

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

Μελέτη, Σχεδίαση και Ανάπτυξη Εφαρμογής Λογισμικού για τον Έλεγχο Εισόδου Υπαλλήλων στο Χώρο Εργασίας τους με Βάση μια Ετικέτα RFID και την Αναγνώριση του Προσώπου τους

ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΗΣ ΕΛΛΑΔΟΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΚΑΒΑΛΑ

2022

Επιβλέπων Καθηγητής: **Θεόδωρος Παχίδης**, Καθηγητής ΔιΠαΕ

**Παναγιώτης Σκλίδας**

**Copyright@ 2022 Τμήμα Πληροφορικής, Διεθνές Πανεπιστήμιο της Ελλάδος,**

<https://www.cs.ihu.gr/>

Το περιεχόμενο της συγκεκριμένης Πτυχιακής Εργασίας αποτελεί πνευματική ιδιοκτησία του συγγραφέα, του επιβλέποντα καθηγητή και του τμήματος Πληροφορικής του ΔΙΠΑΕ και προστατεύεται από τους νόμους περί πνευματικής ιδιοκτησίας (Νόμος 2121/1993 και κανόνες Διεθνούς Δικαίου που ισχύουν στην Ελλάδα).

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπόγραφα ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται με λεπτομέρεια στην πτυχιακή εργασία. Έχω αναφέρει πλήρως και με σαφείς αναφορές όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία, είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

|  |  |
| --- | --- |
| **Επώνυμο** | Σκλίδας |
| **Όνομα** | Παναγιώτης |
| **Αριθμός Μητρώου** | 4360 |
| **Ημερομηνία** | 29/5/2022 |
| **Υπογραφή** |  |

Ευχαριστίες

Περιεχόμενα

[Περίληψη 2](#_Toc107057401)

[Abstract 3](#_Toc107057402)

[Κεφάλαιο 1 Εισαγωγή 4](#_Toc107057403)

[1.1 Τεχνολογία RFID 4](#_Toc107057404)

[1.1.1 Τύποι RFID ετικετών 4](#_Toc107057405)

[1.1.2 Συσκευές ανάγνωσης – εγγραφής ετικετών RFID 4](#_Toc107057406)

[1.1.3 Χαρακτηρίστηκα συστημάτων RFID 5](#_Toc107057407)

[1.2 Συστήματα μηχανικής όρασης 6](#_Toc107057408)

[1.2.1 Convolutional Neural Network 6](#_Toc107057409)

[1.2.2 Χρήσης μηχανικής όρασης 8](#_Toc107057410)

[1.2.2.1 Βιομηχανία μεταφορών 9](#_Toc107057411)

[1.2.2.2 Ιατρική 9](#_Toc107057412)

[1.2.2.3 Βιομηχανία κατασκευών 10](#_Toc107057413)

[1.3 Αναγνώρισης προσώπου 10](#_Toc107057414)

[1.3.1 Αλγόριθμοι αναγνώρισης προσώπου 11](#_Toc107057415)

[1.3.1.1 Haar Cascade 11](#_Toc107057416)

[1.3.1.2 Eigenfaces 12](#_Toc107057417)

[1.3.1.3 Fisherfaces 15](#_Toc107057418)

[1.3.1.4 Local Binary Pattern (LBP) 16](#_Toc107057419)

[1.3.1.5 Αναγνώριση προσώπου με CNN 18](#_Toc107057420)

[1.4 User-friendly εφαρμογές 21](#_Toc107057421)

[1.4.1 Διεπαφή χρήστη 21](#_Toc107057422)

[1.4.2 Εμπειρία χρήστη 22](#_Toc107057423)

[1.5 Επιλογές 22](#_Toc107057424)

[Βιβλιογραφία 23](#_Toc107057425)

# Περίληψη

Η πτυχιακή αυτή εργασία αφορά: 1) Στην έρευνα βιβλιογραφική και μέσω του διαδικτύου α) για την εύρεση συσκευών ανίχνευσης ετικετών (tags) RFID, των διαφορετικών μορφών ετικετών και του σχετικού λογισμικού που τις συνοδεύει καθώς και των ιδιαίτερων χαρακτηριστικών αυτών των συστημάτων. β) Συστημάτων όρασης και των χαρακτηριστικών τους που χρησιμοποιούνται σε ανάλογα συστήματα καταγραφής και ελέγχου. γ) Αλγόριθμων αναγνώρισης προσώπου (διαφορετικές προσεγγίσεις, χαρακτηριστικά, ποσοστά αναγνώρισης, κ.τ.λ.) 2) Την ανάπτυξη εφαρμογής λογισμικού, φιλικής προς το χρήστη, ακολουθώντας τις αρχές της τεχνολογίας λογισμικού με τις εξής απαιτήσεις: α) ανάγνωση μιας ετικέτας RFID μέσω του κατάλληλου συστήματος που θα συνδέεται με την εφαρμογή λογισμικού και καταγραφή της σχετικής πληροφορίας. β) Εισαγωγή, τροποποίηση, διαγραφή των προσωπικών στοιχείων των υπαλλήλων που εργάζονται σε ένα οργανισμό ή επιχείρηση από διαφορετικές μορφές αρχείων και καταγραφή των στοιχείων αυτών. Η εφαρμογή θα μπορεί να διαχειριστεί στοιχεία υπαλλήλων που εργάζονται στον ίδιο οργανισμό ή επιχείρηση αλλά και σε διαφορετικούς χώρους εργασίας. γ) Συλλογή, βελτίωση και καταγραφή της εικόνας του προσώπου υπαλλήλων με τη βοήθεια του κατάλληλου συστήματος όρασης. Αναγνώριση του προσώπου υπαλλήλων με τη βοήθεια αποθηκευμένων εικόνων του προσώπου του υπαλλήλου. δ) Ταυτοποίηση των υπαλλήλων με αντιστοίχιση του κωδικού του RFID tag με τα προσωπικά στοιχεία και την εικόνα του προσώπου τους. Η εφαρμογή θα πρέπει να συνδυάζει όλα αυτά τα διαφορετικά στοιχεία για την αξιόπιστη ταυτοποίηση του κάθε υπαλλήλου. ε) Αυτόματη λειτουργία αρχικής εγγραφής και στη συνέχεια καταγραφής των παρουσιών των υπαλλήλων, των χρόνων καταγραφής (εισόδου – εξόδου), της διάρκειας παρουσίας των στο χώρο εργασίας, κ.τ.λ. σε τύπο αρχείου που επιλέγεται παραμετρικά από το χρήστη και με δυνατότητα διαμόρφωσης του τρόπου καταγραφής των παρουσιών. Να υπάρχει η δυνατότητα μετατροπής σε διαφορετικούς τύπους αρχείου με διαφορετική διαμόρφωση. στ) Έκδοση στατιστικών στοιχείων για τα χρόνο παραμονής των υπαλλήλων στους χώρους εργασίας του οργανισμού ή της επιχείρησης (π.χ. ανά εβδομάδα, μήνα, έτος), τις μετακινήσεις, κ.τ.λ. ζ) Δυνατότητα διαχείρισης του χρόνου εργασίας των υπαλλήλων. 3) Πειραματισμός με το σύστημα για την εύρεση της ταχύτητας αρχικής εγγραφής και στη συνέχεια καταγραφής καθώς και της αξιοπιστίας του γενικότερα ανάλογα με τον αριθμό των εγγραφών, τα ποσοστά αναγνώρισης και ταυτοποίησης, κ.τ.λ.

Η γλώσσα υψηλού επιπέδου μπορεί να είναι η C++, C#, Java. Σκοπός της πτυχιακής αυτής εργασίας είναι η απόκτηση γνώσεων και δεξιοτήτων σε θέματα που αφορούν στην ανάπτυξη μιας ολοκληρωμένης εφαρμογής λογισμικού (ακολουθώντας τις αρχές της Τεχνολογίας Λογισμικού), η απόκτηση γνώσεων σε θέματα που αφορούν επεξεργασία εικόνας, η απόκτηση γνώσεων και εμπειρίας στη σύνδεση και επικοινωνία με εξωτερικές συσκευές καθώς και στον έλεγχό τους με τη βοήθεια λογισμικού.

# Abstract

The present dissertation concerns: 1) In bibliographic research and through internet a) to find RFID tag detection devices, the different forms of tags, and the relevant software that accompanies them as well as the special features of these systems. b) Vision systems and their characteristics, which used in similar recording and control systems. c) Face recognition algorithms (different approaches, features, recognition rates, etc.) 2) The development of a user-friendly software application, based on the principles of software technology with the following requirements: a) reading a RFID tag through the appropriate system, that it will be connected to the software application and recording the relevant information. b) Import, modify and delete the personal data of employees, who are working in an organization or company from different file formats and record these data. The application will be able to manage data of employees, who are working in the same organization or company but also in different workplaces. c) Collection, improvement, and recording of the image of the employee's faces with the use of the appropriate vision system. Identification of the employee's faces via stored images of their faces. d) Identification of employees by matching the RFID tag code with their personal details and face image. The application should combine all these different elements for the reliable identification of each employee. e) Automatic function of initial registration and then recording of the employees presence, the recording times (entry-exit), the duration of their presence in the workplace, etc. in a file type that is selected parametrically by the user and with the ability to configure how to record the employees presence. Further, it should have the ability to convert to different file type with a different formation. f) Issuance of statistics on the employee time of stay at the workplace of the organization or company (e.g. per week, month, year), travel, etc. g) Ability to manage the working time of employees. 3) Experiment with the system to find the speed of initial registration and afterwards the recording as well as its general reliability, depending on the number of registrations, recognition and identification rates, etc.

The high-level language can be C ++, C #, Java. The purpose of this dissertation is to acquire knowledge and skills in subjects related to the development of a complete software application (following the principles of Software Technology), to acquire knowledge related to image processing, to gain knowledge and experience in connecting and communicating with external devices as well as their control with the help of software.

# Κεφάλαιο 1 Εισαγωγή

## 1.1 Τεχνολογία RFID

Η ταυτοποίηση μέσω ραδιοσυχνοτήτων (Radio Frequency Identification - RFID) είναι ένας τύπος ασύρματης επικοινωνίας, ο πρόγονος της οποίας παρουσιάστηκε τον Ιανουάριο του 1973 από τον Mario Cardullo[1,2], λειτουργούσε είτε με ηλεκτρομαγνητική είτε με ηλεκτροστατική σύζευξη στο τμήμα των ραδιοσυχνοτήτων ηλεκτρομαγνητικού φάσματος και χρησιμοποιείται για την αναγνώριση αντικειμένων, ζώων ή ανθρώπων. Η αναγνώριση πραγματοποιείται με την βοήθεια εξειδικευμένων μικροεπεξεργαστών, γνωστών και ως tags, οι οποίοι βρίσκονται ενσωματωμένοι είτε σε κάποια αυτοκόλλητη ετικέτα πάνω σε κάποιο αντικείμενο ή κάποια έξυπνη κάρτα την οποία έχει στην διάθεση του ένας άνθρωπος είτε είναι εμφυτευμένοι σε κάποιο ζώο.

### 1.1.1 Τύποι RFID ετικετών

Οι RFID ετικέτες μπορεί να είναι είτε ενεργητικές (active) είτε παθητικές (passive). Οι ενεργητικές κάρτες πέρα από τον μικροεπεξεργαστή διαθέτουν και μια μπαταρία η οποία παρέχει συνεχώς ρεύμα στον μικροεπεξεργαστή δίνοντας έτσι την ικανότητα στις ενεργητικές ετικέτες να εκπέμπουν το σήμα τους συνεχώς. Αντίθετα, οι παθητικές ετικέτες όπως δηλώνει και το όνομα τους, δεν διαθέτουν μπαταρία και αντλούν την απαραίτητη ενέργεια που χρειάζονται για να μεταδώσουν το σήμα από τον αναγνώστη, έτσι οι παθητικές ετικέτες μεταδίδουν το σήμα μόνο όταν βρίσκονται αρκετά κοντά στην συσκευή ανάγνωσης. Επιπλέον, οι ενεργητικές ετικέτες έχουν την δυνατότητα να μεταδίδουν σε πολλή μεγάλες αποστάσεις (περίπου 150m για τις Ultra High Frequency), ενώ οι παθητικές μπορούν να εκπέμπουν σε σχετικά μικρές αποστάσεις (περίπου 15m για τις Ultra High Frequency)[3]. Τέλος, οι ετικέτες μπορεί να είναι ή μόνο για ανάγνωση (read only tags) ή ανάγνωσης και εγγραφής (read/write tags).



Σχήμα 1.1: (a) Ενεργητικές και (b) παθητικές RFID ετικέτες

### 1.1.2 Συσκευές ανάγνωσης – εγγραφής ετικετών RFID

Για να μπορέσουν να χρησιμοποιηθούν οι ετικέτες πρέπει να έχουμε στην διάθεση μας μια συσκευή εγγραφής – ανάγνωσης RFID (RFID reader/writer). Στο εμπόριο υπάρχουν αρκετά είδη συσκευών ανάγνωσης όπως για παράδειγμα ο RC522 RFID Reader/Writer o οποίος είναι ένα πρόσθετο (module) για το Arduino και για το Raspberry Pi. Στην ίδια κατηγόρια ανήκει και ο PN532 NFC RFID Module Reader Writer[4]. Πέρα από τα modules για μικροεπεξεργαστές τα οποία πρέπει ο χρήστης να συνδέσει με το board της επιλογής του με καλώδια και να τα προγραμματίσει ώστε να εξυπηρετούν καλύτερα τις ανάγκες του, υπάρχουν και έτοιμες λύσεις για αναγνώστες και εγγραφείς στο εμπόριο. Οι αναγνώστες RFID χειρός (Handheld RFID scanner)[5] είναι έτοιμες συσκευές που περιέχουν όλες τις απαραίτητες λειτουργίες που χρειάζεται ένας χρήστης για να μπορεί να γράφει και να κάνει ανάγνωση των πληροφοριών που είναι αποθηκευμένες στις ετικέτες. Επιπλέον, με το συγκεκριμένο είδος αναγνώστη ο χρήστης έχει την δυνατότητα να έχει πάνω του την συσκευή μπορώντας έτσι να την χρησιμοποιεί όπου και αν βρίσκεται. Τέλος, υπάρχουν USB αναγνώστες – εγγραφείς RFID[6] οι οποίοι μπορούν να συνδεθούν με τον υπολογιστή του χρήστη και αφού αυτός κατεβάσει το αντίστοιχο πρόγραμμα μπορεί να διαβάσει και να γράψει στα RFID tags.

|  |  |
| --- | --- |
|  |  |
| (A) | (B) |
|  |  |
| (C) | (D) |

Σχήμα 1.2: (A) RC522 RFID reader/writer. (B) PN532 NFC – RFID reader/writer. (C) Handheld RFID scanner. (D) USB desktop RFID scanner.

### 1.1.3 Χαρακτηρίστηκα συστημάτων RFID

Η τεχνολογία RFID χρησιμοποιήθηκε ως συσκευή διοδίων από τον Mario Cardullo το 1973 μέχρι να καταχωρηθεί ως πατέντα το 1983[2] από τότε μέχρι και σήμερα η συγκεκριμένη τεχνολογία συνεχώς εξελίσσεται. Η βασική χρήση της είναι η αποθήκευση και η μετάδοση της αποθηκευμένης πληροφορία. Η κάθε ετικέτα έχει μοναδικό αναγνωριστικό (unique id) καθιστώντας δυνατή την αναγνώριση εξαρτημάτων, υλικών, εμπορευμάτων λιανικής, ζώων συντροφιάς ή άλλων ειδών. Τα συστήματα υπολογιστών μπορούν να χρησιμοποιήσουν το αναγνωριστικό για να ενημερώσουν αυτόματα τις εγγραφές σε συνδυασμό με άλλα δεδομένα όπως τοποθεσία, θερμοκρασία, ημερομηνία και ώρα. Η αυτόματη αναγνώριση είναι ένα ακόμα χαρακτηριστικό το οποίο βοήθησε στην διάδοση της τεχνολογίας αυτής, καθώς είναι δυνατόν οι ετικέτες να επισυνάπτονται σε αντικείμενα και να μεταδίδουν το αποθηκευμένο αναγνωριστικό στους αναγνώστες. Οι ετικέτες μπορούν να τοποθετηθούν σε σταθερά σημεία εσωτερικά ή εξωτερικά των εμπορευμάτων, ενώ οι αναγνώστες μπορούν να στηθούν σε συγκεκριμένες τοποθεσίες εντός μιας μονάδας επεξεργασίας ή μιας αποθήκης, με τον τρόπο αυτό κάθε φορά που κάποιο στοιχείο περνάει, ο αναγνώστης θα σαρώνει αυτόματα το αναγνωριστικό. Εναλλακτικά, οι αναγνώστες μπορούν να ενσωματωθούν σε κινητά τηλέφωνα, επιτρέποντας τη σάρωση των ετικετών όπου και αν βρίσκεται η οντότητα. Επιπρόσθετα, αρχεία σε ράφια ή αντικείμενα που κινούνται σε μεταφορικούς ιμάντες, μπορούν να εντοπιστούν απλά τοποθετώντας σωστά τη συσκευή ανάγνωσης. Ωστόσο, η ικανότητα ανάγνωσης μιας ετικέτας εξαρτάται από έναν συνδυασμό παραγόντων όπως ο τύπος της ετικέτας και ο αναγνώστης, το υλικό της επιφάνειας και το περιβάλλον. Ένας αναγνώστης RFID μπορεί να συλλέξει δεδομένα από μεγάλο αριθμό ετικετών ταυτόχρονα, με αποτέλεσμα την εξοικονόμηση χρόνου για παράδειγμα το περιεχόμενο μιας ολόκληρης παλέτας μπορεί να σαρωθεί καθώς εξέρχεται από την αποθήκη. Οι παθητικές ετικέτες RFID, όπως έχει προαναφερθεί ενεργοποιούνται από την πηγή ισχύος της συσκευής ανάγνωσης, ώστε να μην απαιτούν μπαταρίες για την χρήση τους. Αυτό σημαίνει ότι οι ετικέτες μπορεί να είναι πολύ μικρές ώστε να μπορούν να ενσωματωθούν σε μικροσκοπικές συσκευές. Έχουν επίσης πολύ μεγάλη διάρκεια ζωής και μερικές μπορούν να επιβιώσουν σε εξωτερικούς χώρους σε όλες τις συνθήκες για αρκετά χρόνια. Συνήθως, οι παθητικές ετικέτες χρησιμοποιούνται για στοιχεία που πρέπει να διαβαστούν εντός εύρους από 1mm έως περίπου 10m. Για μεγαλύτερες αποστάσεις ή πιο εξειδικευμένες εφαρμογές, μπορούν να χρησιμοποιηθούν ενεργές ετικέτες με τη δική τους πηγή ενέργειας, οι οποίες μπορούν να παραμείνουν ενεργές για έως και πέντε χρόνια ανάλογα με τη χρήση και τις περιβαλλοντικές συνθήκες. Τέλος, ένα ακόμα καθοριστικό χαρακτηριστικό το οποίο βοήθησε την τεχνολογία RFID να εδραιωθεί είναι η ανθεκτικότητα των ετικετών στις καιρικές συνθήκες.Αντίθετα με τα barcodes (γραμμωτοί κώδικες) οι ετικέτες RFID μπορούν να χρησιμοποιηθούν ακόμη και στα πιο απαιτητικά περιβάλλοντα[7].

## 1.2 Συστήματα μηχανικής όρασης

Η μηχανική όραση είναι ο τομέας εκείνος της τεχνητής νοημοσύνης (AI) που καθιστά δυνατόν για τους υπολογιστές και τα συστήματα να αντλούν πληροφορίες από ψηφιακές εικόνες, βίντεο και άλλες οπτικές αναπαραστάσεις και να προβαίνουν σε ενέργειες με βάση τις πληροφορίες αυτές[8]. Όπως κάθε εφαρμογή που στηρίζεται στην AI για να λειτουργήσει αποδοτικά χρειάζεται να έχει στην διάθεση της αρκετά δεδομένα ώστε να κάνει πολλές αναλύσεις των δεδομένων αυτών μέχρι να διακρίνει τις διαφοροποιήσεις και τελικά να πάρει κάποια απόφαση, έτσι και στην μηχανική όραση χρειαζόμαστε έναν τεράστιο όγκο δεδομένων.

### 1.2.1 Convolutional Neural Network

Οι περισσότεροι αλγόριθμοι μηχανικής όρασης για να υλοποιηθούν χρησιμοποιούν συνελεκτικά νευρωνικά δίκτυα ή αλλιώς γνωστά ως CNNs. Τα CNNs, όπως και τα νευρωνικά δίκτυα ανατροφοδότησης (feedforward neural networks), μαθαίνουν από τις εισόδους, προσαρμόζοντας τις παραμέτρους για να κάνουν μια επιτυχή πρόβλεψη. Το χαρακτηριστικό που καθιστά τα CNNs ξεχωριστά είναι η ικανότητα τους να αντλούν χαρακτηριστικά από εικόνες, για να το πετύχουν αυτό χρησιμοποιούν συνελεκτικά στρώματα (convolutional layers) και ομαδοποίηση (pooling).

Τα convolutional layers εφαρμόζουν μια σειρά από φίλτρα εικόνας στην εικόνα που έχουν λάβει ως είσοδο, η οποία αναπαρίσταται ως ένας πίνακας (matrix). Οι τελικές εικόνες που θα προκύψουν, παρουσιάζουν διαφορετικές εκδοχές της αρχικής εικόνας καθώς έχουν εξαχθεί διαφορετικά χαρακτηριστικά. Τα φίλτρα εικόνας που χρησιμοποιούνται λέγονται πυρήνες συνέλιξης (convolutional kernels).



Σχήμα 1.3: Πυρήνας συνέλιξης

Στο σχήμα 1.3 βλέπουμε πως λειτουργεί ο πυρήνας συνέλιξης. Το μπλε τετράγωνο αναπαριστά την εικόνα εισόδου, το σκιασμένο πλαίσιο αποτελεί τον πυρήνα συνέλιξης, ο οποίος είναι ένα σύνολο από βάρη που “περνά” πάνω από την εικόνα. Κάθε τιμή του μπλε πίνακα πολλαπλασιάζετε με το αντίστοιχο βάρος και έπειτα όλα τα νούμερα στον σκιασμένο χώρο προστίθενται μεταξύ τους ώστε να προκύψει ο πράσινος πίνακας, ο οποίος ονομάζεται χάρτης χαρακτηριστικών (feature map)[8]. Στη συγκεκριμένη περίπτωση μετά τον πολλαπλασιασμό θα προκύψουν τα νούμερα 0,1,0,2,6,0,0,2,6 που όταν αθροιστούν θα δώσουν τον αριθμό 17.

Συνήθως τα συνελεκτικά επίπεδα έχουν πολλαπλούς συνελεκτικούς πυρήνες με διαφορετικά βάρη ώστε να είναι σε θέση να δημιουργήσουν διαφορετικά feature maps. Έπειτα τα convolutional layers ακολουθούνται από μια συνάρτηση ενεργοποίησης (activation function) συνήθως αυτή η συνάρτηση είναι η ReLU. Η συνάρτηση ενεργοποίηση εισάγεται με σκοπό να δώσει στο νευρωνικό δίκτυο την ικανότητα να παρουσιάσει τα αποτελέσματα με μη γραμμική έκφραση, ώστε να είναι σε θέση συγκρίνοντας τα αποτελέσματα που προκύπτουν να επιλέξει το καλύτερο, βελτιώνοντας έτσι την ακρίβεια του[9].



Σχήμα 1.4: Η συνάρτηση ενεργοποίησης ReLU

Αν χρησιμοποιηθούν πολλοί πυρήνες συνέλιξης τα δεδομένα θα έχουν πολλές διαστάσεις (dimensions) οδηγώντας το νευρωνικό σε overfitting, δηλαδή το νευρωνικό θα τείνει να μην μπορεί να γενικεύσει. Με άλλα λόγια το δίκτυο θα έχει πάρα πολύ καλή απόδοση στα δεδομένα εκπαίδευσης αλλά χαμηλή απόδοση στα δεδομένα ελέγχου ή σε πραγματικά δεδομένα. Για να αντιμετωπιστεί αυτό, χρησιμοποιείται ομαδοποίηση. Η πιο συχνή μέθοδος ομαδοποιήσης που χρησιμοποιείται είναι αυτή της μέγιστης ομαδοποίησης (max pooling) η οποία, όπως φαίνεται και στο σχήμα 1.5, επιλέγει την μεγαλύτερη τιμή από μια γειτονία τιμών.



Σχήμα 1.5: Maxpool

### 1.2.2 Χρήσης μηχανικής όρασης

Η μηχανική όραση επιτρέπει στις μηχανές να προσομοιάσουν την ανθρώπινη αίσθηση της όρασης. Η τεχνολογία αυτή έχει βρει αρκετές εφαρμογές σε όλες τις βιομηχανίες όπως είναι για παράδειγμα η βιομηχανία μεταφορών, η βιομηχανία παραγωγής γεωργικών και κτηνοτροφικών προϊόντων και άλλες[10].

#### 1.2.2.1 Βιομηχανία μεταφορών

Βρισκόμενοι στο έτος 2022 αρκετά αυτοκίνητα έχουν κάποιου είδους αυτόνομη οδήγηση, κάποια χρησιμοποιώντας sonar είναι σε θέση να μεταβάλουν μόνα τους την ταχύτητα χωρίς ανθρώπινη παρέμβαση, άλλα διαθέτοντας κάμερες μπορούν να αναγνωρίσουν και να εμφανίσουν σε μια δευτερεύουσα οθόνη πληροφορίες όπως για παράδειγμα οδικά σήματα και να προειδοποιήσουν τους οδηγούς για πεζούς ή άλλα οχήματα που παρεκκλίνουν της πορείας τους κάνοντας την οδήγηση πιο ασφαλή και εύκολη.



Σχήμα 1.6: Μηχανική όραση στην αυτόνομη οδήγηση (Jeremy Cohen 2019)[11]

#### 1.2.2.2 Ιατρική

H μηχανική όραση βρήκε χρησιμότητα και στην ιατρική καθώς χωρίς την τεχνολογία αυτή οι γιατροί θα αναγκάζονταν να αφιερώνουν αρκετές ώρες στην ανάλυση δεδομένων που σχετίζονται με τους ασθενείς και στην διεκπεραίωση διοικητικών εργασιών. Πλέον αρκετά διαγνωστικά κέντρα και νοσοκομεία χρησιμοποιούν την τεχνολογία αυτή για να κάνουν μια πρώτη πρόβλεψη για την υγεία τον ασθενών σε εξετάσεις με οπτικές απεικονίσεις όπως στις αξονικές και στις ακτινογραφίες. Αφού τα μηχανήματα λάβουν την εικόνα του ασθενή την επεξεργάζονται με αλγόριθμους αναγνώρισης εικόνας (image recognition), οι οποίοι ως έξοδο θα δώσουν την αρχική εικόνα με κάποιες “σημειώσεις” πάνω τους, τις οποίες βλέπει ο θεράπων ιατρός και αναλόγως αποφασίζει για την διενέργεια συμπληρωματικών εξετάσεων ή την χορήγηση κάποιας θεραπείας στον ασθενή.

#### 1.2.2.3 Βιομηχανία κατασκευών

O κατασκευαστικός τομέας υιοθετεί με γοργό ρυθμό την τεχνολογία της μηχανικής όρασης αξιοποιώντας την για επιθεώρηση περιουσιακών στοιχείων υποδομής, για πρόληψη κινδύνου στον χώρο εργασίας ή για προγνωστική συντήρηση. Στο περιβάλλον των κατασκευών είναι σύνηθες τα υλικά να διαβρώνονται και να προκαλείται παραμόρφωση θέτοντας σε κίνδυνο την ζωή των εργατών και οδηγώντας το έργο σε μεγάλες καθυστερήσεις. Οι συσκευές υπολογιστικής όρασης παρακολουθούν τα εισερχόμενα δεδομένα από τα μηχανήματα μέσω καμερών και εντοπίζουν ελαττώματα και άλλες αλλαγές. Όταν εντοπίζουν ένα πρόβλημα, στέλνουν ένα σήμα στο σύστημα, επιτρέποντας στους ανθρώπους χειριστές να λάβουν διορθωτικά μέτρα πριν συμβεί κάποιο ατύχημα ή καταστραφεί ένα περιουσιακό στοιχείο.



Σχήμα 1.7: Μηχανική όραση στην βιομηχανία κατασκευών

## 1.3 Αναγνώρισης προσώπου

Το λογισμικό το οποίο χαρτογραφεί, αναλύει και επιβεβαιώνει την ταυτότητα ενός ατόμου από το πρόσωπο του είτε από φωτογραφία είτε από video καλείται αναγνώριση προσώπου[12]. H αναγνώριση προσώπου είναι μια εφαρμογή της μηχανικής όρασης που χρησιμοποιείται από τις αρχές επιβολής του νόμου και τις κυβερνήσεις[13] για να αναγνωριστούν εγκληματίες, αλλά και από ιδιωτικές εταιρίες οι οποίες θέλουν να προσφέρουν στους χρήστες τους ένα επιπλέον επίπεδο ασφαλείας στις συσκευές τους όπως είναι η Apple με το FaceID και η Microsoft με το Windows Hello ή εταιρίες και οργανισμούς που θέλουν να προστατέψουν της φυσικές τους εγκαταστάσεις ή μέρη αυτών από μη εξουσιοδοτημένη πρόσβαση.

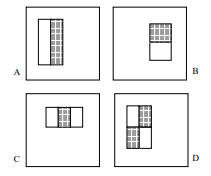
### 1.3.1 Αλγόριθμοι αναγνώρισης προσώπου

Ένας αλγόριθμος αναγνώρισης προσώπου είναι το κομμάτι ενός οποιουδήποτε συστήματος ή λογισμικού το οποίο είναι υπεύθυνο με τις λειτουργίες της ανίχνευσης ή της αναγνώρισης ενός προσώπου[14]. Οι ειδικοί χωρίζουν τους αλγορίθμους σε δύο κατηγορίες: α) τους γεωμετρικούς αλγορίθμους οι οποίοι προσεγγίζουν την αναγνώριση χρησιμοποιώντας διακριτικά χαρακτηριστικά και β) στις φωτομετρικές στατιστικές μεθόδους οι οποίες εξάγουν τιμές από την εικόνα εισόδου, τις οποίες συγκρίνονται με πρότυπα για την εξάλειψη των αποκλίσεων. Οι αλγόριθμοι εκτελούν τρεις βασικές εργασίες: α) ανιχνεύουν τα προσώπα σε φωτογραφία ή σε video πραγματικού χρόνου μέσω καμερών, β) υπολογίζουν το μαθηματικό μοντέλο του προσώπου και γ) συγκρίνουν το μοντέλο που προέκυψε με τα δεδομένα εκπαίδευσης ή με βάσεις δεδομένων ώστε να γίνει η αναγνώριση και η επαλήθευση ενός ατόμου.

#### 1.3.1.1 Haar Cascade

O αλγόριθμος Haar Cascade είναι ένας αλγόριθμός αναγνώρισης αντικειμένων και χρησιμοποιείται για να ταυτοποιήσει αντικείμενα και πρόσωπα σε μια εικόνα ή σε ένα video πραγματικού χρόνου. Προτάθηκε από τους Paul Viola και Michael Jones στην ερευνητική εργασία τους “Rapid Object Detection using a Boosted Cascade of Simple Features”[15] που δημοσιεύθηκε το 2001. H υλοποίηση αυτή βασίζεται στην μηχανική εκμάθηση όπου μια συνάρτηση cascade (καταρράκτη) εκπαιδεύεται από πολλές θετικές (εικόνες που έχουν το αντικείμενο) και αρνητικές εικόνες (όπου δεν υπάρχει το αντικείμενο), και έπειτα χρησιμοποιείται για την ανίχνευση σε άλλες εικόνες.

Ο συγκεκριμένος αλγόριθμος χρειάζεται αρκετές θετικές και αρνητικές εικόνες για να εκπαιδεύσει τον ταξινομητή. Για να εξαχθούν τα χαρακτηριστικά χρησιμοποιούνται τα χαρακτηριστικά Haar (Σχήμα1.8). Τα χαρακτηριστικά αυτά μοιάζουν στην λειτουργία τους με τους convolutional kernels. Κάθε χαρακτηριστικό προκύπτει από την αφαίρεση του αθροίσματος των pixels που βρίσκονται στην λευκή περιοχή από το άθροισμα αυτών που βρίσκονται στην σκιασμένη[16].



Σχήμα 1.8: Χαρακτηριστικά Haar. A-B) Ανίχνευση ακμών. C) Ανίχνευση γραμμών. D) Ένα τεσσάρων-ορθογωνίων χαρακτηριστικό

Για το κάθε πιθανό μέγεθος και πιθανή θέση του πυρήνα συνέλιξης θα υπολογιστεί ένα χαρακτηριστικό, με αποτέλεσμα στο τέλος της διαδικασίας ακόμα και για μια αρκετά μικρών διαστάσεων εικόνα να έχουν παραχθεί υπερβολικά πολλά χαρακτηριστικά. Σύμφωνα με τους ερευνητές μια εικόνα με διαστάσεις 24x24 θα μπορούσε να δώσει μέχρι και πάνω από 180.000 χαρακτηριστικά. Με σκοπό να περιορίσουν τα χαρακτηριστικά, εισήγαγαν την έννοια της ολοκληρωμένης εικόνας (integral image). Μια ολοκληρωμένη εικόνα στην τοποθεσία περιέχει το άθροισμα των pixels που υπάρχουν πάνω από την τοποθεσία: όπου είναι η ολοκληρωμένη εικόνα και είναι η αρχική εικόνα. Κάνοντας χρήση του ακόλουθου ζεύγους συναρτήσεων η integral image μπορεί να υπολογισθεί σε ένα πέρασμα της αρχικής εικόνα.

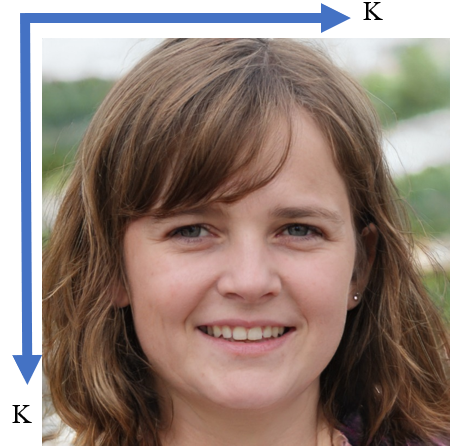
Έπειτα μια παραλλαγή του AdaBoost χρησιμοποιείται για να περιορίσει τα εξαχθέντα χαρακτηριστικά και να εκπαιδεύσει τον ταξινομητή. Για κάθε χαρακτηριστικό βρίσκεται το καλύτερο κατώτερο όριο (threshold) το οποίο θα ταξινομήσει τις εικόνες σε θετικές ή αρνητικές. Στην συνέχεια, επιλέγονται τα χαρακτηριστικά εκείνα που έχουν το χαμηλότερο δείκτη λάθους (minimum error rate), καθώς όταν αυτό ισχύει σημαίνει πως το χαρακτηριστικό αυτό είναι κατάλληλο για χρήση στην ταξινόμηση. Ο τελικός ταξινομητής προκύπτει από το άθροισμα (weighted sum) των ασθενών επιμέρους ταξινομητών. Καλούνται ασθενής διότι δεν μπορούν να ταζινομήσουν από μόνοι τους μια εικόνα. Σύμφωνα με τους συγγραφείς ακόμα και ένας μικρός όγκος χαρακτηριστικών της τάξης κάποιων εκατοντάδων είναι σε θέση να παρέχουν ακρίβεια στην ταξινόμηση περίπου 95%.

Το μεγαλύτερο ποσοστό της επιφάνειας μιας εικόνας δεν περιέχει κάποιο πρόσωπο (non-face region), οπότε οι Viola και Jones εισήγαγαν τον καταρράκτη ταξινομητών (cascade of classifiers). Τα χαρακτηριστικά ομαδοποιούνται σε διαφορετικά στάδια (stages) της διαδικασίας ταξινόμησης. Αν ένα παράθυρο αποτύχει στο πρώτο στάδιο τότε μπορεί να απορριφθεί και με αυτόν τον τρόπο δεν χρειάζεται να εξετασθούν τα υπόλοιπα χαρακτηριστικά. Αν περάσει τότε προχωράει στο δεύτερο στάδιο χαρακτηριστικών και η διαδικασία συνεχίζεται.

#### 1.3.1.2 Eigenfaces

O αλγόριθμος Eigenfaces δημιουργήθηκε από τους Matthew Turk και ο Alex Pentland στην εργασία τους “Face Recognition Using Eigenfaces”[17] το 1991 και στηρίχθηκε στην εργασία των L. Sirovich και M. Kirby με τίτλο “Low-dimensional procedure for the characterization of human face”[18] που προτάθηκε το 1987, για να αναπαραστήσουν αποτελεσματικότερα, οι πρώτοι, τις εικόνες προσώπου χρησιμοποίησαν την ανάλυση πρωτευουσών συνιστωσών (Principal Component Analysis – PCA). Ο αλγόριθμος αυτός προϋποθέτει όλες οι εικόνες που θα χρησιμοποιηθούν (για εκπαίδευση και οι προς εξέταση) πρέπει να έχουν όλες το ίδιο μέγεθος και να γίνουν ένα διάνυσμα (vector) ώστε τα μάτια και το στόμα να ευθυγραμμιστούν. Τέλος, για να είναι όσο το δυνατόν πιο επιτυχημένη και πρέπει να υπάρχουν αρκετές εικόνες ανά άτομα το οποίο θέλουμε να αναγνωρίσουμε.

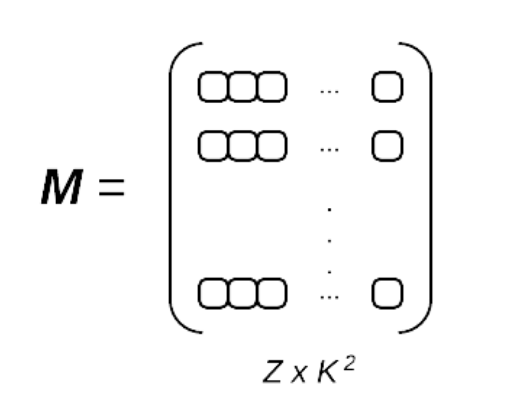
Έστω, μια τετραγωνική εικόνα διαστάσεων



Σχήμα 1.9: Εικόνα προσώπου[19]

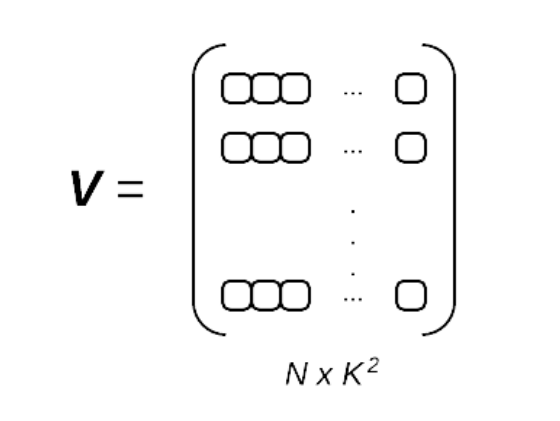
Όταν γίνεται χρήση του αλγορίθμου Eigenfaces κάθε εικόνα προσώπου αντιπροσωπεύεται σε κλίμακα του γκρι και έναν πίνακα στοιχείων pixel (bitmap of pixels) . Όπως αναφέρθηκε προηγουμένως, η εικόνα πρέπει να γίνει διάνυσμα (vector) με διάσταση , στην συγκεριμένη περίπτωση, όπου η εικόνα δηλαδή είναι τετραγωνική, το διάνυσμα θα έχει μήκος , αυτό επιτυγχάνεται κάνοντας flattening την κάθε εικόνα.

Αφού ολοκληρωθεί το flattening όλων των εικόνων της βάσης (dataset) σχηματίζεται ένας πίνακας με “ισοπεδωμένες” εικόνες όπου το σύνολο των εικόνων που υπάρχουν στο dataset. Ο πίνακας περιέχει ολόκληρο το dataset.



Σχήμα 1.10: Πίνακας ισοπεδωμένων εικόνων

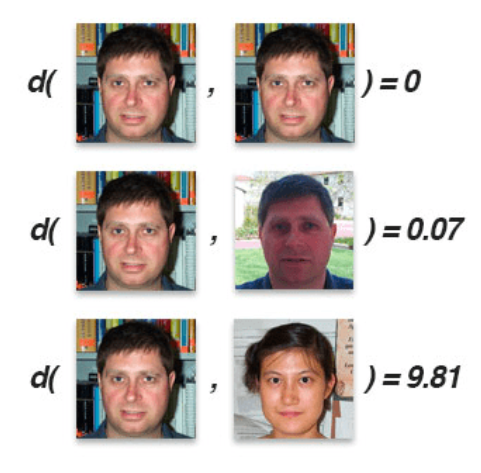
Στον πίνακα εφαρμόζεται η ανάλυση πρωτευουσών συνιστωσών όπως αναλύεται από την Lindsay Smith στο έγγραφο “A tutorial on Principal Components Analysis”[20]. To επόμενο βήμα μετά την δημιουργία του πίνακα είναι να υπολογιστεί ο μαθηματικός μέσος (mean) κάθε στήλης του πίνακα. Έπειτα, αφαιρείται ο μέσος από κάθε στήλη , η διαδικασία αυτή καλείται κεντράρισμα δεδομένων μέσου (mean-centering data). Αφού έχει κεντραριστεί ο πίνακας , υπολογίζεται ο πίνακας συνδιακύμανσης (covariance matrix) . Το επόμενο βήμα είναι η πραγματοποίηση μιας αποσύνθεσης ιδιοτιμών (eigenvalue decomposition) στον πίνακα συνδιακύμανσης με σκοπό να ληφθούν οι ιδιοτιμές και τα ιδιοδιανύσματα (eigenvectors) . Έπειτα, τα eigenvectors ταξινομούνται κατά φθίνουσα σειρά σύμφωνα με την τιμή , και επιλέγονται τα ιδιοδιανύσματα με τις αντίστοιχες μεγαλύτερες ιδιοτιμές. Τέλος, τα δεδομένα εισόδου προβάλλονται στον χώρο σύμφωνα με την εξίσωση (dot product), όπου *a* και *b* το πρώτο και δεύτερο διάνυσμα, *n* η διάσταση του διανυσματικού χώρου (vector space) και *­* είναι συνιστώσες των διανυσμάτων *a* και *b* αντίστοιχα. Ο χώρος οριοθετείται από τα κορυφαία ιδιοδιανύσματα που επιλέχθηκαν, και ονομάζονται ιδιοπρόσωπα (eigenfaces).



Σχήμα 1.11: Ο πίνακας ιδιοδιανυσμάτων V

Έχοντας τα eigenfaces, τα νέα πρόσωπα μπορούν να αναπαρασταθούν λαμβάνοντας το dot product μεταξύ της ισοπεδωμένης εικόνας και τον *N* ιδιοπροσώπων. Έτσι κάθε πρόσωπο μπορεί να αναπαρασταθεί ως ένας γραμμικός συνδυασμός των πρωτευουσών συνιστωσών για παράδειγμα:

Οι Sirovich και Kirby για να πραγματοποιηθεί η αναγνώριση προσώπου πρότειναν τη λήψη της ευκλείδειας απόστασης μεταξύ των προβαλλόμενων αναπαραστάσεων eigenface υλοποιώντας στην ουσία έναν ταξινομητή k-NN.



Σχήμα 1.12: Ευκλείδεια απόσταση στον αλγόριθμο Eigenfaces

Όσο μικρότερη η ευκλείδεια απόσταση, τόσο πιο όμοια είναι τα δύο πρόσωπα, η συνολική ταύτιση βρίσκεται παίρνοντας την ετικέτα που σχετίζεται με το πρόσωπο με τη μικρότερη ευκλείδεια απόσταση. Στο Σχήμα 1.12 η πρώτη απόσταση είναι 0 καθώς “συγκρίνεται” η ίδια φωτογραφία. Στη δεύτερη περίπτωση παρότι είναι διαφορετική εικόνα περιέχει το ίδιο πρόσωπο και για αυτό τον λόγο η ευκλείδεια απόστασή είναι 0.07 και στην τελευταία συγκρίνονται δύο τελείως διαφορετικά πρόσωπα οδηγώντας έτσι την ευκλείδεια απόσταση σε υπερβολικά μεγάλη τιμή 9.81. Συνήθως, πέρα από έναν αλγόριθμό k-NN για να αυξηθεί η επιτυχία της ταυτοποίησης χρησιμοποιείται κάποιος πιο προχωρημένος αλγόριθμος μηχανικής όρασης όπως είναι για παράδειγμα τα SVMs (Support Vector Machines)[21].

#### 1.3.1.3 Fisherfaces

Η γραμμική ανάλυση διακρίσεων (Linear Discriminant Analysis - LDA) εκτελεί μια μείωση διαστάσεων για συγκεκριμένη κλάση (class) και εφευρέθηκε από τον στατιστικολόγο R. A. Fisher, ο οποίος την χρησιμοποίησε για την ταξινόμηση λουλουδιών στην εργασία του “The use of multiple memeters in taxonomic problems”[22] το 1936. Προκειμένου να βρεθεί ο πιο κατάλληλος συνδυασμός χαρακτηριστικών που θα είναι σε θέση να πραγματοποιεί καλύτερο διαχωρισμό των κλάσεων η LDA μεγιστοποιεί την απόσταση μεταξύ των διαφορετικών τάξεων (between-class) και ταυτοχρόνως ελαχιστοποιεί την απόσταση μέσα στην ίδια κλάση (within-class). Με άλλα λόγια, τα αντικείμενα που ανήκουν στην ίδια κλάση δεν πρέπει να έχουν μεγάλη απόσταση μεταξύ τους ενώ οι διαφορετικές κλάσεις είναι όσο πιο μακριά γίνεται[29].

Έστω ένα τυχαίο διάνυσμα με δείγματα που προέρχονται από κλάσεις , . Ο πίνακες διασποράς και υπολογίζονται ως εξής:

.

Όπου *μ* είναι ο συνολικός μέσος και είναι ο μέσος κάθε κλάσης . Ο κλασσικός αλγόριθμός Fisher ψάχνει να βρει μια προβολή,

τέτοια ώστε να μεγιστοποιείται η απόσταση μεταξύ των κλάσεων. Η λύση στην βελτιστοποίηση του προβλήματος δίνεται λύνοντας το Γενικό Πρόβλημα Ιδιοδιανύσματος (General Eigenvalue Problem)

Τέλος, η κατάταξη του είναι στο μέγιστο , όπου *N* τα δείγματα και *c* οι διαφορετικές κλάσεις. Στην αναγνώριση προτύπων το πρόβλημα είναι πως αρκετές φορές τα δείγματα είναι μικρότερων διαστάσεων από τα δεδομένα εισόδου. Αυτό λύνεται αν στα δεδομένα και στα προβαλλόμενα (projecting) δείγματα εφαρμοστεί η ανάλυση πρωτευουσών συνιστωσών (PCA), έπειτα μπορεί να πραγματοποιηθεί LDA στα δεδομένα (reduced data).

.

O πίνακας μορφοποίησης (transformation matrix) *W*, ο οποίος προβάλει ένα δείγμα μέσα στον χώρο των *(c – 1)* διαστάσεων δίνεται από τον τύπο:

#### 1.3.1.4 Local Binary Pattern (LBP)

To 2004 από τους Timo Ahonen, Abdenour Hadid και Matti Pietikäinen προτάθηκε ένας νέος αλγόριθμος για αναγνώριση προσώπου στην εργασία τους με τίτλο “Face Recognition with Local Binary Patterns” [23].



Σχήμα 1.13: Πρώτο βήμα για αναγνώριση προσώπου με LBP

Το πρώτο βήμα δοθέντος ενός προσώπου από ένα σύνολο δεδομένων προσώπων για να χρησιμοποιηθεί ο LBP είναι να χωριστεί η εικόνα σε ισομεγέθη τετράγωνα όπως στο Σχήμα 1.13. Έπειτα, για κάθε τετράγωνο από τα 49 που υπάρχουν υπολογίζεται το Local Binary Pattern ιστόγραμμα (histogram). Ένα ιστόγραμμα δεν λαμβάνει υπόψη του τις χωρικές πληροφορίες (spatial information) σχετικά με τον τρόπο με τον οποίο τα μοτίβα είναι προσανατολισμένα μεταξύ τους. Υπολογίζοντας όμως ένα ιστόγραμμα για κάθε ένα από τα κελία είναι δυνατόν να κωδικοποιηθεί ένα επίπεδο τοπικών πληροφοριών όπως τα μάτια, η μύτη, το στόμα. Επιπλέον, η χωρική κωδικοποίηση δίνει την δυνατότητα να αποδοθούν βάρη στα ιστογράμματα που προέκυψαν για κάθε ξεχωριστό κελί παρέχοντας έτσι μεγαλύτερη διακριτική δύναμη σε πιο διακριτικά χαρακτηριστικά του προσώπου. Για παράδειγμα

|  |  |
| --- | --- |
|  |  |

Σχήμα 1.14: Σύστημα στάθμισης

To ιστόγραμμα LBP για τα λευκά κελία (πχ. μάτια) είναι κατά τέσσερις φορές πιο σημαντικά από τα υπόλοιπα. Η περιοχή του στόματος και των αυτιών (φωτεινό γκρι) συνεισφέρουν δύο φορές περισσότερο. Τα σκούρα γκρι (μέτωπο, εσωτερικά του μάγουλου) συνεισφέρουν μια φορά και τα μαύρα (μύτη, εξωτερικό του μάγουλου) απορρίπτονται και συνεισφέρουν μηδέν φορές.

Οι τιμές στάθμισης που αναφέρθηκαν βρέθηκαν πειραματικά από τους Ahonen, Hadid και Pietikäinen εκτελώντας αλγορίθμους συντονισμού υπερπαραμέτρων πάνω από τα δεδομένα εκπαίδευσης και επικύρωσης – δοκιμής. Τα σταθμισμένα LBP ιστογράμματα συνδέονται μεταξύ τους σχηματίζοντας το τελικό διάνυσμα χαρακτηριστικών.

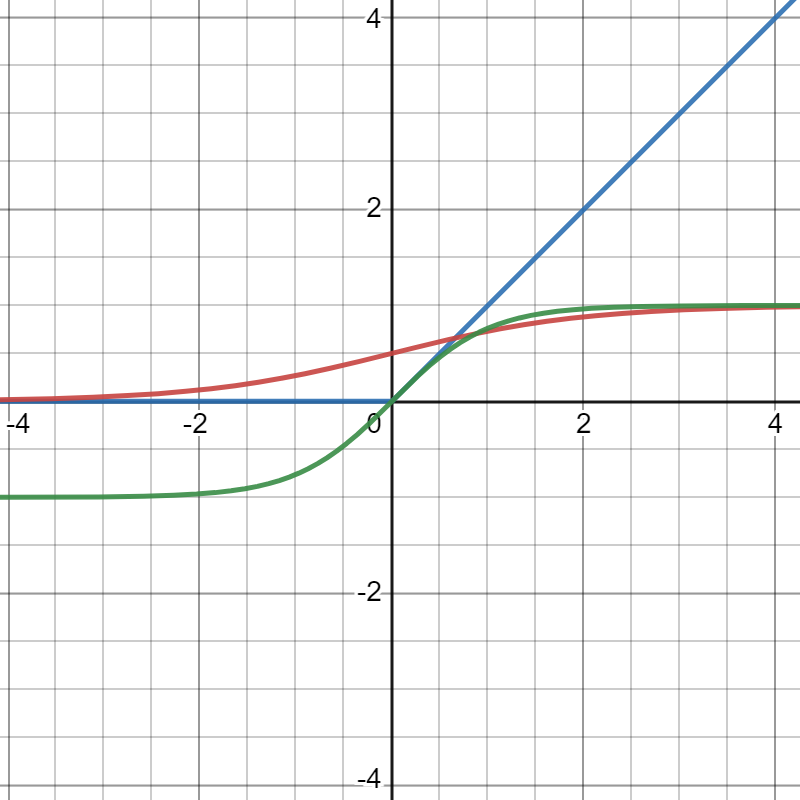
Η αναγνώριση προσώπου πραγματοποιείται χρησιμοποιώντας την απόσταση και έναν ταξινομητή πλησιέστερου γείτονα (nearest neighbor classifier) ως εξής: Αρχικά, παρουσιάζεται ένα πρόσωπο στο σύστημα και εξάγονται τα LBPs. Στη συνέχεια, ζυγίζονται και συνενώνονται με τον ίδιο τρόπο όπως τα δεδομένα εκπαίδευσης. Έπειτα, με εκτελείται ο k-NN με την απόσταση για να βρεθεί το πιο κοινό (όμοιο) πρόσωπο στα δεδομένα εκπαίδευσης. Τέλος, επιλέγεται το όνομα του ανθρώπου που σχετίζεται με το πρόσωπο που έχει την μικρότερη απόσταση[24].

#### 1.3.1.5 Αναγνώριση προσώπου με CNN

Ένα CNN (Convolutional Neural Network) είναι ένα νευρωνικό δίκτυο το οποίο έχει κρυφά επίπεδα (hidden layers) συνέλιξης και ομαδοποίησης (pooling) συνήθως εναλλάξ και προτάθηκε από τον μεταδιδακτορικό ερευνητή Yann LeCun την δεκαετία του 1980s χρησιμοποιώντας ως έμπνευση το neocognitron, ένα απλοϊκό νευρωνικό για αναγνώριση εικόνων που σχεδιάστηκε από τον Ιάπωνα ερευνητή Kunihiko Fukusima[25]. Το ποσοστό επιτυχίας του νευρωνικού δικτιού πέρα από την δομή που έχει σχετίζεται και με την συνάρτηση ενεργοποίησης (activation function) που έχει υιοθετηθεί. Η πιο συχνά χρησιμοποιούμενη συνάρτηση είναι η ReLU (Rectified Linear Unit), η οποία για όλες τις αρνητικές τιμές και το μηδέν επιστρέφει μηδέν ενώ για τις θετικές επιστρέφει δεκαδικό από το μηδέν μέχρι το 1 και δίνεται από των τύπο:

.

Πέρα από την ReLU αρκετά CNNs χρησιμοποιούν είτε την σιγμοειδή (sigmoid) ή την υπερβολική εφαπτομένη (hyperbolic tangent – tanh) με τύπους αντίστοιχα:



Σχήμα 1.15: Συναρτήσεις ενεργοποίησης ReLU (μπλε), Sigmoid (κόκκινη), Tanh (πράσινη)

Για την εκπαίδευση του νευρωνικού χρησιμοποιείται ο αλγόριθμός οπισθοδρόμησης (back propagation) ο οποίος με βάση μια συνάρτηση λάθους (error function) συνήθως το λάθος μέσου τετραγώνου (mean square error – MSE) ενημερώνει τα βάρη στις συνδέσεις του δικτύου ώστε να μπορεί να δώσει πιο ακριβή και σωστά αποτελέσματα. Ο τύπος για την MSE είναι:

όπου με *W* συμβολίζονται ο πίνακας βαρών, με *B* ο πίνακας προκαταλήψεων (bias) του nn (neural network), *aiL* και *tiL* υποδεικνύουν την πραγματική τιμή και την τιμή του *iοστού* νευρώνα στο επίπεδο εξόδου με *NL* νευρώνες που έδωσε το nn αντίστοιχα. Η έξοδος του *iοστού* νευρώνα στον *lοστό* επίπεδο μπορεί να υπολογισθεί ως:

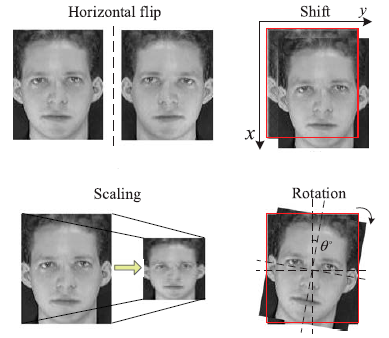
όπου *fil* είναι η συνάρτηση ενεργοποίησης, *bil* είναι το bias του *ail.* Tα βάρη σύνδεσης (connection weight) μεταξύ του *iοστού* νευρώνα του *lοστού* επίπεδο και το *jοστού* νευρώνα του προηγούμενου επιπέδου συμβολίζεται ως *wijl*. Τo νούμερο του *L-*επιπέδου του νευρωνικού συμβολίζεται με *l*. *Nl* είναι το πλήθος των νευρώνων του επιπέδου *l*. Αν χρησιμοποιηθεί η μέθοδος εμπροσθοδρόμησης (forward propagation) τα βάρη του νευρωνικού δίνονται:

Στη λειτουργία συνέλιξης, η λειτουργία μηδενισμού ακμών εφαρμόζεται για την εικόνα εισόδου ώστε να διασφαλιστεί η ορθολογικότητα των αποτελεσμάτων, δηλαδή τα στοιχεία της εικόνας που εισάχθηκε και του πυρήνα συνέλιξης μπορούν να σταθμιστούν (weight) και να αθροιστούν (summated) διαδοχικά. Το CNN μπορεί να θεωρηθεί ως μια κοινή διαδικασία εξαγωγής χαρακτηριστικών όπως τα περισσότερα νευρωνικά δίκτυα που χρησιμοποιούνται για την εξαγωγή χαρακτηριστικών.

Έστω, πως ως εισάγουμε μια τετραγωνική εικόνα με διαστάσεις 100x100 στο νευρωνικό, το οποίο έχει στο κρυφό επίπεδο 100 νευρώνες. Αυτό σημαίνει πως θα υπάρχουν 100x100x100 βάρη σύνδεσης μεταξύ του επιπέδου εισόδου και του πρώτου κρυφού επιπέδου αν όλα τα εικονοστοιχεία συνδέονται με όλους τους νευρώνες. Ο τεράστιος υπολογιστικός όγκος, που σε αυτή την περίπτωση αναπόφευκτα θα προκύψει, θα καθυστερήσει τον χρόνο εκπαίδευσης και πιθανότατα θα μειώσει την απόδοση του νευρωνικού. Αντίθετα, αν κάθε κρυφός νευρώνας είναι συνδεδεμένος σε ένα τοπικό πεδίο της εικόνας διαστάσεων για παράδειγμα 10x10 τα βάρη σύνδεση θα είναι 10x10x100. Τα βάρη σύνδεση μπορούν να μειωθούν και άλλο καθώς όλοι οι νευρώνες ενός πυρήνα συνέλιξης έχουν τα ίδια βάρη οδηγώντας σε περεταίρω μείωση του υπολογιστικού φόρτου και αύξηση της απόδοσης.

Τα στρώματα ομαδοποίησης βρίσκονται συνήθως μετά από τα επίπεδα συνέλιξης και χρησιμοποιούνται κυρίως για τη συμπίεση των δεδομένων χαρακτηριστικών εξόδου των συνελικτικών στρωμάτων. Μετά το επίπεδο συγκέντρωσης, τα βελτιωμένα αποτελέσματα εξόδου μπορούν να μειώσουν την πιθανότητα υπερεκπαίδευσης (over-fitting) στο νευρωνικό δίκτυο. Ουσιαστικά, η συγκέντρωση είναι μια διαδικασία μείωσης της εικόνας, η οποία μπορεί να ταξινομηθεί ως συγκέντρωση μέσης τιμής (mean-pooling), μέγιστη συγκέντρωση (max-pooling), επικάλυψη συγκέντρωση (overlapping-pooling), στοχαστική συγκέντρωση (stochastic-pooling) και συνολική μέση συγκέντρωση (global average pooling).

Το πρόβλημα που εμφανίζεται στο CNN είναι πως απαιτούν αρκετές φωτογραφίες – δεδομένα για να εκπαιδευτούν κάνοντας τα δύσχρηστα για απλές χρήσης όπου δεν μπορούν ή δεν επιτρέπεται η συλλογή πολλών δεδομένων. Λύση στο πρόβλημα αυτό δίνεται από τις μεθόδους αύξησης δεδομένων (Σχήμα 1.16), όπως είναι για παράδειγμα η μετακίνηση (shift), η αλλαγή μεγέθους (scaling), η περιστροφή κατά γωνία (rotation) και η οριζόντια αντιστροφή (horizontal flip).



Σχήμα 1.16: Η τέσσερις μέθοδοι αύξησης δεδομένων

Ένα CNN μπορεί να πετύχει μέχρι και 99.5% ακρίβεια στα δεδομένα ελέγχου[26].

## 1.4 User-friendly εφαρμογές

### 1.4.1 Διεπαφή χρήστη

Ο όρος διεπαφή χρήστη (User Interface - UI) αναφέρεται στην γραφική διάταξη μιας εφαρμογής και αποτελείται από όλα εκείνα τα στοιχεία με τα οποία αλληλεπιδρά ο χρήστης, όπως για παράδειγμα τα κουμπιά, την διάταξη των οθονών και κάθε είδους οπτικού στοιχείου αλληλεπίδρασης.

Την δημιουργία ενός καλού UI την αναλαμβάνουν οι UI designers, οι οποίοι είναι σχεδιαστές γραφικών, ασχολούνται με την αισθητική και εναπόκειται σε αυτούς να βεβαιωθούν ότι η διεπαφή της εφαρμογής είναι ελκυστική, έχει το κατάλληλο θέμα ώστε να ταιριάζει με τον σκοπό και αποφασίζουν για το φαίνεσθαι της εφαρμογής, επιλέγοντας για παράδειγμα συνδυασμούς χρωμάτων, τη γραμματοσειρά που θα φαίνεται το κείμενο. Επιπλέον, πρέπει να αποφεύγουν τα περιττά στοιχεία και να είναι ευανάγνωστη η γλώσσα που χρησιμοποιείται στις ετικέτες και στα μηνύματα έτσι ώστε η διεπαφή να διατηρηθεί απλή και κατανοητή. Δημιουργούν μια διεπαφή που χαρακτηρίζεται από συνέπεια (consistency), αυτό επιτυγχάνεται με την χρήση κοινών στοιχείων τα οποία κάνουν τους χρήστες να αισθάνονται πιο άνετα και καθιστώντας τους πιο αποτελεσματικούς. Μόλις ένας χρήστης μάθει πώς να κάνει κάτι, θα πρέπει να μπορεί να μεταφέρει αυτήν την ικανότητα σε άλλα μέρη του συστήματος (της εφαρμογή ή του ιστοτόπου). Εξετάζουν τις χωρικές σχέσεις μεταξύ των στοιχείων της σελίδας και δομούν τη σελίδα με βάση τη σημασία του περιεχομένου. Η προσεκτική τοποθέτηση των αντικειμένων μπορεί να επιστήσει την προσοχή του χρήστη στις πιο σημαντικές πληροφορίες και μπορεί να βοηθήσει στη σάρωση και την αναγνωσιμότητα της εφαρμογής[27].

### 1.4.2 Εμπειρία χρήστη

Η εμπειρία χρήσης, θετική ή αρνητική που θα έχει ένας χρήστη, (User Experience - UX) από την εφαρμογή καθορίζεται από τον τρόπο που αυτός αλληλεπιδρά με την εφαρμογή. Η εμπειρία χρήστη καθορίζεται από το πόσο εύκολη ή δύσκολη είναι η αλληλεπίδραση με τα στοιχεία της διεπαφής χρήστη που έχουν δημιουργήσει οι σχεδιαστές UI.

Οι σχεδιαστές UX εστιάζουν στο να κατανοήσουν τις ανάγκες, τις ικανότητες και τους περιορισμούς των χρηστών με απώτερο σκοπό την δημιουργία μιας εύκολης και απλής για αυτούς διεπαφής. Επιπλέον, λαμβάνουν υπόψη τούς επιχειρηματικούς στόχους του ομίλου που διαχειρίζεται το έργο. Οι βέλτιστες πρακτικές UX προωθούν τη βελτίωση της ποιότητας της αλληλεπίδρασης του χρήστη και της αντίληψης του για την εφαρμογή και οποιεσδήποτε σχετικές υπηρεσίες[28].

## 1.5 Επιλογές

Για τις ανάγκες αυτής της πτυχιακής εργασίας θα χρησιμοποιηθούν παθητικές ετικέτες RFID καθώς δεν χρειάζεται να εκπέμπουν συνέχεια ή σε μεγάλες αποστάσεις το σήμα τους. Σημαντική σημείωση οι παθητικές ετικέτες που χρησιμοποιήθηκαν είναι NFC που είναι ένα υποσύνολο της τεχνολογίας RFID με συχνότητα 13.56MHz (RFID HF)[30]. Επιπλέον, για την ανάγνωση – εγγραφή των tags έχει αγοραστεί έτοιμη λύση USB RFID/NFC scanner για ηλεκτρονικό υπολογιστή.

Έπειτα από δοκιμές εντοπίστηκε ένα πρόβλημα με το λειτουργικό σύστημα των Windows τα οποία διαθέτουν κάποιες υπηρεσίες (services) ώστε να λειτουργούν σωστά. Ένα service με όνομα “Certificate Propagation” ήταν η αιτία η συσκευή ανάγνωσης να χάνει συχνά την σύνδεση με τον υπολογιστή και να μην λειτουργεί όπως θα έπρεπε[7]. Η συγκεκριμένη υπηρεσία ήταν ο πρόγονος του σύγχρονου “Windows Hello!” και χρησιμοποιούνταν μέχρι την εμφάνιση αυτού για να βοηθά στην πιο εύκολη και γρήγορη σύνδεση του χρήστη σε ένα προστατευμένο υπολογιστή (password protected computer) αντιγράφοντας τα πιστοποιητικά του χρήστη και της ρίζας (root certificates) από έξυπνες κάρτες (RFID/NFC cards) στο χώρο αποθήκευσης πιστοποιητικών χρήστη (user’s certificate store), και ανίχνευε πότε μια έξυπνη κάρτα διαβαζόταν από έναν αναγνώστη, και αν ήταν αναγκαίο εγκαθιστούσε τον αντίστοιχο driver smart card Plug&Play. Η συγκεκριμένη υπηρεσία πρέπει να απενεργοποιηθεί στα σύγχρονα συστήματα για να δουλέψουν σωστά οι περισσότεροι αναγνώστες/εγγραφείς RFID (RFID readers/writers)[8].

# Βιβλιογραφία

1. Cardullo, M., & Parks, W. (1973). *Transponder apparatus and system* (U.S. Patent No. 3,713,148). U.S Patent and Trademark Office. <https://patentimages.storage.googleapis.com/4a/63/c4/1e14dedfdb7bd2/US3713148.pdf>
2. A. (2020, April 9). *Genesis of the Versatile RFID Tag*. RFID JOURNAL. <https://www.rfidjournal.com/genesis-of-the-versatile-rfid-tag>
3. Lagkas, T. (2021, September). *Wireless Sensor Network* [Slides]. Dropbox. <https://www.dropbox.com/sh/lqcltjb5smwyvbq/AADr9rdDo9LZzJI1iwtvTvaya?dl=0&preview=JAUNTY_O2_3_Wireless_Sensor_Networks.pptx>
4. Playful Technology. (2018, December 5). *RFID Roundup!* [Video]. YouTube. <https://www.youtube.com/watch?v=98GXrixOM4c>
5. Amsler, S., & Shea, S. (2021, March 31). *RFID (radio frequency identification)*. IoT Agenda. <https://internetofthingsagenda.techtarget.com/definition/RFID-radio-frequency-identification>
6. JT308 RFID ID Card Reader. (n.d.). Hellas Digital. Retrieved April 15, 2022, from <https://www.hellasdigital.gr/electronics/sensors/rfid-sensors/jt308-rfid-id-card-reader/>
7. RFID Features: *Key Features of RFID* - Corerfid. (2020, December 20). RFID Systems for Manufacturing, Assets, Lifting & Logistics. <https://www.corerfid.com/rfid-technology/what-is-rfid/rfid-features/>
8. Taylor, M. (2021, December 14). *Computer Vision with Convolutional Neural Networks - The Startup*. Medium. <https://medium.com/swlh/computer-vision-with-convolutional-neural-networks-22f06360cac9>
9. W. Hao, W. Yizhou, L. Yaqin and S. Zhili, *"The Role of Activation Function in CNN,"* 2020 2nd International Conference on Information Technology and Computer Application (ITCA), 2020, pp. 429-432, doi: 10.1109/ITCA52113.2020.00096.
10. Rizzoli, A. (2022, February 7). *27+ Most Popular Computer Vision Applications in 2022*. V7. <https://www.v7labs.com/blog/computer-vision-applications>
11. Cohen, J. (2019, November 25). *Computer Vision applications in Self-Driving Cars.* Becoming Human: Artificial Intelligence Magazine. <https://becominghuman.ai/computer-vision-applications-in-self-driving-cars-610561e14118>
12. IBM. (n.d.). *What is Computer Vision?* | IBM. Retrieved February 27, 2022, from <https://www.ibm.com/topics/computer-vision>
13. Face Recognition. (2021, February 15). *Electronic Frontier Foundation*. <https://www.eff.org/pages/face-recognition>
14. RECFACES. (2021, March 25). *Understanding Facial Recognition Algorithms*. <https://recfaces.com/articles/facial-recognition-algorithms>
15. P. Viola, & M. Jones. (2001). *Rapid Object Detection using a Boosted Cascade of Simple Features*. Conference on Computer Vision and Pattern Recognition. <https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf>
16. OpenCV. (n.d.). *OpenCV: Cascade Classifier*. <https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html>
17. M. Turk and A. Pentland. (1991). *Face Recognition Using Eigenfaces.* Conference on Computer Vision and Pattern Recognition. <https://sites.cs.ucsb.edu/~mturk/Papers/mturk-CVPR91.pdf>
18. L. Sirovich and M. Kirby. (1986, November 10). *Low-dimensional procedure for the characterization of human faces*. <https://pyimagesearch.com/wp-content/uploads/2021/05/kirby_1987.pdf>
19. Wang, P. (n.d.). *This Person Does Not Exist*. This Person Does Not Exist. <https://thispersondoesnotexist.com/>
20. L. Smith. (2002, February 26). *A tutorial on Principal Components Analysis*. <http://www.iro.umontreal.ca/~pift6080/H09/documents/papers/pca_tutorial.pdf>
21. Rosebrock, A. (2021, May 12). *OpenCV Eigenfaces for Face Recognition*. PyImageSearch. <https://pyimagesearch.com/2021/05/10/opencv-eigenfaces-for-face-recognition/>
22. R. A. Fisher. (1936). The use of multiple measurements in taxonomic problems. <https://www.comp.tmu.ac.jp/morbier/R/Fisher-1936-Ann._Eugen.pdf>
23. T. Ahonen, A. Hadid & M. Pietikäinen. (2004). *Face Recognition with Local Binary Patterns*. 3021. 469-481. 10.1007/978-3-540-24670-1\_36. <https://www.researchgate.net/publication/221304831_Face_Recognition_with_Local_Binary_Patterns>
24. Rosebrock, A. (2021a, May 8). Face Recognition with Local Binary Patterns (LBPs) and OpenCV. PyImageSearch. <https://pyimagesearch.com/2021/05/03/face-recognition-with-local-binary-patterns-lbps-and-opencv/>
25. Dickson, B. (2020, January 6). What are convolutional neural networks (CNN)? TechTalks. <https://bdtechtalks.com/2020/01/06/convolutional-neural-networks-cnn-convnets>
26. Peng Lu, Baoye Song & Lin Xu. (2021). *Human face recognition based on convolutional neural network and augmented dataset*. Systems Science & Control Engineering, 9:sup2, 29-37, DOI: [10.1080/21642583.2020.1836526](https://doi.org/10.1080/21642583.2020.1836526)
27. User Interface Design Basics | Usability.gov. (n.d.). Usability. <https://www.usability.gov/what-and-why/user-interface-design.html>
28. User Experience Basics | Usability.gov. (n.d.). Usability. <https://www.usability.gov/what-and-why/user-experience.html>
29. OpenCV. (n.d.). OpenCV: Face Recognition with OpenCV. <https://docs.opencv.org/3.4/da/d60/tutorial_face_main.html>
30. The Difference Between NFC and RFID. (2020, April 21). Global Payments Integrated. <https://www.globalpaymentsintegrated.com/en-us/blog/2020/04/21/what-is-the-difference-between-nfc-and-rfid>
31. D. (2021, October 1). Certificate Propagation Service (Windows) - Windows security. Microsoft Docs. <https://docs.microsoft.com/en-us/windows/security/identity-protection/smart-cards/smart-card-certificate-propagation-service>
32. GoToTags. (2021, September 29). Windows Certificate Propagation Service. GoToTags Learning Center. <https://learn.gototags.com/nfc/software/windows/certificate-propagation-service>