανίχνευσης προσώπου παραγόμενα από τους Rowley, Baluja, και Kanade (1998) © 1998 IEEE. Μπορείτε να διακρίνετε το ψευδώς ανιχνευμένο (ένα ορθογώνιο πλαίσιο που δεν περιέχει κανένα πρόσωπο) ανάμεσα σε 57 αληθώς ανιχνευμένα αποτελέσματα;

Οι Sung και Poggio (1998) και οι Rowley, Baluja, και Kanade (1998) παρουσιάζουν δύο από τους πιο πρώιμους ανιχνευτές προσώπου με βάση την εμφάνιση και εισάγουν έναν αριθμό από καινοτομίες που είναι ευρέως χρησιμοποιημένοι σε μεταγενέστερη επεξεργασία από άλλους. Ξεκινώντας, και τα δύο συστήματα συλλέγουν ένα σετ από επιφάνειες με ετικέτες προσώπων (Εικόνα 8) όπως και ένα σετ από τμήματα που έχουν εξαχθεί από εικόνες που δεν περιείχαν πρόσωπα, για παράδειγμα εναέριες φωτογραφίες ή βλάστηση. Οι συλλεγμένες φωτογραφίες ενισχύονται από τεχνητό καθρέφτισμα, περιστροφή ή αλλαγή κλίμακας και

μεταφράζοντας τις φωτογραφίες σε μικρά τμήματα ώστε να φτιαχτεί ανιχνευτής προσώπου λιγότερο ευαίσθητος σε τέτοια φαινόμενα.

Οι επόμενες παράγραφοι, παρέχουν γρήγορες αναλύσεις από έναν αριθμό από πρώιμους ανιχνευτές προσώπου βάσει εμφάνισης, κωδικοποιημένους από μηχανές αλγορίθμων και βασισμένους σε αυτές. Αυτά τα συστήματα παρέχουν μια ενδιαφέρουσα οπτική της προοδευτικής υιοθέτησης και εξέλιξης της μηχανικής μάθησης στην όραση με χρήση υπολογιστή. Πιο λεπτομερείς περιγραφές μπορούν να εντοπιστούν στις πρωτότυπες εργασίες.

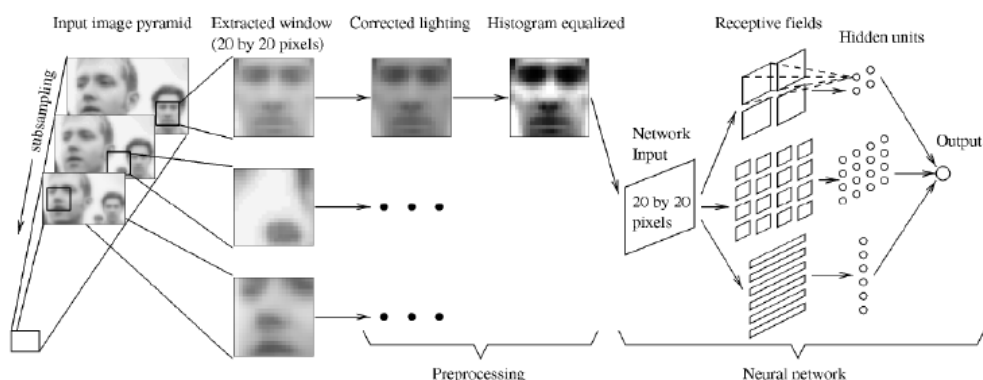
2.1.1 Ομαδοποίηση και PCA

Όταν το πρόσωπο και τα μοτίβα έχουν προ-επεξεργαστεί, οι Sung και Roggio (1998) ομαδοποιούν κάθε ένα από αυτά τα σετ δεδομένων σε έξι ξεχωριστές συστάδες χρησιμοποιώντας ομαδοποίηση K-Means και μετά εφαρμόζουν ανάλυση κύριων Συνιστωσών - Principal Component Analysis (PCA) υποδιαστήματα σε καθεμία από τις 12 συστάδες που προκύπτουν. Την ώρα της ανίχνευσης, οι DIFS and DFFS μέθοδοι μέτρησης οι οποίες αρχικά αναπτύχθηκαν από τους Moghaddam και Pentland (1997) χρησιμοποιήθηκαν για να παράγουν 24 μετρήσεις απόστασης Mahalanobis (δύο ανά συστάδα). Οι τελικές 24 μετρήσεις εισάγονται ως δεδομένα σε ένα πολυστρωματικό αντίληπτρο – multilayer perceptron (MLP), δηλ., ένα ολοκληρωτικά συνδεδεμένο νευρωνικό δίκτυο.

2.1.2 Νευρωνικά Δίκτυα

Αντί να γίνεται πρώτα ομαδοποίηση των δεδομένων και υπολογισμός των διαστάσεων Mahalanobis στις συστάδες, οι Rowley, Baluja, και Kanade (1998) εφαρμόζουν ένα νευρωνικό δίκτυο MLP απευθείας σε τμήματα των 20x20 εικονοστοιχείων διαφόρων εντάσεων της κλίμακας του γκρι, χρησιμοποιώντας μια ποικιλία από διαφορετικών μεγεθών χειροποίητα «πεδία αντίληψης» για να αιχμαλωτίσουν και τις μεγάλης κλίμακας αλλά και τις μικρότερες δομές (Εικόνα 8). Το νευρωνικό δίκτυο που απορρέει, εξαγεί απευθείας την πιθανότητα για ένα πρόσωπο στο κέντρο κάθε επικαλυπτόμενης έκτασης σε μια αναλυτική πυραμίδα.

Από τη στιγμή που αρκετές επικαλυπτόμενες εκτάσεις (και σε χώρο αλλά και ανάλυση) μπορούν να πυροδοτηθούν κοντά σε ένα πρόσωπο, ένα πρόσθετο «συγχωνευτικό» δίκτυο χρησιμοποιείται για να ενώσει επικαλυπτόμενες ανιχνεύσεις. Οι δημιουργοί επίσης πειραματίζονται με την εκπαίδευση διάφορων τύπων νευρωνικών δικτύων και συνδυάζουν τα αποτελέσματά τους. Η Εικόνα 7 δείχνει ένα δείγμα αποτελέσματος από το λογισμικό ανίχνευσης προσώπου τους.



Εικόνα 8: Ένα νευρωνικό δίκτυο για ανίχνευση προσώπου (Rowley, Baluja, and Kanade 1998) © 1998 IEEE. Επικαλυπτόμενες περιοχές εξάγονται από διαφορετικά επίπεδα μιας πυραμίδας και επεξεργάζονται. Ένα νευρωνικό δίκτυο τριών επιπέδων χρησιμοποιείται έπειτα για να ανιχνεύσει περιοχές που είναι πιθανόν να περιέχουν πρόσωπο.

2.1.3 Μηχανές Ανυσματικής Στήριξης

Αντί να χρησιμοποιούν νευρωνικά δίκτυα για την ταξινόμηση τμημάτων, οι Osuna, Freund, και Girosi (1997) χρησιμοποιούν μηχανές ανυσματικής στήριξης – support vector machines (SVM), για να κατηγοριοποιήσουν τα ίδια προ-επεξεργασμένα τμήματα όπως οι Sung και Poggio (1998). Μια SVM ψάχνει για μία σειρά από όρια ξεχωριστών επιπέδων σε χαρακτηριστικό διάστημα μεταξύ διαφορετικών κατηγοριών (στην περίπτωσή μας, εικόνα προσώπου και τμήματα χωρίς πρόσωπο). Σε αυτές τις περιπτώσεις όπου τα όρια γραμμικής ταξινόμησης είναι ανεπαρκή, το τμήμα του γνωρίσματος μπορεί να ανυψωθεί σε γνωρίσματα μεγαλύτερων διαστάσεων χρησιμοποιώντας πυρήνες - *kernels*. Οι SVMs έχουν χρησιμοποιηθεί από άλλους ερευνητές και για ανίχνευση και αναγνώριση προσώπου (Heisele, Ho et al. 2003; Heisele, Serre, και Poggio 2007) όπως επίσης και για γενική αναγνώριση αντικειμένων (Lampert 2008).

2.1.4 Ενίσχυση

Από όλα τα λογισμικά ανίχνευσης προσώπου που αναπτύχθηκαν την δεκαετία του 2000, αυτό που παρουσιάστηκε από τους Viola και Jones (2004) είναι πιθανόν το πιο αναγνωρίσιμο. Η τεχνική τους ήταν ή πρώτη που παρουσίασε την αντίληψη της ενίσχυσης (boosting) στην κοινότητα της μηχανικής όρασης, η οποία περιέχει την εκπαίδευση μιας σειράς από αυξητικά διαφοροποιούμενους απλούς ταξινομητές και μετά αναμιγνύει τα αποτελέσματα τους (Bishop 2006, Ενότητα 14.3; Hastie, Tibshirani, και Friedman 2009, Κεφάλαιο 10; Murphy 2012, Ενότητα 16.4; Glassner 2018, Ενότητα 14.7).

Πιο αναλυτικά, η ενίσχυση περιλαμβάνει την κατασκευή ενός ταξινομητή $h(x)$ ως άθροισμα απλών αδύναμων ταξινομητών – *weak learners*,

$$h(x) = \text{πρόσημο} \left[\sum_{j=0}^{m-1} a_j * h_j(x) \right]$$

Όπου ο καθένας από τους αδύναμους ταξινομητές $h_j(x)$ αποτελεί μια εξαιρετικά απλή λειτουργία της τροφοδοσίας, και συνεπώς, με συνεισφορά περισσότερων (μεμονωμένα), δεν αναμένεται αύξηση στην απόδοση της ταξινόμησης. Στις περισσότερες παραλλαγές της ενίσχυσης οι αδύναμοι ταξινομητές είναι λειτουργίες κατωφλίου,

$$h_j(x) = a_j * [f_j < \theta_j] + b_j * [f_j \geq \theta_j] = \begin{cases} a_j, & \text{αν } f_j < \theta_j \\ b_j, & \text{σε άλλη περίπτωση} \end{cases}$$

Που είναι επίσης γνωστές ως ρίζες αποφάσεων – decision stumps (βασικά, η απλούστερη πιθανή εκδοχή των δέντρων απόφασης – decision trees). Στις περισσότερες περιπτώσεις, είναι επίσης παράδοση (και απλούστερο) να θέσουμε ως a_j και b_j το ± 1 , πχ. $a_j = -s_j$, $b_j = +s_j$, έτσι ώστε μόνο το γνώρισμα f_j , η τιμή κατωφλίου θ_j , και η πολικότητα του κατωφλίου $s_j \in \pm 1$ πρέπει να επιλεγεί¹.

¹ Μερικές παραλλαγές, όπως αυτή των Viola και Jones (2004), χρησιμοποιούν $(s_j, b_j) \in [0, 1]$ και ρυθμίζουν τον αλγόριθμο εκμάθησης αναλόγως.

Σε πολλές εφαρμογές της ενίσχυσης, τα γνωρίσματα είναι απλοί άξονες συντεταγμένων x_k , δηλ., ο αλγόριθμος ενίσχυσης επιλέγει ένα από τα δεδομένα των σύνθετων διανυσμάτων ως το καλύτερο κατώφλι. Στο λογισμικό αναγνώρισης προσώπου των Viola και Jones, τα γνωρίσματα είναι διαφορές των ορθογώνιων περιοχών στο κομμάτι εικόνας που εισάγεται, όπως φαίνεται στην Εικόνα 16. Το πλεονέκτημα της χρήσης αυτών των γνωρισμάτων είναι αυτό, ενώ είναι πιο διακριτά από μεμονωμένα εικονοστοιχεία, ο υπολογισμός είναι εξαιρετικά γρήγορος όταν μια συνολική περιοχή πίνακα έχει προ-υπολογιστεί. Ιδανικά, για το κόστος μιας προ-υπολογιστικής φάσης $O(N)$ (όπου N είναι ο αριθμός των εικονοστοιχείων στην εικόνα), οι επακόλουθες διαφορές των ορθογωνίων μπορούν να υπολογιστούν με $4r$ προσθέσεις ή αφαιρέσεις, όπου $r \in \{2, 3, 4\}$ είναι ο αριθμός των ορθογωνίων στο γνώρισμα.

Το κλειδί της επιτυχίας της ενίσχυσης είναι η μέθοδος της επιλογής των αδύναμων ταξινομητών, με σταδιακά αυξανόμενο ρυθμό, και το επαναζύγισμα των εκπαιδευτικών παραδειγμάτων μετά από κάθε στάδιο. Ο AdaBoost (Adaptive Boosting) αλγόριθμος (Bishop 2006; Hastie, Tibshirani, και Friedman 2009; Murphy 2012) το πετυχαίνει αυτό με το επαναζύγισμα κάθε δείγματος εξετάζοντας εάν είναι σωστά ταξινομημένα σε κάθε στάδιο, και χρησιμοποιώντας σε κάθε στάδιο το σφάλμα της μέσης ταξινόμησης για να αποφασίσει τα τελικά ζυγίσματα α_j μεταξύ των αδύναμων ταξινομητών.

Για περισσότερη αύξηση της ταχύτητας του ανιχνευτή, είναι πιθανό να δημιουργηθεί ένας καταρράκτης - cascade από ταξινομητές, όπου ο κάθε ταξινομητής χρησιμοποιεί ένα μικρό αριθμό ελέγχων (ας υποθέσουμε, έναν AdaBoost ταξινομητή δυο όρων) για να απορρίψει ένα μεγάλο τμήμα από επιφάνειες χωρίς πρόσωπα ενώ προσπαθεί να προσπελάσει όλα τα πιθανά υποψήφια πρόσωπα (Fleuret και Geman 2001; Viola και Jones 2004; Brubaker, Wu et al. 2008).

2.1.5 Δίκτυα μάθησης σε βάθος

Εφόσον οι απαρχές της έρευνας στην ανίχνευση προσώπου τοποθετούνται στις αρχές της δεκαετίας του 2000, η ανίχνευση προσώπου και οι αλγόριθμοι συνεχίζουν να εξελίσσονται και να βελτιώνονται (Zafeiriou, Zhang, και Zhang 2015). Οι ερευνητές πρότειναν να

χρησιμοποιηθούν μια σειρά από καταρράκτες γνωρισμάτων (Li και Zhang 2013), μοντέλα παραμορφώσιμων τμημάτων – deformable parts models (Mathias, Benenson et al. 2014), συγκεντρωτικά γνωρίσματα καναλιών – aggregated channel features (Yang, Yan et al. 2014), και νευρωνικά δίκτυα (Li, Lin et al. 2015; Yang, Luo et al. 2015). Το «Wider Face» ορόσημο (Yang, Luo et al. 2016) περιέχει αποτελέσματα και είναι δείκτης σε πιο πρόσφατες εργασίες, περιέχοντας RetinaFace (Deng, Guo et al. 2020b), που συνδυάζει ιδέες από άλλα πρόσφατα νευρωνικά δίκτυα και ανιχνευτές αντικειμένων όπως τα Πυραμιδικά Δίκτυα Γνωρισμάτων – Feature Pyramid Networks (Lin, Dollar et al. 2017) και RetinaNet (Lin, Goyal et al. 2017), η οποία δίνει επίσης καλές κριτικές για άλλες πρόσφατες εργασίες σχετικές με ανίχνευση προσώπου.

2.2 Αναγνώριση προσώπου

Μεταξύ των διάφορων εργασιών αναγνώρισης που καλούνται να εκτελέσουν οι Ηλεκτρονικοί Υπολογιστές, η αναγνώριση προσώπου είναι από τις εργασίες που έχουν αδιαμφισβήτητη την μεγαλύτερη επιτυχία². Ενώ οι άνθρωποι δεν μπορούν εύκολα να αναγνωρίσουν παρόμοια πρόσωπα που δεν γνωρίζουν (O’Toole, Jiang et al. 2006; O’Toole, Phillips et al. 2009), η δυνατότητα των Υπολογιστών να διακρίνουν πρόσωπα σε ένα μικρό αριθμό μελών οικογενείας ή φίλων, έχει βρει την εφαρμογή της σε εφαρμογές επεξεργασίας φωτογραφιών επιπέδου λιανικού εμπορίου (όχι σε βιομηχανικές εφαρμογές). Η αναγνώριση προσώπου μπορεί να χρησιμοποιηθεί σε πάμπολλες επιπρόσθετες εφαρμογές, όπως η διεπαφή ανθρώπου-υπολογιστή, η αναγνώριση ταυτότητας (Kirovski, Jojic, και Jancke 2004), το ξεκλείδωμα του υπολογιστή με κάποιο λογαριασμό χρήστη (login), το γονικό έλεγχο και την παρακολούθηση ασθενών (Zhao, Chellappa et al. 2003a). Βεβαίως, όπως οτιδήποτε συνδέεται με προηγμένη τεχνολογία, πάντα υπάρχει και η πιθανότητα κακής χρήσης (Chokshi 2019; Ovide 2020).

² Η άμεση αναγνώριση, π.χ. η αναγνώριση γνωστών αντικειμένων όπως τοποθεσίες και επίπεδα αντικείμενα, είναι η έτερη επίσης πολύ επιτυχημένη εργασία γενικής αναγνώρισης. Στον γενικό τομέα της Βιομετρίας, π.χ. της αναγνώρισης ταυτότητας, εξειδικευμένες απεικονίσεις όπως η ίριδα του ματιού ή το δαχτυλικό αποτύπωμα εκτελούνται ακόμα καλύτερα (Jain, Bolle, and Pankanti 1999; Pankanti, Bolle, and Jain 2000; Daugman 2004).

Τα λογισμικά αναγνώρισης προσώπου λειτουργούν καλύτερα όταν τους παρέχονται εικόνες σε μεγάλη κλίμακα από διαφορετικές οπτικές γωνίες, διαφορετικό φωτισμό και εκφράσεις (Pose – Illumination – Expression) όπως στη μελέτη των (Phillips, Moon et al. 2000; Sim, Baker, και Bsat 2003; Gross, Shi, και Cohn 2005; Huang, Ramesh et al. 2007; Phillips, Scruggs et al. 2010). Τα πιο πρόσφατα ευρέως χρησιμοποιούμενα σετ δεδομένων περιλαμβάνουν τα Labeled Faces in the Wild (LFW) (Huang, Ramesh et al. 2007; Learned-Miller, Huang et al. 2016), το YouTube Faces (YTF) (Wolf, Hassner, and Maoz 2011), το MegaFace (Kemelmacher-Shlizerman, Seitz et al. 2016; Nech και Kemelmacher-Shlizerman 2017), και το IARPA Janus Benchmark (IJB) (Klare, Klein et al. 2015; Maze, Adams et al. 2018), όπως φαίνονται χαρακτηριστικά στον «Πίνακα 1: Αναγνώριση προσώπου και σετ δεδομένων ανίχνευσης, προσαρμοσμένο από Maze, Adams et al. (2018)». (Δείτε *Masi, Wu et al. 2018* για επιπρόσθετα σετ δεδομένων που χρησιμοποιούνται για την εκπαίδευση των εφαρμογών.)

Όνομα/Διαδικτυακός τόπος (URL)	Περιεχόμενα/Αναφορές
Βάση δεδομένων CMU Multi-PIE http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie	337 πρόσωπα ατόμων σε διάφορες πόζες <i>Gross, Matthews et al. (2010)</i>
Faces in the Wild http://vis-www.cs.umass.edu/lfw/	5,749 Διασημότητες του Διαδικτύου <i>Huang, Ramesh et al. (2007)</i>
YouTube Faces (YTF) https://www.cs.tau.ac.il/~wolf/ytfaces/	1595 Άνθρωποι σε 3425 Youtube βίντεο <i>Wolf, Hassner, and Maoz (2011)</i>
MegaFace 1M Internet faces http://megaface.cs.washington.edu/	1 Εκατομμύριο πρόσωπα από το Διαδίκτυο <i>Nech and Kemelmacher-Shlizerman (2017)</i>
IARPA Janus Benchmark (IJB) https://www.nist.gov/programs-projects/face-challenges	31,334 faces of 3,531 people in videos <i>Maze, Adams et al. (2018)</i>

Πίνακας 1: Αναγνώριση προσώπου και σετ δεδομένων ανίχνευσης, προσαρμοσμένο από Maze, Adams et al. (2018)

Μερικές από τις πρώιμες προσεγγίσεις στην αναγνώριση προσώπου περιλάμβαναν την εύρεση της θέσης κάποιων χαρακτηριστικών γνωρισμάτων της εικόνας, όπως τα μάτια, τη μύτη ή το στόμα, και τη μέτρηση των αποστάσεων μεταξύ των θέσεων αυτών των χαρακτηριστικών (Fischler και Elschlager 1973; Kanade 1977; Yuille 1991).

Άλλες προσεγγίσεις βασίζονταν στην σύγκριση εικόνων κλίμακας του γκρι (grayscale) προβαλλόμενες σε υποδιαστήματα μικρότερων διαστάσεων, αποκαλούμενα και *eigenfaces* και μοντελοποιώντας από κοινού τις διαφορές σε σχήμα και εμφάνιση (αφαιρώντας τις παραλλαγές

της πόζας) χρησιμοποιώντας μοντέλα ενεργής εμφάνισης, *active appearance models*. Περιγραφές των κλασικών συστημάτων αναγνώρισης προσώπου μπορούν να βρεθούν κάποιες από τις αναφορές αυτής της εργασίας καθώς και στον ιστότοπο Face Recognition³. Η έρευνα στην αναγνώριση προσώπων από τον άνθρωπο, από τους *Sinha, Balas et al. (2006)* αξίζει επίσης μελέτη καθώς περιλαμβάνει έναν αριθμό αναπάντεχων αποτελεσμάτων, όπως τη δυνατότητα των ανθρώπων να αναγνωρίζουν εικόνες οικείων προσώπων πολύ χαμηλής ανάλυσης (Εικόνα 9: Οικεία πρόσωπα, εικόνες χαμηλής ανάλυσης) και τη σημαντικότητα των φρυδιών στην αναγνώριση.



Εικόνα 9: Οικεία πρόσωπα, εικόνες χαμηλής ανάλυσης

2.2.1 Ενεργή εμφάνιση και μοντέλα 3-Διάστατων σχημάτων

Η ανάγκη της χρήσης αρθρωτών ή βασιζόμενων-στην-εμφάνιση χώρων *eigen* (*eigenspaces*) στην αναγνώριση προσώπων, είναι υποπερίπτωση μιας πιο γενικής παρατήρησης, π.χ. η εμφάνιση και η αναγνωρισιμότητα ενός προσώπου εξαρτώνται τόσο από το σχήμα, όσο και από το χρώμα ή την υφή (που είναι και αυτά τα γνωρίσματα που συλλαμβάνουν οι *eigenspaces*). Επιπροσθέτως, όταν έχουμε να χειριστούμε και 3-Διάστατη περιστροφή του προσώπου, η πόζα του ανθρώπινου προσώπου (η γωνία του κεφαλιού) θα πρέπει να παραβλέπεται - διορθώνεται όταν κάνουμε αναγνώριση.

Στην πραγματικότητα, τα πρωταρχικά συστήματα αναγνώρισης προσώπου, όπως αυτά των *Fischler και Elschlager (1973)*, *Kanade (1977)*, and *Yuille (1991)*, έβρισκαν συγκεκριμένα χαρακτηριστικά γνωρίσματα και σημεία σε εικόνες προσώπων και εκτελούσαν αναγνώριση βασιζόμενα στην σχετική θέση των γνωρισμάτων, καθώς επίσης και τις αναμεταξύ τους αποστάσεις. Μεταγενέστερες τεχνικές όπως η ανάλυση τοπικών στοιχείων *local feature analysis (Penev και Atick 1996)* και η ελαστική ομαδοποιημένη ταύτιση γραφημάτων - *elastic*

³ <http://www.face-rec.org/>

bunch graph matching (Wiskott, Fellous et al. 1997) όπου συνδύαζε κοινές αποκρίσεις σε τοπικά φίλτρα σε χαρακτηριστικές περιοχές γνωρισμάτων μαζί με μοντέλα σχημάτων, για να προχωρήσει σε αναγνώριση.

Ένα οπτικά ενδιαφέρον παράδειγμα όπου το σχήμα και η υφή είναι σημαντικά, είναι η δουλειά των Rowland και Perrett (1995), που εντόπισαν τις καμπύλες των χαρακτηριστικών του προσώπου και μετά χρησιμοποίησαν αυτές τις καμπύλες για να ομαλοποιήσουν ή «στρεβλώσουν» την εικόνα σε ένα πιο «κανονικό» σχήμα. Μετά την ανάλυση του σχήματος και του χρώματος σε μια εικόνα και την αναγνώριση αποκλίσεων από το μέσο όρο, μπόρεσαν να συνδέσουν συγκεκριμένες παραμορφώσεις σχήματος και χρώματος με προσωπικά χαρακτηριστικά όπως η ηλικία και το φύλο. Η δουλειά τους αποδεικνύει ότι το σχήμα και το χρώμα επηρεάζουν σημαντικά στην αντίληψη αυτών των χαρακτηριστικών.

Περίπου την ίδια περίοδο, οι ερευνητές της μηχανικής όρασης ξεκινούσαν να χρησιμοποιούν ταυτόχρονα παραμορφώσεις σχημάτων και εισαγωγή υφών, σε μια προσπάθεια να μοντελοποιήσουν την ποικιλία της εμφάνισης του προσώπου ανάλογα με την ταυτότητα ή την έκφραση (Beymer 1996; Vetter και Poggio 1997), αναπτύσσοντας τεχνικές όπως τα Μοντέλα Ενεργού Σχήματος – Active Shape Models (Lanitis, Taylor, και Cootes 1997), τα 3-Διάστατα Μορφοποιούμενα Μοντέλα – 3D Morphable Models (Blanz και Vetter 1999; Egger, Smith et al. 2020) και την ελαστική ομαδοποιημένη ταύτιση γραφικών - *Elastic Bunch Graph Matching* (Wiskott, Fellous et al. 1997).



Εικόνα 10: Κύριες παραλλαγές στα Μοντέλα Ενεργού Σχήματος (Cootes, Edwards and Taylor 2001). Οι τέσσερις εικόνες δείχνουν την επίδραση των ταυτόχρονων αλλαγών των πρώτων τεσσάρων διακυμάνσεων

σε σχήμα και υφή, από το μέσο όρο. Μπορούμε ξεκάθαρα να διακρίνουμε πως το σχήμα του προσώπου και η σκίαση επηρεάζονται ταυτόχρονα.

Τα Μοντέλα Ενεργής Εμφάνισης – Active Appearance Models (AAMs) των Cootes, Edwards και Taylor (2001) μοντελοποιούν τόσο τις παραλλαγές στο σχήμα μιας απεικόνισης s , που κανονικά κωδικοποιείται από την τοποθεσία των γνωρισμάτων κλειδιών σε μια εικόνα, όσο και από τις παραλλαγές στην υφή t , η οποία ομαλοποιείται σε κανονικό σχήμα πριν αναλυθεί. Το σχήμα και η υφή απεικονίζονται ως διακυμάνσεις των μέσων τιμών σχήματος \bar{s} και υφών \bar{t} ,

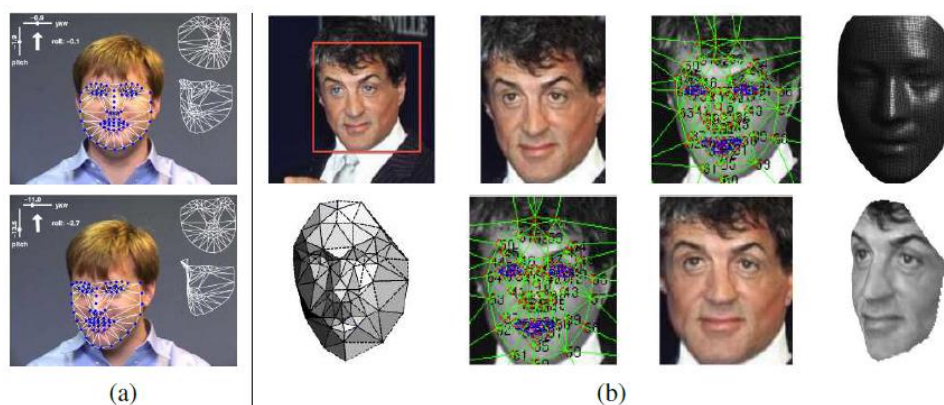
$$\begin{aligned} s &= \bar{s} + U_s a \\ t &= \bar{t} + U_t a, \end{aligned}$$

Όπου τα διανύσματα $eigen$ U_s και U_t έχουν ομαλοποιηθεί (λευκανθεί) ώστε όλα τα διανύσματα στο a να αναπαριστούν μια συγκεκριμένη απόκλιση των μεταβολών που παρατηρείται στα δεδομένα εκπαίδευσης. Επιπρόσθετα αυτών των βασικών παραμορφώσεων, οι παράμετροι των σχημάτων μεταλλάσσονται με καθολική ομοιότητα ώστε να αντιστοιχισθούν με τη θέση, το μέγεθος και τον προσανατολισμό ενός καθορισμένου προσώπου. Ομοίως, η απεικόνιση της υφής περιέχει κλίμακα και αντιστάθμιση για να ταιριάζει καλύτερα σε νέες συνθήκες φωτισμού.

Όπως παρατηρούμε, οι ίδιες παράμετροι εμφάνισης a στις παραπάνω μαθηματικές εξισώσεις ταυτόχρονα ελέγχουν τις παραμορφώσεις σε σχήμα καθώς και υφή από το μέσο όρο, το οποίο έχει νόημα εφόσον θεωρούμε ότι σχετίζονται. Η Εικόνα 10 δείχνει πως τρεις κινούμενες στάνταρ αποκλίσεις σε καθεμία από τις τέσσερις στάνταρ διαστάσεις μας οδηγεί στο να αλλάζουν πολλοί σχετιζόμενοι παράγοντες στην εμφάνιση ενός προσώπου, συμπεριλαμβανομένης της έκφρασης, του φύλου, της ηλικίας και της ταυτότητας.

Παρόλο που τα Μοντέλα Ενεργής Εμφάνισης έχουν κυρίως σχεδιαστεί για να αιχμαλωτίζουν με ακρίβεια τις διαφορές στην εμφάνιση και την παραμόρφωση που είναι

χαρακτηριστικά των προσώπων, μπορούν να προσαρμοστούν στην αναγνώριση προσώπων υπολογίζοντας μια «υποενότητα» ταυτότητας που διαχωρίζει τις διαφορές στην ταυτότητα από άλλες πηγές διαφορετικότητας όπως το φωτισμό, την πόζα και την έκφραση (Costen, Cootes et al. 1999). Η βασική ιδέα, η οποία έχει μοντελοποιηθεί μετά από παρόμοια δουλειά στα eigenfaces (Belhumeur, Hespanha, και Kriegman 1997; Moghaddam, Jebara, και Pentland 2000), είναι να υπολογίζονται ξεχωριστά οι μεταβλητές των αποκλίσεων σε ένα πρόσωπο καθώς και σε διαφορετικά πρόσωπα και μετά να διακρίνονται οι διαφορές τους. Ενώ τα MEE έχουν μερικές φορές χρησιμοποιηθεί απευθείας για αναγνώριση (Banz και Vetter 2003), η κύρια χρήση τους στο πεδίο τη αναγνώρισης είναι η ευθυγράμμιση των προσώπων σε μια «κανονική» πόζα (Liang, Xiao et al. 2008; Ren, Cao et al. 2014) και έπειτα να χρησιμοποιούνται πιο παραδοσιακές μέθοδοι αναγνώρισης προσώπου (Penev και Atick 1996; Wiskott, Fellous et al. 1997; Ahonen, Hadid, και Pietikainen 2006; Zhao και Pietikainen 2007; Cao, Yin et al. 2010).



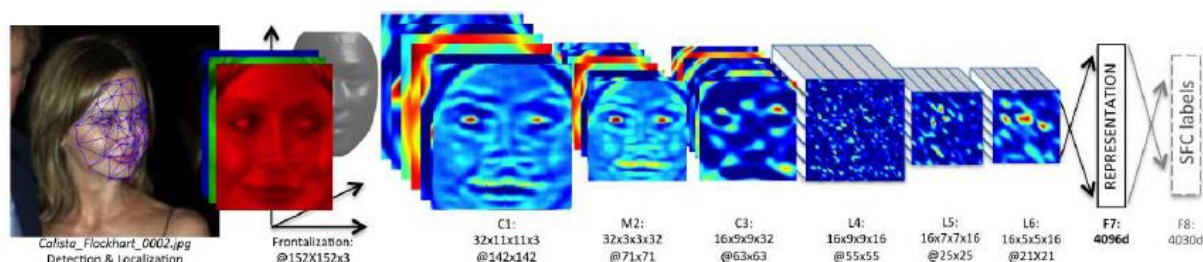
Εικόνα 11: Ιχνηλάτηση προσώπου και frontalization: (a) χρησιμοποιώντας 3-Διάστατα Μοντέλα Ενεργής Εμφάνισης, απεικονίζοντας video frames με την εκτιμώμενες παραμέτρους κλίσης και περιστροφής και το 3-Διάστατο παραμορφωμένο πλέγμα (b) χρησιμοποιώντας 6 και μετά 67 σημεία εμπιστοσύνης στο σύστημα DeepFace, χρησιμοποιούμενα για το frontalization του προσώπου (κάτω σειρά).

Η Ενεργή Εμφάνιση έχει επεκταθεί ώστε να χειρίζεται τον φωτισμό και τις μεταβολές της οπτικής γωνίας (Gross, Baker et al. 2005) όπως και την οδοντική σύγκλιση (Gross, Matthews, και Baker 2006). Μια από τις πιο σημαντικές επεκτάσεις είναι η κατασκευή 3-Διάστατων

μοντέλων σχημάτων (Matthews, Xiao, και Baker 2007), τα οποία είναι πολύ καλύτερα στο να αιχμαλωτίζουν και να εξηγούν όλο το εύρος της μεταβλητότητας των εκφράσεων του προσώπου καθώς αλλάζει η οπτική γωνία. Τέτοια μοντέλα μπορούν να κατασκευαστούν είτε από μονοφθαλμικές ακολουθίες βίντεο (Matthews, Xiao, και Baker 2007) ή από ακολουθίες βίντεο που περιλαμβάνουν διαφορετικές οπτικές γωνίες (Ramnath, Koterba et al. 2008), που παρέχουν μεγαλύτερη ακρίβεια και αξιοπιστία στην αναπαράσταση και την ιχνηλάτηση (Murphy-Chutorian και Trivedi 2009).

2.2.2 Μάθηση σε Βάθος

Παρακινήμενοι από τη δραματική επιτυχία των δικτύων μάθησης σε βάθος στην κατηγοριοποίηση ευρείας κλίμακας εικόνων, οι ερευνητές της αναγνώρισης προσώπου ξεκίνησαν να χρησιμοποιούν ραχοκοκαλιά Βαθέων Νευρωνικών Δικτύων στα συστήματά τους. Οι Εικόνες 11 και 12 δείχνουν δυο σκέλη στο σύστημα DeepFace των Taigman, Yang et al. (2014), το οποίο ήταν ένα από τα πρώτα συστήματα που έδειξε τεράστια πρόοδο χρησιμοποιώντας δίκτυα μαθήσεως σε βάθος. Στο σύστημά τους, χρησιμοποιείται ένα σύστημα frontalization βασισμένο σε σημεία αναφοράς το οποίο εφαρμόζεται και μετατρέπει την αρχική εικόνα σε ένα ωραίο, «περικομμένο» πρόσωπο που κοιτάει μπροστά. Έπειτα, ένα συνδεδεμένο δίκτυο μάθησης σε βάθος (όπου οι πυρήνες συσπείρωσης – convolution kernels μπορούν να διαφοροποιούνται χωροταξικά) τροφοδοτείται σε δυο τελικά πλήρως συνδεδεμένα στρώματα πριν την κατηγοριοποίηση.



Εικόνα 12: Η αρχιτεκτονική της DeepFace (Taigman, Yang et al. 2014) © 2014 IEEE, ξεκινά με ένα στάδιο frontalization, ακολουθούμενο από μερικά μη-συνεστραμμένα φίλτρα και έπειτα από δυο πλήρως συνδεδεμένα επίπεδα, με softmax K-κλάσεων.

Μερικά από τα πιο πρόσφατα λογισμικά αναγνώρισης προσώπου με δίκτυα μάθησης σε βάθος παραλείπουν το στάδιο του Frontalization και χρησιμοποιούν ενίσχυση δεδομένων –

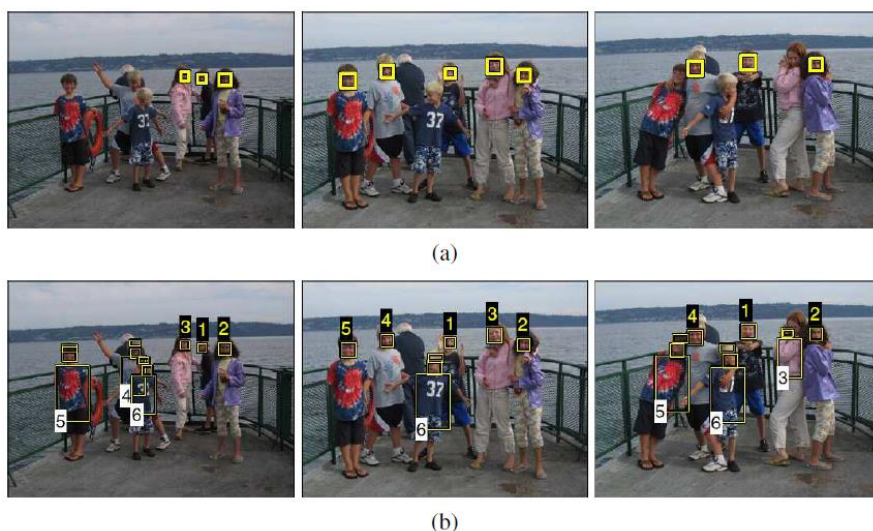
data augmentation για να δημιουργήσουν σύνθετες καταχωρήσεις με μια μεγαλύτερη ποικιλία οπτικών γωνιών – πόζας (Schroff, Kalenichenko, και Philbin 2015; Parkhi, Vedaldi, και Zisserman 2015). Οι Masi, Wu et al. (2018) παρέχουν εξαιρετικό εκπαιδευτικό υλικό από την έρευνά τους στην αναγνώριση προσώπου με δίκτυα μάθησης σε βάθος, περιλαμβάνοντας μια λίστα ευρέως χρησιμοποιούμενων σετ δεδομένων εκμάθησης και δοκιμών, μια συζήτηση για το Frontalization και τα ενισχυμένα σετ δεδομένων και μια ενότητα για τις απώλειες από τα δεδομένα εκπαίδευσης. Οι Schroff, Kalenichenko, Philbin (2015) και οι Parkhi, Vedaldi, Zisserman (2015) χρησιμοποιούν τρίδυμες απώλειες – triplet losses για να κατασκευάσουν ένα χώρο μικρών διαστάσεων που είναι ανεξάρτητος με τον αριθμό των θεμάτων. Πιο πρόσφατα συστήματα χρησιμοποιούν απώλειες αντικρουόμενων – contrastive losses εμπνεόμενες από τη συνάρτηση Softmax. Για παράδειγμα, στο paper ArcFace από Deng, Guo et al. (2019) μετρώνται γωνιακές αποστάσεις στη σφαίρα ενός αντικειμένου και προστίθεται ένα πλαίσιο ώστε να συσταδοποιηθούν οι ομοιότητες. Αυτή η ιδέα έχει επεκταθεί και στην έρευνα αναζήτησης οπτικής ομοιότητας (Bell, Liu et al. 2020) και στην αναγνώριση προσώπου (Huang, Shen et al. 2020; Deng, Guo et al. 2020a).

Προσωπικές συλλογές φωτογραφιών

Επιπρόσθετα της δυνατότητας των ψηφιακών φωτογραφικών μηχανών τελευταίας τεχνολογίας στο να βρίσκουν πρόσωπα ώστε να υποβοηθείται η αυτόματη εστίαση, και των βιντοκαμερών να βρίσκουν πρόσωπα στις τηλεδιασκέψεις ώστε να «κεντράρουν» στο πρόσωπο του συνομιλητή (είτε μηχανικά ή ψηφιακά), η ανίχνευση προσώπου έχει βρει εφαρμογή σε εμπορικά πακέτα και ιστότοπους διαμοιρασμού φωτογραφιών. Βρίσκοντας πρόσωπα και υποβοηθώντας το χρήστη να χαρακτηρίσει (tag) εκείνον ή κάποιο οικείο πρόσωπο, βοηθάει αργότερα στην αναζήτηση ή τον αυτόματο διαμοιρασμό με την οικογένεια ή τα φιλικά πρόσωπα. Είναι γεγονός πως οι ετικέτες σε φιλικά πρόσωπα (friend tagging) σε φωτογραφίες, είναι μια από τις πιο δημοφιλείς λειτουργίες του Facebook.

Μερικές φορές πάντως, η ανίχνευση και αναγνώριση ενός προσώπου είναι δύσκολη, ειδικά αν το πρόσωπο είναι μακριά από την κάμερα, σε κακή οπτική γωνία, ο φωτισμός χαμηλός ή υπάρχουν άλλα εμπόδια που αποκόπτουν τμήμα του προσώπου. Σε αυτές τις περιπτώσεις, συνδυάζοντας την αναγνώριση προσώπου με την ανίχνευση ατόμου καθώς ακόμα και η αναγνώριση των ρούχων μπορεί να γίνουν πολύ αποδοτικές, όπως φαίνεται στην Εικόνα

13 (Sivic, Zitnick, και Szeliski 2006). Συνδυάζοντας την αναγνώριση ατόμων με άλλου τύπου περιεχόμενο, όπως αναγνώριση τοποθεσίας, δραστηριότητας ή εκδήλωσης (λαμβάνοντας υπόψη τη στάση του σώματος, το είδος του ρουχισμού κ.α.), μπορεί να βοηθήσει στην ενίσχυση της απόδοσης (Lin, Karoor et al. 2010).



Εικόνα 13: Ανίχνευση ατόμων και αναγνώριση χρησιμοποιώντας ένα συνδυασμένο μοντέλο προσώπου, μαλλιών και στέρνου (Sivic, Zitnick, και Szeliski 2006) © 2006 Springer. (a) Χρησιμοποιώντας μόνο ανίχνευση προσώπου, πολλά κεφάλια δεν ανιχνεύονται. (b) Το συνδυασμένο μοντέλο προσώπου και ρούχων ανιχνεύει επιτυχώς όλα τα μέλη στη φωτογραφία.

2.2.3 Αναγνώριση αντικειμένων

Εάν μας δινόταν μια εικόνα προς ανάλυση, όπως το ομαδικό πορτραίτο της Εικόνας 13, θα μπορούσαμε να προσπαθήσουμε να εφαρμόσουμε έναν αλγόριθμο αναγνώρισης σε κάθε πιθανό τμήμα αυτής. Τέτοιοι αλγόριθμοι είναι επιρρεπείς σε λάθη όπως και σε πολύ αργή ταχύτητα. Αντ' αυτού, είναι πολύ πιο αποτελεσματικό να κατασκευάσουμε ανιχνευτές ειδικής χρήσης, που η κύρια «δουλειά» τους είναι να βρουν πιθανές περιοχές της εικόνας, που υπάρχουν συγκεκριμένα αντικείμενα ή συνδυασμός αντικειμένων.

Η απαρχή των λογισμικών αναγνώρισης προσώπου, ήταν τα λογισμικά ανίχνευσης. Τέτοιοι αλγόριθμοι έχουν εφαρμοστεί στα σημερινά εμπορικά προϊόντα φωτογραφίας και βιντεοσκόπησης για να ενισχύσουν την αυτόματη εστίαση (autofocus). Σε πιο σύνθετες εφαρμογές αναγνώρισης όπου απαιτείται ανίχνευση πολλών διαφορετικών στοιχείων, όπως αναφέραμε παραπάνω, χρησιμοποιούνται πολυεπίπεδα νευρωνικά δίκτυα μάθησης σε βάθος.

3. Python

Η Python είναι μια γλώσσα προγραμματισμού σχεδιασμένη για λειτουργικότητα και αναγνωσιμότητα. Ένας από τους στόχους του σχεδιασμού της Python είναι ότι το νόημα του κώδικα είναι εύκολα κατανοητό λόγω της πολύ «καθαρής» σύνταξης της γλώσσας. Η γλώσσα Python έχει πολύ συγκεκριμένη σύνταξη και σημασία, που της επιτρέπει να εκφράζει υπολογισμούς και χειρισμούς δεδομένων που μπορούν να πραγματοποιηθούν από τον υπολογιστή.

Η υλοποίηση της Python ξεκίνησε το 1989 από τον Guido van Rossum στο CWI (Εθνικό Ινστιτούτο Ερευνών στην Ολλανδία) σαν διάδοχος της γλώσσας προγραμματισμού ABC (μια ασαφής γλώσσα η οποία έγινε περισσότερο διάσημη από το γεγονός ότι έδωσε κίνητρο για τη δημιουργία της Python!). Ο Van Rossum είναι ο βασικός δημιουργός της Python, και ο συνεχιζόμενος κεντρικός του ρόλος στις αποφάσεις για την πορεία της Python αντικατοπτρίζεται στον τίτλο που του έχει δοθεί από την κοινωνία προγραμματιστών Python, «Καλοπροαίρετος Δικτάτορας για μια Ζωή» (Benevolent Dictator for Life, BDFL).

Η Python είναι μια ερμηνευμένη γλώσσα, δηλ. μια γλώσσα προγραμματισμού όπου τα προγράμματα δεν εκτελούνται ευθέως από την επεξεργαστή του εξυπηρετητή (host) αλλά αντιθέτως εκτελούνται (ή «μεταφράζονται») από ένα πρόγραμμα γνωστό ως «μεταφραστής». Ο πηγαίος κώδικας ενός προγράμματος Python μεταφράζεται ή μεταγλωττίζεται τμηματικά σε μορφή byte από ένα μηχανισμό γλώσσας της Python “Process Virtual Machine”. Αυτός είναι και ο κύριος διαχωρισμός με τη γλώσσα προγραμματισμού “C” όπου ο κώδικας μεταφράζεται σε bytes στον επεξεργαστή του εξυπηρετητή πριν εκτελεστεί.

Η Python είναι μια «δυναμικά συντασσόμενη» γλώσσα προγραμματισμού. Μια γλώσσα αναφέρεται σαν «δυναμικά συντασσόμενη» όταν η πλειοψηφία των ελέγχων γίνεται όταν εκτελείται και όχι όταν μεταγλωττίζεται. Υπάρχουν πολλές δυναμικά συντασσόμενες γλώσσες όπως η JavaScript, Lisp, Lua, Objective-C, Python, Ruby και Tcl.

Τα δεδομένα που χρησιμοποιεί ένα πρόγραμμα της Python για να εκτελέσει διάφορες λειτουργίες πρέπει να περιγράφονται με ακριβή τρόπο. Η περιγραφή αυτή αναφέρεται σαν

«τύπος μεταβλητής». Στην περίπτωση της Python, το γεγονός ότι είναι δυναμικά συντασσόμενη βασικά σημαίνει ότι το μεταγλωττιστής θα καταλάβει για μας τον τύπο της μεταβλητής στην εκτέλεση του προγράμματος, ώστε να μη χρειάζεται να δηλώσουμε εμείς τις μεταβλητές εξαρχής. Το γεγονός ότι η Python είναι δυναμικά συντασσόμενη σημαίνει ότι θα δείξει σφάλμα όταν έχουμε παραβιάσει κάποιο κανόνα γραμματικής/σύνταξης χρησιμοποιώντας μαζί τις μεταβλητές σε μια συνάρτηση. Φυσικά αυτό δε σημαίνει ότι μπορούμε να παραμελήσουμε την πειθαρχία σύνταξης, δηλαδή να αναμειγνύουμε τύπους μεταβλητών ελπίζοντας ότι η Python θα το διορθώσει.

Επίσης, η Python είναι μια γλώσσα αντικειμενοστραφούς προγραμματισμού. Ο αντικειμενοστραφής προγραμματισμός χρησιμοποιεί «αντικείμενα» - δομές δεδομένων που αποτελούνται από πεδία δεδομένων και μεθόδους – για το σχεδιασμό προγραμμάτων υπολογιστών. Για παράδειγμα, ένας πίνακας θα μπορούσε να είναι το «αντικείμενο» όπου θα επεξεργάζεται το πρόγραμμα. Θα μπορούσαμε να ορίζουμε μια κλάση πινάκων και η λειτουργία να ήταν μια πρόσθεση (αναπαριστώντας την κανονική πρόσθεση πινάκων). Ένα από τα βασικά πλεονεκτήματα της γλώσσας προγραμματισμού Python είναι ότι έχει δοθεί έμφαση στην εύκολη σύνταξη και ανάγνωση του κώδικα, γνώρισμα το οποίο επιτρέπει στον προγραμματιστή να χρησιμοποιεί λιγότερες γραμμές για να «στήσει» το λογισμικό ή μια υπορουτίνα, συγκρινόμενη με άλλες γλώσσες προγραμματισμού, όπως η Java και η C++.



Εικόνα 14: Λογότυπα της Python και OpenCV

3.1 Η βιβλιοθήκη OpenCV

Η OpenCV βιβλιοθήκη (Open Source Computer Vision Library), είναι μια βιβλιοθήκη λειτουργιών προγραμματισμού, που στοχεύει κυρίως στην μηχανική όραση πραγματικού χρόνου (real-time) και πιο ειδικά στην επεξεργασία βίντεο ή στατικής εικόνας που προκύπτει

από τις συσκευές σύλληψης ψηφιακής εικόνας (πχ. βιντεοκάμερες). Αρχικά, αναπτύχθηκε από την Intel και αργότερα υποστηρίχθηκε από την Willow Garage (μετέπειτα Itseez) η οποία τελικά αγοράστηκε από την Intel. Η βιβλιοθήκη είναι λογισμικό Ανοικτού Κώδικα και ελεύθερη προς χρήση από όλους, υπό την άδεια ανοικτού κώδικα Apache 2.

Επίσημα λανσαρισμένο το 1999, το ερευνητικό εγχείρημα OpenCV ήταν αρχικά μια πρωτοβουλία της Intel Research για την ανάπτυξη «δύσκολων» υπολογιστικών εφαρμογών, μαζί με άλλα εγχειρήματα όπως τις απεικονίσεις 3D. Το πρώτο διάστημα ανάπτυξης της OpenCV, οι στόχοι που ορίστηκαν περιγράφονταν ως:

- Προηγμένη έρευνα στην υπολογιστική όραση, παρέχοντας όχι μόνο λογισμικό ανοικτού κώδικα, αλλά μια προηγμένη μορφή του, για την εξέλιξη του ήδη υπάρχοντος εξοπλισμού
- Διασπορά της γνώσης, παρέχοντας κοινή υποδομή όπου όλοι οι προγραμματιστές θα μπορούσαν να βασιστούν, με στόχο ο κώδικας να είναι άμεσα διαθέσιμος και διαμοιράσιμος ανά πάσα στιγμή
- Προηγμένες εφαρμογές μηχανικής όρασης που θα ήταν διαθέσιμες στο ευρύ κοινό, κάνοντας τον εξελιγμένο κώδικα διαθέσιμο δωρεάν – κάτω από μια άδεια που δεν απαιτεί περαιτέρω δικαιώματα ή χρηματικό αντίτιμο.

Η πρώτη δοκιμαστική έκδοση της OpenCV έγινε διαθέσιμη στο ευρύ κοινό στο “IEEE Conference on Computer Vision and Pattern Recognition” το 2000, και πέντε εκδόσεις κυκλοφόρησαν ανάμεσα στα έτη 2001 και 2005. Η πρώτη 1.0 (ολοκληρωμένη) έκδοση κυκλοφόρησε το 2006, καθώς και η διάδοχος 1.1 κυκλοφόρησε τον Οκτώβρη του 2008. Η δεύτερη μεγάλη βελτίωση της OpenCV κυκλοφόρησε τον Οκτώβριο του 2009. Η OpenCV2 περιέχει μεγάλες βελτιώσεις στη διεπαφή C++, στοχεύοντας σε πιο εύκολα, πιο ασφαλή μοτίβα πληκτρολόγησης, νέες λειτουργίες καθώς και καλύτερη εφαρμογή των υπάρχοντων, σε επίπεδο απόδοσης. Σήμερα, επίσημες εκδόσεις κυκλοφορούν περίπου ανά 6 μήνες και η ανάπτυξη του κώδικα γίνεται από μια ανεξάρτητη Ρωσική ομάδα προγραμματιστών υποστηριζόμενων από κάποιες εταιρίες του τομέα. Τον Αύγουστο του 2012, η υποστήριξη της OpenCV πέρασε στα χέρια ενός μη-κερδοσκοπικού οργανισμού, του OpenCV.org, που διατηρεί έναν ιστότοπο προγραμματιστών καθώς και έναν για τους απλούς χρήστες.

Τον Ιούλιο του 2020, η OpenCV ανακοίνωσε και ξεκίνησε μια καμπάνια στο KickStarter⁴ για το OpenCV AI Kit, μια σειρά hardware και προσθηκών σε συσκευές Χωρικής Τεχνητής Νοημοσύνης (Spatial AI) που υποστηρίζουν OpenCV.

Η βιβλιοθήκη OpenCV περιέχει και περιλαμβάνει μια σειρά αλγορίθμων μηχανικής μάθησης και όρασης. Οι αλγόριθμοι μπορούν να βοηθήσουν τον προγραμματιστή να αναπτύξει κώδικα για ανίχνευση αντικειμένων ή προσώπων και την αναγνώριση αυτών, για την αναζήτηση όμοιων (ή παρόμοιων) εικόνων σε ένα σετ δεδομένων, για την παρακολούθηση της κίνησης αντικειμένων σε κινούμενη εικόνα, για την αναγνώριση και ψηφιοποίηση χειρονομιών ή τρισδιάστατων μοντέλων αντικειμένων, την αναγνώριση βάθους χρησιμοποιώντας 2 κάμερες (στερεοσκοπική όραση), και ό,τι έχει να κάνει με την επαυξημένη πραγματικότητα – augmented reality. Οι εφαρμογές αυτές μπορούν να χρησιμοποιηθούν σε ευρύ φάσμα στρατιωτικών, βιομηχανικών και ιατρικών εφαρμογών καθώς και συστημάτων ασφαλείας. Ένα από τα πιο σημαντικά πλεονεκτήματά της είναι η ταχύτητα, αφού είναι δυνατό να επεξεργάζεται βίντεο real time, αυξάνοντας την αποτελεσματικότητα μειώνοντας ταυτόχρονα το κόστος (εφόσον διατίθεται δωρεάν).

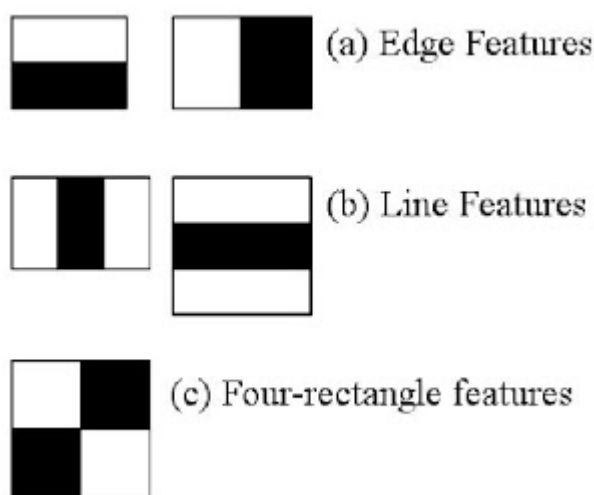
Για να υποστηρίξει μερικές από τις παραπάνω εφαρμογές, η βιβλιοθήκη OpenCV περιλαμβάνει μια βιβλιοθήκη στατιστικής μηχανικής μάθησης που εμπεριέχει ενίσχυση – boosting, μάθηση με Δένδρα Αποφάσεων, αλγόριθμο K-κοντινότερων γειτόνων, πιθανολογικούς ταξινομητές Bayes, τεχνητά νευρωνικά δίκτυα και νευρωνικά δίκτυα μάθησης σε βάθος. Η OpenCV, όταν ξεκίνησε, αναπτύχθηκε με γλώσσα προγραμματισμού C++. Τα τελευταία χρόνια, έχει αναπτυχθεί και με άλλες γλώσσες, όπως η Python και η Java και έτοιμα λογισμικά όπως το Matlab, και υποστηρίζεται από λειτουργικά συστήματα όπως Windows, Mac OS και Linux.

⁴ Η Kickstarter είναι μια εταιρεία κοινής ωφέλειας που εδρεύει στο Μπρούκλιν της Νέας Υόρκης, η οποία διατηρεί μια παγκόσμια πλατφόρμα crowdfunding (χρηματοδότησης από ιδιώτες) επικεντρωμένη στη δημιουργικότητα. Η δηλωμένη αποστολή της εταιρείας είναι να "βοηθήσει στη δημιουργία δημιουργικών έργων στη ζωή". Πηγή: [Wikipedia.org](https://www.kickstarter.com/)

3.2 Ανίχνευση με Haar-Cascade ταξινομητές

Η αναγνώριση αντικειμένων χρησιμοποιώντας ταξινομητές που βασίζονται σε Haar Cascade χαρακτηριστικά, είναι μία αξιόπιστη μέθοδος αναγνώρισης που προτάθηκε από τους Paul Viola και Michael Jones στη δημοσίευσή τους «Γρήγορη αναγνώριση αντικειμένων χρησιμοποιώντας ενισχυμένα χαρακτηριστικά αλληλοεπικάλυψης, Rapid Object Detection using a Boosted Cascade of Simple Features» το 2001. Είναι μια προσέγγιση μηχανικής μάθησης που η λειτουργία της εκπαιδεύεται από πολλές βοηθητικές και μη-βοηθητικές εικόνες. Έπειτα χρησιμοποιείται για να ανιχνεύει αντικείμενα σε άλλες εικόνες.

Παρακάτω, θα δώσουμε ένα παράδειγμα ανίχνευσης προσώπου. Αρχικά, ο αλγόριθμος χρειάζεται πολλές βοηθητικές (εικόνες με πρόσωπα) και μη-βοηθητικές (εικόνες που δεν περιέχουν πρόσωπα) για την εκπαίδευση του ταξινομητή. Έπειτα χρειάζεται να εξάγουμε γνωρίσματα από αυτές. Για να επιτευχθεί αυτό, χρησιμοποιούνται Haar γνωρίσματα που φαίνονται στην παρακάτω εικόνα και λειτουργούν όπως ο συσπειρωτικός πυρήνας. Κάθε γνώρισμα είναι μια μοναδική τιμή που αποκτάται αφαιρώντας αθροίσματα εικονοστοιχείων που δημιουργούν λευκά ορθογώνια, από αθροίσματα εικονοστοιχείων μαύρων τετραγώνων.

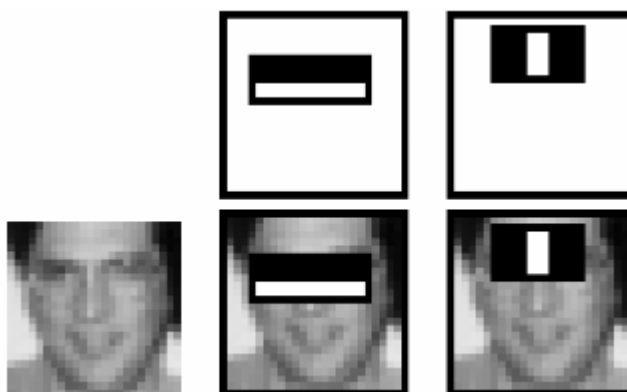


Εικόνα 15: Γνωρίσματα γωνιών, γραμμών και ορθογώνιων

Τώρα όλα τα πιθανά μεγέθη και περιοχές κάθε πυρήνα χρησιμοποιούνται για να τα υπολογίσουμε πολλά διαφορετικά γνωρίσματα. Φανταστείτε την υπολογιστική ισχύ που χρειάζεται για κάτι τέτοιο. Ακόμα και ένα παράθυρο 24x24 εικονοστοιχείων βγάζει αποτέλεσμα πάνω από 160000 γνωρίσματα. Για κάθε υπολογισμό γνωρίσματος, χρειάζεται να βρούμε ένα

άθροισμα εικονοστοιχείων λευκών και μαύρων ορθογωνίων. Για να επιλυθεί αυτό, παρουσίασαν τις εικόνες καίριας σημασίας. Αυτό απλοποιεί τον υπολογισμό αθροισμάτων των εικονοστοιχείων, πόσο μεγάλο πρέπει να είναι το πλήθος των εικονοστοιχείων, σε μια λειτουργία που περιλαμβάνει μόνο τέσσερα εικονοστοιχεία. Το αποτέλεσμα? Ο αλγόριθμος γίνεται πολύ γρήγορος.

Βέβαια, ανάμεσα σε όλα αυτά τα γνωρίσματα που υπολογίσαμε, τα περισσότερα είναι άσχετα με το αποτέλεσμα που θέλουμε να επιτύχουμε. Για παράδειγμα, δείτε την παρακάτω Εικόνα 16. Η πάνω γραμμή δείχνει δυο καλά γνωρίσματα. Το πρώτο γνώρισμα που έχει επιλεγεί φαίνεται να εστιάζει στην ιδιότητα ότι η περιοχή των ματιών είναι συνήθως σκουρότερη από την περιοχή της μύτης και των μάγουλων. Το δεύτερο γνώρισμα που έχει επιλεγεί βασίζεται στην ιδιότητα ότι τα μάτια είναι σκουρότερα από τη βάση της μύτης. Τα ίδια γνωρίσματα αν εφαρμοστούν στα μάγουλα ή σε άλλα σημεία της εικόνας είναι άσχετα. Πώς επιλέγουμε τα καλύτερα γνωρίσματα ανάμεσα σε περίπου 160000+? Επιτυγχάνεται με τον **Adaboost**.



Εικόνα 16: Απλά γνωρίσματα που χρησιμοποιούνται σε ανιχνευτή προσώπου που βασίζεται στην ενίσχυση (Viola και Jones 2004) © 2004 Springer. Το πρώτο και δεύτερο γνώρισμα επιλέγεται από τον Adaboost. Το πρώτο γνώρισμα μετράει τις διαφορές στην ένταση των εικονοστοιχείων μεταξύ των ματιών και των μάγουλων, το δεύτερο μεταξύ των ματιών και της γέφυρας της μύτης.

Για να επιτευχθεί αυτό, εφαρμόζουμε κάθε γνώρισμα σε όλες τις εικόνες εκπαίδευσης. Για κάθε χαρακτηριστικό, ο αλγόριθμος βρίσκει το καλύτερο όριο για να κατηγοριοποιήσει τις εικόνες σύμφωνα με το αν περιέχουν πρόσωπα ή όχι. Προφανώς όμως θα υπάρξουν λάθη. Επιλέγουμε τα γνωρίσματα με το μικρότερο δείκτη λαθών, το οποίο σημαίνει ότι αυτά τα

γνωρίσματα κατατάσσουν τις εικόνες με τον καλύτερο δυνατό τρόπο. Η διαδικασία αυτή δεν είναι τόσο απλή, καθώς στην αρχή δίνεται σε κάθε εικόνα ίδιο «βάρος». Μετά την πρώτη κατάταξη, τα βάρη των εικόνων που έχουν καταταχθεί λάθος αυξάνονται. Κατόπιν η ίδια διαδικασία εκτελείται, όπου υπολογίζονται ξανά οι δείκτες και τα βάρη των εικόνων. Η διαδικασία συνεχίζεται μέχρι να επιτευχθεί ακρίβεια στην αναγνώριση των συγκεκριμένων γνωρισμάτων.

Ο τελικός ταξινομητής είναι ένα άθροισμα αυτών των «αδύναμων» ταξινομητών. Καλούνται αδύναμοι καθώς μόνοι τους δεν μπορούν να ταξινομήσουν τις εικόνες, με συνδυασμό τους όμως επιτυγχάνεται ένας δυνατός ταξινομητής. Η δημοσίευση των προαναφερθέντων Viola & Jones αναφέρει ότι αλγόριθμος με ακόμα και 200 γνωρίσματα, παρέχει ακρίβεια ακόμα και 95% στην αναγνώριση. Το τελικό στήσιμο του αλγόριθμου περιείχε περίπου 6000 γνωρίσματα (φανταστείτε μια μείωση από 1.600.000+ γνωρίσματα σε 6000. Είναι πολύ μεγάλο κέρδος).

Ας υποθέσουμε ότι έχουμε μια εικόνα και λαμβάνουμε περιοχές 24x24 εικονοστοιχείων. Σε αυτές εφαρμόζουμε 6000 γνωρίσματα και προσπαθούμε να αναγνωρίσουμε κάποιο πρόσωπο. Αυτό φυσικά είναι λίγο αναποτελεσματικό και χρονοβόρο, όμως υπάρχει και άλλη λύση σε αυτό το πρόβλημα.

Σε μια εικόνα, η μεγαλύτερη περιοχή της δεν περιέχει κάποιο πρόσωπο. Θα μπορούσαμε να εφαρμόσουμε έναν αλγόριθμο που θα ελέγχει μια φορά αν η περιοχή περιέχει πρόσωπο, και αν όχι, να την αποκλείσουμε την πρώτη φορά ελέγχου ώστε να μη χρειαστεί να την προσπελάσουμε ξανά. Αυτό μας επιτρέπει να εστιάσουμε στην περιοχή που θα μπορούσε να υπάρχει κάποιο πρόσωπο και να χρησιμοποιήσουμε τους πόρους συστήματος και το χρόνο που έχουμε κερδίσει για να ελέγξουμε πιο λεπτομερώς αυτή την περιοχή.

Για την επίτευξη αυτού παρουσιάστηκε η ιδέα του καταρράκτη Ταξινομητών (Cascade of Classifiers). Αντί να εφαρμόσουμε όλα τα 6000 γνωρίσματα σε ένα παράθυρο, ομαδοποιούμε τα γνωρίσματα σε διαφορετικά στάδια ταξινομητών και τα εφαρμόζουμε ένα προς ένα (Τυπικά τα πρώτα στάδια περιέχουν πολύ μικρότερο αριθμό γνωρισμάτων). Εάν ένα παράθυρο αποτύχει το πρώτο στάδιο, το απορρίπτουμε και δεν αξιολογούμε ούτε εφαρμόζουμε

επιπλέον χαρακτηριστικά σε αυτό. Εάν «περάσει», εφαρμόζουμε το δεύτερο στάδιο γνωρισμάτων και συνεχίζουμε τη διαδικασία. Όποιο παράθυρο περάσει όλα τα στάδια είναι μια περιοχή προσώπου.

Ο ανιχνευτής των συγγραφέων είχε 6000+ γνωρίσματα σε 38 στάδια με 1, 10, 25 και 50 γνωρίσματα στα πέντε πρώτα στάδια (τα δυο γνωρίσματα στην παραπάνω εικόνα εκφράζονται σαν τα καλύτερα δυο γνωρίσματα του Adaboost). Σύμφωνα με τους συγγραφείς, κατά μέσο όρο, 10 γνωρίσματα από τα 6000+ αξιολογούνται σε κάθε υποπαράθυρο.

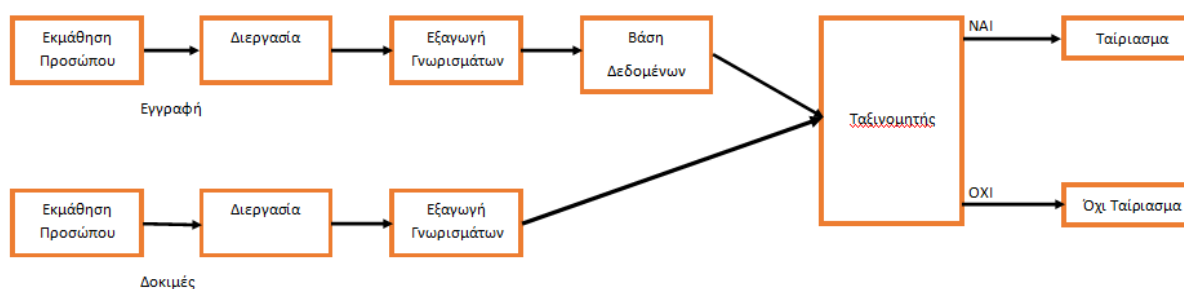
3.3 LBPH Ταξινομητής

Ο αλγόριθμος Τοπικών Δυαδικών Μοντέλων – Local Binary Patterns Histogram (LBPH) προτάθηκε το 2006. Είναι βασισμένος σε έναν τοπικό δυαδικό χειριστή. Χρησιμοποιείται ευρέως στην αναγνώριση προσώπου, λόγω της υπολογιστικής απλότητας του και της δύναμής του στο να βρίσκει τις διαφορές.

Τα βήματα που είναι απαραίτητα για να επιτευχθεί αυτό, είναι:

- Η δημιουργία σετ δεδομένων
- Η εκμάθηση ενός προσώπου
- Η εξαγωγή γνωρισμάτων
- Η Ταξινόμηση

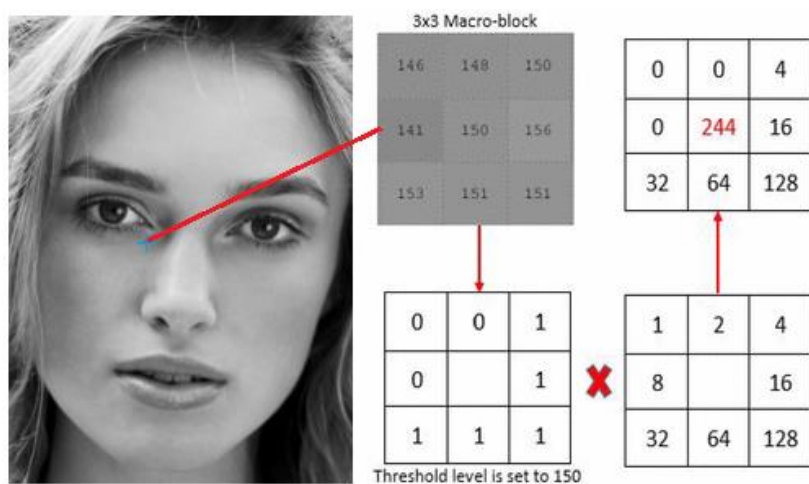
Ο LBPH αλγόριθμος είναι κομμάτι της βιβλιοθήκης OpenCV.



3.3.1 Περιγραφή – Λειτουργία του LBPH

Από την ανακάλυψη του, το LBPH παραμένει ένας εξαιρετικά αποτελεσματικός αλγόριθμος για την ταξινόμηση γνωστών υφών, όπως ένα πρόσωπο. Απαιτεί τέσσερις

ξεχωριστές παραμέτρους για να επεξεργαστεί μια εικόνα, τις εξής: Ακτίνα – Radius (r), Γείτονες – neighbors (n), τον X-άξονα και τον Y-άξονα. Η απεικόνιση της (r) σε σχέση με τους (n) δίνεται στην Εικόνα 17, όπου η επισημασμένη περιοχή είναι ένα 3×3 μακροτετράγωνο - macro-block. Εδώ οι X και Y άξονες αναπαριστούν τις διαστάσεις του πλέγματος γνωρισμάτων, σε κάθετο και οριζόντιο άξονα. Το πρώτο βήμα είναι να εκπαιδύσουμε τον αλγόριθμο και για να το πετύχουμε αυτό, είναι απαραίτητο να χρησιμοποιήσουμε σωστό σετ δεδομένων με εικόνες προσώπων των ανθρώπων που θέλουμε να αναγνωρίσουμε. Για αυτό το υπολογιστικό βήμα είναι επιτακτικό να μετασχηματίσουμε μια εικόνα ενός προσώπου σε ένα σετ των 3×3 macro-τετραγώνων για καλύτερη απεικόνιση, όπως δίνεται στην Εικόνα 17. Κάνοντάς το αυτό, είναι πιθανό να προσδιορίσουμε κάθε γνώρισμα που υπάρχει σε ένα ανθρώπινο πρόσωπο. Κάθε macro-τετράγωνο αποτελείται από 9 εικονοστοιχεία που παίρνουν τιμές στο εύρος από 0 έως 255, εφόσον είναι εντάσεως κλίμακας του γκρι.



Εικόνα 17: Αναπαράσταση ενός 3×3 Macro-block

Το επόμενο βήμα είναι να μετατρέψουμε το macro-block σε δυαδικό μοτίβο. Η τιμή του κεντρικού εικονοστοιχείου είναι ένα κατάφλι για όλα τα γειτονικά του. Οποιαδήποτε τιμή η οποία είναι ίση ή μεγαλύτερη του 150 τίθεται ως 1 και αν είναι μικρότερη, τίθεται ως 0 όπως φαίνεται στην Εικόνα 17. Ο αλγόριθμος συνδέει την δυαδική πληροφορία και την τοποθετεί σε ένα μοτίβο ακολουθώντας τη φορά του ρολογιού. Αυτή η διαδικασία είναι γνωστή ως Γραμμικό Δυαδικό Μοτίβο - Linear Binary Pattern (LBP). Η μαθηματική σχέση παρακάτω επίσης εκφράζει ένα LBP, όπου N είναι ο αριθμός των εικονοστοιχείων εκτός του κεντρικού. Υπάρχουν 8 γειτονικά του κεντρικού εικονοστοιχείου αν τεθεί η ακτίνα (r) = 1. Αυτό το macro-

τετράγωνο γίνεται ένα συρόμενο παράθυρο και εφαρμόζεται σε κάθε 3×3 ζευγάρι μεμονωμένα. Όταν η μαθηματική σχέση εφαρμόζεται σε ένα 3×3 macroblock μοτίβο, το μετασχηματίζει σε δυαδικό μοτίβο με αναφορά το κατώφλι, ή την τιμή του μέσου εικονοστοιχείου. Η μαθηματική σχέση απεικονίζει αυτό το φαινόμενο.

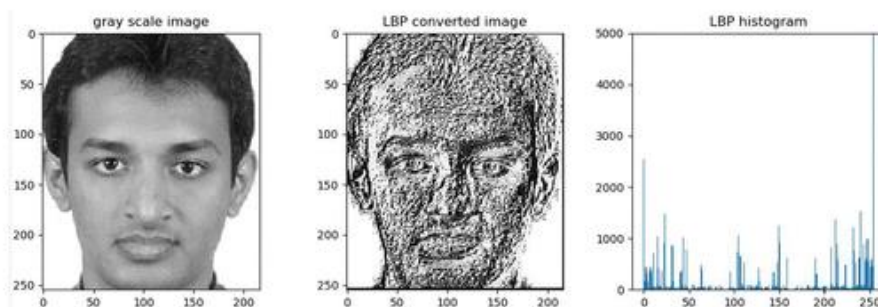
$$LBP = \sum_{p=0}^{N-1} K * (I_p - I_c) * 2^p$$

Όπου

$$K(x) \begin{cases} 1, & \text{αν } x \geq 0 \\ 0, & \text{σε άλλη περίπτωση} \end{cases}$$
$$0 * 2^0 + 0 * 2^1 + 1 * 2^2 + 0 * 2^3 + 1 * 2^4 + 1 * 2^5 + 1 * 2^6 + 1 * 2^7 = 244$$

$$H_j = \sum_{p(x,y)} I * \{F_i(x, y = j)\},$$
$$j = 0, 1, 2 \dots (n - 1)$$

Η Εικόνα 18 δίνει την απεικόνιση ως πίνακα, μετά την εύρεση της κάθε τιμής εικονοστοιχείου χρησιμοποιώντας δυαδικό μοτίβο στην $K(x)$. Ο αλγόριθμος μετά το πολλαπλασιάζει στο δυαδικό μοτίβο και το προσθέτει ώστε να αποκτήσει τον κώδικα του LBP. Αφού οριστικοποιηθεί η εικόνα LBP, χρησιμοποιείται η συνάρτηση H_j για να δημιουργηθεί μια απεικόνιση ιστογράμματος της εικόνας. Η Εικόνα 18 εμφανίζει την εικόνα μετασχηματισμένη ως LBP και το ισοδύναμο ιστόγραμμά της. Κάθε macro-τετράγωνο παράγει ξεχωριστό Ιστόγραμμα και τελικώς, ο αλγόριθμος τα συγκεντρώνει όλα για να παράγει ένα αθροιστικό ιστόγραμμα όλων των macro-τετραγώνων που δημιουργούνται σε μία εικόνα.



Εικόνα 18: (Αριστερά) Εικόνα προσώπου σε κλίμακα του γκρι, (κέντρο) εικόνα εφαρμόζοντας LBP και (δεξιά) Ιστόγραμμα της LBP εικόνας, Πηγή: <https://journals.sagepub.com/doi/full/10.1177/56>

Ακολουθεί η σύγκριση των ιστογραμμάτων της δοκιμαστικής εικόνας και των εικόνων στη βάση δεδομένων και επιστρέφεται η εικόνα με το κοντινότερο ιστόγραμμα. Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας διάφορες τεχνικές, όπως Ευκλείδεια Απόσταση, μέθοδος Χ-Τετραγώνων (chi-square), Απόλυτη τιμή, κ.α. Η Ευκλείδεια Απόσταση υπολογίζεται συγκρίνοντας τα χαρακτηριστικά της εικόνας δοκιμής με δεδομένα που έχουν αποθηκευτεί στο σετ δεδομένων. Η μικρότερη απόκλιση μεταξύ της εικόνας δοκιμής και του πρωτότυπου δίνει το βαθμό ταυτοποίησης, ή αλλιώς το επίπεδο εμπιστοσύνης της μέτρησης – *confidence*. Σαν αποτέλεσμα, λαμβάνουμε την ταυτότητα της εικόνας από τη βάση δεδομένων αν η εικόνα προς δοκιμή αναγνωρίζεται. Ο χειριστής του αλγορίθμου μπορεί έπειτα να ορίσει μια τιμή κατωφλίου εμπιστοσύνης ώστε να εξετάσει εάν το πρόσωπο αναγνωρίστηκε σωστά.

3.3.2 Πλεονεκτήματα της LBPH

Ο LBPH αλγόριθμος μπορεί να αναγνωρίσει πρόσωπα εάν κοιτούν προς το φακό ή είναι σε γωνία προφίλ και δεν επηρεάζεται από τις διακυμάνσεις του φωτισμού, που σημαίνει ότι είναι αρκετά ευέλικτος. Πιο αναλυτικά:

- Η μέθοδος LBPH είναι μία από τις καλύτερες στην περιγραφή υφής
- Είναι στιβαρή μέθοδος όταν χρησιμοποιείται για μονοτονικούς μετασχηματισμούς κλίμακας του γκρι
- Η μέθοδος FischerFaces εμποδίζει τα γνωρίσματα ενός προσώπου από το να κυριαρχήσουν, παρόλα αυτά λαμβάνει υπόψη τις διακυμάνσεις του φωτισμού ως ένα χρήσιμο χαρακτηριστικό. Όμως, οι διακυμάνσεις του φωτισμού δεν είναι ένα χρήσιμο χαρακτηριστικό προς εξαγωγή, καθώς δεν θεωρείται τμήμα του προσώπου.

- Η μέθοδος FisherFaces χρειάζεται μεγαλύτερο χώρο αποθήκευσης δεδομένων του προσώπου και περισσότερο υπολογιστικό χρόνο στην αναγνώριση.
- Με την LBPH μέθοδο, κάθε εικόνα αναλύεται ξεχωριστά, ενώ οι EigenFaces και FisherFaces μέθοδοι βλέπουν κάθε σετ δεδομένων ως σύνολο.
- Η μέθοδος LBPH πιθανότατα θα λειτουργήσει καλύτερα από την FisherFaces σε διαφορετικές περιβαλλοντικές συνθήκες και μεταβολές του φωτισμού. Βέβαια αυτό εξαρτάται από τα σετ δεδομένων εκπαίδευσης και δοκιμών.
- Μπορεί να απεικονίσει τοπικά γνωρίσματα στις εικόνες.
- Η μέθοδος LBPH μπορεί να αναγνωρίσει ευθεία ή πλάγια προς το φακό πρόσωπα.

4. Ερμηνεία των υποχρεωτικών κανονισμών στα UMS πλοία

Διάφορες ρυθμιστικές αρχές σε παγκόσμιο επίπεδο εκδίδουν κανονισμούς και διατάξεις για την εγκατάσταση και λειτουργία εξοπλισμού ελέγχου στα πλοία. Ο IMO (International Maritime Organization) προβλέπει υποχρεωτικούς κανονισμούς για τα πλοία που έχουν τη δυνατότητα περιοδικής επιτήρησης χώρων μηχανοστασίου. Αυτές οι διατάξεις περιέχονται στη Διεθνή Σύμβαση Διατήρησης της Ασφάλειας Ζωής στη Θάλασσα (Safety of Life at Sea – SOLAS). Περαιτέρω υποχρεωτικοί κανονισμοί επιβάλλονται από τις Εθνικές Διατάξεις της χώρας όπου είναι νηολογημένο το πλοίο (Σημαία), που συνήθως είναι ερμηνεία των κανονισμών SOLAS. Οι διεθνείς Νηογνώμονες όπως Lloyd’s Register of Shipping, Det Norske Veritas, American Bureau of Shipping και Bureau Veritas, εκδίδουν αναλυτικούς κανονισμούς ορίζοντας τις ελάχιστες διατάξεις ελέγχου και ασφαλείας που είναι απαραίτητο να εγκατασταθούν σε ένα πλοίο που έχει ανατεθεί προς γνωμάτευση.

Οι Νηογνώμονες επικοινωνούν μεταξύ τους στο φόρουμ IACS (International Association of Classification Societies) για να διαβεβαιώσουν ότι δεν υπάρχουν σημαντικές διαφορές μεταξύ των ελάχιστων προαπαιτούμενων και επίσης ότι συμμορφώνονται με τους κανονισμούς SOLAS κατά τη διαδικασία ταξινόμησης (classification). Σημαντικό είναι να σημειωθεί πως όταν ένα πλοίο έχει χαρακτηριστεί ως UMS – Unmanned Machinery Space (ή αντίστοιχο), η σημαία συνήθως επιτρέπει μειώσεις των ατόμων που επανδρώνουν το μηχανοστάσιο. Αυτή η απαλλαγή επιτρέπεται μόνο όταν ο χαρακτηρισμός UMS παραμένει ενεργός στους περιοδικούς ελέγχους που πραγματοποιεί ο Νηογνώμων, στα συστήματα ελέγχου και ασφαλείας του πλοίου, κατά τη διάρκεια πλεύσης του. Εάν βρεθούν ελαττώματα στην εγκατάσταση ή τις λειτουργίες αυτών των συστημάτων, που επηρεάζουν τον χαρακτηρισμό του ως UMS, ο χαρακτηρισμός ανακαλείται μέχρι να γίνουν οι απαραίτητες ενέργειες επιδιόρθωσης και να επαναληφθεί επιτυχής έλεγχος.

Ο χαρακτηρισμός UMS ορίζεται ως «Χώρος Μηχανοστασίου χωρίς Επιτήρηση» - αυτά τα πλοία που χαρακτηρίζονται ως UMS μπορούν γενικά να έχουν λιγότερο πλήρωμα λόγω του ότι τα συστήματα ελέγχου και ασφαλείας είναι περισσότερο προηγμένα. Γι’ αυτόν ακριβώς το λόγο η

αξιοπλοΐα και η σωστή λειτουργία των αυτόματων συστημάτων ασφαλείας ελέγχεται με πιο σχολαστικές μεθόδους.

Πιο αναλυτικά, κάποια προαπαιτούμενα συστήματα που θα μπορούσαν να χαρακτηρίσουν ένα πλοίο ως UMS, εξηγούνται παρακάτω:

- **Έλεγχος του εξοπλισμού πρόωσης από τη Γέφυρα**

Ένα σύστημα ελέγχου που επιτρέπει το χειρισμό των κύριων μηχανών θα πρέπει να είναι εγκατεστημένο στη γέφυρα του πλοίου. Ο αξιωματικός βάρδιας της γέφυρας πρέπει να έχει εύκολη πρόσβαση στο σύστημα ελέγχου εκτάκτου ανάγκης.

- **Ενοποιημένος έλεγχος**

Ένας εύκολα προσβάσιμος κεντρικός χώρος ελέγχου θα πρέπει να περιέχει τον επαρκή εξοπλισμό και όργανα ελέγχου και επιτήρησης. Ο εξοπλισμός εγκατεστημένος σε αυτό τον χώρο θα πρέπει να έχει τη δυνατότητα επιτήρησης και ελέγχου όλου του κύριων και βοηθητικών μηχανημάτων.

- **Αυτόματο σύστημα Πυρανίχνευσης και σημάτων κινδύνου**

Ένα σύστημα ανίχνευσης και συναγερμού, που αντιδρά άμεσα, θα πρέπει επίσης να είναι ικανό να δίνει έγκαιρη προειδοποίηση πυρκαγιάς σε χώρο μηχανημάτων, και πιο συγκεκριμένα:

- I. Στο Λέβητα
- II. Στην εισαγωγή αέρα της κύριας μηχανής
- III. Στο στροφαλοφόρο της κύριας μηχανής

Αισθητήρες ανίχνευσης νεφελωμάτων καυσίμου θα πρέπει να είναι εγκατεστημένοι σε μηχανές ισχύος 2250kW και παραπάνω, ή όταν η διάμετρος χιτωνίων της μηχανής υπερβαίνει τα 300mm – ειδικά σε επικίνδυνες ζώνες. Οι αισθητήρες αυτοί είναι πολυάριθμοι, τοποθετημένοι σε στρατηγικά σημεία και αποκρίνονται ταχύτατα.

- **Γενικό σύστημα συναγεμίων μηχανμάτων**

Το σύστημα αυτό θα πρέπει να είναι ικανό να απεικονίζει (μέσω μιμικού διαγράμματος) κάθε ανωμαλία των μηχανμάτων στη Γέφυρα του πλοίου, όπως και στους χώρους ενδιαίτησης, συμπεριλαμβανομένης και της καμπίνας του Μηχανικού Βάρδιας, της καμπίνας του Πρώτου Μηχανικού και όλους τους δημόσιους χώρους. Η τροφοδοσία ρεύματος του συστήματος συναγεμίων θα πρέπει να έχει και εφεδρική λήψη από τον πίνακα διανομής εκτάκτου ανάγκης, σε περίπτωση απρόσμενης διακοπής της κύριας τροφοδοσίας, και να υπάρχει μια ειδοποίηση συναγεμίου όταν συμβεί κάτι τέτοιο.

- **Χώρος ελέγχου και περιορισμού πυρκαγιάς**

Επιπρόσθετα των συμβατικών φορητών πυροσβεστήρων, είναι υποχρεωτικό να υπάρχει ένας χώρος ελέγχου και περιορισμού πυρκαγιάς ο οποίος να βρίσκεται απομακρυσμένα από τον χώρο μηχανοστασίου. Αυτός ο χώρος θα πρέπει να περιέχει πίνακες ελέγχου των αντλιών έκτακτης ανάγκης, των γεννητριών, βαλβίδων, εξαεριστήρων, μέσων πυρόσβεσης κλπ.

- **Αυτόματο σύστημα συναγεμίων υπερχειλίσης σεντινών και αντλίες εξυδάτωσης**

Για να διασφαλιστεί η προστασία από υπερχειλίση, οι υδροσυλλεκτικοί χώροι σεντίνας θα πρέπει να επιτηρούνται για υψηλή στάθμη (σε κατάσταση φυσιολογικής κλίσης του πλοίου – even keel). Για τις αυτόματες αντλίες σεντίνας, θα πρέπει να υπάρχει ένας συναγεμμός «μακρόχρονης λειτουργίας» της αντλίας. Αυτό υποδεικνύει υψηλό ρυθμό πλήρωσης της σεντίνας, καθώς η ικανότητα της αντλίας δεν αρκεί για να ανταπεξέλθει στον μεγάλο όγκο νερού. Οι έλεγχοι του συστήματος αντλιών και των απομακρυσμένων βαλβίδων θα πρέπει να είναι εύκολα προσβάσιμοι και πάνω από το πιθανό επίπεδο υπερχειλίσης. Αισθητήρια θα πρέπει να είναι τοποθετημένα σε καίρια σημεία και να παρέχεται χειροκίνητος ή αυτόματος έλεγχος του συστήματος.

- **Αυτόματα εκκινούμενη Η/Ζ εκτάκτου ανάγκης για αναγκαία ηλεκτρικά φορτία**

Αυτό το Ηλεκτροζεύγος είναι σχεδόν πάντα συνδεδεμένο με ξεχωριστό πίνακα διανομής εκτάκτου ανάγκης, τοποθετημένο πάνω από το κύριο κατάστρωμα και μακριά από το Ηλεκτροστάσιο, τον κύριο χώρο παραγωγής ηλεκτρικής ενέργειας του πλοίου. Η κύρια λειτουργία αυτού του Η/Ζ είναι να εκκινεί αυτόματα στην περίπτωση απρόσμενης διακοπής (Blackout) ή κατάστασης «νεκρού πλοίου (dead ship)» σε χρόνο (όχι περισσότερο από) 45 δευτερόλεπτα.

- **Τοπικός (χειροκίνητος) έλεγχος κύριων μηχανημάτων**

Τοπικοί ελεγκτές και εκκινητές μπορούν να δώσουν στο χειριστή δυνατότητα λειτουργίας, όταν η αυτόματη λειτουργία δεν είναι δυνατή ή όταν η χειροκίνητη λειτουργία είναι προτιμότερη για τοπική παρακολούθηση.

- **Αυτόματο σύστημα ελέγχου του καυστήρα**

Το σύστημα του λέβητα ελέγχεται αυτόματα με τη βοήθεια αισθητηρίων στάθμης νερού τροφοδοσίας. Επιπρόσθετα υπάρχει αυτόματο σύστημα ελέγχου καύσης με πολλές ασφαλιστικές διατάξεις, που προστατεύουν από πυρκαγιά ή έκρηξη.

- **Ασφαλιστικές διατάξεις**

- Οι διατάξεις θα πρέπει να βεβαιώνουν την επίτευξη ασφάλειας σε όλες τις περιπτώσεις, καθώς και να είναι ισοδύναμες με επανδρωμένους χώρους μηχανημάτων
- Επαρκώς σχεδιασμένα συστήματα ασφαλείας θα πρέπει να παρέχονται για τον αυτόματο τερματισμό λειτουργίας της κύριας μηχανής, των βοηθητικών μηχανών και του λέβητα, σε περίπτωση κάποιας σοβαρής βλάβης
- Οι εφεδρικές βοηθητικές μηχανές (Η/Ζ) που χρησιμοποιούνται για την παραγωγή ηλεκτρικής ενέργειας θα πρέπει να είναι ικανές να εκκινήσουν αυτόματα και να δεχθούν τα ηλεκτρικά φορτία του πλοίου σε περίπτωση βλάβης της Η/Ζ που λειτουργεί τη

δεδομένη στιγμή. Αυτό μπορεί να επιτευχθεί με την εγκατάσταση συστήματος παρακολούθησης και διαχείρισης ηλεκτρικής ενέργειας.

- Στην περίπτωση των ηλεκτρικών συστημάτων, θα πρέπει να παρέχονται μέσα για την αποκοπή των πλεοναζόντων φορτίων, όπως η «λειτουργία φορτίων προτίμησης» (preferential trip), όπου όσα ηλεκτρικά φορτία δεν θεωρούνται ουσιώδη για την πρόωση και πλοήγηση του πλοίου, βγαίνουν αυτόματα εκτός λειτουργίας (π.χ. μαγειρεία, κλιματισμός, μηχανήματα καταστώματος κ.α.) μέχρι την αποκατάσταση του ηλεκτρικού δικτύου διανομής
- Όλες οι αντλίες που είναι απαραίτητες για τη λειτουργία της κύριας μηχανής θα πρέπει να είναι διαθέσιμες εις διπλούν (θεωρώντας τη μια εφεδρική) και να υπάρχει δυνατότητα αυτόματης εκκίνησης της εφεδρικής σε περίπτωση βλάβης της αντλίας που λειτουργεί (για παράδειγμα μέσω ενός πρεσοστάτη που ανιχνεύει την πτώση πίεσης στο δίκτυο). Κάθε αντλία θα πρέπει να είναι ικανή να εκτελέσει το καθήκον της ανεξάρτητα.
- Το σύστημα πηδαλίου στην περίπτωση του μη-επανδρωμένου μηχανοστασίου θα πρέπει να έχει τη δυνατότητα να απομονώνει και να επαναφέρει τον έλεγχο, σε περίπτωση μεμονωμένης βλάβης, με αυτόματη λειτουργία βαλβίδων απομονώσεως. Τα συστήματα ελέγχου πηδαλίου συνήθως παρέχονται με δυο ή τρία ανεξάρτητα συστήματα που λειτουργούν εφεδρικά.
- Κατ' επέκταση, ο κανονισμός προβλέπει ελάχιστον αριθμό δυο αντλιών πηδαλίου όπου η μια θα τροφοδοτείται από τον πίνακα διανομής εκτάκτου ανάγκης, με τρόπο που θα συνεχίσει να λειτουργεί σε περίπτωση απρόσμενης συσκότισης του κυρίου πίνακα διανομής.
- Θα πρέπει να υπάρχει ένα εγκεκριμένο σύστημα για την βεβαίωση της ασφάλειας του προσωπικού που εισέρχεται σε χώρους μηχανημάτων, κατά τη διάρκεια της UMS λειτουργίας, για τη διερεύνηση συναγερμών, την παρακολούθηση της καλής λειτουργίας των μηχανημάτων ή την διεξαγωγή απαραίτητων επισκευών. Για παράδειγμα, υπάρχει το σύστημα «συναγερμού νεκρού» (dead man alarm system) για το οποίο θα αναφερθούμε αναλυτικά παρακάτω.

- **Τακτικές δοκιμές και συντήρηση των οργάνων/συστημάτων παρακολούθησης**

Περιοδικές επιθεωρήσεις κυκλικά, σε όλο τον εξοπλισμό παρακολούθησης, θα πρέπει να οριστούν, κατά τη διάρκεια ή όχι της επάνδρωσης των χώρων μηχανοστασίου. Σήμερα, η τάση είναι να δημιουργούνται κεντρικά δωμάτια ελέγχου, χρησιμοποιώντας ένα πλήρως ενοποιημένο σύστημα ελέγχου για όλες τις πτυχές της λειτουργίας του πλοίου, συμπεριλαμβανομένων των μηχανών, του φορτίου, της ναυσιπλοΐας και της γενικής διαχείρισης. Σε κάθε σύστημα, τα περισσότερα ελεγχόμενα μέρη θα πρέπει να έχουν έναν ή περισσότερους τρόπους λειτουργίας:

- A. Τοπικό χειροκίνητο έλεγχο
- B. Απομακρυσμένο χειροκίνητο έλεγχο
- C. Αυτόματο έλεγχο

Στον τοπικό έλεγχο συνεπάγεται ότι το σημείο ελέγχου βρίσκεται τοπικά στη μηχανή, ενώ στον απομακρυσμένο έλεγχο βρίσκεται κάπου πιο μακριά, όπως για παράδειγμα στο δωμάτιο ελέγχου. Η λειτουργία μιας βαλβίδας αναρροφήσεως σεντίνας από τον τοπικό μοχλό, στο μηχανοστάσιο, είναι ένα παράδειγμα τοπικής χειροκίνητης λειτουργίας. Εάν στη βαλβίδα υπήρχε μακρύτερος άξονας και μπορούσε να γίνει χειρισμός από το κατάστρωμα, θα θεωρούνταν απομακρυσμένος χειροκίνητος έλεγχος. Και στις δυο παραπάνω περιπτώσεις θα ήταν απαραίτητος ο ανθρώπινος παράγοντας (χειριστής) για τη λειτουργία της βαλβίδας. Εάν η βαλβίδα είχε υδραυλικό κινητήρα με μηχανισμό κίνησης και η βαλβίδα άνοιγε ή έκλεινε σύμφωνα με τη θέση του πλωτήρα ανιχνεύσεως υψηλής στάθμης στη σεντίνα, θα θεωρούνταν αυτόματος έλεγχος. Ο χειριστής είναι περιττός σε αυτή την περίπτωση.

- **Χειρισμός κύριων μηχανημάτων**

Σε πλοία που έχουν δυνατότητα απομακρυσμένου ελέγχου των μηχανημάτων πρόωσης από τη γέφυρα ναυσιπλοΐας και το μηχανοστάσιο, η ταχύτητα, η κατεύθυνση ώθησης και η κλίση της έλικας, όπου εφαρμόζονται, θα πρέπει να είναι εξοπλισμένα με μέσα πλήρους ελέγχου από τη γέφυρα ναυσιπλοΐας σε κάθε στιγμή και κάτω από οποιοσδήποτε συνθήκες πλεύσης. Στην περίπτωση περισσότερων από μιας έλικας που είναι σχεδιασμένες να λειτουργούν ταυτόχρονα, θα πρέπει να ελέγχονται από μια διάταξη.

Ο απομακρυσμένος έλεγχος θα πρέπει να είναι δυνατός από ένα σταθμό ελέγχου κάθε φορά, με ένδειξη ποιος σταθμός είναι ενεργός και στα υπόλοιπα σημεία ελέγχου. Η δυνατότητα μεταφοράς των ελέγχων θα πρέπει να βρίσκεται στο δωμάτιο ελέγχου του μηχανοστασίου και να υπάρχουν διατάξεις που αποτρέπουν διαφορετική εντολή κατά τη μεταφορά. Κάθε διακοπή ή αδυναμία λειτουργίας ενός σημείου ελέγχου θα πρέπει να ενεργοποιεί κάποια ηχητική και οπτική ένδειξη προειδοποίησης. Επιπρόσθετα, ένα τοπικό χειροκίνητο σημείο ελέγχου θα πρέπει να είναι εγκατεστημένο στην κύρια μηχανή που να μπορεί να παρακάμψει τον αυτόματο έλεγχο ανά πάσα στιγμή.

Τα κύρια μηχανήματα πρόωσης θα πρέπει να έχουν διάταξη διακοπής λειτουργίας εκτάκτου ανάγκης ανεξάρτητο από το προαναφερθέν σύστημα ελέγχου, στη γέφυρα ναυσιπλοΐας. Όλες οι εντολές από τη γέφυρα θα πρέπει να είναι ορατές στο δωμάτιο ελέγχου στο μηχανοστάσιο, καθώς και στο τοπικό χειροκίνητο σημείο ελέγχου της μηχανής. Η ταχύτητα της έλικας και η φορά περιστροφής της σε περίπτωση σταθερής έλικας, ή η γωνία περιστροφής της σε περίπτωση μεταβλητής έλικας, θα πρέπει να ενδείκνυνται στη γέφυρα ναυσιπλοΐας.

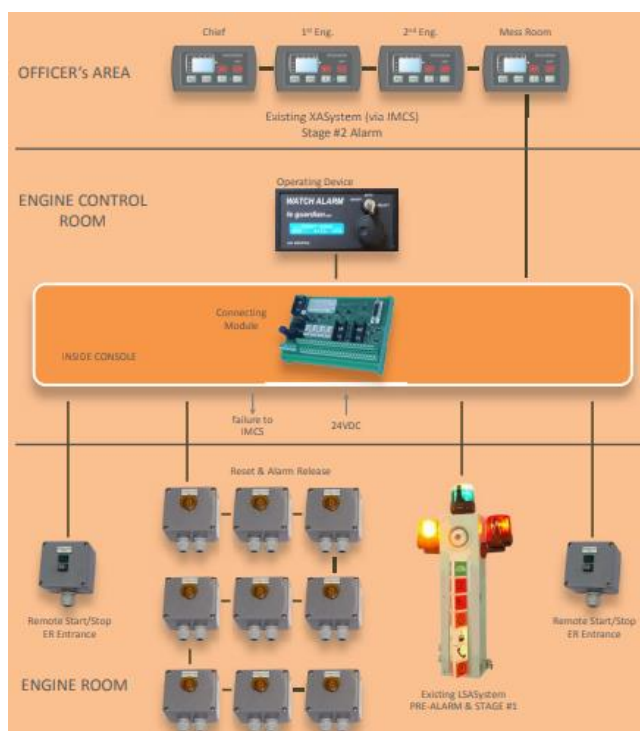
- **Αυτόματη διακοπή λειτουργίας**

Στην περίπτωση σοβαρών δυσλειτουργιών στα μηχανήματα ή στη λειτουργία του λέβητα οι οποίες είναι δυνατό να προκαλέσουν άμεσο κίνδυνο, το εξάρτημα που έχει βλάβη θα πρέπει να τίθεται αυτόματα εκτός λειτουργίας και ταυτόχρονα να ενεργοποιείται οπτική και ηχητική ένδειξη. Το σύστημα πρόωσης θα πρέπει επίσης να διακόπτεται αυτόματα, σε περίπτωση σοβαρής βλάβης, ολικής κατάρρευσης ή όταν ανιχνευθεί μεγάλη πιθανότητα έκρηξης.

Χειροκίνητη παράκαμψη του αυτόματου ελέγχου πρόωσης θα πρέπει να παρέχεται μόνο για την αποτροπή ακούσιας λειτουργίας. Σε τέτοιες περιπτώσεις θα πρέπει να υπάρχει συσκευή προειδοποίησης και ένδειξης της χειροκίνητης παράκαμψης του συστήματος.

4.1 Dead Man Alarm Σύστημα

Το σύστημα «συναγερμού νεκρού» (dead man alarm system), ενεργοποιείται κατά την είσοδο του μηχανικού βάρδιας στον χώρο μηχανοστασίου, και μετά από προκαθορισμένο χρόνο, θα πρέπει να επαναφέρεται χειροκίνητα μέσω μπουτόν που βρίσκονται τοποθετημένα σε διάφορα σημεία εντός του μηχανοστασίου. Σε περίπτωση ανικανότητας το Μηχανικό Βάρδιας να επαναφέρει το σύστημα μέσω των μπουτόν μέσα στον προκαθορισμένο χρόνο, θα ηχήσει συναγερμός στη Γέφυρα του πλοίου και μπορεί να ανατεθεί σε κάποιο άλλο μέλος του πληρώματος να διερευνήσει το λόγο και να κάνει τις απαραίτητες ενέργειες.



Εικόνα 19: Υπόδειγμα εγκατάστασης Dead Man Alarm, Πηγή: <http://www.deadmanalarm.com/sm-dma-watch-alarm/>

Επιγραμματικά παρατίθεται παρακάτω τμήμα των κανονισμών που διέπουν τη δυνατότητα μη-επανδρωμένου μηχανοστασίου σε διάφορους τύπους πλοίων, καθώς και τις δυνατότητες που θα πρέπει να έχει ένα Dead Man Alarm σύστημα.

Καθορισμός Βάρδιας σε μηχανοστάσιο

Ο κύριος κανονισμός ορίζει ότι το μηχανοστάσιο θα πρέπει να είναι συνεχώς επανδρωμένο από τα μέλη του πληρώματος.

Εξαίρεση – λειτουργία με περιοδικά μη-επανδρωμένο μηχανοστάσιο. Για να εγκριθεί ότι το πλοίο μπορεί να ταξιδέψει με περιοδικά μη-επανδρωμένο μηχανοστάσιο, θα πρέπει να πληροί τις παρακάτω προϋποθέσεις:

- Τα ταξινομημένα πλοία θα πρέπει να συμμορφώνονται με τους κανονισμούς που έχουν οριστεί από τον ανάλογο Νηογνώμονα όσον αφορά τα όργανα παρακολούθησης και τα συστήματα αυτοματισμού και να έχουν την ανάλογη σημειογραφία – Class Notation, για λειτουργία με περιοδικά μη-επανδρωμένο μηχανοστάσιο.
- Τα μη ταξινομημένα πλοία θα πρέπει να συμμορφώνονται με τους κανονισμούς του Det Norske Veritas (DNV) για λειτουργία με περιοδικά μη-επανδρωμένο μηχανοστάσιο.
- Πλοία με ισχύ πρόωσης που υπερβαίνει τα 750kW δε θα πρέπει να επιχειρούν με περιοδικά μη-επανδρωμένο μηχανοστάσιο, κάτω από τις παρακάτω συνθήκες:
 - Όταν επιχειρούνται μανούβρες προσέγγισης ή απομάκρυνσης από λιμένα
 - Όταν εισέρχονται στη ζώνη ασφαλείας των εγκαταστάσεων φόρτωσης-εκφόρτωσης (πχ. εγκαταστάσεις διυλιστηρίων)
 - Όταν πλέουν σε νερά με μεγάλη κυκλοφοριακή συμφόρηση (πχ. στενά του Βοσπόρου, Γιβραλτάρ, παράλια της Κίνας)
 - Όταν επικρατεί μειωμένη ορατότητα λόγω κακών καιρικών συνθηκών

Μη επανδρωμένα μηχανοστάσια – Διάρκεια

1. Για επιβατηγά πλοία, ο οργανισμός Norwegian Maritime Directorate θα ορίσει πόσο χρόνο μπορεί το μηχανοστάσιο να παραμένει μη-επανδρωμένο σε κάθε περίπτωση ξεχωριστά, κάτι το οποίο εξαρτάται από τις συνθήκες πλεύσης, κα. Η μη επανδρωμένη πλεύση συνήθως δεν θα πρέπει να ξεπερνά τις 2 ώρες.
2. Για πλοία διαφορετικού τύπου, η πλεύση με μη-επανδρωμένο μηχανοστάσιο θα πρέπει να ορίζεται και να οροθετείται στον χρόνο που θέτει ο ανάλογος Νηογνώμονας που ταξινομεί το πλοίο. Το χρονικό διάστημα μεταξύ δυο επιθεωρήσεων του μηχανοστασίου δε θα πρέπει να ξεπερνά τις 24 ώρες.

Βοηθήματα πλεύσης με μη-επανδρωμένο μηχανοστάσιο και προϋποθέσεις λειτουργίας:

Σε πλοία μεγαλύτερα από 500 GRT (Gross Register Tonnage), η βάρδια ενός και μόνο ατόμου στο μηχανοστάσιο δεν επιτρέπεται, εκτός και αν έχει εγκατασταθεί ένα αυτόματο σύστημα

«συναγερμού νεκρού» ή Dead Man Alarm system. Το σύστημα αυτό θα πρέπει να συμμορφώνεται και να ακολουθεί τις παρακάτω προϋποθέσεις:

- A. Το σύστημα ειδοποίησης θα πρέπει να είναι εγκατεστημένο ώστε να ενεργοποιεί συναγερμό στη γέφυρα του πλοίου, στις καμπίνες των αξιωματικών, στα εστιατόρια και τα δωμάτια ψυχαγωγίας του πλοίου εάν δεν αναγνωρισθεί από το μηχανοστάσιο. Ο συναγερμός θα πρέπει να ενεργοποιείται μετά από προ-ρυθμισμένο χρόνο που δεν θα υπερβαίνει τα 30 λεπτά.
- B. Ένα σήμα προειδοποίησης – Prewarning θα πρέπει να ενεργοποιείται στο μηχανοστάσιο μετά από χρόνο 3 λεπτών και πριν ενεργοποιηθεί ο συναγερμός που προαναφέρθηκε στην παράγραφο A.
- C. Το σύστημα συναγερμού θα πρέπει να τίθεται σε λειτουργία με τους ακόλουθους τρόπους: 1. Αυτόματα όταν κάποιος μηχανικός ή άλλο μέλος του πληρώματος θα πρέπει να εισέλθει στο μηχανοστάσιο, σε περίπτωση που θα χρειαστεί εξετάσει κάποιο άλλο πρόβλημα που παρουσιάστηκε και δόθηκε ειδοποίηση από το σύστημα παρακολούθησης μηχανοστασίου. 2. Χειροκίνητα από τον μηχανικό βάρδιας που θα εισέλθει στο μηχανοστάσιο για ελέγχους ρουτίνας.
- D. Το σύστημα ειδοποίησης/συναγερμού θα πρέπει να απενεργοποιείται από τον μηχανικό αφού ολοκληρώσει την εργασία του και αφού εξέλθει από το μηχανοστάσιο. Όταν το σύστημα μπαίνει αυτόματα σε λειτουργία όπως αναφέρεται στην παράγραφο C., η απενεργοποίηση δε θα πρέπει να είναι δυνατή πριν ο μηχανικός/μέλος του πληρώματος που βρίσκεται στο μηχανοστάσιο αναγνωρίσει την ειδοποίηση συναγερμού τοπικά.
- E. Το παρόν σύστημα ειδοποίησης θα πρέπει να ενεργοποιεί επίσης και το συναγερμό μηχανικού – Engineer’s Alarm.
- Σε συμφωνία με τις απαιτήσεις του *IMO Assembly, Resolution A.481 (XII), Annex 2, Part 7 «Engineering Watchkeeping», υποπαράγραφος 7.3.*
 - Σε συμφωνία με τις απαιτήσεις του *IMO Assembly, Resolution A.686 (17), υποπαράγραφος 7.1 «Personnel Alarm».*
 - Ένα σύστημα συναγερμού το οποίο λειτουργεί από το χώρο ελέγχου μηχανοστασίου (engine control room) ή από την πλατφόρμα ελιγμών, όπως προαναφέραμε, θα πρέπει να είναι διαθέσιμος και να ειδοποιεί το υπόλοιπο

πλήρωμα που βρίσκεται στους χώρους ενδιαίτησης του πλοίου ότι βοήθεια είναι απαραίτητη, *IMO Assembly Resolution A.686 (17)*, υποπαράγραφος 2.3.5.

Σήμερα, τα Dead Man Alarm συστήματα που εγκαθίστανται σε UMS πλοία είναι συμβατικά και η λειτουργία τους είναι πολύ απλή. Επίσης, λόγω του ότι δε χρησιμοποιούν ηλεκτρονικές διατάξεις δεν παρουσιάζονται πολύ συχνά βλάβες, κάνοντας το σύστημα αξιόπιστο. Η πρόταση που καλύπτει αυτή η ερευνητική προσπάθεια, είναι η εξειδίκευση του συστήματος και η πιο ακριβής ενημέρωση της ταυτότητας του ατόμου που ένιωσε αδυναμία κατά τη διάρκεια της περιόδου βάρδιας ή κάποιας επισκευής στο μηχανοστάσιο, με χρήση κάμερας και υπολογιστή με λογισμικό ανίχνευσης και αναγνώρισης προσώπων.

5. Περιγραφή της Εφαρμογής

Το παρόν σύστημα που παρουσιάζεται στην παρούσα διπλωματική διατριβή βασίζεται σε Η/Υ, όπου είναι εγκατεστημένο λογισμικό ανίχνευσης και αναγνώρισης προσώπου σε γλώσσα προγραμματισμού Python και μια οποιαδήποτε απλή web-κάμερα του εμπορίου. Η κάμερα μπορεί να εγκατασταθεί μόνιμα στην είσοδο (ή στις εισόδους) του μηχανοστασίου σε σημείο με καλό φωτισμό και με σχετικά ουδέτερο background, ώστε όταν ενεργοποιηθεί το σύστημα και κάποιο μέλος του πληρώματος περάσει μπροστά από την κάμερα, το λογισμικό που «τρέχει» από πίσω να κάνει ανίχνευση και αναγνώριση προσώπου. Έπειτα, τα στοιχεία του ατόμου που έχει εισέλθει κρατούνται στη βάση δεδομένων, όπου θα μπορούσαν να ενισχυθούν με άλλα γνωρίσματα, όπως:

- Γενικά στοιχεία (ονοματεπώνυμο, ηλικία κλπ.)
- Βαθμός (rank) του ατόμου
- Ώρες εργασίας του ατόμου στο πλοίο τη συγκεκριμένη μέρα ή την τελευταία εβδομάδα, σύμφωνα με τους κανονισμούς MLC (Maritime Labour Convention, 2006)
- Άλλα βιομετρικά στοιχεία και ιατρικό ιστορικό του ατόμου

Όταν λοιπόν το άτομο εισέλθει στο μηχανοστάσιο μπορεί να αρχίσει να μετράει αντίστροφα ο χρόνος, μέχρι η επόμενη κάμερα να τον ανιχνεύσει και να μηδενιστεί εκ νέου η μέτρηση. Η παρούσα διατριβή καλύπτει εφαρμογή με μια κάμερα, φυσικά μπορούν να εγκατασταθούν και περισσότερες μελλοντικά, για εντελώς αυτοματοποιημένη διαδικασία ανίχνευσης και αναγνώρισης, καθώς το μηχανοστάσιο των πλοίων είναι πολυεπίπεδο, με πολλά φυσικά εμπόδια (πχ. Μηχανήματα, κλίμακες κ.α.) τα οποία αποτρέπουν τη συνεχή παρακολούθηση του ατόμου που κινείται ανάμεσα σε διάφορα μηχανήματα ή εργάζεται σε κάποιο συγκεκριμένο.

5.1 Κώδικας σε Python

Στην παρούσα εφαρμογή αξιοποιήθηκαν τα εργαλεία που αναφέρονται παραπάνω, όπως η γλώσσα προγραμματισμού Python σε συνδυασμό με τη βιβλιοθήκη OpenCV, λόγω ευκολίας της εκμάθησης και χρήσης του κώδικα με σκοπό την ανίχνευση και αναγνώριση προσώπου.

Χρησιμοποιήθηκαν επίσης Haar Cascade ταξινομητές εφόσον ο στόχος μας είναι η ανίχνευση με αυξημένη ακρίβεια. Στο κεφάλαιο αυτό αναλύεται ο κώδικας λογισμικού που αναπτύχθηκε για την ανίχνευση και αναγνώριση προσώπου σε πραγματικό χρόνο. Για να επιτευχθεί αυτό χρειάστηκε αποθήκευση των δεδομένων γνωστών προσώπων σε μια βάση δεδομένων από την ίδια τη web-κάμερα, εκπαίδευση του LBP Classifier από αυτά τα δεδομένα και ανίχνευση – αναγνώριση προσώπου σε πραγματικό χρόνο.

5.2 Δημιουργία σετ δεδομένων

Για τη δημιουργία σετ δεδομένων, δημιουργήσαμε ένα φάκελο τοπικά στον υπολογιστή όπου θα καταχωρηθούν φωτογραφίες των οικείων προσώπων προς αναγνώριση. Το όνομα του φακέλου αντιστοιχεί στην ετικέτα του προσώπου (id) και θα τον χρησιμοποιήσουμε στην πορεία για την αναγνώριση του προσώπου.

Για τον κώδικα χρησιμοποιήθηκαν οι εξής εντολές της Python και της OpenCV:

- `face_cascade = cv2.CascadeClassifier (Διαδρομή Αρχείου)`: Συνάρτηση της OpenCV με την οποία καλούνται οι Haar ταξινομητές.
Διαδρομή Αρχείου: Η διαδρομή αρχείου, τοπικά στον υπολογιστή, που βρίσκεται ο ταξινομητής Haar.
- `Gray = cv2.cvtColor (img, cv2.COLOR_BGR2GRAY)`: Συνάρτηση της OpenCV, χρησιμοποιείται για μετατροπή της εικόνας σε κλίμακα του γκρι
`img`: Εικόνα που εξάγεται από βίντεο πραγματικού χρόνου
`cv2.COLOR_BGR2GRAY`: Το χρώμα που θέλουμε να μετατραπεί η εικόνα, στην προκειμένη περίπτωση κλίμακα του γκρι στα πρότυπα της OpenCV
- `Faces = face_cascade.detectMultiScale(gray, 1.5, 5)`: Συνάρτηση της OpenCV που χρησιμοποιείται για την ανίχνευση προσώπων σε πραγματικό χρόνο και διαφορετικό μέγεθος.
(`gray, 1.5, 5`): Εικόνα βίντεο στην οποία ανιχνεύονται τα πρόσωπα. Βλέπουμε ότι έχει μετατραπεί ήδη σε κλίμακα του γκρι.

Παρακάτω βλέπετε τον κώδικα που γράφτηκε σε Python Shell για τη δημιουργία της βάσης δεδομένων:

```
# -*- coding: utf-8 -*-
"""
Created on Sun Apr 12 18:35:40 2020
@author: ilias
"""
#Ρύθμιση παραμέτρων και εισαγωγή βιβλιοθηκών/συναρτήσεων
import numpy as np
import cv2
#Δημιουργία αντικειμένου ανίχνευσης και τοπική θέση στον υπολογιστή
face_cascade =
cv2.CascadeClassifier(r"C:\Users\Administrator\Desktop\DIPLOMATIKI\~\haarca
scades\haarcascade_frontalface_default.xml")
cap = cv2.VideoCapture(0)
id = input(r"enter user id: ")
sampleN=0;
while 1:
    ret, img = cap.read()
    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
    faces = face_cascade.detectMultiScale(gray, 1.5, 5)
    for (x,y,w,h) in faces:
#Καταγραφή εικόνας προσώπου στον φάκελο με το id και όνομα αρχείου τον
αριθμό SampleN
        sampleN=sampleN+1;
cv2.imwrite(r"C:\Users\Administrator\Desktop\Dataset\Ilias_Georgiou."+str(i
d)+ "." +str(sampleN)+ ".jpg", gray[y:y+h, x:x+w])
#Κατασκευή πλαισίου προσώπου
        cv2.rectangle(img, (x, y), (x+w, y+h), (255, 0, 0), 2)
        cv2.waitKey(100)
#Εμφάνιση αποτελεσμάτων
        cv2.imshow('img', img)
        cv2.waitKey(1)
#Τερματισμός προγράμματος σε 20 δείγματα SampleN
        if sampleN > 20:
            break
#Κλείσιμο κάμερας και παραθύρου
cap.release()
cv2.destroyAllWindows()
```

5.3 Ανίχνευση - Αναγνώριση προσώπου

Μετά την προετοιμασία της βάσης δεδομένων και της εκπαίδευσης του μοντέλου, φτάνουμε στο στάδιο της ανίχνευσης και αναγνώρισης προσώπου σε πραγματικό χρόνο, χρησιμοποιώντας ως αντικείμενο προς εξέταση το ζωντανό βίντεο της κάμερας. Για την υλοποίηση αυτού, δημιουργούμε κώδικα για τον αναγνωριστή – recognizer και εν συνεχεία τον εκπαιδούμε χρησιμοποιώντας τα πρόσωπα που έχουν καταχωρηθεί στη βάση δεδομένων και

τις αντίστοιχες ετικέτες τους (ID). Αφού ανιχνευθεί κάποιο πρόσωπο στο παράθυρο του ζωντανού βίντεο κάνουμε μια πρόβλεψη με τη χρήση του αναγνωριστή, και σύμφωνα με την αξιοπιστία δημιουργούμε μια πιο μικρή ζώνη αναγνώρισης έτσι ώστε να βοηθήσουμε την εξαγωγή σωστού αποτελέσματος. Εάν εντός της ζώνης αυτής υπάρχει ταύτιση με κάποιο ID από το σετ δεδομένων μας, εμφανίζεται πλαίσιο με το αν είναι το πρόσωπο οικείο, καθώς και η ταυτότητά του.

Για τον κώδικα χρησιμοποιούμε τη βιβλιοθήκη OpenCV σε Python, καθώς και τους Haar Ταξινομητές και τον LBPH αλγόριθμο που αναφέραμε παραπάνω:

- `Rec = cv2.face.LBPHFaceRecognizer_create (radius, neighbors, grid_x, grid_y)`, όπου:

Δημιουργείται ο ταξινομητής τοπικού δυαδικού μοντέλου, όπου:

Radius = Η ακτίνα του macro-τετραγώνου, σε εικονοστοιχεία γειτονικά του κεντρικού

Neighbors = Πλήθος των σημείων δειγματοληψίας που χρησιμοποιούνται για την κατασκευή του δυαδικού μοτίβου. Όσο πιο πολλά είναι, τόσο πιο αργό γίνεται το λογισμικό μας.

Grid_X = Αριθμός εικονοστοιχείων στον οριζόντιο άξονα. Όσο περισσότερα είναι, τόσο πλουσιότερο είναι το διάλυμα του χαρακτηριστικού που προκύπτει σε διαστάσεις.

Grid_Y = Αριθμός εικονοστοιχείων στον κάθετο άξονα. Όσο περισσότερα είναι, τόσο πλουσιότερο είναι το διάλυμα του χαρακτηριστικού που προκύπτει σε διαστάσεις

- `faces = face_cascade.detectMultiScale(src, scaleFactor, minNeighbors)`, όπου:

src = Πηγή εικόνας όπου θα εφαρμοστεί αναγνώριση προσώπου

scaleFactor = Παράμετρος αλλαγής μεγέθους της εικόνας υπό κλίμακα

minNeighbors = Παράμετρος οι ορίζει το πλήθος των γειτόνων που πρέπει να διατηρηθούν σε κάθε υποψήφιο macro-τετράγωνο.

```
@author: ilias
"""
import numpy as np
import cv2
face_cascade =
cv2.CascadeClassifier(r"C:\Users\Administrator\Desktop\DIPLOMATIKI\~\haarca
scades\haarcascade_frontalface_default.xml")
eye_cascade =
cv2.CascadeClassifier(r"C:\Users\Administrator\Desktop\DIPLOMATIKI\~\haarca
scades\haarcascade_eye_tree_eyeglasses.xml")
cap = cv2.VideoCapture(0)
#Ορισμός ανίχνευσης και αναγνώρισης μέσω LBPH της OpenCV
rec = cv2.face.LBPHFaceRecognizer_create(3,9,1,1);
rec.read(r"C:\Users\Administrator\Desktop\DIPLOMATIKI\Dataset\trainingdata.
yaml")
id=0
font=cv2.FONT_HERSHEY_DUPLEX
while 1:
#Ορισμός του πλαισίου και μετασχηματισμός σε κλίμακα του γκρι
    ret, img = cap.read()
    gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
#Ανίχνευση Προσώπων στο βίντεο, σε πραγματικό χρόνο, με την OpenCV
    faces = face_cascade.detectMultiScale(gray, 1.3, 5)
#Ελεγχος και εξαγωγή, μετά την ανίχνευση, των χαρακτηριστικών του προσώπου
    for (x,y,w,h) in faces:
#Δημιουργία τετραγώνου στο πρόσωπο που ανιχνεύεται
        cv2.rectangle(img, (x,y), (x+w,y+h), (255,0,0),2)
#Σύγκριση των Διανυσμάτων του προσώπου που ανιχνεύεται, με το σετ δεδομένων
και επιστροφή της ταυτότητας του (ID)
            id,conf=rec.predict(gray[y:y+h,x:x+w])
            if id==1:
                id="Ilias Georgiou"
            if id==2:
                id="Argyro K."

        cv2.putText(img,str(id), (x,y-10), font,0.55, (0,255,0),1)

        roi_gray = gray[y:y+h, x:x+w]
        roi_color = img[y:y+h, x:x+w]
        eyes = eye_cascade.detectMultiScale(roi_gray)
        for (ex,ey,ew,eh) in eyes:
            cv2.rectangle(roi_color, (ex,ey), (ex+ew,ey+eh), (0,255,0),2)
#Εμφάνιση αποτελεσμάτων
        print ("found" + str(len(faces)) + "face(s)")
        cv2.imshow('img',img)
#Αναμονή τερματισμού του προγράμματος με την πληκτρολόγηση του γράμματος
"f" και κλείσιμο των παραθύρων
        if cv2.waitKey(1) == ord('f'):
            break

cap.release()
cv2.destroyAllWindows()
```

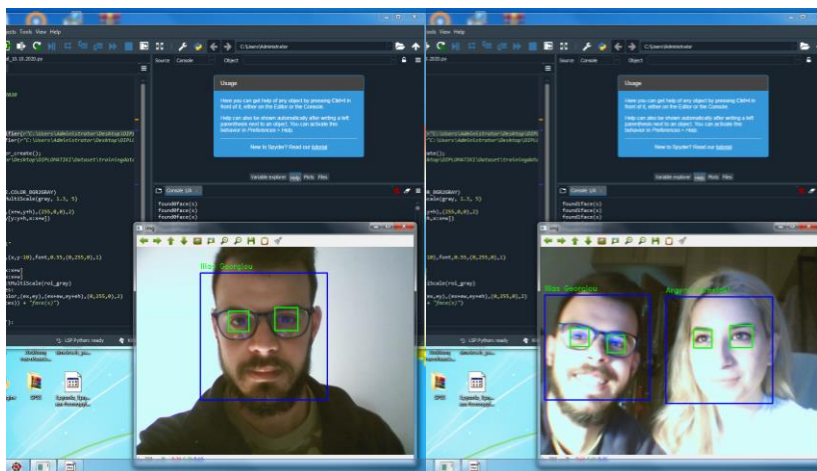
6. Αποτελέσματα – Συμπεράσματα – Μελλοντική έρευνα

Αρχικά, εικόνες οικείων προσώπων χρησιμοποιήθηκαν για την εκπαίδευση του αλγορίθμου με στόχο την αναγνώριση. Ο αλγόριθμος δημιουργεί το δικό του σετ δεδομένων και μετατρέπει τις εικόνες σε κλίμακα του γκρι, ώστε να κάνει την απαραίτητη επεξεργασία και να αποφασίσει αν στο βίντεο πραγματικού χρόνου εμφανίζεται οικείο πρόσωπο. Στην εφαρμογή μας έχουμε εκπαιδεύσει τον αλγόριθμο με 2 πρόσωπα.



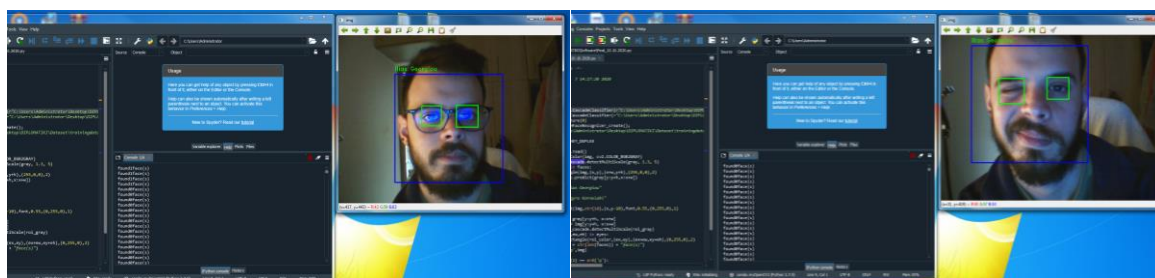
Εικόνα 20: Εικόνες σετ Δεδομένων

Μετά από τη ανάπτυξη του κώδικα έγινε δοκιμαστική εκτέλεση του προγράμματος και αναγνώριση προσώπων, με τα δεδομένα που είχαν αποθηκευτεί σε πρότερο στάδιο στο σετ δεδομένων (εκπαίδευσης του αλγορίθμου). Παρακάτω έχουν εισαχθεί φωτογραφίες από το λογισμικό, με τα πρόσωπα που αναγνωρίζονται:



Εικόνα 21: Ανίχνευση και Αναγνώριση Προσώπου με το λογισμικό, σε συνθήκες κανονικού φωτισμού.

Παρατηρούμε ότι σε συνθήκες κανονικού φωτισμού το λογισμικό παρουσιάζει καλή αξιοπιστία στην αναγνώριση των προσώπων. Επίσης, εμφανίζει με επικάλυψη ενός τετραγώνου και κειμένου την ταυτότητα του ατόμου, κατευθείαν πάνω στο ζωντανό βίντεο και σε παράθυρο του προγράμματος των αριθμό των ατόμων που έχουν ανιχνευθεί.



Εικόνα 22: Ανίχνευση και αναγνώριση προσώπου με γυαλιά/γκριμάτσα

Η εφαρμογή θα μπορούσε να βελτιωθεί περαιτέρω, καθώς ο εξοπλισμός που έχει χρησιμοποιηθεί είναι πολύ απλός (όπως αναφέρθηκε και παραπάνω) με κύριο στόχο την ταχύτητα στην ανίχνευση και αναγνώριση, αλλά και στο χαμηλό κόστος. Χρησιμοποιήθηκαν μόνο 20 εικόνες του κάθε οικείου ατόμου για την εκπαίδευση του αλγορίθμου, με την αύξηση αυτού του αριθμού θα μπορούσαμε επίσης να βελτιώσουμε την αξιοπιστία στην ταύτιση των ετικετών της ζωντανής εικόνας με αυτές του σετ δεδομένων. Το ίδιο θα συνέβαινε επίσης με τον εμπλουτισμό της βάσης δεδομένων με εικόνες αγνώστων προσώπων.

Να σημειωθεί επίσης εδώ ότι όλες οι τιμές των παραμέτρων των εντολών που επιλέχθηκαν, σε εντολές της Python, είναι πειραματικές. Εισάχθηκαν κατ' εκτίμηση λαμβάνοντας υπόψη τη συνεργασία του υπάρχοντος εξοπλισμού με το λογισμικό και τις περιβαλλοντικές συνθήκες που καλούνται να λειτουργήσουν.

Η εφαρμογή επιδέχεται πολλών ειδών βελτιώσεις σε υλικό επίπεδο (hardware) στους τομείς της φορητότητας, του μεγέθους και της υπολογιστικής ισχύος. Με την επίτευξη αυτών θα μπορούσαν να ενταχθούν και περισσότερες από μια κάμερες ώστε να παρακολουθείται κάθε σημείο του μηχανοστασίου. Στο επίπεδο του λογισμικού (software), πέρα από την αύξηση της αξιοπιστίας του με τη βελτιστοποίηση των παραμέτρων των εντολών, θα μπορούσε να προστεθεί και κάποιο τμήμα αναγνώρισης κινούμενου ανθρώπινου σώματος που θα βοηθούσε πολύ σε αυτή την εφαρμογή, καθώς το άτομο που έχει αναγνωριστεί και κινείται στο μηχανοστάσιο του πλοίου είναι αδύνατο να έχει πάντα στραμμένο το πρόσωπό του προς την κάμερα.

7. Αναφορές - Βιβλιογραφία

- <https://www.python.org/>
- <https://opencv.org/>
- International Maritime Organization,
<https://www.imo.org/en/About/Conventions/Pages/ListOfConventions.aspx>
- International Convention for the Safety of Life at Sea (SOLAS),
[https://www.imo.org/en/About/Conventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-\(SOLAS\),-1974.aspx](https://www.imo.org/en/About/Conventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-(SOLAS),-1974.aspx)
- Maritime Labour Convention (MLC), 2006
<https://www.ilo.org/global/standards/maritime-labour-convention/lang--en/index.htm>
- O'Toole, A. J., Phillips, P. J., Jiang, F., Ayyad, J., Penard, N., and Abdi, H. (2009). Face recognition algorithms surpass humans matching faces over changes in illumination.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(9):1642–1646.
- Kirovski, D., Jojic, N., and Jancke, G. (2004). Tamper-resistant biometric IDs. In *ISSE 2004 - Securing Electronic Business Processes: Highlights of the Information Security Solutions Europe 2004 Conference*, pp. 160–175.
- Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. (2003a). *Face recognition: A literature survey*. *ACM Computing Surveys*, 35(4):399–358.
- Chokshi, N. (2019). *Facial recognition's many controversies, from stadium surveillance to racist software*. *New York Times*.
- Phillips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J. (2000). The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104.
- Sim, T., Baker, S., and Bsat, M. (2003). The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618.
- Gross, R., Shi, J., and Cohn, J. F. (2005). Quo vadis face recognition? In *IEEE Workshop on Empirical Evaluation Methods in Computer Vision, San Diego*.
- Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. *Technical Report 07-49, University of Massachusetts, Amherst*.
- Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W. et al. (2010). FRVT 2006 and ICE 2006 large-scale experimental results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):831–846.
- Wolf, L., Hassner, T., and Maoz, I. (2011). Face recognition in unconstrained videos with matched background similarity. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO*.
- Kemelmacher-Shlizerman, I., Seitz, S. M., Miller, D., and Brossard, E. (2016). The MegaFace benchmark: 1 million faces for recognition at scale. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

- Maze, B., Adams, J., Duncan, J. A., Kalka, N., Miller, T., Otto, C., Jain, A. K., Niggel, W. T., Anderson, J., Cheney, J., and Grother, P. (2018). IARPA Janus Benchmark - C: Face dataset and protocol. *In International Conference on Biometrics (ICB)*, pp. 158–165.
- Masi, I., Wu, Y., Hassner, T., and Natarajan, P. (2018). Deep face recognition: A survey. *In Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 471–478.
- Fischler, M. A. and Elschlager, R. A. (1973). The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22(1):67–92.
- Kanade, T. (1977). *Computer Recognition of Human Faces*. Birkhauser, Basel.
- Yuille, A. (1991). Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70.
- Chellappa, R., Wilson, C., and Sirohey, S. (1995). Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740.
- Li, S. Z. and Jain, A. K. (eds). (2005). *Handbook of Face Recognition*, Springer
- Sinha, P., Balas, B., Ostrovsky, Y., and Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962
- Penev, P. and Atick, J. (1996). Local feature analysis: A general statistical theory for object representation. *Network Computation and Neural Systems*, 7:477–500.
- Wiskott, L., Fellous, J.-M., Kruger, N., and von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779.
- Rowland, D. A. and Perrett, D. I. (1995). Manipulating facial appearance through shape and color. *IEEE Computer Graphics and Applications*, 15(5):70–76.
- Beymer, D. (1996). Feature correspondence by interleaving shape and texture computations. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 921–928, San Francisco.
- Vetter, T. and Poggio, T. (1997). Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):733–742.
- Lanitis, A., Taylor, C. J., and Cootes, T. F. (1997). Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):742–756.
- Blanz, V. and Vetter, T. (1999). A morphable model for the synthesis of 3D faces. *In ACM SIGGRAPH 1999 Conference Proceedings*, pp. 187–194.
- Egger, B., Smith, W. A. P., Tewari, A., Wuhler, S., Zollhoefer, M., Beeler, T., Bernard, F., Bolkart, T., Kortylewski, A., Romdhani, S., Theobalt, C., Blanz, V., and Vetter, T. (2020). 3D morphable face models—past, present, and future. *ACM Transactions on Graphics*, 39(5).
- Cootes, T. F. and Taylor, C. J. (2001). Statistical models of appearance for medical image analysis and computer vision. *In Medical Imaging*.

- Costen, N., Cootes, T. F., Edwards, G. J., and Taylor, C. J. (1999). Simultaneous extraction of functional face subspaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 492–497, Fort Collins.
- Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720.
- Moghaddam, B., Jebara, T., and Pentland, A. (2000). Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782.
- Gross, R., Matthews, I., and Baker, S. (2006). Active appearance models with occlusion. *Image and Vision Computing*, 24(6):593–604.
- Ramnath, K., Koterba, S., Xiao, J., Hu, C., Matthews, I., Baker, S., Cohn, J., and Kanade, T. (2008). Multi-view AAM fitting and construction. *International Journal of Computer Vision*, 76(2):183–204.
- Murphy-Chutorian, E. and Trivedi, M. M. (2009). Head pose estimation in computer vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4):607–626.
- Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In *British Machine Vision Conference*, p. 6.
- Huang, Y., Shen, P., Tai, Y., Li, S., Liu, X., Li, J., Huang, F., and Ji, R. (2020). Improving face recognition from hard samples via distribution distillation loss. In *European Conference on Computer Vision (ECCV)*.