Joint Postgraduate Program of Electrical and Computer Engineering (ECE) and & Computer Engineering and Informatics Departments (CEID)

HCI Master
Masters in Human-Computer Interaction – University of Patras

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Postgraduate Thesis
Development of an online platform for real-time facial recognition

Student
Evangelos Michos
Matr. Num: 1041552

Supervisor
Nikolas Karacapilidis, Professor

Rest members of the Supervising Committee
Christos Bouras, Professor
Kyriakos Sgarbas, Professor

Patras, June 2021

Acknowledgements

This thesis is a diploma thesis in the context of postgraduate studies in the Department of Electrical and Computer Engineering (ECE) of the University of Patras. The title of this dissertation is "***Development of an online platform for real-time facial recognition***".

However, before presenting the theoretical background for the subject we studied and the conclusions we came to, I feel the need to warmly thank my mentor, Prof. Nicolas Karacapilidis. His guidance was decisive, while his suggestions and useful comments were undoubtedly valuable supplies during the writing of this paper.

In addition, I would like to say a big thank you to Prof. Christos Bouras and Prof. Kyriakos Sgarbas, whose valuable knowledge, advice and many years of experience helped me greatly in completing my dissertation.

Finally, I would like to thank my parents and my sister from the bottom of my heart for the many months of support they have given me all this time. Undoubtedly, if I did not have their contribution and substantial support in this endeavor of mine, then I would not have successfully completed my postgraduate studies.

Michos Evangelos

"The measure of a man is what he does with power"

Plato

Abstract in English

This Master Thesis targets upon extending the existing work of facial recognition algorithms regarding accuracy and efficiency and at proposing an online platform that can be used from the police forces towards effective real-time human recognition. The main features of the platform will include a) inserting, editing and deleting user and criminal information and b) searching for criminals based on their picture through a livestream camera feed and identify them. The platform will support two different types of users: a) Police employees in the headquarters/precincts and b) Police Administrators, with a higher level of access and also responsible for database maintenance. Regarding the facial recognition algorithm, the approach is to use and extend the Haar Cascade algorithm for real-time recognition, which is widely considered one for the most efficient and used algorithms for that cause. The development of the website is envisioned to be developed following the Model-View-Controller (MVC) architectural pattern, separating the platform in three different logical components. As for the criminal identification, it will made possible through image pattern recognition between the provided criminal's image and snapshots of identified faces from the livestream feed. The platform will include a live feed section, accompanied with different options for video filters, enabling the user to select the best filter, depending on the relevant situation of the physical surroundings for better recognition results. At first, extensive research will be conducted on both the state-of-the-art face recognition algorithm and the platforms dedicated for such causes and the capabilities and features they offer. After the initial research is conducted, the requirements definition will follow, alongside with personas and scenarios development. Direct communication with relevant stakeholders who already have experience from these systems will also occur (e.g. through interviews). At some point, the platform will be in an initial design stage (draft version), where the system should be meticulously studied and optimized towards its efficiency through a Heuristic Evaluation (HE). Then, we will continue with our system evaluation by conducting HE from usability experts and after gathering the relevant feedback and any other improvements to be implemented in the final version of the platform, the website will enter its final stage. A questionnaire will then be disseminated in order to collect further feedback. After gathering the results and performing the necessary statistical analysis, we will then know whether our approach was successful at its goals.

Abstract in Greek

Αυτή η Μεταπτυχιακή Διατριβή στοχεύει στην επέκταση της υπάρχουσας κατάστασης των αλγορίθμων αναγνώρισης προσώπου σχετικά με την ακρίβεια και την αποτελεσματικότητα και στην πρόταση μιας διαδικτυακής πλατφόρμας που μπορεί να χρησιμοποιηθεί από την αστυνομία για αποτελεσματική αναγνώριση σε πραγματικό χρόνο. Τα κύρια χαρακτηριστικά της πλατφόρμας περιλαμβάνουν: α) εισαγωγή, επεξεργασία και διαγραφή πληροφοριών χρηστών και εγκληματιών και β) αναζήτηση εγκληματιών βάσει της εικόνας τους μέσω μιας ροής κάμερας ζωντανής ροής και αναγνώρισή τους. Η πλατφόρμα θα υποστηρίζει δύο διαφορετικούς τύπους χρηστών: α) Αστυνομικοί υπάλληλοι στα κεντρικά γραφεία και β) Αστυνομικοί διαχειριστές, με υψηλότερο επίπεδο πρόσβασης και επίσης υπεύθυνοι για τη συντήρηση της βάσης δεδομένων. Όσον αφορά τον αλγόριθμο αναγνώρισης προσώπου, η προσέγγιση είναι η χρήση και επέκταση του αλγορίθμου Haar Cascade σε λειτουργία πραγματικού χρόνου, ο οποίος θεωρείται ευρέως ένας από τους πιο αποτελεσματικούς και χρησιμοποιούμενους αλγόριθμους για αυτόν τον σκοπό. Η ανάπτυξη της ιστοσελίδας προβλέπεται να αναπτυχθεί σύμφωνα με το αρχιτεκτονικό σχέδιο Model-View-Controller (MVC), χωρίζοντας την πλατφόρμα σε τρία διαφορετικά λογικά στοιχεία. Όσον αφορά την αναγνώριση των κακοποιών, θα καταστεί δυνατή μέσω της αναγνώρισης μοτίβου εικόνας μεταξύ της εικόνας του εγκληματία και των στιγμιότυπων των αναγνωρισμένων προσώπων του από τη ζωντανή ροή. Η πλατφόρμα θα περιλαμβάνει μια σελίδα ζωντανής ροής, που θα συνοδεύεται από διαφορετικές επιλογές για φίλτρα βίντεο, επιτρέποντας στον χρήστη να επιλέξει το καλύτερο φίλτρο, ανάλογα με τη τρέχουσα κατάσταση του φυσικού περιβάλλοντος για καλύτερα αποτελέσματα αναγνώρισης. Αρχικά, θα διεξαχθεί εκτεταμένη έρευνα τόσο στον προηγμένο αλγόριθμο αναγνώρισης προσώπου όσο και στις πλατφόρμες που είναι αφιερωμένες σε τέτοιες αιτίες και στις δυνατότητες και τα χαρακτηριστικά που προσφέρουν. Μετά τη διεξαγωγή της αρχικής έρευνας, θα ακολουθήσει ο ορισμός απαιτήσεων, παράλληλα με την ανάπτυξη σχετικών personas και σεναρίων. Θα πραγματοποιηθεί επίσης άμεση επικοινωνία με σχετικούς ενδιαφερόμενους που έχουν ήδη εμπειρία από αυτά τα συστήματα (π.χ. μέσω συνεντεύξεων). Σε κάποιο σημείο, η πλατφόρμα θα βρίσκεται σε αρχικό στάδιο σχεδίασης (draft έκδοση), όπου το σύστημα θα πρέπει να μελετηθεί σχολαστικά και να βελτιστοποιηθεί προς την αποτελεσματικότητα του μέσω Heuristic Evaluation (HE). Στη συνέχεια, θα συνεχίσουμε με την αξιολόγηση του συστήματός μας διεξάγοντας HE από ειδικούς χρηστικότητας και αφού συγκεντρώσουμε τα σχετικά σχόλια και άλλες βελτιώσεις που θα εφαρμοστούν στην τελική έκδοση της πλατφόρμας, η σελίδα θα εισέλθει στο τελικό στάδιο. Στη συνέχεια, θα διανεμηθεί ένα ερωτηματολόγιο προκειμένου να συλλεχθούν περαιτέρω σχόλια. Αφού συγκεντρώσουμε τα αποτελέσματα και πραγματοποιήσουμε την απαραίτητη στατιστική ανάλυση, τότε θα γνωρίζουμε εάν η προσέγγισή μας ήταν επιτυχής στους στόχους της.

Contents

List of Figures

List of Tables

INTRODUCTION

Human-Computer Interaction (HCI) has been challenged in recent years due to the fact that advanced technology requires adopting new applications and investigations related to connecting with other disciplines, in order to enhance its theoretical knowledge. HCI researchers frequently design research prototypes based on theories from the cognitive and social sciences, anthropology and sociology in addition to computer science. Such researchers focus on HCI research and the analytic approaches and techniques in design practice. Human is in the heart of the HCI scope and the ability to recognize a human being relates directly to this.

In general, face recognition algorithms are often utilized in order to identify or authenticate a human through a camera feed. Accuracy and efficiency of such algorithms are becoming critical prerequisites for applying face recognition within a variety of application domains like surveillance, e-banking, e-government. However, face recognition algorithms must adapt to real-life uncertainties and noisy environments. In today's world, face recognition is more and more utilized in order to search and identify shoplifters, retail criminals, or people with a history of frauds or crimes. This means that their pictures can, after being processed, be matched against a large criminal database and prevent and reduce crime rates by identifying them from installed cameras. The ability to efficiently recognize individuals through a combination of their facial characteristics is certainly an important phenomenon and the goal of facial recognition is something already apparent on our everyday lives.

To this end, this Master Thesis targets upon extending the existing work of facial recognition algorithms regarding accuracy and efficiency and proposing an online platform that can be used from the police forces towards effective real-time human recognition. The main features of the platform will include a) inserting, editing and deleting user and criminal information and b) searching for criminals based on their picture through a livestream camera feed and identify them. The platform will support two different types of users: a) Police employees in the headquarters/precincts and b) Police Administrators, with a higher level of access and also responsible for database maintenance. The approach on facial recognition is to use and extend the Haar Cascade algorithm for real-time purposes, which is widely considered one for the most efficient and used algorithms for that cause. The development of the website will be developed following the Model-View-Controller (MVC) architectural pattern, separating the platform in three different logical components. As for the criminal identification, it will made possible through image pattern recognition between the provided criminal's image and snapshots of identified faces from the livestream feed. The platform will include a live feed section with different options for video filters, enabling the user to select the best filter, depending on relevant situation of the physical surroundings for better recognition.

At first, extensive research will be conducted on both the state-of-the-art face recognition algorithm and the platforms dedicated for such causes and the capabilities and features they offer. After the initial research is conducted, the requirements definition will follow, alongside with personas and scenarios development. Direct communication with relevant stakeholders who already have experience from these systems will also occur. At some point, the platform will be in an initial design stage (draft version), where the system should be meticulously studied and optimized towards its HCI efficiency through a Heuristic Evaluation (HE). Then, we will continue with our system evaluation by conducting HE from usability experts and after gathering the relevant feedback and any other improvements to be implemented in the final version of the platform, the website will enter its final stage. A questionnaire will then be disseminated in order to collect further feedback. After gathering the results and performing the necessary statistical analysis, we will then know whether our approach was successful at its goals.

In **Chapter 1**, readers can have a thorough take on the generic field of HCI. In this Chapter, we take a closer look on all the different definitions of HCI and how HCI is applied in today's world. We will learn about the benefits of applying HCI into the whole process of product development from start to end and the equivalent issues and challenges HCI brings along. Then, we will also gather insights on the expected future of HCI in our world.

In **Chapter 2**, we take a closer look on the field of Vision Computing. We will learn about its categorizations in many fields of today's world (e.g. robotics, neurobiology, signal processing etc.) and how Vision Computing can be applied practically (e.g. 3-D image recreation, social distancing tools, body pose definition etc.). Then, readers will dive into the whole spectrum of recognition algorithms and the relevant subsections that object recognition and face recognition offer.

In **Chapter 3**, we believe that the reader has gained the necessary background knowledge on the HCI and Computer Vision fields required to understand the next stage, which is the aggregation of all phases involved to produce the finalized product, which is the real time face recognition platform. Readers will understand why we conducted interviews, how the personas and scenarios are tied with the development, why the requirements are necessary and after the statistical analysis of all information gathered, we will present the draft version of the platform. On this platform, we also explain how the HE was conducted and how the platform was updated (towards the better) towards its final version.

In **Chapter 4**, we thoroughly present the finalized version of our platform, explaining how the platform changed from its draft version. The final stage of the development was the questionnaire dissemination, which will allow useful insights on whether the human-centered approach during the entire development phase was actually a success or not.

In **Chapter 5**, we conclude this thesis on critically listing all the limitations, challenges and issues encountered and to be generally encountered with such approaches. A relevant future work section also exists, that offers insights on how such platforms can be exploited and extended with many more features to be implemented.

Chapter 1. The field of Human-Computer Interaction

1.1 Introduction

Nowadays, computers and portable devices around us should be designed towards understanding the fact that future user will have very specific tasks in mind before using them, tasks that can be seamless in regards to their everyday work. This can only achieved if such system designers actually know how to think in terms of end user tasks and how they can transform such knowledge and information towards a final product. Yet, a very serious challenge arises, regarding teaching the notion of designing such products for people. Designers are (of course) already human people and users of the development product, so isn't this enough intuitive to help them design the final product? Why should such designers be taught human design principles, since they (as developers) have already a vast majority of using such platforms and at a certain amount, developing such platforms [1]?

The field of *Human-Computer Interaction (HCI)* strives towards dealing with this exact issue of the designer's training throughout the development process. The interfaces designers are tasked to produce are not tasks that can just be "fixed" by applying HCI methodologies at the very end of the development stage (if applied). Undoubtedly, it is not an easy task for designers to constantly design robust and consistent systems that can tackle any type of user mistakes and this is why the development of the product's design should be an continuous integral part and not a last minute choice, supporting the actual user needs, preventing their mistakes, providing adequate underlying information and help them browse as intended. To this end, we can study how efficient such systems can eventually become, upon applying the HCI aspect into the design process.

If one wanted to research for the very first appearance of the HCI term, he would end up looking back in the 1980s, but the truth is that we actually must go way back and surprisingly end up in the Second World War (1939 - 1945). As stated by the author, since the war motivated research on the interaction between human beings and machines towards produce more effective and deadly weaponry, a new wave of research interest arose towards studying the machinery physical characteristics that could affect user performance, ending in the formation of the Ergonomics Research Society in 1949 [1]. As a result, the more widespread computers became in the following years after the war, the more research was put in effect, targeted at studying meticulously the interaction between a human being and the computer, regarding both physical, theoretical and psychological aspects related. In the 1980s, HCI surfaces with the advent of personal computing, like machines such as the Apple Macintosh and Commodore 64, which meant that for the first time, highly sophisticated electronic products were available to the general public for uses like word processing, gaming and accounting [2]. These products were no longer room-sized or extremely expensive, so consequently, a vital need was born to create interaction techniques that was respectively easy and efficient for less experienced everyday users.

As a term, this was first introduced as the *Man-Machine Interaction (MMI)* and was sometimes also referenced as *Human-Machine Interaction (HMI)* and *Computer-Human Interaction (CHI)*, but was later eventually renamed as HCI, since the term better described the ongoing interest in computers and the composition of user population. The HCI field depends on multiple disciplines of course, but is in *Computer Science (CS)* and systems design that it must be ultimately accepted as an area of concern. This is because for the rest of the disciplines, it can be seen as a specialism (albeit one that provides crucial input), yet for systems design, it should be considered as a very important aspect during the design process and from this point of view, it is safe to say that HCI involves the a) design, b) implementation and c) evaluation of interactive systems in the context of the user's tasks.

Later on in this chapter, we will see in detail the existing definitions of HCI, alongside with its application in the world of today and try to anticipate what the future holds for HCI, taking into consideration the challenges that emerge.

1.2 Defining the term

In general, when we discuss about HCI, it is a common mistake to only imagine a single user handling a personal computer. This is not true, as according to [1], the explanation of each of the components of the HCI term is defined as follows:

- *Human:*
  The human can in fact be anyone! He can be (of course) an individual person, or a big group of users, or a bunch of employees in an organization. What they call have in common is that they all deal with some part of a task embedded in the designed system, using the system to get the job done.
- *Computer:*
  The computer can be anything technological, ranging from a generic desktop computer, a laptop, a mobile phone, or even a smartwatch up to a very big control system or a big embedded system. The system could include non-computerized parts, including other people.
- *Interaction:*
  The interaction relates to any type of direct or indirect communication between the human and the computer. Direct interaction relates to feedback dialog and control through the completion of the task, whereas indirect interaction refers to batch processing or intelligent sensors able to control the environment.

Yet, even though the HCI term is universal, there exist multiple definition of the term. For example, according to [3]:

- "*Human–computer interaction (HCI) studies the design and use of computer technology, focused on the interfaces between people (users) and computers. Researchers in the field of HCI observe the ways in which humans interact with computers and design technologies that let humans interact with computers in novel ways*".

Moreover, Alan Dix (see [4]) himself states that:

- "*Human–Computer Interaction (HCI) is the study of the way in which computer technology influences human work and activities*".

The *Association for Computing Machinery (ACM)* defines HCI as (see [5]):

- "*a discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them*"

The more we search, the more we find relevant definitions overlapping one another, so to sum it up, we could say that HCI is simply the multidisciplinary field of study that focuses on the design of computer technology and more specifically, the interaction between humans and computers.

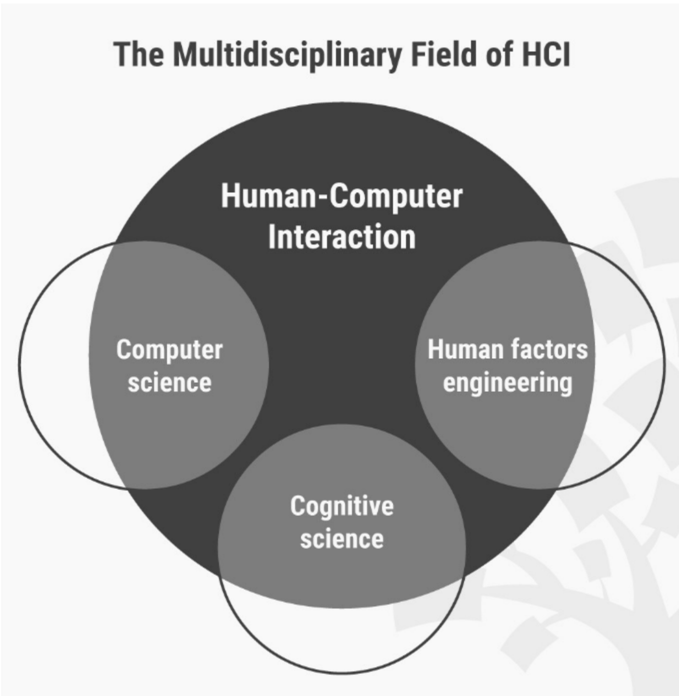


Figure 1. The Multidisciplinary field of HCI [2]

As already stated in **Chapter 1.1**, HCI first surfaced in 1980, yet we had instances of HCI appliances way before in our history. It was evident from its origins that HCI was inevitable to expand in order to incorporate multiple disciplines (e.g. computer and cognitive sciences and human-factors engineering). Soon, the HCI field ended up being the subject of intense academic research, as those who studied and worked in the field saw it as a means to popularize the idea that the interaction between humans and computers should resemble a human-to-human open-ended dialog. As a result, in the beginning, researchers first concentrated on improving the computer usability (e.g. focus on how easy it was to learn and use computers), however with the upcoming exponential rise of technologies of the era like the Internet and the smartphones, using computers increasingly moved away from only-desktop usage and researchers also embraced the mobile world. This is supported also from John M. Carroll, who was the author and founder of the field of HCI, who shared his views on the subject, saying that (see [2]):

- *"…it no longer makes sense to regard HCI as a specialty of computer science; HCI has grown to be broader, larger and much more diverse than computer science itself. HCI expanded from its initial focus on individual and generic user behavior to include social and organizational computing, accessibility for the elderly, the cognitively and physically impaired, and for all people, and for the widest possible spectrum of human experiences and activities*".

We already have saw multiple definitions and so far, we should have a concrete idea of what HCI is, but we have not yet discussed how HCI differentiates from other related fields. To start with, it differentiates from human factors and ergonomics, since HCI focuses much more on users working specifically with computers, instead of working with other kinds of machines or artifacts. HCI also focuses on implementing the software and hardware elements that support interaction between humans and computer, meaning that human factors is in fact a broader term. Additionally, regarding the human factors, HCI also differs in the sense that there is less focuses on repetitive tasks that are work-oriented, giving far less focuses on the physical stress and form produced from the user interface (e.g. keyboard design) [3].

1.3 HCI practitioners VS UX designers

HCI is quite a broad field, which means that it only makes sense to overlap with fields like for example, *User-Centered Design (UCD)*, *User Interface (UI)* and *User Experience (UX)* design.

UCD can be described as an iterative design process, where designers mostly put focus on both the users and their needs during all phases of the design process. Users are involved throughout the whole design process through different research and design techniques, so as to create more usable and accessible products for the end users. For the designers' *Point of View (POV)*, they are allowed to use a variety of different tools and methods of investigation (e.g. interview) and also generative methods (e.g. brainstorming) with the ultimate goal of understanding the needs of the end users. Regarding the process of applying UCD in the product, its iteration consists of four different phases [6].

- Phase One:
  Designers work together in groups in order to understand the context in which the end users will use the system.
- Phase Two:
  The designers altogether identify and list the user requirements.
- Phase Three:
  The designers, based on the requirements gathered, propose and design solutions
- Phase Four:
  The last phase concerns the evaluation stage, where designers assess the outcomes compared to the users' needs and check if the final system design behaves as intended.

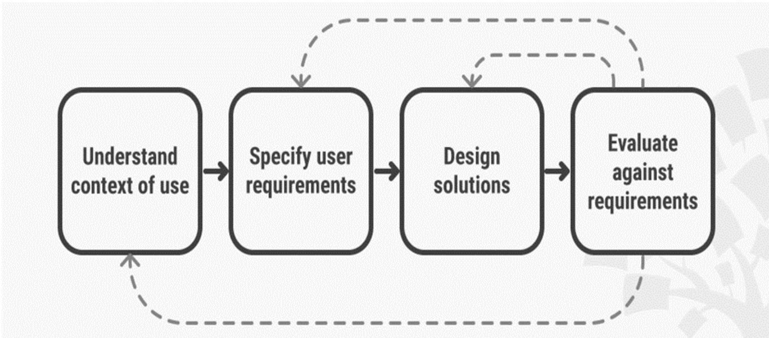


Figure 2. All four phases of the iterative design process [6].

From HCI's POV, the UI is the creative space where the interaction between the user and the machine happens. The ultimate goal is to provide efficient control of the machine's states from the human side, while at the same time, allow feedback from the machine towards the user to help him in the decision-making process. The final UI must be appealing to the end user and also make it easy and enjoyable for him/her to navigate safely and produce the desired results without issues, thus maximizing the usability. What this means is that the operate will need to provide some minimal input in order to achieve the desired result and at the same time, the machine must also minimize the unwanted results to the end user [7].

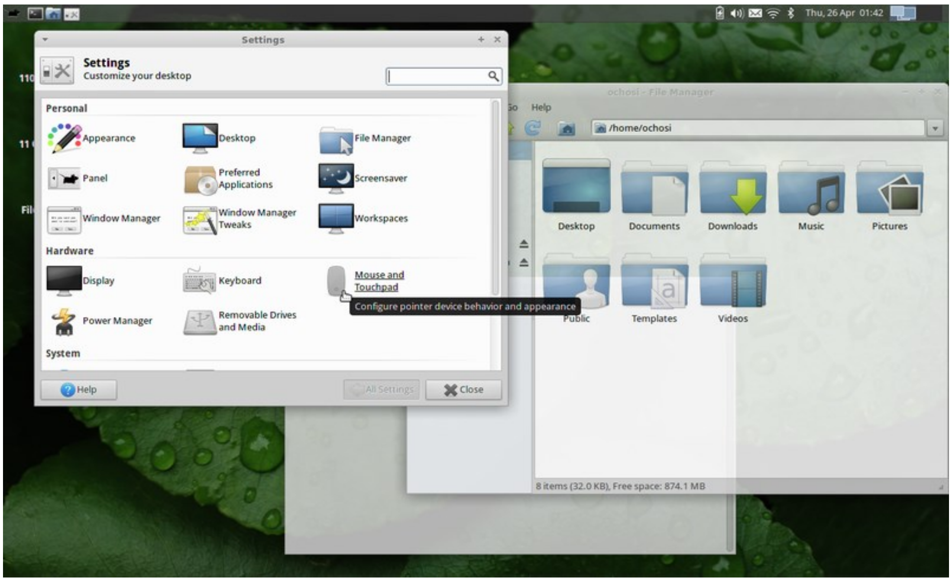


Figure 3. Example of a UI on a UNIX-like operating system [7]

UX design can be described as the process in which design project groups create systems that can provide user-relevant experiences to the end users, involving the design of the whole process of developing and integrating the final system, alongside the relevant aspects of branding, design, usability and functionality. In the market, those products that can offer great UX (e.g. a new smartphone) are designed with not only the product's consumption in mind, but also minding the entire process of acquiring, owning and troubleshooting. In the same context, UX designers don't just focus on creating products that are ultimately considered usable. In fact, they focus on all aspects around UX, like the users' satisfaction, efficiency and fun [8].

Many strongly believe that HCI was the forerunner to UX design, but in fact, there exist fundamental differences between HCI and UX design. The very first difference is that HCI researchers are more focused in academic and scientific research by developing empirical user understanding whereas on the other hand, UX designers mostly emphasize on creating products, services and platforms. In the long term, UX designers mostly consider products that have direct linkage to the HCI findings from users' mindsets and if we consider the wider topic span of HCI, they can choose from a large variety of resources.

Furthermore, it is no secret that UX designers lack the luxury of time HCI researchers usually have. As a result, UX designers must strive forward and stretch beyond the time constraints dictated in order to gain access in more academic and research findings. So, when they can achieve a high efficiency on this aspect of work, they can leverage key insights towards providing and designing the best designs for the end users and in fact, this collaboration with the HCI field can aid designers make meaningful changes in the greater society.

1.4 HCI Applications in our world

Today, computer devices are in virtually everything we touch in our everyday lives. Some of us might still have in our heads images of computer as very large rectangular objects on our work desks or as mobile phones in our pockets. But this notion we possess is not entirely correct, as computers are also in vehicles, thermostats, refrigerators etc. Inevitably, this increase in the size and functionality of computers continues to gather academic interest every year, with the field of HCI becoming an increasingly popular topic, having more HCI-related events in the ACM conferences than ever before. In the upcoming paragraphs, we will take a closer look at some of the most famous examples of HCI applications in the world (information listed below refers only to the most popular cases and is as accurate as listed in [9]).

Also the creator of the typewriter, Christopher Latham Sholes is considered to be the architect behind the design of the QWERTY keyboard layout that helped users at the time to avoid key jams back in 1873. QWERTY explains the order of the first six keys on the top left letter row of the keyboard. Sholes designed the layout in such a way that letter unlikely to often be paired together in typing, sat next to one another on the typewriter. When two keys met when using the typewriter, this could often cause jams and slow down speed. Although this is no longer an issue today, the QWERTY keyboard layout has remained.

The *Electronic Numerical Integrator and Computer (ENIAC)* is widely considered as the world's first general-purpose computer in 1946. ENIAC, also known as the "giant brain", was developed in order to calculate artillery firing tables for the *United States (US)* army forces, weighted more than 27.000 kgs and its first study was to measure the feasibility of the thermonuclear weapon. A few years later, in 1967, the world's very first gaming joystick is invented by Ralph Baueréckert, which was primarily used alongside the release of SEGA's arcade game called "Missile". This game was the very first that used a joystick with a fire button, as the game used the joystick to steer and shoot missiles into incoming playings in the player's screen.

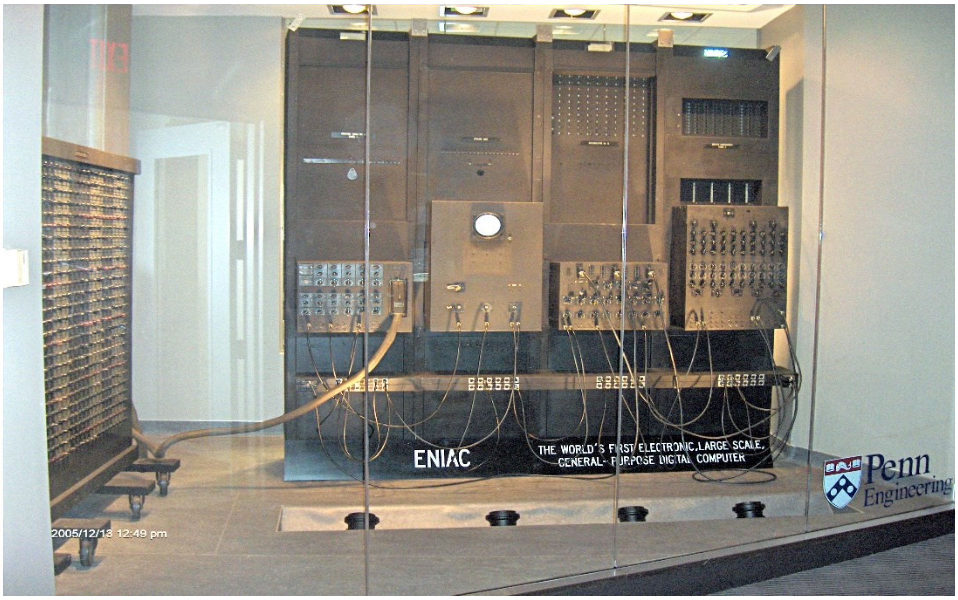


Figure 4. ENIAC, the world's first electronic large scale general-purpose digital computer [10]

A year later (1968), the field of *Virtual Reality (VR)* makes its very first appearance. Even though some question and dispute the history of VR, it is widely accepted that the Sword of Damocles was recognized as the first VR head mounted-display. Ivan Sutherland – the engineer behind this idea – argued that the initial concept was a *three-dimensional (3-D)* display that served the user with a perspective image that would change as the person moved in his surroundings. Even though it was quite primitive in nature, this device was later described as indeed a successful one, as it ensured that the device user could not tell the difference between the real world and the digital world presented in front of him.

Figure 5. A British Army Reserve soldier, wearing a VR headset [11]

In the 80s, the only noticeable HCI application was the invention of 3-D printing in 1984 and was developed so as to serve as a rapid prototyping technique. The device used a photosensitive resin that then polymerized using *Ultraviolet (UV)* light. 14 years later – in 1998 – the first smartwatch launches in the market, designed by Steve Mann. It was powered by the Linux *Operating System (OS)* and it was described by many as the "father of wearable computing", even though its specifications included a 16-bit processor and 128KBs of *Random Access Memory (RAM)*. Its basic functionality evolved around wireless communication both with computers and mobile devices.

In 2006, the Nintendo Wii gaming console comes into our everyday lives, providing natural UIs and gesture recognition mechanisms through its ability to track the player's physical motions and transform them into machine instructions interpretable by the system. The console used gyroscopes and controller-based accelerometers in order to sense the physical rotations, accelerations and tilting, something that was considered a revolutionary invention at the time for the HCI field. Eventually, this system was enhanced and transformed into XBOX Kinect, which is a motion sensing input system that allows gamers to interact with their computer without needing a dedicated controller.

Figure 6. The KINECT device, transforming motion gestures into interpretable information [12]

The last most popular example we will be referencing is the Google's Voice Search Application, which launched in 2011. Even though the application was originally called "Voice Action", it allowed users to interact with their phones through Google voice queries and originally released only on Android OS mobile phones. Back at the time, only six commands where available to the users, namely:

1. *Send text to*
2. *Call*
3. *Go to*
4. *Navigate to*
5. *Directions to*
6. *Map of*

Google Voice has been greatly upgraded to the point that nowadays, user can use their Wi-Fi connection to make voice calls using the application, which can also be installed in iOS or Windows OS.

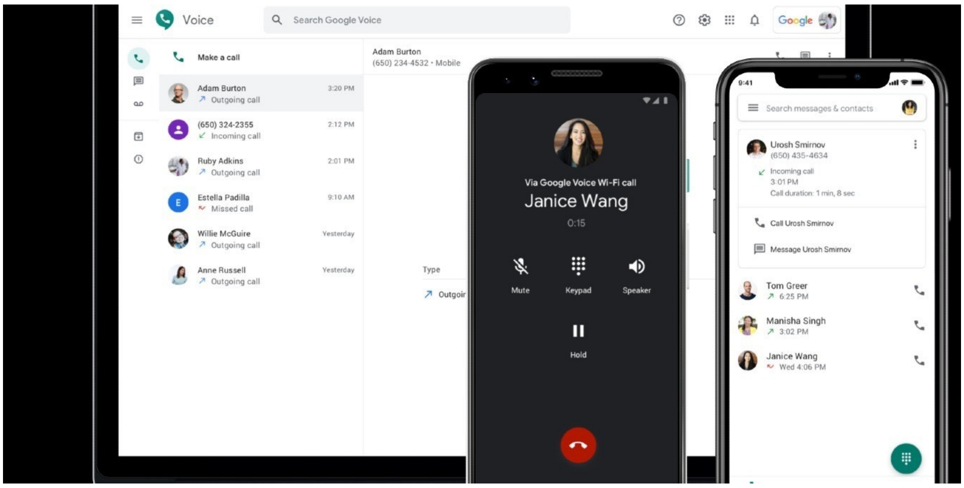


Figure 7. Using Google Voice for voice calls [13]

But enough of applications of the past. The real question that arises is what does the future hold for HCI applications in our everyday lives? A special aspect that has not been given enough emphasis so far but is currently been academically studied is the emotion-sensing field, where test and smell are considered as two under-explored senses in HCI and the future could bring innovative multisensory experiences inside the HCI scope. Indeed, there has been a lot of attention on emotion-sensing technologies that are fueled by *Artificial Intelligence (AI)*, which could very well read facial expressions, skin conductance and eye movements in order to improve concentration and manage stress. This can prove very useful in fatigue systems installed in smart cars of the future, where they can detect whether the driver is tired or on the verge of sleeping on the wheel and can prevent accidents through alarming sound feedback.

Except from the emotion-sensing field of HCI, a second aspect that HCI researchers study more and more for the upcoming future are the flexible smartphones. For example, Samsung was the first company to unveil a curved smartphone back in 2013, but the phone was not 100% flexible in order to allow bend input using haptic technology. But, in the future, this is a big possibility, as newly-created flexible displays use *Organic Light-Emitting diodes (OLED)* from far more flexible materials and since the material is more flexible, it can produce its own light, compared to ordinary mobile phones that are comprised of thick and heavy backlights made from *Liquid Crystal Display (LCD)* glasses and are thus inflexible.

Last but not least, *Brain-Computer Interfaces (BCI)* are becoming more and more popular and many believe that they will soon change the lives of those who unfortunately suffer from paralysis or amputations. Such a technology can support interaction with neural structured through decoding and translating information from a user's mind thoughts into actions and in the near future, this technology could soon be used for thought-to-text translation or even control prosthetic limbs. Such an example among the BCI research community is the Brainternet project that introduced the brain to the *Internet of Things (IoT)* through converting a user's brain into a node that can connect directly to the Internet! More specifically, it converts electroencephalogram signals into an open source brainwave live stream, using *Machine Learning (ML)* classifiers in order to classify the user's movement in the space. The project is still carried out by the University of Wits in Johannesburg, South Africa and those interested, more information can be found at [14].

1.5 Emerging Issues and Challenges

Due to the numerous technological advances around the field of HCI field, many new challenges are emerging accordingly. Such issues are inevitable to arise, as the emerging HCI landscape brings rapid technological evolutions towards more innovative, intelligent and efficient technologies that take into account the increased needs of the current society and the collective expectation from the HCI field. Many researchers have provided state-of-the-art reviews studies, attempting to sum up and categorize all the newly emerging issues and challenges. Some of them stand out, like the survey article in [15], where a grand total of 32 experts that belong in the *Human Computer Interaction International (HCII)* Conference Series, a group that started teaming up back in 2018 and offers the opportunity to all members to independently list and describe what they believe the HCI grand challenges are and how they impact the emerging technologies of the future. Without further ado, we present the findings of this survey and

Figure 8. The 7 grand HCI challenges, as summarized in [15]

1) <u>Human-Technology Symbiosis</u>

The term ‘symbiosis’ is a composite Greek word meaning ‘to live together’ and is an ideal term to be used upon discussing the issues of interactions between two counter-parts: the human part and the intelligent computers part. The very first concern is that even though everybody acknowledges the advanced technological intelligence and the potential towards created automated interfaces, humans should be kept in the loop, controlling and monitoring as supervisors the autonomous systems. Like for example, one could imagine the effect of the human control pilots have inside cockpits and flight decks. Pilots do exercise controlling everything in an outer-loop manner (e.g. monitoring and setting high-level goals) instead of an inner-loop manner (e.g. hands-on monitoring every minute). As a result, the pilot's tasks in such control systems inside the flight deck has changed immensely in the last years. But, to achieve meaningful human control – which has been defined as a research priority to achieve robust AI – the new features of automated systems must offer a) transparency, b) understandability and c) accountability, ultimately building a strong relationship of trust between the user and the automated system. Transparency must exist to allow UIs to make their own outputs visible to the system users, clarify the user's responsibility and promote better behaviors, since the term itself highly relates to system explainability. Since making a system visible does not necessarily mean that it is understandable from the human's side, it is a challenging task when systems are complex. As far as understandability is concerned, it will always be a difficult challenge for designers to keep users inside the loop to determine how the functions and mechanisms are provided and under what circumstanced. Automated systems that can evolve on their own, drawing conclusions from the data already gathered, also face the challenge of being understandable to the end users and can end up being characterized by opacity and unpredictability, which should be dealt accordingly.

2) <u>Human-environment interactions</u>

Undoubtedly, the interactions that will occur in the near future will radically shift and escalate users' location, posture, emotions and intentions, which will constitute input data to a large variety of both visible and invisible technological artifacts inside the environment. Robotic and autonomous systems are inevitable included in such technological environments and naturally, data will be sent and received from both the humans and the technological environment respectively, yet the digital counterpart will coexist with the physical one and will always try to enhance it. Such a challenge will enable evolving the existing design and evaluation methods in order to tackle and eventually provide future innovative technologies to their benefit. For example, some challenges include: a) the computer disappearing as a visible distinctive device, b) the risk of information overload and information inequality, c) the risk of high perceptual and cognitive demands, confusion of frustration, d) the ethics and privacy issues etc. It thus becomes clear that it is very important to understand how innovative interaction possibilities can enrich the surrounding environments that affect users and as technologies evolve exponentially day by day and become smarter and more pervasive, able to understand and anticipate human needs, users should be able to attribute personalities and internationalities to them. As a result, the challenge is to design such interactions as "*designing relations between humans and the world, and, ultimately designing the character of the way in which we live our lives*".

3) Ethics, Privacy and Security

In our days, it is getting more and more difficult to trust anyone and anything. Many factors, such as transparency, usability, collaboration and data privacy affect our trust in today's technological advances. As a result, a challenge today's HCI technologies need to cope with is their behavior towards being beneficial for people and be more than just fully functional or only addressing basic problems or just serving basic human needs. The main concerns refer to privacy, ethical and cybersecurity issues and ultimately, such challenges require the creation of new codes of ethics towards three different directions, namely ethics: a) by design, b) in design and c) for design. This means that is this new code of conduct, users' privacy should be better shielded against attempts of cyber-attacks that target the abundance of information today's (social media) networks have. The field of HCI should always strive towards supporting regulation activities targeting privacy, security, safety and ethics aspects of our lives. Examples of such concerns include: a) the right to not have any public personal identity, b) the right not to be monitored, c) the right to control his personal information that is shared across information systems, d) the right not to be forced to experience targeted marketing etc.

4) Well-being, health and eudaimonia

Today, healthcare technology is used more and more to make our lives easier. For example, in the context of healthcare, there exist many medical and tracking devices that without them, many doctor functionalities would not be possible. In the near future, we should expect more advanced technologies for more intelligent and efficient healthcare environments, like incorporating AI mechanisms and robotic functions. The most recent concrete example is the world's very first remote surgery in China using *Fifth Generation (5G)* networks that offer ultra-high speeds and almost minimal latency in 2019, where a doctor in the city of Sanya managed to insert a stimulation device inside the brain of a patient that suffered from Parkinson's disease, with the doctor being located in the city of Beijing, 3000 kms away from the patient [16]! But beyond the physical aspect of health, HCI technologies should also promote human well-being, focusing not only on the health aspects, but also on the psychological aspects of completing the intended life goals through helping to reduce the issues of stress, depression and mental illnesses. Concrete examples of such challenges include: a) sociological concerns that regard quantified bodies, b) loss of human free agency and autonomy, c) excessive control by caregivers up to violating basic human rights, d) difficulties in measuring human well-being and eudaimonia etc.

5) Accessibility and Universal Access

In a global scale, HCI always focused on the human aspects, on the emerging and innovative environments enhanced by the emerging technologies, in order to improve the quality of life various populations, like for example for people with special abilities. So for these people, the terms of accessibility and universal access are not something new, quite the opposite actually, they tend to gain increasing popularity and become pivotal for the prosperity of future generations. Of course, the advent of emerging environments entails some universal access challenges, like: a) the inadequacy of reactive approaches in distributed technological environments, b) the different perceptions users have on technology, c) the reluctance the industry show due the niche markets, d) the fact that some interaction techniques might prove prohibitive for some users etc. Yet, such inclusive environments also offer new opportunity up for the taking and to elaborate further, reactive approaches to accessibility will eventually fail at efficiently addressing the complexity and scalability requirements of the interactive environments of the future. The greater challenge remains still at large, which is no other than setting the correct requirements for more holistic approaches and ending up constituting a more researched direction for the field of HCI.

6) Learning and Creativity

It should be evident by now that in the field of HCI, the emerging technologies can support innovative learning styles in the contexts of education and creativity. Such technologies have the full potential to support emerging learning styles, due to their exponential growth in the last years and under the influence of the technological pervasiveness in our everyday lives. It is envisioned that the learning process will be personalized and training procedures in physical surroundings will be the exception, as learning will be happening when needed. Such a transformations will also affect tutors, as they will become more intelligent and efficient and adopt to new teaching styles outside of the classroom. The learning process can even evolve to other forms, such as AI-driven learning, targeting tutoring, collaborative learning etc. What this means for HCI is that its emerging technologies will shape the vast majority of the applications of the future, including the learning and teaching processes. The challenge lies behind the fact that such emerging environments must inevitably: a) be unobtrusive, b) avoid losing focus on the learning and creative processes and c) avoid discriminating the learners involved. In this regard, the technologies produced should blend in the support both digital and physical surroundings in the learning and education contexts.

7) Social Organization and Democracy

Last but not least, the aspect of social organization and democracy is a very important challenge to overcome. Current research has been focusing more and more on creating technologies that can help humans solve major problems (e.g. resource scarcity, climate change, poverty etc.). Aspects such as social justice and democracy should still be considered ideals not only desirable, but also actively pursuable and achievable. A wide variety of issues is emerging in this regard, such as: a) new forms of inequalities in the society (e.g. spatial inequalities), b) low involvement of populations of minorities, c) low civic engagement, d) manipulating information by monopolies, e) directed formulation of public opinion etc. Such issues can prove challenging to solve, cut also pave the way for new promises in the distant future. For now, current (and later on, future) decisions and practices will be the ones to decide whether the promises will be fulfilled or not.

1.6 The Future of HCI

In our times, the way we use and interact with devices and computers has significantly changed in the last years, with smartphones and tablets in a position to pave the way for multiple innovative emerging interactions techniques between humans and computers. This is simply because, the computational power such handheld devices has in our days has improved drastically over the last decade, offering more and more efficient interaction methods than ever. To grasp how big the technological gap is between the technology of today and the technology many years ago, the fact that our everyday phone has more computational power than all *National Aeronautics and Space Administration (NASA)* had back in 1969 on the moon landing, is both fascinating and frightening [17].

The first envisioned scenario for the future is the VR landscape, which looks more intriguing day by day and attracts a very large research interest. After Facebook managed to acquire the Oculus Rift head-mounted display and Microsoft also followed up and entered the game with HoloLens, it is a fierce competition that keeps riling up, since such displays might eventually end up replacing the traditional TVs we all have in our houses. These devices can offer a more entertaining experience through the digitalized immersive environments and users can taste the feeling of virtually being in the created surrounding environment. Such examples mostly apply in gaming experiences and online games where multiple gamers can use their avatar and immense themselves in 3-D created worlds, but the applications of HCI-related VR do not stop there. For example, one could imagine a world where we could drive our personal car wearing such displays and see 3-D crosses, signs, feedback and suggestions in front of us in real time. Alternatively, someone else could visualize being inside a restaurant and wearing the display, browsing across all available dishes and seeing images and reviews for the dishes in front of him. As a result, a gradual transition in needed from traditional to emerging innovative HCI VR technologies is needed in the near future, which posed a great challenge to both UI and UX designers, which is no other than the lack of familiarity with these new unique VR interactions mechanisms and in order to stay relevant in the future VR-driven world, such a transition that will truly change the way designers create and interact with interfaces is a must.

Furthermore, voice-guided UIs are also attracting more academic interest and UI/UX practitioners research them more and more. Voice interaction of course, will not be limited to what we have today (e.g. voice commands to browse the TV), but will radically expand and to be more specific, predictions have it that by the year 2025, the adoption rate of speech recognition will be more than 80% (according to [17])! If we give it a thought though, it totally makes sense, if one considers how much smaller the UI screens are today, as technology evolves at a rapid pace. And even though touch interfaces are envisaged to be around for a long time, voice-guided interfaces do promise deeper levels of HCI.

For instance, our university (University of Patras) has already taken significant steps towards this direction through voice-guidance projects like GuideMe [18]. The GuideMe project (an active partner in the project) acknowledged that there is an increasing demand for efficient indoor navigation systems, demand that mainly derive from smart cities, robots and visually impaired people and as a result, proceeded towards the design and development of a navigation system that assists people with special needs using an audio guidance system that incorporates input from a voice recognition system. In its core, the system consists of a device that provides the ability to navigate and route by voice commands, based on the device's location and orientation capabilities. The instructions are based on the device's location and orientation capabilities and this device shall be connected to the server via the user's mobile phone (Android). The suggested voice recognition system is used to guide the user through obstacles in indoor locations. Wireless connection between the user’s mobile phone and the mobile device are made available through low energy consumption *Bluetooth Low Energy (BLE)* protocols. The motivation of the paper is to improve two areas of social life of the blind people and general speaking people with special needs: convenience and security. Specifically, with the use of proposed system, users will feel more comfortable visiting public places such as airports, shopping malls, stations, etc., as they will be guided by the system to reach their destination. At the same time in case of emergencies involving both the user (accidents) and the building (fire, earthquake etc.), the system will inform the users of the exact location of the users, whilst also guiding them to the nearest exit. The ultimate goal is to increase the presence of the population with mobility or other problems in buildings by 20%. For more information on the project, its partners, deliverables/publications, news and others, please visit [18].

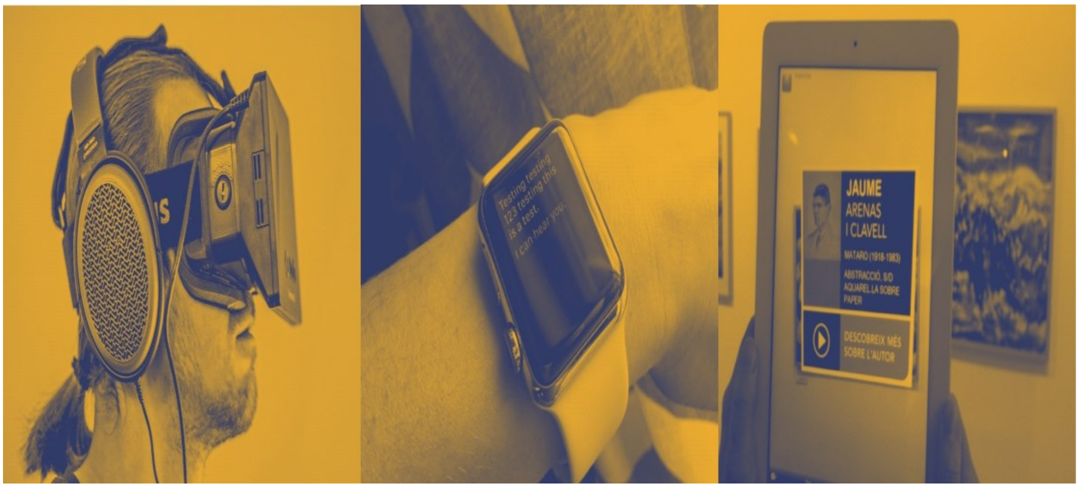


Figure 9. Envisioning the HCI of the future [17]

Speaking of size-shrinking devices, one could not forget to bring up the wearable devices, which are expected to have their prices significantly drop in the near future and will move past just showing the time and a person's pulse rates, become more efficient, adoptable, and user-friendly. The greater visions that exists here is "*injecting monitoring technology in our bodies*" and to further elaborate on this, the University of Patras has once again shown exemplary steps towards investing in the HCI areas of the future. One primary example is the WeSAR (the university is an active partner there too) project, where the project's aim is to design and develop a system for identifying and rescuing individuals, especially those belonging to population groups with a particularly high probability of being lost. The core of the system includes a wearable device, while the navigation algorithms on the mobile device play an important role in its effectiveness. The wearable device offers communication with *Base Stations (BSs)* that are located many miles away from the device location and the device uses low energy consumption communication methods in order to achieve its purposes. The target group includes all the cases where people could get lost, like: a) people with autism spectrum abilities, b) people suffering from dementia and c) children. For more information on the project, its partners, deliverables/publications, news and others, please visit [19].

Chapter 2. Vision Computing

2.1 Introduction

According to [20], the term Computer Vision can be defined as "*an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos*". From engineering's perspective, Computer Vision targets at firstly understanding and then automating tasks that the human visual system can do. This field is comprised of multiple tasks, like:

- Acquiring digital images
- Processing digital images
- Analyzing digital images
- Understanding analyzed images
- Extracting useful data

This field includes multiple well-known sub-categories, like: a) scene reconstruction, b) event detection, c) object tracking, d) learning, e) motion detection, f) visual serving, g) image restoration and many more. All these tasks effectively contribute towards creating either numerical or symbolic information. It is crucial to understand the context of the visual images projected, meaning that is it important to transform visual images to world descriptions that can be understood from the machine's side. From a scientific POV, Computer Vision related to the theory behind all known artificial systems capable of extracting meaningful information from any kind of digital input. Of course, the input can be any kind, ranging from short videos, multiple camera feeds and even multi-dimensional data (like a 3-D scanner). The goal is to apply the known theories to practice and construct smart and meaningful Computer Vision systems.

2.2 The sub-fields of Computer Vision

As already explained above, the field of Computer Vision is comprised of numerous sub-fields. In the following sub-sections, we will take a closer look at each sub-field at a time.

2.2.1 Signal Processing

The sub-field of Signal Processing relates to the focus given towards analyzing, modifying, filtering and synthesizing signal like sound, images or measurements. The methodologies which are incorporated in this sub-field can be used in order to optimize the signal quality as it traverses from the sender to the received. There are many aspects of its that can be improved, like augmenting the transmission storage efficiency, the quality of the received signal, the accurate measurement of the detected components in the final signal, the optimal sampling frequency in order to reconstruct the signal and many more. Signal Processing has numerous applications in our world, like: a) audio signal processing (e.g. speech/music), b) image processing (e.g. digital cameras), c) video processing (e.g. object detection), d) seismology, e) control systems, f) feature extractions (e.g. human face identification) and more. The most popular devices that support this field include analog/digital filters, samplers and signal compressors. This field can be used so that its methodologies and function can be extended so as to support processing two-variable signals or even multi-variable signals, which occurs more often in the recent years. Due to the fact that the very own nature of the input images, there exist multiple ways applicable in the Computer Vision field, yet many of them have no equivalent counterpart in the process of one-variable signal and alongside the perplexity and multidimensionality of the input signal, one could say that the sub-field of Signal Processing can self-create a new sub-field of each own, tackling this issue [21].

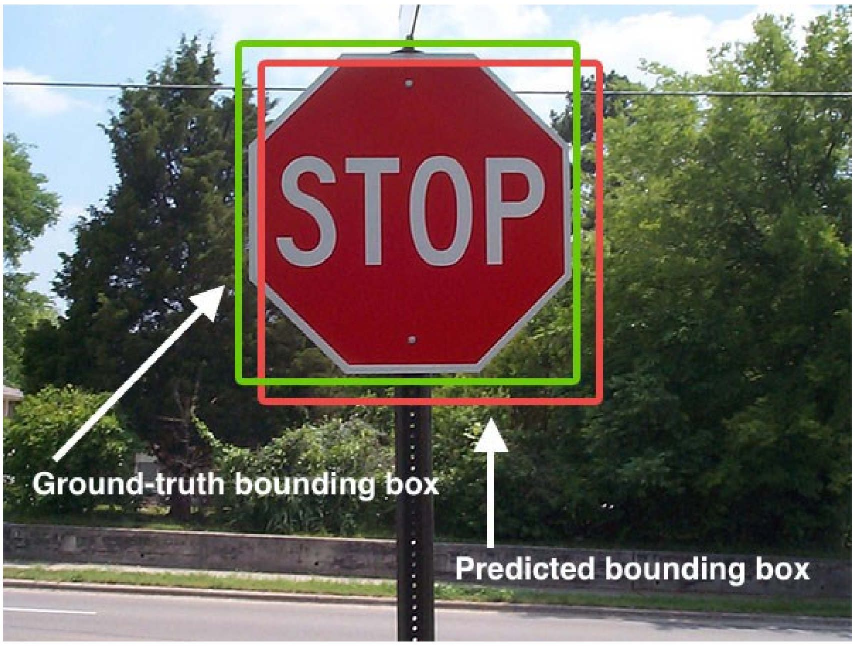


Figure 10. Object detection in Computer Vision [20]

2.2.2 Neurobiology

As a term, Neurobiology relates to "*the study of cells of the nervous system and the organization of these cells into functional circuits that process information and mediate behavior*" [22]. It is widely considered as a sub-category of the fields of biology and neuroscience, yet one would wonder how Computer Vision is involved with the hundred billion neurons inside the human brain. Therefore, to elaborate further, in the last century, there has been an extended academic research interest towards studying the eyes, the neurons and the brain neuron structures that are responsible for processing the visual input and interpreting, either the subject is a human being or some animal. Consequently, multiple but highly complicated descriptions have emerged related to how human or animal vision operates in order to interpret the physical surroundings. The outcome was the creating of a new sub-section within the spectrum of Computer Vision, where there exist highly intelligent and efficient AI systems that are capable of accurately mimicking the whole processing behavior of such biological systems, incorporating all the complexity levels required. Yet, even though the biological vision is an integral part of neurobiology and thus, of the Computer Vision field, they tend to differ on something, even though everyone agrees that information dissemination across these two fields is and will be fruitful for both. Biological vision (and more specifically, the field of biological vision) mostly puts emphasis towards analyzing, interpreting and modeling the physiological processing that are behind the visual understanding that humans and/or animals have, whereas the Computer Vision field studies and describes the applied methodologies in software/hardware levels, behind the AI system created.

2.2.3 Robotics

The technology of robotics is an interdisciplinary section of both science and engineering, devoted towards designing and putting to used physical mechanical robots. The replicated machines are called robots and are created as substitutes for human behavior. It is a field that is constantly gaining ground and as the overall world is expanding, all robotic constructions share some common characteristics. For example, all robots consist of some sort of mechanical production that helps them accomplish the tasks in their surrounding environments. All robots created require electrical components in their design in order to be able to control and power the machinery. Furthermore, all robots are developed in such a way that they incorporate computer programming (else software programming), as they would be unable to interpret the visual stimuli without written commands of software, telling the robot how it needs to react in each case presented to it. So, after having developed the mechanical counterpart of the human behavior, many of the functions required from robots is to be able to move into space, in order to either aid humans (e.g. robots for households), or complete any other given tasks (e.g. a robot sent to Mars to collect samples). In any case, robotic navigation is mostly intended and envisioned to relate to autonomous path planning or deliberation for robotic systems to navigate through an environment. But first, is crucial that a highly detailed representation of the surroundings is served as information to the robot in order for the navigation to be truthful and at least comparable to the navigation mechanisms of a human being [20] [23].

2.2.4 Solid-state Physics

This sub-field of Computer Vision relates to studying rigid matters or solids, through quantum mechanics, crystallography, electromagnetism, and metallurgy. In other words, it studies how large-scale properties of solid material end up, deriving from the properties of their atomic scale, thus forming a theoretical basis of materials science, applied mostly in transistors and semiconductors. It is considered a sub-section of Computer Vision, as the vast majority of systems included in this field rely on image sensors, that detect objects using electromagnetic radiation and to build such sensors, quantum physics are required. Furthermore, the process in which it is determined how light interacts (and is or not projected) with any given surface also requires an advanced level of physics. Optics is also a popular field where physics play an integral part and especially when designing optics systems [20] [24].

2.2.5 Honorable mentions

Even though the aforementioned subfields are considered the most popular ones, there exist many other subfields of Computer Vision that need to be mentioned and studied from a mathematical POV. It goes without saying that many sub-fields of Computer Vision require in-depth analysis of the received information, which can be translated as a prerequisite for knowledge on statistical analysis and geometrical theory. All this backend knowledge and theory can be devoted to implementing Computer Vision, like for instance, how existing mechanisms can be utilized and merged as software/hardware combinations, or how can they be altered to increase aspects like speed, accuracy or efficiency. Lastly, the field of Computer Vision is also apparent in fashion ecommerce, inventory management, patent search, furniture and even the beauty industry.

2.3 Putting Computer Vision into practice

To put it simply, Computer Vision is the technology that allows to traverse from the digital world and interact with the rea world and thus, is considered one of the hottest research field nowadays, become more and more integrated into our everyday lives [25]. In the following subsections, we will see how Computer Vision can be actively put into practice in our personal lives.

2.3.1 Defining Body Posing

The most commonly known definition of body posing is the Human Pose Estimation. This technique is based on accurately capturing the Human Pose Skeleton of a human being, which can be defined as a set of different coordinates that define a human's pose. A pair (two neighboring points) is considered as a limb. In overall, this technique can be used so as to identify and track the key points of a human's pose, given an input stream data (image or video). Human Pose Estimation can be applicable for: a) activity recognition for real-time sports analysis or surveillance system, b) for augmented reality experiences, c) robot training and d) animation and gaming.

One very indicative example of the academic research interest on Human Pose Estimation is the recent work published in 2020 regarding the 'DEEP' *Landing Error Scoring System (LESS)* [26]. The authors developed an injury-risk screening tool that can mostly be used in sports, aiming to evidence that if the LESS scores are automated using Deep Learning and Machine Learning to for video cropping and identification, the accuracy of the toll will be greatly enhanced, given as input different input videos. As input videos, two-dimensional (2-D) videos from double-leg and drop-jump exercises were passed to the LESS program to score calculation. The tracking and identification of the human posing (as see below in *Figure 11*) were made possible by using the OpenPose framework, which was the very first real-time Human Pose Estimation method that can all-together detect human bodies, hands, faces and feet using signal images as input. The framework was developed using a combination of the programming languages Python, C++ and a Unity plugin and is available as open-source on the Github website by visiting [27]. As a library, OpenPose is very much dependent on the CMU Panoptic Studio dataset. The most remarkable features of OpenPose include:

- Real-time 2-D multi-person keypoint detections
- Real-time 3-D single-person keypoint detections
- Calibration toolbox for estimation of distortion, intrinsic, and extrinsic camera parameters
- Single-person tracking for speeding up the detection and visual smoothing

For more information regarding the detailed mathematical model behind how the OpenPose real-time framework operates, please visit [28], where a detailed and thorough analysis is provided.

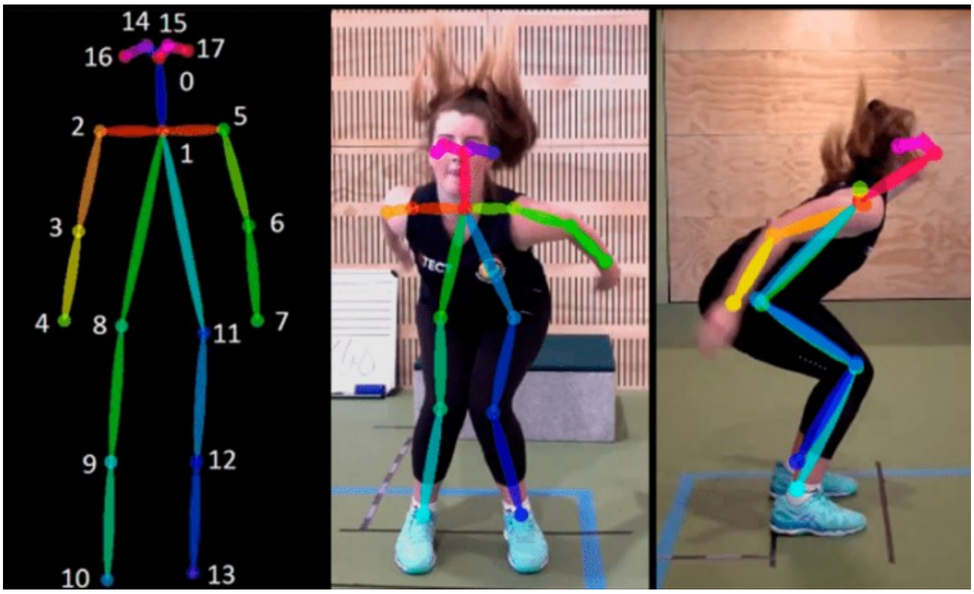


Figure 11. Detecting human posing using the OpenPose framework [26]

Of course, the OpenPose framework is not the only one out there for accurately detecting human posing. Such framework can also operate in today's browser, much like the PoseNet model, which is a standalone model, applicable in any of today's browsers through the TensorFlow package (the source code is available in Github in [29]). According to the documentation, PoseNet can be used to estimate either a single pose or multiple poses (see *Figure 12*). This means that there is a) a version of the algorithm that can detect only one person in an image/video and b) another version that can detect multiple persons in an image/video. These two distinct versions are both needed, because on the one hand, the single person pose detector is faster and simpler, but it also requires that only one subject is presented in the image at a time. The input is a *Red-Green-Blue (RGB)* image that is fed through a *Convolutional Neural Network (CNN)*. Afterwards, either a single-pose or multi-pose decoding algorithm is used to decode poses, pose confidence scores, keypoint positions, and keypoint confidence scores from the model outputs. For those interested in more information behind this real-time Human Pose Estimation, the developers/contributors of PoseNet suggest that readers visit [30]. In fact, as long as the readers of this thesis have a camera installed or built-in in their computer or laptop or any kind of portable device, the developers offer an official free live demonstration by visiting [31]!

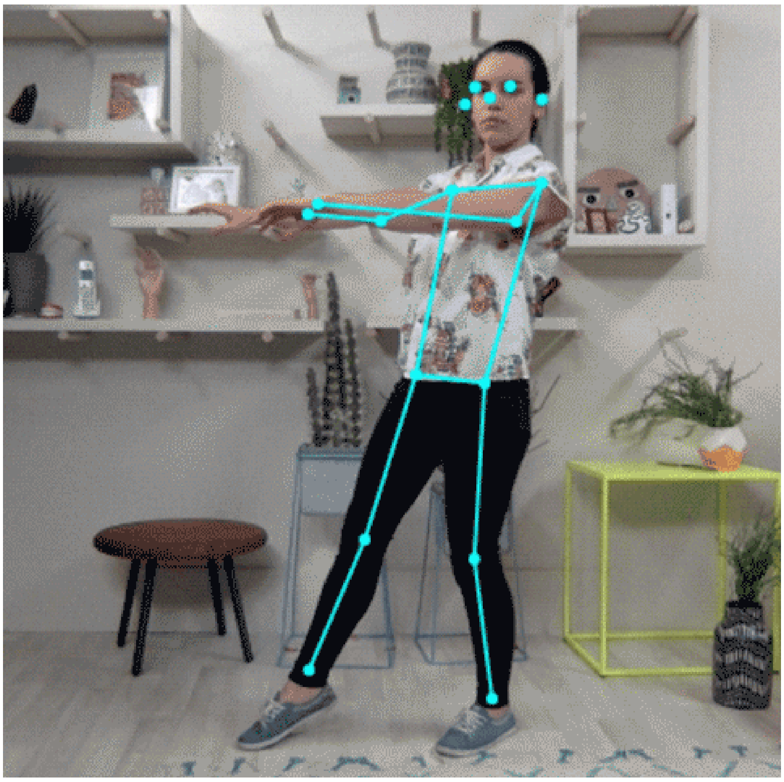


Figure 12. Detecting human posing via the PoseNet browser framework [30]

2.3.2 Image Transformation

Undoubtedly, application that allow us to easily swap faces in a picture or change the gender of the person in the picture are just some of the most scary things free-for-all technology can produce nowadays. For example, the Faceapp application is one of the most (if not the most) popular Computer Vision applications that within a few steps, can easily age or de-age a person, swap genders of a given picture and many more [32].

Most of the times, *Generative Adversarial Networks (GANs)* are used to this end, an approach that targets to generate picture modelling using Deep Learning techniques, like CNNs. More specifically, Generative modelling exists in the field of Machine Learning, involving the automatic discovery and learning of the regularities in the input data so that in can generate new examples that cold as well have been extracted from the original dataset. In fact, GANs can be described as a clever way to train a generative model, in which we classify the problem as a learning problem with two sub-categories: a) the generator model that we train to provide new examples and b) the discriminator model that classifies examples are real (from the original domain) or fake (generated from learning process). Those two models play part in a zero-sum game, until the discriminator model is fooled half the time. This means that the generator is creating plausible and believable examples, which is the goal through the learning process. So to this end, GANs are a rapidly changing field, that they keep on giving more and more believable generated examples through their generative models and cover a wide range of problem domains (mostly image-to-image translation). GAN's applications in today's world include:

- Image to image translation in style transfer and photo in painting
- Image super-resolution
- Text to image generation
- Image editing
- Semantic image to photo translation

As a result of the above, the generated photos are so photorealistic, that even people have trouble telling which of them are fakes and which are the real ones. The scary part in all this is not actually the fact that this is possible, but the fact that such tools are available to everyone with access to a computer or a portable device, allowing them to easily break any kind of ethics code by submitting images without approval of the depicted person. And while it might be ok to use actors that know and agree that their faces are out there publicly shown in movies, articles, videos and newspapers, but what if our face was added in the application's database by our friend, without our consent? How do we know what the application developers will do with all this abundance of images? How are we certain that this information will not be disseminated to third party companies? The truth is that we don't and this is an alarming and emerging issue that must be addressed from such companies in our days, so that our personal information remains safe and secure and we are not exploited without our knowledge or consent.

Figure 13. Easily changing the gender (male -> female) of actor Dave Batista using Faceapp [32]

2.3.3 Social Distancing Tools

By the time this thesis was being written, the whole world was already 1 year tackling the COVID-19 pandemic, forcing all of us to exercise personal hygiene (e.g. hand sanitizers), wearing face masks to avoid spreading germs and exercising social distancing from others in public places. To this end, Computer Vision can help tackle this issue by using it as a vital part towards tracking people in a specific geographical area and identifying whether these people are acting according to the health regulations of social distancing or not.

A social distancing tool is a software tool, capable of real-time detection and tracking and in our cause, it is used in order to check for social distancing violations by detecting each person using a surrounding box. The box is firstly green to signify that no rules have been violated so far by creating a circle with a radius of let's say 1.5 m surrounding the person. If this person comes too close to another person, then these circles will be one inside the other, meaning that the social distancing rules are violated and the borders turn from green to red. An accurate representation of this can be found below in *Figure 14*.

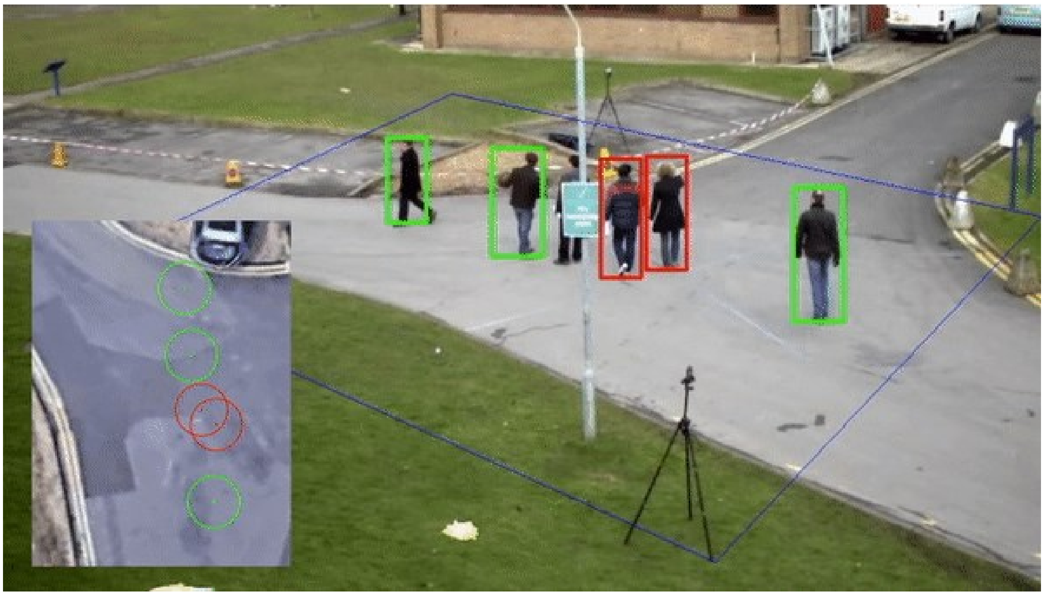


Figure 14. A Real-time social distancing tool camera feed [25]

Such tools (like the one presented above) are accurate enough on their own, yet they can of course be optimized further through Machine Learning techniques. In fact, there exist famous pre-trained object detection models, like *You Only Look Once (YOLO)* [34] and Mask R-CNN [35] to this end. For example, YOLO (currently at version 3 offering improved training and performances like multi-scale predictions) does not abide with the prior detection mechanisms, where they repurpose classifiers or localizers to perform detection and apply the model to an image at multiple locations and scales. High scoring regions of the image are considered detections. What YOLO does is that they apply a single neural network to the full image and this network divides the image into regions and predicts bounding boxes and probabilities for each region. These resulting bounding boxes are weighted by the predicted probabilities. An example of YOLO's detection outcome for object recognition can be seen below in *Figure 15*.

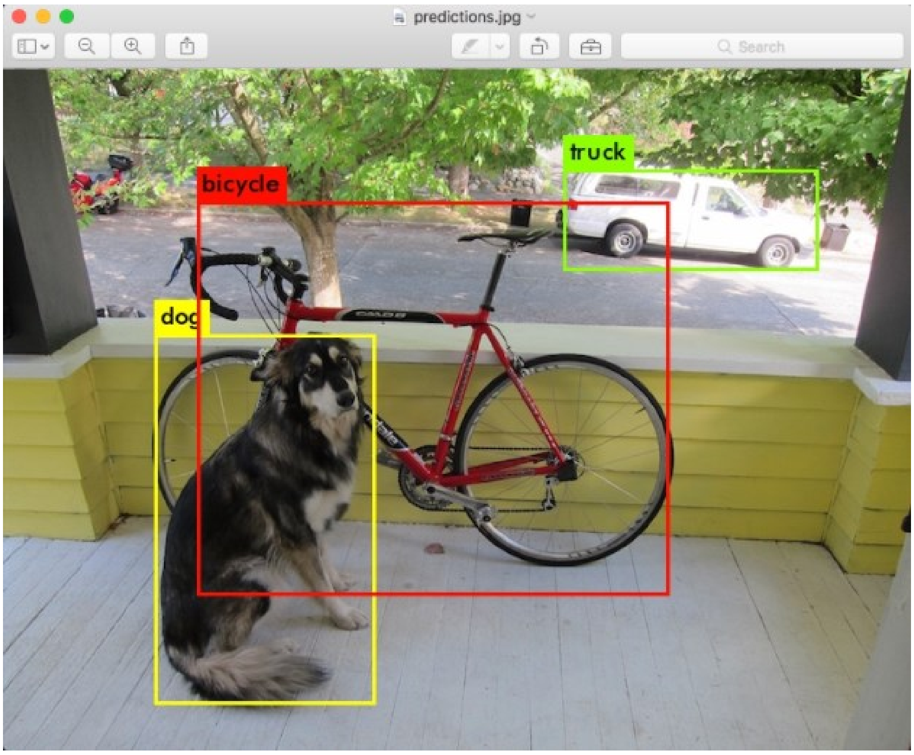


Figure 15. YOLO's object detection prediction outcome [34]

2.3.4 Generating 3-D models from 2-D input

Computer Vision can also be applied in the field of 3-D modelling, and more specifically, in the process of creating 3-D representations out of any 2-D input images. Many of us that use social media (e.g. Facebook), might have stumbled upon someone posting a 3-D image and wondering how did he have the software capability to do this on his/her own? The truth is that, he does not possess that knowledge. Facebook (and many other colossal companies of social media) have developed 3-D expanding features that can be applied on 2-D images that are posted by regular users. The problem lies behind the fact that a 2-D image can never capture the depth of the data, so as a result, the backend software used AI and Machine Learning in order to train a neural network how to best output 3-D structural representations out of simple 2-D photos users upload (even though these photos were taken with a single lens camera). Facebook's announcement specifically stated that "*This advance makes 3-D photo technology easily accessible for the first time to the many millions of people who use single-lens camera phones or tablets*". They also added that it will allow literally everyone to access to such devices the ability to experience decades-old family photos and other treasure images in a whole new 3-D way [36].

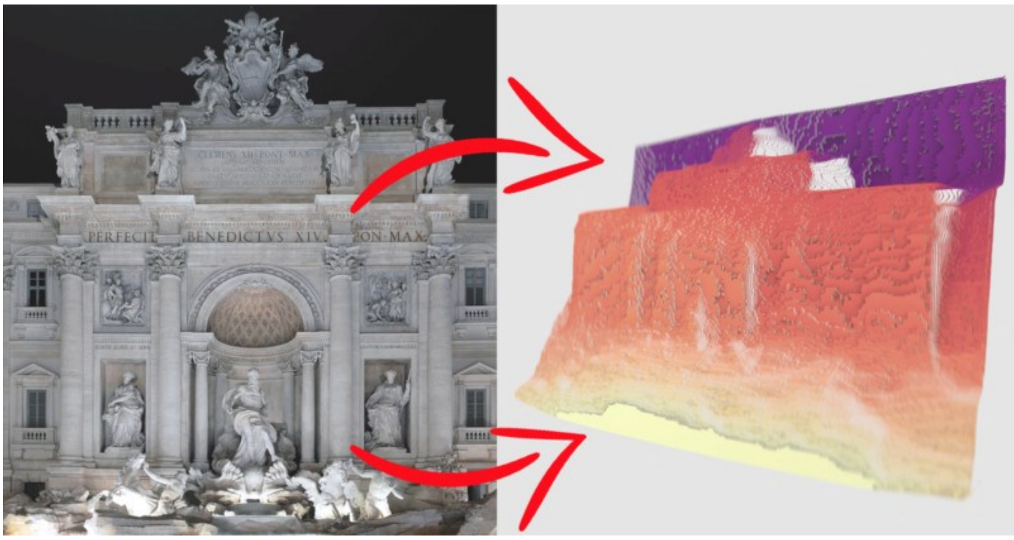


Figure 16. Facebook's 3-D representation of a 2-D image being posted in the feed [36]

The technical part to achieve this result is a bit complicated, yet Facebook could only say that their neural network can estimate the distance of every single pixel from the camera using 4 (four) different methods, which are the following:

- A network architecture built with a set of parameterizable, mobile-optimized neural building blocks.
- Automated architecture search to find an effective configuration of these blocks, enabling the system to perform the task in under a second on a wide range of devices.
- Quantization-aware training to leverage high-performance INT8 quantization on mobile while minimizing potential quality degradation from the quantization process.
- Large amounts of training data derived from public 3-D photos.

The outcome results are actually very convincing, as this methodology seems to work just as fine with any regular 2-D image posted in someone's feed. If anyone reading this thesis attempt to upload a 3-D image on Facebook, they will also notice that the amount of depth data that Facebook's methodology infers is quite staggering. Facebook urges everyone interested in this features to view their AI section of the website devoted to this process of converting 2-D images into 3-D depth representations by reading the information in [37].

Of course, Facebook is not an inventor of this feature, either are they the only one out there able to offer this transformation for free. Researchers at Deep Mind have developed an AI system that also works in a similar mostly approach, namely the *Generative Query Network (CQN)*, which can also perceive images from different angles, just like a human being can [38]. CQN is described - according to its developers - as a framework capable of making machines learn how to interpret 2-D surroundings through data training and produce a 3-D scene that can be navigated through. The framework gains knowledge through observations provided through 2-D images and can perceive the surroundings as plausible scenes with geometrical properties. As a result, the generative network can itself imagine unobserved scenes from new viewpoints with optimal precision. In fact, if the framework is given a scene representation with new camera viewpoints, the developers suggest that it can provide clean images without any prior specifications on the laws of perspective or lighting, thus creating a neural rendered capable of deciding on its own and that keeps getting better and better as long as more information is fed.

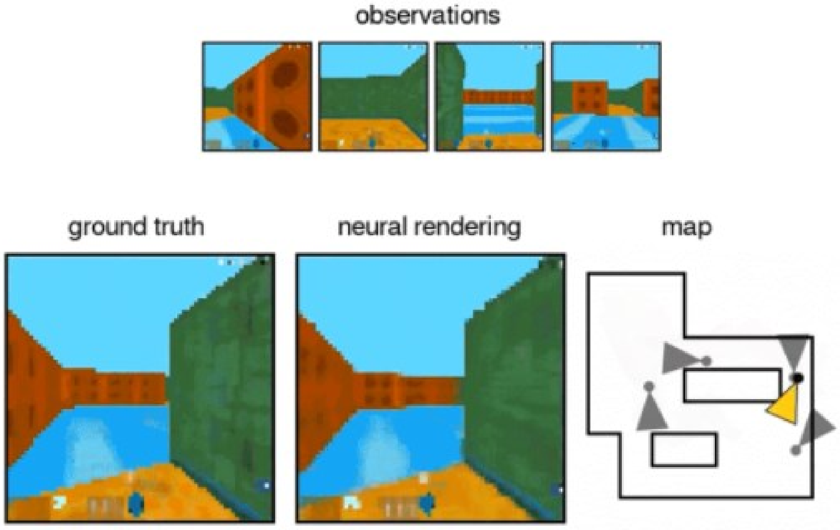


Figure 17. Deep Mind's Neural CQN renderer [38]

2.3.5 The Healthcare sector

Computer Vision has always adequately supported the medical sector and more specifically, in applications regarding medical image processing like *Computer Tomography (CT)* scans, X-rays and more. More recent studies in the Computer Vision field are promising to allow doctor to better understand the underlying technologies by converting such imagery to 3-D models that doctors will be more capable of interpreting [25].

For example, the most immediate example that comes to mind is (once again) the COVID-19 pandemic case. Computer Vision is currently being used in order to detect if a patient is infected by COVID-19 or not, by scanning with X-rays his chest. This comes after a recently published research in 2020 by the Department of Radiology in Wuhan, China (see [39]) which suggest that deep learning methods can be used efficiently to distinguish COVID-19 from community-acquired pneumonia. According to the authors, it was reported that that chest CT may be negative for viral pneumonia of COVID-19 at initial presentation. In fact, unenhanced chest CT may be considered for early diagnosis of viral disease, although viral nucleic acid detection with real-time polymerase chain reaction (RT-PCR) remains the standard of reference. Combating this issue, the authors compared the sensitivity of chest CT with that of viral nucleic acid assay at initial patient presentation. Their concluded that the use of chest CT to screen for COVID-19 in patients with clinical and epidemiologic features compatible with COVID-19 infection, particularly when results of RT-PCR tests are negative. An indicative example is shown below in *Figure 18*, which is provided from the COVID-19 dataset from the authors' research and which has a public domain license for usage. The dataset can be found directly from the Kaggle website in [40].

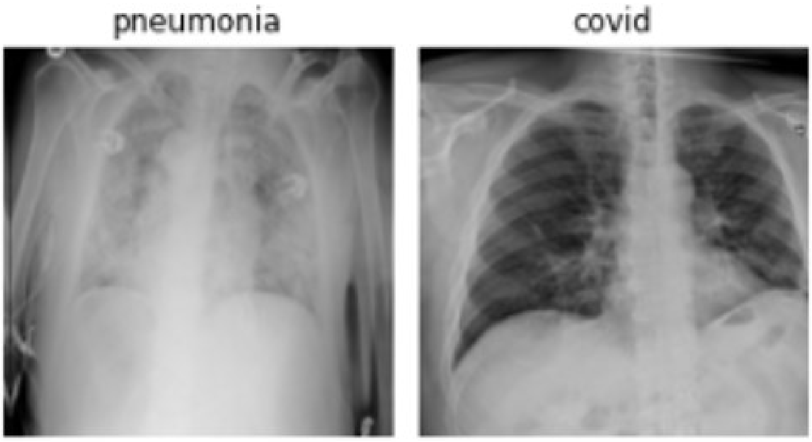


Figure 18. Comparing pneumonia-infected lungs with COVID-infected lungs [40]

2.4 Recognition Algorithms

In the following sub-sections, we will take a detailed look of the state-of-the-art algorithms regarding both object and facial recognition. Even though this thesis is targeted solely on real-time face recognition, it is imperative that the readers have a more comprehensive and overall view on the subject regarding the process of recognition, either it relates to objects or humans.

2.4.1 Object Recognition Algorithms

The technique of object recognition describes a collection of related tasks of the Computer Vision field, which involve object identification in digital images provided to the equivalent system. In most cases, when someone (e.g. a practitioner) refers to the process of recognizing an object, they actually refer/mean the process of detecting an object. To this end, according to [41], we can define four (4) different sub-fields of the object recognition process, which are the following:

1. <u>Image Classification:</u>
   Takes as input a single object image (e.g. a 2-D photograph) and outputs a class label (e.g. one or more integers mapped to class labels). In simpler terms, it predicts the object type, given any type of image.

2. <u>Object Detection:</u>
   Takes as input an image that contains many objects and outputs the equivalent bounding boxes (e.g. point/width/height definition) and a class label for each box. In simpler terms, it can locate object presence and mark it with a bounding box and assign a class to the located objects.

3. Object Localization:
   Takes as input an image that contains many objects and outputs the bounding boxes (e.g. point/width/height definition). In simpler terms, it can locate object presence and mark it with a bounding box, but as opposed to object detection, it cannot classify the object detected.

4. Object Segmentation:
   In this case, instances of recognized objects are indicated by highlighting the specific pixels of the object instead of a coarse bounding box.

The aforementioned sub-fields of the object recognition process can be visualized as depicted below in *Figure 19*. Meanwhile, we are observing an exponential research interest on these sub-sections and for the most part of it, responsible for this is the *ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*, a competition devoted to the identification and evaluation of algorithms for object detection and image classification at large scale, allowing: a) researchers to compare progress in detection across a wider variety of objects by taking advantage of the quite expensive labeling effort and b) measuring the Computer Vision progression so far regarding large scale image indexing for retrieval and annotation. Furthermore, every year the challenge is scheduled, there is also a corresponding workshop event, with the aim of presenting the methodologies and results of the challenge and where the most successful and innovative participant entries are being presented. More information on ILSVRC can be found by visiting the official webpage of the competition in [42].

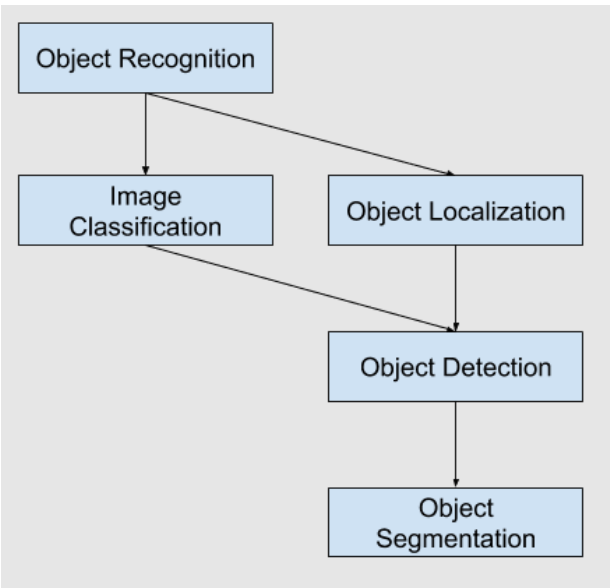


Figure 19. The subfields of Object Detection [42]

A comprehensive example that tackles the Object Localization and the Object Detection aspects is the following, which is straight from the IRLVRC. In this example, we can see the differences between: a) the original ground truth (on the left) and the levels of accuracy (on the right) and b) the ground truth of Single-object Localization as opposed to the ground truth of the case of Object Detection. It should be noted at this stage that the methodology of measuring the performances is different across all sub-sections mentioned above. This means that:

- Regarding image classification, the intended model's performance is evaluated using the mean classification error across the predicted class labels.
- Regarding single-object localization, the intended model's performance is evaluated using the distance between the expected and predicted bounding box for the expected class.
- Regarding object recognition, the intended model's performance is evaluated using the precision and recall across each of the best matching bounding boxes for the known objects in the image.

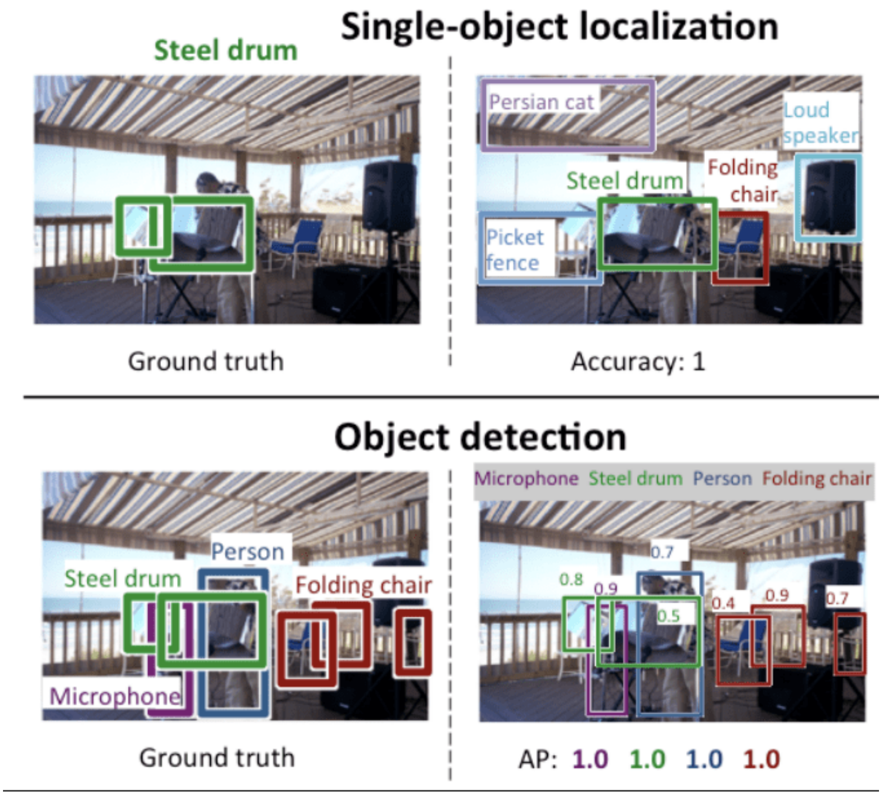


Figure 20. Comparing the cases of Single-object Localization and Object Detection [42]

Up to this point, readers are expected to have a basic understanding of the problem of object localization and object detection, so it is time to dig further in the abyss of the relevant Deep Learning models that are the top performers in the field.

2.4.1.1 The R-CNN framework

The R-CNN framework was firstly introduced to us from a research work back in 2014 from the University of Berkeley by Ross Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik titled "*Rich feature hierarchies for accurate object detection and semantic segmentation*" (this work can be accessed from [43]). That work proposed a simple and scalable detection algorithm that improved mean average precision by more than 30% relative to the previous best result on the canonical PASCAL VOC 2012 dataset, achieving a precision of 53.3%. To achieve this, they applied high-capacity CNNs to bottom-up region proposals in order to localize and segment objects. When labeled training data was scarce, the applied supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance boost. This approached was named by the authors as R-CNN, which stands for regions with CNN features. Regarding the evaluations, the approach was demonstrated on benchmark datasets, achieving then state-of-the-art results on the VOC-2012 dataset and the 200-class ILSVRC-2013 object detection dataset.

It is widely believed that this might as well be the first very large-scale (and of course successful) application of CNNs towards solving the problems of Object localization, detection and segmentation. The R-CNN framework consists of three (3) different modules, which are the following:

1. Region Proposal:
   This module is responsible for generating independent categorical region proposals (e.g. proposals for bounding boxes).

2. Feature Extractor:
   This module is responsible for extracting features from the candidate regions (e.g. through using a CNN).

3. Classifier:
   This module is responsible for classifying features to the pre-defined classes (e.g. a linear *Support Vector Machine (SVM)* problem).

The architectural blocks that support the R-CNN approach are presented below in *Figure 21*. This technique proposes either candidate regions or bounding boxes in the input image through selective search, even though the design is flexible enough to allow other region candidates to be proposed and used. The feature extractor that R-CNN is comprised of 4.096 element vectors as output, which accurately describes the image contents that are fed to a linear SVM for classification and this approach is considered quite simple and straightforward towards solving the problems of object localization and recognition.

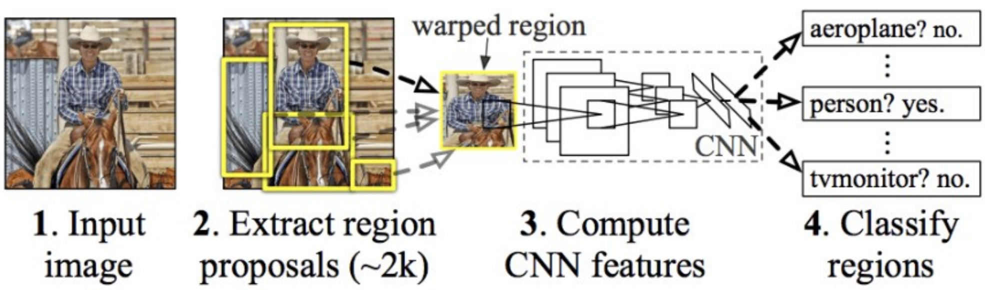


Figure 21. The R-CNN features [41]

As always, one should expect that there exist some downsides in this approach. More specifically, regarding the computational time, it is a very slow approach, since it requires a CNN-based feature extraction pass on each of the candidate regions generated by the region proposal algorithm. To this end, the research work produced by its authors [43] suggests that the model operates just above approximately 2.000 proposed regions per image during testing, something that we can also see in *Figure 21* above. Lastly, as far as the software implementation is concerned, the Python and MATLAB programming languages were used to develop the R-CNN framework and the original source code is publicly uploaded on a Github repository for those interested and can be accessed from [44].

2.4.1.2 The Fast R-CNN framework

So far, we saw how the original C-RNN framework operates in detail and given its success, it should not be a surprise that this work has been extended in the latter years towards the best. In this sub-section, we will describe the Fast R-CNN framework, which stems from Ross Girshick (he was also part of the team that developed the original R-CNN) during his tenure at Microsoft Research, where he extended his work through a new scientific research titled just "*Fast R-CNN*" in 2015 (his work can be accessed from [45]). This new work built upon the existing work a year ago (R-CNN was presented for the first time in 2014, as already mentioned) in order to efficiently classify object candidates using deep CNNs. In contrast to the original R-CNN, Fast R-CNN employs several innovations to improve training and testing speed while also increasing detection accuracy. More specifically, according to the author, Fast R-CNN trains the very deep VGG16 network 9 times faster than R-CNN, is 213 times faster at test-time, and achieves a higher precision on PASCAL VOC 2012. Compared to SPPnet, Fast R-CNN trains VGG16 3 times faster, tests 10 times faster and is more accurate. This framework was developed using the Python and C++ programming languages and as its predecessor, it is available as an open-source project in Github and be can be accessed from [46].

In more detail, Fast R-CNN is in fact a single model instead of a linear pipeline. As input, the model receives the 2-D photograph and creates a set of region proposals, which are later on passed through a CNN. The model then used a pre-trained CNN (e.g. VGG-16) to extract the features from the proposals. The output of the CNN is a layer called "*Rol Pooling*", which is responsible for extracting specific features for each candidate region respectively. This output is later on interpreted by a fully connected layer. After this stage, the information is split into two different outputs, namely: a) the softmax layer that predicts the class and b) the linear output that creates the bounding box. The aforementioned procedure is repeated for search region that is of interest and this approach is depicted visually in *Figure 22* below.

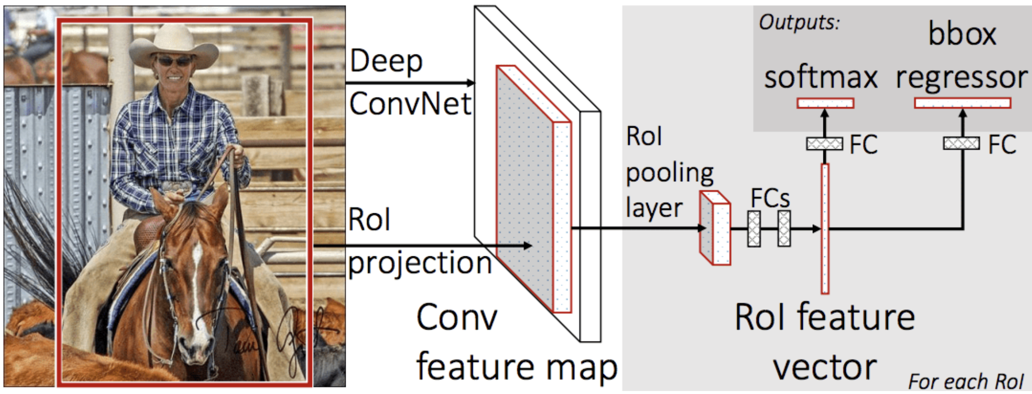


Figure 22. The Fast R-CNN features[45].

Ultimately, the Fast R-CNN framework model is significantly faster to train and to make predictions, but one downside is that it still requires a set of candidate regions to be proposed along the 2-D image which is provided as input. Furthermore, this framework also has several other limitations. The procedure of training is a pipeline of multiple sates which includes operating three (3) separate models and also, training a deep CNN in such an abundance of region proposals is very slow in time and expensive in space. Lastly, the procedure of object detection is slow, since the too many predictions on the CNN for so many region proposals is also slow.

2.4.1.3 The Faster R-CNN framework

As expected, this framework once again was extended further in 2016, where Ross Girshick, Shaoqing Ren, Kaiming He and Jian Sun submitted an extended tech report titled "*Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*". In this contribution, they introduced a *Region Proposal Network (RPN)*, which is capable of sharing full-image convolutional feature with the detection network. This enables region proposal that do not cost in terms of computations. So, what an RPN does is that is a fully convolutional network that simultaneously predicts object bounds and objectness scores at each position. It can generate region proposals of high quality, which are then used by the Fast R-CNN framework for detecting features. This contribution merges RPN and Fast R-CNN through sharing the convolutional features. As it turned out, in ILSVRC-2015 and COCO 2015 competitions, both Faster R-CNN and RPN were the foundations of the 1st-place winning entries in several tracks [47].

As effectively expressed over, the design of the structure was the reason for a considerable lot of the primary spot results accomplished on both the ILSVRC-2015 and MS COCO-2015 item acknowledgment and discovery rivalry errands. All the more explicitly, the framework model of the Faster R-CNN structure was planned to propose and redfine the area proposition as a component of the preparation cycle, which are then utilized with a Fast R-CNN model in a solitary model plan. These enhancements both lessen the quantity of area recommendations and speed up the test-time activity of the model to approach continuous with then cutting edge execution. The architecture is depicted below in *Figure 23*.

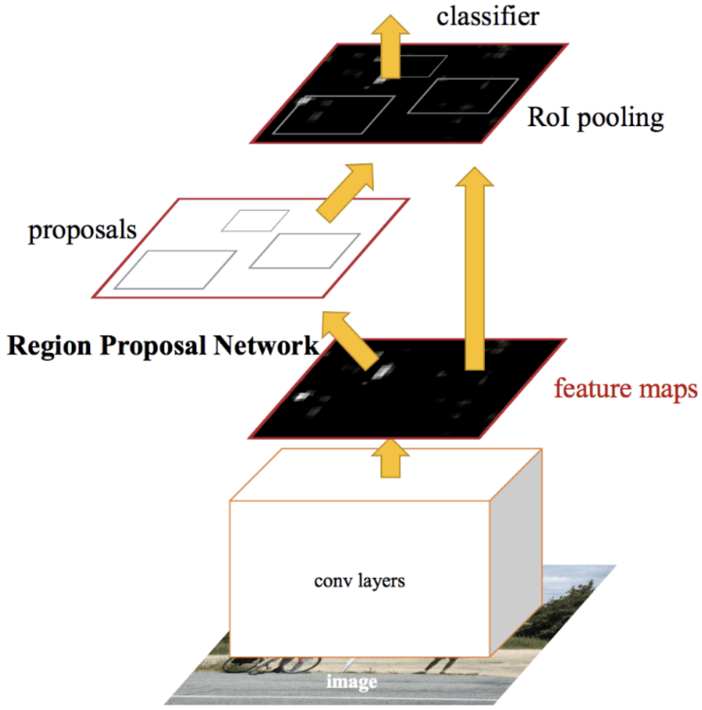


Figure 23. The Faster R-CNN features [47]

The RPN works like this: RPN acquires the output of a pre-trained deep CNN, and then forwards a small network over the feature map. This results in the generation of multiple region proposals and later on, a class prediction for each. Region proposals are basically bounding boxes, based on anchor boxes that were designed in order to accelerate and improve the proposal of regions. The class prediction is binary and can indicate whether an object appears or not inside the image. A procedure of alternating training is used where both sub-networks are trained at the same time, although interleaved. As a result, the parameters in the deep CNN can be programmed for both tasks simultaneously. Regarding the programming languages, once again, the framework was developed using the Python and C++ programming languages and as its predecessor, it is available as an open-source project in Github and be can be accessed from [48].

2.4.1.4 The YOLO framework

We came across the YOLO framework before in this thesis (please refer to the subsection **2.3.3. Social Distancing Tools**), but we did no thorough analysis whatsoever. The YOLO framework was firstly introduced in 2016 (a few years after the appearance of R-CNN) again through a research work from Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi (note that Ross Girshick is the only author who was involved in all object recognition scientific papers so far) under the title "*You Only Look Once: Unified, Real-Time Object Detection*". This scientific work can be accessed from [49], whereas the source code is publicly accessibly at Github in [50].

At the time, the common approach on object detection was to re-purpose classifiers so at to proceed with the detection process. The authors took a different path, by framing object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. YOLO is capable of using a single neural network that can predict the bounding object boxes and the probabilities of the equivalent classes from the 2-D input images in just one evaluation and since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. According to the developers, YOLO can be considered as a quite fast tool for object detection, as it can process images in real-time in 45 *Frames-Per-Second (FPS).* Not only the YOLO object recognition approach is very fast, it is also far less likely to make false predictions on object detection in areas where nothing exists, but on the other hand, it make more localization errors. In overall, YOLO is able of interpreting and learning – of course through a training phase – about general representations and shapes of objects and can without any problem outperform any well-known object detection mechanism (e.g. R-CNN).

Concerning more intensive look on the cycle of acknowledgment, the model works by initial parting the info picture into a network of cells, where every cell is liable for foreseeing a jumping box if the focal point of a bouncing box falls inside the cell. Every framework cell predicts a bouncing box including the x, y facilitate and the width and stature and the certainty. A class forecast is likewise founded on every cell. For instance, a picture might be partitioned into a 7×7 framework and every cell in the lattice may foresee 2 jumping boxes, bringing about 94 proposed bouncing box forecasts. The class probabilities map and the bouncing boxes with confidences are then consolidated into a last arrangement of jumping boxes and class names. The picture taken from the paper beneath sums up the two yields of the model. This can be seen below in *Figure 24*.

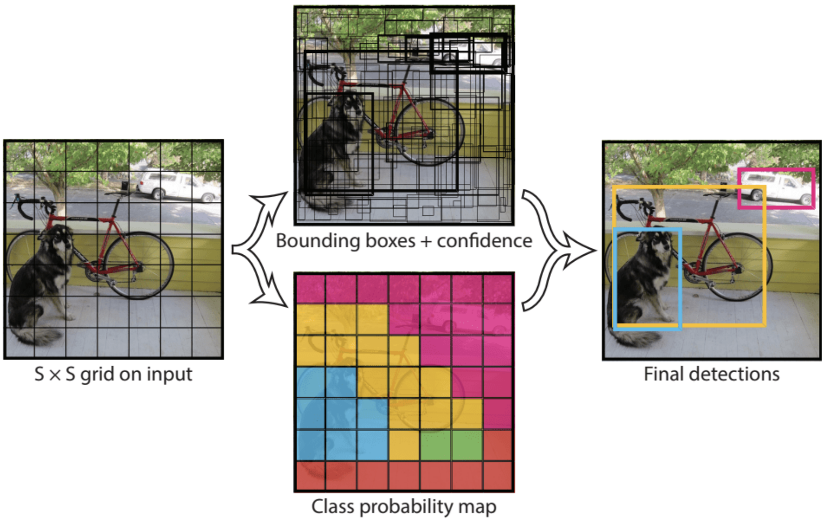


Figure 24. The YOLO procedure towards object detection [49]

2.4.1.5 The YOLOv2 and YOLOv3 frameworks

As R-CNN, the first edition of YOLO received some noticeable enhancements from its native version. Two different versions came after the original YOLO version, namely the YOLOv2 [51] and YOLOv3 [52] respectively. YOLOv2 was introduced in 2016 and was capable of training over two object recognition datasets simultaneously, which resulting in a predicting capability of 9.000 object classes, hence some called it the YOLO9000 version. This version also features many important enhancements, like the inclusion of batch normalization and the capability of high resolution input 2-D images. YOLOv2 reminds us of Faster R-CNN in the sense that it uses bounding boxes with equivalent shapes and sizes that are created during the training process. Depending on the efficiency and the density of the training dataset and the overall training process, selecting the bounding boxes for the input image requires pre-processing using a K-means analysis. The prediction that includes the predicted bounding boxes can only allow small changes so as to have a less impactful effect on the predictions, which can offer a more stable model. So, rather than predicting position and size directly, offsets are predicted for moving and reshaping the pre-defined anchor boxes relative to a grid cell and dampened by a logistic function [41].

Regarding the YOLOv3 version suggested in 2018, there are minor improvements from the YOLOv2 version (yet extensive if compared with the original YOLO framework). The improvements include incorporating a more deep feature detection network, alongside with several minor changes regarding the representation of the output image and the bounding boxes.

2.4.2 Face Recognition Algorithms

Several approaches exist related to face recognition via feature extraction. They fall under three different categories: a) Geometric-based feature extraction, b) Appearance-based feature extraction and c) Template-based feature extraction.

2.4.2.1 Geometric-based feature extraction

To start with, Geometric-based feature extraction are mainly based on two different algorithms that use the above approach: a) *Active Shape Model (ASM)* and b) *Active Appearance Model (AAM)*. The ASM and AAM have been deeply studied, tested and evaluators from the authors in [53].

In the ASM calculation, the state of an item is addressed utilizing tourist spots as a chain of sequential characteristics focuses where each point existent in many pictures is thought of (for example area of the human mouth). To cover the face shape and its subtleties, a sufficient number of quality focuses should be given (the creators recommend that an aggregate of 68 attribute focuses can satisfactorily clarify the state of a human face). At that point, by gathering all milestones, the structure the face shape as vectors utilizing the x and y coordinates. By aligning shapes with each other using transformations, the Euclidean distance average between the shape points is significantly reduced and thus, the ASM begins searching for facial landmarks from the mean shape aligned to the place and size of the face specified from the detector. The next steps include re-iterations by proposing temporary face shapes and confirming this shape to a universal shape model. The last step is re-iterating the complete search on each level in an image pyramid, beginning from the lowest level up to the optimal image resolution.

The ASM has been greatly used in the research spectrum of vision computing. In [54], the authors suggested improvements on the classical ASM approach in order to increase the model performance regarding face alignment. Their improvements included better face models, edge enhancements in the image, better landmark classification in the face and automatically adjustable 2-D profile based on the input image. The authors of [55] proposed an ASM segmentation scheme steered by optimal local face features using non-linear classifiers, which is a far different approach from the original ASM approach where the derivative profiles are normalized and linear. In [56], the authors used ASM for face recognition and presented all shapes with a landmark set that forms a *Point Distribution Model (PDM)*, an approach that provided a series of improvements in the original ASM, such as the search profile width, image noise reduction, landmark grouping and zero mouth shape distortion.

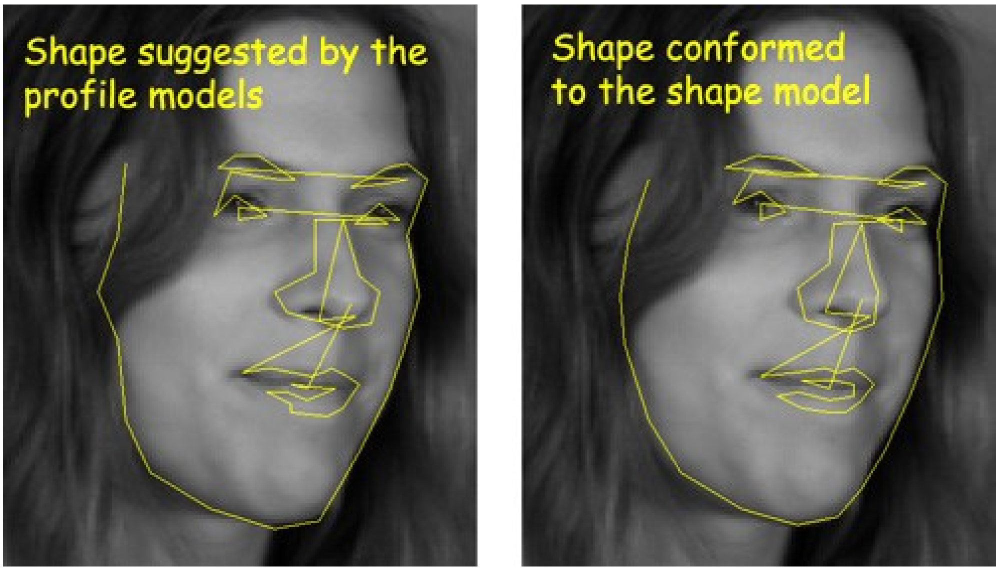


Figure 25. The ASM model [57]

The AAM algorithm was firstly introduced back in 2004 from the authors of [58]. The AAM algorithm is a generative, non-linear, parametric model and is still today one of the most efficient and powerful model-based object tracking and detection algorithm. The algorithm builds the texture and shape of the specified object in order to produce an overall set of realistic images. While the ASM algorithm captures very little texture variation and mainly used shape models, the AMM approach considers detailed textures, utilized to form generative models to produce photorealistic images. This means that while ASM matches shape models, AAM matches a complete full appearance model to an image. AAM decides which shapes are valid using a pre-defined training set and uses the decisions to generate similar examples to the training set. As a result, this approach provides a strong modelling and overall ability in capturing the raised complexity, due to its high-dimensional texture representation.

In [59], the authors demonstrate a novel method for image interpretation using the AAM algorithm. They learned the relationship between the model parameters and the residuals errors between the algorithm's training image and the synthesized result and matched the image using the residuals and thus, locating deformable objects. In [60], the authors acknowledge that AAM algorithms are either yet robust but slow and either fast yet less able to generalize and proceeded to suggest a simple optimization framework which can efficiently unify and revised the most important optimization problems and solution in AAMs. The authors of [61], the authors studied the generalization problem AAM faces when applied to unseen images or objects and developed a novel *Locality-constraint Active Appearance Model (LC-AAM)* that tackles this issue through a learning approach using a strong localized shape and appearance, before the image input.

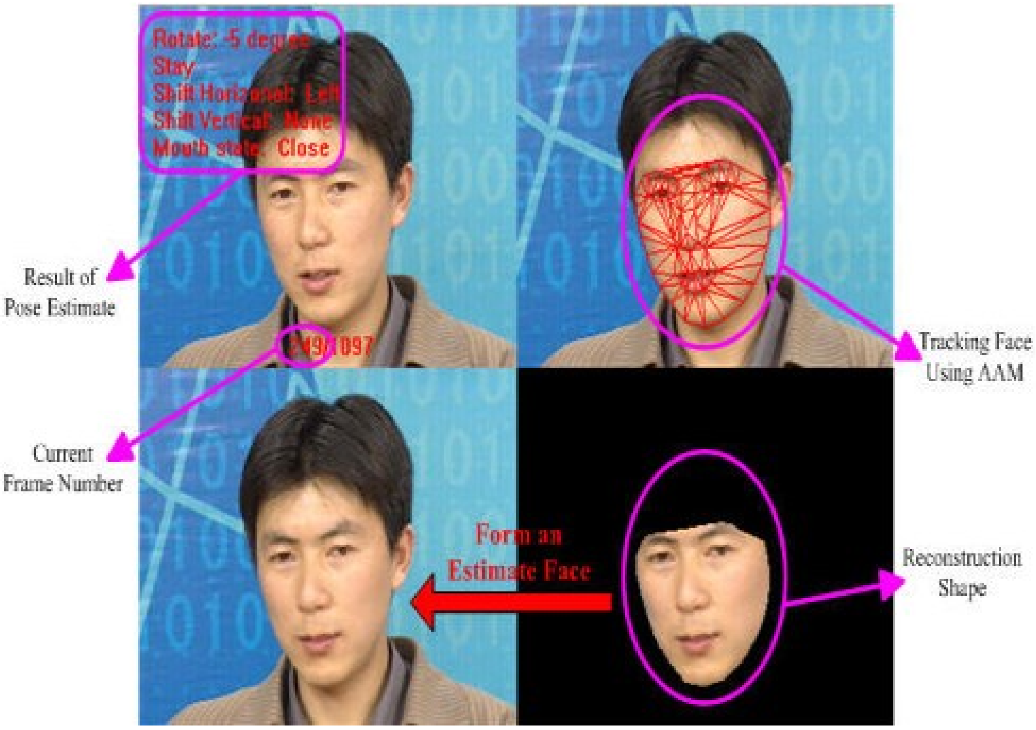


Figure 26. The AAM model [62]

2.4.2.2 Appearance-based feature extraction

Appearance-based feature extraction techniques are usually performed using color-based imaged and appearance-based classification. The algorithms that are included in this category are: a) Haar Cascade, b) *Principal Component Analysis (PCA)* and c) *Local Binary Pattern (LBP)*.

The Haar Cascade algorithm is one of the most used approaches towards face recognition. In [63], the authors suggest the Haar cascade face detection algorithm, extended using three additional classifiers, such as skin hue matching, eyes detection and mouth detection. The authors used the extra classifiers in order to get rid of the wrongly accepted faces that did not represent humans, at each stage of the process. For example, firstly they used the face skin hue matching for discarding non-human faces, then the eye detection classifier and lastly the mouth detection classifier. Results revealed that in both training and test sets, this approach offers "state-of-the-art performances", as they authors state and increased levels of actual human face detection. The authors of [64] performed an all-around evaluation of the available Haar Cascade classifiers used for human face detection. This greatly helps researchers towards choosing the best classifiers for the particular needs of face recognition. This contribution focused on evaluating facial classifiers regarding the facial features recognized in the found faces and on a later stage, the authors proposed a method of variable scores assigned to each facial feature. The authors used the FEI and the Yale face databases and evaluated 10 face classifiers in total.

The research work behind the Haar Cascade algorithm does not stop here. In [65], the authors applied the Haar Cascade algorithm to detect safety equipment inside safety management systems for multiple working environments. Even though many ways exist in order to keep the workers safe in a working site (e.g. human supervisors, computer supervisors, smoke-flake detection systems), it was decided to follow an state-of-the-art approach on the matter through the Haar Cascade algorithm. The algorithm was used so as to build four (4) different classes for recognizing safety equipment and after that, the algorithm was used in order to calculate a score that determines the dangerousness of the equivalent working environment, based on a) the safety equipment and b) the working environment and if deemed necessary, the system gives warning signals. The authors of [66] presented 3 main contributions on rapid object detection through the classifiers offered from the Haar Cascade algorithm. Firstly, they used the Integrral Images as an image representation technique so as for the face detection algorithm to recognize human faces more quickly. Secondly, they adapted the AdaBoost learning algorithm so at to yield more efficient classifiers and lastly, the combined the most complex classifiers so as to quickly discard background regions of the image. The usage of object detection using the Haar Cascade algorithm also caught the interest of researchers like in [67], where a case study was performed regarding vehicle detection. The study addressed the possibility of using object detection and its benefits towards providing assistance in semi-enclosed areas and proposed a system that will augment our everyday lives through the features offered from the Haar Cascade algorithm.

The LBP approach is a texture-based feature extraction technique, introduced to the research community from the authors of [68]. The LBP works with the eight (8) neighbors of a pixel and uses the center pixel value as a threshold. If the neighboring pixel has a higher gray value than the center pixel, then the binary one (1) is assigned to that pixel, otherwise we assign the binary zero (0) to that pixel. After that, the LBP code for the center pixel is calculated by concatenating the 8 binary numbers and forming a binary sequence of zeros and ones. The next step is to extend the LBP to neighbors of other sized, by creating a circle radius from the center pixel. The goal then is to receive the values of the sampling points on the circle edge and compare them with the center pixel. This approach is called *Extended Local Binary Pattern (ELBP)*. The last step of the LBP is to construct the feature vector from the image, after calculating the vector for every pixel. LBP will then divide the face image into multiple regions and create a histogram with all possible labels per region. The histogram represents a pattern and includes the number of the region appearance. Using the histograms, the feature vector is constructed by merging the regional histograms to one and only histogram.

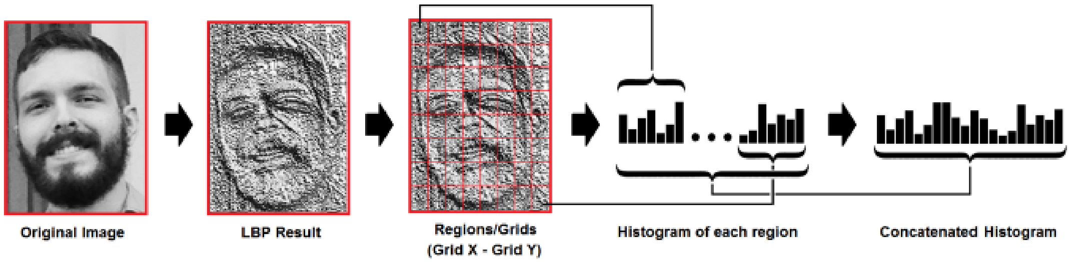


Figure 27. The LBP model [71]

The authors of [69] achieved fast pre-processing, feature extraction and matching that improves current face recognition systems using the LPB and PCA algorithms. The PCA algorithm is a classic feature extraction technique mostly used for pattern recognition and vision computing matters. This approach begins by using eigenfaces to construct the training set images and transform them into the final eigenfaces (only the highest eigenvalues are considered) so as to calculate the equivalent weight for each face image (eigenvector). The next stage is the so-called Recognition Stage, where we provide in image for recognition and the whole process is re-iterated and applied to a testing set for face recognition, alongside with the calculation of the Euclidean distance between them. In this stage, the minimum Euclidean distances are considered with a specific threshold. If the distance is less than the threshold, the result is a face recognition match, otherwise the algorithm provides no recognition. The authors of [70] performed a survey on the non-parametric LBP algorithm (which is considered one of the post powerful descriptor of local structures) and its applications in facial image analysis. They researched the many areas of image processing and computer vision and presented a survey of all LBP approaches and extensions and also reviewed the LBP-based facial image analysis. The paper concludes by drawing attention on open questions related to LBP descriptions (e.g. facial description and its components and neighboring operators) and acknowledges that even though these questions have been discussed from different authors, their results were user database-dependent of that time.

2.4.2.3 Template-based feature extraction

Template-based feature extraction techniques are usually performed by determining the facial features during recognition. The algorithms that are included in this category are: a) Deformable templates and b) Gravity center template.

The deformable templates approach can be used for both face and object recognition. A typical approach for object recognition is presented by the authors of

[72], who propose a general object localization and retrieval scheme which is based on object shapes using deformable templates. They describe the object shape from a prototype template consisting of the equivalent contour/edges and a set of deformation transformation on the template. Their method can be used to retrieve objects from a variety of shapes from images with high complex backgrounds and can perform well, invariant to the location, the rotations and object scales. On the other hand, a common approach for face recognition is presented in [73], where the authors take a specific feature of interest and describe it by using a parameterized template, which is then used to interact with the image by altering the parameter values. The final parameters are later used as descriptors for this feature. As a result, the deformable templates can be used in real images as features. This means that algorithms can be formed by extracting e.g. the eye contours from the face image, using the deformable templates and then convert the digital color face image into a binary image that represents the eyes.

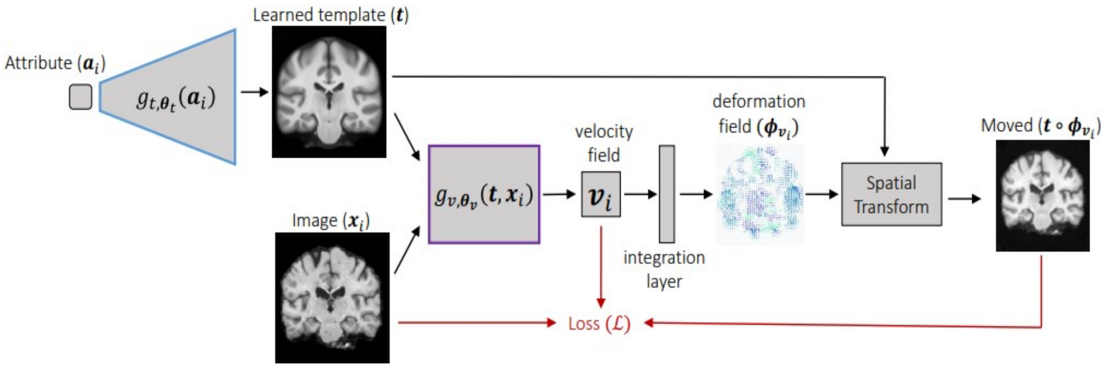


Figure 28. The Deformable Templates model [74]

The Gravity Center template, when utilized for face acknowledgment, works by removing the essential edges and locales around the first picture's organs and afterward the picture is prepared. During the preparing stage, the picture is checked to distinguish the organs. This methodology has been executed by the creators of [75] and the outcomes showed that the proposed strategy is very quick, productive, reliable and the last state of the face highlights is satisfactorily portrayed. In [76], the authors proposed a model that locates and tracks a driver’s mouth movement while driving through a *Charge-Coupled Device (CCD)* camera using the Gravity Center template. On the first stage, the drive’s face is recognized through the Gravity Center template. On the second stage, the driver’s left and right mouth corners are discovered using grey projection. Lastly, by making use of the Gabor wavelets, the driver’s texture features of the mouth are extracted and used.

2.5 Selecting the face recognition algorithms

As we saw in **2.4.2.2 Appearance-based feature extraction**, the Haar Cascade algorithm is one of the most used approaches towards face recognition. Firstly introduced back in 2001 by Paul Viola and Michael Jones (see [66]), where the authors proposed the first ever Object Detection Framework for Real Time Face Detection in Video Footage. The creators introduced three (3) primary commitments on fast object detection through the classifiers offered from the Haar Cascade calculation. Right off the bat, they utilized the Integral Images as a picture portrayal method so concerning the face identification calculation to perceive human faces all the more rapidly. Furthermore, they adjusted the AdaBoost learning calculation so at to yield more effective classifiers (roughly 6.000) and in conclusion, the consolidated the most intricate classifiers to rapidly dispose of foundation locales of the picture. This work was done some time before the DL techniques had even begun, yet it's an amazing work in contrast with the incredible models that can be worked with the cutting edge DL techniques. The calculation is as yet discovered to be utilized all over, it has completely prepared models accessible on GitHub, it's quick and it's really precise (so it meets precisely to our needs).

Haar Cascade is an Object Detection Algorithm used to distinguish faces in a picture or a real-time video. The calculation is given a ton of positive pictures comprising of appearances, and a great deal of negative pictures not comprising of any face to prepare on them [79]. The model made from this preparation is accessible at the OpenCV GitHub vault (see [78]), consisting of all the relevant models (face or eye detection, upper/lower body detection, license plate detection and more) being stored as *eXtensible Markup Language (XML)* files and understandable by OpenCV methods. The first contribution to the research from their work was the introduction of the Haar feature shown below in *Figure 29* that make it easy to discover the edges or the lines in any image, or to select areas where there is a sudden change in the pixel intensity.

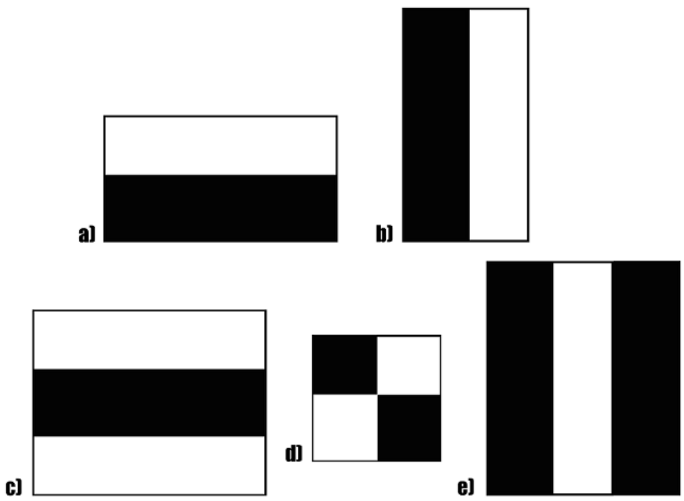


Figure 29. The Haar Cascade features [66]

Now, let's look at *Figure 30*. The square shape on the left is an example portrayal of a picture with pixel esteems that range from 0.0 to 1.0, though the square shape at the middle is a Haar feature which has every one of the white pixels on the left and every one of the dark pixels on the right. At that point, the Haar estimation is done by discovering the distinction of the normal of the pixel esteems at the dark district and the normal of the pixel esteems at the white locale. On the off chance that the thing that matters is near 1, there is an edge recognized by the Haar feature. The more obscure regions in the Haar feature are pixels with values 1 and the lighter regions are pixels with values 0. Each of these is liable for discovering one specific component in the picture. Like an edge, a line or any design in the picture where there is an abrupt difference in forces. the general target is to find the amount of all the picture pixels that exist in the more obscure space of the Haar feature and the amount of all the picture pixels that show up in the lighter space of the Haar feature and afterward figure their distinction. If the picture has an edge isolating dull pixels on the privilege and light pixels on the left, at that point the Haar worth will be more like 1. That implies, we say that there is an edge identified if the Haar feature is more like 1. In the model seen underneath, there is no edge as the Haar feature is a long way from 1 (as should be obvious, equivalent to - 0.02).

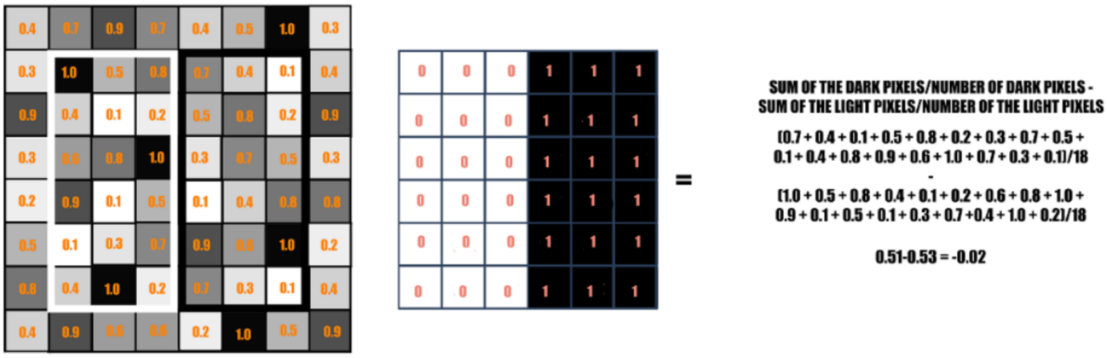


Figure 30. Calculating the Haar value [79]

The example we presented covers roughly only one representation of a specific Haar feature that separated a vertical edge. Of course, there exist many other Haar features as well that can detect edges in other directions and image structures. So for the Haar algorithm to detect an edge at any place on a given image, the Haar feature has to be applied on the whole image. As depicted below in *Figure 31*, we are testing a personal image of the author of this thesis, where at that exact moment, the Haar feature a) (see *Figure 29*) is being applied in the image.

Figure 31. Applying the Haar Feature a)

At this stage of the algorithm, the Haar feature will continuously traverse the image from the top left towards the bottom right in search of a certain feature (more precisely, the search traversal is pixel by pixel). Depending on the feature each one is looking for, there exists three (3) different categories: a) 2 rectangle features, b) 3 rectangle features and c) 4 rectangle features. The first case is responsible for discovering the edges horizontally or vertically, as depicted above in *Figure 31*. The second case is responsible for discovering if there is a lighter region that is surrounded by darker regions on each side. The third case is responsible for discovering changes in pixel intensities across diagonals.

At this stage, a major issue arises. The Haar features need to traverse the whole image, which including too many calculations and this must be done repetitively for all Haar features, a task that is a heptic operation for even high performance machines. More specifically, the Haar features traversal requires 18 pixel value additions for a single rectangle on either side. To this end, the Integral image concept was introduced, which is calculated from the original image so that each pixel is the sum of all pixels in its left and above from the original image. This is illustrated in *Figure 32*, each pixel inside the Integral image is basically the sum of all the pixels in its left and above. Consequently, the last pixel at the bottom right corner of the Integral Image will be the sum of all the pixels in the original image.

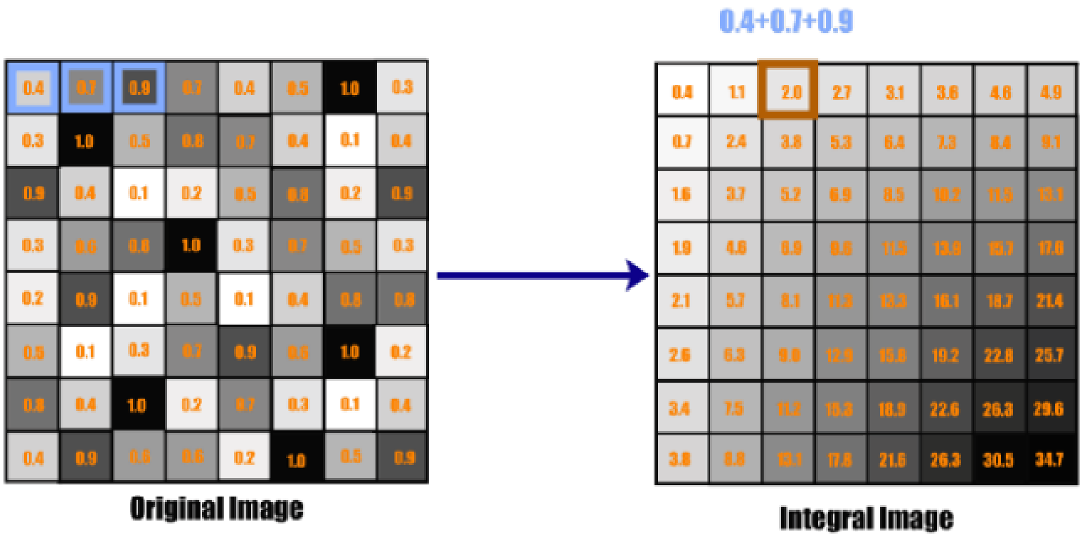


Figure 32. Creating the Integral image [79]

Using the Integral images concept, now we only need 4 constant value additions each time for each feature size (before this technique, 18 value additions were necessary). This of course greatly reduced complexity. As we can see in *Figure 33* below, there is no edge in the vertical direction as the Haar value is - 0.02, which is not even close near to the value of 1.00, which is our optimal value.

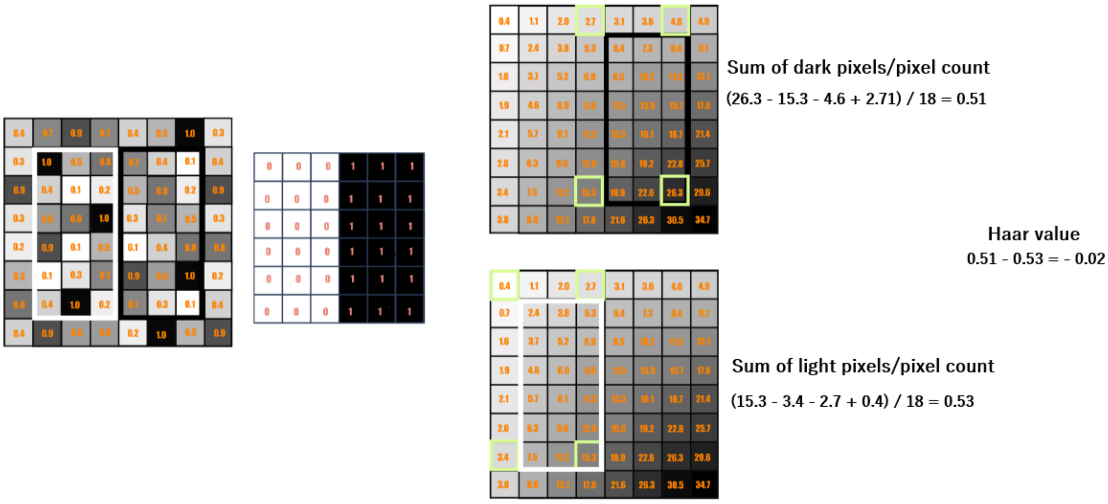


Figure 33. Haar value calculation from Integral image [79]

Now, at this stage of the algorithm, we can comprehend how the Haar features can aid us in discovering certain facial characteristics, like eyebrows or lips. Here comes the help of the Adaboost training technique, which reduced the overall feature set from the 180.000 to only 6.000, which are the features that work best with facial features and offer better performances. The remaining 6.000 features will again run on the training images to detect the facial features, but running all these features is again a tiresome task and as a result, there exists an additional cascade level where not all features need to run on all windows. Simply, if a feature fails in a particular window (24x24), then we safely assume that the facial features are not present there and the algorithm shall move on to neighboring windows. Thus, we end up saving a lot of processing time, as the irrelevant windows are discarded quickly. This is illustrated in *Figure 34* below.

Figure 34. The additional Cascading window

So, this is how the underlying detection procedure using the Haar features that takes place in stages. When the window is at a non-face region, only the first stage with two rectangle features runs, discarding the window before the second stage begins. Just a single window which really contains a face would run both stages and distinguishes the face.

Chapter 3. Designing the Platform

3.1 Selecting a human-centered framework

In order to help designers across the world to tackle their everyday struggles and issues on their running projects, it is vital for designers to be provided with frameworks that will enable them to use them in their work and better dictate the process of delivering the services. One of the most well known frameworks for innovation is the design methodology called the 'Double Diamond'. As well as highlighting the design process, Design Council's framework for innovation also includes the key principles and design methods that designers need to take, and the ideal working culture needed in order to achieve significant and long-lasting positive change in the future. This is not a linear process, as the arrows show in *Figure 35*. Many of the organizations learn something more about the underlying problems which can send them back to the beginning. Making and testing very early stage ideas can be part of discovery and in an ever-changing and digital world, no idea is ever 'finished', as our world is constantly getting feedback on how products and services are working and iteratively improving them [80].

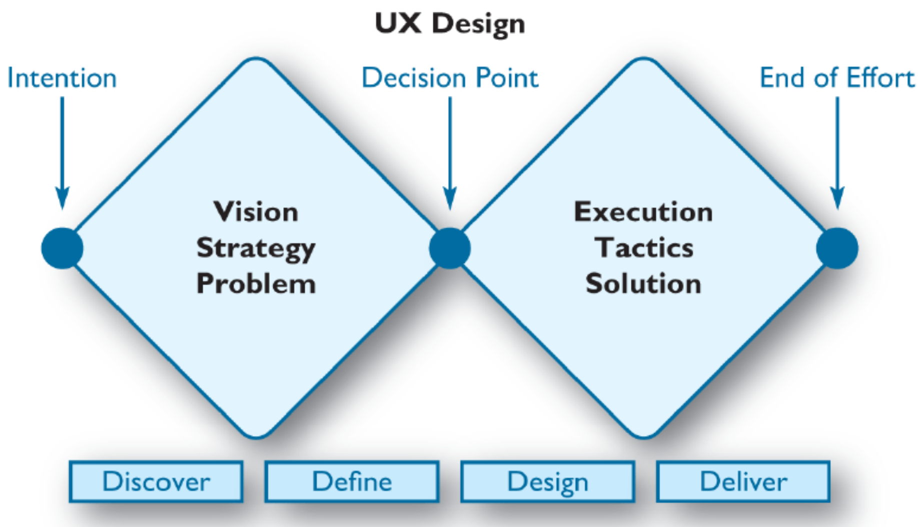


Figure 35. The Double Diamond model [80]

This framework represents a process in which an issue is explored more widely through divergent thinking and then designers use the knowledge gathered in order to take focused actions, which is the convergent thinking. By following the Double Diamond framework during the design process, the principles dictate that we: a) put people first, b) communicate visually and inclusively, c) collaborate with people and d) always iterate. The model is structured in four main points of interest:

- <u>Discover:</u>
  This helps the people involved to better understand instead of assume what is the actual problem that requires solving, though speaking and spending time with the affected people who experience the problem
- <u>Define:</u>
  This helps developers define the challenge through the knowledge gathered through the discovery phase.
- <u>Develop:</u>
  Here, the people are encouraged to offer different answers to the clearly-defined problem, so as to seek inspiration and co-design using a wide range of people who experience the problem.
- <u>Deliver:</u>
  The last phase includes testing the offered solutions at a small scale and rejecting the solutions that will probably fail.

Now, based upon the Double Diamond and its much needed iterations throughout the discovery, definition, design and delivery process and taking into account the development of the platform addressed in this thesis, we list the following key concerns that we will be tackling, namely a) Design, b) Technologies, c) People and d) Activities. Designers need to understand the people who will use their systems and products, the activities that people want to undertake and the contexts in which those activities take place. Additionally, they also need to know about the features of interactive technologies and how to approach designing interactive systems [81].

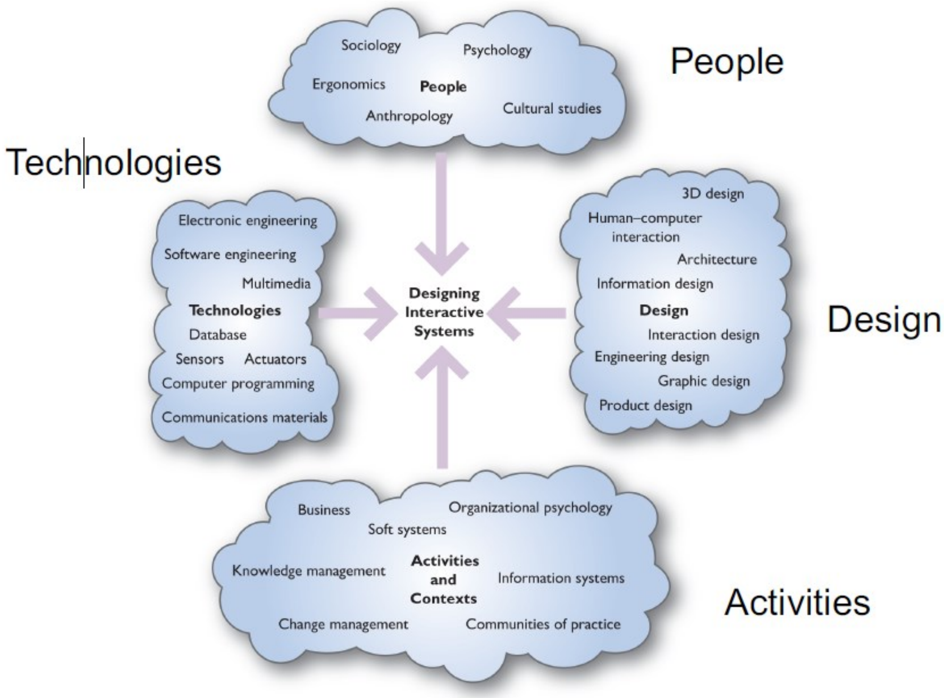


Figure 36. The disciplines that contribute to interactive systems design [81]

Our goal is that our platform is at its core a human-centered approach. This means that as designers, we must: a) think about what people want to do rather than what the technology can do, b) design new ways to connect people with people, c) involve people in the design process, and d) design for diversity and be inclusive. Our aim is to make the system as usable and useful as possible, by focusing on the users, their needs and their requirements and by applying human factors, ergonomics, usability knowledge and human-centered methodologies.

One may ask at this point whether this whole human-centered approach is worth it in the end, or is it just a waste of time. Of course, one can easily understand that it is expensive and the reason behind that it that it involves observing people, talking to people and trying ideas out with people, and all this takes time. So, human-centered approaches are an additional cost to any project, so from their perspective, businesses rightly ask whether taking so much time to talk to people, produce prototype designs and so on is worthwhile. But in the end, the answer is a fundamental 'yes', as a human-centered design approach for interactive systems is advantageous for a number of reasons. The approach helps the final product, as it will enhance effectiveness, efficiency, user satisfaction, user experience, accessibility, sustainability and human well-being.

To this end, we must select a design framework that is human-centered and iterative at its processes (like the Double Diamond model). The aim of human-centered interactive systems design is to arrive at the best combination of the *People-Activities-Context-Technologies (PACT)* elements with respect to a particular domain. In our approach, we need to get the right mix of technologies to support the activities being undertaken by people in different contexts, meaning that A PACT analysis is useful for both analysis and design activities: understanding the current situation, seeing where possible improvements can be made or envisioning future situations. In order to perform a PACT analysis, we must study the variety of people, activities, contexts and technologies in our domain (platform). To do so, we will be using a variety of envisionment techniques (e.g. personas) in combination with working with people and interacting with them (e.g. interview). So, during the design phase, we must abide with the PACT analysis framework and to do so, with must pay focuses on the issues addressed in this human-centered framework, namely:

- For people, we need to consider the physical, mental and social contrasts and how those distinctions change in various conditions and after some time.
- For activites, we need to consider the intricacy of the action (engaged or ambiguous, straightforward or troublesome, barely any means or many), the worldly highlights (recurrence, pinnacles and box, constant or interruptible), helpful highlights and the idea of the information.
- For contexts, we need to consider the physical, social and authoritative setting.
- For technologies, we focus on input, yield, correspondence and substance.

In our case (development of a live-stream platform for criminal identification), we may narrow it down to the following categorization, as seen below in *Table 1*.

Table 1. Applying the PACT analysis model for the design of the face recognition platform

PACT category	Platform Specification
People	▪ The main groups targeted are police force employees in desk duties inside police precincts (Criminals are used without their knowledge through the cameras). ▪ Language differences may occur (need for language support) ▪ Visual disabilities may occur (need for inclusive design for color-blindness) ▪ No other stakeholders can possibly gain access (e.g. cleaning staff, password required to login)
Activities	▪ The overall purpose it for the system to detect in real time criminals through camera feeds installed (criminal pictures in relevant database). ▪ Login Credentials required from users. ▪ Employees can add/edit/delete a criminal's information in the system from the database. ▪ Employees can report issues, problems or concerns found in the platform directly. ▪ Employees can change the visuals on the platform according to their liking (light or dark theme, small, medium or big font size, different language support). ▪ Employees can open any camera feed in real time and apply different camera filters to the existing camera feed (e.g. Gray, Redish, Invert, Blur, and more). ▪ System Administrators have top level access and can also add/edit/delete other employees' information from the database.
Context	▪ Physically, all aforementioned activities take place indoors and access to the platform should be restricted for inside the police precincts (not safe to allow IP access from external sources outside the precinct). For the criminals' POV, the identification procedure takes place outdoors, depending on the location of the installed cameras. ▪ Socially, it can happen during the equivalent working hours of the employees and/or administrators (no time constraints in accessing the platform). ▪ Organizationally, the context refers to security/identification and who has access to which platform features.
Technologies	▪ Must be accessible from other media like tablets or mobile phones inside the precinct (responsiveness must be ensured). ▪ Communication with a central database is necessary to validate the login credentials, and control the information existing on employees (from the admins) and criminals (admins and employees). ▪ Must be online and accessible at all times inside the precinct, regardless of browser and operating system. ▪ Must be obvious on how to be used from people unfamiliar with such systems (need for action feedback, alert messages, confirmations, help sections and more).

3.2 Personas and Scenarios

At this stage of the design process and always following the PACT model throughout the development phase, we must describe our target audience/group and reflect their abilities, status, needs, visions, background history demographics. This means that we need personas. Personas are fictional characters, which we the designers create based upon our research in order to represent the different user types that might use our service, product, site, or brand in a similar way. By creating personas, we can better understand users' needs, experiences, behaviors and goals. It is a means that will help our design, as it recognizes that different people have different needs and expectations and keeps designers away from the notion that if they like it, the final users will also like it. Keeping the system design relevant to the personas' needs and goals will make the design task less complex and messy, it will guide our ideation process and help us achieve the ultimate goal of creating a platform offering the best UX and QoS possible [82].

The approach of developing personas and using them to center the design on them, as opposed to designing products, services, and solutions based upon the preferences of the design team, it has become standard practice within many human-centered design disciplines. This is because as mentioned above, personas are fictional characters and not real people, based on real data collected from multiple individuals. This can help designers be less cold on the human factor and add more of a human touch in the design of the final product, as constructing personas will help us ask the right questions and answer those questions in line with the users we are designing for.

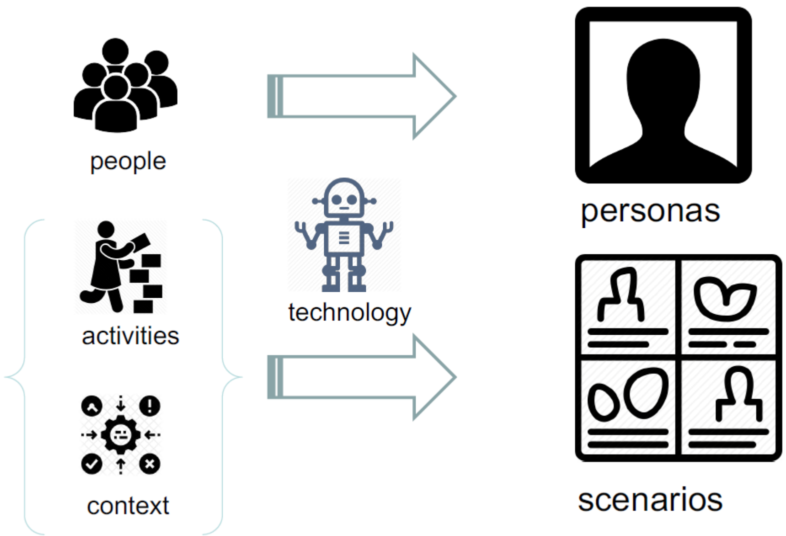


Figure 37. Transforming the PACT model into personas and scenarios [81]

To this end, based upon the research conducted so far and the data collected, we will be presenting the personas that reflect our target users in our platform application. We must assure that our design strictly follows the needs and goals of these personas, otherwise we will make the common mistake that many designers make, which is that they develop their personas but never take them really into account during the design process. Personas, combined with their respective usage scenarios (we will see this later on) will help us transform the PACT model and use it in order to make an actual impact in our design (see *Figure 37*).

<u>Persona #1:</u> *Manolis is a highly valued member of the police precinct in the city of Patras. He has been working as a police worker for over 20 years and has a great experience on any issues that arize. He is married and he has two kids, ages 10 and 12*

respectively. He lives with his family in the city of Patras and would never like to live the city, as he has spent his whole life here. He is not very adept to new changes in his life and likes to perform the same actions repetitively as he is sure that no mistake will occur.

Manolis Fouras | Police Employee

Responsibilities:

- *Desk Job, dealing with any reports or issues that arize (e.g. tracking stolen phones)*
- *When needed, regular patrolling the city to check for any misconducts*
- *Due to his experience, supervises many of his co-workers*

Behavioral Characteristics:

- *Will rarely use new technologies (not very adept with emerging technologies)*
- *Prefers to stick to old school practices*
- *Likes to feel in control of things*
- *Gladly helps co-workers to perform better in their everyday tasks*
- *Likely to go on daily walks to keep up with his physique*
- *Enjoys being praised on his work*

Key Drivers and Motivation:

- *Being a police employee was his dream from a young age*
- *Hopes to gain a promotion and a salary raise in the near future*
- *Experiences a healthy working environment around him*

Needs and Goals:

- *Wants to keep his job stability to provide the means for this family*
- *Would like to be awarded for his effors and expertise through future promotions*
- *Would like a higher salary due to his commitment to the cause*

Profile Attributes

Age: 40 years old

Experience: 20 years

Personal Information:

Manolis is a highly valued member of the police precinct in the city of Patras. He has been working as a police worker for over 20 years and has a great experience on any issues that arize. He is married and he has two kids, ages 10 and 12 respectively. He lives with his family in the city of Patras and would never like to live the city, as he has spent his whole life here. He is not very adept to new changes in his life and likes to perform the same actions repetitively as he is sure that no mistake will occur.

Figure 38. Persona #1: Manolis Fouras (Police Employee)

Persona #2: *Dimitra is a new addition in the police precinct in the large city of Athens. She is relatively new in the field, but is eager to keep learning and exploring new possibilities that can help her perform better in her work. She lives alone in Athens, along with a pet (dog) to keep her company. She enjoys using the latest tech available (expensive mobile phones, tablets, TVs) and likes to change ways of living on in order for her life not to boring.*

Dimitra Adamantiou | Police Employee

Responsibilities:

- *Desk Job, dealing with any reports or issues that arize (e.g. tracking stolen phones)*
- *No patrol duties assigned to her so far*
- *Under supervision due to the absence of experience*

Behavioral Characteristics:

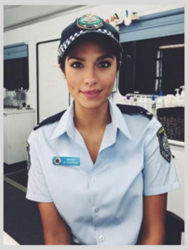
- *Enjoys latest tech innovations (phones, tablets, TVs)*
- *Eager to keep learning day by day*
- *Eats healty and goes for regular jogging and bicycle riders*
- *Prefers to commute by train and not buy car or motorcycle*
- *Respects her work and wants to impress in the field*

Key Drivers and Motivation:

- *Needs the job in order to sustain herself living alone*
- *Hopes to gain a promotion and a salary raise in the near future*
- *Experiences a healthy working environment around her*

Needs and Goals:

- *Wants to keep her job stability to be able to live alone in Athens*
- *Would like to be awarded for her actual contribution and not for just being a woman*
- *Would change field of work to find a job that offers more excitement and opportunities*



Profile Attributes

Age: 25 years old

Experience: 5 years

Personal Information:

Dimitra is a new addition in the police precinct in the large city of Athens. She is relatively new in the field, but is eager to keep learning and exploring new possibilities that can help her perform better in her work. She lives alone in Athens, along with a pet (dog) to keep her company. She enjoys using the latest tech available (expensive mobile phones, tablets, TVs) and likes to change ways of living on in order for her life not to boring.

Figure 39. Persona #2: Dimitra Adamantiou (Police Employee)

For the transformation of the PACT model towards the design phase to be a success, the personas themselves are not enough. In HCI, a persona is most typically combined with one or more scenarios. In simple terms, a scenario is a brief story that describes how and why the persona would use the product to complete a specific task in a specific context. The purpose of a scenario is to help us visualize how a target user would interact with the final product (in our case, the online face recognition platform) in real-life. This can help with determining the functions and features of the product for design requirements, as well as with developing tasks for usability testing [82]. To this end, we will describe an indicative scenario for both personas explained above.

Manolis Fouras | Police Employee

Scenario: Attempting to change a criminal's records in the platform

- *Manolis is not very familiar with new software or innovative tech designs.*
- *The platform has been installed inside the polic precinct for testing purposes.*
- *Manolis is asked to alter the records of a criminal using the platform and change the criminal's list of crimes and add newly found crimes).*
- *He struggles to identify where he must go in the website.*
- *He asks helps from some of this young co-workers who are better with such computer skills.*
- Ultimately, he succeeds at the task set to him.
- Yet, he complains that it would have been faster if he hadn't spent all this time searching the platform and just adding notes in the criminal's physical printed file and store it in the precinct's records.

Behavioral Characteristics:

- *Will rarely use new technologies (not very adept with emerging technologies)*
- *Prefers to stick to old school practices*
- *Likes to feel in control of things*
- *Gladly helps co-workers to perform better in their everyday tasks*

Figure 40. Scenario for Persona #1: Manolis Fouras (Police Employee)

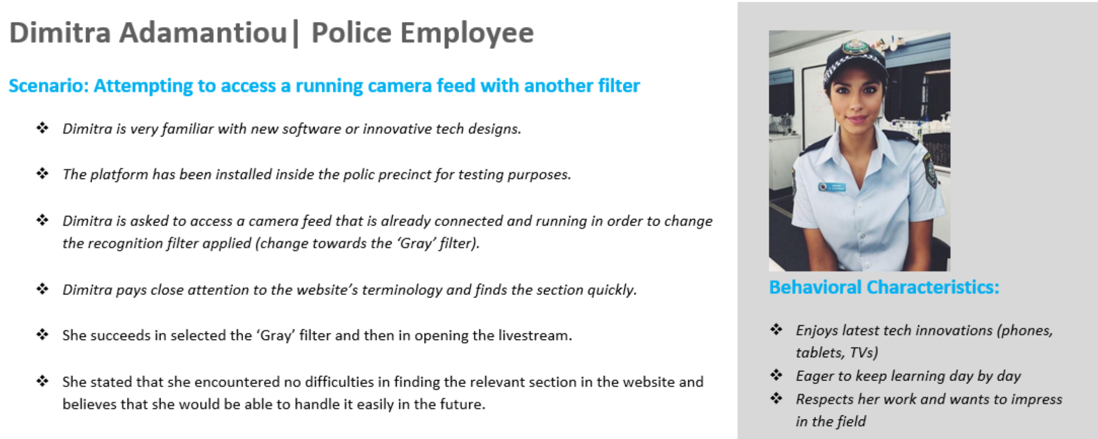


Figure 41. Scenario for Persona #2: Dimitra Adamantiou (Police Employee)

3.3 Requirements

At this stage, before our creative design of the platform can start and take shape and form, we must develop a clear understanding a clear understanding of the people involved with our platform, the activities that will take place inside the platform, the context of the activities and the implications of the technologies included. This refers to the PACT analysis model and we have already performed this analysis (see section **3.1 Selecting a human-centered framework**). As a result, from this understanding, we as designers of the platform can generate the requirements for the system that is to be designed. The focus of this procedure is to understand what people might want to do, how they will do it and why they want to do it (problem encountered are taken into consideration too). So, at this stage of the design phase, we will perform a requirements analysis. This will help us during the later stages of design by transforming these requirements into a service platform, which is the ideation process and requires a relevant creative leap from the developers of the platform.

The first phase of requirements will categorize them into two (2) different categories, functional and non-functional requirements. The functional requirements refer to want our platform must have, whereas the non-functional requirements refer to a quality that our platform should have and are both crucial factors regarding the acceptability and usability of our final product.

Table 2. Functional and Non-Functional Requirements

Functional	Non-Functional
▪ Platform Database ▪ Server Acquisition ▪ IP cameras	▪ Responsiveness support in other media screens ▪ Change theme support (Light/Dark/Colorblind) ▪ Change font size Support ▪ 'My profile' settings ▪ Credentials authorization ▪ Livestream support ▪ Edit user/criminal records

3.4 Interviews

For us to fully understand what future customers might need from the platform and challenges they might face, it is crucial to talk to them through interviews. Since what we were tasked to do is an innovative future design (especially in Greece, such inter-connnected platforms do not exist in any police precinct), we decided to conduct a semi-structured interview, because we strongly believed that even though pre-prepared questions could easily apply, new topics could arise during the conversation with the stakeholders and we should be in a position to build upon such new topics without being too strongly attached to the questions drafted beforehand. Using this approach, we ultimately managed to obtain more rich data, suggestions and problems through the semi-structured interviews we performed and of course, we constructed a basis of some pre-defined questions applicable to all stakeholders so that we could draw feedback and perform some analysis. The semi-structured interview was the method we used in the discovery phase. We believed that it would also be to our advantage if we managed to observe people in their working environment without having to formally research them, but of course, this is something that is not applicable inside a police precinct for security reasons.

We interviewed 10 people in total inside the local police precinct in the author's city (city of Patras). Luckily, the author had a personal contact already working inside the precinct, which made the procedure of communicating and making police officers agreed to be interviewed faster and easier. We made a direct discussion with these individuals, we sat down with them in order to understand their needs. During the interview, we made the questions to the participant and asked for permission to record the sessions. This was because there was no other person to write down notes during the interviews (one person could not do both of them effectively) and as a result, notes were extracted during re-hearing the recordings. The conversation with each participant helped understand the needs, practices, concerns, preferences and attitudes of the people who might interact with our system. The discussion started with a set of questions which may be similar to questions in a fully structured interview and then we were free to let the conversation go where it may. For example, if we considered that the interviewee mentioned something important, interesting or something that needs more explanation, we asked for them to tell more about that. It goes without saying that no interviews were held with any kind of criminals.

In order to be able to distract employees from their work during times where they did not have much desk work, after consulting the author's personal contact, we visited the police precinct two different times at 14:00, so that we would minimize the risk of heavy workloads during this hour. Moreover, we made sure that at least one other employee would be available at that time, so that even if something unexpected came up, there would be support from another employee. A representative example of one of the semi-structured interviews we performed is the following, where we interviewed a 38 years old employee at the precinct (the interview was conducted in Greek and is presented below in English translation):

Interviewer: So, would you say that your precinct has the equipment required?

Employee: Well, I wouldn't say that, no. We have old computers and they all run on old software.

Employee: Would you like your precinct to get state-of-the-art equipment?

Interviewer: I don't know, it's a difficulty question to answer to. One the one hand, I have been using the computer for many years now and even though I do not do very heavy tasks on complicated things, I am still not very familiar with new software or websites. From the other hand, even though it would be difficult in the beginning, I guess new features could be in the new software that could help us in our jobs rights? Like automation tasks for example?

Interviewer: Well yes, the core idea is something like that. Instead of you looking at the camera feeds in order to try and identify a criminal, imagine an interconnected network of cameras, where you get automatic feedback whether a criminal has automatically been identified and on top of that, you may use a specific website in order to change the information of that criminal instead of searching physical files in storage rooms.

Employee: This sounds quite innovating for us here in Patras and certainly in Greece. Up until now, we had to rely on the peoples' quality of the cameras installed inside their shops or in the streets and since not all of them tend to buy high quality camera infrastructures, we have big problems in the identification procedure, as we have to compare ourselves the camera screenshots and the pictures we have in the printed files.

Interviewer: Exactly this limitation is one of the main motivations behind this platform contribution. By creating a website and not an application that would run in your desktops only, not only you can use the website regardless of the operating system of your desktop (e.g. Windows XP - 10, iOS, Debian Linux), but you will also be able to access the website through various media inside your precinct. For example, you will be able to login through your phone, or through a tablet.

Employee: This sounds quite cool and innovative, this could actually prove quite useful for us in the future. As long it is accompanied by a comprehensive user manual, we will be doing ok with these…(*laughs).

3.5 Statistical Analysis

In this section, we present the most important aspects of the information gathered from the interviews (we remind that they happened on two different days in order to attract more stakeholders). Firstly, we present a statistical analysis regarding the age of the stakeholders that were interviewed, revealing that the vast majority of the stakeholders interviewed were of ages 40 to 59.

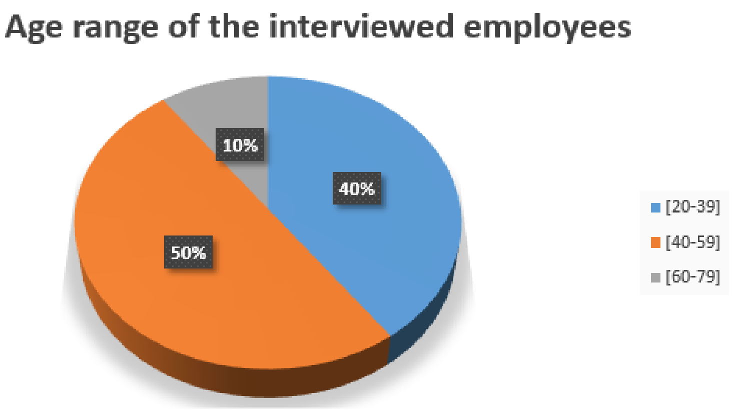


Figure 42. Pie Chart for the interviewed stakeholders' age

The likelihood of employees still relying on their existing software equipment was also a necessary factor for our research. During the analysis, we discovered that most of the interviewed stakeholders (60%) stated that they would still rely on their existing equipment, whereas only the 40% of them would trust state-of-the-art equipment and under certain circumstances.

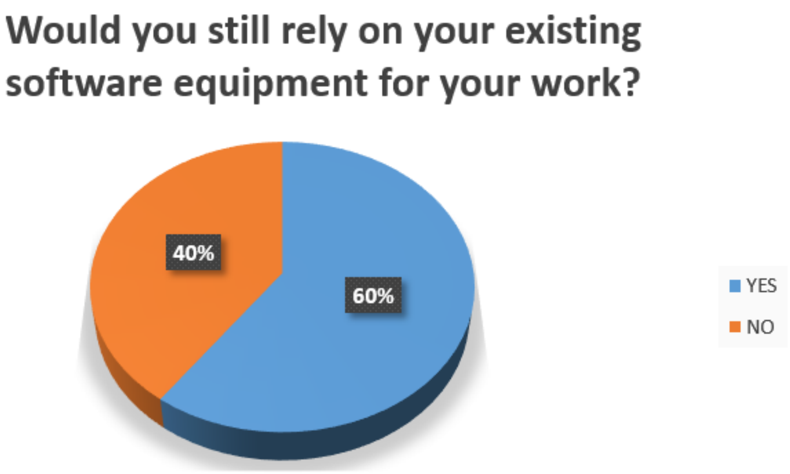


Figure 43. Pie Chart for the employees that would still want to use their existing software equipment

One also very important aspect that requires thorough interpretation is the concerns and fears reported from the stakeholders. This was one of the most important aspects that we took under consideration when we started the design of the final product. All challenges reported below were the basis when we started designing our product. As we can see, the majority of the interviewed stakeholders state that the most common problems would be that they do not believe that the new they would be as effective as presented to them and that such approaches can only work in highly technologically developed countries. Some other concerns raised were that people themselves are not good with computers (so any kind of new technology would greatly make their job even harder), they do not like new technologies and that it would be very difficult to get all local people to agree in this interconnected approach with the envisioned camera network.

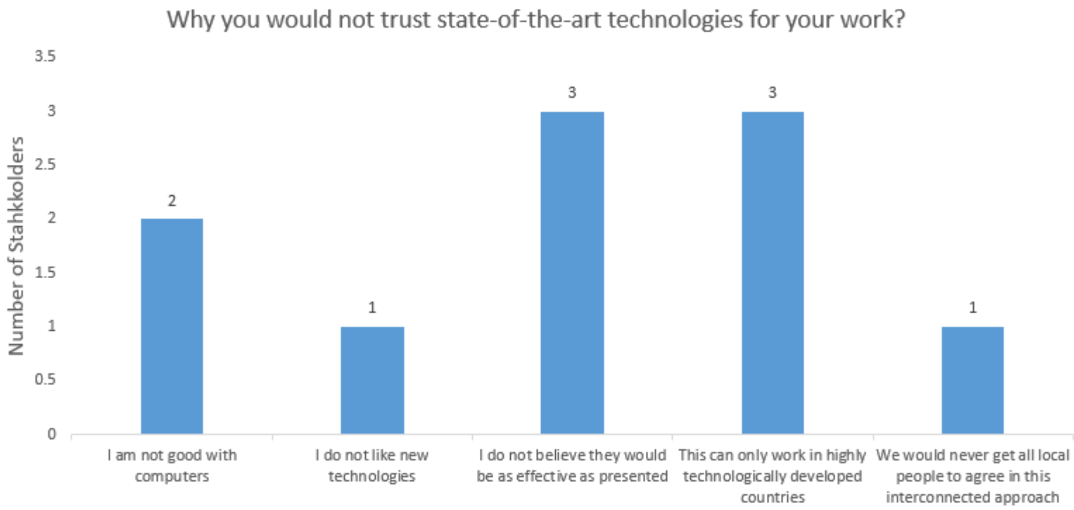


Figure 44. Bar Chart for the employees fears towards state-of-the-art technologies for the precinct

All this rich information will prove crucial towards our next steps in the design phase. Such information must be our guide and we must remain focused towards our main goal, which is the development of not just any platform, but a human-centered platform, caring for the needs and goals of all stakeholders that will ultimately be the end users.

3.6 Draft Version

Using all the information gathered so far (relevant research on the topic, accordance with the PACT framework, keeping in line with our personas and scenarios, feedback from the interviews), we are presenting the draft version of the platform developed at this phase of the iterative design process. This can be interpreted as a 'mockup' for the final platform, meaning that not all functionalities appear at this stage of development, many additions and revisions will still occur in the latter stages of development and this serves only as a general guide for our ongoing work on the livestream face recognition platform.

3.6.1 General Information

All existing face recognition approaches can be used, extended and efficiently deployed for a wide variety of application in our everyday lives, ranging from outdoor face recognition for crime detection up to indoor object recognition and pattern matching. Our platform contribution is an extension of the existing facial identification approaches by proposing an online platform/website that can be used from the police forces. This approach builds upon existing work and suggests not just an algorithm, but a complete systemic approach through a crime detection platform to be used in police headquarter/precincts. The platform considers the Haar Cascade algorithm for the facial recognition detection and extends it by performing real-time recognition from the connected cameras in the system. The platform supports two different types of users, namely a) police employees in the headquarters/precincts and b) police administrators, with a higher level of access and also responsible for database maintenance. The main features of the platform include:

- inserting, editing and deleting user information
- inserting, editing and deleting criminal records
- applying 6 different filters in the livestream cameras (e.g. gray, sepia, redish, blur, inverted)
- searching for criminals based on their picture through a livestream camera feed and identifying them

As for the criminal identification, it is made possible through image pattern recognition between the provided criminal's image and snapshots of identified faces from the livestream feed. The Live Feed section offers six different options for video filters, enabling the user to select the best filter, depending on the relevant situation of the physical surroundings (e.g. broad light, dark environment). A large proportion of the research interest is driven towards extending the Haar Cascade algorithm in order to produce more accurate or faster recognition. Yet, little work has been conducted regarding the best real time recognition filter for human detection, especially regarding the usage of Haar Cascade for the recognition procedure. The live feed facial recognition approach can be easily explained by looking at *Figure 45* below.

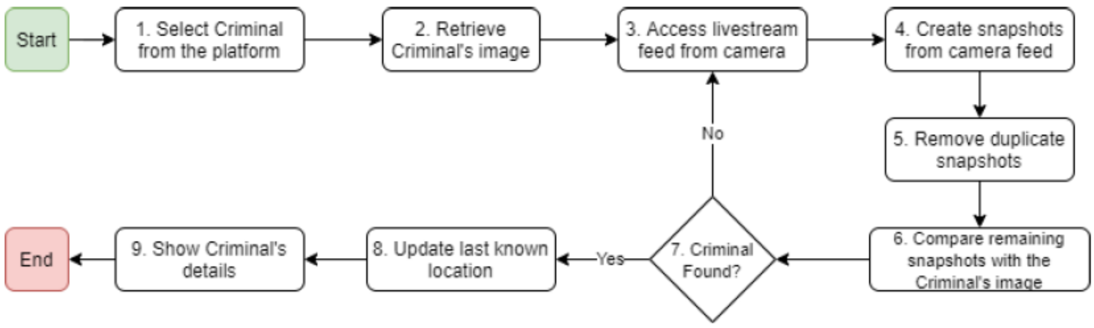


Figure 45. The live feed facial recognition approach

3.6.2 Technical Specifications

In the table below, the technical specifications for the livestream platform are listed, offering adequate information on the tools that made this platform come online possible.

Table 3. The technical specifications for the platform

Technical Specifications	
Integrated Development Environment (IDE)	InteliJ PyCharm 2020.3.5 Community Version
Core programming language	Python 3.6.0
Secondary programming languages	HyperText Markup Language (HTML), Cascading Style Sheets (CSS), Boostrap 4.0.0

WEB framework	Flask 1.1.2
Relational Database Management System (RDBMS)	MySQL (Xampp)
Server Type	Apache (localhost)
IP Camera Models	Smartphone Xiami Redmi Note 7, Tablet Huawei Mediapad T3

3.6.3 Building the Database

The database has been constructed by using three different tables, the '*users*' table, the *'criminals'* table and the *'contact'* table. The '*users*' table holds information regarding the name of the users, their email and their subject. The *'criminals'* table holds information regarding their demographics (e.g. name, age, height, eye color), a short bio of their criminal history, their photograph and their last known location. The *'contact'* table holds information for any issues or concerns raised or encountered by the users of the platform. The SQL structure of these tables is presented in the following figures below.

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	contact_id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	first_name	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
3	last_name	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
4	email	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
5	subject	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More

Figure 46. The 'contact' table

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	criminal_id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	full_name	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
3	age	int(11)			Yes	NULL			Change Drop More
4	height	float			Yes	NULL			Change Drop More
5	weight	int(11)			Yes	NULL			Change Drop More
6	eye_color	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
7	biography	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
8	portrait	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
9	last_location	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More

Figure 47. The 'criminals' table

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	user_id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
2	username	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
3	password	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
4	email	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
5	full_name	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
6	role	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More
7	avatar	varchar(255)	utf8_general_ci		Yes	NULL			Change Drop More

Figure 48. The 'users' table

3.6.4 Deploying the draft website

The project structure of our platform looks like this:

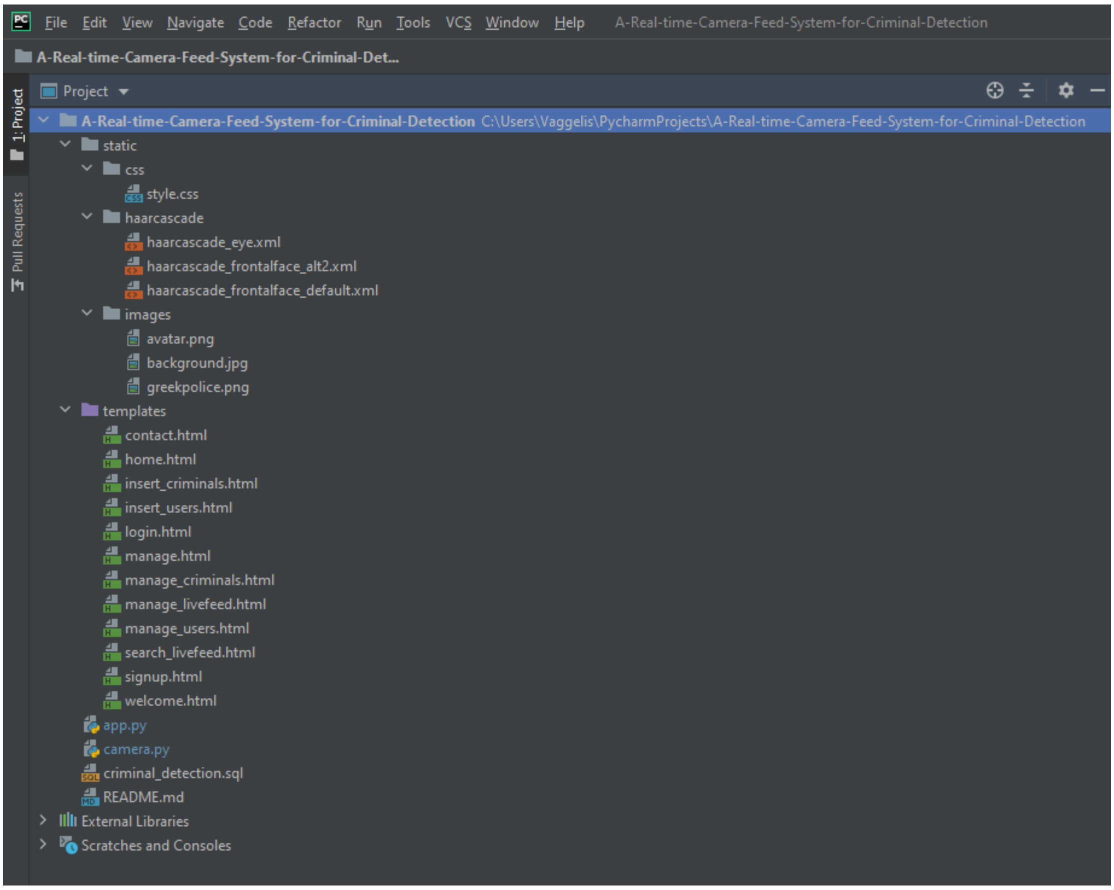


Figure 49. The project structure of the draft website platform

The steps to deploy the draft website are the following:

1. **Creating the database**
   You may login in any MySQL database management panel and create a new database. You may use the '*criminal_detection*' name for your database. Make sure you select the UTF-8 encoding (collation *utf8_general_ci*). Both *app.py* and *camera.py* files include the following line for a connection to the MySQL database (You may update this according to your localhost setup and credentials):

```
mydb = MySQLdb.connect(
db="criminal_detection",
host="localhost",
user="root",
passwd=""
)
```

2. **Populating the database**
   Select your database (e.g. the *criminal_detection*) and import the *criminal_detection.sql* file, which includes the creation of the tables and the data population

3. **<u>Running the *app.py* file</u>**
   This is the main file which includes the necessary app routings towards the HTML pages, alongside with the relevant data processing. The *camera.py* file is only called in the livestream page.

4. **<u>That's it! You have successfully deployed the website...</u>**

When launching the website for the first time, we are greeted with the welcoming page of the platform. As stated in previous sections, login credentials are required to enter the platform.

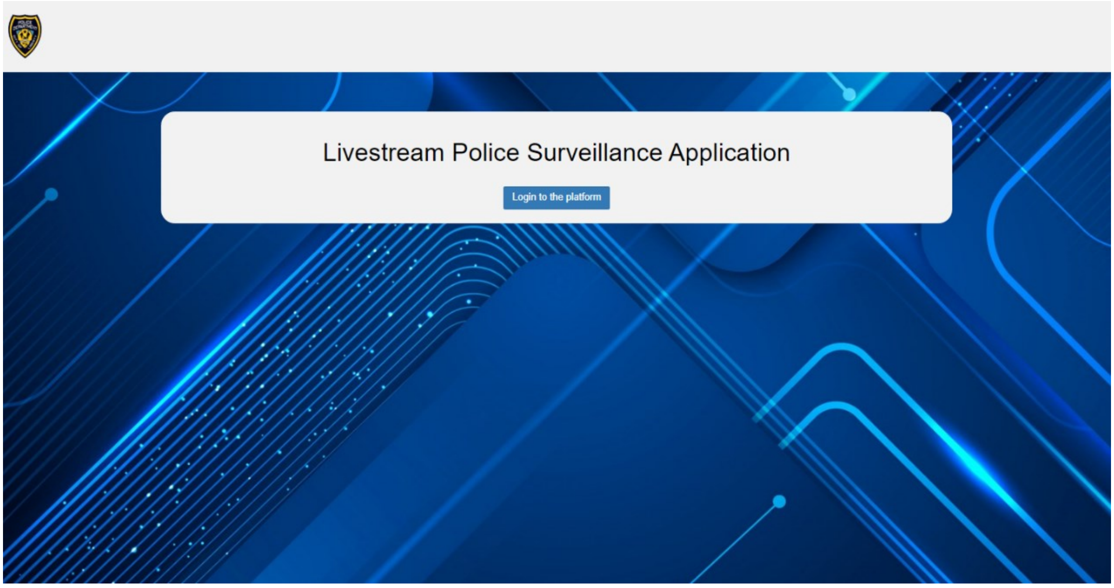


Figure 50. The welcome page

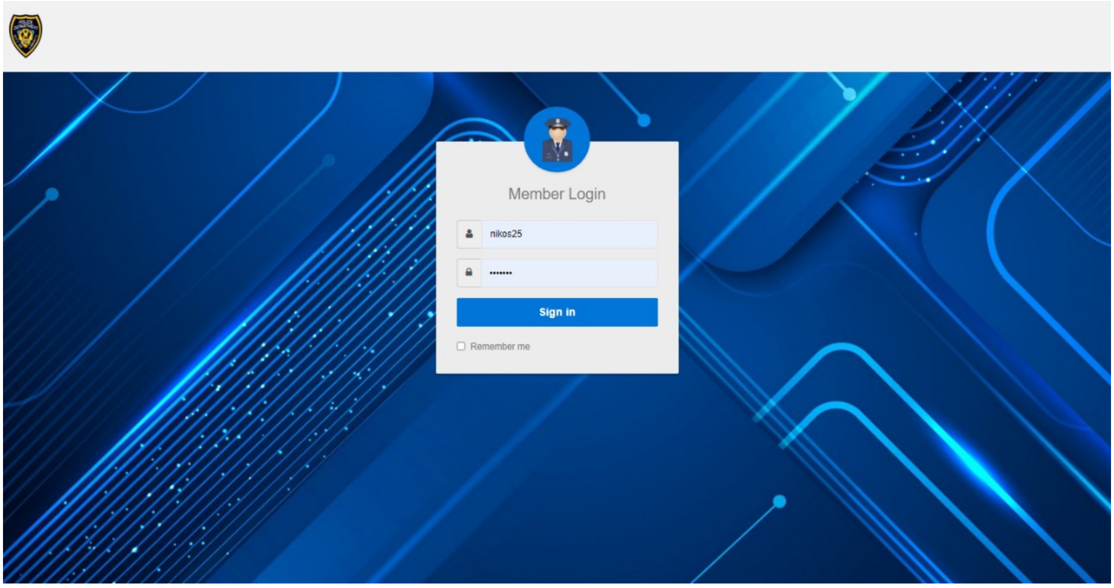


Figure 51. The login page

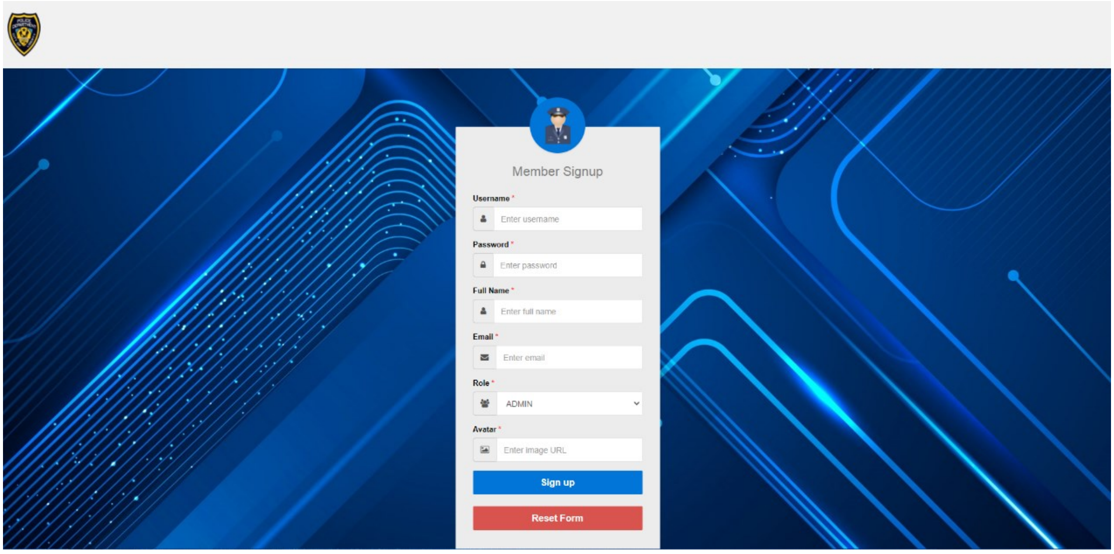


Figure 52. The signup page

Upon correctly entering our user credentials, we can now view the main page (Home page) of the website. As we can see below, we are introduced with relevant information on what we can do in the platform and we can access the platform's features either from the hyperlinks in the '*Features*' section or from the website header. Users can use all pages to change the language from the right-hand size of all containers.

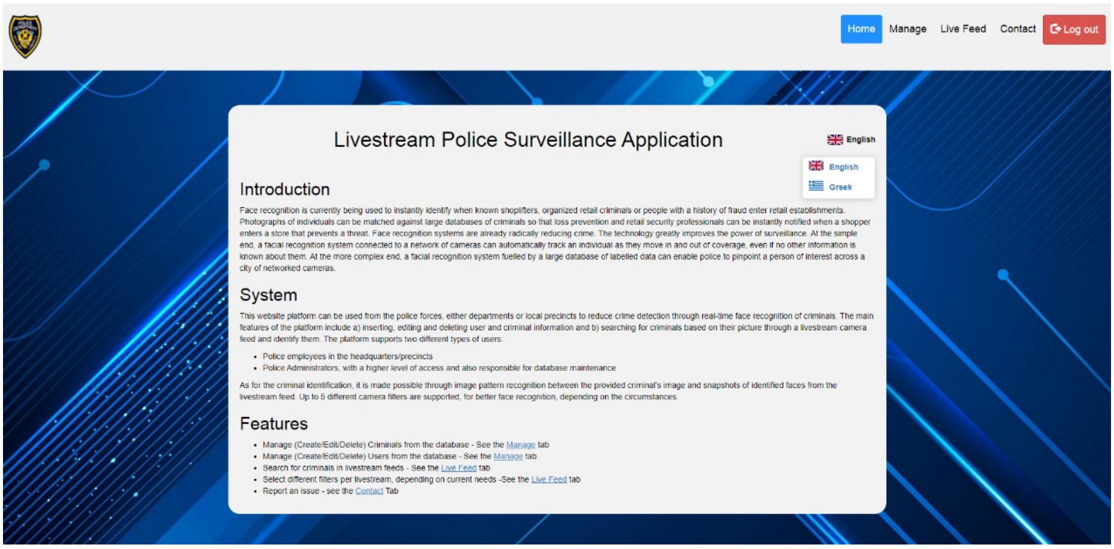


Figure 53. The 'Home' page

Let's start by accessing the Management Panel on the website. In this section, according to the pre-defined specifications, only the system administrators can add, edit or delete currently registered employees on the platform, as letting the employees themselves have that kind of access would be a high level security breach and potentially harmful for the precinct. The rights to manage (add/edit/delete) criminals in the database is available to any kind of user (admin/employee).

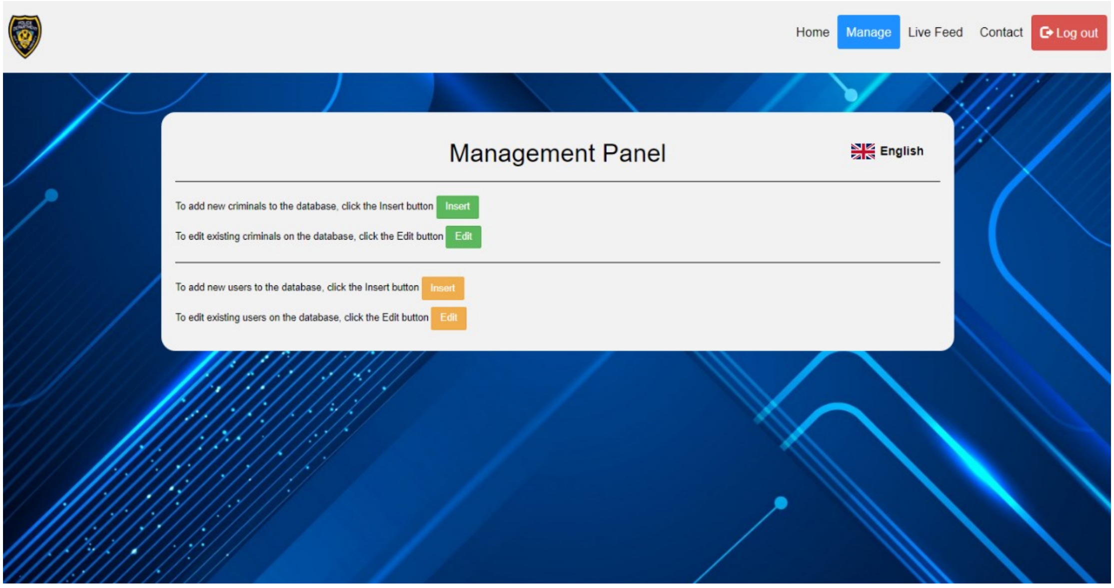


Figure 54. The 'Manage' page

The UI design for the insertion of either users or criminals follows the same core approach. Most of the fields for the criminals are required in the form in order to be submitted, except from his last known location, which can be unknown at this stage.

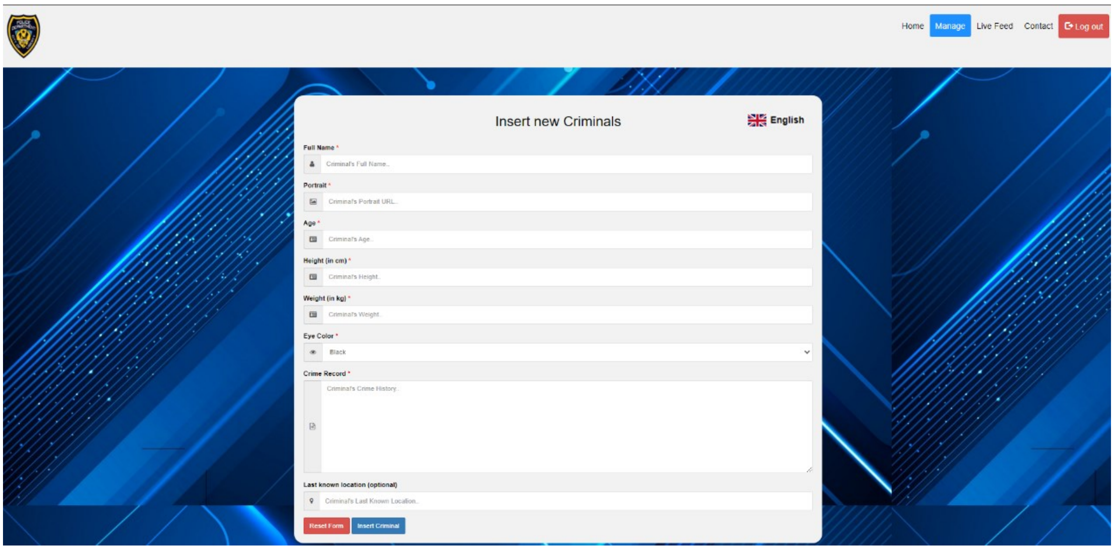


Figure 55. Inserting new criminals in the database

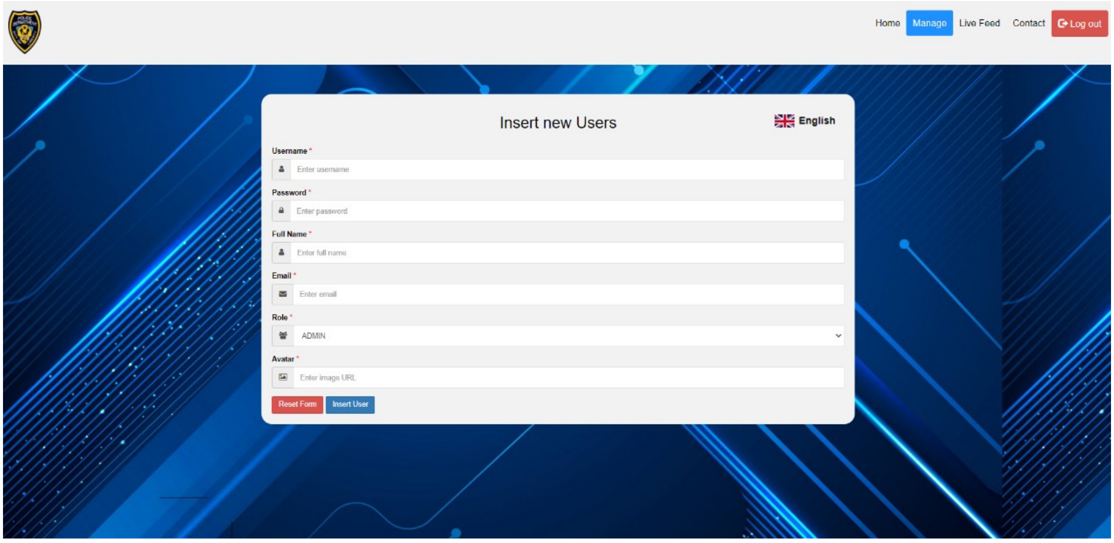


Figure 56. Inserting new users in the database

Upon insert a criminal or a user in the database, we are greeted with a confirmation message that everything was ok with the new record and we are automatically headed towards the equivalent management page (either for criminals or users), where we can either edit or delete existing records. For example, we attempted to create a new employee with some sample tests as input and as we can see, the record appears directly in the website (of course, all passwords of other employees are hidden for obvious reasons, yet can be changes by the administrators).

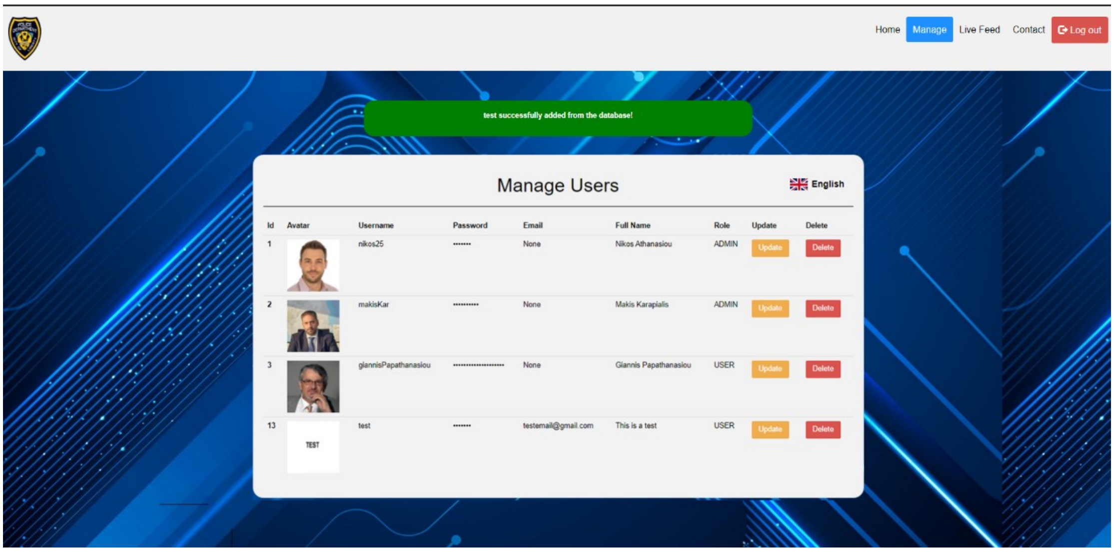


Figure 57. Inserting a new user 'test' in the database

Now, let's suppose that we want to edit this newly created user 'test' and update his records. Upon pressing on the '*Update*' button in his row, a modal (in simpler terms, a popup message) appears with the user's information, allowing us to make any changes we want. The exact same approach also appears in the criminal's relevant pages. Upon updating his records, again a relevant message pops up, informing us whether everything went OK with the update or not.

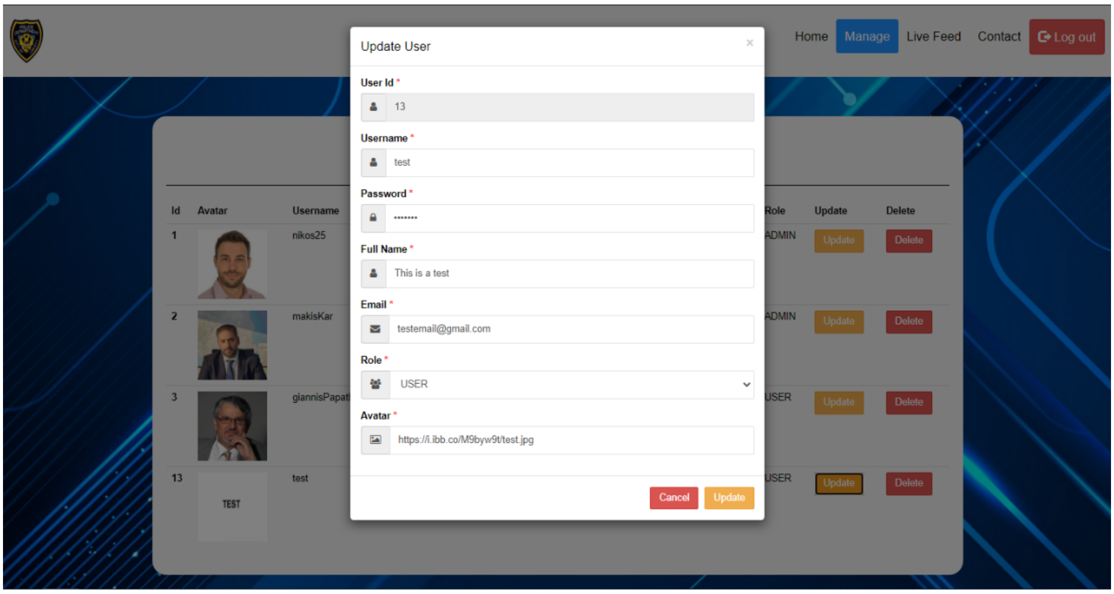


Figure 58. Updating the user 'test' in the database

The '*Contact*' page following quite a simplistic approach, requiring only the name, surname email and subject to be sent to the system administrators and recorded in the database for any issues/concerns encountered in the platform.

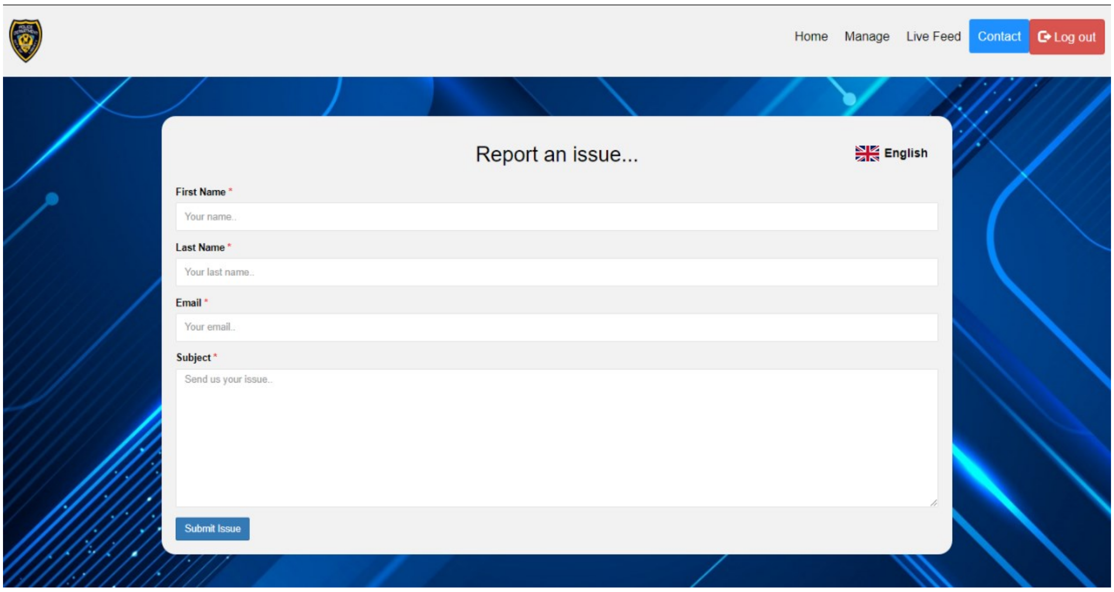


Figure 59. The 'Contact' page

The criminal identification is made possible through image pattern recognition between the provided criminal's image and snapshots of identified faces from the livestream feed. The Live Feed section offers six different options for video filters, enabling the user to select the best filter, depending on the relevant situation of the physical surroundings (e.g. broad light, dark environment).

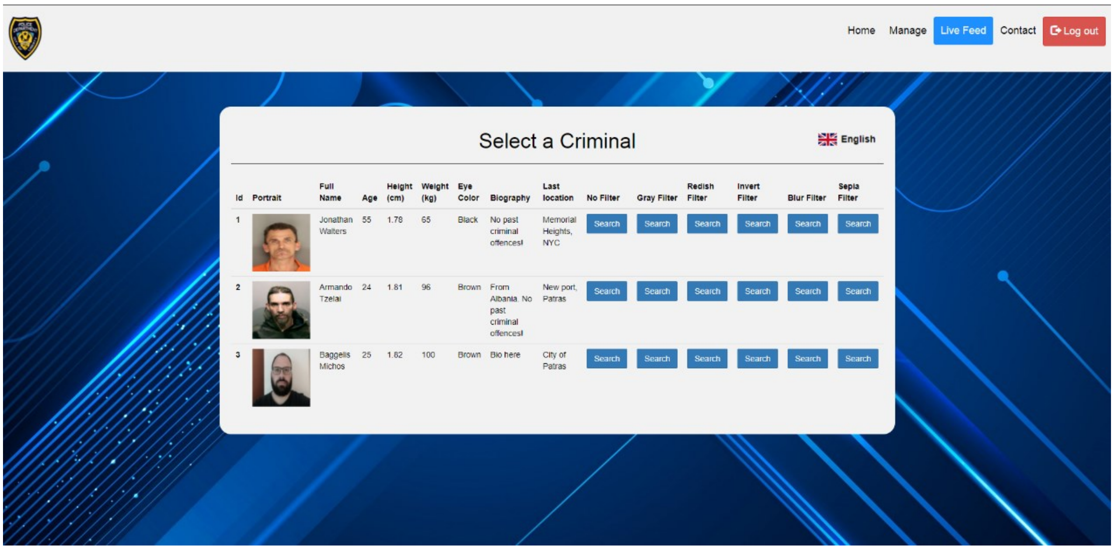


Figure 60. The 'Live Feed' page, selecting a criminal

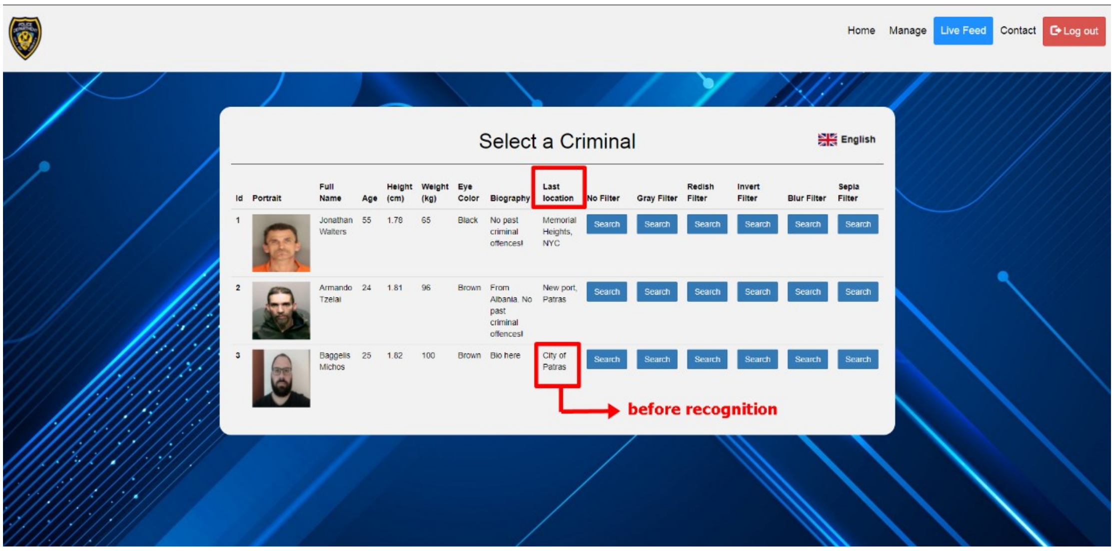


Figure 61. The 'Live Feed' page, selecting a criminal

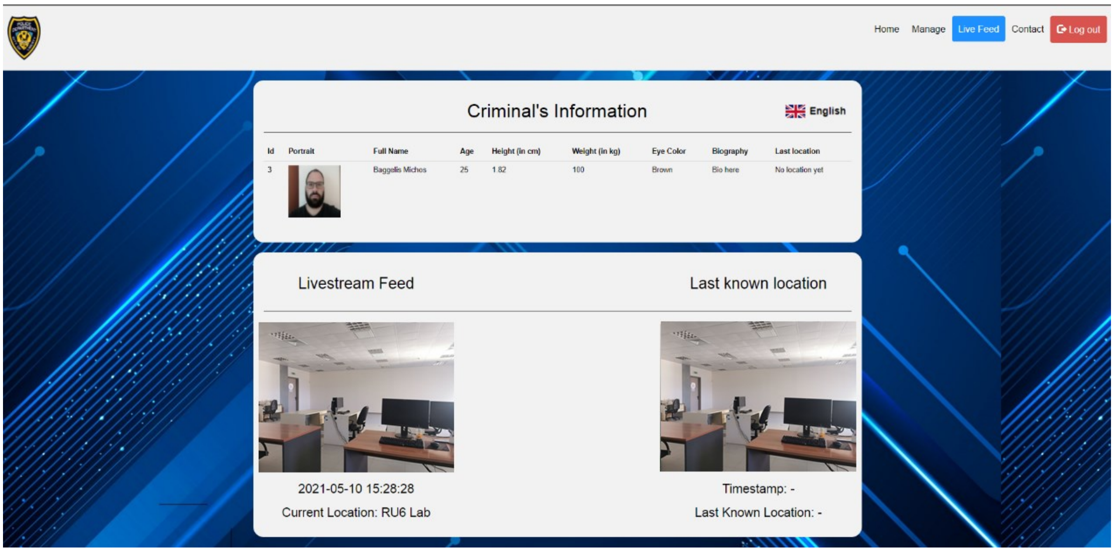


Figure 62. The 'Live Feed' page with the live camera feed, before recognition

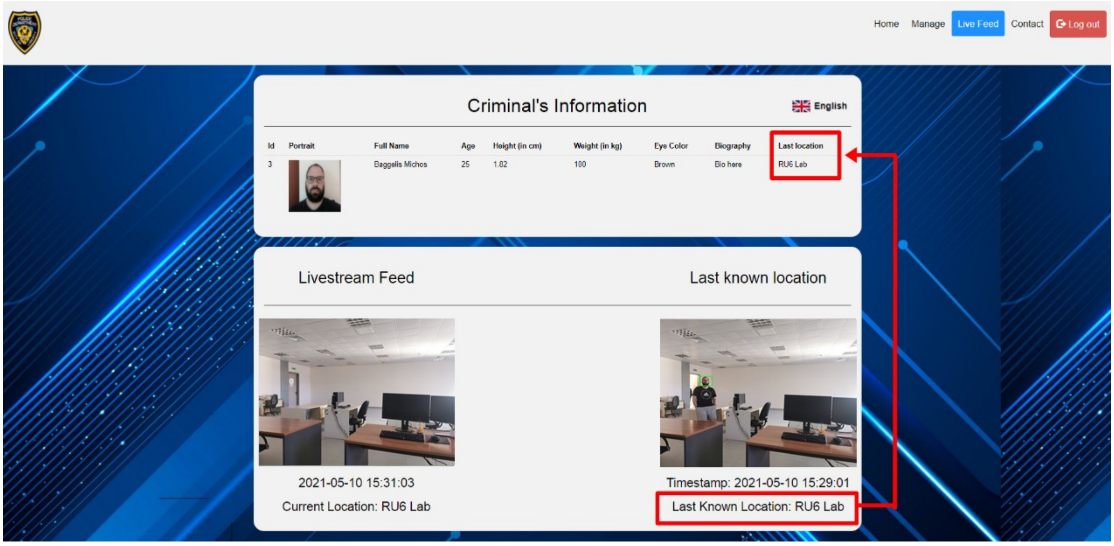


Figure 63. The 'Live Feed' page with the live camera feed, after recognition

Regarding the face recognition part of our platform, each time a criminal is recognized, a snapshot is taken directly from that camera feed. The six (6) different filters that can be used for face recognition are the ones depicted in *Figure 64*. All snapshot are categorized in the specific folder created for that criminal, where snapshots are saved inside the folder and re-categorized depending on the camera that took the snapshot. This helps us, as we can retrieve his last known location by knowing the location of the camera that took the screenshot. Lastly, the image name contains the date and time and filter used, as seen below in *Figure 65*.

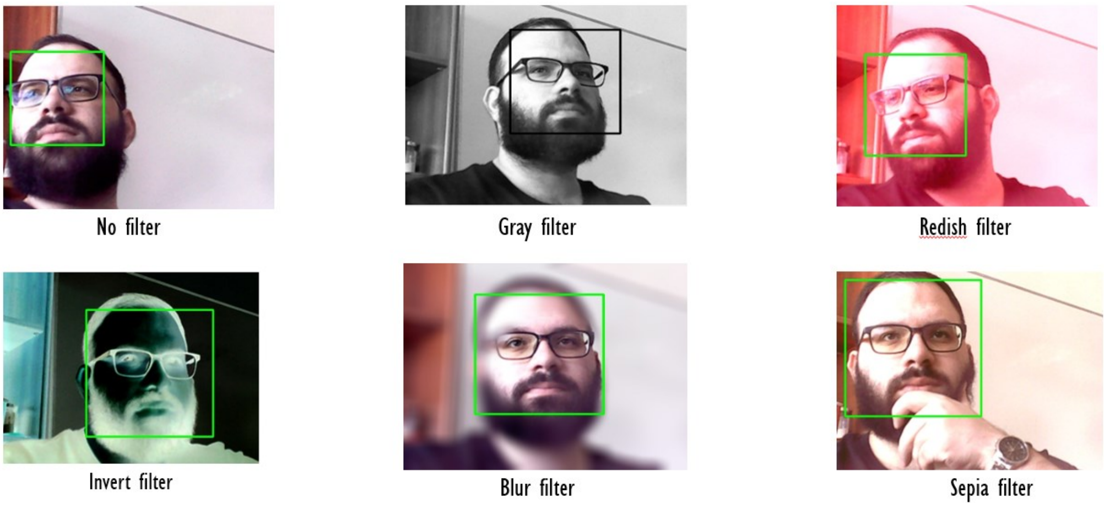


Figure 64. The applicable filters in all livestream cameras

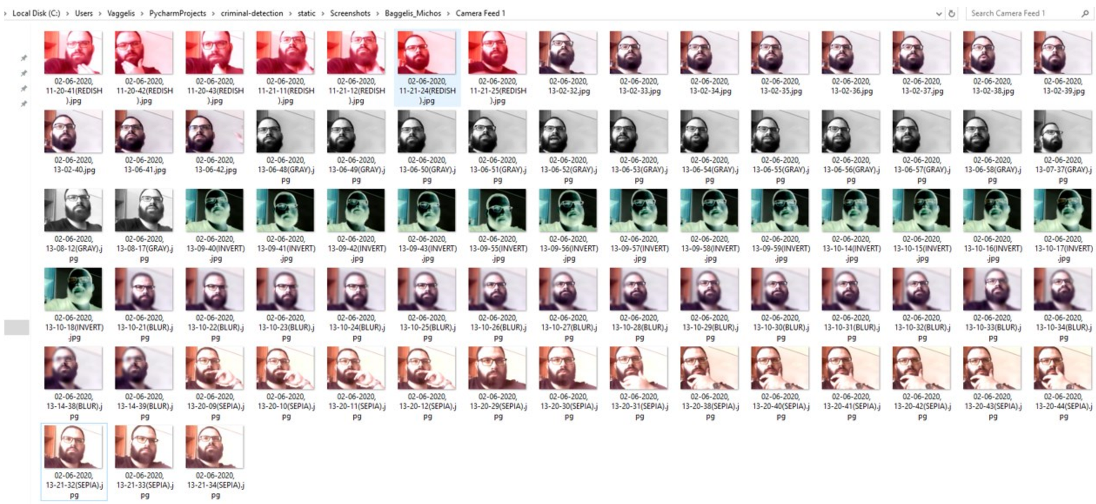


Figure 65. Saving facial recognition snapshots (Date + Filter in file)

3.7 Heuristic Evaluation

A *Heuristic Evaluation (HE)* was conducted in order to evaluate the draft version of the platform. Through this HE, we will enumerate the findings from the study of the HCI evaluators gathered, which offer their insights and recommendations that aim to enhance the user's experience for our final platform.

In general, HE is a usability engineering method that requires no actual users. Instead, a small set of evaluators examine an interface according to its compliance with Jakob Nielsen's ten usability principles, or heuristics. In this way, usability problems in an interface can be diagnosed extremely quickly and effectively. According to Nielsen, there are ten (10) in total heuristics that can improve the final product [83].For the complete list of these heuristics, see **Annex I – HEURISTIC EVALUATION REPORT**. There are many ways in which a rating scale can be the source of biases. Careful consideration of the properties of a scale can reduce these biases and thereby improve the accuracy of ratings. We follow the severity scale proposed by Nielsen, which includes five points associated with numerical values: 0 not a problem; 1 cosmetic problem; 2 minor problem; 3 major problem; and 4 usability catastrophe, as seen below.

Severity Rating	Meaning
0	*Team does not agree that issue impacts system usability*
1	*Cosmetic problem only; need not to be fixed unless extra time is available on project*
2	*Minor usability problem; fixing this should be given low priority*
3	*Major usability problem; important to fix, so should be given high priority*
4	*Usability catastrophe; imperative to fix before product can be released*

Figure 66. The different types of Severity Rate and their Meaning

Based on the aforementioned information, we created a basic template for the HE, which was disseminated to the experts to conduct the analysis. The template can be found in **Annex I** (see the last pages of this thesis for all Annexes). According to Nielsen, 3-5 usability evaluators are enough to identify 80% of the existing usability problems and adding even more evaluators would not identify significantly more problems, while the most serious ones are already identified. Thus, we managed to convince four (4) HCI evaluators (aged 21-27) that agreed to perform the HE process on our draft version of the website. Each evaluator individually and independently evaluated the application, so that they would not influence one another. The ten (10) heuristic mentioned above were used from the evaluators as a way to identify usability problems and to document them in a common way, enabling us (the receiving end) to have a short list in our hands and avoid extremely long lists of more specific guidelines. Each of the evaluators filled in his own evaluation report template, matching every finding to the relevant heuristics violated and assigned a severity score to each finding. Additionally, each evaluator also suggested ways in order to tackle the issues mentioned. As the designers of the platform, the merged information received from the evaluators provided us with a weighted list of findings (according to their severity), alongside with propositions on how to resolve the issues mentioned. By conducting the HE, we were able to analyze the draft platform's usability, functionality and design in a standardized structure. This proved that heuristic analysis is still an extremely valuable tool that should definitely be utilized early and often for all testing phases of the UX process. To close this, we strongly believe that if all developers pay attention to the issues discovered and give effort into fixing them (using the provided guidelines or not),final products will greatly improve, as far as UX and usability are concerned.

- **<u>Issue 1:</u>**

<u>Description:</u>
I did not like the fact that the dropdown to change the language was in every page, it was repetitive and annoying and sometimes distracted me. I do not see the reason why this option was on every page and not just edit it one time.

<u>Screenshot of the issue:</u>
See *Figure 53*.

<u>Recommendation:</u>
Developers could just let the user change the language either on the bottom of the page, or by using a Settings category.

<u>Heuristic Problem Category:</u>
1☐ 2☐ 3☐ 4☐ 5☐ 6☐ 7☐ 8☒ 9☐ 10☐

<u>Problem Severity</u>
(1) - Cosmetic problem only; need not to be fixed unless extra time is available on project

<u>Designer's Solution:</u>

In the refined version of the platform, we added a '*My Settings*' section on the top-right corner of the website. The header of the website now includes a welcoming message to the user (did not exist in the draft version) and by clicking on it, users can select the '*My Settings*' section or they can simply log out of the platform. In order to help user reduce the amount of time needed for settings tasks, the 'My Settings' and '*My Profile*' sections appear inside the same webpage, to reduce time needed to change between them.

- **Issue 2:**

Description:
There is no point in having different pages to insert users/criminals and to edit users/criminals. Developers could just add a "Add new user/criminal" button in the relevant pages and use a pop up window for the information.

Screenshots of the issue:
See *Figure 56* and *Figure 57*.

Recommendation:
Developers could merge these relevant user/criminal pages together as one to save up space and keep the relevant information together in only one page.

Heuristic Problem Category:
1☐ 2☐ 3☐ 4☐ 5☐ 6☐ 7☐ 8☒ 9☐ 10☐

Problem Severity
(1) - Cosmetic problem only; need not to be fixed unless extra time is available on project

Designer's Solution:
We took under consideration the following issue and re-designed the website. Now, the management panels include the Insert Users/Criminals option in the same page with the update/delete Users/Criminals, which in overall, is indeed a better approach.

- **Issue 3:**

Description:
There is no kind of confirmation alert when attempting to delete a record. This is a serious flaw, as the button can be pressed accidentally.

Screenshots of the issue: -

Recommendation:
Developers could add a confirmation pop up message when users attempt to delete records from the table.

Heuristic Problem Category:

1☐ 2☐ 3☐ 4☐ 5☒ 6☐ 7☐ 8☐ 9☐ 10☐

Problem Severity:
(3) – Major usability problem; important to f ix, so should be given high priority

Designer's Solution:
We acknowledge that this can prove a major future flaw in the system. We have added conformational popups upon pressing the 'Delete' buttons in the management panel.

➢ **Issue 4:**

Description:
There is no kind of input validation anywhere in the forms. This can prove very problematic in the future, as the only feedback we are getting is whether the form was submitted successfully or not.

Screenshots of the issue:
See *Figure 57*.

Recommendation:
Developers could add relevant input validation in their forms.

Heuristic Problem Category:
1☐ 2☐ 3☐ 4☐ 5☒ 6☐ 7☐ 8☐ 9☐ 10☐

Problem Severity:
(3) – Major usability problem; important to f ix, so should be given high priority

Designer's Solution:
We acknowledge that this can prove a major future flaw in the system. We have added validation in all relevant forms in the website to ensure that the correct information is entered at all times. The validation occurs as the user enters the information and not when pressing the submit button, in order to reduce the time needed to submit an error-free form.

➢ **Issue 5:**

Description:
I expected that there would be somewhere a My Profile section where I would be able to edit my personal information or check the site's settings. I did not manage to find this anywhere.

Screenshots of the issue: -

Recommendation:

Developers could add this feature in the final product, where registered users could see their information, edit them themselves and also include settings on the website.

Heuristic Problem Category:
1☐ 2☐ 3☐ 4☐ 5☐ 6☒ 7☐ 8☐ 9☐ 10☐

Problem Severity:
(3) – Major usability problem; important to f ix, so should be given high priority

Designer's Solution:
In the refined version of the platform, we added a '*My Settings*' section on the top-right corner of the website. The header of the website now includes a welcoming message to the user (did not exist in the draft version) and by clicking on it, users can select the '*My Settings*' section or they can simply log out of the platform. In order to help user reduce the amount of time needed for settings tasks, the 'My Settings' and '*My Profile*' sections appear inside the same webpage, to reduce time needed to change between them.

- **Issue 6:**

Description:
Some forms in the website could be re-arranged in order to save up space. For example, the insert criminal page goes all the way to the bottom, where some entries could be placed in the same row.

Screenshots of the issue:
See *Figure 55*.

Recommendation:
Developers could re-design such pages and include more input fields in the same row. This would greatly increase the input time, due to the lack of scrolling.

Heuristic Problem Category:
1☐ 2☐ 3☐ 4☐ 5☐ 6☐ 7☐ 8☒ 9☐ 10☐

Problem Severity
(1) - Cosmetic problem only; need not to be fixed unless extra time is available on project

Designer's Solution:
We re-designed the relevant insert pages. Since the are now not webpages as they where before but they have been moved into the management panels, they are now pop-up windows allowing faster input completion and more concrete design so as to disable the amount of scrolling needed in the previous forms for insertion.

➢ **Issue 7:**

Description:
Some of the forms in the website could be colored respectively in order to signify the type of action that takes place. For example, since the buttons to insert users in the management panel are green (see Figure 54), then the form to insert should include green elements, the update modal to include orange elements and the deletion procedures to include red elements. Keeping conformity with these colors will help users understand the type of action that takes place, as they can mnemonically link insertions with green, updates with orange and deletions with red.

Screenshots of the issue:
See *Figure 54, Figure 56* and *Figure 58.*

Recommendation:
Developers could keep color conformity and re-color the forms accordingly.

Heuristic Problem Category:
1☐ 2☐ 3☐ 4☒ 5☐ 6☐ 7☐ 8☐ 9☐ 10☐

Problem Severity:
(1) - Cosmetic problem only; need not to be fixed unless extra time is available on project

Designer's Solution:
We took this under consideration and agree that this would help users. As a result, we re-colored this form to help users link the color to specific action that take place in the platform (green – insert, orange – update, red – delete).

➢ **Issue 8:**

Description:
I pressed on the signup page in order to register, but then I remembered that a test account was already created for the testing purposes. There was no option to go back to the login page, so I had to manually enter the original website URL in order to go to the start page and from there to select to login. This wasted enough of my time to be considered an issue.

Screenshots of the issue:
See *Figure 52*.

Recommendation:
Developers could add a relevant button in the signup page that directs to the login page, in case users already have an account. This would save them time from going first to the welcome page and then to the login page.

Heuristic Problem Category:
1☐ 2☐ 3☒ 4☐ 5☐ 6☐ 7☐ 8☐ 9☐ 10☐

Problem Severity:
(2) – Minor usability problem; fixing this should be given low priority

Designer's Solution:
We added a button in the signup page that directs to the login page, in case users already have an account. thus saving time from having to go first to the welcome page and then to the login page.

- **Issue 9:**

Description:
It was quite odd to find that the management tables were not sortable. An application as this should offer the choice to users to sort the tables on any columns required. The fact that for the testing purposes, only 3-4 records existed in the table does not mean that this won't be a problem in the future.

Screenshots of the issue:
See *Figure 57*.

Recommendation:
Developers could make all tables sortable.

Heuristic Problem Category:
1☐ 2☐ 3☐ 4☒ 5☐ 6☐ 7☐ 8☐ 9☐ 10☐

Problem Severity:
(1) - Cosmetic problem only; need not to be fixed unless extra time is available on project

Designer's Solution:
We added a specific JavaScript plugin making the tables sortable where required. The plugin can be found in [84].

Chapter 4. The finalized platform

4.1 Fine-tuning the platform

After taking into consideration all the information gathered from the HE feedback, we incorporated all aspects and issues mentioned into our final product. The process of fine-tuning the website also included perfecting several aspects of the website that even through were not mentioned from the HE, we considered that it was of high importance to be optimized. In this sub-section, the finalized version of the platform is presented in detail.

4.1.1 Fine-tuning the database

Not many changes occurred regarding the database created fro the draft version of the platform. The database still consists of three different tables, the '*users*' table, the *'criminals'* table and the *'contact'* table. The *'contact'* table has remained the same, as no changes were deemed necessary. The '*criminals*' table was fine-tuned by adding the criminal's gender as a required information in the database. The '*users*' table was the one that required the most changes, as many new fields of information were added, like the gender, a short biography, the work phone and mobile phone and the personal website settings (theme, font size, language). For the personal settings of all users, the default choices (unless of course changed by the user) are the selection of the Light theme as a starting theme, the font size set to medium and language set to English. The SQL structure and schema of these tables is presented in the following figures below.

	#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
☐	1	contact_id	int(11)			No	*None*		AUTO_INCREMENT	Change Drop More
☐	2	first_name	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
☐	3	last_name	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
☐	4	email	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
☐	5	subject	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More

Figure 67. The finalized 'contact' table

	#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
☐	1	criminal_id	int(11)			No	*None*		AUTO_INCREMENT	Change Drop More
☐	2	full_name	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
☐	3	age	int(11)			Yes	*NULL*			Change Drop More
☐	4	height	float			Yes	*NULL*			Change Drop More
☐	5	weight	int(11)			Yes	*NULL*			Change Drop More
☐	6	eye_color	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
☐	7	biography	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
☐	8	portrait	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
☐	9	last_location	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
☐	10	gender	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More

Figure 68. The finalized 'criminals' table

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
1	**user_id**	int(11)			No	*None*		AUTO_INCREMENT	Change Drop More
2	**username**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
3	**password**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
4	**email**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
5	**full_name**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
6	**gender**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
7	**biography**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
8	**work_phone**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
9	**mobile_phone**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
10	**role**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
11	**avatar**	varchar(255)	utf8_general_ci		Yes	*NULL*			Change Drop More
12	**theme**	varchar(255)	utf8_general_ci		Yes	Light Theme			Change Drop More
13	**language**	varchar(255)	utf8_general_ci		Yes	English			Change Drop More
14	**fontsize**	varchar(255)	utf8_general_ci		Yes	14			Change Drop More

Figure 69. The finalized 'users' table

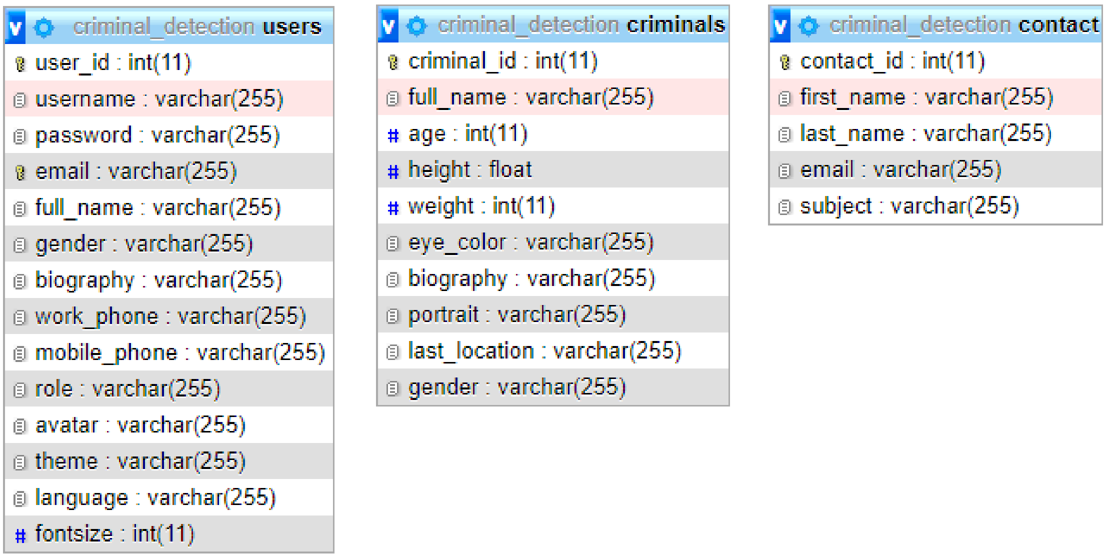


Figure 70. The finalized database schema

4.1.2 Why use Github?

During the development of this platform, the GitHub version control system was used. The reason why we used GitHub is because for such large scale development, we need a version control system. More specifically, version control is the management system that manages all changes made in our project from start to end. Changes can include adding new files, editing or deleting files. Every time we upload ('*push*') a newer version into the repository, the version control system captures a snapshot of all changes made, which allows us to be able to access different versions of the platform at any time if we need to. Such an approach that keeps track a large collection of information allows tracking of the entire history of changes made in the code and developers can thoroughly examine the differences between each version.

To this end, GitHub is such a version control system where developers can store their code in the GitHub servers in files and folders that are called '*Repositories*' and track them continuously. It is an open-source version control and collaboration platform for program developers. GitHub helps programmers collaborate with each other (if needed) that work on similar projects and allows code sharing easily. GitHub is free and open-source, allowing us to download the source code of the repository everywhere we are. Chances of losing data is extremely low, and most of the operations included offer huge speed benefits. Furthermore, the '*Issues*' functionality can be taken advantage of by the developers. Issues are a great way to keep track of tasks, enhancements, and bugs for the projects themselves.

For anyone interested of downloading the source code from GitHub, the project can be downloaded ('*forked*') from the following URL in GitHub: https://github.com/vagg777/Msc-Thesis-Website.

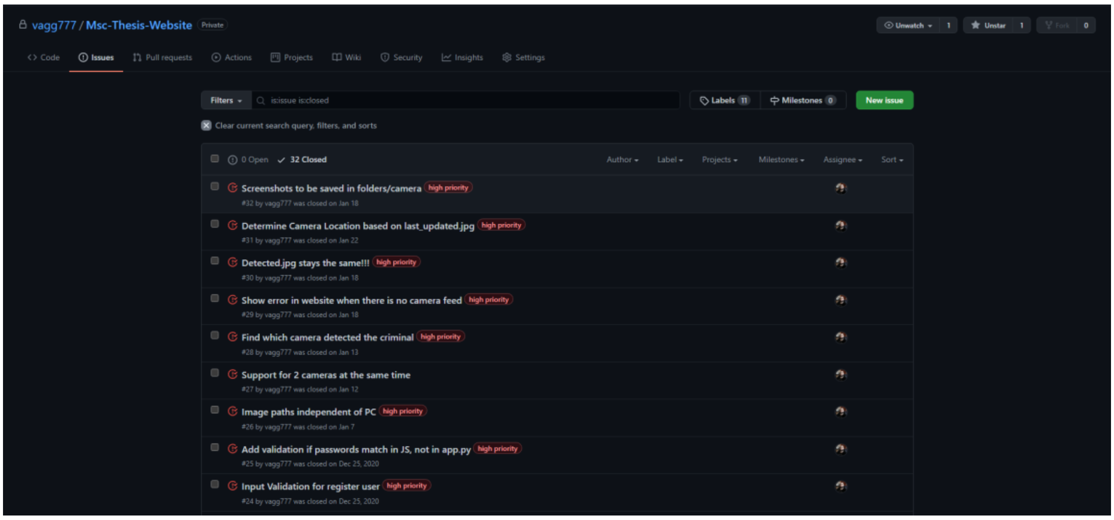


Figure 71. Closed Issues during the development stage

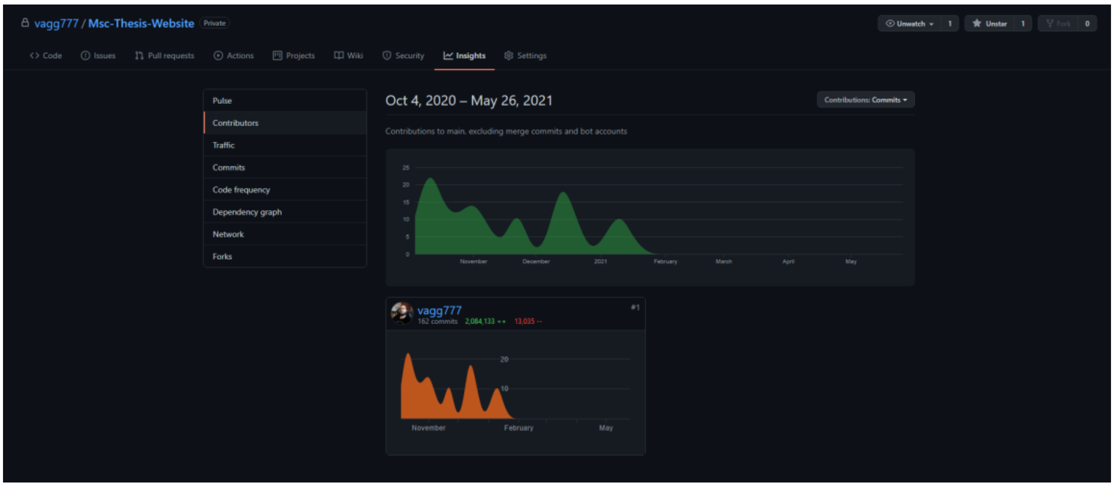


Figure 72. Contribution Analytics during the development stage

4.1.3 Fine-tuning the backend code

Of course, changes occurred in the backend software code (Python) towards better readiness and comprehensiveness from those who might want to see the underlying code of the platform. Any inline CSS appearing inside the HTML files has been removed as a good practice and all relevant CSS is loaded from the equivalent external files. Any hardcoding techniques have been discarded. The finalized project structure is as follows below in *Figure 73*. One can easily notice that any changes occurred are only for the best. For example, there is a dedicated external JavaScript folder containing all external scripts that are loaded from the equivalent pages (e.g. for input validation in forms), instead of them appearing in each HTML page respectively. The CSS folder now contains separate styling depending on the theme selection (light/dark) and the font size selected (small/medium/large). There is now a dedicated SQL folder where the export of the SQL base is stored for backup purposes.

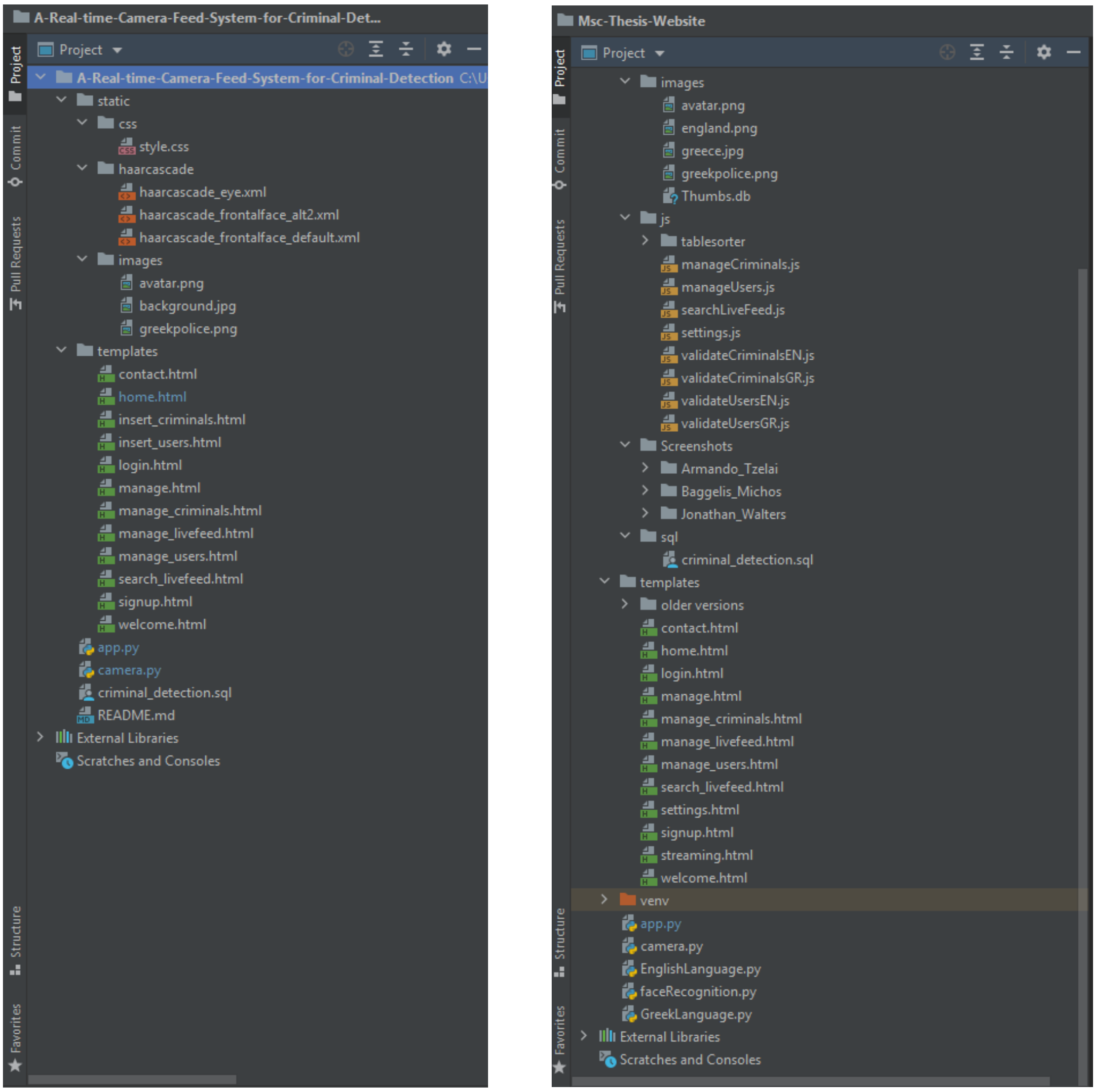


Figure 73. The finalized project structure (left: Draft version, right: Final version)

Due to re-designing the website, for the different languages in the website and in order for the website internationalization to be efficient in terms of coding, we created two different Python files, called *EnglishLanguage.py* and *GreekLanguage.py*. These files serve as 'translation libraries', storing the translated variables under the respective Python classes. These greatly helps designing the website, as now there is now need to create each webpage two times and hardcode all information in the HTML file, but we can simply include the equivalent Python file (English or Greek) where needed and access the variables after created the respective class initializations through Python objects. This also helps our website in the distant future, as in order to insert a new language in the webpage, it is not necessary to re-design anything and the only thing needed is a translation on the respective Python file! The decision on which language is to be imported in the website happens in the *checkUserSettings()* function, inside the *app.py* file (see *Figure 76*). The respective translations (only the ones needed) are passed as object into the HTML files (see *Figure 77*).

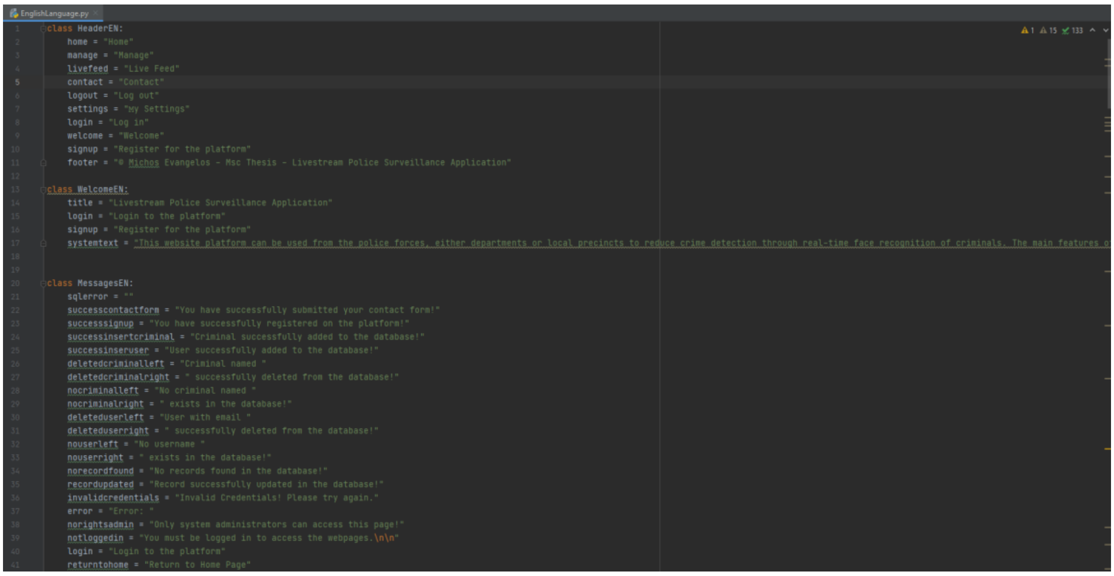


Figure 74. English Translations

GreekLanguage.py

```python
class HeaderGR:
    home = "Κεντρική Σελίδα"
    manage = "Διαχείριση"
    livefeed = "Ζωντανή Μετάδοση"
    contact = "Επικοινωνήστε μαζί μας"
    logout = "Έξοδος"
    settings = "Οι Ρυθμίσεις μου"
    login = "Συνδεθείτε"
    welcome = "Καλωσήρθες"
    signup = "Εγγραφείτε στη πλατφόρμα"
    footer = "© Μίχος Ευάγγελος - Μεταπτυχιακή Εργασία - Αστυνομική Πλατφόρμα Παρακολούθησης σε Πραγματικό Χρόνο"

class WelcomeGR:
    title = "Αστυνομική Πλατφόρμα Παρακολούθησης σε Πραγματικό Χρόνο"
    login = "Συνδεθείτε στη πλατφόρμα"
    signup = "Εγγραφείτε στη πλατφόρμα"
    systemtext = "Αυτή η πλατφόρμα ιστότοπου μπορεί να χρησιμοποιηθεί από τις αστυνομικές δυνάμεις, είτε από τμήματα είτε από τοπικούς χώρους για τη μείωση της ανίχνευσης εγκλημάτων μέσω της αναγνώρισ

class MessagesGR:
    sqlerror = ""
    successcontactform = "Υποβάλλατε τη φόρμα με επιτυχία"
    successsignup = "Ολοκληρώσατε την εγγραφή σας με επιτυχία!"
    successinsertcriminal = "Έγινε προσθήκη του κακοποιού με επιτυχία στη βάση!"
    successinseruser = "Έγινε προσθήκη του χρήστη με επιτυχία στη βάση!"
    deletedcriminalleft = "Ο κακοποιός με όνομα "
    deletedcriminalright = " διεγράφη με επιτυχία από τη βάση!"
    nocriminalleft = "Δεν υπάρχει κακοποιός με όνομα "
    nocriminalright = " στη βάση δεδομένων!"
    deleteduserleft = "Ο χρήστης με email "
    deleteduserright = " διεγράφη με επιτυχία από τη βάση!"
    nouserleft = "Δεν υπάρχει χρήστης με όνομα "
    nouserright = " στη βάση δεδομένων!"
    norecordfound = "Δεν βρέθηκαν εγγραφές στη βάση!"
    recordupdated = "Οι αλλαγές αποθηκεύτηκαν επιτυχώς στη βάση!"
    invalidcredentials = "Λάθος στοιχεία εισόδου. Προσπαθήστε ξανά!"
    error = "Πρόβλημα: "
    norightsadmin = "Μόνο οι διαχειριστές συστήματος μπορούν να έχουν πρόσβαση σε αυτή τη σελίδα!"
    notloggedin = "Πρέπει να συνδεθείτε για να έχετε πρόσβαση στις ιστοσελίδες.\n\n"
    login = "Συνδεθείτε στη πλατφόρμα"
    returntohome = "Επιστροφή στη Κεντρική Σελίδα"
    yourchanges = "Οι αλλαγές σας αποθηκεύτηκαν επιτυχώς!"
```

Figure 75. Greek Translations

app.py

```python
def checkUserSettings(theme, language, fontsize):
    global welcome
    global login
    global messages
    global header
    global home
    global insertUser
    global manageUser
    global insertCriminal
    global manageCriminal
    global manageLivefeed
    global manage
    global settings
    global contact
    global signup
    global site_fontsize
    global site_language
    global site_theme
    if site_language == "Greek":
        welcome = WelcomeGR()
        signup = SignupGR()
        login = LoginGR()
        messages = MessagesGR()
        header = HeaderGR()
        home = HomeGR()
        insertUser = InsertUserGR()
        manageUser = ManageUserGR()
        insertCriminal = InsertCriminalsGR()
        manageCriminal = ManageCriminalGR()
        manageLivefeed = ManageLivefeedGR()
        manage = ManageGR()
        settings = SettingsGR()
        contact = ContactGR()
    else:
        welcome = WelcomeEN()
        signup = SignupEN()
        login = LoginEN()
        messages = MessagesEN()
        header = HeaderEN()
        home = HomeEN()
        insertUser = InsertUserEN()
        manageUser = ManageUserEN()
```

checkUserSettings()

Figure 76. Deciding on the user settings and accessing the respective translation classes

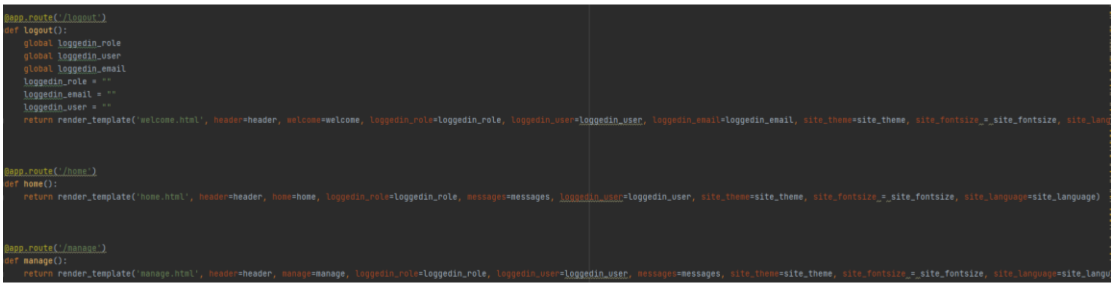


Figure 77. Passing the translation object into the HTML webpages

4.1.4 How Live Streaming works

In general, streaming can be defined as a mechanism in which the server provides the response to a request using chunks. Streaming is useful for the following two (2) very important reasons: very large responses and real time data. Regarding the very large responses, it is considered inefficient to construct a memory response that is only returned to a client and an alternative approach would be to construct the response into the disk and then send the file with the *flask.send_file()* function, yet this adds I/O complications. As a result, it is highly advisable for the responses to be send in small chunks, of course providing that the data can be generated in small portions. Regarding the real-time data, a request could need to return data from a real time source (e.g. real-time video). Such a technique is largely used in security cameras in order to stream live video feed to web browsers.

In order to implement streaming in Python using the Flask framework, we must use Flask’s native support for streaming responses through the generator functions, which are basically special functions that can be interrupted and resumed at any time. A generator function is able to return multiple results in a sequence, thus Flask takes this into advantage through generator functions in order to implement the streaming features. An extension of this is the use of multipart Responses. Streaming can be uses in order to have each chunk replace the previous one in the page, enabling streams to play or animate in a browser. This technique enables each chunk in the stream to be simply an image, and then we get the whole video live-stream in the browser through repetitive chunks of images and in order to do that, Flask uses the multipart responses. Multipart responses consist of headers that include one of the multipart content types and are followed by the parts, separated by a boundary marker and each having its own part specific content type. For the purposes of our platform, we need each new chunk to replace the previous one as single-instance images, thus we will use the ‘*multipart/x-mixed-replace*’ content.

In the following example below, one can understand how the aforementioned can be implemented in Python and viewed in a simple *index.html* page. This is a straightforward HTML page with only a title and an image tag. The *src* attribute of the image tag points to the application's second route. The streaming answer is returned by the */my_video_feed* route. The URL to this route appears in the *src* attribute of the image tag since this stream returns the images that will be shown on the web page. Since most/all browsers support multipart responses, the browser will automatically keep the image element updated by displaying the stream of JPEG images in it. The */my_video_feed* route's generator function is called *gen()* and takes a Camera class instance as an input. The mimetype argument is specified as indicated above, with the content type *multipart/x-mixed-replace* and a boundary of "frame." The *gen()* function enters a loop, returning frames from the camera as response chunks indefinitely. By calling the camera, the function requests that it supply a frame. The *get_frame()* method returns with this frame prepared as a response chunk with an image/jpeg content type. All that remains is for us to combine the IP camera URL and parse them as input, resulting in successive frames/chunks that can then be displayed in our platform's livestream page part.

Python file

```
from flask import Flask, render_template, Response
from camera import Camera

app = Flask(__name__)

@app.route('/')
def home():
    return render_template('home.html')

def gen(camera):
    while True:
        frame = camera.get_frame()
        yield (b'--frame\r\n' b'Content-Type: image/jpeg\r\n\r\n' +
frame + b'\r\n')

@app.route('/my_video_feed')
def my_video_feed():
    return  Response(gen(Camera()),  mimetype='multipart/x-mixed-
replace; boundary=frame')

if __name__ == '__main__':
    app.run(host='127.0.0.1')
```

HTML file

```
<html>

<head>
       <title>Flask Livestream Example</title>
</head>

<body>
    <h1>Video Streaming Demonstration</h1>
    <img src="{{ url_for('my_video_feed') }}">
</body>

</html>
```

4.1.5 The finalized website

To deploy the website online, the procedure is exactly the same as described in **3.6.4 Deploying the draft website**. When launching the website for the first time, we are greeted with the welcoming page of the platform. As stated in previous sections, login credentials are required to enter the platform.

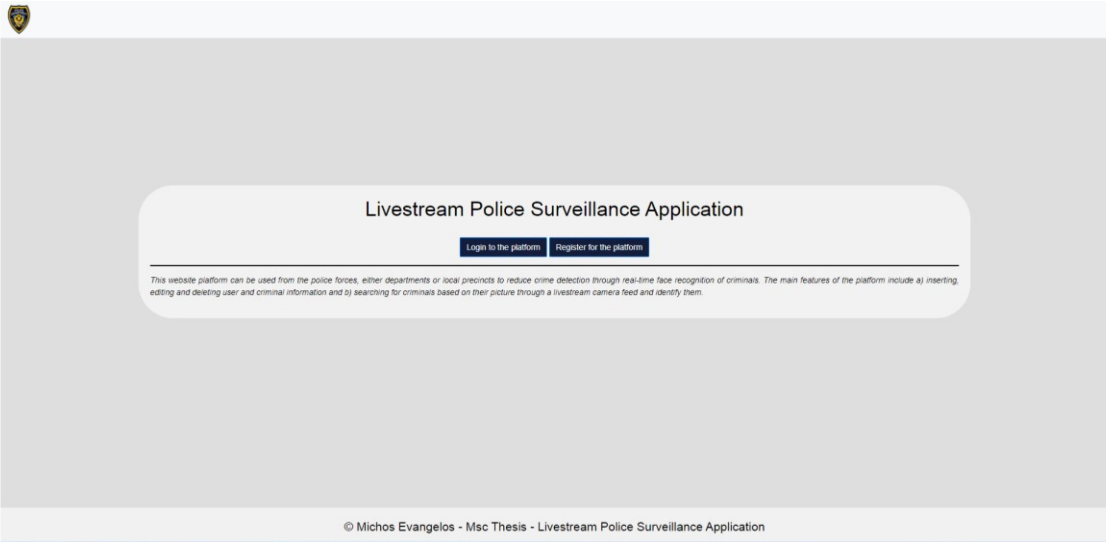


Figure 78. The finalized welcome page

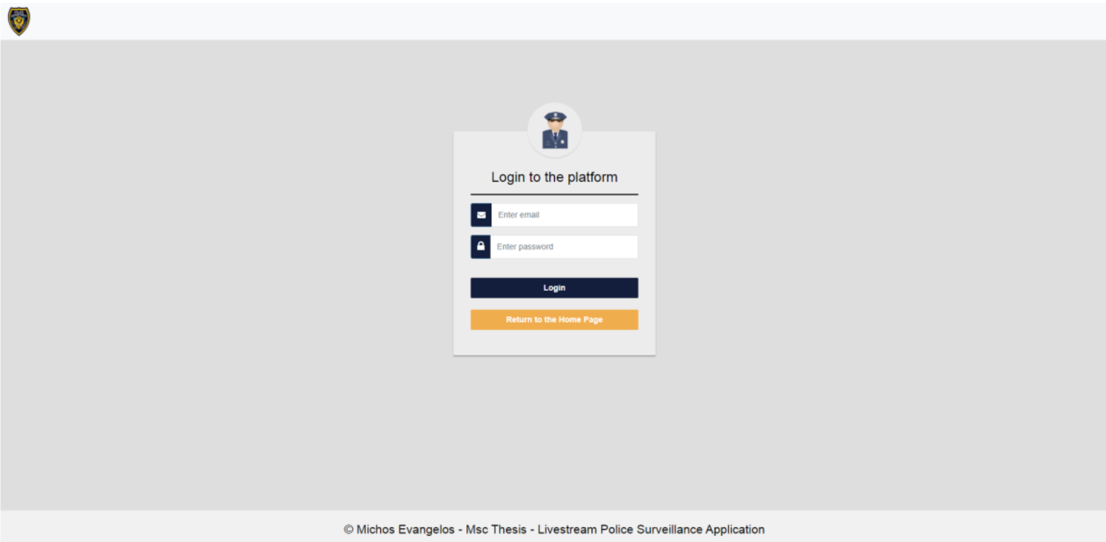


Figure 79. The finalized login page

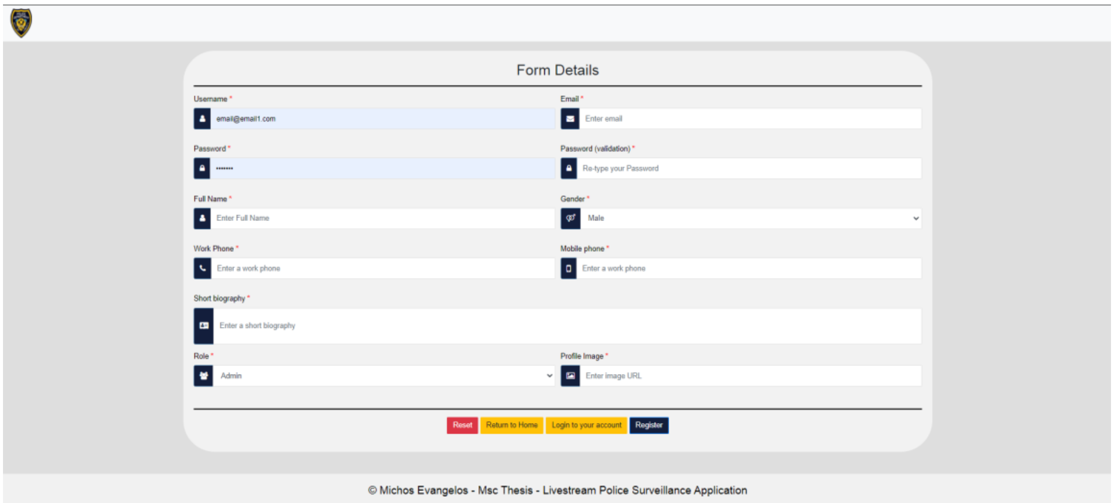


Figure 80. The finalized signup page

As one may see, the finalized version contains the relevant fixes that were reported during the HE from the evaluators. For example, we have added the 'Return to Home' button requested in '*Issue 8*' that would re-direct users directly to the Login page to save time and also we have changed the UI of the signup form and included more than one field per row in order to reduce form scrolling, as requested in HE '*Issue 7*'. To compare these designs with the draft version of the platform, see the old versions in *Figure 50, Figure 51* and *Figure 52*.

Upon correctly entering our user credentials, we can now view the main page (Home page) of the website. As we can see below, we are introduced with relevant information on what we can do in the platform and we can access the platform's features either from the hyperlinks in the 'Features' section or from the website header. As one may see, the option of changing language in each container was deemed repetitive and irritating during HE '*Issue 1*' and was moved into the Settings panel, which is on the top-right corner of the website. The management panel has experienced relevant refinement in order to be better understood from the users. This means that relevant information now appears under the relevant choices, informing users on what they can do by pressing the relevant buttons in that section. To compare these designs with the draft version of the platform, see the old versions in *Figure 53* and *Figure 54.*

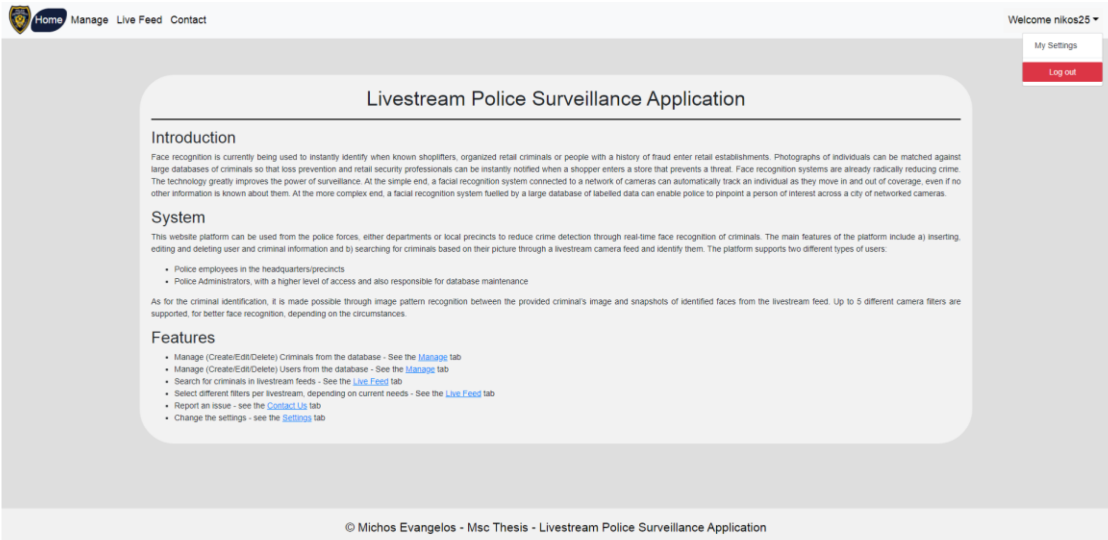


Figure 81. The finalized Home page

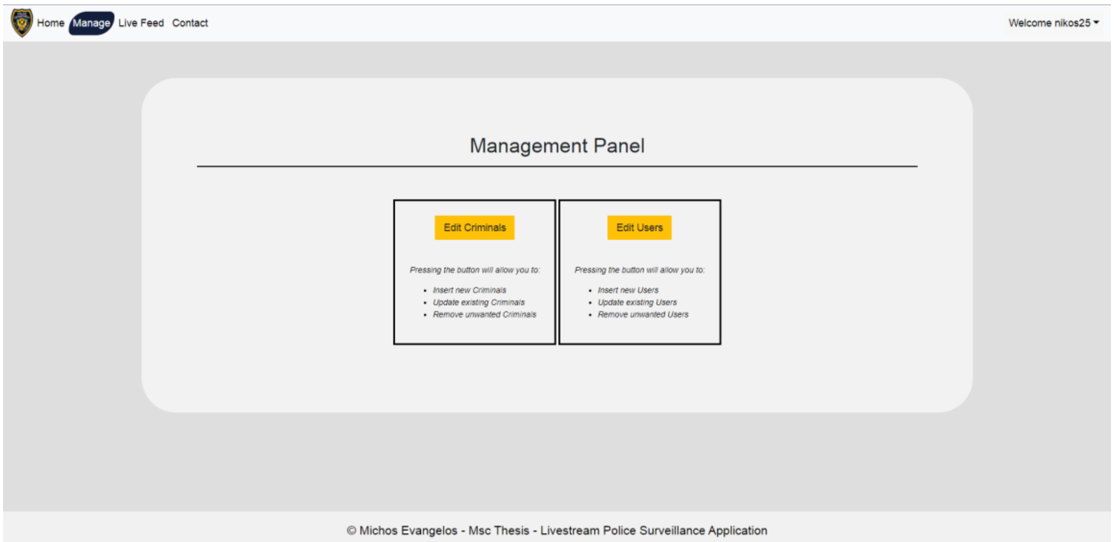


Figure 82. The finalized Management page

For example, we decide to access the management panel for the criminals. As requested in HE ‘*Issue 9*’, now such tables are all sortable. Additionally, the ‘Insert’ button has been added in the management panel, as requested in HE ‘*Issue 2*’, which resulted in totally discarding the approach of having separate pages for inserting user/criminals and separate pages for updating/deleting them. Additionally, as requested in HE ‘*Issue 7*’, we have updated the design for the pop-up windows regarding insertion/updating/deletion of users/criminals and linked green colors with insertions, orange colors with updating and red colors with deletions. By keeping conformity with these colors, we will help users understand the type of action that takes place, as they can mnemonically link colors with actions. As one may observe in the figures below, the header color of the pop-up and the relevant submit action button are colored the same, exactly following the aforementioned approach. The conformational message that appears in *Figure 86* comes from HE ‘*Issue 3*’ and we have added this feature into the final design of the website. To compare these designs with the draft version of the platform, see the old versions in *Figure 56, Figure 57* and *Figure 58.*

Figure 83. The finalized management page for Criminals (same approach for users)

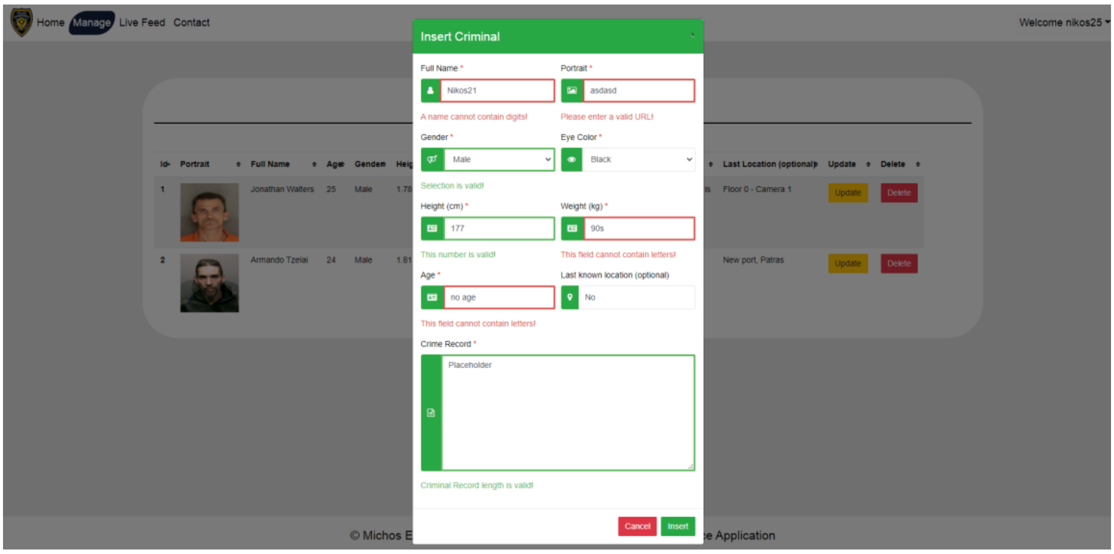


Figure 84. Color conformity for inserting a criminal + JavaScript validation

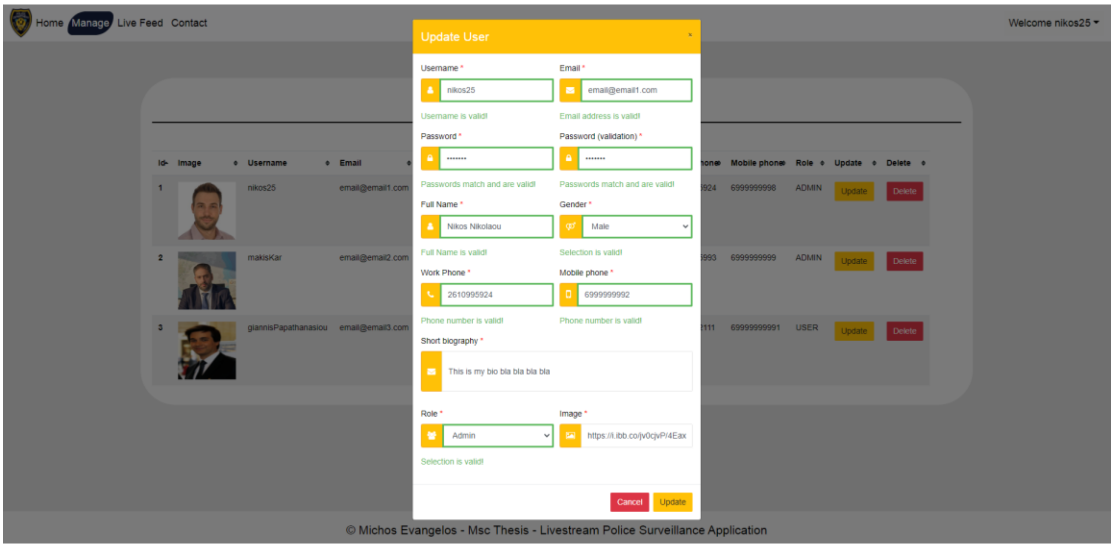


Figure 85. Color conformity for updating a user + JavaScript validation

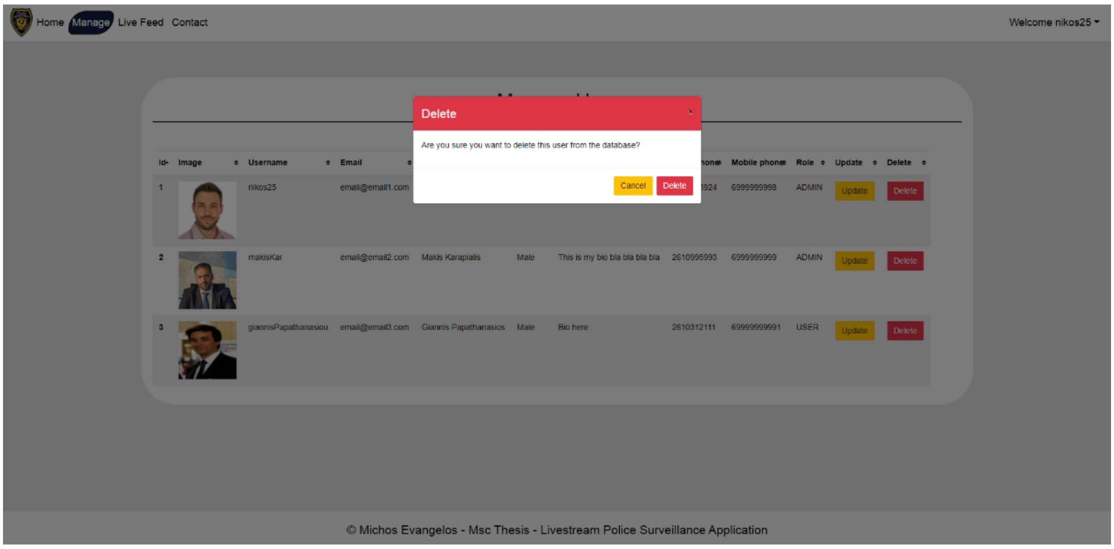


Figure 86. Color conformity and confirmational message upon requesting to delete a user

The 'Contact' page still follows the same simplistic approach as its draft version, requiring only the name, surname email and subject to be sent to the system administrators and recorded in the database for any issues/concerns encountered in the platform. To compare these designs with the draft version of the platform, see the old version in *Figure 59*.

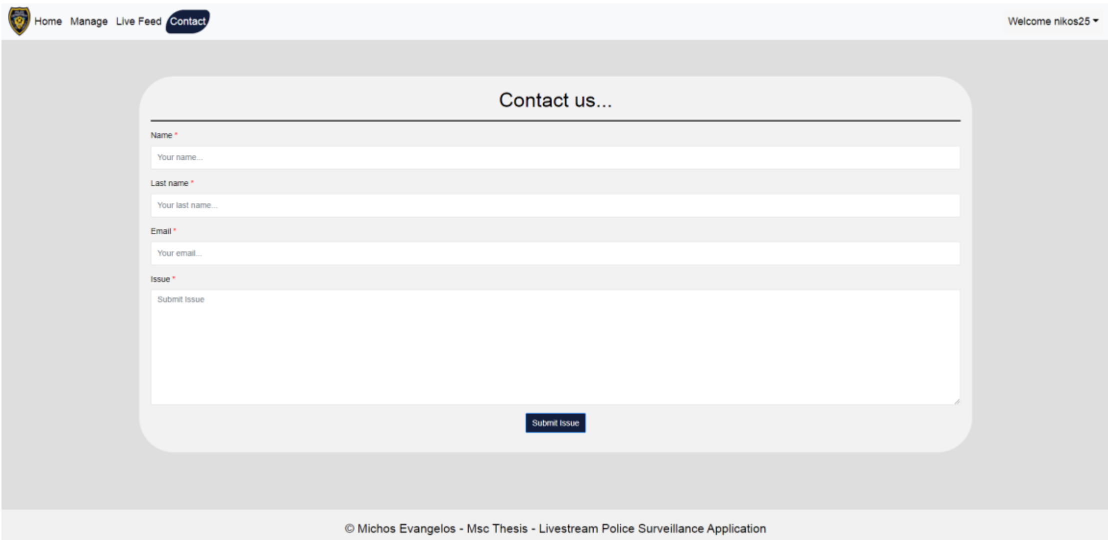


Figure 87. The finalized Contact page

Much like the draft version, the criminal identification is made possible through image pattern recognition between the provided criminal's image and snapshots of identified faces from the livestream feed. The Live Feed section offers six different options for video filters, enabling the user to select the best filter, depending on the relevant situation of the physical surroundings (e.g. broad light, dark environment). Again, each time a criminal is recognized, a snapshot is taken directly from that camera feed. The six (6) different filters that can be used for face recognition and all snapshot are categorized in the specific folder created for that criminal, where snapshots are saved inside the folder and re-categorized depending on the camera that took the snapshot. This helps us, as we can retrieve his last known location by knowing the location of the camera that took the screenshot. Lastly, the image name contains the date and time and filter used. A feature added in the finalized version of the platform is that now, the screenshots include a running timestamp on the image itself, which was previously not the case with the draft version (see *Figure 90*).

For the IP cameras' software, we went with the solution of installing an relevant application on tablets and mobile phones rather than buying relevant equipment. This is because the approach of the application is easily transferable to many devices and the only requirement is that the OS is the same. Yet, even if the OS are different, the final approach of the website with the IP cameras only requires the camera's URL in order to 'grab' the video feed and it is up to the camera owners to provide the URL feed. One of the best choices out there (for Android OS) is the IP Webcam application, which is free on the GooglePlay Store [85]. As the developer of the application explains:

"*IP Webcam turns your phone into a network camera with multiple viewing options. View your camera on any platform with VLC player or web browser. Stream video inside WiFi network without internet access. Optional Ivideon cloud broadcasting is supported for instant global access. Two-way audio supported in tinyCam Monitor on another android device. Use IP Webcam with third-party MJPG software, including video surveillance software, security monitors and most audio players.*

Features include:

• Video upload to Dropbox, SFTP, FTP and Email using Filoader plugin

• Several web renderers to choose from: Flash, Javascript or built-in

• Video recording in WebM, MOV, MKV or MPEG4 (on Android 4.1+)

• Audio streaming in wav, opus and AAC (AAC requires Android 4.1+)

• Motion detection with sound trigger, Tasker integration.

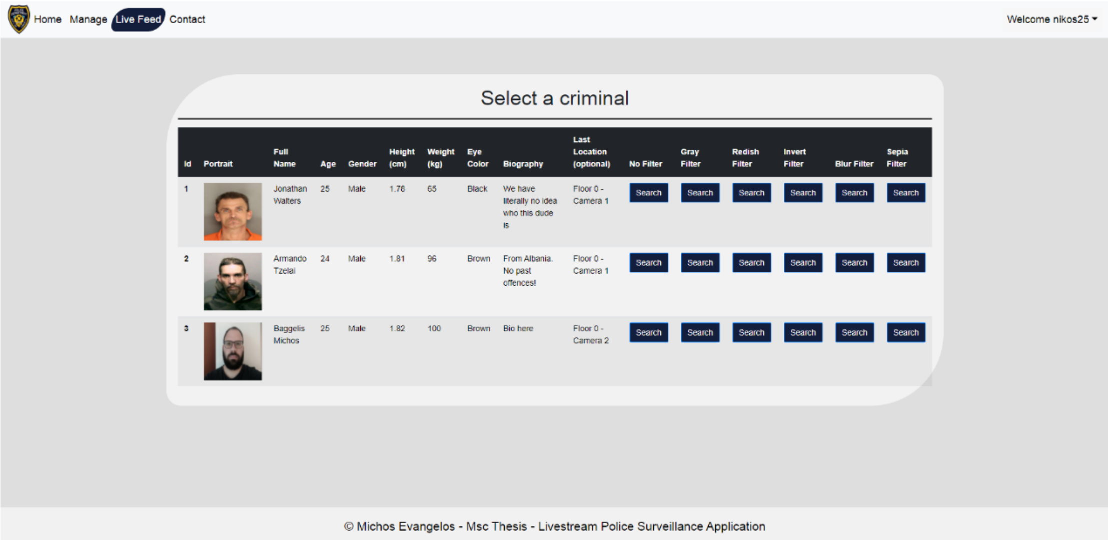
• Date, time and battery level video overlay.

• Sensor data acquisition with online web graphing.

• Videochat support (video stream only for Windows and Linux via an universal MJPEG video streaming driver)

• Cloud push notifications on motion and sound, cloud recording for motion-triggered records, online video broadcasting powered by Ivideon.

• Extensive baby and pet monitor features: night vision, motion detection, sound detection."



Id	Portrait	Full Name	Age	Gender	Height (cm)	Weight (kg)	Eye Color	Biography	Last Location (optional)	No Filter	Gray Filter	Redish Filter	Invert Filter	Blur Filter	Sepia Filter
1		Jonathan Walters	25	Male	1.76	65	Black	We have literally no idea who this dude is	Floor 0 - Camera 1	Search	Search	Search	Search	Search	Search
2		Armando Tzelal	24	Male	1.81	96	Brown	From Albania. No past offences!	Floor 0 - Camera 1	Search	Search	Search	Search	Search	Search
3		Baggelis Michos	25	Male	1.82	100	Brown	Bio here	Floor 0 - Camera 2	Search	Search	Search	Search	Search	Search

Figure 88. The finalized Live Feed page

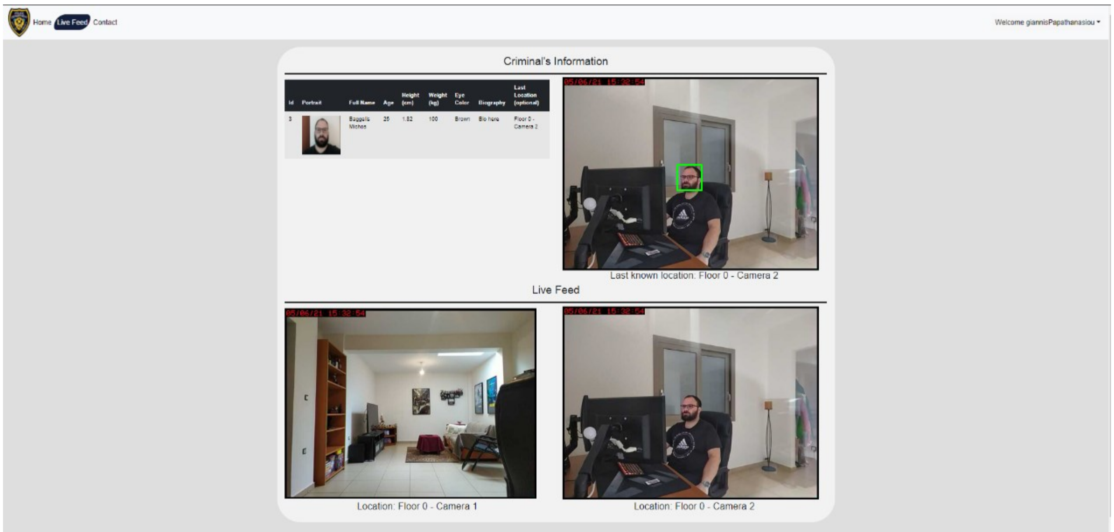


Figure 89. The finalized Live Feed page

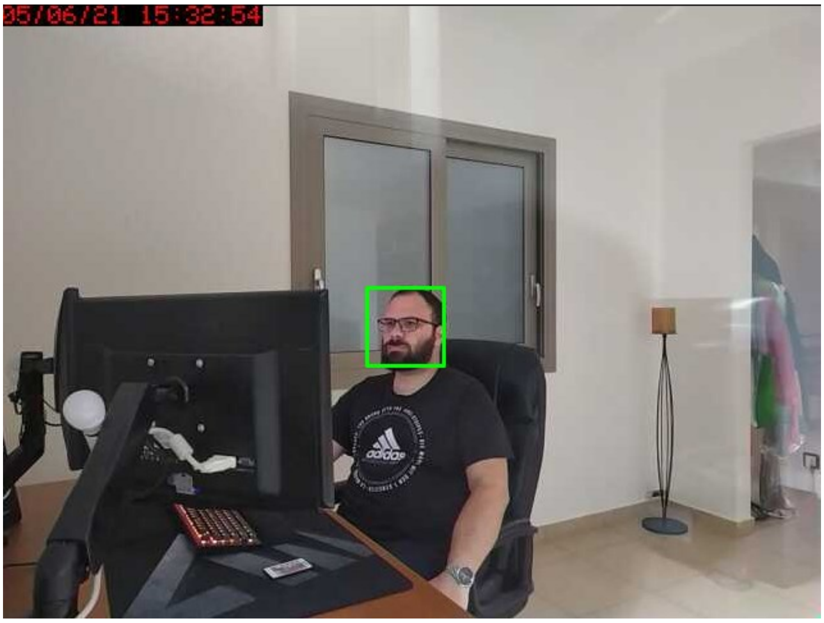


Figure 90. Timestamping all criminal screenshots (see top-left corner)

Of course, the platform could not be launched without including error handling. For example, as mentioned previously for our platform, the platform supports two different types of users, namely a) police employees in the headquarters/precincts and b) police administrators, with a higher level of access and also responsible for database maintenance. The main features of the platform include:

- inserting, editing and deleting user information
- inserting, editing and deleting criminal records
- applying 6 different filters in the livestream cameras (e.g. gray, sepia, redish, blur, inverted)
- searching for criminals based on their picture through a livestream camera feed and identifying them

This creates a necessity to include error handling in order to not allow users access to pages that they should not have access to (e.g. simple employees accessing information on other employees). Such error handling examples are depicted below.

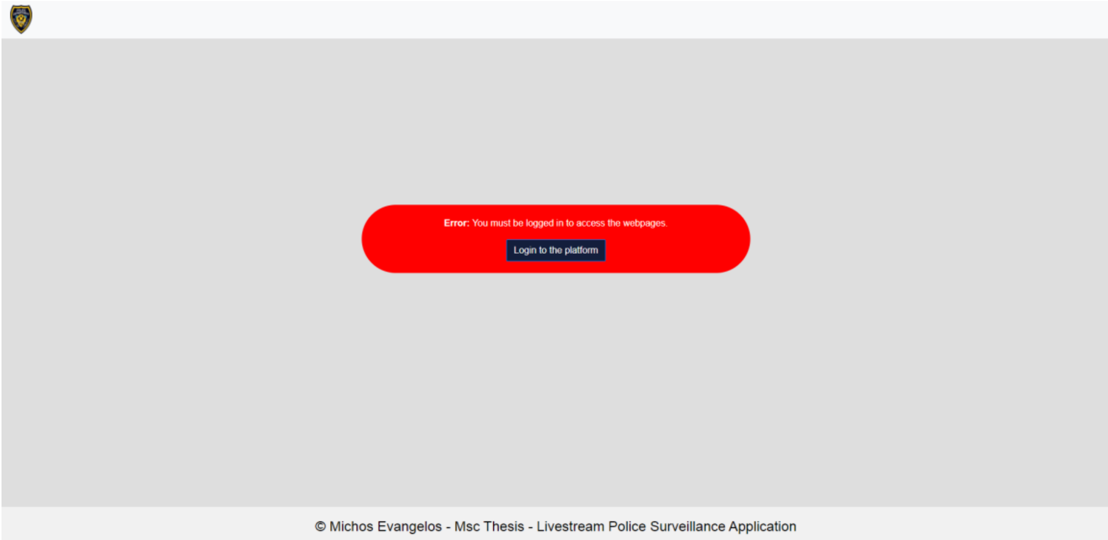


Figure 91. Error at attempting to enter the website without any credentials

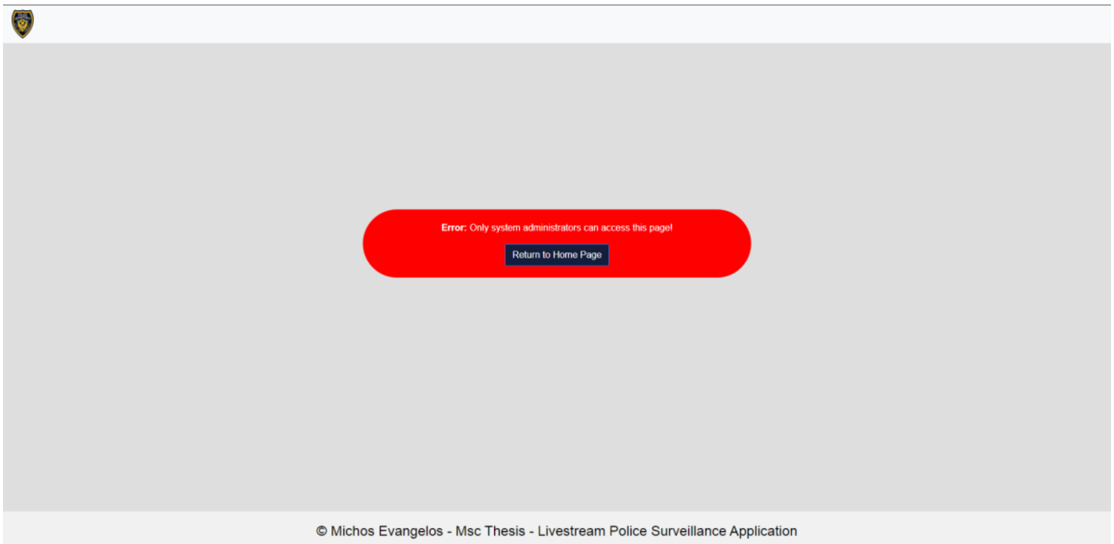


Figure 92. Error for non-admin users attempting to check employees' information

The newest section in our design, the Settings page, can be found on the top-right corner of the website, under the welcoming message to the user. In the refined version of the platform, we added a '*My Settings*' section on the top-right corner of the website. The header of the website now includes a welcoming message to the user (did not exist in the draft version) and by clicking on it, users can select the '*My Settings*' section or they can simply log out of the platform. In order to help user reduce the amount of time needed for settings tasks, the '*My Settings*' and '*My Profile*' sections appear inside the same webpage, to reduce time needed to change between them. The language issue reported in HE now appears only in the '*My Settings*' section and does not appear in all pages, as was the case with the draft version.

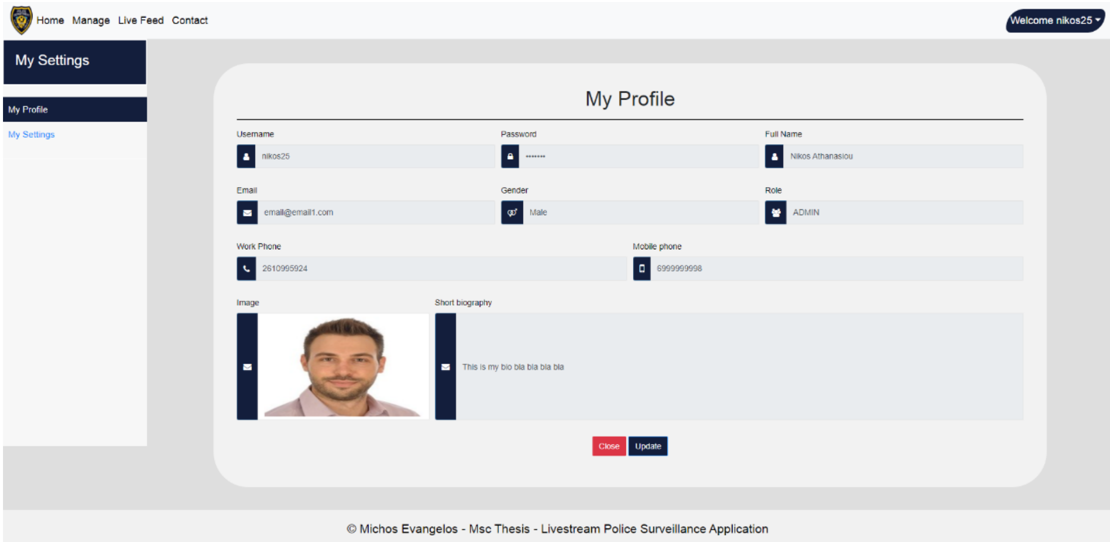


Figure 93. The newly added My Profile section

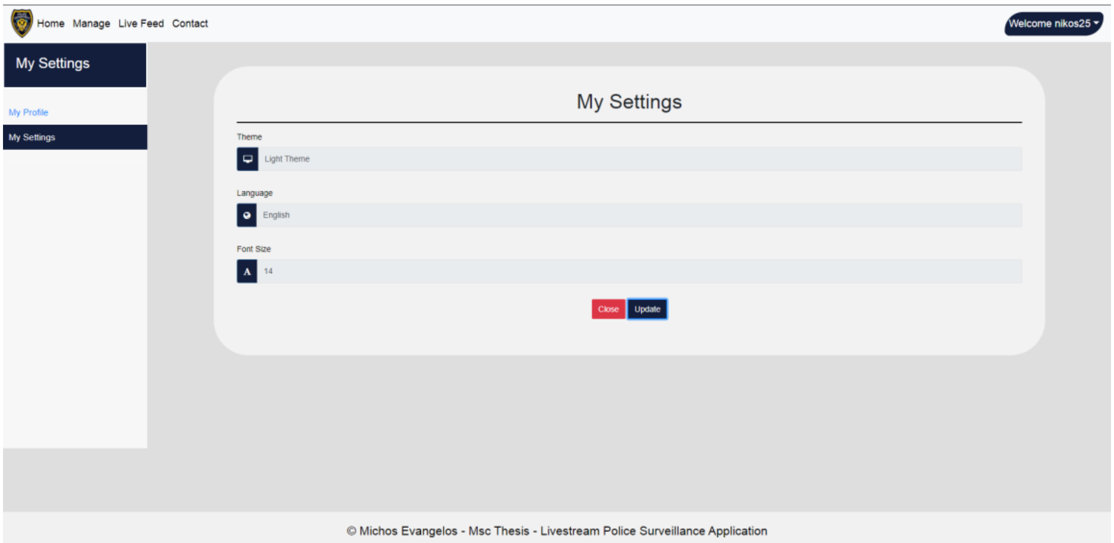


Figure 94. The newly added My Settings section

Supposing we want to edit the settings in the website and change the language to Greek, the theme to Dark and the font size to small. Some examples of this combination (Dark theme – font size 12 – Greek language) are depicted below.

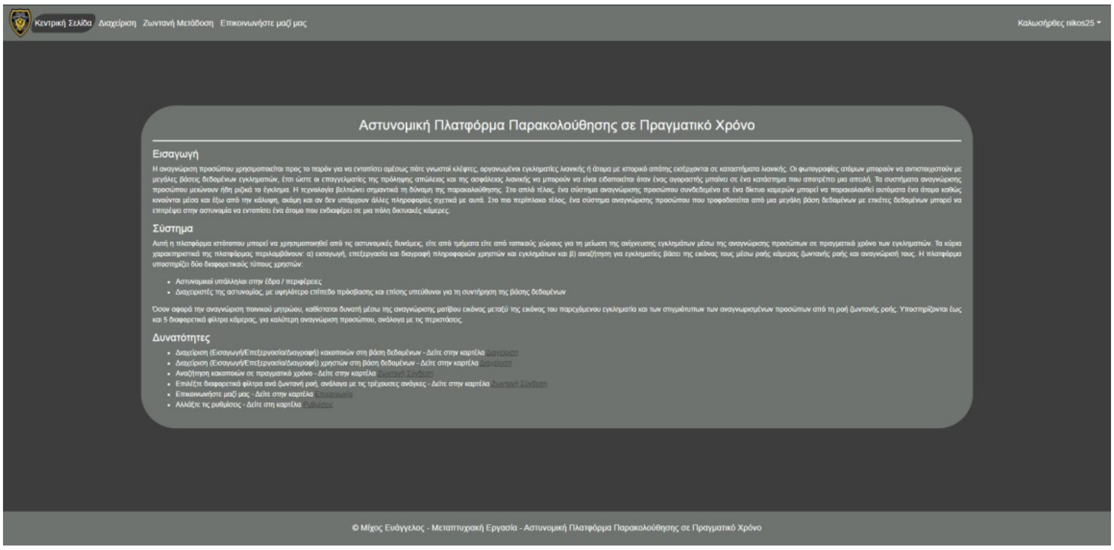


Figure 95. Changing the website settings

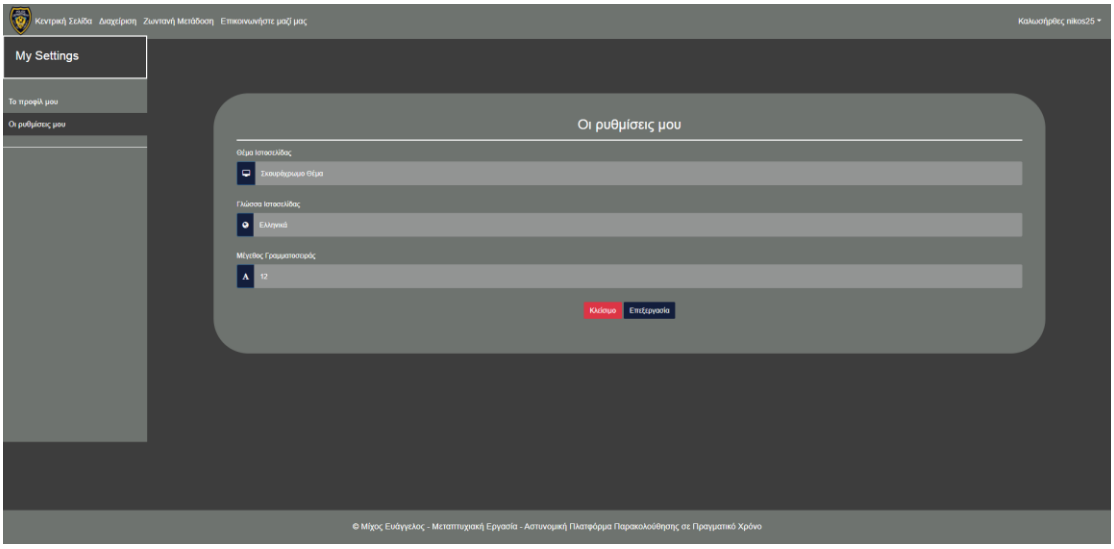


Figure 96. Changing the website settings

Lastly, as already mentioned in the Requirements of the platform during the early stages of design (see **3.3 Requirements**), the site's responsiveness into different device platforms was also envisioned. To this end, the use of Boostrap's CSS allowed easy styling of the webpage for the cases where the platform was accessed from either phones or tables. Some screenshots of the site's appearance on mobiles is depicted below.

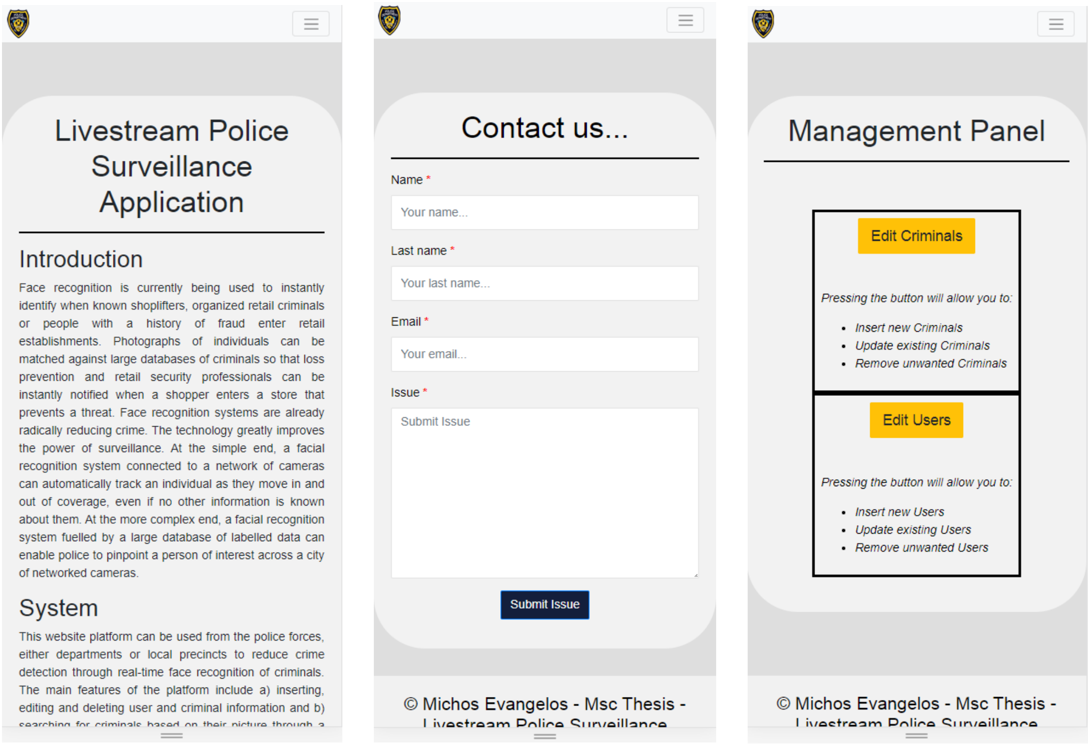


Figure 97. Platform mobile responsiveness

4.2 Evaluation of the platform

In order for the evaluation procedure to be valid and reliable, the platform had to be tested from police officers from a local police department. The reason for starting this process is that this profession is accustomed and aware of technologies that identify criminals, they know what characteristic a system like the proposed one of this thesis needs to have and can propose recommendations for future development. For this reason, this particular group was selected as the evaluation team.

In order to extract data from the evaluation process and mostly specific information for the characteristics of the system, three types of questionnaires were used.

1. The first one is based on close type **questions** in which the respondent has to choose between "YES" or "NO" as potential answers. This particular questionnaire consists of general questions about the system and aims to obtain some information about the user's interaction with the platform.

2. Then follows a questionnaire that is based on Likert scale that starts from 1 (Strongly Disagree) and ends with 5 (Strongly Agree). The purpose of this questionnaire is to record impressions from certain elements that can affect the user experience on the website.

3. The last section ends with some open-ended questions in which respondents can give their own answer and describe how they feel or comment a figure of the platform that they like the most or not.

The reason why we used Likert scale (which by the way is one of the most widely used question types in a survey) in the questionnaire for the specific questions regarding the platform design is as follows. According to psychology, a Likert scale question is a psychometric scale where questions based on this scale are used in a survey. In a Likert scale survey, respondents don't choose just between ‘YES’ or ‘NO’, but there are specific choices based on 'agreeing' or 'disagreeing' on a particular survey question. Likert scale survey questions are essential in measuring a respondent's opinion or attitude towards a given subject and is an integral part of market research. Likert scale is typically a five, seven, or nine-point agreement scale used to measure respondents' agreement with various statements. In general, a series of statements each designed to view a construct from a slightly different perspective is leveraged. This technique's power is that it works across disciplines—it is just as applicable to a social science construct as it is a marketing one. Likert scale questionnaires offer the following advantages:

- Easy to design and apply on the survey
- Can include items not related to expression
- Provide a range of opinions to the respondents
- Produces accurate and quality information, reducing measurement error
- Allows to perform the necessary analysis to achieve the research objective
- Can be compared with previous service evaluations

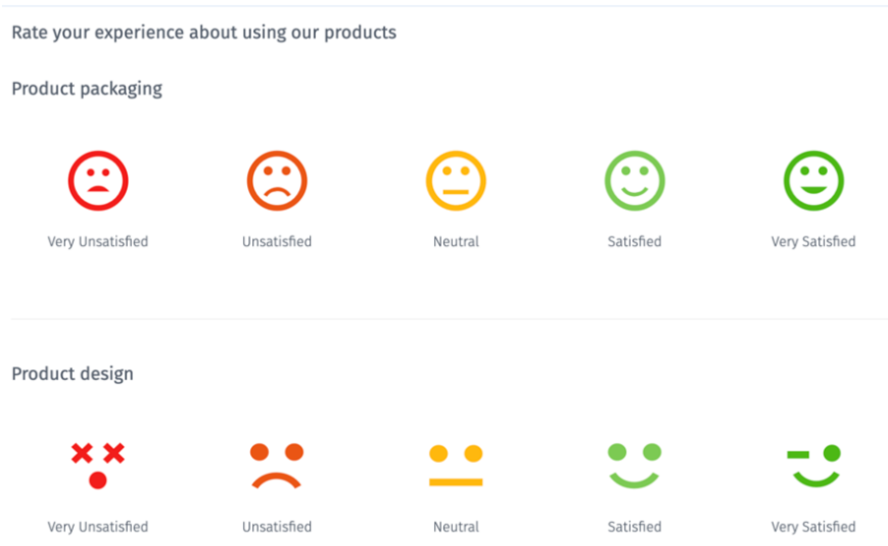
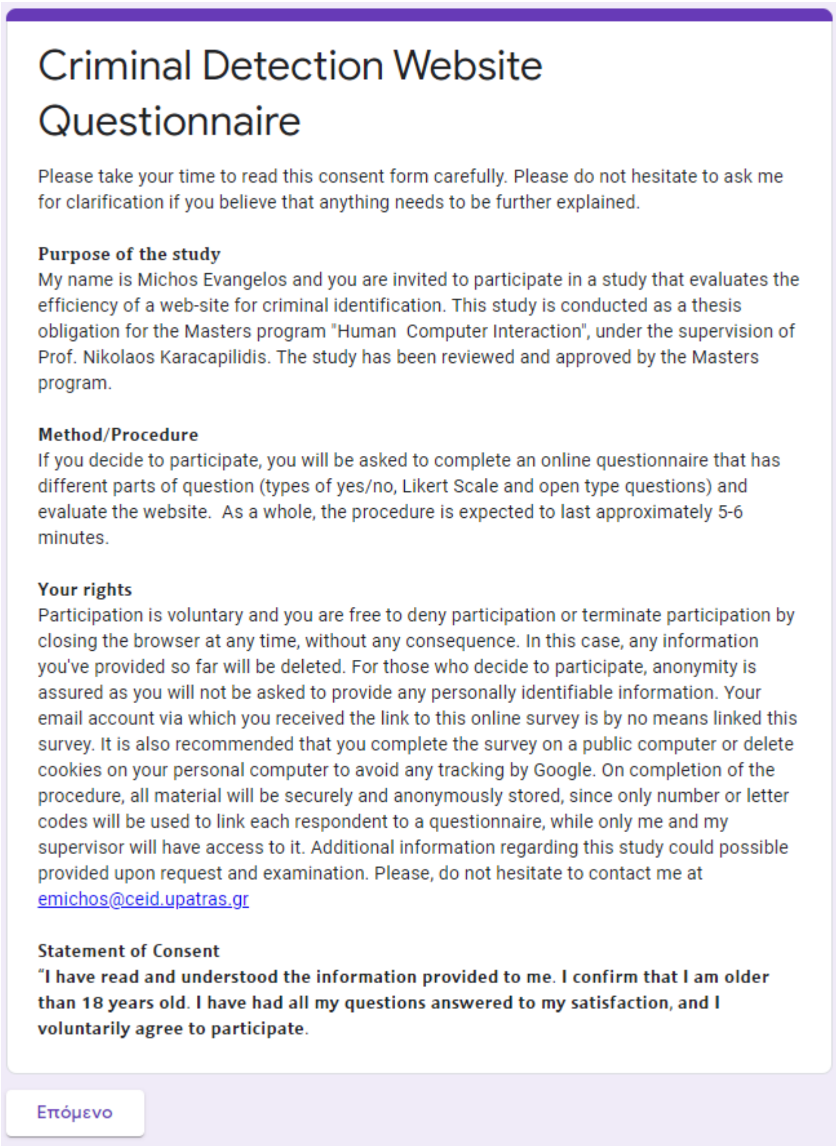


Figure 98. Purpose of a Likert scale questionnaire [89]

The full list of the questions that were included in the questionnaire can be found by visiting **Annex II – Usability Questionnaire**. The purpose of the work was not to make a statistical analysis, but to use the questionnaire as an additional mean of evaluation. After all, the number of participants as it was mentioned above it was relatively small and equal to 30, by the time the submission deadline was closed. As a result, the following will tackle the most important and interesting questions of the questionnaire. The questionnaire can be accessed by clicking here:

https://docs.google.com/forms/d/1v91rRXIZKu4VzCPzEosD9B9Z0zSoBzDjcsQI_hdqwg4



Criminal Detection Website Questionnaire

Please take your time to read this consent form carefully. Please do not hesitate to ask me for clarification if you believe that anything needs to be further explained.

Purpose of the study
My name is Michos Evangelos and you are invited to participate in a study that evaluates the efficiency of a web-site for criminal identification. This study is conducted as a thesis obligation for the Masters program "Human Computer Interaction", under the supervision of Prof. Nikolaos Karacapilidis. The study has been reviewed and approved by the Masters program.

Method/Procedure
If you decide to participate, you will be asked to complete an online questionnaire that has different parts of question (types of yes/no, Likert Scale and open type questions) and evaluate the website. As a whole, the procedure is expected to last approximately 5-6 minutes.

Your rights
Participation is voluntary and you are free to deny participation or terminate participation by closing the browser at any time, without any consequence. In this case, any information you've provided so far will be deleted. For those who decide to participate, anonymity is assured as you will not be asked to provide any personally identifiable information. Your email account via which you received the link to this online survey is by no means linked this survey. It is also recommended that you complete the survey on a public computer or delete cookies on your personal computer to avoid any tracking by Google. On completion of the procedure, all material will be securely and anonymously stored, since only number or letter codes will be used to link each respondent to a questionnaire, while only me and my supervisor will have access to it. Additional information regarding this study could possible provided upon request and examination. Please, do not hesitate to contact me at emichos@ceid.upatras.gr

Statement of Consent
"I have read and understood the information provided to me. I confirm that I am older than 18 years old. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate.

Επόμενο

Figure 99. Beginning of the questionnaire

Starting with the relevant demographic information (see *Figure 100*, *Figure 101* and *Figure 102*), we can see that the vast majority of the employees that answered our questionnaire are between 40 and 49 years old (30%), whereas the ages between 20-29 and 50-59 come in second tied at 23,3%. The biggest age gaps are the 18-19 year olds and the 60 years old or older ages, that contributed the least. The majority of the educational level of the submitted responses have a postgraduate degree (30%), whereas the 2nd place take submissions that had a Bachelor's Degree (23.3%). Approximately 23.3% of the employees that submitted the questionnaire have been working in that precinct for about '10-19 years' and right after that, the 20% of them has been working for '20-29 years'. The least submitted choice was the '30 years and more' and the '4-6 years' selection.

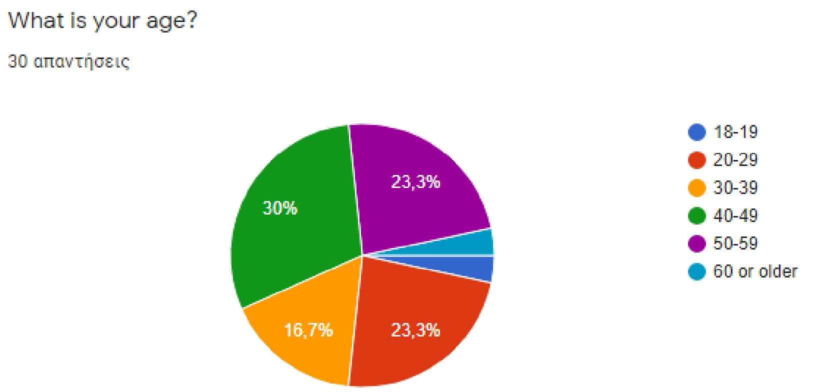


Figure 100. Questionnaire Q1

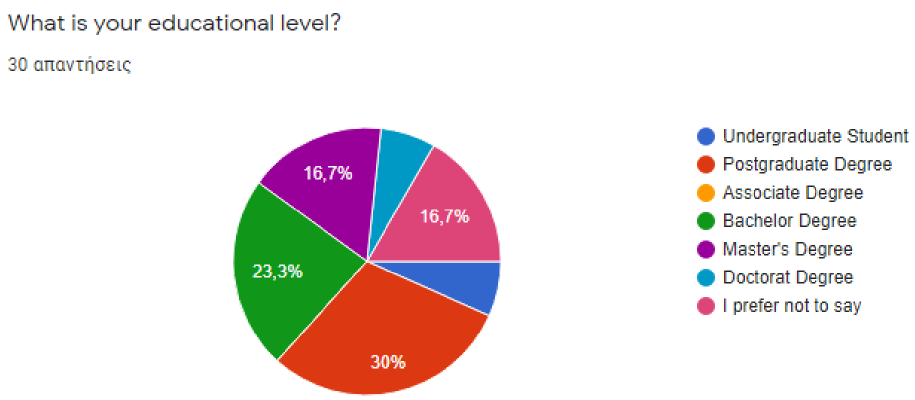


Figure 101. Questionnaire Q3

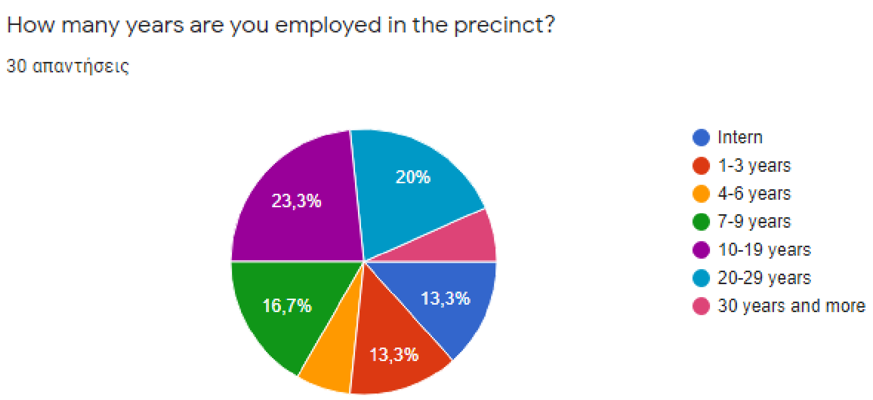


Figure 102. Questionnaire Q4

Up next are the most important usability evaluation questions regarding the general impression given to the employees by the website (see *Figure 103 - Figure 106*). As we can see, the vast majority of the employees stated that the purpose and scope of the website was clear (86.7%), whereas the 90% of them revealed that they felt that the website endorsed interactivity. Regarding the website's responsiveness in different platforms (as stated in the requirements early on during the development, the website was envisioned to also operate on other forms of media, like phones and tablets), 56.7% of them stated that they tested the website with such media (mostly mobile phones), whereas 36.7% of them did not (even) check if the website was supported in such media. Regarding the language support, the hopeful amount of 76.7% % stated that the platform could be understood from people of different nationality, meaning that they found the setting to change the language in the settings. On the other hand, the 23.3% of them either wanted to be in a neutral position or did not understand that they could change the language.

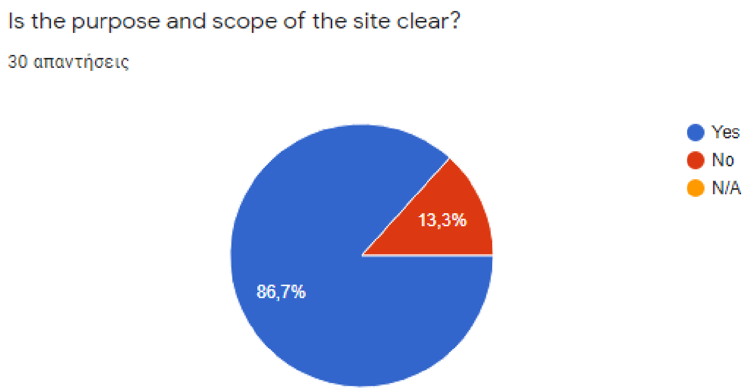


Figure 103. Questionnaire Q6

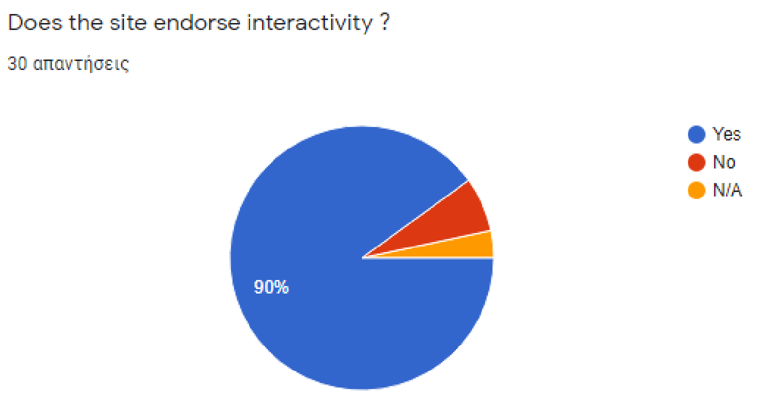


Figure 104. Questionnaire Q9

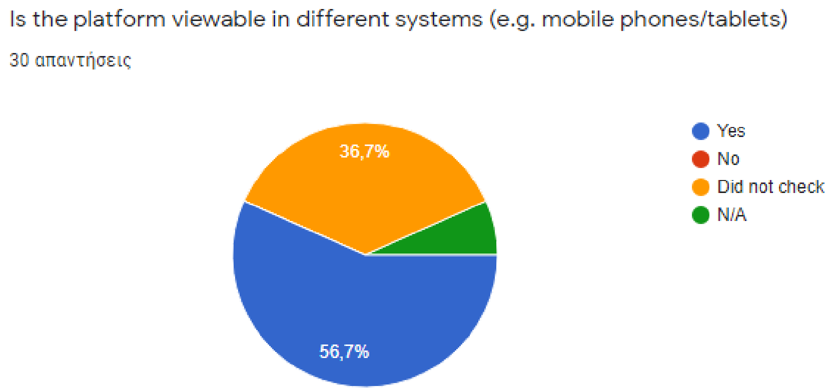


Figure 105. Questionnaire Q11

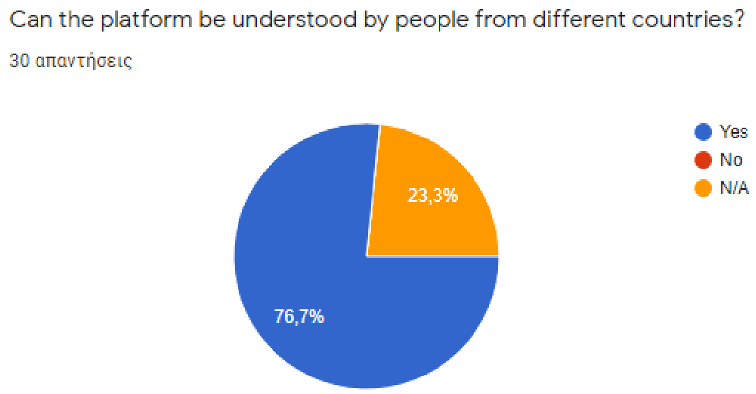


Figure 106. Questionnaire Q12

Next, we will look at the specific evaluation questions on the website's design and interactivity. As we can see below (*Figure 107 - Figure 112*), around 66.7% of the employees agree (either softly or strongly) that the management panel to add/edit/delete users and criminals was appealing. The percentage of the employees that remained neutral is the same with those that disagreed and is close to 23 – 26%. 66.7% of the people that responded believe that it is possible to recover from errors encountered in the platform (ranging from DB query errors up to accessing webpage that are restricted to employees). Regarding the question of whether they felt that the website was offering innovative things, the 53.3% of them consider the platform's features innovative and it goes without saying that they are implying that the live-stream face recognition was the most innovative feature among them, whereas the 33.3% of them remained neutral. We are proud to see that 80% of the employees felt that a) the website was easy to use and b) they can efficiently keep up with their work, something that is extremely important as the employee age in such precinct ranges from somewhere around 20 to 60+ years old and it is nearly impossible to please everyone. Half (50%) of the employees believe that they can turn up productive in the distant future with this platform, while 16.7% of them disagrees and 33.3% of them remains neutral so far.

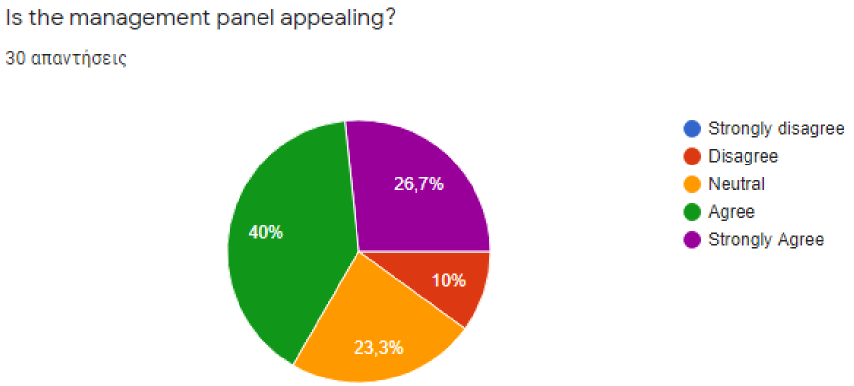


Figure 107. Questionnaire Q15

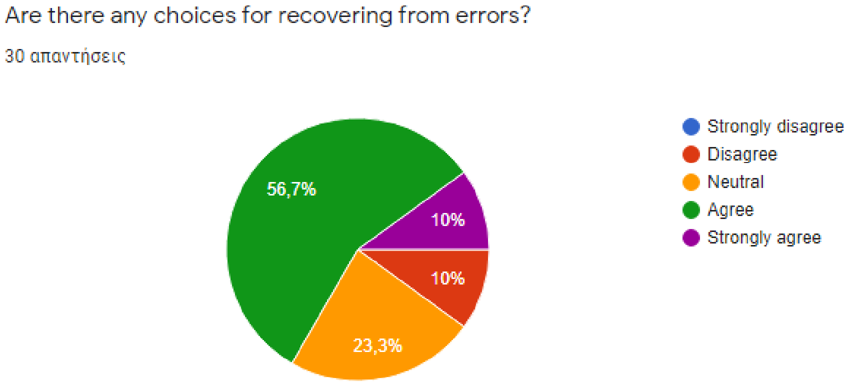


Figure 108. Questionnaire Q16

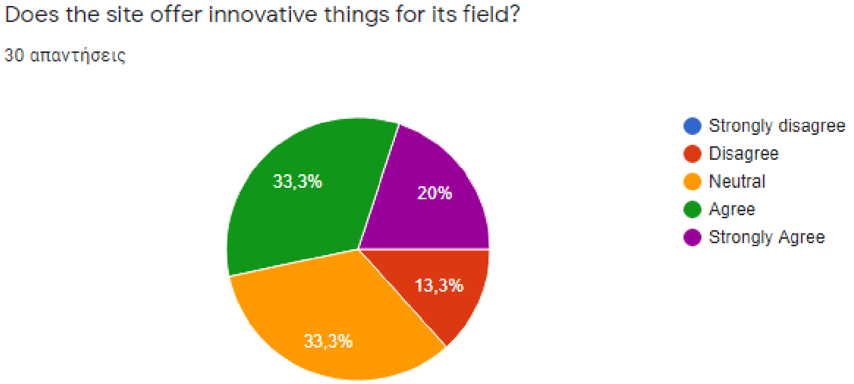


Figure 109. Questionnaire Q20

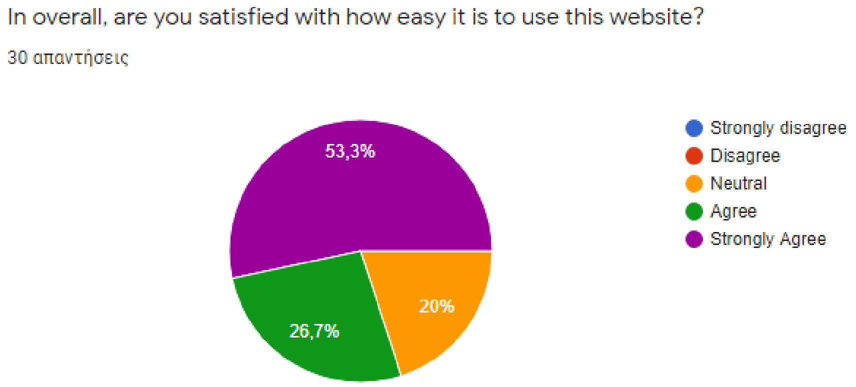


Figure 110. Questionnaire Q21

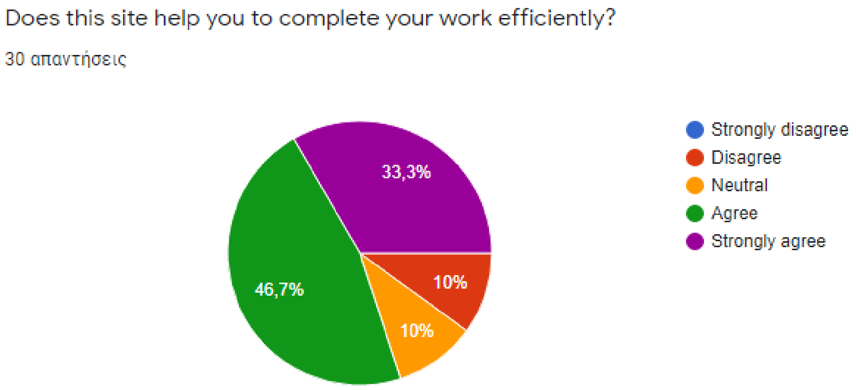


Figure 111. Questionnaire Q23

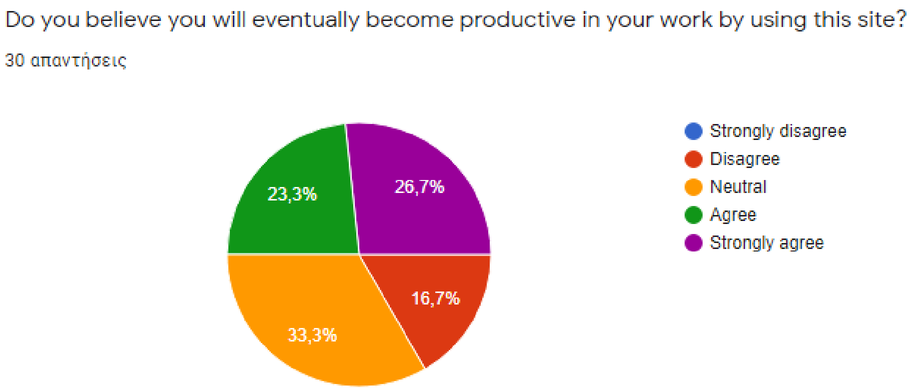
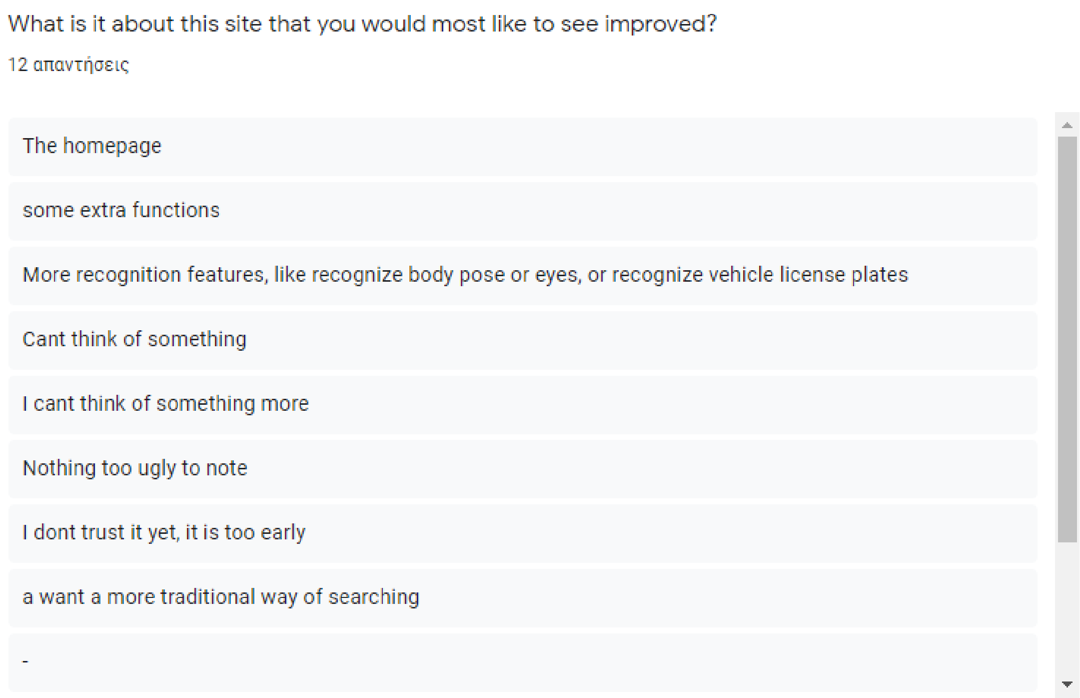


Figure 112. Questionnaire Q24

The latest questions in the questionnaire are open type questions, where employees could openly express their honest opinions on the platform without restrictions (see *Figure 113 - Figure 116*). On whether they would continue to use the website on a regular basis, 14 of the 22 submission were positive, while the remaining 8 of them where still not convinced about the platform due to the country it is being implemented mostly. Regarding the question on what site feature they would like to see most improved, the responses (12 in total) vary between users that were happy with the design and could not think of something to add and those that very vaguely suggested changes like in the home page, or another way of searching for criminal in the livestream page. Regarding the question of what changes they would generally suggest for the website, we received 13 responses and the requests varied between larger databases for criminals and more recognition features (license plate, human pose and stolen clothes), while once again, others had nothing to note. On the question of whether this platform could effectively help the fight towards crime, almost 9 out of the 13 responses (69%) responded positively where the rest 4 responses (31%) were still not convinced that such a platform could work (the country plays a very important factor on this). Lastly, we asked employees to use only one word to rate the whole platform and to this end, 17 out of the 18 response (94%) were positive, using words like insightful, appealing, promising, innovative, useful and satisfying, whereas only one person (6%) judged the website as complicated.


What is it about this site that you would most like to see improved?

12 απαντήσεις

The homepage

some extra functions

More recognition features, like recognize body pose or eyes, or recognize vehicle license plates

Cant think of something

I cant think of something more

Nothing too ugly to note

I dont trust it yet, it is too early

a want a more traditional way of searching

-

Figure 113. Questionnaire Q30

What changes or additional features would you suggest for this website?

13 απαντήσεις

A bigger database for criminals

a section only for criminals that are identified from the caera feed

a different design approach

More recognition features, like recognize body pose or eyes, or recognize vehicle license plates

Cant think of something

I cant think of something more

Nothing in particular

I am not a person that likes technology

So far, nothing

Figure 114. Questionnaire Q31

Do you believe that this site can contribute to the fight against crime ?

13 απαντήσεις

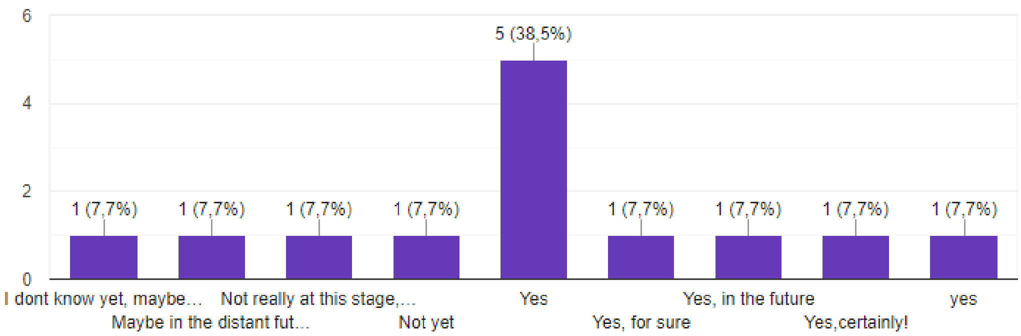


Figure 115. Questionnaire Q32

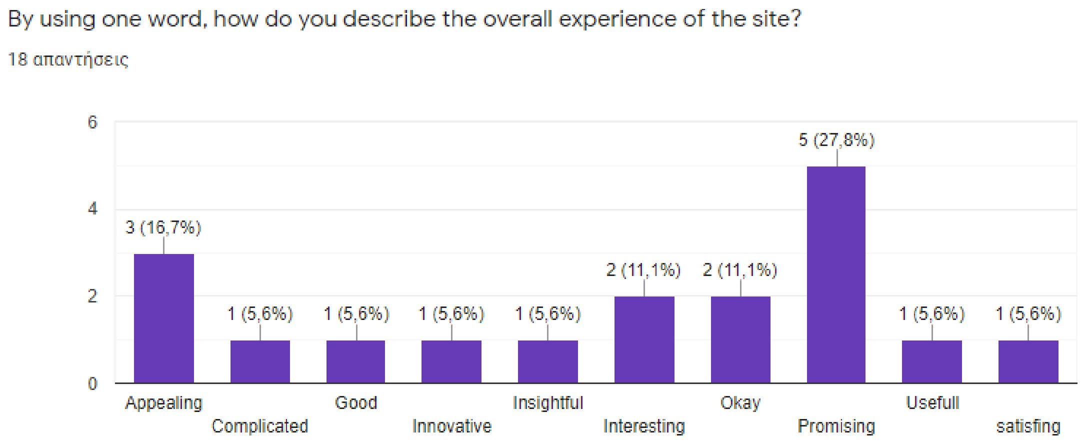


Figure 116. Questionnaire Q33

Chapter 5. Thesis Closure

5.1 Limitations

So far, our approach could seem ideal for the interested parties. Yet, significant limitations and concerns exist when facial recognition applies for criminal identification.

The very first fear that should be generated in everyone's head is the fact that it may not work well enough. To be fair, this is indeed an issue, since technology is always prone to false positives and false negatives and particularly when utilized with noise imagery from the Close-circuit television (CCTV) cameras that have been already installed years or even decades ago and the state has never considered updated the camera infrastructure in the city. And this may not have been a great issue so far, but when face recognition cameras begin to be used in order to arrest, convict or imprison people, then a possibly faulty basis can certainly cause real harm.

What is even worse is the fact that errors that occur are not evenly distributed, as facial recognition systems have been regularly found in the past to be inaccurate at identifying people with darker skin [86]. This can of course be solved with a new camera-system infrastructure, yet this demands an adequate amount of funds from the state, set aside the fact that as technology advances are facial recognition becomes more and more included in our everyday lives, people might eventually think that facial recognition inherently undermines freedom by enabling perfect surveillance of everyone, all the time.

Furthermore, it has been revealed that even though police tend to praise the different technologies used in order to improve future investigations, many agencies prefer to keep their methods as secret as possible [86]. This results in people getting arrested or convicted without them knowing they were recognized through face identification from cameras and exactly because the police does not treat facial recognition as adequate evidence in court, this approach sometimes does not turn up in public documents and has not been many times a subject of judicial rulings.

Last but not least, people know very little about facial recognition technology, set aside how efficient it will be in reducing crime rates. When systemic federal regulations or permitting processes are not apparent, police can only gather information from stories, interviews, public reports and investigative reports. This could mean that even though a police department or a police precinct may publicly acknowledge the fact that facial recognition is utilized from their side for criminal apprehension, they might not collect or share any king of tangible metrics or analytics on how effective is their approach, leaving everyone on the outside.

5.2 Conclusion

We saw that in today's world, face recognition is more and more utilized in order to search and identify shoplifters, retail criminals, or people with a history of frauds or crimes. This means that their pictures can, after being processed, be matched against a large criminal database and prevent and reduce crime rates by identifying them from installed cameras. The ability to efficiently recognize individuals through a combination of their facial characteristics (e.g. their eyes) is certainly an important phenomenon and the goal of recognizing peers through their faces is something already apparent on our everyday lives.

To this end, we decided to build a website platform from scratch that targets a) upon extending the existing work of facial recognition algorithms regarding accuracy and efficiency and b) at proposing an online platform that can be used from the police forces towards effective real-time human recognition. The main features of the platform included a) inserting, editing and deleting user and criminal information and b) searching for criminals based on their picture through a livestream camera feed and identify them. The platform supported two different types of users: a) Police employees in the headquarters/precincts and b) Police Administrators, with a higher level of access and also responsible for database maintenance. Regarding the facial recognition algorithm, we used the Haar Cascade algorithm, which is widely considered one for the most efficient and used algorithms for that cause. The development of the followed the MVC architectural pattern, separating the platform in three different logical components. As for the criminal identification, it will made possible through image pattern recognition between the provided criminal's image and snapshots of identified faces from the livestream feed. The platform included a live feed section, accompanied with different options for video filters, enabling the user to select the best filter, depending on the relevant situation of the physical surroundings for better recognition results.

The stages from starting to envision the final product up until its final form included many re-iterations, re-designing and re-inventing. After extensive research both on facial recognition algorithms and relevant platforms, we gathered a first idea that was later on enriched with information gathered from the interviews conducted. Following the PACT model throughout our design, we developed relevant personas and scenarios that recognized that different people have different needs and expectations and kept us designers away from the notion that if we like it, the final users will also like it. After creating the very first draft version of the platform, HE was conducted which helped us enumerate the findings from the study of the HCI evaluators gathered, which offered their insights and recommendations that finally enhanced the user's experience for our final platform. After taking into consideration all the information gathered from the HE feedback, we incorporated all aspects and issues mentioned into our final product. The process of fine-tuning the website also included perfecting several aspects of the website that even through were not mentioned from the HE, we considered that it was of high importance to be optimized. Finally, we were able to launch the finalized version of our platform. To this end, an online questionnaire was disseminated to all employees of the same precinct used for the interviews in order to evaluate the final version of the platform. The reason for starting this process is that this profession is accustomed and aware of technologies that identify criminals, they know what characteristic a system like the proposed one of this thesis needs to have and can propose recommendations for future development. Results were very positive, as the vast majority of employees stated with their responses that they found that the website appealing, insightful, innovative, easy to use and acknowledged the different features offered. Amongst the most interesting features stated by the employees for future development include 1080p camera support for better recognition, implementation of such websites from other regions, ability to also recognize license plates, human poses, clothes, eyes and alternative ways to search for criminals.

5.3 Future Work

In the distant future, it is expected that facial recognition will transform into an even-larger part of the biometrics market and digital transformations around the world. This means that the vast majority of the diverse facial recognition markets will grow faster than expected, combined with the COVID-19 pandemic crisis, which also led to more face recognition systems and other biometric techniques relevant to the subject. As it is expected digitization and digitalization will accelerate in several areas because of the pandemic, some inevitably will see increasing usage of face recognition technologies. According to [87], it is expected that facial recognition hardware modules like the Face ID on the latest iPhone models will be the fastest-growing smartphone biometric, while at the same time, we will be experiencing an outrage of mobile devise ranging from 96 million devices in 2019 towards over 800 million devices in 2024. According to forecasts coming from the same study, face recognition hardware will continue to grow over 50% annually and eventually, these applications will be included on almost 1.3 billion smartphones, with companies like iProov and Mastercard already offering face recognition authentication strong enough to alloy for online payments and other relevant user-authentication services. As a result, there exists many other scenarios where such authentication will be needed for payments or other services in applications that are consumer-oriented (let us not forget that face recognition hardware is already used for building access on other areas than the hotel industry too).

In EU, there had been calls to ban relevant face recognition applications for some time as not all of them adhere to the GDPR's personal data protection rules and in fact, in some countries, they have been banned. Yet, this does not mean that they will go away in the near future, as they keep on being developed and becoming more and more accurate. More powerful special micro-controllers and processors, better images with better cameras and on-chip processing and edge computing for more intelligence in and close to the cameras, 3D facial recognition, and more accurate face recognition algorithms thanks to neural network algorithms are just some of the evolutions from the past few years. According to the report in [88], the overall accuracy of such face recognition technologies has dramatically improved from 2013 to 2018, stating that '*the accuracy gains stem from the integration, or complete replacement, of prior approaches with those based on deep convolutional neural networks*'. Face recognition technologies keep evolving and creating an industrial revolutions, as the relevant algorithms are becoming even more tolerant on poor quality images (which is the case in the majority of the countries that have relevant surveillance systems). The main reasons why facial recognition isn't going to go anywhere:

- demand is high (and not just from governments and law enforcement, as said there are applications in marketing, retail, the protection of critical facilities, etc.) and the COVID-19 pandemic increased demand,
- all countries and supranational organizations use it and will increasingly do so although the ways in which and reasons for what differ, as do debates with different cultures and attitudes,
- although there are more accurate biometric measurements it's far easier to use – at scale – and offers far more opportunities for those using it, while avoiding contact biometrics and thus hygiene concerns,

As we already learned during the completion of this thesis, Haar Cascade is an Object Detection Algorithm used to identify faces in an image or a real time video. The algorithm is given a lot of positive images consisting of faces, and a lot of negative images not consisting of any face to train on them. The model created from this training consists of all the relevant models (face or eye detection, upper/lower body detection, license plate detection and more) and this leaved an open research field for future work regarding possible optimizing eye detection or licence plate detection for stolen vehicles or body detection.

Looking at the current state in our country (Greece), the market seems small. But, predicting its size and evolution seems like a very challenging task, given all the debates and uncertainties ahead. The scope of the approach of the research obviously doesn't show the full' market value' beyond solutions and services and as mentioned, there is a massive market for facial recognition infrastructure. To this end, enabling police forces with having such powerful tools in their hands to fight crime will only benefit in the near future. The platform can be extended in multiple ways, like for example the incorporation of e-mail support where emails can be automatically sent by the system regarding the criminal's last found location. The platform can also be populated with an very high amount of IP cameras connected and the only thing required for this is the URL of the online camera (and of course repeating a few lines of HTML code). Yet, whether our country will follow such approaches already implemented in various other countries still remains to be seen…

REFERENCES

[1] Human-Computer Interaction, 3rd Edition.

[2] Interaction Design Foundation, '*Human-Computer Interaction (HCI)*'. Retrieved November 2020.

[3] https://en.wikipedia.org/wiki/Human%E2%80%93computer_interaction

[4] L.Liu, M. Tamer Özsu, '*Encyclopedia of Database Systems*', Springer Science and Business Media, 2009.

[5] T. T. Hewett, R. Baecker, S. Card, T. Carey, J. Gasen, M. Mantei, G. Perlman, G. Strong, and W. Verplank, '*ACM SIGCHI Curricula for Human-Computer Interaction*'. Technical Report. Association for Computing Machinery, New York, NY, USA, 1992.

[6] Interaction Design Foundation, '*User Centered Design*'. Retrieved December 2020.

[7] https://en.wikipedia.org/wiki/User_interface

[8] Interaction Design Foundation, '*User Experience (UX) Design*'. Retrieved December 2020.

[9] Getsmarter.com, '*14 Human Computer Interaction examples*', 2018.

[10] https://en.wikipedia.org/wiki/ENIAC

[11] https://en.wikipedia.org/wiki/Head-mounted_display

[12] D. Cardinal, '*Making gesture recognition work: Lessons from Microsoft Kinect and Leap*', Extremetech.com, 2013.

[13] Maketecheasier.com, '*Supercharging iOS Calling with Google Voice*', 2019.

[14] School of Electrical and Information Engineering, '*Brainternet*', University of the Witwatersband, Johannesburg. Retrieved January 2021.

[15] C. Stephanidis, et al. '*Seven HCI Grand Challenges*', International Journal of Human-Computer Interaction, 35 (14), pp. 1229-1269, 2019.

[16] C. Frost, '*5G is being used to perform remote surgery from thousands of miles away, and it could transform the healthcare industry*', BusinessInsider.com, 2019.

[17] E. Bieller, '*What's the Future of Human-Computer Interaction (HCI)?*', CareerFoundry.com, 2018.

[18] http://www.guideme-project.upatras.gr/

[19] https://www.wesar-project.upatras.gr/

[20] https://en.wikipedia.org/wiki/Computer_vision

[21] https://en.wikipedia.org/wiki/Signal_processing

[22] ScienceDaily, '*Neurobiology*', ScienceDaily.com. Retrieved January 2021.

[23] Builtin, '*Robotics Technology*', Builtin.com. Retrieved January 2021.

[24] https://en.wikipedia.org/wiki/Solid-state_physics

[25] S. Saxena, '*5 Exciting Computer Vision Applications with relevant datasets*', Analyticsvidhya.com, 2020.

[26] K. Hébert-Losier, I. Hanzlíková and L. Streeter, '*The 'DEEP' Landing Error Scoring System*', Applied Sciences, 10(3), 892, 2020.

[27] https://github.com/CMU-Perceptual-Computing-Lab/openpose

[28] GeeksforGeeks, '*OpenPose: Human Pose Estimation Method*', GeeksforGeeks.com. 2020.

[29] https://github.com/tensorflow/tfjs-models/tree/master/posenet

[30] Medium, '*Real-time Human Pose Estimation in the Browser with TensorFlow.js*', Medium.com, 2018.

[31] https://storage.googleapis.com/tfjs-models/demos/posenet/camera.html

[32] N. Drum, '*People are turning Marvel actors into women and the results are interesting*', Comicbook.com 2019.

[33] J. Brownlee, '*A Gentle Introduction to Generative Adversarial Networks (GANs)*', 2019.

[34] https://pjreddie.com/darknet/yolo/

[35] K. He, G. Gkioxari, P. Dollár and R. Girshick, '*Mask R-CNN*', in IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(2), pp. 386-397, 2020.

[36] DL Cade, '*Facebook now lets you turn any 2D photo into a 3D Image using AI*', PetaPixel.com, 2020.

[37] Facebook AI, '*Powered by AI; Turning any 2D photo into 3D using convolutional neural nets*', ai.facebook.com, 2020.

[38] S. Ali Eslami and D. J. Rezende, '*Neural scene representation and rendering*', DeepMind.com, 2020.

[39] V. Fang et al., '*Sensitivity for Chest CT for COVID-10: Comparison to PT-PCR*', Radiology, 296(2), 2020.

[40] https://www.kaggle.com/bachrr/covid-chest-xray

[41] J. Brownlee, '*A Gentle Introduction to Object Recogntion with Deep Learning*', Machine Learning Mastery, 2019.

[42] ImageNet, '*ImageNet Large Scale Visual Recognition Challenge (ILSVRC)*', ImageNet.com. Retrieved March 2021.

[43] R. Girshick, J. Donahue, T. Darrell and J. Malik, '*Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation*', IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587, 2014.

[44] https://github.com/rbgirshick/rcnn

[45] R. Girshick, '*Fast R-CNN*', IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448, 2015.

[46] https://github.com/rbgirshick/fast-rcnn

[47] S. Ren, K. He, R. Girshick and J. Sun, '*Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*', Computer Vision and Pattern Recognition, 2016.

[48] https://github.com/rbgirshick/py-faster-rcnn

[49] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, '*You Only Look Once: Unified, Real-Time Object Detection*', IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.

[50] https://github.com/pjreddie/darknet/wiki/YOLO:-Real-Time-Object-Detection

[51] J. Redmon and A. Farhadi, '*YOLO9000: Better, Faster, Stronger*', IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517-6525, 2017.

[52] J. Redmon and A. Farhadi, '*YOLOv3: An Incremental Improvement*', Computer Vision and Pattern Recognition, 2018.

[53] M. Iqtait, and F. Mohamad, '*Feature extraction for face recognition via Active Shape Model (ASM) and Active Appearance Model (AAM)*', IOP Conference Series Materials Science and Engineering. 2018.

[54] T. Le, and V. Truong, '*Face Alignment Using Active Shape Model And Support Vector Machine*'. International Journal of Biometric and Bioinformatics, 2011.

[55] B. van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny and M. A. Viergever, '*Active shape model segmentation with optimal features*', in IEEE Transactions on Medical Imaging, 21(8), pp. 924-933, 2002.

[56] H. Lu, and F. Yang, '*Active Shape Model and its Application to Face Alignment',* in Subspace Methods for Pattern Recognition in Intelligent Environment. Studies in Computational Intelligence, vol 552. Springer, Berlin, Heidelberg, 2014.

[57] https://en.wikipedia.org/wiki/Active_shape_model

[58] T.F. Cootes, and C.J. Taylor, '*Statistical models of appearance for computer vision*'. 2004.

[59] T.F. Cootes, G.J. Edwards, and C. Taylor, '*Active Appearance Models. Pattern Analysis and Machine Intelligence*', IEEE Transactions, pp. 681 – 685, 2001.

[60] G. Tzimiropoulos, and M. Pantic, '*Fast Algorithms for Fitting Active Appearance Models to Unconstrained Images*'. International Journal of Computer Vision, 122, pp. 17–33, 2017.

[61] X. Zhao, S. Shan, X. Chain, and X. Chen, '*Locality-Constrained Active Appearance Model*', in Computer Vision – ACCV 2012. Lecture Notes in Computer Science, vol 7724. Springer, Berlin, Heidelberg, 2013.

[62] IIAU-Lab, '*Active Appearance Model*'. Retrieved March 2021.

[63] L. Cuimei, Q. Zhiliang, J. Nan and W. Jianhua, '*Human face detection algorithm via Haar cascade classifier combined with three additional classifiers*', 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), pp. 483-487, 2017.

[64] R. Padilla, C. Filho, and M. Costa, '*Evaluation of Haar Cascade Classifiers for Face Detection*', 2012.

[65] Le T.H. Phuc, H. Jeon, N. Truong, T.N. Nguyen, and J.J. Hak, '*Applying the Haar-cascade Algorithm for Detecting Safety Equipment in Safety Management Systems for Multiple Working Environments*', Electronics, 8(10), 1079, 2019.

[66] P. Viola and M. Jones, "*Rapid Object Detection using a Boosted Cascade of Simple Features*", IEEE Conference Computer Vision Pattern Recognition. 1. I-511, 2001.

[67] S. Soo, '*Object detection using Haar-cascade Classifier*', Institute of Computer Science, 2014.

[68] T. Ojala, M. Pietikainen and D. Harwood, '*A comparative study of texture measures with classification based on feature distributions*', Pattern Recognition vol. 29, 1996.

[69] A. Gaikwad, '*LBP and PCA based on Face Recognition System*', 5, pp. 368-373, 2018.

[70] D. Huang, C. Shan, M. Ardabilian, Y. Wang and L. Chen, '*Local Binary Patterns and Its Application to Facial Image Analysis: A Survey*', in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41(6), pp. 765-781, 2011.

[71] K. Salton do Prado, '*Face Recognition: Understanding LBPH Algorithm*', Towardsdatascience.com, 2017.

[72] A. K. Jain, Yu Zhong and S. Lakshmanan, '*Object matching using deformable templates,*' in IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(3), pp. 267-278, 1996.

[73] Alan L. Yuille, Peter W. Hallian, and David S. Cohen, '*Feature Extraction from Faces Using Deformable Templates*', International Journal of Computer Vision, 8(2), 99-111, 1992.

[74] A.V. Dalca, M. Rakic, J. Guttag and M.R. Sabuncu, '*Learning Conditional Deformable Templates with Convolutional Networks*', 33rd Conference on Neural Information Processing Systems, 2019.

[75] J. Miao, W. Gao, Y. Chen, and J. Lu, '*Gravity-Center Template Based Human Face Feature Detection*', in Advances in Multimodal Interfaces - ICMI 2000. Lecture Notes in Computer Science, vol 1948. Springer, Berlin, Heidelberg, 2001.

[76] X. Fan, B. Yin and Y. Sun, '*Yawning Detection for Monitoring Driver Fatigue*', 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, pp. 664-668, 2007.

[77] P. Viola & M. Jones, "*Rapid Object Detection using a Boosted Cascade of Simple Features*", IEEE Conference Computer Vision Pattern Recognition. 1. I-511, 2001.

[78] https://github.com/opencv/opencv/tree/master/data/haarcascades

[79] G. S. Behera, '*Face Detection with Haar Cascade*', TowardsDataScience.com, 2020.

[80] DesignCouncil, '*What is the framework for innovation?*', DesignCouncil.com. Retrieved April 2021.

[81] D. Benyon, '*Designing Interactive Systems*', 4th Edition, Pearson Publications, 2019.

[82] R.F. Dam and T.Y. Siang, '*Personal – A Simple Introduction*', Interaction Design Foundation, 2021.

[83] J. Nielsen, '*10 Usability Heuristics for User Interface Design*', Nielsen Norman Group, 2020.

[84] https://mottie.github.io/tablesorter/docs/

[85] https://play.google.com/store/apps/details?id=com.pas.webcam&hl=el&gl=US

[86] J. Schuppe, '*How facial recognition became a routine policing tool in America*', NBC News. Retrieved May 2021.

[87] S. Smith, '*Facial Recognition Hardware to feature on over 800m mobiles by 2024, but software will win out*', JuniperResearch, 2019.

[88] https://www.nist.gov/system/files/documents/2019/11/20/frvt_report_2019_11_19_0.pdf

[89] https://www.questionpro.com/article/likert-scale-survey-questions.html

Annex I – HEURISTIC EVALUATION REPORT

The heuristic evaluation report used by the experts to evaluate the draft platform.

Name:
Surname:
Email:

Severity Rating	Meaning
0	*Team does not agree that issue impacts system usability*
1	*Cosmetic problem only; need not to be fixed unless extra time is available on project*
2	*Minor usability problem; fixing this should be given low priority*
3	*Major usability problem; important to fix, so should be given high priority*
4	*Usability catastrophe; imperative to fix before product can be released*

Heuristic	Difficulties	Severity Rating	Recommendations
1. Visibility of system status *The system should always keep users informed about what is going on, through appropriate feedback within reasonable time. The system should always inform users, they should not have to guess or remember things.*			
2. Match Between system and the real world *The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order. The system language as easy and plain for the simple user*			

to understand (extremely important for error messages).			
<u>3. User control and freedom</u> *Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo. Users should be able to redo and undo actions and go back or exit.*			
<u>4. Consistency and standards</u> *Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow the platform conventions.*			
<u>5. Error prevention</u> *Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action. We should try to prevent errors from happening through careful design, so users can't break the system.*			
<u>6. Recognition rather than recall</u> *Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.*			
<u>7. Flexibility and efficiency of use</u> *Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions. Shortcuts allow experts to speed up the experience for commonly used features.*			
<u>8. Aesthetic and minimalist design</u> *Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their*			

relative visibility. Irrelevant information should not be included in the system design, only the must have features should be included.			
9. Help users recognize, diagnose, and recover from errors *Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution. Inform users when an error has occurred, tell them what the problem is and how to solve it.*			
10. Help and documentation *Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large. Users should be provided with additional helping information for better understandings of user tasks.*			

Annex II – USABILITY QUESTIONNAIRE

The online Google Forms usability questionnaire disseminated to evaluate the final version of the platform

Criminal Detection Website Questionnaire

Please take your time to read this consent form carefully. Please do not hesitate to ask me for clarification if you believe that anything needs to be further explained.

Purpose of the study
My name is Michos Evangelos and you are invited to participate in a study that evaluates the efficiency of a web-site for criminal identification. This study is conducted as a thesis obligation for the Masters program "Human Computer Interaction", under the supervision of Prof. Nikolaos Karakapilidis. The study has been reviewed and approved by the Masters program.

Method/Procedure
If you decide to participate, you will be asked to complete an online questionnaire that has different parts of question (types of yes/no, Likert Scale and open type questions) and evaluate the website. As a whole, the procedure is expected to last approximately 5-6 minutes.

Your rights
Participation is voluntary and you are free to deny participation or terminate participation by closing the browser at any time, without any consequence. In this case, any information you've provided so far will be deleted. For those who decide to participate, anonymity is assured as you will not be asked to provide any personally identifiable information. Your email account via which you received the link to this online survey is by no means linked this survey. It is also recommended that you complete the survey on a public computer or delete cookies on your personal computer to avoid any tracking by Google. On completion of the procedure, all material will be securely and anonymously stored, since only number or letter codes will be used to link each respondent to a questionnaire, while only me and my supervisor will have access to it. Additional information regarding this study could possible provided upon request and examination. Please, do not hesitate to contact me at emichos@ceid.upatras.gr

Statement of Consent
"I have read and understood the information provided to me. I confirm that I am older than 18 years old. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate.

Επόμενο

Μην υποβάλετε ποτέ κωδικούς πρόσβασης μέσω των Φορμών Google.

Αίτημα για πρόσβαση με επεξεργασία

Criminal Detection Website Questionnaire

* Απαιτείται

Demographic Questions

General Information about your personal demographics

What is your age? *

- 18-19
- 20-29
- 30-39
- 40-49
- 50-59
- 60 or older

What is your gender? *

- Male
- Female
- Other
- I prefer not to say

What is your educational level? *

- ◯ Undergraduate Student
- ◯ Postgraduate Degree
- ◯ Associate Degree
- ◯ Bachelor Degree
- ◯ Master's Degree
- ◯ Doctorat Degree
- ◯ I prefer not to say

How many years are you employed in the precinct? *

- ◯ Intern
- ◯ 1-3 years
- ◯ 4-6 years
- ◯ 7-9 years
- ◯ 10-19 years
- ◯ 20-29 years
- ◯ 30 years and more

Πίσω   Επόμενο

Criminal Detection Website Questionnaire

* Απαιτείται

Evaluation Questions (Reply with Yes or No)

Some general questions regarding your understanding of the platform

Is it clear who is the authority of the site? *

- ◯ Yes
- ◯ No
- ◯ N/A

Is the purpose and scope of the site clear? *

- ◯ Yes
- ◯ No
- ◯ N/A

Can the site be used from everyone independent of their profession ? *

- ◯ Yes
- ◯ No
- ◯ N/A

Is the information presented objectively? *

- Yes
- No
- N/A

Does the site endorse interactivity ? *

- Yes
- No
- N/A

Does the site include permissions or licensing for the use of copyrighted images and multimedia? *

- Yes
- No
- N/A

Is the platform viewable in different systems (e.g. mobile phones/tablets) *

- Yes
- No
- Did not check
- N/A

Can the platform be understood by people from different countries? *

- Yes
- No
- N/A

Πίσω | Επόμενο

Μην υποβάλετε ποτέ κωδικούς πρόσβασης μέσω των Φορμών Google.

Criminal Detection Website Questionnaire

* Απαιτείται

Evaluation questions (Reply with Strongly Disagree -> Strongly Agree)

Specific Evaluation Questions on the website's design and interactivity

Is the design of live feed page of criminals confusing or overwhelming? *

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly agree

Are the menu choices (dropdowns) ordered in the most logical way? *

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

Is the management panel appealing? *

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Are there any choices for recovering from errors? *

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Did you understand whether your role was admin or employee? *

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Can it be understood from different levels of employees in the police department? *

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

Does the site offer alternative solutions for recognizing a criminal? *

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Does the site offer innovative things for its field? *

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

In overall, are you satisfied with how easy it is to use this website? *

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly Agree

Was it simple to navigate through the site? *

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly agree

Does this site help you to complete your work efficiently? *

- ◯ Strongly disagree
- ◯ Disagree
- ◯ Neutral
- ◯ Agree
- ◯ Strongly agree

Do you believe you will eventually become productive in your work by using this site? *

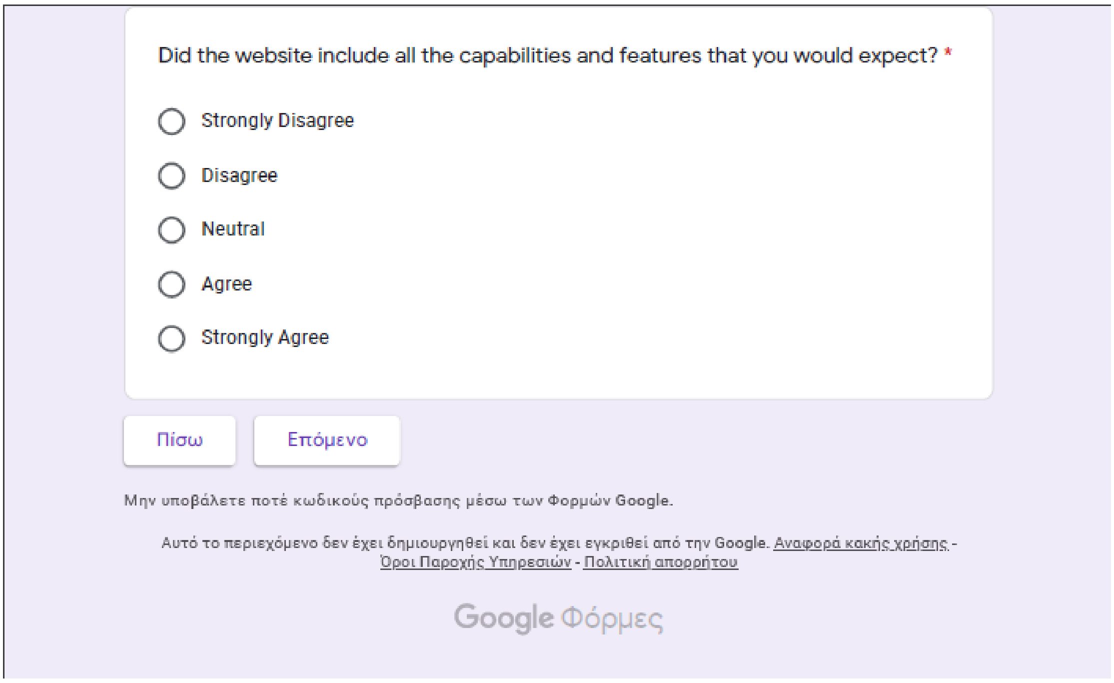
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Do you find the important information easily? *

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Was the interface of the website clear? *

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly Agree


Did the website include all the capabilities and features that you would expect? *

Strongly Disagree

Disagree

Neutral

Agree

Strongly Agree

Πίσω

Επόμενο

Μην υποβάλετε ποτέ κωδικούς πρόσβασης μέσω των Φορμών Google.

Αυτό το περιεχόμενο δεν έχει δημιουργηθεί και δεν έχει εγκριθεί από την Google. Αναφορά κακής χρήσης - Όροι Παροχής Υπηρεσιών - Πολιτική απορρήτου

Google Φόρμες

Criminal Detection Website Questionnaire

Evaluation questions (Open Type)

Optional Responses regarding your overall feel after using the website

How likely are you to use this site on a regular basis?

Η απάντησή σας

What features would influence your decision to continue using this website?

Η απάντησή σας

What is it about this site that you would most like to see improved?

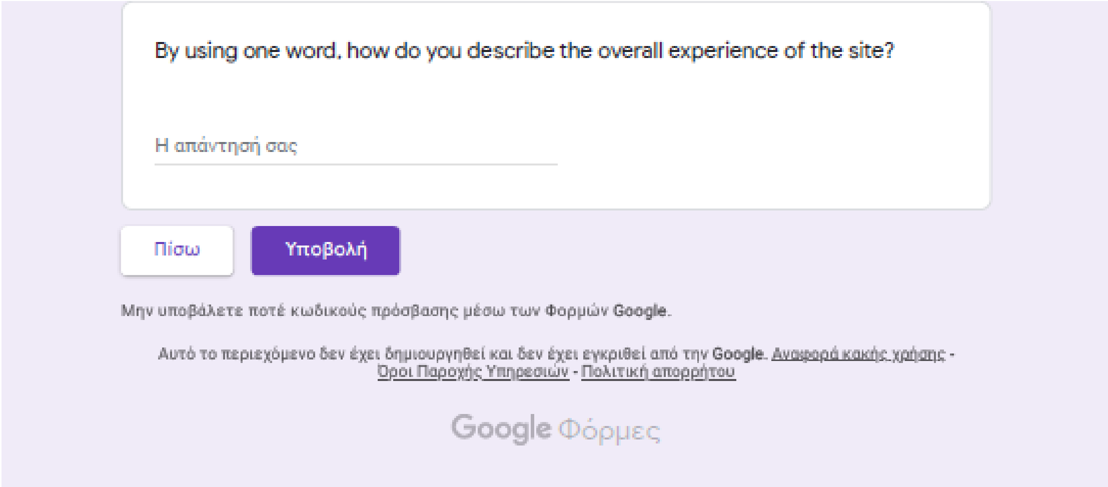
Η απάντησή σας

What changes or additional features would you suggest for this website?

Η απάντησή σας

Do you believe that this site can contribute to the fight against crime ?

Η απάντησή σας



By using one word, how do you describe the overall experience of the site?

Η απάντησή σας

Πίσω

Υποβολή

Μην υποβάλετε ποτέ κωδικούς πρόσβασης μέσω των Φορμών Google.

Αυτό το περιεχόμενο δεν έχει δημιουργηθεί και δεν έχει εγκριθεί από την Google. Αναφορά κακής χρήσης - Όροι Παροχής Υπηρεσιών - Πολιτική απορρήτου

Google Φόρμες

Annex III - ABBREVIATIONS

Abbreviations used in this dissertation, accompanied by their explanation.

Abbreviation	Explanation
2-D	Two-Dimensional
3-D	Three-Dimensional
5G	Fifth Generation
AAM	Active Appearance Model
ACM	Association for Computer Machinery
AI	Artificial Intelligence
ASM	Active Shape Model
BCI	Brain-Computer Interface
BLE	Bluetooth Low Energy
BS	Base Stations
CCD	Charge-Coupled Device
CHI	Computer-Human Interaction
CNN	Convolutional Neural Network
CQN	Generative Query Network
CS	Computer Science
CSS	Cascading Style Sheets
CT	Computer Tomography
ELBP	Extended Local Binary Pattern
ENIAC	Electronic Numerical Integrator and Computer
FPS	Frames Per Second
GAN	Generative Adversarial Network
HCI	Human-Computer Interaction
HCII	Human-Computer Interaction International
HE	Heuristic Evaluation
HMI	Human-Machine Interaction
HTML	HyperText Markup Language
IDE	Integrated Development Environment
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IoT	Internet of Things
KLM	Keystroke-level model
LBP	Local Binary Pattern
LC-AAM	Locality-constraint Active Appearance Model
LCD	Liquid Crystal Display
LESS	Landing Error Scoring System
ML	Machine Learning
NASA	National Aeronautics and Space Administration
OLED	Organic Light-Emitting Diode
OS	Operating System
PACT	People-Activities-Context-Technologies

PCA	Principal Component Analysis
PDN	Point Distribution Model
POV	Point of View
RAM	Random Access Memory
RDBMS	Relational Database Management System
RGB	Red-Green-Blue
RPN	Region Proposal Network
SVM	Support Vector Machine
UCD	User-Centered Design
UI	User Interface
US	United States
UV	UltraViolet
UX	User Experience
VR	Virtual Reality
XML	eXtensible Markup Language
YOLO	You Only Look Once