1290

UNIVERSIDADE DE COIMBRA

Rui Filipe de Arvins Barbosa

Accuracy analysis of region-based 2D and 3D Face Recognition

COMPARISON OF NASAL AND MASK-WEARING OCULAR REGIONS

Dissertação no âmbito do Mestrado Integrado em Engenharia Electrotécnica e de Computadores, ramo de especialização em computadores orientada pelo Professor Doutor Nuno Miguel Mendonça da Silva Gonçalves e apresentada à Faculdade de Ciências e Tecnologia da Universidade de Coimbra no Departamento de Engenharia Electrotécnica e de Computadores.

Outubro de 2020

1 2 90

UNIVERSIDADE D COIMBRA

Accuracy Analysis of Region-Based 2D and 3D Face Recognition

Comparasion of Nasal and Mask-Wearing Ocular Regions

Rui Filipe de Arvins Barbosa

Dissertação para obtenção do Grau de Mestre em

Engenharia Electrotécnica e de Computadores

Orientador: Nuno Miguel Mendonça da Silva Gonçalves

Júri

Presidente: Jorge Manuel Moreira de Campos Pereira Batista

Vogal: Nuno Miguel Mendonça da Silva Gonçalves
Paulo José Monteiro Peixoto

Outubro de 2020

I want to thank me for doing all this hard work. I want to thank me for having no days off. I want to thank me for never quitting.

- Calvin Cordozar Broadus Jr.

Agradecimentos

Gostaria de agradecer ao professor Nuno Gonçalves pelo privilégio da sua orientação, pela disponibilidade, dedicação e apoio na elaboração deste trabalho.

Aos membros da VISteam que direta ou indiretamente estiveram ligados a esta pesquisa pelas suas sugestões e ajuda prestada durante a realização da presente dissertação.

Ao Instituto de Sistemas e Robótica por todos os meios disponibilizados e pelo ambiente de trabalho profissional proporcionado.

Sou muito grato a todos os meus familiares e amigos pelo incentivo recebido ao longo destes anos e pela experiência de vida transmitida ajudando-me a manter o foco sobre o que realmente é importante.

A todos,
Muito Obrigado

Abstract

The evolution of FR systems enable the incorporation of 3 dimensions analysis combined with the 2 dimensions methods already developed was largely a consequence of the development occurred in the available data acquisition technology and FR improved mechanisms such as the use of Machine Learning (ML) algorithms.

Despite FR systems recent achievements, new habits changes such as the generalization of face covering, as a consequence of COVID-19 pandemic present a new challenge to FR algorithms. The majority of the methods have not been tested in this new reality making an updated survey over FR fundamental to understand if they can be reused or appear obsolete.

In this work a classic feature extractor algorithm developed by Emambakhsh & Evans et al. [1] based on spherical patches working along with a designed and personalized NN is applied with the objective to demonstrate the importance of the nasal region for 3D FR algorithms, as stated in the article.

In order to adapt the research to a new reality, tests were performed to prove that algorithms focused on the ocular region reach similar values of success when compared with the nasal region in order to overcome the nose occlusion consequence of using face coverings due to the COVID-19 pandemic. A second version of the FR system built for the first goal demonstration was implemented having been proved that the ocular region effectively has comparable accuracy in the 3D domain.

Keywords

FR with mask, 2D and 3D FR, 3D Face Validation, 3D Face Verification.

Resumo

A evolução dos sistemas de FR permitiu incorporar análises 3D combinadas com os métodos 2D já desenvolvidos, foi em grande parte consequência do desenvolvimento da tecnologia de aquisição de dados disponível e mecanismos de FR aprimorados, como o recurso a ML.

Apesar das últimas conquistas dos sistemas de FR, recentes mudanças de hábitos, como a generalização da utilização da máscara como consequência da pandemia de COVID-19, representam um novo desafio para os algoritmos de FR. A maioria dos métodos não foi testada nesta nova realidade, tornando um estudo atualizado sobre FR fundamental para entender se poderão ser reutilizados ou se encontram obsoletos.

Neste trabalho um modelo clássico de deteção de features desenvolvido por Emambakhsh & Evans et al. [1], baseado em patches nasais esfericos que combinados com uma NN projetada e personalizada é projetado, com o objetivo de analisar possíveis aplicações sobre uma população multicultural e diversificada, como foi proposto pelo artigo.

Para adequar a tese à nova realidade, foram realizados testes para comprovar que algoritmos focados na região ocular alcançam valores de sucesso semelhantes quando comparados com a região nasal de forma a superar a oclusão da mesma como consequência da utilização de máscara devido à pandemia COVID. Uma segunda versão do sistema FR inicialmente implementado para demonstrar o primeiro objetivo foi projetada tendo sido demonstrado que estes efetivamente mantêm uma precisão comparável no domínio 3D.

Palavras Chave

Reconhecimento Facial com máscara, Reconhecimento Facial 2D e 3D, Validação Facial 3D, Verificação Facial 3D

Contents

List of Acronyms

CV Computer Vision

ISR Instituto de Sistemas e Robótica

INCM Imprensa Nacional-Casa da Moeda

ML Machine Learning

DNN Deep Neural Network

NN Neural Network

FR Facial Recognition

SVM Support Vectors Machine

PCA Principal Component Analysis

ML Machine Learning

HOG Histogram of Oriented Gradients

1

Introduction

Contents

1.1 Context

This master thesis was undertaken in the context of the TrustFaces project. This project is financed by the Imprensa Nacional-Casa da Moeda (INCM).

The TrustFaces project aims at a partnership for research and development in the area of reliable labels for the certification of products and identification (ID) documents, namely in the authenticity certification area, more specifically in the creation of stamps and encoded images of faces of people, with applications in personal ID documents.

Developing a reliable Facial Recognition (FR) system, robust enough to maintain good scores despite environments variations or facial expression changes, has become a point of interest for an increasing number of organizations. The cumulative technological achievements based on new accuracy and precision levels made Computer Vision (CV) a powerful, useful and widespread tool imposing itself as one of the major areas in Computer Science nowadays. This growing popularity makes possible to surpass the theoretical and research state to a practical one, revealing itself as a new unexplored market.

The number of applications for solving real-world problems in scene reconstruction, event detection, object recognition, image restoration and machine learning since the use of Neural Network (NN) with multiple processing layers to understand representations of data from multiple levels of feature extraction [2], turn CV an attractive investment field for public and private corporations as well as big companies in the technological industry and official government entities.

FR application becomes almost universally useful, from surveillance systems and tracking people searching for possible fugitives or abducted people, thus becoming an auxiliary tool to resolve criminal investigations. Business and finances can also represent new applications areas with FR becoming a more popular choice in payment services, maximizing security and minimizing fraud.

Notwithstanding the proved potential that FR technology presented the COVID-19 pandemic has led to an unprecedented generalization of facial coverings while concomitantly accelerating the use of digital surveillance tools focusing on those that do not require any contact for biometric recognition. This new situation should be used as an opportunity to enhance FR performance with dataset upgrades, new algorithms and new test over the old ones. Taking these factors into account some modifications were made to our tests where mask models were included to prove our methodology performance.

On a social point of view FR manifest a good “status quo” between what is socially acceptable and secure against what is reliable. Questions such as whether facial recognition should be used as legal evidence considering that all systems are still flawed, are still unanswered by the academic community. This debate was conducted by the Ethics and Governance of Artificial Intelligence project [3], reminding the general public for the problems attached to privacy invasion, exploring technical and societal risks and ethics, and governance issues for Artificial Intelligence.

The next question will be when the use of this technology by official government institutions is scheduled after the private companies have taken the first step. Never forgetting existing documents and procedures, facial validation undeniably has the potential to lighten identity verification processes to a new level. To move towards this goal, a study of the State-of-the-Art techniques recent developed is imperative.

1.2 Motivation

FR is the practical consequence of every human having a unique face. Nowadays we can use technological advancements to achieve a complete method that uses it as an element of personal security in a reliably design. More than signature, fingerprint or even voice, our face is a fundamental identification form. Advancements in 2D or 3D methods make possible to capture, analyze and calculate data with the desired precision and accuracy.

The research conducted by Shakir F. Kak & Firas Mahmood Mustafa & Pedro Valente in the Eurasian Journal of Science & Engineering [4], present to the academic community a quick recap on the FR technological applications.

Table 1.1 FR practical scenarios applications.

Fields	Scenario of applications
Security	Terrorist alert Secure flight boarding systems Stadium audience scanning Computer security
Face ID	Driver licenses Entitlement programs National ID
Face Indexing	Labeling faces in the video
Access Control	Border-crossing control Facility and vehicle access
Multimedia Environment	Face-based search and video segmentation Event detection
Smart Cards Application	Stored value security User authentication
Face Databases	Face indexing and classification Automatic face labeling
Surveillance	Advanced video surveillance and CCTV Control Nuclear plant or Power grid surveillance Park surveillance and neighborhood watch

Comparing the scan between face and iris, as an example, the iris scanning is very invasive, even though the iris provides excellent accuracy and precision with a quick validation. Even with 2D recognition having a very high success rate, it would be interesting to improve the rates for particular cases of types of images, since most algorithms are trained with non-representative populations.

When we focus on the 3D case, it is not fully known how much this type of data improves recognition in specific populations, and this research seeks to answer these questions directly.

Some of the proposed solutions still show drawbacks as consequence of the parameters that it will need to compute, such as processing and storing quality data, 3D model size and model position angle and environment conditions, and 3D data acquisition extra sensors as Stereo System (a classic model of 3D cameras), Multi-view, Structured-Light 3D Scanner, Time-of-Flight (ToF), Light-Field and Plenoptic cameras are been used to estimate the 3D model.

Analyzing the economic perspective on FR allied to the new applications developed, the perspectives are also positively expectant. The research complete by the association UNLOCKING POTENTIAL in partnership with the University of Exeter reveals that the global market for FR is consistently growing, estimated to be at around 4.05 billion USD as of November 2017 and estimated to increase to around 7.76 billion USD in 2022, an estimated 92 percent increase, with North America and Europe countries showing on the largest market share, followed by Asia the most rapidly growing market, with the world's largest market in surveillance [5].

So the application and individual use as an official and certified element of security and validation, in a utopic perspective it's almost perfect. In this research area accuracy becomes a superlative factor on the next algorithms developed. Even passenger transportation will be positively affected with major benefits, as FR has already been deployed in airports and train stations to save travelers time from checking in or helping travelers to pay for their fares.

1.3 Goals

Future FR systems are expected to bring improvements in the way data are processed and the search engines are designed. Such improvements are related to higher data quality, bigger datasets, improved algorithms, and advanced machine learning classifiers brought by the need to upgrade the score results over hostile environments aligned with the 3D FR algorithm's benefits.

Most FR software in use today relies on two dimensional technology, namely photographs or video images. These images are easy to obtain, by a simple picture or using frame from a security camera video, per example.

As previously exposed, with the need to have more clear idea over the potential of 2D and 3D FR techniques and respective advantages and disadvantages, the first goal set for this research project is to structure a robust FR system based on Mehryar Emambakhsh & Adrian Evans article [1] in order to prove the importance of the nose region in 3D FR systems. To achieve this goal, two altered versions of the algorithm provided in the article were tested in order to compare the results of the method stated in the research with the entire 3D model of the face

As a direct consequence of the current pandemic reality, it was decided to extend the objectives of our research in order to test a method that was not compromised by the use of a mask. The proposal presented in this work is to shift the focus from the nasal area, now being often covered, to the ocular zone in terms of analysis according to 3D algorithms.

Therefore, to achieve the objectives set out above, it is proposed to develop comparative methods and register all variables in the model environment that may affect the desired results, and then study how to overcome the problems they represent.

1.4 Chapter outline

This chapter introduces the topic of the thesis, the motives that led to investigate and enhance the knowledge in this particular theme and describes the main goals proposed to achieve with this work. Chapter two introduces the current state of the art in which recent 2D and 3D FR algorithms are analyzed. The datasets considered for the practical approach are quickly summarized in the next chapter. It is in the last three chapters where our practical methodology, experimental results and conclusions of this master thesis report is outlined.

2

Facial Recognition Fundamentals

Contents

The process of labeling a face as recognized or unrecognized is the simplest description for a FR system, going through detection, feature extraction and recognition stages defining the process pipeline. This chapter initially presents a brief resume of the essential steps of FR systems followed by a research over the state of the art methods used in the feature extraction and Machine Learning (ML) stages present in FR algorithms, as these were the stages where we focused on most of our research.

2.1 Facial Recognition Process

2.1.1 Detection

The majority of FR system process begin with the detection of the face on the input data. The ability to locate the facial region is critical to every FR system so variables as multiple facial expressions and head orientations, different ethnicity or gender and illumination variations must be considered when FR system are developed. To overcome such problems a multiple set of sensors incorporating RGB, depth and thermal are used to collect new data and delivering extra information to FR methods improving robustness and reliability of the system.

A major breakthrough in face detection appeared with the Viola-Jones detection framework, motivated primarily by the problem of face detection. Viola Jones classified the images according with simple features using three different types of features: square features, three-square features, and a four-square feature. The features value's is set as the difference between black and white regions.

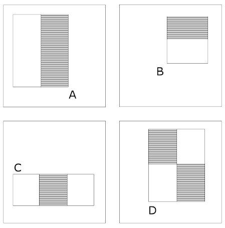


Figure 2.1: Example rectangle features shown relative to the enclosing detection window.

For the classifier to work properly, the size of the image in the training set must be the same as the size of the input image used for object detection.

Histogram of Oriented Gradients (HOG) algorithms were first proposed by N. Dalal et al. [6] and have been pioneered in face detection in the recent decade. In this article it is present a window-based feature set in which HOG feature vectors are extracted and used in a pre-trained binary classifier depending on face detection success.

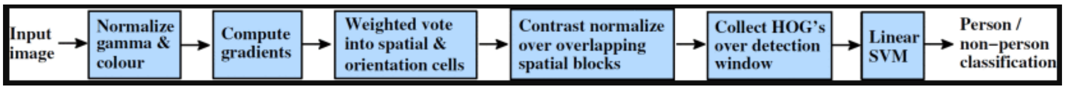

Input image
Normalize gamma & colour
Compute gradients
Weighted vote into spatial & orientation cells
Contrast normalize over overlapping spatial blocks
Collect HOG's over detection window
Linear SVM
Person / non−person classification

Figure 2.2: Overview of the feature extraction and object detection pipeline, in this particular case for people detection.

The HOG representation has several advantages. It is capable of capture edge or gradient structure very characteristic of local shape with an easily controllable degree of invariance to local geometric and photometric transformations as translations or rotations make little difference if they are much smaller that the local spatial or orientation bin size.

The Principal Component Analysis (PCA) is also one of the most successful techniques that have been used in image recognition and compression after Sirovich et al. [7] exploit a PCA based classifier to represent the images of human faces. It is a statistical method with the purpose to reduce large dimensionality of the data space to smaller intrinsic dimensionality feature space , by compute large 1-D vector of pixels constructed from 2-D facial image into the compact principal components of the feature space.

2.1.2 Feature Extraction

As stated previously FR algorithms are susceptible to head orientation, partial occlusion, facial expression, and light condition denominated external effects. To minimize these effects on the performance of the algorithm, removing unwanted features such as shadow or excessive illumination and to reduce error, facial images are pre-processed to make their recognition friendly.

The principal objective of this stage is to extract features from images or 3D models. It is based on the principle that each face is unique and characterized by its structure, shape and size. Several techniques such as HOG [8] , Eigenface, independent component analysis, linear discriminant analysis [9], scale-invariant feature transform [10], Haar wavelets, Fourier transforms and local binary pattern [11] techniques are widely used to extract the facial features.

2.1.3 Machine Learning Algorithms

ML can be defined as a branch of Artificial Intelligence and Computer Science that is an ever-evolving method with the goal of understand the structure of data by fitting it into understandable models while increasing its own performance with each iteration. In ML in general, tasks are classified into categories describing how learning part is processed.

Supervised Learning is a type of ML algorithm that is used to discover known patterns on unknown data. If someone gives a ML algorithm some images of different objects with the classification of the objects on each image sample, it is expected the algorithm learn how to say which type of object is in a given image that was not presented to the algorithm during its training stage.

On the other hand Unsupervised Learning is another type of ML algorithm used to discover unknown patterns on known data. For instance, if someone have a database with the shopping list of every client that shops on that supermarket, they could apply an unsupervised learning to understand what kind of products the clients are more likely to buy together.

2.1.3.A Neural Networks

Artificial NN is one group of algorithms used for ML that models the data using graphs of Artificial Neurons. Those neurons are a mathematical model that simulate the process of the neurons work inside a brain, connected to each other, and the strength of their connections to one another is assigned a value based on their strength: inhibition (maximum being -1.0) or excitation (maximum being $+1.0$). There are three types of neutrons in an ANN denominated input, hidden and output nodes.

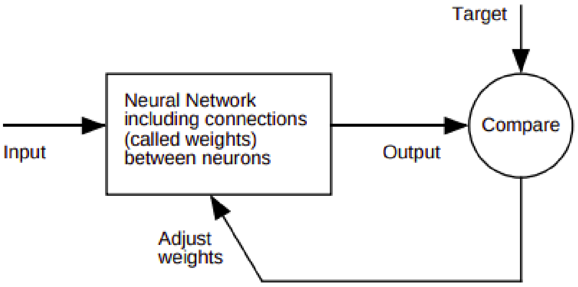


Figure 2.3: NN minimalist diagram, present in Matlab NN Toolbox user's guide.

Commonly Neural Networks (NN's) are adjusted or trained, so that a particular input leads to a specific target output following a learning algorithm. The first layer has input neurons which send data via synapses to the next layer of neurons, and then via more synapses to the following layers until the signal reach the output neurons. More complex systems will have more layers of neurons with some having increased layers of input neurons and output neurons. These synapses parameters are called ”weights” with responsibility to manipulate the data in the calculations. A

NN is typically defined by the interconnection pattern between the different layers of neurons, the learning process for updating the weights of the interconnections and the activation function.

Back-propagation methods are one of the most popular learning algorithm class, responsible to find the optimal weights among neurons. Every learn algorithm has unique characteristics that must be reminded when the NN is being designed as every class is suitable for certain specific tasks. Descending gradient is a classical back-propagation algorithm for training neural networks.

It is one of the most used in FR state of the art systems. If $E(\omega)$ is defined as the error function as function of the weights ω of the NN, the learning algorithm looks for a global minimum of this error function. It can be expressed as given the weights of the network $\omega(0)$ for the instant $n = 0$ and a learning factor μ, the direction of greatest variation of the error function is calculated, which is given by the gradient $\Delta E(\omega)$.

$$\omega(n+1) = \omega(n+1) - \mu \cdot \Delta E(\omega) \tag{2.1}$$

In the next section it will be presented a state of the art survey of the existing FR techniques and ML algorithms describing the related advanced research of deep learning and FR.

2.2 Sate of the Art

In this section the research advancements of FR algorithms and ML methods will be described. We followed the research conducted by Jin Bo et al. [12]:

2.2.1 Face Recognition

FR refers to the identity's identification or verification of the subjects collected from faces in images or videos in the case of 2 dimension or point clouds or mesh-grids in 3D case. Face identification is the task of matching a given face image to one in a database of faces represented by a one-to-many mapping while face verification, on the other hand, is the task of comparing two candidate faces and verifying if the result is a match represented by a one-to-one mapping

The 2D FR strand can use recognition algorithms to identify facial features by analysing relative position, size, and/or shape of the eyes, nose, cheekbones, jaw and extracting landmarks or texture information from an image of the subject's face. Another possibility will be normalizing a gallery of images and then compress the data, only just saving the one that is useful. On the other hand, 3D image processing reveals the potential to improve the recognition and validation performances presented by 2D FR when it is keep in perspective the principal problems faced by 2D FR technology.

A smart integration between the texture image and facial surface is the next step. As new methods are being developed, Deep Neural Network (DNN) have marked themselves as one of the principal classification techniques [2]. DNN have scored top performers on a wide variety of applications including image classification and FR.

2.2.2 Appearance based

Appearance based algorithms use image pixel data as a whole for recognition. Direct Correlation, Eigen-face and Fisher-face belong to this class of methods. Direct correlation uses direct comparison of image pixels of two facial images. Unlike direct correlation algorithm Eigen-face and Fisher-face do not use facial images in their original image space, these algorithms reduce the image to the most discriminating factor and make their comparison between images in a reduced dimension image space.

Usually the information of interest can be found in a lower dimension than the original dimension. The dimensionality reduction approach brings out useful information that can be revealed in lower dimensions.

2.2.3 Active appearance

Active Appearance Model algorithms contain statistical information of an image shape and texture variation, built during a training phase. Is a FR algorithm class for matching a statistical model of object shape and appearance to a new image. A set of images, together with coordinates of landmarks that appear in all the images, is provided to the training supervisor algorithm.

This algorithm class uses the difference between the current estimate of appearance and the target image to drive an optimization process. By taking advantage of the least squares techniques, it can match to new images very quickly.

2.2.4 Bayesian model

The Bayesian Model algorithms compute a real-valued function defined on a set of events in a probability space of similarity derived from a Bayesian Analysis of the difference between face images. They are among the simplest Bayesian network models.

Naïve Bayes classifiers are a group of probabilistic classifiers based on Bayes' theorem with strong independence assumptions between the features. The assumptions may be naive at some point, presenting a drawback to the classifier performance. The next equation present the Bayes's theorem where theorem A and B are events and $P(B) \neq 0$:

$$P(A \mid B) = \frac{P(B \mid A)P(A))}{P(B))} \tag{2.2}$$

P$(A \mid B)$ represent a conditional probability: the probability of event A occurring given that B is certain. $P(B \mid A)P(B \mid A)$ is also a conditional: the likelihood of event B occurring given that A is certain. $P(A)$ and $P(B)$ are the marginal probabilities of observing A and B.

2.2.5 Texture based

Texture based algorithms extract textual features from face images by separate the face into several regions. Local Binary Pattern (LBP) is an example of Texture based algorithms, where LBP features are extracted to generate a feature vector.

The LBP operator was originally designed for texture description. The operator delivers a label to every pixel of the image by thresholding the 3x3 neighborhood of each pixel and considering the result as a binary value. The histogram of the labels can be used as a texture descriptor

T. Ahonen & A. Hadid & M. Pietikainen present at el. [13] a novel and efficient facial image representation based on LBP texture features where the face image is divided into several regions from while the LBP feature distributions are extracted and concatenated into an enhanced feature vector to be used as a face descriptor.

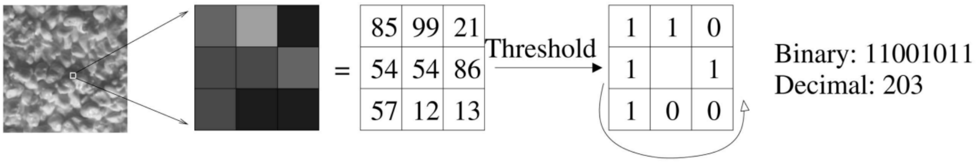


Figure 2.4: The basic LBP operator.

New FR systems have been proposed such as DeepID [14] for face verification was proposed by Yi Sun. The network input's size is 39×31×k for rectangle patches, and 31×31×k for square patches, where k = 3 for color patches and k = 1 for gray patches. The features are built on top of the feature extraction hierarchy of deep ConvNets and are summarized from multi-scale mid-level features. It is trained on the CelebFaces+ dataset, which contains 202,599 face images samples of 10,177 subjects. It achieved 97.45 percent verification accuracy with only weakly aligned faces on Labeled Faces in the Wild dataset. DeepID2 [14] and

DeepID2+ [15] proposed in the same year added verification supervisory signals to reduce interpersonal variations. DeepID3 networks proposed at 2015 rebuilt from basic elements of the VGGnet [16] and GoogLeNet. During training, joint face identification-verification supervisory signals are added to the final feature extraction layer as well as a few intermediate layers of each network. In order to learn a richer pool of facial features, weights in higher layers of some of DeepID3 networks are unshared. In the testing process, DeepID3 [17] is trained on the same dataset as DeepID2+. As a result, DeepID3 improves accuracy for face verification from 99.47 to 99.53 percent and rank-1 accuracy for face identification from 95.0 to 96.0 percent on Labeled Faces in the Wild dataset.

In 2015 VGG-Face was proposed by Omkar M. Parkhi et al. [18]. This model is trained with 2.6 million images of 2.6 thousand people for FR and verification. VGG-Face is based on the VGG-Very-Deep-16 CNN architecture to represent a face image as a vector of scores. The first eight blocks are convolutional layers followed by ReLU rectification layers and max-pooling layers, and the last three blocks are called Fully Connected layers also followed by ReLU rectification layers.

2.2.6 Neural Networks

FR techniques have shifted from traditional methods to deep learning methods in these years. Two different typologies of NN according with the directions where information flow inside the network.

A feed-forward network is a non-recurrent network where the signals can only travel in one direction. Input data is delivered onto a process layer. Each processing element makes its computation based upon a weighted sum of its inputs as previously described. The new calculated values then become the new input values that feed the next layer. This process continues until it has gone through all the layers and determines the output. A threshold transfer function is sometimes used to quantify the output of a neuron in the output layer.

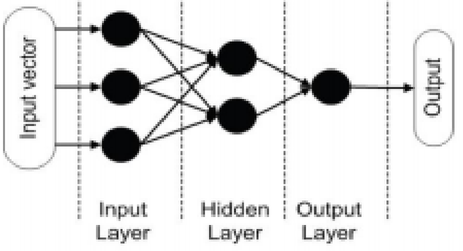


Figure 2.5: Simple feed forward topology where the information flows from inputs to outputs. Each black dot represents a single neuron.

On the other hand, Recurrent NN represent a simple and recurrent topology where some of the information flows in both directions. They are similar to feed forward NN with no limitations regarding back loops. In these cases, information is no longer transmitted only in one direction, but it is also transmitted backwards. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior.

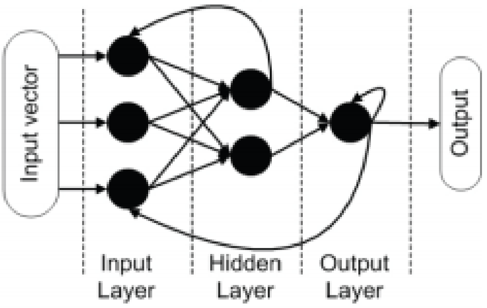


Figure 2.6: Simple recurrent NN topology where the information flows from inputs to outputs. Each black dot represents a single neuron.

2.2.7 Types of Artificial Neural Networks

In terms of architecture there are three main types of artificial NN's. They are Single and Multi-Layer Feed Forward Network, both following a feed-forward topology. The last one is Recurrent Network type. Other types of networks are Delta-Bar-Delta, Hopfield, Vector Quantization, Counter Propagation, Probabilistic, Hamming, Boltzman, Associative Memory, Spacio-Temporal Pattern, Adaptive Resonance, Self Organizing Map, Recirculation, among others [19].

Single Layer Feed Forward represent all the networks in which the input layer of source nodes is connected to an output layer of neurons. In this type of NN single layer is a reference to the output layer of computation nodes as the next figure illustrates:

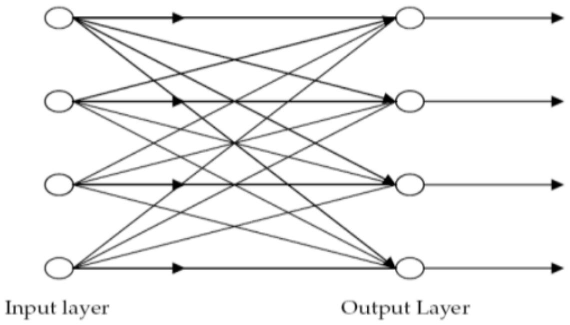


Figure 2.7: Single Layer Feed Forward Network.

The second type of network presented contains of at least one hidden layer whose nodes are denominated as hidden neurons. They have the function to interact between the external input and network. The output of the neurons in the output layer of network constitutes the overall response of network to the activation pattern supplied by source nodes in the first layer.

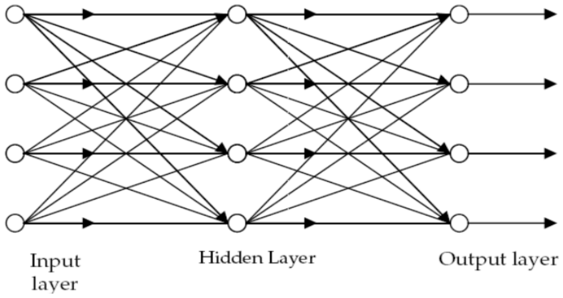


Figure 2.8: Multi-layer Feed Forward Network.

Recurrent Network are feed forward NN having one or more hidden layers with at least one feedback loop. The feedback may be a self feedback. In this case the output of neuron is given back to its own input.

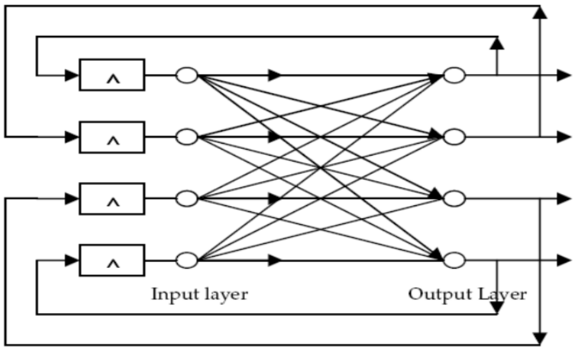


Figure 2.9: Recurrent Connected Network.

Beside the structure there are five characteristics of Artificial NN which are basic and important for this technology which are the ability of parallel processing, distributed memory, fault tolerance, ability collective solution and learning ability.

Beyond classify subjects NN can also be applied on the pre-processing stage by estimating depth from single images with DNN. In order to estimate monocular depth based on focal length Lei He propose on et al. [20] an effective NN to predict accurate depth, which achieves competitive performance as compared with the state of the art methods, and further embedding the focal length information into the proposed model. In addition, focal length is embedded in the network by the encoding mode.

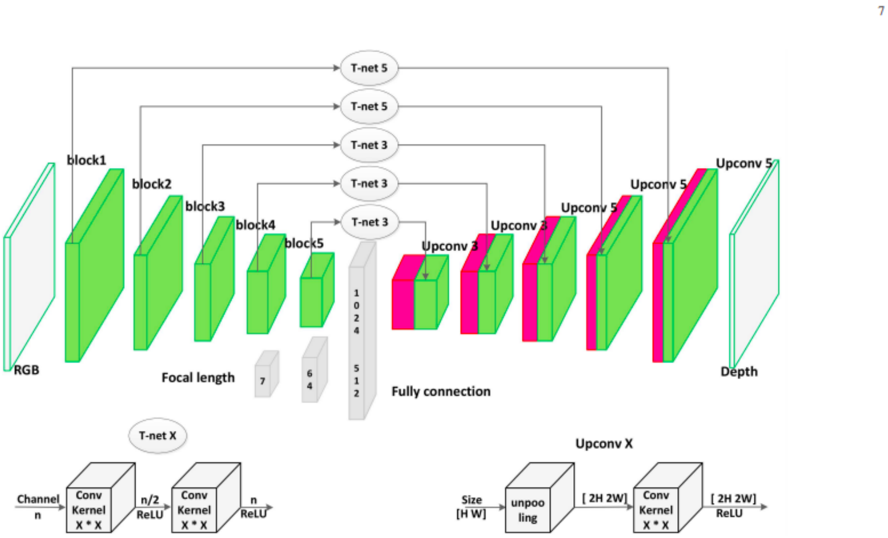


Figure 2.10: Lei He proposed network architecture.

The proposed network is composed of four parts: the first part is built on the pre-trained VGG models, followed by the global transformation layer and upsampling architecture to produce depth with high resolution, the third part effectively integrates the middle-level information to infer the structure details, converting the middle-level information to the space of the depth, and the last part embeds the focal length into the global information.

The deep learning success was recognized by the scientific community as consequence of the flattening of manifold-shaped data in higher layers of neural networks [21]. Euclidean distance miss to capture the relation between two points on a manifold. As a direct result high-dimensional data relay in the proximity of a low dimensional manifold. Deep learning enables the extraction of hidden variations factors and unfold manifold-shaped data.

By 2012, the deep convolutional NN structure AlexNet [22] was proposed by the University of Toronto team at the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). AlexNet is structured with five convolutional layers and three fully connected layers. It integrates multiple technologies such as data enhancement and ReLU which is linear for all positive values, and zero for the negative ones. Data enhancement includes create new test images by translations and horizontal reflections and altering the intensities of the RGB channels applied over the original training images.

Two years later GoogleNet [23] of Google Inc. introducing Inception Module with 22 layers achieved excellent results in ILSVRC, their structures are more complex in structure than AlexNet.

Due to the expense of data acquisition and costly annotation, it is very difficult to construct a large-scale dataset. The idea of transfer learning is overcoming the isolated learning methods and utilizing knowledge acquired for one task to solve similar ones.

Many details of how Deep Convolutional NN models work still remain a mystery. Matthew D. Zeiler and Rob Fergus [24], from New York University, let us aware that the lower layers are to capture generic features, while the higher ones learn source task specific features through deconvolution method in 2014.

3

Databases

Contents

3.1 Generic Database Characteristics

In order to train new FR algorithms that will be designed to validate the subjects the use of training and test samples present on databases is imperative. The results will improve proportionally depending on the number and variety of samples present in it.

This samples are available on multiple databases. An interpersonal variability database need to be carefully design in order to achieve a flawless FR program. Raphaël Weber, Catherine Soladié and Renaud Seguier on the VISAAP 2018 present a conference paper where the main points for a reliable database are debated [25].

Table 3.1 Characteristics of a database.

Category	Characteristic
. Population	# of subjects
	women/men %
	Age range
	Ethnic group(s)
. Modalities	Available modalities
. Data acquisition hardware	# of cameras
	Resolution
	FPS
. Experimental conditions	Background
	Lightning
	Occlusions
	Head pose
. Experimental protocol	Method of acquisition
	Available expressions
. Annotations	Facial features
	Action units (FACS)
	Emotional labels
	Emotional dimensions

Following the survey carried out by this FAST team members it is defined characteristics of population as the number of subjects, gender and ethnic group distribution and age range of the subjects. The choice of population influences the interpersonal variability: shape and texture of the face varies with identity, gender, age and ethnic group. The mean opening of the eyes differs between Asians and Caucasians is one illustrative example.

Modalities refer to the nature of the acquired signals. Databases can be distinguished according to the number of modalities: uni-modal vs. multi-modal (for two or more). Historically, the first databases are uni-modal with 2D video or images of the face. The available modalities are facial expression (2D or 3D), audio, body movement and physiological signals.

The focus here is the data acquisition hardware for image and video. The three characteristics consider are the number of cameras, camera resolution and frame per second. Experimental conditions include the background and lightning condition as well as head pose variation and occlusions cases.

Experimental protocol describes the expressive/emotional content available of the database and the method of acquisition of the samples from the subjects. Is possible to distinguish three kinds of databases: posed, spontaneous and in the wild where the experimental protocol varies from one kind to another.

The annotations are meta-data provided with the database giving low-level information (facial features or action units) or high-level (emotional labels or emotional dimensions) to FR system. The choice of annotations depends on the problem the database is meant to tackle since they will be used as ground truth as emotional labels are aimed at facial expression recognition and action units annotations are aimed at action units recognition and emotional dimensions are aimed at emotional dimension estimation.

Facial features such as facial landmarks or LBP could make a database more efficient since they may be used to quickly compute a specific point without computing it.

3.2 3D Databases Specifications

During the last decade research about FR has shifted from 2D to 3D face representations. The need for 3D data has resulted in various databases available for 3D FR and occasionally 3D expression analysis focused on recognition, containing a limited range of expressions and head poses. After the analysis carried out, it was determined that Bosphorus Database will be the one selected for comparing all the methods on our practical approaches. This database was analyzed in [26], a conference paper present in Biometrics and Identity Management in 2008.

Table 3.2 3D Databases content details

Database	Subjects	Samples	Total	Expressions	Pose	Occlusion
FRGC v.2	466	1-22	4007	Available	NA	NA
BU-3DFE	100	25	2500	Available	NA	NA
ND2006	888	1-63	13450	Available	NA	NA
YORK	350	15	5250	Available	Available	NA
CASIA	123	15	1845	Available	NA	NA
GavabDB	61	9	549	Available	Available	NA
3DRMA	120	6	720	NA	Available	NA
Bosphorus	105	31-53	4666	Available	Available	Available

Bosphorus is a wildly used database in 3D FR. Composed by 105 subjects, where 18 subjects have a large beard or mustache and 15 subjects with short facial hair. The majority of the subjects are aged between 25 and 35.

There are 60 men and 45 women in total, and most of the subjects are Caucasian. Also, 27 professional actors/actresses are incorporated into the database. Up to 54 face scans are available per subject, but 34 of these subjects have 31 scans. Thus, the number of total face scans is 4652 with each one been manually labeled for 24 facial landmark points such as nose tip, inner eye corners and some others.

3.3 Bosphorus Database

Bosphorus follow the basic emotions theory created by Paul Ekman. This theory assumes the existence of six discrete basic emotions. The emotions mention et al. [27] are anger, fear, disgust, surprise, joy and sadness at different levels of facial expression. Occlusions are also present, as previously mentioned, the types present are simulated with hands glasses and hair.

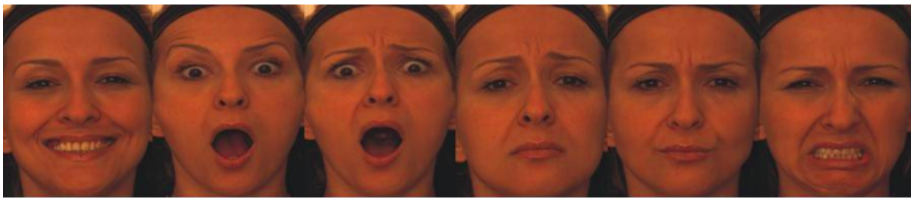


Figure 3.1: 6 emotions types present on Bosphorus.

For occlusion with glasses, multiple eyeglasses were used so that each subject could select at random one of them. Finally, if the subject's hair was long enough, their faces were also scanned with hair partly occluding the face. The subject to subject variation of occlusions is more pronounced when compared to expression variations as while one subject occludes his mouth with the whole hand, another one may occlude it with one finger.

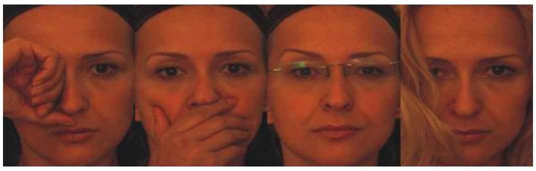


Figure 3.2: Occlusions examples present on Bosphorus.

Bosphorus contain 18 subjects with beard/moustache and short facial hair is available for 15 subjects. The majority of the subjects are aged between 25 and

35. There are 60 men and 45 women with most percentage of the subjects being Caucasian. Up to 54 face scans are available per subject, but 34 of these subjects have 31 scans. Thus, the number of total face samples is 4652.

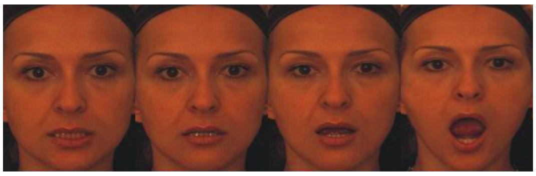


Figure 3.3: Lower face action units example with lower lip part highlighted present on the first article.

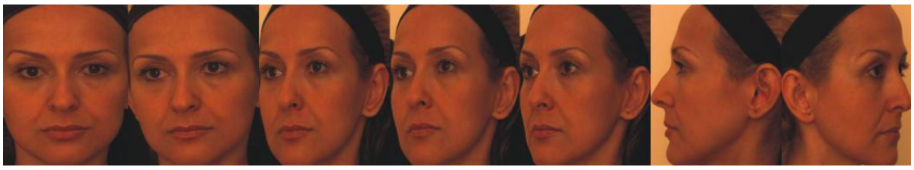


Figure 3.4: Head poses with rotations on yaw axis: +10°, +20°, +30°,+45°, +90°, -45° and -90° present on the first article.

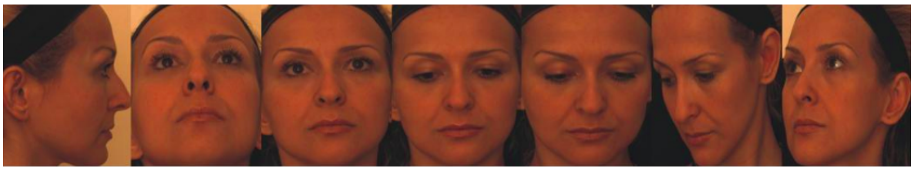


Figure 3.5: Head poses with rotations on pitch axis upwards, slight upwards, slight downwards, downwards; right-downwards and right-upwards present on the first article.

There are three types of head poses which correspond to seven yaw angles, four pitch angles, and two cross rotations which incorporate both yaw and pitch.

4

Methodology

Contents

After some consideration it was defined that an algorithm divided in 4 parts: pre-processing, feature extraction, training and classification will be implemented, following all the considerations mentioned until this point over the multiple program stages.

4.1 3D Methodology

A FR hybrid approach was designed in order to identify all the subjects present on Bosphorus facial database containing 3D face database that includes a rich set of expressions, systematic variation of poses and different types of occlusions presented. This database present facial expressions composed of judiciously selected subset of Action Units as well as six basic emotions simulation samples. Finally a rich set of head pose variations are available with different types of face occlusions included [28].

4.1.1 Pre-Processing

The quality of the acquired data is a critical point in a FR system. Due to 3D digitizing system and setup conditions, significant noise may occur as documented et al. [28]. To reduce noise the solution found was filtering with three different filters the data collected from the facial database. Commonly occurring problems during image acquisition and face reconstruction are noise due to movement and depth errors present on beard and eyebrows causing spiky noise.

4.1.1.A Filtering

In this work, the data collected is filtered in cascade by Savitzky–Golay, Gauss and bilateral filters. The next four models will represent the filter effect over the subject's 3D facial model. In the first sample it is possible to notice some of the previous errors mentioned in the previous chapter.

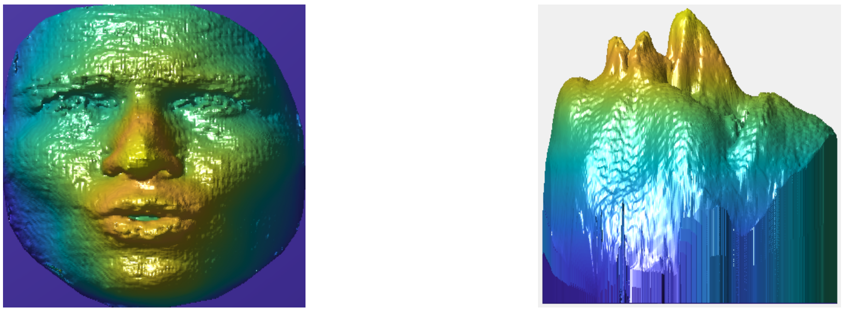


Figure 4.1: 3D generic original facial model.

The filter technique proposed in this thesis uses Savitzky–Golay filter in the first place. A Savitzky–Golay is a digital filter which can be applied to a set of digital data points to smooth the data while increasing the precision of the data without distorting the data values. This is achieved with a convolution by fitting successive sub-sets of adjacent data points with a low-degree polynomial by the method of linear least squares. When data points are equally spaced, an analytical solution to the least-squares equations can be found, in the form of a single set of ”convolution coefficients” that can be applied to all data sub-sets, to give estimates of the smoothed data values at the central point of each sub-set.

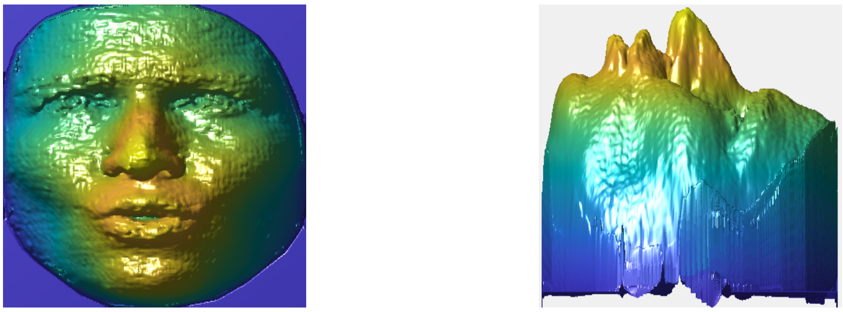


Figure 4.2: 3D facial model filtered with Savitzky–Golay Filter.

The second filter to be applied to the 3D model is the Gaussian. It is a widely used effect in computer graphics algorithms, typically to reduce image noise. The output model is the result of blurring a model by a 3D Gaussian smoothing kernel. Gaussian smoothing is also used as a pre-processing stage in computer vision algorithms in order to enhance image structures at different scales—see scale space representation and scale space implementation.

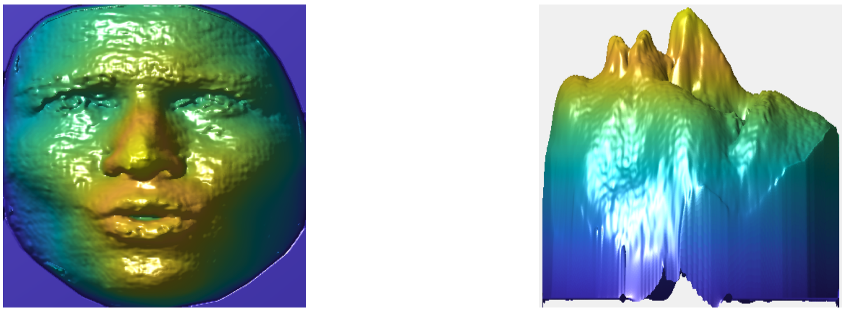


Figure 4.3: Golay and Gaussian Filter.

Bilateral filter is the last filter to be applied on this algorithm. It is a non-linear, edge-preserving and noise-reducing smoothing filter for images. Bilateral filter replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels preserving sharp edges. This weight can be based on a Gaussian distribution. Crucially, the weights depend not only on Euclidean distance of pixels, but also on the radiometric differences such as color intensity, depth distance among others.

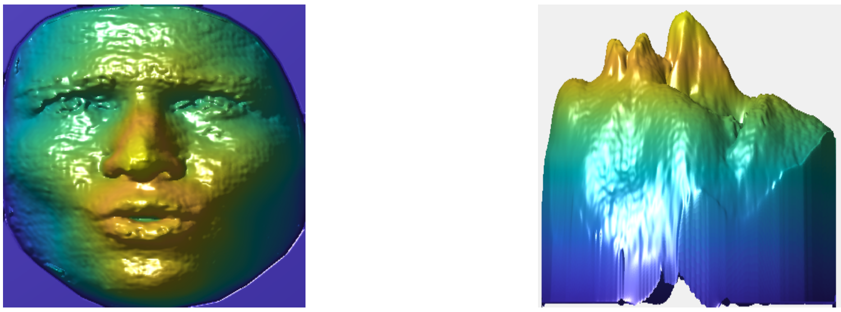


Figure 4.4: Golay + Gaussian + Bilateral Filters.

The bilateral filter follows this equation:

$$I^{\text{filtered}}(x) = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|) \tag{4.1}$$

where W_p, representing the normalize term, is defined as:

$$W_p = \sum_{x_i \in \Omega} f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|) \tag{4.2}$$

Although the filters used can cause changes on the data initially collected, we consider the advantages they present after calculating the new 3D models, such as the correction of certain three-dimensional models with facial hair, represent an added value for the FR system developed .

4.1.1.B Dataset Preparation

To have a better way to evaluate our 3D approach in a real-world scenario it was decided to expand the method on three different approaches instead of the original two initial idealized. The first one is a replica of the algorithm emphasizing the nasal zone [1]. The second methodology developed was an extension of the previously mentioned method for the entire facial region in order to evaluate this method as a global method.

Due to the most recent world pandemic crisis we decided to develop a new methodology. In order to simulate face occlusions derived from wearing a mask, models were manipulated in order to simulate a mask, denying the algorithm and subsequently to the NN potentially critical information for the classification of subjects and validation of their identity. Although Bosphorus Database already includes facial occlusions, these cases were not present in the dataset.

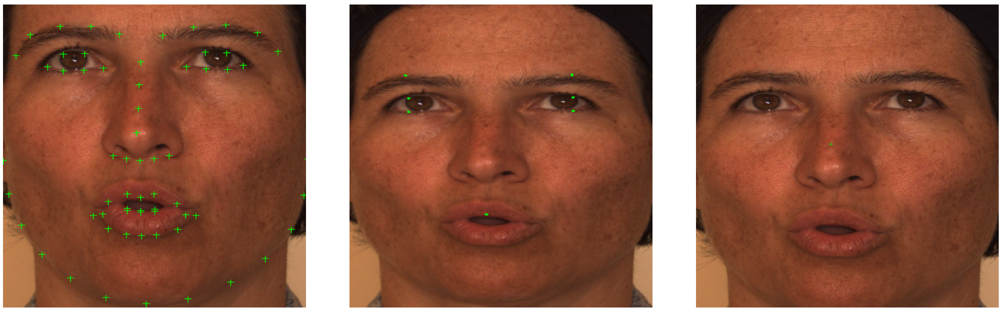


Figure 4.5: Facial landmarks detected by Dlib.

In order to remove unwanted facial regions a Dlib face detection and alignment algorithm adapt to MATLAB was applied. Dlib is a general purpose cross-platform software library written in C++. It is a cross-platform package for threading, networking, numerical operations, machine learning, computer vision, and compression, placing a strong emphasis on extremely high-quality and portable code developed by Davis E. King [29], where we took advantage of 68-point facial landmark classifier model to detect specific facial landmarks.

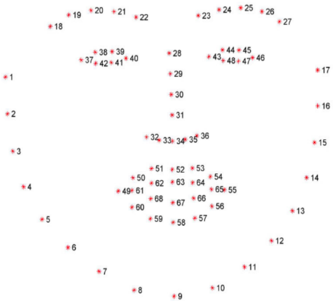


Figure 4.6: 68 facial landmark coordinates from the iBUG 300-W dataset.

The final image will be used as texture of the 3D facial model of the same subject, delimiting the 3D nose region that will be used on the feature extraction section.

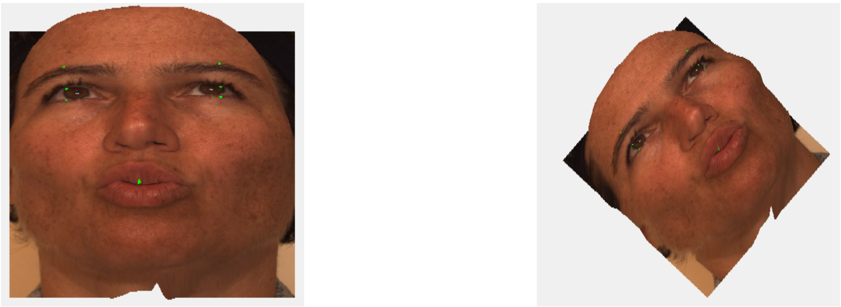


Figure 4.7: Texture applied over the 3D model.

The 3D models resultants of the three methodologies are exposed in the following figures. The first set of models is the complete facial model of the subjects. Is followed by the facial model that simulate region of interest of the example provided by Emambakhsh & Mehryar et al. [1]. The last group represent the mask models.

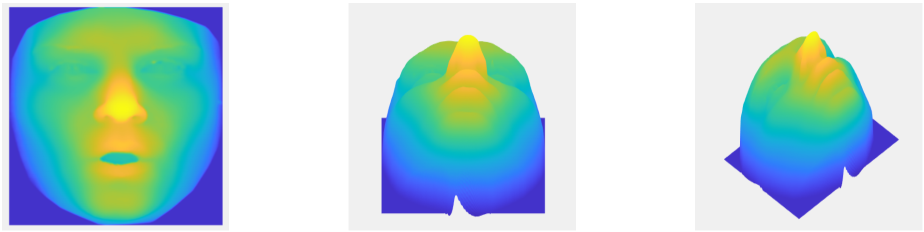
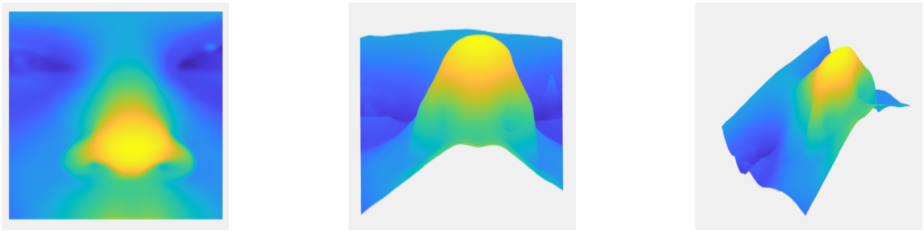


Figure 4.8: Complete Face Version 3D model overview.



(a) 3D Facial Model (Nose Region Version).

(b) Regular Grid Displayed over 3D Facial Model.

(c) Spheres centered over every grid point.

Figure 4.9: Nose Region Version 3D model overview.

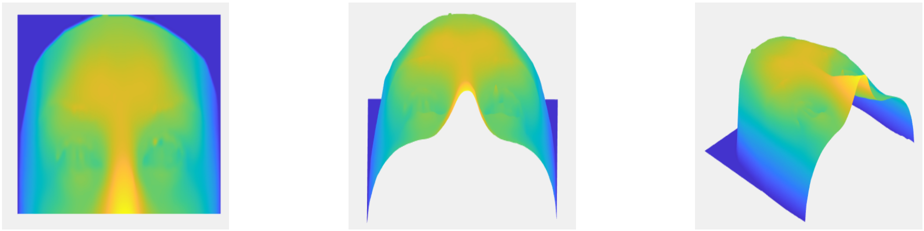


Figure 4.10: Ocular Version 3D model overview.

4.1.2 Feature Extractions

We used a similar approach of Emambakhsh & Mehryar article et al. [1] for the feature extraction based in the wavelet filter. The model face input was loaded using the function provided by Bosphorus, prepared to deal with RGB-D input data.

Based on previous work [1], Emambakhsh & Mehryar continue their survey, resulted in the article published in 2017, with a statement proposition of a new local-approach focused on the nose region for 3D FR et al. [30]. A coarsely nose tip location detected is the first goal. This step was described in the previous section, were an independent approach was designed.

The second step is to apply a regular grid to the resized 3D model. A set of spherical patches are localized over the nasal region to providing feature descriptors to the classification and validation algorithm. These feature descriptors provide the ability to evaluate the potential of overlapping spherical regions on the nasal surface, when used as feature vectors making them more robust against facial expressions.

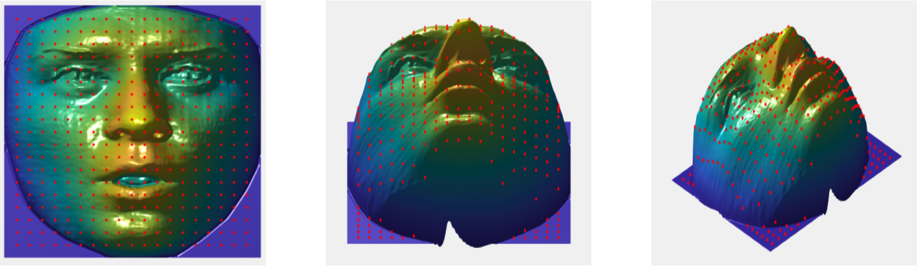


Figure 4.11: Regular Grid over complete facial model.

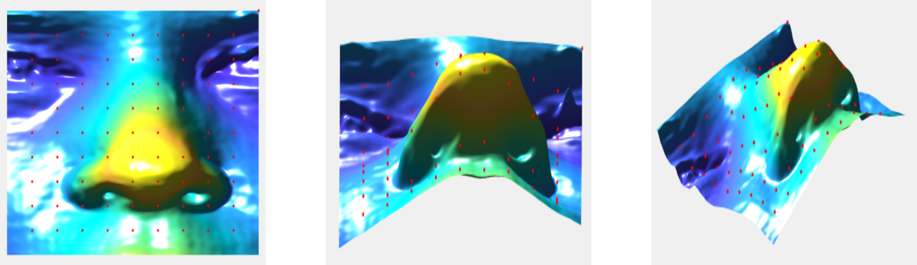


Figure 4.12: Regular Grid over the nasal region.

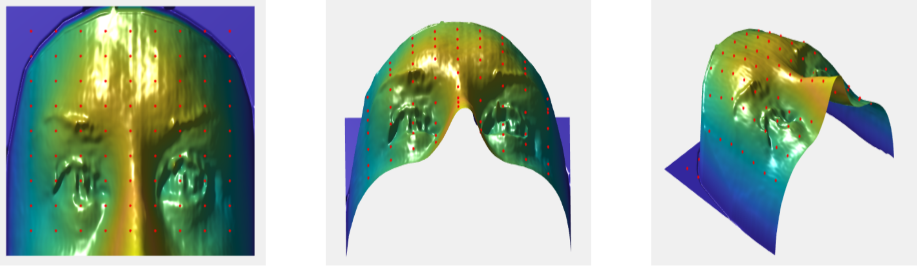


Figure 4.13: Regular Grid over the ocular region.

The feature space creation procedure is initialized by applying the wavelets in different orientations and scales. Subsequently all normals are computed on the maximum of absolute values of the filtered images per scale. At this stage, the feature descriptors are applied and normalized histograms are concatenated for all descriptors.

$$N = [N_x, N_y, N_z] \tag{4.3}$$

the normals are

$$n = [n_x, n_y, n_z] = \bigtriangledown N \tag{4.4}$$

where

$$n_x \circ n_x + n_y \circ n_y + n_z \circ n_z = 1 \tag{4.5}$$

with $\circ$ and 1 representing the Hadamard product operator that is a binary operator and a matrix of ones. In order to reduce noise sensitivity of the normal vectors and enable the extraction of multi-resolution directional region-based information from the nasal region, instead of calculating the normal vectors directly from the nose surface, they are derived from the Gabor wavelet filtered depth map.

The feature descriptors are used to define the region of interest containing a set of normal vectors from the Gabor wavelets filters. The resulting feature vectors histograms for the X, Y and Z maps are concatenated to create the feature space.

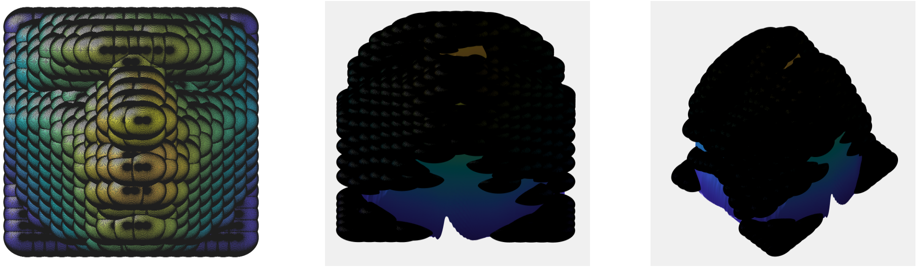


Figure 4.14: Spherical patches applied over the grid (first methodology).

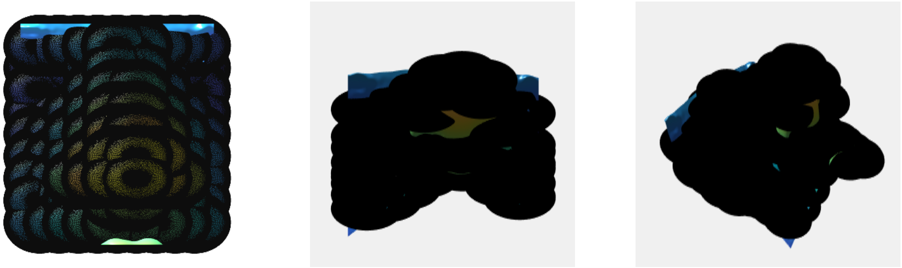


Figure 4.15: Spherical patches applied over the nose region cropped (second methodology).

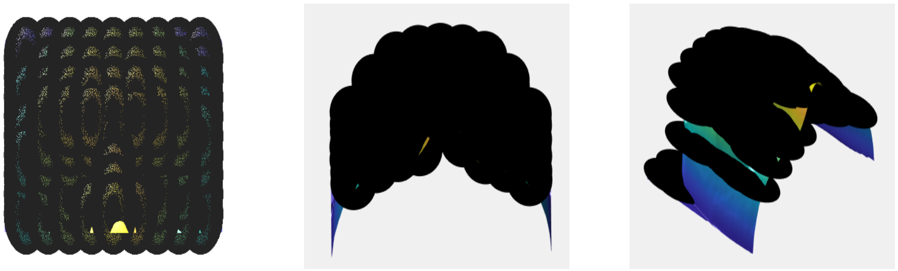


Figure 4.16: Spherical patches applied over the ocular region (third methodology).

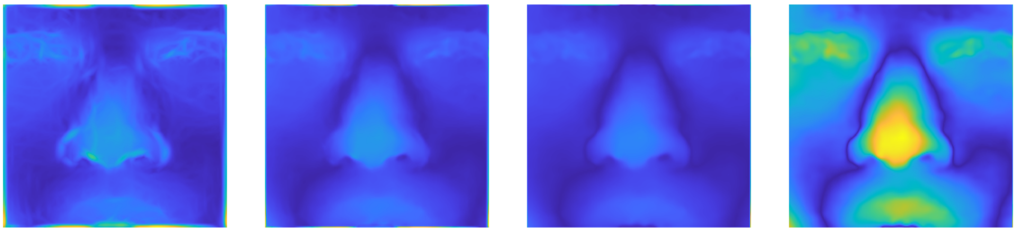
This approach selects the normals with maximal concentration of within-class scatter while at the same time maximize between-class distribution by using spherical patches as feature descriptors to extract histograms of the normal maps computed over the Gabor-wavelet images. Emambakhsh & Mehryar develop a function with the objective to compute the Gabor-wavelets using, with different orientations and scales, values of the 3D nose model represented by a $[M{*}N{*}3]$ matrix of multi-resolution directional region-based information, instead of calculating the normal vectors directly from the surface, they are derived from the Gabor wavelet filtered depth map et al. [31].

The discrete Fourier transform of the resampled Gabor wavelet $G_{s,o}$ for the s^{th} scale and o^{th} orientation level (s = 1, 2, ..., s_m and o = 1, 2, ..., o_m) is calculated and its zero frequency component is set to zero. The result Hadamard product of $G^f_{s,o}$ and the Fourier transform of N_z is calculated after and the absolute value of its inverse Fourier transform is calculated for each scale and orientation, i.e.

$$N^f_{z_{s,o}} = |\mathscr{F}^{-1}\{\mathscr{F} \circ G^f_{s,o}\}| \tag{4.6}$$

The maximum of all the corresponding elements of the filtered images is computed over all orientations for each scale s:

$$s : \{NGm_s | \forall_i, j, o : NGm_s(i,j) \geqslant N^f_{z_{s,o}}(i,j)\} \tag{4.7}$$



(a) Scale = 1.

(b) Scale = 2.

(c) Scale = 3.

(d) Scale = 4.

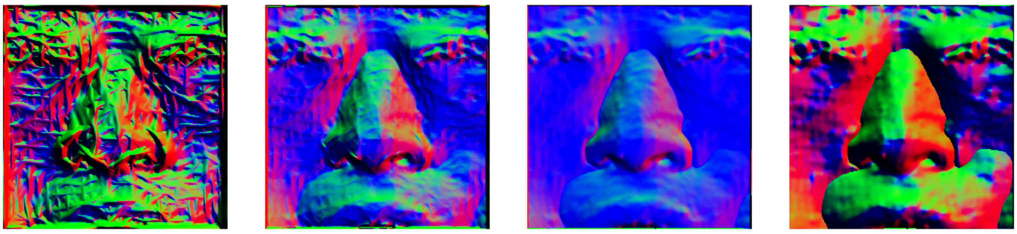
Figure 4.17: Gabor-wavelet images.

Finally, the normal vectors of the resulting per scale maximal map NGm_s is calculated using the aligned nose coordinate maps N_x and N_y,

$$\begin{cases} n_s = \bigtriangledown [N_x, N_y, NGm_s] \\ s = 1, 2, ..., s_m \end{cases} \tag{4.8}$$

with n_s = [N_{x_s}, N_{y_s}, N_{y_s}] representing a block matrix containing the normal vectors for the s^{th} scale level.

This specific function developed by Emambakhsh & Mehryar gets the matrices containing the horizontal and vertical resolution and computes the normal maps for each Gabor wavelet scale maps.



(a) Scale = 1. (b) Scale = 2. (c) Scale = 3. (d) Scale = 4.

Figure 4.18: Normal maps for each Gabor wavelet scale maps.

4.1.3 Feature Selection

The feature selection step selects subsets of feature vectors extracted more robust against facial expressions from the spherical patches. For a generic feature descriptor and n different Gabor wavelets scales s_1, s_2, . . . , s_m, the feature vector is calculated by

$$\begin{cases} F = [F_{s1}, F_{s2}, ..., F_{sn}], \\ F_{s_k} = \left[F_{x_{s_k}}, F_{y_{s_k}}, F_{z_{s_k}}\right], \end{cases} \tag{4.9}$$

where $F_{x_{s_k}}$, $F_{y_{s_k}}$ and $F_{z_{s_k}}$ are features of the s^k_{th} scale, for the x, y and z surface normal components. For K feature descriptors the feature set of the normal maps are represented by the concatenation of K different histograms, of length h_l from the feature descriptors, giving

$$\begin{cases} F_{x_{s_k}} = \left[H_{x_{1,sk}}, H_{x_{2,sk}}, ..., H_{x_{k,sk}}\right] \\ F_{y_{s_k}} = \left[H_{y_{1,sk}}, H_{y_{2,sk}}, ..., H_{y_{k,sk}}\right] \\ F_{z_{s_k}} = \left[H_{z_{1,sk}}, H_{z_{2,sk}}, ..., H_{z_{k,sk}}\right] \end{cases} \tag{4.10}$$

In previous equation, $H_{x_{i,sk}}$, $H_{y_{i,sk}}$ and $H_{z_{i,sk}}$ are the normalised histograms computed using the $i^t h$ feature descriptor (i = 1, ..., K) for the s^k_{th} scale (k = 1, ..., n) on the normal map $n_{s_k}^{th}$, which is computed using the previous equation.

4.2 2D Methodology

A 2D classic HOG based FR system was designed in order to understand how different local features methods performed in two dimensions FR and compare them with the 3D scores obtained in the first part of our practical methodology. The next figure shows the overall system design which covers the entire workflow that includes setting up of the database, feature extraction using HOG features, building up of classifier model and feature matching.

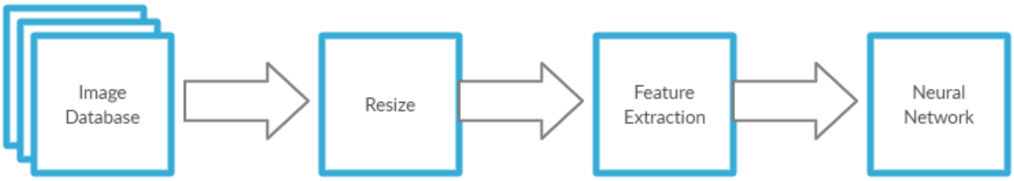


Figure 4.19: 2D FR pipeline.

The approaches to solve the FR problem have been diverse and numerous. HOG features are one of the famous handcrafted features extraction methods in use on 2D FR systems. It has proven to be an effective descriptor for pattern recognition in general and in human detect by N. Dalal et al. [32] and FR in particular [33] as demonstrated by O. Déniz. HOG descriptors are extracted from a regular grid to compensate for errors in face feature detection due to occlusions, pose and illumination variations.

In order to obtain uniform feature vectors and consequence of HOG features being variance to the input image size, an initial resize of all images set is necessary. The image dataset is composed with three different 2D models, simulating once more the tree cases that our survey explore: complete facial model, ocular or nasal region based. To simulate the mask model, in the 2D modality, a separate set of inputs was created. First a set of facial landmarks was computed using a Dlib face detection algorithm and with this set of points a polygon was created. The last part of the algorithm is responsible for fill up the area marked. The next figure illustrate the mask simulation process of the algorithm.

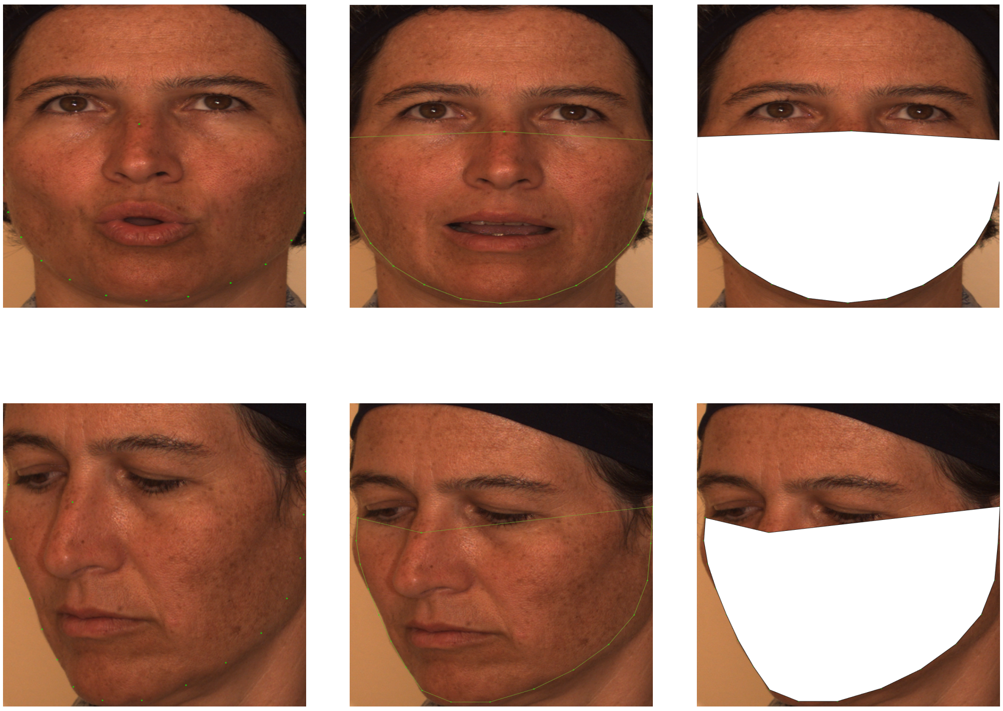


Figure 4.21: Mask training subjects creation process.

The pre-processing stage of the FR 2D version system is finished when all dataset is resized as consequence of HOG features being proportional variant to different images sizes.

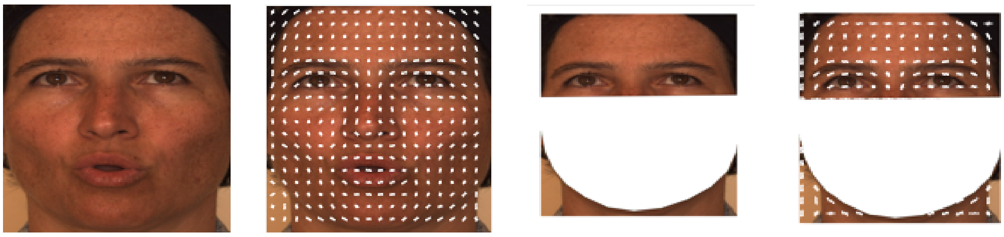


Figure 4.22: HOG feature localization.

HOG features are based in the sum of gradient directions over the pixel of a small spatial region and in the subsequent construction of a 1D histogram whose concatenation supplies the features vector to be considered for further purposes. Let L be the intensity of gray-scale level function describing the image analyzed. The image is divided into cells of size $N \times N$ pixels and the orientation $\theta_{x,y}$ of the gradient in each pixel is computed by the following equation:

$$\theta_{x,y} = tan^{-1}\left(\frac{L(x,y+1)-L(x,y-1)}{L(x,y+1)-L(x,y-1)}\right) \tag{4.11}$$

The orientation of all pixels is computed and accumulated in an M-bins histogram of orientations. Finally, all cell histograms are concatenated in order to create the final features vector that will be used by the NN.

4.3 Training and Process

Classification is defined by A. Hajraoui et al. [34] as the step that enable the classification of the feature vector of the person to recognize. It is treatment requires the introduction of a comparison algorithm or classification which provides at its output a score of similarity or distance between this characteristic vector and the reference features vector of the database. This score is compared subsequently to a decision threshold fixed in advance to provide a final decision on identity.

The automatic data classification is a branch of the data analysis which has resulted in numerous and diverse algorithms. It is used to group data into classes so that the data of the same class is as similar as possible and the classes are the most distinct possible.

Although the objective of this research is not to exhaustively present the existing methods for multi-class classification in the context of FR, the main methodologies will briefly present in the next sub-chapter.

4.3.1 Neural Network

As previously mentioned a NN is a complex combination of basic objects called formal neurons. These have an activation function that allows to influence other neurons. The connections between the neurons, which is called synaptic connections, spread the activity of neurons with a characteristic weighting connection. With their machine learning capability from data modeling the problem to solve NN's have proven themselves as proficient classifiers and are particularly well suited for addressing non-linear problems. Given the non-linear nature of real-world phenomena, like processing chain of automatic face recognition system, NN's are certainly a good candidate for solving the problem.

4.3.2 Preparing the Data

When training multi-layer NN, the general approach is to first divide the data into three subsets. The first subset is the training set representing 70 percent of the overall subjects, which is used for computing the gradient and updating the network weights and biases. These are presented to the NN during training and the NN is adjusted according to its error.

The second subset is the validation set representing 15 percent of the overall subjects, used to measure the NN generalization and to halt training when generalization stops improving. This subgroup is presented to the NN during training and the NN is adjusted according to its error. The validation error normally decreases during the initial phase of training, as does the training set error. However, when the NN begins to overfit the data, the error on the validation set typically begins to rise. The NN weights and biases are saved at the minimum of the validation set error.

The test subset represent 15 percent of the overall subjects. It is not used during training process, but it is used to compare different models by measuring the NN performance during and after training. It is a completely independent test of NN generalization.

4.3.3 Building the Neural Network Classifier

After the data has been collected and divided the next step is to create a NN that will learn to identify and classify all of 104 subjects. A feed-forward NN that depend on a softmax activation function in an output layer that assigns a probability for each class, was designed to accomplish this objective. Feed-forward NN is a NN where the connections between the nodes do not form cycles or loops so the information only flow in one direction, from the input nodes forward through hidden nodes and finally to the output nodes.

A multi-layer perceptron is a subclass of feed-forward artificial NN consisting of at least three layers sets of nodes where the first correspond to an input layer, followed by a variant number of hidden layers that converge all the data over an output layer who is represented at the last set. Except for the input nodes, each node is a neuron that uses a nonlinear activation function.

To train the customized NN by updating weight and bias values it was used as a resource a Scaled Conjugate Gradient which is a supervised learning algorithm for feed-forward NN based on the idea to combine the model-trust region approach with the conjugate gradient approach. It is a training algorithm representing the class of conjugate gradient methods, developed by Moller [35] and was designed to avoid the time-consuming line search at each iteration

For classification problems the softmax layer and then a classification layer must follow the final fully connected layer. The output unit activation function is the softmax function:

$$y_r(x) = \frac{exp(a_r(x))}{\sum_{j=1}^{k} exp(a_j(x))} \tag{4.12}$$

where $0 \leq y_r \leq 1$ and $\sum_{j=1}^{k} y_j = 1$.

The softmax function is the output unit activation function after the last fully connected layer for multi-class classification problems:

$$P(c_r|x,\theta) = \frac{P(c_r|x,\theta)P(c_r)}{\sum_{j=1}^{k} P(c_r|x,\theta)P(c_r)} = \frac{exp(a_r(x,\theta))}{\sum_{j=1}^{k} exp(a_r(x,\theta))} \tag{4.13}$$

where $0 \leq P(c_r|x,\theta) \leq 1$ and $\sum_{j=1}^{k} P(c_r|x,\theta) = 1$.

This training routine may require more iterations to converge when compared with other conjugate gradient methods but the number of computations in each iteration is significantly reduced because no line search is performed. Also, the conjugate gradient method require only a little more storage than simpler training algorithms, so they are often chose for networks with a large number of weights.

5

Experiments and Discussion

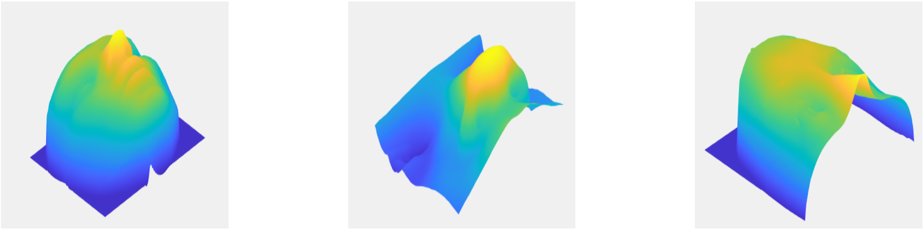
Contents

This section explains first how the test environment was set up followed by the experimental observations recorded depending on the 3D or 2D modality and configuration of the NN. In order to prove the importance of the nasal zone and then the ocular region in 3D FR system, a multiple dimension analyses was performed with various NN designs, with the goal to understand how every process stage can influence our overall success to validate a subject identity.

The first test were performed in order to understand our method performance in three dimensions, with the NN input sets being exclusive to one of the three modalities presented in the previous chapter. In the second section is presented the 2D research of our FR system with two different NN training modes with the first test performed under the same circumstances of the 3D modality and the last ones with different percentages of mask wearing subjects present in the training set.

5.1 3D Experimental Results

The first experimental results are focused on the nasal region with the intention to simulate and reproduce the Emambakhsh & Evans thesis in [1]. In this article it is proposed that the nasal region can be used to extract features with similar success level when compared with algorithms that use the complete face model. The second model represent the complete 3D model of the face. In this specific case the algorithm proposed on [1] was applied over the entire face. The third example highlight the ocular region. The purpose of this experiment is to quantitatively evaluate the performance of a FR system based on the ocular region when compared with the first two models.



(a) Complete Face Version 3D model overview

(b) Regular Grid Displayed over 3D Facial Model.

(c) Ocular Version 3D model overview.

Figure 5.1: Nose Region Version 3D model overview.

Multiple configurations were applied to the neuronal network with the numbers of neurons varying between 100 and 1000 and the layers 5 and 10 in order to understand the best possible design. Finally, the number of subjects was changed in order to understand if in any way the network could be affected by an over fitting of input data.

The FR system evaluation method used to compare multiple NN designs performance is the percentage fraction values of correct classified NN predictions of the subjects identities. The experimental results are summarized in Tables 5.1 to 5.4, detailing the portion of subjects correctly classified by the FR system.

Table 5.1 Experimental results for **105 and 90 subjects**

		105 Subjects			90 Subjects		
Neurons	Layers	Nose	All Face	Ocular	Nose	All Face	Ocular
100	5	91.92	90.53	**92.19**	**92.36**	90.78	91.71
100	10	**90.36**	87.13	90.00	**89.96**	85.38	87.38
1000	5	**93.34**	90.35	91.59	89.88	**91.33**	89.86
1000	10	91.61	89.48	**92.21**	91.1	90.75	**91.42**

An evaluation of the tables shows first that the nose region can be used as effectively as the whole 3D face model since it is noticeable that in front of the battery of tests performed it has a higher success rate in most test cases, confirming of Emambakhsh & Evans cited article conclusion and highlighting the importance of the nasal region in face of 3D FR systems.

Table 5.2 Experimental results for **75 and 60 subjects**

		75 Subjects			60 Subjects		
Neurons	Layers	Nose	All Face	Ocular	Nose	All Face	Ocular
100	5	**93.41**	92.31	92.49	**93.32**	92.10	92.89
100	10	**90.89**	84.95	89.93	**91.97**	89.11	90.79
1000	5	**91.78**	89.12	91.23	91.43	89.66	**92.05**
1000	10	91.10	85.54	**91.89**	91.80	91.64	**92.58**

The analysis of the 3D methodology was finished with the introduction of the ocular model. As initially mentioned, this specific model was added in order to understand how the NN would react to the fact that the nasal zone and a large percentage of the face are hidden, and also in order to understand how important is the ocular zone to 3D algorithms. An evaluation of tables of experiments shows that, for the majority of the NN configurations, better results were achieved with ocular zone than with all face model.

When compared to the nasal model, the ocular region presents a similar performance, even superior in some cases. This fact is a result of the variation of the model's texture and depth in the ocular zone as it happens with the nasal zone.

Table 5.3 Experimental results for **45 and 30 subjects**

		45 Subjects			30 Subjects		
Neurons	Layers	Nose	All Face	Ocular	Nose	All Face	Ocular
100	5	**94.82**	93.69	94.47	95.36	94.07	**96.14**
100	10	93.32	92.94	**93.87**	95.14	93.47	**95.45**
1000	5	**94.60**	91.87	92.77	**94.53**	94.21	88.11
1000	10	**93.44**	92.63	92.57	**93.80**	93.00	93.25

As a consequence of the evaluation of the ablation studies we performed, it can be observed that the best possible configuration is close to 100 neurons in 5 layers. This result was consistently obtained for the different 3D modalities and NN configurations. Such conclusion is based on the fact that in 21 tests performed, 17 have been with this configuration, corresponding approximately to a total of 80.95 percent of the tests presented in this chapter.

The number of subjects and the way they influence the system was also a point of interest during this research. With the view to understand the effect that different populations of tests have on the NN, we reduced the subjects by 15 for each test set, corresponding to an approximate variation of minus 15 percent of the original dataset size in each set.

Table 5.4 Experimental results for **15 subjects**

		15 Subjects		
Neurons	Layers	Nose	All Face	Ocular
100	5	**96.89**	94.44	96.29
100	10	95.81	94.03	**96.29**
1000	5	**96.51**	94.31	95.21
1000	10	95.35	92.82	**96.29**

It is in the last table that the best results are found in 11 of the 12 possible configurations of the NN when comparing the different subjects tested. This percentage values, representing the percentage of samples correctly classified, could be result of overfitting caused due to an overly complex model with too many parameters.

5.2 2D Experimental Results

After performing the 3D analysis of our practical approach a 2D FR system was designed in order to have a 2D performance baseline to compare with the first results. As stated before it is based in 2D HOG features FR system using the same NN configurations presented in the previous section. Another resemblance when compared with the first experiment is the three 2D model used in this FR system version simulating again the complete model of the face followed by a nose region and an ocular zone approaches.

Table 5.5 Experimental results for **105 and 90 subjects**

		105 Subjects			90 Subjects		
Neurons	Layers	Nose	All Face	Ocular	Nose	All Face	Ocular
100	5	**96.02**	91.77	93.38	**96.72**	91.88	93.1
100	10	**94.94**	90.12	91.73	**94.61**	87.04	89.53
1000	5	**94.39**	90.59	92.94	**96.78**	90.61	91.9
1000	1000	**94.71**	92.22	92.42	**94.97**	91.83	92.35

In the 2D test set contrary to what happened in the 3D experiments, with the best results scored being split between the nose and ocular modalities, the better scores are present every time in the nose region dedicated modality, demonstrating again the importance of this specific region in FR systems in both dimensions.

Table 5.6 Experimental results for **60 and 30 subjects**

		60 Subjects			30 Subjects		
Neurons	Layers	Nose	All Face	Ocular	Nose	All Face	Ocular
100	5	**96.76**	93.73	94.52	**97.9**	94.71	97.21
100	10	**96.12**	88.84	93.36	**97.62**	93.06	94.58
1000	5	**96.76**	91.28	92.69	**96.97**	92.61	94.12
1000	1000	**96.58**	90.99	91.11	**94.95**	93.47	89.62

The new analysis of the same ablation studies previously performed show once more that the best possible configuration is close to 100 neurons in 5 layers with 11 of the 15 different tests consistently reaching a better score in this NN configuration.

When the number of subjects and its influence over 2D FR systems was considered nose regions and complete face models show once again the best performance for 15 subjects while the ocular zone shifted his better score for the 30 subjects case.

Table 5.7 Experimental results for **15 subjects**

		15 Subjects		
Neurons	Layers	Nose	All Face	Ocular
100	5	**97.78**	95.46	97.17
100	10	**98.26**	94.47	96.06
1000	5	**97.62**	95.17	96.69
1000	1000	**97.31**	94.47	90.24

Differently to our initial expectation, the best performance was achieved by using the 2D face images when the subjects wear face masks when compared with 2D face images with the all face, as it was expected after our literature review that face mask test subjects would worsen the results achieved by generic FR systems.

5.3 Further 2D Experiments

In order to understand why our NN presented such scores a new test set was designed. This time the training population was changed to understand how different NN training could lead to different scores. This time the training population start just with complete 2D face models to be applied both on mask and all face 2D models. During the follow-up of the tests, we reduced the percentage of complete 2D facial models' subjects used for training the NN by 25 percent while mask models were added to the training set in the same proportion. The test population used is exclusively 100 percent of All Face or Mask models. The results are present in the next tables.

Table 5.8 Experimental results

			Success	
Training	Neurons	Layers	Test All Face	Test Mask
100% All Face	100	5	91.77	**94.79**
	100	10	90.12	**93.12**
	1000	5	90.59	**94.23**
	1000	10	92.22	**94.41**
75% All Face + 25% Mask	100	5	96.54	**96.83**
	100	10	**95.67**	94.82
	1000	5	**96.75**	96.49
	1000	10	**96.54**	96.49
50% All Face + 50% Mask	100	5	**95.99**	92.64
	100	10	**95.02**	90.73
	1000	5	**94.09**	92.06
	1000	10	**95.63**	92.16
25% All Face + 75% Mask	100	5	**92.29**	85.27
	100	10	**89.35**	16.18
	1000	5	**92.00**	85.08
	1000	10	**91.41**	85.69

Once again the best results are presented in the Mask modality, when the training population is composed with only All Face models and for the first case of training population being constituted by 75% of complete 2D facial model and 25% of subjects with simulated masks. However, this time a change occurs for the last test cases with the best results to be presented in the column where the test subjects are composed of complete 2D facial models.

On an overall analysis over these tests it can be stated that the bests results are presented with the All Face test population as initially was expected although the results presented in the initial part of the last tests show a better rate according to the initial 2D tests. This point will be further analyzed in the next chapter.

After finishing the set of previous tests, two questions continued over the influence of the size of the NN and on whether it would be able to extract any more non-intuitive features from the masked 2D model that could be altering the results initially envisioned. With this objective in mind, we changed our FR algorithm at two different points: A new kind of simulated mask was adapted to the models present in Bosphorus dataset, as shown in the next image.

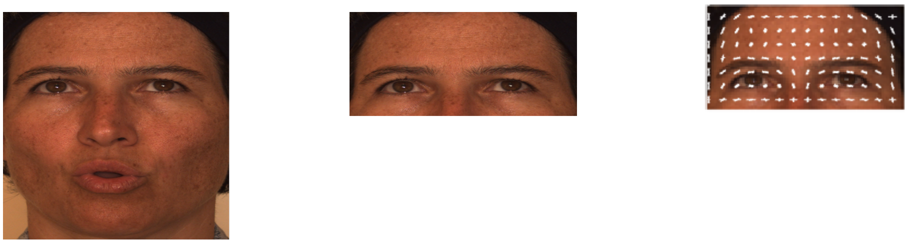


Figure 5.2: HOG feature localization on new 2D model.

With this model it was hopped to exclude the hypotheses of non-intuitive features being computed by the NN. To resolve the first point, our NN was resized so that it is possible to exclude network overfitting since 2D data is smaller in size when compared with the 3D case, as a possible cause of unexpected values in the success rate for the two modalities under analysis. This time the NN vary in size between 1 to 5 layers and between 10 to 100 neurons. The success rate is detailed in the next tables.

Table 5.9 Experimental results for training with **100% All Face** 2D models.

			Success	
Training	Layers	Neurons	Test All Face	Test Mask
100% All Face	1	10	71.97	**80.78**
	3	10	62.51	**74.74**
	5	10	1.97	**5.32**
	1	50	94.02	**96.40**
	3	50	91.86	**95.21**
	5	50	89.69	**93.68**
	1	100	94.75	**96.54**
	3	100	93.59	**95.05**
	5	100	92.93	**94.87**

Once again the best results are presented in the Mask modality, when the training population is composed with only complete 2D facial models and for some punctual cases on the next configurations presented, confirming the results observed in the initial mask simulation version and NN design.

Table 5.10 Experimental results for training with **75% All Face and 25% Mask** 2D models.

			Success	
Training	Layers	Neurons	Test All Face	Test Mask
75% All Face + 25% Mask	1	10	77.89	**89.41**
	3	10	78.42	**83.32**
	5	10	**37.54**	31.83
	1	50	**98.14**	97.60
	3	50	**97.21**	96.89
	5	50	**96.41**	94.13
	1	100	98.06	**98.09**
	3	100	**97.82**	97.71
	5	100	**97.58**	97.19

Another parallel analysis between the extra 2D experimental results, we can see that again the best results are present for the training combination composed of 75% of the subjects with the full face and 25% who use a mask, reinforcing the importance of training diversified even when applied to specific test cases.

Table 5.11 Experimental results for training with **50% All Face and 50% Mask** 2D models.

			Success	
Training	Layers	Neurons	Test All Face	Test Mask
50% All Face + 50% Mask	1	10	**91.51**	72.76
	3	10	**52.33**	38.45
	5	10	**26.96**	11.07
	1	50	**96.83**	94.86
	3	50	**95.41**	93.44
	5	50	**95.30**	80.10
	1	100	**96.72**	95.45
	3	100	**96.79**	94.30
	5	100	**96.32**	91.96

One more point of agreement between both analyzes was the deterioration of performance as number of hidden layers is increased while, on the other hand, when comparing different numbers of neurons with an equal number of layers it is observed that the FR system improves the success rate as neurons are added.

Table 5.12 Experimental results for training with **25% All Face and 75% Mask** 2D models.

			Success	
Training	Layers	Neurons	Test All Face	Test Mask
25% All Face + 75% Mask	1	10	**85.12**	56.80
	3	10	10.47	**17.95**
	5	10	**18.23**	2.00
	1	50	**94.47**	91.72
	3	50	**91.82**	88.34
	5	50	**90.47**	75.11
	1	100	**94.59**	91.42
	3	100	**93.47**	91.18
	5	100	**92.53**	87.13

As it is possible to observe in the previous tables in a general evaluation the results agree with the first set of extra 2D experiments, being possible to exclude the possibility of the network overfitting influence performance. As for the mask case, we can state that it is not the shape of the mask that may be manipulating the success results, since in the examples where both were tested with the same NN design [100 Neurons with 5 Layers], 3 in 4 cases the value was not lower. Finally, if we take the color into account, future tests should be carried out to understand how it can influence the system. Some of these details will be debated in the next chapter.

6

Conclusions

Contents

This thesis addresses a modified version of the Emambakhsh & Evans algorithm [1], with the aim of questioning the importance of the nasal zone in 3D algorithms first and reproduce the results that confirm the statement proposed in the article.

With the course of the research and taking into account the current context of pandemic it was decided to expand our objectives. Although we have effectively confirmed the importance of the nose region in FR with the widespread use of the mask, most of it is obstructed by face coverings for data collection instruments. The solution found to fit this new pandemic reality was to shift the focus from the nasal area to the ocular area.

This new approach proved that it can compete with the approach initially tested, both in 2D and 3D, managing to compete with the nasal zone in terms of percentage of success with the advantage that it can be used today even if the subject to be tested has a mask.

As referred in the end of the last chapter contrary to what it was initially expected, NN with test populations composed exclusively with subjects wearing simulated masks should present a lower score when compared with complete 2D models tests. As presented in a recent report of NIST [36] masked subjects can influence in some cases the overall performance of the FR system when it is taken in account different mask models and mask colors.

It is also important to mention that the NN applied to the 2D features extracted was the same as in the 3D case in order to maintain the same specifications in first two tests although in the 2D test it could lead in some cases the NN to overfit and not reach the best possible results in unseen data, as the feature vectors were considerably smaller.

A feature that we wanted to test was the scalability of our FR system. Given this idea, we can say that the difference found between 15 subjects and 105 is significant but it is not expressive. Furthermore, when we took in consideration the cases where the number of neurons is equal to 1000 the difference stays between 3 to 4 points in the 3D scenario and 2 to 3 points for the 2D.

Although the overall results were satisfactory and according to the theories initially conceived, it is not possible to hide that the percentage of success would be expected to improve when applied to a database with a higher ratio of samples per subject as some FR algorithms success rate are based trained and tested models using datasets several times larger than Bosphorus dataset, furthermore if taking into account the fact that the tests involve models containing facial occlusions and expressions changes, escaping the stigma of trained and tested FR systems in ideal environment conditions.

6.1 Future Work

Although some interesting results were achieved with the 3D results it should be noted that these tests were performed separately by the NN, that is, the network only tests the modality for which it was trained.

It would be interesting to complete the tests where the network would train with a modality and tested against a different set of 3D models in order to prove whether the current FR systems could be reused, or whether they would have to undergo a new training phase with 3D models adapted to the current pandemic reality as a consequence of COVID-19.

A detail worthy of future interest is that it was only tested for the two versions of the 2D simulated masks, in the sub-chapter 5.3, with an equal design of the NN, in order to form a more detailed conclusion on FR systems that ignore the hidden area as consequence of mask cover or take advantage of the jaw line for features extraction stage, and if possible expand the research to 3 dimensions.

More than testing the adaptability of systems to changes in routines caused by recent pandemic events, the different masks must undergo an analysis process by the FR algorithms, since different designs and colors can lead to variable results.

Another suggestion for future work is the construction of a larger and robust comparative model between 2D and 3D algorithms in order to register the benefits that we can derive from this new technology in comparison with the classic methods currently implemented.

Currently, FR is in exponential evolution. We can say that machine learning algorithms are largely responsible for this upgrade in the results obtained. If possible, to complete the set of research on FR systems, start testing with different machine learning algorithms to understand what is the best method we can use taking into account parameters such as accuracy and speed.

Bibliography

[1] M. Emambakhsh and A. Evans, "Nasal patches and curves for expression-robust 3d face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 995–1007, 2017.

[2] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 10 2018, pp. 471–478.

[3] Y. Zeng, E. Lu, Y. Sun, and R. Tian, "Responsible facial recognition and beyond," 2019.

[4] S. Kak, F. Mustafa, and P. Valente, "A review of person recognition based on face model," vol. 4, pp. 157–168, 01 2018.

[5] U. P. . U. of Exeter, *Facial Recognition Technology Market Research*, 2019 (accessed May 30, 2020). [Online]. Available: https://www.unlocking-potential.co.uk/wp-content/uploads/2019/06/Facial-Recognition-Technology-Market-Research.pdf

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, pp. 886–893 vol. 1.

[7] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America. A, Optics and image science*, vol. 4, pp. 519–24, 04 1987.

[8] Q. Wang, D. Xiong, A. Alfalou, and C. Brosseau, "Optical image authentication scheme using dual polarization decoding configuration," *Optics and Lasers in Engineering*, vol. 112, pp. 151 – 161, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0143816618309357

[9] M. Annalakshmi, S. Roomi, and A. Naveedh, “A hybrid technique for gender classification with slbp and hog features,” *Cluster Computing*, vol. 22, 01 2019.

[10] V. A, D. Hebbar, V. S. Shekhar, K. N. B. Murthy, and S. Natarajan, “Two novel detector-descriptor based approaches for face recognition using sift and surf,” *Procedia Computer Science*, vol. 70, pp. 185 – 197, 2015, proceedings of the 4th International Conference on Eco-friendly Computing and Communication Systems. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050915032342

[11] T. Napoléon and A. Alfalou, “Pose invariant face recognition: 3d model from single photo,” *Optics and Lasers in Engineering*, vol. 89, 07 2016.

[12] B. Jin, L. Cruz, and N. Gonçalves, “Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis,” *IEEE Access*, vol. 8, pp. 123 649–123 661, 2020.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, “Face description with local binary patterns: Application to face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[14] Y. Sun, X. Wang, and X. Tang, “Deep learning face representation from predicting 10,000 classes,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1891–1898.

[15] Y. Sun, X. Wang, and X. Tang, “Deep learning face representation by joint identification-verification,” 2014.

[16] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.

[17] Y. Sun, D. Liang, X. Wang, and X. Tang, “Deepid3: Face recognition with very deep neural networks,” 2015.

[18] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition,” in *BMVC*, 2015.

[19] M. Kasar, D. Bhattacharyya, and T.-H. Kim, “Face recognition using neural network: A review,” *International Journal of Security and Its Applications*, vol. 10, pp. 81–100, 03 2016.

[20] L. He, G. Wang, and Z. Hu, “Learning depth from single images with deep neural network embedding focal length,” *IEEE Transactions on Image Processing*, vol. PP, 03 2018.

[21] P. P. Brahma, D. Wu, and Y. She, “Why deep learning works: A manifold disentanglement perspective,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 10, pp. 1997–2008, 2016.

[22] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” *Neural Information Processing Systems*, vol. 25, 01 2012.

[23] C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[24] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional networks,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 818–833.

[25] R. Weber, C. Soladié, and R. Seguier, “A survey on databases for facial expression analysis,” 01 2018.

[26] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, and L. Akarun, “Bosphorus database for 3d face analysis,” 01 2008, pp. 47–56.

[27] P. Ekman, “Basic emotions,” in *Handbook of Cognition and Emotion*, T. Dalgleish and M. J. Powers, Eds. Wiley, 1999, pp. 4–5.

[28] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, “Bosphorus database for 3d face analysis,” in *Biometrics and Identity Management*, B. Schouten, N. C. Juul, A. Drygajlo, and M. Tistarelli, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 47–56.

[29] “matlab-dlib-facetrack,” https://github.com/davisking, accessed: 2020-07-06.

[30] M. Emambakhsh, A. N. Evans, and M. Smith, “Using nasal curves matching for expression robust 3d nose recognition,” in *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–8.

[31] Tai Sing Lee, "Image representation using 2d gabor wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, 1996.

[32] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, pp. 886–893 vol. 1.

[33] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1598 – 1603, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865511000122

[34] A. Hajraoui, M. Sabri, and M. Fakir, "Face recognition: synthesis of classification methods," *International Journal of Computer Science and Information Security*, vol. 14, 03 2016.

[35] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525 – 533, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608005800565

[36] K. K. H. Mei L. Ngan, Patrick J. Grother, *NIST Interagency/Internal Report (NISTIR)*, 7 2020.