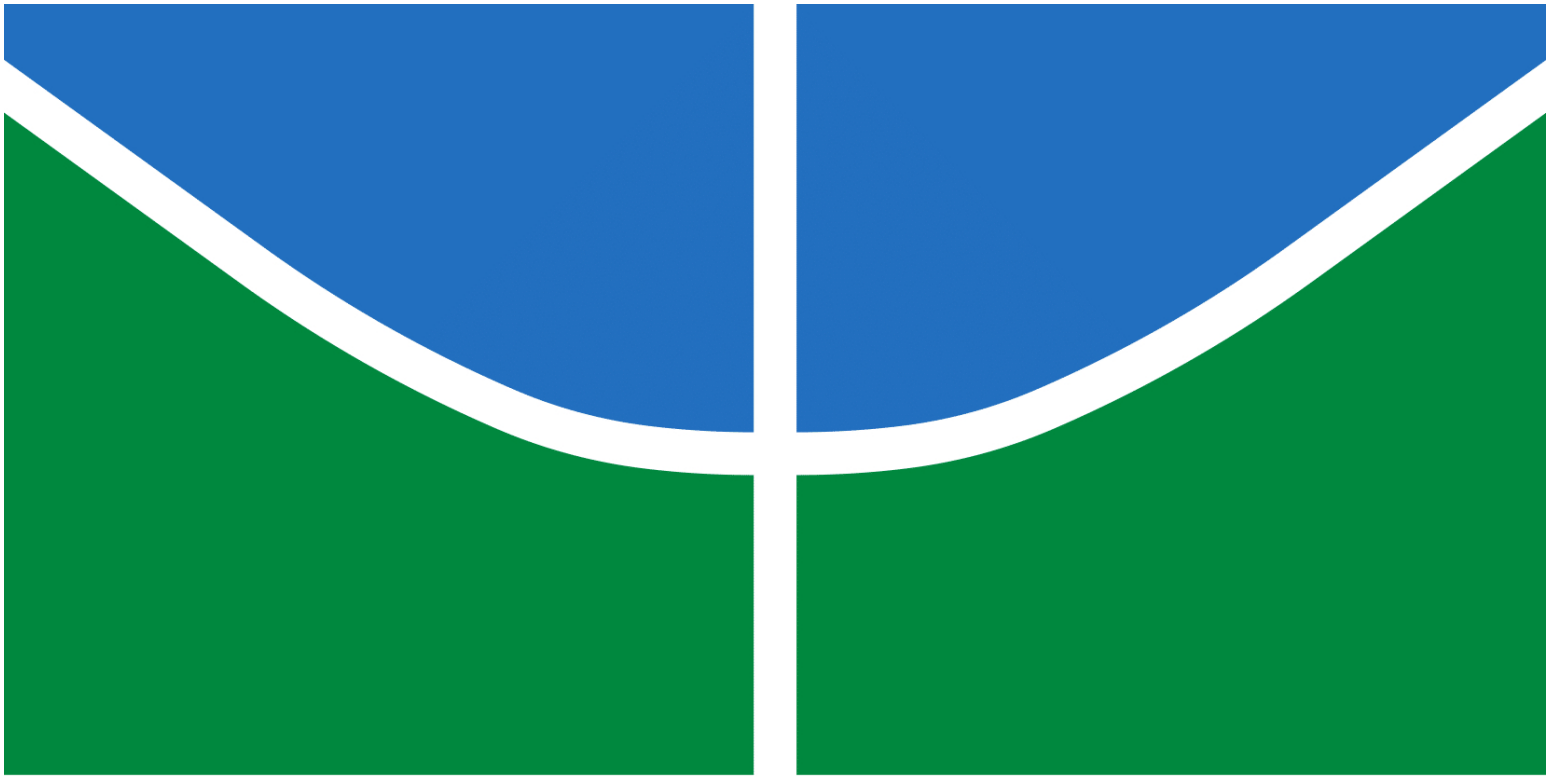


Master's Thesis

Contributions on 3D Biometric Face Recognition for point clouds in low-resolution devices

Luís Felipe de Melo Nunes

Brasilia, February 18 of 2020

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSITY OF BRASILIA
TECHNOLOGY COLLEGE
Mechatronics Systems Postgraduate Program (PPMEC)

Master's Thesis

Contributions on 3D Biometric Face Recognition for point clouds in low-resolution devices

Luís Felipe de Melo Nunes

Thesis submitted as partial requirement to obtain

title as Master in Mechatronic Systems

Examiners

Prof. Dr. Flávio de Barros Vidal, CIC/UnB _____
Advisor

Prof. Eduardo Peixoto F. da Silva _____
Examiner

Prof. José Maurício S. T. da Motta _____
Examiner

Brasília, February 18 of 2020

Dedication

I dedicate this work to my family, especially my parents for all the support given to me, my girlfriend and all the help offered by her, my advisor for opportunity and knowledge, and for all those close to me.

Luís Felipe de Melo Nunes

Acknowledgment

I thank God first for all the strengths bestowed and for guiding my path.
I thank my family for the support and teachings, specially my father Edilton, my mother Karla, my brother João Victor and my sister Julia.
I thank Luana for all the help, companionship, concerns and love, especially in times of difficulty and for always being by my side.
I thank Prof. Flávio de Barros Vidal for guiding me on this project, showing determination and concern with its progress, and instructing in a pleasant and friendly manner.
I thank all my friends for their encouragement, help and strength.

Luís Felipe de Melo Nunes

RESUMO

Recentemente, diversos processos de automação fazem uso de conhecimentos relacionados a visão computacional, utilizando-se das informações digitalizadas que auxiliam na tomada de decisões destes processos. O estudo de informações 3D é um assunto que vem sendo recorrente em comunidades de visão computacional e atividades gráficas. Uma gama de métodos vem sendo propostos visando obter melhores resultados de performance, em termos de acurácia e robustez. O objetivo deste trabalho é contribuir com métodos de reconhecimento facial em dispositivos de baixa resolução de núvens de ponto. Neste trabalho realiza-se um processo de reconhecimento facial em uma base de dados contendo 31 sujeitos, em que cada sujeito apresenta 3 imagens de profundidade e 3 imagens de cor (RGB). As imagens de cor são utilizadas para detecção facial por uso de um Haar Cascade, que permite a extração dos pontos da face da imagem de profundidade formando uma nuvem de pontos 3D. Da nuvem de pontos foram extraídas a intensidade normal e a intensidade do índice de curvatura de cada ponto permitindo a formação de uma imagem bidimensional, intitulada de mapa de curvatura, a partir da qual extrai-se histogramas utilizados no processo de reconhecimento facial. Junto com os mapas de curvature, Um novo método de correspondência é proposto por meio da adaptação do algoritmo clássico de Bozorth, formando uma representação 3D de marcos faciais em nuvens de ponto de baixa resolução para prover um descritor dos pontos chaves da nuvem e extrair uma representação única de cada indivíduo. A validação é realizada e comparada com uma técnica de linha de base para reconhecimento facial 3D. O manuscrito apresentado provê multiplos cenários de teste (faces frontais, acurácia, escala e orientação) para ambos métodos atingindo uma acurácia de 98.92% no melhor caso dos mapas de curvature e uma acurácio de 100% no melhor caso do algoritmo clássico de Bozorth adaptado.

Palavras Chave: Reconhecimento Facial, Nuvens de Pontos 3D, Haar Cascade, Normal e Curvatura, Mapa de Curvatura, F-Measure, Equal Error Rate.

ABSTRACT

Recently, many automation processes make use of knowledge related to computer vision, exploiting digital information in the form of images or data that assists the decision-making of these processes. 3D data recognition is a trending topic in computer vision and graphics tasks. Many methods had been proposed for applications on 3D, expecting a better performance in accuracy and robustness. The main goal of this manuscript is to contribute with face recognition methods for low-resolution point cloud devices. In this manuscript, a face recognition process was accomplished in a 31 subject database, using colorful images (RGB) and depth images for each subject. The colorful images are utilized for face detection by a Haar Cascade algorithm, allowing the extraction of facial points in

the depth image and the generation of a face 3D point cloud. The point cloud is used to extract the normal intensity and the curvature index intensity of each point, allowing the confection of a bidimensional image, entitled curvature map, of which histograms are obtained to perform the facial recognition task. Along with the curvature maps, a novel matching method is proposed by an adaptation of the classic Bozorth's algorithm, forming a net-based 3D representation of facial landmarks in a low resolution point cloud in order to provide a descriptor of the cloud key points and extract an unique representation for each individual. The validation was fulfilled and compared with a baseline technique for 3D face recognition. The presented manuscript provide multiple testing scenarios (frontal faces, accuracy, scale and orientation) for both methods, achieving an accuracy of 98.92% in the best case of the curvature maps and an 100% accuracy in the best case of the classic Bozorth's algorithm adaptation.

Keywords: Facial Recognition, 3D Point Clouds, Haar Cascade, Normal and Curvature, Map Curvature, F-Measure, Equal Error Rate.

SUMMARY

FIGURES LIST

TABLES LIST

Symbols List

Latin Symbols

v_j or e_j	Covariance Matrix eigenvectors
x	Total Vicinity Points
S	Similarity Function
C	Curvature Index Feature Vector
d	Cartesian Distance

Greek Symbols

σ	Curvature Index
λ_j	Covariance Matrix eigenvalues
δ	Difference Function
β	Angular Differences
Ω	Set of Three Dimensional Data

Non-dimensional Groups

n	Number of Face Regions
bins	Histogram *bins* Count
X, Y, Z	Cartesian System Coordinates
I	Identity
M_j	Covariance Matrix
score	Similarity Score
i, j, k, m	Iterators
T or Th	Tolerance Threshold

Overwritten

$\bar{}$	Centroid
$\vec{}$	Vector
T	Transpose

Acronyms

3DSC	*3D Shape Context*
AdaBoost	*Adaptive Boosting*
ANN	*Artificial Neural Networks*
BA	*Batch Adjustment*
CNN	*Convolutional Neural Networks*
DNN	*Deep Neural Networks*
EBGM	*Elastic Bunch Graph Matching*
EER	*Equal Error Rate*
FAR	*False Acceptance Rate*
FBG	*Face Bunch Graph*
FCM	*Face Curvature Map*
FDA	*Fisher Discriminant Analysis*
FERET	*Face Recognition Technology*
FN	*False Negative*
FP	*False Positive*
FRVT	*Face Recognition Vendor Test*
HMM	*Hidden Markov Model*
HOG	*Histogram of Oriented Gradients*
IA	Inteligência Artificial
ICA	*Independent Component Analysis*
IRF	*Intrinsic Reference Frames*
ISS	*Intrinsic Shape Signatures*
LBP	*Local Binary Pattern*
LDA	*Linear Discriminant Analysis*
LFW	*Labeled Faces in the Wild*
LRF	*Local Reference Frames*
LUT	*Lookup Table*
LVQ	*Learning Vector Quantization*
MCS	*Multiple Classifier System*
ML	*Machine Learning*
MLP	*Multi-Layer Perceptron*
MVS	*Multi-View Stereo*
NIST	*National Institute of Standards and Technology*
PBA	*Partial Batch Adjustment*
PCA	*Principal Components Analysis*
PCL	*Point Cloud Library*
PDBNN	*Probabilistic Decision-Based Neural Networks*
PWC	*Pair Wise Coupling*
RBF	*Radial basis function*
RDF	*Random Decision Forest*

RGB	*Red Green Blue*
RNN	*Recurrent Neural Networks*
SFM	*Struct From Motion*
SGD	*Stochastic Gradient Descent*
SIFT	*Scale-Invariant Feature Transform*
SVM	*Support Vector Machines*
TN	*True Negative*
TP	*True Positive*
VTK	*Visual Tool-Kit*

Chapter 1

Introduction

The increasing need to monitor and restrict access to information or environments has led to significant efforts towards the development of a variety of security mechanisms, such as biometric systems [1]. In addition to applications related to access control, there are also others associated with civil identification and criminal investigation. To correctly identify a user, biometric systems must rely on traits that present sufficient levels of universality, distinctiveness, permanence, collectability, acceptability, and circumvention [2].

1.1 Motivation

The proposed work presented in this manuscript aims to provide techniques that grants viable face recognition performance in devices that generate low-resolution depth data, such as modern cellphone cameras and sensors, instead of outperforming current commercial applications outstanding performance [3]. This work is partially published in [4], presenting the description and application of the first proposed method, and a second article is being prepared for publication containing the second proposed method contained in Chapter 4.

1.2 Context of the Study

The advance of technology provides a favorable environment for the emergence of new fields of study or previously presented approaches (that contained many drawbacks due to the lack of structure and resources for a proper usage in the time they were presented) [5]. The acquisition, composition, computation and exploitation of 3D data is included in these mentioned approaches [5, 6, 7].

According to the survey study performed in [5], 3D data acquisition is provided by early proposed techniques and machinery such as laser scanners (that provide high accuracy although presenting a high cost of acquisition and operation), structured light systems [8, 9], stereo vision systems [10, 11] and contact scanners [12]. The advance of this topic and the multiple interest of

application of this type of data led to the emergence of low cost sensors for data acquisition like Microsoft Kinect [13] and techniques available for mobile phones [14].

The usage of 3D data in modern society are found in diverse areas and applications. This kind of technology is found in entertainment environments (video games [13], cinema and television [15]), tracking and pattern recognition applications [14], construction industry and civil engineering [16, 17], forensics and law enforcement [18] and biometric systems for recognition [4, 5, 19, 20, 21, 22, 23, 24].

The research and development of 3D data applications, products and services are promoted by a few cluster of companies and big enterprises (*e.g* Microsoft) encouraging the emergence of events and conferences for presentation and publication of related works, like the LowCost 3D international workshop [25] and the Optical 3D Metrology conference [26].

1.2.1 Biometric System

A biometric system primarily consists of a pattern recognition system operating through the biometric data acquisition of a subject and extracting a group of features from these data that may correspond to previously mapped features of a known subject [27]. This type of application is employed in various social areas currently and it is continuously studied at the academic level, generating enhancement to known methods and arising new techniques.

Biometric systems represent a modern relevant field of study that includes the research of physical and behavioral assets of human beings in order to provide distinction and unique recognition [28]. Biometric systems are constantly studied in order to improve society's abilities of individual recognition exploiting characteristics of universality (every person should present such characteristics), uniqueness (each individual presents unique features, separating an individual from another), permanence (measuring a feature variance along age and time) and collectability (presenting features of easy acquisition) [28]. In a modern society revolved by technology and data storage, the access to personal information in terms of security, authentication and practicality should present alternatives to current methods (such as user passwords, identification cards and personal assistance) guaranteeing safe solutions, leading to the application of biometric systems for identification and verification of identity, granting relevance to the topic [29, 30, 31].

Among the various ways of performing biometrics, it is possible to highlight the facial recognition. Undoubtedly, facial recognition is the most natural and common form of biometric routinely used by humans and one of the most promising areas in biometrics research [32]. In general, facial recognition algorithms uses facial shape and their spatial relationships to perform individuals recognition [1]. A human being can recognize a human face in an unfamiliar environment in approximately 100-200 ms, while in the early years of the 21st century this task used to be challenging to a computer [33]. Nowadays a computer outperforms the human capability of facial recognition [3]. It is true that in the last decade, the reliability of face recognition algorithms has been improved. However, in unconstrained environments, problems such as uncontrolled illumination, head pose, facial expression, and partial occlusion are still a bottleneck to these algorithms to achieve higher

efficiency [32].

Diverse contributions recently published displays the application of commonly known and innovative techniques of face recognition to accomplish the recognition task in low resolution 3D inputs. The works of various authors [34, 22, 21, 35] present the possibility of performing face recognition in low resolution 3D inputs like depth maps (with assistance of color images in few cases) adapting previously attempted methods in color-based recognition approaches. These approaches inspires the adaptation of classic biometric algorithms to solve a 3D data face recognition task expecting efficient results even for algorithms that were not originally developed for the specific task of biometric face recognition (adapting from other biometric approaches such as fingerprint recognition).

Following the hypothesis that an adaptation of a classic biometric algorithm of another recognition task that contain similar principles may present efficient results in the task of 3D face recognition, this manuscript propose an adaptation of Bozorth's fingerprint matching algorithm [36], forming a net-based 3D representation of facial landmarks in a low resolution point cloud in order to provide a descriptor of the cloud key points and extract an unique representation for each individual.

The majority of traditional research and commercial use of facial recognition systems are focused on methods that explore 2D (two-dimensional) images of human faces [19]. These methods, in general, are based on feature extraction that does not take into account the 3D shape of faces, especially about the depth. This work presents methods based on Point Cloud, curvature map and the linking net distribution of landmarks in reference to the normal vector of these landmarks (adapted from Bozorth fingerprint matching and labeled Bozorth 3D) to perform face recognition. To the best of our knowledge, although some works have already addressed the challenge of 3D facial recognition [37, 19], the solution presented in this paper is the only one that uses Point Cloud data and the FCM method or a variation of the fingerprint matching algorithm developed by Bozorth [36], applied to a public dataset of 3D face images acquired by a low-cost sensor device, to perform the task.

1.3 Manuscript Organization

In this work, a review of relevant concepts and the explanation of significant algorithms for the proposed methodology is provided in Chapter 2, elucidating the core techniques applied in the methodology. Next, an overview of Face Recognition concepts, involved tasks, and published techniques and methods is developed in Chapter 3 to contextualize the field of study and present state-of-art algorithms. These chapters supply the basic required knowledge for the best understanding of the proposed methodology.

The following chapter (Chapter 4) describes the methodology proposed by this work, showing the task description, the elements of the used data set, and the data processing required to achieve the desired output. Pipelines are provided depicting the following course of action of the methodology, aiming to grant clear understanding. The next chapter (Chapter 5) provides the

testing scenarios elaborated to measure the performance of the proposed methodology and discuss the viability and drawbacks. Finally, Chapter 6 concludes the researching emphasizing the best results achieved, difficulties faced, and the future works available for the current approach.

Chapter 2

Main Concepts and Techniques

The proposed implementation to perform the three dimensional biometric facial recognition is based in point clouds, three dimensional geometrical assets and published techniques that solves minor tasks along the methodology. The explanation of these concepts is important in order to obtain a better understanding of the proposed methodology steps.

2.1 Point Clouds

According to [38], the surface of a three dimensional object is generally represented by triangular meshes formed by a known mathematical estimation, interpolation or approximation of a cluster of organized points, in terms of computer graphics.

Multiple applications in manufacturing, medicine, geography, design, and other areas of knowledge require the scanning of three dimensional objects of certain complexity, aiming to incorporate them to an assisted processing, usually performed by computer [38]. In order to supply the demands of such areas, multiple measuring techniques were improved to easily produce a great amount of points composing the objects surface. This cluster of points representing a three dimensional object in a coordinate system is labeled point clouds, as rendered in the example of Figure 2.1.

Recently, multiple methods were proposed and developed for applications point cloud based, aiming to structure and dispose practicality to handle the data. Some of these methods are:

- Surface reconstruction based on triangular meshes [38];
- Detection and Extraction of 3D features;
- Surface estimation;
- Surface normal and curvature estimation;
- 3D data Filtering;
- Depth image formation;

- Pattern and Objects recognition [39];
- Objects detection and segmentation [40];
- Visualization Methods;

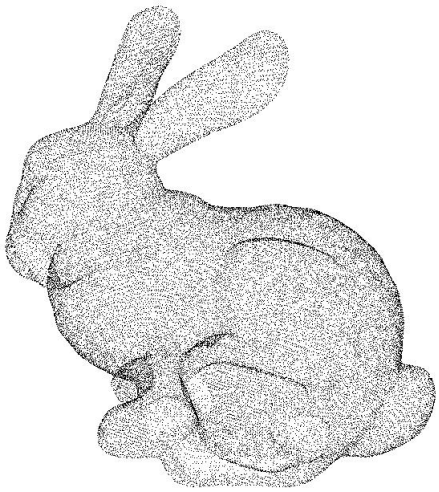


Figure 2.1: An example of Point Cloud [41].

Point Clouds are represented by points scattered spatially, generally represented by a Cartesian coordinate system. In a three dimensional coordinate system, the points' position is described in terms of the Cartesian coordinates *X, Y* and *Z*. In a point cloud environment, the properties of the cloud may encapsulate additional information to each point (other than position) such as normal vector components, intensity level, color information (RGB, YUV, and others), gradient intensity, among other information that the cloud manager consider relevant [42].

2.2 Algorithms

This sections is responsible for the exposure of the algorithms hereinafter. These algorithm are necessary for minor tasks contained in the methodology, making their understanding fundamental for this work. There are three main algorithms shown in this section, the Viola-Jones face detector [43] responsible for data pre-processing, the Intrinsic Shape Signatures (ISS) used for point cloud feature extraction [44] and the Bozorth Algorithm [36] responsible for the web formation and the matching process.

2.2.1 Viola-Jones Algorithm (Haar Cascade)

The object detection based on Haar-like features cascade classifiers is an efficient method proposed by Paul Viola and Micheal Jones [43], composing a machine learning algorithm where a cascade function is trained by an image data set, generating a model that can be applied to detect the same objects on other images.

2.2.1.1 Haar-like Features

The called Haar-like features are descriptors used in digital images for object detection. These features were applied on the first real time face detection application [45]. These descriptors usually represent a color or gradient pattern of pixel values like the example displayed in Figure 2.2. These descriptors received this nomenclature due to its intuitive similarity with Haar wavelets [45].

The algorithm developed by Viola-Jones is commonly known as Haar Cascade, composing a real time detection method based on integral images, permitting the evaluation of features effectively in terms of time and a scale invariant analysis. The structure of the classifiers is a relevant aspect of the algorithm, supporting the selection of pertinent features, since the number of Haar-like features in any sub-window of the image can be considerably large, surpassing even the pixel total. The performance of the algorithm is enhanced by the third and last step that consists in the successive combination of complex classifiers in a cascade formation, discarding unimportant features that permits the algorithm to compute the facial detection only in promising regions of the image.

The algorithm's classifiers are based on simple feature-values, due to its higher efficiency compared to pixel value operations. The value of a Haar-like feature is computed as the difference between the sums of pixel values inside sub-windows, with at least a common (adjacent) vertex or edge, or the pixel value sum difference between clusters of sub-windows (usually represented by color like shown in Figure 2.2). Three rectangular shaped Haar-features are used for classification in this algorithm, where the first one consists of two rectangular regions side by side either vertically and horizontally (depicted by blocks A and B in Figure 2.2), the second one is formed by three rectangular regions disposed side by side and horizontally and the feature value is composed by the difference between the center region and the edge regions (depicted by block C in Figure 2.2), and the third and last feature is formed by two clustered regions (formed by four rectangles) disposed diagonally (depicted by block D in Figure 2.2). This type of feature is crucial for the operation of the Viola-Jones algorithm that explores the shading caused by the facial structure and the facial landmarks.

2.2.1.2 Integral Images

Rectangular features are rapidly computed in an image representation labeled integral image. An integral image consists in an image transformation in which every pixel of the integral image stores the sum of all the pixel values located in the top and/or left of its current (x,y) position. The integral image allows a faster rectangular Haar-like feature value computation, since the sum of pixels of each region is simply calculated, being reduced to a combination of differences between the vertices values instead of a full traversal of the region with iterative sum computations.

2.2.1.3 Learning and Classification

The algorithm uses Machine Learning (ML) to train a classifier. Using the feature set displayed in Figure 2.2, the extraction of a feature values set together with a set of training images (containing

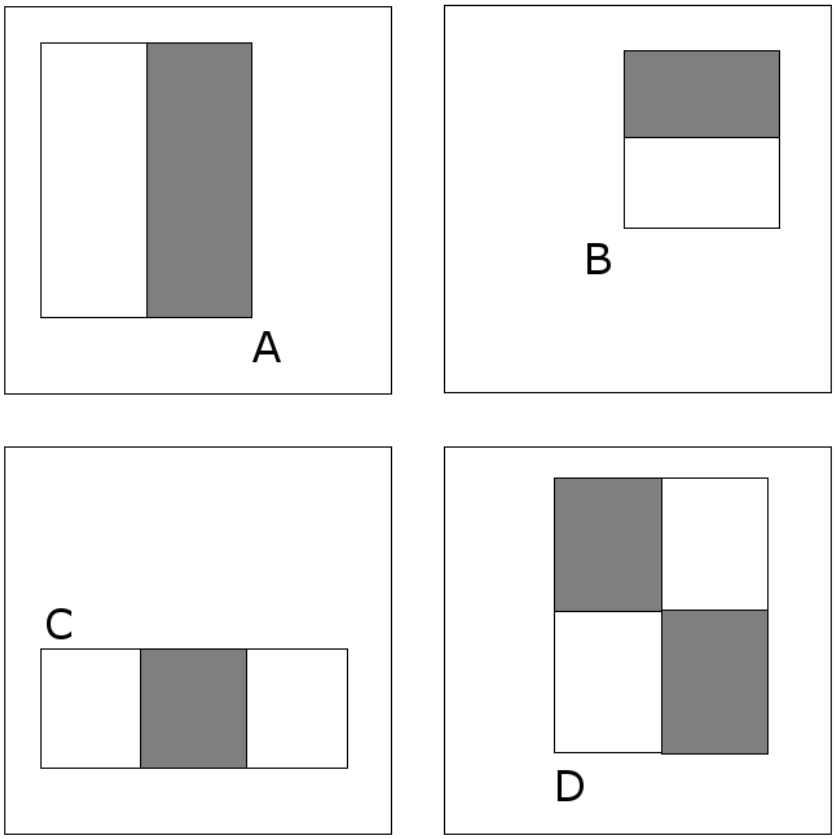


Figure 2.2: Rectangular Haar-like Feature. These features value are determined by the difference between the sum of pixel values in the white rectangles and the sum of pixel values in the gray rectangles [43].

face and non-faces images) the algorithm set a classifier training scenario. In Viola-Jones paper a variant of *AdaBoost* [46] is applied for the feature set selection and as classifier.

Early experiments shows that a frontal face classifier built on 200 features presents a detection rate of 95% with a ratio of 1 false positive for each 14084 [43].

In the face checking process, the *AdaBoost* starts selecting a feature focused on the eyes and the nose/cheeks, since the eyes region is usually darker than the nose/cheeks region, due to the presence of eye lids, eye lashes and the iris color, generating a high value gradient, next the *AdaBoost* select a feature focused on the eyes and the nose bone, that presents the same principle of the last feature. The Figure 2.3 presents an example of these features application.

2.2.1.4 Classifier Cascade

The general form of the detection process is depicted as a degenerative decision tree, called by the term "cascade". This degenerative decision tree is displayed in Figure 2.4. The positive classification of the first classifier activates the feature set evaluation in a second classifier with tougher rejection criteria, repeating the process in cascade for a defined number of classifiers,

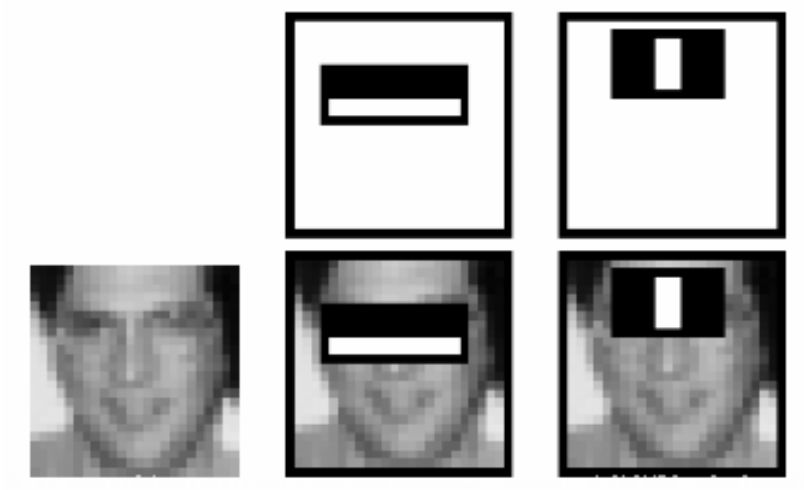


Figure 2.3: Features used by *AdaBoost* for face checking. The first feature is a horizontal rectangular feature that searches for the region of eyes and nose/cheeks. The second feature is a vertical rectangular feature that searches the region of eyes and nose/bone [43].

resulting as output a super refined feature set with high probability of representing a face. The rejection of the feature set in any level of the classifier results in the immediate rejection of the sub-window.

Usually the classifiers that receive a higher number of features tend to reach higher detection rates and less false positive rates, despite of the processing cost elevation that leads to slower output rates. Therefore the principal aspects to be defined are the cascade size, stipulating the number of classifying stages, and the rejection threshold within each stage. These aspects represent a considerable problem to maximize the efficiency of the classifier.

An Equal Error Rate (ERR) approach is implemented by Viola-Jones in order to obtain an optimal operation setting for the algorithm. The number of stages is increased until no further significant change is noticed in the face checking procedure, then the threshold of these stages are regulated until the false positive rate is similar to the detection rate. The results presented in [43, 45] show that the algorithm presents high efficiency in terms of detection rate and the execution time, turning it eligible for real time facial detection applications.

2.2.2 Intrinsic Shape Signatures

The Intrinsic Shape Signatures (ISS) is a shape descriptor and a key point extractor developed by Zhong [44] relying on region-wise quality measurements. The algorithm selects points based on the smallest eigenvalue (representing large variations on principal directions) and filters duplicates through successive eigenvalues ratio checking (discarding points with a similar spread along principal directions) [47]. This algorithm provides invariant frames assisting in the elaboration of consistent and discriminating feature vectors for subject description.

The ISS is based on 3D Shape Context (3DSC), described as a 3D histogram representing the

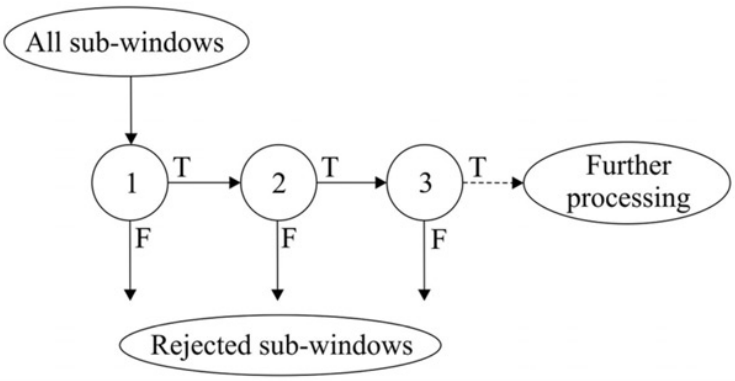


Figure 2.4: Pipeline of the Degenerative Decision Tree of classifiers (Classifier Cascade). The sub-windows are used as input in the first classifier, continuing through the process until a false classification occurs or the data reaches the last classifier, heading to the next processing step (if it exists) [43].

spatial distribution of data points in a surrounding support sphere at an oriented basis point, in which the north pole is aligned with the surface normal, leading to a reference irresolution. To solve this irresolution Zhong [44] proposed the eigen analysis of local points, obtaining four distinguishing frames derived from Principal Component Analysis (PCA) subspaces, with each frame transforming into another through a 180° rotation along one of its axes. This solution generates symmetrical ambiguities that must be eliminated in order compute highly discriminating shape descriptors.

The ISS algorithm operates directly on point clouds, presenting a viable algorithm for further exploitation in Chapter 4.

2.2.2.1 Definition

An Intrinsic Shape Signatures consists of an intrinsic reference frame enabling both view-invariant feature extraction and fast pose registration, and a highly discriminating feature vector encoding the 3D shape characteristics [44]. The ISS are composed by the intrinsic reference frame and a set of basis vectors obtained from the eigenvectors.

2.2.2.2 Intrinsic Reference Frames

The idea behind the Intrinsic Reference Frames (IRF) is to generate shape descriptors view-independent similarly to Local Reference Frames (LRF) [48]. The surface normal alone is not enough to compose a 3D coordinate system, therefore other basis vectors are generated by algebraic combination of the point eigenvectors.

Zhong defines an IRF F_i at a basis point p_i with a supporting radius r_{frame} using the eigen analysis of the point covariance matrix as follows:

1. Compute a weight for each point p_i inversely related to the number of points in its spherical neighborhood of radius $r_{density}$, used to compensate for uneven sampling of 3D points,

highlighting sparsely sampled regions contributions over dense sampled regions contributions;

2. Compute a weighted covariance matrix $cov(p_i)$ for p_i using all points p_j within a vicinity of distance r_{frame};

3. Compute its eigenvalues $\{\lambda_i^1,\lambda_i^2,\lambda_i^3\}$ in the order of decreasing magnitude and their eigenvectors $\{e_i^1,e_i^2,e_i^3\}$;

4. Use p_i as the origin, use e_i^1, e_i^2 and their cross product $(e_i^1 \otimes e_i^2)$ as the $x-$, $y-$ and $z-$axes to define a 3D coordinate system F_i at p_i labeled intrinsic reference frame. The coordinate system generated is displayed in Figure 2.5(a).

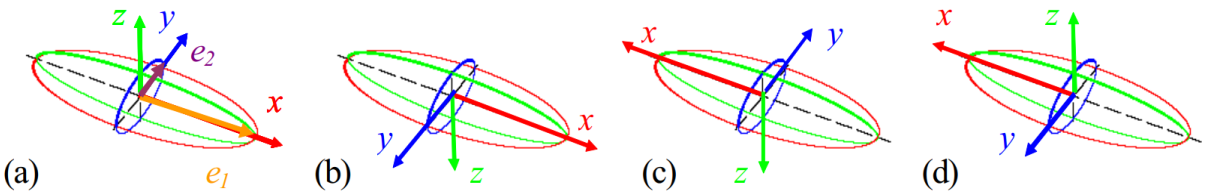


Figure 2.5: Four intrinsic reference frames computed from the covariance matrix eigenvectors $\{e_i^1,e_i^2,e_i^3\}$ of decreasing magnitude from eigenvalues [44].

The generated IRF presents a 180^o orientation ambiguity, since the eigenvectors from the covariance matrix computes a direction based on vicinity point distribution, leaving two options for the orientation of each axis that results in four variants of IRF uniquely defined at a basis point as shown in Figure 2.5, *i.e.*, F_i as described above, and three others obtained by a 180^o rotation along $x-$, $y-$ and $z-$axes (respectively shown in Figure 2.5(b)-(d)). The same ambiguity also exists for the surface normal (unless a view-point is established, contradicting the view-invariant purpose of the algorithm). The IRF is a generalization of the surface normal for view independent shape feature extraction, however it defines unique and highly discriminating representations for local 3D shape patches.

2.2.2.3 Key Point Selection

In [47] the filtered points of ISS algorithm contain highly discriminating shape representation, containing a large variation along every principal direction. Tombari [48] defined a pipeline representing the steps of the ISS key point extraction (see Figure 2.6), based on the covariance matrix computation, whose eigenvalues are used as measure for a pruning step to discard points with a similar spread along principal directions, where a repeatable LRF cannot be defined, and the eigenvalue of smallest magnitude is used as saliency, concluding the extraction with a Non-Maxima Supression (NMS) over the saliency.

The pruning step consists in a threshold verification of the spread along principal directions. Only points whose ratio between two successive eigenvalues are below a threshold are considered eligible to contain discriminant local description [47]. Equation 2.1 represents the spread tolerance checking to prune the original input set.

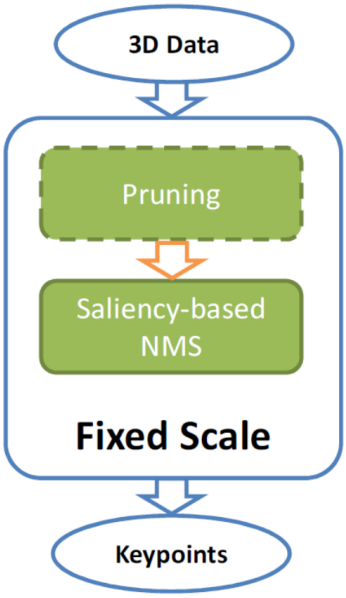


Figure 2.6: Pipeline of the Key Point Extraction algorithm using the ISS. The 3D Data consists in a point cloud data, which points are pruned by a threshold evaluation over the vicinity spread along principal directions, followed by a saliency-based NMS to retrieve discriminating points described as key points[48].

$$\frac{\lambda_1(p)}{\lambda_2(p)} < Th_{12} \quad \wedge \quad \frac{\lambda_2(p)}{\lambda_3(p)} < Th_{23} \tag{2.1}$$

considering p a random input point and $\{\lambda_1(p),\lambda_2(p),\lambda_3(p)\}$ the eingenvalues of its covariance matrix in decreasing order.

The output of the pruning represent candidates of efficient key points, presenting a highly discriminative local shape description. In order to check the veracity of the point efficiency, a saliency-based NMS is applied over the magnitude of the last eigeinvalue, that represents the saliency of the point. This final operation returns only points that contain a large variation along principal directions and consequently describe discriminating elements.

2.2.3 Bozorth Minutiae Matching Algorithm

The Bozorth Algorithm is a fingerprint matching system proposed by the National Institute of Standards and Technology (NIST) in cooperation with the Federal Bureau of Investigation (FBI) that uses detected fingerprint minutiae to determine if two fingerprints are from the same finger of a subject. The system can analyze fingers two at a time or run in a batch mode comparing a single finger (probe) against a large database of fingerprints (gallery) [36].

A Basic fingerprint image is formed by ridges, valleys, cores, deltas, pores, and other elements that represents the fingerprint surface. Fingerprints minutiae are discrete features composed by points in a finger's friction skin where a ridge termination (labeled ridge *ending*) or splitting

(labeled ridge *bifurcation*) is present [36]. The minutiae features simply formed by the ridge element orientation and the point location where it occurs. The Figure 2.7 demonstrates an example of a bifurcation and a ridge ending.

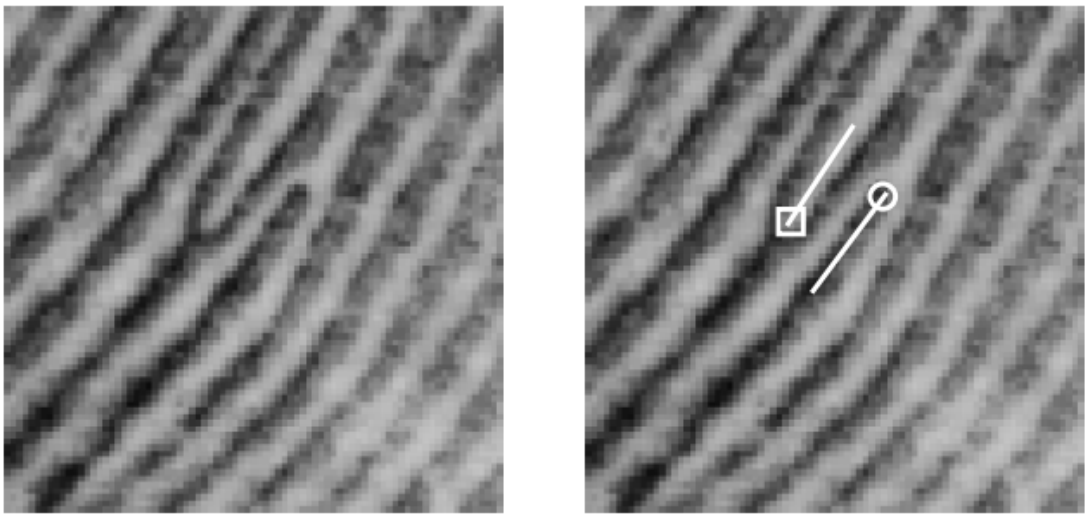


Figure 2.7: Example of two minutiae detected. The squared minutiae represents a bifurcation point and the circled one represent a ridge ending. In both cases the white line represents the orientation of the minutiae [36].

The algorithm is described as a pipeline composed by three steps [49], in which the first step consists in the construction of an Intra-Fingerprint Comparison Table that stores pairs of minutiae linked by a distance vector, the following step is the construction of an Inter-Fingerprint Compatibility Table that compute matches between minutiae pairs in the Intra-Fingerprint Comparison Table of two different fingerprints storing the minutiae index of each pair and the orientation difference between the distance vector of both tables, and the finishing step traverses the Inter-Fingerprint Compatibility Table to check for duplicates and matching quality.

The first step receives a set of minutiae from each picture to create a linked point web labeled as Bozorth Composition, in which the minutiae are linked following parameters specifications, leading to a Intra-Fingerprint Comparison Table, that represents a linked web of the minutiae that allows the matching of two input fingerprints. The Intra-Fingerprint Comparison Table stores the distance magnitude between two minutiae, the index of each minutiae, the angular difference between the distance vector orientation and each minutiae orientation, and the angle of the distance vector. Figure 2.8 presents the elements contained in the comparison table generated by a link between two minutiae. Each minutiae is linked to others minutiae within a predefined radial distance.

After computing the Intra-Fingerprint Comparison Table for each input fingerprint (either two fingerprint for verification or a probe and gallery for identification), the Inter-Fingerprint Compatibility Table is generated, comparing the table elements of both fingerprints minutiae web, representing an eligible matching table between two fingerprints. This compatibility table stores all the matching minutiae pairs between two webs being compared, where the matching criteria is described as a distance tolerance checking, followed by an angular tolerance checking applied

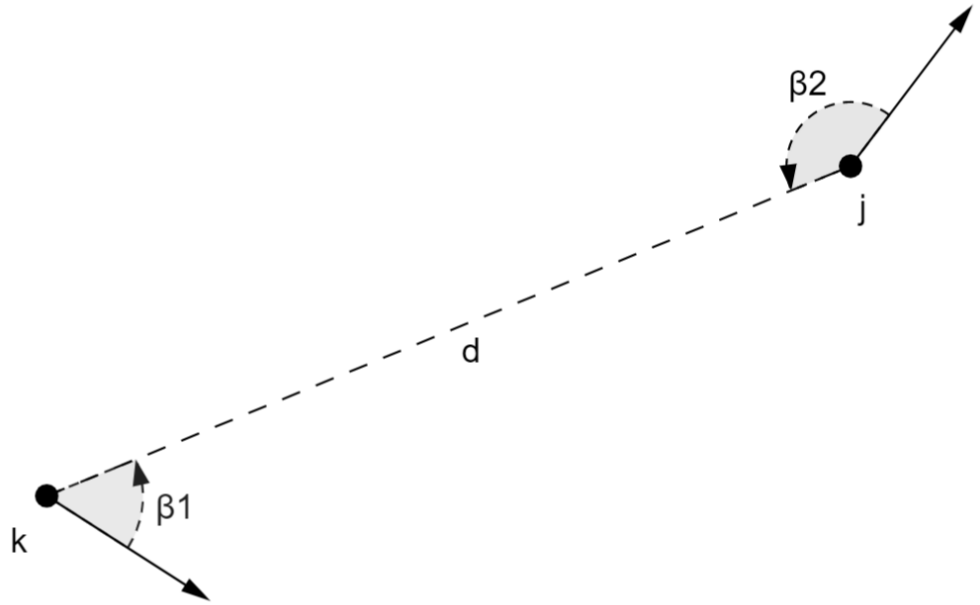


Figure 2.8: Intra-Fingerprint minutiae link. A two-minutiae link example, where the distance vector is represented by d, while k and j are the source and destiny minutiae index, respectively, and $\{\beta 1, \beta 2\}$ are the respective angular difference between the minutiae (k and j) vector orientation and the distance vector orientation (Adapted from [49]).

to the distance vector orientation and the other angular elements stored in the Intra-Fingerprint Comparison Table. If all the elements are within the threshold tolerance, the angular difference between both distance vectors and the minutiae index (from the probe and the gallery) are stored in the compatibility table, indicating a match between the current minutiae pairs. The result of this procedure is a set of matching minutiae pairs between the probe and the current gallery fingerprint being compared, that leads to a final step of match score computation.

The final step measures the quality of the matched pairs computed in the Inter-Fingerprint Compatibility Table, outputting a numerical score of matches that fit the algorithm's tolerances. Figure 2.9 presents a full example of Bozorth's original algorithm.

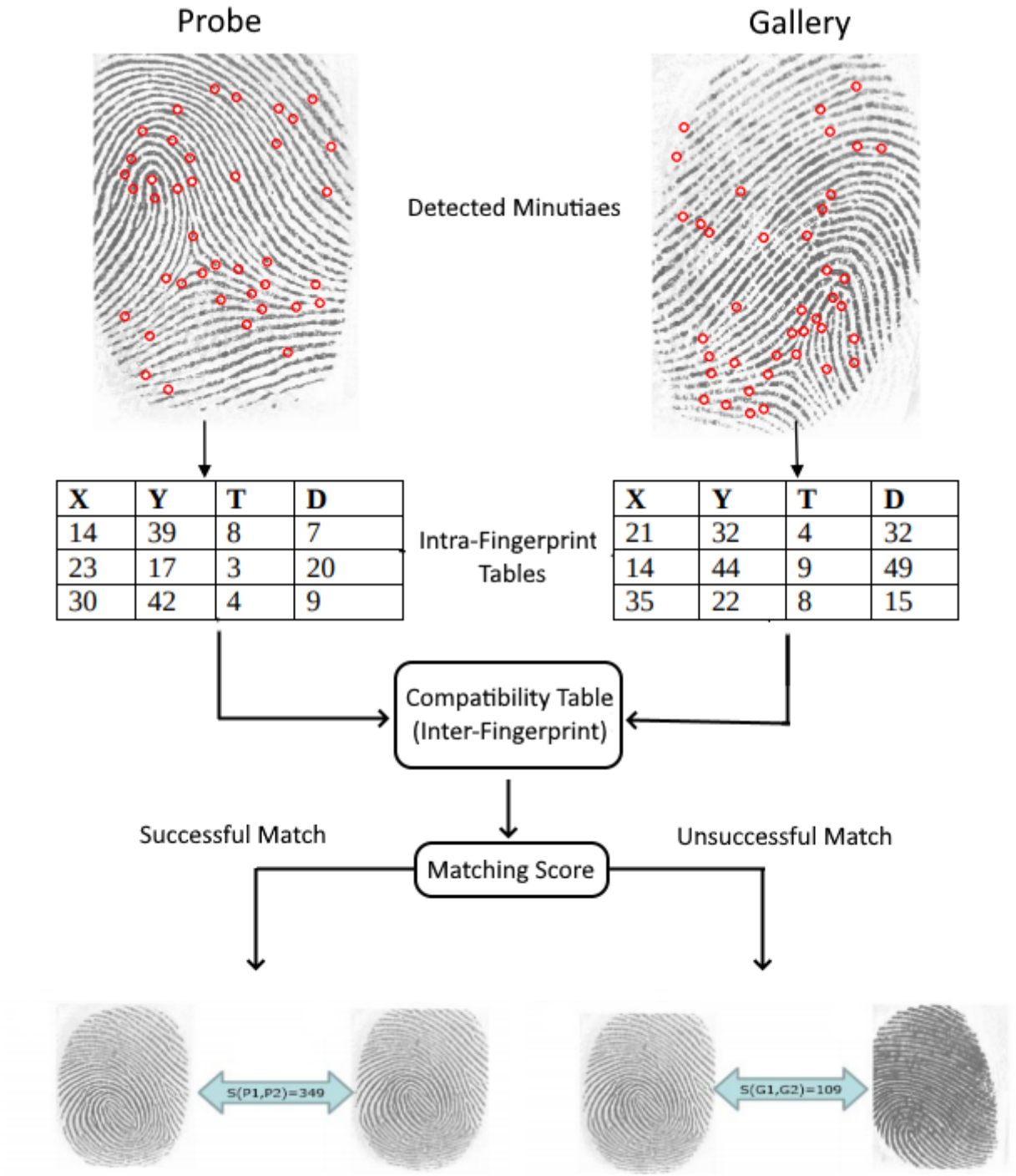


Figure 2.9: Example of original Bozorth's minutiae matching process (Adapted from [49]).

Chapter 3

Related Works

Since the advance of computer vision, many face recognition related techniques were proposed and developed over the years [5]. This chapter presents an overview of the face recognition study scenario, granting information over the theme and proposed methods along the years.

3.1 Face Recognition

The facial recognition process consists in techniques of body measurement and feature analysis (biometrics) to determine a person's identity through data correlation [5]. The biometrics features are a type of resource frequently studied and researched in modern days for multiple purposes, specially automation.

The technologies based on biometrics features includes identification based in physiological characteristics (such as face, fingerprints, fingers geometry, hands geometry, veins, palm shape, iris, or other elements) and behavioral traits (like gait, signature, or tapping pattern) [50]. The facial recognition tends to offer advantages in comparison with other biometrics recognition methods, primarily because most of these other technologies require a voluntary action of the user, either by placing hands or fingers in a physical sensor or by positioning or approaching in an specific pose a capture spot (like in iris or retina scanning), avoiding any external stimulus that might jeopardize the captured data quality. The facial recognition permits a passive data capture, without any explicit action or participation performed by the user. However, according to [51] facial recognition viability is bounded to a constricted environment, since the recognition process is sensitive to extreme variations on illumination, pose and expression. However, recent works presenting deep learning architectures reduce drastically the effects of extrinsic factors sensitivity [3, 52, 53, 54].

3.1.1 Applications

Facial Recognition is primarily applied in two task scenarios:

1. Verification (Unique-wise Comparison): This task is one for one matching verification, in

which an unknown input face claims an identity, either returning a positive or a negative output.

2. Identification (Group-wise Comparison): This task is one for many matching verification, in which an unknown input face is compared to a gallery of stored known individuals' face, retrieving either the identity of the person or a matching rejection.

Numerous fields of modern society are evolving in terms of technological infrastructure, automating services for practicality and faster attendance. The facial recognition is an automation asset that is exploited in the following areas:

- Security - Building access control, airports and seaports, ATM, border control, network and computer safety, e-mail and multimedia workstation authentication;
- Surveillance - CCTVs monitoring (alerting authorities in case of outlaw recognition). The facial recognition on CCTV resource was used in 2001 in a Super Bowl game in Florida [55];
- General Identity Verification - electoral register, banking operations, e-commerce, newborn identification, passports, drivers license, employees recognition;
- Criminal Justice Systems - police booking/mug-shot systems, post-event analysis, forensics;
- Police Investigation Assistance - missing persons, illegal foreigners and criminal search;
- Smart-cards - authenticity check;
- Multimedia Environments with adaptive human-computer interface - context aware systems for childcare and old people's centers;
- Video Indexing - labeling faces throughout a video; and
- Witness Face Reconstruction;
- Social Media Indexing;

In addition to these applications, the underlying techniques in face recognition led to modifications to related fields of computer vision, *e.g.* gender classification, expression recognition and facial features detection and tracking [5]. Each of these cases mentioned are valuable for various dominions, like the medicine field for intensive treatments, a driver's tiredness condition monitoring by eyes tracking, and even the face recognition can be assisted by these techniques.

The face recognition is also used along with other biometrics algorithms, such as speech, iris, fingerprint, ears and gait recognition in order to enhance the recognition result and the algorithms performance [20, 56, 57, 58].

3.1.2 General Difficulties

Face Recognition consists in an arduous specific object recognition case. The main challenge of the face recognition is defined by faces' small visual variability in general, even in its most natural form (frontal vision). In 1998 face images were described as a dense group in images space that leads to an inability to obtain accurate discrimination and high success rate through standard pattern recognition techniques [59], but nowadays many published approaches achieve results overcoming human recognition capability [3].

Furthermore, human faces are not an unique rigid object, and there are numerous factors that interferes with facial visual appearance. The source of variation are categorized as extrinsic or intrinsic [60]. Intrinsic factors accrue from faces physical nature, being view-independent. This class of factors is represented by subsets called intrapersonal and interpersonal, respectively representing unique person's self appearance variability (caused by age, facial expression, cosmetics, facial hair, glasses, and others) and the appearance variability among others (*e.g.* ethnicity and gender). Extrinsic factors are derived from interactions between the environment and the observer. The elements of the environment that may cause changes in a person's appearance includes illumination, pose, scale and data parameters (resolution, focus, noise, and others).

3.1.3 Main Face Recognition Methods

The acquisition method of facial images depends on the applications goal since there are numerous sensors possibilities and scenarios with different constriction levels. For example, a surveillance scenario usually retrieves the images from security cameras presenting a low level of constraining, while a building access control or an ATM identity validation should present common cameras in a highly constraining environment. Some application may even use 3D sensors for another layer of security against fraud [61]. The data type of image that the sensor retrieves generate three categories of face recognition according to [5], those that operate on intensity images, video sequences and other sensory data (like 3D images and infrared).

3.1.3.1 Intensity Images

There are two main categories that represent face recognition from intensity images, feature-based and holistic. The first one explores the correlation of local features, while the second one handle a global analysis of the image.

3.1.3.2 Feature-Based Recognition

Feature-based approaches primarily process an input data, retrieving crucial and discriminating facial elements (labeled as fiducial marks), such as eyes, nose, mouth, chin bottom, cheekbone and others. Through these features data correlation (usually represented by geometrical relation) statistical techniques are applied in order to match correspondences between faces.

The first automated face recognition techniques developed were mostly feature-based. Kanade[62] applied simple image processing methods to extract a feature vector of 16 facial parameters, composed of distance, area and angular proportions, using euclidean distances as comparison element, that generated a 75% peak performance in a 20 subject database, containing 2 images of each. Brunelli and Poggio [63], used Kanade's approach as reference to compute 35 geometric features (Figure 3.1) from 4 images of each 47 subjects contained in their data set, generating a 90% recognition ratio and a 100% accuracy over a simple template-matching model.

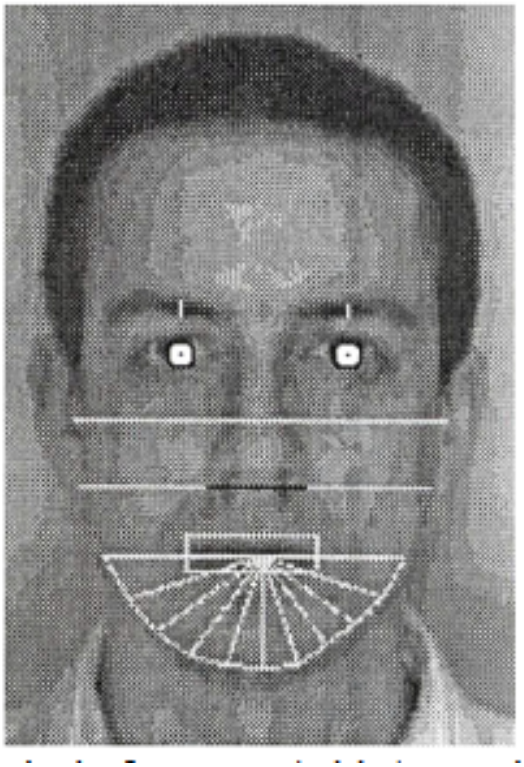


Figure 3.1: Geometric Features computed by Brunelli and Poggio. These features explore the correlation of some fiducial landmarks contained in a human face [63].

Along the years, new approaches based on sophisticated feature extraction techniques were elaborated, exploring deformable templates, Reisfeld symmetric operator [64], Hough transforms [65], Graf morphological operators [66], Elastic Bunch Graph Matching (EBGM)[67], Gabor Filters, data structure correspondences and Histograms of Oriented Gradients (HOG) [68]. All these methods are strongly heuristics-based, constraining the image search by geometric limitations and the performance by template tolerance checking. Cox *et al.* [69] reported a performance of 95% in a data set composed by 685 images and subjects using a feature vector of 30 dimensions derived from 35 facial landmarks, however all the landmarks selection was manual, that contributes to a better performance compared to automated detections. Wiskott *et al.* [67] developed a dynamic data struct-based solution, constructing a full connected graph for each face choosing a set of fiducial points labeled by Gabor Filters' responses applied to a window around it, and each arch of the graph is labeled with the distance between the points. After a reasonable number of graphs are computed, graphs of new faces are obtained automatically by EBGM. The comparison of graphs leads to a

similarity score, which the highest value represents the probe identity. This solution provided a recognition ratio of 98% for first-rank classification and up to 99% for the first 10 ranks using a gallery composed by 250 subjects. The method was enhanced to allow pose variation (Figure 3.2), presenting the same performance. Despite of the great performance, a manual construction of at 70 graphs were necessary in order to obtain viable EBGM outputs. Campadelli and Lanzarotti [70] proposed an enhancement that does not require the manual placement of the graphs, using deformable templates proposed in [71] for automatic fudicial landmarks detection. Other solution substitutes the Gabor features with a graph correspondence strategy and HOGs application.

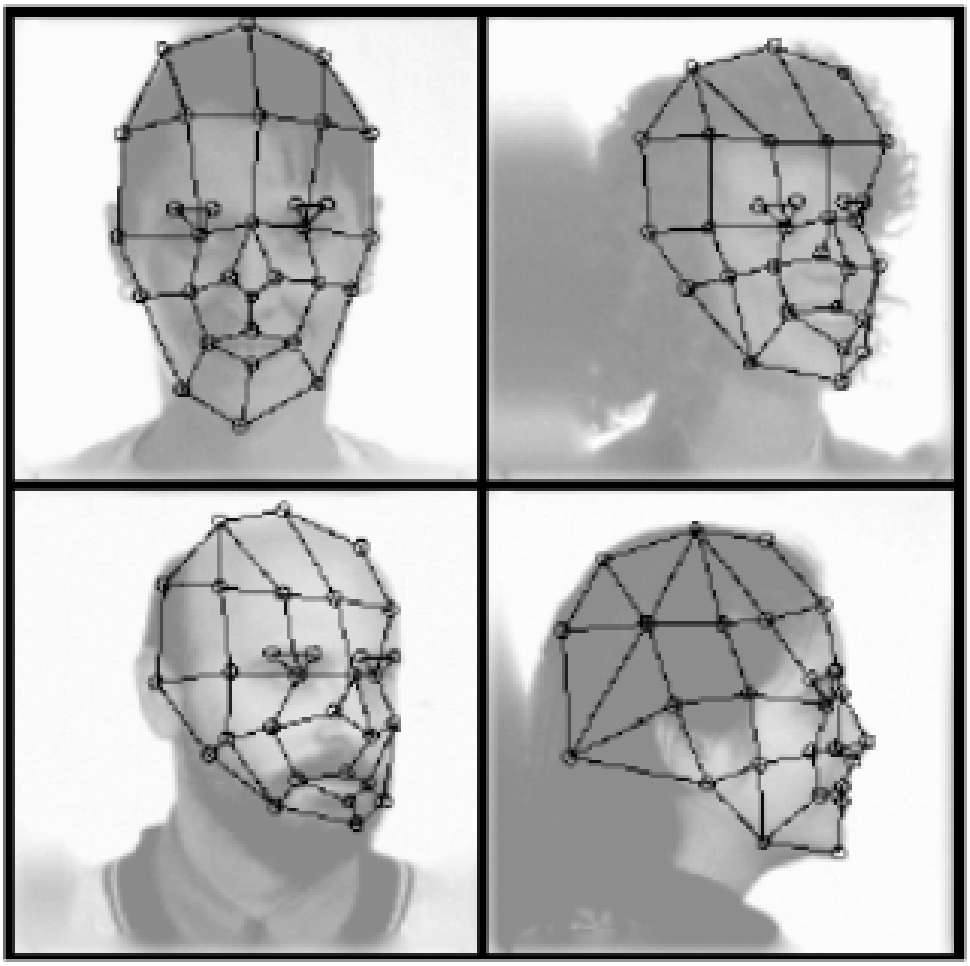


Figure 3.2: Full connected Graphs proposed by Wiskott *et al.* enhanced with pose variation [72].

Mainly face recognition algorithms are based in the most common form of face, represented by a frontal view. However, considerable efforts were devoted to recognition over faces' profile, generally reducing the feature extraction to a simple one dimensional problem. Kaufman and Breeding [73] reported a 90% recognition rate using faces' profile of 10 subjects. Liposcal and Loncaric [74] obtained a 90% accuracy rate using a subspace filter to extract a 20-dimensional feature vector from the profile pictures correlating the data through euclidean distance in a 30 subject data set. Harmon *et al.* [75] achieved recognition accuracies up to 96% using a data set

formed by 112 subjects, exploiting a 17-dimensional feature vector to describe the profile images, corresponding the vectors through euclidean distance.

3.1.3.3 Benefits and Liabilities

The main benefit offered by feature-based techniques accrue to the fact that the feature extraction precedes any matching analysis, and these methods are usually robust to pose variations. Feature-based methods can also present invariability to scale, orientation and brightness. In terms of processing cost, usually the facial data is represented in a compressed state (feature vectors) that results in a fast comparison computation.

Despite of the benefits that feature-based recognition provides, these techniques contain obstacles related to landmark detection, that requires either the manually setting of points or the implementation of a robust complex algorithm, and the landmarks (or features) selection based on its discriminating capacity. These algorithms usually depicts a complex scenario hindering the achievement of an optimal case.

3.1.3.4 Holistic

Holistic approaches try to identify faces using the global description of the image, focusing mainly in the individuals appearance instead of the correlation between the facial landmarks. These approaches are categorized in two subsets: statistical approaches and Artificial Intelligence (AI)

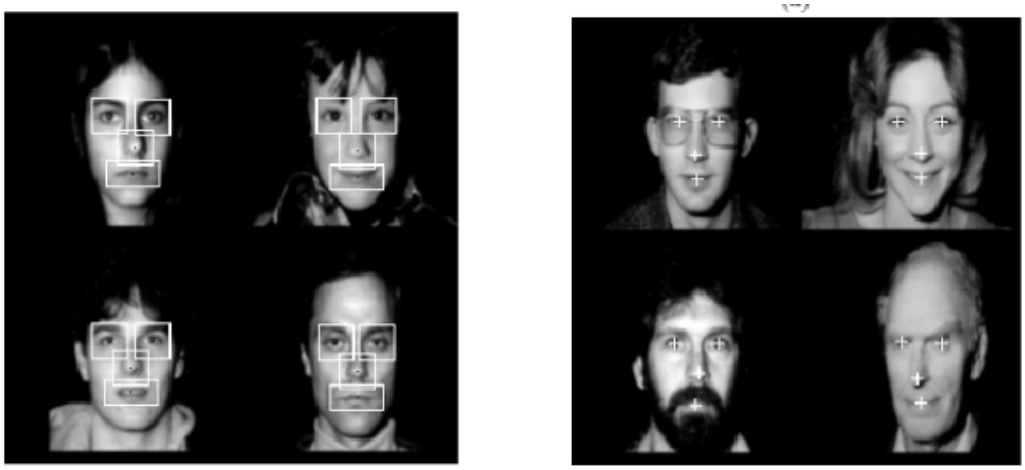
3.1.3.5 Statistical Approaches

Statistical approaches usually represents the image as a two-dimensional vector of intensity values extracted from the pixels. Since the recognition is based on appearance, the correspondence is obtained through direct comparison of these vectors values or another a representation derived from them. According to [76], this approach is proven functional, however its performance is dependent on environmental constraints (illumination, scale, background and pose stability) and sensor capability (noise intensity), besides presenting high computational cost due to the classification in a high dimensional space. Most techniques aim to reduce the image representation into sub-spaces, focusing on the conservation of significant discriminating data.

Sirovich and Kirb [77] were the first researchers to apply a Principal Component Analysis (PCA) [78, 79] to obtain compact facial image representations. The method demonstrates that any image image can be efficiently represented in eigenpictures coordinate spaces, and that any face can be reconstructed through a small collection of its eigenpictures and their corresponding projections. Through Sirovich and Kirb's research, Turk and Pentland [80, 81] reported that projection along eigenfeatures coordinates could be corresponded in order to classify and recognize features. Exploiting this concept, a facial recognition system that builds Eigenfaces was developed, corresponding the eigenvectors of dominant eigenvalues of facial covariance matrix to determine

an image input identity. The Eigenfaces is an image representation that drastically reduces the dimensionality of the original space. The method was tested in a scenario composed by 2500 images of 16 subjects with different variations of orientation, scale, illumination and resolution, achieving recognition rates of 96%, 85% and 64% for lightning, orientation and scale variation, respectively. The algorithms capabilities was extended in several ways in [82], being tested in a data set with 3000 subjects and over 7562 images.

Turk and Pentland's approach was enhanced to a modular "eigenfeatures" system to deal with localized variations in facial appearance [82]. The description of the face in an eigen space is augmented by additional higher resolution details in terms of fiducial landmarks (Figure 3.3). This enhancement provide a slight upgrade in performance, producing better results than the ones reported in the basic eigenfaces approach. The Figure 3.4 depicts an example of this outperforming. PCA applications functions correctly in cases of a single image per subject, however in cases of multiple images per subject (according to Belhumeur *et al.* [51]) choosing the projection that maximizes the total dispersion results in conservation of unwanted variations due to illumination and facial expressions.



(a) Examples of facial features training templates.

(b) Detection Result.

Figure 3.3: Training of feature detection template-based and detection result. These features are used to enhance Turk and Pentland's approach [82].

Moses *et al.* [83] states that "the variations between the images of the same face due to illumination and lighting direction are almost always larger than image variations due to a change in face identity" as shown in Figure 3.5. In order to overcome this liability, a method using Fisher's Linear Discriminant Analysis (LDA) [84] is proposed to maximize the ratio of the between-class scatter and the within-class scatter, supposedly obtaining a better classification scenario than PCA. Conducting numerous tests in 330 images of 5 subjects, the Fisherfaces' (label granted for the method) reported results indicates that this method is more efficient to handle the recognition over simultaneous manipulated variations of illumination and expression. Swets and Wang [85] previously announced similar results when the same procedures were applied not only in face images, but in general objects as well (presenting an accuracy of 90% in a data set of 504 classes).

(a) Test Inputs.

(b) Eigenfaces matches.

(c) Eigenfeatures matches.

Figure 3.4: Matching results of test images for the eigenfaces and eigenfeatures approaches [82].

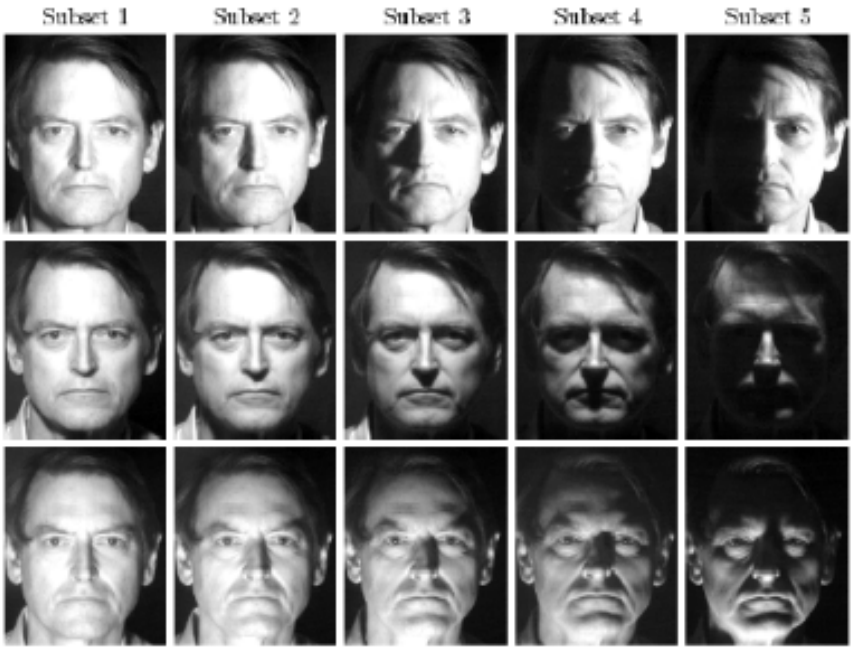


Figure 3.5: Facial appearance variation towards different source illumination angles [51].

One main drawback of PCA and LDA techniques consists in the fact that their analysis occurs merely on the euclidean structure of the space, failing to discover underlying structure if the face

images lie on a non-linear description of the space image, since these algorithms' nature present a linear discriminant approach. Since face images possibly lie in non-linear sub-manifolds [86, 87], non-linear techniques were proposed to focus on this theme, *e.g.* Isometric Feature Mapping (ISOMAP) [87], Laplacian Eigenmaps [88], Embedded Manifold [89] and Laplacianfaces [90].

The eigenvectors found by PCA depends exclusively in pair-wise relations between pixels, however existent methods provide base vectors that rely on higher-order relations among pixels. Independent Component Analysis (ICA) [91] is a generalization of PCA that aim to decompose an independent image representation rather than an uncorrelated one. Bartlett *et al.* [92] applied ICA under two architectures, one treat images as random variables and pixels as outcome, and the other is the inverse representation. Both ICA representation outperformed the PCA results for expression and daylight variation (along the whole day). A classifier combined both the ICA and PCA representation achieving a better performance. Other researches [93, 94] indicates that ICA outperforms PCA in most circumstances.

Besides the mostly known sub-space methods and their variations, other sub-space representations were proposed. Foon *et al.* [95] integrated diverse wavelet transforms and non-negative matrix factorization claiming to obtain higher verification rates in comparison to eigenfaces. In [96], an intraclass sub-space is built and the classification distance-based, where weighted distances between faces and each intraclass sub-space.

3.1.3.6 Machine Learning (ML) and Artificial Intelligence (AI)

Machine Learning and AI consist in techniques in which classifiers can learn new concepts and draw useful conclusions from the acquired knowledge [97]. In terms of face recognition, the approaches based on these elements learns either from select data extracted from known faces (training set in a supervised learning) how to classify random faces identity or from pattern within the data determines labels by itself and insert a random face in a determined label (unsupervised learning). A combination of both cases can be applied on different subsets of data creating a semi-supervised environment [98].

In the method proposed by [99], fifty principal components were extracted and an auto-associative neural network reduced its representation to five dimensions, in which a standard multi-layer perceptron (MLP) classify this final representation. Despite of the favorable results perceived, the training and testing data set composition was relatively simple, without variations in illumination, pose or scale.

Weng *et al.* [100] proposed an hierarchical neural network trained by unconventional techniques (instead of the standard descendant gradient) for classification, reporting favorable results in a data set composed of 10 subjects.

Lawrence *et al.* [101] reported a recognition rate of $96,2\%$ in a data set of 400 images of 40 subjects using a hybrid neural network composition that combined local image sampling, a self-organizing map neural network for space dimensionality reduction and a convolution neural network. The application of eigeinfaces [80, 81] led to an accuracy of $89,5\%$ in the same data set.

The substitution of the self-organizing map for a Karhunen–Loève transform and the convolution neural network for a MLP resulted in recognition rates of $94,7\%$ and 60% respectively. Figure 3.6 displays all the pipeline possibilities proposed in this work.

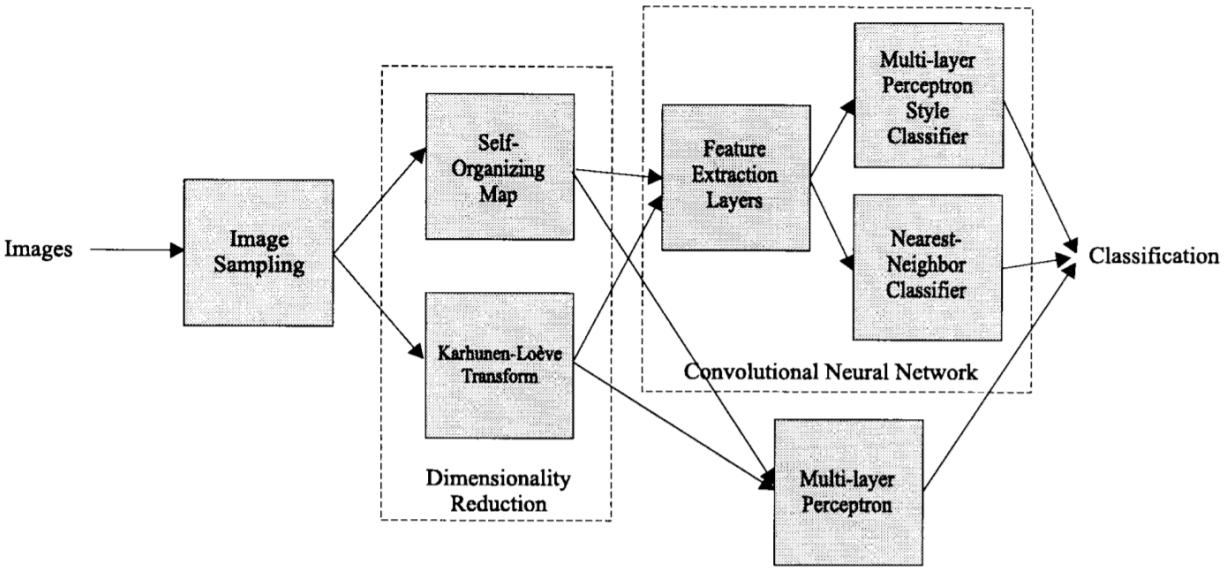


Figure 3.6: Pipeline of Lawrence *et al.* methods. The pipeline presents the combination of techniques available to reach classification [101].

Li and Yin [102] introduced a system using wavelet transforms to decompose the image to three levels. The Fisherfaces [51] method is applied to each of the three low-frequency levels. This method fused individual classifiers using a Radial Basis Function (RBF) neural network, outperforming the Fisherfaces and the individual classifiers results in a 40 subject composed data set. Elyan and Demerial [103] proposed a PCA-based approach, in which the projection vectors were classified using a feed forward neural network, outperforming an eigenfaces method combined with a nearest-neighbor classifier.

Zhang *et. al* [104] proposed a solution that learns a similarity function describing the level of confidence that two images belong to the same person, similar to [105]. Facial features are selected generating Local Binary Pattern (LBP) histograms of sub-regions of the image and the Chi-square distances between the correspondent LBP histograms are chosen as discriminating features. The Adaboost [46] algorithm is applied to select highly-efficient LBP features and define a similarity function based on the linear combination of weak learners LBP feature-based. Experimental results reported from the use of FERET (Face Recognition Technology) [106] frontal face images an outstanding recognition rate higher than $97,9\%$ using less features than other similar methods proposed, like the ones presented in [107].

Melin *et al.* [108] divided the image faces in three region (eyes, nose and mouth) assigning a neural network module to each region. A fuzzy Sugeno integral was used to merge the output of each module and achieve a conclusion upon the input image identity. The tests were performed in a small data set formed by 20 subjects, reporting that the modular solution outperformed the monolitic network.

A technique based on problem decomposition (dividing the recognition into a one against one comparison) was developed to create multiple binary classifiers. Each binary classifier represents a single class (subject), presenting as outcome if the input belongs to that class. The output of all binary classifiers combined grants a global overview of the identity classification. For binary classifiers presenting probabilistic outputs, Pair-Wise Coupling (PWC) can be used to cluster these outputs into a set of posterior probabilities, however if a testing image does not belong to either classes related to the binary classifier, then the output of that classifier is meaningless and has a chance to interfere negatively in the global result. Based on this technique and aiming to correct meaningless classifiers an algorithm proposed labeled PWC-Correcting Classifier (PWC-CC) [109] trains a classifier to separate the outcome of a class separation classifier from all other classes, *e.g.* for each binary classifier separating a class c_i from a class c_j a new classifier is trained to separate these classes from all the other known classes. Despite of PWC-CC being more consistent and efficient than PWC, the algorithm still presents drawbacks. A new PWC-CC method (NPWC-CC) [110] is proposed and their report states that NPWC-CC outperforms the PWC-CC on tests realized on ORL data set. In [111], the optimal PWC (O-PWC) approach is presented demonstrating better recognition rates than PWC method. Both the NPWC-CC and the O-PWC methods used Support Vector Machines (SVM) as binary classifiers, and the posterior probabilities were computed using Platt's suggested method [112]. Support Vector Machines (SVM) is considered one of the most effective machine learning techniques for pattern classification, and other researchers proposed methods using SVM as binary classifiers [113, 114, 115, 116].

Another machine learning tool exploited to perform face recognition task are the Hidden Markov Models (HMM) [117]. Samaria and Harter [118] used a one-dimensional HMM to achieve a peak accuracy of 87% in the ORL data set, and a pseudo two-dimensional HMM [119] to enhance performance to 95% accuracy rate. Nefian and Hayes [120] reached a recognition rate around 98% in the same training and testing scenario using embedded HMM face models, claiming to achieve a significantly faster and scale invariant algorithm compared to Samaria's [119] implementation.

3.1.3.7 Multiple Classifier Algorithms

The performance of any classifier is susceptible to diverse factors data related such as distribution, variation and composition, however an specific classifier's performance might be more sensitive or invariant to a factor in comparison to others [5]. Due to this variant sensitivity between classifiers, a trend emerged to combine individual classifiers in order to aggregate their complementary information in order to compose a robust system that overcome (even if partially) the classifiers drawbacks. These algorithms are nominated as Multiple Classifier Systems (MCSs) [121] and are extensively researched. An example of this approach is proposed by Lu *et al.* [122], that merged results from PCA, ICA and LDA using an strategy to integrate a RBF neural network and sum rule (depicted in Figure 3.7). Marcialis and Roli [123, 124, 125] combined the results of PCA and LDA algorithms.

Achermann and Bunke [126] used simple fusion rules, like majority voting, rank sum, Baye's combination rule to integrate weighted responses from three classifiers based on profile and frontal

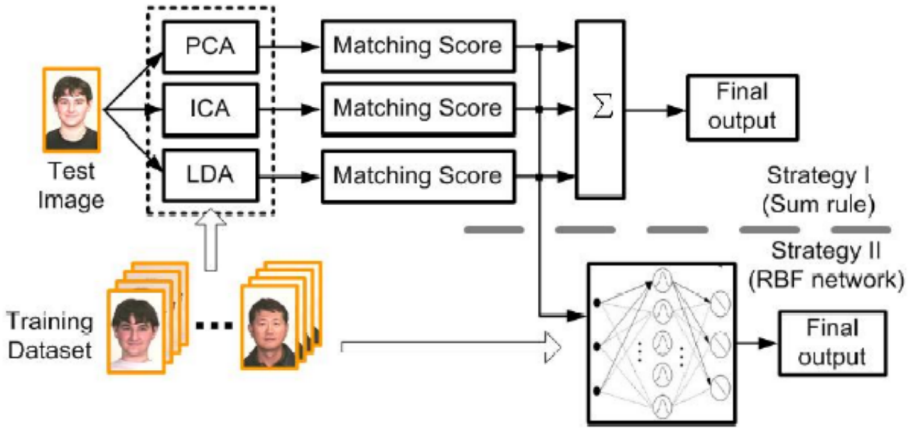


Figure 3.7: Classifier combination proposed by [122].

view faces. Tolba and Abu-Rezq [127] fused the decision from RBF and LVQ networks. Wan *et al.* [128] proposed a hybrid model of SVM and HMM. Kwak and Pedrycz [129] applied Fisherfaces methods in three distinct subsets from the face images integrating the result through Choquet's fuzzy integral.

3.1.3.8 Deep Structured Algorithms

Until modern researches, most machine learning and signal processing techniques exploited shallow-structured architectures that typically contained one or two non-linear feature transformations [130, 131]. Shallow architecture proved to be efficient in the solution of well-defined controlled problems, however its modeling limitation and representing power may fail to perform in complex real world application related to natural signals, like human speech, sounds and natural languages, natural images and visual scenes.

Human information processing mechanism (like vision and audio) suggest that a deep architecture is needed to extract complex elements and construct a rich sensory input representation. Therefore, it is expected that advanced researches focus on the development of efficient and effective deep learning algorithms to handle favorably complex signal processing tasks [131]. These methods represent the state of the art of face recognition study field, surpassing the human recognition capability in terms of accuracy and speed [3].

The deep learning concept originated from neural network researches aiming to create bio-inspired methods (methods inspired in the theories of human body functioning) to realize a digital learning process, also known as artificial neural networks (ANN) [132]. An ANN is based on a cluster of connected units called artificial neurons. Each connection (synapse) between neurons can transmit a signal, that may suffer transformation in each traversed neuron in the propagation path. Neurons and synapses may present a weight associated to them in the learning process, intensifying or diminishing the signal intensity forwarded. Deep Neural Networks (DNN) is a type of ANN that contains numerous hidden layers between the input and the output layer [132, 133]. The DNN have potential for modeling complex non-linear relations, clustering layers of data transformation

and classification layers. Usually DNN are feedforward networks, where data flow exclusively from the input to the output direction without returning to previous layers, but there are Recurrent Neural Networks (RNN), in which dataflow has no unique direction.

Sun *et al.* [134] elaborates a hybrid network of Convolutional Neural Networks (CNN) and a classification RBM (restricted Boltzmann machine) trained to perform verification. The lower part of the network presents 12 groups of 5 CNNs (each responsible for a part of the face image), presenting four convolutional layers and a max-pooling layer. After defining an average output and reorganizing the data in groups, the data reach the last layer represented by a classification RBM with two outputs that indicates same or different class for the pairs. The CNNs and RBM are trained separately and later the whole model is jointly fine-tuned using back-propagation. Using a training data set labeled "CelebFaces" containing 87,628 images of 5,436 celebrities they reported $91.75 \pm 0.48\%$ accuracy on the LFW in the *unrestricted with label-free outside data* protocol and $92.52 \pm 0.38\%$ following the *unrestricted with labeled outside data* protocol. Later they proposed the DeepID method [135] consisting in four convolutional layers, followed by a max-pooling layer, the DeepID-layer consisting in 160-dimensional fully connected layer (connected to the third and fourth convolution layers) and a softmax layer. This method globally align faces using five landmarks as reference, then trains sixty CNNs (which output is reduced using PCA) followed by a joint Bayesian model for classification. The reported accuracy was of $96,05\%$ using CelebFaces data set, and $97.45 \pm 0.26\%$ using CelebFaces+[135] data set (which has 202,599 face images of 10,177 celebrities from the web.) under the unrestricted with labeled outside data protocol. A DeepID2 method is proposed [136] built based on DeepID [135] using cross-entropy loss and a verification signal for network training, using PCA to reduce the face representation to a 180-dimensions representation. A total of seven networks are trained using different subsets o selected patches, combining the joint Bayesian scores with an SVM, achieving $99.15 \pm 0.13\%$ accuracy under the *unrestricted with labeled outside data* protocol. A DeepID2 enhanced version was proposed in [137] labeled as DeepID2+ increasing the number of feature maps to 128 in the four convolutional layers, the DeepID size to 512 dimensions and expanding the training set merging CelebFaces+ and WDRef[138] data sets. They report an accuracy of $99.47 \pm 0.12\%$ on *unrestricted with labeled outside data* LWF protocol, using the joint Bayesian model from 25 trained networks. Another enhancement is proposed in the DeepID2+ pipeline labeled DeepID3 [54], using a deeper network Inception layers [139] and stacked convolution layers. Similar to DeepID2+, they include unshared weights in later convolutional layers, max-pooling in early layers and the addition of joint identification-verification loss functions to branched-out fully connected layers from each pooling layer in the network. They train two networks, one using the stacked convolution and the other using the proposed Inception layer [139], combining the feature from both networks into a 30000-dimensional vector, reduced to a 300-dimensional representation using PCA, followed by learning a joint Bayesian model, achieving a $99.53 \pm 0.10\%$ verification accuracy on LFW *unrestricted with labeled outside data* protocol.

Face++ is an algorithm proposed by Fan *et al.* [140] creates a new structure nominated pyramid CNN conducting supervised training of a deep neural network. A four-level Siamese network was applied on four face landmarks, concatenating the outputs for classification. This method provided a verification accuracy of $97,3\%$ on the LWF *unrestricted with labeled outside data* protocol. The

same group proposed a ten-layer network trained on 5 million labeled faces of 20000 subjects achieving an identification accuracy of $99.50 \pm 0.36\%$ [141].

Facebook developed a face recognition algorithm labeled DeepFace [142] that performed a 3D face frontalization method and trained a DNN for classification. 3D frontalized RGB face images are taken as input in a networks composed by a convolutional filters layer, followed by a max-pooling layer, another convolutional filters layer, localy connected layers and fully connected layers. Three different input image types (3D aligned RGB, grayscale with gradient magnitude and orientation and 2-D aligned RGB) are used, and their scores are combined using a kernel support vector machine (SVM). Using the restricted protocol, this reaches 97.15% accuracy. Under the unrestricted protocol, they train a Siamese network (initially using their own SFC data set, followed by two epochs on LFW pairs), reaching 97.25% after combining the Siamese network with the above ensemble. Finally, adding four randomly-seeded DeepFace networks to the ensemble a final accuracy of $97.35 \pm 0.25\%$ is reached on LFW following the unrestricted with labeled outside data protocol.

Google team also proposed a method labeled FaceNet [53], basically mapping the images in compact euclidean spaces (128-dimensional representation) followed by a *triplet loss* training, creating an output capable of classifying identity through euclidean distance. The structural model of FaceNet is depicted in Figure 3.8. The FaceNet algorithm fulfills the verification task (through threshold checking between two mapped entries), identification (represented by a k-Nearest Neighbor classifying problem) and clustering (return a group of input-alike images, using agglomerating clustering or k-Means). The method discuss the use of two main architectures of Deep Convolutional Neural Network, one of them being Inception-based [139]. The neural networks are trained using Stochastic Gradient Descent (SGD) with standard back-propagation and AdaGrad [53]. The stated performance presents an accuracy of $98.87 \pm 0.15\%$ on *unrestricted with labeled outside data* LFW protocol (using central image crops) and an accuracy of $99.6 \pm 0.09\%$ using a proprietary face detector layer on the same protocol in both architectures.

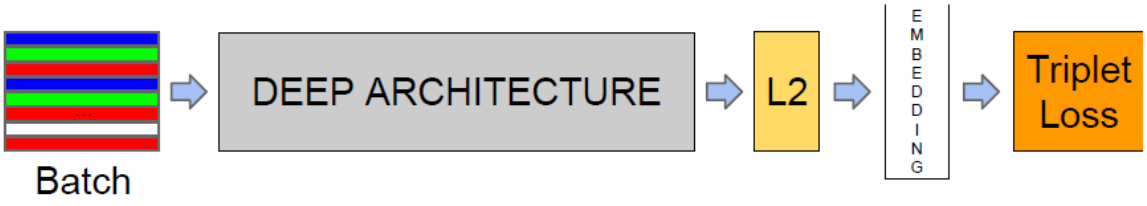


Figure 3.8: Structural model of Google's FaceNet Algorithm [53].

3.1.3.9 Advantages and Disadvantages

The main advantage of holistic approaches consists in not destroying any information originally present on the image, altering the global image representation instead of focusing in limited regions around landmarks [69]. However this same property has drawbacks due to the fact that most of these approaches follow the initial hypothesis that all pixels from the original image are

equally important. Consequently, these applications turn computationally expensive and require a high degree of correlation between the testing and training sets (if the method requires a learning process), that may lead to high sensitivity to illumination, pose, scale and other significant image variations. Many of these algorithms were modified and enhanced to compensate these variations' interference and dimensionality reduction techniques have been explored to diminish processing cost and discard irrelevant content (presents the collateral effect of potentially losing discriminating information [143]). In general, these methods achieve better performance rates compared to feature-based techniques. A FERET assessment [144, 106] indicates that probabilistic eigenfaces [105], Fisherfaces [51] and EBGM [67] (that despite of being a feature-based algorithm, its performance depends on holistic neural networks for feature detection) were the best techniques in the epoch for face recognition applications.

3.1.4 Hybrid Approaches

Due to the dimension of possibilities and researches performed along the years, the natural course of the facial recognition approaches were to merge feature-based and holistic techniques together into a single solution. Some works mentioned are actually classified as hybrid, for example the EBGM [67] that uses holistic for feature extraction and based on feature relations classifies the input identity, or the eigenfeatures [82] that uses holistic classification over sub-images formed around features. The main idea of hybrid approaches consists in presenting at least one step that extract features and their correlation, and one step based on a global analysis of data or sub-data [145].

Heisele [146] developed two holistic methods and a component-based approach (hybrid between feature-based and holistic), applying SVMs as classifiers. The global methods used sub-space representation of square sub-windows of size 40, converting into feature vectors using the gray scale pixel values of the region. The component-based approach used an automatic feature detection method to acquire 14 features to compose a feature vector for SVM classification; The data set tested contained a total of 9567 images of 5 subjects with variation in pose, illumination and scale. One training and testing scenario was formed using a training set of 974 images and a testing set of 8593 images, achieving recognition rates of 93% with the component-based approach and 78% in the best performing holistic approach. Another test was conducted with a training set of 1383 images, achieving recognition rates of approximately 90% and 68% for the component-based approach and the best performing holistic approach, respectively.

A Probabilistic Decision-based Neural Networks (PDBNN) was used in the work of Lin *et al.* [147] for both face detection and recognition, and feature detection (eyes). An experiment using a data-set composed of 40 subjects containing 150 images of each subject with orientation and pose variation without background elements interference tested the algorithm in two circumstances, the first one use the whole data set (labeled as original set) providing a recognition rate of $84,64\%$, while the second scenario uses a data set (labeled valid-set) containing the images that correctly detected faces and eyes location representing a total of 2176 images that generated a recognition rate of $98,34\%$. In both cases the training set presented 20% of the whole data-set.

A compilation of information towards the main approaches research in this field of study is presented in Table 3.1, granting an overview of general techniques, the benefits commonly present in the methods and the usual drawbacks.

Table 3.1: Main Techniques and Approaches for Face Recognition.

Approach	**Features**	**General Techniques**	**Benefits**	**Disadvantages**
Feature-based	Locals	Geometric Relations between fiducial points (distance, area, orientations, etc), Dynamic-Link Structures (Graph, Grids).	Relatively Robust to pose variation, rotation and can be invariant to scale, illumination. Compact representation and fast processing	Rely on efficient automatic feature detection. Hardship in arbitrary choice of discriminating features.
Holistic	Global	Sub-spaces representation (PCA, LDA, ICA), static classifiers, machine learning and AI (neural networks, SVM, HMM, deep architectures).	Preserves image originality. High generalization capability (dimensional reduction and enhancement).	May present high sensibility to environment variations. High processing cost and may require training. Dimensional reduction may cause discriminating information loss.
Hybrids	Locals and Global	Combination between holistic and feature-based techniques (component-based, eigenfeatures).	Select significant features presenting low sensitivity to appearance variation. Avoid irrelevant information.	Depends on appropriate training for feature detection. Generalization capability reduced.

3.2 FERET and LFW

The Face Recognition Technology (FERET) initiative from the National Institute of Standard Technology (NIST) and the Labeled Faces in the Wild (LFW) initiative are benchmarks developed to standardize the face recognition research environment and stimulate face recognition researches granting a fair comparative scenario for algorithms testing [3, 148, 144].

3.2.1 FRVT

The FERET presents the Face Recognition Vendor Test (FRVT) that presents a competition for both identification (one against one comparison) and verification (one against many comparison), besides related topics such as gender classification [149]. The FRVT measures advancement in accuracy and speed of face recognition algorithms enrolled in galleries containing at least 10 million identities. The evaluation primarily uses standardized portrait images, and quantifies how accuracy depends on subject-specific demographics and image-specific quality factors.

The FRVT 2018 evaluation used four datasets (frontal mugshots, profile views, webcam photos and wild images). The primary dataset is comprised of 26.6 million reasonably well-controlled live portrait photos of 12.3 million individuals. The three smaller datasets contain more unconstrained photos: 3.2 million webcam images; 200 thousand side-view images; and 2.5 million photojournalism and amateur photographer photos. These datasets are sequestered at NIST, meaning that developers do not have access to them for training or testing. The last dataset, however, consists of images drawn from the internet for testing purposes so while it is not truly sequestered, its composition is unknown to the developers [52]. The evaluation was executed in three phases.

3.2.2 LFW

The Labeled Faces in the Wild project [148] was elaborated with the main goals of stimulating face recognition research in unconstrained images, providing an easily handled database with low access conditions, composed by protocols to standardize assessments encouraging fair and meaningful comparisons and grant full access to results for easy comparison and replication of the algorithms.

The LFW presents an specific set of image pairs, previously organized for training purposes, alongside a label identifying if the pair contains images of the same or different subjects. The training specification and instructions are represented by protocols. Originally LFW proposed two protocols of *image-restricted* and *unrestricted* images. The unrestricted protocol allows the creating of new training pairs using combination of the already conceded pairs. Since many researchers started to introduce additional training data (from outside LFW) to their methods aiming to achieve better performance, new protocols were elaborated [3, 150], being those Unsupervised, Image-Restricted with no outside data, Unrestricted with no outside data, Image-Restricted with label-free outside data, Unrestricted with label-free outside data, and Unrestricted with labeled outside data.

The work elaborated by [3] grants a discussion about the importance of studying restricted protocols, bringing up topics like the utility of the methods for other tasks, the comparison of statistical efficiency and assymptotic optimality, and the comparison of automated methods and human efficiency.

3.3 Face Recognition from Other Sensory Inputs

A face recognition research is mainly focused in the identification of individuals in 2D intensity images, however various studies have been directed to the exploration of other modalities, *e.g.* 3D images, depth images, infrared and others [5].

3.3.1 Infra-red

Infra-red images tend to present low sensitivity to illumination variation, leading to an exploitation of these images in face detection and recognition tasks. Despite this fact, [151] states that since infra-red facial images reveals veins and tissue structure that are ubiquitous to each subject (like fingerprints), visible spectrum face recognition techniques should provide favorable results when applied to such images. However there are multiple factors that discourage this type of application, *e.g.* the substantial cost of thermal sensors, the low resolution and high noise sensitivity of images, the lack of infra-red images' data sets, the opacity property of infra-red radiation against glass (leading to occlusion in subjects wearing glasses and other glass assets) [152], and infra-red images sensitivity to environmental temperature changes and subjects' metabolic processes [153].

In [151], the eigenface technique [80, 81] is applied to a data set containing 288 low resolution images (manually aligned) of 24 subjects in 3 different poses. The recognition rates reported were of 96% for frontal faces and 45 degrees orientation variation and 100% to profile pictures. Wilder *et al.* [154] compared the performance of three algorithms in a data set of visible spectrum and infra-red images of 101 subjects, concluding that there wasn't significant improvements in recognition taxes from one modality to another.

Socolinsky *et al.* [155] tested the application of eigenfaces [80, 81] and ARENA [156] in a data set containing visible spectrum and infra-red images of 91 subjects (captured under varied conditions of illumination, facial expression, glasses usage) using a sensor that capture images in both modalities simultaneously (Figures 3.9,3.10, and 3.11) claiming that infra-red images significantly outperformed the visible spectrum images in all classification experiments conducted. Selinger and Socolinsky [153] used the same data set and tested the performance of four know algorithms (PCA,LDA,LFA and ICA) achieving the same conclusion previously stated, although the authors admitted that the superiority of infra-red approach may occur due to the fact that infra-red interference factors (temperature variation, ventilation, and others) were not measured nor manually varied during the experiments.

Chen *et al.* [157] collected diverse image data sets (both visible spectrum and infra-red) of 240 subjects under expression variation and illumination condition (not all images were acquired simultaneously, presenting time lapses of ten weeks). The purpose of this data set was to study the facial appearance change effects along time in the performance of eigenfaces algorithm in both modalities, concluding that both modalities present similar accuracy rates, however the visible-spectrum images surpassed infra-red modality whenever a significant time lapse was present between train and test image acquisition. They attributed this low performance to variation in temperature patterns on the same subject (from both environment and subjects' metabolic variations) and to infra-red images sensibility to eyes manual location. They also discovered that the FACEIT software [158] provided better results than eigenfaces in both modalities, however these results were outperformed by a combination of two classifiers using the sum rule [159]. Posterior experiments [160] in a bigger data set preserved the same performance hierarchy. Other approaches proposed the merge of visible spectrum and infra-red images to enhance the performance in recognition rates [152, 161, 162, 163].

Figure 3.9: Samples extracted from data set in [155].

Figure 3.10: First five visible eigenfaces [155].

Figure 3.11: First five LWIR eigenfaces (infra-red) [155].

3.3.2 3D Model/Data Methods

The main argument in favor of 3D information usage for facial recognition resides on the exploration of shape features, based on curvature of face elements (like cheeks, jaw line, and

shape of the forehead) without interference from illumination variation, orientation and background confusion present in 2D systems [164, 165, 166, 145]. Another fact to use depth data is the advance of technology, permitting direct manners of introduction and registration of complex shape information for digital analysis. A consequence and drawback of this approach consists in the complexity and cost in computational terms.

3.3.2.1 3D data acquisition

There are multiple forms of generating 3D data, varying from sensory reading and image-based modeling reconstruction. Tibbalds [165] describe the following 3D model acquisition:

- Scanning Systems: Scanners laser-based are producing accurate models, however the commercial cost of these are usually expensive (despite of recent low commercial cost sensors like Kinect [13]).
- Structured Light Systems : Use stereo-graphic vision principles to determine depth data. The main advantage is the equipment composition, requiring only cameras and a projection system.
- Stereo-vision Systems: These systems aim to extract 3D information from two or more 2D images from the same object with a baseline displacement (or angular variation). They are limited by the object's image feature extraction capability, the feature extraction algorithm and its output quality and the feature matching algorithm to allow conclusive correspondences.
- Reverse rendering/shade from shading: Focus on the model construction using illumination and object's physical properties knowledge.

The mostly used technique based on 3D model reconstruction from image consists in the multi-view stereo (MVS) reconstruction. The goal of these techniques is to achieve complete 3D model reconstruction of an object taken from know camera viewpoints [167]. The work in [167] focus on methods that retrieve a dense object model from calibrated view, and according to it the multi-view stereo algorithms are roughly categorized in four classes, voxel-based approaches [168, 169], deforming polygonal meshes based algorithms [170, 171], multiple depth maps (merging them in a single 3D model) [172] and patch-based methods [11] [10]. Figure 3.12 depicts a pipeline of patch based a MVS model reconstruction proposed by [11]. According to [11] MVS can be used to obtain objects, scenes and crowded scenes reconstruction. In [10], a Struct from Motion (SFM) approach is presented to estimate camera view-points using a Batch Adjustment (BA) or a Partial Batch Adjustment (PBA, faster than BA) permitting a dense point cloud generation using a patch-based reconstruction.

In the past years, great interest arose in 3D face recognition supporting the elaboration of new techniques. One of the first approaches described in [7], where the surface principal curvatures are calculated through depth data and further used to locate fiducial landmarks (*e.g.* nose, forehead,

Figure 3.12: An overall approach of a patch-based MVS model reconstruction. From left to right: A sample input image; detected features; reconstructed patches after the initial matching; final patches after expansion and filtering; and the mesh model [11].

eyes, neck, chin, and others). The faces are normalized in a standard position and re-interpolated onto a regular cylindrical grid. The spatial volume between two normalized surfaces is used as similarity measure. Experiments using 8 subject faces led to recognition rates of 97% for individual landmarks (features) and 100% for the whole face. Another approach described in [173] uses profile instead of frontal images, capturing 3D data through triangulation and comparing the acquired data. The method requires user cooperation and background constraints to provide efficient performance. [164] uses 3D data to normalize results from a facial detection algorithm to properly construct an environment for face recognition, using the 3D as complementary enhancing data.

Castellani *et al.* [174] approximate face depth images (estimated through stereoscopic analysis using B-Splines Multi-level [175] and SVMs) using the numerical approximation coefficients for classification. Other techniques [166, 176, 177] first project 3D data into 2D intensity images and then process these new representation as standard intensity images. Varying recognition approaches like 3D local feature-based recognition [178], local and global geometrical cues [179], profile analysis [180, 181, 182, 183] were also proposed. In [5], methods using PCA and HMM are noticed.

Diverse approaches proposed the integration of 2D texture and 3D shape information. These methods use PCA in intensity images [184, 185, 186] and profile intensity images [187], Iterative Closest Points (ICP [188]) [189, 190], Gabor wavelets [191], Local Feature Analysis [183] and other known methods of data representation, extraction and correspondence. Wang *et al.* [191] extracted 3D shapes templates from depth images and texture templates from gray scale images, applying PCA separately in both template modalities for spatial reduction, concatenating both texture vectors and classifying the identity using SVM.

In Goswami *et al.* [34] a method is proposed to extract an entropy map from the depth map and the RGB image of a person and a saliency map from the RGB image, computing a histogram of gradient (HOG) from these maps and classifying them by a Random Forest (RDF). Other work from Goswami *et al.* [22] has presented improvements adding a geometric attribute computation from depth map fiducial points, creating the called RISE (entropy and saliency maps) and ADM (geometric attribute relation) descriptors.

In Hu *et al.* [192] was proposed a face recognition for a user tracking robotics application, using the depth map from head detection and the RGB image for recognition by illumination normalization, head pose correction and face space projection. Bormann *et al.* [193] implements a similar algorithm to Hu *et al.* algorithm, using Fisherfaces[51] space parametrization, a Support Vector Machine (SVM) and Nearest Neighbor techniques for classification. Zhou *et al.* [21] proposed a three-dimensional face recognition using 7 feature points and a two-level Cascade Classifier, formed by a Decision Tree Classifier in the first level, and an improved Euclidian Distance classifier in the second level. Saleh and Edirisinghe [194] proposed an Eigenface-based method, training models with eigenfaces applied to the normal images and depth images, under different illumination conditions. Chowdhury *et al.* [35] proposed a method based on machine learning, that trains a Neural Network to reconstruct the depth map from a color image, using the color image and the real depth map as input elements, and classifying the reconstructed depth map through another multi-class neural network.

The work in [19] presents a table containing varied methods of 3D facial recognition, describing the number of subjects in the data set, the size of the utilized images, the total number of images, the data type utilized, the main applied technique and the recognition rate of each.

Russ *et al.* [23] algorithm applies a Haussdorff distance in range images, the experiment used a data set composed of 200 subjects and 398 images in total. The reported recognition rate was of 98% not presenting any false positive occurrences (erroneously classifying an input). A drawback mentioned in the research is the execution time needed due to the algorithm's high computational cost. Lee *et al.* extracted geometrical features (curvature, length, angle) from geometrically localized fiducial landmarks. A testing scenario was composed using two sensors for image capturing, the first sensor consist of an structure light sensor (Genex 3D FaceCam) obtaining test images and the second sensor is a full laser scan (Cyberware) to obtain model images (since laser scan provide high detail and shape quality of scanned surfaces). Two classification architectures are presented, the first uses curvature values extracted from landmarks as correspondence and the second applies a SVM into a feature vector. The first classification scenario uses a 20 subjects data set claiming a recognition rate of 95%, while the SVM-based scenario uses a data set composed by 100 subjects claiming a rank-1 recognition rate of 96%.

Li *et al.* [24] presented a method based in descriptor correlation of 3D fiducial landmarks. The proposed algorithm is inspired by SIFT (Scale-Invariant Feature Transform) descriptor methodologies aiming to develop a robust algorithm to deal with face recognition difficulties (occlusions, expression and pose variation). The methodology is divided in three steps as depicted in Figure 3.13. Initially key points are detected in the point clouds, selecting the points through maximum and minimum local curvature values (principal curvature), then key point local descriptors are generated in form of histograms storing shape and gradient (extracted through canonical direction attribution in a quasi-daisy configuration), finally a dictionary gallery of descriptors is formed and a Fine-Grained Matcher is used to correlate the descriptors of testing images with the formed training dictionary. The data set used by Li *et al.* is composed of 4666 3D faces of 105 subjects (61 men and 44 women), presenting 34 facial expression, 13 poses and 4 occlusions for each subject. They reported a recognition rate of $98,8\%$, $99,2\%$ and $91,1\%$ for expression variation, occlusions

and pose variation, respectively. A scenario containing the whole data set provided a recognition rate of $96,6\%$. The algorithm generalization is tested in other data set.

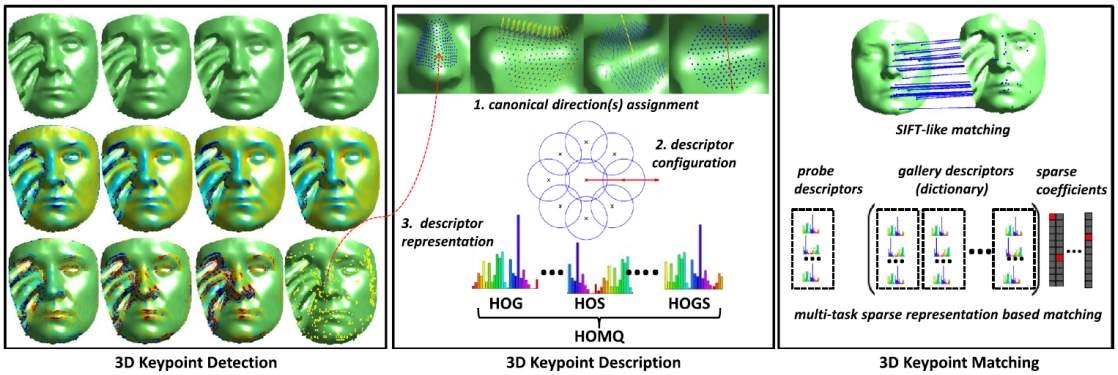


Figure 3.13: General Representation of Li *et al.* [24] methodology. Initially key points are detected (left block), then local descriptors of these points are built: attributing canonical directions, descriptor configuration and histogram representation (middle block), and finally a key point correspondence is made: dictionary construction and descriptors comparison (right block) [24].

Chapter 4

Methodology

This chapter presents the proposed methodology applied in this facial recognition approach. The following method is composed of two main algorithm pipelines, which both contain common stage of data preparation, followed by classification criteria, and validation process. The commons steps are related to the database pre-processing, downsampling the database to obtain only the data of interest. Next, each algorithm presents a feature extraction process followed by a matching procedure that provides a basis for the classification of an input subject.

During the whole elaboration of this work and the solution to the three-dimensional face recognition problem, the analysis of raw features and aspects of two-dimensional images (that are also present in the database) were avoided, to focus in a solution independent of any standard two-dimensional image data, exclusively exploiting the three-dimensional data (despite of using it to filter facial data). The database studied in this case consist of point clouds obtained from a low resolution device, that should be carefully treated in order to avoid noise and interference from external factors that may jeopardize the data description and feature identification in the pre-processing step. The database pre-processing corresponds with facial data segmentation, building a three-dimensional space represented in a point cloud structure. In both algorithms here presented, this step is equally implemented, loading the data stored on file into a data structure containing only facial data.

Succeeding the pre-processing comes the feature extraction step of both methods' pipelines. That is the point of the methodology in which the techniques diverge since each procedure extracts different features for data processing and classification criteria. The first approach exploits the curvature projection of each point stated in the facial point cloud, composing a RGB image of curvature projection entitled Face Curvature Map (FCM) [4]. The second approach selects a subset sample of points, treating those as key points, either through salience or through a voxel grid uniform spacing, and estimates their normal vector component through surface estimation defined by neighborhood distribution to connect these key points in pairs (under various topology, including Bozorth's algorithm composition [36]), projecting their components in planes to exploit their spacial correlation through angular and distance differentiation. The stochastic validation finishes both pipelines based on F-measure analysis [195][196].

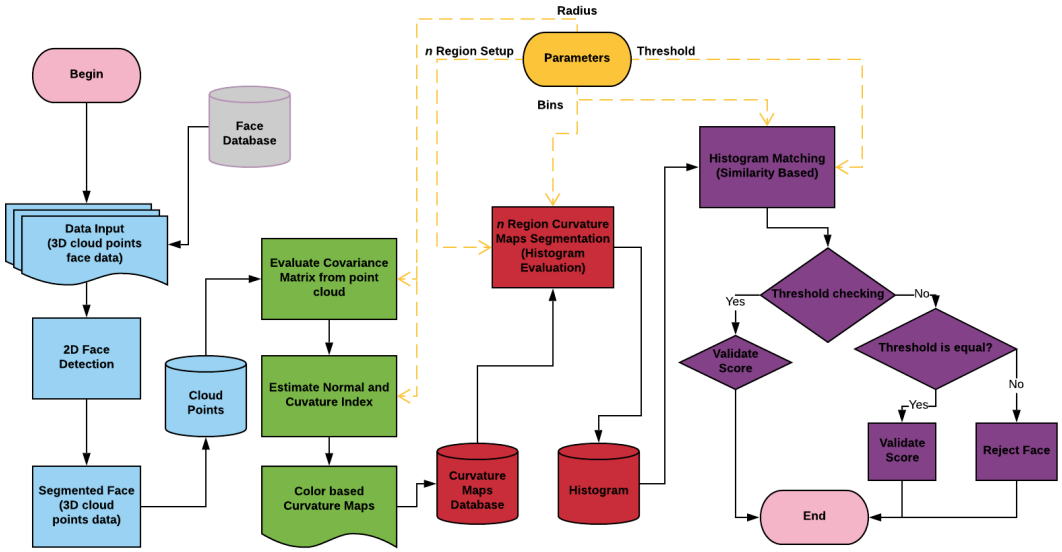


Figure 4.1: Flowchart of FCM pipeline. The light-blue flow corresponds to the data input and pre-processing to obtain the facial point cloud structure. The following green colored processes refer to the feature extraction procedure, including the Covariation Matrix calculation and the curvature index estimation, assembling the curvature maps. The red flow consists of the face curvature map analysis, constructing the pixel values histograms for classification. The purple flow presents the classification and validation steps, ending the pipeline.

The feature extraction and data processing are followed by the classification (which the procedure is eccentric to each pipeline). The FCM-based approach generates histograms of the subjects, storing the count of pixels values in determined regions of the FCM in interval bins, and comparing these histograms to obtain a difference score between FCMs. The Bozorth's based approach generates a matching score between the data structure that contains the pairs of points, through spatial and angular distribution analysis of these pairs, permitting induction of the subject's identity. Both pipelines finish the methodology with a validation procedure, over stochastic evaluation that applies thresholds over the scores obtained in the previous steps, measuring the performance of both algorithms to deal with the task of recognizing a person in the presented scenarios.

To provide a better understanding of both pipelines, an overview flowchart of the main procedures in both pipelines was elaborated (FCM flowchart presented in Figure 4.1 and Bozorth's Approach flowchart presented in Figure 4.17), showing the sequential application of each step, the input of each process and the main parameters involved in them.

4.1 Database Pre-Processing and Feature Extraction

This section is focused on detailing the procedures common to both pipelines developed, describing the database elements followed by the operations applied on specific items to prepare data for later feature extraction (that will also be covered in details further in this section). The pre-processing step loads the point clouds into memory and filter data to obtain only the points

that belong to the face. Figure 4.2 presents the operations of the pre-processing stage.

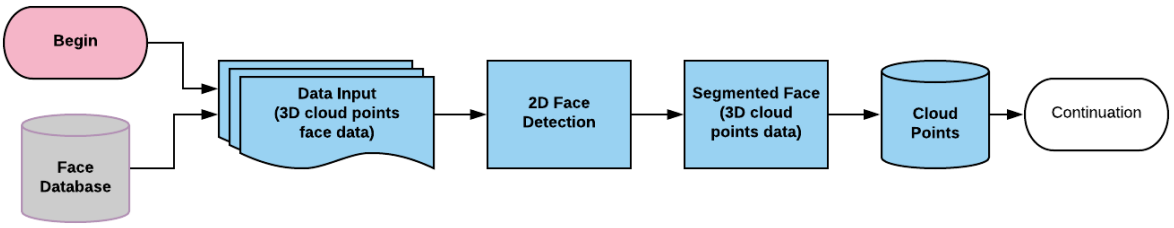


Figure 4.2: Data pre-processing steps. The first step consists of a wrapper that receives input data files to convert into whole scene point clouds. The second step applies the Viola-Jones [43] algorithm in the RGB images to detect the face location. The last step filters the points in the region defined from the previous step to generate a facial point cloud.

In the FCM pipeline, these initial steps head until the curvature maps formation, which will, later on, be used for data comparison and classification of the subjects, being represented by the light-blue and green operations in its flow chart. In the Bozorth's pipeline, the light-blue steps related to the pre-processing are precisely the same, only differentiating in the green operations that lead to normal vector estimation and ending on key points extraction.

4.1.1 Database

The database presented in this work was used to check the functionality of both pipelines. The data is provided by [197]. This database is composed of thirty-one subjects in thirteen different head orientations as well as four different facial expressions (representing mixed emotions), presenting three different samples for each one of these cases mentioned. The provided database was captured within the same environment conditions using the Microsoft Kinect device (the first version of the device) [13], and later published after smoothing the data through pre-processing (spikes removal, gaps and holes fulfillment throughout the face surface).

The obtained samples previously mentioned are composed of an RGB image and the depth information of the scene captured from Kinect stored in a data file (extension dat). The depth values are represented in a numeric interval of [-1,4095], presenting a negative element (-1) to inform an error in the depth acquisition in the current pixel. The neutral element(0) to report depth values that are too close to the sensor, and the maximum value (4095) to inform that in such pixel, there was either an object in the limit capacity of the sensor (or that was beyond the reach of the sensor). The RGB images mentioned are sized in 1280x960 pixels, while the depth maps are sized in 640x480 pixels, half the size of the RGB images. Figure 4.3 displays an example of the database elements described.

This database was developed initially for facial detection in-depth analysis, adding the possibility to find faces even if the person presents a different expression than a neutral look. However, it is applied to verify the capability of the developed pipelines to execute an identity recognition towards the low-resolution point cloud data.

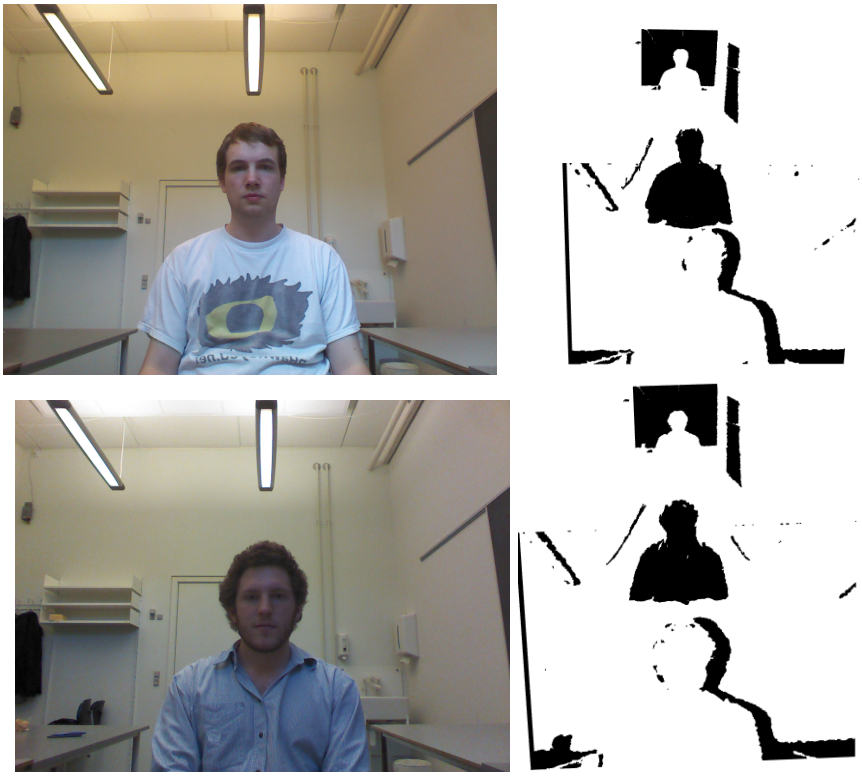


Figure 4.3: Samples of the database elements before the pipelines pre-processing. This illustration provides two subjects in a frontal position, showing the RGB image to the left and the Depth map representation (rendered as point clouds composed of black points) to the right [197].

4.1.2 Facial Segmentation and Point Cloud Assembly

The database contains useless information (e.g. objects in the environment or the entire upper-body of the subjects), thus a data filtering procedure must be performed to obtain only the region of interest of the depth information (facial data). Since both the RGB image and the depth information are taken from the same scene, they are correlated, being only differentiated by a translation baseline. This information is exploited to delimit the facial region of the depth map using the RGB image facial region. The first step of the pre-processing consists of removing the irrelevant and corrupted data from the depth map, getting rid of the reading errors from the sensor, and the limit values (too close and too far) obtained, remaining only the data represented in Figure 4.4. This filtering process is referent to the first block of operations of Figure 4.2.

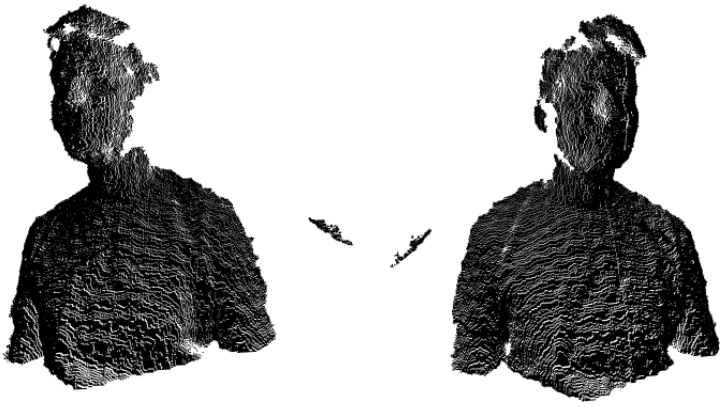


Figure 4.4: Multiple views of the relevant data obtained after the first pre-processing phase (represented as point clouds). This filters the bad capture errors (bad readings, maximum and minimum range) from the sensor.

To obtain the face location in the RGB image, the Viola-Jones algorithm [43] (also knows as a Haar-cascade) is applied defining an area which the face is contained (represented in Figure 4.5). Thenceforth a scale transformation is applied to this region, along with the baseline translation, to determine the depth data correspondent to the face of the subject, obtaining the data shown in Figure 4.6. In the pipeline, this method is represented by the second operation block of the section present in Figure 4.2.

All results from this pre-processing step are defined as a subset(S_{face}) of three-dimensional data points delimited by the horizontal and vertical spatial dimensions containing the face of the individual/subject to be recognized/verified and described by Equation 4.1.

$$S_{face} \subset \Omega \in \mathbb{R}^3 \tag{4.1}$$

where Ω is a set of three-dimensional data in spacial domain $\mathbb{R}^3$. The subset S_{face} consists of points $p_k(x, y, z)$ where $k = 1 \dots m$, and m is the total number of points composing this subset.

After discovering the facial region in the depth map, it is possible now to assemble the point cloud referent to that region. The assembly is assisted by the Point Cloud Library [42], which

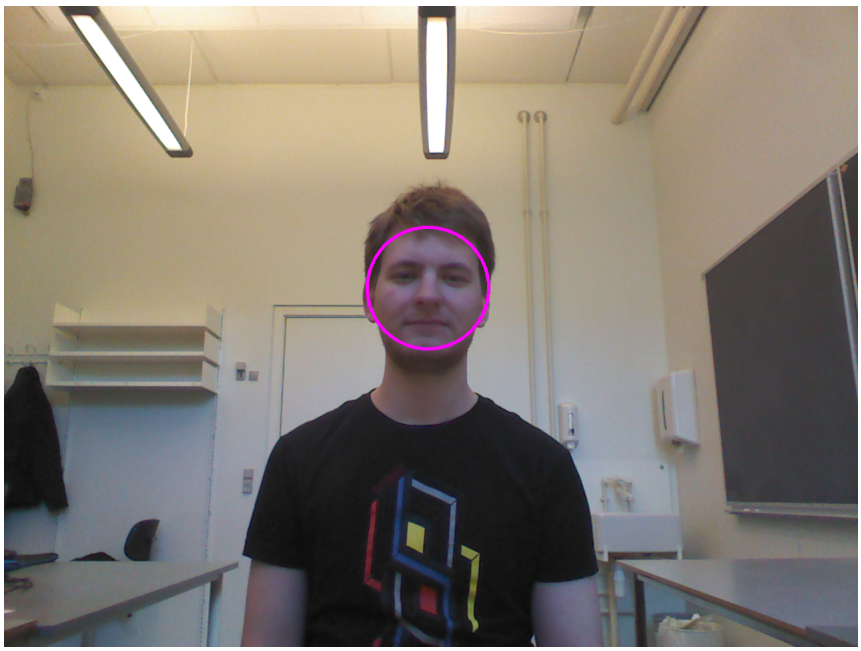


Figure 4.5: Facial Segmentation on RGB Image using the Viola-Jones [43] algorithm. This image represents one of the subjects from the database [197].

provides an appropriate data structure to load the desired data on memory and supplies the user with many built-in functions focused on point cloud data structure. It assists in tasks such as visualization, filtering, feature extraction, key point acquisition, registration, segmentation, etc. For each subject, the point cloud is built point by point, since the parser defined by the library couldn't identify the data distribution of the depth map files. The output of such operation is the point cloud of the subjects, which is represented by the third and fourth blocks in Figure 4.2.

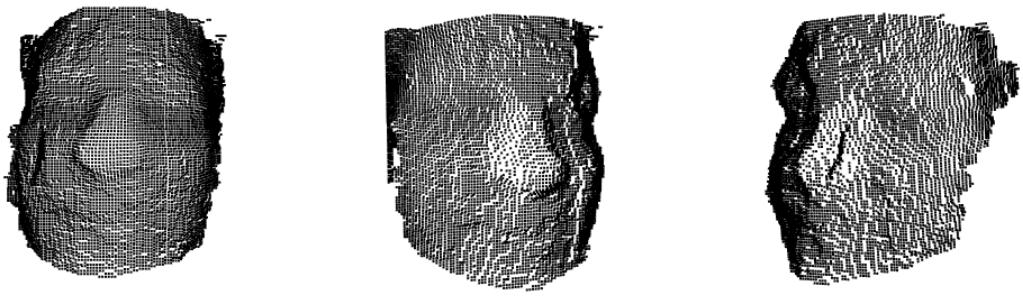


Figure 4.6: Multiple viewpoints from the remaining relevant data of Figure 4.4 after the facial segmentation.

To verify if the correct facial data was acquired, the visualization module of the Point Cloud Library was used to render the remaining points, obtaining results like Figure 4.6.

4.1.3 Points Neighborhood and Covariance Matrix

When it comes to point cloud data structure, a single point itself does not present enough information to reach spatial conclusions of a region [198]. However, when a group of points in an enclosed defined region is analyzed together, spatial features and descriptions may be estimated, providing information about curvature, surface shape, local geometrical features, local spatial distribution, and other spatial information. Due to the importance of these local characteristics, a geometrical region is defined around each point to establish a neighborhood and compute the spatial distribution of the cloud.

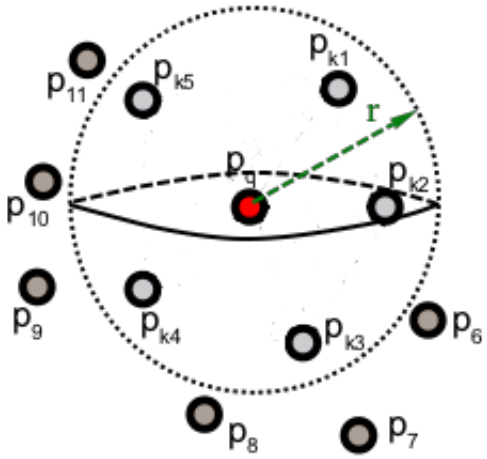


Figure 4.7: Spherical Neighborhood defined around a sample point. p_q represents the current point, while p_{ki} symbolize the neighbor points inside the spherical vicinity of radius r and p_i neighbor points outside the vicinity.

The neighborhood geometry chosen was a sphere around each point to obtain a determined radial limited description, establishing this way the symmetrical bound around its center (Figure 4.7). To determine the spatial distribution of a point's neighborhood the covariance matrix definition is applied, as mentioned in the pipeline of Figure 4.1. The covariance matrix is calculated using the centroid of the sphere as a reference, achieving the spatial variation within its limits. The covariance matrix is relevant to both pipelines since they are used either in the curvature index estimation as in the normal vector component estimation. Figure 4.8 illustrates the covariance matrix and the feature extraction operations in the pipeline.

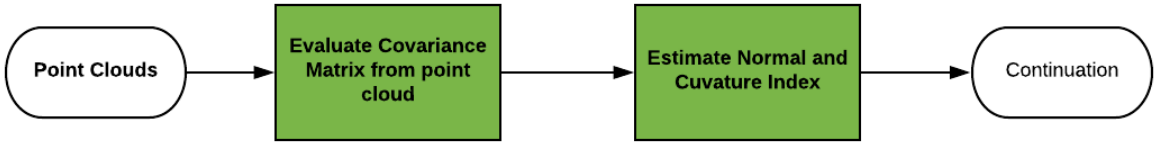


Figure 4.8: After the conception of the Point clouds, the pipelines follow to the covariance matrix calculation and the feature extraction (normal vector and curvature estimation).

The covariance matrix is mathematically expressed by Equation 4.2, consisting of the square deviation of the points towards the centroid of the neighborhood of the current point being analyzed, providing a 3x3 matrix of deviation towards the main axes representing the local three-dimensional space. The matrix calculated can specify the local surface around the point and extract important features crucial for the implemented methodology.

$$M_n = \frac{1}{x}\sum_{i=1}^{x}(p_i - \bar{p})^T \cdot (p_i - \bar{p}), \tag{4.2}$$

in which:

M_n is the covariance Matrix of neighborhood n;

x is the number of points inside the neighborhood n;

$\bar{p}$ represents the spatial vector of the centroid (dimension 1x3);

p_i represents the spatial vector of a point contained in the neighborhood (dimension 1x3).

The matrix has crucial information to obtain discriminating features, possessing high value on its eigenvalues and deviation elements. The eigenvalues of the covariance matrix can be mathematically expressed as Equation 4.3.

$$M_n \cdot \vec{v_j} = \lambda_j \cdot \vec{v_j}, \quad j \in \{0, 1, 2\} \tag{4.3}$$

in which:

$\vec{v_j}$ are the eigenvectors of the covariance matrix M_n.

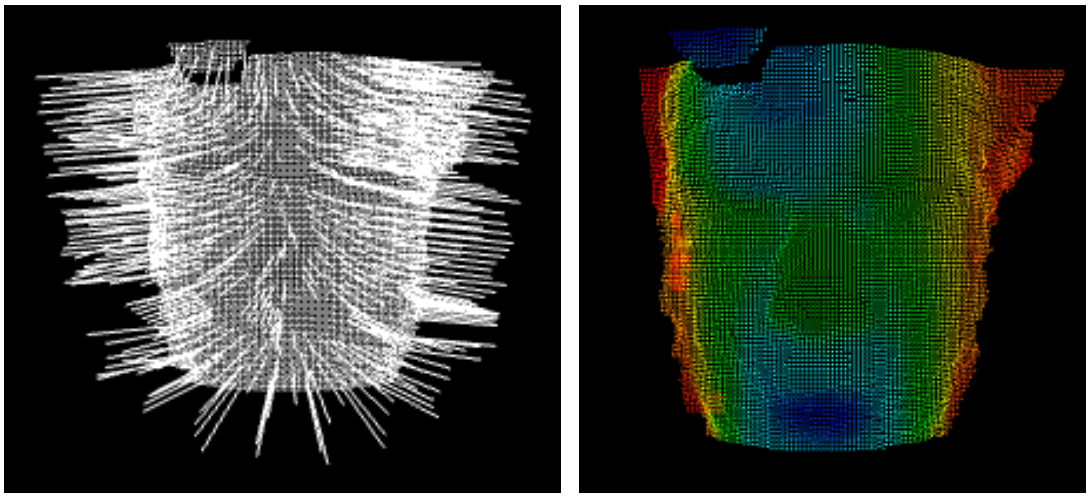
λ_j are the eigenvalues of the covariance matrix M_n.

j refers to the index of the eigenvalue/eigenvector. A 3x3 matrix is composed by three eigenvectors/eigenvalues, allowing the values of j to assume any value of the set {0,1,2}.

4.1.4 Normal Estimation and Curvature Index

As previously mentioned, the normal surface vectors and the shape of the surface are crucial information to identify a pattern [41]. For the estimation of the normal vectors, a tangent plane is estimated through least squares to fit the surface description of each neighborhood. It provides the normal vector spherical coordinate angles [199], which through the manipulation of the covariance matrix eigenvectors, generates the normalized (using the vector magnitude as a maximum value) components of the normal vector [41].

The curvature index is obtained through eigenvalues correlation, verifying the ratio of spatial deviation in the principal coordinate planes [198]. The curvature index is calculated by Equation 4.4 written below.



(a) Normals (b) Curvature

Figure 4.9: Rendering of Normal Vectors (the parametrization chosen doesn't render all the vectors) and Curvature Index rendering.

$$\sigma_k = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}, \tag{4.4}$$

in which σ_k is the curvature index of the point k, the λ's are the three eigenvalues of the covariance matrix, and the λ_0 is the eigenvalue with lowest magnitude. The normal estimation and curvature index rendering result are shown in Figure 4.9.

Since the spherical radius previously defined determines the neighborhood of each point, this radius is critical to the estimation of the curvature index and the normal vectors. Figure 4.10 shows an example of how the size of the neighborhood may interfere in the values of the curvature index of curvature. In the normal vector estimation, the variation is less aggressive, not being able to perceive visually (in the visualizer just as Figure 4.9).

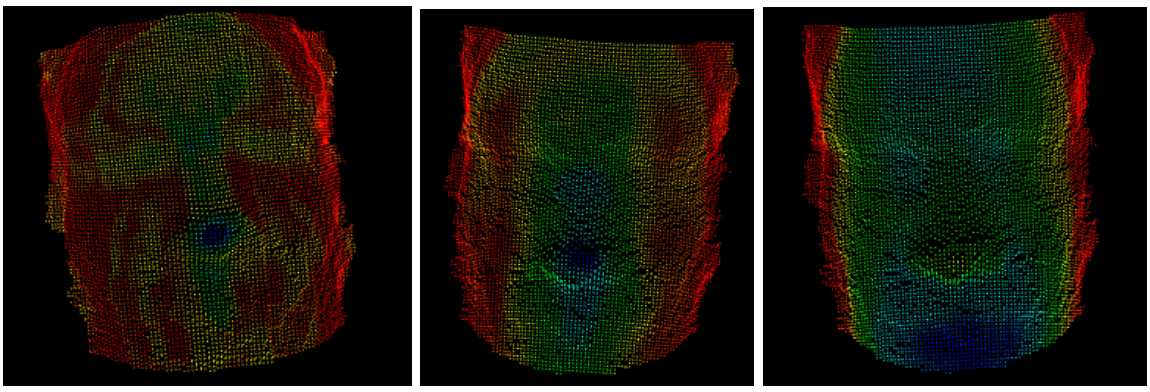


Figure 4.10: The difference in curvature rendering between the outputs of the same subject for different spherical radius set for the neighborhood: The left image is generated with the lower radius value, while the right image is created with the highest radius value between the three images.

4.2 Face Curvature Map Pipeline

This section introduces the stage in which both pipelines diverge since the data pre-processing is finished. This section will develop on the remaining process of the FCM pipeline shown in Figure 4.1, detailing the handling of the pre-processed data, the output of each method, and the progress of the whole application to reach the identification of a facial input data. In short terms, this section will explain how the face curvature maps are generated, followed by a gray color transformation and a n-section division of the curvatures maps to extract the pixel value histograms. It ends with the histograms comparison, to calculate a differentiation score between subjects, and a validation process of the recognition over score ranking. Figure 4.11 shows the steps before validation.

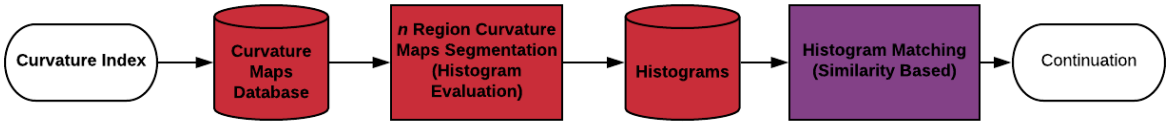


Figure 4.11: FCM operations before the validation steps. After the Feature Extraction the FCMs are formed, being divided in regions and converted to gray scale to form histograms that are compared to generate a matching score.

4.2.1 Face Curvature Map Extraction

The Face Curvature Maps (FCM) consist of a projection transformation based on the viewpoint of the three-dimensional space represented by the point clouds, resulting in an image (RGB), where each pixel represents a point, and the color is related to the curvature index value, describing it as a curvature-based method. The curvature index are normalized values (obtained in the Equation 4.4) in which the bottom limit (zero value) represents the minimal curvature of the local face surface (approaching or being equal to a plane surface), and the top limit (value equals one) represents the maximum curvature of the local surface (the region with greater distribution).

The curvature maps idea came from the analysis of the Visualize class of the PCL, in which the curvature index is represented by colors (translated by a Look Up Table), attributing a color for each point, just like in Figure 4.9.

The PCL uses a data array of the VTK (Visual Tool-Kit) Library to store the values of the curvature index of each point. Through this array, it is possible to create a visualization table that normalizes the values of the curvature based on its range, returning the color levels for each of the RGB spectrum colors (red, green and blue). In virtue of this function, a planar projection of the points from any viewpoint can generate an RGB image, which allows a frontal projection of the face to be transmuted into a frontal RGB facial image, as shown in Figure 4.12. This method should be avoided in view-points that occur overlaps of points.

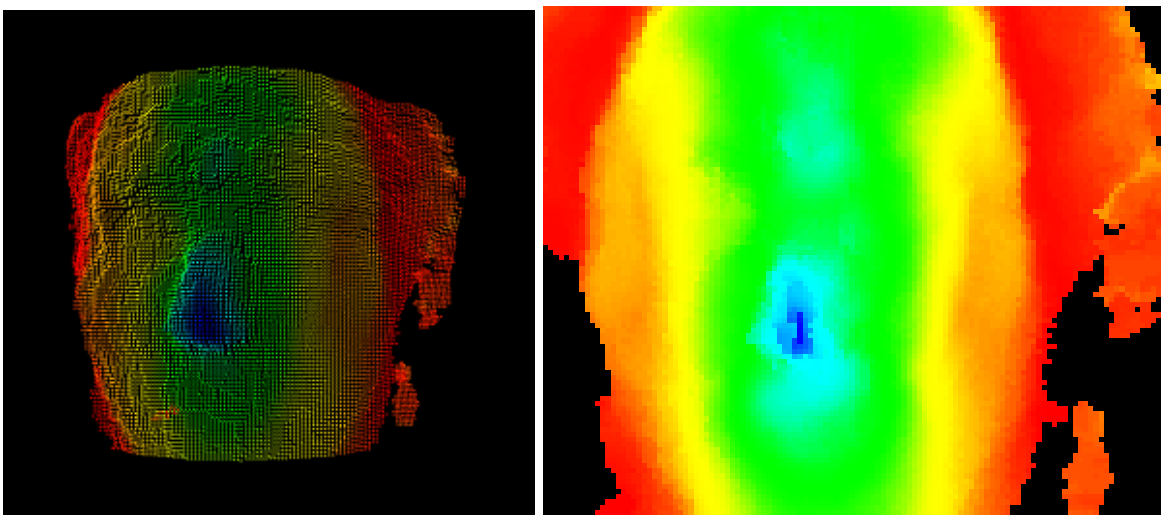
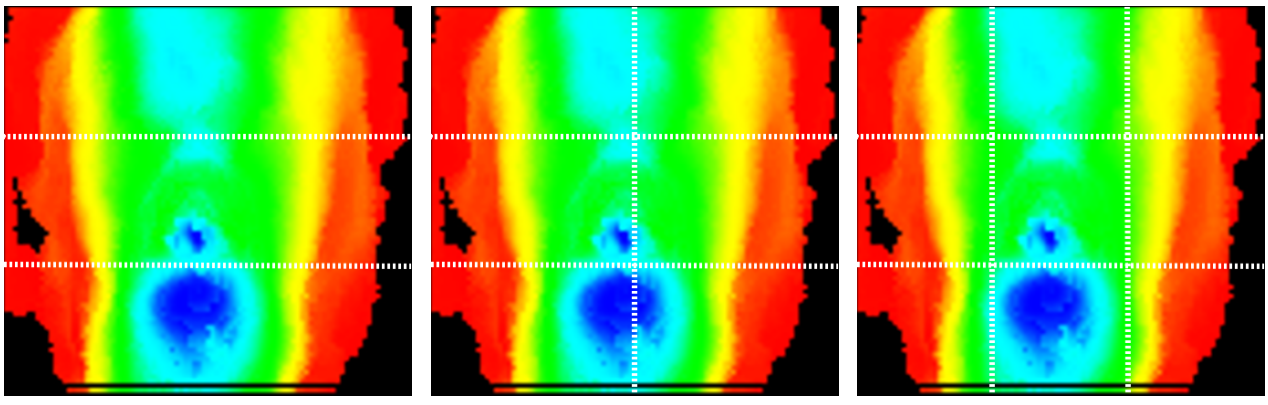


Figure 4.12: Transposition of the Curvature Index Values to a Curvature Map: The left image is a point cloud rendering in which the point colors represents the curvature intensity, while the right image is the image projection generated from the curvature color map.

4.2.2 Division of the Face Curvature Map

The FCMs are a global representation of the subjects' face surface curvature. Despite a noticeable visual discrepancy of the subjects' FCM, two different subjects may present similar average pixel count distribution of their FCMs, that could generate high similarity scores and jeopardize the classification results. It happens due to the scattered variability of local features, so to avoid this circumstance, the FCM is divided in sub-windows to focus on local features and guarantee a better discriminating description of the subject. The number of regions (either for height and width) is defined as an input parameter (set as a dynamic variable label n) by the user. The height and width division can also be settled by separate variables with the proper parameterization. Figures 4.13a, 4.13b, and 4.13c give a better understanding of the FCM division.



(a) 3 region division (b) 6 region division (c) 9 region division

Figure 4.13: Division of the Curvature Map into equally spaced areas.

4.2.3 Gray Scale Transformation

The grayscale transformation of the FCM consists of a simplification process for further histogram extraction and classification of the subject. The grayscale transformation converts the data extraction from a multi-channel problem (Red, Green, and Blue) to a single-channel problem (grayscale value), reducing the time complexity of the comparison of FCMs. Figure 4.14 shows an example of RGB to grayscale transformation.

The OpenCV Library [200] adopted to provide such transformation describes the process mathematically as written in Equation 4.5.

$$RGB[A]\ to\ Gray : Y \longleftarrow 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \tag{4.5}$$

in which Y is the grayscale value of the pixel and R, G, and B are the pixel values of the red, green and blue channels, respectively.

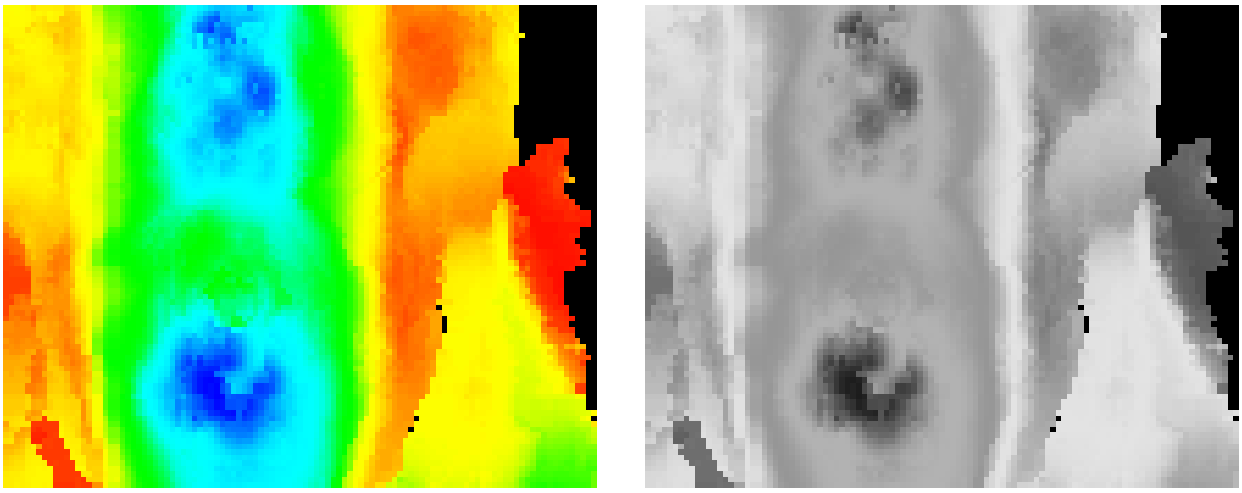


Figure 4.14: Transformation of the FCM from RGB to gray scale. This transformation is performed using Equation 4.5.

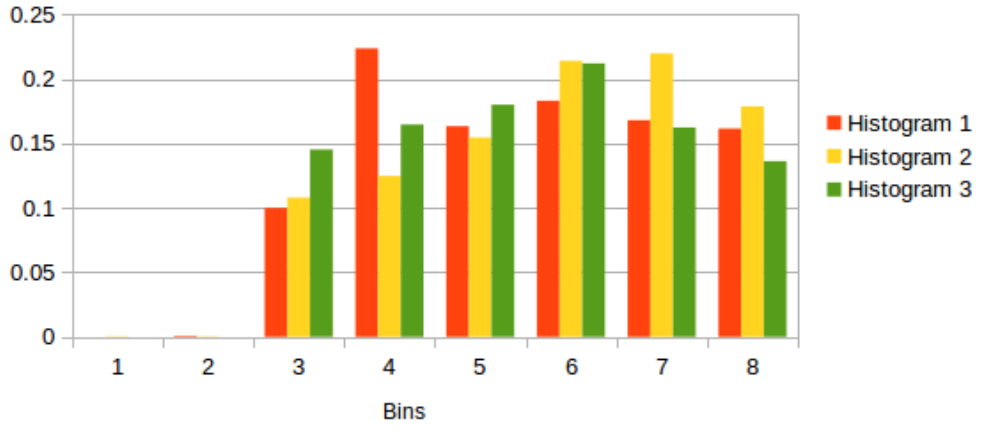
The division of the FCM and the grayscale transformation of the RGB concludes the last processes directly applied on the FCMs, being followed by a pixel value count to obtain the histograms for classification.

4.2.4 Histogram Construction

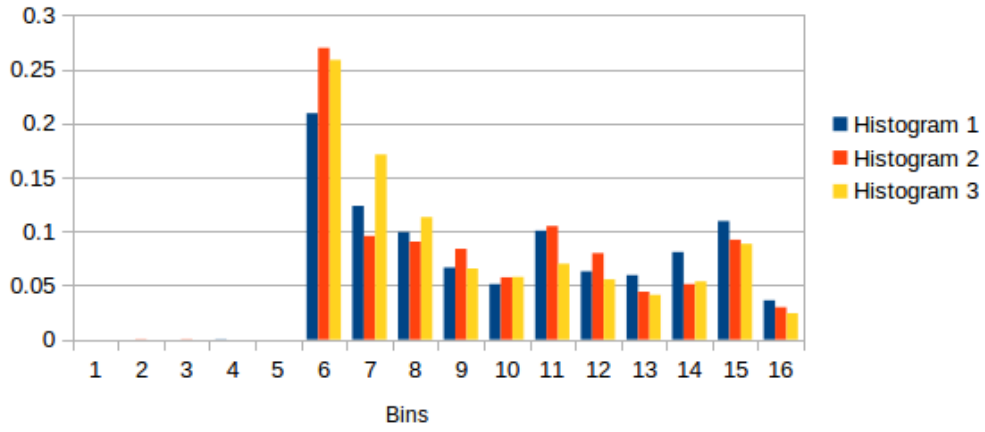
To transform the visual information provided by the grayscale FCMs into a better data-driven numerical representation, histogram of normalized pixel values count is extracted for each area obtained in the division process. The normalized pixel value count are stored into a parametrized number of bins (set as input into a dynamic variable labeled as *bins*). The bins correspond to the number of groups that the grayscale range (that fluctuates from 0 to 255) will be represented, so whenever a pixel value is found inside the range of a group, the bin represented by this group is incremented. To provide equally sized group values the total number of bins should be represented

by a power of 2. After filling all the bins with the respective pixel values, the histogram is finished with the normalization of the bins counts about the total pixel count of the image to avoid deception in the classification process due to different-sized samples (FCMs). The histograms are submitted to a similarity matching function, leading to a numerical score that represents the discrepancy between histograms.

Figure 4.15 presents two examples of different histograms extracted from random FCMs in a three region division, offering a visual notion of the constitution of a histogram of 8 bins and a histogram of 16 bins.



(a) 8 bins Histogram for a 3 region division



(b) 16 bins Histogram for a 3 region division

Figure 4.15: Histogram structure of three regions from the FCM. Each bin of the histogram represents a numerical interval of values from the grayscale, while the vertical component is the normalized count (occurrence frequency) of the values contained in such intervals

4.2.5 Histogram Comparison, Scores and Classification

A human being giving a glance at the FCMs in Figure 4.16 could easily distinguish both subjects achieving the same conclusion as presented, in which the first two images resemble the same subject, while the second one does not. The expectation upon the histogram comparison is

to obtain the same result through a numerical representation, being able to classify two curvature maps as the same subject or different subjects. For numerical evaluation, a histogram comparison is portrayed as a minimization function between the frequencies (normalized pixel count values) represented by Equation 4.6. This function returns a numerical score representing the discrepancy of the histograms, implying that the lower the score is, the higher the resemblance is.

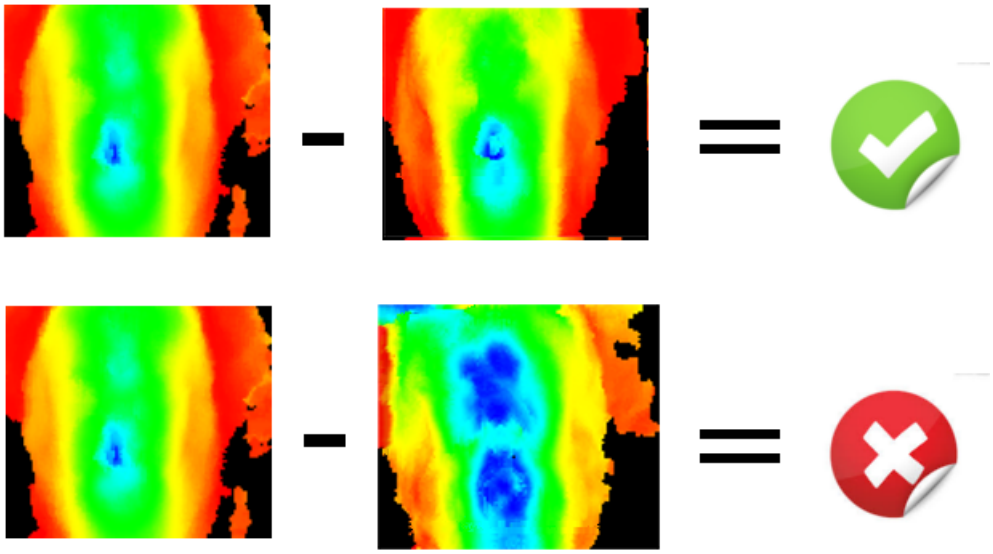


Figure 4.16: Classification process between FCMs. Visually there is a clear similarity between the first pair of subjects, while there is a clear variance between the second pair.

$$score_{(i,j)} = \sum_{m=1}^{n} \sqrt{\sum_{k=1}^{bins} (H_{in}^{k}[m] - H_{(i,j)}^{k}[m])^2} \tag{4.6}$$

in which:

H_{in} refers to an input histogram;

$H_{(i,j)}$ refers to one of the j-th samples of i-th subjects known as a database member.

k represents the current (k-th) bin of the histograms;

$bins$ are the total number of bins contained in the histograms;

m represents the current (m-th) histogram since each region of the FCM produces a histogram; and

n represents the total number of divisions (and the number of histograms) the curvature maps were submitted.

Since the score represents the average discrepancy between histograms of two FCMs, the lower similarity score should be chosen as the subject's identity if the unique task consists in establishing an identity to the input based on known subjects (it is not the case in this application). To discard

input data that does not represent a person or represent a subject which the identity is unknown (the subject does not belong to the database gallery), a threshold is applied to set an upper bound to the lower score, implying that if the discrepancy of the input subject is above the expected, it should be discarded, bypassing a false recognition. This process is also essential in the validation step.

4.3 Bozorth Method

This Section introduces the Bozorth's algorithm (as previously mentioned and explained in Chapter 2). It consists in a fingerprint matching algorithm implemented by the National Institute of Standards and Technology (NIST) [36], based on fingerprints minutiae that represent specific edge features of fingerprint lines (as mentioned by the NIST[36] and Madalla[49]). This pipeline resembles Bozorth's algorithm by selecting specific key points (meant to surrogate the fingerprint minutiae) through two different possible feature extraction applied on a point cloud context. This one is followed by a topology selection of the features linkage and the matching process proposed by Bozorth's algorithm, which is finished in a validation procedure, as shown in Figure 4.17.

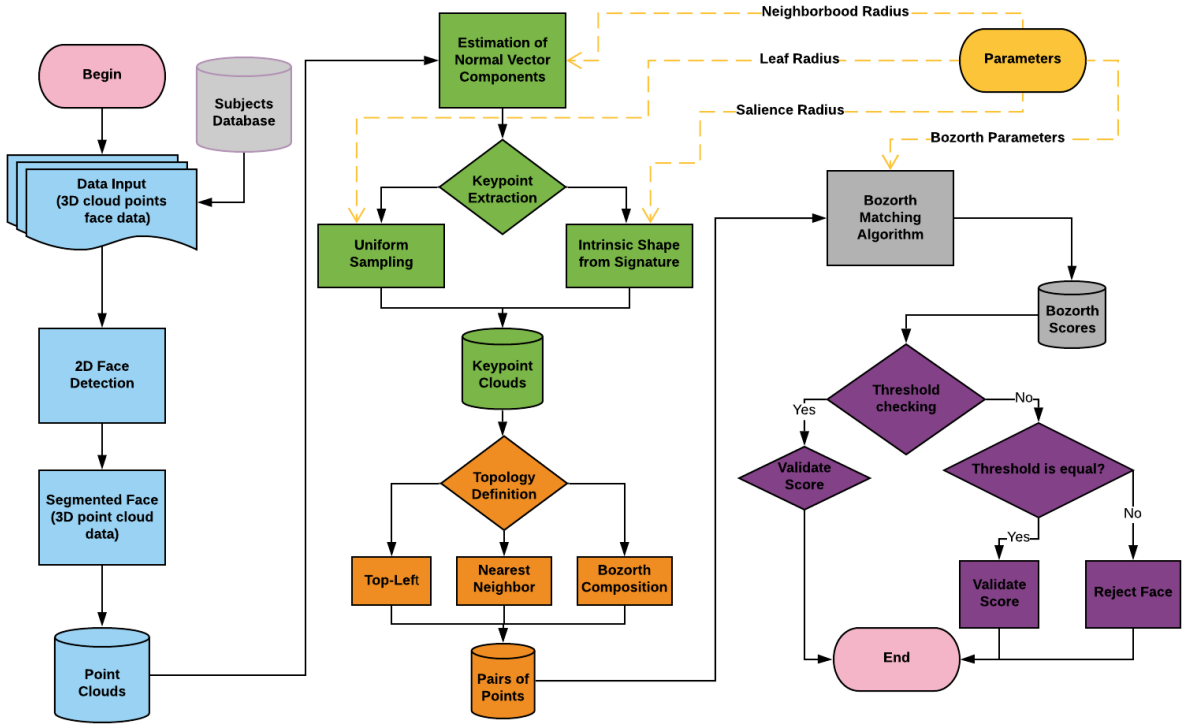


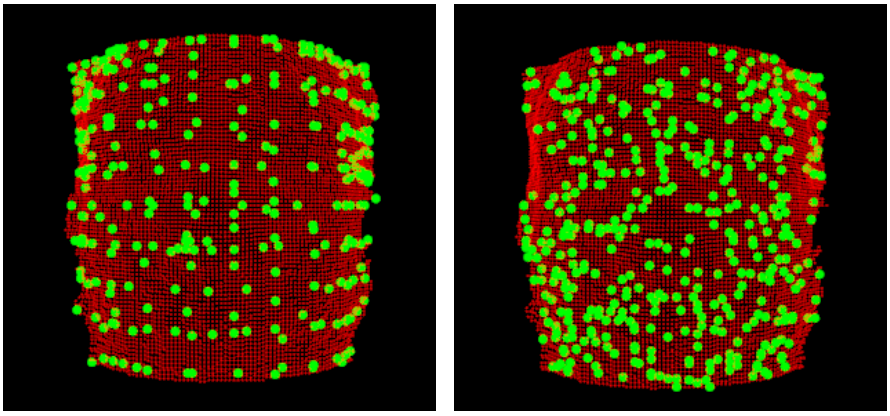
Figure 4.17: Flowchart of the Bozorth Pipeline. As previously mentioned the blue operations represent the pre-processing steps, the green operations are related with the feature extraction ending in the key point selection, the orange operations are related to the choice of topology of the key point linkage, the gray operations represents the matching algorithm, followed by the ending process of validation in purple.

4.3.1 Key Point Selection

Bozorth's pipeline follows a similar path to FCM pipeline after the data pre-processing, extracting features to propose a score computation. However, FCM pipeline detains a global structure of the face (despite the features describing the local environment of the point) for analysis, while Bozorth's pipeline diverges from this premise, choosing a select group of points expecting to provide only discriminating salient information to the matching algorithm. A key point extraction is proposed based either on uniform sampling (by the application of a voxel grid) or a salience approach (detailed by the Intrinsic Shape from Signature (ISS) algorithm [44]).

Both the uniform sampling and the ISS present their unique input parameters, and coincidentally the primary setting consists of the radial argument. In uniform sampling, there is a parameter labeled *leaf radius*, which determines the volume of the grid by electing the number of points that must be contained in this same volume. The ISS presents the salience radius, which determines (alongside other parameters) points that offer a high salient value to the cloud (as previously mentioned in Equation 2.1 in Chapter 2). The radius parameters are both set dynamically to provide practicality in testing and validation. Figure 4.18 provides an example of both key point selection and visual evaluation of them. Each of the selected key points contains their spatial location and the normal vector components. crucial for the matching procedure.

Bozorth's approach states that the points are linked in pairs to exploit their angular distribution in consonance with the local feature direction. In order to do this, linking strategies must be defined, leading to the choice of a topology between the points. The next step consists of defining a topology for the points connection.



(a) Uniform Sampling (b) Intrisic Shape from Signature

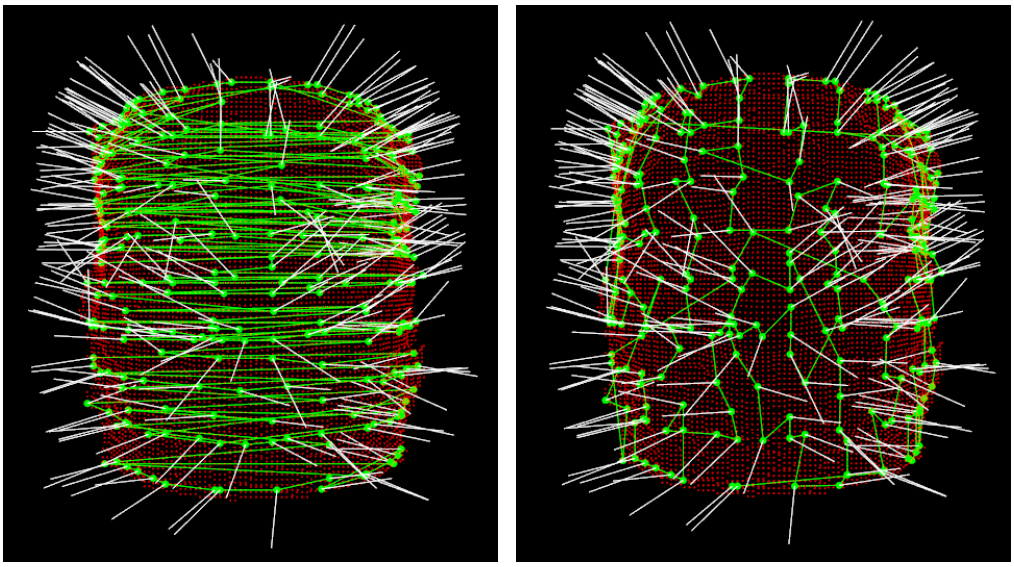
Figure 4.18: Facial point cloud rendered in red and key points detected rendered in green from a frontal perspective.

4.3.2 Topology Selection

The necessity to obtain a connection between pairs of points brings up the uncertainty about the way these pairs should be connected. There are many possibilities available which presents

different concerns about where to begin (which point), how to proceed to the next point, how many pairs a single point may bare and how to avoid duplicates. To detain a vast understanding of the behavior that may be caused by the topology choice throughout the matching process, three different structure were proposed for pairing: two of them constitute a single paired net (where a source point may present only one destination point). and the other constitute a multi-paired net (where each source point may present more than one destination pair, applying the proper arrangement to avoid pairs duplicates). The single-paired topologies are the top-left and nearest neighbors organization, while the multi-paired topology is represented by a clust defined in Bozorth's original work (Bozorth compostion) [36].

The Top-Left topology starts from the point that is located geometrically farthest in the top and to the left of the cloud, linking it to a destination point that represents the closest one to the right until no point can be found to the right which will connect this point to the one that comes closest in the bottom direction and farthest to the left, repeating the process until there are no more available points. This process results in a single link per source point, generating a single paired web. The Nearest Neighbor topology can start from any point since the principle in this topology is to grow a radius around the point until it encloses one of the neighbors (and destination point). Once again, these topologies provide a single link per source point. Figure 4.19 provides a visual representation of the web formation in both these topologies.



(a) Top-Left Topology (b) Nearest Neighbor Topology

Figure 4.19: Single Paired Web Topologies. The facial points clouds are rendered in red, The key points and the link between them are rendered in green, and the normal vector of each key point is rendered in white.

The Bozorth Composition topology original scope receives a sorted table containing the pixel position and the angular direction of the fingerprint minutiae with the intent to avoid linking points that were already treated previously (despite the implementation of data critics in case of input modeling error). The adapted structure of key points elaborated to exploit Bozorth's method in

this application is typified the same way as the original input of Bozorth, sorting the input data in a table containing the spatial information of the point and the angular description of the local normal, aiming to acquire the best performance and similarity to the original application.

In this composition, the points are linked through a distance threshold, similarly to the neighborhood radius, setting a destination pair to every point within the radial limit of a source point, forming this way a multi-linked web. For each point linked, a content table is generated (labeled as web in the original algorithm), storing the distance module between the points, the points' index, and the angular difference between the points' normal vector direction and the distance vector direction. This table of contents contains all the information necessary to proceed to the matching algorithm.

4.3.3 Bozorth Matching and Score Evaluation

The matching process suffered minor adjustments to work in the elaborated scenarios of this work. Bozorth's matching algorithm consists of a few phases of comparison between two input web of pairs, establishing thresholds between the angular variation stored in the content table to define a correlation between the webs and a numerical matching score. The higher the matching score, the higher is the similarity between the two webs, and most probably the closest to being the same subject.

Before describing the alterations made to the matching algorithm and the score acquisition, the matching process must be explained. It consists of three cascades of tolerance checking between both webs, where every pair of the input web is compared through brute force with all the points of the gallery web. The first content of the table to be verified is the pairs that are within the same distance range, diminishing the number of pairs that may represent a match. It is followed by the angular tolerance checking of the source point in the remaining pairs, once again diminishing the number of pairs that may represent a match. And then, the last verification is made, which is the angular checking of the destination point, that usually leaves just one pair as match or none, applying data critics in case of mutual matching within other pairs. Equations 4.7, 4.8, 4.9 represents all the tolerance checking steps described in [49]. The matching score represents the number of pairs between the webs that contain a high correlation in the distance and feature orientation.

$$\Delta d(ds(Ip_k), ds(Gp_{i,j}) < T(d) \tag{4.7}$$

$$\Delta\beta(\beta 1(Ip_k, \beta 1(Gp_{i,j}) < T(\beta) \tag{4.8}$$

$$\Delta\beta(\beta 2(Ip_k, \beta 2(Gp_{i,j}) < T(\beta) \tag{4.9}$$

where Δd and $\Delta\beta$ represent the difference among these functions, Ip is the input's pairs array, Gp is the gallery's (database) pairs array, k is the pair index of the input, j is the pair index of the gallery subject, and i is the subject's index in the database.

4.3.3.1 Adaptation

The matching algorithm itself has not suffered many alterations. The main element to give proceeding to the matching process consist of the input structure and the three-dimensional scenario, which implies in the addition of more information to be analyzed. Figure 4.20 presents a pipeline with the original Bozorth's matching algorithm in comparison to the adapted Bozorth's matching algorithm.

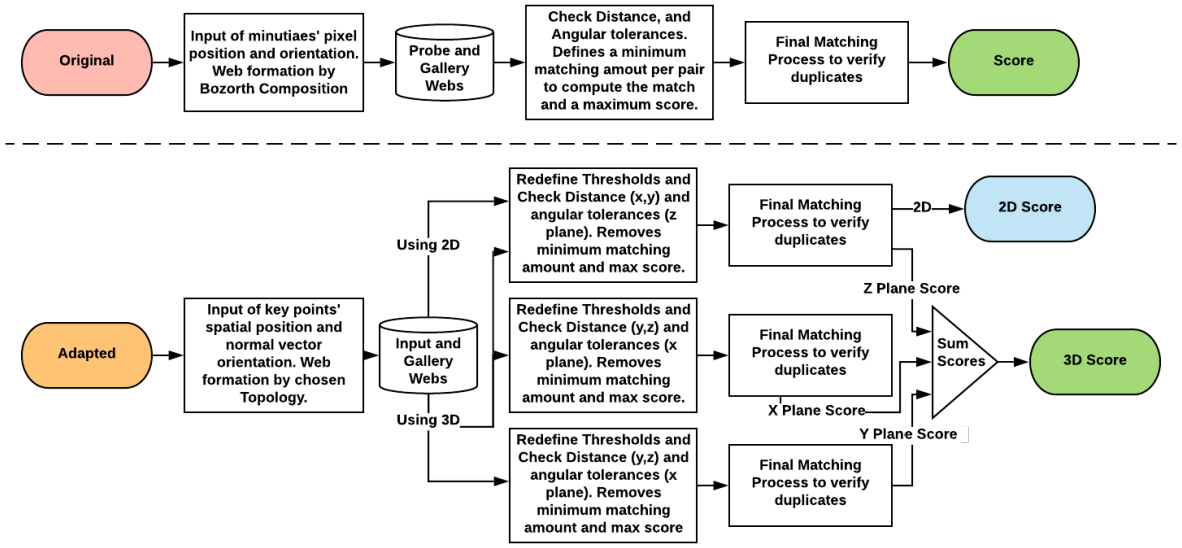


Figure 4.20: Flowchart detailing the original and adapted algorithms' pipeline. The original algorithm is composed of a single path formed by an web formation from input key points, followed by the matching process between a probe and the gallery webs that results in a final score, finished by a score validation removing duplicates. The adapted algorithm is similar until the matching process, which happens in three different projections, followed by score validation for each projection with duplicates removal, and final score computation with the sum of each projection score.

4.3.3.2 Table of Content

When it comes to data amount, the three-dimensional scenario provided by the point cloud environment is more plentiful compared to the original scenario (based on a grayscale fingerprint image). Figure 4.21 displays all the variables that could compound the initial content table of the algorithm, presenting more information than the original method illustrated in Figure 2.8.

A content table involving all the information that the three-dimensional space provides (displayed in Figure 4.21) would contain the complete information for a single execution of the matching process, all at once. However, it would create a complex scenario to apply an eccentric threshold for each one of the angular and spatial discrepancies. Avoiding this scenario, the three-dimensional space formed by the points is manipulated to divide the content into projections of coordinate planes (similarly to the FCM frontal viewpoint projection). The pairing is evaluated then into

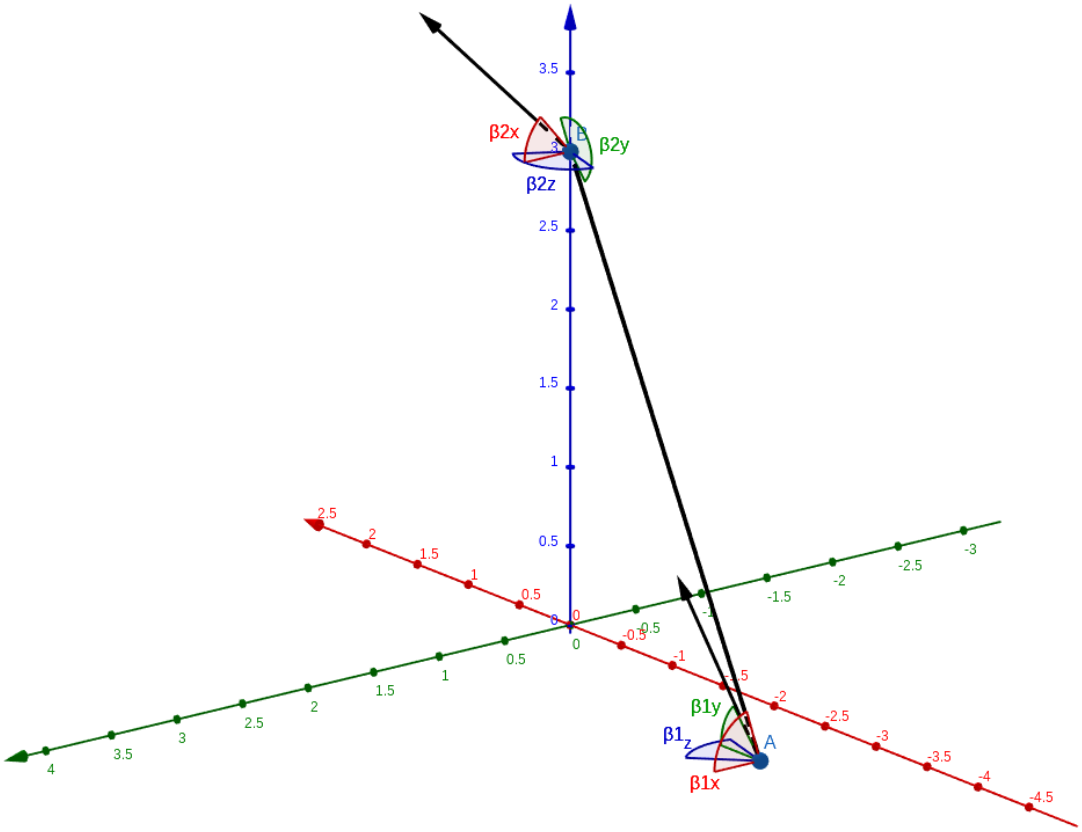
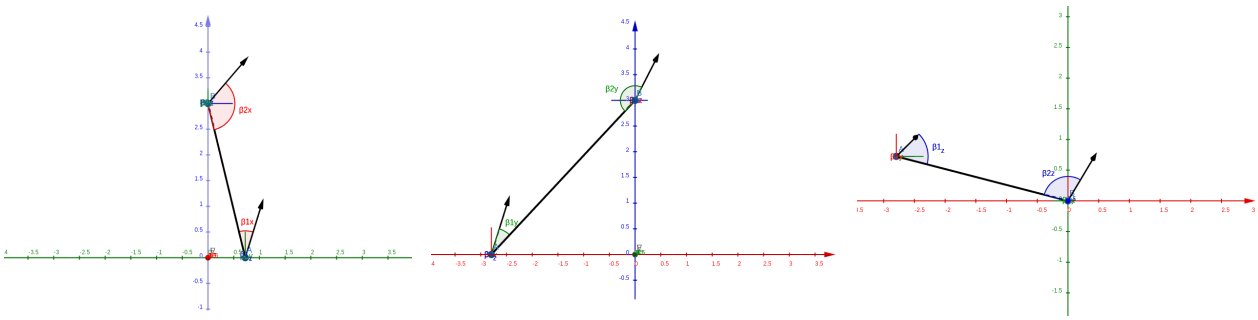


Figure 4.21: Example of two key points link, displaying the normal and distance vectors, the main coordinate axes (z - red, y - green, z - blue) and the angular difference of the normal and distance vector in relation to the coordinate planes (in their respective colors).

three projections (representing the coordinates planes), preserving the original structure of the algorithm and allowing the application of a variant threshold for each projection, leading to eccentric scores on each view-point. Figure 4.22 displays the three projections obtained from the example of Figure 4.21.



(a) X Plane Projection (b) Y Plane Projection (c) Z Plane Projection

Figure 4.22: Projection of the coordinate planes illustrated in Figure 4.21.

The original matching algorithm is applied to each projection, using the tolerance verification

expressed in Equations 4.7, 4.8, and 4.9, but with a few minor changes compared to the original algorithm. The first change consists of the limitations of grouping to a source minutiae (in the adapted scenario as a source point). It has a minimum and a maximum number of matches allowed, limiting expansion to avoid minutiae conglomerations in a small area (ignoring low significant minutiae). These limitations were removed in the adaptation to elevate the flexibility in cases of few key points. The second change consists in the angular and distance tolerance variation for each projection, providing a better suitability for the projections' diversity. The last adaptation (which in this case is not a modification) is the final score confection, obtained by the sum of matches (score) of each projection.

TBozorth's method was tested in a frontal projection as well, simulating the process of FCM in a frontal projection (z plane), to provide a "two-dimensional" result for the proposed resolutions.

Equally to FCM pipeline, Bozorth's pipeline ends in the acquisition of the scores, which in this case define the identity of the input subject for the higher score obtained. For terms of validation, a minimum score is also set to define if the input subject should be treated as an unknown identity or random object, discarding the input. A minimum bound can be set for each of the projections to avoid a single plane similarity as well.

The ending of the matching process in both pipelines results once again in a common process. In this case, is the validation of the algorithm, necessary to observe the capability of these proposed methods, and the effects of parameterization for both cases.

4.4 Validation

The Validation process consists in verifying the capability of the proposed method through a statistical metric, establishing test scenarios that allow the count of hits and misses in subject classification. An important topic to be discussed for classification is the similarity match, which was in both pipelines described as a matter of score comparison. The statistical metric chosen is an F-Measure [196] approach, which calculates the optimal parameters through an Equal Error Rate (EER) analysis to provide a biometric evaluation [201].

4.4.1 Similarity Matching

In this case, the focus is to solve a face verification problem using a Similarity Matching schema presented by [2]. It can be formally defined as: Given an input vector of curvature indexes features C_v extracted from the 3D face data and an alleged identity I, determine if (I, C_v) belongs to the class f_1 or f_2, where f_1 indicates that the alleged identity is true and f_2 that it's false. C_v is compared with C_I, as the vector of biometric features of the individual I, to determine its class. Thus

$$(I, C_v) \in \begin{cases} f_1, & \text{if } S(C_v, C_I) \geq t \\ f_2, & \text{otherwise} \end{cases} \tag{4.10}$$

where S is a function that measures the similarity score between the vectors C_v and C_I, and t is the predefined threshold. $S(C_v, C_I)$ is called a similarity matching score between the biometric features of the individual and the alleged identity. The identification problem can be formally defined as: given as entry a vector of features C_v, determine if the identity I_k, where $k \in \{1, 2, ..., N, N+1\}$. Here $I_1, I_2, ..., I_N$ are the identities already in the system, and I_{N+1} indicates the rejected case, where no identity is compatible with the users. Thus

$$C_v \in \begin{cases} I_k, & \text{if } \max\limits_k \{S(C_v, C_{Ik})\} \geq t, k = 1, 2, ..., N \\ I_{N+1}, & \text{otherwise} \end{cases} \tag{4.11}$$

where C_{Ik} is the vector of biometric features corresponding to the identity I_k, and t is a predefined threshold.

In both pipelines presented, the feature vectors are reduced to a matching score. The determination of the subjects, identity and rejection, through threshold permits to resume the validation through the F-Measure metric.

4.4.2 F-Measure

The F-Measure consists of the evaluation metrics disposed to measure a method's quality based on the hits and misses of the classification [196]. The members of the database are known due to the nomenclature of the subject's data files, which contains the numerical index assigned to each subject.

The F-Measure elements rely on the concepts of truth and falsity along with positivity and negativity, guaranteeing two classes of hits and misses. This hits and misses criteria establish the following cases:

- True Positive (TP) - The input subject is not rejected, and the similarity matching returns the right identity;
- False Positive (FP) - The input subject is not rejected, and the similarity matching returns the wrong identity;
- True Negative (TN) - The input subject is rejected, and the subject does not belong to the database;
- False Negative (FN) - The input subject is rejected, and the subject belongs to the database.

The validation process in both pipelines is represented by a brute force checking of identity towards all the subjects present in the database. However, since a biometric system needs to present a rejection state, 30% of the subjects were randomly selected to be labeled as an unknown identity, forcing the negative state of the evaluation. This validation demands a repetition process due to the randomness of the subjects' state, which led to a fixed setting group in the EER phase to provide equal conditions between the algorithms appraisal and parameter checking.

The F-Measure calculation relies on two principles of this statistical analysis, which are Precision and Recall. The Precision is represented by Equation 4.12 expressing the fraction of the positive cases, representing the select elements provided with a classification. The Recall is represented by Equation 4.13 expressing the relevant elements rejected. The F-Measure is calculated by Equation 4.14 that obtains the ration of two times the product of the Precision and the Recall and the sum of these properties.

$$Precision = \frac{TP}{TP + FP} \tag{4.12}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.13}$$

$$F\text{-Measure} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4.14}$$

The Equal Error Rate stipulates the optimal operation parameters that minimize the two types of error (misses) determined by the classifier. The functions that demonstrate the progression of these misses are the False Acceptance Rate (FAR) and the False Rejection Rate (FRR), which are both respectively expressed by Equations 4.15 and 4.16. The EER is the point in which both functions present the same value (Equation 4.17), or in a visual term, it is represented by the crossing point of the EER curved defined by the plot of both the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) in a single graph.

$$FAR = \frac{FP}{n} \tag{4.15}$$

$$FRR = \frac{FN}{n} \tag{4.16}$$

$$EER = FAR \parallel FRR, \text{ if FAR} = \text{FRR} \tag{4.17}$$

where n refers to the number of recognition attempts.

Chapter 5

Results

This chapter is responsible for explaining the testing scenarios built and the validation result towards each of these scenarios. Each scenario provides constraints related to the data distribution and selection, allowing to obtain a perception about the application performance capability to the presented data set and the viability in a real general scenario.

In terms of parameters, the results of validation clarify the behavior of the algorithms against the variation of their control variables, setting a dynamic environment to determine the best point of operation and expose the best performance in a specific scenario. The parameters tested follows the order of appearance in the algorithm's flowchart, and whenever the optimal value of a parameter is defined that value is set as constant to acquire the optimal value of following parameters.

The result is mostly presented in the form of the Equal Error Rate (ERR) that provides a robust and valid biometric evaluation, and the F-Score of the optimal point extracted from EER to grant a statistical measurement of the performance [28]. As discussed in the work of Jain [28], biometric systems don't necessarily present a machine learning classification method, thus the determination of the optimal parameters' values in a data set is acquired through the application of the Equal Error Rate (EER). The presented methodology doesn't contain any training of classifiers for validation.

The main scenario consists of a frontal face position identification, performing a vast combination of the pipeline possibilities to determine the best solution and a comparison between the algorithms and parameters. This scenario is divided into subsections in a few cases to determine one at a time, the optimal settings. The last scenario is applied only to the FCM and Bozorth 3D algorithm, checking the ability to deal with lesser orientation changes.

The result of the provided algorithms is also compared with other approaches of 3D facial recognition that were related to the database (either cited or partially used by the proposed work) to provide a comparative measurement.

5.1 Settings and Materials

The programming was focused on structures to deal with image processing and point cloud manipulation, and to take advantage of already implemented tools. The following libraries were used:

- Point Cloud Library (PCL)
- OpenCV

being the OpenCV mainly applied to handle the image processing (*e.g.* the face detection, the FMC grayscale conversion, and pixel value count), and the PCL provided the instrumental necessary for building and manipulation of the point cloud operations.

5.1.1 Database

Despite the previous explanation of the database in Chapter 4, the database is one crucial element to the following scenarios, worth mentioning again. The database consists of the VAP RGB-D Face Database by [197], which contain a total of 31 subjects in 13 different head orientation low-resolution clouds, obtained after reconstruction (holes and gap treatment) and smoothing (spikes removal). Not all the orientations were used in the following scenarios. The selection of the database focused in low resolution devices outputs of facial data that could be transformed or translated to point cloud data, restricting the number of available databases that fulfilled this characteristic, besides the approval of the database's owner for such application.

5.2 Test Scenarios

This section is about the test scenarios developed, exposing the results of each algorithm to a specific constraint level. The main objective of these scenarios is to determine the capability and viability of these algorithms to be applied in a similar real-life scenario and define the impact of obstacles to their performance.

The first Scenario presents a frontal face constraint, in which the biometric evaluation is performed, applying the EER to determine the best point of operation of the determined parameters, followed by an F-Measure observation towards this optimal environment.

The second Scenario presents the same constraint, in which the accuracy evaluation is performed to obtain a comparative metric towards other developed approached towards the same topic of three-dimensional facial recognition.

The Third Scenario applies a scale constraint towards the methods, to observe if a variation in scale to the point cloud (the number of points remains the same, preserving density) brings noticeable variation to the performance.

The Fourth and last scenario applies a lesser orientation variation, adding four orientation and its respective point clouds to the gallery, verifying the performance in the optimal point obtained through EER.

5.2.1 First Scenario - Frontal Biometric Evaluation

In this scenario, each subject - in a recognition (classification application) system - can be treated as a class. The used method to provide a possible identification (intra-personal score minimization) is analog to a classifier described in Chapter 4. The Equal Error Rate (EER)[202] can be defined as an objective, threshold-independent measure of the classifier's performance for statistical pattern recognition, which is used to evaluate this classifier and commonly used to evaluate biometric systems.

The rejection criterion established to the ERR was based on the score obtained in each pipeline method. The first criterion defined is the maximum threshold, avoiding false recognition from unknown subjects, and refusing badly acquired sensor outputs. The second criterion is characterized by a threshold interval limitation between the two best enroll matches, if they do not belong to the same subject (since each subject has three images samples), allowing the system to presume doubt between two subjects, and possibly reject the input.

During this test scenario, a random group of subjects was selected to establish a rejection base to allow the biometrics evaluation, due to the necessity of refusing unknown subjects or non-facial inputs, guaranteeing the negative condition concepts of the F-Measure (true and false negatives). This group was fixed during all the different cases in each scenario to compare the algorithms under the same circumstances, including the data distribution.

These scenarios implement the identification issue of the biometrics, in which an input subject is compared one-by-one with all the known subjects of the gallery. The one-to-all comparison, diverging from the verification issue, where there is usually an identity claim, and their comparison is one-to-one. So to process all the scores and classify each input, a brute force comparison is made towards every subject in the gallery database, traversing through the pipeline repeatedly.

5.2.1.1 FCM Biometric Evaluation

To evaluate and minimize the errors of the proposed methodology, the most influential variables were selected (radius, bins). Then EER was applied to each one of these variables, defining the optimal value for them. The parameters are fixed in the optimal value once defined, and they are tested in the sequential order of appearance in the pipeline.

The radius, responsible for the vicinity description of a point and, consequently, the curvature intensity was analyzed in step intervals of 5, generating different curvature maps and score for each radius value from minimization function. In Figure 5.1 it is possible to visualize the rates values of false acceptance and false rejection obtained for each radius values, obtaining an interception at the radius of 26.67, resulting in an ERR of $3,58\%$ for acceptance and rejection. The interception

is calculated by approaching the FRR and FAR to lines and equal the obtained expressions to find the interception radius and the value of FAR and FRR at that point.

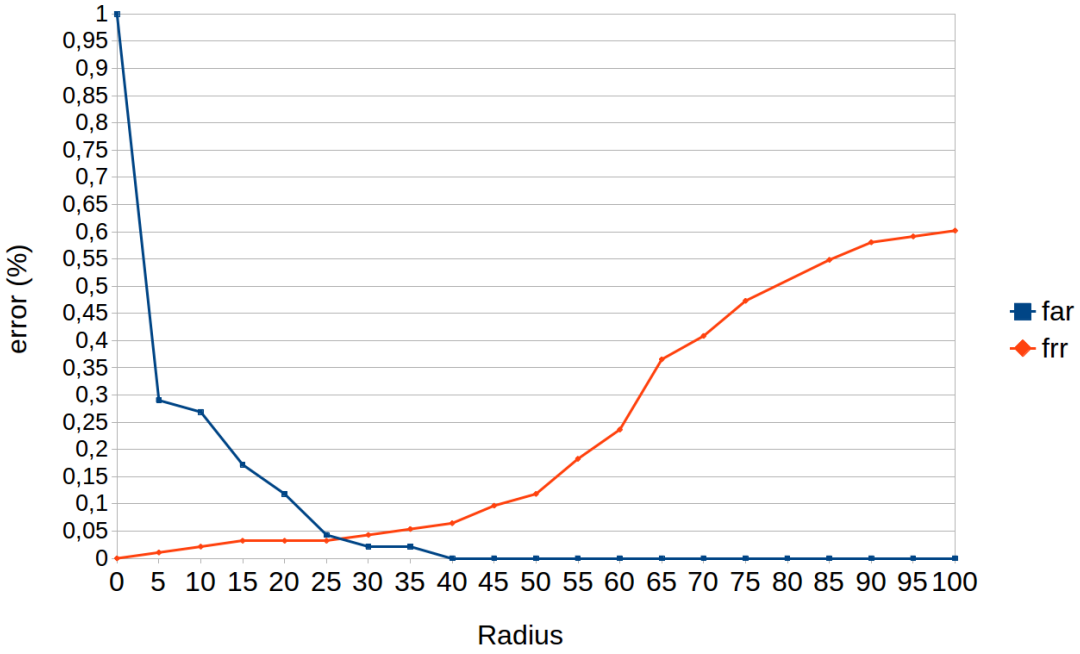


Figure 5.1: Equal Error Rate of Vicinity Radius. The blue curve depicts the False Acceptance Rate (FAR), representing the ratio between false positives and total comparison numbers. The orange curve depicts the False Rejection Rate (FRR), representing the ratio between false negatives and total comparison numbers. The equal error occurs in the intersection point of the curves between the 25 and 30 radius values.

The behavior observed in Figure 5.1 is expected when it comes to the radius progression. As the radius presents a low value, the curvature values are almost randomly scattered due to the resolution, which causes mass confusion without many discrepancies to be threshold limited, resulting in false positives. As the radius increases massively, the neighborhood established involves a lot of features that approach the surface descriptions to a whole face, which brings the similarity between FCMs closer. It leads to mass rejection due to threshold limitations.

Once the radius is the most independent and the first required variable to be defined in the flowchart of the proposed methodology (Figure 4.1), all remaining variables will use this optimal radius value fixed as a reference to define their optimal values from the EER method.

The subsequently analyzed variable is the number of bins of the intensity histograms obtained from the curvature maps. The ERR was applied with the same previously criterion, fixing the optimal radius value obtained during this analysis, changing only the number of bins used to represent the histogram resulting in the chart displayed in Figure 5.2.

Although it is expected to obtain a better classifier result using a higher number of bins representation (and consequently intensity distribution), that doesn't guarantee the best discriminating value among subjects. In Figure 5.2, values from 8 up to 16 bins achieved an error of $2,5\%$. Still,

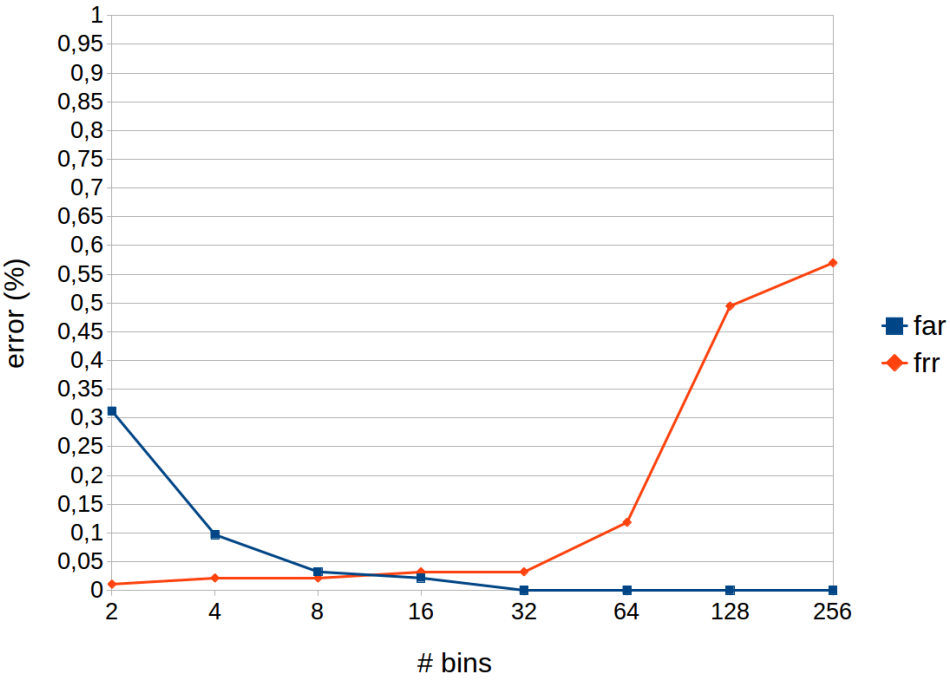


Figure 5.2: Equal Error Rate of the number of bins. The blue curve illustrates the False Acceptance Rate (FAR) and the orange curve illustrates the False Rejection Rate (FRR). The equal error occurs in the intersection point between 8 and 16 bins.

the number of bins is restricted to values that represent a power of 2 (to guarantee the equal numerical distribution of the intensity values [ranged from 0 to 255]) and located in the exact average of 8 and 16, both of these values are optimal for the number of bins in this application. The use of values greater than 16 bins provided a higher disparity between the images of the same subject, possibly due to the information loss (caused by mainly by filtering process) and the occlusions filling estimation done by [197] preprocessing, which causes a higher rejection rate and consequently false rejection as well.

In both cases, the division of the FCM was set in a fixed size of 3x3, because the n-region division did not provide the required pattern of FFR and FAR for the biometric analysis.

5.2.1.2 Bozorth's Algorithm - Biometric Evaluation

Bozorth's Pipeline presents the same approach for biometric evaluation as described by the FCM Pipeline. Still, there is some minimal variation related to the parameter composition and the different structural combinations that the feature organization may assume. Bozorth's algorithm presents a mix of possible features due to the feature extractor that offers the choices of Uniform Sampling and Intrinsic Shape From Signature. The topology of the web has three representations, the Top-Left, Nearest Neighbor, and Bozorth Composition (2D and 3D), producing a considerable number of setups to determine the optimal operation point. Each of these cases is based on an extrinsic parameter that will be discussed in each specific case.

Each of the combinations allowed by Bozorth's Pipeline will be treated separately in items of this subsection, comparing and discussing their results by the end. The first approaches will be towards the 2D applications, further covering the 3D composition.

5.2.1.3 Uniform Sampling - Top Left Topology

The Uniform Sampling utilized was implemented by the PCL library, receiving as an input parameter a search radius that acquires the average number of points. It determines the number of leaves to build a Kd-Tree and determine the volume of each grid in the voxel (thus entitled Leaf Radius). The higher the volume of the grids, the smaller will be the number of key points.

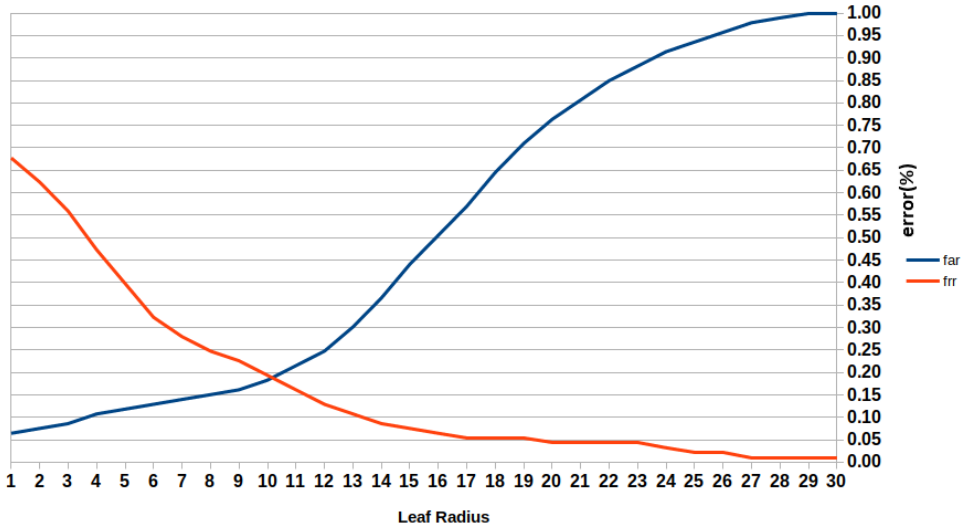


Figure 5.3: EER of the Uniform Sampling with a Top-Left topology. The equal error occurs in the intersection point between the leaf radius of values 10 and 11.

Due to the behavior of the uniform sampling, the FAR and FRR should be inverted, because a low-value radius will provoke a high number of key points that result in high rejection. In contrast, a high-value radius will result in a small number of key points resulting in mass confusion. Figure 5.3 shows the EER obtained for the proposed configuration, with a leaf radius step increase of 1 unit, the intersection point is found in the leaf radius equals to $10, 17$ presenting an error of $18, 84\%$ for false acceptance and rejection. The crossing point is calculated equally to the FCM procedure.

Compared to the FCM, the uniform sampling with a top-left topology showed lower performance and capability to deal with the recognition issue. This leads to the next configuration of Bozorth's pipeline, composed by the uniform sampling and the nearest neighbor topology.

5.2.1.4 Uniform Sampling - Nearest Neighbor Topology

Continuing with the Uniform Sampling feature extractor, the following validation is also focused on the leaf radius to determine the voxel grid division.

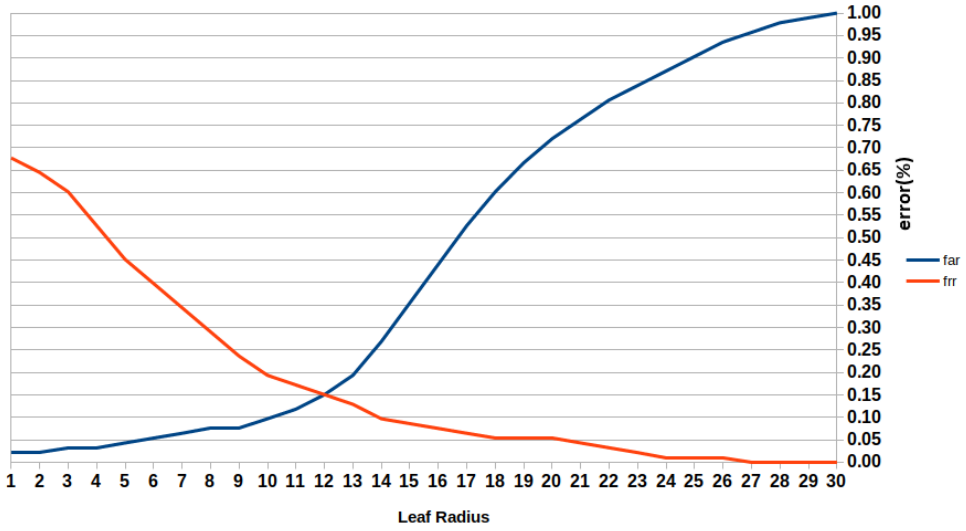


Figure 5.4: EER of the Uniform Sampling with a Nearest Neighbor topology. The equal error occurs in the intersection point in the value 12 of leaf radius.

Figure 5.4 presents the EER of the Uniform Sampling and the nearest neighbor topology, with FRR and FAR curves similar to the ones displayed in the Top-Left topology (Figure 5.3). Still, the overall performance was better, containing an intersection point in the exact value of 12 for the leaf radius, reaching an error of $15,1\%$. In this topology, there was a difference of $3,74\%$ in the equal error, presenting a better performance in terms of biometrics, but still, a worse performance than the one noticed in the FCM scenarios.

The last composition involving the Uniform Sampling as a feature extractor consists of the combination with the Bozorth composition topology, which is the only one multi-linked based web.

5.2.1.5 Uniform Sampling - Bozorth Composition

Finishing the Uniform Sampling combinations, there are both combinations of the Bozorth Composition, which can be applied in a 2D manner (treating only the xy plane projecting the z coordinate) and in a 3D way. The Bozorth Composition varies from both of the previous combinations due to the multi-linked web, permitting a higher score range, and possibly a more discriminating scenario. The 2D approach is introduced first in this subsection.

The 2D approach of Bozorth composition produces the results exhibited in Figure 5.5. The multi-linked web outperformed the single-linked webs presenting an equal error rate of $12,9\%$ when the leaf radius is equal to 12. The presence of more links between the points (permitting the computation of a higher amount of matches) has provided a better discriminating environment, which led to a reduction of the errors (false positive and negative). Despite this error reduction, the two-dimensional Bozorth composition still unperformed the FCM algorithm, with a difference of about 10% in the optimal operation points.

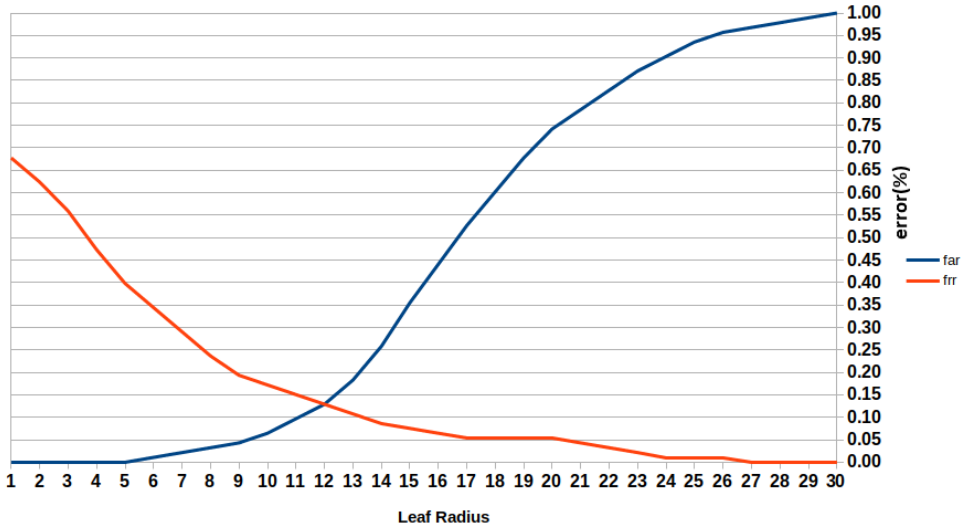


Figure 5.5: EER of the Uniform Sampling with a Bozorth Composition (2D) topology. The equal error occurs in the intersection point of leaf radius equal 12.

The main problem the Uniform Sampling might generate is the absence of discriminating points, extracting information that does not represent salient facial locations. Due to this discrepancy of performance towards the previously tested pipeline, the three-dimensional Bozorth composition was skipped for this feature extraction, being later evaluated in the Intrinsic Shape Signatures feature extractor configuration, that focuses on the structural environment of the cloud's point to obtain key points rich in local geometrical components, that may lead to better performance in terms of biometrics.

5.2.1.6 Intrinsic Shape Signatures - Top Left Topology

The PCL implementation of the Intrinsic Shape Signatures (ISS) requests a salient radius to determine the covariance matrix, using the eigenvectors as base for pruning points that cannot define a repeatable Local Reference Frame (LRF) and define the salience module based on the eigenvalues, finished by a non-maxima suppression as stated by [48]. The biometric evaluation is analyzed towards the salient radius, responsible for the selection of the key points that contain significant variations in each principal direction, possibly describing a fiducial facial point.

The ISS main parameter consists of the salient radius, which causes variation in the key points selection and number of key points retrieval, which tends to rise along with the radius.

The combination of the ISS feature extractor and the top left topology provided the EER result displayed in Figure 5.6. The resulting chart shows the same behavior previously noticed in the FCM, where the FAR starts at a high level and decays along with the parameter increase. In contrast, the FRR begins at a low level and increases along with the parameter expansion, inverse to the Uniform Sampling behavior. The Salient Radius is increased in a step of $0,2$, providing the optimal point of radius equals to $4,8$ with an error of $9,7\%$ in this operation point.

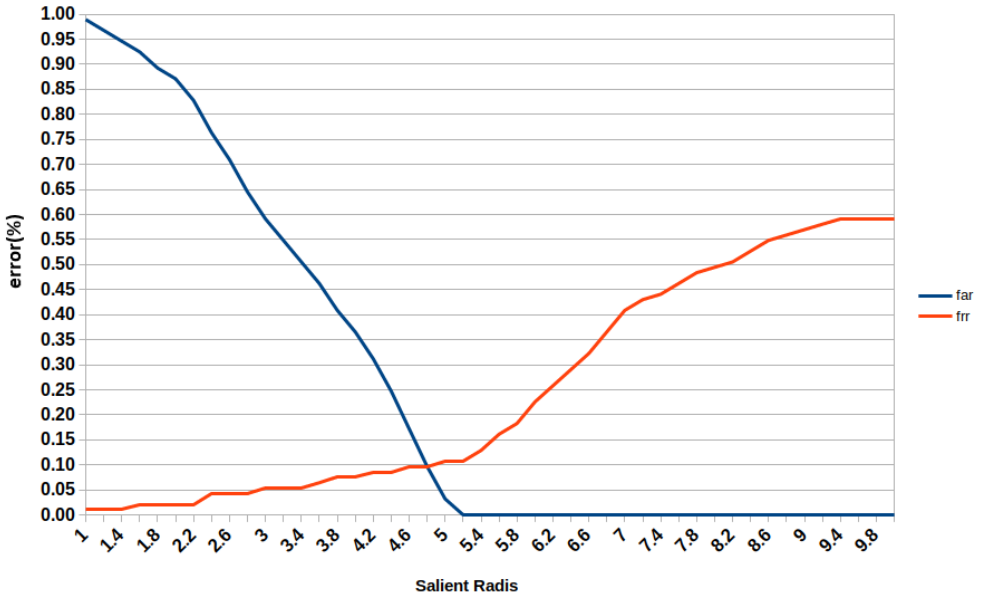


Figure 5.6: EER of the Intrinsic Shape Signatures with a Top Left topology. The equal error occurs in the intersection point of salient radius equal $4,8$.

The outcome of this evaluation outperformed all the compositions tested of the Uniform Sampling, endorsing the argument that the Uniform Sampling operating in a distributed manner (and not on an important key point selection) does not provide a viable contribution to this recognition scenario.

The following configuration related to the single-linked web for the ISS feature extractor consists of the application of the Nearest Neighbor topology.

5.2.1.7 Intrinsic Shape Signatures - Nearest Neighbor Topology

Resuming the evaluation of the whole Bozorth pipeline, the following assessment consists of the web of points linked in the nearest neighbor topology.

This configuration followed the same standards in terms of the radius increasing step of the previous topology, generating the results of Figure 5.7. The optimal point reported occurs in the intersection point of radius equals to $5,5$ that leads to an error of $5,8\%$. Regardless of the feature extractor, the nearest neighbor topology outperformed the top left topology (considering the same feature extractor applied) is these scenarios, turning the nearest neighbor the best option among the single-linked web topologies.

The following topology consists of the multi-linked Bozorth composition, either applied in a two and three-dimensional structure.

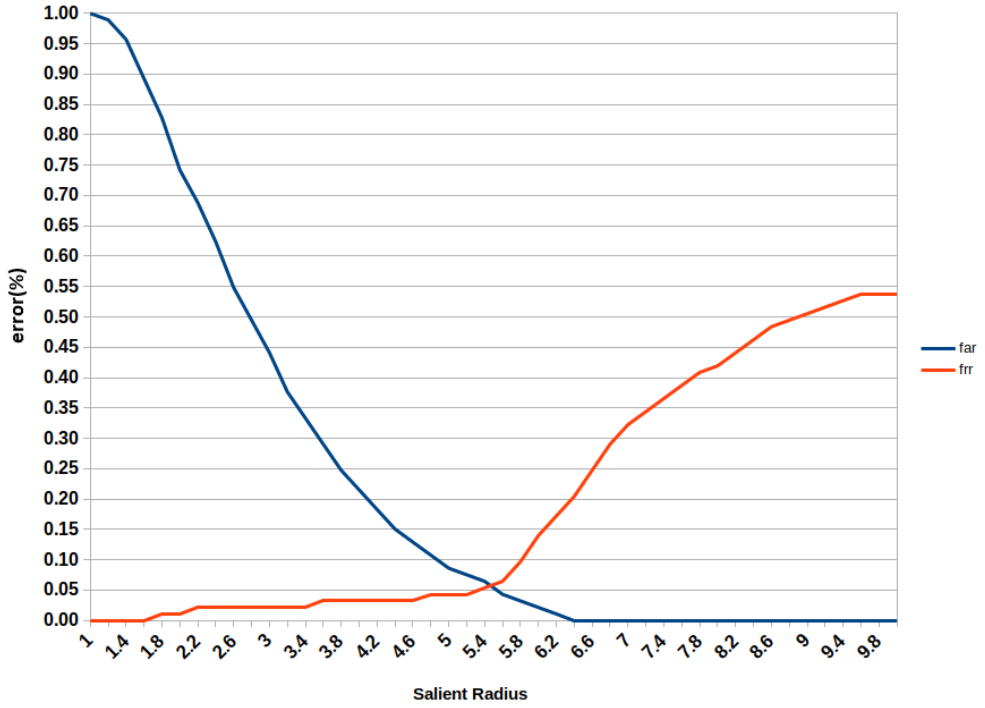


Figure 5.7: EER of the Intrinsic Shape Signatures with a Nearest Neighbor topology. The equal error occurs in the intersection point between the salient radius of values $5,4$ and $5,6$.

5.2.1.8 Intrinsic Shape Signatures - Bozorth Composition (2D)

This initial analysis of the Bozorth Composition using the Intrinsic Shape Signatures as a feature extractor explores the two-dimensional approach defined in the pipeline.

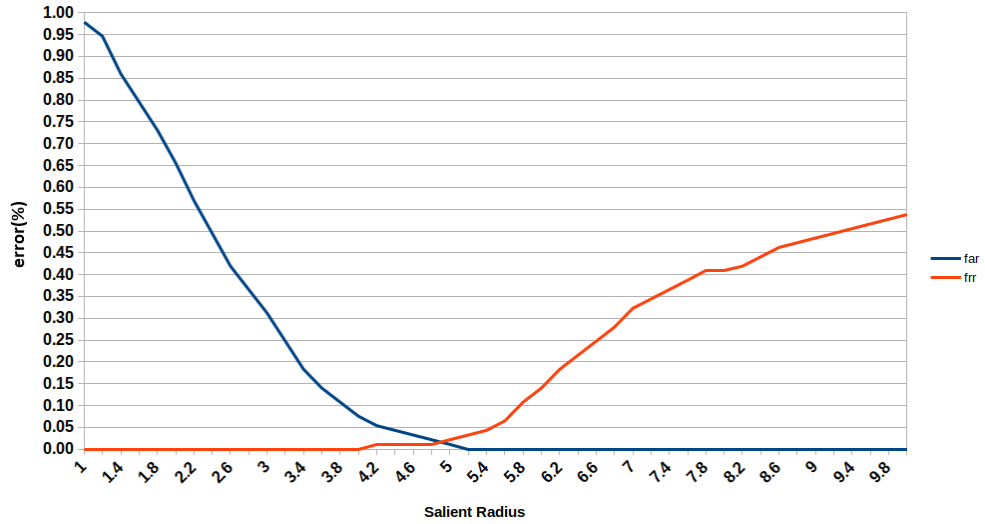


Figure 5.8: EER of the Intrinsic Shape Signatures with Bozorth Composition (2D). The equal error occurs in the intersection point between the leaf radius of values $4,8$ and 5.

The Equal Error Rate analysis of the Bozorth Composition (2D) is displayed in Figure 5.8.

In Figure 5.8 the stated intersection point occurs when the radius value is $4,9$ stating an equal error of $4,85\%$. This analysis provided the best performance about the Bozorth's pipeline settings tested to this point. Compared to the results stated in the FCM, the performance did not reach the same level of operation error. However, the discrepancy towards the FCM result has dropped compared to the Uniform Sampling analysis, reaching an error discrepancy of $2,35\%$, closing in the FCM's pipeline performance.

During this assessment of Bozorth's pipeline, the parameters regarding Bozorth's algorithm remained unchanged to its original application, which could not be suitable for an application with different standards. To sketch a more appropriate algorithm specifically to the data type and arrangement and the recognition issue itself, the biometric evaluation is also measured against the main effect parameter noted in the matching process, which is the angular threshold, redefining the value that defines the angular difference tolerance between two pairs comparison, as stated in Chapter 4.

The angular threshold (TXS) consists of the squared value of the max angle discrepancy allowed (to include either positive and negative variations) between the angles of two pairs being compared, discarding the matching possibility if any of the angular components overpass the value.

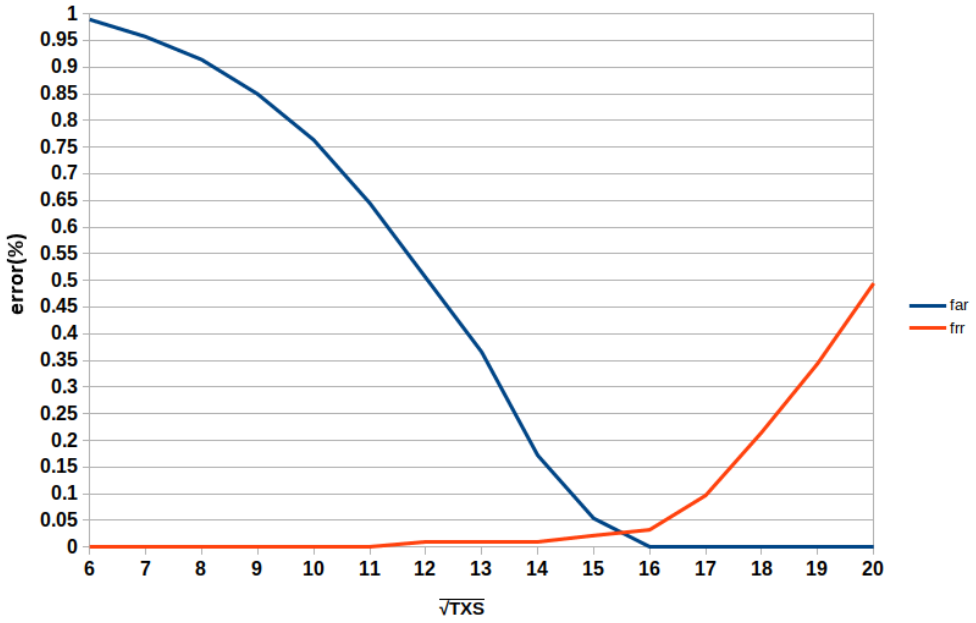


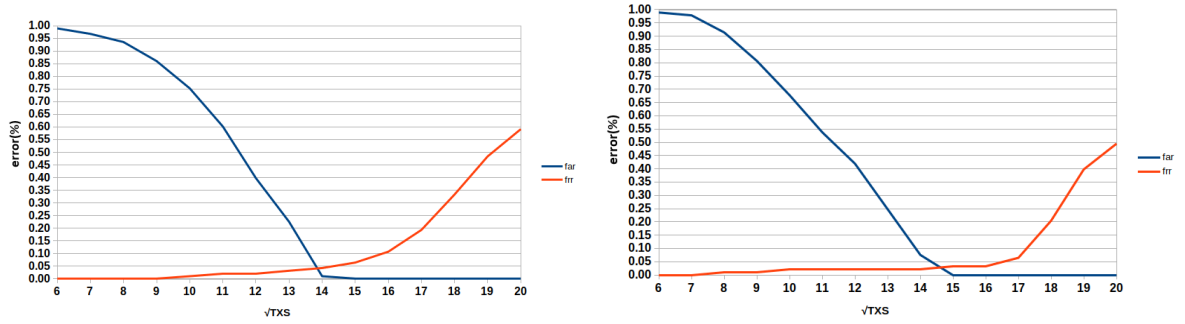
Figure 5.9: EER of the Instrinsic Shape Signatures with Bozorth Composition (2D) and parameters redefinition. The equal error occurs in the intersection point where the angular tolerance presents values between 15 and 16 degrees.

The EER showed in Figure 5.9 is the result of the combination of the optimal point of Figure 5.8 and the variation of the angular parameter $\sqrt{TXS}$. This parameter redefinition guaranteed an error reduction to the value of $2,71\%$ in the intersection point of $15,5$, diminishing the discrepancy towards the FCM error to the value of $0,21\%$, which describes a similar performance of both pipelines.

This configuration exploits only the Z coordinate direction plane projection (considering only the variations in the X and Y coordinates direction), making the subjects' description in a single view projection, this way, performing only in a two-dimensional manner (thus entitled 2D). The projection in the other two different coordinate planes combined with this analyzed projection creates a new three-dimensional complete description of the subjects', which consists of the Bozorth Composition 3D, finishing the last possible setting provided by Bozorth's pipeline.

5.2.1.9 Intrinsic Shape Signatuures - Bozorth Composition (3D)

The evaluation of the three-dimensional version of Bozorth Composition consists of joining the result obtained from the already computed two-dimensional settings with the other two coordinate planes projection results, applying the equal error rate to define the optimal parameters of each coordinate plane. The parametrization of the X and Y plane projections also rely on their extrinsic angular threshold (TXS).



(a) Y Plane Projection

(b) X Plane Projection

Figure 5.10: EER of the Bozorth Composition (3D) for the parameter redefinition of the remaining coordinate plane projections (Y and X).

The application of the EER in all the coordinate plane projections of the Bozorth 3D provided the optimal angular threshold for each of the projections. Figure 5.10 delivers the intersection points of both plane Y and X projections, presenting an error of $4,1\%$ in the optimal point of $13,86$ and $2,82\%$ at the optimal point of $14,62$, respectively.

The conclusion of the parameters redefinition finishes the setup of the three-dimensional Bozorth composition to exploit the best performance in the salient radius variation of the feature extractor ISS.

The application of the EER on the coordinate plane projections did not provide any cases without errors. However, the behavior noted when building the EER analysis for each case is that the errors occurred in different indexes in the force brute comparison, meaning that the errors which happened in each of the projections are different, presenting extrinsic variant subjects recognition difficulty in each projections. This analysis implies that the sums of the scores of all the coordinates projections may overcome their singular errors. The result of such investigation is displayed in Figure 5.11 in which appears two cases of intersection point (radius equals $5,2$ and $5,4$)

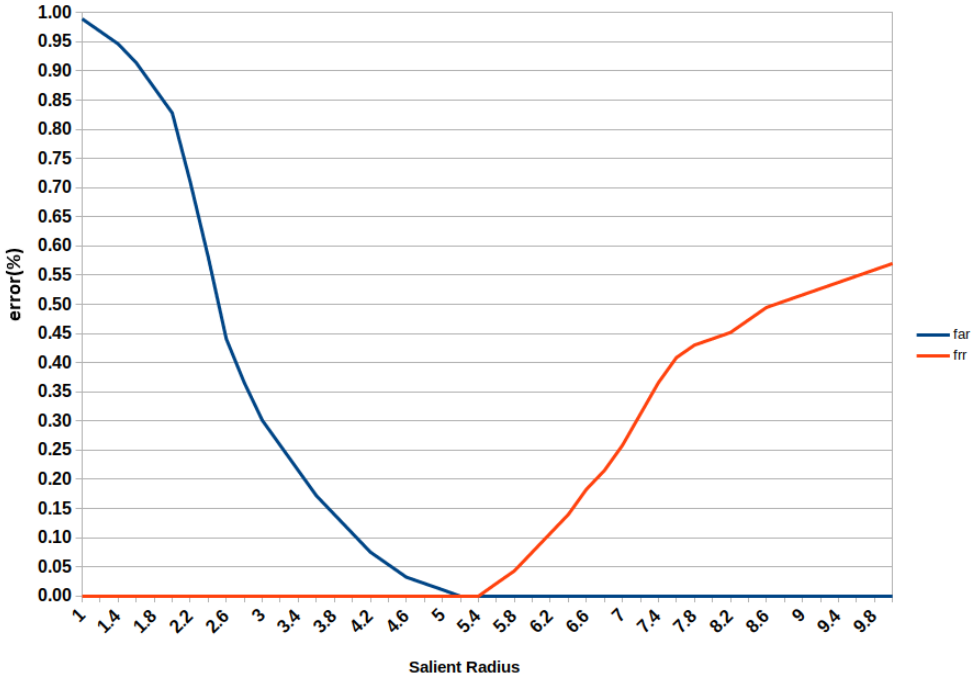


Figure 5.11: EER of the Intrinsic Shape Signatures with Bozorth Composition (3D) after parametrization. The equal error occurs along the interval of $5,2$ and $5,4$ presenting a perfect score with null error.

with a perfect matching score, meaning there are no errors for this scenario in such configuration, beating the performance observed in the FCM pipeline application and thus presenting the best performance among the ones presented.

Since the optimal points of each configuration are acquired with the EER biometric evaluation, the average F-Measure is computed through these optimal points to determine the final performance score of each case.

5.2.1.10 F-Measure

The F-Measure consists of a statistical analysis of binary classification, providing a performance score based on the classifier precision and recall capabilities. In this evaluation, the F-Measure is applied to the optimal points obtained from the EER analysis. In terms of Bozorth's pipeline, the F-Measure is applied only to the ISS feature extractor configurations since all the Uniform Sampling configurations performance are outperformed in this current scenario.

The error value obtained from the EER simulates an approximation of the classification performance, since a lower error indicates that a lower number of false statements of the F-Measure are occurring, implying that a higher precision and recall are generated. This effect is noticed by the F-Measures calculated in Figure 5.12, in which the performance of the proposed methods is presenting a similar outcome of the comparison through the EER values. Due to the randomness of the select set to test the rejection state for biometric evaluation, the F-Measure is computed for

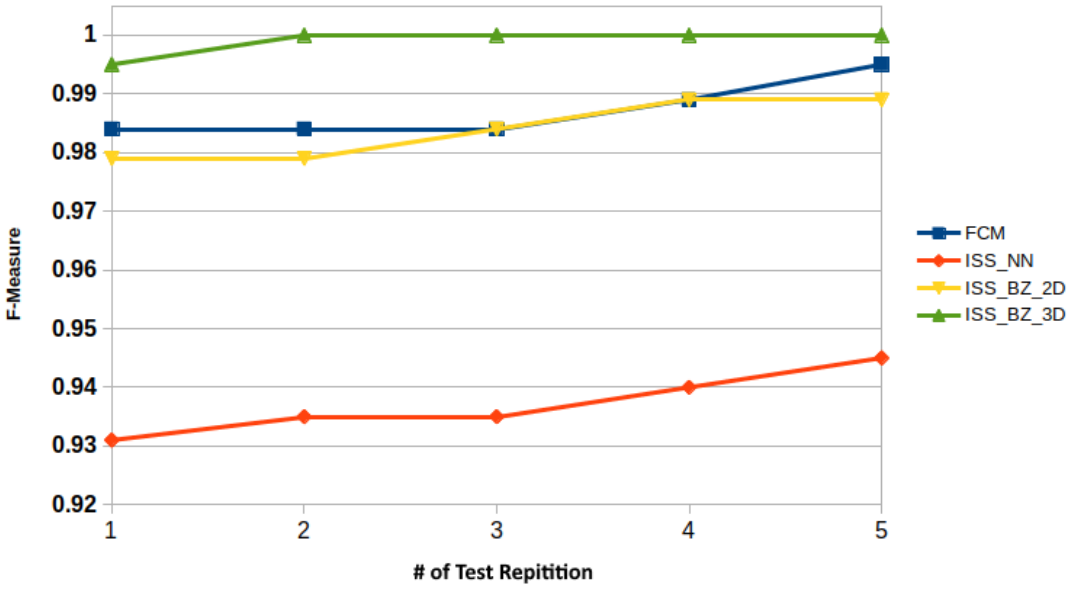


Figure 5.12: F-Measure of the best performers obtained in 5 measurements. The last-placed algorithm (in orange) consists in the Bozorth pipeline using ISS and Nearest Neighbor topology, followed by the ISS with the 2D Bozorth Composition (in yellow) as third best, along with the FCM pipeline (in blue), and ending with the best performance algorithm composed by the ISS and the 3D Bozorth Composition (in green).

five different sets, randomly generated, to verify that the classifier provides a stable classification score, maintaining the same behavior around an expected value related to the error calculated in the validation. This analysis shows that the three-dimensional Bozorth composition, along with the ISS feature extractor, guarantees the best performance among the proposed methods, followed by the FCM and the two-dimensional Bozorth composition + ISS presenting a similar overall score, in sequence with the ISS feature extractor in a nearest neighbor topology with a considerable score discrepancy from the previous configurations. The other Bozorth pipelines settings are not included in this analysis due to the lower performance shown in the biometrics evaluation.

The computation of the optimal operation points for each of the classifiers, followed by the calculation of the F-Measure, concludes the biometrics validation and this first scenario of testing. In this scenario, multiple classification settings were tested, demonstrating that three classifiers generate high-interest performance, and a perfect score for this application, indicating that the proposed methods are significant for a similar real scenario application.

5.2.2 Second Scenario - Accuracy

The following scenario provides a comparison between published methods and the proposed methodology, projected to provide a comparative validation with related works.

To obtain a real performance evaluation, this scenario was developed to compare the performance of the proposed methodology with other state-of-the-art techniques related to the used data

type and the face recognition task. This evaluation is focused on the Rank-1 Accuracy [203] of the facial recognition process, presented as the most usual evaluation method found in the techniques used in the comparison.

The Rank-1 Accuracy is obtained by a ratio of the relevant samples from the recognition process (true positives and true negatives) and the total subject's enrollments in the database. From extensive tests used to define parameters of the proposed methodology, presented in a previous test scenario, the best recognition results achieved an accuracy of 100% followed by 98.92%. These results are compared with the best performance of others state-of-the-art techniques as described in Table 5.1.

Table 5.1: The best Rank 1 Accuracy of the face recognition algorithms related to the VAP RGB-D database. (The external results are mentioned in the cited publications).

Data Type	**Method**	**Rank 1 - Accuracy (%)**
RGB Images + Depth Map	Goswami et al.[22]	80.6
RGB Images + Depth Image	Hu et al. [192]	90.0
RGB Images + Depth Image	Bormann et al. [193]	96.0
RGB Images + Depth Map	Zhou et al. [21]	95.9
Depth Map	Saleh and Edirisinghe [194]	96.67
Point Cloud	**2D Bozorth Composition Method**	**97.85**
RGB Images + Depth Map	Chowdhury et al. [35]	98.71
Point Cloud	**FCM Method**	**98.92**
Point Cloud	**3D Bozorth Composition Method**	**100.00**

For a fair comparison, the selected techniques were obtained related to the database presenting the identification task in facial recognition as well, based on different techniques. All presented methods in Table 5.1 are either implemented methods in this work or related proposed methods by other authors previously mentioned in Chapter 3.

The proposed methodology implemented is the unique approach from Table 5.1 that handles point clouds as data type, using exclusively the geometric information provided by the depth and pixel components.

5.2.3 Third Scenario - Scale Sensitivity

The scale is considered a relevant constraint related to biometrics and image processing. The variation on the scale can interfere with feature extraction or data correlation, probably causing classification errors. This scenario consists on the application of a scale transformation in a subset of the data to notice the interference of such transformation in the algorithm performance. This scenario is applied to the best performance configuration from each presented pipeline.

The scenario environment consists in the application of a random scale factor in the numeric interval $[0.5, 0.8]$ to a random subset composed by 30% of the database (including the rejection subset), followed by an F-Measure computation in the optimal point of each algorithm, equivalent

to the first scenario. The concept of this scenario is to analyze the variation of the classification score, estimating the interference caused by the scale transform. Both algorithms are compared with the same resultant random subsets to guarantee a fair comparison.

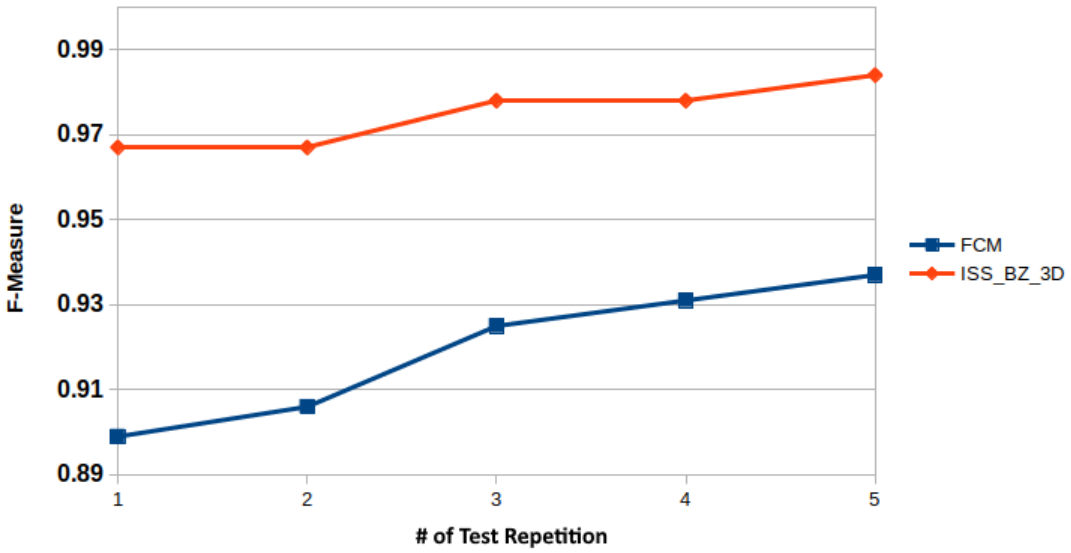


Figure 5.13: F-Measure of the algorithms in the optimal operation point with scale variation along the data set. The FCM performance is represented by the blue line and squared dots, while the Bozorth Composition 3D is represented by the orange line with diamond dots.

According to Figure 5.13, the FCM algorithm suffered a higher interference to scale transform compared to the Bozorth Composition 3D. This result may imply that the curvature index of the facial surface present a higher sensitivity to the scale factor than the normal vector estimation since the FCM algorithm is curvature-based and Bozorth composition is normal vector-based, or either the classifier applied in the FCM pipeline suffers a higher interference than the Bozorth matching algorithm since the matching of Bozorth's algorithm is highly dependent of the angular tolerance. A scale transformation preserves these angular components, as long as the normal vector presents the same orientation or negligible variance.

This scenario permits us to conclude that the FCM pipeline has a higher sensitivity to a scale factor rather than Bozorth's pipeline for this application. Thus Bozorth's pipeline is more scale-invariant (despite a drop in performance being noticed). This scenario also presents evidence that the curvature index has a higher sensitivity than the normal vector estimation.

5.2.4 Fourth Scenario - Orientation Sensitivity

This last scenario has the objective of evaluating the interference that facial orientation variation generates in the proposed methods, determining if these methods are invariant to orientation or not. To provide these results, a biometric evaluation is performed once again to determine the optimal parameters of each algorithm, since the data set presents fewer constraints and includes new inputs for each subject. The scenario consists of adding the four lesser orientation variation (upwards, downwards, left and right) towards the frontal position contained in the database to

a new data set and analyzing if it occurs changes to the performance. It is valid to remember that since the sensor capturing position was fixed, lesser occlusions occur in orientation variation, including another factor that may degrade the performance of the algorithms.

5.2.4.1 FCM Pipeline

The same procedures realized in the first scenario are repeated to acquire the optimal operation point to the new data set content.

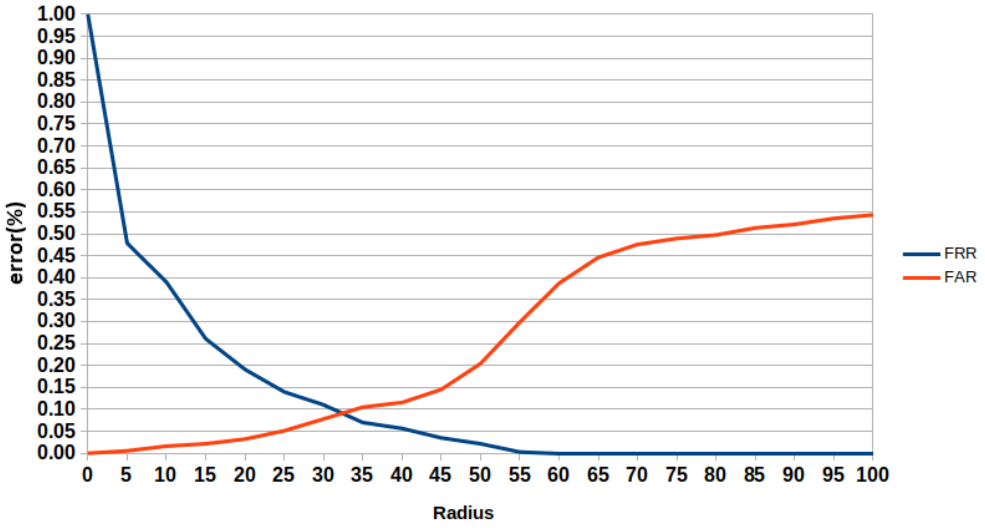


Figure 5.14: EER of the FCM algorithm for the vicinity radius with orientation variation among the data set.

The new optimal operation point for the FCM algorithm consists in the intersection point displayed in Figure 5.14, in which the radius value is equal $32,31$ generating an equal error of $9,15\%$. This data acquisition assumed the previous bins optimal value, which coincidentally matched the previous evaluation result, obtaining a higher error for the previous and next possible bins values. It is noted in this EER that the performance of the FCM has lowered with the orientation variation among the data set.

5.2.4.2 Bozorth Pipeline (Bozorth Composition 3D)

Similarly to the FCM Pipeline evaluation, the procedure realized are the same as the first scenario's, aiming to acquire the optimal operation point of the salient radius for the newly formed data set.

The intersection point and new optimal operation salient radius shown in Figure 5.15 presents the value of $5,2$ generating an equal error of $3,5\%$. The same assumption is made towards the angular thresholds, noticing minor variations in the angular tolerance values that are discarded later in the algorithm due to the threshold value handling as an integer (truncating the decimal part). The result of this EER analysis indicates that the Bozorth Composition 3D presents sensitivity to

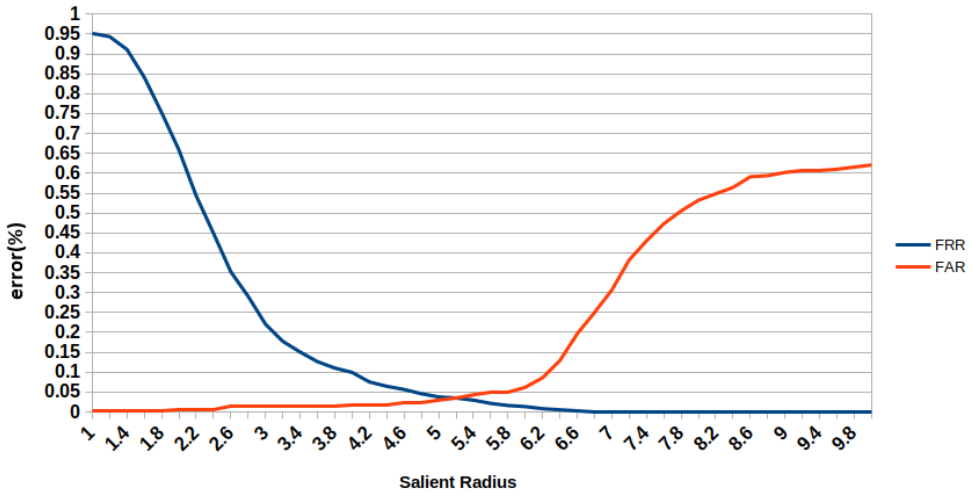


Figure 5.15: EER of the Bozorth Composition 3D for the salient radius with orientation variation among the data set.

orientation variation since there is no perfect classification result, as noted in the frontal orientation biometrics evaluation.

In both cases, the orientation variation application resulted in the deterioration of the performance, implying that the algorithms contain sensitivity to the orientation.

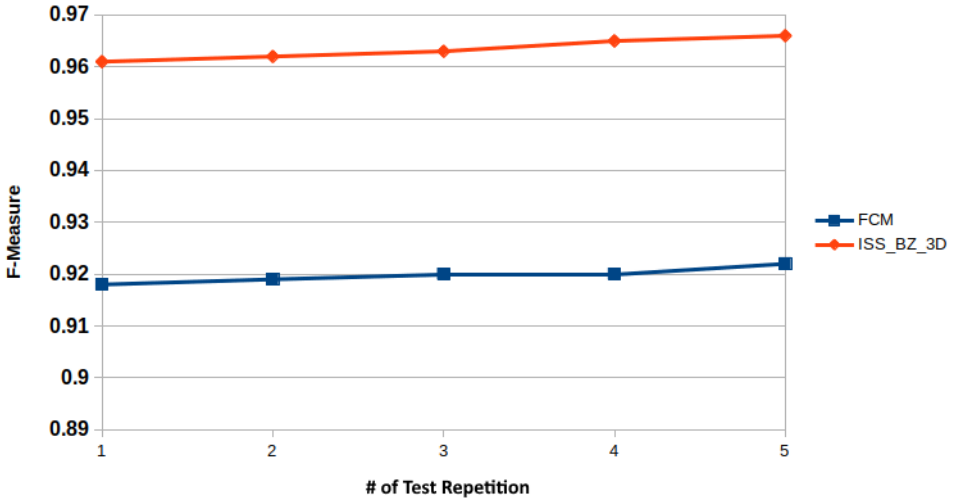


Figure 5.16: F-Measure of the FCM and Bozorth Composition 3D algorithms in an orientation variation scenario. The blue line with squared dots represents the FCM performance, while the orange line with diamong dots represents the Bozorth Composition 3D.

Analyzing the F-Measure values computed for both algorithms, a sensitivity to orientation variation is noticed, since the values presented in 5.16 are lower than the previous scenarios, and consequently containing a loss in performance. The performance variation seen in the FCM algorithm is higher, displaying a maximum score variation of $0,07$. At the same time, the performance of the Bozorth Composition 3D presented a maximum score variation of $0,03$, thus concluding

that the FCM sensitivity to orientation is higher.

Despite the loss in performance, Bozorth's algorithm continues to represent a viable method for biometrics application, providing an F-Measure score above the $0,95$, guaranteeing a high probability of proper recognition, while the FCM turns out to be a viable application in frontal face constrained environments.

5.2.5 Evaluation Conclusions

The conclusion of these scenarios completes the validation process of both pipelines. This process allowed the measurement of the performance for each pipeline configuration, determining the optimal operational point in each case and the capability to deal with the facial recognition issue. Between both pipelines, Bozorth's approach stated a better performance overall with the 3D configuration, presenting better results (including a perfect score in the first scenario) and lower sensitivity to impediments such as scale and orientation. An interesting observation towards the proposal of two pipelines and their evaluation consists in the difference of feature usage since the FCM implementation is curvature-based while the Bozorth's algorithm is normal vector-based. The evaluation also allowed the comparison of the proposed methodology results with other related works, grating a comparative measure, and the application viability of the proposal for the recognition issue.

5.3 Main Difficulties

During the proposed methodology implementation, some difficulties were evidenced related to the data structure and quality, and the pre-processing steps.

Regarding the pre-processing, an observed difficulty occurred in the automated facial detection applying the Viola-Jones algorithm. The facial detection failed to deliver a proper result in a few cases, either returning multiple faces (as displayed in Figure 5.17) or returning an empty result (no faces were detected). The first case occurrence is solved following the hypothesis that the subject's face is centered in the image due to the manner the database was developed, thus acquiring the closest face to the center of the image as the subject's face, and the second case is solved by using a successful detection of the same subject as guide to the subject's face location in the failed attempt, although this solution doesn't work in all scenarios (some orientation available returned no faces for most subjects, adding an impediment for the whole database usage).

Concerning the database, the precision of the Kinect V1 along with the subject capture distance (around 2 meters) and the filtering algorithm (spike and noise removal) during the base confection generated low-resolution clouds, presenting holes and sparse representation in a few face locations, creating wrong value computations in the covariance matrix formation and, consequently, feature estimation and extraction.

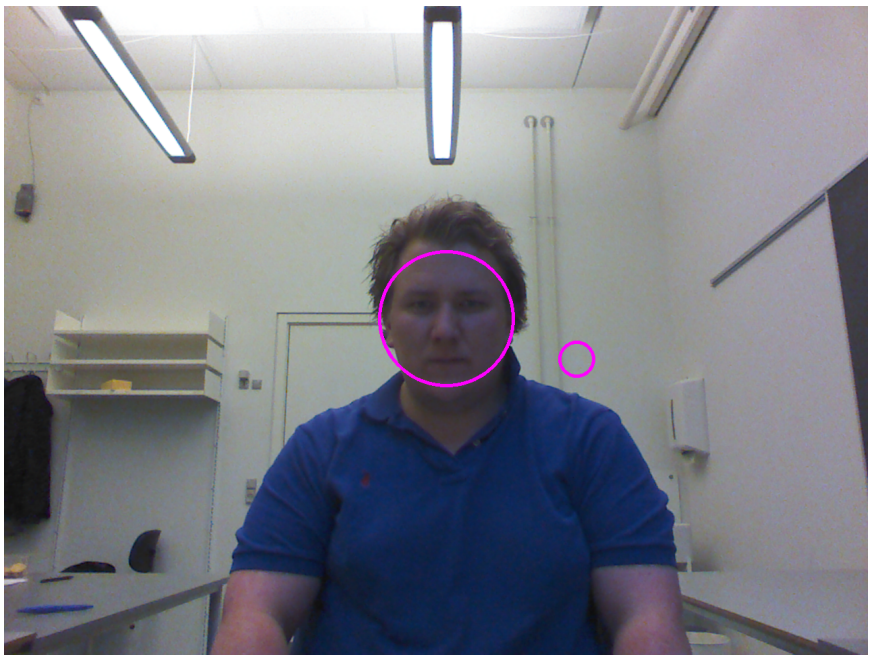


Figure 5.17: Multiple faces detected in the methodology pre-processing using the Viola-Jones [43] face detection algorithm. This image was produced based on a subject image from the database [197].

Chapter 6

Conclusions and Further Works

The proposed and developed methodology aimed to perform a 3D face recognition exploring features extracted from 3D point clouds representing facial data. To solve this task, curvature and normal vector estimations were applied, transposing this information into colored face curvature maps (FCMs) or linked point-based webs to verify the identity of an input data for each subject in the data set using similarity functions or geometrical-based matched scores (Bozorth's algorithm).

The possible combinations of algorithms were tested in four different scenarios, checking the capability against frontal faces, scale and pose (orientation) variation, and comparing rank-1 accuracy against other related-works. Between the two proposed pipelines, Bozorth's pipeline outperformed the Face Curvature Map (FCM) pipeline in all proposed scenarios. In the first scenario, the best Bozorth's solution granted a case with no errors in classification. In contrast, FCM's best result provided an Equal Error Rate (EER) of $2,5\%$, which led to an inferior F-Measure score than Bozorth's solution. In terms of accuracy, Bozorth's solution stands out with a 100% accuracy rate case, while FCM's solution granted the best $98,92\%$ accuracy rate.

The third and fourth scenarios tested the capabilities of these best performers to face recognition known drawbacks, scale, and pose (orientation) variation. Once again, the proposed architecture of Bozorth's pipeline outperformed the FCM's pipeline proposition in both scenarios. In terms of scale sensibility, FCM presented a higher sensitivity to scale, showing an average F-Measure score of about $0,91\%$. At the same time, Bozorth's variation achieved an average F-Measure score of about $0,98\%$, displaying a better capability of handling scale variation. In terms of orientation sensibility, the same hierarchy is noticed, with EERs of $9,15\%$ and $3,5\%$ for the FCM and Bozorth's adaptation, respectively, resulting in F-Measure scores of about $0,92$ and $0,96$ for FCM and Bozorth's adaptation, respectively.

Although the data set application not being specifically suitable for face recognition tasks due to information low-density, it permitted to test the viability of the proposed methodology in low resolution constrained scenarios, allowing the generalization of the proposed approach for low-resolution similar-composition data sets. This data set also contains a low number of subjects for identity classification, creating an uncertainty towards the generalization for bigger data sets. The testing provided was not performed on other data sets due to availability issues (related to the

data restriction of translation to low resolution point clouds).

The presented work displayed two main contributions to the 3D facial recognition field of study. The first contribution consist of a curvature-based facial mapping of point clouds and a simple histogram descriptor for comparative analysis. The second contribution consists of the adaptation of a classic biometric algorithm properly developed for fingerprint matching tasks. This adaptation portrayed a highly efficient algorithm for 3D facial recognition over low resolution point clouds showing the capability of adapting classic biometric algorithms to solve different biometric tasks over more complex biometric data.

6.1 Future Perspective

The proposed solutions' outcome evinces that for similar environment-constrained scenarios, the methodology could handle a task of face recognition identification with favorable results. However, enhancements towards the proposed methodology permit a better generalization of the algorithms and fewer constraints application in exchange for performance.

Enhancements towards pose and expression variation may be proposed to fully explore the provided data set content, generating a pose and expression invariant system. In terms of feature structuring, comparison and matching better representation methods or coordinate projections might be chosen, searching for spatial reduction techniques or invariant discriminating descriptors. For the FCM pipeline, a better matching algorithm and feature representation (current is intensity value histogram) should enhance the performance achieved.

One important task to be developed with the FCM algorithm it the expansion of data set or different data set testing scenarios application, acquiring more data to define the generalization capability of the algorithm and more evidence towards its drawbacks. A data set creation was one of the planned steps for this work (using a Struct From Motion approach), however diverse factors prevented the idea to be implemented.

Finally, to provide a better independent 3D system, a face detection algorithm towards point cloud data could be developed and integrated into the pre-processing steps, removing the necessity of intensity-based images for facial region extraction.

References

[1] A. Jain, L. Hong, and S. Pankanti, "Biometric identification," *Communications of the ACM*, vol. 43, no. 2, pp. 90–98, 2000.

[2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[3] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled faces in the wild: A survey," in *Advances in face detection and facial image analysis*, pp. 189–248, Springer, 2016.

[4] L. F. de Melo Nunes, C. Zaghetto, and F. de Barros Vidal, "3d face recognition on point cloud data - an approaching based on curvature map projection using low resolution devices.," 2018.

[5] R. Jafri and H. R. Arabnia, "A survey of face recognition techniques.," *Jips*, vol. 5, no. 2, pp. 41–68, 2009.

[6] R. Chellappa, C. L. Wilson, S. Sirohey, *et al.*, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[7] G. G. Gordon, "Face recognition based on depth maps and surface curvature," in *San Diego,'91, San Diego, CA*, pp. 234–247, International Society for Optics and Photonics, 1991.

[8] S. Zhang, *High-resolution, real-time 3-D shape measurement.* PhD thesis, Citeseer, 2005.

[9] R. A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, and J. Nissano, "Structured light using pseudorandom codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 322–327, 1998.

[10] C. Wu, "Towards linear-time incremental structure from motion," in *2013 International Conference on 3D Vision-3DV 2013*, pp. 127–134, IEEE, 2013.

[11] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 8, pp. 1362–1376, 2009.

[12] M. Vermeulen, P. Rosielle, and P. Schellekens, "Design of a high-precision 3d-coordinate measuring machine," *Cirp Annals*, vol. 47, no. 1, pp. 447–450, 1998.

[13] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE MultiMedia*, vol. 19, pp. 4–10, Apr. 2012.

[14] V. A. Prisacariu, O. Kähler, D. W. Murray, and I. D. Reid, "Real-time 3d tracking and reconstruction on mobile phones," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 5, pp. 557–570, 2014.

[15] T. F. Rodriguez, "3d data representation, conveyance, and use," Dec. 9 2010. US Patent App. 12/727,092.

[16] S. Larsson and J. Kjellander, "Motion control and data capturing for laser scanning with an industrial robot," *Robotics and Autonomous Systems*, vol. 54, no. 6, pp. 453–460, 2006.

[17] A. Bewley, R. Shekhar, S. Leonard, B. Upcroft, and P. Lever, "Real-time volume estimation of a dragline payload," in *2011 IEEE International Conference on Robotics and Automation*, pp. 1571–1576, IEEE, 2011.

[18] O. P. Unit, "Crime scene documentation." FBI Laboratory.

[19] K. W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition," *Computer vision and image understanding*, vol. 101, no. 1, pp. 1–15, 2006.

[20] T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland, "Multimodal person recognition using unconstrained audio and video," in *Proceedings, International Conference on Audio- and Video-Based Person Authentication*, pp. 176–181, Citeseer, 1999.

[21] W. Zhou, J.-x. Chen, and L. Wang, "A rgb-d face recognition approach without confronting the camera," in *Computer and Communications (ICCC), 2015 IEEE International Conference on*, pp. 109–114, IEEE, 2015.

[22] G. Goswami, M. Vatsa, and R. Singh, "Rgb-d face recognition with texture and attribute features," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 10, pp. 1629–1640, 2014.

[23] T. D. Russ, M. W. Koch, and C. Q. Little, "3d facial recognition: a quantitative analysis," in *Security Technology, 2004. 38th Annual 2004 International Carnahan Conference on*, pp. 338–344, IEEE, 2004.

[24] H. Li, D. Huang, J.-M. Morvan, Y. Wang, and L. Chen, "Towards 3d face recognition in the real: A registration-free approach using fine-grained matching of 3d keypoint descriptors," *International Journal of Computer Vision*, vol. 113, no. 2, pp. 128–142, 2015.

[25] I. L. U. . Photogrammetry & Geomatics Group, "6th international workshop lowcost 3d - sensors, algorithms, applications." INSA Strasbourg.

[26] I. Strasbourg, "Optical 3d metrology." National Institute of Applied Sciences (INSA).

[27] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[28] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society*, vol. 479. Springer Science & Business Media, 2006.

[29] N. B. Sukhai, "Access control & biometrics," in *Proceedings of the 1st annual conference on Information security curriculum development*, pp. 124–127, ACM, 2004.

[30] S. Benziane and A. Benyettou, "An introduction to biometrics," *International Journal of Computer Science and Information Security*, vol. 9, no. 4, pp. 40–47, 2011.

[31] D. D. Zhang, *Automated biometrics: Technologies and systems*, vol. 7. Springer Science & Business Media, 2013.

[32] J. Soldera, G. Schu, L. R. Schardosim, and E. T. Beltrao, "Facial biometrics and applications," *IEEE Instrumentation & Measurement Magazine*, vol. 20, no. 2, pp. 4–10, 2017.

[33] S. Haykin and N. Network, "A comprehensive foundation," *Neural Networks*, vol. 2, no. 2004, p. 41, 2004.

[34] G. Goswami, S. Bharadwaj, M. Vatsa, and R. Singh, "On rgb-d face recognition using kinect," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pp. 1–6, IEEE, 2013.

[35] A. Chowdhury, S. Ghosh, R. Singh, and M. Vatsa, "Rgb-d face recognition via learning-based reconstruction," in *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*, pp. 1–7, IEEE, 2016.

[36] K. Ko, "Users guide to export controlled distribution of nist biometric image software (nbis-ec)," tech. rep., 2007.

[37] H. Patil, A. Kothari, and K. Bhurchandi, "3-d face recognition: features, databases, algorithms and challenges," *Artificial Intelligence Review*, vol. 44, no. 3, pp. 393–441, 2015.

[38] L. Linsen, *Point cloud representation*. Univ., Fak. für Informatik, Bibliothek, 2001.

[39] K. Lai and D. Fox, "Object recognition in 3d point clouds using web data and domain adaptation," *The International Journal of Robotics Research*, vol. 29, no. 8, pp. 1019–1037, 2010.

[40] G. Sithole and G. Vosselman, "Automatic structure detection in a point-cloud of an urban landscape," in *Remote Sensing and Data Fusion over Urban Areas, 2003. 2nd GRSS/ISPRS Joint Workshop on*, pp. 67–71, IEEE, 2003.

[41] R. B. Rusu, "Semantic 3d object maps for everyday manipulation in human living environments," *KI-Künstliche Intelligenz*, vol. 24, no. 4, pp. 345–348, 2010.

[42] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pp. 1–4, IEEE, 2011.

[43] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–511, IEEE, 2001.

[44] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3d object recognition," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pp. 689–696, IEEE, 2009.

[45] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[46] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *European conference on computational learning theory*, pp. 23–37, Springer, 1995.

[47] S. Filipe and L. A. Alexandre, "A comparative evaluation of 3d keypoint detectors in a rgb-d object dataset," in *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. 1, pp. 476–483, IEEE, 2014.

[48] F. Tombari, "Keypoints and features," 2013.

[49] S. Maddala, S. R. Tangellapally, J. S. Bartunek, and M. Nilsson, "Implementation and evaluation of nist biometric image software for fingerprint recognition," in *ISSNIP Biosignals and Biorobotics Conference 2011*, pp. 1–5, IEEE, 2011.

[50] A. Jain, R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society*, vol. 479. Springer Science & Business Media, 2006.

[51] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," in *European Conference on Computer Vision*, pp. 43–58, Springer, 1996.

[52] P. J. Grother, M. L. Ngan, and K. K. Hanaoka, "Face recognition vendor test (frvt) part 2: Identification," tech. rep., 2019.

[53] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[54] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *arXiv preprint arXiv:1502.00873*, 2015.

[55] D. McCullagh, "Call it super bowl face scan i," *Wired, Feb*, vol. 2, 2001.

[56] M. Viswanathan, H. S. Beigi, A. Tritschler, and F. Maali, "Information access using speech, speaker and face recognition," in *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)*, vol. 1, pp. 493–496, IEEE, 2000.

[57] A. K. Jain, K. Nandakumar, X. Lu, and U. Park, "Integrating faces, fingerprints, and soft biometric traits for user recognition," in *International Workshop on Biometric Authentication*, pp. 259–269, Springer, 2004.

[58] S. Ben-Yacoub, J. Luttin, K. Jonsson, J. Matas, and J. Kittler, "Audio-visual person verification," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 1, pp. 580–585, IEEE, 1999.

[59] C. Nastar and M. Mitschke, "Real-time face recognition using feature combination," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 312–317, IEEE, 1998.

[60] S. Gong, S. J. McKenna, and A. Psarrou, *Dynamic vision: from images to face recognition*. Imperial College Press, 2000.

[61] L. Akarun, B. Gökberk, and A. A. Salah, "3d face recognition for biometric applications," in *2005 13th European Signal Processing Conference*, pp. 1–5, IEEE, 2005.

[62] T. Kanade, "Picture processing system by computer complex and recognition of human faces," *Doctoral dissertation, Kyoto University*, vol. 3952, pp. 83–97, 1973.

[63] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE transactions on pattern analysis and machine intelligence*, vol. 15, no. 10, pp. 1042–1052, 1993.

[64] D. Reisfeld, *Generalized symmetry transforms: attentional mechanisms and face recognition*. PhD thesis, Tel Aviv University, 1994.

[65] M. Nixon, "Eye spacing measurement for facial recognition," in *29th Annual Technical Symposium*, pp. 279–285, International Society for Optics and Photonics, 1985.

[66] H. P. Graf, T. Chen, E. Petajan, and E. Cosatto, "Locating faces and facial parts," in *Proc. First Int'l Workshop Automatic Face and Gesture Recognition*, pp. 41–46, 1995.

[67] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 775–779, 1997.

[68] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using hog–ebgm," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1537–1543, 2008.

[69] I. J. Cox, J. Ghosn, and P. N. Yianilos, "Feature-based face recognition using mixture-distance," in *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pp. 209–216, IEEE, 1996.

[70] P. Campadelli and R. Lanzarotti, "A face recognition system based on local feature characterization," in *Advanced Studies in Biometrics*, pp. 147–152, Springer, 2005.

[71] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," in *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR'89., IEEE Computer Society Conference on*, pp. 104–109, IEEE, 1989.

[72] L. Wiskott, J.-M. Fellous, N. Krüger, and C. v. d. Malsburg, "Intelligent biometric techniques in fingerprint and face recognition," ch. Face Recognition by Elastic Bunch Graph Matching, pp. 357–396, Boca Raton, FL, USA: CRC Press, Inc., 1999.

[73] G. J. Kaufman and K. J. Breeding, "The automatic recognition of human faces from profile silhouettes," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 2, pp. 113–121, 1976.

[74] Z. Lipoščak and S. Loncaric, "A scale-space approach to face recognition from profiles," in *International Conference on Computer Analysis of Images and Patterns*, pp. 243–250, Springer, 1999.

[75] L. Harmon, M. Khan, R. Lasch, and P. Ramig, "Machine identification of human faces," *Pattern Recognition*, vol. 13, no. 2, pp. 97–110, 1981.

[76] R. J. Baron, "Mechanisms of human facial recognition," *International Journal of Man-Machine Studies*, vol. 15, no. 2, pp. 137–178, 1981.

[77] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Josa a*, vol. 4, no. 3, pp. 519–524, 1987.

[78] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.

[79] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 2013.

[80] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pp. 586–591, IEEE, 1991.

[81] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[82] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pp. 84–91, IEEE, 1994.

[83] Y. Moses, Y. Adini, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," in *European conference on computer vision*, pp. 286–296, Springer, 1994.

[84] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[85] D. L. and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.

[86] Y. Chang, C. Hu, and M. Turk, "Manifold of facial expression," in *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, p. 28, IEEE Computer Society, 2003.

[87] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[88] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in neural information processing systems*, pp. 585–591, 2002.

[89] S. Yan, H. Zhang, Y. Hu, B. Zhang, and Q. Cheng, "Discriminant analysis on embedded manifold," in *European Conference on Computer Vision*, pp. 121–132, Springer, 2004.

[90] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 3, pp. 328–340, 2005.

[91] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.

[92] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on neural networks*, vol. 13, no. 6, pp. 1450–1464, 2002.

[93] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, "Recognizing faces with pca and ica," *Computer vision and image understanding*, vol. 91, no. 1, pp. 115–137, 2003.

[94] C. Liu and H. Wechsler, "Comparative assessment of independent component analysis (ica) for face recognition," in *International conference on audio and video based biometric person authentication*, Citeseer, 1999.

[95] N. H. Foon, A. T. B. Jin, and D. N. C. Ling, "Face recognition using wavelet transform and non-negative matrix factorization," in *Australasian Joint Conference on Artificial Intelligence*, pp. 192–202, Springer, 2004.

[96] W. Liu, Y. Wang, S. Z. Li, and T. Tan, "Nearest intra-class space classifier for face recognition," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 4, pp. 495–498, IEEE, 2004.

[97] D. W. Patterson, *Introduction to artificial intelligence and expert systems*. Prentice-hall of India, 1990.

[98] R. C. Schank and P. Childers, "The cognitive computer on language, learning, and artificial intelligence," 1984.

[99] J. Kim, J. Choi, and J. Yi, "Face recognition based on locally salient ica information," in *International Workshop on Biometric Authentication*, pp. 1–9, Springer, 2004.

[100] J. Yi, J. Kim, J. Choi, J. Han, and E. Lee, "Face recognition based on ica combined with fld," in *International Workshop on Biometric Authentication*, pp. 10–18, Springer, 2002.

[101] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 98–113, 1997.

[102] B. Li and H. Yin, "Face recognition using rbf neural networks and wavelet transform," in *International Symposium on Neural Networks*, pp. 105–111, Springer, 2005.

[103] A. Eleyan and H. Demirel, "Face recognition system based on pca and feedforward neural networks," in *International Work-Conference on Artificial Neural Networks*, pp. 935–942, Springer, 2005.

[104] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu, "Boosting local binary pattern (lbp)-based face recognition," in *Advances in biometric person authentication*, pp. 179–186, Springer, 2004.

[105] B. Moghaddam, C. Nastar, and A. Pentland, "A bayesian similarity measure for direct image matching," in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, vol. 2, pp. 350–358, IEEE, 1996.

[106] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi, "The feret evaluation methodology for face-recognition algorithms," in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 137–143, IEEE, 1997.

[107] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European conference on computer vision*, pp. 469–481, Springer, 2004.

[108] P. Melin, C. Felix, and O. Castillo, "Face recognition using modular neural networks and the fuzzy sugeno integral for response integration," *International journal of intelligent systems*, vol. 20, no. 2, pp. 275–291, 2005.

[109] M. Moreira and E. Mayoraz, "Improved pairwise coupling classification with correcting classifiers," in *European conference on machine learning*, pp. 160–171, Springer, 1998.

[110] H. Li, F. Qi, and S. Wang, "Face recognition with improved pairwise coupling support vector machines," in *International Work-Conference on Artificial Neural Networks*, pp. 927–934, Springer, 2005.

[111] Z. Li and S. Tang, "Face recognition using improved pairwise coupling support vector machines," in *Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on*, vol. 2, pp. 876–880, IEEE, 2002.

[112] J. Platt *et al.*, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in large margin classifiers*, vol. 10, no. 3, pp. 61–74, 1999.

[113] H. Li, S. Wang, and F. Qi, "Automatic face recognition by support vector machines," in *International Workshop on Combinatorial Image Analysis*, pp. 716–725, Springer, 2004.

[114] G. Dai and C. Zhou, "Face recognition using support vector machines with the robust feature," in *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, pp. 49–53, IEEE, 2003.

[115] O. Déniz, M. Castrillon, and M. Hernández, "Face recognition using independent component analysis and support vector machines," *Pattern recognition letters*, vol. 24, no. 13, pp. 2153–2157, 2003.

[116] K. Jonsson, J. Matas, J. Kittler, and Y. Li, "Learning support vectors for face verification and recognition," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pp. 208–213, IEEE, 2000.

[117] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[118] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on*, pp. 138–142, IEEE, 1994.

[119] F. S. Samaria, *Face recognition using hidden Markov models*. PhD thesis, University of Cambridge, 1994.

[120] A. Nefian and M. Hayes, "Face recognition using an embedded hmm," in *IEEE Conference on Audio and Video-based Biometric Person Authentication*, pp. 19–24, 1999.

[121] F. Roli and J. Kittler, *Multiple Classifier Systems: Third International Workshop, MCS 2002, Cagliari, Italy, June 24-26, 2002. Proceedings*, vol. 2364. Springer Science & Business Media, 2002.

[122] X. Lu, Y. Wang, and A. K. Jain, "Combining classifiers for face recognition," in *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on*, vol. 3, pp. III–13, IEEE, 2003.

[123] G. L. Marcialis and F. Roli, "Fusion of lda and pca for face recognition," *Department of Electrical and Electronic Engineering, University of Cagliari, Piazza díArmi*, 2002.

[124] G. L. Marcialis and F. Roli, "Fusion of lda and pca for face verification," in *Proceedings of the International ECCV 2002 Workshop Copenhagen on Biometric Authentication*, ECCV '02, (London, UK, UK), pp. 30–38, Springer-Verlag, 2002.

[125] G. L. Marcialis and F. Roli, "Fusion of appearance-based face recognition algorithms," *Pattern Analysis and Applications*, vol. 7, no. 2, pp. 151–163, 2004.

[126] B. Achermann and H. Bunke, *Combination of classifiers on the decision level for face recognition*. Citeseer, 1996.

[127] A. S. Tolba and A. Abu-Rezq, "Combined classifiers for invariant face recognition," *Pattern Analysis & Applications*, vol. 3, no. 4, pp. 289–302, 2000.

[128] Y. Wan, S. Ji, Y. Xie, X. Zhang, and P. Xie, "Video program clustering indexing based on face recognition hybrid model of hidden markov model and support vector machine," in *International Workshop on Combinatorial Image Analysis*, pp. 739–749, Springer, 2004.

[129] K.-C. Kwak and W. Pedrycz, "Face recognition: A study in information fusion using fuzzy integral," *Pattern Recognition Letters*, vol. 26, no. 6, pp. 719–733, 2005.

[130] L. Deng, D. Yu, *et al.*, "Deep learning: methods and applications," *Foundations and Trends® in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[131] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014.

[132] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[133] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[134] Y. Sun, X. Wang, and X. Tang, "Hybrid deep learning for face verification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1489–1496, 2013.

[135] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1891–1898, 2014.

[136] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in neural information processing systems*, pp. 1988–1996, 2014.

[137] Y. Sun, X. Wang, and X. Tang, "Deeply learned face representations are sparse, selective, and robust," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2892–2900, 2015.

[138] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3025–3032, 2013.

[139] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[140] H. Fan, Z. Cao, Y. Jiang, Q. Yin, and C. Doudou, "Learning deep face representation," *arXiv preprint arXiv:1403.2802*, 2014.

[141] E. Zhou, Z. Cao, and Q. Yin, "Naive-deep face recognition: Touching the limit of lfw benchmark or not?," *arXiv preprint arXiv:1501.04690*, 2015.

[142] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1708, 2014.

[143] Q. Yang and X. Tang, "Recent advances in subspace analysis for face recognition," in *Advances in Biometric Person Authentication*, pp. 275–287, Springer, 2004.

[144] P. J. Phillips, S. Z. Der, P. J. Rauss, and O. Z. Der, *FERET (face recognition technology) recognition algorithm development and test results.* Army Research Laboratory Adelphi, MD, 1996.

[145] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[146] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2, pp. 688–694, IEEE, 2001.

[147] S.-H. Lin, S.-Y. Kung, and L.-J. Lin, "Face recognition/detection by probabilistic decision-based neural network," *IEEE transactions on neural networks*, vol. 8, no. 1, pp. 114–132, 1997.

[148] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," 2008.

[149] M. Ngan, P. J. Grother, and M. Ngan, *Face recognition vendor test (FRVT) performance of automated gender classification algorithms.* US Department of Commerce, National Institute of Standards and Technology, 2015.

[150] G. B. Huang and E. Learned-Miller, "Labeled faces in the wild: Updates and new reporting procedures," *Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep*, pp. 14–003, 2014.

[151] R. G. Cutler, *Face recognition using infrared images and eigenfaces.* Citeseer, 1996.

[152] A. Gyaourova, G. Bebis, and I. Pavlidis, "Fusion of infrared and visible images for face recognition," in *European Conference on Computer Vision*, pp. 456–468, Springer, 2004.

[153] A. Selinger and D. A. Socolinsky, "Appearance-based facial recognition using visible and thermal imagery: a comparative study," tech. rep., DTIC Document, 2006.

[154] J. Wilder, P. J. Phillips, C. Jiang, and S. Wiener, "Comparison of visible and infra-red imagery for face recognition," in *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, pp. 182–187, IEEE, 1996.

[155] D. A. Socolinsky, L. B. Wolff, J. D. Neuheisel, and C. K. Eveland, "Illumination invariant face recognition using thermal infrared imagery," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–527, IEEE, 2001.

[156] T. Sim, R. Sukthankar, M. Mullin, and S. Baluja, "Memory-based face recognition for visitor identification," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pp. 214–220, IEEE, 2000.

[157] X. C. P. J. F. Kevin and W. Bowyer, "Visible-light and infrared face recognition," in *Workshop on Multimodal User Authentication*, p. 48, Citeseer, 2003.

[158] Identix, "Identix inc.: Electronic documentation.."

[159] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[160] X. Chen, P. J. Flynn, and K. W. Bowyer, "Ir and visible light face recognition," *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 332–358, 2005.

[161] R. Singh, M. Vatsa, and A. Noore, "Hierarchical fusion of multi-spectral face images for improved recognition performance," *Information Fusion*, vol. 9, no. 2, pp. 200–210, 2008.

[162] J. Heo, S. G. Kong, B. R. Abidi, and M. A. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, pp. 122–122, IEEE, 2004.

[163] R. Singh, M. Vatsa, and A. Noore, "Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition," *Pattern Recognition*, vol. 41, no. 3, pp. 880–893, 2008.

[164] T. S. Jebara, *3D pose estimation and normalization for face recognition.* PhD thesis, McGill University, 1995.

[165] A. D. Tibbalds, *Three dimensional human face acquisition for recognition.* PhD thesis, University of Cambridge, 1998.

[166] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in *Signal processing and its applications, 2003. Proceedings. Seventh international symposium on*, vol. 2, pp. 201–204, IEEE, 2003.

[167] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, pp. 519–528, IEEE, 2006.

[168] G. Vogiatzis, P. H. Torr, and R. Cipolla, "Multi-view stereo via volumetric graph-cuts," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 391–398, IEEE, 2005.

[169] S. N. Sinha, P. Mordohai, and M. Pollefeys, "Multi-view stereo via graph cuts on the dual of an adaptive tetrahedral mesh," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.

[170] Y. Furukawa and J. Ponce, "Carved visual hulls for image-based modeling," in *European Conference on Computer Vision*, pp. 564–577, Springer, 2006.

[171] A. Zaharescu, E. Boyer, and R. Horaud, "Transformesh: a topology-adaptive mesh-based approach to surface evolution," in *Asian Conference on Computer Vision*, pp. 166–175, Springer, 2007.

[172] M. Lhuillier and L. Quan, "A quasi-dense approach to surface reconstruction from uncalibrated images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 3, pp. 418–433, 2005.

[173] C. Beumier and M. Acheroy, "Automatic 3d face authentication," *Image and Vision Computing*, vol. 18, no. 4, pp. 315–321, 2000.

[174] U. Castellani, M. Bicego, G. Iacono, and V. Murino, "3d face recognition using stereoscopic vision," in *Advanced Studies in Biometrics*, pp. 126–137, Springer, 2005.

[175] S. Lee, G. Wolberg, and S. Y. Shin, "Scattered data interpolation with multilevel b-splines," *IEEE transactions on visualization and computer graphics*, vol. 3, no. 3, pp. 228–244, 1997.

[176] G. Pan, Z. Wu, and Y. Pan, "Automatic 3d face verification from range data," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 3, pp. III–193, IEEE, 2003.

[177] C. Xu, Y. Wang, T. Tan, and L. Quan, "Automatic 3d face recognition combining global geometric features with local shape variation information," in *Automatic face and gesture recognition, 2004. Proceedings. Sixth IEEE international conference on*, pp. 308–313, IEEE, 2004.

[178] Y. Lee, H. Song, U. Yang, H. Shin, and K. Sohn, "Local feature based 3d face recognition," in *International Conference on Audio-and Video-based Biometric Person Authentication*, pp. 909–918, Springer, 2005.

[179] F. R. Al-Osaimi, M. Bennamoun, and A. Mian, "Integration of local and global geometrical cues for 3d face recognition," *Pattern Recognition*, vol. 41, no. 3, pp. 1030–1040, 2008.

[180] J.-Y. Cartoux, J.-T. LaPresté, and M. Richetin, "Face authentification or recognition by profile extraction from range images," in *Interpretation of 3D Scenes, 1989. Proceedings., Workshop on*, pp. 194–199, IEEE, 1989.

[181] T. Nagamine, T. Uemura, and I. Masuda, "3d facial image analysis for human identification," in *Pattern Recognition, 1992. Vol. I. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on*, pp. 324–327, IEEE, 1992.

[182] C. Li and A. Barreto, "Profile-based 3d face registration and recognition," in *International Conference on Information Security and Cryptology*, pp. 478–488, Springer, 2004.

[183] Y. Wu, G. Pan, and Z. Wu, "Face authentication based on multiple profiles extracted from range data," in *International Conference on Audio-and Video-Based Biometric Person Authentication*, pp. 515–522, Springer, 2003.

[184] J.-G. Wang, K.-A. Toh, and R. Venkateswarlu, "Fusion of appearance and depth information for face recognition," in *International Conference on Audio-and Video-Based Biometric Person Authentication*, pp. 919–928, Springer, 2005.

[185] F. Tsalakanidou, D. Tzovaras, and M. G. Strintzis, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters*, vol. 24, no. 9, pp. 1427–1435, 2003.

[186] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2d and 3d facial data," in *ACM Workshop on Multimodal User Authentication*, pp. 25–32, Citeseer, 2003.

[187] C. Beumier and M. Acheroy, "Face verification from 3d and grey level clues," *Pattern recognition letters*, vol. 22, no. 12, pp. 1321–1329, 2001.

[188] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Robotics-DL tentative*, pp. 586–606, International Society for Optics and Photonics, 1992.

[189] T. Papatheodorou and D. Rueckert, "Evaluation of automatic 4d face recognition using surface and texture registration," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pp. 321–326, IEEE, 2004.

[190] X. Lu and A. K. Jain, "Integrating range and texture information for 3d face recognition," in *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, vol. 1, pp. 156–163, IEEE, 2005.

[191] Y. Wang, C.-S. Chua, and Y.-K. Ho, "Facial feature detection and face recognition from 2d and 3d images," *Pattern Recognition Letters*, vol. 23, no. 10, pp. 1191–1202, 2002.

[192] N. Hu, R. Bormann, T. Zwölfer, and B. Kröse, "Multi-user identification and efficient user approaching by fusing robot and ambient sensors," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pp. 5299–5306, IEEE, 2014.

[193] R. Bormann, T. Zwölfer, J. Fischer, J. Hampp, and M. Hägele, "Person recognition for service robotics applications," in *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference on*, pp. 260–267, IEEE, 2013.

[194] Y. Saleh and E. Edirisinghe, "Novel approach to enhance face recognition using depth maps," in *Systems, Signals and Image Processing (IWSSIP), 2016 International Conference on*, pp. 1–4, IEEE, 2016.

[195] F. Crestani, M. Lalmas, and C. J. van Rijsbergen, *Information Retrieval: Uncertainty and Logics: Uncertainty and Logics: Advanced Models for the Representation and Retrieval of Information*, vol. 4. Springer Science & Business Media, 1998.

[196] Y. Sasaki *et al.*, "The truth of the f-measure," *Teach Tutor mater*, vol. 1, no. 5, pp. 1–5, 2007.

[197] R. Hg, P. Jasek, C. Rofidal, K. Nasrollahi, T. B. Moeslund, and G. Tranchet, "An rgb-d database using microsoft's kinect for windows for face detection," in *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on*, pp. 42–46, IEEE, 2012.

[198] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pp. 3212–3217, IEEE, 2009.

[199] G. Hetzel, B. Leibe, P. Levi, and B. Schiele, "3d object recognition from range images using local feature histograms," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 2, pp. II–II, IEEE, 2001.

[200] G. Bradski, "Opencv library," *Dr. Dobb's Journal of Software Tools*, 2000.

[201] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 306–311, IEEE, 2010.

[202] E. Trentin and M. Gori, "A survey of hybrid ann/hmm models for automatic speech recognition," *Neurocomputing*, vol. 37, no. 1, pp. 91–126, 2001.

[203] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "On the best rank-1 and rank-(r1,r2,. . .,rn) approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 1324–1342, Mar. 2000.