***UNIVERSITY OF MOHAMED BOUDIAF – M'SILA***


*Thesis submitted to the*

Faculty of Mathematics and Computer Science

*In partial fulfillment of the requirements for the degree of*

**Master** in **Computer Science**

By

**Chouiter Djamal**

# Implementation of A Smart Student Attendance System based on Facial Recognition

*In front of the jury composed of:*

| | | |
|---|---|---|
| Prof. Benouis Mouhamed | University of M'Sila | President |
| Prof. Heraguemi KamelEddine | University of M'Sila | Reporter |
| Prof. Amraoui NourEddine | University of M'Sila | Examiner |

*June, 2022*

# Acknowledgments

Firstly, I am grateful to Almighty Allah for my health and well-being that was necessary to complete this project work. Secondly, I would like to express my sincere gratitude to my supervisor, **Dr. KAMELEDDINE HERAGUEMI** for his valuable guidance, continuous support and patience. Last but not least I want to thank myself for all this hard work. Finally, I would like to thank my family for their unceasing encouragement, appreciable patience, support and attention. I also thank the honorable students who participated in this work.

# Abstract

Face recognition is very important technologies today, especially after extensive performance upgrades over the past decade, and these systems are now popular in areas such as security and commerce. In the traditional method, it is difficult to manage a large number of students in the classroom, since it takes time and has a high risk of error when entering data into the system, so it is not recommended. But the real difficulty lies in implementing an accurate, real-time attendance system. Real-time facial recognition is a convenient way to deal with the large number of students who attend every day. In this paper, a real-time presence detection system was created, and using the OpenCV library, the video is read as frames. This was developed this system based on the dlib model using deep learning techniques, which has an accuracy of 99.37, on the Labeled Faces in the Wild benchmark dataset, combined with the library simple Face Recognition library in python. For face detection, we used Histogram of Oriented Gradients (HOG). And for face recognition, we apply Euclidean distance calculation to identify the input face, our system is capable of recognizing multiple faces, and it will be a successful way to manage student attendance.

***Keywords: Face Detection, Face Recognition, Deep Learning, OpenCv, Dlib, HOG, Face Recognition Library.***

# Table of Contents

# List of figures

# List of tables

# Introduction

Every university needs a robust and stable system to record pupil attendance. Each university has its own practice, some universities use papers to manually record attendance during lectures, and some universities use biometric systems as the attendance system. The traditional system of manually memorizing student names is a waste of time, especially in the period of rapid-fire development of new technologies, as this system can take a long-time during attendance. The app can ameliorate attendance as it reduces teachers'workload, increases productivity and reduces mortal error, plus it offers better data security and easy stoner operation.

Attendance tracking using facial recognition is a smart way of attendance operating system. Facial recognition is more accurate and faster, and is extensively used along compared, with other biometrics similar as fingerprints, iris, autographs, etc.

The main purpose of this project is to make a face recognition- grounded attendance monitoring system for educational institutions to ameliorate the being attendance system and upgrade it to be more efficient and effective than what was preliminarily used manually that uses face recognition technology.

This document contains four chapters. The first chapter has two parts, the first introduces the basic concepts of biological systems in general, whereas, the other deals with the main concepts of face recognition. In the second chapter, we will introduce various algorithms to deal face recognition, also we will touch on some models of deep learning. In the third chapter we will explain in detail the mechanism of face recognition using the deep learning model Dlib, last chapter we will present the Implementation, testing and results obtained.

.

# Chapter 01

# Biometrics and Facial Recognition system

## 1.1. Introduction

Nowadays we are talking more and more about the insecurity in various sectors, as well as the IT. This means that tools need to be implemented in count this trend. Verifying the identity of individuals and identifying them is one of the means of ensuring this security, which is represented in the identity card or passport, or what this person knows, which is the password or PIN (personal identification number), however, these elements can be forgotten, stolen or forged. To circumvent these restrictions, another means of security has been developed that is not represented in the information that an individual possesses or knows, but is represented in the material information of that person, this new method of identifying individuals is called **Biometrics**.

This chapter includes two parts, the first is devoted to the generality of biometric systems, and the second will only deal with face detection and recognition. We begin by providing some general information about biometrics, such as: its definition, operation, different types of biometrics, application areas and characteristics, and measuring the performance of biometric systems. Then we move on to the second part of this chapter devoted to facial recognition.

## 1.2. Biometrics

Biometrics can be defined as “the automatic recognition of a person using distinctive traits”. Another definition of biometrics is “any automatically measurable, robust and distinctive physical characteristics or personal traits that can be used to identify an individual or to verify an individual's purported identity” [7]**.**

The need for this system whose function depends on accurate identification of the individual in the context of multiple applications has reinforced the importance of biometrics in modern society, in fact, it provides more security and convenience which generates huge economic advantages and closes major security holes such as password.

### 1.2.1. Biometric system

A biometric system is basically a system that acquires biometric data from an individual, extracts a set of characteristics from that data and then compares it to a set of data previously stored in a database so that it can finally perform an action or decide based on the result of this comparison [13].

#### 1.2.1.1. Modules of a biometric system

A typical biometric system can be represented by four main modules:

- **Capture module:** responsible for the acquisition of an individual's biometric data (this can be a camera, a fingerprint reader, a security camera...).
- **Feature extraction module:** which takes as input the biometric data acquired by the capture module and extracts only the relevant information in order to form a new representation of the data. Ideally, this new representation is supposed to be unique for each person and relatively invariant to intra-class variations.
- **Matching module:** it compares the extracted feature set with the model stored in the system database and determines the degree of similarity (or dissimilarity) between the two.
- **Decision module:** verifies a user's asserted identity or determines a person's identity based on the degree of similarity between the extracted features and the stored pattern(s).

### 1.2.1.2. Architecture of a biometric system

Biometric systems are increasingly used. In general, a system for recognizing people can be broken down into two phases, the enrollment phase (creation of the database) and the recognition phase.

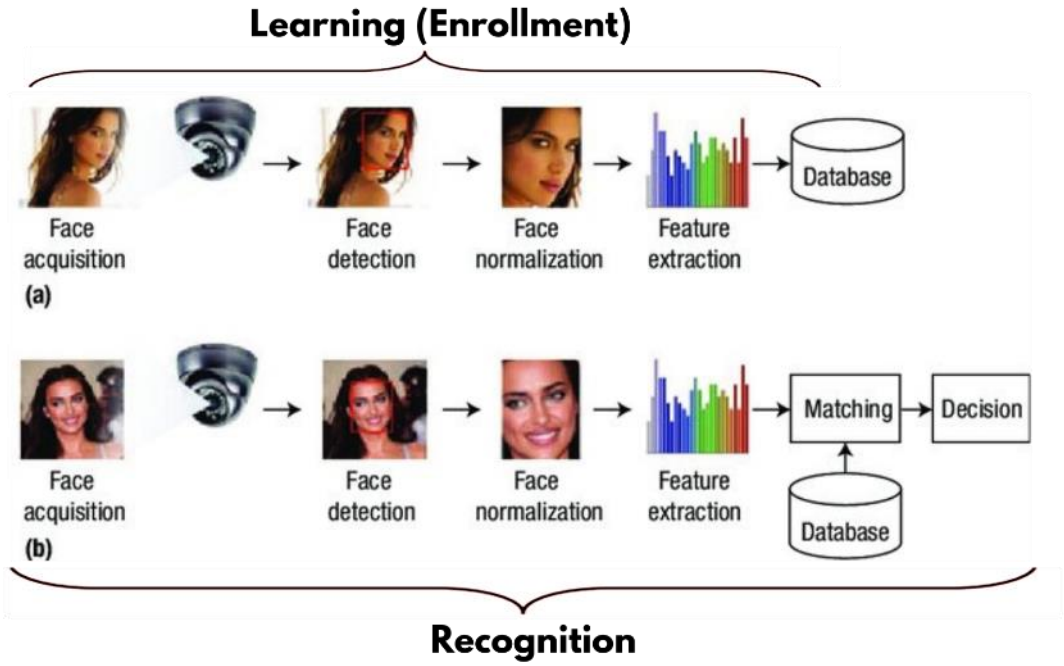



**Fig. 1-1:** Architecture of a biometric system [29].

#### 1.2.1.2.1. Learning (Enrollment)

In the event that someone comes to the system for an identification or authentication request, a biometric modality of the person in question (voice, face, iris, retina, fingerprint, etc.) is captured by an appropriate device, then digitized. The essential information is then detected and normalized then extracted from the acquired data to be modeled using mathematical functions. The model obtained will be saved in a database of system users, which will be used for their recognition.

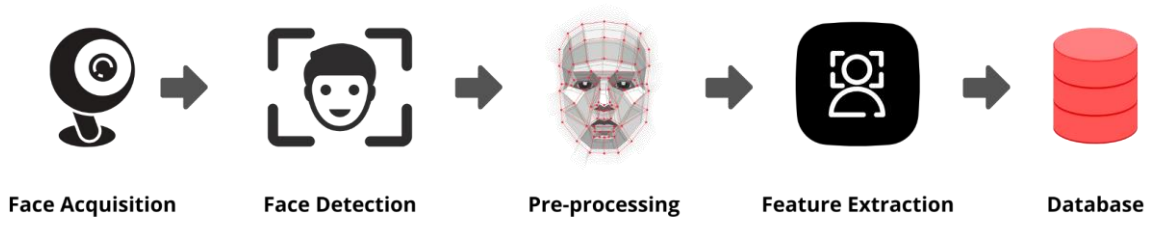



**Fig. 1-2:** Learning (Enrollment) process in a biometric system.

#### 1.2.1.2.2. Recognition

A biometric modality of the person who wishes to be recognized is captured by an acquisition device. The relevant parameters of the individual in question will then be extracted. The next step depends on the mode of recognition, identification or authentication. The face recognition systems can operate basically in two modes:

- **Verification or authentication of a facial image:**

One-to-one (1: 1): In this configuration, biometrics is used to verify a person's identity. For example, physical access to a secure space in a building can be ensured by fingerprint, or access to a bank account or an ATM can be guaranteed by iris recognition.

Biometric authentication requires comparing a previously registered biometric sample (Biometric Template) to another newly captured biometric sample (for example, one captured during a login) [3].

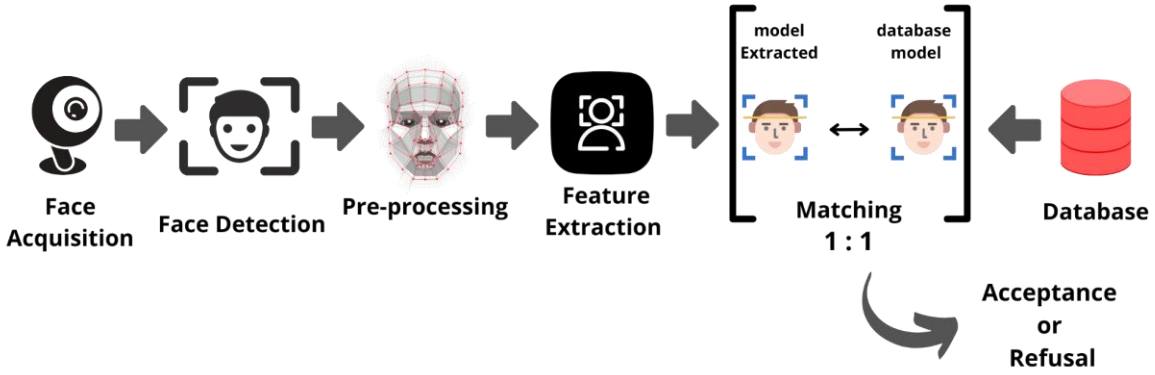



**Fig. 1-3:** Verification process in a biometric system.

- **Identification or facial recognition:**

One-to-many (1: N): Biometrics can be used to determine a person's identity, even without their consent. For example, the scanning of a crowd with a camera and the use of face recognition technology can help in determining the subject matter, in comparison with profiles stored in one or more reference databases [3].

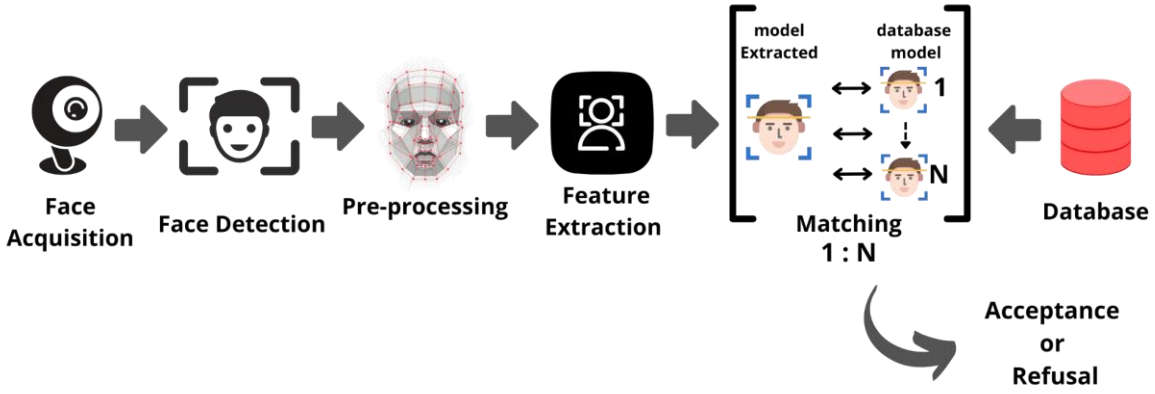



**Fig. 1-4:** Identification process in a biometric system.

### 1.2.2. Biometric Modalities

The multiplicity of human biometric personalities has led to the emergence of many authentication systems, each of which depends on a morphological, behavioral or biological character, and among these systems there are those that have proven their reliability and performance, and others are still developing, as the (Fig 1-5) shows.

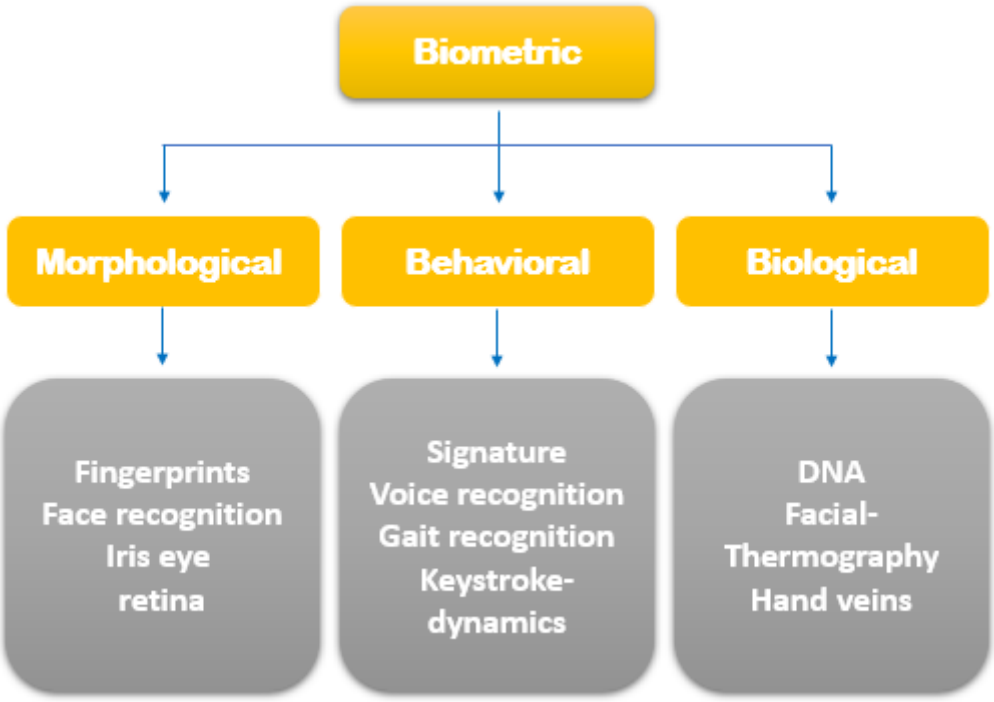



**Fig. 1-5**: Global scheme of biometrics.

### 1.2.2.1. Morphological biometrics

Involve the structure of your body. More physical traits like your eye, fingerprint, or the shape of your face can be mapped for use with security scanners.


**Fig. 1-6:** Morphological biometrics.

### 1.2.2.2. Behavioral biometrics

Use the behaviors of a person, which are characteristic, and are learned and acquired over time, in order to identify him. They can include: voice, signature, gait, keystroke.

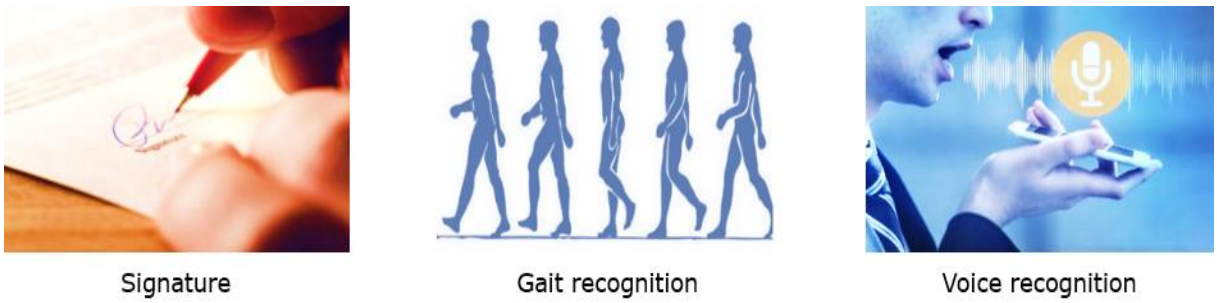



**Fig. 1-7**: Behavioral biometrics.

### 1.2.2.3. Biological biometrics

The use of traits at the genetic and molecular level. These may include features such as DNA or blood, which can be evaluated with a sample of body fluids.

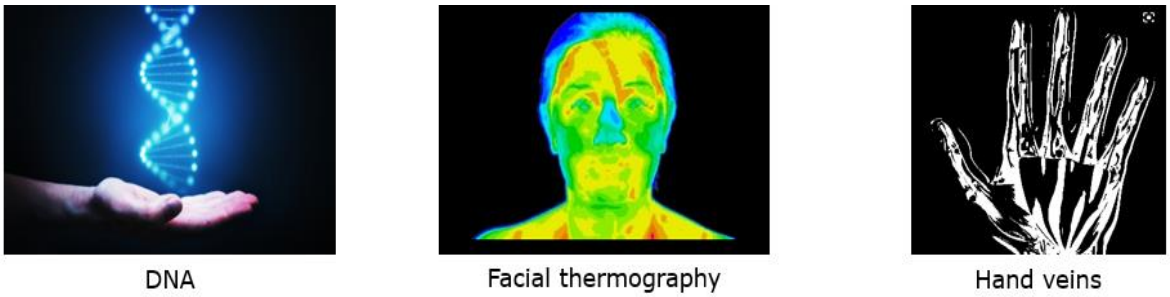



**Fig. 1-8**: Biological biometrics.

The main limitations related to biometrics are due to the ergonomics and acceptance of certain modalities. If identification based on behavioral biometrics or biological biometrics in general is not well perceived by the public, identification based on morphlogy particulary on fingerprints and facial modalities, is less intrusive. These methods have the advantage of being natural to human, while providing a sufficient level of safety for a large number of applications. In addition, the necessary camera hardware is currently being integrated into most embedded systems.

### 1.2.3. Assessment of Biometric Characteristics

Biometric characteristics are an alternative solution to the old means of identity verification. The advantage of these biometric characteristics is that they are universal, i.e. present in all the people to be identified. On the other hand, they are measurable and unique: no two people can have exactly the same characteristic. They are also permanent, which means that they vary little or not over time.

For collected characteristics to be qualified as modalities biometric, they must be [9]:

- **Universality:** which measures the degree to which the characteristic can be found in the majority of people;
- **Uniqueness:** which measures the degree to which the characteristic is unique among different people;
- **Permanence**: which measures the characteristic's resistance to change due to advancing age, illness and/or accidents;
- **Collectability**: which measures how easy and convenient it is to capture and measure the characteristic;
- **Acceptability**: which measures peoples' willingness to accept a biometric system based on that characteristic;
- **Circumvention**: which measures how easy it is to use fraudulent techniques in order to fool a biometric system based on that characteristic.
- **Performance**: which measures factors such as the speed and accuracy of the capturing of the characteristic.

and this (Table 1-1) below shows a performance comparison of the different biometric modalities mentioned above according to the following characteristics:

Universality (Univ), Uniqueness (Unic), Permanence (Pm), Collectability (Col), Acceptability (Acc), Circumvention (Cir), Performance (Pf).

| Modality | Univ | Unic | Pm | Col | Acc | Cir | Pf |
|---|---|---|---|---|---|---|---|
| DNA | Yes | Yes | Yes | Weak | Weak | No | ***** |
| Gait recognition | Yes | No | Yes | Weak | No | Weak | * |
| Voice recognition | Yes | Yes | Weak | Yes | Yes | No | *** |
| Iris eye | Yes | Yes | Yes | Yes | Weak | No | ***** |
| Face recognition | Yes | No | Weak | Yes | Yes | Weak | **** |
| Fingerprint | Yes | Yes | Yes | Yes | Medium | No | **** |
| Signature | Weak | NO | NO | Yes | Yes | No | ** |

**Table 1-1**: A comparison of the different modalities.

This table shows that no characteristic is therefore ideal and that they may be more or less suitable for particular applications. For example, DNA-based analysis is one of the most effective techniques for verifying an individual's identity or identifying them. However, it cannot be used for logical or physical access control for reasons of computation time, but also, because no one would be willing to give a little blood to do the verification. The choice of modality is thus made according to a compromise between the presence or absence of some of these properties according to the needs of each application. It should be noted that the choice of the biometric modality may also depend on the local culture of the users. In Asia, methods requiring physical contact such as fingerprints are rejected for hygienic reasons while non-contact methods are more widespread and accepted such as face recognition.

#### 1.2.3.1. The best biometric solution

Despite the existence of several biometric modalities, there is no perfect biometric system. Each biometric technology has acceptable or unsuitable advantages and disadvantages depending on the application. It is necessary to consider the environment in which they are used, the level of security sought, the ability to adapt to changes (being compatible with temporary or permanent modifications by a user), market logic and performance analysis.

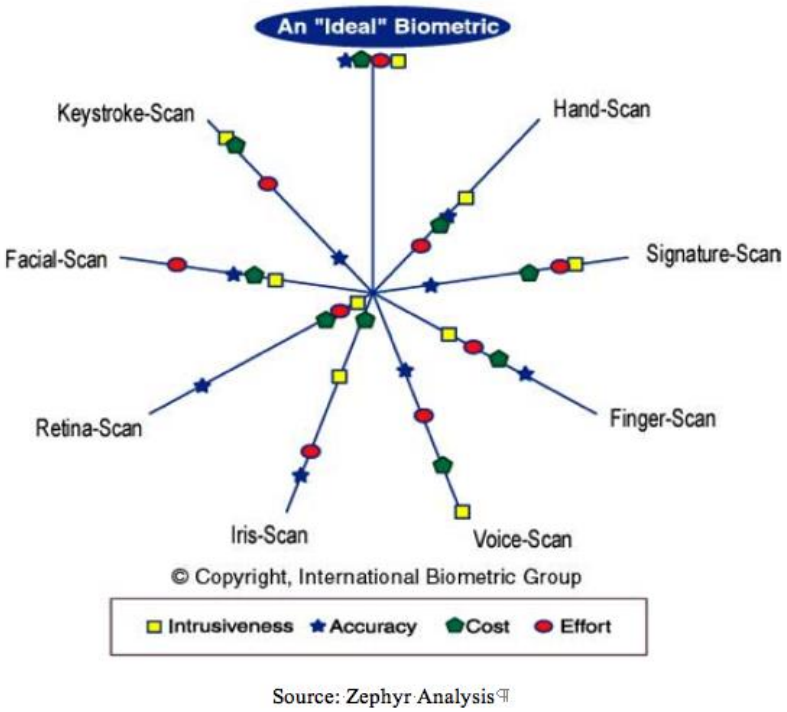
On the one hand, the International Biometrics Group IBG (International Biometric Group) in 2003 has carried out a comparison of the different biometric technologies called Zephyr Analysis. The results of this comparison are shown in (Fig 1-9) this comparison is based on four (04) main criteria [10]:

**Intrusiveness**: level of perception by the user of the test as intrusive

**Accuracy**: effectiveness of the method (ability to identify someone)

**Cost**: cost of the technology (readers, sensors, etc.)

**Effort**: effort required by the user




Source: Zephyr Analysis

**Fig. 1-9**: A comparison of the main biometric technologies by the company IBG (International Biometric Group)

### 1.2.4. Applications of biometrics

Today, the main applications are the production of identity documents, access control to secure sites, border control, access to networks, information systems and workstations, electronic payment, electronic signature and even data encryption. This list is not exhaustive, and new applications will certainly see the light of day. Biometric techniques are applied in several areas and their scope potentially covers all areas of security where it is necessary to know the identity of people. Applications can be divided into three main groups [6]:

1- **Commercial applications:** It includes applications such as computer network access, electronic data security, e-commerce, internet access, ATM (automated teller machine), credit card, physical access control, the mobile phone, the PDA (personal digital assistant), the management of the registers, the study of distances, etc …
2- **Government applications:** This category includes applications such as the national identification card, driving license, social security, passport control, etc.
3- **Legal applications:** It covers applications such as corpse identification, criminal investigation, terrorist identification, missing children, etc.


**Fig. 1-10**: Biometric apps.

We also use biometrics in:

- **Access control to premises:**
  - Computer rooms.
  - Sensitive sites (research service, nuclear site).
- **Communication equipment:**
  - Access terminals.
  - Mobile phones.
- **Information system:**
  - Launch of the operating system.
  - Network access.
  - Transaction (financial for banks, data between companies).
- **Miscellaneous machinery and equipment:**
  - ATM machine.
  - Sensitive location (shooting club, police).
  - Control of members in private clubs.
  - Control of attendance times.

- **State/Administration:**

  - Judicial file.
  - Social services (securing regulations).
  - Electronic voting system.

### 1.2.5. Evaluation of a biometric system

The performance of a biometric identification system can be measured mainly using three criteria: its accuracy, its efficiency (speed of execution) and the volume of data that must be stored for each person. We will focus in this section on the first aspect. As we have seen previously, identification and verification are different operating modes. They therefore require different precision measurements. The figure below shows a diagram FAR and FRR:

**1- False Acceptance Rate or FAR**

The false acceptance rate, or FAR, is the measure of the likelihood that the biometric security system will incorrectly accept an access attempt by an unauthorized user. In the binary system, this outcome is referred to as a false positive. That is, a false positive is a false acceptance.

**2- False Rejection Rate or FRR**

The false rejections rate, or FRR, is the measure of the likelihood that the biometric security system will incorrectly reject an access attempt by an authorized user. In the binary system, this outcome is referred to as a false negative. That is, a false negative is a false rejection.

**3- Equal Error Rate or EER**

The equal error rate is a biometric security system algorithm used to predetermine the threshold values for its FAR and FRR. When rates are the same, the common value is called EER. The lower the equal error rate value, the higher the accuracy of the biometric system.

- **The Threshold**

Is the value that we put to decide if we will consider the prediction of the model as correct. Any algorithm automatically sets a fixed threshold of 0.6, if not specified manually.

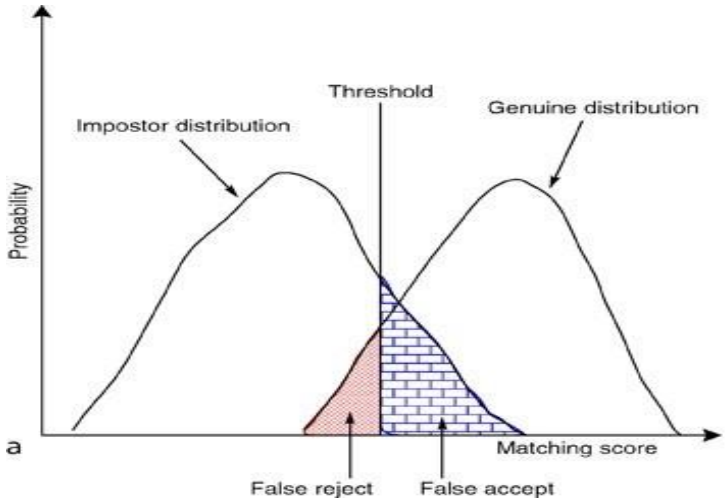



**Fig. 1-11**: FAR and FRR Diagram [1].

### 1.2.5.1. Degree of impact of FAR and FRR on security levels

FAR and FRR are in a state of equilibrium. If you lower the FAR, the FRR level will rise, and vice versa. The false acceptance rate is responsible for security, while the false rejection rate is related to convenience for the end-user, as we can see in (Fig 1-12). Ultimately, you will have to decide which is more important, usability or security.

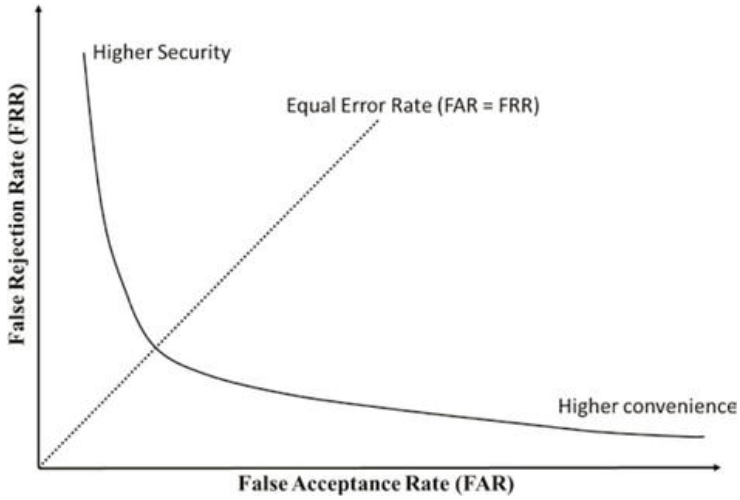



**Fig. 1-12**: The graph of FAR, FRR, and ERR in reciever operating ROC curve.

There are two ways to measure the biometric system performance, according to the mode (authentication or identification):

- **Verification assessment**

In the case of authentication mode, then the ROC (Receiver Operating Characteristic) curve is used. This curve draws the false rejection rate depending on the false acceptance rate.

The more this curve fits the mark shape the more the system is efficient with a high Recognition Rate (RR).

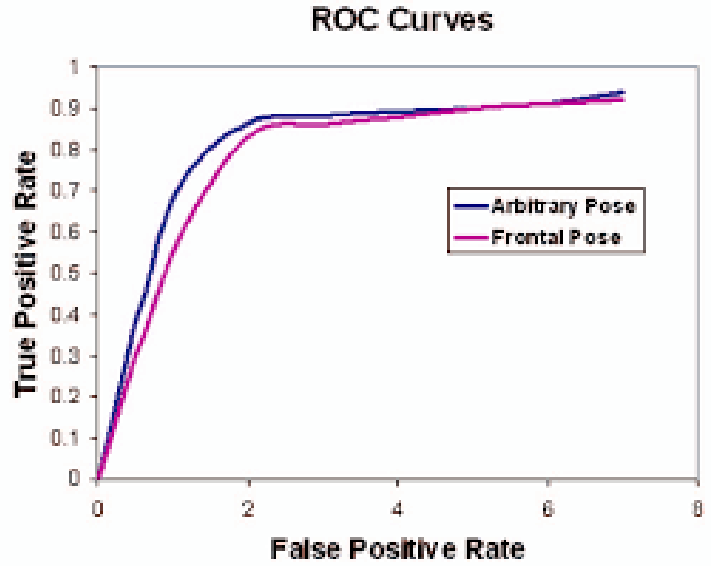



**Fig. 1-13**: ROC curves of our method for the frontal and arbitrary pose experiments [8].

- **Identification assessment**

In the case of identification mode, the CMC (Cumulative Match Characteristic) curve is used. The CMC curve provides the percentage of recognized individuals according to a variable called rank, as shown in (Fig 1-14). A system is said to recognize at the rank 1 when the nearest image is selected as the recognition result, and a system is said to recognize at the rank 2 when it selects, among two images, the one that best matches the input image. Subsequently, the more the rank is high the more the correspondent recognition rate is related to a low security level.

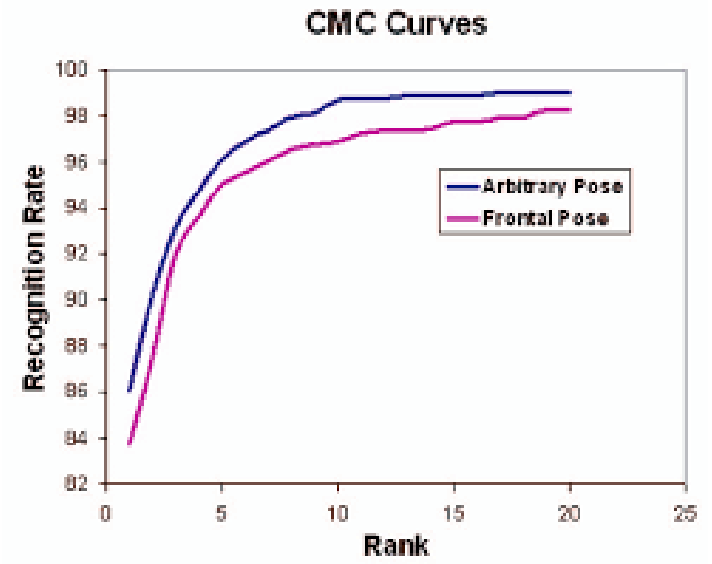



**Fig. 1-14**: CMC curves of our method for the frontal and arbitrary pose experiments [8].

### 1.2.6. The biometrics market

This report is essential reading for institutions deploying biometric technology, investors in biometric companies, or developers of biometric solutions. Revenue in the biometrics industry, including forensic and public sector applications, is growing rapidly. Much of the growth will be attributable to access control to information systems (computer/network) and e-commerce, although public sector applications continue to be an essential part of the industry

The turnover of emerging markets (access to information systems, electronic commerce and telephony, physical access and surveillance) is only expected to exceed the turnover of more mature sectors (criminal identification and identification of citizens).

The global biometrics technology market size was valued at USD 14.40 billion in 2017. The technology presents several advantages such as high level of security in the private, public, and commercial sectors. It records unique human characteristics, such as retina, fingerprints, DNA, and voice patterns, for authorization. Furthermore, the market is gaining traction owing to spiraling use of the technology across several verticals in different sectors for enhancing security.

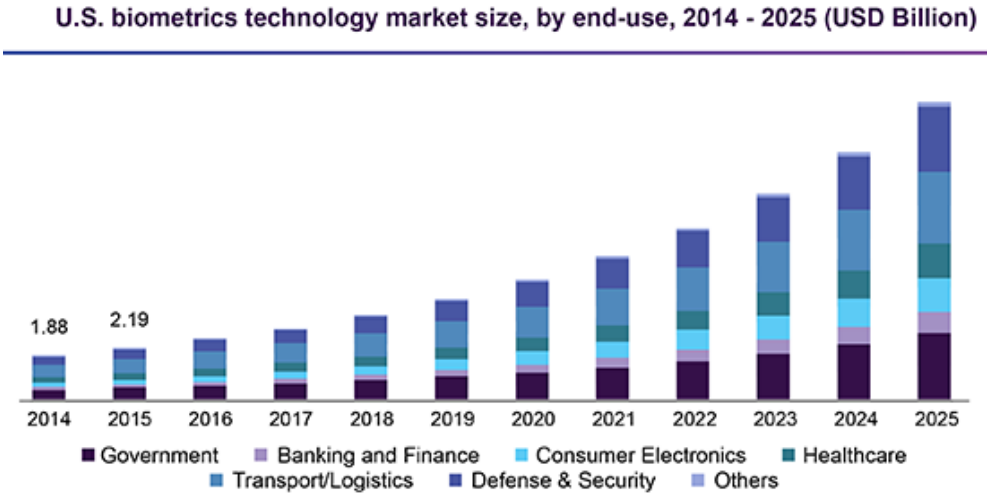



**Fig. 1-15**: Evolution of the international biometrics market [11].

#### 1.2.6.1. Market share by technology

On the basis of application, the market has been segmented into face, hand geometry, voice, signature, iris, automatic fingerprint identification system (AFIS), non-AFIS, and others. Facial recognition is among the most widely used biometric technologies. Contact biometric

technologies such as hand vein and palm fingerprint biometric recognition are expected to register significant growth over the forecast period.

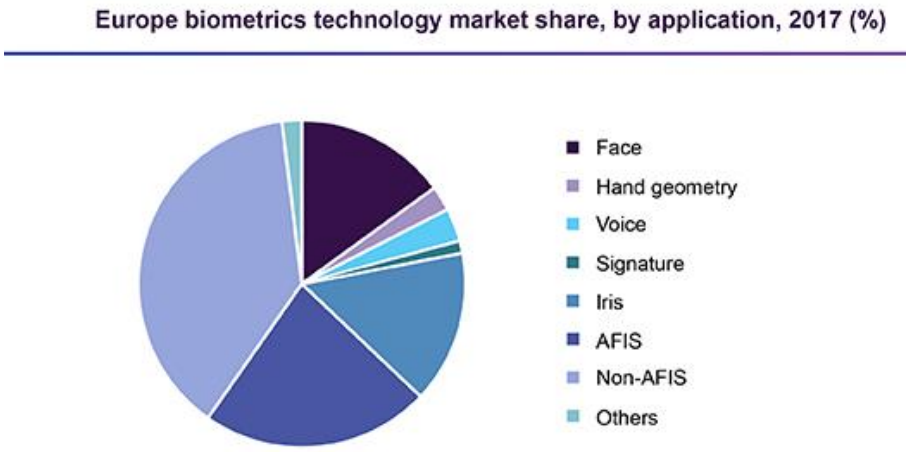



**Fig. 1-16**: Market shares of the different biometric methods [11].

## 1.3. Facial Recognition

Facial recognition is a task that humans perform naturally and effortlessly in their daily lives. The wide availability of powerful and inexpensive computers as well as embedded computing systems has generated enormous interest in the automatic processing of digital images and videos within many applications, including biometric identification, surveillance, human interaction -machine and multimedia data management.

Facial recognition, as one of the basic biometric technologies, has taken an increasingly important place in the field of research, this being due to rapid advances in technologies such as digital cameras, the Internet and mobile devices, all associated with ever-increasing security needs. Facial recognition has several advantages over other biometric technologies: it is natural, non-intrusive and easy to use.

Indeed, face recognition has several advantages, among which we can mention:

- **Short time:** This is one of the fastest biometric modalities. One can talk about real-time application because you have to go through the biometric system only once.
- **High security:** Let us take the example of a company that is checking the identities of people at the entry; such a biometric system allows not only employees to check presence at the time, but also any visitor can be added to the biometric system. Therefore, this system does not provide access to individuals not included in the system.
- **Automatic system**: This system works automatically without being controlled by a person.

- **Easy adaptation**: It can be easily used in a company. It only requires the installation of the capturing system (camera).
- **High success rate:** This system has achieved high recognition rates, especially with the emergence of three-dimensional technology, which makes it very difficult to cheat. Subsequently, this gives confidence to the system users.
- **Acceptance in public places:** It allows getting gigantic databases and, thus, improving the recognition performance.

Among the six biometric attributes (face, voice, eye, hand, signing, fingers) considered by [4], facial features mark a compatibility score in the MRTD system ("Machine-Readable Travel Documents") based on several evaluation factors, like enrollment, renewal data, required materials and user perception [31]. This score is shown in (Fig 1-17).

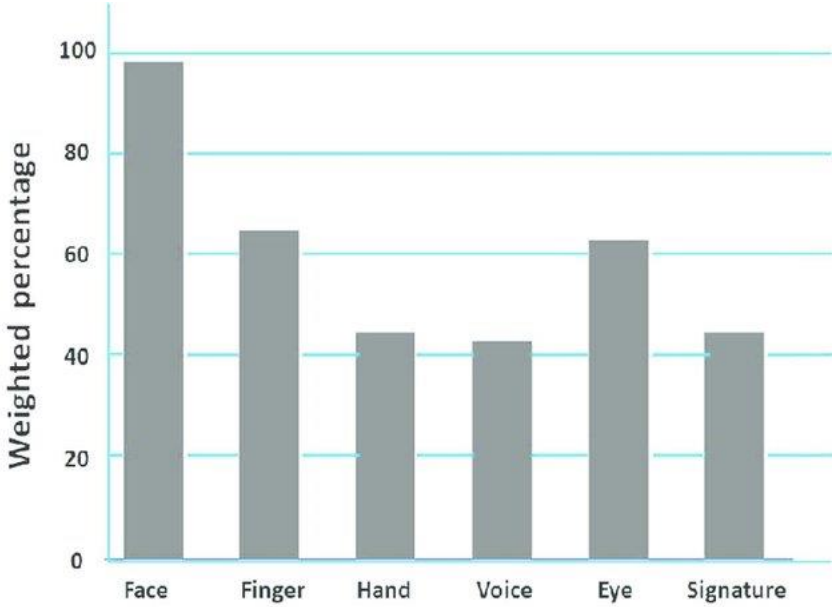



**Fig. 1-17**: Compatibility scores for different biometric technologies in an MRTD system.

### 1.3.1. Facial recognition system

Facial recognition systems are automated systems capable of identifying individuals according to the characteristics of their face such as the distance between the eyes, the bridges of the nose, the corners of the lips, the ears, the chin, etc. These characteristics are analyzed and then compared to an existing database in order to identify a person or verify their identity.

In a face recognition system, since its acquisition, the image follows a very specific process to arrive at determining or verifying the identity of the face wearer.

The basic operating principle of the facial recognition system (Fig 1-18) can be summarized in four stages: first the image of the individual is captured by the camera and then comes the second stage which is the pre-processing (that is, the face is detected and normalized). Then the third stage, which is recognition (Features are extracted, classified and matched). And finally, comes the decision-making stage.

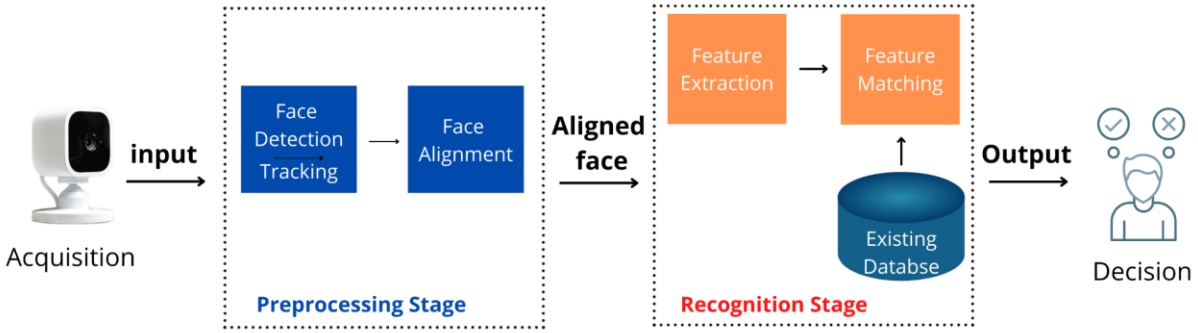



**Fig. 1-18**: Block diagram of a general face recognition system.

In the following we will detail each step of the facial recognition system:

### 1.3.1.1. Acquisition systems

The acquisition system is usually equipped with a sensor that allows users to obtain a specific function (for example, a microphone to record sound and a camera to capture a photo, etc.). A camera allows us to have a 2D image of the face from a 3D scene as shown

in (Fig 1-19):

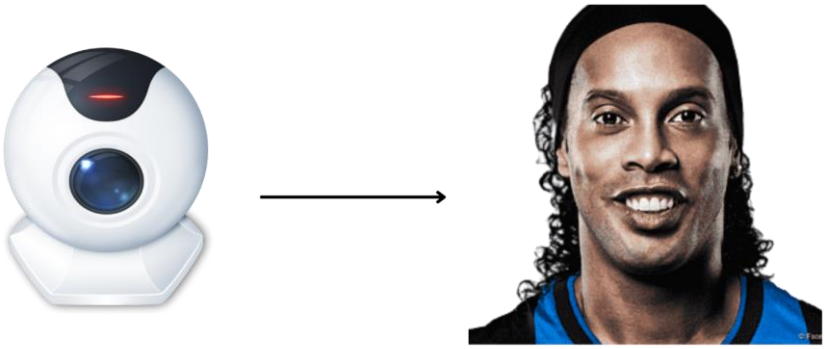


**Fig. 1-19**: Example of image acquisition.

#### 1.3.1.2. Pre-processing (Detection and Alignment)

This part of the preprocessing step is responsible for identifying and tracking the faces in the selected image or video file. Once this process is complete, we know for sure that there is a face in the given input, and it can be processed further. The problem with facial recognition is exacerbated because the faces in a particular photo or video do not follow any guidelines. The person may move a lot, or look away from the camera, which makes the problem of face detection more difficult. This is where face alignment comes in: it tells us where the facial lines are in the selected photo/video and what the facial features are as shown in (Fig 1-20):


**Fig. 1-20**: Example of face detection and alignment.

#### 1.3.1.3. Recognition (Extraction and Matching)

During this phase of the process, the individual features of the face, such as eyes, nose, chin, lips, etc., are extracted in the form that can be used by the algorithms in the next step. At this point, the computer has collected enough reliable data to uniquely distinguish a face.

In this step, the inputs received from the feature extraction are compared to the given database to derive the identity of the person. This phase is also known as classification, as the algorithm may need to categorize faces instead of identifying them individually.

#### 1.3.1.4. Decision making

The decision is part of the system in which we decide whether the individual belongs to all faces or not. In this phase, the identification system consists of finding the model corresponding to the face taken from those stored in the database, in this case what is its identity. Therefore, the resolution is the culmination of this process, it can be evaluated at the recognition rate (reliability), determined by the resolution rate of the decision.

### 1.3.2. Main difficulties of facial recognition

For the human brain, facial recognition is a high-level visual mission. Although people can detect objects and recognize them in a scene without too much trouble, creating an automatic system that performs these tasks is a serious challenge. This challenge is much greater than the conditions for obtaining very different images. The difference between the subjects is limited by the physical similarity between the individuals. On the other hand, the difference in this subject is greater. This can be attributed to several factors that we analyze below.

#### 1.3.2.1. Illumination change

When you want to take a photo, always take into consideration the lighting of the seine, since the appearance of a face in a photo is linked to the lumination, and the performance of the face recognition system (verification or identification) can be degraded when an image and taken under different lighting conditions to those used during registration, is sometimes more critical than the physical difference between individuals [26].

**Fig. 1-21**: Example of dimming lighting [2].

#### 1.3.2.2. Pose variation

The face recognition rate drops considerably when pose variations are present in the images. This difficulty has been demonstrated by evaluation tests developed on the FERET and FRVT databases [15]. Pose variation is considered a major problem for facial recognition systems. When the face is in profile in the image plane (orientation < 30°), it can be normalized by detecting at least two facial features (passing through the eyes). However, when the rotation is greater than 30°, geometric normalization is no longer possible. (see Fig 1-22).

**Fig. 1-22**: Examples of pose variations [2].

### 1.3.2.3. Facial expressions

Another factor that affects facial appearance is facial expression (see Fig 1-23). The deformation of the face which is due to facial expressions is localized mainly on the lower part of the face. The facial information located in the upper part of the face remains almost invariable. It is generally sufficient to carry out an identification. However, since the facial expression changes the appearance of the face, it necessarily leads to a decrease in the recognition rate. Face identification with facial expression is a difficult problem that is still relevant and remains unsolved. The temporal information provides significant additional knowledge that can be used to solve this problem [27].

**Fig. 1-23**: Examples of variation in expressions [2].

### 1.3.2.4. Presence or absence of structural components

The presence of structural components such as a beard, mustache, or glasses can greatly modify facial characteristics such as the shape, color, or size of the face. In addition, these components can hide basic facial features causing failure of the recognition system. For example, opaque glasses make it difficult to distinguish the shape and color of the eyes, and a mustache or beard changes the shape of the face.

### 1.3.2.5. Occultations

Faces can be partially obscured by other objects that cover the face. Indeed, in an image that contains a group of people, for example, the face can partially hide other faces.

## 1.4. Conclusion

This chapter consists of two parts. In the first, we dealt with the basic concepts of biometrics by recalling the different methods and applications of biometric systems. In the second section, we introduced the working method of the facial recognition system, which are the different techniques spread in the literature for facial recognition, we also presented the architecture of facial recognition systems and the various difficulties encountered by these systems such as brightness degradation, mode change, face masking, etc. In general, face recognition and biometrics have made great progress in the past decade, especially with the development of the latest technology

It is considered the use of artificial intelligence principles and algorithms based on deep learning. In the next chapter, we will discuss our work on processing facial recognition technology using various algorithms.

# Chapter 02

# State of the Art Face Recognition

## 2.1. Introduction

The extraction of facial features is an essential step in face recognition systems. Many face recognition methods have been proposed in recent years. It is an open research axis attracting researchers from different disciplines: psychology, pattern recognition, neural networks, artificial vision and computer graphics the features that are used for face recognition are eyes, mouth, face shape (outline), etc.

In this chapter we will give a classification of face recognition methods according to the type of input data of the system is adopted, we distinguish three main classes of methods (local, Holistic and hybrid). Next, we will determine the operating principle of the most representative methods, as well as a comparison between them and we will discuss the different models of deep learning at the end of the chapter.

## 2.2. Face Recognition Methods

The use of the face as mentioned earlier has its own drawbacks which include the limitations of changing lighting and therefore poses for the facial recognition system to have a high recognition rate requires a robust algorithm. However, not all of these method groups are able to handle limitations such as lighting or pose variations. Many systems are implemented to identify the human face in 2D or 3D images. These systems are categorized into three ways based on the detection and recognition method: (1) the local approach, (2) the comprehensive (microspace) approach, and (3) the mixed approach. The first method is classified according to certain features of the face, without considering the entire face. The second method uses the

whole face as input data and then projects in a small subspace or at the link level. The third method uses local and global features to improve the accuracy of face recognition [20]. shown as in (Fig 2-1).

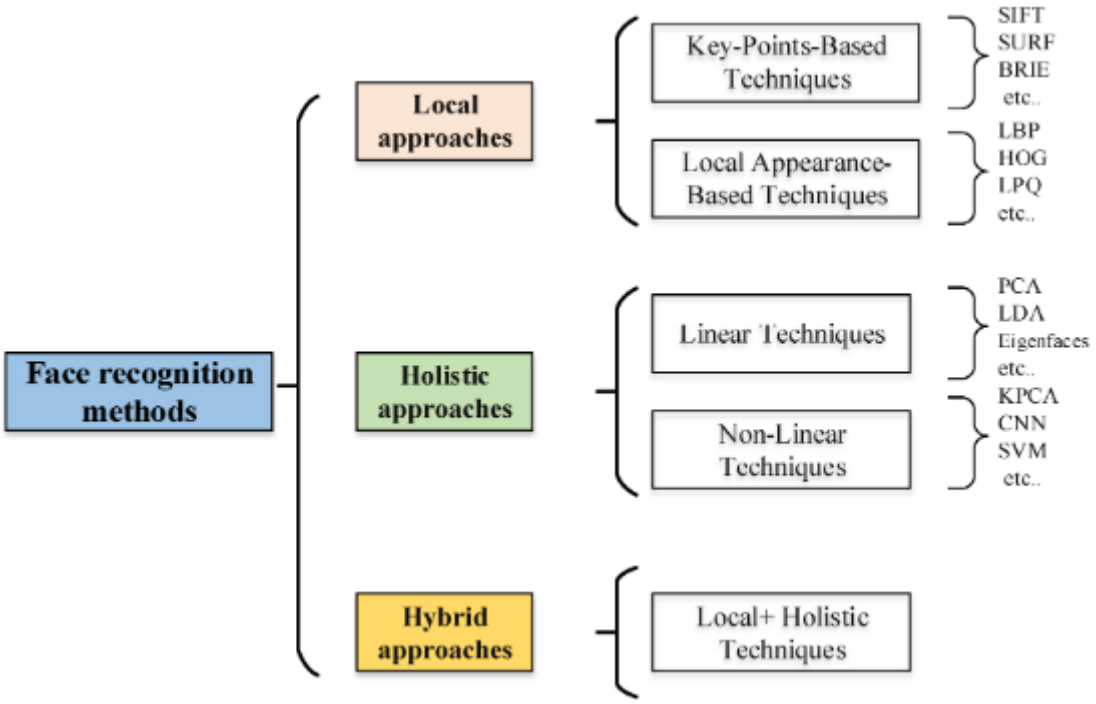



**Fig. 2-1**: Face recognition methods [20].

### 2.2.1. Local Approach

Local methods, based on models, use a priori knowledge that we have about the morphology of the face and are generally based on characteristic points of the face. These methods provide another approach to take nonlinearity into account by constructing a local feature space and using appropriate image filters, so that face distributions are less affected by various changes.

The advantage of these methods is that they take into account the particularity of the face as a natural shape to be recognized, in addition they use a reduced number of parameters and they are more robust to the problems posed by variations in illumination, pose and facial expression, but their difficulty arises when it comes to taking into consideration several views of the face as well as the lack of precision in the "extraction" phase of the points constitute their major drawback [5]. Some of these approaches include:

### 2.2.1.1. Local Appearance-Based Techniques

#### 2.2.1.1.1. Local Binary Pattern Histogram (LBPH)

Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number [34]. Perhaps the most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes caused for example, by illumination variations, especially when combined with HOG descriptor graphs, it greatly improves detection performance and the steps of this algorithm (LBPH) represent in:

1. **Parameters**: the LBPH uses 4 parameters:
   - **Radius:** the radius is used to build the circular local binary pattern and represents the radius around the central pixel. It is usually set to 1.
   - **Neighbors:** the number of sample points to build the circular local binary pattern. It is usually set to 8.
   - **Grid X:** the number of cells in the horizontal direction. It is usually set to 8.
   - **Grid Y:** the number of cells in the vertical direction. It is usually set to 8.
2. **Training the algorithm**

Algorithm training. To do this, you must use a dataset that contains images of the faces of the people you will recognize. Also assigning an identifier (may be a number or a person's name) to each image, so the algorithm will use this information to identify an input image and give an output.

3. **Applying the LBP operation**

The first computational step of the LBPH is to create an intermediate image that describes the original image in a better way, by highlighting the facial characteristics**,** by using the algorithm the concept of a sliding window, based on the radius and neighbor parameters. The (Fig 2-2) below shows this procedure:

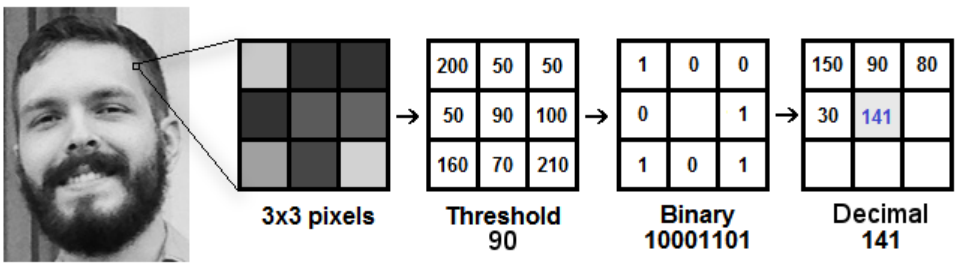



**Fig. 2-2:** An example of procedure LBP Operator [34].

**4. Extracting the histograms**

With the image created in the last step, the Grid X and Grid Y parameters can be used to split the image into multiple grids, as shown in (Fig 2-3):

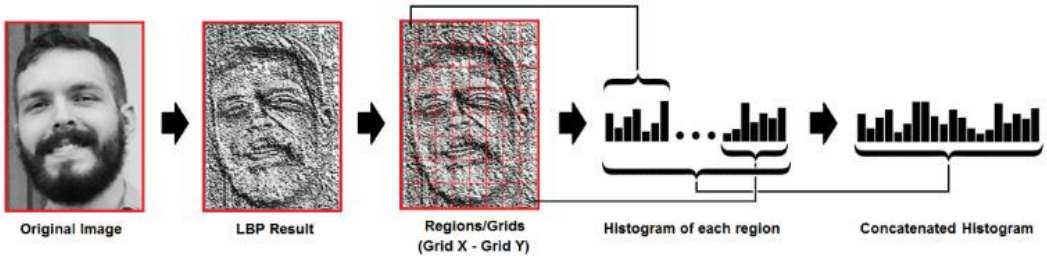



**Fig. 2-3:** Histogram Extraction [34].

**5. Performing the face recognition**

In this step, the algorithm has already been trained. Each histogram generated is used to represent each image from the training data set. It is sufficient to compare the histogram of the input images with these processed images and return the image to the nearest histogram.

**2.2.1.1.2. Histogram of Oriented Gradients (HOG)**

The H.O.G (Histogram of Oriented Gradients) is a feature descriptor used in computer vision It is a feature descriptor used in computer vision to manipulate images by calculating the gradient direction of its local area and, then Features are shaped by which emotional features can be effectively extracted. This got traction after Navneet Dalal and Bill Triggs published a paper called Histograms of Oriented Gradients for human detection in 2005. This was powerful and state of the art way of doing object detection before the deep learning era. HOGs are widely known for their use in pedestrian detection. At present, HOG is mainly combined with an SVM (Support Vector Machine) classifier, which is mainly used for image recognition.

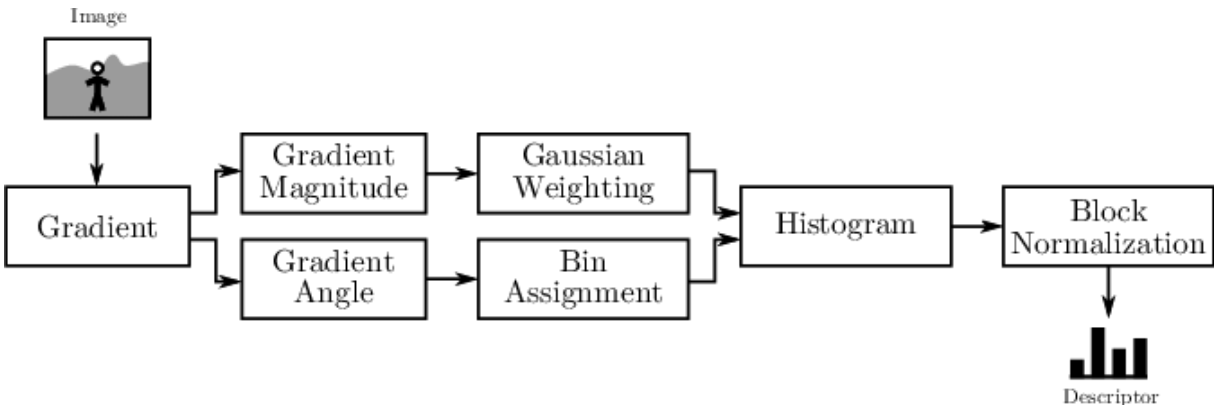



**Fig. 2-4**: Calculation steps of the HOG descriptor [24].

Actually, the HOG description operator can be obtained by the following four steps:

**1. Gradient calculation**

Firstly, two Sobel filters and expression images are convoluted to calculate the vertical and horizontal gradient maps. The vertical edge operator is $[-1,0,1]^T$, and the horizontal edge operator is $[-1,0,1]^T$. In particular, gamma and smooth normalization operations can be omitted.

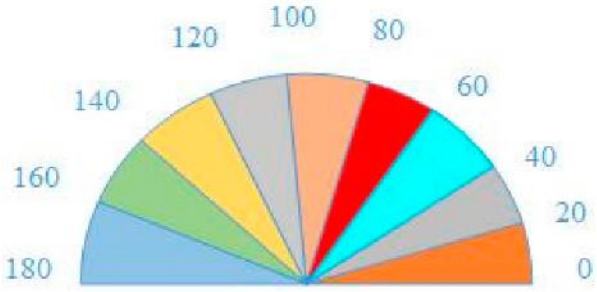



**Fig. 2-5**: The sobel operator [22].

**2. Calculation of amplitude and direction**

The amplitude and direction maps are calculated based on the vertical and horizontal gradient maps in step 1. Assuming *dx* and *dy* represent the gradient values in the horizontal and vertical maps, the amplitude and gradient of the pixel can be obtained according to equation.

$$\textbf{\textit{Magnitude}} = \sqrt{(\boldsymbol{dx})^2 + (\boldsymbol{dy})^2}$$

$$\textbf{\textit{Orientation}} = \boldsymbol{tan^{-1}}\left(\frac{\boldsymbol{dy}}{\boldsymbol{dx}}\right)$$

**3. Unit quantization**

The emotional face image is divided into several small unites. The value range of gradient direction is 0~180, which is equally divided into 9 intervals, each of which is 20 degree.

The gradient amplitude is used as the weight of projection (i.e., mapped to a certain direction interval).

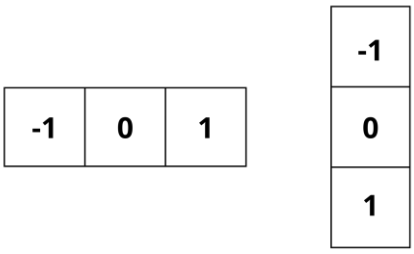



**Fig. 2-6**: Nine Bins.

**4. Block normalization**

In most cases, uneven illumination will affect the amplitude of the gradient, resulting in different value ranges, and local contrast normalization can improve the robustness due to illumination changes and improve performance. The normalization process can be obtained by equation.

$$v \rightarrow \sqrt{v \,/\, (\|v\|_1 + \varepsilon)}$$

### 2.2.1.2. Key-Points-Based Techniques

#### 2.2.1.2.1. Scale-Invariant Feature Transform (SIFT)

SIFT is an algorithm used to detect and describe the local features of an image. This algorithm is widely used to link two images by their local descriptors, which contain information to make a match between them. The main idea of the SIFT descriptor is to convert the image into a representation composed of points of interest. The algorithm is realized in four steps:

1. detection of the maximum and minimum points in the space-scale.
2. Location of characteristic points.
3. Assignment of orientation.
4. descriptor of the characteristic point.

### 2.2.2. Holistic Approach

They are based on well-known statistical analysis techniques. It is not necessary to mark certain characteristic points of the face (like the centers of the eyes, the nostrils, the center of the mouth, etc.) except for normalizing the images. In these methods, face images (which can be seen as arrays of pixel values) are processed.

globally and are usually transformed into vectors, which are easier to manipulate. The main advantage of global methods is that they are relatively quick to implement and that the basic calculations are of medium complexity. On the other hand, they are very sensitive to variations in lighting, pose and facial expression. This is easily understood since the slightest variation in the conditions of the ambient environment leads to inevitable changes in the values of the pixels which are processed directly. These methods mainly use an analysis of face subspaces. The use of subspace modeling techniques has advanced facial recognition technology significantly.

#### 2.2.2.1. Linear Techniques

The most popular linear techniques used for face recognition systems are Eigenfaces (Principal Component Analysis - PCA) technique, Fisherfaces (Linear Discriminative Analysis - LDA) technique, and Independent Component Analysis (ICA).

##### 2.2.2.1.1. Linear Discriminative Analysis (LDA)

Linear Discriminant Analysis or Normal Discriminant Analysis or Discriminant Function Analysis is a dimensionality reduction technique that is commonly used for supervised classification problems. It is used for modelling differences in groups i.e. separating two or more classes. LDA focuses primarily on projecting the features in higher dimension space to lower dimensions. This can be achieved in three steps [12].

1. **Firstly**, calculating the separability between classes (the distance between the mean of different classes). This is called the between-class variance.

$$S_b = \sum_{i=1}^{g} N_i(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

2. **Secondly**, calculating the distance between the mean and sample of each class. It is also called the within-class variance.

$$S_w = \sum_{i=1}^{g} (N_i - 1)S_i = \sum_{i=1}^{g} \sum_{j=1}^{N_i} \left(x_{i,j} - \bar{x}_i\right)\left(x_{i,j} - \bar{x}_i\right)^T$$

3. **Finally**, constructing the lower-dimensional space (maximizes the between-class variance and minimizes the within-class variance). P is considered as the lower-dimensional space projection, also called Fisher's criterion.

$$P_{LDA} = arg\,\max_P \frac{|P^T S_b P|}{|P^T S_w P|}$$

### 2.2.2.2. Nonlinear Techniques

#### 2.2.2.2.1. Support Vector Machine (SVM)

This approach is called linear classification and it is an approach where the objective is to find the best separator level, i.e., the ultrathin level which provides the highest margin distance between the two nearest point of the two classes (called functional margin). This approach generally ensures that the higher the margin the lower the generalization error of the classifier [30]. as can be seen from (Fig 2-7) below:

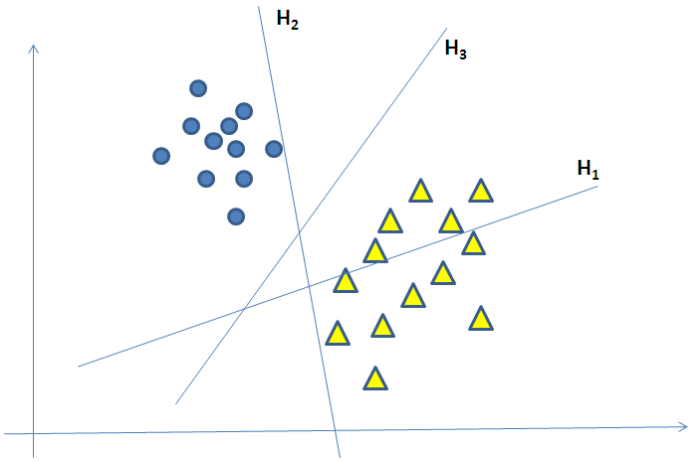



**Fig. 2-7**: Representation separation hyperplanes [30].

### 2.2.3. Hybrid Approach

Hybrid methods jointly use global and local features of faces by combining the advantages of both methods, in order to improve the performance of face recognition. Holistic features and local features are different properties hoping to combine the advantages of both methods to improve classification.

| Variation factors | Local characteristics | Holistic characteristics |
|---|---|---|
| *Illuminations* | *very sensitive* | *Sensitive* |
| *Expressions* | *Not sensitive* | *Sensitive* |
| *Pose* | *Sensitive* | *very sensitive* |
| *Occlusion* | *Not sensitive* | *very sensitive* |

**Table 2-1**: Table qualitatively summarizes the difference between the two types of characteristics.

## 2.3. Deep Learning Model

Deep Learning is a machine learning field concerned with utilising Artificial Neural Networks (ANNs) to solve computer vision tasks such as image classification, object detection, and pose estimation.

Various configurations of ANNs such as convolutional neural networks (CNN), recurrent neural networks (RNN), deep neural networks (DNN) can extract features from various data formats such as text, images, videos etc.

### 2.3.1. Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN), also known as CNNs, is a deep learning algorithm that recognizes and classifies features from an image for computer vision tasks. The general architecture of CNN is described in (Fig 2-8). It is a multi-layer neural network build to process visual inputs and perform various computer vision-related tasks like image classification, segmentation, and object detection for various applications for example medical imaging. and generally composed of the following layers.

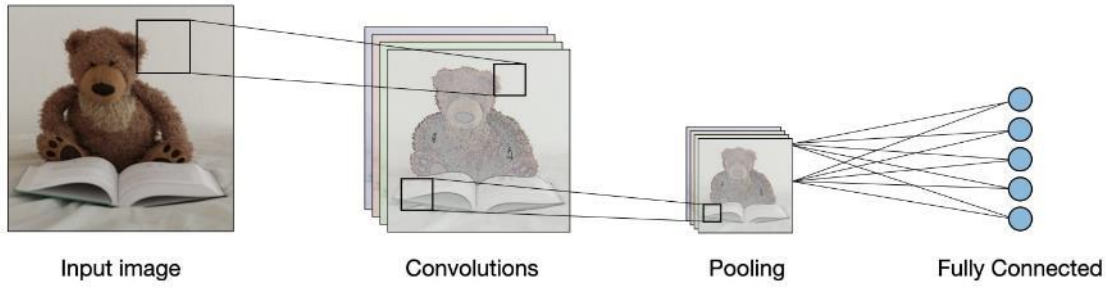



**Fig. 2-8**: Architecture of CNN [19].

The convolution layer and the pooling layer can be find-turned with respect to hyperparameters that are described in the next sections.

- **Convolution layer (CONV):** The convolution layer (CONV) uses filters that perform convolution operations as it is scanning the input I with respect to its dimensions. Its hyperparameters include the filter size F and stride S. The resulting output O is called Feature map or Activation map**.**
- **Pooling (POOL):** The pooling layer (POOL) is a downsampling operation, typically applied after a convolution layer, which does some spatial invariance. In particular, max and average pooling are special kinds of pooling where the maximum and average value is taken, respectively.

- **Fully Connected (FC):** The fully connected layer (FC) operates on a flattened input where each input is connected to all neurons. If present, FC layers are usually found towards the end of CNN architectures and can be used to optimize objectives such as class scores.

#### 2.3.1.1. LeNet-5

This CNN architecture comprises of 7 layers (Fig 2-9), and takes an input image of size 32×32 pixel. it was used to recognize the hand-written digits on a check. The primary units in each layer of convolution are an activation function called sigmoid and a pooling layer. In each layer of convolution, there are a filter of size 5×5. These layers conduct mapping operations on the inputs of two-dimensional feature vectors. The 1st convolution layer contains six output, but the 2nd comprises of 16. Each pooling filter of size 2×2 with stride size of 2 reduces the dimension.

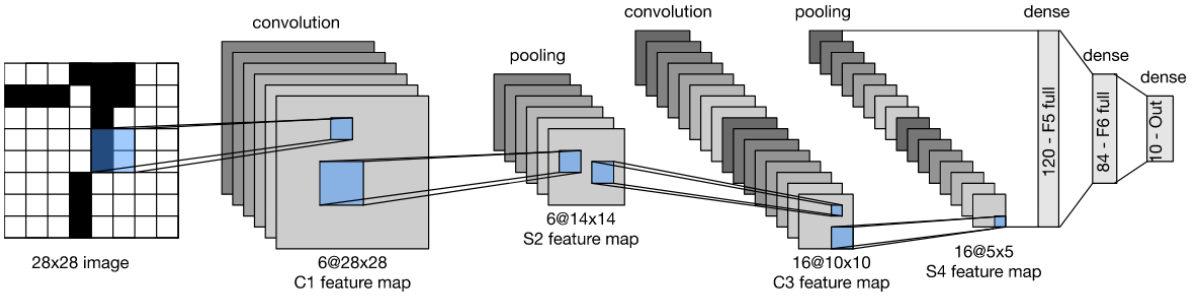



**Fig. 2-9**: Architecture of LeNet-5 [25].

#### 2.3.1.2. AlexNet

The Alexnet has eight layers with learnable parameters. The model consists of five layers with a combination of max pooling followed by 3 fully connected layers and they use Relu activation in each of these layers except the output layer. Using Relu as an activation function speeds up the training process by about six times, the model is trained on the imagenet dataset. The imagenet dataset contains approximately 14 million images across a thousand categories.

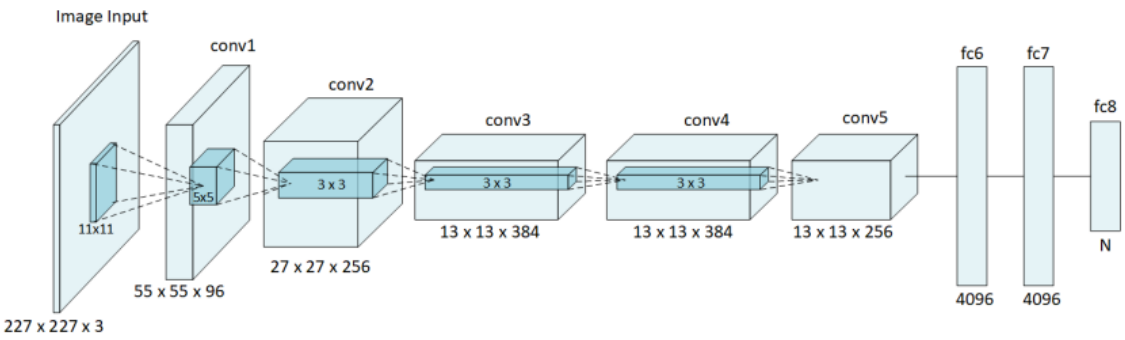



**Fig. 2-10**: Architecture of AlexNet [32].

### 2.3.1.3. GoogLeNet

The GoogLeNet architecture consists of 22 layers (27 layers including pooling layers), and part of these layers are a total of 9 inception modules (The Inception module is a neural network architecture that leverages feature detection at different scales). The GoogLeNet architecture solved most of the problems that large networks faced, mainly through the Inception module's utilisation. as shown in (Fig 2-11).

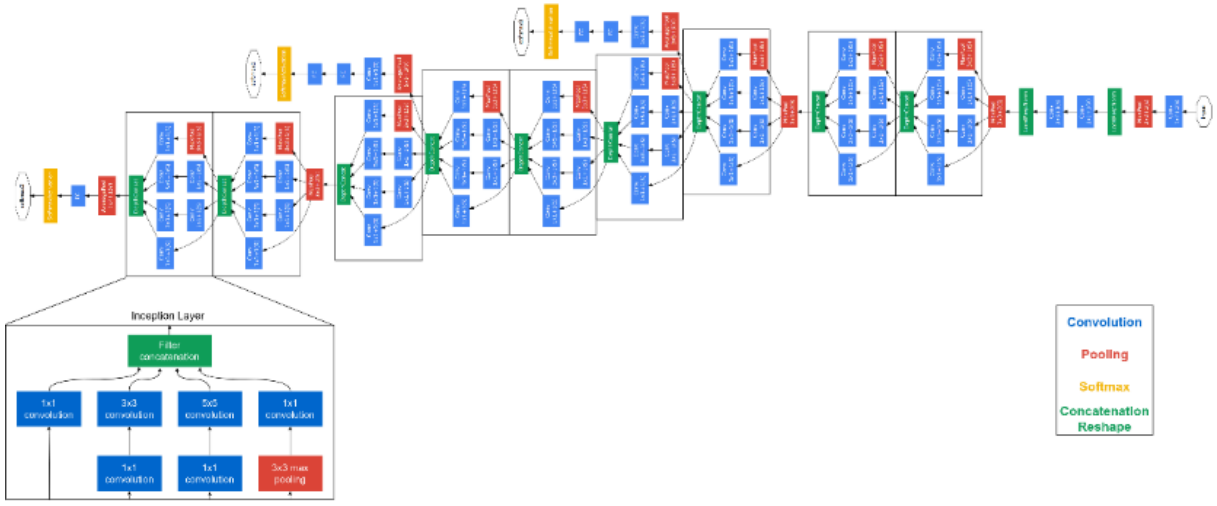



**Fig. 2-11**: Architecture of GoogLeNet.

### 2.3.2. FaceNet

FaceNet is a neural network for face recognition, verification, and aggregation. It is a 22-layer deep neural network that directly trains its output to be a merging of 128 dimensions.

They proposed an approach in which it generates a high-quality face mapping from the images using deep learning architectures such as ZF-Net and Inception. Then it used a method called triplet loss as a loss function to train this architecture.

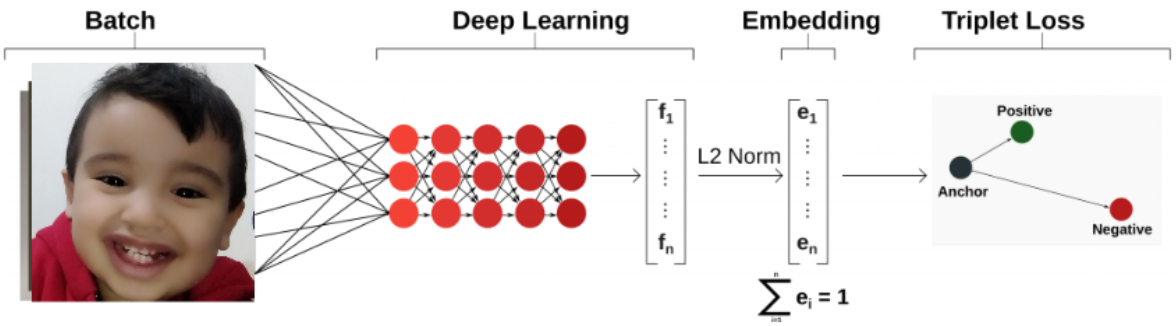



**Fig. 2-12**: Architecture of FaceNet.

### 2.3.3. Dlib

Dlib is a modern C++ toolkit containing machine learning algorithms and tools for creating complex software in C++ to solve real world problems. It is used in both industry and academia in a wide range of domains including robotics, embedded devices, mobile phones, and large high-performance computing environments.

Dlib is mainly inspired from a ResNet-34 model. Davis E. King modified the regular ResNet structure and dropped some layers and re-build a neural network consisting of 29 convolution layers. It expexts 150x150x3 sized inputs and represent face images as 128 dimensional vectors.

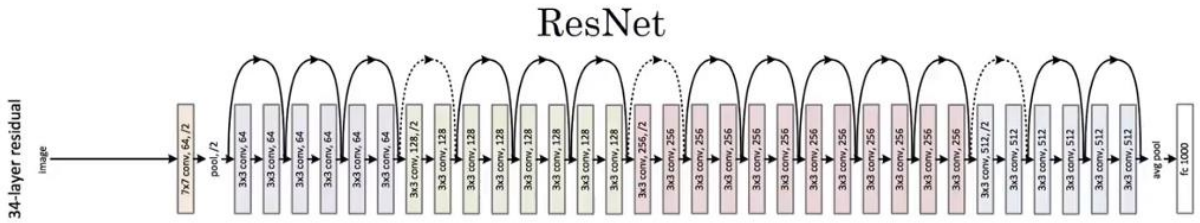



**Fig. 2-13**: Architecture of ResNet.

## 2.4. Conclusion

In this chapter, we discussed some of the existing methods for recognizing faces in its three types: Local Method, Holistic and Hybrid, then we touched on the best and most accurate part of this field, which is deep learning, we shared some of its models such as Convolution Neural Networks, with its deferent families, LeNet-5, AlexNet and GoogLeNet. And then we touched on the two models FaceNet and Dlib, which are indispensable in terms of speed, ease, and high accuracy, and they achieved the latest results in many facial recognition data sets in recent year.

# Chapter 03

## Facial recognition using Deep Learning

### 3.1. Introduction

Facial recognition is one of the biometrics used for identification. It has been for a long time a very interesting area that has attracted the interest of several researchers for being non-intrusive, very popular and not expensive.

Face recognition was around before the advent of deep learning. Earlier, classical feature descriptors and linear classifiers were a really good solution for face detection. Currently, for face detection, deep learning models are probably the best, since deep learning is a very data-intensive task and we may not always have such a huge amount of data to work on in the case of face recognition, so with the advances, recognition has become on the faces is more practical and feasible.

A face recognition system can be divided into four stages, namely face detection, face alignment, face encoding and face recognition, as shown in (Fig 3-1).

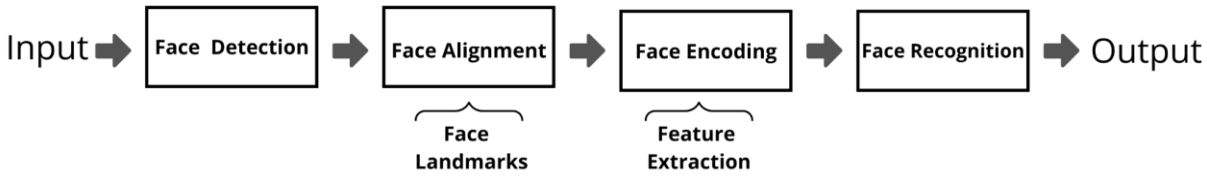



**Fig. 3-1**: Face recognition process using deep learning.

In this chapter, we discussed the mechanism and processing of facial recognition by the Dlib model, which is based on deep learning, which dealt with four stages, the first stage is face detection, comes after real-time video capture of course, which is the basic stage in this processing, we used as Histogram of Oriented Gradients, or HOG. Then the face alignment stage, in which the facial Landmarks 68 were extracted. Then the pre-processing stage, by means of a deep neural network, 128 measurements are extracted for each face, to be encoded. In the final stage, the entered face is compared with the faces in the data set, by calculating the Euclidean distance, which determines the closest distance between the faces.

## 3.2. Architecture of a facial recognition system

A well-known method called HOG (Histogram of Oriented Gradients) is used to find faces. After face detection, the face image is cropped and further preprocessed to remove the effects of insufficient light, face tilt, and skew. The cropped image is preprocessed using a facial feature estimation algorithm. The algorithm locates 68 landmarks on the cropped image and uses a simple affine transformation to preprocess the image using rotation, shearing, and scaling to optimally center the cropped image's eyes and mouth. Feeding the preprocessed image into a

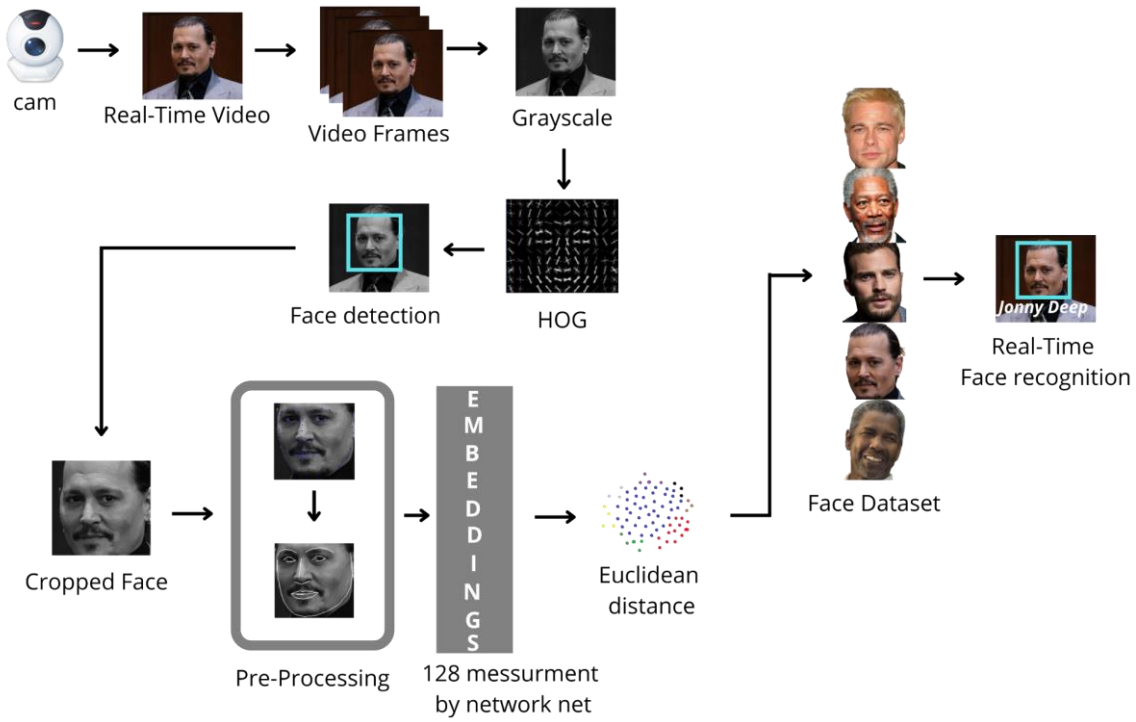



**Fig. 3-2**: Process model in real time.

pretrained network extracts features from the image and generates 128 embeddings, which are measurements of faces. The features of the preprocessed images are generated using a neural network that generates a feature vector with 128 embeddings. The final step is to classify the

images by measuring the closest match among the 128 embeddings by comparing them to database images. Then calculate the Euclidean distance and choose the smallest distance corresponding to the desired face. The entire architecture and its detailed framework are shown in (Fig 3-2).

### 3.2.1. Face detection

#### 3.2.1.1. Face detection by HOG - (Histogram of Oriented Gradients)

Face separation is the first and most important task after capturing real-time face images to remove unwanted and redundant information, such as background, from the images. We will use a method called Histogram of Oriented Gradients (HOG) to detect face images. HOG is basically a feature descriptor performed for image processing and computer vision techniques. After capturing an image of the face in frames from live video, convert this image to grayscale, then HOG focuses on the structure of the object to extracts the information of the edges magnitude as well as edges orientation. and then the facial part in the input image is extracted. Finally, the face image is cropped, as shown in (Fig 3-3).

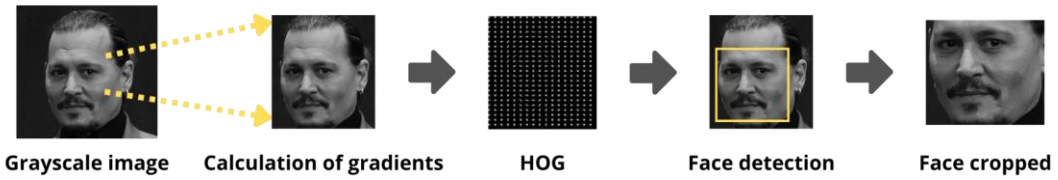



**Fig. 3-3**: Face detection using HOG.

To find the faces in an image, it is enough to make the image gray because we do not need color data to find the faces, then each pixel in the image is analyzed with the pixels surrounding it, and knowing how dark the current pixel is compared to the surrounding pixel directly to draw an arrow in the darker direction. After generalizing it to the whole image, each pixel will be replaced by an arrow. These arrows are called **Gradients**, showing the flow from light to dark, this is what the (Fig 3-4) shows.

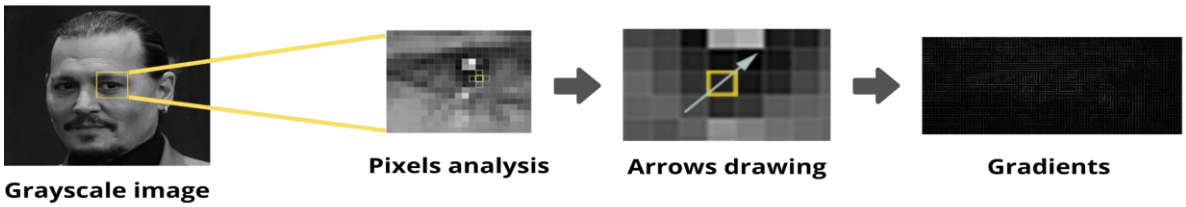



**Fig. 3-4**: Steps of making gradients.

This calculation is repeated for the entire grayscale image, and we'll end up with an image of gradients. The next step is to calculate the strongest gradients in 16 x 16 windows of pixels and replace the gradients in that window with the strongest gradient. This will result in a very simple representation that captures the basic structure of the face in a simple way.

To locate the face in the input image captured in the real-time video, we only located that part of the image which looks remarkably like a known HOG pattern and crop that part of the input image, as a result we get the face image cropped, as shown in (Fig 3-5).

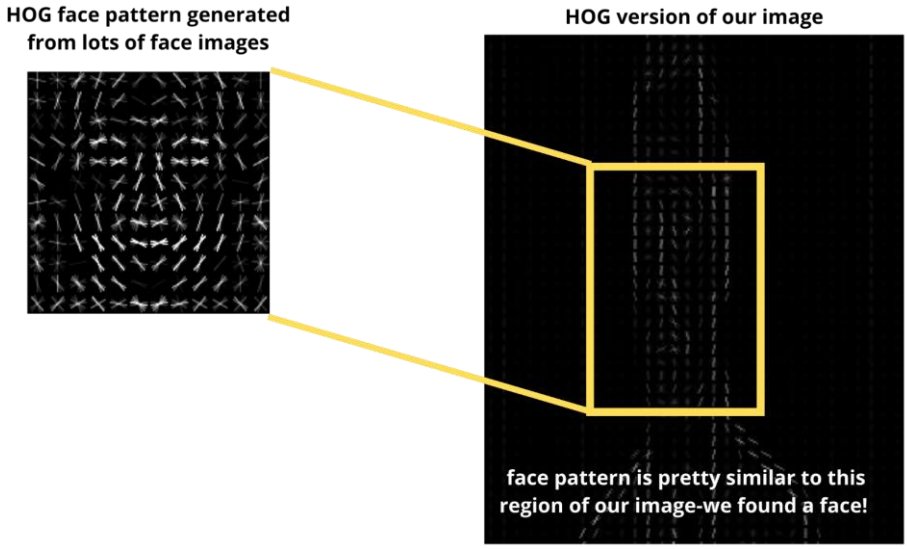



**Fig. 3-5**: Extracting a new face pattern that is very similar to the general face pattern.

### 3.2.2. Face Alignment

The process continues with posing and projecting faces, in order to deal with faces that are in different directions or angles. There are many ways to do this, but we'll use the approach that Vahid Kazemi and Josephine Sullivan pioneered in 2014, using an algorithm face landmark estimation, that a machine learning algorithm that able to detect 68 specific human face landmarks on a face image [23], as shown in (Fig 3-6), which consists of two steps:

- Localize the face in the image.
- Detect the key facial structures on the face ROI.

The basic idea is to extract the 68 specific points (called **Landmarks)** located on each face these landmarks locate the eyes, nose, chin, lips, and eyebrows etc., Then we will train a machine learning algorithm to be able to find these 68 specific points on any face.

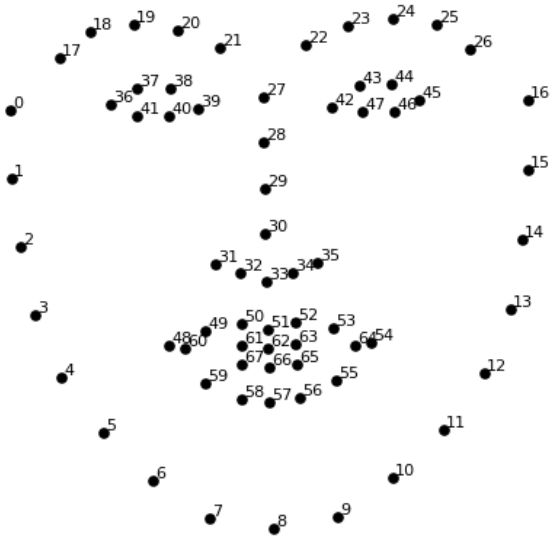


**Fig. 3-6**: The 68 landmarks detected by dlib library. This image was created by Brandon Amos of CMU who works on OpenFace.

As the face is captured in real-time, the image captured can have faces turned in different direction. To deal with such situations, we wrapped each picture so that our system can locate the eyes and lips in a sample place [18], as shown in (Fig 3-7).

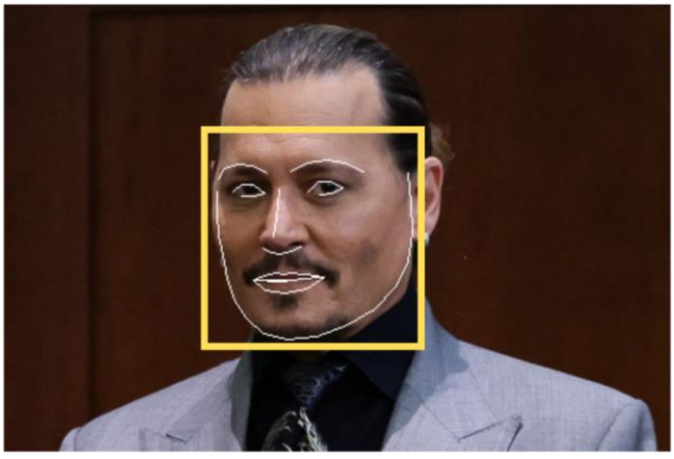


**Fig. 3-7**: The locating of 68 face landmarks.

Now that the eyes and mouth are known, the image is simply modified by rotating, resizing, and cropping while keeping the lines parallel, so that the eyes and mouth are centered as best they can. These transformations are called **Affine Transformations**, while avoiding any 3D

convolutions as this can distort the image, in the latter the eyes and mouth are centered at roughly the same position in the image regardless of how the face is transformed.

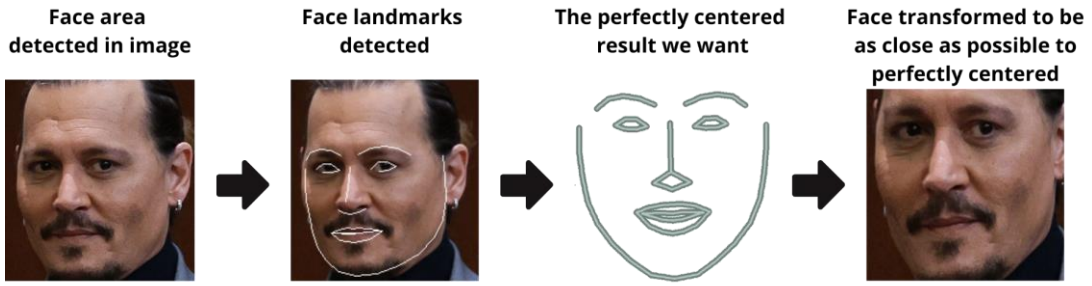



**Fig. 3-8**: Implementation of the affine transformation.

### 3.2.3. Face Encoding

The next important step is to extract the features from the exactly centered image. The best way to get the unique features of any facial image is to measure the face. The dimensions of each face are different. This task can be difficult to achieve if performed with the traditional method of feature extraction.

We know that deep learning needs to train a network by inputting a single image and outputting a classification/label for that image, in our work it's not the same thing, instead of trying to output a single label (or even the coordinates/bounding box of objects in an image), we output a feature vector with Real value. That's what it's called **Deep Metric Learning**.

For the Dlib facial recognition network, the output feature vector is 128-d (i.e., a list of 128 real-valued numbers) that is used to quantify the face.

Training the network is done using triplets at the same time:

- ❖ Loading a training face photo (anchor photo) of a known person
- ❖ Loading another (positive) photo of the same person
- ❖ Loading a different (negative) picture of another person

After the measurements of each of the three images are generated and considered by the algorithm, this algorithm slightly tweaks the weights of our neural network according to the result, making the image anchor closer to the positive image, compared to any other image called the negative image. After a lot of repetitive training, the neural network learns to reliably generate 128 measurements for each person. It is important to choose triplets for 128 measurements.

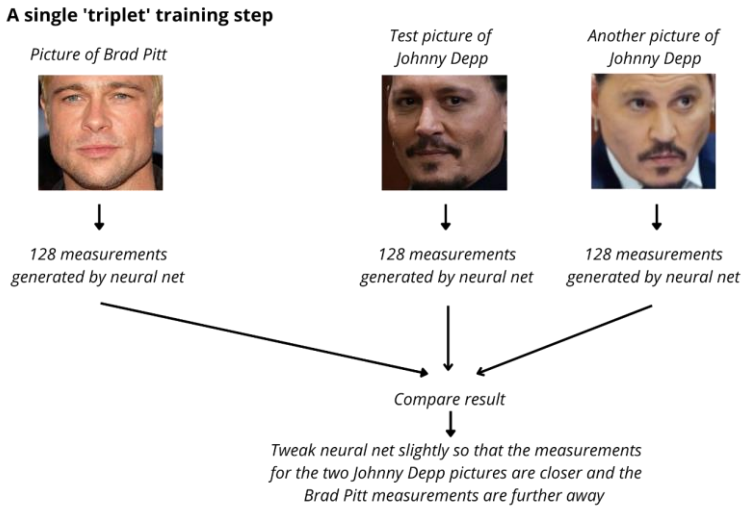



**Fig. 3-9**: Generation of 128-dimensional data from triplet.

After repeating this step millions of times for millions of images of thousands of different people, the neural network learns to reliably generate 128 measurements for each person. Machine learning experts call these measurements of every individual face **Embedding**. Training on face embedding using a large set of images called **Dataset** will improve the accuracy and decrease the error rate eventually. This process of training a convolutional neural network to output face embeddings requires a lot of data and computer power. But once the network has been trained, it can generate measurements for any face, even ones it has never seen before! So, this step only needs to be done once. After running images of faces through this pre-trained grid to get 128 measurements for each face, as the Fig (3-10) shows:

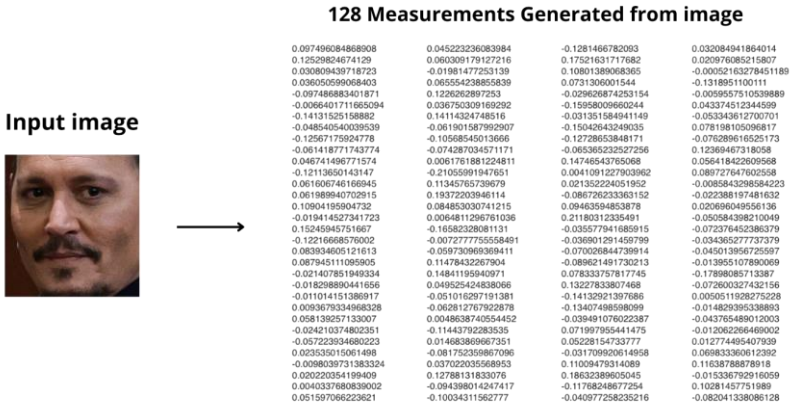



**Fig. 3-10**: Example of 128 measurements extracted from human face.

### 3.2.4. Face recognition

The process continues with the most important step of finding people's names by comparing 128 measures of an unknown face, with all measures of faces in the data set. And based on the

result to classify a particular face as stranger or used. The face_recognition library uses a default value equal 0.6 called **Threshold**, where after comparing faces a value of 0 is returned for a face's mismatch, and a value of 1 for a face's compatibility.

The crux of most data science problems is to define the distance or similarity function between the given observations. The similarity measure would be used to measure how close or how far the observations are in a given n-dimensional space. Many distance functions are available to use for a given problem. We will suffice with the **Euclidean Distance**, to get the person's name. This is what the (Fig 3-11) below shows:

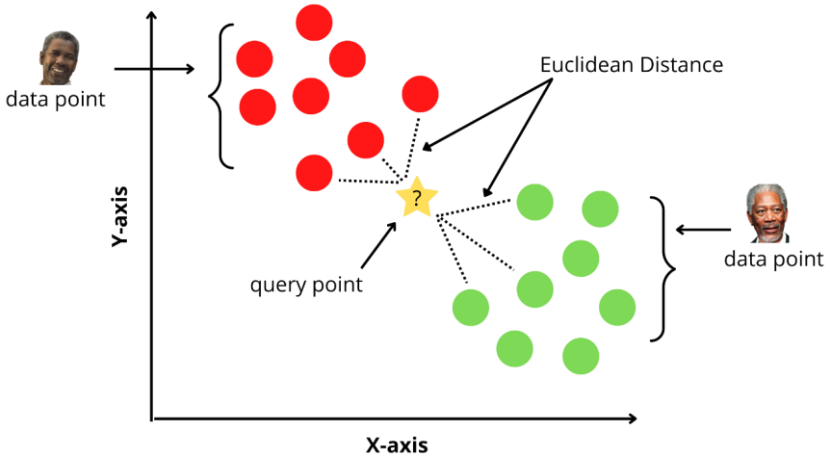



**Fig. 3-11**: The final step in get names.

**Euclidean distance (p = 2)**: This is the most widely used distance measure, and is limited to real-valued vectors. Using the formula below, it measures a straight line between the query point and the other point being measured [28].

$$E_d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}$$

## 3.3. Conclusion

Dlib achieved state-of-the-art results in the many benchmark face recognition dataset such as Labeled Faces in the Wild (LFW). Dlib is a versatile and well-diffused facial recognition library, with perhaps an ideal balance of resource usage, accuracy and latency, suited for real-time face recognition in mobile app development. It's becoming a common and possibly even essential library in the facial recognition landscape, and, even in the face of more recent contenders, is a strong candidate for your computer vision and facial recognition or detection framework.

# Chapter 04

# System Implementation and Testing

## 4.1. Implementation

### 4.1.1. Training dataset

A Dlib Face Recognition Network model with 29 convolutional layers, an optimized version of the well-used ResNet-34 network. This model was trained on Three Million faces across various datasets, including Face Scrub, Oxford's VGG set. In other words, it learns how to find face representations with 3M samples. Then, he tested the built for **Labeled Faces in the Wild (LFW)** dataset which is accepted as a baseline for face recognition researches. He got 99.38% accuracy. On the other hand, human beings hardly have 97.53% score on same dataset. This means that Dlib face recognition model can compete with the other state-of-the-art face recognition models and human beings as well. To test our Dlib models, we decided to use a real

**Fig. 4-1**: Sample of the prepared dataset.

student dataset for face verification and recognition. This data set contains 35 different images (Fig 4-1). The images are collected headdresses, and have been categorized as the students' pictures to which the head covering belongs. Unlike a typical facial database, where images of people are taken in a sterile environment, often with a solid background, images come from real-world environments.

### 4.1.2. Performance of AI model

#### 4.1.2.1. Confusion matrix

A confusion matrix is a way of assessing the performance of a classification model. It is a comparison between the ground truth (actual values) and the predicted values emitted by the model for the target variable [33].

A confusion matrix is useful in the supervised learning category of machine learning using a labelled data set. As shown below, it is represented by a table. This is a sample confusion matrix for a binary classifier (i.e. 0-Negative or 1-Positive) [33].

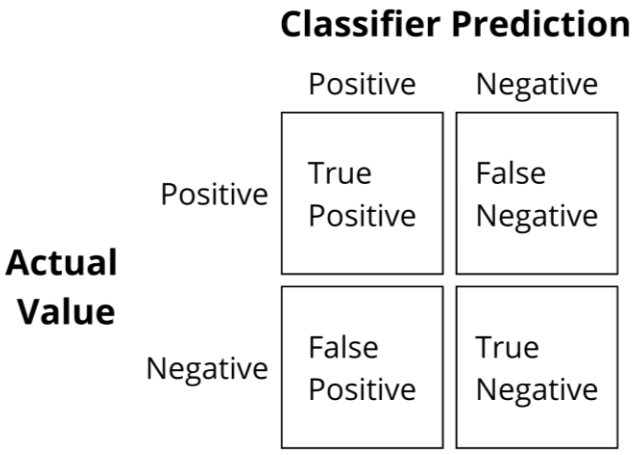



**Fig. 4-2**: Confusion Matrix.

The confusion matrix is represented by positive and a negative class. The positive class represents the not-normal class or behavior, so it is usually less represented than the other class. The negative class, on the other hand, represents normality or a normal behavior.

Here are the four quadrants in a confusion matrix:

- **True Positive (TP)** is an outcome where the model correctly predicts the positive class.
- **True Negative (TN)** is an outcome where the model correctly predicts the negative class.

- **False Positive (FP)** is an outcome where the model incorrectly predicts the positive class.
- **False Negative (FN)** is an outcome where the model incorrectly predicts the negative class.

#### 4.1.2.1.1. Confusion metrics

✓ **Accuracy**

(all correct / all) = TP + TN / TP + TN + FP + FN

This is the general accuracy of the model.

✓ **Misclassification**

(all incorrect / all) = FP + FN / TP + TN + FP + FN

Misclassification states how many cases were not classified correctly.

✓ **Percision**

(true positives / predicted positives) = TP / TP + FP

Percision states, this is a class-level metric.

### 4.1.2.2. Evaluation FAR and FRR

Evaluation is a main part for any project, in Biometric system there are some special evaluation parts used for security reasons and these parts are:

**FAR**: False Acceptance Rate

**FRR**: False Rejection Rate

- **False Acceptance Rate or FAR**

The FAR is expressed as a percentage of situations in which are user gets a false positive result. To calculate the FAR value, you need to divide the sum of imposter scores falling below the threshold by the total number of imposter scores.

FAR = imposter scores falling below threshold / all imposter scores.

imposter scores exceeding threshold = FP

all imposter scores = TP+TN+FP+FN

FAR = FPR = FP/(TP+TN+FP+FN)

- **False Rejection Rate or FRR**

The FRR is expressed as a percentage of situations in which are user gets a false negative result. To calculate the FRR value, you need to divide the sum of genuine scores falling above the threshold by the total number of genuine scores.

FRR = genuines scores falling above exceeding threshold / all genuine scores

genuines scores exceeding threshold = FN

all genuine scores = TP+TN+FP+FN

FRR = FNR = 100 - FN/(TP+TN+FP+FN)

## 4.2. Hardware and Software Tools

### 4.2.1. Hardware

Processor: Intel Core i5 5300U CPU 2.30 GHz– 2 Core(s) – 4 Logical Processor(s).
RAM: 4 GB DDR3.
GPU: Intel HD Graphics 5500.
OS: Linux: "Kali Linux".
Programming Language: Python.

### 4.2.2. Modules

#### 4.2.2.1. OpenCV (Open-Source Computer Vision)

OpenCV (Open-source Computer Vision) is a library which is used for computer vision. [21] OpenCV comes with a trainer as well as detector. [21] OpenCV was worked to give a good infrastructure to PC vision applications and to quicken the utilization of machine perception in business.

It has C++, Python and Java interfaces and backings Windows, Linux, Mac OS, iOS and Android. The library has more than 2500 improved algorithms, which incorporates a far-reaching set of great machine learning algorithms. [21] These algorithms can be utilized to distinguish and perceive faces, recognize objects, group human activities in recordings, track camera developments, track moving articles, extricate 3D models of objects. The library is utilized broadly in organizations, research groups and by administrative bodies.

Installation command: $ ***pip install opencv-python***

#### 4.2.2.2. NumPy

NumPy is the fundamental package for logical computing with Python. It contains in addition to other things:

- Intense N-dimensional array object.
- Tools for integrating C/C++ code.

Other than its undeniable logical uses, NumPy can likewise be utilized as a proficient multi-dimensional container of non-specific information. This enables NumPy to flawlessly and quickly coordinate with a wide range of databases.

#### 4.2.2.3. Dlib

Dlib is an Open-Source C++ toolkit. It contains various machine learning algorithms and tools for creating complex software. Dlib used to solve real-world problems. It is useful in industry and academia including robotics, embedded devices, mobile phones, and large high-performance computing environments [14].

Our network architecture Dlib for face recognition is inspired on ResNet-34 from the Deep Residual Learning for Image Recognition paper by He et al., but with fewer layers and the number of filters reduced by half.

Installation command: $ ***pip install dlib***

Installation command: $ ***pip install cmake***

- ***cmake*** *because dlib was developed in C based programming language.*

#### 4.2.2.4. Face-recognition library

Face recognition module is used to recognize and manipulate faces from Python or from command line from the world's simplest face recognition module. Built using dlib's state-of-the-art face recognition built with deep learning. This provides a simple face recognition command line tool that lets you do face recognition on a folder of images from the command line [16].

Installation command: $ ***pip install face-recognition***

#### 4.2.2.5. PIL

PIL is the python Imaging Library which provides the python interpreter with image editing capabilities. The Image module provides a class with the same name which is used to represent a PIL image.

Installation command: $ ***pip install pillow***

#### 4.2.2.6. Imutils

A series of convenience functions to make basic image processing operations such as translation, rotation, resizing, skeletonization, and displaying Matplotlib images easier with OpenCV and python.

Installation command: $ ***pip install imutils***

#### 4.2.2.7. Os

The OS module in Python provides functions for interacting with the operating system. OS comes under Python's standard utility modules. This module provides a portable way of using operating system-dependent functionality. The *os* and *os.path* modules include many functions to interact with the file system.

#### 4.2.2.8. Matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible [17].

Installation command: $ ***pip install matplotlib***

#### 4.2.2.9. Scikit-image

scikit-image is a collection of algorithms for image processing. It is available free of charge and free of restriction.

Installation command: $ ***pip install scikit-image***

## 4.3. Methodology

In this part, we will discuss the architecture of the attendance system based on facial recognition. (Fig 4-3) shows all the sequence of attendance marks from taking photos to generating the attendance file.

### 4.3.1. Attendance system

The steps involved in the proposed attendance system is described below:

- The registration system will be considered when the students of a session are registered and admitted. Each student's image will be captured and processed. The face portion will be extracted from each image and stored in the database. The image database will relate to the face detection and recognition system. The stored images will be taken by the face recognition model for comparison.
- The camera installed in the classroom will continuously detect faces of the students. Whenever a student passes in front of the camera, his face will be detected. To avoid false positive rate, each student's face will be taken by multiple frames. At least 10 frames will be taken for each face.

- For identifying the face in real time, we have used deep face model. Dlib can handle multiple stages of face detection and recognition tasks like detect, align and detect landmark of the faces, and verify. A Dlib face recognition network model with 29 convolutional layers, an optimized version of the well-used ResNet-34 network. This model was trained on three million faces across various datasets, including Face Scrub, Oxford's VGG set. He got 99.38% accuracy after tested the built for labeled faces in the wild (LFW) data set.
- When the image is read from the frame of a camera, the image colors are in order of BGR (blue, green, red). Therefore, the BGR image is converted first to RGB. Before converting, the face ROI is extracted. After that, the resize, operation is conducted on these images. The best resizing process is to make each image as 250*250.
- After the face is detected using the HOG descriptor, comes the student recognition step, the system compares this image to the image stored in the database from the registration system. If the image is compared and matched with the stored image, that student will get attendance.
- After marking attendance, the system will generate report as PDF, the teacher can download it.

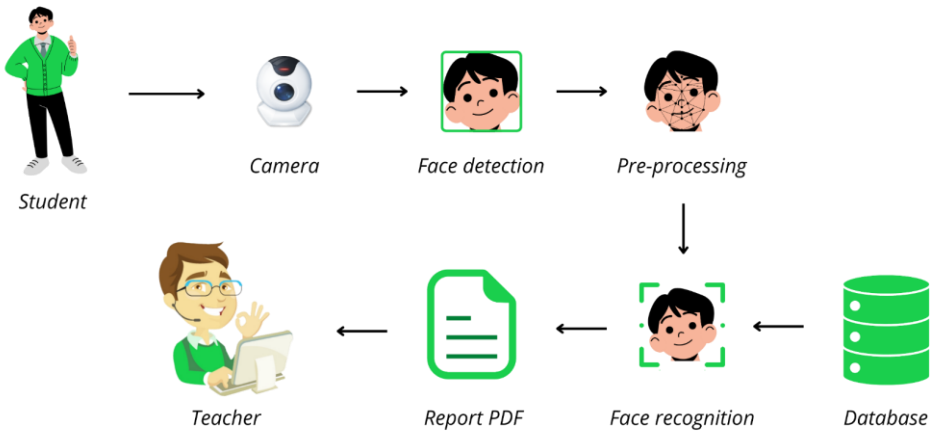



**Fig. 4-3**: Proposed facial recognition-based attendance system architecture.

## 4.4. Testing and Result

### 4.4.1. Face detection

Though in this part we will not just test the frontal face detection but also different angles of the image and see where our model will perform well and where not along with that, we will be computing the total time taken by the HOG detector model to detect the faces in the image.

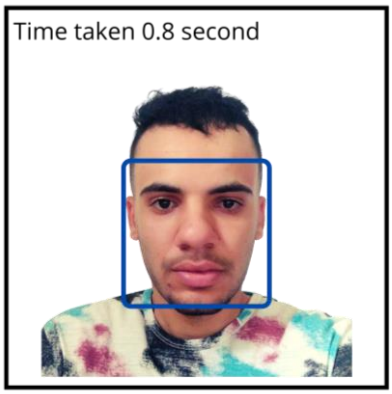



**Fig. 4-4:** Reading sample images of frontal face recognition.

In the above output, you can see that our model has perfectly detected the face in 0.8 seconds which shows that along with being accurate it is faster as well.

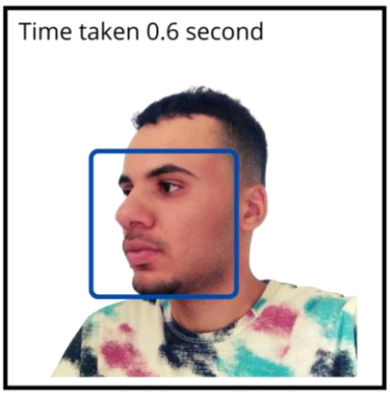



**Fig. 4-5:** Reading a sample picture with a slanted face.

So, from the above output, we can conclude that the HOG face detection model not only detects the frontal face but side tilted side face as well with ease and efficiency, and that too pretty fast. As this HOG algorithm is trained with the purpose to detect a face size of at least 80×80 so whenever we think of detecting a face smaller than that we need to up sample the image which

will increase the resolution of the image. though the computing time will also increase after this process, as shown below in (Fig 4-6).


**Fig. 4-6:** Reading face sample images smaller than 80x80.

- ✓ Dlib HoG is the fastest method on CPU. But it does not detect small sized faces (80×80). So, if you know that your application will not be dealing with very small sized faces (for example a selfie app), then HoG is a better option. These descriptors are powerful to detect faces with occlusions, pose and illumination changes because they are extracted in a regular grid.

### 4.4.2. Pre-processing

#### 4.4.2.1. Face Landmarks

Added a 5-points face landmarking model in 2017, that is over 10x smaller than the 68-point model (9.2MB versus 99.7MB, respectively), runs faster (by 8-10%), more efficient, and works with both HOG and CNN generated face detections, as seen in (Fig 4-7).

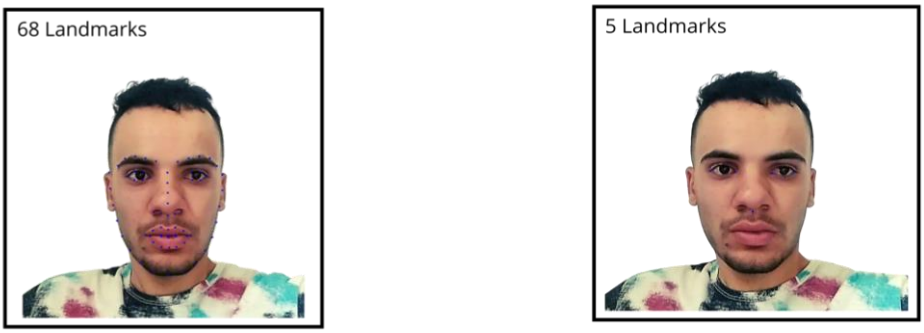



**Fig. 4-7**: A comparison of the dlib 68-point facial landmarks (right) and the 5-point facial landmarks (left).

Figure above visualizes the difference between dlib's new 5-point facial landmark detector versus the original 68-point detector. While the 68-point detector localizes regions along the eyes, eyebrows, nose, mouth, and jawline, the 5-point facial landmark detector reduces this information to:

- **2 points** for the left eye
- **2 points** for the right eye
- **1 point** for the nose

✓ The primary usage of the 5-point facial landmark detector will be face alignment.

#### 4.4.2.2. Face Alignment

Face alignment can be considered a form of "data normalization". They are often used to improve the accuracy of facial recognition algorithms, including deep learning models. It is very common to align faces in a data set before training a machine learning model by scaling. and the faces are scaled so that the size of the faces is nearly identical. Can be achieve this using an affine transform as shown in (Fig 4-8).

For face alignment, the 5-point facial landmark detector can be considered a drop-in replacement for the 68-point detector — the same general algorithm applies:

- Compute the 5-point facial landmarks
- Compute the center of each eye based on the two landmarks for each eye, respectively
- Compute the angle between the eye centroids by utilizing the midpoint between the eyes
- Obtain a canonical alignment of the face by applying an affine transformation

**Fig. 4-8:** Facial alignment using facial landmarks.

✓ While the 68-point facial landmark detector may give us a slightly better approximation of the eye centers, in practice we'll find that the 5-point facial feature detector works just as well. However, it cannot be used in all situations.

### 4.4.3. Face recognition success rate of AI model with marking attendnace

The success rate of face recognition model is calculated using selected training and testing set, the AI model is trained with 35 images for students and testing is done using testing set which consist of 40 number of images and by showing the images to the camera as input for the face recognition model to recognise the students. the AI model shows as unknown if it is not recognized, and if the face is trained, the AI model recognizes the person's face. The figures below are some of our results obtained during the testing phase and have been tabled for ease of understanding.

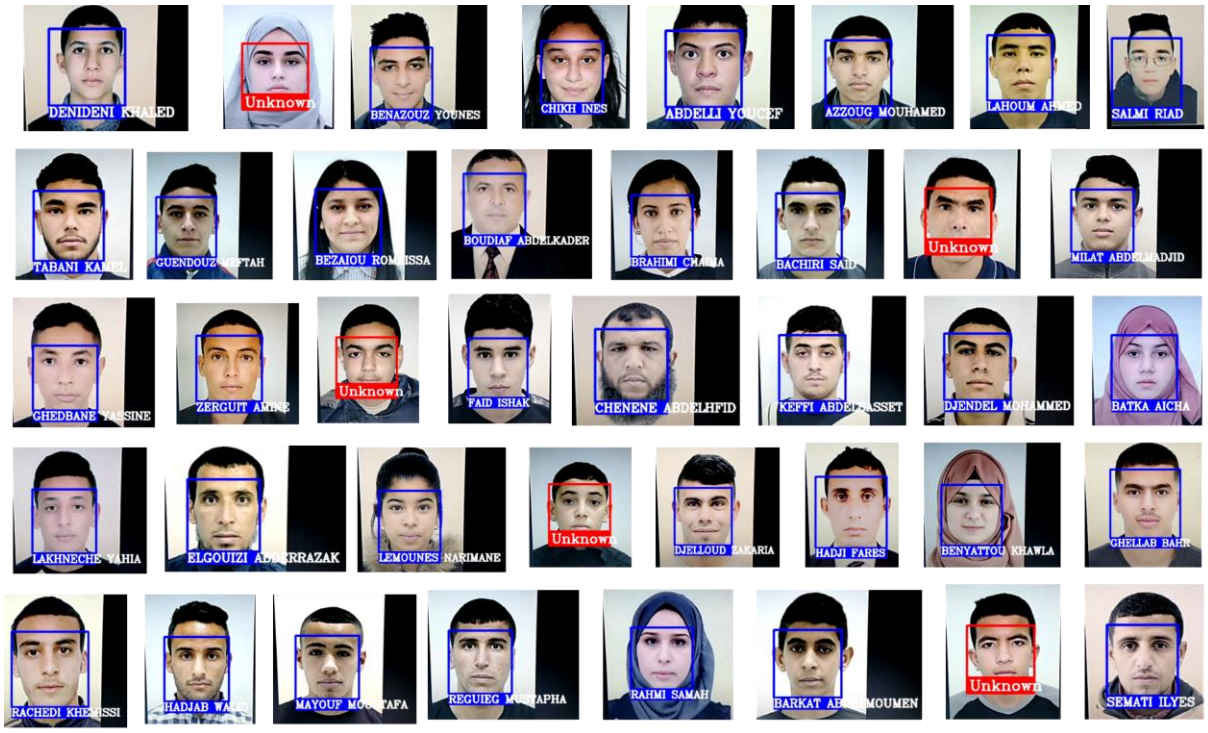



**Fig. 4-9**: AI model Recognizing faces.

#### 4.4.3.1. Confusion matrix results

Since Dlib facilitates free and open-source face recognition. The accuracy ranges from 99.38%. By utilizing the standard LFW Benchmark, and by predicting the images are success or not, the accuracy is calculated. For calculating the details like accuracy, precision, misclassification and error rate through confusion matrix. We considered a training dataset of 40 pictures and perform the experiment by utilizing known and unknown faces. If the picture is not trained in then, output is ‘unknown’. If the face detected is ‘unknown’ then it is considered to be fail (0) otherwise it is considered as successful attempt (1).

- **True positive:** If the observed face matches the predicted face by the model, it is true positive.
- **True negative:** Actually, the face is not trained and our model also predicts as unknown.
- **False positive:** The model detects the face and recognizes the face, but the prediction is wrong then it is called false positive.
- **False Negative:** The model predicts the image that is trained in as unknown.

  The confusion matrix created on the train data set is:

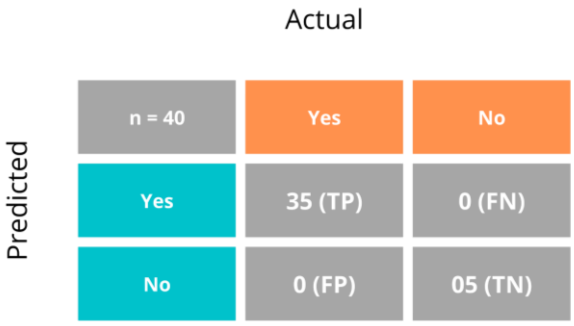



**Fig. 4-10**: Confusion matrix of AI model.

The above matrix shows the actual and predicted values which resulted in:

- True positive: 35
- True negative: 05
- False positive: 0
- False negative: 0

**Accuracy:** This is the percentage of times a classifier is correct.

Accuracy = (TP + TN) / (TP + TN + FP + FN) = 1.

Hence, we have achieved an accuracy of 100%.

**Misclassification:** This is the percentage of times a classifier is incorrect.

Misclassification = FP + FN / TP + TN + FP + FN = 0

Here, we got a 0% misclassification.

**Precision:** This is the rate at which the desirable predictions turn out to be correct.

Precision = TP / (FP + TP) = 1.

We have determined a precision of 100%.

- ✓ The Secret to Better Face Recognition Accuracy: **Thresholds**
- ✓ To reduce False positives, a threshold value equal 0.5 is preferred (tolerance = 0.5)

#### 4.4.3.2. FAR and FRR assessment results

**False Accept:** System claims a pair of pictures are a match, when they are actually pictures of different individuals.

**False Accept Rate (FAR):** Frequency that the system makes False Accepts.

In Our Example: FAR of 0% system will make 0 false accept for every 35 imposter attempts.

**FAR** = 0%.

**False Reject:** System claims a pair of pictures are a mismatch, when they are actually pictures of the same individual.

**False Reject Rate (FRR):** Frequency that the system makes False Rejects.

**ID Rate** = 100% minus FRR:

In our Example: FRR of 0% or Identification rate of 100% system will reject 0 matches for every 5 authorized attempts.

**FRR** = 100%.

#### 4.4.3.3. Showing attendance report as PDF

## List of Students - 2022-05-17

**Module : IHM WEB**

**Groupe : 01**

**Date : 2022-05-17 – 08:00:00 ---- 10:00:00**

**Prof: KamelEddine HERAGUEMI**

| N. Inscreption | Name | Observation |
|---|---|---|
| 161635090750 | ABDELLI YOUCEF | |
| 162135101400 | AZZOUG MOUHAMED | |
| 171735097309 | BACHIRI SAID | |
| 212135080533 | BARKAT ABDELMOUMEN | |
| 212135096762 | BATKA AICHA | |
| 212135079787 | BENAZOUZ YOUNES | |
| 212135087589 | BENYATTOU KHAWLA | |
| 212135095002 | BEZAIOU ROMAISSA | |
| 212135099199 | BOUDIAF ABDELKADER | |
| 212135096693 | BRAHIMI CHAIMA | |
| 212135093572 | CHENENE ABDELHFID | |
| 212135082244 | CHIKH INES | |
| 212135082100 | DENIDENI KHALED | |
| 212135096511 | DJELLOUD ZAKARIA | |
| 212133050619 | DJENDEL MOHAMMED | |
| 212135101195 | ELGOUIZI ABDERRAZAK | |
| 191935076129 | FAID ISHAK | |
| 212135080981 | GHEDBANE YASSINE | |
| 212135096245 | GHELLAB BAHR | |
| 161635096545 | GUENDOUZ MEFTAH | |
| 212135094825 | HADJAB WALID | |
| 161635101583 | HADJI FARES | |
| 212135055737 | KEFFI ABDELBASSET | |
| 212135101842 | LAHOUM AHMED | |
| 212135083246 | LAKHNECHE YAHIA | |
| 212135088497 | LEMOUNES NARIMANE | |
| 172135097151 | MAYOUF MOUSTAFA | |
| 212135086444 | MILAT ABDELMADJID | |
| 212133050028 | RACHEDI KHEMISSI | |
| 131335074046 | RAHMI SAMAH | |
| 212135101949 | REGUIEG MUSTAPHA | |
| 212135082107 | SALMI RIAD | |
| 212135087021 | SEMATI ILYES | |
| 212135079677 | TABANI KAMEL | |
| 162135100625 | ZERGUIT AMINE | |

**Table 4-1**: Attendance list.

# Conclusion

Face recognition is a challenging problem and isn't 100% accurate in the field of image processing and computer vision, which has received a great deal of attention over the past years because of its several applications in various domains. Although research efforts have been conducted roundly in this area, achieving mature face recognition systems for operating under constrained conditions, they're far from achieving the ideal of being suitable to perform adequately in all various situations that are general encounter by operations, such as lighting conditions, orientations, scale and facial expressions. Yet, as digital culture migrates to ‘Realtime’ biometric identity verification through mobile and IoT devices, reliability has to extend beyond these restrictions — so rigorous work is being done to train these systems to be more effective in "real-world" conditions presently considered unfavorable for accuracy. This paper aims to exploit the use of face recognition technology in other scientific and daily life applications such as attendance recording that we have touched upon. Also, many algorithms are still being proposed to improve the performance of these systems as future directions. Finally, this paper concludes by arguing that the next step in the evolution of face recognition algorithms will require radical and courageous steps forward in terms of the face representations/descriptors, as well as the learning algorithms used.

# Bibliography


[1] Stan Z. Li, Anil Jain « Encyclopedia of Biometrics » 2009.

[2] R. Gross, J. Shi, J. Cohn "Quo vadis Face Recognition?", June 2001.

[3] Y. Kabbara « Characterization of X-ray images of the hand by mathematical els: application to biometrics » Paris East University; Lebanese University, ;.

[4] Hietmeyer, R. Biometric identification promises fast and secure processing rline passengers. Int. Civ. Aviat. Organ. J. **2000**, 17, 10–11.

[5] G. Guo, S.Z. Li, K. Chan, Face Recognition by Support Vector Machines, . Of the IEEE International Conference on Automatic Face and Gesture ognition, 26-30 March 2000, Grenoble, France, pp. 196-201

[6] B. Ibtissam « Etude et mise au point d'un procédé biométrique multimodale la reconnaissance des individus » Université Oran Mohamed Boudiaf, 2015.

[7] John D. Woodward, Jr., Christopher Horn, Julius Gatune, and Aryn mas, "Biometrics A Look at Facial Recognition", documented briefing by ND Public Safety and Justice for the Virginia State Crime Commission, 2003.

[8] Pradeep Buddharaju, Ioannis T. Pavlidis, Senior Member, IEEE, Panagiotis myrtzis, and Mike Bazakos «Physiology-Based Face Recognition in the mal Infrared Spectrum ».

[9] A. Jain, R. Bolle and S. Pankanti, Biometrics: Personal Identification in vorked Society, New York: Springer, 2006.

[10] S. Hocquet, "Authentification biométrique adaptative Application à la umique de frappe et à la signature manuscrite", Université François Rabelais s, 2007.

[11] https://www.grandviewresearch.com/

[12] https://www.knowledgehut.com/

[13] Anil. k. jain, P. Flynn, A. Ross, « Handbook of biometrics », Springer, 2007.

[14] http://dlib.net/

[15] D. Blackburn, M. Bone, P. J Phillips. “Face recognition vendor test 2000”. ı. rep. http://www.frvt.org, 2001.

[16] https://face-recognition.readthedocs.io/

[17] https://matplotlib.org/

[18] Geitgey, A. “How to do Modern Face Recognition with Deep Learning “, )17.

[19] https://stanford.edu/

[20] Y. Kortli, M. Jridi, A. Al Falou, M. Atri “Face Recognition Systems: A 'ey”, 7 January 2020.

[21] https://pypi.org

[22] https://learnopencv.com/

[23] V. Kazemi, J. Sullivan, “One Millisecond Face Alignment with an :mble of Regression Trees.”

[24] M. Hahnle, F. Saxen, M. Hisung, U. Brunsmann, K. Doll “FPGA-based .-Time Pedestrian Detection on High-Resolution Images” June 2013

[25] https://d2l.ai/

[26] Nilsson, K., and Bigun, J. “Localization of corresponding points in :rprints by complex filtering”. Pattern Recognition Letters ,2135–2144,2003.

[27] R. Gross, J. Shi, J. Cohn. Quo Vadis Face Recognition? Third Workshop on ıirical Evaluation Methods in Computer Vision, December, 2001.

[28] https://www.ibm.com/

[29] Z. Akhtar, A. Rattani «A Face in any Form: New Challenges and ortunities for Face Recognition Technology», april 2017.

[30] https://www.intechopen.com/

[31] Morizet, M. Reconnaissance Biométrique Par Fusion Multimodale du ge et de lÍris., ParisTech, Paris, France, 2009.

[32] M. Hemmer, Huynh V. Khang, Kjell G. Robbersmyr, Tor I. Waag, Thomas Meyer“Fault Classification of Axial and Radial Roller Bearings Using ısfer Learning through a Pretrained Convolutional Neural Network ”, 19 ɛmber 2018 .

[33] https://blogs.oracle.com/

[34] https://towardsdatascience.com/